INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1348 USA
800-521-0600

UMI®
RICE UNIVERSITY

An Approach to Modeling a Multivariate Spatial-Temporal Process

by

Mary Anne Calizzi

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Approved. Thesis Committee:

Katherine B. Ensor, Professor. Chair
Department of Statistics

Dennis D. Cox. Professor
Department of Statistics

Matthew P. Fraser, Assistant Professor
Department of Environmental Science and Engineering

Houston, Texas

May, 2000
Abstract

An Approach to Modeling a Multivariate Spatial-Temporal Process

by

Mary Anne Calizzi

Although modeling of spatial-temporal stochastic processes is a growing area of research, one underdeveloped area in this field is the multivariate space-time setting. The motivation for this research originates from air quality studies. By treating each air pollutant as a separate variable, the multivariate approach will enable modeling of not only the behavior of the individual pollutants but also the interaction between pollutants over space and time. Studying both the spatial and the temporal aspects of the process gives a more accurate picture of the behavior of the process. A bivariate state-space model is developed and includes a covariance function which can account for the different cross-covariances across space and time. The Kalman filter is used for parameter estimation and prediction. The model is evaluated through the prediction efforts in an air-quality application.
Acknowledgments

This thesis would not have been possible without the endless support of my husband, Peter Calizzi, and my parents, Dewey and Teresa Sifford. I also appreciate all of the guidance offered by my advisor, Kathy Ensor; her help with this project is invaluable. I thank the members of my committee for their advice and for the time they spent with me on this thesis.

I gratefully acknowledge the assistance of M. W. Hemphill of the Texas Natural Resource Conservation Commission, the Houston Regional Monitoring Network, Loren Hopkins, and David Andrews with advice and data collection.

This research was funded by NSF grants GER-9355087 and DMS-9704648.
Contents

Abstract ii
Acknowledgments iii
List of Illustrations vi

1 Background 1
1.1 Spatial-Temporal Background 2
1.2 Spatial Statistics 8
1.3 Multivariate Time Series 9
1.4 State-Space Modeling 10

2 Model Development 12
2.1 Bivariate Model 12
2.1.1 State-Space Equations 13
2.1.2 Prediction 14
2.2 Covariance Model 22
2.3 Preliminary Parameter Estimation 24
2.4 Maximum Likelihood Estimation Through the Kalman Filter 26

3 Application to Houston Air Quality 29
3.1 Data Description 31
3.2 Transformations 36

4 Preliminary Analysis 40
4.1 Model Application 40
4.2 Results ................................................................. 46
  4.2.1 Predicted Surfaces ........................................... 46
  4.2.2 Cross-Validation ............................................. 50

5 Model Application and Results ................................. 56
  5.1 Model Application ................................................. 56
  5.2 Predicted Surfaces ............................................... 59
  5.3 Cross-Validation .................................................. 60

6 Conclusions and Other Issues ................................ 75

A Notation .................................................................. 77

B S-Plus Programs ...................................................... 78
  B.1 Kalman Filter ......................................................... 78
  B.2 Input File .............................................................. 81
  B.3 Other Functions ..................................................... 82

Bibliography ............................................................ 84
Illustrations

3.1 Map of 22 Houston-area monitoring stations.................. 31
3.2 Time series plots at three monitoring stations................. 33
3.3 Time series plots of median \( O_3 \) and \( NO_X \) levels across all sites ... 34
3.4 Kriged surface of median \( O_3 \) levels across time.................. 35
3.5 Boxplots of transformed ozone by day............................ 37
3.6 Bivariate time series of transformed observations.............. 38

4.1 Time series plots of average transformed \( O_3 \) and transformed \( NO_X \) levels across all sites.................................................. 41
4.2 Auto- and cross-correlation plots............................... 42
4.3 Partial auto-correlation plots.................................... 43
4.4 Variogram fits for the 2 time series............................ 45
4.5 Surfaces of predicted \( O_3 \) and standardized differences at one time point...................................................... 48
4.6 Surfaces of predicted \( NO_X \) and standardized differences at one time point...................................................... 49
4.7 Cross-validation results for \( O_3 \) and \( NO_X \) at Site D........... 51
4.8 Cross-validation results for \( O_3 \) and \( NO_X \) at Site J........... 52
4.9 Cross-validation results for \( O_3 \) at Site E.......................... 53
4.10 Cross-validation results for \( O_3 \) at Site U.......................... 54

5.1 Map of 7 Bivariate Sites........................................... 57
5.2 Predicted surface plot for $O_3$. .......................................... 61
5.3 Contour plot of standardized differences. ................................. 62
5.4 Predicted surface plot for $NO_X$. ........................................ 63
5.5 Contour plot of standardized differences. ................................. 64
5.6 Cross-validation results for $O_3$ and $NO_X$ at Site F ............... 66
5.7 Cross-validation results for $O_3$ and $NO_X$ at Site G ............... 67
5.8 Cross-validation results for $O_3$ at Site E ............................... 68
5.9 Cross-validation results for $O_3$ and $NO_X$ at Site D ............... 71
5.10 Cross-validation results for $O_3$ and $NO_X$ at Site F .............. 72
5.11 Cross-validation results for $O_3$ and $NO_X$ at Site G .............. 73
Chapter 1

Background

The modeling of spatial-temporal stochastic processes is a growing area of research. Modern technology has fueled this growth through increased computing power and advanced techniques used in the observation of spatial-temporal processes.

Few of the existing models have a broad range of application: research in the space-time setting is usually aimed at a very specific situation. Most spatial-temporal models include elements from classical geostatistics, such as Stein [21] and Posa [17], or time series analysis, such as Rouhani and Wackernagel [19] and Haslett and Raftery [7]. However, the modeling efforts generally cannot be labeled as simple extensions of geostatistical or time series models. Thus, the literature presents a variety of methods for analyzing spatial-temporal problems with few standard or widely accepted approaches.

One underdeveloped area in this field of research is the multivariate space-time setting, which is the focus of this paper. The motivation for this research originates from air quality studies. Although the modeling of air pollutants has often been studied, the multivariate perspective is rarely considered. In general, observations of air pollutants are made over space and time. By treating each pollutant as a separate variable, the multivariate approach will enable modeling of not only the behavior of the individual pollutants but also the interaction between pollutants over space and time. This interaction may enhance our understanding of the entire air-quality picture. The application used in this study is described in Chapter 3.

Ultimately, we want to develop strategies for analyzing such observations which will facilitate the complex problem of reducing air pollution. In this paper, our
contribution to air quality improvement efforts is the development of a multivariate spatial-temporal model with a covariance function which can account for the different cross-covariances across space and time. In the model we combine the spatial and temporal components of multiple variables in a manner which has not been done before now.

The first chapter of this thesis provides a literature review of related research in spatial-temporal modeling. Additionally, this chapter includes basic background theory for spatial statistics, time series analysis, and state-space modeling. Chapter 2 outlines the newly-developed bivariate state-space model for application to air quality data. Additionally, equations are developed for prediction through Kalman filtering, and issues surrounding covariance modeling and parameter estimation are addressed. Next, the air-quality application is described in Chapter 3. Descriptions of the data and transformations of the data are included as well.

Analysis of the model begins in Chapter 4, which contains a preliminary application of the model. This preliminary analysis does not address prediction with the Kalman filter. The full modeling effort is evaluated in Chapter 5. The model is evaluated by comparing bivariate spatial-temporal predictions to the actual observations and to the predictions from the preliminary analysis. Finally, further uses for the model are discussed in Chapter 6.

1.1 Spatial-Temporal Background

The study of spatial-temporal processes is sometimes approached solely from the spatial viewpoint. However, drawing on time series theory in addition to spatial statistical methodology can enhance the analysis. Ignoring the time domain constitutes a failure to use all the information available from the data.
In this section, we review some of the recent research efforts in spatial-temporal modeling. Although this is not a complete review of existing literature, we do present examples of the different prevailing approaches to analyzing spatial-temporal data. We are particularly concerned with multivariate modeling, as this is the setting for our air-quality application. Because the multivariate spatial-temporal approach appears infrequently in the literature, univariate models are also reviewed.

Recent work in multivariate-spatial-temporal modeling includes that of Mardia and Goodall [12] and Rouhani and Wackernagel [19]. Mardia and Goodall present various ideas on how to model spatial-temporal data in a multivariate setting. The different models are formed by making varying assumptions about the mean function of the multivariate process. One simple approach presented in their paper is to use a repeated measures model, assuming the data are equally spaced in time. This is generalized to a factored model which separates the covariance into the product of a time-dependent term and a time-independent (spatial) term. The temporal covariance term comes from a time series process.

The multivariate modeling in Rouhani and Wackernagel [19] is quite different. The authors note that hydrological data sets often consist of a "few scattered clusters of observations, each with a long time series." They choose to treat observations from each spatial location as a separate time series, and these time series are correlated. Thus the observed process is viewed as a multivariate time series where each univariate series represents the observations at one location. In other words, let the spatial-temporal set of observations from \( N \) locations and \( T \) time intervals, \( \{z_i(t_\alpha) : i = 1, \ldots, N; \alpha = 1, \ldots, T\} \), be samples from a set of \( N \) variables. Then the variables can be considered a realization of a set of random functions in one dimension:

\[
\{Z_i(t) : i = 1, \ldots, N\}.
\]
The covariance structure is modeled through variograms at different temporal scales; the method relies on the nested variogram modeling in Journel and Huijbregts [10].

This approach is fundamentally different from ours in the multivariate component. Here the multivariate aspect of the modeling comes from the multiple spatial locations: only one variable is measured at those sites. In our setting we have more than one variable as well as multiple spatial locations.

The authors also present a brief summary of problems encountered in space-time data analysis. In addition to the scattered clusters, other issues include spatial nonstationarity, periodicity in the time series, scaling problems between space and time, and the frequent disparity between the amount of spatial information and the amount of temporal information available in a data set. Rouhani and Wackernagel argue that their approach, which focuses on the dimension which is richer in information, is better suited to deal with some of the problems in spatial-temporal modeling.

Univariate spatial-temporal modeling is more common. Applications with environmental data appear in journals across many disciplines, including hydrology, meteorology, and geology. Haslett and Raftery [7], for example, use spatial-temporal modeling to estimate wind power in Ireland. The model includes information from seasonality, spatial correlation, short-memory temporal dependence, and long-memory temporal dependence. Both point and interval estimators are developed.

The mean function is estimated through generalized least squares, which incorporates the spatial covariance structure. Fractional differencing is used to model the long-memory dependence. Additionally, the model assumes second-order spatial-temporal stationarity. Finally, Haslett and Raftery discuss features which are not included in the model such as height of monitoring stations, wind direction, and prior expert information or knowledge.
Oehlert [15] studied sulfate concentrations measured at a network of monitoring locations. The study region is tiled into rectangles, and all sites within a rectangle are assumed to have the same mean and trend. The model is designed to estimate the mean and trend for an individual site as well as for larger regions. Using an idea similar to Rouhani and Wackernagel [19], Oehlert treats the sites (or rectangles) as multiple time series which are spatially correlated. First the mean and trend for each site are estimated using ordinary least squares. Then the means and trends for all rectangles are estimated using the individual estimates as well as a discrete smoothness prior for trend.

In Huang and Cressie [8], spatial-temporal modeling is applied to snow water equivalent data: the model is compared to the purely spatial model which was normally used for this situation. The basic model for the observed spatial-temporal process is

$$Z_t(s) = S_t(s) + \varepsilon_t(s); \ s \in \mathbb{R}^2.$$ 

where $Z_t(s)$ is the observed process at time $t$ and location $s$, $S_t(s)$ is the unobserved state process, and $\varepsilon_t(s)$ is a white-noise process.

The temporal dependence in the unobserved state process $S_t(s)$ is modeled as an autoregressive process with order $p$. Predictions and prediction errors for $S_t(s)$ are obtained through the Kalman filter algorithm. Kalman filtering is advocated because it allows for easy updating of $\hat{S}_t(s)$ as new observations are made. The spatial dependence and spatial-temporal interaction appear in the covariance modeling and the estimation of covariance parameters. More detail on this approach will be presented in Chapter 2.

Studies which model air quality data include Cox and Ensor [2] and Carroll et al. [4]. The Carroll paper presents a study of human exposure to ambient ozone. The study location is Houston, Texas (Harris County, specifically). The pattern of
population exposure over 14 years is summarized. The ozone model has a deterministic trend component and a Gaussian random field component. The deterministic part relies on time, temperature, and "possibly other meteorological data." The random field includes spatial and temporal correlation. Census data are used in the assessment of population exposure to ozone.

Due to the large amount of data generated in this type of monitoring, the authors present a fast method for model fitting. For instance, in fitting the random component, the deterministic trend is assumed to be known: the least squares estimates are used as the known quantities. And for prediction, a moving window is used to limit the number of observations that are included in the calculations.

