MODELING AND COMPUTATIONS OF MULTIVARIATE DATASETS IN SPACE AND TIME

by

SAMUEL SETH DEMEL

B.S., Sam Houston State University, 2006
M.S., Sam Houston State University, 2008

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2013

Abstract

Spatio-temporal and/or multivariate dependence naturally occur in datasets obtained in various disciplines; such as atmospheric sciences, meteorology, engineering and agriculture. There is a great deal of need to effectively model the complex dependence and correlated structure exhibited in these datasets. For this purpose, this dissertation studies methods and application of the spatio-temporal modeling and multivariate computation.

First, a collection of spatio-temporal functions is proposed to model spatio-temporal processes which are continuous in space and discrete over time. Theoretically, we derived the necessary and sufficient conditions to ensure the model validity. On the other hand, the possibility of taking the advantage of well-established time series and spatial statistics tools makes it relatively easy to identify and fit the proposed model in practice. The spatio-temporal models with some ARMA discrete temporal margin are fitted to Kansas precipitation and Irish wind datasets for estimation or prediction, and compared with some general existing parametric models in terms of likelihood and mean squared prediction error.

Second, to deal with the immense computational burden of statistical inference for multiple attributes recorded at a large number of locations, we develop Wendland-type compactly supported covariance matrix function models and propose multivariate covariance tapering technique with those functions for computation reduction. Simulation studies and US temperature data are used to illustrate applications of the proposed multivariate tapering and computational gain in spatial cokriging.

Finally, to study the impact of weather change on corn yield in Kansas, we develop a spatial functional linear regression model accounting for the fact that weather data were recorded daily or hourly as opposed to the yearly crop yield data and the underlying spatial autocorrelation. The parameter function is estimated under the functional data analysis framework and its characteristics are investigated to show the influential factor and critical period of weather change dictating crop yield during the growing season.

MODELING AND COMPUTATIONS OF MULTIVARIATE DATASETS IN SPACE AND TIME

by

SAMUEL SETH DEMEL

B.S., Sam Houston State University, 2006
M.S., Sam Houston State University, 2008

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2013

Approved by:

Major Professor
Dr. Juan Du

Copyright

Abstract

Spatio-temporal and/or multivariate dependence naturally occur in datasets obtained in various disciplines; such as atmospheric sciences, meteorology, engineering and agriculture. There is a great deal of need to effectively model the complex dependence and correlated structure exhibited in these datasets. For this purpose, this dissertation studies methods and application of the spatio-temporal modeling and multivariate computation.

First, a collection of spatio-temporal functions is proposed to model spatio-temporal processes which are continuous in space and discrete over time. Theoretically, we derived the necessary and sufficient conditions to ensure the model validity. On the other hand, the possibility of taking the advantage of well-established time series and spatial statistics tools makes it relatively easy to identify and fit the proposed model in practice. The spatio-temporal models with some ARMA discrete temporal margin are fitted to Kansas precipitation and Irish wind datasets for estimation or prediction, and compared with some general existing parametric models in terms of likelihood and mean squared prediction error.

Second, to deal with the immense computational burden of statistical inference for multiple attributes recorded at a large number of locations, we develop Wendland-type compactly supported covariance matrix function models and propose multivariate covariance tapering technique with those functions for computation reduction. Simulation studies and US temperature data are used to illustrate applications of the proposed multivariate tapering and computational gain in spatial cokriging.

Finally, to study the impact of weather change on corn yield in Kansas, we develop a spatial functional linear regression model accounting for the fact that weather data were recorded daily or hourly as opposed to the yearly crop yield data and the underlying spatial autocorrelation. The parameter function is estimated under the functional data analysis framework and its characteristics are investigated to show the influential factor and critical period of weather change dictating crop yield during the growing season.

Table of Contents

Table of Contents vi

List of Figures ix

List of Tables xi

Acknowledgements xiii

Dedication xiv

Preface xvi

1 Spatio-temporal covariance modeling in continuous space and discrete time 1

1.1 Introduction . . . 1

1.2 Simplifying and validating the space-time process . . . 3

1.3 Moving-average-type temporal margin . . . 5

1.3.1 Moving average simulation study . . . 10

1.3.2 Kansas daily precipitation data . . . 13

1.4 ARMA-type temporal margin . . . 22

1.4.1 Asymmetric Covariance Functions . . . 26

1.5 Irish Wind Data Analysis . . . 27

1.5.1 Symmetric Covariance Models . . . 27

1.5.2 Asymmetric Covariance Models . . . 33

1.6 Chapter 1 Appendix . . . 36

1.6.1 Proof of **Theorem 1.3.1** . . . 36

1.6.2 Proof of **Theorem 1.3.2** . . . 38

1.6.3 Proof of **Theorem 1.4.1** . . . 40

2 Multivariate Tapering **45**

2.1 Introduction . . . 45

2.2 Multivariate tapering . . . 49

2.2.1 Preliminary results of covariance matrix functions . . . 50

2.2.2 Wendland type of compactly supported covariance matrix functions . . . 54

2.3 Bivariate exponential case . . . 59

2.4 Bivariate Matérn case . . . 62

2.5 USA Multivariate Climate Data . . . 70

2.6 Chapter 2 Appendix . . . 75

2.6.1 Proof of Theorem 2.2.1 . . . 75

2.6.2 Proof of Theorem 2.2.2 . . . 75

2.6.3 Proof of Theorem 2.2.5 . . . 76

2.6.4 Proof of Theorem 2.2.6 . . . 76

2.6.5 Proof of Corollary 2.2.3 . . . 76

2.6.6 Proof of Corollary 2.2.4 . . . 76

2.6.7 Additional Bivariate Exponential Results . . . 77

2.6.8 Additional Bivariate Matérn Results . . . 79

3 Functional space-time modeling **87**

3.1 Introduction and Background . . . 87

3.2 Methodology . . . 91

3.2.1 B-Splines . . . 91

3.2.2 Functional linear models for scalar responses . . . 92

3.2.3 The spatial functional linear model . . . 95

3.3 Kansas corn and weather data analysis . . . 98

3.4 Ten year weather patterns . . . 102

Conclusion **112**

Bibliography 118

List of Figures

1.1 Moving Average Plots 11
1.2 Kansas Counties Map 14
1.3 Seasonal Component of Kansas Precipitation 15
1.4 ACFs of Kansas Counties 16
1.5 Kansas Precipitation Marginal Plots 17
1.6 Empirical precipitation space-time correlations 18
1.7 PACF of Ireland wind staitons 29
1.8 Irish Wind marginal plots 30
1.9 Symmetric comparison plots 30
1.10 Asymmetric comparison plots 34

2.1 Location orientation covariance plots 58
2.2 Variable orientation covariance plots 59
2.3 Matérn with Wendland 1 covariance plots 59
2.4 Exponential Askey $\gamma = 0.1$ Graph 61
2.5 Matérn : $\nu_1 = 0.25$, $\nu_2 = 0.25$ and Tapers Graph 65
2.6 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.5$ and Tapers Graph 66
2.7 Matérn : $\nu_1 = 1.5$, $\nu_2 = 1.5$ and Tapers Graph 68
2.8 Matérn : $\nu_1 = 2.5$, $\nu_2 = 2.5$ and Tapers Graph 69
2.9 US Precipitation Taper Compare 72
2.10 US Precipitation Matérn 73
2.11 US Precipitation Askey Tapered 73
2.12 US Precipitation Wendland 1 Tapered 74
2.13 US Precipitation Wendland 2 Tapered 74
2.14 Exponential Askey $\gamma = 0.5$ Graph 77

2.15 Exponential Askey $\gamma = 0.3$ Graph 78
2.16 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.25$ and Tapers Graph 80
2.17 Matérn : $\nu_1 = 0.5$, $\nu_2 = 1.5$ and Tapers Graph 81
2.18 Matérn : $\nu_1 = 0.5$, $\nu_2 = 2.5$ and Tapers Graph 83
2.19 Matérn : $\nu_1 = 1.5$, $\nu_2 = 0.5$ and Tapers Graph 84
2.20 Matérn : $\nu_1 = 1.5$, $\nu_2 = 2.5$ and Tapers Graph 86

3.1 Weather stations of Kansas 88
3.2 Finney County Weather 1990-2011 89
3.3 Corn Yield in Kansas 90
3.4 Yearly Variograms 99
3.5 Precipitation $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$ 100
3.6 Maximum Temperature $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$ 101
3.7 Minimum Temperature $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$ 102
3.8 Kansas Agriculture Districts 103
3.9 Kansas Rural and Urban Areas 103
3.10 Rural Functional Boxplot of Yield 104
3.11 Urban Functional Boxplot of Yield 105
3.12 Rural Functional Boxplot of Precipitation 105
3.13 Urban Functional Boxplot of Precipitation 106
3.14 Rural Functional Boxplot of Maximum Temperature 106
3.15 Urban Functional Boxplot of Maximum Temperature 107
3.16 Rural Functional Boxplot of Minimum Temperature 107
3.17 Urban Functional Boxplot of Minimum Temperature 108
3.18 3D Functional Surface Boxplot of Yield 109
3.19 3D Functional Surface Boxplot of Precipitation 110

List of Tables

1.1 Moving Average Parameter Estimates 12
1.2 Moving Average RMSEs 13
1.3 Likelihood Comparisons 13
1.4 Kansas Precipitations RMSE Statistics 19
1.5 Kansas Precipitation RMSEs for Counties 20
1.6 Kansas Precipitations CRPS Statistics 21
1.7 Kansas Precipitation CRPSs for Counties 21
1.8 Irish Wind Parameter Estimates 31
1.9 Irish Wind: RMSEs; symmetric case 32
1.10 Irish Wind: Log Likelihoods; symmetric case 32
1.11 Irish Wind: RMSEs; asymmetric case 35
1.12 Irish Wind: Log Likelihoods; asymmetric case 35

2.1 Exponential Askey $\gamma = 0.1$ MSPE and Time 61
2.2 General Results of Multivariate Tapers 63
2.3 Matérn : $\nu_1 = 0.25$, $\nu_2 = 0.25$ and Tapers MSPE 64
2.4 Matérn : $\nu_1 = 0.25$, $\nu_2 = 0.25$ and Tapers Times 64
2.5 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.5$ and Tapers MSPE 65
2.6 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.5$ and Tapers Times 66
2.7 Matérn : $\nu_1 = 1.5$, $\nu_2 = 1.5$ and Tapers MSPE 67
2.8 Matérn : $\nu_1 = 1.5$, $\nu_2 = 1.5$ and Tapers Times 67
2.9 Matérn : $\nu_1 = 2.5$, $\nu_2 = 2.5$ and Tapers MSPE 68
2.10 Matérn : $\nu_1 = 2.5$, $\nu_2 = 2.5$ and Tapers Times 69
2.11 US Precipitation Times 71
2.12 US Precipitation MSPEs 71

2.13 Exponential Askey $\gamma = 0.5$ MSPE and Time 77
2.14 Exponential Askey $\gamma = 0.3$ MSPE and Time 78
2.15 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.25$ and Tapers MSPE 79
2.16 Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.25$ and Tapers Time 79
2.17 Matérn : $\nu_1 = 0.5$, $\nu_2 = 1.5$ and Tapers MSPE 80
2.18 Matérn : $\nu_1 = 0.5$, $\nu_2 = 1.5$ and Tapers Times 81
2.19 Matérn : $\nu_1 = 0.5$, $\nu_2 = 2.5$ and Tapers MSPE 82
2.20 Matérn : $\nu_1 = 0.5$, $\nu_2 = 2.5$ and Tapers Times 82
2.21 Matérn : $\nu_1 = 1.5$, $\nu_2 = 0.5$ and Tapers MSPE 83
2.22 Matérn : $\nu_1 = 1.5$, $\nu_2 = 0.5$ and Tapers Times 84
2.23 Matérn : $\nu_1 = 1.5$, $\nu_2 = 2.5$ and Tapers MSPE 85
2.24 Matérn : $\nu_1 = 1.5$, $\nu_2 = 2.5$ and Tapers Times 85

Acknowledgments

I would like to acknowledge and extend my heartfelt gratitude to my Major Professor Dr. Juan Du. She has guided me for the past four years on a journey through both space and time. Her positive attitude and genius intellect will be something that I strive for every day. She will be a ever lasting light of inspiration throughout the rest of my life and career. Without her this dissertation would not have been possible.

I would like to thank my committee members, Dr. Gary Gadbury, Dr. Weixing Song, Dr. Marcelus Caldas, and Dr. Nathan Albin for providing different avenues of thought when looking at problems and help editing. Also to Dr. Gadbury and Dr. Song for some of the best classes I had through my graduate career. Thank you all for taking the time to read my dissertation and listen to both my preliminary exam and my defense.

In addition, a thank you to Dr. Cecil Hallum for introducing me to the world of statistics, and finding the diamond in the rough. I would like to thank all my professors for the wonderful classes and years of knowledge. Thanks to all my classmates for the great discussions and hours of studying. A very special thanks to Pam Schierer and Teresa Zerbe for all they do, especially making a good ole Texas boy feel at home. Many thanks to my family and friends who kept encouraging me through out my studies.

And to God, who makes all things possible.

Dedication

This is dedicated to my loving wife, Nicole. For all the love and support she has given to me over the past seven years. For believing that I could accomplish this goal even when I though I could not. For making me smile, laugh, and have a good time when things were down. For finding my marbles everytime I lost them in the space-time domain. Babe, all my love goes to you. I'm Yours Forever

And to my parents, love you Mom and Dad.

Preface

Space time data sets are often collected at monitored discrete time lags, which are normally viewed as a component of time series. Valid and practical covariance structures are needed to model these types of data sets in various disciplines, such as environmental science, climatology, and agriculture. In Chapter 1 we propose two classes of spatio-temporal functions whose discrete temporal margins are some celebrated autoregressive and moving average (ARMA) models, and obtain necessary and sufficient conditions for them to be valid spatio-temporal covariance functions. An asymmetric version of this type of model is also provided to account for the space-time irreversibility property. It is found that the existing tools in analyzing time series and spatial data can be efficiently applied to identify and fit the proposed model in practice. To show the application of these models, a simulation study and two data analyses are presented. Kansas daily precipitation data is fitted with a spatio-temporal model that has a MA(1) temporal margin. Daily wind speed data from Ireland is fitted with a spatio-temporal model with AR(2) temporal margin. In all the analyses, the performance of proposed models are compared with some general existing parametric models in terms of likelihood, mean squared prediction error, and continuous rank probability score.

Spatial covariance tapering is one of the techniques to mitigate the computational burden for certain attributes observed on a large number of locations. If multiple attributes are further recorded at each of these location, as is often seen in geophysical and atmospheric sciences, the added complexity of the covariances among different attributes at different locations presents even bigger challenges to statistical approaches in estimation and prediction. In Chapter 2 we investigate multivariate generalization of tapering techniques. We derive classes of covariance matrix functions and develop multivariate covariance tapering with those functions whose entries are compactly supported. In particular, we generate Wendland-type compactly supported covariance functions for the multivariate case with different degrees of smoothness. After normalization, those functions are employed as a

tapering matrix function to multiply element-wise the original matrix function so that the resulting covariance matrices can be efficiently manipulated using sparse matrix techniques. A simulation study is conducted to show the computational gain and provide guideline on choosing appropriate tapering matrix functions in applications of multivariate spatial prediction. Moreover, an illustration of applying multivariate tapering is presented using USA climate data to show how tapering can be used in data analysis.

Finally in Chapter 3, we study the impact of weather change on corn yield in Kansas. Weather elements such as maximum and minimum temperatures and precipitation dictate the crop yield during the growing season. To use this information effectively, we develop a spatial functional linear regression model accounting for the fact that weather data were recorded daily or hourly as opposed to the yearly crop yield data and the underlying spatial processes are autocorrelated. The parameter function is estimated under the functional data analysis framework and its characteristics are investigated to show the influential factor and critical period of weather change affecting crop yield during the growing season. The effect of climate change in larger time scale is explored using functional box plot. We also investigate the behavior of a space/space-time dependence structure that can model the error term of the functional regression.

Chapter 1

Spatio-temporal covariance modeling in continuous space and discrete time

1.1 Introduction

In the ever growing world of data analysis one area, spatio-temporal statistics, has allowed us to model our world as a whole. The world we live in is four dimensional, three dimensions making up space, and the fourth dimension being time. Often times appealing to computational convenience, we study space and time separately not taking into account their possible interaction. However, many environmental processes are dependent on the space-time interaction, and examples can be found in climatology, agriculture, and other environmental studies. See Cressie and Huang (1999), de Luna and Genton (2002), Wikle and Royle (2005), Xu et al. (2005), Le and Zidek (2006), Pearce et al. (2006), among others. Combining the spatial and temporal processes together creates the spatio-temporal process, which can be viewed as a random field given by, $\{Z(\mathbf{s};t), \mathbf{s} \in \mathcal{S}, t \in \mathcal{T}\}$, where $\mathcal{S} = \mathbb{R}^d$ and $\mathcal{T} = \mathbb{R}$ or $\mathbb{Z}$. From this notation the spatial marginal of the spatio-temporal process is given by, $\{Z(\mathbf{s};0), \mathbf{s} \in \mathbb{R}^d\}$, i.e. observing the spatial process at a fixed time. The temporal margin of the process looks at the temporal process at a fixed location and is given by, $\{Z(0;t), t \in \mathcal{T}\}$. Understanding these two marginals is the starting point for addressing the the main interests of space-time statistics.

In spatio-temporal data analysis there are two basic questions of interest: predicting at a

new space-time location, and the development of valid spatio-temporal covariance functions. There have been many recent developments in space-time modeling, mainly dealing with the construction of space-time covariance functions. When the second-order moments of the random field exist, its covariance function is defined by

$$C(\mathbf{s}_1, \mathbf{s}_2; t_1, t_2) = \text{Cov}(Z(\mathbf{s}_1; t_1), Z(\mathbf{s}_2; t_2)), \ (\mathbf{s}_1; t_1), \ (\mathbf{s}_2; t_2) \in \mathcal{S} \times \mathcal{T} \tag{1.1}$$

where $\mathcal{S} = \mathbb{R}^d$ and $\mathcal{T} = \mathbb{R}$ or $\mathbb{Z}$. Currently many research efforts use $\mathcal{T} = \mathbb{R}$, to mention a few; Haslett and Raftery (1989), Gneiting (2002b), Ma (2003), Stein (2005a) and Gneiting et al. (2007). Within all of these references the authors build valid space-time covariance structures under the framework of continuous space and time. They also mention a need to contend with difficulties arising from the fact that time data are usually measured at discrete time points and normally viewed as a component of time series. However, the attempts in the spatio-temporal modeling directly with discrete time series margin, i.e. $\mathcal{T} = \mathbb{Z}$, are limited. Most of these works were either based on spectral representation or stochastic equations (see e.g., Storvik et al. (2002), Stein (2005b)). In these cases the model choice is very difficult to justify in practice.

The goal of this chapter is to construct a valid spatio-temporal covariance function over continuous space and discrete time. This model will provide an intuitive approach which takes advantage of existing time series and spatial statistics techniques; and will provide an interpretable model that is easier to apply in practice. The remainder of this chapter is organized as follows: Section 1.2 explains the simplifications of stationarity, separability, and full symmetry that are usually assumed in order to preform data analysis. Section 1.3 discusses the conditions for space-time covariance functions with a moving average while Section 1.4 discusses ARMA type temporal margin. Also included is a simulation study of the moving average case, a data analysis of Kansas Daily Precipitation data, and a data analysis of the classic Irish Wind dataset introduced by Haslett and Raftery (1989).

1.2 Simplifying and validating the space-time process

When trying to model the covariance of a spatial or spatio-temporal process, we tend to make some simplifications and often justifiable assumptions, such as stationarity, separability, and symmetry. Stationarity is the most common one of these assumptions, in many cases without this assumption analysis can not be preformed. In the case of space-time processes there are two types of stationarity involved, spatial and temporal. The covariance function 1.1 has spatial stationarity if $C(\mathbf{s}_0, \mathbf{s}_0 + \mathbf{s}; t, t)$, written as $C(\mathbf{s}_0, \mathbf{s}_0 + \mathbf{s}; t)$, depends only on the space lag $\mathbf{s}$; it has temporal stationarity if $C(\mathbf{s}, \mathbf{s}; t_0, t_0 + t)$, written as $C(\mathbf{s}; t_0, t_0 + t)$, depends only on the time lag t. A random field is said to be stationary in both space and time if its mean $\mathrm{E}(Z(\mathbf{s}; t))$ is constant for all $(\mathbf{s}; t)$ and its covariance function $C(\mathbf{s}_0, \mathbf{s}_0 + \mathbf{s}; t_0, t_0 + t)$ depends only on the space lag $\mathbf{s}$ and time lag t for all $(\mathbf{s}_0; t_0) \in \mathbb{R}^d \times \mathcal{T}$, which allows us to write $C(\mathbf{s}; t)$. With the stationary assumption the covariance function of certain lag does not change when shifted in space and/or time.

In the history of this field the initial idea was to develop valid covariance functions for both space and time independently of each other then multiply them together. This created what is known today as the separable covariance function. A random field Z is said to have a separable covariance function if there exists a valid spatial covariance function $C_S(\mathbf{s}_1, \mathbf{s}_2)$ and a valid temporal covariance function $C_T(t_1, t_2)$, such that $\mathrm{Cov}(Z(\mathbf{s}_1; t_1), Z(\mathbf{s}_2; t_2)) = C_S(\mathbf{s}_1, \mathbf{s}_2) \cdot C_T(t_1, t_2)$. Gneiting (2002b), Stein (2005a) and Gneiting et al. (2007) explain that the separable model, although simple and computationally attractive, does not really capture the interaction between space and time. When applying this covariance function, space and time are considered to be independent of each other, which in many cases is not physically justifiable. This limits the modeling on the actual behavior of underlying processes. Some further discussion on this matter involving estimation and prediction can be found in Stein (2005a) and references therein.

Symmetric versus asymmetric models is a more recent discussion in space-time modeling. Asymmetric models allow the covariance function to take into account underlying

directional space-time effects caused by atmospheric, environmental, and geophysical processes. Examples of such processes are, prevailing winds, ocean currents, climate patterns, or thermal waves. Symmetric models often do not have the structure to include these effects; therefore, the asymmetric or not fully symmetric (Gneiting (2002b); Gneiting et al. (2007)) covariance function is introduced. A space-time process Z has a fully symmetric covariance function if $\text{Cov}(Z(\mathbf{s}_1;t_1), Z(\mathbf{s}_2;t_2)) = \text{Cov}(Z(\mathbf{s}_1;t_2), Z(\mathbf{s}_2;t_1))$ for all space-time coordinates $(\mathbf{s}_1;t_1)$ and $(\mathbf{s}_2;t_2) \in \mathcal{S} \times \mathcal{T}$, Gneiting et al. (2007). With the above mentioned processes this condition does not always hold true, Stein (2005a) and Gneiting et al. (2007) construct covariance functions that account for these asymmetric effects in the Irish Wind data. Further discussion of the asymmetric case is left to Section 1.4.1.

Another simplification that can be made to spatio-temporal covariance functions is when C is compactly supported. Having a compactly supported covariance function allows a threshold to be set on the space-time lag, meaning the $\text{Cov}(Z(\mathbf{s}_1;t_1), Z(\mathbf{s}_2;t_2)) = 0$ whenever the lag in space and/or the lag in time exceeds the threshold. This leads to computational efficiency for prediction, estimation, and simulation, especially for large datasets (Gneiting (2002a)). More of this discussion is left to Chapter 2. Figure 4.1 on page 155 in Gneiting et al. (2007) provides a very nice schematic illustration giving the relationships between separable, fully symmetric, stationary, and compactly supported covariance functions.

