Abstract

LEE, HYEYOUNG. Reparametrized Dynamic Space-Time Models and Spatial Model Selection. (Under the direction of Sujit Ghosh.)

Researchers in diverse areas such as environmental and health sciences are increasingly facing working with space-time data. Often the dimension of space-time data sets can be very large and moreover, space-time processes are often complicated in that the dependence structure across space and time is non-trivial, often non-separable and non-stationary in space and/or time. Hence, space-time modeling is a challenging task and in particular parameter estimation can be problematic due to the high dimensionality. We propose a reparametrization approach to fit dynamic space-time models with an unstructured covariance function. Our modeling contribution is to present unconstrained reparametrization for a covariance matrix in dynamic space-time models. Using this unconstrained reparametrization method, we are able to implement the modeling of a high dimensional covariance matrix that automatically maintains the positive definiteness constraint. We illustrate the use of this reparametrization method by applying our model to a set of atmospheric nitrate concentration data. We also consider the problem of model selection for spatial data. The issue of model selection in spatial models has rarely been addressed in the literature, though it is very important. To address this problem, we consider selection criteria such as the Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC). The performance of these selection criteria are examined using Monte Carlo simulations. In particular, the ability of these criteria to select the correct model is evaluated.

Reparametrized Dynamic Space-Time Models and Spatial Model Selection

by

Hyeyoung Lee

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh

2006

APPROVED BY:

Sujit K. Ghosh Chair of Advisory Comittee	David A. Dickey
Montserrat Fuentes	Jerry M. Davis

To my family and friends

Biography

Hyeyoung Lee was born in Seoul, South Korea in 1973 to Myungkul Lee and Siyeon Ryu. Hyeyoung earned her BS of Statistics in 1996 and her MS in 1998 from Sookmyung Women's University in Seoul. She worked as an instructor in Statistics at Sookmyung Women's University in 1998 and 2000. She also worked as a researcher in Chemometrics in 1999. In 2000 she entered graduate school at North Carolina State University in Raleigh. She received her Ph.D. in Statistics in 2006. While she was at NC State, she worked on projects with Environmental Protection Agency (EPA). She also did an internship at EPA.

Acknowledgements

First of all, I would like to thank my family and friends for their love and support. I am very fortunate to have such wonderful people. They have always encouraged me to achieve my goals and helped me to get through tough times. I would have not finished my study at NC State without them.

I also would like to thank my advisor, Dr. Ghosh, for his guidance. He has been of great help for me to write this thesis. I have learned so much from him.

My committee members, Dr. Dickey, Dr. Fuentes, and Dr. Davis, have been very helpful. I would especially like to thank them for their comments on this thesis. Dr. Davis has been very supportive and understanding. He helped me get the data from EPA for my research and also helped me understand chemistry and meteorology things.

Dr. Holland and Prakash at EPA have been also very helpful. Dr. Holland provided a data set from EPA for my research. Prakash was very important in interpreting the results for chemistry and meteorological variables. His time and efforts are greatly appreciated.

I would also like to thank many graduate students for their friendship and encouragement.

Additionally, I express thanks to EPA and our department for their financial support.

Contents

List of Tables vii

List of Figures viii

1 Introduction 1
- 1.1 Statistical Models for Space-Time Data . . . 3
- 1.2 Model Selection Methods . . . 9

2 A Reparametrization Approach for Dynamic Space-Time Models 17
- 2.1 Introduction . . . 17
- 2.2 Dynamic Linear Models . . . 19
- 2.3 Dynamic Space-Time Models (DSTM) . . . 21
- 2.4 Model Fitting . . . 24
- 2.5 A Reparametrization Method . . . 26
- 2.6 Comparison with Space-Time Kalman Filter . . . 30

3 Performance of Information Criteria for Spatial Models 33
- 3.1 Information-Theoretic Criteria for Model Selection . . . 34
  - 3.1.1 Akaike Information Criterion (AIC) . . . 35
  - 3.1.2 Corrected Akaike Information Criteria (AICc) . . . 36
  - 3.1.3 Bayesian Information Criterion (BIC) . . . 37
- 3.2 Spatial Models . . . 38
  - 3.2.1 Stationary Processes . . . 39
  - 3.2.2 Nonstationary Processes . . . 41
- 3.3 A Simulation Study . . . 42
  - 3.3.1 Covariance Models . . . 42
  - 3.3.2 Data Generation Processes . . . 44
  - 3.3.3 Results . . . 45
  - 3.3.4 Findings and Future Research . . . 65

4 Application to Total Nitrate Concentration Data 67

4.1 Introduction . . . 67

4.2 The CASTNET Data . . . 68

4.3 Background Chemistry for Total Nitrate . . . 71

4.4 Exploratory Data Analysis (EDA) . . . 73

4.5 Statistical Models . . . 75

4.5.1 Linear Regression Models (LRM) . . . 76

4.5.2 Reparametrized Dynamic Space-Time Models (RDSTM) . . . 77

4.6 Results . . . 79

4.6.1 LRM . . . 79

4.6.2 RDSTM . . . 82

4.7 Findings and Future Research . . . 97

Reference 100

A Tables and Plots 105

A.1 Table and Plots for EDA in Chapter 4 . . . 105

List of Tables

3.1 Number of Parameters and Parameters in Each Model 44
3.2 True Parameter Values for Each Data Generation Process 46
3.3 Frobenius Distance between Covariance Functions of Models 46
3.4 Penalty Given by Each Criterion to Four Models with Two Different Sample Size 48
3.5 Percentage of Correct Decisions When data are generated from M_1 . 49
3.6 Percentage of Correct Decisions When data are generated from M_2 . 51
3.7 Percentage of Correct Decisions When data are generated from M_3 . 55
3.8 Percentage of Correct Decisions When data are generated from M_4 . 59

4.1 Spearman Rank Correlation Coefficients between Variables 74
4.2 ANOVA Table for LRM 80
4.3 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for SO_4 85
4.4 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Residual NH_4 86
4.5 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for O_3 88
4.6 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Wind Speed 90
4.7 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Relative Humidity 92
4.8 Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Precipitation 93

A.1 Total Number of Weeks N for Each Site 106

List of Figures

3.1 Box Plots for Each Criterion Obtained by Fitting Each Model to D_1 . 50
3.2 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{21} 53
3.3 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{22} 54
3.4 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{31} 57
3.5 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{32} 58
3.6 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{41} 61
3.7 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{42} 62
3.8 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{43} 63
3.9 Box Plots for Each Criterion Obtained by Fitting Each Model to D_{44} 64

4.1 Locations of Stations 69
4.2 Histogram for TNO_3 and $\log(TNO_3)$ 77
4.3 Space-Time Correlogram for the Residual of LRM 81
4.4 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for SO_4 84
4.5 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Residual NH_4 86
4.6 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for O_3 87
4.7 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Wind Speed 89
4.8 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Relative Humidity 91
4.9 Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Precipitation 93
4.10 Spatial Correlogram and Fitted Exponential Correlation Model . . . 96

A.1 Box Plots for HNO_3 by Siteid, Year and Month 107
A.2 Box Plots for NO_3 by Siteid, Year and Month 108
A.3 Box Plots for TNO_3 by Siteid, Year and Month 109
A.4 Box Plots for SO_4 by Siteid, Year and Month 110

A.5 Box Plots for NH_4 by Siteid, Year and Month 111
A.6 Box Plots for O_3 by Siteid, Year and Month 112
A.7 Box Plots for Temperature by Siteid, Year and Month 113
A.8 Box Plots for Dew Point Temperature by Siteid, Year and Month . . 114
A.9 Box Plots for Solar Radiation by Siteid, Year and Month 115
A.10 Box Plots for Wind Speed by Siteid, Year and Month 116
A.11 Box Plots for Relative Humidity by Siteid, Year and Month 117
A.12 Box Plots for Precipitation by Siteid, Year and Month 118

Chapter 1

Introduction

Researchers in diverse areas such as environmental and health sciences are increasingly faced with working with data that are observed over space and time. These data arise because of advances in data collection methods, which use state of the art equipment for collecting data from such platforms as radars and satellites. Increased computational resources aid in the analysis of these data. The use of space-time data can be found in many applications. For example, we investigate a set of space-time data that involves the analysis of nitrate concentrations and their relations to meteorological data such as temperature, relative humidity, wind speed and precipitation and chemical data such as sulfate, ammonium, and ozone as predictors. This data set was collected at selected monitoring sites on a weekly basis over several years in the eastern part of United States. We describe some general features of this spatial/temporal data set in Chapter 4.

Traditionally spatial data are classified into one of three types: point referenced data, areal data and point pattern data. Different data types correspond to different mechanisms of data collection procedures. Let $D \subseteq \mathbb{R}^d \, (d \geq 2)$ denote a domain in the space where we collect data and $s \in D$ represent the location of a site. *Point referenced data* are often referred to as geostatistical data and typically arise when the spatial location s varies continuously over a fixed study region D. For example, in our application to an air pollution problem, data are collected over a domain in the eastern U.S. The data are called *areal* or *lattice data* where the fixed domain D is partitioned into a finite number of areal units with well defined boundaries such as postal codes or counties. Here an observation is thought to be associated with an areal unit of non-zero volume as opposed to a particular location point. Spatial *point pattern data* arise when an event of interest (e.g., an outbreak of a disease) occurs at random locations. In this case, the domain D is random and its index set gives the spatial point pattern. In some cases this information might be supplemented by additional covariate information at the event locations. See Chapter 1 of Schabenberger and Gotway (2005) for additional discussion on spatial data types.

Similar to spatial data, time series data can be categorized into continuous or discrete. *Continuous* time series data are observed at every instant of time (e.g., lie detectors), and *discrete* time series data are usually observed at regularly spaced intervals (e.g., weekly share prices and daily rainfall). Notice that frequently observed discrete time series data can be used to approximate a continuous time series.

In this thesis, we focus on space-time data, which are a combination of spatial point referenced data and discrete time series data. Thus, for our application we only consider geostatistical (point-referenced) data observed over a discrete time grid.

1.1 Statistical Models for Space-Time Data

In recent years, there has been widespread attention in the statistical literature given to models for space-time data (Mardia et al., 1998; Kyriakidis and Journel, 1999; Wikle and Cressie, 1999; Stroud et al., 2001; Gelfand et al., 2005). Often, in modeling space-time data, it is of interest to predict the time evolution of a response variable over a given spatial domain. To this end, statistical models are employed to obtain accurate predictions of a response variable, such as nitrate concentrations.

Following Banerjee et al. (Chapter8, 2004), the general form of models for space-time data can be defined as

$$Z(s_i, t) = Y(s_i, t) + \epsilon(s_i, t),\ i = 1, \cdots, n,\ t = 1, \cdots, m, \tag{1.1}$$

where $Z(s_i, t)$ represents the observed response variable, $Y(s_i, t)$ represents the underlying space-time process and $\epsilon(s_i, t)$ represents the error process which is assumed to be a white noise process. The space-time process $Y(s_i, t)$ can be expressed as

$$Y(s_i, t) = \mu(s_i, t) + \omega(s_i, t), \tag{1.2}$$

where $\mu(s_i, t)$ is a mean process and $\omega(s_i, t)$ is a zero mean space-time process. We generally assume the ϵ-process to be independent of the ω-process. The mean process is usually modeled using parametric or nonparametric regression with a set of observed covariates $\mathrm{x}(s_i, t)$. A parametric covariance structure is typically assumed for the $\omega(s_i, t)$ process. The space-time covariance function is defined as

$$C(s_1, s_2; t_1, t_2) = Cov[\omega(s_1, t_1), \omega(s_2, t_2)]. \tag{1.3}$$

The space-time process $\omega(s,t)$ is said to be *stationary* if

$$C(s_1, s_2; t_1, t_2) = C(s_1 - s_2; t_1 - t_2) = C(d; \tau), \tag{1.4}$$

where $d = s_1 - s_2$ and $\tau = t_1 - t_2$ denote the separation vectors, which means the space-time covariance function is a function of the separation vectors. A stationary process is said to be *isotropic* if

$$C(d; \tau) = C(||d||; |\tau|), \tag{1.5}$$

that is, the covariance function depends on the separation vectors only through their lengths $||d||$ and $|\tau|$. Processes which are not isotropic are called *anisotropic.* In the literature isotropic processes are popular because of their simplicity and interpretability. An isotropic process $\omega(s,t)$ is said to be *separable* if

$$C(||d||; |\tau|) = C_s(||d||)C_t(|\tau|). \tag{1.6}$$

Suitable forms for the functions $C_s(\cdot)$ and $C_t(\cdot)$ are available in the literature. A popular choice for $C_s(\cdot)$ is the Matern covariance (Matern, 1986) function and the ARMA (Box and Jenkins, 1976) covariance function for $C_t(\cdot)$.

Various approaches have been proposed to model space-time processes (Kyriakidis and Journel, 1999). One can consider the space-time problem from a multivariate geostatistical perspective, which requires that the space-time covariance functions be specified (Cressie and Huang, 1999; Gneiting, 2002; Schmidt and O'Hagan, 2003; Stein, 2005). This approach has been limited in that the known class of valid space-time covariance functions is quite small, and such covariance functions are often not realistic for complicated dynamical processes. In addition, high dimensionality of these space-time models can prohibit practical implementation.

Space-time models are often constructed by combining traditional time series techniques with methods from spatial statistics. In the time series context, popular approaches include ARIMA models (Box, Jenkins and Reinsel, 1994) for stationary data, and dynamic linear models (West and Harrison, 1997), which allow for nonstationary components such as temporal trends and seasonality.

Early attempts to develop space-time models assumed temporal stationarity. In an early Bayesian application, Handcock and Wallis (1994) considered the space-time modeling of winter temperature data observed over a region in the northern United States. They employed stationary Gaussian process models with an AR(1) model for the time series at each location and carried out separate spatial analyses to study global warming in each year. Carroll et al. (1997) again used stationary Gaussian processes, assuming a separable form for the space-time covariance function to study ground level ozone. Their model combines trend terms incorporating temperature and hourly or monthly effects, and an error model in which the correlation in the residuals is a nonlinear function of time and space, in particular the spatial structure is a function of the lag between observations. Brown et al. (2000) considered the space-time modeling of rainfall data using a non-separable model. They showed that this model is well suited to a wide range of realistic problems which will be poorly fitted by separable models.

More recently, researchers have developed space-time models that allow for nonstationary components. Guttorp et al. (1994) modeled the spatial covariances of hourly ozone levels using the Sampson and Guttorp (1992) nonparametric spatial covariance approach. They allowed the parameters of the model to vary as a function of time of day. Other approaches involving hierarchical Bayesian models include Wikle et al. (1999) and Waller et al. (1997). Wikle et al. (1999) analyzed monthly maximum atmospheric temperatures and Waller et al. (1997) used generalized linear models to map lung cancer rates in Ohio.

There is also much recent space-time modeling, which employs a Markov random field structure in the form of conditionally autoregressive (CAR) specifications. Waller et al. (1997) studied disease mapping, and Gelfand et al. (1998) looked at single family home sales. Pace et al. (2000) worked with simultaneous autoregressive (SAR) models extending them to allow temporal neighbors as well as spatial neighbors.

Researchers have also found the dynamic linear model (DLM), or state-space, framework convenient for analyzing spatial time series. Many of the authors took a Bayesian approach, often relying on Markov Chain Monte Carlo (MCMC) simulation for posterior inference. MCMC methods for DLM are extensively described by West and Harrison (1997). A DLM framework is a common approach used in environmental problems, which are commonly temporally rich in data to update uncertainties about general model parameters as observations are made. Examples include Shaddick and Wakefield (2002) and Stroud et al. (2001). Shaddick and Wakefield (2002) modeled multiple-pollutant data sets measured over time at multiple sites within a dynamic linear modeling framework. The modeling was carried out to provide exposure for a study investigating the health effects of air pollution. In Stroud et al. (2001) a DLM was used principally in the time domain but only a restricted number of spatial points were modeled. Inferences for intermediate points were then made using *kriging*. A particular advantage of this approach is that data are not restricted to a lattice framework and the updates have a relatively low computational load using the Kalman filtering and smoothing algorithms. Kriging is a common interpolation technique used for making spatial predictions at locations where measurements have not been obtained. Points which are close together have a higher spatial correlation than points that are far apart. Predictions at unmeasured locations are based on a weighted average of measured locations, where weights are assigned to all points based on the rate of information decay when moving away from a measurement point. Sansó and Guenni (1999) proposed a DLM with unknown covariance parameters to model rainfall data. Tonellato (1997) developed a state-space model with both stationary and nonstationary temporal components, which was applied in Tonellato (1998) to an Irish wind power prediction problem. Huerta et al. (2004) modeled ozone concentrations over Mexico City and carried out spatial as well as temporal interpolation and prediction using DLM. Banerjee et al. (2005) discussed univariate space-time dynamic models and multivariate spatial dynamic models. They allowed for nonlinear mean structure and non-stationary association structure in modeling space-time data.

State-space approaches in a non-Bayesian setting include Huang and Cressie (1996) and Mardia et al. (1998). Huang and Cressie (1996) modeled snow water in time and space using a separable dynamic model. Mardia et al. (1998) proposed a kriged Kalman filter and outlined a likelihood-based estimation strategy.

From a methodological point of view, space-time data require substantial work, as we must develop models that account for both spatial and temporal correlations. Such modeling also carries an obvious associated increase in computational complexity. Also, the dimension of space-time data sets can be very large, and moreover, space-time processes are often complicated in that the dependence structure across space and time is non-trivial, often non-separable and non-stationary in space and/or time. Hence, space-time modeling is a challenging task which requires the manipulation of large data sets and the ability to fit realistic and complex models. In particular, parameter estimation can be problematic due to the high dimensionality. Parameter identifiability is often a difficult task with high dimensional models. Several modeling strategies have been proposed to address this problem. Many authors have considered ways of reducing the computational load of the multivariate DLM. A common approach is to reduce the number of dimensions, for instance by constructing the update using summary variables. Wikle and Cressie (1999) developed a space-time Kalman-filter that achieves dimension reduction by decomposing the state-process into sets of basis functions and time series. To avoid a high computational load they used an empirical Bayesian method for the estimation of model parameters rather than a fully Bayesian hierarchical approach. However, the additional variability in estimating parameters is ignored in such a method. Stroud et al. (2001) specified very simple random walk dynamics, while Xu and Wikle (2004) discussed efficient estimation approaches via the expectation-maximization (EM) algorithm for the parameter and covariance matrices with high dimensionality in dynamic space-time models.

In Chapter 2 of this thesis, we propose a reparametrization approach to fit dynamic space-time models with an unstructured covariance function. That is, our models are not limited by the stationarity and isotropy restrictions for the covariance function. Our modeling contribution is to present an unconstrained reparametrization method to be used for a covariance function within dynamic space-time modeling framework. Using this unconstrained reparametrization method, we are able to implement the models to fit high-dimensional data without many restrictive assumptions. We take a Bayesian approach and use Markov chain Monte Carlo (MCMC) simulation techniques to obtain the parameter estimates followed by predictive inference. Recent developments in MCMC computing allow fully Bayesian analyses of complex multilevel models for dynamic space-time data. Pole et al. (1994) and West and Harrison (1997) are good references on the dynamic models from a Bayesian point of view.

1.2 Model Selection Methods

Model choice is a crucial issue in statistical data analysis. Researchers typically consider a number of plausible models in statistical applications, and hence model comparison is required to identify the "best " model among several candidate models. Model comparison enables the selection of a suitable model based on a given data set and other modeling information. A variety of model selection methods are available in the literature. We discuss these methods in this section.

Hypothesis testing is probably the most frequently used method in model selection. In particular, sequential hypothesis testing has most often been used. Sequential testing can be performed via one of the forward, backward, and stepwise methods. The forward method begins with no variables in the model. Then variables are added one by one to the model until no remaining variable produces a significant F statistic. The variable that has the largest F statistic is added to the model. The F

statistic calculated for each variable reflects the variable's contribution to the model if it is included. The backward method begins with all of the variables included in the model. Then the variables are deleted from the model one by one until all the variables remaining in the model produce significant F statistics. At each step, the variable showing the smallest contribution to the model is deleted. The stepwise method modifies the forward and backward methods and differs in that variables can be added or deleted at each step. This stepwise process ends when none of the variables outside the model has a significant F statistic and every variable in the model is significant.

However, hypothesis testing is defined only for nested models. A model M_1 is called nested under M_2 if M_1 is a special case of model M_2. Also, Akaike (1974) stated that hypothesis testing, in general, performs very poorly when used for model selection.

Cross-validation is one of the methods that also has been suggested for model selection (Stone, 1974). The holdout method is the simplest kind of cross validation. The data set is separated into two sets, called the training set and the testing set. The training set is used for model fitting, and the testing set is used for model validation. One predicts the data in the testing set using the fitted model based on the training set. Then a best model is selected by a chosen criterion such as minimum squared prediction error. However, this method depends heavily on how the division is made. K-fold cross validation is one way to obtain improvement over the holdout method. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other $k-1$ subsets are put together to form a training set. Then the average error across all k trials is computed. The advantage of this method is that it matters less how the data gets divided, however it becomes computationally very intensive. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of doing this is that one can independently choose how large each test set is and how many trials you average over.

The adjusted coefficient of multiple determination (R^2) has been used in model selection for classical multiple linear regression analysis. This is computed as

$$1 - (1 - R^2)\left(\frac{n-1}{n-p}\right),$$

where R^2 is the usual coefficient of multiple determination (Draper and Smith, 1981). Under this method, one selects the model in which this adjusted statistic is largest. McQuarrie and Tsai (1998) found this approach to be very poor in model selection.

Mallows' C_p (Mallows 1973) statistic is well known for variable selection, but limited to multiple linear regression problems with normal errors. Mallows' C_p is computed as

$$C_p = \frac{SS_{res}}{MS_{res}} - n + 2p,$$

where SS_{res} is the residual sum of squares for the model with $p-1$ variables, MS_{res} is the residual mean square when using all available variables, n is the number of observations, and p is the number of variables used for the model plus one. The general procedure to find an adequate model by means of the C_p statistic is to calculate C_p for all possible combinations of variables. The model with the lowest C_p value which is approximately equal to p is the most "adequate " model.

Information criteria have played an important role in model selection. These criteria are based on information theory. The basis for the information-theoretic approach to model selection is the Kullback-Leibler (K-L) information (Kullback and Leibler, 1951) given by,

$$I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)dx,$$

where θ represents parameters in model g. Here, $I(f,g)$ can be interpreted as the "information lost when the model g is used to approximate full reality or truth f" (Burnham and Anderson, Section 2.1, 2002). $I(f,g)$ also can be interpreted as a "distance " from the approximating model g to truth f. We seek to find a candidate model that minimizes $I(f,g)$ over the candidate models. However, $I(f,g)$ cannot be used directly because it requires knowledge of $f(x)$ and the parameters in models $g(x|\theta)$.

Akaike's information-theoretic approach has led to a number of methods having desirable properties for the selection of the best approximating models in practice. Akaike (1973) provided a way to estimate relative, expected $I(f,g)$ based on the empirical log-likelihood function. He found that the maximized log-likelihood value was a biased estimate of the relative, expected K-L information, and that under certain conditions this bias was approximately equal to p, the number of parameters in the approximating model g. His method, Akaike's Information Criterion (AIC) is based on this finding. AIC is defined as,

$$\text{AIC} = -2\log\left[L(\hat{\theta}|X)\right] + 2p,$$

where $\log[L(\hat{\theta}|X)]$ is the value of the log-likelihood at its maximum point, and p

denotes the number of paramters in the model.

Takeuchi (1976) derived a general method to get from K-L information to AIC. He derived an asymptotically unbiased estimator of the relative, expected K-L information without special conditions. Takeuchi's Information Criterion (TIC) has a more general bias adjustment term,

$$\text{TIC} = -2\text{log}\left[L(\hat{\theta}|X)\right] + 2\text{tr}\left(J(\theta)I(\theta)^{-1}\right),$$

where $J(\theta)$ represents the variance matrix of the first-order derivatives, $I(\theta)$ represents minus the expected value of the matrix of second-order derivatives of the log-likelihood with respect to the parameter θ and "tr " denotes the matrix trace function. AIC is an approximation to TIC, where $\text{tr}(J(\theta)I(\theta)^{-1}) = p$. TIC requires large sample sizes to estimate the elements of the two $p \times p$ matrices in the bias-adjustment term. This criterion is known to be useful when the candidate models are not particularly close approximations to f (Burnham and Anderson, 2002).

The method for small sample approximations, called Corrected AIC (AICc), was proposed by Sugiura (1978) and Hurvich and Tsai (1989). They pointed out that AIC may perform poorly if the number of parameters are too large in relation to the size of the sample. AICc is defined as,

$$\text{AICc} = -2\text{log}\left[L(\hat{\theta}|X)\right] + 2p\left(\frac{n}{n-p-1}\right),$$

where n is sample size. Unless the sample size is large with respect to the number of estimated parameters, use of AICc is recommended.

Information criteria are applicable across a very wide range of statistical models. Burnham and Anderson (2002) recommends the use of AIC and AICc because they are easy to compute and quite effective in many applications. In most practical situations, AIC and AICc are very useful approximations to the relative K-L information.

QAIC and QAICc, based on quasi-likelihood theory, have been derived for appropriate model selection when count data are found to be overdispersed. If overdispersion is found in the analysis of count data, the nominal log-likelihood function must be divided by an estimate of the overdispersion to obtain the correct log-likelihood. The principles of quasi-likelihood suggest simple modifications to AIC and AICc (Lebreton et al., 1992). QAIC and QAICc are defined as,

$$\begin{aligned} \text{QAIC} &= -2\text{log}\left[L(\hat{\theta}|X)/\hat{c}\right] + 2p, \\ \text{QAICc} &= -2\text{log}\left[L(\hat{\theta}|X)/\hat{c}\right] + 2p\left(\frac{n}{n-p-1}\right), \end{aligned}$$

where $\hat{c}$ is an estimated overdispersion factor.

AIC, AICc, QAIC, and QAICc are estimates of the relative K-L distance between truth f and the approximating model g. These criteria were motivated by the concept that the truth is very complex and that no true model exists. Thus, one could only approximate truth with a model, say g. Given a good set of candidate models for the data, one could estimate which approximating model is best. TIC allows that the set of candidate models does not include f or any model similar to f (Burnham and Anderson, 2002).

Several criteria have been developed, based on the assumption that a true model exists, that it is one of the candidate models being considered, and that the model selection goal is to select the true model. These criteria are derived to provide a consistent estimator of the dimension (p) of this true model, and the probability of selecting this true model approaches one as sample size increases. The best known of these dimension-consistent criteria is the Bayesian Information Criterion (BIC), which was derived by Schwarz(1978) in a Bayesian context. It is defined as,

$$\text{BIC} = -2\text{log}\left[L(\hat{\theta}|X)\right] + p\text{log}(n),$$

where p is the dimension of the model and n is sample size. BIC arises from a Bayesian viewpoint with equal prior probability on each model and very vague priors on the parameters, given the model. BIC is not an estimator of relative K-L information. Bozdogan (1987) provides a review of many of the other dimension-consistent criteria.

Spiegelhalter et al. (2002) have developed a Deviance Information Criterion (DIC) from a Bayesian perspective that is analogous to AIC. DIC is one of the common methods for model comparison in the Bayesian framework. It is defined as,

$$\text{DIC} = -2\text{log}\left[L(\bar{\theta}|X)\right] + 2p_{DIC},$$

where $\bar{\theta} = E[\theta|X]$ denotes the posterior mean of θ given the data X, and p_{DIC} denotes the effective number of parameters given by $p_{DIC} = E[D(\theta)|X] - D(E[\theta|X])$, where $D(\theta) = -2\text{log}[L(\theta|X)]$ denotes the deviance of the model. DIC seems to behave like AIC rather than like BIC (Spiegelhalter et al., 2002). One disadvantage of DIC is that it usually requires computationally intensive methods like Markov Chain Monte Carlo (MCMC) approach to compute both of its components.

Another measure that can be used as a model selection tool in a Bayesian framework is the predictive criterion suggested by Laud and Ibrahim (1995), and further developed by Gelfand and Ghosh (1998). This method assesses the predictive performance of a model in terms of prediction accuracy of a replicate of the observed data while still being loyal to the original data. A simplified version of this predictive criterion computes the predictive mean of the squared difference between the observed data and a replicate of data under the model, and chooses the model with the smallest value. The predictive criterion can also be decomposed into two terms (Gelfand and Ghosh, 1998). The first term can be viewed as a goodness of fit term with the mean of the posterior predictive distribution of the replicate replacing the maximum likelihood estimate of the mean, and the second term is a variance term that can be thought of as a penalty function. Gelfand et al. (1998) adopt this criterion to select a model for the analysis of residential sales data based on a variety of space-time models.

The principle of parsimony provides a basis for model selection. As the goal of model selection using information criteria is to select parsimonious models, models that minimize the criterion are selected.

We further investigate information criteria, such as AIC, AICc, and BIC, for model selection in Chapter 3. We discuss comparisons of these criteria and explore their performance using Monte Carlo simulations. In particular, we evaluate their ability to select the correct model in a spatial modeling context.

Chapter 2

A Reparametrization Approach for Dynamic Space-Time Models

2.1 Introduction

Statistical modeling of time series processes is usually based on classes of dynamic models. The term *dynamic* relates to changes in such processes due to the passage of time. Following West and Harrison (1997), in a Bayesian framework, forecasting problems through dynamic modeling are structured using four fundamental principles:

(i) sequential model definition,

(ii) structuring using parametric models,

(iii) probabilistic representation of information about parameters, and

(iv) forecasts derived as probability distributions.