The authors choose to look at exposure to young children, since this is a population group which can be particularly sensitive to high ozone levels. Also, the lack of mobility for this group makes the use of population density "more meaningful" than it would be for a more mobile group. Census data provides population densities for the years 1980 and 1990. Linear interpolation is utilized to obtain population densities for the years in between, and extrapolation is similarly used for the years after 1990. Ozone levels in general decreased across the region for the time period studied. And the authors find that ozone exposure for children in Houston has decreased "substantially" over time. They also discuss the placement of additional ozone monitoring stations in order to make more accurate assessments.

Cox and Ensor also study ozone levels in the Houston area. The goal of this project, though, is to assess "the accuracy of the existing monitoring network for capturing the daily maximum" ozone level. Additionally, the authors use the data from the monitoring stations to predict ozone levels at locations within the study region which are not monitored. The observed ozone process is decomposed into two parts. First, the underlying ozone surface is modeled using a 3-dimensional spatial
process, treating time as the third dimension. Additionally, an error term accounts for measurement error and microscale variation. The spatial-temporal model for ozone is a constant-mean spatial process, $\zeta(\xi)$, with generalized covariance given by $\text{Cov}[\zeta(\xi), \zeta(\eta)] = -\alpha \|\xi - \eta\|$. where $\alpha$ is an estimated scale parameter. The model depends on time and spatial location as well as temperature, wind speed, and wind direction. The modeled process is used in prediction of ozone levels at locations for which ozone is not observed. In these ozone models, temperature is used as a substitute for sunlight, which is necessary for the formation of ozone. Other meteorological information, such as wind measurements, is also used in modeling the transport of air pollutants.

Simple models for describing spatial-temporal processes have been suggested by Stein [21] and Posa [17]. Although simple models can be quite appealing, these models do make some strong assumptions in order to achieve that simplicity. For instance, Stein models the process as “the sum of a random field fixed in time plus a second independent random field that varies both spatially and temporally.” but he assumes no spatial-temporal interaction. The author notes that this model is not appropriate for modeling a process when measurements are made frequently in time.

Posa assumes that the range of the spatial correlation does not change with time, an idea that is similar to the one proposed by Mardia and Goodall. With an additional stationarity assumption, the proposed covariance model is:

$$C(h; t) = A_0(t)S(h).$$

where $A_0(t)$ is the variance of the spatial-temporal process and $S(h)$ is the spatial correlation. So the variance $A_0(t)$ accounts for the temporal variation. On the other hand, the spatial correlation $S(h)$ is a function of distance $h$ and is time
independent. Posa applies the model to a data set of monthly dissolved oxygen measurements across a region.

Currently, there is no widely accepted general approach to spatial-temporal modeling. Most of the methods are designed to work in a highly specific situation. The methods described in this section, however, provide insight for spatial-temporal modeling in other contexts as well. Yet we recognize that adapting methods to new applications can be difficult due to the specific assumptions required by the original model.

As technology develops for measuring spatial-temporal processes, data sets are becoming larger and more prevalent. Spatial-temporal analysis is usually computationally intensive, and larger data sets only underscore this issue. The good news is that advancement of computer technology should enable us to keep up with analysis of large data sets. Very little software is available for general application to spatial-temporal data: this may be explained in part by the lack of a unified approach to modeling spatial-temporal processes. Although with advances in research, we anticipate that the availability of software will soon improve.

1.2 Spatial Statistics

In this section we briefly present background information on spatial statistics, a key component in our model development. The term "spatial statistics" encompasses the statistical analysis and modeling of data obtained from different spatial locations. The variation across spatial locations is of fundamental interest and motivates research in spatial statistics.

A description of the basic spatial model [3, page 8] follows. Let $s$ be any $d$-dimensional location in $D$, where $D \subset \mathbb{R}^d$. Then the spatial random process is
\{Z(s) : s \in D\}. A realization of the random process is written as \{z(s) : s \in D\}. The subset \(D\), or sampling region, is generally assumed to be fixed (not random).

The variation in a spatial process is often illustrated by the use of a variogram \(2\gamma(\cdot)\), where \(\text{var}[Z(s_1) - Z(s_2)] = 2\gamma(s_1 - s_2)\) for all \(s_1, s_2 \in D\). For an intrinsically stationary process, the variogram is only a function of \(h = s_1 - s_2\), and so a plot of \(|| h ||\) vs. \(\gamma(h)\) provides a graphical representation of the variation. A variety of variogram estimation techniques exist as do a number of theoretical variogram models.

Many of the currently used methods in spatial statistics were developed in a geostatistical setting. That is, the work originated from applications in earth science. Cressie [3] and Isaaks and Srivastava [9] contain details on the major topics in this field.

1.3 Multivariate Time Series

A second important component in our model development is the study of time series data. Modeling multiple individual time series as a multivariate time series is very useful: such joint modeling allows for the description of dependence within each series as well as dependence between individual series. In fact, multivariate time-series modeling may provide the opportunity to improve forecasting for individual time series. Improvements may come from the additional information provided by the other series [18].

Much of the general univariate time series theory can be extended to multivariate time series. Let the time series \(Y_t\) be an \(m\)-dimensional time series vector. In this paper when a time series is referred to as stationary, assume the definition for second-order stationarity is used. So if \(Y_t\) is a stationary time series, the first and second moments must be finite and must not depend on time [1]. In other words,
the mean $E[Y_t] = \mu$ is constant and the covariance $E[(Y_t - \mu)(Y_{t+1} - \mu)^T]$ depends only on the time lag $l$.

The vector autoregressive, or AR($p$), process has a representation similar to the univariate case. The AR($p$) process is

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \eta_t$$

where $A_1, \ldots, A_p$ are $m \times m$ parameter matrices and $\eta_t$ is an $m$-dimensional white-noise process.

1.4 State-Space Modeling

We will develop our spatial-temporal model using a state-space representation. Such a representation is advantageous because it can be applied to a "wide range of time series models" [6]. One advantage of the state-space representation is that it provides a set of updating equations with which we easily obtain the Kalman filter. Kalman filtering has prediction and smoothing capabilities. Additionally, a state-space model and the Kalman filter permit maximum likelihood estimation of unknown model parameters. Under normality assumptions, the prediction equations that make up the Kalman filter algorithm are also used to build the likelihood function.

The general state-space model consists of a state (or transition) equation, which has a first-order Markov structure, and an observation equation, which relates the observed process and the unobserved state process. Assume we apply the state-space representation to a multivariate time series $Z_t$. Then the state equation may be written as

$$Y_t = \Phi_t Y_{t-1} + \eta_t$$
where $\mathbf{Y}_t$ is an $m \times 1$ vector and a first-order Markov process: $\Phi_t$ is an $m \times m$ matrix of coefficients; and $\eta_t$ is a white-noise process. The observation equation is defined as

$$\mathbf{Z}_t = Q_t \mathbf{Y}_t + \varepsilon_t.$$  

Here, $\mathbf{Z}_t$ is an $n \times 1$ vector of observations of the multivariate time series at time $t$, $Q_t$ is an $n \times m$ matrix, and $\varepsilon_t$ is a zero-mean error process, which is uncorrelated with $\eta_t$.

To see how this state-space modeling applies to the vector autoregressive process, consider the following example from Fuller [5]. Let $\mathbf{Y}_t$ be an $m$-dimensional, first-order autoregressive process. Then the AR(1) process is the state equation:

$$\mathbf{Y}_t = A_1 \mathbf{Y}_{t-1} + \eta_t.$$  

where $A_1$ is the AR(1) coefficient and $\eta_t$ is a zero-mean vector process. Assume the process we observe is perturbed by measurement error. Then the observation equation is

$$\mathbf{Z}_t = \mathbf{Y}_t + \varepsilon_t,$$  

where $\varepsilon_t$ is an $m$-dimensional, zero-mean vector process. The vector of random variables $\varepsilon_t$ represents the measurement error and is uncorrelated with $\eta_t$.

In summary, a state-space representation is a useful way to combine all the elements of our model. That is, we can incorporate spatial and temporal components into a multivariate model in state-space form, which leads to methods for prediction and parameter estimation. In Chapter 2, we develop a bivariate state-space model with a bivariate autoregressive structure in the temporal domain and an exponential variogram structure in the spatial domain. Also in Chapter 2, we derive the Kalman filter to obtain minimum mean squared error predictors and prediction errors for this model.
Chapter 2

Model Development

We will follow the spatial-temporal modeling effort in Huang and Cressie [8] as a basis for our multivariate spatial-temporal model. Although the theory extends to a general multivariate model, for now we consider the details of the model only for the bivariate case. In this chapter, the model is defined with a state-space formulation. We also describe the spatial-temporal prediction, develop the covariance modeling, and outline estimation methods for the model parameters.

2.1 Bivariate Model

Assume that two processes are observed at a set of locations or "sites". For any location \( s \) and time \( t \), we may observe both processes or only one process. Assume the observed processes are \( Z_1(s, t) \) and \( Z_2(s, t) \), where \( s \in D \) and \( t = 0, 1, 2, \ldots \)

Suppose that the observations are modeled as a partial realization of the bivariate Gaussian random process

\[
Z(s, t) : s \in D
\]

which can be decomposed into the sum of a state process and an error process. As suggested in Cressie [3, pages 112-113], we write the decomposition as

\[
Z(s, t) = S(s, t) + \varepsilon(s, t). \tag{2.1}
\]

where \( S(\cdot, \cdot) \) is intrinsically stationary in the spatial dimension and second-order stationary in the time dimension. The error process \( \varepsilon(\cdot, \cdot) \) is a zero-mean white noise process with covariance matrix \( \Sigma_e \) and independent of \( S(\cdot, \cdot) \). Note that the observed, state, and error processes are two-dimensional.
In order to model temporal dependence on the past, represent the state process as a vector autoregressive process with order \( p \) and dimension 2:

\[
S(s, t) = \sum_{i=1}^{p} A_i S(s, t - i) + \eta(s, t)
\]  

(2.2)

The coefficients \( A_1, \ldots, A_p \) are \( 2 \times 2 \) matrices. The selection of the order \( p \) will be based on the data. For temporal stationarity in the univariate setting, the coefficients of the AR(\( p \)) are chosen so that the roots of the characteristic equation are less than one in absolute value. The multivariate analog to this is that the roots of

\[
det[\lambda^p I - \lambda^{p-1} A_1 - \cdots - A_p] = 0
\]

must be less than one in absolute value [22]. The process \( \eta(s, t) \) is a zero-mean temporal white noise process with spatial and temporal stationarity. Spatial stationarity is explained in Section 1.2.

Spatial dependence is modeled through the covariance structure. Specifically, the covariance of \( \eta(s, t) \) is modeled with a spatial covariance function. In multivariate time series, the covariance of \( \eta_t \) is usually estimated nonparametrically rather than modeled as a function of the spatial distance. If the spatial component were not important to our study, we might consider using well-known estimation techniques such as maximum likelihood or least squares methods for covariance estimation. However, we chose a modeling approach that captures the spatial dependence in the process. This modeling is detailed in Section 2.2.

2.1.1 State-Space Equations

Now consider the bivariate model for observations at \( n \) locations put in the state-space representation. First we must identify the observation and state equations. The observation vector consists of all observations of the process for a single time \( t \)
and for all locations $s_1, \ldots, s_n$:

$$Z(t) \equiv [Z(s_1, t)^T, \ldots, Z(s_n, t)^T]^T$$

$$= [Z_1(s_1, t), Z_2(s_1, t), \ldots, Z_1(s_n, t), Z_2(s_n, t)]^T$$

This vector has length $2n$. The observation equation defines the relationship between the observed process, $Z(s, t)$, and the unobserved state process, $S(s, t)$:

$$Z(t) = S(t) + \varepsilon(t). \quad (2.3)$$

where $S(t) \equiv [S(s_1, t)^T, \ldots, S(s_n, t)^T]^T$ is the corresponding vector of the state variables at the $n$ sites. The state equation for $S(s, t)$ is defined in Equation 2.2. It describes the transition of the state process from one time period to the next. Recall that the white noise process $\varepsilon(s, t)$ is independent of $S(s, t)$.

State-space modeling does not require that the state and observation vectors have the same dimension. Recall from Section 1.4 that the general state-space form allows for an $n \times m$ matrix in the observation equation, assuming that the observation vector is of length $n$ and the state vector is of length $m$. In our air pollution application, however, it makes sense to assume $Z$ and $S$ have the same dimension. We can think of the state process $S$ as the portion of the observed process which is undisturbed by noise.