Regardless the simplifications made to the covariance function, it must still have nonnegative definiteness to be valid. This is the biggest challenge for all space-time covariance functions. A spatio-temporal covariance function is nonnegative definite if for any $a_1, \ldots, a_k \in \mathbb{R}$ and any k locations and time points, the following inequality holds

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(\mathbf{s}_i, \mathbf{s}_j; t_i, t_j) \geq 0. \tag{1.2}$$

For the proposed models in this chapter proofs for nonnegative definiteness are left to the Chapter 1 Appendix in Section 1.6.

As mentioned earlier many well- established covariance functions build models on the framework of continuous space and continuous time. However, in practice space-time data

are often collected at monitored discrete time lags, and it is natural to assume the space-time process $Z(\mathbf{s}, t)$ inhabits $\mathbb{R}^d \times \mathbb{Z}$. Using the framework of continuous space and discrete time opens the door to a new class of covariance functions that would take advantage of classical time series modeling techniques. To model most spatial-temporal data, the usual starting point is breaking the problem into two parts, the time series and the spatial component. Typically, we start at a fixed location and analyze the time series, usually with an ARMA-type model and then analyze the spatial field at a fixed time. In many research fields, experimenters use this approach in looking at time on one hand and space on the other. There is a a wealth of knowledge in the data exploration of these two processes; why not construct a model that takes advantage of this front end analysis efficiently? This is the motivation behind the construction of the covariance functions introduced in this chapter. To begin we first look at a spatio-temporal process where the temporal margin has a moving-average-type model.

1.3 Moving-average-type temporal margin

The moving average model in time series is a building block of more complex model structures. It is only fitting that we begin our discussion of discrete temporal margins with this type of structure. Namely, we start to investigate the permissibility of real-valued functions $\psi_0(\mathbf{s}_1, \mathbf{s}_2)$ and $\psi_1(\mathbf{s}_1, \mathbf{s}_2)$, $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$, such that the function

$$C(\mathbf{s}_1, \mathbf{s}_2; t) = \begin{cases} \psi_0(\mathbf{s}_1, \mathbf{s}_2), & t = 0, \\ \psi_1(\mathbf{s}_1, \mathbf{s}_2), & t = \pm 1,\ \mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}, \\ 0, & \text{otherwise}, \end{cases} \tag{1.3}$$

in the domain of $\mathcal{S} \times \mathbb{Z}$ that is stationary in time is a covariance function. A basic feature of (1.3) is that its temporal margin at a fixed location $\mathbf{s}$,

$$C(\mathbf{s}, \mathbf{s}; t) = \begin{cases} \psi_0(\mathbf{s}, \mathbf{s}), & t = 0, \\ \psi_1(\mathbf{s}, \mathbf{s}), & t = \pm 1, \\ 0, & \text{otherwise}, \end{cases}$$

is a stationary first-order moving average model, provided that $|\psi_1(\mathbf{s}, \mathbf{s})| \leq \frac{1}{2}\psi_0(\mathbf{s}, \mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, under which (1.3) is a spatio-temporal function whose temporal margin is a first-order

moving average model.

Note here that the covariance function given in (1.3) is not stationary and as a result, to prove its nonnegative definiteness the well- known Bochner's Theorem (see Chapter 1 Appendix 1.35) can not be directly applied. However, (1.3) can be rewritten as

$$\begin{aligned} C(\mathbf{s}_1,\mathbf{s}_2;t) &= \tfrac{\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)}{2}\cdot\begin{cases}1, & t=0,\\ \frac{1}{2\alpha}, & t=\pm1,\\ 0, & t=\pm2,\ldots,\end{cases} \\ &\quad + \tfrac{\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)}{2}\cdot\begin{cases}1, & t=0,\\ -\frac{1}{2\alpha}, & t=\pm1,\\ 0, & t=\pm2,\ldots,\end{cases}\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}, \end{aligned} \tag{1.4}$$

which is a product-sum of purely spatial and purely temporal functions

$$\frac{\psi_0(\mathbf{s}_1,\mathbf{s}_2)\pm 2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)}{2},\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},\ \text{and}\ \begin{cases}1, & t=0,\\ \pm\frac{1}{2\alpha}, & t=\pm1,\\ 0, & t=\pm2,\pm3,\ldots,\end{cases} \tag{1.5}$$

where the latter are correlation functions of stationary first-order moving average models and α is a constant not less than 1. Based on the decomposition, (1.5), we can use Bochner's Theorem to obtain the conditions for (1.3) to be a valid spatio-temporal covariance function. Also for ease of use the conditions will be given in terms of the functions $\psi_0(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$. To do this we need two lemmas.

Lemma 1.3.1. *Let α be a constant greater than or equal to 1. If $\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, are spatial covariance functions on $\mathcal{S}$, then (1.3) is a spatio-temporal covariance function on $\mathcal{S}\times\mathbb{Z}$ that is stationary in time.*

Lemma 1.3.2 discusses the relationship that exists between, $\psi_0(\mathbf{s}_1,\mathbf{s}_2)\pm2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)\pm2\beta\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, when α and β are two nonnegative constants.

Lemma 1.3.2. *If α is a positive constant, and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, are spatial covariance functions on $\mathcal{S}$, then so are $\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\beta\psi_1(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\beta\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, for any $0\le\beta\le\alpha$.*

In fact, when both $\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$ and $\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)$, $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, are spatial covariance functions,

$$\psi_0(\mathbf{s}_1,\mathbf{s}_2)=\frac{\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)}{2}+\frac{\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)}{2},\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},$$

is also a covariance function on $\mathcal{S}$. So are

$$\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\beta\psi_1(\mathbf{s}_1,\mathbf{s}_2)=\left(1-\frac{\beta}{\alpha}\right)\psi_0(\mathbf{s}_1,\mathbf{s}_2)+\frac{\beta}{\alpha}\{\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)\},$$

and

$$\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\beta\psi_1(\mathbf{s}_1,\mathbf{s}_2)=\left(1-\frac{\beta}{\alpha}\right)\psi_0(\mathbf{s}_1,\mathbf{s}_2)+\frac{\beta}{\alpha}\{\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\alpha\psi_1(\mathbf{s}_1,\mathbf{s}_2)\},\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}.$$

Lemma 1.3.2 leads us to search for what values of α could be necessary and sufficient for (1.3) to be a spatio-temporal covariance function. Actually the only choice is $\alpha=1$, as Theorem 1.3.1 indicates.

Theorem 1.3.1. *The function (1.3) is a spatio-temporal covariance function on $\mathcal{S}\times\mathbb{Z}$ if and only if both*

$$C_+(\mathbf{s}_1,\mathbf{s}_2)=\psi_0(\mathbf{s}_1,\mathbf{s}_2)+2\psi_1(\mathbf{s}_1,\mathbf{s}_2),\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},$$

and

$$C_-(\mathbf{s}_1,\mathbf{s}_2)=\psi_0(\mathbf{s}_1,\mathbf{s}_2)-2\psi_1(\mathbf{s}_1,\mathbf{s}_2),\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},$$

are spatial covariance functions on $\mathcal{S}$.

Corollary 1.3.1. *When (1.3) is a spatio-temporal covariance function on $\mathcal{S}\times\mathbb{Z}$, it is easy to verify that for any constant θ with $|\theta|\le 1$,*

$$C(\mathbf{s}_1,\mathbf{s}_2;t)=\begin{cases}\psi_0(\mathbf{s}_1,\mathbf{s}_2), & t=0,\\ \theta\psi_1(\mathbf{s}_1,\mathbf{s}_2), & t=\pm1,\ \mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},\\ 0, & \textit{otherwise},\end{cases}\tag{1.6}$$

is a spatio-temporal covariance function on $\mathcal{S}\times\mathbb{Z}$.

Note that $\psi_0(\mathbf{s}_1, \mathbf{s}_2)$ is a purely spatial covariance function on $\mathcal{S}$, since it equals $C(\mathbf{s}_1, \mathbf{s}_2; 0)$, the spatial margin of $C(\mathbf{s}_1, \mathbf{s}_2; t)$. While (1.3) can be used for non-stationary covariance functions, Corollary 1.3.2 shows the simplification when the field has stationarity.

Corollary 1.3.2. *Let $\psi_0(\mathbf{s})$ and $\psi_1(\mathbf{s})$ be real-valued functions defined on $\mathcal{S}$. Then*

$$C(\mathbf{s}; t) = \begin{cases} \psi_0(\mathbf{s}), & t = 0, \\ \psi_1(\mathbf{s}), & t = \pm 1,\ \mathbf{s} \in \mathcal{S}, \\ 0, & \textit{otherwise}, \end{cases} \tag{1.7}$$

is a stationary covariance function on $\mathcal{S} \times \mathbb{Z}$ if and only if both $\psi_0(\mathbf{s}) + 2\psi_1(\mathbf{s})$ and $\psi_0(\mathbf{s}) - 2\psi_1(\mathbf{s})$ are stationary covariance functions on $\mathcal{S}$.

As a benefit of this corollary, it suffices to check the validity of two strictly spatial functions $\psi_0(\mathbf{s}) + 2\psi_1(\mathbf{s})$ and $\psi_0(\mathbf{s}) - 2\psi_1(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$, in order to verify the spatio-temporal function (1.7).

Example 1.3.1. Let us take a look of the permissible domain of the parameter θ that makes the function

$$C(\mathbf{s}; t) = \begin{cases} \exp(-\|\mathbf{s}\|), & \mathbf{s} \in \mathbb{R}^d,\ t = 0, \\ \theta\|\mathbf{s}\| \exp(-\|\mathbf{s}\|), & \mathbf{s} \in \mathbb{R}^d,\ t = \pm 1, \\ 0, & \text{otherwise}, \end{cases}$$

a valid covariance function on $\mathbb{R}^d \times \mathbb{Z}$. By the Corollary 1.3.2,

$$\exp(-\|\mathbf{s}\|) \pm 2\theta\|\mathbf{s}\| \exp(-\|\mathbf{s}\|)$$

must be covariance functions in $\mathbb{R}^d$. For this, a necessary and sufficient condition is $|\theta| \leq \frac{1}{2d}$ (see, e.g., Cambanis et al. (1981)).

Now that the basic structure of a spatio-temporal covariance function with a moving-average-type temporal margin has been established, we construct the spatial component to be more specific. Using Theorem 1.3.1 as well as Corollary 1.3.2, we impose the well- known Matérn spatial margin and give the conditions to create a valid covariance function.

The spatial covariance model $(\alpha\|\mathbf{s}\|)^\nu K_\nu(\alpha\|\mathbf{s}\|)$, $\mathbf{s} \in \mathbb{R}^d$, was proposed in von Kármán (1948) for $\nu = \frac{1}{3}$ in $\mathbb{R}^3$, constructed in the plane in Whittle (1954) via the stochastic partial

differential equation, and proposed in Matérn (1960) in the general form, where α is a positive constant, and $K_\nu(x)$ stands for the modified Bessel functions of the second kind of order ν (Gradshteyn and Ryzhik (2000)). It reduces to $\sqrt{\frac{\pi}{2}}\exp(-\alpha\|\mathbf{s}\|)$ and $\sqrt{\frac{\pi}{2}}(1+\alpha\|\mathbf{s}\|)\exp(-\alpha\|\mathbf{s}\|)$, $\mathbf{s}\in\mathbb{R}^d$, when $\nu=\frac{1}{2}$ and $\nu=\frac{3}{2}$, respectively. Note when $\nu=\frac{1}{2}$ the Matérn model becomes the exponential spatial covariance function. The parameter ν is a smoothness parameter that controls the degree of the differentiability of the model. The following theorem determines a spatio-temporal covariance function whose spatial margin is a linear combination of von Kármán-Whittle-Matérn models and whose temporal margin is a first-order moving average.

Theorem 1.3.2. *Assume that ν, α_k, and β_k $(k=1,2)$ are constants with $\nu>0$, $0<\alpha_1<\alpha_2$ and $-\frac{1}{2}\le\beta_1<\beta_2<\frac{1}{2}$. A necessary and sufficient condition for the function*

$$C(\mathbf{s};t)=\begin{cases}\theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|)+(1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|), & t=0,\\ \theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|)\beta_1+(1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|)\beta_2, & t=\pm1,\ \mathbf{s}\in\mathbb{R}^d,\\ 0, & \textit{otherwise},\end{cases} \tag{1.8}$$

to be a stationary correlation function on $\mathbb{R}^d\times\mathbb{Z}$ is that the constant θ satisfies

$$\left\{1-\frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\right\}^{-1}\ \le\ \theta\ \le\ \left\{1-\frac{\alpha_1^{2\nu}(1+2\beta_1)}{\alpha_2^{2\nu}(1+2\beta_2)}\right\}^{-1}. \tag{1.9}$$

The reason why β_1 and β_2 are so restricted in Theorem 1.3.2 is to presume the validity of the temporal margin of (1.8)

$$\begin{cases}1, & t=0,\\ \theta\beta_1+(1-\theta)\beta_2, & t=\pm1,\\ 0, & \text{otherwise},\end{cases}$$

in case θ equals 0 or 1. When β_2 approaches $\frac{1}{2}$ from its left-hand side, the lower bound of θ in (1.9) tends to zero so that (1.8) reduces to a separable spatio-temporal model, which is the product of a spatial model and a temporal model.

The lower bound of θ in (1.9) depends on the dimensional parameter d, but the upper bound does not; in contrast, the upper bound of θ in (1.9) depends on the smoothness parameter ν and the lower bound does not. Since the interval $[0,1]$ is only a subset of θ's

permissible domain (1.9) and β_1 or β_2 may be negative, the function (1.8) is flexible to represent spatio-temporal positive and negative correlations.

Taking $\nu = \frac{1}{2}$ in (1.8) and omitting the constant, yields

Corollary 1.3.3. The function

$$C(\mathbf{s};t) = \begin{cases} \theta \exp(-\alpha_1\|\mathbf{s}\|) + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|), & t = 0, \\ \theta \exp(-\alpha_1\|\mathbf{s}\|)\beta_1 + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2, & t = \pm 1,\ \mathbf{s} \in \mathbb{R}^d, \\ 0, & \textit{otherwise}, \end{cases} \tag{1.10}$$

is a stationary correlation function on $\mathbb{R}^d \times \mathbb{Z}$ if and only if the constant θ satisfies

$$\left\{1 - \frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\right\}^{-1} \leq \theta \leq \left\{1 - \frac{\alpha_1(1+2\beta_1)}{\alpha_2(1+2\beta_2)}\right\}^{-1}. \tag{1.11}$$

For proof of Corollary 1.3.3 see Demel and Du (2011). The above is an example of a spatio-temporal covariance function with an exponential spatial margin and a moving average (MA(1)) temporal structure. This example is used in Section 1.3.1 to create a simulation study of the performance of this covariance function and applied to study the Kansas weather dataset.

1.3.1 Moving average simulation study

Using Theorem 1.3.2 and spatio-temporal correlation function of Corollary 1.3.3 we simulate one hundred realizations from a Gaussian spatio-temporal process; with a constant mean of zero and twenty uniformly distributed locations in $\mathbb{R}^2$ on the unit square at one hundred time points in $\mathbb{Z}$. The spatial marginal is an exponential structure and the temporal margin is a MA(1). The model used for this simulation is from Corollary 1.3.3 with parameters; $\alpha_1 = 0.005$, $\alpha_2 = 0.009$, $\beta_1 = 0.3$, $\beta_2 = 0.4$, and $\theta_1 = 0.3$.

In the simulations we compare the likelihoods of our model to the models introduced by Gneiting (2002b). Gneiting has a separable model given by,

$$C_{G.SEP}(\mathbf{s};t) = [\exp(-c\|\mathbf{s}\|)] \cdot \left[(1 + a|t|^{2\alpha})^{-1}\right],\ \mathbf{s} \in \mathbb{R}^d,\ t \in \mathbb{R}, \tag{1.12}$$

and a non-separable model given by,

$$C_{G.NSEP}(\mathbf{s};t) = \frac{1}{1+a|t|^{2\alpha}}\left(\exp\left(-\frac{c\|\mathbf{s}\|}{(1+a|t|^{2\alpha})^{\beta/2}}\right)\right),\ \mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{R}. \tag{1.13}$$

Note that Gneiting's models are in terms of correlation functions and not covariance functions. To have a covariance function one would multiply by the overall location variance σ^2. To start the process of evaluating the parameters we first look at time and space independently of each other. As an example, look at the Autocorrelation Functions (ACFs) for nine of the twenty locations. From these plots as well as plots for other locations it is determined that at a fixed location time has a MA(1) structure. Spatially there is evidence in the variogram, Figure 1.1 to suggest an exponential spatial structure.

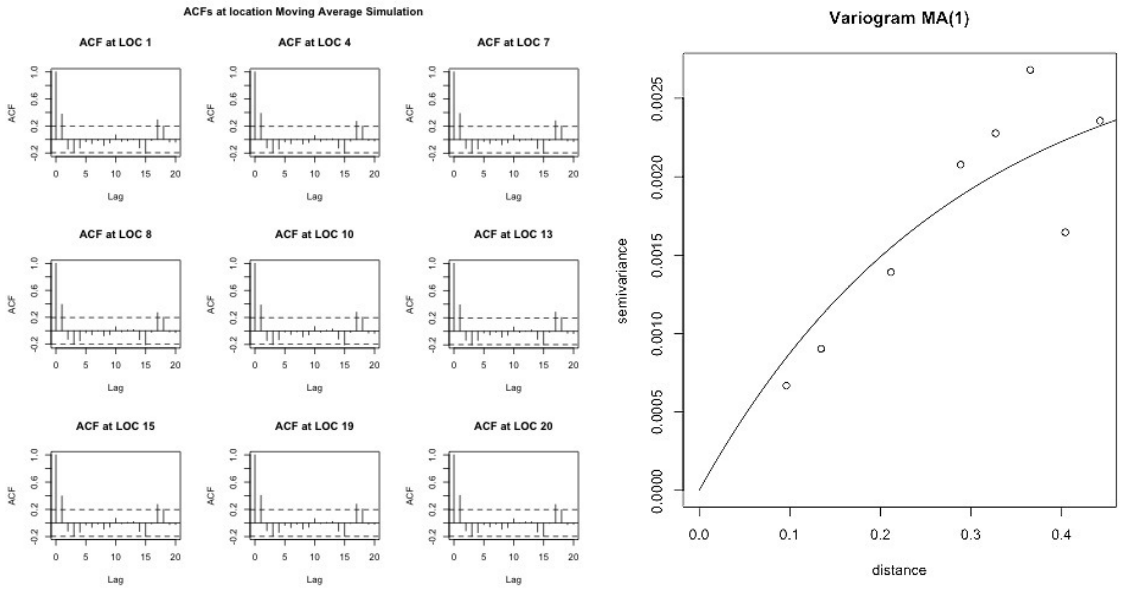

Figure 1.1: Autocorrelation functions of nine of twenty locations (left) and spatial variogrom at a fixed time (right)

Using weighted least squares procedure (Cressie and Huang (1999)) ,

$$W(\theta) = \sum_{i,j}\sum_{u=1}^{2}\left(\frac{\hat{C}(h_{ij};u) - C(h_{ij};u|\theta)}{1 - C(h_{ij};u|\theta)}\right)^2, \tag{1.14}$$

where $C(h_{ij};u|\theta)$ is the correlation function for the model and $\hat{C}(h_{ij};u)$ is the empirical space- time correlation. Here we minimize $W(\theta)$ on θ by summing over all locations and the

first two time points since after a lag time of one the correlation is zero. Table 1.1 shows the resulting parameter estimates.

Table 1.1: *Parameter Estimates and standard errors*

Parameter	TRUE	$C_{MA(1)}$
α_1	0.005	0.003728 (0.0004841)
α_2	0.009	0.009186 (0.0014624)
β_1	0.300	0.278291 (0.0383801)
β_2	0.400	0.374883 (0.1149921)
θ_1	0.300	0.250020 (0.0001042)

Table 1.1 is generated by finding the parameter estimates for all one hundred realizations and taking an average. Notice that the TRUE parameters and estimated parameters are virtually the same, which should be expected. Looking at the standard errors we can tell that the estimates are stable and that the model is doing well estimating the parameters. Another conclusion to draw from this analysis is that the weighted least squares estimates provide effective estimates that are accurate with ease of computation. Gneiting's models and the proposed model are compared by their ability to predict the process and they respected likelihoods. To compare prediction capabilities of the models, the data was split into two parts; the first seventy time points are the training dataset and the last thirty are the prediction dataset. Using two days of lag to predict the next day, Table 1.2 was created by finding the average root mean square error, (RMSE). Here the root mean square error is the average squared difference between the testing observed values and the forecasted values based on the training dataset.

Note that from Table 1.2 we can see that the proposed model has the smallest RMSE at all locations, and has slightly better predictions than Gneiting's models when the temporal margin is MA(1).

Our next comparison comes by analyzing the ratio between the likelihoods of the models versus the true likelihood. Notice that Table 1.3 shows that the model (1.10) is much closer to the true likelihood than Gneiting's models.

Table 1.2: *Average RMSEs predicting the next day based on two days of lag*

Model	Loc.1	Loc.2	Loc.3	Loc.4	Loc.5	Loc.6	Loc.7	Loc.8
G.SEP	1.053	1.054	1.053	1.054	1.053	1.053	1.054	1.054
G.NSEP	1.053	1.054	1.053	1.054	1.053	1.053	1.054	1.054
$C_{MA(1)}$	1.042	1.042	1.042	1.043	1.042	1.041	1.043	1.042

Model	Loc.9	Loc.10	Loc.11	Loc.12	Loc.13	Loc.14	Loc.15	Loc.16
G.SEP	1.054	1.053	1.053	1.054	1.053	1.054	1.053	1.053
G.NSEP	1.054	1.053	1.053	1.054	1.053	1.054	1.053	1.053
$C_{MA(1)}$	1.042	1.042	1.042	1.042	1.042	1.043	1.042	1.041

Model	Loc.17	Loc.18	Loc.19	Loc.20
G.SEP	1.054	1.054	1.054	1.053
G.NSEP	1.054	1.054	1.054	1.053
$C_{MA(1)}$	1.043	1.042	1.042	1.042

Table 1.3: *Likelihood Comparisons*

Model versus TRUE	Average Likelihood Ratio
Gneiting Separable	0.02951883
Gneiting Non Separable	0.08797862
$C_{MA(1)}$	0.70498240

This simulation study suggests that the proposed model, which takes into account the discreteness of the time points has an advantage over Gneiting's model when the temporal margin is MA(1). Now we will fit Kansas daily precipitation data using the developed space-time covariance function with a MA(1) temporal margin.

1.3.2 Kansas daily precipitation data

When dealing with raw daily weather data there are numerous issues that must be dealt with, one being missing data observations. This missing data can occur for many reasons; funding loss in a station, missed observation, data not entered, and much more. Using space-time modeling to fill in these gaps is one way of solving this problem. In Chapter 3

this analysis is used to model precipitation for Kansas counties to see how weather plays a vital role is the production of agriculture.