Suppose that the time origin, $t = 0$, represents the current time, and that the existing information available is denoted by D_0, the initial information set. This represents all the available relevant starting information which is used to form ones initial views about the future. In forecasting ahead to any time $t_1 > 0$, the primary objective is calculation of the forecast distribution for $(Y_{t_1}|D_0)$, where Y_{t_1} is the observation at time t_1. Generally, we denote by D_t the information set available at time t. Thus statements made at time t about any random quantities of interest are based on D_t. In particular, forecasting ahead to time $t_2 > t_1$ involves consideration of the forecast distribution for $(Y_{t_2}|D_{t_1})$. Observing the value of Y_{t_2} at time t_2 implies that D_{t_2} includes both the previous information set D_{t_1} and the observation Y_{t_2}, that is, $D_{t_2} = \{Y_{t_2}, D_{t_1}\}$, which represents information updating. The sequential focus is emphasized through the use of statistical models for the development of the series into the future described via distributions for Y_{t_2}, $Y_{t_3}(t_3 > t_2), \cdots$, conditional on past information D_{t_1}. From now on we assume that the temporal observations are collected at regular time interval denoted by $t = 1, 2, \cdots$. Focusing on one-step ahead, the forecaster's views are then structured in terms of a parametric model,

$$p(Y_t|\boldsymbol{\theta}_t, D_{t-1}),$$

where $\boldsymbol{\theta}_t$ represents the parameter vector at time t. Indexing $\boldsymbol{\theta}_t$ by t indicates that the parameterization is dynamic. The model parameter $\boldsymbol{\theta}_t$ provides the means by which information relevant to forecasting the future is summarized and used in forming forecast distributions. The learning process involves sequentially revising the state of knowledge about such parameters. At time t, historical information D_{t-1} is summarized through a prior distribution for future model parameters. The prior density

$p(\boldsymbol{\theta}_t|D_{t-1})$ and the posterior density $p(\boldsymbol{\theta}_t|D_t)$ provide a concise, coherent and effective transfer of information on the time series process through time. An ultimate goal is attained by directly applying probability laws. That is,

$$p(Y_t, \boldsymbol{\theta}_t|D_{t-1}) = p(Y_t|\boldsymbol{\theta}_t, D_{t-1})p(\boldsymbol{\theta}_t|D_{t-1}),$$

from which the relevant one-step forecast may be deduced as the marginal

$$p(Y_t|D_{t-1}) = \int p(Y_t, \boldsymbol{\theta}_t|D_{t-1})d\boldsymbol{\theta}_t.$$

Inference for the future Y_t is now a standard statistical problem of summarizing the information encoded in the forecast distribution.

2.2 Dynamic Linear Models

The most widely known and used subclass in dynamic models are the Gaussian Dynamic Linear Models, referred to more simply as Dynamic Linear Models (DLM) (West and Harrison, 1997), where the normality and linearity are assumed for sequential model definitions and structured parameters, respectively. The DLM framework has been a common approach used in space-time data, which are commonly temporally rich, such as environmental problems to update uncertainties about general model parameters as observations are made. The DLM can be seen as a generalization of regression models that allow changes in parameter values over time and provide a very flexible framework that permits smooth and abrupt changes in the time series generating the process. In this section we discuss traditional DLM in the time series context.

Let $\boldsymbol{y}_t$ be a $n \times 1$ vector of observations at time t. $\boldsymbol{y}_t$ is modeled conditionally based on a $p \times 1$ vector, $\boldsymbol{\theta}_t$, called the state vector, through a measurement equation or a *observation equation.* In general, the elements of $\boldsymbol{\theta}_t$ are not observable, but are generated by a first-order Markovian process, resulting in a transition equation or *evolution equation.* Therefore, we can describe the above framework for $t = 1, 2, \cdots,$ as

$$\text{Observation equation}: \quad \boldsymbol{y}_t = F_t\boldsymbol{\theta}_t + \boldsymbol{\upsilon}_t, \quad \boldsymbol{\upsilon}_t \sim N(\boldsymbol{0}, \Sigma_t^{\upsilon}),$$

$$\text{Evolution equation}: \quad \boldsymbol{\theta}_t = G_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \Sigma_t^{\eta}),$$

where F_t and G_t are $n \times p$ and $p \times p$ matrices, respectively. The first equation is the observation equation, where $\boldsymbol{\upsilon}_t$ is a $n \times 1$ vector of serially uncorrelated Gaussian variables with mean zero and a $n \times n$ covariance matrix, Σ_t^{υ}. The second equation is the evolution equation with $\boldsymbol{\eta}_t$ being a $p \times 1$ vector of serially uncorrelated zero-mean Gaussian disturbances and Σ_t^{η} the corresponding $p \times p$ covariance matrix. The model is completed with a Gaussian prior on the initial state: $\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, C_0)$, with $\boldsymbol{\theta}_0$ independent of $\boldsymbol{\upsilon}_t$ and $\boldsymbol{\eta}_t$. We further assume that $\boldsymbol{\eta}_t$ and $\boldsymbol{\upsilon}_t$ are independent.

F_t and G_t are referred to as system matrices which are allowed to change over time. The matrix F_t is usually specified by the design of the problem at hand, while G_t is specified through modeling assumptions; for example, $G_t = I_p$, the $p \times p$ identity matrix would render a vector autoregressive (VAR) random walk of order one $VAR(1)$ process for $\boldsymbol{\theta}_t$. A random walk prior is a natural choice when no prior information is available. As a result, the system is linear, and for any time point t, $\boldsymbol{y}_t$ can be expressed as a linear combination of the present $\boldsymbol{\upsilon}_t$ and the present and past $\boldsymbol{\eta}_t$'s.

The evolution variance matrix Σ_t^η can be specified either explicitly or through a discount factor $\delta \in (0, 1]$, which defines $\Sigma_t^\eta = \frac{1-\delta}{\delta} P_t$, with $P_t = Var(G_t\boldsymbol{\theta}_{t-1}|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{t-1})$. When p is large, the discount method is usually preferred, since it requires elicitation of only one parameter instead of $p(p + 1)/2$. The observation variance matrix Σ_t^v is usually assumed to be a diagonal matrix. We will show that our proposed reparametrization method allows one to relax these restrictive assumptions.

2.3 Dynamic Space-Time Models (DSTM)

We adapt the above DLM framework to space-time data. The approach taken here is to view the data as arising from a time series of spatial processes.

Suppose the process $Z(s, t)$ is observed on a finite number of sites labeled as $s_1, ..., s_n$ at each time t, where $t = 1, 2, \cdots, m$. Consider the $n \times 1$ vector time series $\boldsymbol{Z}_t = (Z(s_1, t), ..., Z(s_n, t))^T$ at time t. For each t, the DLM is usually formed by an observation equation and an evolution equation. An observation equation describes the relationship between the observation $(\boldsymbol{Z}_t)$ and the regressors (X_t) that takes the form of a multivariate regression process,

$$\boldsymbol{Z}_t = X_t\boldsymbol{\beta_t} + \boldsymbol{\nu_t}, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \Sigma_t^\nu) \tag{2.1}$$

where X_t is an $n \times p$ observed design matrix and $\boldsymbol{\beta}_t$ is a $p \times 1$ vector of regression coefficients or state parameters. An evolution equation describes the dynamics of the vector of regression coefficients or state parameters $\boldsymbol{\beta}_t$ through time,

$$\boldsymbol{\beta}_t = G_t\boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \Sigma_t^\omega) \tag{2.2}$$

where G_t is a $p \times p$ evolution matrix. There are several ways to model the G_t's. The most common assumption is that the G_t's are structurally known, possibly up to some finite number of parameters. In this thesis, we do not make any structural assumption about the G_t's but we assume that $G_t = G$ for all t and that G follows a matrix-valued normal distribution with mean G_0 and variance-covariance parameters Ω_0 and Σ_0^G. That is, $G \sim MN_{p\times p}(G_0,\, \Omega_0,\, \Sigma_0^G)$. We also assume that the $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ error vectors are independent and have multivariate normal distributions with mean $\mathbf{0}$ and variance-covariance matrices Σ_t^ν and Σ_t^ω, respectively. The model is completed with a normal prior for the initial state, $\boldsymbol{\beta}_1 \sim N(\boldsymbol{\beta}_0, \Sigma_0^\omega)$, where $\boldsymbol{\beta}_0$ is known.

Equivalently, the model can be written using hierarchical specifications as follows:

$$\begin{aligned}
\boldsymbol{Z}_t|\boldsymbol{\beta}_t &\sim N(X_t\boldsymbol{\beta}_t, \Sigma_t^\nu),\\
\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, G &\sim N(G\boldsymbol{\beta}_{t-1}, \Sigma_t^\omega),\\
G &\sim MN(G_0, \Omega_0, \Sigma_0^G),\\
\boldsymbol{\beta}_1 &\sim N(\boldsymbol{\beta}_0, \Sigma_0^\omega),
\end{aligned}$$

where the matrix valued normal distribution $MN(G_0, \Omega_0, \Sigma_0^G)$ has the probability density function given by

$$p(G|G_0, \Omega_0, \Sigma_0^G) = (2\pi)^{-p^2/2}|\Omega_0|^{-p/2}|\Sigma_0^G|^{-p/2} exp\left(-\frac{1}{2}\text{tr}\left[{\Omega_0}^{-1}(G-G_0){\Sigma_0^G}^{-1}(G-G_0)^T\right]\right).$$

This Bayesian hierarchical approach not only helps organize our thinking about the model, but also fully accounts for all sources of uncertainty without making substantial structural assumptions like, spatial stationarity, isotropy, etc.

In order to keep our illustrations simple, first we consider the following simplified DLM. Specifically, we assume that $\Sigma_t^\nu = \Sigma^\nu$ and $\Sigma_t^\omega = \Sigma^\omega$ in equation (2.1) and (2.2).

That is, variance-covariance matrices Σ^{ν} and Σ^{ω} do not change over time (and hence are static). Our simplified DLM can be written as,

$$\begin{aligned} \boldsymbol{Z}_t &= X_t\boldsymbol{\beta_t} + \boldsymbol{\nu}_t, \;\; \boldsymbol{\nu}_t \sim N(\mathbf{0}, \Sigma^{\nu}), && (2.3) \\ \boldsymbol{\beta_t} &= G\boldsymbol{\beta_{t-1}} + \boldsymbol{\omega}_t, \;\; \boldsymbol{\omega}_t \sim N(\mathbf{0}, \Sigma^{\omega}), && (2.4) \end{aligned}$$

and $\boldsymbol{\beta}_1 \sim N(\boldsymbol{\beta}_0, \Sigma^{\omega}_0)$. Here, Σ^{ν} and Σ^{ω} are unstructured variance-covariance matrices and for the ease of exposition, we initially assume that $\boldsymbol{Z}_t$ and $\boldsymbol{X}_t$ are observed at all time points t. Later we discuss how to relax this assumption if some observations are missing.

The Bayesian model is completed with the specification of a prior distribution for the parameters. These include the data-model covariance matrix Σ^{ν} and the error covariance matrix Σ^{ω}. Prior specifications for Σ^{ν} and Σ^{ω} can be tricky as these matrices are usually high-dimensional (e.g., Σ^{ν} is $n \times n$) and they need to be positive definite (pd). It is customary to use the inverse Wishart distributions to model such covariance matrices. Note that our models require no restrictive assumptions such as stationarity and isotropy as Σ^{ν} is left unstructured. However, if such assumptions are deemed necessary, we can easily incorporate them in our modeling framework.

As we mentioned earlier, inference for dynamic models is made sequentially by obtaining the prior predictive and updated distributions for the state parameters $\boldsymbol{\beta}_t$ for each time t. The prior predictive distributions are respectively obtained by

$$\begin{aligned} p(\boldsymbol{\beta}_t|D_{t-1}) &= \int p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})p(\boldsymbol{\beta}_{t-1}|D_{t-1})d\boldsymbol{\beta}_{t-1} \\ \\ p(\boldsymbol{Z}_t|D_{t-1}) &= \int p(\boldsymbol{Z}_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|D_{t-1})d\boldsymbol{\beta}_t \end{aligned}$$

and the updated distribution is obtained by Bayes' theorem as

$$p(\boldsymbol{\beta}_t|D_t) \propto p(\boldsymbol{Z}_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|D_{t-1})$$

where D_t represents all the available information upto time t.

2.4 Model Fitting

The most popular computing tools in Bayesian practice today are Markov chain Monte Carlo (MCMC) methods. This is due to their ability to enable inference from posterior distributions of arbitrarily large dimension, essentially by reducing the problem to one of recursively solving a series of lower-dimensional problems. Like traditional Monte Carlo methods, MCMC methods work by producing not a closed form for the posterior, but a sample of values $\{\theta^{(l)},\ l = 1, \cdots, B\}$ from this distribution. While this obviously does not carry as much information as the closed form itself, a histogram or kernel density estimate based on such a sample is typically sufficient for reliable inference. Moreover such an estimate can be made accurate merely by increasing the Monte Carlo sample size B. However, unlike traditional Monte Carlo methods, MCMC algorithms produce correlated samples from the posterior, since they arise from recursive draws from the path of a Markov chain, the stationary distribution of which is the same as the posterior.

Suppose our model features p parameters, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$. To implement the Gibbs sampler, we must assume that samples can be generated from each of the full conditional distributions $\{p(\theta_i|\boldsymbol{\theta}_{j\neq i}, \boldsymbol{y}), i = 1, \cdots, p\}$ in the model, where $\boldsymbol{y}$ denotes the set of observed data. Under mild conditions, the collection of full conditional distributions uniquely determine the joint posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, and hence all marginal posterior distributions $p(\theta_i|\boldsymbol{y})$, $i = 1, \cdots, p$. Given an arbitrary set of starting values $\{\theta_2^{(0)}, \cdots, \theta_p^{(0)}\}$, the algorithm for the Gibbs sampler proceeds as follows:

Step 1 Draw $\theta_1^{(l)}$ from $p(\theta_1|\theta_2^{(l-1)}, \theta_3^{(l-1)}, \cdots, \theta_p^{(l-1)}, \boldsymbol{y})$

Step 2 Draw $\theta_2^{(l)}$ from $p(\theta_2|\theta_1^{(l)}, \theta_3^{(l-1)}, \cdots, \theta_p^{(l-1)}, \boldsymbol{y})$

$\vdots$

Step p Draw $\theta_p^{(l)}$ from $p(\theta_p|\theta_1^{(l)}, \theta_2^{(l)}, \cdots, \theta_{p-1}^{(l)}, \boldsymbol{y})$

The cycles from **Step 1** to **Step p** are repeated for $l = 1, \cdots, B$. Under mild regulatory conditions that are generally satisfied for most statistical models (see Geman and Geman, 1984), one can show that $\{\theta_1^{(l)}, \cdots, \theta_p^{(l)}\}$ converges in distribution to a draw from the true joint posterior distribution $p(\theta_1, \cdots, \theta_p|\boldsymbol{y})$. This means that for l sufficiently large, say l_0, $\{\boldsymbol{\theta}^{(l)}, l = l_0 + 1, \cdots, B\}$ is a sample from the true posterior, from which any posterior quantities of interest may be estimated. For example, a histogram of the $\{\theta_i^{(l)}, l = l_0 + 1, \cdots, B\}$ provides a simulation-consistent estimator of the marginal posterior distribution for θ_i, $p(\theta_i|\boldsymbol{y})$. The time from $l = 0$ to $l = l_0$ is commonly known as the burn-in period, and posterior estimates are obtained using $\{\boldsymbol{\theta}^{(l)}, l = l_0 + 1, \cdots, B\}$.

We fit our model using a Markov chain Monte Carlo (MCMC) procedure known as the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990) via the `WinBUGS` software which is a window-based software package for Bayesian analysis using MCMC methods. The software can be downloaded from the website

`http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml`.

The updating scheme in the dynamic space-time model may not be easy to implement when the G matrix is completely unknown. Further these types of multivariate updating schemes can be very unstable and time consuming when the dimensions are very large. In order to avoid such numerical instabilities and to accelerate model fitting, we describe an equivalent univariate scheme for the aforementioned DLM using a reparametrization method in the next section. As by-products of this reparametrization method, we obtain several extensions of the usual DLM.

2.5 A Reparametrization Method

In this section, we describe the proposed reparametrization method for a covariance matrix. Modeling a covariance matrix Σ is difficult because (i) it is a high dimensional parameter and (ii) it is restricted to be positive definite. Pourahmadi (1999) introduced an unconstrained parameterization procedure to model a temporal covariance matrix. The Cholesky decomposition of the inverse of a covariance matrix is used to associate a unique unit lower triangular and a unique diagonal matrix with each covariance matrix. The entries of the lower triangular matrix and the log of the diagonal matrix are completely unconstrained and have interpretations similar to regression coefficients and prediction variances, respectively, when regressing a measurement on its predecessors. Using the Cholesky decomposition and the ensuing unconstrained and statistically meaningful reparameterization, Daniels and Pourahmadi (2002) provides a convenient and intuitive framework for developing conditionally conjugate prior distributions for covariance matrices, and show their connections with generalized inverse Wishart priors. However, to the best of our knowledge this type of reparameterization of covariance matrices has been used only to model temporal processes, taking advantage of the natural ordering of time. We extend these methodologies to spatial and temporal processes and show several extensions.

For our DLM framework, we define two lower triangular matrices T^{ν} and T^{ω} and two diagonal matrices D^{ν} and D^{ω} such that $T^{\nu}\Sigma^{\nu}T^{\nu T} = D^{\nu}$ and $T^{\omega}\Sigma^{\omega}T^{\omega T} = D^{\omega}$. Such decompositions of positive definite matrices are unique. More precisely, let T^{ν} and T^{ω} be the lower triangular matrices with 1's as their diagonal entries and $-\phi_{ii'}$, $i > i'$ and $-\psi_{kk'}$, $k > k'$ as their lower triangular entries, respectively. Also, let D^{ν} and D^{ω} be diagonal matrices with entries $\sigma_1^{\nu 2}, \cdots, \sigma_n^{\nu 2}$ and $\sigma_1^{\omega 2}, \cdots, \sigma_p^{\omega 2}$, respectively. We now re-express the equations (2.3) and (2.4) using the entries of their lower triangular and diagonal matrices.

Let $Z_{it} = Z(s_i, t)$, $i = 1, \cdots, n$, $t = 1, \cdots, m$ and $X_{itk} = X_k(s_i, t)$, $k = 1, \cdots, p$. Notice that $\boldsymbol{Z}_t = (Z_{1t}, \cdots, Z_{nt})^T$ and $\{X_t\}_{n\times p} = ((X_{itk}))_{1\leq i\leq n, 1\leq k\leq p}$, which appear in (2.3). Then we can represent the DLM defined by (2.3) and (2.4) as follows. The observation equation can be written as,

$$
\begin{aligned}
Z_{it} &= \sum_{k=1}^{p} \beta_{kt} X_{itk} + \sum_{i'=1}^{i-1} \phi_{ii'} Z_{i't} + \nu_{it}, && (2.5)\\
Z_{1t} &= \sum_{k=1}^{p} \beta_{kt} X_{1tk} + \nu_{1t}, && (2.6)
\end{aligned}
$$

where $i = 2, \cdots, n$, $t = 1, \cdots, m$, $E[\nu_{it}] = 0$, $E[\nu_{it}^2] = \sigma_i^{\nu 2}$, and $E[\nu_{it}\nu_{i't}] = 0$, $i \neq i'$. Notice that $\boldsymbol{\nu}_t = (\nu_{1t}, \cdots, \nu_{nt})^T$ as in (2.3). The evolution equation can now be written as,

$$\begin{aligned} \beta_{kt} &= \sum_{k'=1}^{p} \beta_{k't-1} g_{kk'} + \sum_{k'=1}^{k-1} \psi_{kk'} \beta_{k't} + \omega_{kt}, && (2.7) \\ \beta_{1t} &= \sum_{k'=1}^{p} \beta_{k't-1} g_{1k'} + \omega_{1t}, && (2.8) \end{aligned}$$

where $k = 2, \cdots, p$, $t = 2, \cdots, m$, $E[\omega_{kt}] = 0$, $E[\omega_{kt}^2] = \sigma_k^{\omega 2}$, and $E[\omega_{kt}\omega_{k't}] = 0$, $k \neq k'$, and initial state equation can be written as,

$$\beta_{k1} = \beta_{k0} + \sum_{k'=1}^{k-1} \psi_{kk'} \beta_{k'1} + \omega_{k1}, \qquad (2.9)$$

where $k = 2, \cdots, p$. Notice that $\boldsymbol{\omega}_t = (\omega_{1t}, \cdots, \omega_{nt})^T$ as in (2.2). The model is completed with

$$\beta_{11} = \beta_{10} + \omega_{11}.$$

Here, notice that no structural constraints are required for the elements of T^ν, T^ω, D^ν, and D^ω (i.e., $\phi_{ii'}, \psi_{kk'} \in \mathbb{R}$ and $\sigma_i^{\nu 2}, \sigma_k^{\omega 2} \in (0, \infty)$). In particular, for our applications we may specify a prior distribution for these parameters as,

$$\begin{aligned} \phi_{ii'} &\sim N(\phi_0, \sigma_\phi^2), \;\; 1 \leq i' < i \leq n, \\ \psi_{kk'} &\sim N(\psi_0, \sigma_\psi^2), \;\; 1 \leq k' < k \leq p, \\ \sigma_{\nu i}^2 &\sim IG(a_\nu, b_\nu), \;\; i = 1, \cdots, n, \\ \sigma_{\omega k}^2 &\sim IG(a_\omega, b_\omega), \;\; k = 1, \cdots, p, \\ g_{kk'} &\sim N(g_0, \sigma_g^2), \;\; k, k' \in 1, \cdots, p, \end{aligned}$$

where ϕ_0, σ_ϕ^2, ψ_0, σ_ψ^2, a_ν, b_ν, a_ω, b_ω, g_0 and σ_g^2 are all known values that can be used to quantify the prior information if available, otherwise we can use values that would generate a set of vague priors. Here $N(a,b)$ denotes a normal distribution with mean a and variance b, and $IG(a,b)$ denotes an inverse gamma distribution with mean $\frac{b}{a-1}$ for $a > 1$ and variance $\frac{b^2}{(a-1)^2(a-2)}$ for $a > 2$. For instance, we choose these values in such a way that these will have minimal impact on the posterior inference of the parameters. Other prior distributions can also be adapted for our framework very easily.

Using this reparametrized univariate scheme we can avoid numerical instabilities due to high dimensionality. The routine handling of missing data is also apparent by using the reparametrization scheme. If an observation is missing, we just sample it from its full conditional distribution. We illustrate the use of this reparametrization method by applying our model to a set of nitrate concentration data in Chapter 4.

The advantages of our proposed reparametrized dynamic space-time models (RDSTM) can be summarized as follows:

i) Numerical stability: RDSTM avoids numerical instabilities caused by multivariate updating scheme when the dimensions are very large.

ii) Routine handling of missing data: RDSTM allows missing data to be imputed from its full conditional distribution.

iii) Implementation using WinBUGS: RDSTM can be implemented using WinBUGS, while multivariate scheme of DSTM cannot.

Our proposed method can be extended to develop a very flexible class of space-time models that are not subject to structural restrictions. Several possible extensions are as follows:

i) Dynamic modeling of covariance function Σ^{ν} and Σ^{ω}: we can allow Σ^{ν} and Σ^{ω} to depend on t, that is, $\phi_{ii'}$'s, $\psi_{kk'}$'s, $\sigma_i^{\nu 2}$'s, and $\sigma_k^{\omega 2}$'s depend on t and thus making these parameters dynamic as well.

ii) Relaxing the Gaussian assumption of distributions for $\boldsymbol{\nu}_t$'s and $\boldsymbol{\omega}_t$'s: we can assume other than Gaussian distributions such as t-distributions for $\boldsymbol{\nu}_t$'s and $\boldsymbol{\omega}_t$'s.

iii) Extension of the first-order Markovian assumption: we can assume $\boldsymbol{\theta}_t$ is generated by a higher order Markovian process rather than a first-order in evolution equations in (2.7) and (2.8).

There are several space-time models available in the literature and it is almost impossible to compare our proposed model to all such models. But we provide a brief description of a model that closely resembles ours and we compare our models to that proposed by Wikle and Cressie (1999).

2.6 Comparison with Space-Time Kalman Filter

Wikle and Cressie (1999) introduced a dimension reduced approach to space-time Kalman filtering (STKF). Their goal is to achieve space-time prediction using the dimension reduction. The model is of the form,

$$\begin{aligned}\boldsymbol{Z}_t &= \boldsymbol{y}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim N(\boldsymbol{0}, R),\\ \boldsymbol{y}_t &= \Phi \boldsymbol{a}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\boldsymbol{0}, V),\\ \boldsymbol{a}_t &= H\boldsymbol{a}_{t-1} + J\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, Q),\end{aligned}$$

where $H = JB$, $J = (\Phi'\Phi)^{-1}\Phi'$, and $n \times K$ matrices Φ and B are defined as $\Phi \equiv [\boldsymbol{\phi}(\boldsymbol{s}_1), \cdots, \boldsymbol{\phi}(\boldsymbol{s}_n)]'$ and $B \equiv [\boldsymbol{b}(\boldsymbol{s}_1), \cdots, \boldsymbol{b}(\boldsymbol{s}_n)]'$. Here, $\boldsymbol{\phi}(\boldsymbol{s}) \equiv [\phi_1(\boldsymbol{s}), \cdots, \phi_K(\boldsymbol{s})]'$, $\boldsymbol{a}_t \equiv [a_1(t), \cdots, a_K(t)]'$, and $\boldsymbol{b}(\boldsymbol{s}) \equiv [b_1(\boldsymbol{s}), \cdots, b_K(\boldsymbol{s})]'$. $\{\phi_j(\boldsymbol{s}) : j = 1, \cdots, K\}$ are a complete and orthonormal sequence of deterministic spatial functions, $\{a_j(t) : t = 1, 2, \cdots.\}$ is a random time series for each $j = 1, 2, \cdots, K$, and $\{b_j(\boldsymbol{s}) : j = 1, \cdots, K\}$ are unknown but nonstochastic parameters. Also, n denotes number of locations and K denotes the key to dimension reduction.

Then the model can be written as,

$$\begin{aligned}\boldsymbol{Z}_t &= \Phi \boldsymbol{a}_t + \boldsymbol{v}_t^*, \quad \boldsymbol{v}_t^* \sim N(\boldsymbol{0}, R + V),\\ \boldsymbol{a}_t &= H\boldsymbol{a}_{t-1} + \boldsymbol{\eta}_t^*, \quad \boldsymbol{\eta}_t^* \sim N(\boldsymbol{0}, JQJ^T).\end{aligned}$$

That is, it has the same form as our dynamic space-time model (DSTM) except that Φ does not change over time and the variance matrices $R + V$ and JQJ^T are also assumed to be static over time.

Wikle and Cressie (1999) used an empirical Bayesian technique for computing efficiency, because a fully Bayesian hierarchical approach requires highly intensive computational resources for Kalman filtering. But doing so, some statistical precision (and hence efficiency) is lost which can be obtained with the fully Bayesian approach. However, our RDSTM achieves both computing efficiency by using the reparametrized univariate scheme and statistical efficiency by using the fully hierarchical approach. Further they used standard method of moments to estimate R, V, Q, and B for computational efficiency and used such plugging in estimator within the Kalman filter. This again leads to underestimation of uncertainty. Moreover, they did not use maximum likelihood estimators because high dimensionality causes difficulty to implement such procedures. As we mentioned earlier, our RDSTM approach are not limited by such approximations due to high dimensionality. In addition, it is not clear if the positive definiteness is ensured for the estimated covariance matrices obtained by such moment based methods. Our RDSTM guarantees that the covariance matrices are positive definite and the estimates are obtained with high efficiency. Finally, it is well-known that a fully hierarchical method of estimation provides the true measure of uncertainty as compared to two-stage methods that uses plugging in estimates from the first stage. We demonstrate the applicability of our RDSTM for a large data set of nitrate concentrations in Chapter 4. We will also provide several ways to fully utilize the estimates obtained from the RDSTM in the context of atmospheric sciences.

Chapter 3

Performance of Information Criteria for Spatial Models

Model selection is an important part of statistical analysis. Many researchers have examined this issue and various methods for selecting the “best ” model have been suggested, as we discussed in Section 1.2. Information criteria have been widely used in model selection. However, there have been few investigations of the performance of these criteria in a spatial modeling context. Hoeting et al. (2006) considered the problem of model selection for geostatistical data. They explored the effect of spatial correlation on model selection using the AIC applied to a geostatistical model. Their simulation results showed that the use of AIC for a geostatistical model is superior to the often used traditional approach of ignoring spatial correlation in the selection of explanatory variables. In particular, few studies have compared the performance between different information criteria like Akaike Information Criterion (AIC) (Akaike, 1973) and Bayesian Information Criterion (BIC) (Schwarz, 1978). Hence, little is known about the relative performance of different information criteria. Currently, there is no consensus on the best criterion for spatial model selection. Of particular interest is how these different criteria perform with various spatial covariance models. We explore this issue via Monte Carlo simulations in this chapter. The purpose of this study is to examine the performance of different information criteria for use in spatial covariance model selection. We compare the performance of traditional information criteria such as AIC, BIC, and Corrected AIC (AICc) (Sugiura, 1978; Hurvich and Tsai, 1989). This comparison is made using various spatial covariance models ranging from stationary isotropic to nonstationary models.

The remainder of this chapter is organized as follows. Section 3.1 introduces various information criteria used for model selection such as AIC, BIC and AICc. In Section 3.2, we describe various spatial covariance models such as stationary isotropic and anisotropic, and nonstationary models that will be used as an illustration to generate data from a specific covariance model. Finally, in Section 3.3., we present results from simulations which compare the performance of AIC, BIC and AICc with regard to their ability to identify the true model among various spatial covariance models.

3.1 Information-Theoretic Criteria for Model Selection

A variety of information-theoretic criteria have been used in model selection as we mentioned in Section 1.2. A substantial advantage in using information criteria is that they are valid for nonnested models. Traditional likelihood ratio tests are defined only for nested models, and this represents a limitation in the use of hypothesis testing in model selection.