### 2.1.2 Prediction

We use the bivariate state-space model given in Equation 2.1 to predict the state process at spatial locations where no observation is made and at future times. Using Kalman filtering we can obtain the minimum mean squared error predictor (MMSEP) of the state process and prediction errors, assuming known parameters.

In this section we compute the optimal spatial-temporal predictor of the state process at some spatial location $s_0$ and any time $t_0$, given the observations up to
time \( t \). In other words, estimate \( \mathbf{S}(s_0, t_0) \) by calculating

\[
\hat{\mathbf{S}}(s_0, t_0|t) \equiv E[\mathbf{S}(s_0, t_0)|\mathbf{Z}(1), \ldots, \mathbf{Z}(t)],
\]

where \( \mathbf{Z}(j) \) is the vector (length \( 2n \)) of all observations at time \( j \). First, we develop the updating (or prediction) equations that are used in the Kalman filter algorithm. Then we explain how these equations are used in the calculation of the predictor and prediction error. The notation introduced and defined in this section is also summarized in Appendix A.

Assume the bivariate state-space model (2.1 and 2.2) holds: then all joint distributions are multivariate normal. We will specify the mean vector and covariance matrix for the joint distribution of interest. The mean and covariance formulas will be used to build the updating equations. Consider the joint distribution of \( \mathbf{S}(\ldots) \) at two locations \( s \) and \( r \) and times \( t - p + 1, \ldots, t \) and \( \mathbf{Z}(t) \). In other words, the distribution of

\[
[\mathbf{S}(s, t - p + 1)^T, \ldots, \mathbf{S}(s, t)^T, \mathbf{S}(r, t - p + 1)^T, \ldots, \mathbf{S}(r, t)^T, \mathbf{Z}(t)^T]^T
\]

conditional on \( (\mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1)) \) is multivariate normal with mean

\[
[\hat{\mathbf{S}}(s, t - p + 1|t - 1)^T, \ldots, \hat{\mathbf{S}}(s, t|t - 1)^T, \hat{\mathbf{S}}(r, t - p + 1|t - 1)^T, \ldots, \hat{\mathbf{S}}(r, t|t - 1)^T, \hat{\mathbf{Z}}(t|t - 1)^T]^T,
\]

where \( \hat{\mathbf{S}} \) is the expectation defined in the previous paragraph and

\[
\hat{\mathbf{Z}}(t|t - 1) \equiv E[\mathbf{Z}(t)|\mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1)].
\]

Additionally, the covariance matrix for this multivariate distribution is

\[
\begin{bmatrix}
\mathbf{\Sigma}_{tt - 1}(s,s) & \mathbf{\Sigma}_{tt - 1}(s,r) & \mathbf{G}(s,t)^T \\
\mathbf{\Sigma}_{tt - 1}(r,s) & \mathbf{\Sigma}_{tt - 1}(r,r) & \mathbf{G}(r,t)^T \\
\mathbf{G}(s,t) & \mathbf{G}(r,t) & \mathbf{F}(t)
\end{bmatrix}
\]
There are three types of elements in this matrix: $\mathbf{\Sigma}_{t|t-1}(\cdot, \cdot)$, $G(\cdot, t)$, and $F(t)$. Each element in the covariance matrix represents the appropriate sub-matrix, considering the vector of random variables in the joint distribution. For instance, $\mathbf{\Sigma}_{t|t-1}(s, r)$ is defined as the covariance of these 2 vectors: $[\mathbf{S}(s, t - p + 1)^T \ldots \mathbf{S}(s, t)^T]^T$ and $[\mathbf{S}(r, t - p + 1)^T \ldots \mathbf{S}(r, t)^T]^T$, conditional on $(\mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1))$. So the definition of this term is

$$\mathbf{\Sigma}_{t|t-1}(s, r) \equiv \text{cov} \left[ \begin{pmatrix} \mathbf{S}(s, t - p + 1) \\ \vdots \\ \mathbf{S}(s, t) \end{pmatrix}, \begin{pmatrix} \mathbf{S}(r, t - p + 1) \\ \vdots \\ \mathbf{S}(r, t) \end{pmatrix} \right] | \mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1).$$

The other elements are defined as follows:

$$F(t) \equiv \text{cov}[\mathbf{Z}(t)|\mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1)]$$

$$G(s, t) \equiv \text{cov}[\mathbf{Z}(t), (\mathbf{S}(s, t - p + 1)^T \ldots \mathbf{S}(s, t)^T)^T|\mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1)].$$

Now through multivariate normal theory (see, for example, Searle [20]), we take the distribution just defined for the vector in (2.4) and formulate the distribution of a subset of that vector. Then we get that the joint distribution of

$$[\mathbf{S}(s, t - p + 1)^T \ldots \mathbf{S}(s, t)^T \mathbf{S}(r, t - p + 1)^T \ldots \mathbf{S}(r, t)^T]^T$$

given $(\mathbf{Z}(1), \ldots, \mathbf{Z}(t))$ is also multivariate normal with mean

$$\begin{pmatrix} \hat{\mathbf{S}}(s, t - p + 1|t - 1) \\ \vdots \\ \hat{\mathbf{S}}(s, t|t - 1) \\ \hat{\mathbf{S}}(r, t - p + 1|t - 1) \\ \vdots \\ \hat{\mathbf{S}}(r, t|t - 1) \end{pmatrix} + \begin{pmatrix} G(s, t)^T \\ G(r, t)^T \end{pmatrix} F^{-1}(t)[\mathbf{Z}(t) - \hat{\mathbf{Z}}(t|t - 1)]$$
and covariance matrix
\[
\begin{bmatrix}
\Sigma_{t|t-1}(s,s) & \Sigma_{t|t-1}(s,r) \\
\Sigma_{t|t-1}(r,s) & \Sigma_{t|t-1}(r,r)
\end{bmatrix} = \begin{bmatrix} G(s,t)^T \\ G(r,t)^T \end{bmatrix} F^{-1}(t)[G(s,t),G(r,t)]
\]

Finally, this conditional distribution is used to formulate a vector of predictors of the state process:
\[
\begin{bmatrix}
\hat{S}(s, t-p+1|t) \\
\vdots \\
\hat{S}(s, t|t)
\end{bmatrix} = E \left[ \begin{bmatrix} S(s, t-p+1) \\
\vdots \\
S(s, t) \end{bmatrix} \right] Z(1), \ldots, Z(t)
\]
\[
= \begin{bmatrix}
\hat{S}(s, t-p+1|t-1) \\
\vdots \\
\hat{S}(s, t|t-1)
\end{bmatrix} + [G(s, t)]^T F^{-1}(t)[Z(t) - \hat{Z}(t|t-1)] \tag{2.5}
\]

and the covariance matrix:
\[
\Sigma_{t|t}(s, r) \equiv \text{cov} \left[ \begin{bmatrix} S(s, t-p+1) \\
\vdots \\
S(s, t) \end{bmatrix}, \begin{bmatrix} S(r, t-p+1) \\
\vdots \\
S(r, t) \end{bmatrix} \right] Z(1), \ldots, Z(t)
\]
\[
= \Sigma_{t|t-1}(s, r) - [G(s, t)]^T F^{-1}(t)[G(r, t)] \tag{2.6}
\]

Equations 2.5 and 2.6 are the first of the updating equations for the Kalman filter algorithm. These equations define the state vector and the error matrix in terms of observations from the current time step and estimates from the previous time step. The elements in these 2 equations are defined and simplified below in Equation 2.7 through Equation 2.11. The seven equations in (2.5) through (2.11) make up the complete set of updating equations.

Along with the notation already established in this section, the following definitions are necessary for the simplification of the updating equations:
\[
Q \equiv \begin{pmatrix}
0 & 0 & \cdots & I_{p-2} \\
& \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_1 \\
A_p & A_{p-1} & \cdots & A_1
\end{pmatrix}.
\]

\[Q[a:b] \equiv \text{subset of } Q \text{ from row } a \text{ through row } b.\]

\[A_* \equiv [A_p \cdots A_1].\]

\[\Sigma_{t_1,t_2|s}(r) \equiv \text{cov}[S(s,t_1), S(r,t_2)|Z(1), \ldots, Z(t)].\]

\[\Sigma_{t_0,t|s}(r) \equiv \text{cov}[S(s,t_0), S(r,t)|Z(1), \ldots, Z(t)].\]

Note that \(Q\) is a matrix of size \(2p \times 2p\) and \(A_*\) is a matrix of size \(2 \times 2p\). The covariance terms are \(2 \times 2\) matrices.

Now the following five updating equations (2.7 through 2.11) define the elements of Equations 2.5 and 2.6. Prediction of the state vector at \(p\) time points is given by

\[
\begin{bmatrix}
\dot{S}(s, t-p+1|t-1) \\
\vdots \\
\dot{S}(s, t|t-1)
\end{bmatrix}
= E
\begin{bmatrix}
S(s, t-p+1) \\
\vdots \\
S(s, t)
\end{bmatrix}
\begin{bmatrix}
Z(1), \ldots, Z(t-1)
\end{bmatrix}
\]

\[
= E \begin{bmatrix}
Q
\end{bmatrix}
= Q
\begin{bmatrix}
\dot{S}(s, t-p|t-1) \\
\vdots \\
\dot{S}(s, t-1|t-1)
\end{bmatrix}.
\]

The one time step ahead predictor of the process is found through

\[
\dot{Z}(s, t|t-1) \equiv E[Z(s, t)|Z(1), \ldots, Z(t-1)]
\]

\[
= E[S(s, t) + \varepsilon(s, t)|Z(1), \ldots, Z(t-1)]
\]

\[
= E[S(s, t)|Z(1), \ldots, Z(t-1)]
\]
\[
\begin{align*}
&= E \left[ \sum_{l=0}^{p-1} A_{p-l} \mathbf{S}(s, t - p + l) + \eta(s, t) | \mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1) \right] \\
&= \sum_{l=0}^{p-1} A_{p-l} \hat{\mathbf{S}}(s, t - p + l | t - 1).
\end{align*}
\] (2.8)

Updating the conditional covariance matrix of the state process at two different spatial locations yields

\[
\begin{align*}
\Sigma_{t|t-1}(s, r) & \equiv \text{cov} \\
&= \text{cov} \left[ \begin{pmatrix} \mathbf{S}(s, t - p + 1) \\ \vdots \\ \mathbf{S}(s, t) \end{pmatrix}, \begin{pmatrix} \mathbf{S}(r, t - p + 1) \\ \vdots \\ \mathbf{S}(r, t) \end{pmatrix} \right] | \mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1) \\
&= \text{cov} \left[ \begin{pmatrix} \mathbf{S}(s, t - p) \\ \vdots \\ \mathbf{S}(s, t - 1) \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ \eta(s, t - 1) \end{pmatrix} \right] + \text{cov} \left[ \begin{pmatrix} \mathbf{S}(r, t - p) \\ \vdots \\ \mathbf{S}(r, t - 1) \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ \eta(r, t - 1) \end{pmatrix} \right] | \mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1) \\
&= \text{cov} \left[ \begin{pmatrix} \mathbf{S}(s, t - p) \\ \vdots \\ \mathbf{S}(s, t - 1) \end{pmatrix}, \begin{pmatrix} \mathbf{S}(r, t - p) \\ \vdots \\ \mathbf{S}(r, t - 1) \end{pmatrix} \right] | \mathbf{Z}(1), \ldots, \mathbf{Z}(t - 1) \right] Q^T \\
&\quad + \text{cov} \left[ \begin{pmatrix} 0 \\ \vdots \\ \eta(s, t - 1) \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ \eta(r, t - 1) \end{pmatrix} \right] \\
&= Q \Sigma_{t-1|t-1}(s, r) Q^T + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{C}^*(s, r) \end{pmatrix}
\end{align*}
\] (2.9)
The only unknown quantity here is the covariance of the spatial-temporal additive noise, namely $C^n(s, r) \equiv \text{cov}[\eta(s, t), \eta(r, t)]$. This quantity is determined through covariance modeling developed in Section 2.2.