In this analysis daily precipitation across the state of Kansas is considered from 1990 to 2011. The data was collected from the National Oceanic and Atmospheric Administration (NOAA) on each of the 105 counties of Kansas. Over the state of Kansas there are 1123 stations, gray points in Figure 1.2, that supply data through the specified time range. First each county was aggregated by taking a daily average across the county's weather stations. This was done both to normalize the data and provide stationarity. Now each county has a time series of daily precipitation measurements in millimeters. Since an average is taken over the county, the time series is then allocated to the centroid point of that county, shown by the red squares in Figure 1.2. The resulting dataset is 105 time series, one for each

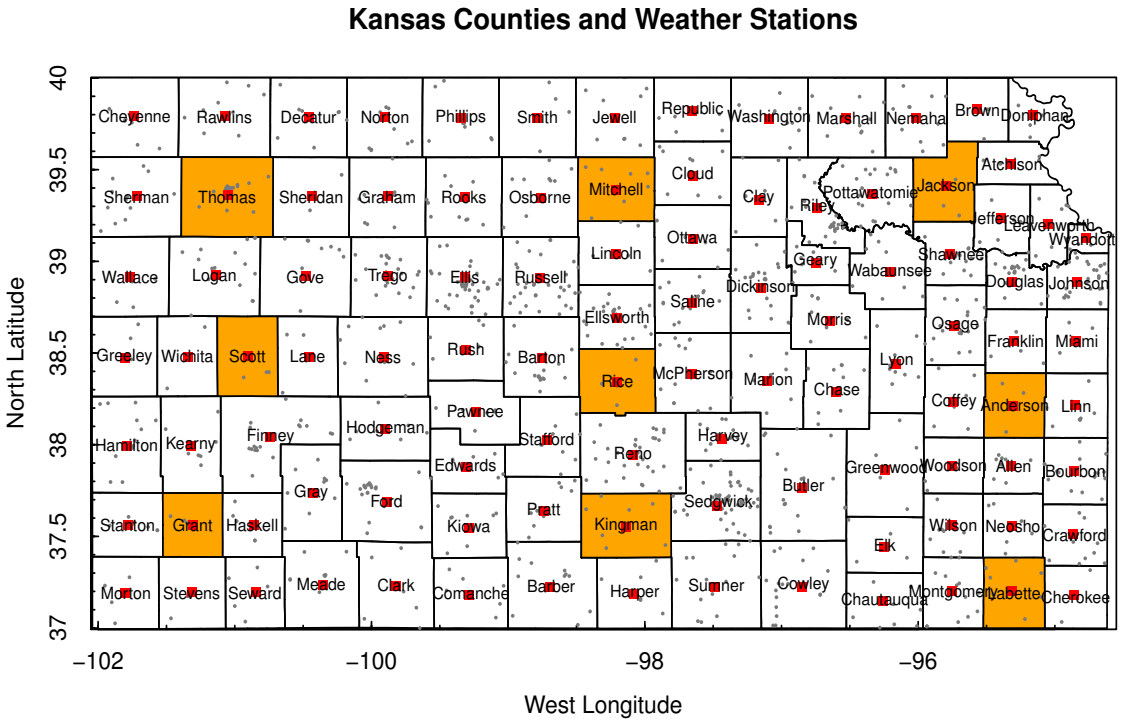

Figure 1.2: Kansas Counties Map: Red squares are centroid points, gray points are stations, orange highlighted counties sampled to display ACFs.

county, of daily precipitation data in millimeters over 8030 daily time points from January

1, 1990 to December 31, 2011. The dataset was then split into the first fifteen years to fit the model and the last five years to test the models predictive capabilities.

To begin the analysis we look at each process independently. It was observed that the margins of both space and time are approximately normally distributed, which allows the use of Gaussian theory for both the temporal process and spatial process when modeling and predicting.

Based on the previous works of Haslett and Raftery (1989), Gneiting (2002b), Stein (2005a) and Gneiting et al. (2007), the seasonal trend was fit and removed using annual harmonic regression, (see Figure 1.3). Also the spatial trend was extracted by removing

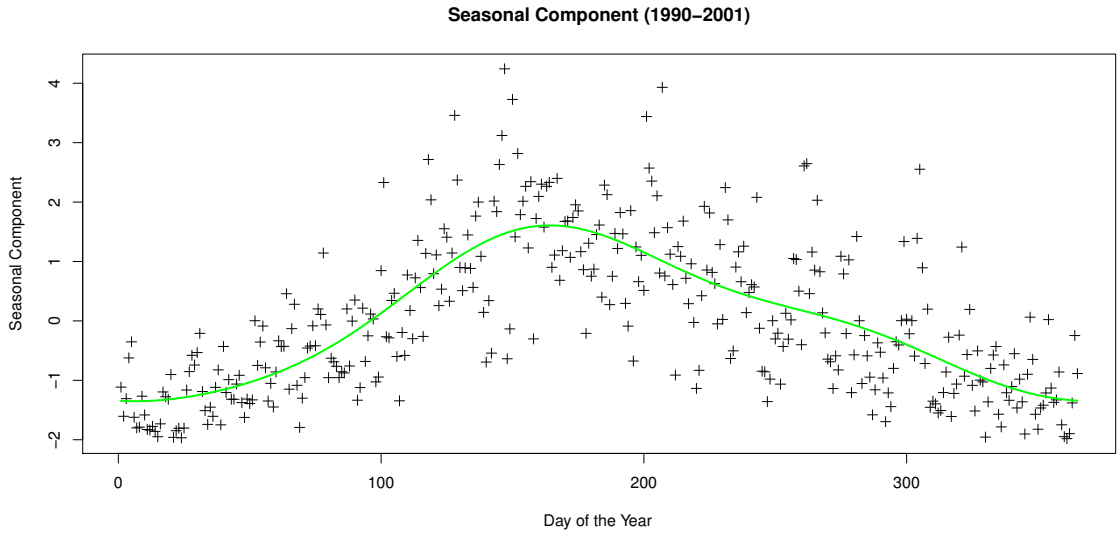

Figure 1.3: Seasonal Component of Kansas daily precipitation using annual harmonic regression.

the station specific means. To start the analysis of the space-time process an exploratory analysis of both space and time margins are done.

Referencing the orange highlighted counties of Figure 1.2, the ACFs of those counties are given in Figure 1.4. Notice the highlighted counties are equally spaced North to South and East to West, creating a small systematic sample of the state of Kansas. This small sample will be used to show a finer scale of results later in the analysis. Using time series analysis techniques, examining all Kansas counties indicates that the temporal marginal

can be modeled with a MA(1) process. This result is represented by ACFs of the nine highlighted counties. Figure 1.5 shows the fitted empirical time correlations for two days of lag.

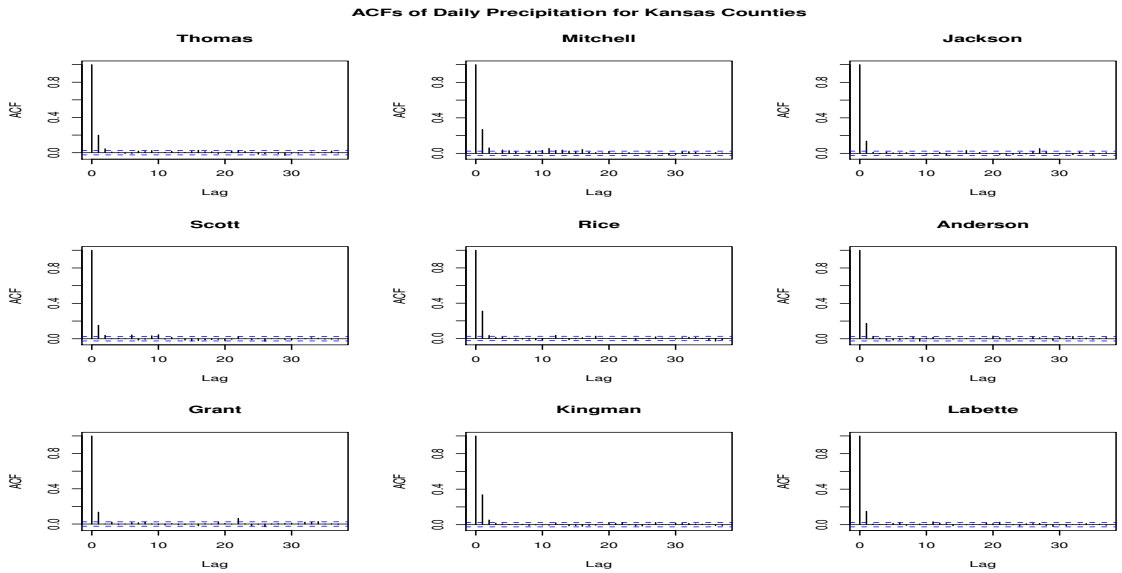

Figure 1.4: Autocorrelation functions of nine chosen counties over Kansas

Exploring the spatial component results in an exponential type structure. Figure 1.5 shows the empirical spatial correlation with the fitted mixture exponential model as proposed by Corollary 1.3.3. Based on the moving average simulation study and literature, namely Gneiting (2002b) and Gneiting et al. (2007), Cressie's weighted least squares (1.14) was used to fit parameter estimates. Fitting space and time independent of each allows for reasonable starting parameter values when fitting the overall space-time model. The proposed model with added measurement error term ν known as the nugget effect is given by,

$$C(\mathbf{s};t) = \begin{cases} (1-\nu)\left\{\theta\exp(-\alpha_1\|\mathbf{s}\|) + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\right\} + \nu\delta_{\mathbf{s}=0}, & t=0, \\ (1-\nu)\left\{\theta\exp(-\alpha_1\|\mathbf{s}\|)\beta_1 + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2\right\} & \\ \quad +\nu\left\{\theta\beta_1 + (1-\theta)\beta_2\right\}\delta_{\mathbf{s}=0}, & t=\pm 1,\ \mathbf{s}\in\mathbb{R}^d, \\ 0, & \text{otherwise}, \end{cases} \tag{1.15}$$

Using Cressie's WLS (1.14) to fit (1.15) the following parameter estimates where obtatined. $\nu = 0.322$, $\alpha_1 = 0.009$, $\alpha_2 = 0.003$, $\theta = 0.230$, $\beta_1 = -0.495$, and $\beta_2 = 0.495$. Notice that in

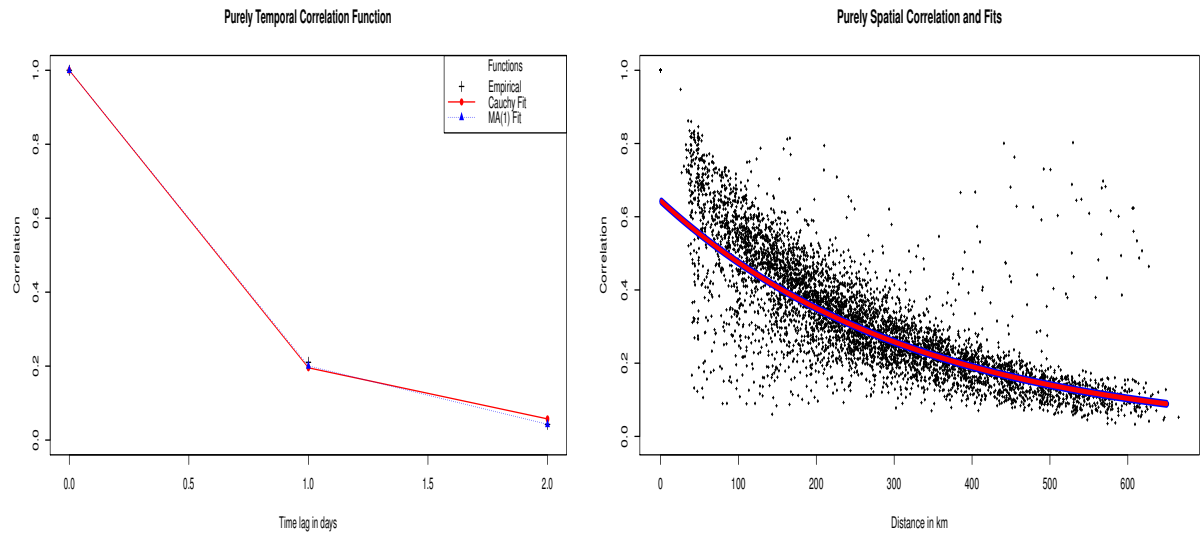

Figure 1.5: Kansas precipitation marginal plots: (left) time and (right) space.

this case $\alpha_1 > \alpha_2$, this cause a slight modification to the inequality (1.10) where the α_1 and α_2 change positions in the inequality.

To show the performance of the proposed model (1.15), comparisons are made with popular models in Gneiting (2002b) and Gneiting et al. (2007). Using Gneiting's separable model

$$C_{G.SEP}(\mathbf{s};t) = \{(1-\nu)\exp(-c\|\mathbf{s}\|) + \nu\delta_{\mathbf{s}=0}\} \cdot \left\{(1+a|t|^{2\alpha})^{-1}\right\}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{R}, \quad (1.16)$$

and Gneiting's non separable model

$$C_{G.NSEP}(\mathbf{s};t) = \frac{1-\nu}{1+a|t|^{2\alpha}} \left\{ \exp\left(-\frac{c\|\mathbf{s}\|}{(1+a|t|^{2\alpha})^{\beta/2}} \right) + \frac{\nu}{1-\nu}\delta_{\mathbf{s}=0} \right\}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{R}, \quad (1.17)$$

the same fitting procedures where used. First Gneiting's models where explored marginally in the same manner as the proposed model. A Cauchy type model was used to fit the temporal correlation, suggested by the empirical correlation function in Figure 1.5. The red dotted line shows the Cauchy fit, which fits the empirical time correlation rather well. An exponential spatial structure also seemed appropriate for Gneiting's model as seen in Figure 1.5. Notice that both the mixture exponential in blue and Gneiting's exponential fit in red are essential the same. Cressie's WLS was used to find overall parameter estimates

for both the separable and non- separable models.

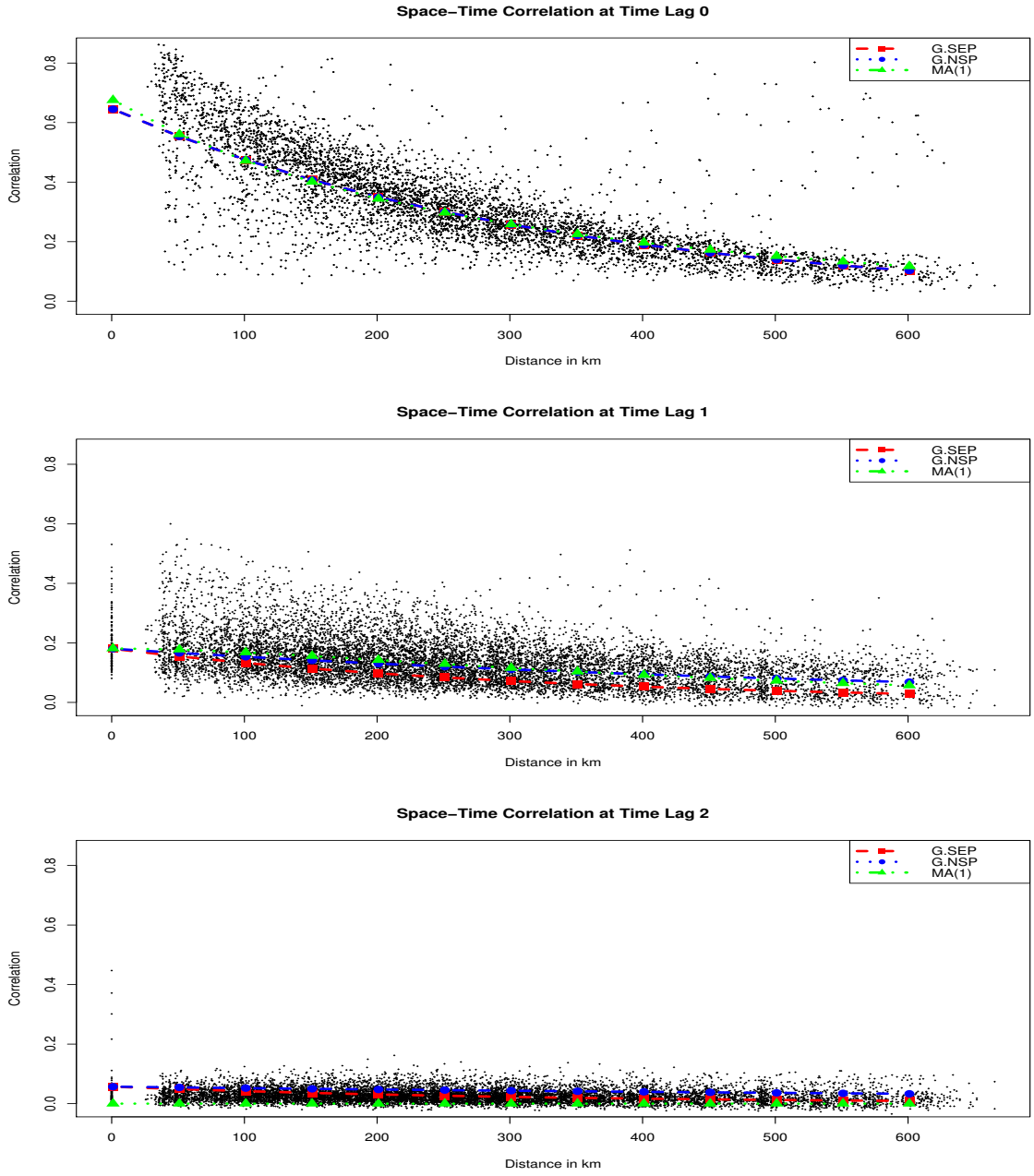

Figure 1.6: Empirical precipitation space-time correlations and fitted models based on 0, 1, 2 days of lag. Gneiting's separable model denoted by G.SEP (red), non separable model denoted by G.NSP (blue), and proposed model denoted by MA(1) (green).

Figure 1.6 shows the empirical space time correlations for 0, 1, and 2 days of lag with three fitted space-time models: Gneiting's separable model is represented by the blue line,

Gneiting's non-separable model is denoted by the red line, and the proposed model is the green line. Notice that after two days of lag the space-time correlation is close to zero. This fact encourages us to do one day ahead prediction based on two days of lag.

One comparison measurement used is the root mean-square error (RMSE). The RMSE is calculated by,

$$RMSE = \left(\frac{1}{T} \sum_{t=1}^{T} (f_t - o_t)^2 \right)^{1/2}; \tag{1.18}$$

where T is the number of time points in the testing dataset, 1825 in this analysis, o_t is the observed at time point t, and f_t is the forecasted at time point t. The RMSE was calculated for each county. Table 1.4 gives the average RMSE over all counties. Notice that all the models have a similar RMSE around 6.8, meaning on the average there is only 6.8 millimeters of error in predicting the next day's precipitation. The proposed model denoted $C_{MA(1)}$ does have the lowest average RMSE and standard deviation. The low standard deviation over counties suggests the models are fairly accurate in one day ahead prediction based on two days of lag. The 95% confidence intervals give a upper and lower bound for prediction errors across counties with the proposed model having the narrowest width. The low count indicates how many counties have the lowest RMSE per model. Sixty-five counties, 62% of counties, are better modeled by the proposed model. This shows that the proposed model performs best for most of the counties.

Table 1.4: *Kansas Precipitation RMSE Statistics*

Measure	G.SEP	G.NSEP	$C_{MA(1)}(\mathbf{s};t)$
AVG. RMSE	6.899	6.879	6.861
STD. DEV.	1.863	1.833	1.817
95% C.I.	(6.54, 7.24)	(6.52, 7.23)	(6.51, 7.21)
Low Count	29 (28%)	11 (10%)	65 (62%)

Table 1.5 shows a snap shot of all the counties to show the RMSEs for the counties highlighted in Figure 1.2. Here again we see that for most counties the $C_{MA(1)}$ model preforms constantly better then than Gneiting's models.

Table 1.5: *Kansas Precipitation Average RMSE for selected Counties*

County	G.SEP	G.NSEP	$C_{MA(1)}(\mathbf{s};t)$
Anderson	8.787	8.784	8.844
Grant	4.150	4.164	4.179
Jackson	7.035	7.083	7.129
Kingman	6.977	6.977	6.954
Labette	10.143	9.982	9.917
Mitchell	6.098	6.074	6.058
Rice	7.539	7.513	7.427
Scott	5.041	4.963	4.894
Thomas	4.691	4.699	4.660

The continuous ranked probability score, Gneiting et al. (2007), is used to compare predictive distributions $F_t = N(\mu_t, \sigma_t^2)$. Let $I(y \geq x)$ be an indicator function that takes the value 1 when $y \geq x$ and 0 otherwise. The continuous ranked probability score is defined as,

$$crps(F, x) = \int_{-\infty}^{\infty} (F(y) - I(y \geq x))^2 \, dy. \tag{1.19}$$

However, if $F_t = N(\mu_t, \sigma_t^2)$ has a normal distribution with mean μ and variance σ^2 the equation (1.19) can be evaluated as

$$crps(N(\mu, \sigma^2), x) = \sigma \left(\frac{x-\mu}{\sigma} \left(2\Phi \left(\frac{x-\mu}{\sigma} \right) - 1 \right) + 2\phi \left(\frac{x-\mu}{\sigma} \right) - \frac{1}{\sqrt{\pi}} \right). \tag{1.20}$$

For each county we find the crps for each time point in the testing dataset and then take an average over the scores to have an average CRPS for each county given by,

$$CRPS = \frac{1}{T} \sum_{t=1}^{T} crps(N(f_t, \sigma_t^2), o_t), \tag{1.21}$$

and the smaller the CRPS is the better. Here again $T = 1825$, the number of time points in the testing dataset, f_t is the forecast at time t, σ_t^2 is the variance of the forecast at each time point, and o_t is the observed at time point t of the test data. Table 1.6 shows the mean CRPS of all counties and again the proposed model has the lowest value. Also the model with MA(1) temporal structure has the most counties with the lowest CRPS with 89 out of

105 counties given by the Low Count, while the separable model can not work as well for all counties.

Table 1.6: *Kansas Precipitation CRPS Statistics*

Measure	G.SEP	G.NSEP	$C_{MA(1)}(\mathbf{s};t)$
AVG. CRPS	9.720	9.644	9.580
STD. DEV.	3.913	3.887	3.840
95% C.I.	(8.96, 10.48)	(8.89, 10.40)	(8.84, 10.32)
Low Count	0 (0%)	16 (15%)	89 (85%)

Table 1.7 gives the individual county average for the highlighted counties in Figure 1.2. Notice that the nine highlighted counties have the lowest CPRS when the model with a MA(1) temporal margin is chosen. This indicates again that the proposed model is constantly performing better overall.

Table 1.7: *Kansas Precipitation Average CRPS for selected Counties*

County	G.SEP	G.NSEP	$C_{MA(1)}(\mathbf{s};t)$
Anderson	14.827	14.726	14.583
Grant	5.797	5.757	5.701
Jackson	13.279	13.189	13.071
Kingman	9.559	9.486	9.386
Labette	17.862	17.732	17.595
Mitchell	7.507	7.447	7.368
Rice	8.692	8.622	8.512
Scott	4.337	4.275	4.340
Thomas	5.592	5.557	5.507

Based on this analysis, the proposed model (1.15) preforms slightly better than Gneiting's models when the temporal margin of the space-time process can be modeled with a MA(1) structure. This suggests that taking into account the discreteness of the times series does help to improve the predictability of the model and the time series analysis tools bring convenience in model identification. Also the addition of the mixture on two exponential spatial covariance functions aids in the flexibility of the model. Now we will make the ex-

tension to include some ARMA type temporal margins in Section 1.4. Once the model is established we use the famous Irish Wind dataset to compare to Gneiting existing results from Gneiting (2002b) and Gneiting et al. (2007).