Most information criteria have a form that consists of two terms. In general, the first term is the negative log-likelihood, multiplied by two, of the data calculated with the maximum likelihood estimates of the parameters; it represents the amount of information required to describe the data. The second term differs between different information criteria; it represents the amount of information required to describe the model. The second term can be interpreted as a penalty for model complexity; it thus increases as the number of parameters in the model increases. The basic principle of model selection using information criteria is to select statistical models that simplify the description of the data and model. Specifically, information methods emphasize minimizing the amount of information required to express the data and the model. This results in the selection of models that are the most parsimonious or efficient representations of observed data. To select parsimonious models, models minimizing information criteria are selected. In the following sections, we describe commonly used information criteria such as AIC, AICc, and BIC.

3.1.1 Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) (Akaike, 1973) is one of the most well known information criteria used in model selection. AIC is an estimate of relative, expected Kullback-Liebler distance (KLD) (Kullback and Liebler, 1951). The KLD is a quantity which measures the discrepancy from the approximating model to truth. AIC is an estimate of KLD between a fitted model and the true model. AIC is defined as

$$\text{AIC} = -2\log\left[L(\hat{\theta}|X)\right] + 2p, \tag{3.1}$$

where $\log L(\hat{\theta}|X)$ represents the log-likelihood function of the maximum likelihood estimator (MLE), $\hat{\theta}$, given the observed data X, and p is the dimension of the parameter θ. The first term can be interpreted as a measure of lack of model fit, while the second term can be interpreted as a penalty for increasing the dimension of the model. Recall that the second term is the asymptotic bias-correction term. It is the result of deriving an asymptotic estimator of relative expected KLD. In application, one computes AIC for each of the candidate models and selects the model with the smallest value of AIC. Models producing smaller values of AIC can be thought of as having a smaller difference from the true model, where the true model is unknown. AIC provides a simple and effective means for the selection of the best approximating model to the true model (Burnham and Anderson, 2002). With regard to general linear models, AIC is known to perform relatively well for small samples, however the criterion is large-sample inconsistent, which means AIC does not tend to select the true model in large samples (Hurvich and Tsai, 1990).

3.1.2 Corrected Akaike Information Criteria (AICc)

As an approximately unbiased estimator of the expected KLD of a fitted model, AIC has been shown to be strongly negatively biased in small samples (Sugiura, 1978; Hurvich and Tsai, 1989). Hurvich and Tsai (1989) derived a bias-corrected version of AIC, AICc. They argued that AICc should be used in place of AIC, when the dimension of the model is large relative to sample size or when n is small, for any p. The AICc is defined as

$$\begin{aligned}\text{AICc} &= -2\log\left[L(\hat{\theta}|X)\right] + 2p\left(\frac{n}{n-p-1}\right)\\ &= -2\log\left[L(\hat{\theta}|X)\right] + 2p + \frac{2p(p+1)}{n-p-1}\\ &= \text{AIC} + \frac{2p(p+1)}{n-p-1},\end{aligned} \tag{3.2}$$

where n is the sample size and p is the number of parameters in the model. AICc has an additional bias-correction term compared to AIC, which is adjusted to the parameter complexity p and the sample size n. However, if n is large with respect to p, then this additional bias-correction is negligible and AIC should perform well. Burnham and Anderson (2002) advocated the use of AICc, in particular, when the ratio $n/p < 40$ for the model with the largest value of p. If n/p is sufficiently large, then AIC and AICc are similar and will tend to select the same model. They also mentioned that AICc should be used in practice, because AICc converges to AIC as n gets large, with p fixed.

3.1.3 Bayesian Information Criterion (BIC)

Along with AIC, Bayesian Information Criterion (BIC) (Schwarz, 1978) is currently among the most commonly used information criteria in model selection. BIC is usually explained in terms of Bayesian theory, especially as an approximation of the Bayes factor, which is the ratio of the marginal likelihoods for two models. Recall that unlike AIC, BIC is not an estimate of relative expected KLD. BIC is the dimension-consistent criterion as we discussed in Section 1.2.4. BIC is defined as

$$\text{BIC} = -2\text{log}\left[L(\hat{\theta}|X)\right] + p\text{log}(n), \tag{3.3}$$

where log $\left[L(\hat{\theta}|X)\right]$ again represents the log-likelihood function of $\hat{\theta}$, which is the maximum likelihood estimator (MLE) based on the observed data X; p is the number of parameters in the model, and n is the sample size. The first term of BIC is same as that of AIC. However, the second term penalizes the model with increased model complexity, or larger p, and sample size as well. AIC and BIC differ only by the coefficient multiplying the number of parameters, in other words, by how strongly they penalize large models. In general, models chosen by BIC are more parsimonious than those chosen by AIC. As usually used, one computes the BIC for each model and selects the model with the smallest criterion value. In contrast to AIC, BIC is large sample model consistent, that is, BIC tends to choose the true model in large samples. However, BIC has also known to perform poorly in small samples in the context of general linear models (Hurvich and Tsai, 1990).

3.2 Spatial Models

We consider various point-referenced data models in this section. In particular, we illustrate the performance of information criteria using models that range from stationary to nonstationary models.

3.2.1 Stationary Processes

Consider a random process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, where D is a fixed subset of $\Re^d$. Assume that the random process $Z(\cdot)$ satisfies

$$E(Z(\boldsymbol{s})) = \mu, \qquad \text{for all } \boldsymbol{s} \in D \quad \text{and} \tag{3.4}$$

$$Cov(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)) = C(\boldsymbol{s}_i - \boldsymbol{s}_j), \qquad \text{for all } \boldsymbol{s}_i, \boldsymbol{s}_j \in D. \tag{3.5}$$

That is, the mean does not depend on $\boldsymbol{s}$ and the covariance is a function only of the increment $\boldsymbol{s}_i - \boldsymbol{s}_j$. Then $Z(\cdot)$ is said to be a *second-order* or *weak stationary* process. Furthermore, if $C(\boldsymbol{s}_i - \boldsymbol{s}_j)$ is a function of $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ only, that is, the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, then $C(\cdot)$ is called *isotropic*. An isotropic process assumes that the correlation structure between sites is circular which indicates the correlation depends only on the distance between sites.

One frequently used isotropic covariance function is the exponential model. Here the covariance between measurements at two locations is an exponential function of the distance between two locations,

$$Cov(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)) = \sigma^2 \text{exp}(-\phi \|\boldsymbol{s}_i - \boldsymbol{s}_j\|) + \tau^2 I(i = j),\ \sigma^2 > 0, \phi > 0, \tau^2 > 0, \tag{3.6}$$

where $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ is the distance between sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, and I denotes the indicator function. Here σ^2 and ϕ are positive parameters called the partial sill and the decay or inverse range parameter, respectively. When $i = j$, $d_{ij} = 0$ and $C(d_{ij}) = Var(Z(s_i))$ is often expanded to $\tau^2 + \sigma^2$, where $\tau^2 > 0$ is called a nugget effect, and $\tau^2 + \sigma^2$ is called the sill.

Many other parametric models for the isotropic covariance function are also commonly used (Schabenberger and Gotway, 2005, Section 2.1). Isotropic processes are popular because a number of relatively simple parametric forms are available.

If dependence between $Z(\boldsymbol{s}_i)$ and $Z(\boldsymbol{s}_j)$ is a function of both the distance and the direction of $\boldsymbol{s}_i - \boldsymbol{s}_j$, then the process $\boldsymbol{Z}$ is called anisotropic. Hence, the covariance function, $C(\boldsymbol{s}_i - \boldsymbol{s}_j)$ is no longer purely a function of distance between two spatial locations, $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$.

Sometimes the anisotropy can be corrected by a linear transformation of the increment vector $\boldsymbol{s}_i - \boldsymbol{s}_j$. This anisotropy is known as *geometric anisotropy* and gives elliptical contours for the correlation. Specifically, the geometric anisotropy is corrected by (i) a rotation of the coordinate system to align the major and minor axes of the elliptical contours, and (ii) a compression of the major axis to make the contours spherical. Following Schabenberger and Gotway (2005, p.151), the anisotropy matrix A is thus defined as,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \mathrm{cos}\gamma & -\mathrm{sin}\gamma \\ \mathrm{sin}\gamma & \mathrm{cos}\gamma \end{bmatrix}, \tag{3.7}$$

where λ and θ are the anisotropy ratio for compression and the anisotropy angle for rotation, respectively. Here λ equals the ratio of the ranges in the directions of the major and minor axes of the elliptical contours. Geometric anisotropy is common for processes that evolve along particular directions. For example, airborne pollution emitted from an industrial plant will likely evolve along the wind directions (Schabenberger and Gotway, 2005, p.151).

In general, geometric anisotropy can be incorporated in the isotropic model by correcting distances. For instance, we can incorporate geometic anisotropy in the exponential model (3.6),

$$Cov(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)) = \sigma^2 \text{exp}(-\phi \|A(\boldsymbol{s}_i - \boldsymbol{s}_j)\|) + \tau^2 I(i = j), \tag{3.8}$$

where A is the anisotropy matrix in (3.7).

3.2.2 Nonstationary Processes

We consider a class of parametric nonstationary covariance models proposed by Hughes-Oliver et al. (1998). They incorporate nonstationarity in the covariance model driven by a point source, (e.g., the center of a wafer in semiconductor processing). Their covariance model for a point source at location $\boldsymbol{c}$ is

$$Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \text{exp}\{-\phi h_{ij} \text{exp}\{\alpha|c_i - c_j| + \beta \text{min}[c_i, c_j]\}\} + \tau^2 I(i = j), \tag{3.9}$$

where $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|, c_i = \|\boldsymbol{s}_i - \boldsymbol{c}\|$, and $c_j = \|\boldsymbol{s}_j - \boldsymbol{c}\|$. Here c_i and c_j are the distances of sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ from the point source $\boldsymbol{c}$, respectively, and α, $\beta \geq 0$. This covariance model is nonstationary because the correlation between sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ depends on the distances between sites and the point source through c_i and c_j.

The covariance model (3.9) can be thought of as a generalization of the exponential model for an isotropic process. Note that when $\alpha = \beta = 0$ in (3.9), the covariance model (3.9) reduces to the exponential model (3.6). Here (3.9) assumes that the effects of the point source are circular, that is, point source isotropy. Hence the correlation depends only on the distance between sites and on the distance between a site and the point source.

We can also incorporate point source anisotropy in the nonstationary point source isotropic model in a similar way as shown in (3.8). Schabenberger and Gotway (2005, p.423) presented point source anisotropy incorporated in the model (3.9),

$$Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \mathrm{exp}\{-\phi h^*_{ij} \mathrm{exp}\{\alpha|c^*_i - c^*_j| + \beta \mathrm{min}[c^*_i, c^*_j]\}\} + \tau^2 I(i = j), \quad (3.10)$$

where $h^*_{ij} = \|A(\boldsymbol{s}_i - \boldsymbol{s}_j)\|$, $c^*_i = \|A_c(\boldsymbol{s}_i - \boldsymbol{c})\|$, $c^*_j = \|A_c(\boldsymbol{s}_j - \boldsymbol{c})\|$ and A, A_c are the anisotropy matrices in (3.7).

3.3 A Simulation Study

In this simulation study, we evaluate and compare the performance of information criteria presented in Section 1.1. Of particular interest is how these criteria perform with different spatial covariance models. Specifically, we compare the performance of these information criteria with regard to their ability to discriminate the true model under various spatial covariance models, parameter values, and sample sizes.

3.3.1 Covariance Models

We consider following four different forms of exponential models for spatial covariance functions.

i) Σ_1: Exponential Isotropic Model,

$$\Sigma_{ij} = Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \mathrm{exp}\{-\phi h_{ij}\} + \tau^2 I(i = j),$$

ii) Σ_2: Exponential Anisotropic Model,

$$\Sigma_{ij} = Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \mathrm{exp}\{-\phi h^*_{ij}\} + \tau^2 I(i = j),$$

iii) Σ_3: Exponential Point Source Isotropic Model,

$$\Sigma_{ij} = Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \text{exp}\{-\phi h_{ij} \text{exp}\{\alpha |c_i - c_j|\}\} + \tau^2 I(i = j),$$

iv) Σ_4: Exponential Point Source Anisotropic Model,

$$\Sigma_{ij} = Cov[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = \sigma^2 \text{exp}\{-\phi h^*_{ij} \text{exp}\{\alpha |c^*_i - c^*_j|\}\} + \tau^2 I(i = j),$$

where $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, $h^*_{ij} = \|A(\boldsymbol{s}_i - \boldsymbol{s}_j)\|$, $c_i = \|\boldsymbol{s}_i - \boldsymbol{c}\|$, $c^*_i = \|A_c(\boldsymbol{s}_i - \boldsymbol{c})\|$, $c_j = \|\boldsymbol{s}_j - \boldsymbol{c}\|$, $c^*_j = \|A_c(\boldsymbol{s}_j - \boldsymbol{c})\|$, and A, A_c are anisotropy matrices in (3.7).

We use a Gaussian process with the above covariance models to enable likelihood inference. The most convenient assumption would be a multivariate normal distribution for the observed data. That is, suppose we have observations $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \cdots, Z(\boldsymbol{s}_n))'$ at known locations $\boldsymbol{s}_i$, $i = 1, \cdots, n$. We then assume that

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \Sigma(\boldsymbol{\theta})), \tag{3.11}$$

where N_n denotes the n-dimensional normal distribution, with mean $\mathbf{0}$ and covariance $(\Sigma(\boldsymbol{\theta}))$, where $(\Sigma(\boldsymbol{\theta}))$ takes one of the four forms described above.

Now we consider the following four different spatial models for our simulation studies.

i) M_1: Stationary Isotropic Model,

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \Sigma_1(\boldsymbol{\theta})), \quad \boldsymbol{\theta} = (\sigma^2, \phi, \tau^2),$$

ii) M_2: Stationary Anisotropic Model,

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \Sigma_2(\boldsymbol{\theta})), \quad \boldsymbol{\theta} = (\sigma^2, \phi, \tau^2, \lambda, \gamma),$$

Table 3.1: Number of Parameters and Parameters in Each Model

Model	p	$\boldsymbol{\theta}$
M_1	3	σ^2, ϕ, τ^2
M_2	5	$\sigma^2, \phi, \tau^2, \lambda, \gamma$
M_3	4	$\sigma^2, \phi, \tau^2, \alpha$
M_4	6	$\sigma^2, \phi, \tau^2, \gamma, \lambda, \alpha$

iii) M_3: Nonstationary Point Source Isotropic Model,

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \Sigma_3(\boldsymbol{\theta})), \quad \boldsymbol{\theta} = (\sigma^2, \phi, \tau^2, \alpha),$$

iv) M_4: Nonstationary Point Source Anisotropic Model,

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \Sigma_4(\boldsymbol{\theta})), \quad \boldsymbol{\theta} = (\sigma^2, \phi, \tau^2, \gamma, \lambda, \alpha).$$

Note that M_1, M_2, and M_3 are nested under M_4. That is, M_1, M_2, and M_3 are special cases of M_4. Specifically, M_4 reduces to M_1 when $\alpha = 0$ and $A = I$ in the covariance function Σ_4. Also, M_4 reduces to M_2 when $\alpha = 0$ and M_3 when $A = I$. Table 3.1 summarizes the number of parameters p in each model along with parameters $\boldsymbol{\theta}$ for each model.

3.3.2 Data Generation Processes

Using the method presented in Cressie (1993, Section 3.6) to simulate point-referenced data, we simulated the spatial process at n locations, $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n$, following a multivariate normal distribution with mean vector $E(\boldsymbol{Z}) = \mathbf{0}$, and covariance matrix $Cov(\boldsymbol{Z}) = \Sigma_i, \quad i = 1, \cdots, 4$, as presented in Section 3.3.1. We used the Cholesky decomposition which allows the covariance matrix, Σ_i, to be decomposed as the matrix product $\Sigma_i = L_i L_i'$, where L_i is a lower triangular $n \times n$ matrix. Then we simulated $\boldsymbol{Z}$, which satisfies the mean $\mathbf{0}$ and the covariance Σ_i through the relation $\boldsymbol{Z} = L_i \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon(\boldsymbol{s}_1), \cdots, \epsilon(\boldsymbol{s}_n))'$ and $\epsilon(\boldsymbol{s}_i)$'s are iid with a standard normal distribution. We also simulated irregularly spaced n locations, $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n$, distributed uniformly on the square $[0, 100]^2$.

Simulated data were generated under 18 different conditions created by varying four factors of interest: the true model ($M = M_1, M_2, M_3, M_4$), the true parameter value for nonstationarity ($\alpha = 5, 10$) and for anisotropy ratio ($\lambda = 5, 10$), and sample size ($n = 50, 100$). From the combinations of the true parameter values in the models, we created nine sets of data as given in Table 3.2. D_1 was generated from model M_1, D_{21}, D_{22} from model M_2 with different λ, D_{31}, D_{32} from model M_3 with different α, and $D_{41}, D_{42}, D_{43}, D_{44}$ from model M_4 with the combination of different λ and α. We assumed the point source to be located at the origin $\boldsymbol{c} = (0, 0)$ for model M_3 and M_4. These nine data sets were generated with two different sample sizes. 100 data sets were replicated for each of the nine scenarios, and hence a total of 1800 data sets were generated for our simulation study.

3.3.3 Results

First, we compared the covariance functions of nine scenarios with $n = 50$ given in Table 3.2 by computing the Frobenius distances between these nine covariance functions. Frobenius distance can be used to measure the distance between two matrices and to indicate the difference between these matrices. Suppose $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are square matrices with the same dimension, then the Frobenius

Table 3.2: True Parameter Values for Each Data Generation Process

DGP	α	γ	λ
D_1	0	0	1
D_{21}	0	$\pi/4$	5
D_{22}	0	$\pi/4$	10
D_{31}	5	0	1
D_{32}	10	0	1
D_{41}	5	$\pi/4$	5
D_{42}	5	$\pi/4$	10
D_{43}	10	$\pi/4$	5
D_{44}	10	$\pi/4$	10

Table 3.3: Frobenius Distance between Covariance Functions of Models

	Σ_1	Σ_{21}	Σ_{22}	Σ_{31}	Σ_{32}	Σ_{41}	Σ_{42}	Σ_{43}	Σ_{44}
Σ_1	0	8.350	9.745	10.972	11.032	11.120	11.132	11.128	11.133
Σ_{21}		0	2.165	4.810	4.887	4.951	4.979	4.970	4.980
Σ_{22}			0	3.407	3.487	3.413	3.448	3.437	3.450
Σ_{31}				0	0.802	1.721	1.694	1.692	1.689
Σ_{32}					0	1.767	1.741	1.707	1.705
Σ_{41}						0	0.302	0.224	0.312
Σ_{42}							0	0.119	0.076
Σ_{43}								0	0.092
Σ_{44}									0

distance between these two matrices is calculated as,

$$F(A,B) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}-b_{ij})^2}. \tag{3.12}$$

Thus the closer $F(A,B)$ is to 0, then the more these two matrices A and B are similar. Also notice that $F(A,B) = 0$ if and only if $A = B$.

Table 3.3 presents the Frobenius distances between the covariance functions of nine scenarios generated using sample size $n = 50$. Here, Σ_i represents the covariance function of D_i, $i = 1, 21, 22, 31, 32, 41, 42, 42, 44$. The Frobenius distances are small between the covariance functions generated from the same model with different parameter values, in particular, between the covariance functions of nonstationary models. Σ_1 seemed very different from all of other covariance functions. Σ_{22} was a little closer than Σ_{21} to the covariance functions of the nonstationary models. That is, the stationary anisotropic model with a large anisotropy ratio ($\lambda = 10$) seems to be closer than the stationary anisotropic model with a small anisotropy ratio ($\lambda = 5$) to the nonstationary model. The distances of the stationary covariance functions from the nonstationary point source isotropic covariance functions are very similar to those from the nonstationary point source anisotropic covariance functions. Unlike in the stationary models, whether the covariance function is isotropic or anisotropic seemed not to make much difference in the nonstationary models. The stationary anisotropic model appeared closer to the nonstationary models than to the stationary isotropic model. The distances between nonstationary point source isotropic models and nonstationary point source anisotropic models were the smallest among the distances between different models. Similar comparative features are observed between covariance functions generated using the sample of size $n = 100$.

Each of the nine scenarios was modeled using one of four different models, M_1, M_2, M_3, M_4. Each dataset was thus modeled with one correct model and three incorrect models. Models were fit using the `R` statistical software which uses the `optim` function to maximize the likelihood. AIC, AICc, and BIC were calculated for each dataset using the formula given in (3.1), (3.2), and (3.3), respectively. Table 3.4 presents the penalty used by each criterion for each model. BIC uses a penalty almost twice as large as that of AIC and AICc. The penalties used by AIC and AICc are not much

Table 3.4: Penalty Given by Each Criterion to Four Models with Two Different Sample Size

	$n = 50$				$n = 100$			
	M_1	M_2	M_3	M_4	M_1	M_2	M_3	M_4
AIC	6	10	8	12	6	10	8	12
BIC	11.736	19.560	15.648	23.472	13.816	23.026	18.421	27.631
AICc	6.522	11.364	8.889	13.953	6.250	10.638	8.421	12.903

different. Only the penalty of AIC does not depend on the sample size n.

We examined whether AIC, AICc, and BIC can correctly identify the true underlying model when various spatial models are fit to a particular data set which is generated from one of the models fitted. Results are summarized in Tables 3.5-3.8. Each table presents the percentage of times one of the four models is chosen by an information criterion. For example, the first row and second column of Table 3.4 presents the percentage of times that model M_1 is selected based on AICc when D_1 is fitted. The last row of each table represents the total percentage of times that the true model is not picked by the corresponding criteria, which can be called an 'Error '. In each table, the true model is marked by '* ' for convenience.

We expected the true model to be chosen most of the time, when the data are fitted to various models including the true model. Results from our simulation study indicated that AIC, AICc, and BIC performed well for some specific spatial models, however these criteria performed poorly as well for some other spatial models.

Table 3.5 presents results for D_1 generated from the stationary isotropic model, M_1. All criteria performed well for both $n = 50$ and $n = 100$. Especially BIC performed very well. BIC picked the true model 99% of the time for $n = 50$ and

Table 3.5: Percentage of Correct Decisions When data are generated from M_1

Model Fit	DGP: D_1					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1^*	90	99	92	91	100	92
M_2	6	0	5	5	0	4
M_3	4	1	3	4	0	4
M_4	0	0	0	0	0	0
Error	10	1	8	9	0	8

100% for $n = 100$. AIC and AICc chose the correct model 90% and 92% of the time for $n = 50$ and 91% and 92% of the time for $n = 100$. The performances of AIC and AICc were similar. Note that M_4 was never picked by all criteria. Overall BIC performed better than AIC and AICc in selecting the stationary isotropic model M_1.

Figure 3.1 shows results of each criterion obtained by fitting four models to D_1. In each plot, the X-axis represents the model fitted and the Y-axis represents the criterion obtained by fitting the models. For all criteria, the medians for M_4 were larger than those for other models in both sample sizes. This explains the reason why M_4 was never picked by all criteria as shown in Table 3.5. Also, note that the median difference between M_1 and other models were larger in BIC than in AIC and AICc. This supports the better performance of BIC than AIC and AICc as shown in Table 3.5.

Table 3.6 summarizes the results for D_{21} and D_{22}. D_{21} and D_{22} were generated from the stationary anisotropic model, M_2, with different parameter values for $\lambda = 5$ and $\lambda = 10$, respectively, and with the same parameter value $\gamma = \pi/4$. Each criterion performed similarly in D_{21} and D_{22} except that AIC and AICc picked M_3 more often

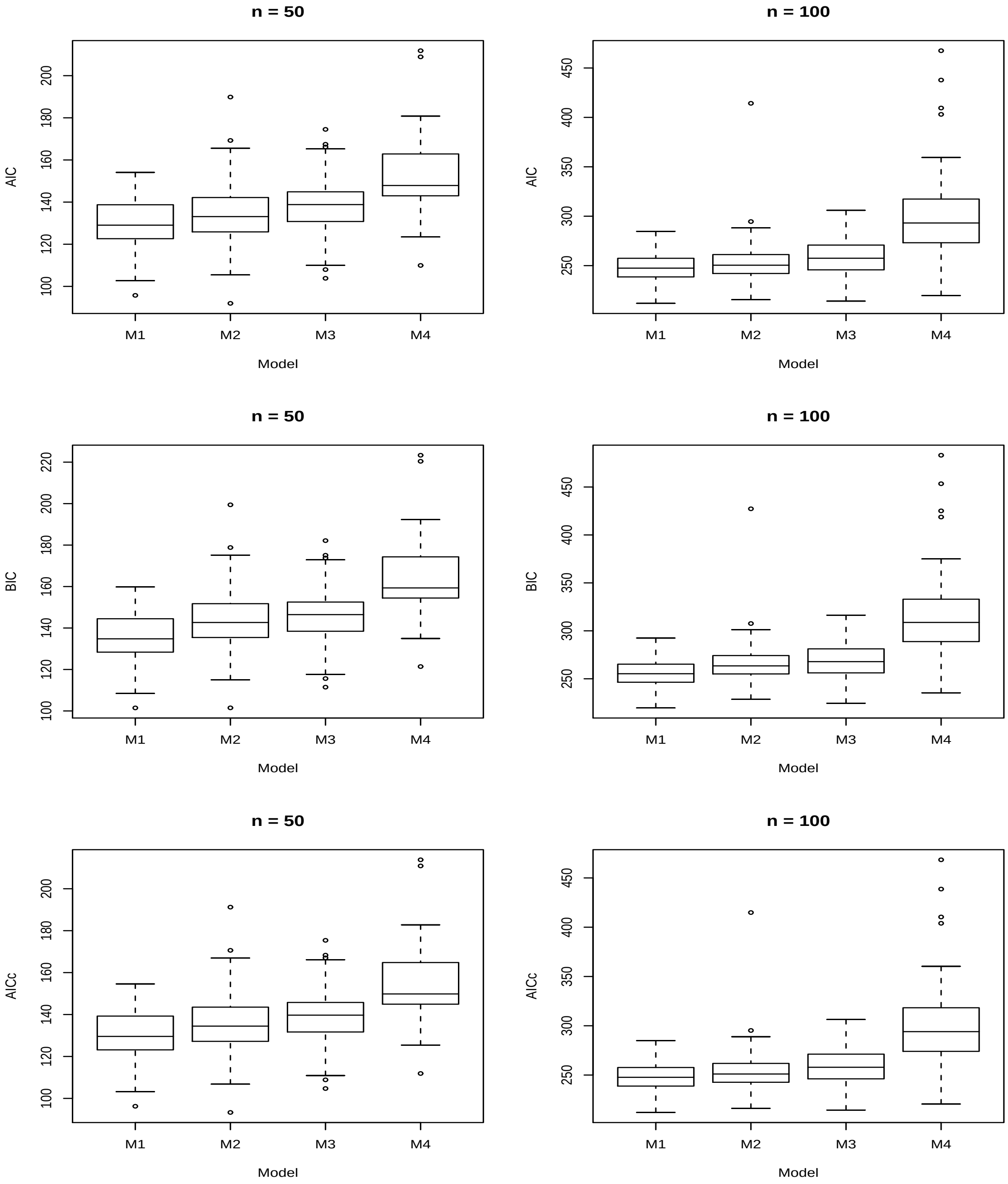

Figure 3.1: Box Plots for Each Criterion Obtained by Fitting Each Model to D_1

Table 3.6: Percentage of Correct Decisions When data are generated from M_2

Model Fit	DGP: D_{21}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	50	84	61	20	50	23
M_2^*	32	7	23	71	45	70
M_3	10	8	9	4	5	4
M_4	8	1	7	5	0	3
Error	68	93	77	29	55	30

Model Fit	DGP: D_{22}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	49	83	56	14	41	15
M_2^*	27	6	19	69	50	69
M_3	21	10	23	8	7	8
M_4	3	1	2	9	2	8
Error	73	94	81	31	50	31

in D_{22}. All criteria did not perform well when the sample size was $n = 50$. Especially BIC performed poorly. BIC picked the true model, M_2, only 7% of the time in D_{21} and 6% in D_{22}. AIC performed better than AICc for $n = 50$. This is counter to the idea that AICc is designed to perform well for small sample sizes. When $n = 50$, all criteria more often selected M_1 instead of the true model, M_2. All criteria tended to pick the parsimonious model, that is, the simpler model even though the true model is more complex. As sample size increased to 100, the performance of all the criteria improved. All criteria selected the correct model M_2 most often except BIC for D_{21}. AIC and AICc were successful in choosing the correct model and the performances of them were similar. For AIC, the success rate of picking the true model increased from 32% to 71% under D_{21} and from 27% to 69% for D_2. Also, the performances of AICc increased by 47% for D_{21} and 50% for D_{22}. While AIC and AICc performed well with the sample size $n = 100$, BIC still tended to pick parsimonious model, M_1. BIC correctly picked the true model 45% for D_{21} and 50% for D_{22}.

We found that the results from D_{21} and those from D_{22} were similar but slightly different. When $n = 50$, all criteria selected M_3 more often in D_{22} than in D_{21}, and chose M_2 less often in D_{22} than in D_{21} except BIC for $n = 100$. This occurrence makes sense because the covariance function for D_{22} was closer than the covariance function for D_{21} to the covariance functions of M_3. However, based on the Frobenius distance given in Table 3.3, it still does not make sense to choose M_1 more often than other models. When $n = 100$, all criteria also picked the correct model M_2 less often and picked M_3 and M_4 more often in D_{22} than in D_{21}. We think it is because that the covariance function of D_{22} was closer than that of D_{21} to the covariance functions of M_3 and M_4. BIC picked M_2 more often than M_1 in D_{22}. The reverse occurred in D_{21}. Overall, AIC performed better than AICc and BIC in choosing the stationary anisotropic model M_2.

Figures 3.2-3.3 support the results shown in Table 3.6. When $n = 50$, the medians of AIC and AICc for M_2 tended to be slightly larger than those for M_1, while the median of BIC for M_2 was much larger than that for M_1. This resulted in for all criteria selecting M_1 more often than M_2. As n increased to 100, the medians of AIC and AICc for M_2 tended to be much smaller than for M_1, while the median of BIC for M_2 was slightly smaller than that for M_1. This resulted that the performance of all criteria increased when $n = 100$.