Next consider the covariance of the observed process conditional on the past:

$$
F(t) \equiv \text{cov}[Z(t)|Z(1), \ldots, Z(t-1)]
$$

$$
= \text{cov}[\bar{Z}(t) + \bar{\xi}(t)|Z(1), \ldots, Z(t-1)]
$$

$$
= \text{cov}[\bar{Z}(t)|Z(1), \ldots, Z(t-1)] + \text{cov}[\bar{\xi}(t)]
$$

$$
= \text{cov}[(S(s_1, t)^T, \ldots, S(s_n, t)^T)^T|Z(1), \ldots, Z(t-1)] + \sigma_t^2 I
$$

$$
= \begin{bmatrix}
\Sigma_{t|t-1}(s_1, s_1) & \cdots & \Sigma_{t|t-1}(s_1, s_n) \\
\vdots & \ddots & \vdots \\
\Sigma_{t|t-1}(s_n, s_1) & \cdots & \Sigma_{t|t-1}(s_n, s_n)
\end{bmatrix} + \sigma_t^2 I
$$

$$
= \begin{bmatrix}
A(\Sigma_{t-1|t-1}(s_1, s_1)A^T & \cdots & A(\Sigma_{t-1|t-1}(s_1, s_n)A^T \\
\vdots & \ddots & \vdots \\
A(\Sigma_{t-1|t-1}(s_n, s_1)A^T & \cdots & A(\Sigma_{t-1|t-1}(s_n, s_n)A^T
\end{bmatrix} + \sigma_t^2 I. \ (2.10)
$$

Finally, we have the covariance between the observed and state process conditional on the past, namely

$$
G(s, t) \equiv \text{cov}[Z(t), (S(s, t - p + 1)^T, \ldots, S(s, t)^T)^T|Z(1), \ldots, Z(t-1)]
$$

$$
= \text{cov}[(\bar{Z}(t) + \bar{\xi}(t)), (S(s, t - p + 1)^T, \ldots, S(s, t)^T)^T|Z(1), \ldots, Z(t-1)]
$$

$$
= \begin{bmatrix}
\Sigma_{t,t-p+1|t-1}(s_1, s_1) & \cdots & \Sigma_{t,t-p+1|t-1}(s_1, s_n) \\
\vdots & \ddots & \vdots \\
\Sigma_{t,t-p+1|t-1}(s_n, s_1) & \cdots & \Sigma_{t,t-p+1|t-1}(s_n, s_n)
\end{bmatrix}
$$

$$
= \begin{bmatrix}
A(\Sigma_{t-1|t-1}(s_1, s_1)Q[1:2]^T & \cdots & A(\Sigma_{t-1|t-1}(s_1, s_n)Q[2p-1:2p]^T \\
A(\Sigma_{t-1|t-1}(s_n, s_1)Q[1:2]^T & \cdots & A(\Sigma_{t-1|t-1}(s_n, s_n)Q[2p-1:2p]^T
\end{bmatrix}.
$$

(2.11)
To implement the Kalman filter algorithm, we start with observations of the state process and initial predictors. The predictors are recursively updated using the updating equations. The result of this updating is a vector of conditional means (2.5), which is the minimum mean squared error predictor (MMSEP) of the state vector [6]. This recursive procedure allows for prediction of the state process at future times. Specifically, for observations up to time \( t_0 \), we can predict the process one time step ahead at any location \( s_0 \). Note that we do need observations at the location \( s_0 \) in order to predict the state process at that site, but initial conditions (predictions) at \( s_0 \) are necessary. The formulation of the one step ahead predictor is summarized in the following result.

**Result 1**

Assume the state-space model holds. For any location \( s_0 \in D \) and data \( Z_1, \ldots, Z_{t_0} \) the one step ahead MMSEP for \( S(s_0, t_0 + 1) \) is given by

\[
\hat{S}(s_0, t_0 + 1|t_0) = A_1 \hat{S}(s_0, t_0|t_0) + A_2 \hat{S}(s_0, t_0 - 1|t_0) + \cdots + A_p \hat{S}(s_0, t_0 - p + 1|t_0).
\]

The predictor is obtained from the updating equations. Equation 2.5 through Equation 2.11.

Additionally, the mean square prediction error is

\[
\Sigma_{t_0 + 1|t_0}(s_0, s_0) = A_\ast \Sigma_{t_0|t_0}(s_0, s_0) A_\ast^T + C^n(s_0, s_0)
\]

where \( \Sigma_{t_0|t_0} \) is obtained from updating equations. \( C^n(s_0, s_0) \) is a specified spatial covariance model, and \( A_\ast \equiv [A_p \cdots A_1] \).

We can verify this result in 2 steps. First the MMSEP’s \( \hat{S}(s_0, t_0|t_0) \ldots, \hat{S}(s_0, t_0 - p + 1|t_0) \) and the mean squared prediction error \( \Sigma_{t_0|t_0}(s_0, s_0) \) are obtained through
the Kalman filter, as explained in this section. Then the relationship to the one-
step-ahead predictor and its prediction error is derived in the following [8]:

\[
\hat{S}(s_0, t_0 + 1 | t_0) = E[S(s_0, t_0 + 1 | Z(1), \ldots, Z(t_0))]
\]
\[
= E[A_1 S(s_0, t_0) + \cdots + A_p S(s_0, t_0 - p + 1) + \eta(s_0, t_0 + 1 | Z(1), \ldots, Z(t_0))]
\]
\[
= A_1 \hat{S}(s_0, t_0 | t_0) + \cdots + A_p \hat{S}(s_0, t_0 - p + 1 | t_0)
\]
\[
= (2.12)
\]

\[
\Sigma_{t_0+1|t_0}(s_0, s_0) = \text{var}[A(t_0 - p + 1 | t_0)^T \ldots S(s_0, t_0 | t_0)^T]^T
\]
\[
+ \eta(s_0, t_0 + 1 | Z(1), \ldots, Z(t_0))]
\]
\[
= A \cdot \text{var}[(S(s_0, t_0 - p + 1 | t_0)^T \ldots S(s_0, t_0 | t_0)^T | Z(1), \ldots, Z(t_0)].A^T
\]
\[
+ \text{var}[\eta(s_0, t_0 + 1)]
\]
\[
= A \cdot \Sigma_{t_0|t_0}(s_0, s_0) A^T + C^\eta(s_0, s_0).
\]
\[
(2.13)
\]

2.2 Covariance Model

Describing the covariance of a spatial-temporal model can be handled in a variety of
times. We will follow the Huang and Cressie approach, which is to select a covariance
model then estimate the parameters of this model.

In multivariate time series, the covariance of \( \eta_t \) is usually estimated nonpara-
dometrically rather than modeled as a function of the spatial distance. If the spatial
component were not important to our study, we might consider using well-known
estimation techniques such as maximum likelihood or least squares methods for co-
variance estimation. However, we chose a modeling approach that will capture the
spatial dependence in the process.

The development of a covariance model for a bivariate spatial-temporal process
could easily lead to a complicated structure. It is in our best interest to keep the
model simple but still capture the critical structure: even a simple covariance model for such a process may have many parameters to estimate.

The covariance of $S$ and $\eta$ are defined, respectively, as

$$C_h^{(S)}(s, r) \equiv \text{cov} [S(s, t), S(r, t + h)]$$

for $h = \ldots -1, 0, 1, \ldots$ and

$$C^\eta(s, r) \equiv \text{cov} [\eta(s, t), \eta(r, t)].$$

Each matrix has dimension $2 \times 2$. The relationship is derived by first post-multiplying Equation 2.2 by $S(r, t + h)$. Then taking expectations results in

$$C_j^{(S)}(s, r) = \sum_{l=1}^{p} A_l C_j^{(S)}(s, r) + \delta_j C^\eta(s, r), \quad (2.14)$$

where $j = 0, 1, 2, \ldots$. The indicator function $\delta_j$ takes the value one only when $j = 0$ and is zero otherwise. Note that this means that the second term is zero for any time lag other than $j = 0$.

We will assume that the process $\eta$ has no cross-covariance, i.e.

$$C^\eta(s, r) = \begin{bmatrix} C^{\eta_1}(s, r) & 0 \\ 0 & C^{\eta_2}(s, r) \end{bmatrix},$$

where $C^{\eta_1}$ and $C^{\eta_2}$ represent the covariances for the two components of the bivariate error process. This means that we are assuming that $\eta_1$ and $\eta_2$ are uncorrelated, certainly a simplifying assumption in covariance modeling. At the beginning of this section we discussed the need to use a simple structure for the covariance model. Additional models for the cross-covariance would lead to more parameters to estimate, which could give us problems with overparameterizing the model. Although we do not model the cross-correlation directly, the cross-covariance is still represented in the updating equations of the Kalman filter algorithm.
Furthermore, we assume the same two-parameter exponential model for each univariate spatial covariance function:

\[ C_{nj}(s, r) = \sigma_{j1} \exp \left\{ -\sigma_{j2} \| s - r \| \right\} \]  

(2.15)

for \( j = 1, 2 \). Using this model, which is one of the simpler spatial covariance models, simplifies the overall modeling effort without losing key structure. The covariance model then has a total of four parameters which must be estimated. A three-parameter exponential model was also considered, but it was not appropriate for our application.

### 2.3 Preliminary Parameter Estimation

In addition to the four parameters associated with the covariance model, other parameters to be estimated include the temporal parameters \( \Sigma_r \) and \( A_1, \ldots, A_p \). Preliminary estimates of the parameters are obtained through traditional time series and spatial methods. In the subsequent section, we describe the full maximum likelihood approach, which utilizes the Kalman filter.

To obtain preliminary parameter estimates, first consider the parameters of the AR(p) process. Assume the Yule-Walker equations are used for estimation of the AR parameters, although computer packages such as S-plus make other estimation techniques readily available as well. The Yule-Walker equations define the relationship between the autocovariance and the AR coefficients, namely, \((C_1(s, s), \ldots, C_p(s, s))\) and \((A_1, \ldots, A_p)\) in Equation 2.14. The autocovariance function estimator, as defined in Brockwell and Davis [1], may be used to estimate \( C_k(s, s) \). This estimator averages over \( n \) locations and \( t_0 \) time periods and is given by

\[
\hat{C}_k(s, s) = \frac{1}{n t_0} \sum_{i=1}^{n} \sum_{t=1}^{t_0-k} (Z(s_i, t) - \bar{Z})(Z(s_i, t + k) - \bar{Z})^T.
\]  

(2.16)
Averaging over all spatial locations may not be the ideal method to estimate the autocovariance; however, it is an adequate estimate for our analysis. The estimates \( \hat{C}_k^{(S)} \) for \( k = 0, 1, \ldots, p \) are used in the Yule-Walker equations (2.14) to obtain estimates of \( A_1, \ldots, A_p \).

As in Huang and Cressie [8], we will assume that \( \Sigma_e \), the variance of the error process \( \varepsilon(s, t) \), is zero. These air pollutant observations are generally considered to be high-quality data [4], and our preliminary study of our data set reinforces this notion. Thus we conclude that the measurement error is negligible, and we will not explicitly estimate the variance of this error process.

Now the only parameters remaining are those for \( C^{(n)} \), the covariance of the spatial noise process. To estimate the parameters \( \{\phi_{ab}; a = 1, 2; b = 1, 2\} \), we fit an empirical variogram, which is an estimated variogram based on the data, to the following model. The variogram of \( Z(\cdot, t) \) between \( s \) and \( r \) is

\[
\text{cov}[Z(s, t) - Z(r, t)] = \text{cov}[S(s, t) + \varepsilon(s, t)] + \text{cov}[S(r, t) + \varepsilon(r, t)]
\]

\[
-2\text{cov}[S(s, t) + \varepsilon(s, t), S(r, t) + \varepsilon(r, t)]
\]

\[
= 2\text{cov}[S(s, t)] + 2\text{cov}[\varepsilon(s, t)] - 2\text{cov}[S(s, t), S(r, t)]
\]

\[
= 2C_0^{(S)}(s, s) + 2C_0^{(r)}(s, s) - 2C_0^{(S)}(s, r).
\]

This variogram is estimated using the parameter estimates we have calculated thus far, specifically, \( \hat{A}_1, \ldots, \hat{A}_p \), and \( \hat{C}_k^{(S)} \). The first term may be estimated with the autocovariance estimator (2.16). The second term is estimated by \( \hat{\Sigma}_e \), which is assumed to be zero. The final term is obtained through recursively solving the Yule-Walker equations: the general form of the solution is

\[
C_0^{(S)}(s, r) = B C_0^{(n)}(s, r) B^T.
\]  (2.17)
where $B$ is a function of the temporal coefficients, $A_1, \ldots, A_p$, and is obtained through solving the Yule-Walker equations. Then the variogram estimate is

$$\hat{\gamma}(s, r) = \begin{cases} 
2\hat{C}_0^{(S)} - 2B\text{cov}[\eta(s, r)]B^T & s \neq r \\
0 & s = r
\end{cases} \quad (2.18)$$

Fitting an empirical variogram to Equation 2.18 results in parameter estimates denoted $\{\hat{o}_{ab}; a = 1.2, b = 1.2\}$.

Estimation of all of the model parameters without the use of the Kalman filter has now been addressed. The parameter estimates may be combined with the results of the previous sections to form predictions. The results of these predictions are presented in Chapter 4. We now consider maximum likelihood estimation using the Kalman filter.