1.4 ARMA-type temporal margin

The ARMA structure is widely used in time series analysis. In this section we extend the spatio-temporal covariance function with discrete temporal margin to include such a structure. By using mixture method, the following Theorem 1.4.1 produces a spatio-temporal covariance function, in which the marginal temporal process has an autoregressive or autoregressive-moving average structure. The following theorem will provide a sufficient and necessary condition for the proposed function to satisfy (1.2), and therefore become a valid covariance spatio-temporal function. The proof is based on the well-known Bochner's Theorem (Rudin (1963)) and is given in the Chapter 1 Appendix 1.6.

Theorem 1.4.1. *Assume that* ν, $\alpha_1, \alpha_2, \beta_1$, *and* β_2 *are constants with* $\nu > 0$, $0 < \alpha_1 < \alpha_2$ *and* $-1 < \beta_1 < \beta_2 < 1$. *A necessary and sufficient condition for the function*

$$C(\mathbf{s};t) = \theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|)\beta_1^{|t|} + (1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|)\beta_2^{|t|},\ \mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{Z}, \quad (1.22)$$

to be a stationary covariance function on $\mathbb{R}^d \times \mathbb{Z}$ *is that the constant* θ *satisfies*

$$\left\{1 - \frac{\alpha_2^d}{\alpha_1^d}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1} \le \theta \le \left\{1 - \frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right\}^{-1}. \quad (1.23)$$

The spatial margin of (1.22) is obtained by taking $t = 0$,

$$C(\mathbf{s};0) = \theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|) + (1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|),\ \mathbf{s}\in\mathbb{R}^d.$$

The temporal margin of (1.22) is

$$C(\mathbf{0};t) = \theta\beta_1^{|t|} + (1-\theta)\beta_2^{|t|},\ t\in\mathbb{Z},$$

with the restriction (1.23) for the parameter θ, and is a linear combination of correlation functions of two univariate first-order autoregressive (AR) time series, which includes families of correlation functions of stationary AR(1), AR(2), and ARMA (2, 1) time series.

Roughly speaking, α_i's can be looked at as the scaling parameter for the spatial covariance exponential model. The β_i's are the corresponding coefficients when fitting an autoregressive time series and θ plays as a balancing parameter based on strength of both space and time interaction. If $\beta_1 = \beta_2$ then the temporal margin has an AR(1) structure and (1.22) becomes a separable model. In the case when the time lag is zero the spatial marginal is given by,

$$C(\mathbf{s}; 0) = \theta \exp(-\alpha_1 \|\mathbf{s}\|) + (1 - \theta) \exp(-\alpha_2 \|\mathbf{s}\|), \ \mathbf{s} \in \mathbb{R}^d,$$

which is a mixture of two spatial Matérn models. In the special case where $\alpha_1 = \alpha_2$ the spatial margin is reduced to a single Matérn model. Example 1.4.1 yields results when $\nu = \frac{1}{2}$; exponential case. When applying this model we can use time series techniques to fit independent time series for each location developing ARMA order and starting values for β_1, β_2, and θ, so the final parameter estimation can be achieved by maximum likelihood estimation or weighted least square estimation. For the spatial aspect we can use spatial statistics procedures to find starting values for α_1 and α_2. The advantage here is that we can employ well-established time series techniques, such as ACF and PACF to determine the model patterns and orders since the temporal margin is treated as ordinary time series. This is worth mentioning due to increasing demand on the statistical technique for the model selection and justification given all these different theoretical models developed in continuous space and time, e.g. Gneiting (2002b), Ma (2003), and Gneiting et al. (2007).

The permissible domain of θ in (1.22) contains the interval [0, 1] as a subset, in which case (1.22) is a convex combination of two separable space-time covariance functions. For other values of θ, (1.22) is the difference of two separable space-time covariance functions. The left-hand bound of (1.23) involves the dimensional parameter d, while its right-hand bound does not. When d gets large, the interval of (1.23) becomes narrow, and when $d \to \infty$, the left-hand bound of (1.23) tends to 0. The right-hand bound of (1.23) depends on the smoothness parameter ν for the spatial component, while the left-hand bound does not. When ν is large so that the model is spatially smooth, the interval of (1.23) becomes

narrow.

An embedding problem arises when the whole or a part of the space-time domain of a covariance function is over a lattice. More precisely, the question is: can we embed the covariance function (1.22) into a covariance function whose space-time domain is $\mathbb{R}^d \times \mathbb{R}$? To this end, it seems better to compare (1.22) with the following covariance function whose temporal domain is $\mathbb{R}$, which differs from that of (1.22),

$$C(\mathbf{s};t) = \theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|)\beta_1^{|t|} + (1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|)\beta_2^{|t|},\ \mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{R}, \quad (1.24)$$

where β_1 and β_2 are restricted to be positive with $0<\beta_1<\beta_2<1$. Theorem 3 of Ma (2005) with $\nu_2=\frac{1}{2}$ gives the permissible condition of θ as

$$\left\{1-\frac{\alpha_2^d}{\alpha_1^d}\frac{\ln\beta_2}{\ln\beta_1}\right\}^{-1} \le \theta \le \left\{1-\frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{\ln\beta_1}{\ln\beta_2}\right\}^{-1}. \quad (1.25)$$

It is easy to verify that

$$\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)} \le \frac{-\ln\beta_1}{-\ln\beta_2},\ 0<\beta_1<\beta_2<1.$$

Thus, the permissible interval of θ in (1.23) is included in that in (1.25), namely,

$$\begin{aligned}&\left[\left\{1-\frac{\alpha_2^d}{\alpha_1^d}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1},\ \left\{1-\frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right\}^{-1}\right]\\ \subset\ &\left[\left\{1-\frac{\alpha_2^d}{\alpha_1^d}\frac{\ln\beta_2}{\ln\beta_1}\right\}^{-1},\ \left\{1-\frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{\ln\beta_1}{\ln\beta_2}\right\}^{-1}\right].\end{aligned}$$

This means that a stationary random field on $\mathbb{R}^d\times\mathbb{Z}$ with covariance (1.22) can be embedded into a stationary random field on $\mathbb{R}^d\times\mathbb{R}$ with covariance (1.24) when β_1 and β_2 are both positive. However, this is not allowable if β_1 or β_2 are negative, in which case (1.24) would not be real-valued. It is unclear whether (1.22) can be embedded into a real-valued, stationary covariance function on $\mathbb{R}^d\times\mathbb{R}$ when β_1 or β_2 is negative. Moreover, some of the models for discrete time can not be expanded to continuous time which covers the discrete one as a constrained version on discrete domain, which is so-called embedding problem. For instance,

the autocorrelation function (ACF) of AR(1) process is $\beta^{|h|}$ with $-1 < \beta < 0$ and $h \in \mathbb{Z}$, but β^t, $t \in \mathbb{R}$ is not a valid ACF for continuous case.

The following is a special case when $\nu = \frac{1}{2}$, producing a spatio-temporal covariance function with an exponential spatial margin and a ARAM-type temporal margin.

Corollary 1.4.1. *Assume that $\alpha_1, \alpha_2, \beta_1$, and β_2 are constants with $0 < \alpha_1 < \alpha_2$ and $-1 < \beta_1 < \beta_2 < 1$. A necessary and sufficient condition for the function*

$$C(\mathbf{s}; t) = \theta \exp(-\alpha_1 \|\mathbf{s}\|)\beta_1^{|t|} + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2^{|t|}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{Z}, \tag{1.26}$$

to be a stationary covariance function on $\mathbb{R}^d \times \mathbb{Z}$ is that the constant θ satisfies

$$\left\{1 - \frac{\alpha_2^d}{\alpha_1^d}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1} \leq \theta \leq \left\{1 - \frac{\alpha_1}{\alpha_2}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right\}^{-1}. \tag{1.27}$$

The complete proof of Corollary 1.4.1 is given in Demel and Du (2011). This corollary is used for the Irish Wind data analysis in Section 1.5 with a nugget effect extension.

Another advantage of the proposed model is that the simple structure of the model gives intuitive meaning for each component, which eases the cumbersome task of determining the appropriateness of the model. The ARMA process can be easily interpreted and techniques for estimating parameters are well- understood. This leads us to use these techniques to find starting values to fit the over space-time covariance function presented. All these are not readily shared in general continuous case. For example, Gneiting (2002b) states that his model relies on two functions which must be complete monotone or that the derivatives are monotone. A researcher has to both determine these functions, verify their conditions, and obtain that the resulting covariance function is positive definite. There is no clear defined method proposed by Gneiting for determining these functions, which can cause some difficulties in actual application. Although this is a challenge to the researcher, Gneiting's model is very versatile in fitting spatio-temporal data. Our proposed model presented can serve as an attempt in seeking a more straightforward approach to study spatial-temporal data where at each location the temporal process can be modeled with a ARMA type covariance structure.

1.4.1 Asymmetric Covariance Functions

As mentioned earlier the need of an asymmetric covariance model comes into play when taking into account the underlying space-time effects caused by natural occurring forces. To construct stationary functions that are not fully symmetric, which means $C(\mathbf{s},t) = C(\mathbf{s},-t) = C(-\mathbf{s},t) = C(-\mathbf{s},-t)$ is not necessarily true. Gneiting et al. (2007) used the general idea of a Lagrangian reference frame (May and Julien (1998)). This Lagrangian framework can be thought of as modeling the center of an air, water, or thermal mass. The resulting spatio-temporal random field has stationary covariance, $C(\mathbf{s};t) = EC_S(\mathbf{s}-\mathbf{V}t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$, where C_S is a valid spatial covariance function and $\mathbf{V} \in \mathbb{R}^d$ is a random velocity vector. This random velocity vector has various choices depending on the physicality of the data and can be justified when included in the model. When analyzing the Irish Wind Data, Gneiting et al. (2007), noted that Ireland has a prevailing westerly wind and that the simplest case is when $\mathbf{V} = v$ is constant and represents the mean or prevailing wind." Applying this knowledge, they form a special case of the the Lagrangian which has the form,

$$C_{LGR}(\mathbf{s},t) = \left(1 - \frac{1}{2v}|s_{long} - vt|\right)_+, \tag{1.28}$$

where the spatial separation vector $\mathbf{s} = (s_{long}, s_{lat})'$ has longitudinal lag (east-west) component s_{long} and latitudinal lag (north-south) component s_{lat}, and $v \in \mathbb{R}$ is a longitudinal velocity. Recall that Gneiting's Model assumes a continuous frame for time while the proposed model in Corollary (1.4.1) has a discrete time domain. Based on the fact that a correlation function on $\mathbb{R}$ restricted to a discrete domain subset of itself, say $\mathbb{Z}$, will still be a valid correlation function on that discrete domain. Therefore, to apply the Lagrangian reference frame and enrich our model to accommodate asymmetry property, we only have to change the time domain from $\mathbb{R}$ to $\mathbb{Z}$. Then make the convex combination of the built model and the asymmetric model (1.28) to create

$$\begin{aligned} C(\mathbf{s},t) = (1-\lambda)\left\{\theta\exp(-\alpha_1\|\mathbf{s}\|)\beta_1^{|t|} + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2^{|t|}\right\} \\ + \lambda\left(1 - \frac{1}{2v}|s_{long} - vt|\right)_+ . \; 0 < \lambda < 1. \end{aligned} \tag{1.29}$$

Now we are ready to apply and compare the models above using the Irish wind dataset.

1.5 Irish Wind Data Analysis

To compare the proposed model (1.22), (i.e. (1.4.1)) with existing models presented by Gneiting (2002b), and Gneiting et al. (2007), we use the Irish wind dataset first analyzed by Haslett and Raftery (1989). This is a spatio-temporal dataset that measures average daily wind speeds taken from twelve synoptic meteorological weather stations in Ireland from year 1961 to 1978. As in Haslett and Raftery (1989), Gneiting (2002b), Stein (2005a,b), and Gneiting et al. (2007), the following steps were taken to clean the data. A square root transformation is taken to stabilize the variance over both stations and time periods so that the margins of space and time are approximately normally distributed. This allows us to use Gaussian theory for both the temporal process and the spatial process. The location of Rosslare was removed for stationarity reasons, as recommended by Haslett and Raftery (1989). The seasonal trend was fit and removed using annual harmonic regression. To extract the spatial trend the station specific means where removed. The final resulting dataset becomes time series of velocities for eleven meteorological stations, to which we will apply the proposed models (1.32) and (1.29). For evaluating the ability to predict, the data are split into a training dataset from years 1961 to 1970 and a test dataset from years 1971 to 1978. These steps were also taken by Haslett and Raftery (1989), Gneiting (2002b), Stein (2005a,b), and Gneiting et al. (2007).

1.5.1 Symmetric Covariance Models

To start the analysis we first compare the symmetric cases of the proposed models. To achieve a good fit of the data Gneiting (2002b), (see, Figure 5 there in), noted that the space-time correlation approaches zero after a lag of three days; we will utilize this in fitting our model. The symmetric spatio-temporal covariance functions that we will compare are

Gneiting's separable and non-separable models given by,

$$C_{G.SEP}(\mathbf{s};t) = \{(1-\nu)\exp(-c\|\mathbf{s}\|) + \nu\delta_{\mathbf{s}=0}\} \cdot \left\{(1+a|t|^{2\alpha})^{-1}\right\}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{R}, \quad (1.30)$$

and

$$C_{G.NSEP}(\mathbf{s};t) = \frac{1-\nu}{1+a|t|^{2\alpha}} \left\{ \exp\left(-\frac{c\|\mathbf{s}\|}{(1+a|t|^{2\alpha})^{\beta/2}} \right) + \frac{\nu}{1-\nu}\delta_{\mathbf{s}=0} \right\}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{R}, \quad (1.31)$$

and our proposed model denoted by $C_{SM}(\mathbf{s};t)$ and given by,

$$\begin{aligned} C_{SM}(\mathbf{s};t) = (1-\nu) &\left\{ \theta \exp(-\alpha_1\|\mathbf{s}\|)\beta_1^{|t|} + (1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2^{|t|} \right\} \\ &+ \nu \left\{ \theta\beta_1^{|t|} + (1-\theta)\,\beta_2^{|t|} \right\} \delta_{\mathbf{s}=0}, \ \mathbf{s} \in \mathbb{R}^d, \ t \in \mathbb{Z}, \end{aligned} \quad (1.32)$$

Again note that the above functions are in terms of a correlation function. This is done for the ease of coding and computation. Here ν is a spatial nugget effect coefficient that can also account for measurement error or micro variability; and δ is an indicator function that equals one only when spatial lag is zero. Using the same techniques mentioned in Section 1.4 we first look at space and time independently of each other. From Figure 1.7 we can see that it is reasonable that the temporal margin for each station could possibly have an AR(1) or AR(2) structure. Recall that by allowing $\beta_1 = \beta_2$ in (1.32) we obtain an AR(1) temporal margin structure. For comparison we fit both the AR(1) and AR(2) cases. Compared to the autoregressive structure of proposed model (1.32) which accounts for the discrete time points, Gneiting (2002b) uses a Cauchy type function to model the temporal margin without formal justification other than using empirical time correlation. Again note that Gneiting's methods have to chose a function to fit both the temporal process and the spatial process; and verify the nonnegative definiteness. The tools from time series allow us to make the choice of AR(1) or AR(2). Both methods use weighted least squares to fit the temporal process in order to find starting values for temporal related parameters for the covariance functions. Notice in Figure 1.8 that both Gneiting's Cauchy and the presented AR(2) fit the empirical temporal correlation quite well. To fit (1.32) spatially we used the $\alpha_1 = \alpha_2$ simplification to obtain a single exponential. Gneiting also uses a exponential structure for

fitting the spatial correlation. Looking at Figure 1.8 the spatial correlation plot, notice that Gneiting's and the proposed model fit are identical.

Based on the simulation study of the moving average case, parameter estimates were obtained using Cressie's weighted least squares procedure. Gneiting also used this approach in fitting his models. After we find the parameter estimates for the models we compare the empirical correlations with the fitted correlations for each model. The empirical correlation is calculated by finding the cross-correlation over time between all stations. The parameter estimates for Gneiting's models (1.30) and (1.31), as well as the proposed model (1.32) are given in Table 1.5.1.

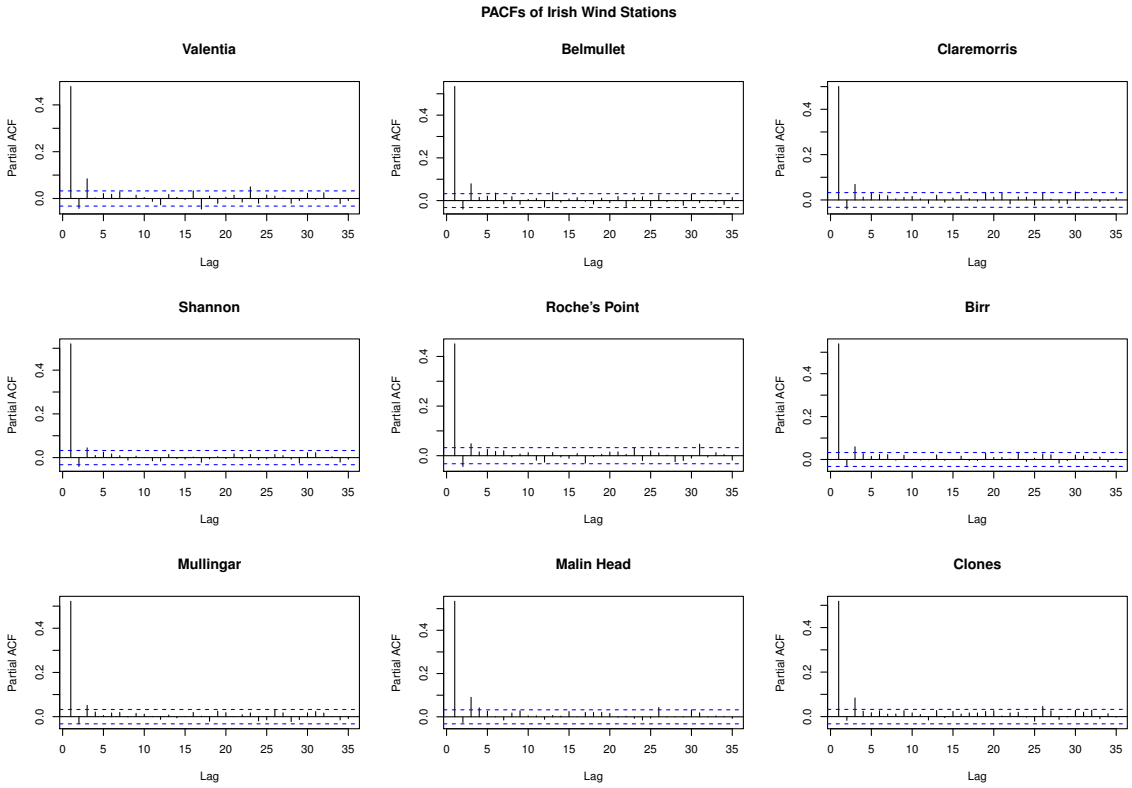

Figure 1.7: Partial autocorrelation function of selected locations in the Irish wind data.

From Figure 1.9 we can evaluate plots of fitted correlation versus empirical correlation. For spatial correlation which is given by $C(\mathbf{s}; 0)$ (black). Using the second set of four plots in Figure 1.9 we can see that the models fit the empirical correlations quite well at a time lag of 0. At a time lag of one (red) we see that all models start to under predict the empirical

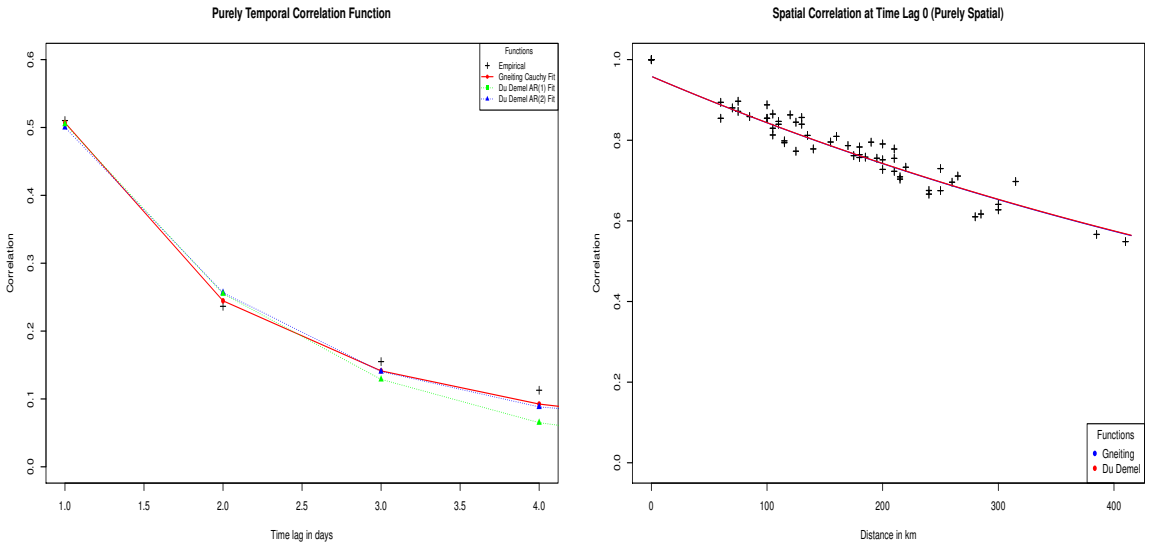

Figure 1.8: Marginal temporal fits at fixed location (left) and spatial correlation of stations at fixed time (right) and . Plot based of years (1961-1970)

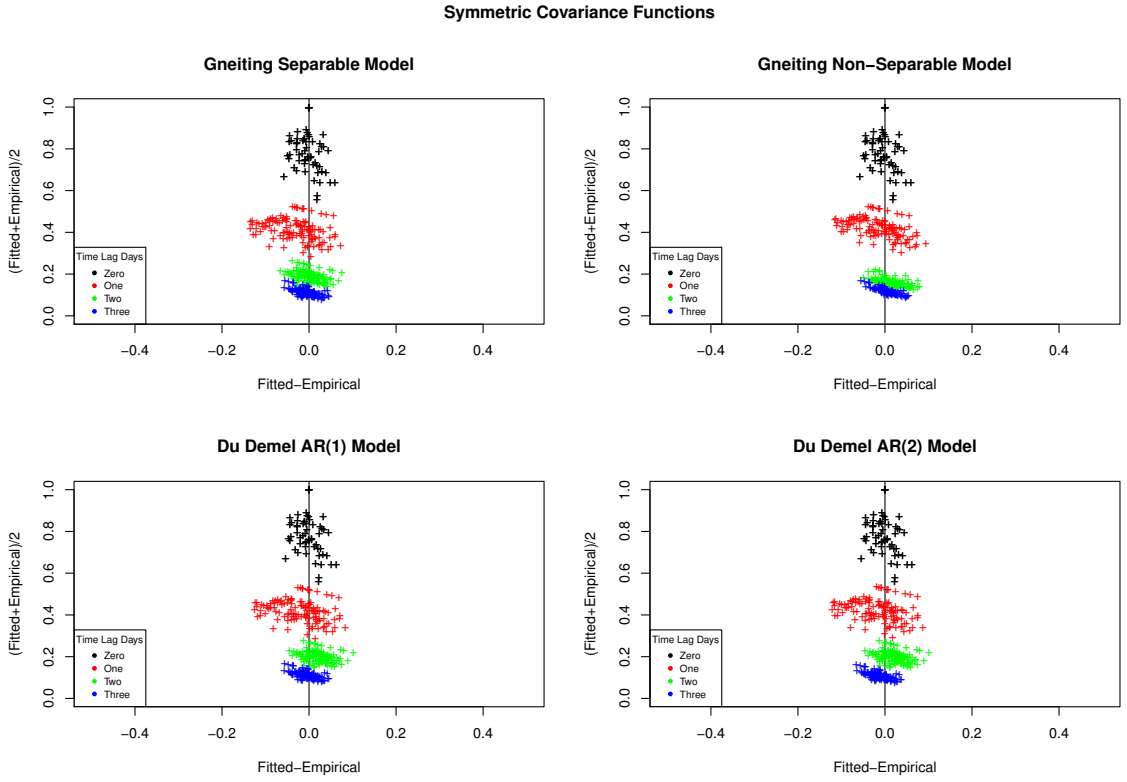

Figure 1.9: Left set of four: Empirical space-time correlations versus model fitted space-time correlation. Right set of four: Fitted - Empirical versus (Fitted + Empirical)/2. Plot based of years (1961-1970)

Table 1.8: *Irish Wind Parameter Estimates*

G.SEP	G.NSEP	$C_{SM}(\mathbf{s};t)$ AR(1)	$C_{SM}(\mathbf{s};t)$ AR(2)
$\nu = 0.04153$	$\nu = 0.04153$	$\nu = 0.04267$	$\nu = 0.04312$
$c = 0.00128$	$c = 0.00128$	$\alpha_1 = 0.00126$	$\alpha_1 = 0.00127$
$a = 0.97200$	$a = 0.97200$	$\theta = 1.4318$	$\theta = 1.07324$
$\alpha = 0.83400$	$\alpha = 0.83400$	$\beta_1 = 0.52002$	$\beta_1 = -0.16108$
	$\beta = 0.68060$		$\beta_2 = 0.48623$

[1] G.SEP is Gneiting's separable model (1.30)
[2] G.NSEP is Gneiting's non-Separable model (1.31)
[3] $C_{SM}(\mathbf{s};t)$ AR(1) is model (1.32) taking $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$
[4] $C_{SM}(\mathbf{s};t)$ AR(2) is model (1.32) taking $\alpha_1 = \alpha_2$

correlation. The second set of plots confirm this conclusion. As the time lag increases to two and three, (green and blue), we see the models start to behave better, but still have a lack of fit when compared to the empirical. To evaluate the prediction capabilities of the models the root-mean-square errors (RMSEs) were calculated between the fitted based on three days of lag and the test data set. When comparing these results the smaller the better. Each column corresponds to a given model, here the Empirical Model use the training empirical correlations to predict the future correlations. Roughly this is how well the data can predict itself through the calculated empirical correlations. Gneiting noted in his papers that the empirical model does have the smallest RMSE overall, but these predictions are only valid at the given locations. We can not use this model to predict at unknown locations. Recall that under model (1.32) in the AR(1) case, the model becomes separable, and comparing it to Gneiting's separable model (1.30) we find the results comparable. The same is true when looking at model (1.32) in the AR(2) case and Gneiting's non-separable model (1.31). Again the presented model is an alternative to Gneiting's that is easier to apply in that the researcher does not have to chose unknown functions to build the covariance structure.