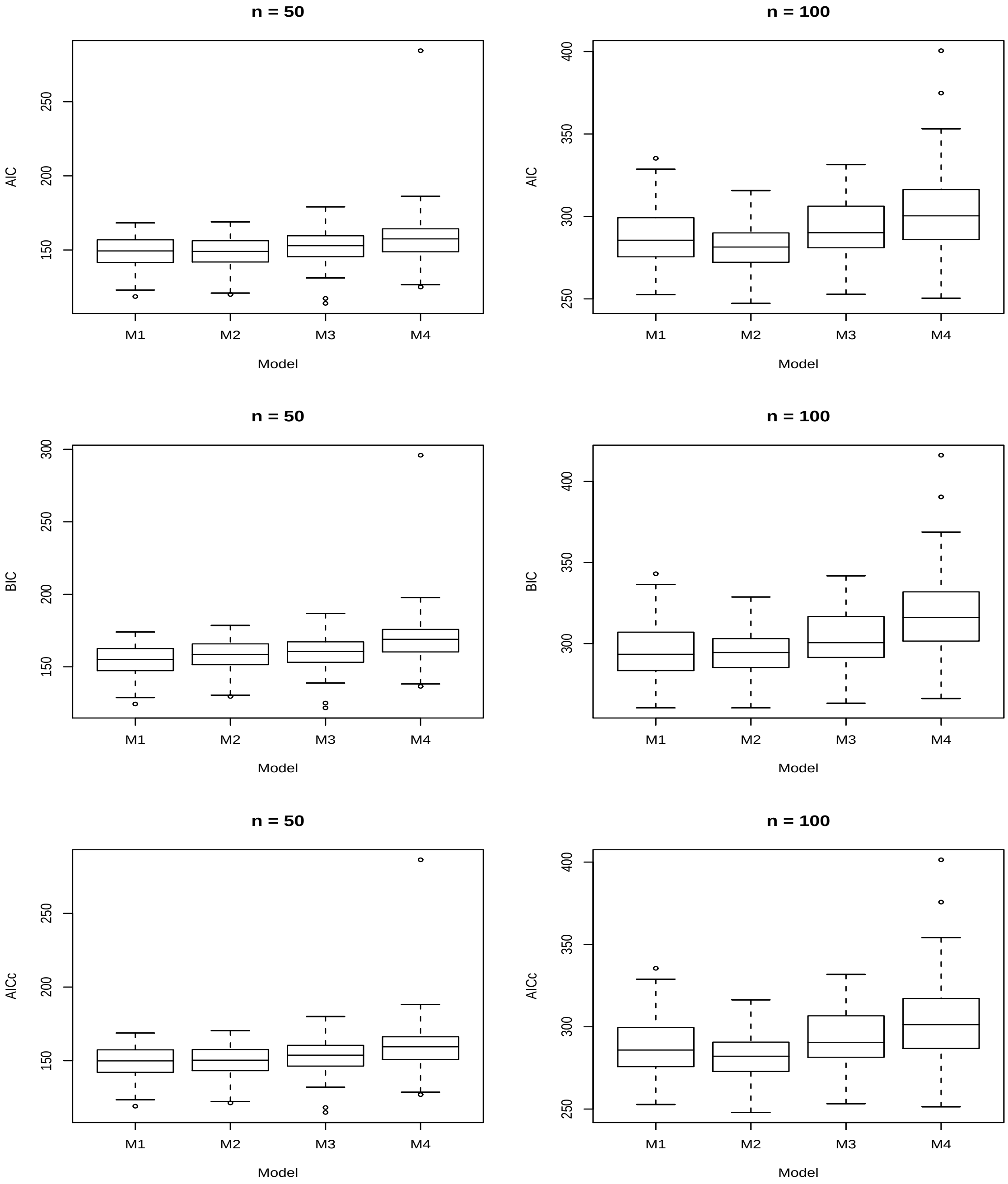

Figure 3.2: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{21}

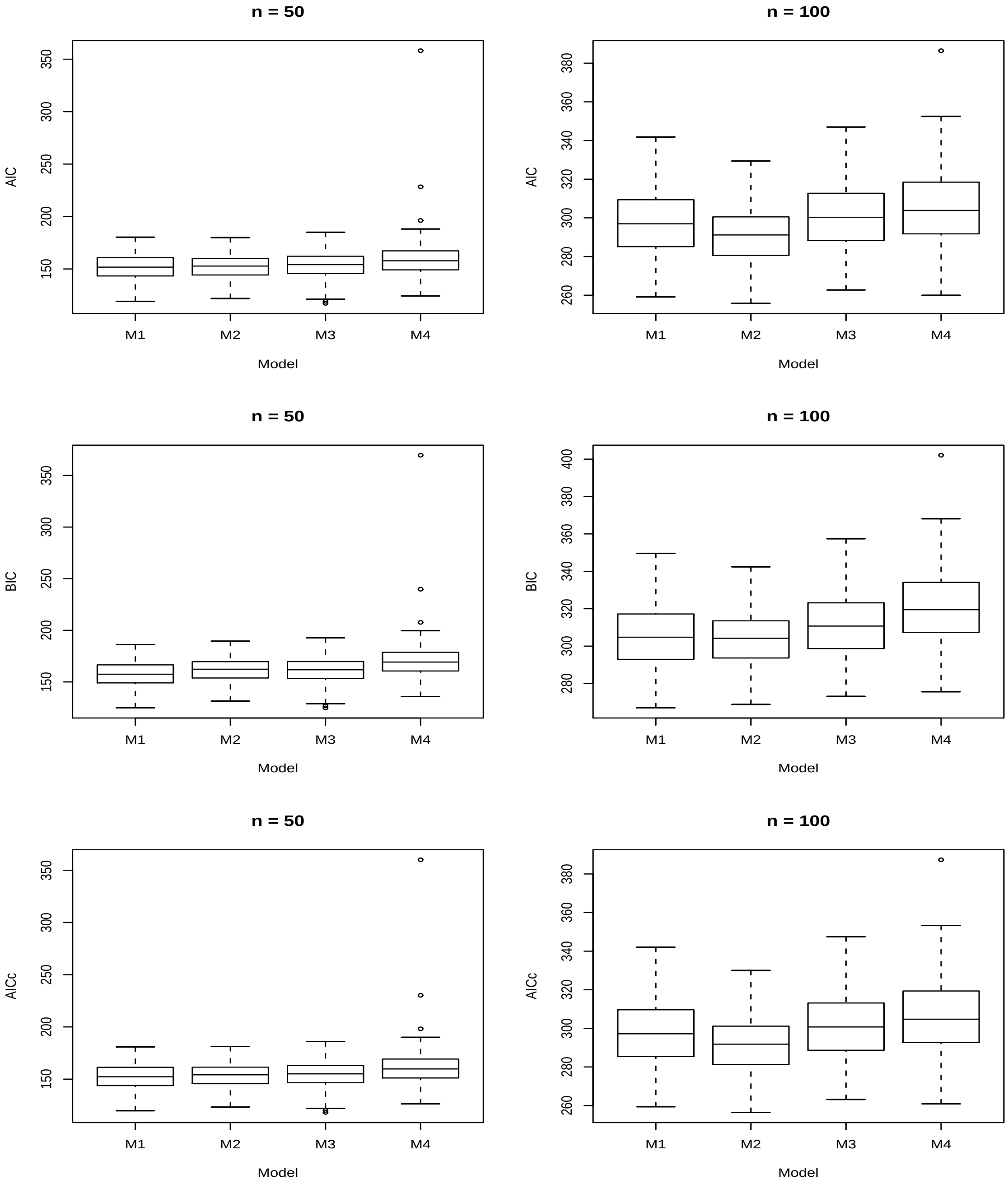

Figure 3.3: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{22}

Table 3.7: Percentage of Correct Decisions When data are generated from M_3

Model Fit	DGP: D_{31}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	24	49	28	6	17	6
M_2	3	1	2	3	1	3
M_3^*	73	50	70	91	82	91
M_4	0	0	0	0	0	0
Error	27	50	30	9	18	9

Model Fit	DGP: D_{32}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	23	61	35	9	18	9
M_2	4	0	2	4	0	4
M_3^*	73	39	63	87	82	87
M_4	0	0	0	0	0	0
Error	27	61	37	13	18	13

The results for D_{31} and D_{32} are given in Table 3.7. D_{31} and D_{32} were generated from the nonstationary point source isotropic model, M_3, with different parameter values for $\theta = 5$ and $\theta = 10$, respectively, and with the same parameter value $\lambda = 1$. Overall, all criteria performed well with the exception of BIC when the sample size was $n = 50$. For $n = 50$, BIC picked both the wrong model, M_1, and the correct model, M_3, with almost the same percentages (49% and 50%, respectively) for D_{31}, while the wrong model, M_1, was picked with a higher percentage (61%) for D_{32}. This indicated that BIC tended to select a simpler model, M_1, than the true model, M_3, when $n = 50$. In contrast, AIC and AICc identified the true model relatively well for $n = 50$. AIC selected the correct model 73% of the time for both D_{31} and D_{32}, and AICc chose the true model 70% and 63% of the time for D_{31} and D_{32}, respectively. AIC performed better than AICc for the small sample size. For $n = 100$, the performance of all criteria improved. The performances of AIC and AICc were same, and these criteria performed better than BIC. Overall, the error rates decreased with increasing sample size for both D_{31} and D_{32}. All criteria performed better in D_{31} than D_{32}. The performance of each criterion appeared to have a similar pattern under D_{31} and D_{32} except that BIC and AICc picked M_1 more often in D_{31} than in D_{32} for $n = 50$. Note that M_4 was never picked by all criteria given all the values of θ and n considered, even though the covariance function of M_4 was closer than that of M_1 and M_2 to the covariance function of M_3 in terms of the Frobenius distance in Table 3.3.

Figures 3.4-3.5 show that the median of BIC for M_3 was similar to that for M_1 in both D_{31} and D_{32} when $n = 50$. This supports the result that BIC picked M_1 often even though the true model was M_3. However, when $n = 100$, the median for M_3 was smaller than that for M_1. This resulted in the high performance of BIC for $n = 100$. We also found that the median of all criteria for M_4 was much larger than those for other models. This explains the result that M_4 was never picked by all criteria as shown in Table 3.7.

Table 3.8 illustrates the results for D_{41}, D_{42}, D_{43}, and D_{44} which were generated from the nonstationary point source anisotropic model, M_4, with different parameter values of θ and λ as shown in Table 3.2. As given in Table 3.8, the performances of the criteria did not vary much across four data sets. All criteria performed very poorly in selecting the true model, M_4, and tended to choose a simpler model than the real model. In particular, BIC did not pick the true model even one time out of 100

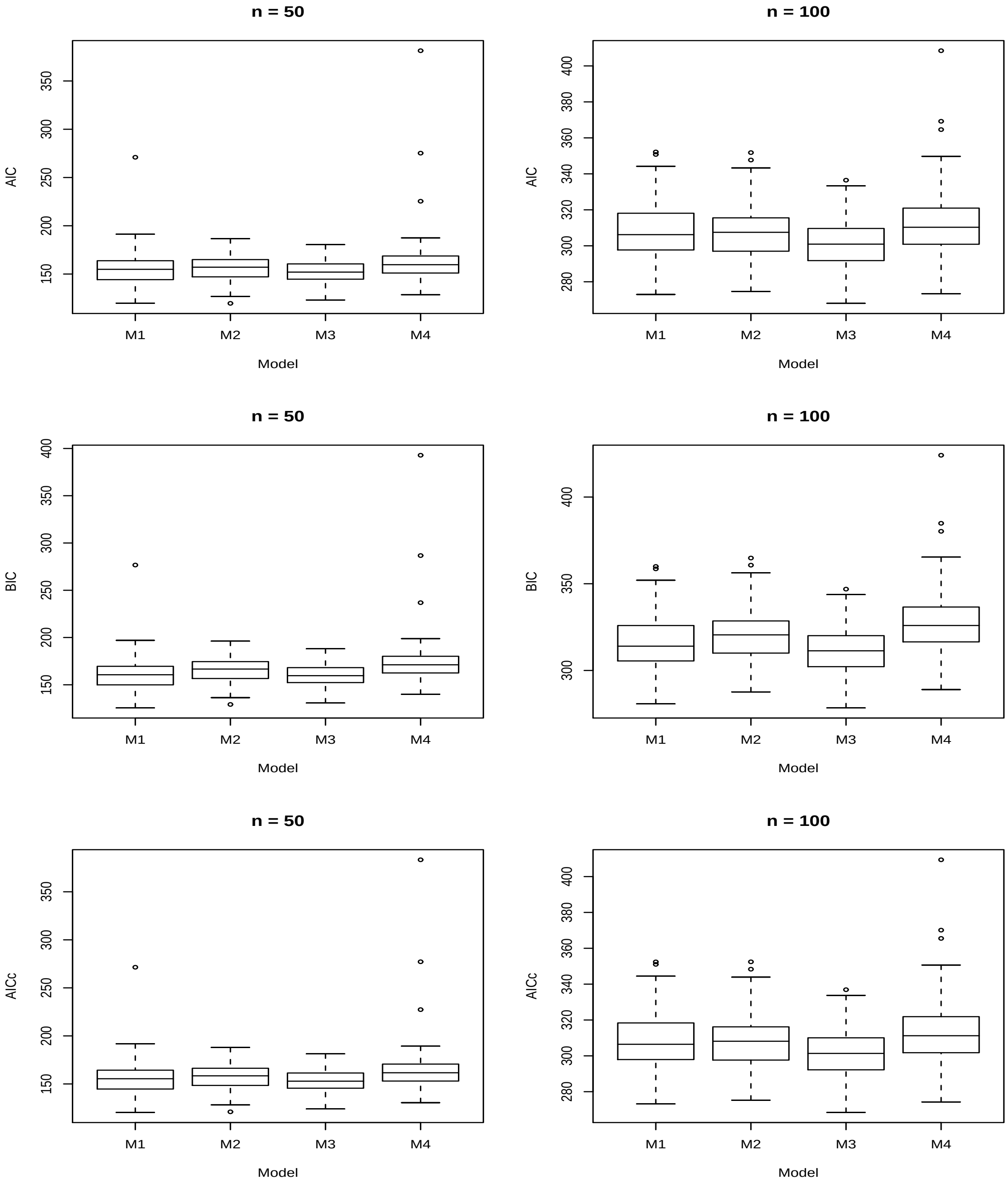

Figure 3.4: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{31}

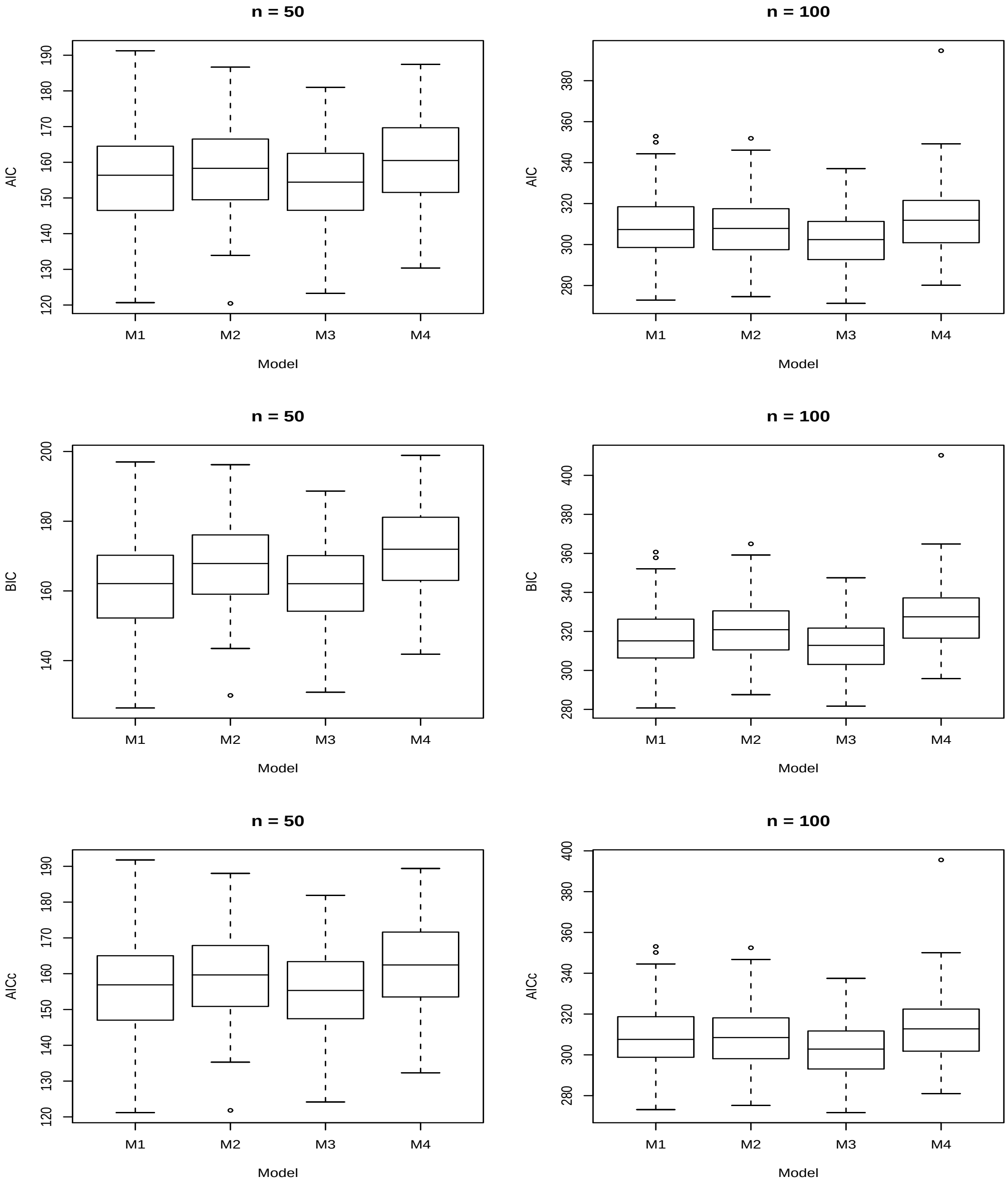

Figure 3.5: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{32}

Table 3.8: Percentage of Correct Decisions When data are generated from M_4

Model Fit	DGP: D_{41}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	44	76	50	11	47	12
M_2	6	9	4	24	4	22
M_3	46	24	43	57	49	60
M_4^*	4	0	3	8	0	6
Error	96	100	97	92	100	94

Model Fit	DGP: D_{42}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	41	72	49	9	45	9
M_2	8	0	2	21	5	19
M_3	52	28	49	56	50	60
M_4^*	2	0	0	14	0	12
Error	98	100	100	86	100	88

Model Fit	DGP: D_{43}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	34	71	41	14	46	16
M_2	8	1	5	19	4	18
M_3	58	28	54	57	50	59
M_4^*	0	0	0	10	0	7
Error	100	100	100	90	100	93

Model Fit	DGP: D_{44}					
	$n = 50$			$n = 100$		
	AIC	BIC	AICc	AIC	BIC	AICc
M_1	42	75	50	11	49	13
M_2	8	1	4	16	3	12
M_3	49	24	46	64	48	67
M_4^*	1	0	0	9	0	8
Error	99	100	100	91	100	92

replications. Even though AIC performed better than AICc and BIC, AIC did not perform well. AIC only picked the true model less than 5% of the time when $n = 50$ and less than 15% when $n = 100$ for all of four data sets. For $n = 50$, all criteria selected M_1 and M_3 most of the time. BIC picked M_1 more often (more than 70% of the time) than AIC and AICc. AICc picked M_1 more often and selected M_3 less often than AIC. For $n = 100$, AIC and AICc picked M_3 more often, and BIC picked M_1 and M_3 with similar percentage. The performances of AIC and AICc were similar. In this case, it seemed that AIC and AICc made more sense than BIC in model selection for M_4. Selecting M_1 more often than other models seems unreasonable based on the Frobenius distance, because the covariance function of M_1 was much different from that of M_4. The covariance functions of M_2 and M_3 were much closer than that of M_1 to the covariance function of M_4 as shown in Table 3.3. Overall, AIC performed better than AICc and BIC in selecting M_4.

Figures 3.6-3.9 show that the medians of all criteria for M_4 were larger than those for other models. This explains why all criteria hardly picked the true model, M_4. In particular, the median of BIC for M_4 was much larger than that for other models. This supports the poor performance of BIC as we already mentioned. For AIC and AICc, M_1 and M_3 had similar median values when $n = 50$, and M_3 had the smallest median value when $n = 100$. This supports the finding that AIC picked M_1 and M_3 with similar percentages when $n = 50$, and M_3 was picked most often when $n = 100$. Similarly, the median of BIC for M_1 was the smallest when $n = 50$, and M_1 and M_3 had similar median values for BIC when $n = 100$. This resulted in M_1 being selected by BIC most often when $n = 50$, and M_1 and M_3 were chosen by BIC with similar