### 2.4 Maximum Likelihood Estimation Through the Kalman Filter

The Kalman filter has a variety of uses. The basic purpose is to estimate the state vector $S_t$. The algorithm can be formulated to solve one of three types of estimation problems. The filtering problem is the estimation of $S_t$ given observations up to time $t$. The smoothing problem is the estimation of missing values, i.e., estimating $S_t$ given observations up to time $t_0$ where $t_0 > t$. And finally, the third problem is the type we consider here. The prediction problem is the estimation of $S_t$ given observations up to time $t - 1$. This section outlines the general procedure for parameter estimation through the Kalman filter.
The updating equations necessary for the Kalman filter are developed in Section 2.1.2. The updating equations give us conditional expectations of the state process,

\[
\begin{bmatrix}
\hat{S}(s, t_0 - p + 1|t_0) \\
\vdots \\
\hat{S}(s, t_0|t_0)
\end{bmatrix}.
\]

These state estimates are used to calculate the one-step-ahead predictors and prediction errors at one time point, described in Result 1. The predictor is

\[
\hat{S}(s_0, t_0 + 1|t_0) = A_1 \hat{S}(s_0, t_0|t_0) + A_2 \hat{S}(s_0, t_0 - 1|t_0) + \cdots + A_p \hat{S}(s_0, t_0 - p + 1|t_0).
\]

The recursions of the Kalman filter are designed so that for each \( t \in (1, \ldots, t_0) \), all of the prediction equations are updated. At each time step, calculations are based on observations up to the current time and on the estimates from the previous time step. The iteration ends at time \( t_0 \) with the vector of state estimates and the covariance matrix for one spatial location. The next step is calculation of the one-step-ahead predictor \( \hat{S}(s_0, t_0 + 1|t_0) \) and the prediction errors. Recall that the estimated state vector and the covariance matrix are formulated in Equation 2.5 and Equation 2.6: the other updating equations are given in Equation 2.7 through Equation 2.11.

The maximum likelihood estimates (MLE's) for the covariance model parameters (2.15) are obtained using this Kalman filter algorithm. Assuming the bivariate Gaussian model, we build the likelihood function using conditional probability functions [6]. Then the log-likelihood is

\[
\log L = -\frac{N_t}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{t_0} \log |F'_t| - \frac{1}{2} \sum_{i=1}^{t_0} \nu_t^T F_i^{-1} \nu_t
\]

where \( \nu_t = Z_t - \hat{S}_{s,t|t-1} \), often called the innovations. We choose the negative log-likelihood as the objective function; it is minimized using the built-in S-Plus function
“nlminb”. The results of the maximum likelihood estimation for our application are presented in Section 5.1.
Chapter 3

Application to Houston Air Quality

Air quality is currently an important area of study in environmental science, and ambient ozone is a major concern in the study of air quality improvement. High ozone levels are associated with smog and related adverse (health) effects [11, 23]. and thus ozone is one of the criteria pollutants for which a National Ambient Air Quality Standard (NAAQS) must be met. Many areas of the country far exceed the NAAQS for ozone. This application concerns one of those regions, the greater metropolitan area of Houston, Texas. The Houston area has consistently ranked second nationally for excessive ozone levels [4]. Recently levels in Houston bypassed those in the Los Angeles area, which has historically ranked first in the nation for excessive ozone levels. A map of the Houston study region with monitoring stations marked is given in Figure 3.1.

As a result of the Clean Air Act of 1970, ambient ozone levels have been monitored across the United States for almost 30 years. Ozone is a secondary pollutant, and thus it is formed by other contaminants which also contribute to the air pollution problems in metropolitan areas. Observations of other relevant pollutant levels are also available. By using the additional information provided by measurements of other pollutants, we hope to offer improvements to current air pollution modeling efforts. Because air contaminants are typically measured over a spatial region and monitored over time, this setting is an ideal application for multivariate spatial-temporal modeling.

This study focuses on two air pollutants, ambient ozone \((O_3)\) and nitrogen oxides \((NOX)\). Ozone is a secondary pollutant produced by photochemical reactions
involving $NO_X$ and volatile organic compounds ($VOC$s), both of which are primary pollutants [2, 23]. A secondary pollutant is one which is "not emitted directly into the air, but is the result of chemical reactions in the ambient air" [16]. The primary pollutant $NO_X$ is produced "during the combustion of fossil fuels" [11]; thus, sources of $NO_X$ emissions include industries and automobiles. Some of the adverse effects of high levels of these pollutants are human health effects, such as respiratory disease, and ecological problems in the form of damage to vegetation and animals [23].

Because $NO_X$ contributes to the formation of $O_3$, a relationship between the two clearly exists. We want to explore the relationship between these two pollutants. Pollutant measurements are taken over time at several locations around Houston. So, we have a bivariate data analysis problem with spatial-temporal correlation. Other meteorological measurements, such as temperature, wind speed and wind direction, are included as covariates.

Development of multivariate spatial-temporal methods is necessary for modeling the joint behavior of the pollutants. The advantage of this modeling approach is that it incorporates spatial information, temporal information, and interactions.

To work toward the greater goal of improving air quality analysis, we will use modeling results to calculate predictions of pollutant levels along with prediction errors. Predictions will include spatial interpolation, i.e. predicting at locations where no measurements are taken. Another facet of the prediction is forecasting pollutant levels at future times.

The observations used in this study were taken over an 18 day period from August 28, 1997 to September 14, 1997. Ozone levels are measured at 22 locations in the Houston area, and $NO_X$ is measured at 8 of those locations. Figure 3.1
displays the monitoring stations in addition to highlighting major highways, the ship channel and the Gulf of Mexico. Water masses are denoted in gray.

![Map of 22 Houston-area monitoring stations. Both $O_3$ and $NO_X$ are observed at the 8 sites marked with asterisks; only $O_3$ is observed at the other 14 sites.]

**Figure 3.1** Map of 22 Houston-area monitoring stations. Both $O_3$ and $NO_X$ are observed at the 8 sites marked with asterisks; only $O_3$ is observed at the other 14 sites.

### 3.1 Data Description

Observations at three of the monitoring stations over a two-week period are plotted in Figure 3.2. These time series plots represent three sites which measure both $O_3$ and $NO_X$. Each individual series represents hourly observations of one pollutant at a single monitoring station. The first plot displays the $O_3$ observations at three
sites and the second plot displays the $NO_X$ observations at the same three sites: vertical lines separate the days. The hourly median $O_3$ and $NO_X$ levels for the same two-week time period are plotted in Figure 3.3. The medians are calculated across all sites. The diurnal pattern in ozone is clearly displayed in both figures. The diurnal pattern is explained in part by sunlight, which is a component of the reactions that create ambient ozone. The $NO_X$ pattern is not as obvious, but large peaks tend to occur in the late morning hours and smaller peaks in the early evening hours.

Observations at the Galveston Island monitoring station (site C) are atypical for observations in this region. We learned that the chemical processes leading to the formation of $O_3$ are different at that site, due in part to its geographical location. So the assumptions we make about the relationship between $NO_X$ and $O_3$ do not apply at this site. Therefore, the Galveston Island site is omitted from the analysis. This reduces the number of $O_3$ sites to 21 and $NO_X$ to 7.

Recall that the entire data set consists of hourly observations for 18 days. With 28 individual time series, the total number of observations is over 12,000. Missing values do occur throughout the data set at an overall rate of approximately 5% of the total observations. The proportion of missing values for $O_3$ is approximately 4%, while the rate is slightly higher for $NO_X$, more than 8%. Data imputation is described in Section 3.2.

A contour plot based on the median values of ambient ozone is given in Figure 3.4. First, for each of the 21 sites, the median $O_3$ level across time is calculated. Then this set of medians is used to obtain kriging predictions across the study region. The plot represents the surface of kriging predictions.

Available meteorological data include temperature, relative humidity, wind speed and wind direction. Because ozone is produced by photochemical reactions, the
Figure 3.2  Time series plots of $O_3$ and $NO_X$ levels measured in ppb at three monitoring stations for a two-week period (August 28, 1997 - September 10, 1997). Vertical lines separate the days. From top to bottom, the sites in each plot are D, H, and B.
Figure 3.3  Time series plots of median $O_3$ and $NO_X$ levels measured in ppb across all sites for a two-week period (August 28, 1997 - September 10, 1997).
Figure 3.4 Kriged surface of median $O_3$ levels across time, measured in ppb.
amount of sunlight is a key factor in ambient ozone levels. Additionally, the amount of sunlight and observed temperature are strongly correlated, and so the latter variable is used in this analysis. Temperature observations across the region are averaged for each hour. The hourly average temperature is used to extract the diurnal behavior in the observed $O_3$ concentrations. Because relative humidity is highly correlated with $O_3$ levels [13], this variable is also included in the analysis. Ozone levels and relative humidity levels exhibit an inverse relationship.

### 3.2 Transformations

All analysis is performed on transformations of the $O_3$ and $NO_X$ observations. Ozone is first transformed by taking the square root of the observed values and then subtracting the hourly means and dividing by the hourly standard deviations. Assume $X_{st}$ is the square root of the observed ozone level at location $s$ and time $t$. then

$$y_{st} = \frac{X_{st} - \overline{X}_t}{s_{X_t}},$$

where $\overline{X}_t$ is the sample mean $O_3$ level and $s_{X_t}$ is the sample standard deviation of $O_3$ across all sites at time $t$.

Next, the simultaneous effects of temperature and relative humidity are removed through the following regression model

$$Y_{st} = \beta_0 + \beta_1 T_t + \beta_2 I_t T_t + \beta_3 I_t H_t + \zeta_{st},$$

where $T_t$ and $H_t$ are the temperature and relative humidity, respectively, at time $t$. The indicator variable $I_t$ is one for the hours from 6 a.m. to 6 p.m. and is zero otherwise. The residuals of the regression $\zeta_{st}$ are used in the analysis as the transformed $O_3$ values. As noted in Menezes and Shively [13], the interaction of temperature and humidity is not significant.
The transformed $O_3$ series is described by boxplots of $O_3$ levels for each day in Figure 3.5. The fluctuation across days seems to show no specific pattern. Figure 3.6 presents the transformed $O_3$ and $NO_X$ observations for a seven day period at two locations. Although some evidence of a diurnal pattern still remains, there is no longer a dominant deterministic cycle.

![Graph showing boxplots of residual ozone by day](image)

**Figure 3.5** Boxplots of transformed ozone by day.

The $NO_X$ transformations consist of a square root transformation and a standardization by subtracting the overall mean of the square root $NO_X$ values and dividing by the standard deviation of all the values. The increase in $NO_X$ levels during the period of 6 a.m. to 8 a.m. is due to the incoming traffic flow during the morning rush hour. The traffic patterns are not explicitly modeled in this analysis, although we may address this issue in future applications.
In Figure 3.6 time series plots are displayed for two locations, one internal to
the city (Site H) and the second on the Northwest corner of the region (Site E).
We observe from this plot that the $NO_X$ levels in downtown Houston are generally
higher than in the northwest area. This is consistent with intuition in that the
locations with elevated $NO_X$ emissions are the heavily industrial areas (east and
southeast regions, in this case) and areas of high traffic volume.

![Site H](image)

![Site E](image)

**Figure 3.6** Bivariate time series of transformed observations at 2
monitoring stations. Solid lines are $O_3$ and dotted lines are $NO_X$.

The Kalman filter prediction algorithm allows for missing values; however, our
initial analysis does not include the use of the Kalman filter. Because the number
of missing values is a small fraction of the total number of observations, we chose
a simple data imputation approach to deal with missing values. Even when the Kalman filter is implemented, we continued the analysis with the data set which includes the imputed values. For both $NO_x$ and $O_3$, data imputation is utilized with the transformed observations. For each pollutant, the missing values are replaced with the mean of the transformed observations.

To incorporate wind behavior, a final transformation was considered but not used. The spatial coordinates may be transformed from UTM (Universal Transverse Mercator) to Lagrangian coordinates, as in Cox and Ensrar [2]. The Lagrangian coordinate system adjusts for wind behavior by not following a fixed Eulerian system [23]. Instead, the coordinate system follows the wind trajectory, and an air parcel may be considered fixed. Because wind behavior is a primary factor in the transport of air pollutants, use of the Lagrangian system would allow for the inclusion of this important information. For our study period, though, wind was not a factor, and so we did not implement a transformation of the spatial coordinates.
Chapter 4

Preliminary Analysis

Exploratory data analysis (EDA) and a preliminary application of the model presented in Chapter 2 will provide some insight into the behavior of the spatial-temporal processes before Kalman filtering is used. This simpler analysis helps identify potential issues of concern before the Kalman filter is even considered.

4.1 Model Application

In this section we describe the initial application of the bivariate spatial-temporal model using the Houston data. We will apply the parameter estimation techniques described in Section 2.3. Some simplifications are made in order to obtain a preliminary perspective on how well the model works. All analysis is performed on the transformed observations.