We also calculated the likelihood using the three days of lag and then averaging over each set of three days, which can be seen in Table 1.10. In all cases the log likelihoods are comparable. So for the Irish wind data the proposed models in the AR(1) and AR(2) cases

Table 1.9: *Irish Wind: RMSEs for one-day ahead predictions based on three previous days*

Station	G.SEP	$C_{SM}(\mathbf{s};t)$AR(1)	G.NSEP	$C_{SM}(\mathbf{s};t)$AR(2)	Empirical
Valentia	0.50052	0.50040	0.50100	0.50074	0.49960
Belmullet	0.49496	0.49553	0.49534	0.49609	0.49396
Claremorris	0.49093	0.49180	0.49157	0.49178	0.48589
Shannon	0.46753	0.46724	0.46834	0.46740	0.45405
RochesPoint	0.48321	0.48337	0.47871	0.48428	0.46475
Birr	0.47713	0.47684	0.47597	0.47744	0.46239
Mullingar	0.42742	0.42750	0.42427	0.42747	0.41433
MalinHead	0.49635	0.49663	0.49191	0.49727	0.47913
Kilkenny	0.43887	0.43937	0.43584	0.43969	0.41412
Clones	0.48634	0.48680	0.48389	0.48710	0.46574
Dublin	0.44986	0.44973	0.44528	0.44903	0.42705

Table 1.10: *Irish Wind: Log Likelihoods; symmetric case*

G.SEP	$C_{SM}(\mathbf{s};t)AR(1)$	G.NSEP	$C_{SM}(\mathbf{s};t)AR(2)$	Empirical
19.1049	19.1574	18.6119	19.2153	19.664738

are very competitive with Gneiting's models. Note that Gneiting's models are very versatile, but our models provide nearly identical results with a procedure that is easier to apply and more intuitive when the temporal margin is close to an AR(1), AR(2), or ARMA(2, 1) structure. The ease of the application does not come from fitting, both Gneiting's and the presented models are fitted the same way. The ease comes from using the exploratory analysis to gain parameter estimates for the models. The presented model also utilizes basic time series techniques to justify the use of an AR(1), AR(2), or ARMA(2, 1) structure. The model proposed can use classical ARMA interpretation of the temporal related parameters, while this is not a straightforward manner in some theoretical continuous space-time models.

1.5.2 Asymmetric Covariance Models

Now we will take into account asymmetric cases and compare Gneiting's general stationary model,

$$C_{G.STAT}(\mathbf{s};t) = \frac{(1-\lambda)(1-\nu)}{1+a|t|^{2\alpha}}\left\{\exp\left(-\frac{c\|\mathbf{s}\|}{(1+a|t|^{2\alpha})^{\beta/2}}\right)+\frac{\nu}{1-\nu}\delta_{\mathbf{s}=0}\right\} + \lambda\left(1-\frac{1}{2v}|s_{long}-vt|\right)_{+} \quad \mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{R}, \tag{1.33}$$

versus the asymmetric case of the proposed model denoted by $C_{ASYM}(\mathbf{s};t)$, and given by,

$$\begin{aligned} C_{ASYM}(\mathbf{s};t) = (1-\lambda)\Big[&(1-\nu)\left\{\theta\exp(-\alpha_1\|\mathbf{s}\|)\beta_1^{|t|}+(1-\theta)\exp(-\alpha_2\|\mathbf{s}\|)\beta_2^{|t|}\right\} \\ &+\ \nu\left\{\theta\beta_1^{|t|}+(1-\theta)\,\beta_2^{|t|}\right\}\delta_{\mathbf{s}=0}\Big] \\ &+\ \lambda\left(1-\frac{1}{2v}|s_{long}-vt|\right)_{+} \quad \mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{Z}. \end{aligned} \tag{1.34}$$

In the asymmetric case we see the addition of two new parameters λ, the weight parameter associated with the Lagrangian reference, and v is the speed in which the westerly wind systems move, measured in kilometers per day. Here we use the existing parameter estimates that we found in the symmetric case and use weighted least squares and obtain fitted estimates $\widehat{\lambda} = 234$ kilometers per day and $\widehat{v} = 0.0573$ as did in Gneiting et al. (2007).

From Figure 1.10 we can evaluate plots of fitted correlation versus empirical correlation. Spatial at a time lag of zero (black) all the models fit rather well for both Gneiting's and our model (1.34). At a time lag of one (red) we see that Gneiting's model (1.33) and the model (1.34) fits are about the same. At a lag of two Gneiting's model is slightly better than model (1.34). At three days the model both asymmetrically models seem to have a similar fit. Again we conclude that the model (1.34) is competitive to Gneiting's methods. To compare prediction capability the RMSEs were calculated between the fitted based on three days of lag and the test dataset. When comparing these results the smaller RMSE indicates better prediction.

Notice that again in Table 1.5.2 we see similar results as the symmetric case, in that our model (1.34) is comparable to Gneiting's model (1.33). If we compare Tables 1.5.1 and

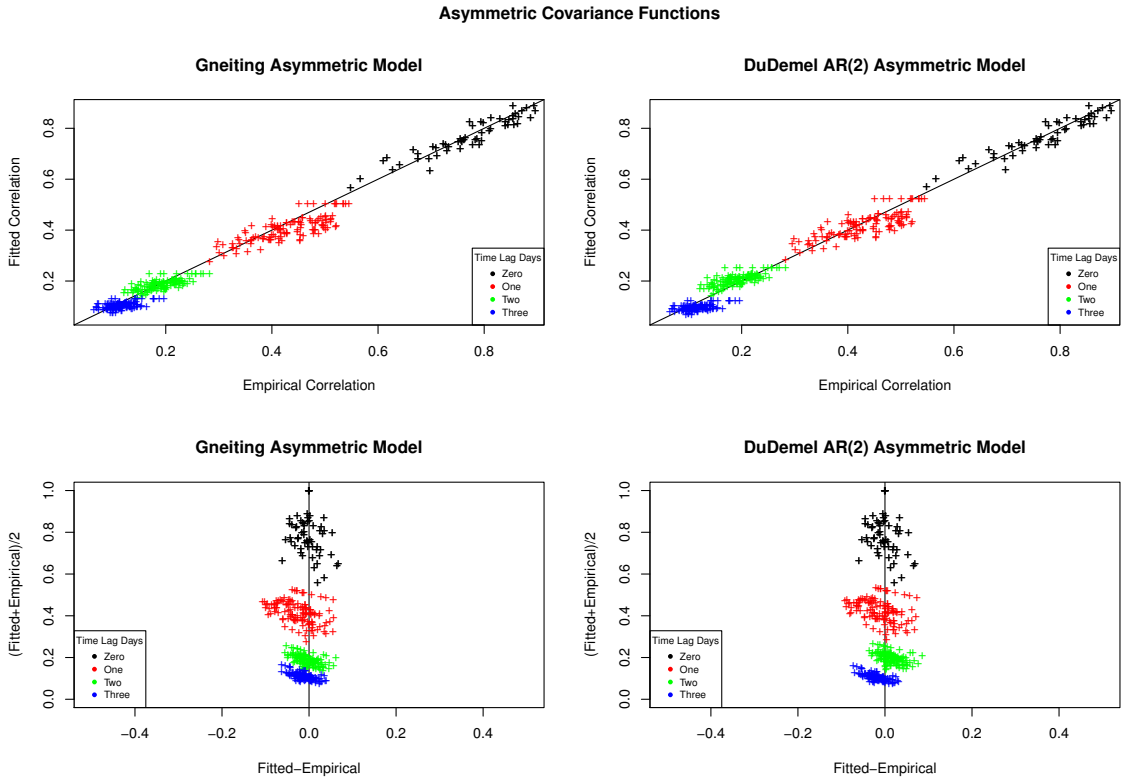

Figure 1.10: Upper: Empirical versus fitted correlation Gneiting's stationary model (1.33) (left) and $C_{ASYM}(\mathbf{s};t)$ AR(2) (right). Lower: Fitted - Empirical versus (Fitted + Empirical)/2. Plot based of years (1961-1970)

1.5.2, notice that Gneiting's model (1.33) has the lowest RMSEs, but our model (1.34) is very close. When applying the asymmetric case with the proposed model (1.34) note that the temporal margin is no longer autoregressive in its nature. This could be why our model (1.34) does not improve over the model (1.32). We also calculated the likelihood using the three days of lag and then averaging over each set of three days, which can be seen in Table 1.12, in which case our model (1.34) again is very comparable to Gneiting's.

After analyzing both the symmetric and asymmetric cases we can see that our models (1.32) and (1.34) are highly comparable to Gneiting's models (1.30), (1.31), and (1.33) on all accounts. Gneiting's models are very versatile, but the models we proposed here provide nearly identical predictions and likelihoods with a procedure that is easier to apply and more

Table 1.11: *Irish Wind: RMSEs for one-day ahead predictions based on three previous days*

Station	G.STAT	$C_{ASYM}(\mathbf{s};t)$ AR(2)	Empirical
Valentia	0.4989789	0.5041726	0.4995971
Belmullet	0.4945723	0.4982614	0.4939611
Claremorris	0.4896983	0.4938673	0.4858922
Shannon	0.4657336	0.4694637	0.4540456
RochesPoint	0.4737622	0.4857265	0.4647548
Birr	0.4724635	0.4793497	0.4623907
Mullingar	0.4193231	0.4291690	0.4143287
MalinHead	0.4878435	0.4984323	0.4791251
Kilkenny	0.4294301	0.4401993	0.4141247
Clones	0.4788818	0.4879050	0.4657351
Dublin	0.4400079	0.4494564	0.4270464

[1] G.STAT is Gneiting's General Stationary (Asymmetric) Model (1.33)
[2] $C_{ASYM}(\mathbf{s};t)$ AR(2) ASYM is model (1.34) taking $\alpha_1 = \alpha_2$

Table 1.12: *Irish Wind: Log Likelihoods; asymmetric case*

MODELS	G.STAT	$C_{ASYM}(\mathbf{s};t)$ AR(2)	Empirical
Log likelihood	18.267072	18.484894	19.664738

intuitive when the temporal margin is of an ARMA type time series. The advantage here is that our model takes into account the fact that the data are collected in time and are often considered as a realization of a time series. There are well developed techniques that can be borrowed to aid the choice of the models rather than trying to find an unknown function that fits the empirical time correlation without much justification. In this sense we believe our proposed models is in the direction of trying to utilize convenient techniques in time series and spatial statistics for practical continuous space and discrete time data modeling, which covers a even larger class of models which may not have continuous counterparts. In addition, the relative ease of interpretation of parameter estimation stems from being able to use those properties studied for time series.

1.6 Chapter 1 Appendix

The following contains proofs for theorems discussed in Chapter 1

Bochner's Theorem. *Suppose that C is a continuous on $\mathbb{R}^d \times \mathbb{R}$. Then C is a covariance function if and only if it is of the form*

$$C(\mathbf{s};t) = \int\int e^{i(\mathbf{s}'\boldsymbol{\omega}+t\tau)} dF(\boldsymbol{\omega},\tau), (\mathbf{s};t) \in \mathbb{R}^d \times \mathbb{R}, \tag{1.35}$$

where F is a finite, non-negative measure on $\mathbb{R}^d \times \mathbb{R}$.

This allows us to use Fourier transforms to aid us in the proofs for nonnegative definiteness for the proposed covariance functions.

1.6.1 Proof of Theorem 1.3.1

With Lemma 1.3.1 ; we only need to show the necessary part. Being a covariance function on $\mathbb{R}^d$, (1.3) is positive definite; that is, the inequality

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(\mathbf{s}_i, \mathbf{s}_j; t_i - t_j) \geq 0$$

holds for every positive integer n, any real numbers a_k, and any $\mathbf{s}_k \in \mathbb{R}^d, t_k \in \mathbb{N}$, $k = 1, \ldots, n$. In particular, for two arbitrary positive integers n and m, choose arbitrarily points $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and points $t_1, \ldots, t_m$, from which to form nm pairs $\mathbf{s}_i$ and t_j, and then for corresponding coefficients take the products $a_i b_j$, $i = 1, \ldots, n, j = 1, \ldots, m$. We obtain

$$\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{m}\sum_{j'=1}^{m} a_i a_{i'} b_j b_{j'} C(\mathbf{s}_i, \mathbf{s}_{i'}; t_j - t_{j'}) \geq 0. \tag{1.36}$$

Define

$$C_1(t) = \begin{cases} 1, & t = 0, \\ \frac{1}{2}, & t = \pm 1, \\ 0, & t = \pm 2, \pm 3, \ldots, \end{cases} \quad \text{and } C_2(t) = \begin{cases} 1, & t = 0, \\ -\frac{1}{2}, & t = \pm 1, \\ 0, & t = \pm 2, \pm 3, \ldots, \end{cases}$$

which are the covariances of the first-order moving average processes. Then inequality (1.36) is rewritten as, with a factor $\frac{1}{2}$ dropped,

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} b_j b_{j'} C_1(t_j - t_{j'}) \\ + &\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} b_j b_{j'} C_2(t_j - t_{j'}) \geq 0. \end{aligned} \tag{1.37}$$

Particularly selecting $t_j = j, j = 1, \ldots, m$, and letting b_j take values from the sets $\{\cos(\omega), \ldots, \cos(m\omega)\}$ and $\{\sin(\omega), \ldots, \sin(m\omega)\}$ respectively, with $\omega \in \mathbb{R}$, (1.37) becomes

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \cos(j\omega)\cos(j'\omega) C_1(j - j') \\ + &\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \cos(j\omega)\cos(j'\omega) C_2(j - j') \geq 0, \end{aligned}$$

and

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \sin(j\omega)\sin(j'\omega) C_1(j - j') \\ + &\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \sin(j\omega)\sin(j'\omega) C_2(j - j') \geq 0. \end{aligned}$$

Adding the last two inequalities together yields

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \cos((j - j')\omega) C_1(j - j') \\ + &\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \sum_{j=1}^{m}\sum_{j'=1}^{m} \cos((j - j')\omega) C_2(j - j') \geq 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned}&\left\{1 + \left(1 - \frac{1}{m}\right)\cos\omega\right\} \sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) \\ + &\left\{1 - \left(1 - \frac{1}{m}\right)\cos\omega\right\} \sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \geq 0. \end{aligned}$$

Letting m tend to infinity, we obtain

$$(1 + \cos\omega) \sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i,\mathbf{s}_{i'}) + (1 - \cos\omega) \sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_-(\mathbf{s}_i,\mathbf{s}_{i'}) \geq 0. \tag{1.38}$$

It then follows by choosing $\omega = 0$ in (1.38) that

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} a_i a_{i'} C_+(\mathbf{s}_i, \mathbf{s}_{i'}) \geq 0,$$

which means that $C_+(\mathbf{s}_1, \mathbf{s}_2)$ is positive definite. Similarly, taking $\omega = \pi$ in (1.38) yields that $C_-(\mathbf{s}_1, \mathbf{s}_2)$ is positive definite. This concludes the proof. ■

1.6.2 Proof of Theorem 1.3.2

We have

$$\begin{aligned}\psi_0(\mathbf{s}) \pm 2\psi_1(\mathbf{s}) &= \theta(\alpha_1\|\mathbf{s}\|)^\nu K_\nu(\alpha_1\|\mathbf{s}\|)(1 \pm 2\beta_1) \\ &\quad +(1-\theta)(\alpha_2\|\mathbf{s}\|)^\nu K_\nu(\alpha_2\|\mathbf{s}\|)(1 \pm 2\beta_2),\ \mathbf{s} \in \mathbb{R}^d.\end{aligned}$$

Notice that

$$\int_{\mathbb{R}^d} e^{-\imath\mathbf{s}'\boldsymbol{\omega}}(\alpha_k\|\mathbf{s}\|)^\nu K_\nu(\alpha_k\|\mathbf{s}\|)d\mathbf{s} = c_0\alpha_k^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_k^2)^{-\nu-\frac{d}{2}},\ \boldsymbol{\omega} \in \mathbb{R}^d,$$

where c_0 is a positive constant not related to α_k (see Eq. (4.130) of Yaglom (1987)), $k = 1, 2$. We obtain the Fourier transforms of $\psi_0(\mathbf{s})+2\psi_1(\mathbf{s})$ and $\psi_0(\mathbf{s})-2\psi_1(\mathbf{s})$ positively proportional to

$$f_1(\boldsymbol{\omega}) = \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{-\nu-\frac{d}{2}}(1+2\beta_1) + (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}(1+2\beta_2),$$

and

$$f_2(\boldsymbol{\omega}) = \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{-\nu-\frac{d}{2}}(1-2\beta_1) + (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}(1-2\beta_2),\ \boldsymbol{\omega} \in \mathbb{R}^d,$$

respectively. Hence, it suffices to show that inequalities (1.9) are necessary and sufficient for $f_1(\boldsymbol{\omega}) \geq 0$ and $f_2(\boldsymbol{\omega}) \geq 0$, $\boldsymbol{\omega} \in \mathbb{R}^d$.

Suppose that $f_1(\boldsymbol{\omega}) \geq 0$ and $f_2(\boldsymbol{\omega}) \geq 0$ hold for every $\boldsymbol{\omega} \in \mathbb{R}^d$. Since $0 < \alpha_1 < \alpha_2$ and $0 \leq 1 + 2\beta_1 \leq 1 + 2\beta_2$, it follows from

$$0 \leq \lim_{\boldsymbol{\omega}\to\infty}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{\nu+\frac{d}{2}} f_1(\boldsymbol{\omega}) = \theta\alpha_1^{2\nu}(1+2\beta_1) + (1-\theta)\alpha_2^{2\nu}(1+2\beta_2) \tag{1.39}$$

that

$$\theta \leq \left\{1 - \frac{\alpha_1^{2\nu}(1+2\beta_1)}{\alpha_2^{2\nu}(1+2\beta_2)}\right\}^{-1},$$

and from

$$0 \leq f_2(\mathbf{0}) = \theta\alpha_1^{-d}(1-2\beta_1) + (1-\theta)\alpha_2^{-d}(1-2\beta_2), \tag{1.40}$$

that

$$\theta \geq \left\{1 - \frac{\alpha_2^{d}(1-2\beta_1)}{\alpha_1^{d}(1-2\beta_2)}\right\}^{-1}.$$

Consequently, inequalities (1.9) are necessary for $f_1(\boldsymbol{\omega}) \geq 0$ and $f_2(\boldsymbol{\omega}) \geq 0$, $\boldsymbol{\omega} \in \mathbb{R}^d$.