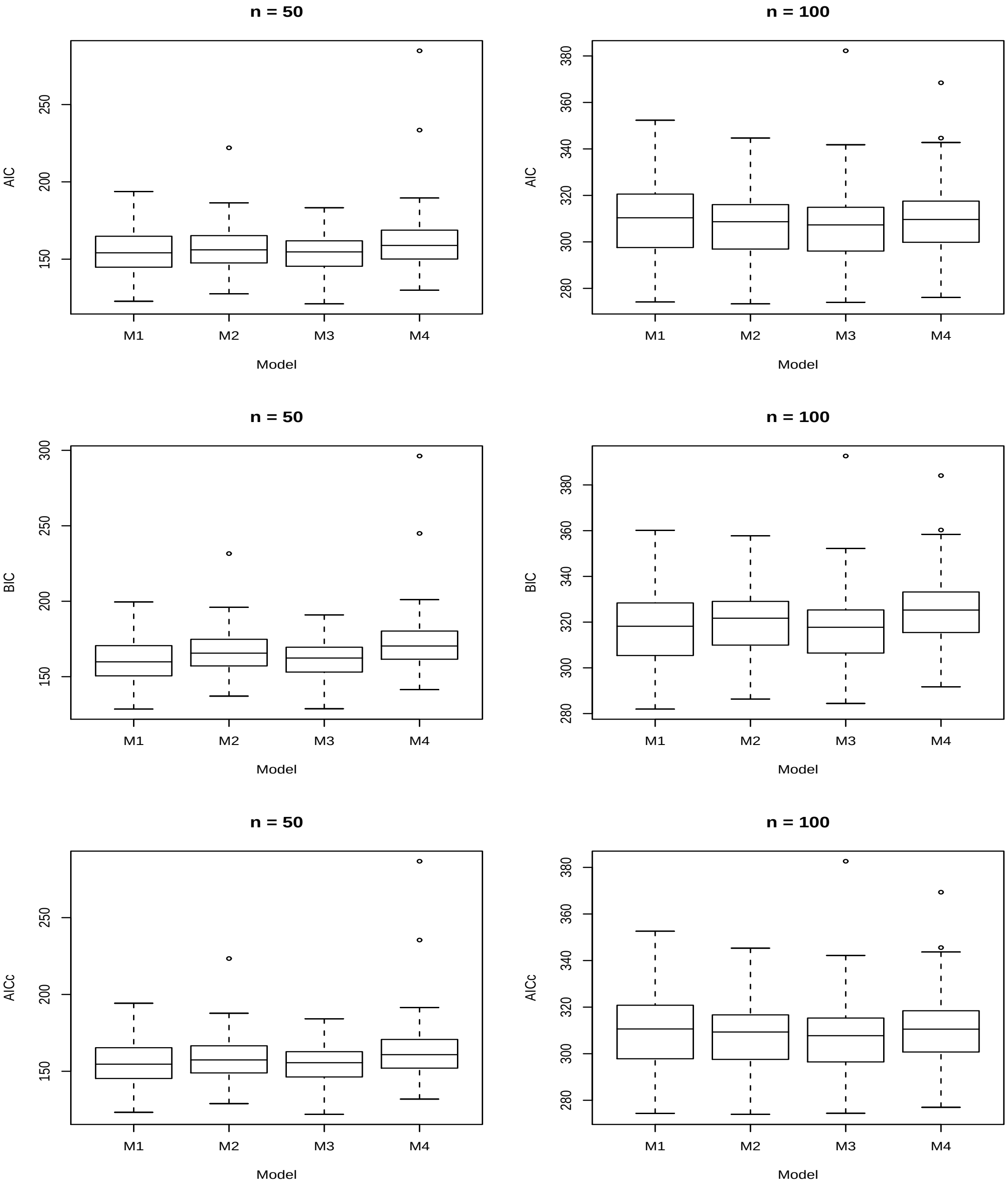

Figure 3.6: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{41}

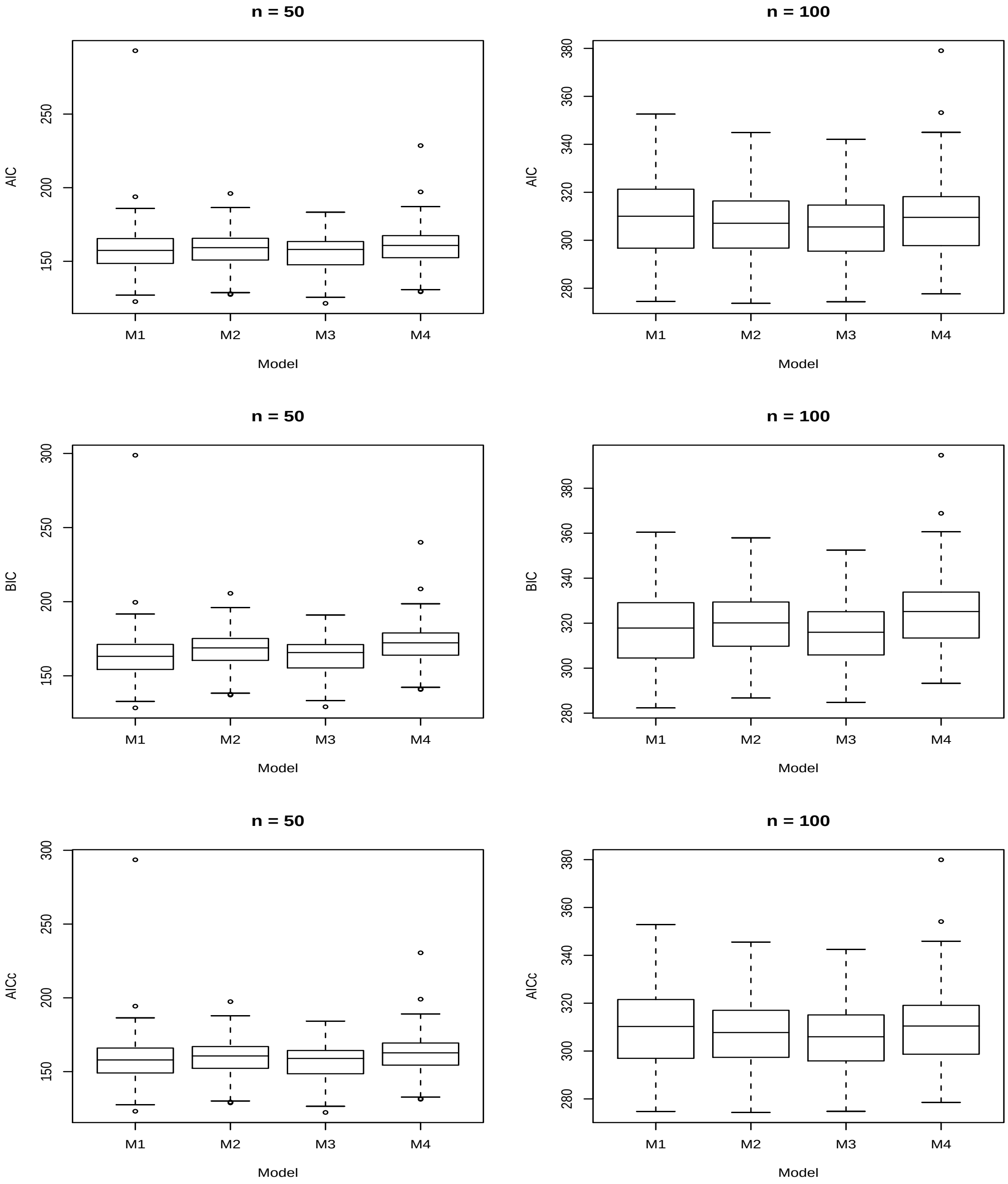

Figure 3.7: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{42}

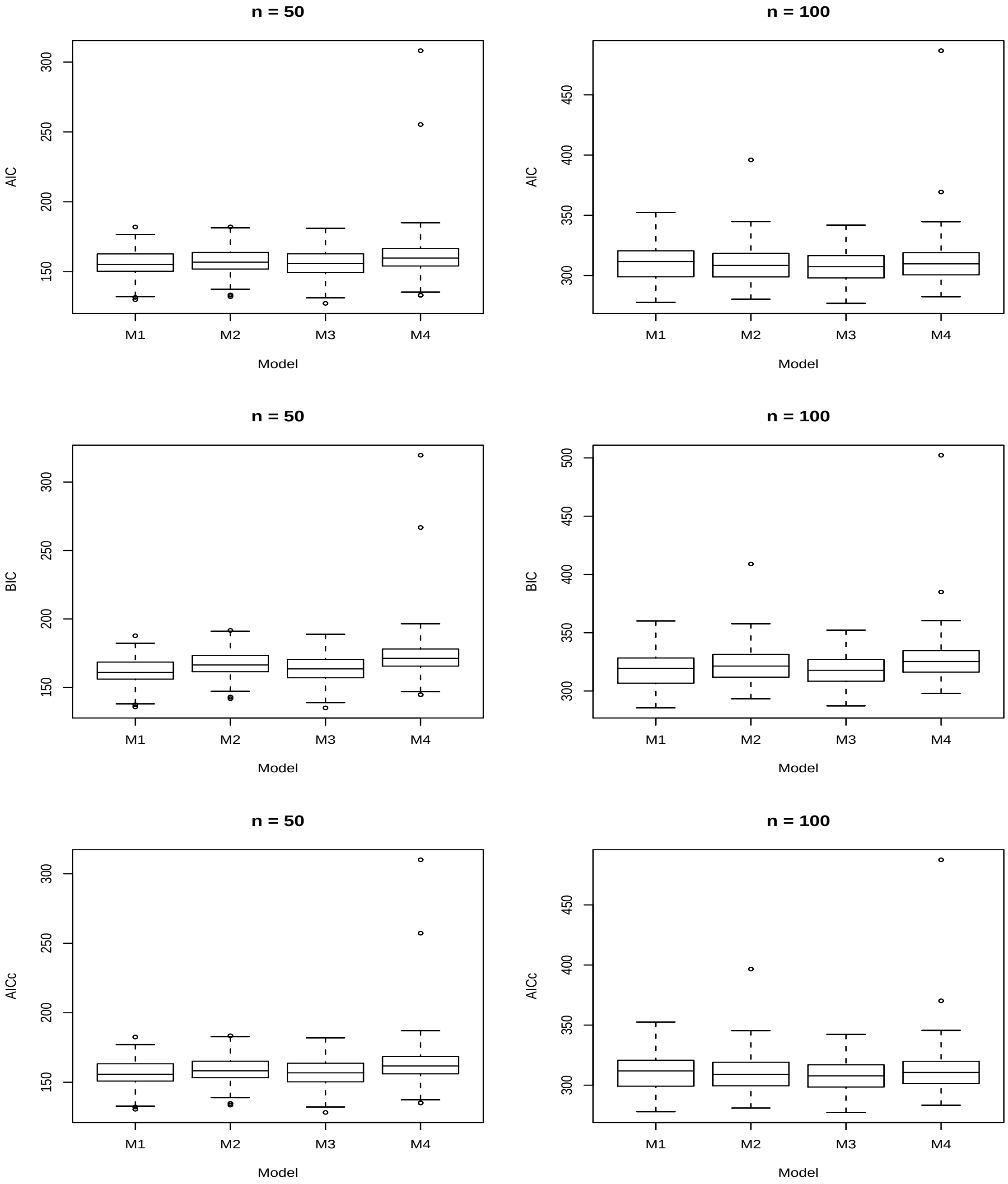

Figure 3.8: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{43}

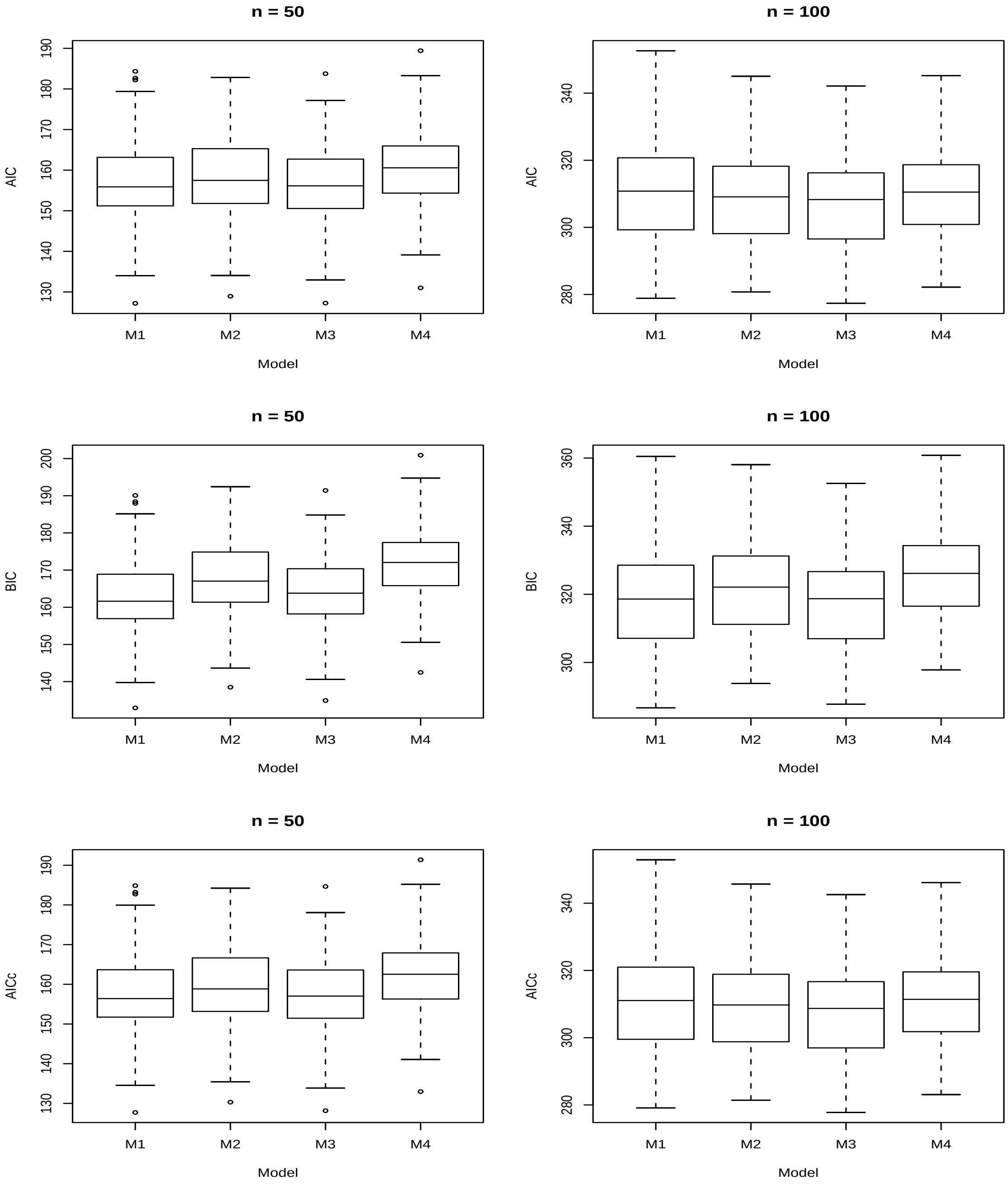

Figure 3.9: Box Plots for Each Criterion Obtained by Fitting Each Model to D_{44}

percentages when $n = 100$.

3.3.4 Findings and Future Research

We investigated how information criteria such as AIC, AICc, and BIC perform in the spatial model selection problems via simulations. The results are summarized as follows:

- BIC was superior to AIC and AICc when the true model was the stationary isotropic model. When the sample size was large (e.g., $n = 100$), BIC perfectly picked the true model. BIC also performed very well even though the sample size was small (e.g., $n = 50$). AIC and AICc also performed well.
- When the true model was the stationary anisotropic model, all criteria did not perform well for $n = 50$. Especially BIC performed poorly. As n increased to 100, the performance of all criteria improved. AIC and AICc performed well for $n = 100$, however BIC did not perform well even for the large sample size.
- AIC performed better than AICc and BIC for $n = 50$, and both AIC and AICc outperformed BIC for $n = 100$, when the true model was the nonstationary point source isotropic model. BIC picked the stationary isotropic model most often when $n = 50$, however it picked the correct model most of the time when $n = 100$. AIC and AICc performed well for both $n = 50$ and $n = 100$. The error rates for all criteria decreased as sample size increased.
- All criteria performed poorly when the true model was the nonstationary point source anisotropic model. AIC performed better than AICc and BIC. BIC never

picked the true model even when the sample size was $n = 100$. In contrast, AIC and AICc picked the true model more often when $n = 100$.

- Our results indicate that the performance of the criteria to select the true model generally improved with increase of sample size, despite differences in performance among the criteria.

- From the results obtained from simulations, we found that the performance of the criteria depends on sample size and model complexity, but not parameter values. Hence, it would be worthwhile to investigate further simulation studies with large sample sizes, e.g., $n = 500$ and $n = 1000$ and other stationary and nonstationary models.

Chapter 4

Application to Total Nitrate Concentration Data

4.1 Introduction

The release of different types of hazardous emissions has been instrumental in polluting the atmosphere over the last few decades, and hence atmospheric deposition has become a major topic of concern in environmental studies. The U. S. Environmental Protection Agency (EPA) established the Clean Air Status and Trends Network (CASTNET) to monitor air pollutant emissions and pollutant deposition (National Research Council, 2004). The U. S. EPA simulates concentrations of a variety of atmospheric pollutants using the Community Multiscale Air Quality (CMAQ) model, and then compares these simulated values with the observed values. One of the pollutants simulated most poorly is particulate nitrate (NO_3^-). Recent studies indicate that the model performance for particulate nitrate depends strongly on the model performance for total nitrate TNO_3 (particulate nitrate plus gaseous nitric acid

(HNO_3)). One of the goals of the scientists working with this model is to improve the predictive accuracy of simulated values of total nitrate, that is, to simulate total nitrate values which are "close " to the observed total nitrate values. To this end, one would like to explore the empirical relationship that may exist between the observed values of total nitrate and other observed variables. We can then use this empirical information to modify, if needed, the routines in CMAQ that simulate TNO_3 values. To estimate this empirical relationship in the observed data, we employ the RDSTM proposed in Chapter 2. We use total nitrate as the response variable and consider the following variables as explanatory or predictor variables within a regression model: sulfate (SO_4), ammonium (NH_4), ozone (O_3), temperature, relative humidity, wind speed, precipitation, solar radiation, and dew point temperature. We expect these variables to be reasonably good predictors of total nitrate. The RDSTM allows us to estimate dynamic relationships that are allowed to vary with weeks or months between total nitrate and the chemical species and meteorological variables.

4.2 The CASTNET Data

All of the data for this study were obtained from the U. S. EPA CASTNET sites. A complete description of this network can be found at: `http : //www.epa.gov/castnet`. Figure 4.1 shows the locations of the stations used in this study. We used 33 stations in the eastern U. S. These sites were selected on the basis of the extent of NO_X ($NO_2 + NO$) emissions, where NO represents nitrous oxide. Note that all the CASTNET sites are located in rural locations.

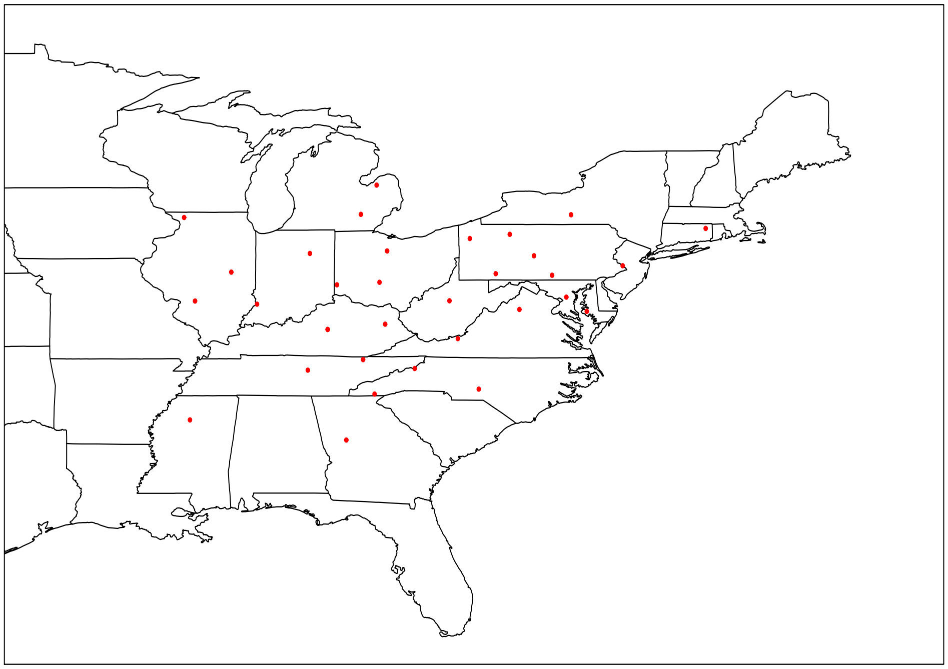

Figure 4.1: Locations of Stations

Our data were collected between January, 1997 and July, 2004, and a total of 394 weeks are available during this period. The chemical species used in this study were nitric acid (HNO_3) ($\mu mol/m^3$), nitrate (NO_3^-) ($\mu mol/m^3$), sulfate (SO_4) ($\mu mol/m^3$), ammonium (NH_4) ($\mu mol/m^3$) and ozone (O_3) (ppb). Nitric acid and nitrate were summed to get total nitrate ($\mu mol/m^3$). Residual ammonium was also used, which is calculated as ($NH_4 - 2 \times SO_4$). The factor of two is necessary as it takes two ammonium molecules to neutralize one sulfate molecule. Sulfate and ammonium usually appear in combination as $(NH_4)_2SO_4$, which is the ammonium sulfate. Sulfate and ammonium are higher in the summer and lower in the winter. Residual ammonium can equal zero or be positive or negative. When this covariate is equal to zero all the ammonium has been used to neutralize all the sulfate; there is no residual ammonium. When the value is negative, there is insufficient ammonium to neutralize the sulfate, which often occurs in the summer. When the value is positive there is more ammonium than is needed to neutralize the sulfate and this extra ammonium is available to form particulate nitrate. Positive values of residual ammonium often occur in the winter. Thus the residual ammonium covariate tells us the status of ammonium after considering its use in the neutralization of sulfate. In the statistical analysis one could use ammonium, sulfate and residual ammonium. However, since residual ammonium is a known linear combination of the other two, only one of the other two should be used. By using sulfate and residual ammonium, one is prevented from having two ammonium variables in the statistical analysis, which also makes sense thermodynamically.

Meteorological variables are observed on site at each of the CASTNET stations. In this study we used temperature (T) ($^{\circ}C$), relative humidity (RH) (%), dew point temperature ($\mathrm{T_d}$) ($^{\circ}C$), solar radiation (SR) (Wm^{-2}), wind speed (WS) (m/s) and precipitation (P) ($mm/week$). Dew point temperature is calculated from temperature and relative humidity. All of the meteorological variables were measured hourly. With the exception of ozone, which was measured hourly, the chemical species were averaged over a week from Tuesday to Tuesday. To conform to this weekly pattern, the meteorological (T, RH, $\mathrm{T_d}$, SR, and WS) variables were averaged over the same period and the daily maximum O_3 values were averaged over each week. The precipitation data were summed over the same period.

We now present some background material on the chemical processes of some of the variables used in this study, which in turn provides more insight to explain the variation of total nitrate as a function of some of the selected chemical species.

4.3 Background Chemistry for Total Nitrate

The brief discussions on the chemistry that we present here is very limited and for additional information on the chemical reactions that are presented below, see Seinfeld and Pandis (1998). During the day, HNO_3 is produced by the following chemical reaction:

$$NO_2 + OH \rightarrow HNO_3$$

The emission and subsequent oxidation of nitrous oxide (NO) is responsible for the generation of the nitrogen dioxide (NO_2) in this equation. The above equation indicates that the hydroxyl radical (OH) is also important in the formation of HNO_3. Measurements of that species are not readily available for use in the statistical model. In its place, we have used ozone, O_3, which through one path in the photolysis process, leads to OH. Thus O_3 is a good surrogate for OH. At night HNO_3 is produced by the following reactions:

$$NO_2 + O_3 \rightarrow NO_3 + O_2$$

$$NO_2 + NO_3 \rightarrow N_2O_5$$

$$N_2O_5 + H_2O(\text{on a particle surface}) \rightarrow 2HNO_3$$

In above, N_2O_5 denotes the dinitrogen pentoxide and NO_3 denotes the nitrate radical. At night O_3 is important because it reacts with NO_2 to the formation of NO_3 which is one of the key elements in the nighttime production of HNO_3. The NO_3 then reacts with NO_2 to form dinitrogen pentoxide. These first two reactions occur in the gas phase. In the third equation, N_2O_5 in the gas phase reacts with liquid water to form dissolved HNO_3. The last reaction occurs on particle surfaces that have become coated with water because of the high relative humidities. Thus we expect that O_3 and relative humidity to be important in the formulation of HNO_3 and then total nitrate TNO_3.

So far, we have described the mechanisms by which TNO_3 is produced in the atmosphere. The mechanisms of removal are also important in predicting the atmosphere concentrations of TNO_3. In general, TNO_3 may be removed from the atmosphere by wet deposition (i.e., rain out) and dry deposition. The speed at which TNO_3 is removed by dry deposition depends on whether the nitrate exists predominantly in the gas phase or the particulate phase. The dry deposition velocity of HNO_3 is significantly faster than that of particulate nitrate. The distribution of TNO_3 between gaseous HNO_3 and particulate nitrate depends mainly on three variables: temperature, relative humidity, and residual ammonium. At low temperatures and/or high relative humidities, the TNO_3 favors the particulate phase, and vice versa. At high concentrations of residual ammonium, total nitrate also favors the particulate phase. Because particulate nitrate is removed less efficiently then gaseous HNO_3, it follows that the atmospheric concentration of total nitrate will be enhanced under conditions of low temperature, high relative humidity, or high residual ammonium. This background chemistry of total nitrate is used to select appropriate predictors in addition to some exploratory data analysis, as we describe in the following section.

4.4 Exploratory Data Analysis (EDA)

We performed some exploratory data analysis (EDA) as a preliminary part of the research. The total number of weeks available varies from station to station because of missing data, but they are in fairly equal proportions. Table A.1 presents the total number of weeks available for each site. The observed values range from a high of 394 weeks (all weeks observed) to 361 weeks.

Figures A.1-A.12 summarize the variations of all the chemical species and the meteorological variables used in this study across locations, years, and months. Overall, all the variables did not seem to vary across years, however significant variations were observed in all of these variables across sites and months. This indicated that year does not seem to have a significant effect on TNO_3. In contrast, location and month appeared to be important factors to describe these variables. Among the chemical variables, HNO_3, NO_3^-, and TNO_3 presented some variations across sites. Recall that TNO_3 is the sum of HNO_3 and NO_3^-. Among the meteorological variables, the largest variation across locations was shown in wind speed. All the chemical species and most of the meteorological variables seemed to have a seasonal pattern. HNO_3, SO_4, NH_4, and O_3 seem to have higher values during summer. In contrast, NO_3^-, TNO_3 and residual ammonium appeared low in the summer. Among meteorological variables, wind speed was low while temperature, dew point temperature, and solar radiation were high during summer, as expected. Note that the pattern of temperature and dew point temperature looked very similar to each other. This is not a surprise as dew point temperature is derived analytically from temperature and

Table 4.1: Spearman Rank Correlation Coefficients between Variables

	TNO_3	SO_4	NH_4	O_3	SR	T	T_d	WS	RH	P
TNO_3	1.00									
SO_4	0.17	1.00								
NH_4	0.43	0.80	1.00							
O_3	-0.06	0.61	0.32	1.00						
SR	-0.09	0.51	0.25	0.81	1.00					
T	-0.20	0.67	0.34	0.79	0.79	1.00				
T_d	-0.23	0.68	0.37	0.70	0.69	0.97	1.00			
WS	0.42	-0.34	-0.07	-0.24	-0.20	-0.36	-0.37	1.00		
RH	-0.14	0.17	0.19	-0.22	-0.27	0.06	0.22	-0.15	1.00	
P	-0.28	-0.01	-0.13	-0.02	-0.08	0.11	0.19	-0.06	0.32	1.00

relative humidity. Precipitation did not seem to have much variation across months.

We know that NO_3^- is highest during the winter months (Warneck, 2000). It occurs as NH_4NO_3 in solid or aqueous phase. On the other hand, SO_4 reaches its highest values in the summer (Warneck, 2000).

SO_4 occurs in the solid or aqueous phase as ammonium sulfate ($(NH_4)_2SO_4$).

The data from the CASTNET sites shows that the peak in TNO_3 generally occurs in the winter of the year. NH_4 and SO_4 both show summer maximum values as does O_3. For residual NH_4, positive values generally occur in the winter when SO_4 is at its lowest, while negative values occur in the summer when SO_4 values are at their highest.

Table 4.1 summarizes Spearman rank correlation coefficient between variables in the data. Spearman rank correlation is a nonparametric measure of the association between two variables based on the rank of the observed values of the two variables. It is known to be more robust than ordinary correlation coefficient that measures only a linear relationship. The formula is

$$Corr(X,Y) = \frac{\sum(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum(R_i - \bar{R})^2 \sum(S_i - \bar{S})^2}}.$$

where R_i is the rank of the i^{th} value of variable X, S_i is the rank of the i^{th} value of variable Y, $\bar{R}$ is the mean of the R_i values, and $\bar{S}$ is the mean of the S_i values. Table 4.1 indicates high correlations between SO_4 and NH_4 (0.80), between O_3 and SR (0.87), between O_3 and T (0.79), and between SR and T (0.79). The correlations are moderate between SO_4 and O_3 (0.61), between O_3 and $\mathrm{T_d}$ (0.70), and between SR and $\mathrm{T_d}$ (0.69). T and $\mathrm{T_d}$ seem to be very highly correlated (0.97), as expected. Note that SO_4, residual NH_4, and WS are positively correlated to TNO_3, whereas O_3, SR, T, $\mathrm{T_d}$, RH, and P are negatively correlated to TNO_3.

4.5 Statistical Models

Atmospheric pollutants are known to depend on location, season, chemical species and meteorological conditions. Using flexible statistical models, we explore the relationship between the response variable TNO_3 and a set of covariates appropriately chosen from the chemical species and meteorological variables mentioned in Section 4.2. In addition we also explore the effects of location and seasonality on TNO_3. A few Linear models are considered to identify such relationships.

First, linear regression models (LRM) are used as a part of the preliminary data analysis. In this analysis, we select explanatory variables for our final analysis using the stepwise variable selection method via `SAS PROC REG`. In the stepwise method, variables are added one by one to the model, and to be added in the model a variable should produce an F statistic which is significant at a specific level, 0.15. Rawlings (1998) recommends 0.15 in the context of stepwise regression. After a variable is added, the stepwise method checks all the variables already included in the model and deletes any variable for which an F statistic is not significant at a specific level. After all the necessary deletions are completed, another variable is then added to the model. This stepwise process ends when none of the variables to be added has a significant F statistic. We then fit the linear regression model based on the covariates chosen from this stepwise variable selection method. Second, we fit the RDSTM proposed in Section 2.4 using these selected covariates. The RDSTM allows the regression coefficients for the covariates to vary with time (week). Hence, RDSTM can explain the dynamic relationship between the response variable and the covariates over time (week). This is an advantage of using RDSTM over the traditional LRM. Note that the regression coefficients are static (fixed) over time in LRM.

4.5.1 Linear Regression Models (LRM)

We use log transformed total nitrate as a response variable. Figure 4.2 suggests that the total nitrate is skewed to the right, and a logarithmic transformation would probably make the normality and homoscedastic variance assumptions plausible for a response variable in the linear regression models. We also use standardized chemical species and meteorological variables as covariates which are transformed to have the empirical mean 0 and the variance 1. In addition, longitude and latitude are considered as covariates to include location effects in the model. We then regress $\log(TNO_3)$ on the covariates adjusted for the time (year and month) effect. The linear regression

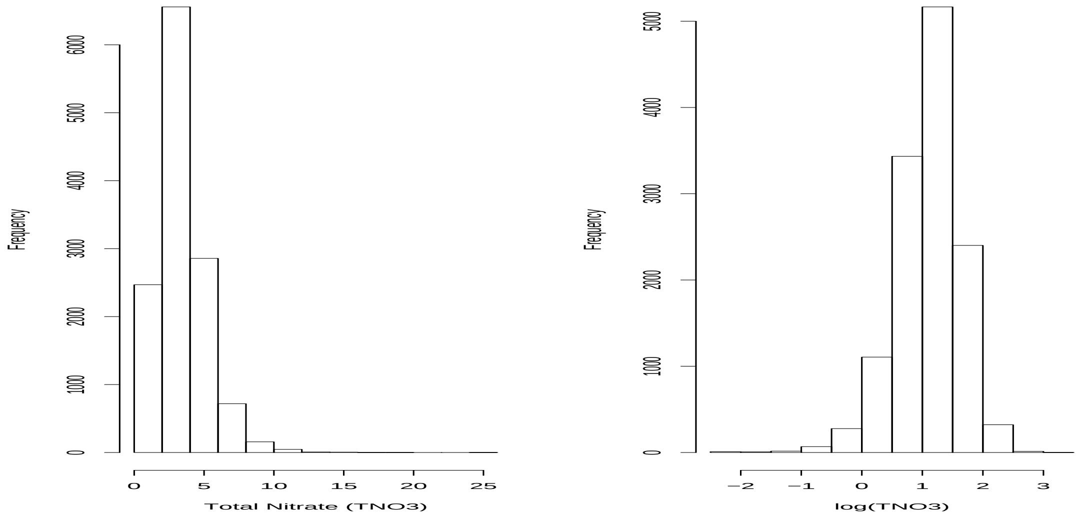

Figure 4.2: Histogram for TNO_3 and $\log(TNO_3)$

model used is of the form,

$$Y_{ijkl} = \mu_0 + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \mathbf{x}_{ijkl}^T \boldsymbol{\beta} + \varepsilon_{ijkl}, \tag{4.1}$$

where $\varepsilon_{ijkl} \overset{iid}{\sim} N(0, \sigma^2)$ for $i = 1997, \cdots, 2004$; $j = 1, \cdots, 12$; $k = 1, \cdots, 33$; $l = 1, \cdots, n_{ijk}$ (number of measurements at month j of year i at location k). Here, Y_{ijkl} denotes the $\log(TNO_3)_{ijkl}$, and α_i and γ_j are the fixed effects of year i and month j, respectively. $(\alpha\gamma)_{ij}$ is the interaction effect of year i and month j. $\mathbf{x}_{ijk}$ is a vector of the covariates measured on the k^{th} week of month j of year i, and $\boldsymbol{\beta}$ is a vector of the regression coefficients for the covariates.

4.5.2 Reparametrized Dynamic Space-Time Models (RDSTM)

We now consider the reparametrized dynamic space-time model (RDSTM) proposed in Section 2.4. Suppose Z_{it} denotes $\log(TNO_3)_{it}$ and X_{itk} the k^{th} covariate at site i and time t (week), where $i = 1, \cdots, n = 33$, $t = 1, \cdots, m = 394$, and $k = 1, \cdots, p = 7$. Notice that $\boldsymbol{Z}_t = (Z_{1t}, \cdots, Z_{nt})^T$ and $\{X_t\}_{n\times p} = ((X_{itk}))_{1\leq i\leq n, 1\leq k\leq p}$. Then following the discussion in Section 2.4, the RDSPM consists of the observation equation that can be written as,

$$\begin{aligned} Z_{it} &= \sum_{k=1}^{p} \beta_{kt} X_{itk} + \sum_{i'=1}^{i-1} \phi_{ii'} Z_{i't} + \nu_{it}, \\ Z_{1t} &= \sum_{k=1}^{p} \beta_{kt} X_{1tk} + \nu_{1t}, \end{aligned}$$

where $i = 2, \cdots, n$, $t = 1, \cdots, m$. The evolution equation can now be written as,

$$\begin{aligned} \beta_{kt} &= \sum_{k'=1}^{p} \beta_{k't-1} g_{kk'} + \sum_{k'=1}^{k-1} \psi_{kk'} \beta_{k't} + \omega_{kt}, \\ \beta_{1t} &= \sum_{k'=1}^{p} \beta_{k't-1} g_{1k'} + \omega_{1t}, \end{aligned}$$

where $k = 2, \cdots, p$, $t = 2, \cdots, m$ and initial state equation can be written as,

$$\beta_{k1} = \beta_{k0} + \sum_{k'=1}^{k-1} \psi_{kk'} \beta_{k'1} + \omega_{k1},$$

where $k = 2, \cdots, p$. The model is completed with

$$\beta_{11} = \beta_{10} + \omega_{11}.$$

Using this univariate reparametrized scheme we avoid numerical instabilities due to high dimensionality that could occur in a multivariate scheme. Also, this allows missing data to be imputed from its full conditional distribution. In addition, the RDSTM does not require simplifying assumptions like stationarity, isotropy etc. for the spatial part by allowing the $\phi_{ii'}$'s to be completely unstructured.

4.6 Results

We now present the results obtained by fitting LRM and RDSTM to our data set described in Section 4.2.

4.6.1 LRM

We considered all the chemical species and the meteorological variables described in Section 4.2 as potential covariates to be included in the final model. Among these variables, SO_4, residual NH_4, O_3, wind speed, relative humidity, and precipitation were selected to be used as covariates in the model. These variables were chosen using the stepwise variable selection method at 0.15 level of significance. `SAS PROC REG` was used to fit this model. We did not include temperature and dew point temperature even though they were also selected by the stepwise method. This is because of the fact that these variables are highly correlated with O_3 as can be seen in Table 4.1. It is well-known that highly correlated variables cause a multicollinearity problem when included together in the model. Multicollinearity exposes the redundancy of variables and the need to remove variables from the analysis. The higher the multicollinearity, the greater the difficulty in partitioning out the individual effects of the independent variables.

Table 4.2 summarizes the result of fitting the LRM. All the year, month, their interaction, location, chemical species, and meteorological covariates were significant at the 0.05 level. R^2 of this model is 0.6374. Linear regression estimates of the covariates were 0.348 (SO_4), 0.270 (residual NH_4), 0.082 (O_3), 0.105 (wind speed), -0.073

Table 4.2: ANOVA Table for LRM

Source	DF	SS	MS	F-value	p-value
Model	98	2027.075	20.684	206.74	<.0001
Error	11624	1162.981	0.100		
Total	11722	3190.056			

Variable	DF	Estimate	S.D.	t-Value	p-value
latitude	1	0.054	0.001	37.91	<.0001
longitude	1	-0.008	0.001	-10.74	<.0001
SO_4	1	0.348	0.005	74.07	<.0001
ResiNH_4	1	0.270	0.005	57.88	<.0001
O_3	1	0.081	0.007	11.81	<.0001
WS	1	0.104	0.004	28.85	<.0001
RH	1	-0.072	0.004	-18.57	<.0001
PR	1	-0.026	0.003	-8.01	<.0001

(relative humidity), and -0.026 (precipitation). While SO_4, residual NH_4, O_3, and wind speed have positive effects, relative humidity, and precipitation have negative effects on log(TNO_3). That is, log(TNO_3) increases as SO_4, residual NH_4, O_3, or wind speed increase, and as relative humidity or precipitation decrease when other effects are kept at a fixed level. Compared to other variables, SO_4 and residual NH_4 seem to have a stronger relationship with log(TNO_3). Note that O_3 was negatively correlated to log(TNO_3) and residual NH_4 was most highly correlated to log(TNO_3) in terms of the Spearman rank correlation given in Table 4.1. This shows that the relationship between TNO_3 and other variables used as covariates changed when they are considered together with other variables in the model.