Estimation of the model parameters is performed in a sequential fashion. First an aggregate time series for each variable, $O_3$ and $NO_X$ in the transformed space, is obtained by averaging all spatial values within each time segment, i.e. computing $\overline{Z}_1(\cdot,t)$ and $\overline{Z}_2(\cdot,t)$. A segment of the aggregated series is shown in Figure 4.1. An autoregressive model is fit to this new bivariate series using the Yule-Walker Equations in S-Plus. The AR(p) coefficients are estimated from the model fitting.

Figure 4.2 shows the autocorrelation and cross-correlation plots, with 95% confidence bands, for the aggregate series. Transformed $O_3$ is “Series 1” and transformed $NO_X$ is “Series 2.” The correlation should be strong only for a few hours, given the nature of the quick formation and destruction of $O_3$ and $NO_X$. Yet the $O_3$ plot does
Figure 4.1 Time series plots of average transformed $O_3$ and transformed $NO_X$ levels across all sites. The solid line is $O_3$; the dashed line is $NO_X$. 
show high auto-correlation for several lags. Additionally, the $NO_X$ auto-correlation plots suggests that there may be some periodicity which has not been removed. The failure to remove such trends is an issue discussed at the end of this paper.

**Multivariate Series : Spatial Average**

![Graphs showing auto- and cross-correlation plots for the aggregate time series.](image)

**Figure 4.2** Auto- and cross-correlation plots for the aggregate time series. $O_3$ is “Series 1”: $NO_X$ is “Series 2”. Horizontal lines are 95% confidence bounds.

A bivariate AR(2) model is selected for the process. Order two is the recommended order using the AIC criterion and from direct observation of the partial auto-correlation plots in Figure 4.3. Also, the estimates of higher order coefficients were quite small. The cross-correlation plots in Figure 4.2 indicate very little cross-correlation between the two series. (Note the scale of the y-axis and the location
Figure 4.3  Partial auto-correlation plots for the aggregate time series. $O_3$ is "Series 1": $NO_x$ is "Series 2". Horizontal lines are 95% confidence bounds.
of the confidence bands.) This is not surprising since we know that NOX levels influence O3 levels instantaneously: there may be little influence at time lags other than lag 0. We will continue to consider the relationship between the pollutants, as well as the influence of other variables, throughout this analysis.

The AR(2) coefficients are estimated using the Yule-Walker equations and the procedures outlined in Section 2.3: the estimates are

\[
\hat{A}_1 = \begin{bmatrix}
1.317 & -0.009 \\
0.110 & 1.252
\end{bmatrix}
\]

\[
\hat{A}_2 = \begin{bmatrix}
-0.424 & 0.005 \\
-0.083 & -0.568
\end{bmatrix}
\]

The small values in the off-diagonals again show only a slight influence of one variable on the other. This suggests that the bivariate analysis may not add much extra information in this particular application. So, the diagonal elements of \( \hat{A}_1 \) and \( \hat{A}_2 \) have the greatest effect on the estimate of \( \mathbf{S}(s, t) = A_1 \mathbf{S}(s, t-1) + A_2 \mathbf{S}(s, t-2) \). In the estimation of the state vector at time \( t \), the state vector at time \( t-1 \) has mainly positive coefficients, while the state vector at time \( t-2 \) has mainly negative coefficients.

Next, the spatial covariance parameters are estimated through the steps described in Section 2.3. Recall that the variogram is defined as

\[
\text{var}[Z(s, t) - Z(r, t)],
\]

which simplifies to

\[
2\text{cov}[\bar{S}(s, t)] + 2\text{cov}[\bar{e}(s, t)] - 2\text{cov}[\bar{S}(s, t), \bar{S}(r, t)].
\]

The first covariance term is estimated directly from the data using the autocovariance function estimator given in Equation 2.16. The second term is zero, as estab-
lished in Section 2.3. The final covariance term is obtained from Equation 2.17. We use the two-parameter exponential model for $C_{nj}$ defined in Equation 2.15.

With $\hat{A}_1$, $\hat{A}_2$, and a model for $C_{nj}$ for $j = 1, 2$, we now have all of the information for estimation of the variogram. In order to estimate $\theta_{j1}$ and $\theta_{j2}$, this variogram estimate is fitted to an aggregate empirical variogram. An empirical variogram is computed for each hour and then the median values at each distance (across time) are used to make up the empirical variogram.

For computational reasons, a univariate approach is used for the variogram fitting and parameter estimation. The empirical semi-variograms ($\frac{1}{2}$) and corresponding fitted semi-variograms are plotted in Figure 4.4. The $O_3$ variogram shows that the fit works well locally, i.e. for shorter distances. For $NO_X$ the best fit involved assuming a constant spatial covariance, resulting in the constant variogram.

![Variogram fits for the 2 time series. The first plot is the $O_3$ semi-variogram and the second is $NO_X$.](image)

Figure 4.4 Variogram fits for the 2 time series. The first plot is the $O_3$ semi-variogram and the second is $NO_X$.

Kalman filtering has not yet been applied in this case. Instead, predictions are calculated with a quicker approach, using the temporal AR(2) relationship. Predictions at time $t$ and any location $s$ are based on data from the two previous
time points:

\[ \tilde{S}(s, t) = \hat{A}_1 \tilde{S}(s, t - 1) + \hat{A}_2 \tilde{S}(s, t - 2) \]

Kriging is used for spatial interpolation, so \( \tilde{S}(s, t - 1) \) and \( \tilde{S}(s, t - 2) \) are each obtained through ordinary kriging, a purely spatial predictor. So, for example, only observations from time \( t - 1 \) are used to obtain \( \tilde{S}(s, t - 1) \), and it is not necessary to have observations at the location \( s \) to get this estimate. In other words, the spatial prediction is separate from the temporal prediction in this preliminary analysis.

### 4.2 Results

Consider prediction of the spatial surface at time \( t \). Observations from the 2 previous time points are needed to predict an AR(2) realization. First, kriging predictions are obtained at times \( t - 1 \) and \( t - 2 \). These kriged values are then used in the bivariate AR model to estimate the value of the process at time \( t \).

In this section, we will analyze the preliminary predictions two ways. First, we look at prediction of the entire spatial surface for a given hour. Second, we perform a cross-validation exercise in which the state vector is predicted at one site when observations from that site are not used. The predictions are compared to the observations. In all cases the predictions are calculated in the transformed space. Then the transformation is reversed so that the plots are in the original units of parts per billion (ppb).

#### 4.2.1 Predicted Surfaces

The first plot in Figure 4.5 shows contours of the predicted \( O_3 \) surface at 4 p.m. on Monday, September 8, 1997. The surface is a result of both kriging and AR(2) prediction. Observations at times \( t - 1 \) and \( t - 2 \) are used in kriging to estimate
\( S(s, t - 1) \) and \( S(s, t - 2) \) for every \( s \) in the spatial grid. Then \( \hat{S}(s, t) \) is calculated for each \( s \). The result is a surface for \( O_3 \), shown in Figure 4.5, and a surface for \( NO_X \), shown in Figure 4.6. The kriged surface of the observed values at this time \( t \) (not shown) reveals two local maxima in the northeast area, where the contour plot in Figure 4.5 shows one maximum.

Now consider the difference between the surfaces for the spatial-temporal prediction and the purely spatial prediction. The differences are calculated by subtracting the spatial-temporal prediction from the spatial prediction at each location on the spatial grid. The second plot in Figure 4.5 depicts the standardized differences between the purely spatial prediction of \( O_3 \) levels and the spatial-temporal prediction for \( O_3 \). The plot is “standardized” in the sense that the the differences are divided by the standard deviation of the differences. Even though this calculation reduces the variability in the plot of differences, we still get a picture of some general characteristics of this set of differences.

The elevated differences in the northeast region are in the same general spatial area as the high ozone levels. Much of this variation is due to the difference in a unimodal prediction surface (spatial-temporal model) and a bimodal prediction surface (spatial model). Differences along the boundary may be attributed to lack of sufficient data in this region.

Results of the predicted \( NO_X \) surface at the same time point are shown in Figure 4.6. In the plot of standardized differences, all of the differences are positive. Thus the spatial prediction is never less than the spatial-temporal prediction in this particular case. We have no confidence in \( NO_X \) predictions in the southern part of the Houston region or the lower part of the plot, due to the absence of any nearby observations.
Figure 4.5  Predicted $O_3$ and standardized differences at one time point. The differences are calculated by subtracting the spatial-temporal prediction from the spatial prediction at each location in space.
Figure 4.6  Predicted $NO_X$ and standardized differences at one time point. The differences are calculated by subtracting the spatial-temporal prediction from the spatial prediction at each location in space.
4.2.2 Cross-Validation

A preliminary evaluation of these estimation methods is carried out through cross-validation. The cross-validation exercise is performed by removing data from one site and predicting at that site for a 24 hour period. The predicted time series is then compared to the observed series. Each site is checked in this way. As we would expect, the $O_3$ predictions are generally better than $NO_X$. (Only 7 spatial sites for $NO_X$ are available.)

Results of the cross-validation at Site D are shown in Figure 4.7. This monitoring station is located north of the downtown area. The observed $O_3$ and $NO_X$ levels are indicated by solid lines. The predicted levels are indicated by the dashed lines. The observed series generally falls within one standard deviation of the predicted series, although the predictor does tend to overestimate observed $O_3$ levels. The predictions of $NO_X$ at this site are also slightly overestimating the observed levels. Figure 4.8 suggests that the predictions at Site J are more successful. Cross-validation results for other sites are shown in Figure 4.9 and Figure 4.10.

Summary statistics for the cross-validation exercise are listed in Table 4.1. The statistics for $O_3$ include cross-validation at all 21 sites; the $NO_X$ statistics include cross-validation at all 7 $NO_X$ sites. The mean absolute deviation (MAD) between the predicted and the observed is calculated as well as the 10% trimmed mean of the absolute deviations (10% MAD). The square root of the mean squared error (RMSE) is also reported. For the twenty-four hour period encompassing our cross-validation exercise, the MAD for the $O_3$ predictions is 9.18 ppb whereas $O_3$ levels for this period range from 0 ppb to 150 ppb. Clearly, we are doing a good job of capturing the behavior of the ozone field. Our MAD for the $NO_X$ predictions, 17.28 ppb, is not as encouraging, however, in that it is almost twice that for the $O_3$
Figure 4.7 Cross-validation results for $O_3$ and $NO_x$ at Site D. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation above and below the predicted values. The results are shown in the original units, ppb.
Figure 4.8  Cross-validation results for $O_3$ and $NO_x$ at Site J. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation above and below the predicted values. The results are shown in the original units, ppb.
Figure 4.9 Cross-validation results for O₃ at Site E. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation. The results are shown in the original units, ppb.
Figure 4.10 Cross-validation results for $O_3$ at Site U. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation. The results are shown in the original units, ppb.
predictions. The $NO_X$ levels range from 0 ppb to 180 ppb during this same period. With seven spatial monitoring sites for $NO_X$ we expect elevated prediction errors.

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>10% MAD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_3$</td>
<td>9.18</td>
<td>7.95</td>
<td>13.36</td>
</tr>
<tr>
<td>$NO_X$</td>
<td>17.28</td>
<td>14.37</td>
<td>26.83</td>
</tr>
</tbody>
</table>

**Table 4.1** Cross-validation statistics in units of ppb.
Chapter 5

Model Application and Results

This chapter presents the results of prediction through Kalman filtering in the air quality application. The first section explains the application of the bivariate spatial-temporal model with the Kalman filter. The subsequent sections in this chapter present our evaluation of the application, including surface predictions, cross-validation exercises, and comparison to the performance of the earlier analysis. As in Chapter 4, the transformed observations are used in all parameter estimation and prediction. For the plots, though, the transformations are reversed and the predictions are displayed in units of ppb.

5.1 Model Application

To use Kalman filtering in this air quality application, we build on the analysis in Section 4.1. where the bivariate AR(2) model is selected. The Kalman filter algorithm is written in the S-Plus computer language. The main program is in Appendix B. A template for the input is also given in Appendix B. The input for the algorithm includes initial conditions for the state vector and covariances, past observations, and parameter estimates. The program is written under the assumption that both variables are observed at each location. Thus, except where noted, only observations from the seven bivariate monitoring stations are used in the prediction.

The set of transformed observations described for the preliminary analysis in Chapter 4 is used here also. In Figure 5.1, the seven sites marked by letters are the
sites used in this analysis. The dashed lines define the subset of the original grid that is used for building a surface of Kalman filter predictions in Section 5.2. In this application, eighteen time points (hours) are always used to make up the set of past observations. This number was chosen by checking the performance of the algorithm: including additional past observations does not affect the predictions.