On the other hand, we are going to show that, under inequalities (1.9), $f_1(\boldsymbol{\omega}) \geq 0$ and $f_2(\boldsymbol{\omega}) \geq 0$ hold for every $\boldsymbol{\omega} \in \mathbb{R}^d$. While this is obviously true if $0 \leq \theta \leq 1$, it remains to consider the cases

$$\left\{1 - \frac{\alpha_2^{d}(1-2\beta_1)}{\alpha_1^{d}(1-2\beta_2)}\right\}^{-1} \leq \theta \leq 0, \text{ and } 1 \leq \theta \leq \left\{1 - \frac{\alpha_1^{2\nu}(1+2\beta_1)}{\alpha_2^{2\nu}(1+2\beta_2)}\right\}^{-1}.$$

Case $\left\{1 - \frac{\alpha_2^{d}(1-2\beta_1)}{\alpha_1^{d}(1-2\beta_2)}\right\}^{-1} \leq \theta \leq 0$: In this case, $1-\theta$ is positive and (1.40) is valid. Since $0 < \alpha_1 < \alpha_2$ implies

$$\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2 + \alpha_2^2}\right)^{\nu+\frac{d}{2}} \geq \left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}},$$

we obtain

$$\begin{aligned} f_2(\boldsymbol{\omega}) &= \theta\alpha_1^{-d}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}}(1-2\beta_1) + (1-\theta)\alpha_2^{-d}\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2 + \alpha_2^2}\right)^{\nu+\frac{d}{2}}(1-2\beta_2) \\ &\geq \{\theta\alpha_1^{-d}(1-2\beta_1) + (1-\theta)\alpha_2^{-d}(1-2\beta_2)\}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}} \\ &\geq 0, \end{aligned}$$

and since $0 \leq 1+2\beta_1 < 1+2\beta_2$,

$$\begin{aligned} f_1(\boldsymbol{\omega}) &\geq \{\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{-\nu-\frac{d}{2}} + (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}\}(1+2\beta_1) \\ &= \left\{\theta\alpha_1^{-d}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}} + (1-\theta)\alpha_2^{-d}\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2 + \alpha_2^2}\right)^{\nu+\frac{d}{2}}\right\}(1+2\beta_1) \\ &\geq \{\theta\alpha_1^{-d} + (1-\theta)\alpha_2^{-d}\}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}}(1+2\beta_1) \\ &\geq 0, \ \boldsymbol{\omega} \in \mathbb{R}^d, \end{aligned}$$

where the last inequality is obtained from

$$\theta \geq \left\{1 - \frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\right\}^{-1} \geq \left(1 - \frac{\alpha_2^d}{\alpha_1^d}\right)^{-1}.$$

Case $1 \leq \theta \leq \left\{1 - \frac{\alpha_1^{2\nu}(1+2\beta_1)}{\alpha_2^{2\nu}(1+2\beta_2)}\right\}^{-1}$: In this case, θ is positive and (1.39) is valid. Thus

$$\begin{aligned} f_1(\boldsymbol{\omega}) &\geq \{\theta\alpha_1^{2\nu}(1+2\beta_1) + (1-\theta)\alpha_2^{2\nu}(1+2\beta_2)\}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}} \\ &\geq 0, \end{aligned}$$

and

$$\begin{aligned} f_2(\boldsymbol{\omega}) &\geq \{\theta\alpha_1^{2\nu} + (1-\theta)\alpha_2^{2\nu}\}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}(1-2\beta_2) \\ &\geq 0,\ \boldsymbol{\omega} \in \mathbb{R}^d, \end{aligned}$$

where the last inequality is due to

$$\theta \leq \left\{1 - \frac{\alpha_1^{2\nu}(1+2\beta_1)}{\alpha_2^{2\nu}(1+2\beta_2)}\right\}^{-1} \leq \left(1 - \frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\right)^{-1}.$$

This concludes the proof ■

1.6.3 Proof of Theorem 1.4.1

Notice that, for $k = 1, 2$,

$$\sum_{t=0}^{\infty} \beta_k^{|t|} e^{-\imath t\omega_0} = \frac{1-\beta_k^2}{1+\beta_k^2 - 2\beta_k\cos\omega_0},\ \omega_0 \in [-\pi, \pi],$$

and

$$\int_{\mathbb{R}^d} e^{-\imath\mathbf{s}'\boldsymbol{\omega}}(\alpha_k\|\mathbf{s}\|)^\nu K_\nu(\alpha_k\|\mathbf{s}\|)d\mathbf{s} = c_0\alpha_k^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_k^2)^{-\nu-\frac{d}{2}},\ \boldsymbol{\omega} \in \mathbb{R}^d,$$

where c_0 is a positive constant not related to α_k (see Eq. (4.130) of Yaglom (1987)). We obtain the Fourier transform of (1.22) that is positively proportional to

$$\begin{aligned} &f(\boldsymbol{\omega};\omega_0) \\ &= \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{-\nu-\frac{d}{2}}\frac{1-\beta_1^2}{1+\beta_1^2-2\beta_1\cos\omega_0} + (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}\frac{1-\beta_2^2}{1+\beta_2^2-2\beta_2\cos\omega_0}, \\ &= \{\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1^2)(1+\beta_2^2-2\beta_2\cos\omega_0) + (1-\theta)\alpha_2^{2\nu} \\ &\times(\|\boldsymbol{\omega}\|^2 + \alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2^2)(1+\beta_1^2-2\beta_1\cos\omega_0)\}\prod_{k=1}^{2}(1+\beta_k^2-2\beta_k\cos\omega_0)^{-1} \\ &= \{g_1(\boldsymbol{\omega}) - 2g_2(\boldsymbol{\omega})\cos\omega_0\}\prod_{k=1}^{2}(1+\beta_k^2-2\beta_k\cos\omega_0)^{-1},\ \boldsymbol{\omega} \in \mathbb{R}^d,\ \omega_0 \in [-\pi,\pi], \end{aligned}$$

where

$$g_1(\boldsymbol{\omega}) = \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1^2)(1+\beta_2^2)+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2^2)(1+\beta_1^2),$$

and

$$g_2(\boldsymbol{\omega}) = \theta\alpha_1^{2\nu}(1-\beta_1^2)\beta_2(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}+(1-\theta)\alpha_2^{2\nu}(1-\beta_2^2)\beta_1(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}},\ \boldsymbol{\omega}\in\mathbb{R}^d.$$

By Bochner's theorem, (1.22) is a stationary covariance function on $\mathbb{R}^d\times\mathbb{Z}$ if and only if its Fourier transform, $f(\boldsymbol{\omega};\omega_0)$, is nonnegative, or equivalently,

$$g_1(\boldsymbol{\omega})-2g_2(\boldsymbol{\omega})\cos\omega_0\geq 0,\ \boldsymbol{\omega}\in\mathbb{R}^d,\ \omega_0\in[-\pi,\pi]. \tag{1.41}$$

Moreover, inequality (1.41) holds for all $\boldsymbol{\omega}\in\mathbb{R}^d$ and $\omega_0\in[-\pi,\pi]$ if and only if

$$g_1(\boldsymbol{\omega})-2g_2(\boldsymbol{\omega})\geq 0,\ g_1(\boldsymbol{\omega})+2g_2(\boldsymbol{\omega})\geq 0,\ \boldsymbol{\omega}\in\mathbb{R}^d. \tag{1.42}$$

This is true because, on one hand we obtain (1.42) from (1.41) by simply taking $\omega_0=0$ and π in (1.41), and on the other hand, inequalities (1.42) imply

$$g_1(\boldsymbol{\omega})\geq 2|g_2(\boldsymbol{\omega})|\geq 2g_2(\boldsymbol{\omega})\cos\omega_0,\ \boldsymbol{\omega}\in\mathbb{R}^d,\ \omega_0\in[-\pi,\pi].$$

Notice that inequalities (1.42) are the same as

$$\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1^2)(1-\beta_2)^2+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2^2)(1-\beta_1)^2\geq 0,$$

and

$$\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1^2)(1+\beta_2)^2+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2^2)(1+\beta_1)^2\geq 0,$$

for all $\boldsymbol{\omega}\in\mathbb{R}^d$, or equivalently,

$$\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1+\beta_1)(1-\beta_2)+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1+\beta_2)(1-\beta_1)\geq 0, \tag{1.43}$$

and

$$\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1)(1+\beta_2)+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2)(1+\beta_1)\geq 0. \tag{1.44}$$

Hence, it suffices to show that inequalities (1.23) are necessary and sufficient for (1.43) and (1.44) to hold.

Suppose that (1.43) holds for any $\boldsymbol{\omega} \in \mathbb{R}^d$. Multiplying $(\|\boldsymbol{\omega}\|^2 + \alpha_1^2)^{\nu+\frac{d}{2}}$ both sides of (1.43) and then letting $\boldsymbol{\omega}$ tend to infinity, we obtain

$$\theta\alpha_1^{2\nu}(1+\beta_1)(1-\beta_2) + (1-\theta)\alpha_2^{2\nu}(1+\beta_2)(1-\beta_1) \geq 0, \tag{1.45}$$

or

$$\theta + (1-\theta)\frac{\alpha_2^{2\nu}}{\alpha_1^{2\nu}}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)} \geq 0,$$

which results in

$$\theta \leq \left(1 - \frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right)^{-1},$$

since $0 < \alpha_1 < \alpha_2$ and $(1+\beta_1)(1-\beta_2) \leq (1-\beta_1)(1+\beta_2)$. On the other hand, substituting $\boldsymbol{\omega} = \mathbf{0}$ in (1.44) yields

$$\theta\alpha_1^{-d}(1-\beta_1)(1+\beta_2) + (1-\theta)\alpha_2^{-d}(1-\beta_2)(1+\beta_1) \geq 0, \tag{1.46}$$

and thus

$$\theta \geq \left\{1 - \frac{\alpha_2^{d}}{\alpha_1^{d}}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1}.$$

Therefore, we conclude that inequalities (1.23) are necessary for (1.43) and (1.44) to hold.

Next we are going to show that, under inequalities (1.23), (1.43) and (1.44) hold for any $\boldsymbol{\omega} \in \mathbb{R}^d$. While this is obviously true if $0 \leq \theta \leq 1$, it remains to consider the cases

$$\left\{1 - \frac{\alpha_2^{d}}{\alpha_1^{d}}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1} \leq \theta \leq 0, \text{ and } 1 \leq \theta \leq \left(1 - \frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right)^{-1}.$$

Case $\left\{1 - \frac{\alpha_2^{d}}{\alpha_1^{d}}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1} \leq \theta \leq 0$: In this case, $1-\theta$ is positive and (1.46) is valid. Since $0 < \alpha_1 < \alpha_2$ implies

$$\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2 + \alpha_2^2}\right)^{\nu+\frac{d}{2}} \geq \left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2 + \alpha_1^2}\right)^{\nu+\frac{d}{2}},$$

inequality (1.44) follows from

$$\begin{aligned}
& \theta\alpha_1(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1)(1+\beta_2) \\
+ \;& (1-\theta)\alpha_2(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2)(1+\beta_1) \\
= \;& \theta\alpha_1^{-d}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2\alpha_1^2}\right)^{\nu+\frac{d}{2}}(1-\beta_1)(1+\beta_2) \\
+ \;& (1-\theta)\alpha_2^{-d}\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2+\alpha_2^2}\right)^{\nu+\frac{d}{2}}(1-\beta_2)(1+\beta_1) \\
\geq \;& \{\theta\alpha_1^{-d}(1-\beta_1)(1+\beta_2)+(1-\theta)\alpha_2^{-d}(1-\beta_2)(1+\beta_1)\}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2+\alpha_1^2}\right)^{\nu+\frac{d}{2}} \\
\geq \;& 0.
\end{aligned}$$

Since $(1+\beta_1)(1-\beta_2) < (1-\beta_1)(1+\beta_2)$, inequality (1.43) follows from

$$\begin{aligned}
& \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1+\beta_1)(1-\beta_2) \\
+ \;& (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1+\beta_2)(1-\beta_1) \\
\geq \;& \{\theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}+(1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}\}(1+\beta_1)(1-\beta_2) \\
= \;& \left\{\theta\alpha_1^{-d}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2+\alpha_1^2}\right)^{\nu+\frac{d}{2}}+(1-\theta)\alpha_2^{-d}\left(\frac{\alpha_2^2}{\|\boldsymbol{\omega}\|^2+\alpha_2^2}\right)^{\nu+\frac{d}{2}}\right\}(1-\beta_2)(1+\beta_1) \\
\geq \;& \{\theta\alpha_1^{-d}+(1-\theta)\alpha_2^{-d}\}\left(\frac{\alpha_1^2}{\|\boldsymbol{\omega}\|^2+\alpha_1^2}\right)^{\nu+\frac{d}{2}}(1-\beta_2)(1+\beta_1) \\
\geq \;& 0,
\end{aligned}$$

where the last inequality is obtained from

$$\theta \geq \left\{1-\frac{\alpha_2^d}{\alpha_1^d}\frac{(1-\beta_1)(1+\beta_2)}{(1+\beta_1)(1-\beta_2)}\right\}^{-1} \geq \left\{1-\frac{\alpha_2^d}{\alpha_1^d}\right\}^{-1}.$$

Case $1 \leq \theta \leq \left(1-\frac{\alpha_1}{\alpha_2}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right)^{-1}$: In this case, θ is positive and (1.45) is valid. We deduce inequality (1.43) from

$$\begin{aligned}
& \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1+\beta_1)(1-\beta_2) \\
+ \;& (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1+\beta_2)(1-\beta_1) \\
\geq \;& \{\theta\alpha_1^{2\nu}(1+\beta_1)(1-\beta_2)+(1-\theta)\alpha_2^{2\nu}(1+\beta_2)(1-\beta_1)\}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}} \\
\geq \;& 0,
\end{aligned}$$

and inequality (1.44) from

$$\begin{aligned} & \theta\alpha_1^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_1^2)^{-\nu-\frac{d}{2}}(1-\beta_1)(1+\beta_2) \\ + \quad & (1-\theta)\alpha_2^{2\nu}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2)(1+\beta_1) \\ \geq \quad & \{\theta\alpha_1^{2\nu}+(1-\theta)\alpha_2^{2\nu}\}(\|\boldsymbol{\omega}\|^2+\alpha_2^2)^{-\nu-\frac{d}{2}}(1-\beta_2)(1+\beta_1)\geq 0, \end{aligned}$$

where the last inequality is due to

$$\theta \leq \left(1-\frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\frac{(1-\beta_2)(1+\beta_1)}{(1+\beta_2)(1-\beta_1)}\right)^{-1} \leq \left(1-\frac{\alpha_1^{2\nu}}{\alpha_2^{2\nu}}\right)^{-1}.$$

This completes the proof of the theorem ■

Chapter 2

Multivariate Tapering

2.1 Introduction

In this world with ever-growing data availability, the size of datasets is always increasing as technology advances and many of these datasets require spatial or spatio-temporal analysis in environmental monitoring, climatology, hydrology, and many other fields. One of the challenges in dealing with those correlated large datasets is to calculate the inverse of the covariance matrices, which is the keystone for finding the best linear unbiased prediction or kriging/co-kriging, and is vital to maximum likelihood estimation or Bayesian inference. In spatial and spatio-temporal statistics this becomes a common obstacle in computation since, as the number of locations increases on a spatial domain, the covariance matrix increases. For example, if a space has n locations the covariance matrix would be $n \times n$, where n could be in the hundreds if not thousands or even larger. It is well documented that calculating the predictions as well as likelihoods can be computationally infeasible for large datasets, requiring $O(n^3)$ operations. Covariance tapering is one of approaches that can be used to mitigate this challenge.

To start let us identify where covariance tapering will aid in calculations. Let the second order stationary Gaussian process $Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$ have a zero mean and an isotropic covariance function $K(h; \theta, \sigma^2)$, where σ^2 is the variance of the process and θ is the parameter that controls how fast the covariance function decays over the lag distance $h = \mathbf{s}_i - \mathbf{s}_j$. Given n

observations, $\mathbf{Z}_n = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$, the log-likelihood is

$$l_n = -\frac{n}{2} log(2\pi) - \frac{1}{2} log[\det[\mathbf{V}_n(\theta, \sigma^2)]] - \frac{1}{2} \mathbf{Z}_n^\top [\mathbf{V}_n(\theta, \sigma^2)]^{-1} \mathbf{Z}_n, \tag{2.1}$$

where $\mathbf{V}_n(\theta, \sigma^2)$ denotes the covariance matrix of $\mathbf{Z}_n$, whose $(i, j)th$ entry is $K(\mathbf{s}_i - \mathbf{s}_j; \theta, \sigma^2)$. Here we would need the inverse to find maximum likelihood estimators. On the other hand, the best linear unbiased prediction, i.e. interpolation, at an unobserved location $\mathbf{s}^*$ is given by

$$\hat{Z}(\mathbf{s}^*) = \mathbf{c}^{*\top} \mathbf{V}_n(\theta, \sigma^2)^{-1} \mathbf{Z}_n, \tag{2.2}$$

where $\mathbf{c}^* = (K(\mathbf{s}^* - \mathbf{s}_i))_{n \times 1}$. Apparently, the inverse of the covariance matrix is needed for both parameter estimation and for making predictions, as well as mean square prediction errors (MSPE).

The principle of tapering is to keep the covariance approximately unchanged at small distance lags and reduce the covariance to zero at large distance lags. To implement the idea, let K_{tap} be an isotropic correlation function of compact support, i.e., $K_{tap}(h) = 0$ if $h > \gamma$ for some $\gamma > 0$. Then the tapered covariance function $\widetilde{K}$ is the product of K and K_{tap}

$$\widetilde{K}(h; \theta, \sigma^2) = K(h; \theta, \sigma^2) \times K_{tap}(h), \tag{2.3}$$

and the tapered covariance matrix is a Hadamard product $\widetilde{\mathbf{V}}_n = \mathbf{V}_n(\theta, \sigma^2) \circ \mathbf{T}_n$, where $\mathbf{T}_n$ has the (i, j)th element as $K_{tap}(\|\mathbf{s}_i - \mathbf{s}_j\|)$. This causes the tapered covariance matrix to have a high proportion of zeros resulting in a sparse matrix. Inverting a sparse matrix is much more efficient computationally than trying to invert the original covariance matrix of the same dimension (see, e.g., Pissanetzky (1984), Gilbert et al. (1992), and Davis (2006)). Once the tapered covariance function $\widetilde{K}$ is calculated it would be used for spatial interpolation and estimation as if it were the correct covariance function. In the case of the finding MLEs, one would maximize the corresponding log-likelihood

$$\tilde{l}_n = -\frac{n}{2} log(2\pi) - \frac{1}{2} log[\det[\widetilde{\mathbf{V}}_n]] - \frac{1}{2} \mathbf{Z}_n^\top [\widetilde{\mathbf{V}}_n]^{-1} \mathbf{Z}_n, \tag{2.4}$$

and for interpolation,

$$\hat{Z}(\mathbf{s}^*) = \widetilde{\mathbf{c}}^{*\top}\widetilde{\mathbf{V}}_n(\theta, \sigma^2)^{-1}\mathbf{Z}_n. \tag{2.5}$$

To measure the accuracy of the BLUP we use the MSPE. The MSPE$(\mathbf{s}^*, K)$ has the form,

$$MSPE(\mathbf{s}^*, K) = K(\mathbf{s}^*, \mathbf{s}^*) - \mathbf{c}^{*\top}\mathbf{V}_n(\theta, \sigma^2)^{-1}\mathbf{c}^*, \tag{2.6}$$

where $K(\mathbf{s}^*, \mathbf{s}^*)$ is the variance of the random variable in question at the new location $\mathbf{s}^* \in \mathcal{S}$. If the BLUP is calculated under the tapered covariance function $\widetilde{K}(h; \theta, \sigma^2)$, the actual MSPE has the form,

$$MSPE(\mathbf{s}^*, \widetilde{K}) = K(\mathbf{s}^*, \mathbf{s}^*) - 2\widetilde{\mathbf{c}}^{*\top}\widetilde{\mathbf{V}}_n^{-1}\mathbf{c}^* + \widetilde{\mathbf{c}}^{*\top}\widetilde{\mathbf{V}}_n^{-1}\mathbf{V}_n\widetilde{\mathbf{V}}_n^{-1}\widetilde{\mathbf{c}}^*. \tag{2.7}$$

Intuitively, being that the taper is a correlation function, when the taper is sufficiently close to 1 for spatial locals that are within a small distance from each other, causes the behavior of the original covariance structure to remain almost unchanged. Specifically, we need to study how to choose the taper without sacrificing the richness of the modeling of the underlying original covariance structure.

For spatial interpolation the key component is the behavior the covariance function has at the origin, i.e. a small neighborhood around the location in question. Furrer et al. (2006) show that if the appropriate taper is chosen, the fixed-domain asymptotic optimality of prediction for the Matérn model is not affected, in the sense that the ratio of (2.6) and (2.7) converges to 1 under fixed-domain asymptotics. Kaufman et al. (2008) provided results that show the tapered MLE is consistent for micro-ergodic parameters in the Matérn covariance under the fixed-domain asymptotic framework with θ fixed. Du et al. (2009) gave general conditions that ensure tapering does not affect the efficiency of the MLEs. Note here that the results stated are under the univariate case where only one attribute is of interest.

The computational burden is even heavier if multiple attributes are concerned other than large number of locations are observed. The important technique to elevate this computational intensity is through likelihood approximation. Stein et al. (2004) developed

techniques for likelihood approximation over a regular complete lattice and in doing this the number of calculations are drastically reduced. Also Fuentes (2007) presented a version of Whittle's approximation for irregularly spaced data and came to the same reduction in calculations. Both cases transform the original process into the spectral domain and as a result the calculations under the approximation require $O(mlog_2m + n)$ rather than n^3 for the exact likelihood. While we are engaged to approximate likelihood in the spatial domain directly by extending tapering technique.

Unlike the univariate case, the multivariate case is a new research effort across the spatial and spatio-temporal statistics field. Many spatial datasets have more than just one variable of interest and often times there is a particular variable of interest and the remaining variables can further help in the prediction of the primary variable. Just like the univariate case where a taper itself must be correlation function, here not only is there the covariance function for each variable that needs to be tapered, there is an addition of the cross-covariance function between each of the variables to be concerned. For example, bivariate process has two covariance functions and one cross-covariance function. For large amounts of data cokriging requires the solution of a large linear system based on the covariance and cross-covariance matrices, and "it is impossible to solve the linear system with direct methods", Sun et al. (2012). Recently Furrer and Genton (2011) proposed aggregation-cokriging for highly-multivariate spatial datasets to reduce computation. In which, the secondary variables are weighted based on the correlation with the primary variable to create a linear combination to reduce the linear system size. However, the challenge is how to efficiently find the weights that are appropriate for the secondary variables.

The goal of this chapter is to establish multivariate tapering functions with different degrees of smoothness and show how the behavior of the original covariance function is affected. The continuation of this discussion will look at multivariate tapering in Section 2.2, a simulation study for an Askey-type multivariate taper on a bivariate exponential covariance structure in Section 2.3 will demonstrate the effectiveness of tapering in the

multivariate case. Section 2.4 examines a simulation study when the Wenland-type taper is applied to a multivariate Matérn covarinance structure. An application to USA climate data compares the behaviors of both Askey and Wendland type tapers.

2.2 Multivariate tapering

The multivariate spatial data can be viewed as realization of multivariate random fields. An m-variate stochastic process or random field $\{\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_m(\mathbf{x}))', \mathbf{x} \in \mathbb{D}\}$ is a family of real random vectors on the same probability space, where the index set $\mathbb{D}$ could be a temporal domain like $\mathbb{R}$, or a spatial domain like $\mathbb{R}^d$, with d as a natural number. When all of its components have second-order moments, $\{\mathbf{Z}(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$ is called a second-order vector (or multivariate) random field, and its covariance matrix (function) is defined by

$$\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{E}\{(\mathbf{Z}(\mathbf{x}_1) - \mathrm{E}\mathbf{Z}(\mathbf{x}_1))(\mathbf{Z}(\mathbf{x}_2) - \mathrm{E}\mathbf{Z}(\mathbf{x}_2))'\}, \ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}.$$

Its diagonal entry $C_{ii}(\mathbf{x}_1, \mathbf{x}_2)$, the covariance function of the ith component random field $\{Z_i(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$, is called a direct covariance (function), and its off-diagonal entry $C_{ij}(\mathbf{x}_1, \mathbf{x}_2)$ $(i \neq j)$, the covariance between the ith component random field $\{Z_i(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$ and the jth component random field $\{Z_j(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$, is called a cross covariance (function), $i, j = 1, 2, \ldots, m$. Moreover, $\{\mathbf{Z}(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$ is said to be a (weak, second-order) stationary or homogeneous random field, if its mean function $\mathrm{E}\mathbf{Z}(\mathbf{x})$, $\mathbf{x} \in \mathbb{D}$, is a constant vector, and its covariance matrix function $\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2)$ depends only on the lag $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}$. In such a case, we simply write $\mathbf{C}(\mathbf{x}_1 - \mathbf{x}_2)$ for $\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2)$ for simplicity of notations. The validity of the multivariate covariance and cross-covariance is warranted by the following property.

Given a real $m \times m$ matrix function $\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2), \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}$, it is shown in Ma (2011) that there is an m-variate second-order random field $\{\mathbf{Z}(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$ with mean zero and with $\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2)$ as its covariance matrix if and only if $\{\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2)\}' = \mathbf{C}(\mathbf{x}_2, \mathbf{x}_1)$ and the inequality

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i' \mathbf{C}(\mathbf{x}_i, \mathbf{x}_j) \mathbf{a}_j \geq 0 \tag{2.8}$$

holds for every positive integer n, any $\mathbf{x}_k \in \mathbb{D}$, and any $\mathbf{a}_k \in \mathbb{R}^m$, $k = 1, \ldots, n$.