We used the space-time correlogram to check whether there are any spatial and/or temporal dependence left in the residuals of the model. This correlogram provides a measure of spatial and temporal correlations by describing how data are related with

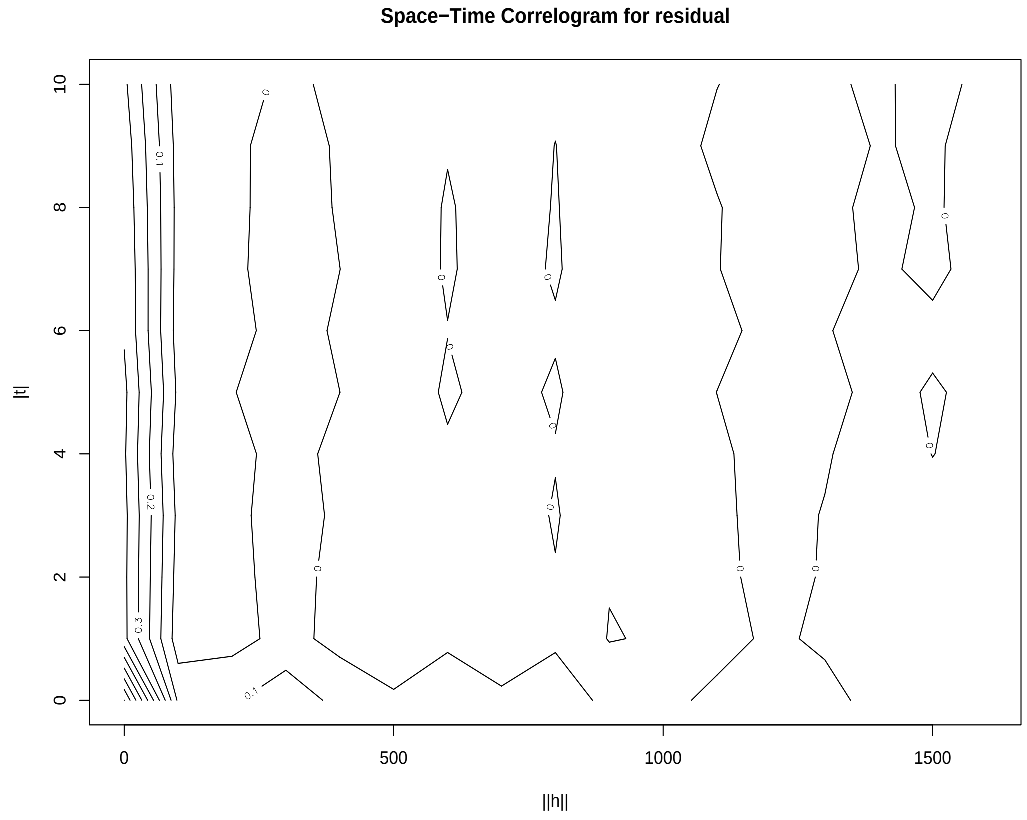

Figure 4.3: Space-Time Correlogram for the Residual of LRM

regard to distance and time lag. In general, two observations close in time or in space are likely to have similar values. Figure 4.3 displays the space-time correologram of the residual of the model (4.1). The X-axis represents distance between sites and the Y-axis represents time lag in weeks. This correlogram indicates that correlation existed in the residuals at a range of about 100 units in space and 1 unit in time. That is, residuals of observations which are closely located in space and time are autocorrelated to each other. This violates the independence assumption for errors. Hence, LRM seems not to be an adequate model for our data.

4.6.2 RDSTM

Our results from the RDSTM were obtained numerically using a Markov chain Monte Carlo (MCMC) procedure via the `WinBUGS` software. Usually a complete data set (i.e., no missing data) is required to fit the multivariate version of DSTM. However, as we already mentioned in the EDA (Section 4.4), our data have missing values. Our proposed RDSTM overcomes this limitation of a regular DSTM and performs imputations using Gibbs sampling. Gibbs sampling provides a natural solution by imputing values for the missing data at each iteration, sampling from their full conditional distribution given the available data. Regression coefficients are then updated conditional on the imputed values. We assumed that each of the standardized covariates, when missing, follows a standard normal distribution, that is, $X_{itk}^{miss} \sim N(0, 1)$.

We analyzed the data using vague priors (i.e., proper priors with large variance) on parameters to have minimal impacts on the posterior inference. We assigned independent $N(0, 10^3)$ priors to $\phi_{ii'}$, $\psi_{kk'}$ and $g_{kk'}$, and independent $G(10^3, 10^3)$ priors to $1/\sigma_{\nu i}^2$ and $1/\sigma_{\omega k}^2$. Here, we may use other priors for the parameters, however we do not expect that the results would be much different. We obtained 10,000 iterates using a single chain from the MCMC sampler. The first 5000 iterates were discarded as a part of the Markov chain burn-in period, and all the posterior summaries reported were based on Monte Carlo estimates from the remaining 5000 iterates. The number of burn-in and final MCMC sample sizes were chosen using trace plots for parameters by diagnosing for their convergence performances to stationary region. We examined trace plots of the sampled values versus iteration to look for evidence of when the simulation appears to have stabilized to a stationary distribution.

For every covariate, the RDSTM provides posterior estimates of the dynamic regression coefficient that changes with each week from January, 1997 to July, 2004, a total of 394 weeks. This is in sharp contrast to the linear regression coefficients. It is of interest to know how the covariates are dynamically related to total nitrate. Notice that a static regression coefficient might turn out to be insignificant when the significant effects in positive and negative directions might vary with time. On the other hand, the RDSTM can provide the dynamic nature of the regression coefficient (see Figures 4.4-4.9).

We first checked how many weeks in each month have a significant positive/negative effect on total nitrate. We used 95% equal-tail credible intervals to see whether the dynamic regression coefficients $\boldsymbol{\beta}_t$ are significant in the sense that these intervals do not contain the zero. To this end, we counted two separate numbers of significant weeks in each month for all the covariates. The first count provides the number of weeks which have significant positive coefficients (i.e., the lower limit of 95% interval is positive) and the other the significant negative coefficients (i.e., the upper limit is negative). The left-side plots in Figures 4.4-4.9 summarize this result. The right-side plots in Figure 4.4-4.9 present the box plots for the posterior medians of significant coefficients for each covariate in each month. Here, '* ' indicates the regression coefficient from LRM. These plots explain how strongly each covariate is related to total nitrate across months and thus provides a better interpretation of the regression coefficients in contrast to static LRM regression coefficients. We also computed the mean and the standard deviation of these posterior medians to summarize our findings numerically. These are given in Tables 4.3-4.8. Here, N_{sig} represents the number

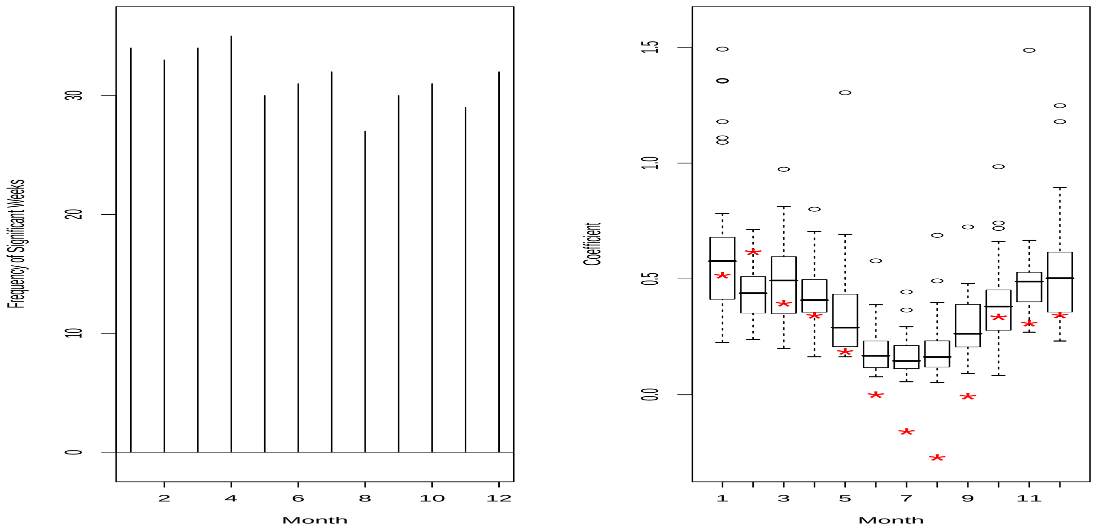

Figure 4.4: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for SO_4

of significant weeks, and N_{tot} represents the total number of weeks. Using these dynamic regression coefficients, we can find which covariate has the biggest effect on the response variable, total nitrate at what time of the year.

In the left plot of Figure 4.4, sulfate seems to have a positive relationship with total nitrate uniformly over all months. This follows from the fact that in rural areas power plants are the main source of both SO_X (SO_3 plus SO_2) and NO_X (NO plus NO_2). Also, both pollutants would build up during stagnant meteorological conditions and be diluted during periods of high winds. In addition to the counts of significant weeks, The right plot of Figure 4.4 shows that the relationship and associated uncertainties between total nitrate and sulfate that varies with months. Sulfate seems to have a stronger relationship with total nitrate along with higher uncertainty during winter months. During the winter sulfate levels are low, and hence we have high level

Table 4.3: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for SO_4

Month	N_{sig}	N_{tot}	Mean	S.E.
January	34	34	0.646	0.322
February	33	33	0.441	0.116
March	34	35	0.496	0.167
April	35	35	0.425	0.134
May	30	34	0.354	0.224
June	31	35	0.194	0.106
July	32	35	0.169	0.085
August	27	31	0.202	0.136
September	30	30	0.290	0.137
October	31	31	0.405	0.194
November	29	29	0.502	0.212
December	32	32	0.528	0.236

of residual ammonium. This results in increase of total nitrate because the residual ammonium reacts with nitrate. Table 4.3 shows the mean and the standard deviation of the posterior medians of the significant regression coefficients. The mean ranges from 0.169 (July) to 0.646 (January) and such differences are significant as evident from non-overlapping box plots (e.g., compare the box plots of June-September to the rest of the box plots).

The left plot in Figure 4.5 indicates that residual ammonium is also positively related to total nitrate with a distinct monthly pattern. Residual ammonium appears less significant in the summer (from June to September) compared to other seasons. In particular, only 9 out of 35 weeks in July are significant. Finding out patterns like this is another advantage which is achieved by using RDSTM. LRM does not provide this kind of result. Based on the right plot in Figure 4.5, similar to sulfate, residual ammonium also seems less strongly related to total nitrate during the summer

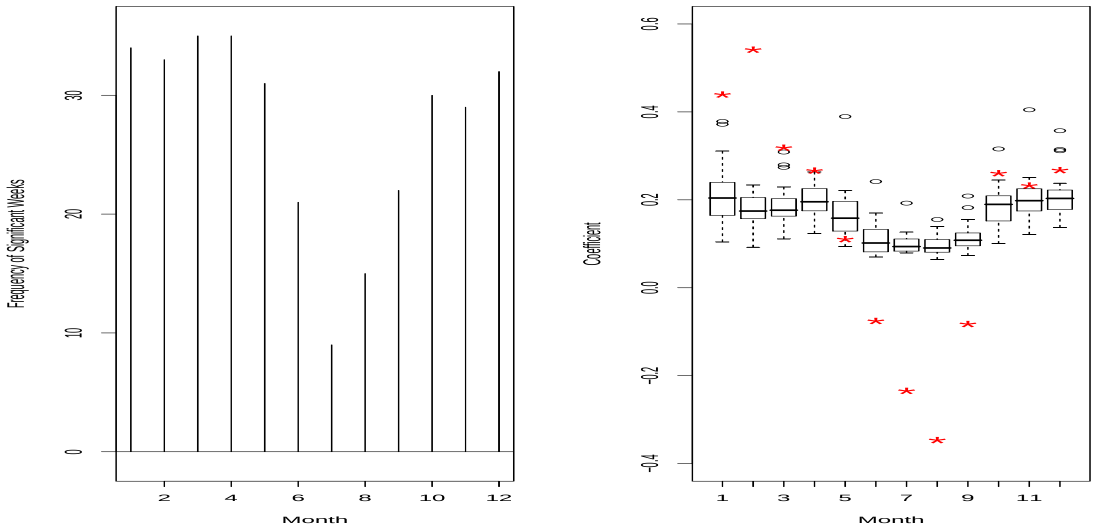

Figure 4.5: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Residual NH_4

Table 4.4: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Residual NH_4

Month	N_{sig}	N_{tot}	Mean	S.E.
January	34	34	0.213	0.063
February	33	33	0.177	0.035
March	35	35	0.186	0.041
April	35	35	0.195	0.037
May	31	34	0.165	0.056
June	21	35	0.113	0.041
July	9	35	0.107	0.036
August	15	31	0.098	0.027
September	22	30	0.117	0.033
October	30	31	0.184	0.047
November	29	29	0.204	0.049
December	32	32	0.207	0.047

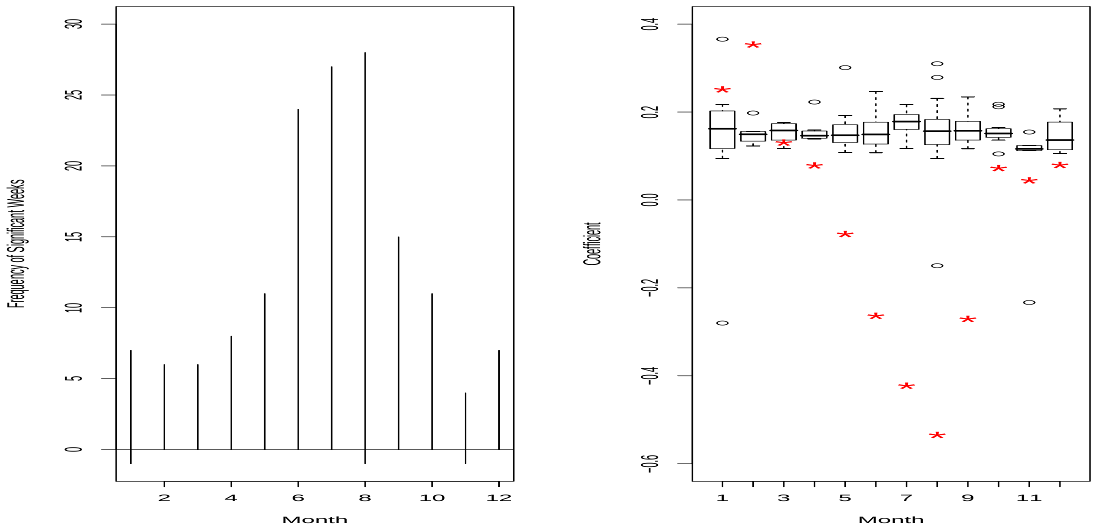

Figure 4.6: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for O_3

months from June to September than during other months. Recall that in the summer sulfate concentrations are high and thus the residual ammonium values will be near zero or even negative, while in the winter sulfate values are low and the residual ammonium values will be positive. Thus in the summer a low number of weeks in each month are significant probably because of the lack of free ammonium to produce ammonium nitrate, NH_4NO_3. In the winter the excess residual ammonium is free to produce more ammonium nitrate, which yields a stronger relationship between residual ammonium and total nitrate. Table 4.4 summarizes the posterior medians of the significant regression coefficients for residual ammonium in each month. The smallest regression coefficient is 0.098 (August) and the largest is 0.213 (January).

The left plot in Figure 4.6 reveals that June, July, and August have more significant weeks than other months for ozone. This tells us that ozone clearly has a

Table 4.5: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for O_3

Month	N_{sig}	N_{tot}	Mean	S.E.
January	8	34	0.131	0.185
February	6	33	0.151	0.026
March	6	35	0.153	0.023
April	8	35	0.156	0.028
May	11	34	0.160	0.054
June	24	35	0.155	0.036
July	27	35	0.175	0.026
August	29	31	0.155	0.078
September	15	30	0.159	0.031
October	11	31	0.158	0.032
November	5	29	0.055	0.162
December	7	32	0.147	0.040

more significant effect on total nitrate during summer time than other times, which is expected. As shown in Figure A.6, ozone displays a strong summer peak. Because of the increased strength of the incoming solar radiation during summer, the photolysis of ozone takes place during this time. The photolysis process leads to the formation of OH, which is critical to the daytime formation of total nitrate. In addition, total nitrate formation during the summer is dominated by daytime production, whereas nighttime production dominates in the winter. Since ozone is mainly related to daytime production even though it is also related to nighttime production as well, it correlates with total nitrate less frequently during the winter than during the summer. Again, this fact can not be learned by fitting LRM. The right plot in Figure 4.6 shows that the significant regression coefficients for ozone are not much different from month to month except November. Ozone seems to have a weaker relationship with total nitrate in November than in other month. Summaries of the posterior medians

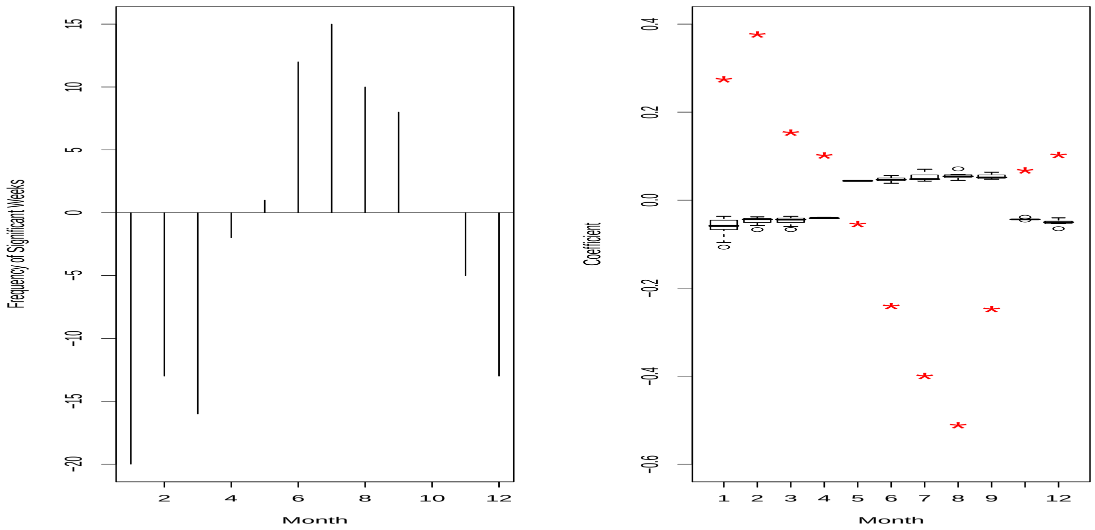

Figure 4.7: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Wind Speed

of significant regression coefficients across months are given in Table 4.5. The mean values range from 0.055 (November) to 0.175 (July).

The RDSTM results for the meteorological variables are also very interesting. Wind speed seems to have a very interesting pattern for total nitrate. Figure 4.7 indicates that wind speed has a positive effect during summer months from May to September and negative effect during other months except October. No significant week is observed in October. Also, note that not many weeks are significant in each month. The most significant month is January, where 20 out of 34 weeks have significant regression coefficients. All of the other months have a smaller number of significant weeks than January. This is an important advantage over using LRM, which does not allow us to see this type of dynamic effect. Here, the negative correlations during winter months make sense. We believe that this is a result of a dilution

Table 4.6: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Wind Speed

Month	N_{sig}	N_{tot}	Mean	S.E.
January	20	34	-0.060	0.018
February	13	33	-0.047	0.008
March	16	35	-0.046	0.009
April	2	35	-0.041	0.002
May	1	34	0.044	.
June	12	35	0.047	0.005
July	15	35	0.052	0.009
August	10	31	0.055	0.007
September	8	30	0.054	0.005
November	5	29	-0.043	0.002
December	13	32	-0.050	0.006

effect. Wind spreads the total nitrate over a large spatial area thus reducing its concentrations. Hence, it would yield lower total nitrate concentration when the wind speeds are high and higher concentrations when the wind speeds are low.

In the summer, we think that an *entrainment effect* may be important. In the summer with higher evaporation rates, the soil surface is generally drier and the aerosol on the soil surface is more easily entrained by the wind. The positive relationship that is indicated between wind speed and total nitrate would seem to support this idea. The higher wind speed would tend to entrain more aerosol from the surface into the lower atmosphere yielding higher concentrations of total nitrate, while lower wind speed would have little effect on the entrainment process thus keeping total nitrate concentrations lower. Table 4.6 shows that the magnitude of such effects seems very similar across months. The only difference over months is that wind speed is positively related to total nitrate from May to September and negatively related during other months. The mean of the significant regression coefficients varies from -0.060

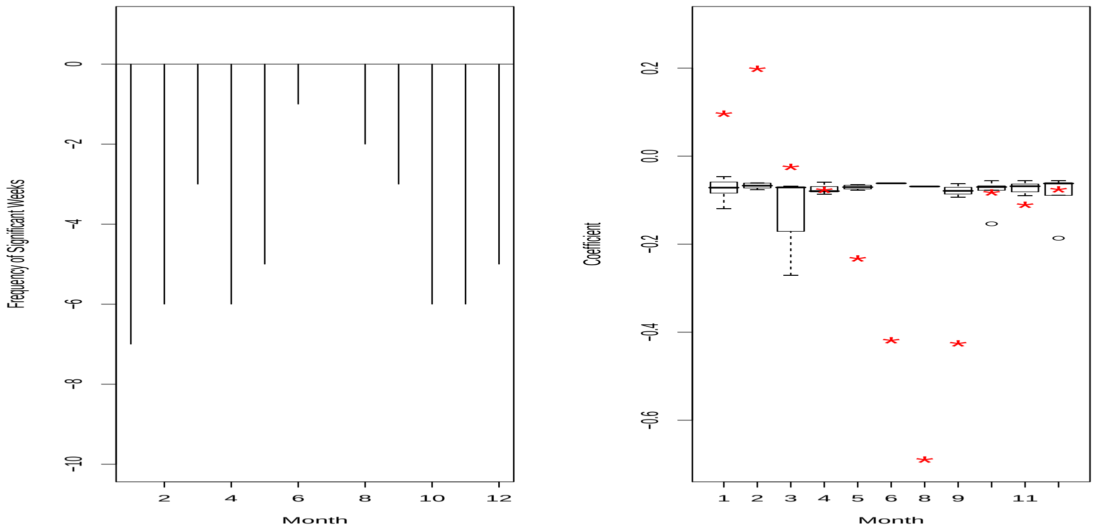

Figure 4.8: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Relative Humidity

(January) to 0.055 (August).

The pattern that we see for relative humidity is not very intuitive. Figure 4.8 demonstrates that relative humidity has negative effects on total nitrate for all months except July. No significant week is observed in July. It also shows that relative humidity is less significant during summer months than other months. However, its effect does not seem to be significant in terms of the number of significant weeks because only a few weeks are significant in each month. Specifically, only 7 out of 34 weeks are significant in January which is the most significant month for relative humidity. Again, this type of result cannot be discovered by fitting LRM. Based on the chemical processes involved, we expected to see an increase of total nitrate with relative humidity at all times of the year, however the RDSTM results point us to just the opposite effect. This is an area that will need more investigations.

Table 4.7: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Relative Humidity

Month	N_{sig}	N_{tot}	Mean	S.E.
January	7	34	-0.075	0.025
February	6	33	-0.068	0.006
March	3	35	-0.137	0.116
April	6	35	-0.076	0.010
May	5	34	-0.071	0.005
June	1	35	-0.062	.
August	2	31	-0.069	0.001
September	3	30	-0.078	0.015
October	6	31	-0.083	0.036
November	6	29	-0.071	0.012
December	5	32	-0.091	0.055

Table 4.7 indicates that the relationship between total nitrate and relative humidity is uniform across the months except March. The mean regression coefficient of March is -0.137 and it is almost twice as large as those of other months. The mean regression coefficients for the other months vary from -0.091 to -0.062. It is instructive to look at the ratio $\frac{HNO_3}{(HNO_3+NO_3^-)}$. Data presented in Warneck (2000, Figure 9.14) shows this ratio to be low in winter and high in summer. This would tend to indicate that the winter period is a time of high aerosol NO_3^- levels and lower HNO_3 levels, while the reverse is true in the summer. Our results are somewhat similar to these. This ratio is a function of relative humidity, temperature, and residual ammonium. The mean relative humidity does not change much from season to season as shown in Figure A.12. However, outside of the summer months there are more instances of low relative humidities. RDSTM results indicate that relative humidity is negatively related to total nitrate, and the relation is particularly stronger in the winter months. This negative relationship is strongest in March, but for only about a third of the weeks.

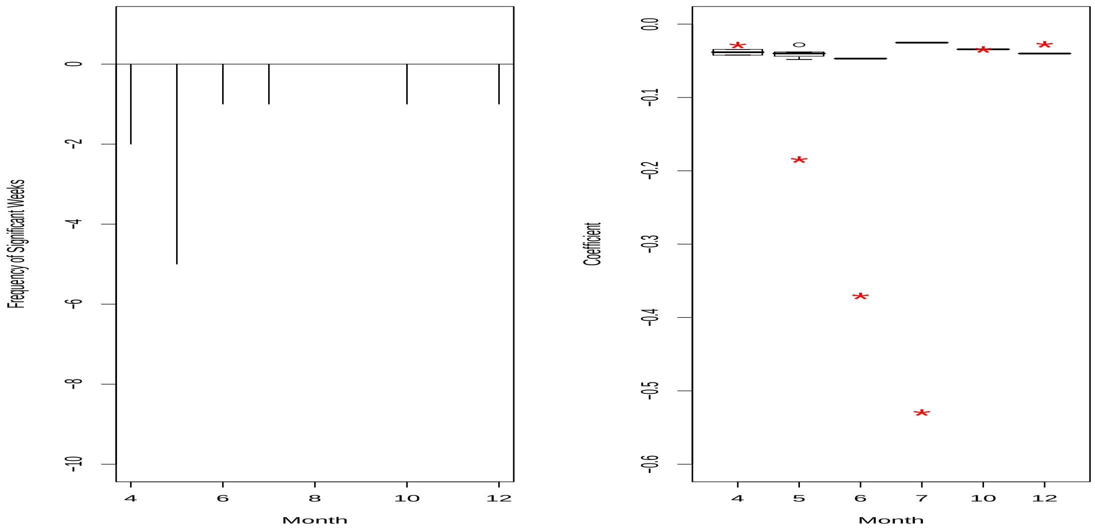

Figure 4.9: Frequencies and Box Plots for Posterior Medians of Significant Weeks in Each Month for Precipitation

In the summer, this relationship, while negative, is very weak. Currently we do not have a formal scientific justification of this peculiar behavior, and it is a focus of our future research.

As shown in Figure 4.9 precipitation has a negative effect on total nitrate. We think it occurs because of precipitation scavenging processes, and hence precipitation

Table 4.8: Mean and S.D. of Posterior Medians of Significant Regression Coefficients for Precipitation

Month	N_{sig}	N_{tot}	Mean	S.E.
April	2	35	-0.038	0.006
May	5	34	-0.040	0.007
June	1	35	-0.047	.
July	1	35	-0.025	.
October	1	31	-0.034	.
December	1	32	-0.040	.

is negatively related to total nitrate during all seasons. Precipitation acts to cleanse the atmosphere of all the pollutants. Precipitation seems the most insignificant factor on total nitrate among the covariates based on the number of significant weeks. Only 11 weeks are significant out of 394 weeks over 7 years. No significant week is observed in January, February, March, August, September, and November. Among these 11 significant weeks, 5 weeks occurred in May. Table 4.8 shows the mean of these significant coefficients which appear to be similar across months. The values range from -0.047 (June) to -0.025 (July).

Overall, by comparing the number of significant weeks and the magnitude of regression coefficients for each covariate shown in Tables 4.3-4.8, we conclude that sulfate has the strongest effect on total nitrate among the covariates used. Sulfate is significant over all months, and its effect on total nitrate differs significantly over month to month. In particular, high effects occur during winter months from November to January, and low effects occur during the summer months from June to August. Ozone has the second largest effect on total nitrate during summer month from June to September, then residual ammonium has the second largest effect during other months. Also, ozone appears more significant during summer, whereas residual ammonium appears more significant during the other months. Wind speed, relative humidity, and precipitation have relatively small impacts on total nitrate. In addition they are not as significant as other covariates, such as sulfate, residual ammonium, and ozone, in terms of the number of significant weeks and also in terms of the magnitude of regression coefficients. Notice that the magnitude of the significant regression coefficients corresponding to chemical species are almost 10-20 times than those that correspond to meteorological variables.