**Figure 5.1** Map of Houston-area monitoring stations. The letters represent the sites at which both pollutants are observed. The other sites, not used in this part of the analysis, are also marked. The dashed line marks the region for the predicted surfaces.

The initial conditions are chosen with consideration for computational simplicity and for available observations. Because the observations are standardized, assigning zero to the initial state vector is a logical choice. And the initial variances are
assigned the value one and the covariances are assigned zero. This choice works
well for computational reasons. And assigning other values to the covariance matrix
does not have much of an effect on the final outcome; the updating structure is more
influential than the starting values in this AR(2) case.

The MLE's for the spatial covariance model parameters are obtained through
the Kalman filter, as described in Section 2.4. These parameters from Equation 2.15
are \( \theta_{j1} \) and \( \theta_{j2} \), for \( j = 1, 2 \), where \( j = 1 \) represents \( O_3 \) and \( j = 2 \) represents \( N O X \).
The computational complexity of the Kalman filtering program is substantial, and
the minimization of the log-likelihood is time consuming. So, for practical reasons
such as the limitations of S-Plus, we retain the preliminary estimates for the other
parameters, namely \( A_1 \) and \( A_2 \); these estimates are given in Section 4.1.

Recall that the objective function is written so that we can use an S-Plus mini-
mization routine. The negative log-likelihood function is easily calculated by using
the Kalman filter program in Appendix B. Only a few additions to this program
are necessary for calculation of the objective function. Starting values for the mini-
mization are taken from Section 4.1, the preliminary estimates of \( \theta_{j1} \) and \( \theta_{j2} \) for
\( j = 1, 2 \).

Ideally we would like to use all available observations to calculate the MLE's.
This could not be done now, because the computational complexity was too great.
In order to make the computational problem manageable, we first limited the data
set down to a period of seven days. Further size reductions were needed, so 2
sites (E and B) were removed from the set of observations. These 2 sites are
the most outlying locations. Thus elimination of these sites excludes the largest
spatial distance from the set of distances. Because the variograms in the preliminary
analysis (Figure 4.4) flatten out after short distances, we chose to remove the 2 sites
which would eliminate the largest distance.
Finally, with this reduced set of hourly observations from five sites over seven days, the negative log-likelihood function is minimized. The MLE's for the covariance model parameters $\{\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}\}$ are $\{0.70, 3.25, 0.035, 1.00\}$.

5.2 Predicted Surfaces

One way to assess the prediction algorithm is to look at the series of predictions across space and at one time point. Here we will predict over the smaller region indicated in Figure 5.1 and for the same Monday afternoon hour used in Section 4.2.1. Figure 5.2 shows the predicted $O_3$ surface for one hour. This predicted surface is calculated from the one-step-ahead predictor, which means that past observations are used in the calculation.

This surface can be compared to one created by current observations. Observations from the seven sites at the current time are kriged. Because the seven bivariate sites are used in spatial-temporal prediction, only those seven sites are used in the kriging, for comparison purposes. The purely spatial kriging predictor provides spatial interpolation across the grid. The covariance functions used in the kriging are based on the covariance modeling in Section 2.2 and the parameter estimation in Section 2.4.

Figure 5.3 is a contour plot of the standardized differences between the kriged observations and the spatial-temporal predictions for $O_3$. The differences are calculated by subtracting the spatial-temporal prediction from the spatial prediction at each location across the entire spatial grid on which the surfaces are plotted. For the plot, the differences are divided by the standard deviation of the differences. The negative numbers in the lower right part of the plot indicate a region where the kriged surface dips noticeably and the spatial-temporal surface is smooth.
The $O_3$ differences, without dividing by the standard deviation, range from -31 ppb to 12 ppb. The mean difference is -9.6 ppb. Less than 6% of the differences are positive; so in general, the spatial-temporal predicted surface has slightly higher levels than the kriged surface.

The spatial-temporal predictions for $NO_X$ at the same hour are shown in Figure 5.4, and the contour plot of the standardized differences is in Figure 5.5. The lower half of the contour plot indicates a region where the kriged surface has a higher peak. The spatial-temporal surface tends to have higher levels in the upper half of the plot. The range of differences (without removing the standard deviation) is smaller for $NO_X$, from -21 ppb to 13 ppb. The mean difference is -9.1 ppb. Again, only a small percentage of the differences are positive.

The surfaces of predictions suggest that the Kalman recursions and one-step-ahead predictor work well. This predictor relies on observations only from the past. And the resulting predictions are not too different from the purely spatial predictions, which are based on current observed levels.

### 5.3 Cross-Validation

Cross-validation is performed at each of the 7 sites where both pollutants are observed. For each site, all observations from the site are removed. The levels of the two pollutants are predicted at the site based on observations from the other 6 sites. The predicted levels are then compared to the observed. In this case, predictions are calculated for a 24-hour period on a Monday.

Results of the cross-validation at Site F and Site G are shown in Figure 5.6 and Figure 5.7, respectively. In each plot, the solid line represents the observations and the dashed line represents the predictions. The $O_3$ predictions at both sites are very close to the observed values. The $NO_X$ predictions at Site G tend to
Figure 5.2  Predicted $O_3$ surface at one time point. The results are shown in the original units. ppb.
Figure 5.3 Contour plot of standardized differences between the kriged $O_3$ observations and the spatial-temporal predictions at one time point.
Figure 5.4  Predicted $NO_x$ surface at one time point. The results are shown in the original units, ppb.
Figure 5.5  Contour plot of standardized differences between the kriged $NO_X$ observations and the spatial-temporal predictions at one time point.
underestimate the observed \( \text{NO}_X \) levels. The \( \text{NO}_X \) predictions at Site F are close to the observations, except for a 4-hour peak in the observed levels that the predictor doesn’t catch.

Prediction of \( O_3 \) levels at Site E, shown in Figure 5.8, show some improvement over the preliminary predictions in Figure 4.9. Site E is the far Northwest location; there are no other monitoring sites nearby. While almost all the observations fall within one standard deviation of the MMSE predictions at this site, only about two-thirds of the observations fall within one standard deviation of the preliminary predictions.

These results are compared to prediction and cross-validation under two other scenarios. First, implementing kriging as a spatial interpolator allows us to use observations from a time \( t \) to form predictions at any location for that time \( t \). Again for cross-validation, all observations at a site are removed. Then at each hour, observations from the remaining 6 sites are kriged, and the levels of the two pollutants at the removed site are predicted based on the kriging. The covariance functions used in the kriging are based on the covariance modeling in Section 2.2. Recall that the spatial covariance functions for each pollutant is \( C^\text{nu}(s, r) = \sigma_j \exp\left\{-\sigma_{j2}\|s - r\|\right\} \) for \( j = 1, 2 \). Here the estimates of the covariance parameters are taken from the preliminary estimation first described in Section 2.3. Second, the predictions from the state-space model and Kalman filter are compared to the preliminary analysis in Section 4.2.2. Note that the earlier analysis does include \( O_3 \) observations from all 21 sites.

Table 5.1 and Table 5.2 list the summary statistics for the three different sets of predictions of \( O_3 \) and \( \text{NO}_X \). All three cross-validations compare predicted and observed levels for the same 24-hour period. Predictions and observations at all seven sites for the 24 hours are combined to calculate the mean absolute deviation
Figure 5.6 Cross-validation results for $O_3$ and $NO_X$ at Site F on Monday. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation above and below the predicted values. The results are shown in the original units, ppb.
Figure 5.7  Cross-validation results for $O_3$ and $NO_X$ at Site G on Monday. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation above and below the predicted values. The results are shown in the original units, ppb.
Figure 5.8  Cross-validation results for $O_3$ at Site E. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. Vertical lines indicate one standard deviation. The results are shown in the original units, ppb.
(MAD) between the predicted and observed values, the 10% trimmed MAD and the root mean squared error (RMSE).

The spatial-temporal prediction through the Kalman filter generally shows a slight improvement over the purely spatial (kriging) prediction and the simpler spatial-temporal prediction. The MAD for the $O_3$ predictions is 8.44 ppb for the Kalman predictions, 9.10 ppb for the kriging predictions and 9.18 ppb from the preliminary analysis. The range of $O_3$ observations during this 24-hour period is 0 ppb to 134 ppb. Given the large range of the observations, a MAD of 8.44 ppb is good. The preliminary analysis suggested that the $NO_X$ predictions would not work as well as those for $O_3$. That is certainly true in this case. The $NO_X$ MAD of 17.32 ppb and RMSE of 29.31 more than double the $O_3$ statistics. The $NO_X$ observations during this time period range from 0 ppb to 180 ppb.

\[
\begin{array}{|c|c|c|}
\hline
&MAD&10\%MAD&RMSE \\
\hline
\text{Kalman}&8.44&7.25&12.90 \\
\text{Kriging}&9.10&7.76&13.73 \\
\text{Preliminary}&9.18&7.95&13.36 \\
\hline
\end{array}
\]

*Table 5.1* Cross-validation statistics in ppb for $O_3$ on Monday.

\[
\begin{array}{|c|c|c|}
\hline
&MAD&10\%MAD&RMSE \\
\hline
\text{Kalman}&17.32&13.52&29.31 \\
\text{Kriging}&17.62&14.94&26.16 \\
\text{Preliminary}&17.28&14.37&26.83 \\
\hline
\end{array}
\]

*Table 5.2* Cross-validation statistics in ppb for $NO_X$ on Monday.

To get a bigger picture of the performance of the MMSE predictors, we look at a week-long cross-validation. The 7-day period starts with the Monday used in
the cross-validation above and runs through the following Sunday. The week-long results for 3 sites are depicted in Figure 5.9 through Figure 5.11. These plots show that the characteristics of the 1-day predictions are carried through the entire week: that is, the $O_3$ predictions are generally good, while the $NO_X$ predictions do not turn out so well.

Summary statistics for the longer cross-validation are presented in Table 5.3. Only the MMSE predictors are calculated in this exercise. The $O_3$ statistics are slightly higher than the statistics for the predictions for the one-day cross-validation. This suggests that the performance of the predictor for that one day (Monday) is better than the overall performance for the week. On the other hand, the $NO_X$ statistics in Table 5.3 are lower than the statistics for the Monday-only predictions. The RMSE, for instance, is 24.81 ppb, which is compared to 29.31 ppb in Table 5.2. Even though we still see that $NO_X$ predictions are somewhat disappointing, the results of the longer cross-validation show that the general performance of the $NO_X$ predictors is not as poor as the earlier predictions suggest.

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>10% MAD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_3$</td>
<td>9.91</td>
<td>8.79</td>
<td>13.99</td>
</tr>
<tr>
<td>$NO_X$</td>
<td>15.26</td>
<td>12.22</td>
<td>24.81</td>
</tr>
</tbody>
</table>

**Table 5.3** Cross-validation statistics for $O_3$ and $NO_X$ for one week.

In summary, we see from the results in this chapter that the bivariate spatial-temporal model shows promise. The $O_3$ predictions have very reasonable error rates. Improvements to the model, though, should include an improvement in the accuracy of the $NO_X$ prediction. The large errors in the $NO_X$ results may be explained in part by the transformation of the original observations. Failing to capture the temporal behavior of $NO_X$ may have lead to difficulties throughout the
Figure 5.9  Cross-validation results for $O_3$ and $NO_x$ at Site D. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. The results are shown in the original units, ppb.
Figure 5.10  Cross-validation results for $O_3$ and $NO_x$ at Site F. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. The results are shown in the original units, ppb.
Figure 5.11 Cross-validation results for $O_3$ and \textit{NO}_X at Site G. The solid line represents the observed values and the dashed line represents the values predicted in the cross-validation. The results are shown in the original units, ppb.
analysis. An example of such temporal behavior is the effect of increased traffic on $NO_X$ levels, which was mentioned in Chapter 3. In retrospect, we realize that giving more attention to $NO_X$ behavior and to detrending the $NO_X$ observations could lead to better predictions under the given model. Further discussion of these results and of future work us included in the following chapter.
Chapter 6

Conclusions and Other Issues

The modeling effort presented here is a good foundation for the development of multivariate spatial-temporal models for air quality applications. We have developed a new model which includes the spatial and temporal components of multiple variables. We have demonstrated that including all components in the model is feasible, although difficult. Because the modeling is complicated, some simplification is necessary for the sake of practical limitations. Yet we have shown that meaningful results can be achieved without the great over-simplification that is often seen in similar studies.

The value of this study extends beyond the prediction results we presented here. This model is an excellent starting point for development in spatial-temporal modeling as well as for more advanced analysis of air quality data. In this chapter, we discuss some of the issues surrounding further model development. Then we suggest other uses beyond prediction for this modeling approach in air quality studies.