In order for the tapered covariance function to maintain the richness of the original processes, the multivariate tapering matrix function $\mathbb{C}_{tap}$ must be a compactly supported correlation matrix function, i.e. a correlation matrix functions whose entries are compactly supported. Like the univariate case the technique of tapering is to use the Hadamard product, $\widetilde{\mathbf{C}}(\mathbf{s}) = \mathbf{C}(\mathbf{s}) \circ \mathbf{C}_{tap}(\mathbf{s})$. Again, instead of using the original covariance matrix function to find MLEs and do cokriging we use $\widetilde{C}$. However the challenge is finding and forming compactly supported covariance function, in fact, "it seems that there exist few results on multivariate compactly supported covariance functions," Zhang and Du (2008). This motivates us to create flexible compactly supported multivariate covariance functions that can be used as tapers that would allow for both computation efficiency and prediction efficiency.

2.2.1 Preliminary results of covariance matrix functions

Multivariate or vector random field modeling is an on-going contemporary research topic in its own right in the spatial statistics. With multiple variables of interest vector random field requires both direct covariance for each variable and cross-covariances between pairs of variables. The following two lemmas provide some fundamental tools to construct a valid covariance matrix function (Du and Ma (2011)). Recall that $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard or Schur product of two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size, which is the entrywise product of $\mathbf{A}$ and $\mathbf{B}$.

Lemma 2.2.1. *If* $\mathbf{C}_1(\mathbf{x}_1, \mathbf{x}_2)$ *and* $\mathbf{C}_2(\mathbf{x}_1, \mathbf{x}_2)$ *are* $m \times m$ *covariance matrix functions, and* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *are* $m \times m$ *positive definite matrices, then there is an* m*-variate spherically invariant random field with covariance matrix* $\mathbf{A}_1 \circ \mathbf{C}_1(\mathbf{x}_1, \mathbf{x}_2) + \mathbf{A_2} \circ \mathbf{C}_2(\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}$.

Note that the particular case where $\mathbf{A}_k = \alpha_k \mathbf{1}$ and α_k $(k = 1, 2)$ are nonnegative constants, we obtain that $\alpha_1 \mathbf{C}_1(\mathbf{x}_1, \mathbf{x}_2) + \alpha_2 \mathbf{C}_2(\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{D}$, is also a covariance matrix, where all entries of the matrix $\mathbf{1}$ equal 1. In other words, the set of covariance matrices is

a convex set.

Lemma 2.2.2. *If $g(\boldsymbol{\omega})$ is a nonnegative function in $\mathbb{R}^d$, $\mathbf{C}(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\omega})$ is an $m \times m$ covariance matrix for every fixed $\boldsymbol{\omega} \in \mathbb{R}^d$, then there is an m-variate spherically invariant random field with direct and cross covariances*

$$\int_{\mathbb{R}^d} \cos(\boldsymbol{\omega}'(\mathbf{x}_1 - \mathbf{x}_2)) C_{ij}(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\omega}) g(\boldsymbol{\omega}) d\boldsymbol{\omega}, \ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d, \ i, j = 1, \ldots, m,$$

assuming that the above integrals exist.

One class of vector random fields can be constructed such that both the direct and cross-covariance functions have the form of an exponential variogram. Here with the above lemmas we have the following theorem.

Theorem 2.2.1. *If $\Theta = (\theta_{ij})$ is an $m \times m$ conditionally negative definite matrix with positive entries, then the $m \times m$ matrix function with entries*

$$C_{ij}(\mathbf{s}) = \theta_{ij}^{-\frac{1}{2}} \exp(-\theta_{ij}^{\frac{1}{2}} \|\mathbf{x}\|), \tag{2.9}$$

where $\|\mathbf{x}\|$ is the Euclidean norm of $\mathbf{x} \in \mathbb{R}^d$, and the matrix function has the form

$$\begin{pmatrix} \theta_{11}^{-\frac{1}{2}} \exp(-\theta_{11}^{\frac{1}{2}} \|\mathbf{x}\|) & \theta_{12}^{-\frac{1}{2}} \exp(-\theta_{12}^{\frac{1}{2}} \|\mathbf{x}\|) & \ldots & \theta_{1m}^{-\frac{1}{2}} \exp(-\theta_{1m}^{\frac{1}{2}} \|\mathbf{x}\|) \\ \theta_{21}^{-\frac{1}{2}} \exp(-\theta_{21}^{\frac{1}{2}} \|\mathbf{x}\|) & \theta_{22}^{-\frac{1}{2}} \exp(-\theta_{22}^{\frac{1}{2}} \|\mathbf{x}\|) & \ldots & \theta_{2m}^{-\frac{1}{2}} \exp(-\theta_{2m}^{\frac{1}{2}} \|\mathbf{x}\|) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{m1}^{-\frac{1}{2}} \exp(-\theta_{m1}^{\frac{1}{2}} \|\mathbf{x}\|) & \theta_{m2}^{-\frac{1}{2}} \exp(-\theta_{m2}^{\frac{1}{2}} \|\mathbf{x}\|) & \ldots & \theta_{mm}^{-\frac{1}{2}} \exp(-\theta_{mm}^{\frac{1}{2}} \|\mathbf{x}\|) \end{pmatrix}, \tag{2.10}$$

is a stationary covariance matrix function.

By changing the Euclidean norm to the ℓ_1-norm we obtain another valid covariance matrix function given in Theorem 2.2.2.

Theorem 2.2.2. *If $\Theta = (\theta_{ij})$ is an $m \times m$ conditionally negative definite matrix with positive entries, then the $m \times m$ matrix function with entries*

$$C_{ij}(\mathbf{s}) = \theta_{ij}^{-\frac{d}{2}} \exp(-\theta_{ij}^{\frac{d}{2}} |\mathbf{x}|), \tag{2.11}$$

where $|\mathbf{x}|$ *is the* ℓ_1*-norm of* $\mathbf{x} \in \mathbb{R}^d$*, and the matrix function has the form*

$$\begin{pmatrix} \theta_{11}^{-\frac{d}{2}} \exp(-\theta_{11}^{\frac{d}{2}}|\mathbf{x}|) & \theta_{12}^{-\frac{d}{2}} \exp(-\theta_{12}^{\frac{d}{2}}|\mathbf{x}|) & \dots & \theta_{1m}^{-\frac{d}{2}} \exp(-\theta_{1m}^{\frac{d}{2}}|\mathbf{x}|) \\ \theta_{21}^{-\frac{d}{2}} \exp(-\theta_{21}^{\frac{d}{2}}|\mathbf{x}|) & \theta_{22}^{-\frac{d}{2}} \exp(-\theta_{22}^{\frac{d}{2}}|\mathbf{x}|) & \dots & \theta_{2m}^{-\frac{d}{2}} \exp(-\theta_{2m}^{\frac{d}{2}}|\mathbf{x}|) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{m1}^{-\frac{d}{2}} \exp(-\theta_{m1}^{\frac{d}{2}}|\mathbf{x}|) & \theta_{m2}^{-\frac{d}{2}} \exp(-\theta_{m2}^{\frac{d}{2}}|\mathbf{x}|) & \dots & \theta_{mm}^{-\frac{d}{2}} \exp(-\theta_{mm}^{\frac{d}{2}}|\mathbf{x}|) \end{pmatrix}, \tag{2.12}$$

which is a stationary covariance matrix function.

Proofs of Theorems 2.2.1 and 2.2.2 are given in Appendix 2.6.

Du et al. (2012) also introduced the hyperbolic vector random fields, which provides direct and cross-covariances matrix functions when the random variables of interest have a generalized hyperbolic distribution. This structure provides a basis for modeling more complex processes. Just as Examples 2-5 in Du and Ma (2011) illustrate, the next theorem demonstrates how naturally a conditionally negative definite matrix gets involved in our covariance matrix construction. The Matérn model is a very common model to use in the univariate case, Du et al. (2012) construct a valid Matérn multivariate covariance matrix function as given in the following theorem.

Theorem 2.2.3. *Let* α *and* ν_{ij} $(i, j = 1, \dots, m)$ *be positive constants. A necessary and sufficient condition for the existence of an* m*-variate stationary hyperbolic random field on* $\mathbb{D}$ *with direct and cross covariances*

$$C_{ij}(\mathbf{x}) = \frac{1}{2^{\nu_{ij}}\Gamma\left(\nu_{ij} + \frac{d}{2}\right)} (\alpha\|\mathbf{x}\|)^{\nu_{ij}} K_{\nu_{ij}}(\alpha\|\mathbf{x}\|), \ \mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \dots, m, \tag{2.13}$$

is that the $m \times m$ *matrix with entries* ν_{ij} *is conditionally negative definite.*

Each direct or cross covariance in (2.13) is a univariate Matérn . Particularly, in (2.13) taking $\nu_{ij} = \frac{1}{2}(\nu_i + \nu_j)$ yields a covariance matrix function with direct and cross covariances

$$\frac{1}{2^{\frac{1}{2}(\nu_i+\nu_j)}\Gamma\left(\frac{1}{2}(\nu_i + \nu_j) + \frac{d}{2}\right)} (\alpha\|\mathbf{x}\|)^{\frac{1}{2}(\nu_i+\nu_j)} K_{\frac{1}{2}(\nu_i+\nu_j)}(\alpha\|\mathbf{x}\|), \ \mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \dots, m.$$

Multiplying this matrix function with two positive definite matrices, one with entries $2^{\frac{1}{2}(\nu_i+\nu_j)}$ and the other with entries $\Gamma\left(\frac{1}{2}(\nu_i+\nu_j)+\frac{d}{2}\right)$, it results in a covariance matrix function in the next corollary,

Corollary 2.2.1. *If $\nu_1,\ldots,\nu_m$ are positive constants, then there exists an m-variate stationary hyperbolic random field on $\mathbb{D}$ with direct and cross covariances*

$$C_{ij}(\mathbf{x}) = (\alpha\|\mathbf{x}\|)^{\frac{1}{2}(\nu_i+\nu_j)} K_{\frac{1}{2}(\nu_i+\nu_j)}\left(\alpha\|\mathbf{x}\|\right), \ \mathbf{x}\in\mathbb{R}^d, \ i,j=1,\ldots,m. \tag{2.14}$$

The following corollary is obtained from Theorem 2.2.3 by letting $\nu_{ij} = \max(\nu_i,\nu_j)$, $i,j=1,\ldots,m$.

Corollary 2.2.2. *If $\nu_1,\ldots,\nu_m$ are positive constants, then there exists an m-variate stationary hyperbolic random field on $\mathbb{D}$ with direct and cross covariances*

$$C_{ij}(\mathbf{x}) = \frac{1}{2^{\max(\nu_i,\nu_j)}\Gamma\left(\max(\nu_i,\nu_j)+\frac{d}{2}\right)}(\alpha\|\mathbf{x}\|)^{\max(\nu_i,\nu_j)} K_{\max(\nu_i,\nu_j)}\left(\alpha\|\mathbf{x}\|\right), \\ \mathbf{x}\in\mathbb{R}^d, \ i,j=1,\ldots,m. \tag{2.15}$$

Gneiting et al. (2010) also presents a multivariate Matérn covariance matrix function. They discuss the necessary and sufficient conditions for a parsimonious and a full bivariate Matérn model. The parsimonious multivariate Matérn model gives each marginal covariance function,

$$C_{ii}(h) = \sigma_i^2 M(h|\nu_i,\alpha_i), \ i=1,\ldots,p, \tag{2.16}$$

and cross covariance function given by,

$$C_{ij}(h) = C_{ji}(h) = \rho_{ij}\sigma_i\sigma_j M(h|\nu_{ij},\alpha_{ij}), \ 1\le i\ne j\le p. \tag{2.17}$$

In (2.16) $\sigma_i>0$, $\nu_i>0$ is a smoothness parameter, and $\alpha_i>0$ is a scaling parameter. Note that in (2.16) and (2.17),

$$M(h|\nu,\alpha) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\alpha\|h\|)^{\nu}K_{\nu}(\alpha\|h\|);$$

where K_ν is a modified Bessel function of the second kind. In (2.17) $\alpha_{ij} > 0$ is a common scaling parameter of the cross covariance, $\nu_{ij} = \frac{1}{2}(\nu_i + \nu_j),\ 1 \leq i \neq j \leq p$, and

$$\rho_{ij} = \beta ij \frac{\Gamma(\nu_i + (d/2))^{1/2}}{\Gamma(\nu_i)^{1/2}} \times \frac{\Gamma(\nu_j + (d/2))^{1/2}}{\Gamma(\nu_j)^{1/2}} \times \frac{\Gamma(\frac{1}{2}(\nu_i + \nu_j))}{\Gamma(\frac{1}{2}(\nu_i + \nu_j) + d/2)},\ 1 \leq i \neq j \leq p;$$

where the matrix $(\beta_{ij})_{i,j=1}^p$ with diagonal elements $\beta_{ii} = 1$ for $i = 1, \ldots, p$ and off-diagonal elements β_{ij} for $1 \leq i \neq j \leq p$ is symmetric and nonnegative definite. Notice that the multivariate Matérn covariance model of Gneiting et al. (2010) is simply the Hardmard product of that in Corollary 2.2.1 and a positive definite matrix given by ρ_{ij}. In contrast to those in Gneiting et al. (2010), the covariance matrix structures here are displayed as neat as possible.

2.2.2 Wendland type of compactly supported covariance matrix functions

In this section we develop multivariate tapering matrix functions by extending Wendland type of correlation function to multivariate case, where each entry is compactly supported. Here a compactly supported function in $\mathbb{R}^d$ is a function whose values are zero outside a compact set in $\mathbb{R}^d$. The celebrated univariate Wendland type of tapering function is constructed by using the fractional descente, i.e. $\phi_{\nu,\kappa} = I^k \phi_{\nu,0}$, where $I\phi(t) = \int_t^\infty u\phi(u)du / \int_0^\infty u\phi(u)du$ and $\phi_{\nu,0}$ is given by the following Askey function with $\alpha = 1$.

The Askey function is given by,

$$C(\mathbf{x}) = \left(1 - \frac{\|\mathbf{x}\|}{\alpha}\right)_+^\nu, \tag{2.18}$$

where α is a positive constant, $\nu \geq \left[\frac{d}{2}\right] + 1$, $[x]$ denotes the largest integer that is not greater than x, and $x_+ = \max(x, 0)$, $x \in \mathbb{R}$; see Askey (1973) and Letac and Rahman (1986). Wendland 1 function is given by,

$$K_{wend1}(\|\mathbf{x}\|; \gamma) = (1 - \frac{\|\mathbf{x}\|}{\gamma})_+^4 (1 + 4\frac{\|\mathbf{x}\|}{\gamma}),\ \gamma > 0, \tag{2.19}$$

which second differentiable at zero, and Wendland 2 function is given by

$$K_{wend2}(\|\mathbf{x}\|; \gamma) = (1 - \frac{\|\mathbf{x}\|}{\gamma})_+^6 (1 + 6\frac{\|\mathbf{x}\|}{\gamma} + \frac{35\|\mathbf{x}\|^2}{3\gamma^2}),\ \gamma > 0, \tag{2.20}$$

which is fourth differentiable at zero, see Wendland (1995).

These functions become the building blocks to create multivariate tapering functions. The frist multivariate taper discussed is the Askey-type which is derived by the mixture of (2.18).

Theorem 2.2.4. *Let $d \geq 2$, $\nu \geq \left[\frac{d}{2}\right] - 1$, and $\nu_k \geq 0$ $(k = 1, \ldots, m)$. If an $m \times m$ matrix with entries $g_{ij}(u)$ is positive definite for every fixed $u \in [0, 1]$ and the function $g_{ij}(x)$ is continuous on $[0, 1]$, then there is an m-variate Gaussian or elliptically contoured random field with direct and cross covariances*

$$C_{ij}(\mathbf{x}) = \begin{cases} \int_0^1 (u - \|\mathbf{x}\|)_+^{\nu} \, g_{ij}(u) du, & \|\mathbf{x}\| \leq 1, \\ 0, & \|\mathbf{x}\| > 1, \end{cases} \tag{2.21}$$

$$\mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \ldots, m.$$

The proof of Theorem 2.2.4 can be found in Du and Ma (2012).

Example 2.2.1. Let α be a constant between 0 and 1. In (2.21) taking $m = 2$ and $g_{11}(u) = u, g_{12}(u) = g_{21}(u) = \min(u, \alpha)$, and $g_{22}(u) \equiv \alpha, \alpha \in [0, 1]$, yields

$$\begin{aligned} C_{11}(\mathbf{x}) &= \frac{1}{\nu + 1}(1 - \|\mathbf{x}\|)_+^{\nu+1} - \frac{1}{(\nu + 1)(\nu + 2)}(1 - \|\mathbf{x}\|)_+^{\nu+2}, \\ C_{12}(\mathbf{x}) &= C_{21}(\mathbf{x}) \\ &= \frac{\alpha}{\nu + 1}(1 - \|\mathbf{x}\|)_+^{\nu+1} - \frac{1}{(\nu + 1)(\nu + 2)}(\alpha - \|\mathbf{x}\|)_+^{\nu+2}, \\ C_{22}(\mathbf{x}) &= \frac{\alpha}{\nu + 1}(1 - \|\mathbf{x}\|)_+^{\nu+1}, \ \mathbf{x} \in \mathbb{R}^d. \end{aligned}$$

Example 2.2.1 will be used in a simulation study later to show the performance of the Askey multivariate taper when used in conjuncture with an exponential and Matérn bivariate covariance structures. Later in the simulation study we find that the Askey taper does not perform well when the smoothness of the original covariance increases at origin. One way this happens is when the covariance has a Matérn structure with smoothness parameter $\nu > 0.5$. Notice that the Askey tapers are not differentiable at origin, the need

for smoother multivariate tapers arises. The following theorem gives a general format to construct nonstandardized tapering matrix function, whose smoothness at origin can be controlled by the input univariate compactly supported covariance function, the proof is given in the Appendix.

Theorem 2.2.5. *Let $K(\|\mathbf{x}\|)$ be a univariate compacted supported and isotropic covariance function, with the supporting range 1. If an $m \times m$ matrix with entries $g_{ij}(u)$ is positive definite for every fixed $u \in [0,1]$ and the function $g_{ij}(x)$ is continuous on $[0,1]$, then there is an m-variate Gaussian or elliptically contoured random field with direct and cross covariances*

$$C_{ij}(\mathbf{x}) = \begin{cases} \int_0^1 K(\frac{\|\mathbf{x}\|}{u}) g_{ij}(u) du, \ \ \|\mathbf{x}\| \leq 1, \\ \\ 0, \ \ \|\mathbf{x}\| > 1, \end{cases} \tag{2.22}$$

$$\mathbf{x} \in \mathbb{R}^d, \ \ i, j = 1, \ldots, m.$$

Wendland-type compact supported correlation functions are usually serve as tapering functions, which possesses a desirable smoothness parameter making them have higher order differentiability at zero, and have flexibility to monitor the degree to which they alter the original covariance matrix functions at zero. This makes the Wendland-type taper an preferable candidate to create a multivariate taper for smoother processes. To generate the well- known Wendland tapering functions (Wendland (1995)) which have been adopted in literature (e.g. Furrer et al. (2006), Kaufman et al. (2008), and Du et al. (2009)) to multivariate, we use univariate Wendland tapering function in the equation (2.22). Note that the Wendland functions are constructed as a series of compacted supported functions with different differentiability at zero by using the fractional descente. With this substitution we obtain the following multivariate extension of Wendland compactly supported covariance function, the standardized version will be called the Wendland tapering matrix function.

Theorem 2.2.6. *Let $\nu \geq \frac{d+1}{2}, k = 0, 1, 2, \ldots$. If an $m \times m$ matrix with entries $g_{ij}(u)$ is positive definite for every fixed $u \in [0,1]$ and the function $g_{ij}(x)$ is continuous on $[0,1]$, then there is an m-variate Gaussian or elliptically contoured random field with direct and cross*

covariances

$$C_{ij}(\mathbf{x}) = \begin{cases} \int_0^1 u^k I^k \phi_{\nu,0}(\frac{\|\mathbf{x}\|}{u}) g_{ij}(u) du, & \|\mathbf{x}\| \leq 1, \\ 0, & \|\mathbf{x}\| > 1, \end{cases} \tag{2.23}$$

$$\mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \ldots, m.$$

Using Theorem 2.2.6 and the Wendland 1 equation (2.19) we have Corollary 2.2.3, which will be used in our simulation study of the bivariate Matérn covariance.

Corollary 2.2.3. Let α be a constant between 0 and 1. In (2.23) taking $\nu = 3, k = 1$ and taking $m = 2$ and $g_{11}(u) = u, g_{12}(u) = g_{21}(u) = \min(u, \alpha)$, and $g_{22}(u) \equiv \alpha, \alpha \in [0, 1]$, yields

$$\begin{aligned} C_{11}(\mathbf{x}) &= (1 - \|\mathbf{x}\|)_+^5 \left(\frac{1}{7}(1 - \|\mathbf{x}\|)^2 + (1 - \|\mathbf{x}\|)\|\mathbf{x}\| + \|\mathbf{x}\|^2 \right), \\ C_{12}(\mathbf{x}) &= C_{21}(\mathbf{x}) \\ &= \alpha(1 - \|\mathbf{x}\|)_+^5 \left(\frac{1}{6}(1 - \|\mathbf{x}\|) + \|\mathbf{x}\| \right) \\ &\quad + (\alpha - \|\mathbf{x}\|)_+^5 \left(\frac{1}{7}(\alpha - \|\mathbf{x}\|)^2 + (\alpha - \|\mathbf{x}\|)(\|\mathbf{x}\| - \frac{1}{6}) + \|\mathbf{x}\|^2 - \|\mathbf{x}\| \right), \\ C_{22}(\mathbf{x}) &= \alpha(1 - \|\mathbf{x}\|)_+^5 \left(\frac{1}{6}(1 - \|\mathbf{x}\|) + \|\mathbf{x}\| \right). \end{aligned}$$

Also using Wendland 2 equation we obtain the multivariate version of Wendland 2 in Corollary 2.2.4.