Finally as a part of stochastic relations, we checked to see if the covariance function obtained from the RDSTM (using the $\phi_{ii'}$'s and $\sigma_i^{\nu 2}$) is stationary in nature. To this end, we computed the correlogram to examine spatial dependence. This correlogram provides a measure of spatial autocorrelation across distances. In the spatial analysis, generally, correlations at short distances (under 100km) between locations are important to identify the characteristics (e.g., stationary, nonstationary, isotropic, or anisotropic) of the covariance function. The stations observed in our data are very sparse as shown in Figure 4.1, and unfortunately only a few observations are available at short distances. Hence our data might not be good for the spatial analysis. The correlogram computed based on our data is plotted in Fig 4.10. Here, the X-axis represents the distance in kilometers between two locations, and the Y-axis represents the correlation between two data values observed at two different locations. We also computed three different directional correlograms to see if the process appears to be isotropic. The first correlogram plot shows that the correlation has a low median value (0.1) at a distance 200km, and it decreases to zero as the distance between two locations increases. The underlying process seems to be stationary because the correlations computed at fixed distance do not appear to be significantly different. That is, the correlations appear to depend only on distances between two locations. Also the underlying process seems isotropic because three different directional correlograms look similar, which means the correlations are not changing significantly with directions. So we considered a variety of the stationary isotropic processes for the covariance model of the RDSTM. Among them we found that the exponential

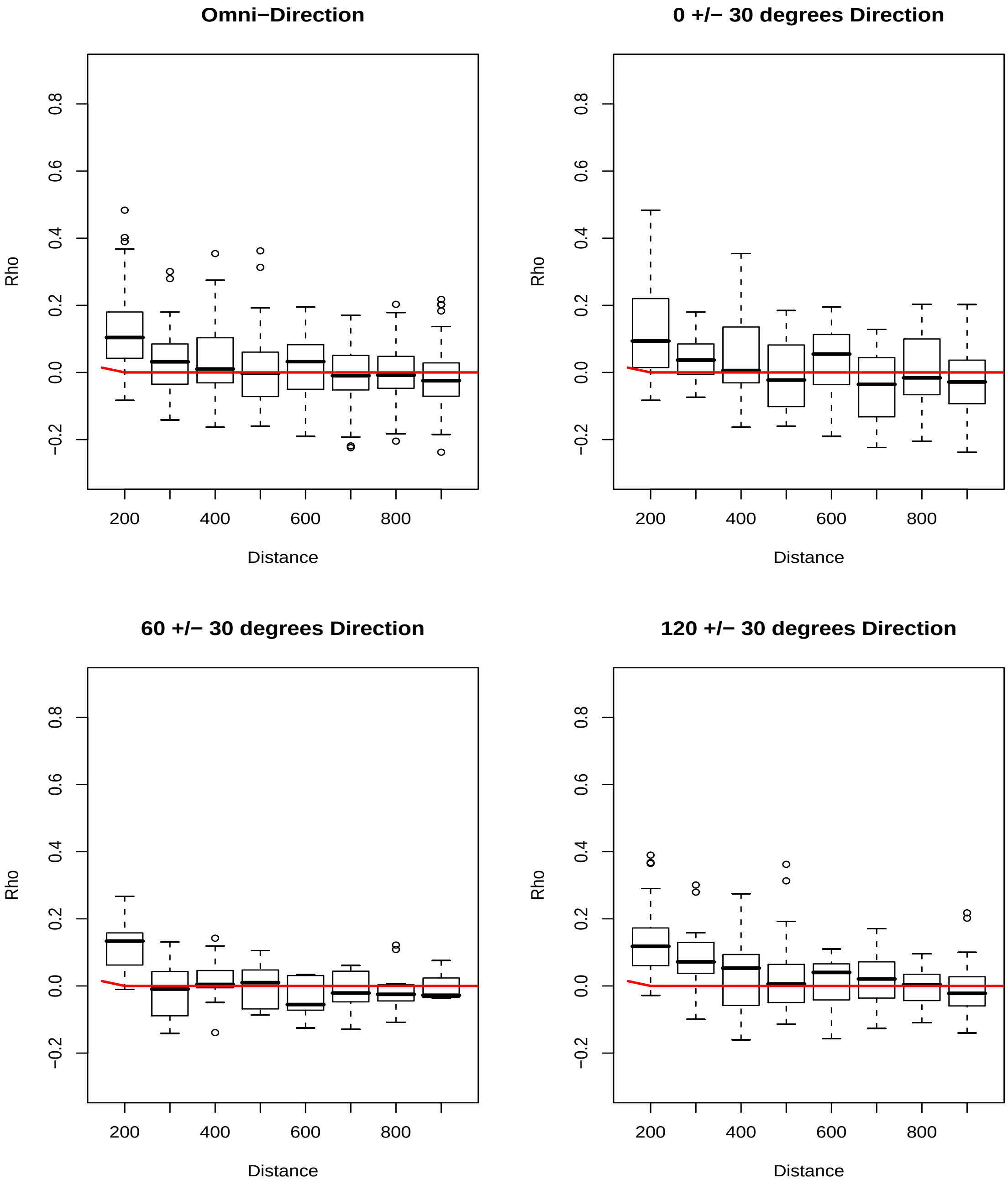

Figure 4.10: Spatial Correlogram and Fitted Exponential Correlation Model

covariance model with parameter values of $\tau^2 \approx 0.029$, $\sigma^2 \approx 0.008$, $r \approx 20$ has the smallest distance (0.031) (in terms of Frobenius distance) from the estimated covariance matrix of RDSTM. Hence the exponential covariance model seems appropriate to fit the covariance function of RDSTM. The fitted exponential correlation model (solid line) is shown along with the correlogram in Fig 4.10 .

4.7 Findings and Future Research

We used LRM and RDSTM to explain the relationship between total nitrate, other chemical species (sulfate, residual ammonium, and ozone), and meteorological variables (wind speed, relative humidity, and precipitation). From these analyses, we have learned quite a few interesting facts; these are summarized as below:

- Among the covariates we considered, sulfate is most strongly related to total nitrate both in LRM and in RDSTM. The relationship of sulfate to total nitrate is weaker during the summer than during other months.

- We found that residual ammonium has a stronger relationship with total nitrate than ozone in LRM. However, RDSTM results indicate that ozone is more strongly related to total nitrate than residual ammonium during summer and less strongly related to total nitrate during winter. Clearly ozone plays an important role in the formation of total nitrate during summer, whereas residual ammonium has a stronger role during winter, which makes sense chemically.

- RDSTM suggests that wind speed is positively related to total nitrate during summer months, and negatively related to total nitrate during other months. We think wind appears to act as an entrainment agent in the summer, and thus increases total nitrate. Wind speed is more highly related to total nitrate than ozone in LRM, however RDSTM results show the reverse.

- RDSTM results seem to indicate that relative humidity is not that significant as compared to the other chemical species covariates described above in the sense that they have a lesser number of significant weeks. However, in the LRM, relative humidity seems more significant than ozone based on the t-value given in Table 4.2. We expected a positive relationship between relative humidity and total nitrate, but RDSTM suggests a counter intuitive result. We do not have much scientific insights for such a counter intuitive result, and hence further research is needed to come up with a scientific explanation.

- Precipitation seems to have an insignificant effect on total nitrate in RDSTM. Both LRM and RDSTM indicate a very weak negative relationship with total nitrate. This relationship makes sense because precipitation acts to cleanse the atmosphere of total nitrate.

- The residuals of LRM appear to be spatially and temporally autocorrelated. RDSTM takes care of such spatial and temporal correlations in the model through the dynamic regression coefficients. For this data it appears that the covariance function of RDSTM can be modeled using a stationary isotropic process, and that the exponential covariance model is most appropriate among various stationary isotropic covariance models.

- The results from RDSTM could be used to diagnose the problems that CMAQ has with the simulations of total nitrate. As a first step in this effort, we have applied RDSTM to atmospheric measurements and obtained a dynamic relationship between total nitrate and the covariates over time. In the future, RDSTM may be applied to obtain predictive values of TNO_3 based on atmospheric data at grid locations used in CMAQ. By using model comparison methods (Fuentes and Raftery, 2005) to compare the predictions with CMAQ model values, if needed, improvements can be done to CMAQ in order to obtain more realistic predictions.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (eds.), *Proceedings of the Second International Symposium on Information Theory.* Budapest: Akademiai Kiado, pp.267-281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC*, **19**, 716-723

Allen, D. M. (1970). Mean square error of prediction as a criterion for selecting variables. *Technometrics*, **13**, 469-475.

Banerjee, S., Carlin, B. P., and Gelfand, A. E.(2004). *Hierarchical modeling and analysis for spatial data*, Chapman & Hall/CRC.

Banerjee, S., Gamerman, D., and Gelfand, A. E.(2005). Spatial process modelling for univariate and multivariate dynamic spatial data, *Environmetrics*, **16**, 465-479.

Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, revised ed.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*, Englewood Cliffs: Prentice Hall, 3rd ed.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, **52**, 345-370.

Brown, P. E., Karesen, K. F., Roberts, G. O., and Tonellato S. (2000). Blur-generated non-separable space-time models, *Journal of the Royal Statistical Society, Series B*, **62**, 847-860.

Burnham, K. P. and Anderson D. R. (2002). *Model selection and inference: A practical information-theoretic approach*, Springer: New York, 2nd ed.

Carroll, R., Chen, R., George, E., Li, T., Newton, H., Schmiediche, H., and Wang, N. (1997). Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association*, **92**, 392-415.

Cressie, N. A. C. (1993). *Statistics for spatial data.* Wiley: New York, revised edition.

Cressie, N. A. C. and Huang, H.-C. (1999). Classes of nonseparable, spatiotemporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330-1340.

Daniels, M. J. and Pourahmadi M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, **89**, 553-566.

Draper, N. R. and Smith, H. (1981). *Applied regression analysis.* Second edition. John Wiley and Sons, New York, NY.

Fuentes, M. and Raftery, A.E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics*, **66**, 36-45.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach, textitBiometrika, **95**, 1-11.

Gelfand, A. E., Ghosh, S. K., Knight, J. R., and Sirmans, C. F. (1998). Spatio-Temporal Modeling of Residential Sales Data. *Journal of Business & Economic Statistics*, **16**, 312-321.

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

Handcock, M. and Wallis, J. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, **89**, 368-390.

Higdon, D. (1999). A process convolution approach to modeling temperatures in the north Atlantic Ocean. *Environmental and Ecological Statistics*, **5**, 173-190.

Hoeting, J. A., Davis, R. A., Merton, A. A., and Thomspon, S. E. (2006). Model Selection for Geostatistical Models, *Ecological Applications*, **16:1**, 87-98.

Huang, H.-C. and Cressie, N. A. C. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, **22**, 159-175.

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

Huerta, G., Sanso, B., and Stroud, J. R. (2004). A spatio-temporal model for Mexico city ozone levels. *Journal of the Royal Statistical Society*, Series C, **53**, 231-248.

Hughes-Oliver, J. M., Lu, J.-C., Davis, J. C., and Gyurcsik, R. S. (1998) Achieving uniformity in a semiconductor fabrication process using spatial modeling. *Journal of the American Statistical Association*, **93**, 36-45.

Hurvich, C. M. and Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297-307.

Hurvich, C. M. and Tsai, C-L. (1990). The impact of model selection on inference in linear regression. *American Statistician*, **44**, 214-217.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22(1)**,79-86.

Kyriakidis, P. C. and Journel, A. G. (1999). Geostatistical space-time models: a review. *Mathematical Geology*, **31**(6), 651-684.

Lebreton,J-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monograph*, **62**, 67-118.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, **12**, 591-612.

Mardia, K. V., Goodall, C., Redfern, E. J., and Alonso, F. J. (1998). The kriged Kalman filter (with discussion). *Test*, **7**, 217-285.

Matern, B. (1986). *Spatial variation.* Berlin: Springer-Verlag.

National Research Council (2004). *Air Quality Management in the United States*, The National Academic Press.

Pole, A., West M. and Harrison, P. J. (1994). *Applied Bayesian forecasting and times series analysis*, Chapman and Hall: New York.

Pourahmadi M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, **86**, 677-690.

Rawlings, J. O., Pantula, S. G., and Dickey, D. A. (1998). *Applied regression analysis: a research tool.* Springer: New York, 2nd ed.

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. (1999). Venezuelan rainfall data analyzed using a Bayesian space-time model. *Applied Statistics*, **48**, 345-362.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*, CRC Press.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society*, Series B, **65**, 745-758.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

Seinfeld, J. H. and Pandis, S. N. (1998). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley: New York.

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society*, Series C, **51**, 351-372.

Spiegelhalter, D. J., Best, N., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society*, Series B, **64**, 583-639.

Stein, M. L. (2003). Space-time covariance functions. *Journal of the American Statistical Association*, **100**, 310-321.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society*, Series B, **39**, 111-147.

Stroud, J. R., Müller, P., and Sanso, B. (2001). Dynamic models for spatiotemporal data. *Journal of Royal Statistical Society, Series B*, **63**, 673-689.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods*, **A7**, 13-26.

Takeuchi, K. (1976). Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematic Sciences)*, **153**, 12-18. (In Japanese).

Tonellato, S. (1997). Bayesian dynamic linear models for spatial time series, Technical report (Rapporto di riceria 5/1997), Dipartimento di Statistica, Universita CaFoscari di Venezia, Venice, Italy.

Waller, L., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607-617.

Warneck, P. (2000). *Chemistry of the natural atmosphere.* New York: Academic Press.

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

Wikle, C., Berliner, M., and Cressie, N. (1999). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, **5**, 117-154.

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

Xu, K. and Wikle, C. (2005). Estimation of Parameterized Spatio-Temporal Dynamic Models. *Ecological and Environmental Statistics*, to appear.

Appendix A

Tables and Plots

A.1 Table and Plots for EDA in Chapter 4

Table A.1: Total Number of Weeks N for Each Site

ID	Siteid	N	Percent
1	ABT147	392	3.04
2	ALH157	394	3.06
3	ANA115	387	3.00
4	ARE128	394	3.06
5	BEL116	366	2.84
6	BVL130	393	3.05
7	BWR139	361	2.80
8	CDR119	391	3.03
9	CKT136	394	3.06
10	CND125	393	3.05
11	COW137	391	3.03
12	CTH110	394	3.06
13	CVL151	394	3.06
14	DCP114	394	3.06
15	ESP127	394	3.06
16	GAS153	388	3.01
17	KEF112	393	3.05
18	LRL117	393	3.05
19	LYK123	392	3.04
20	MCK131	393	3.05
21	MCK231	394	3.06
22	MKG113	394	3.06
23	OXF122	394	3.06
24	PNF126	393	3.05
25	PSU106	393	3.05
26	SAL133	390	3.02
27	SHN418	390	3.02
28	SPD111	394	3.06
29	STK138	392	3.04
30	UVL124	394	3.06
31	VIN140	394	3.06
32	VPI120	388	3.01
33	WSP144	394	3.06