First, in the model development, more research may be devoted to covariance issues. Recall that we chose to model the spatial covariance rather than estimate it directly. In a future study, a comparison of our work to an application with an estimated covariance would be interesting. An additional issue with covariances is the direct modeling of the spatial-temporal interaction between pollutants.

Next, we learned quite a bit from the actual application of the air pollutant data. In this study we were not very successful in $NO_X$ prediction. And for the $O_3$ prediction, it is not clear that the presence of the $NO_X$ variable contributed much to the $O_3$ predictions: additional trend analysis for $NO_X$ may mitigate this issue
We also learned through this air quality study that the ratio of $NO_2$ to $NO$ has a stronger correlation with $O_3$ than does $NO_X$. We originally chose $NO_X$ based on the chemical reactions that form $O_3$, past research, and available data. Observed $NO_X$ is the sum of observed $NO$ and $NO_2$. This new information about the ratio of $NO_2$ to $NO$ may be useful for future air quality modeling efforts.

Even though we can find many ways to continue development of this model, we also should consider extended uses for the modeling. For instance, improved prediction could be helpful in the area of outlier detection. Outlier detection is used to determine unusual observations or "extreme events," and it may be used to determine entire sites which are behaving unusually. Exercises similar to the cross-validation in Chapter 5 would be applicable to outlier detection problems. A second use in the air quality field is the assessment of personal exposure to pollutants. In this case, predicted pollutant levels are used to estimate exposure. Naturally, improved predictions could improve the assessment of personal exposure. Thus, the model presented here provides a foundation for future model development and analysis of multivariate spatial-temporal processes, particularly in air quality research.
Appendix A

Notation

This appendix contains a summary of the notation found in Section 2.1.2, which outlines prediction under the bivariate spatial-temporal model. The small numbers in parentheses after each equation indicate the size of the vector or matrix.

\[ \hat{Z}(s, t_0 | t) \equiv E[Z(s, t_0) | Z(1), \ldots, Z(t)]_{(2 \times 1)} \]

\[ \hat{Z}(t_0 | t) \equiv E[Z(t_0) | Z(1), \ldots, Z(t)]_{(2n \times 1)} \]

\[ \Sigma_{1,2}(s, r) \equiv \text{cov}[S(s, t_1), S(r, t_2) | Z(1), \ldots, Z(t)]_{(2 \times 2)} \]

\[ \Sigma_{1,0}(s, r) \equiv \text{cov}[S(s, t_0), S(r, t_0) | Z(1), \ldots, Z(t)]_{(2 \times 2)} \]

\[ \Sigma_{0,1}(s, r) \equiv \text{cov} \left[ \begin{pmatrix} S(s, t_0 - p + 1) \\ \vdots \\ S(s, t_0) \end{pmatrix}, \begin{pmatrix} S(r, t_0 - p + 1) \\ \vdots \\ S(r, t_0) \end{pmatrix} | Z(1), \ldots, Z(t) \right]_{(2p \times 2p)} \]

\[ F(t) \equiv \text{cov}[Z(t) | Z(1), \ldots, Z(t - 1)]_{(2n \times 2n)} \]

\[ G(s, t) \equiv \text{cov}[Z(t), (S(s, t - p + 1)^T, \ldots, S(s, t)^T) | Z(1), \ldots, Z(t - 1)]_{(2n \times 2p)} \]

\[ Q \equiv \begin{pmatrix} 0 & 0 \\ \vdots & I_{p-2} \\ 0 & v \\ A_p & A_{p-1} & \cdots & A_1 \end{pmatrix}_{(2p \times 2p)} \]

\[ Q[a : b] \equiv \text{subset of } Q, \text{ from row } a \text{ through row } b \]

\[ A_* \equiv [A_p \cdots A_1]_{(2 \times 2p)} \]
Appendix B

S-Plus Programs

B.1 Kalman Filter

# This program runs through the Kalman algorithm.
#
# INPUTS: See the input file
#
# OUTPUTS: Vector (length 4) of state estimates & cov matrix (4x4)
#

kf.predict <- function(...){
  #------
  # Initializing
  G1.eqn <- NULL; G2.eqn <- NULL; N.incr <- NULL
  F.eqn <- NULL; F.inv <- NULL
  ti <- NULL; ind1<- NULL; ind2 <- NULL
  t.last <- ncol(observ)

  #------
  # Determine whether the site of interest ("site") is already in our
  # list of sites by checking the matrix of x-y locations. If it is,
  # newsite is False and tab1 tells us where in the list "site" is.
  # Establish site2 (called "r" in notes) location in the list of sites.
  # Assume site2 comes from given list.

  tab1 <- NULL
  tab2 <- NULL
  newsite <- T
  for (i in 1:ns) {
    if(site[1] == sites.xy[i,1] && site[2] == sites.xy[i,2]) {
      newsite<-F
      tab1 <- i
    }
    if(site[2,1] == sites.xy[i,1] && site2[2] == sites.xy[i,2]) {
      tab2 <- i
    }
if( !is.null(tab1) & !is.null(tab2)){
    break
    NULL }

#-----
#Assign input values to variables used in the loop
Spred.array <- s.initial.array
sig.array.incr <- sig.initial.array

if(newsite) { Spred <- s.initial
    Vpred <- sig.initial
    N.incr <- newsite.array }
else { Spred <- s.initial.array[, , tab1]
    Vpred <- sig.initial.array[, , (tab1+ns*(tab2-1))]
}

#-----
#Prepare error covariance matrix so the dimensions are correct
#for calculations in the loop.
bigsigep <- matrix(0, ncol=ns*2, nrow=ns*2)
seq1 <- seq(1, (2*ns-1), by=2)
seq2 <- seq(2, 2*ns, by=2)
diag(bigsigep) <- rep(diag(sigep), ns)
bigsigep[cbind(seq1, seq2)] <- sigep[1, 2]
bigsigep[cbind(seq2, seq1)] <- sigep[2, 1]

#Prepare spatial covariance
d <- as.numeric(dist(rbind(site, site2)))
ceta <- ceta.fun(d)
cmat <- rbind(matrix(0, 2, 4), cbind(matrix(0, 2, 2), ceta))

d.array <- array(0, dim=ns*ns)
if(newsite) newd <- array(0, dim=ns)
for(k in 1:ns){
    if(newsite)
        newd[k] <- as.numeric(dist(rbind(sites.xy[k, ], site)))
    for(j in 1:ns){
        d.array[(j+ns*(k-1))] <-
            as.numeric(dist(sites.xy[c(j, k), ]))
        NULL }

ceta.array <- array(apply(d.array, 1, ceta.fun), dim=c(2, 2, ns*ns))
cmat.array <- array(apply(ceta.array, 3, func), dim=c(4, 4, ns*ns))
if(newsite) { newceta <- array(apply(newd,1,ceta.fun),dim=c(2,2,ns))
    newcmat <- array(apply(newceta,3,func),dim=c(4,4,ns))}

# Other sequences defined here to be used in the loop.
Gseqs <- cbind(seq(1,((ns-1)*ns+1),by=ns),seq(ns,ns*ns,by=ns))
seqns <- as.array(1:ns)

# For each run through the loop, all of the updating equations
# are calculated.
#-------------------
# Start the loop:
#-------------------

for(ti in 1:t.last){

# 1st updating equation in Appendix B
s.incr <- Q *%*% Spred

# 3rd updating equation in Appendix B
sig.incr <- Q *%*% Vpred *%*% t(Q) + cmat

# 2nd updating equation in Appendix B
Zhat <- as.vector(apply(Spred.array,3,funz))

# 4th updating equation in Appendix B
F.eqn <- Fmat(ns,sig.array.incr) + bigsigep
F.inv <- solve(F.eqn)

# 5th updating equation in Appendix B
# First calculate G for all of the sites.
# Then pick out appropriate G(s) to match site and site2.
# Calculate Gmat for site1 if it is a newsite.

G.array <- array(apply(Gseqs,1,funG,garray=sig.array.incr),
    dim=c(2*ns,4,ns))

if(newsite) G1.eqn <- Gmat(N.incr)
else G1.eqn <- G.array[,,tab1]

G2.eqn <- G.array[,,tab2]

# Define Zt, the vector of observations at time t
Zt <- obsmat[,ti]

#The 2 main predicting equations, see Chapter 3
Spred <- s.incr + t(G1.eqn) %*% F.inv %*% (Zt - Zhat)
Vpred <- sig.incr - t(G1.eqn) %*% F.inv %*% G2.eqn

#Increment/update arrays for the next time through the loop
Spred.array <- array(
  apply(seqns,1,funs,Spred.array,G.array,F.inv,Zt,Zhat),
  dim=c(4,1,ns))
prev.sigma <- sig.array.incr
if(newsite) prev.N <- N.incr
for(ind2 in 1:ns){
  if(newsite){ N.incr[,,ind2] <-
    ( Q%*%prev.N[,,ind2]%*%t(Q) + newcma[,,ind2] -
      t(G1.eqn) %*% F.inv %*% G.array[,,ind2])}
  for(ind1 in 1:ns){
    sig.array.incr[,,(ind1+ns*(ind2-1))] <-
      (Q%*%prev.sigma[,,(ind1+ns*(ind2-1))]%*%t(Q) +
       cma.array[,,(ind1+ns*(ind2-1))]
      - t(G.array[,,ind1]) %*% F.inv %*% G.array[,,ind2] )
    NULL}
  NULL}
NULL
}
#-------------------
#End the loop.
#-------------------

list(S.hat = Spred, cov.mat = Vpred)
}

B.2 Input File

#Set up initial conditions here.
#Completing this file gives all the input needed for the main program.

ns <- #number of sites
sites.xy <- #ns x 2 matrix of location coordinates;
#column1=x column2=y
site <- #vector of length 2: (x,y); site of interest
site2 <- #vector of length 2: (x,y); in sites.xy
s.initial.array <- #4x1xns array of initial states, for each site
sig.initial.array <- #array(covdata,dim=c(4,4,ns*ns))
#array of all initial 4x4 sigma for all
#location pairs; order matrices as
#(s1,s1),...(sn,s1),(s1,s2),...(sn,sn)

#----------------------------------------
#---The next 3 are necessary only when "site" is not in the list of
#---sites.xy (observation sites)
#vector of length 4: (var1,var2) at t=0 & t=-1
s.initial <-
sig.initial <- #4x4 covariance matrix assoc. with s.initial
#----------------------------------------

newsite.array <- #array(somedata,dim=c(4,4,ns)
#Array of 4x4 matrices, order as
#(s1,site),...(sn,site)

#----------------------------------------
A1 <- #2x2 matrix of vector AR(2) coefficients
A2 <- #2x2 matrix
Astar <- cbind(A2,A1)
Q <- rbind(cbind(matrix(0,2,2),diag(c(1,1))),Astar)

ceta.fun <- function(d){
  #Fill in model; defined outside;
  #fcn of distance; should output a 2x2 matrix
  
sigep <- #2x2 matrix of measurement error covariance

  obsmat <- #2n x t matrix of observations

B.3 Other Functions

#This file contains the smaller functions called
#in the kfpredict function.

#----------------------------------------
#F(t)
funx <- function(x){ Astar*x*t(Astar) }

Fmat <- function(ns, sig.array.incr){
  M <- sig.array.incr
  Mprime <- apply(M,3,funx)
  Mprime <- array(Mprime,dim=c(2,2,ns*ns))
  hold <- NULL; hold2 <- NULL
  for(i in 1:ns){
    hold <- c(Mprime[,, (1 + (i-1)*ns):(i*ns)])
    hold2 <- cbind(hold2,matrix(hold,byrow=T,ncol=2))
    NULL }
  return(hold2)
}

#G(s,t)
fun1 <- function(x){Astar*x*t(Q[1:2,])}
fun2 <- function(x){Astar*x*t(Q[3:4,])}

Gmat <- function(N){
  Na <- apply(N,3,fun1)
  Nb <- apply(N,3,fun2)
  final <- cbind(as.vector(Na[1:2,]),as.vector(Na[3:4,]),as.vector(Nb[1:2,]),as.vector(Nb[3:4,]))
  return(final)
}

#other
funsig <- function(x,cmat,G1.eqn,F.inv,G2.eqn){
  Q*x*t(Q) + cmat - t(G1.eqn)*F.inv*G2.eqn
}
funG <- function(x,garray){ Gmat(garray[,]x[1]:x[2]) }
funs <- function(seqns,Spred.array,G.array,F.inv,Zt,Zhat){
  (Q %*% Spred.array[,]seqns) + t(G.array[,]seqns)*F.inv*(Zt-Zhat))
}
func <- function(x){rbind(matrix(0,2,4),cbind(matrix(0,2,2),x))}

#------------------------
Bibliography