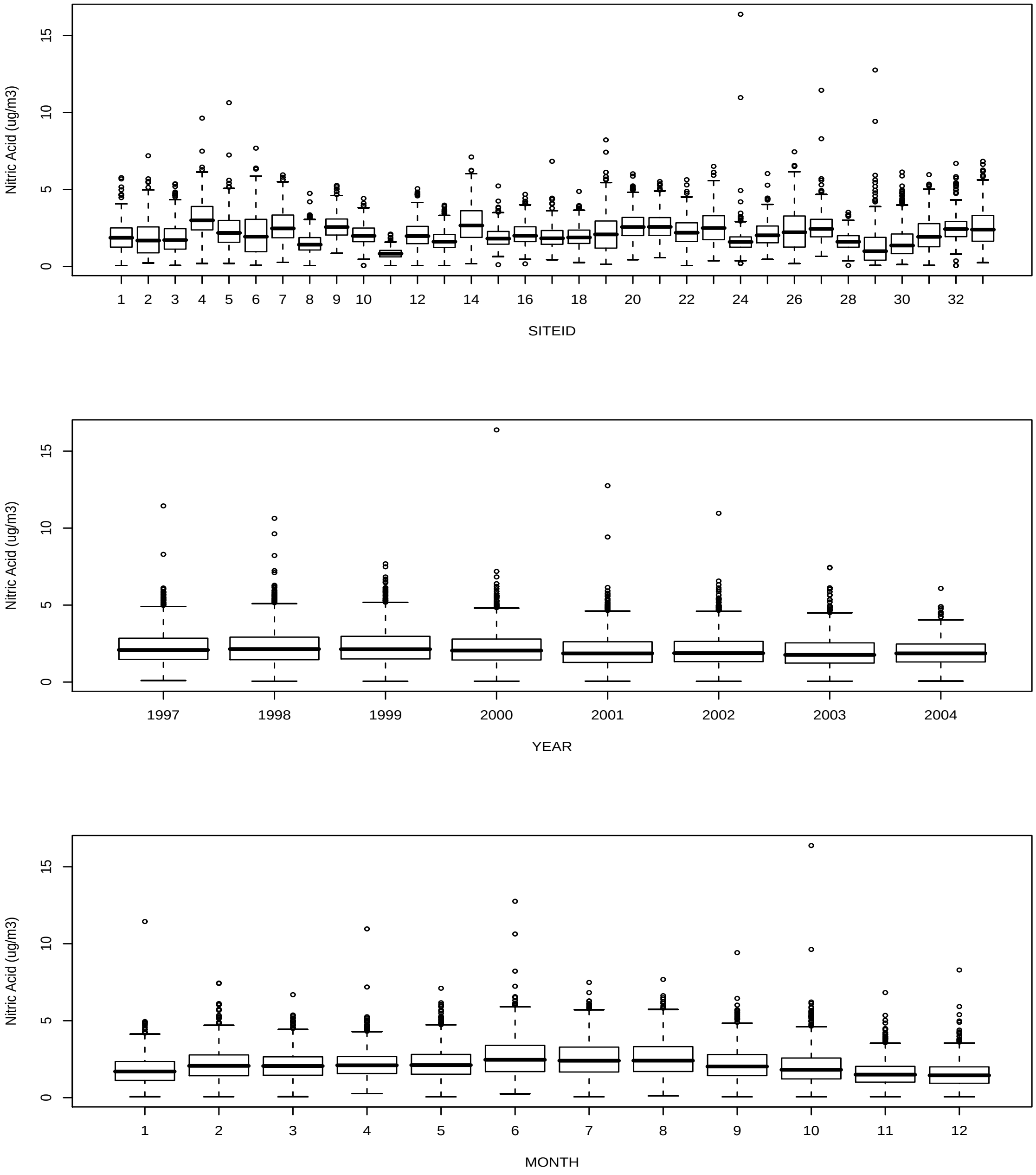

Figure A.1: Box Plots for HNO_3 by Siteid, Year and Month

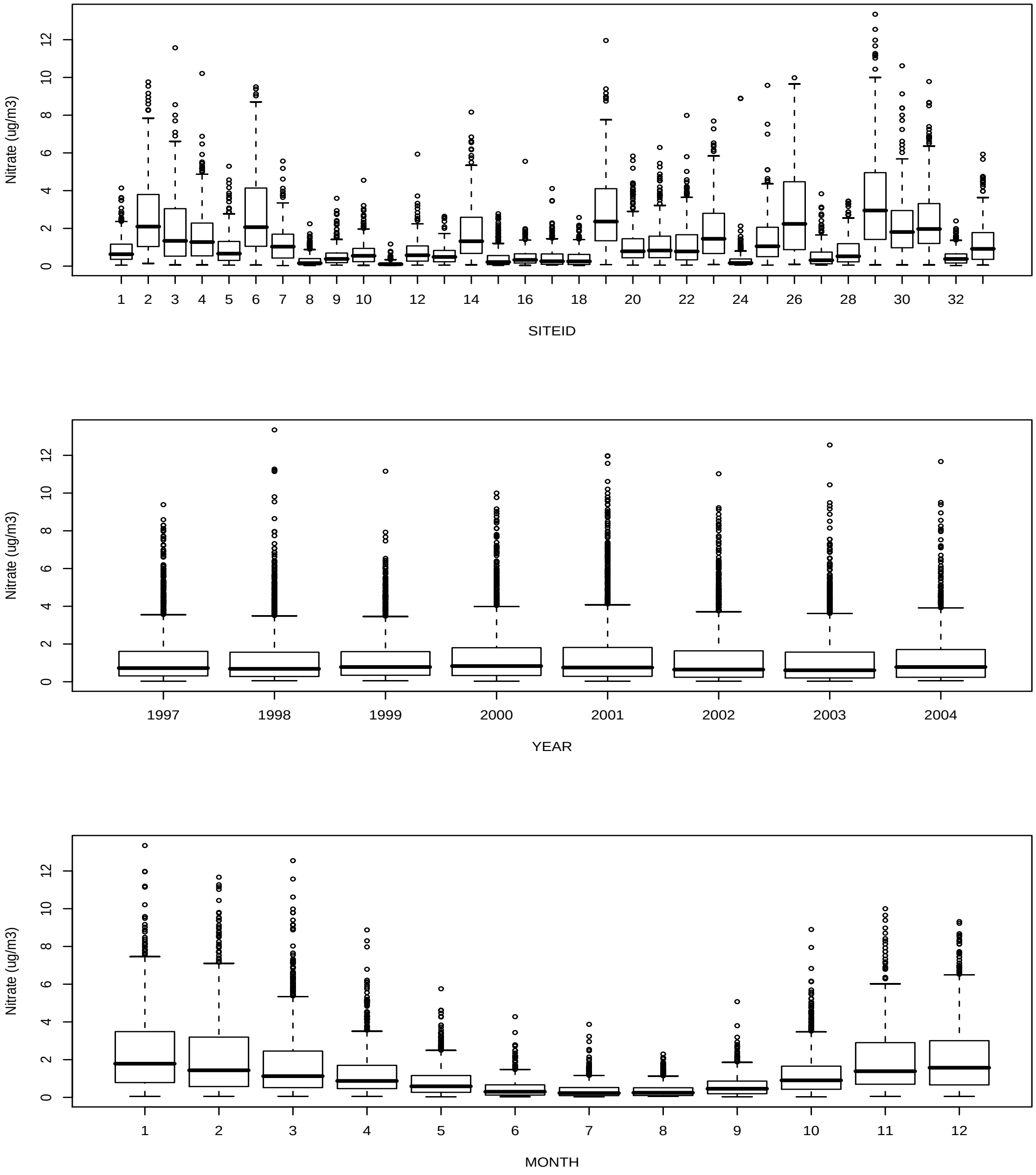

Figure A.2: Box Plots for NO_3 by Siteid, Year and Month

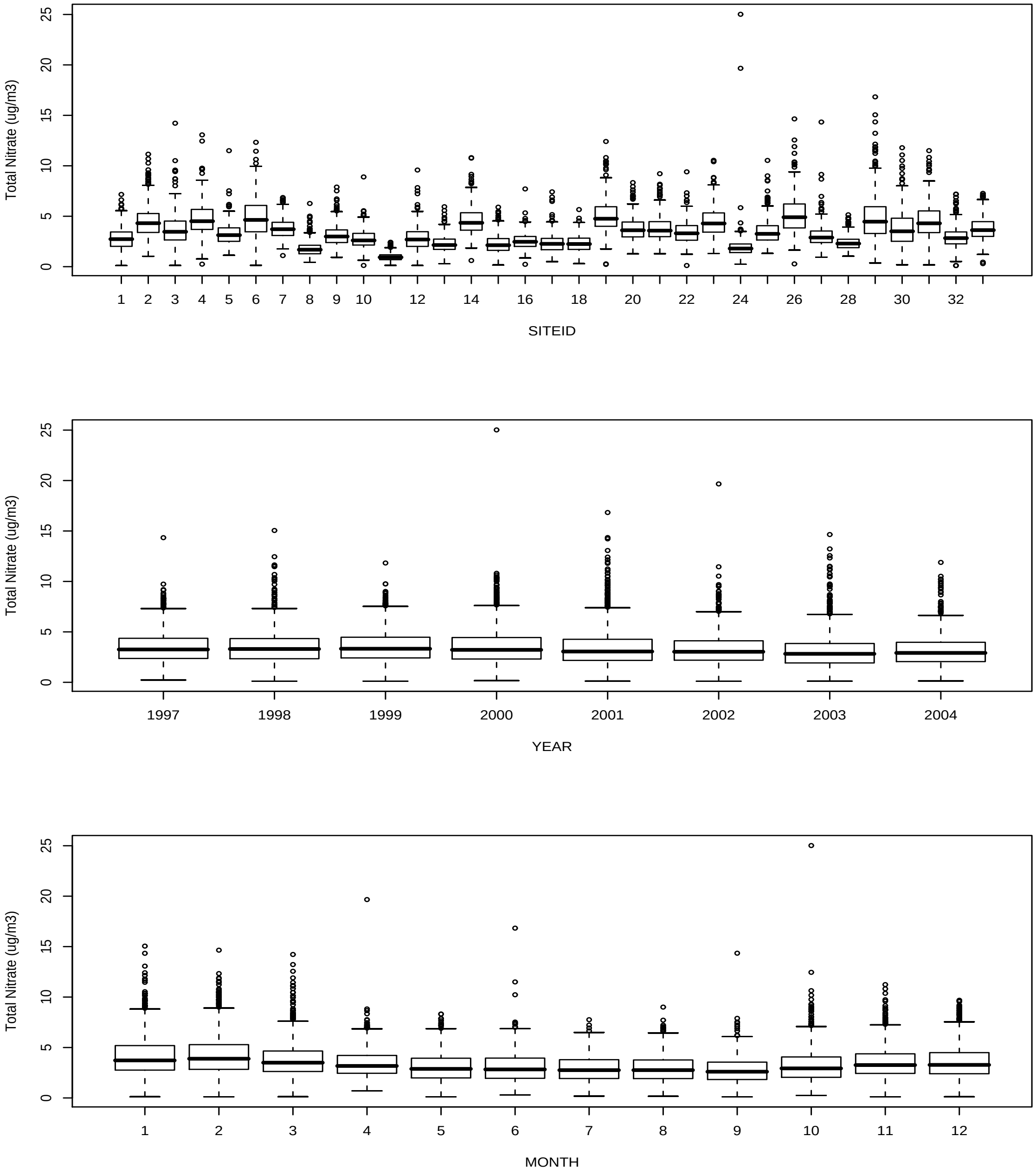

Figure A.3: Box Plots for TNO_3 by Siteid, Year and Month

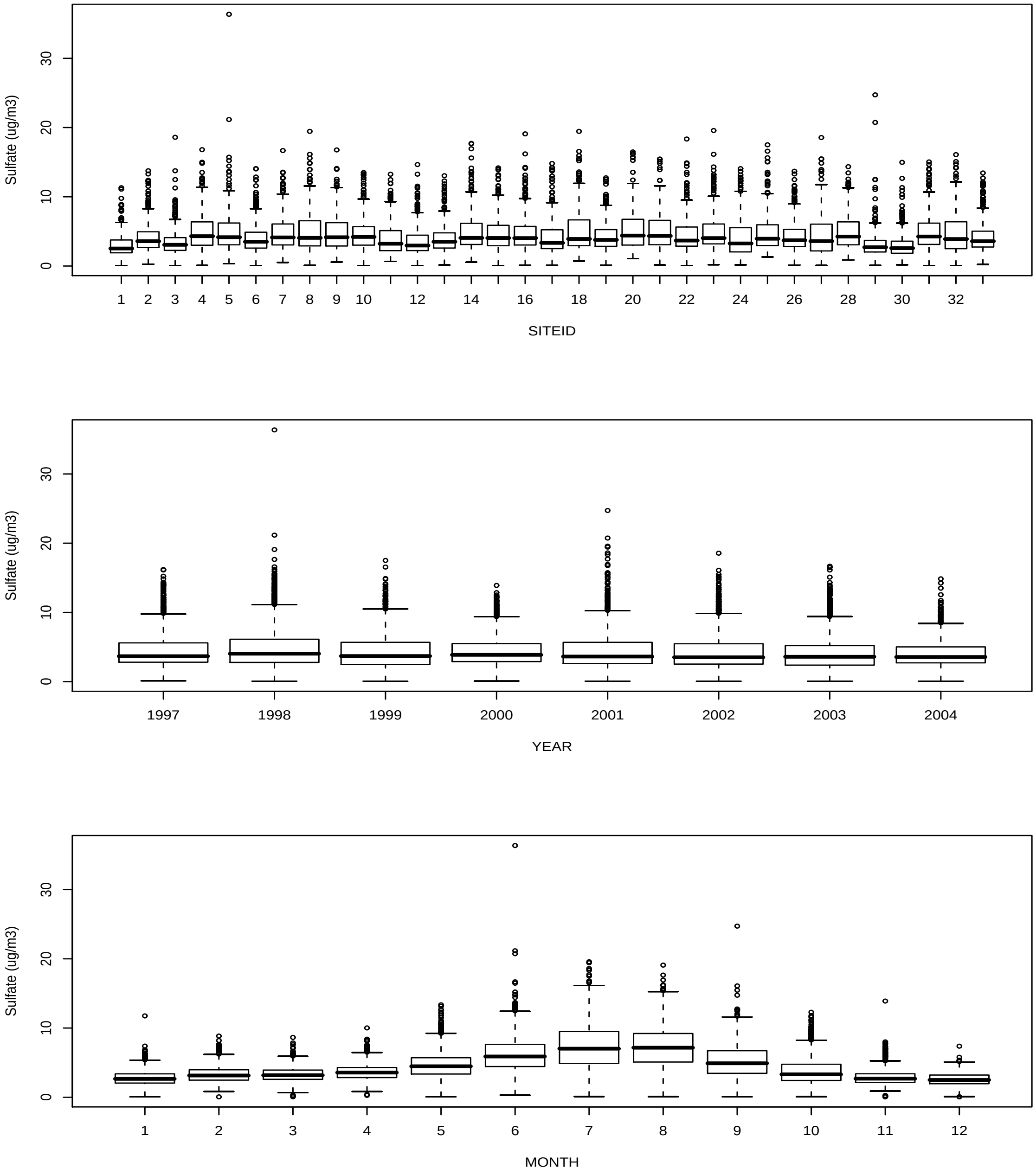

Figure A.4: Box Plots for SO_4 by Siteid, Year and Month

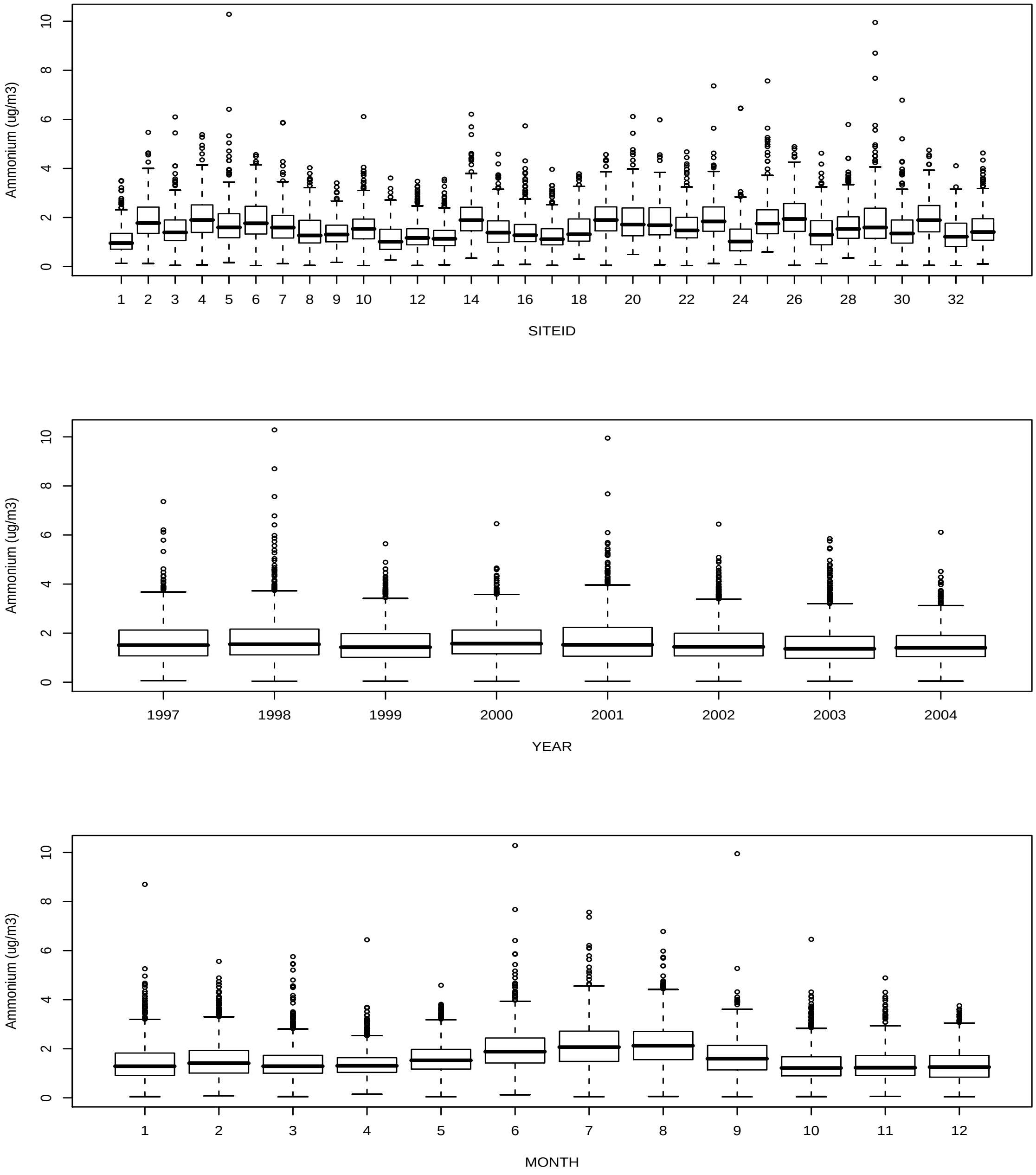

Figure A.5: Box Plots for NH_4 by Siteid, Year and Month

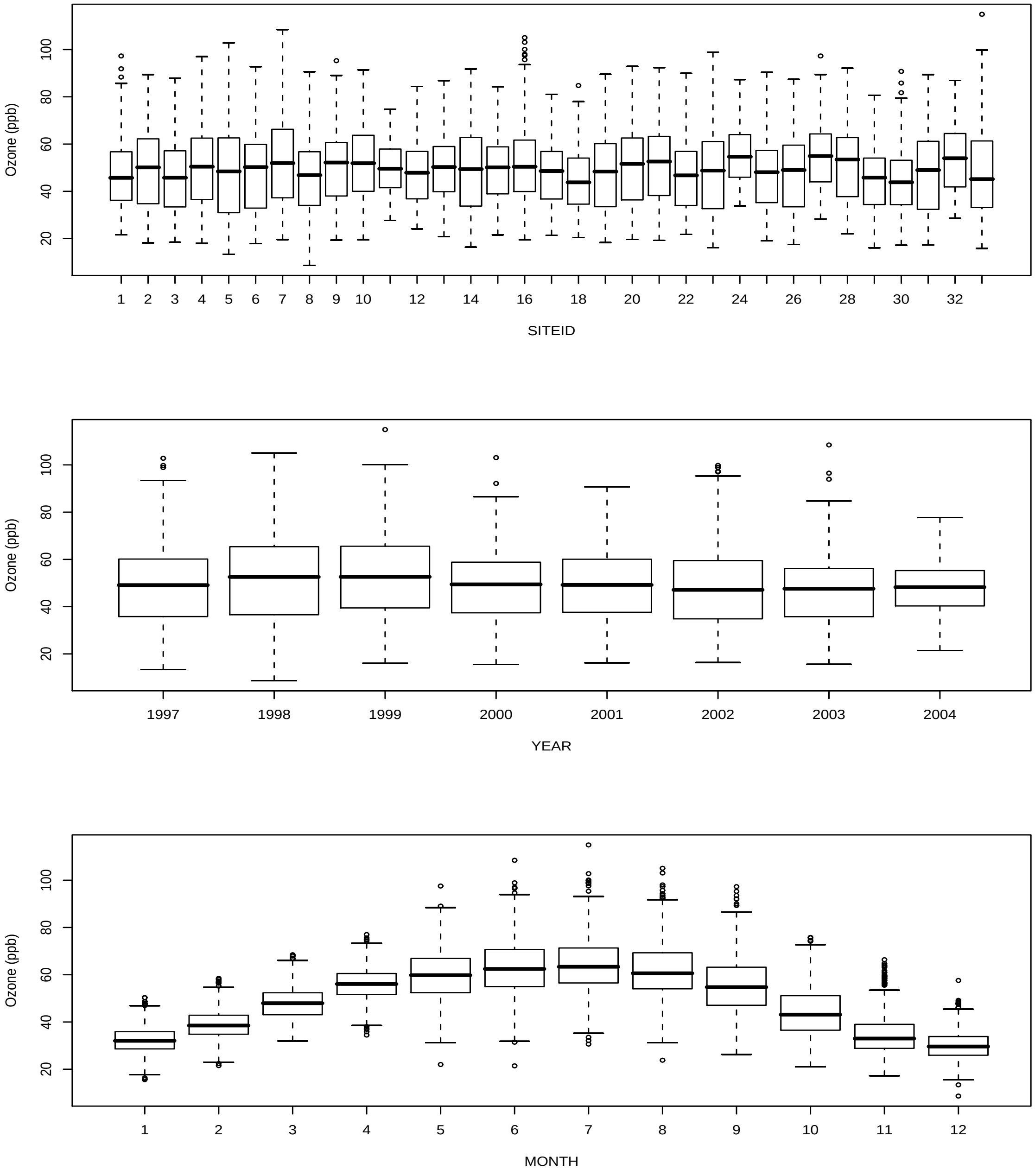

Figure A.6: Box Plots for O_3 by Siteid, Year and Month

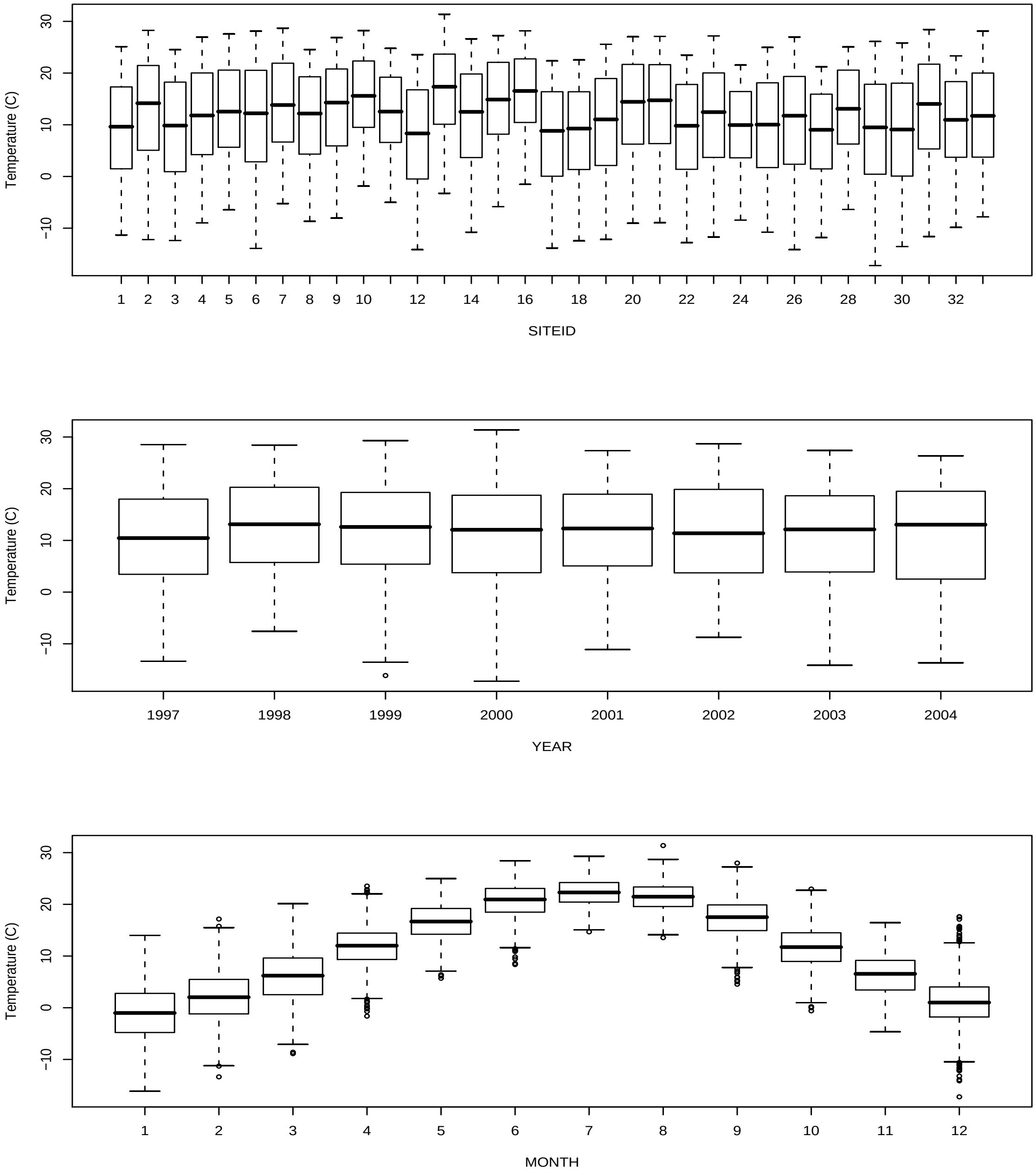

Figure A.7: Box Plots for Temperature by Siteid, Year and Month

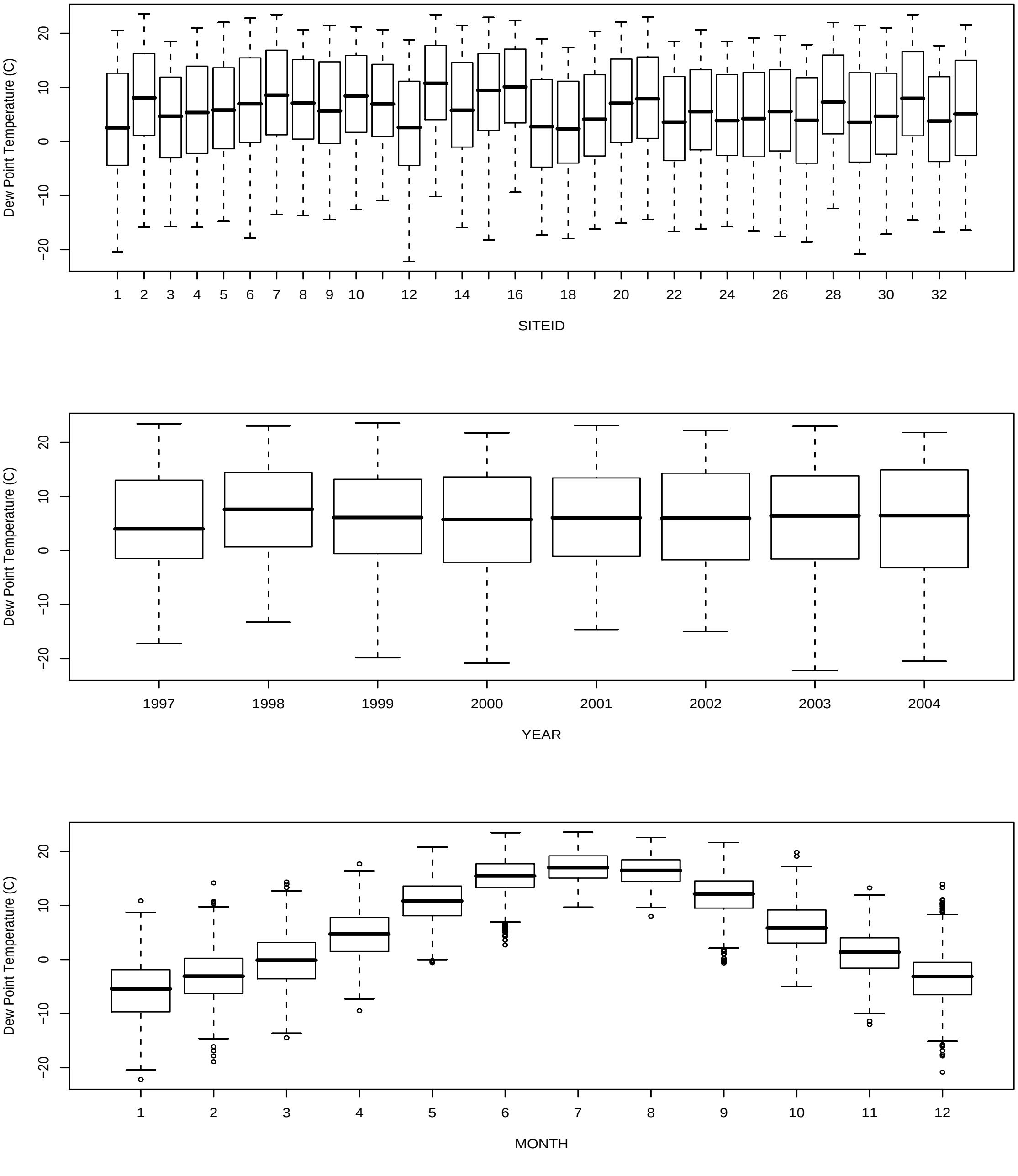

Figure A.8: Box Plots for Dew Point Temperature by Siteid, Year and Month

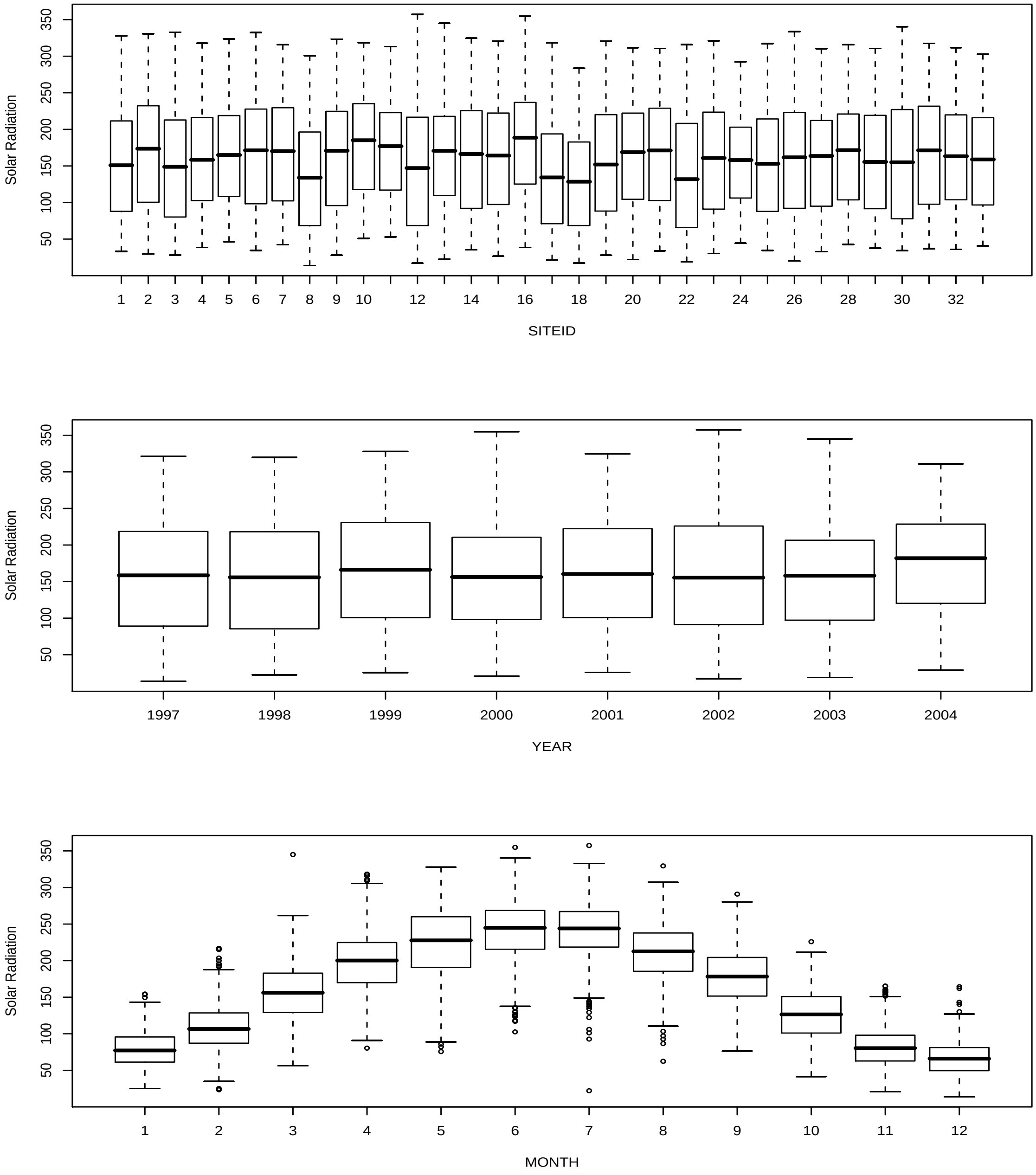

Figure A.9: Box Plots for Solar Radiation by Siteid, Year and Month

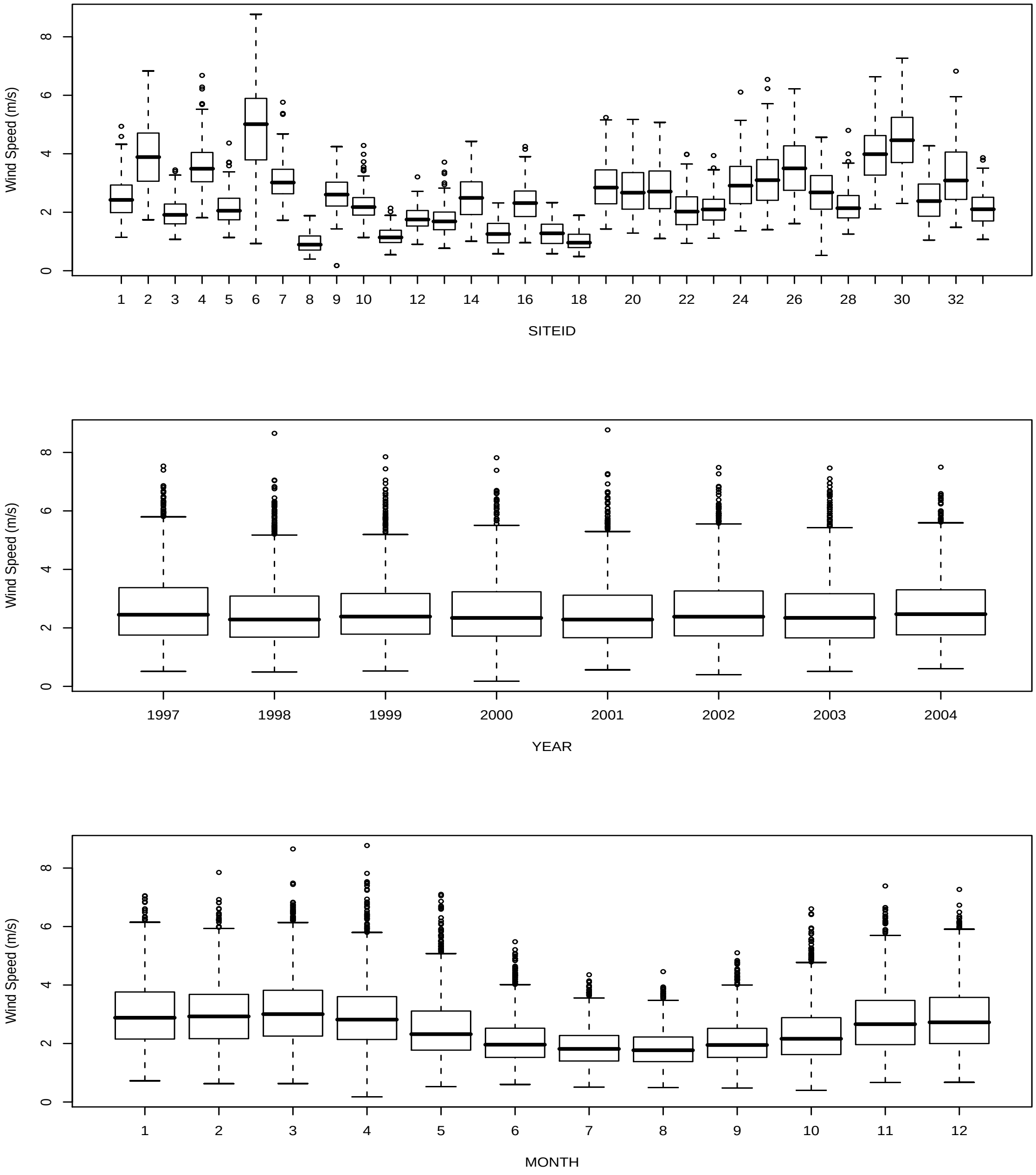

Figure A.10: Box Plots for Wind Speed by Siteid, Year and Month

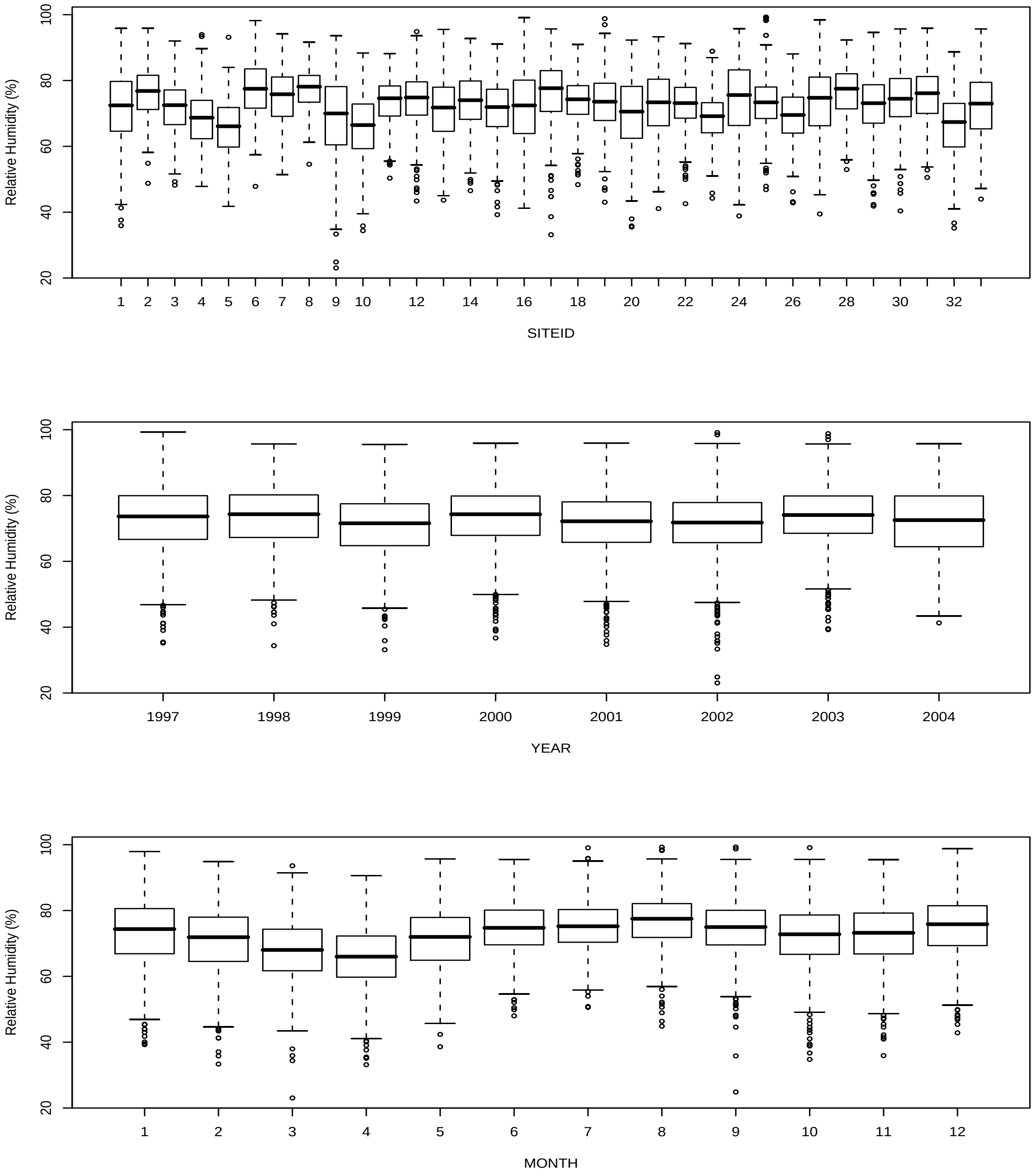

Figure A.11: Box Plots for Relative Humidity by Siteid, Year and Month

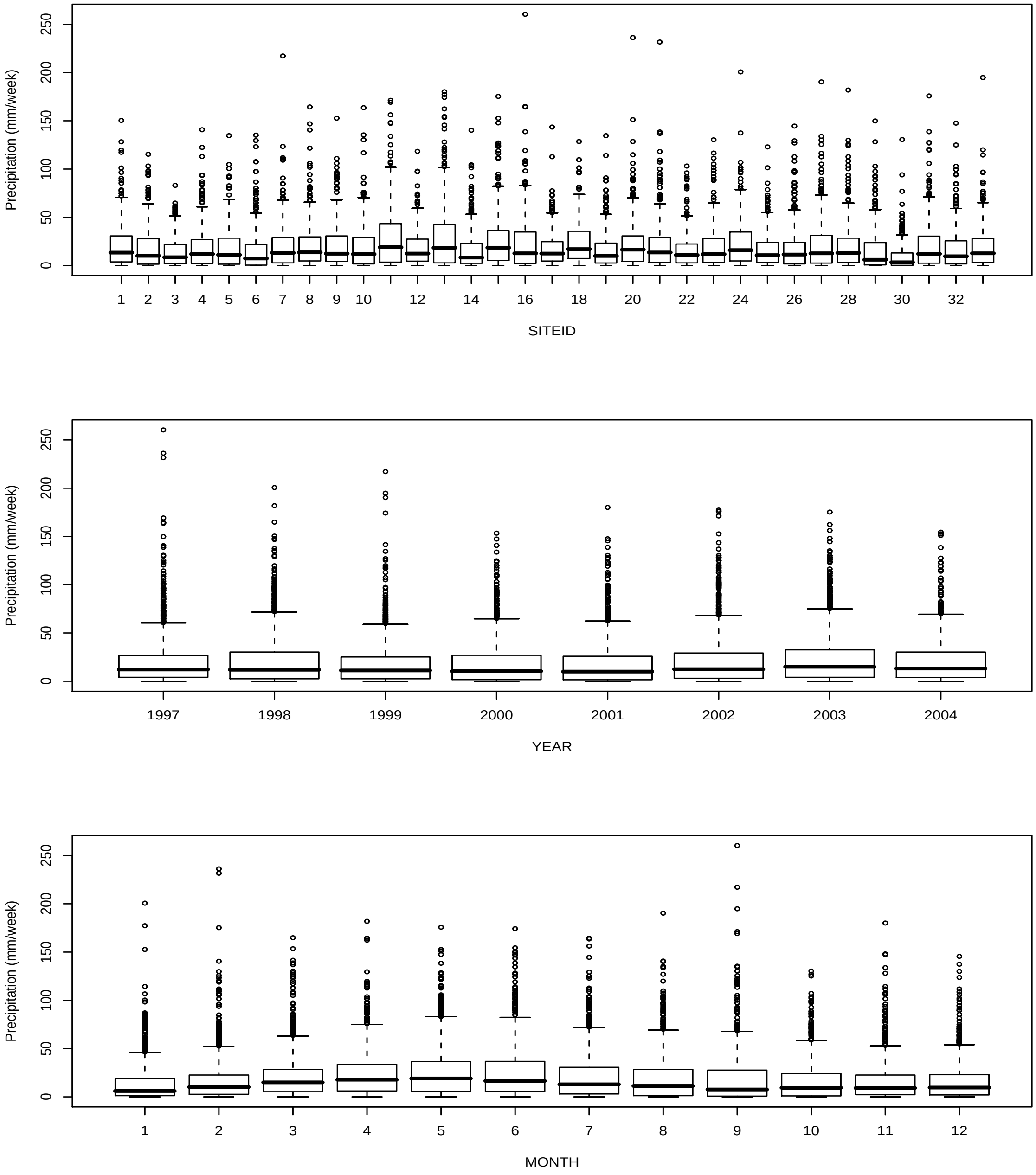

Figure A.12: Box Plots for Precipitation by Siteid, Year and Month