Corollary 2.2.4. Let α be a constant between 0 and 1. In (2.23) taking $\nu = 3, k = 2$ and taking $m = 2$ and $g_{11}(u) = u, g_{12}(u) = g_{21}(u) = \min(u, \alpha)$, and $g_{22}(u) \equiv \alpha, \alpha \in [0, 1]$, yields

$$\begin{aligned} C_{11}(\mathbf{x}) &= (1 - \|\mathbf{x}\|)_+^7 \left(\frac{1}{10}(1 - \|\mathbf{x}\|)^3 + (1 - \|\mathbf{x}\|)^2\|\mathbf{x}\| + \frac{10}{3}(1 - \|\mathbf{x}\|)\|\mathbf{x}\|^2 + \frac{8}{3}\|\mathbf{x}\|^3 \right), \\ C_{12}(\mathbf{x}) &= C_{21}(\mathbf{x}) \\ &= \alpha(1 - \|\mathbf{x}\|)_+^7 \left(\frac{1}{9}(1 - \|\mathbf{x}\|)^2 + (1 - \|\mathbf{x}\|)\|\mathbf{x}\| + \frac{8}{3}\|\mathbf{x}\|^2 \right) \\ &\quad + (\alpha - \|\mathbf{x}\|)_+^7 \left(\frac{1}{10}(\alpha - \|\mathbf{x}\|)^3 + (\alpha - \|\mathbf{x}\|)^2(\|\mathbf{x}\| - \frac{1}{9}) \right. \\ &\quad \left. + \frac{10}{3}(\alpha - \|\mathbf{x}\|)(\|\mathbf{x}\|^2 - \|\mathbf{x}\|) + \frac{8}{3}(\|\mathbf{x}\|^3 - \|\mathbf{x}\|^2) \right), \\ C_{22}(\mathbf{x}) &= \alpha(1 - \|\mathbf{x}\|)_+^7 \left(\frac{1}{9}(1 - \|\mathbf{x}\|)^2 + (1 - \|\mathbf{x}\|)\|\mathbf{x}\| + \frac{8}{3}\|\mathbf{x}\|^2 \right), \ \mathbf{x} \in \mathbb{R}^d. \end{aligned}$$

When using the above theorems and corollaries there are two ways to arrange the covariance structure, a variable arrangement or a location arrangement. Figure 2.1 and 2.2 give an example of Theorem 2.2.1, the exponential covariance model, with the Askey-type taper of Theorem 2.2.4. These images were created using 100 locations are placed on a line for 0 to 1. From variable orientation plots one can clearly see the two direct covariance matrices and the cross-covariance matrix. The location orientation of Figure 2.1 can be seen as a layering of the three covariance functions, each 4×4 block creating a mini covariance matrix for each location starting for the upper left hand corner. From the variable orientation of Figure 2.2 notice that along the diagonal would represent a distance lag of $h = 0$ where the correlation should be the strongest. As the distance lag increases the correlation decays towards zero. Also after the taper the tapered covariance remains to have the same correlations along the diagonal, but note that the off diagonals are set to zero.

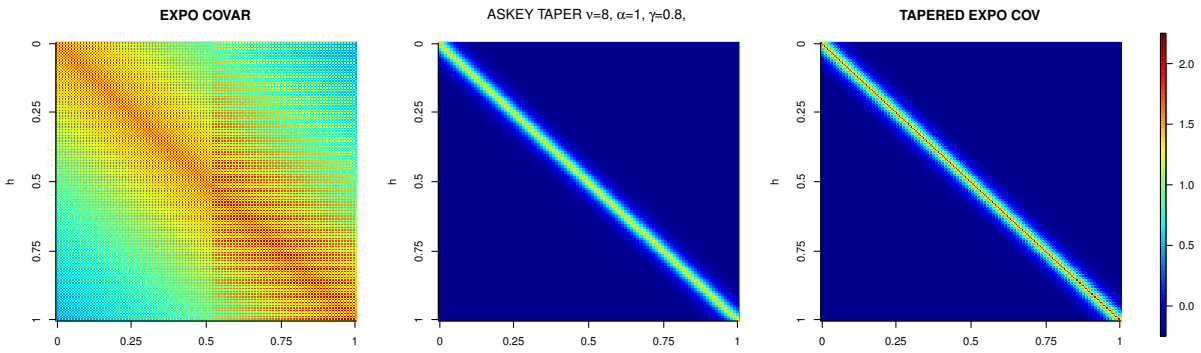

Figure 2.1: Location orientation covariance plot of exponential covariance, Askey taper, and tapered covariance.

Figure 2.3 is an example of the Matérn model from Theorem 2.2.3 with the Wendland 1 taper of Example 2.2.3. Notice that the Wendland 1 taper is much smoother than that of the Askey taper. This smoothness plays a critical part in the simulation studies to come.

In Section 2.3 we simulate the effect of the Askey-type taper on a vector random field with an exponential covariance structure as in Theorem 2.2.1 and in Section 2.4 we investigate the Askey-type, Wendland 1 and Wendland 2 tapers with the Matérn model of Theorem 2.2.3. The goal of the simulations is to study the effect of multivariate tapering has on

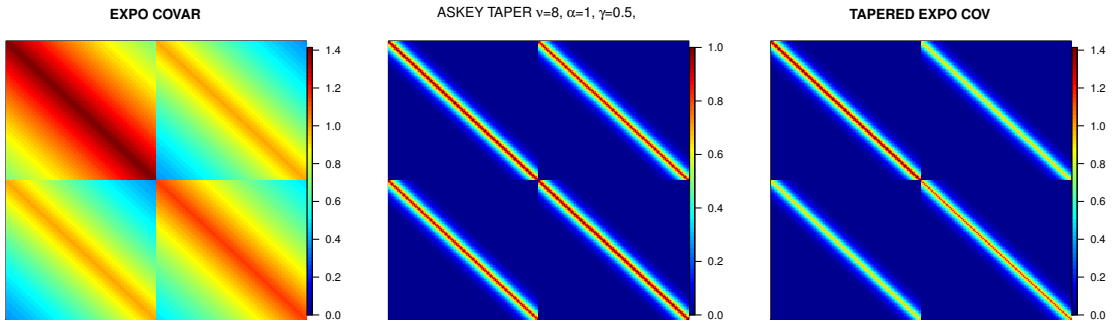

Figure 2.2: Variable orientation covariance plot of exponential covariance, Askey taper, and tapered covariance.

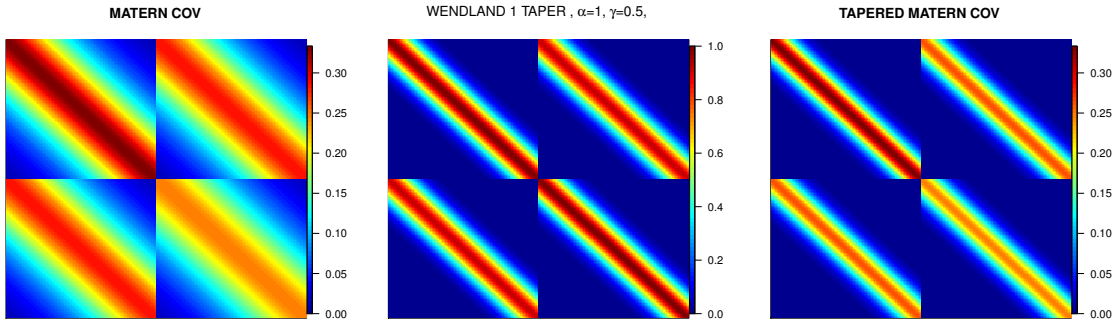

Figure 2.3: Variable orientation covariance plot of Matérn covariance, Wendland 1 taper, and tapered covariance.

cokriging and computation efficiency.

2.3 Bivariate exponential case

In this section we show how Theorem 2.2.4 and 2.2.1 can be used to create a multivariate valid covariance matrix function as well as how tapering can be used to perform efficient computations. We simulate samples from the original multivariate random field with mean zero based on Theorem 2.2.1. The exponential covariance parameters are $\theta_{11} = .5$, $\theta_{22} = 1$, and $\theta_{12} = \theta_{21} = 1$, formulating the Θ matrix. The covariance is constructed over the unit square with n uniformly distributed locations. Here in the simulation n is allowed to increase to show, in the case of a fixed-domain, that the MSPE is hardly affected and the time for computation is drastically improved.

The taper used for the exponential covariance structure comes from standardization of Example 2.2.1, where $\alpha = 1$, $\nu = 7$, and $\gamma \in 0.5, 0.3, 0.1$. For each γ, i.e. the tapering range, we provide the simulation results for MSPE and time improvement, note the smaller the γ is the more tapered the covariance will become. For each simulation we predict 50 points along the vector form $(0, 0)$ to $(1, 1)$ and 50 points along the vector from $(0, 1)$ to $(1, 0)$. The MSPE is calculated for each point then averaged to find an average MSPE. The time taken to calculate cholesky decomposition is the measure we use to compare times. Because of very similar results, the $\gamma = 0.1$ simulation will be looked at in detail and other results will be left to the appendix of this chapter. For the $\gamma = 0.1$ case this amounts to basically reducing the information of the covariance structure 90%. From Figure 2.4 we can see that at 2000 locations the taper starts to take effect by decrease time to calculate the inverse of the covariance matrix as well as produce the same MSPE as the true exponential covariance structure. Table 2.1 shows a more detailed look at the actual measurements taken. Notice as the number of locations increase that the MSPE of the tapered covariance converges to the MSPE of the true. This is analogous to the results of Furrer et al. (2006) in the univariate case. Looking at the times for cholesky decomposition, the true covariance spends a lot of time in calculation where the tapered covariance does not. In the case of 8100 locations, a 16200×16200 covariance matrix is produced because two variables are under consideration. The original covariance takes 61.12 minutes to calculate the cholesky decomposition and the tapered covariance takes 27.46 seconds. The calculations here were done using R (R Core Team (2012)) on the Kansas State University Beocat with a node having 16 GB of RAM and a 4x Quad-Core AMD Opteron 8350 Processor. Results in Appendix 2.6 are very similar to the results shown here, the difference being γ is larger allowing for more information. This allows the taper to be more accurate at a smaller number of locations.

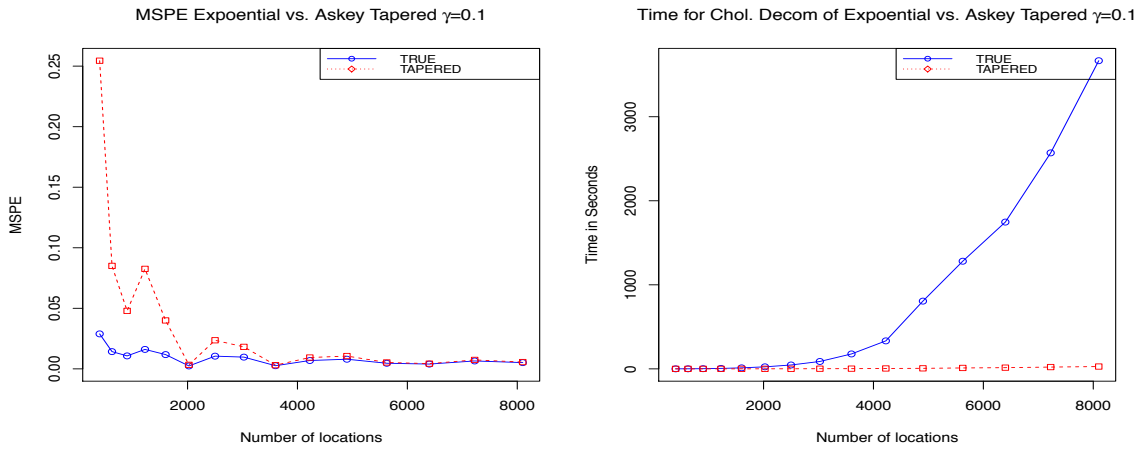

Figure 2.4: Results Graph: Exponential and Askey taper, tapered with $\gamma = 0.1$

Table 2.1: *MSPE and Time for Exponential and Askey with* $\gamma = 0.1$

NUM.LOCS	EXP.MSPE	EXP.TIME	Askey.TIME	Askey.MSPE
400	0.0288870	0.30	0.14	0.2544790
625	0.0142470	0.72	0.15	0.0850270
900	0.0107040	2.14	0.18	0.0478780
1225	0.0160650	5.37	0.18	0.0825650
1600	0.0117850	11.90	0.25	0.0399870
2025	0.0024041	24.07	0.77	0.0033075
2500	0.0104930	46.03	0.89	0.0236090
3025	0.0096929	87.92	1.91	0.0181450
3600	0.0026003	177.54	2.44	0.0030367
4225	0.0068860	332.88	4.60	0.0093571
4900	0.0079591	806.05	6.21	0.0105690
5625	0.0046308	1280.80	10.84	0.0052771
6400	0.0039480	1746.50	14.29	0.0043060
7225	0.0065214	2569.80	21.08	0.0073742
8100	0.0051621	3666.50	27.46	0.0055965

[1] NUM.LOCS is total number of locations
[2] EXP.MSPE is the MSPE using the true covariance
[3] EXP.TIME seconds to calculate Choleski decomp. for the exponential
[4] Askey.TIME seonds to calculate Choleski decomp. for the tapered
[5] Askey.MSPE is the MSPE using the tapered covariance
[6] Refer to Figure 2.4 for graph of results

2.4 Bivariate Matérn case

Here we used the multivariate Matérn spatial covariance model, taken from Du et al. (2012), with two variables of interest. The tapers used are the Askey tapering function of Example 2.2.1 and both Wendland tapering function of Examples 2.2.3 and 2.2.4. For each simulation we take n uniformly distributed locations on a 100 x 100 unit square. Our simulations are concerned with how the taper performs when the smoothness for the multivariate Matérn increases or decreases. The reason for the 100 x 100 unit square is because of the smoothness of the Matérn .

As the smoothness increases, in order to do calculations the field must be larger to avoid singularity in the covariance matrix. For these examples the following parameters where used over all comparisons; the Matérn Covariance parameters are $\alpha_{matern} = 1$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\phi_{i,j} = 0.5$, ν_1 and ν_2 change to modify the smoothness of the two Matérn processes. The change in ν_1 and ν_2 will be noted for each table. Some cases are left to Appendix 2.6 because of similar results. For all tapers the tapering range $\gamma = 10$ resulting in a 90% reduction of information from the original covariance structure, and $\alpha_{taper} = 0.5$. With other simulations we noticed that α does not change the outcome of the MSPE and α is never small tuning parameter depended on each situation. Based on smaller simulations of the Matérn , we used a taper $\nu_{taper} = 8$ for the Askey-type taper to account for the smoothness of the bivariate Matérn processes.

To calculate MSPE, cokriging was done on 500 uniformly distributed points in the 100×100 square. As a result notice that in all cases the MSPE of the Askey-type taper is not performing well because this taper is not smooth enough to capture the behavior of the Matérn at short distances. However, when the number of location goes to 1000 and beyond both Wendland 1 and 2 tapers perform well against the true covariance structure producing fast accurate cokriging. Tables 2.3, 2.5, 2.7, and 2.9, show that as the number of locations increase the MSPE of the Wendland tapered covariances converge to the MSPE of the true covariance. Tables 2.4, 2.6, 2.8, and 2.10, give the results for the time it takes to calculate

the cholesky decomposition for the inversion of the covariance matrix. Again we see that tapering provides faster faster calculations of the inverse. The calculations here were done using R (R Core Team (2012)) on the Kansas State University Beocat with a node having 16 GB of RAM and a 4x Quad-Core AMD Opteron 8350 Processor.

In conclusion we can see that in general as the smoothness of the Matérn increases the smoother the taper has to be. Also notice that tapering range is 10 which is 10%of the domain range. This indicates that by removing 90% of the data from the covariance structure we are still able to make accurate predictions using our tapers instead of the full covariance structure. The results are similar to the Furrer, Genton, Nychka (2006) in that as the smoothness of the Matérn increases we must use a taper with appropriate smoothness. Time savings becomes apparent around 1500 locations and the MSPE for both Wendlands is very close to the TRUE Matérn MSPE at less than 200 locations. Although we used a tapering range of 10, increasing would decrease the MSPE further to the TRUE, but raise time taken to perform calculations. In practice one should make the taper as large as possible, but choose so that time can be saved. Both Wendland 1 and Wendland 2 are very close in all simulations, so based on the complexity of the tapers one might chose the simpler of the two tapers. We formulate a guideline for choosing tapers suggested by the simulation studies given in Table 2.2.

Table 2.2: *General Results of Multivariate Tapers*

Taper	True Covariance Structure	Valid taper for
Askey	Bivariate Exponential	Not valid for Bivariate Matérn
Wendland 1	Bivariate Matérn	$\nu_1, \nu_2 \leq 1.5$
Wendland 2	Bivariate Matérn	$\nu_1, \nu_2 \leq 2.5$

Table 2.3: *MSPE Results for Matérn :* $\nu_1 = 0.25$, $\nu_2 = 0.25$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9999	1.0000	1.0000	1.0000
421	0.9985	0.9998	0.9989	0.9990
761	0.9928	0.9983	0.9938	0.9939
1201	0.9865	0.9964	0.9878	0.9881
1741	0.9636	0.9831	0.9650	0.9652
2381	0.9418	0.9676	0.9431	0.9432
3121	0.9181	0.9481	0.9191	0.9192
3961	0.8931	0.9250	0.8939	0.8940
4901	0.8839	0.9173	0.8847	0.8847

[1] NUM.LOCS is total number of locations
[2] MAT.MSPE is the MSPE using the true covariance
[3] ASK.MSPE is the MSPE for Askey taper
[4] WEND.1.MSPE is the MSPE for Wendland type one taper
[5] WEND.2.MSPE is the MSPE for Wendland type two taper
[6] Refer to Figure 2.5 for graph of results

Table 2.4: *Time (in sec.) Results for Matérn :* $\nu_1 = 0.25$, $\nu_2 = 0.25$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.01	0.00	0.01	0.00
181	0.17	0.00	0.00	0.00
421	0.59	0.01	0.01	0.02
761	3.53	0.05	0.05	0.06
1201	13.78	0.22	0.21	0.23
1741	41.91	0.69	0.67	0.76
2381	107.29	1.54	1.58	1.65
3121	242.64	3.60	3.51	3.52
3961	494.24	7.59	7.65	7.71
4901	936.71	12.76	12.93	12.78

[1] NUM.LOCS is total number of locations
[2] MAT.TIME time to calculate cholesky decomposition for Matérn
[3] ASK.TIME time to calculate cholesky decomposition for Askey
[4] WEND.1.TIME time to calculate cholesky decomposition for Wendland 1
[5] WEND.2.TIME time to calculate cholesky decomposition for Wendland 2
[6] Refer to Figure 2.5 for graph of results

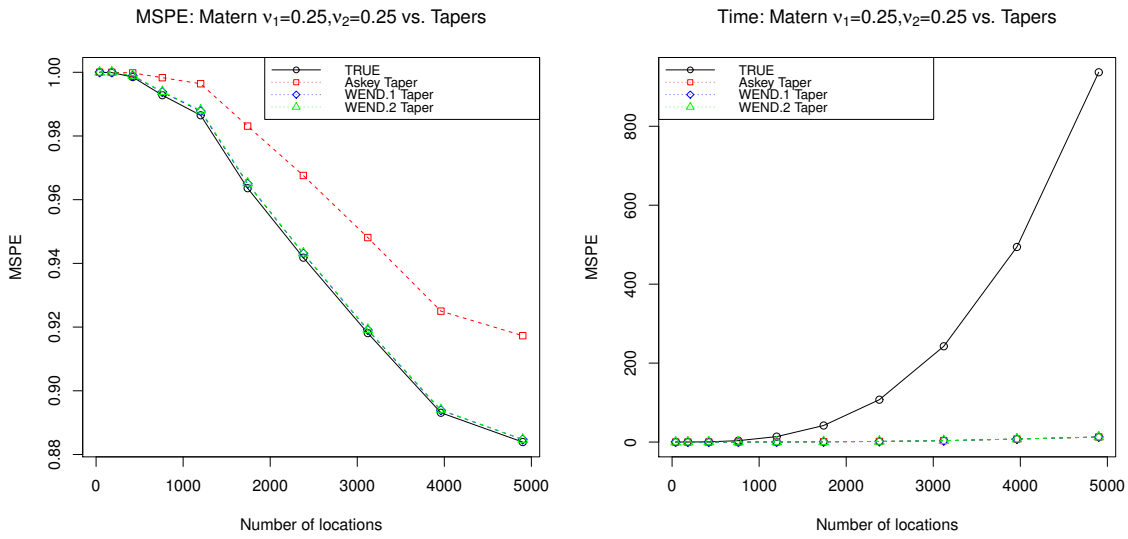

Figure 2.5: Graphical results of comparing the Matérn : $\nu_1 = 0.25$, $\nu_2 = 0.25$ and Tapers

Table 2.5: *MSPE Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 0.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9996	1.0000	0.9999	0.9999
421	0.9925	0.9991	0.9945	0.9949
761	0.9676	0.9922	0.9717	0.9724
1201	0.9424	0.9839	0.9475	0.9483
1741	0.8686	0.9359	0.8722	0.8726
2381	0.8084	0.8857	0.8108	0.8110
3121	0.7507	0.8288	0.7522	0.7523
3961	0.6965	0.7689	0.6974	0.6975
4901	0.6802	0.7486	0.6808	0.6809

[1] NUM.LOCS is total number of locations
[2] MAT.MSPE is the MSPE using the true covariance
[3] ASK.MSPE is the MSPE for Askey taper
[4] WEND.1.MSPE is the MSPE for Wendland type one taper
[5] WEND.2.MSPE is the MSPE for Wendland type two taper
[6] Refer to Figure 2.6 for graph of results

Table 2.6: *Time (in sec.) Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 0.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0000	0.0100	0.0100
181	0.1900	0.0000	0.0000	0.0100
421	0.5900	0.0100	0.0200	0.0100
761	3.5300	0.0600	0.0500	0.0600
1201	13.8500	0.2200	0.2100	0.2300
1741	41.9200	0.6600	0.6500	0.7600
2381	107.1300	1.5600	1.5800	1.6400
3121	241.3700	3.5100	3.4400	3.4800
3961	494.6700	7.7400	7.7500	7.7100
4901	937.7300	12.7900	13.0700	12.8900

[1] NUM.LOCS is total number of locations
[2] MAT.TIME time to calculate cholesky decomposition for Matérn
[3] ASK.TIME time to calculate cholesky decomposition for Askey
[4] WEND.1.TIME time to calculate cholesky decomposition for Wendland 1
[5] WEND.2.TIME time to calculate cholesky decomposition for Wendland 2
[6] Refer to Figure 2.6 for graph of results

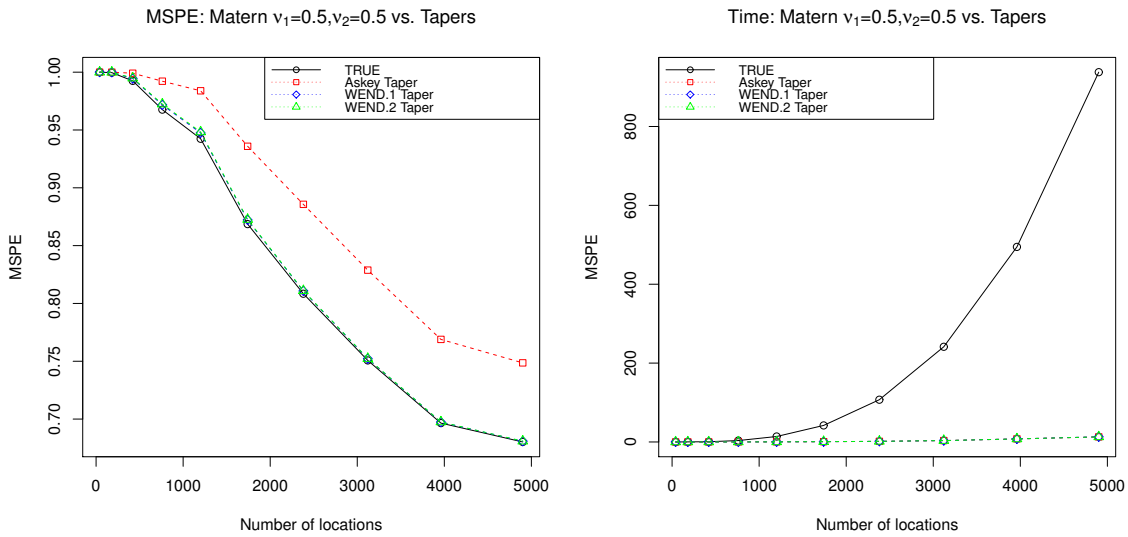

Figure 2.6: Graphical results of comparing the Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.5$ and Tapers

Table 2.7: *MSPE Results for Matérn :* $\nu_1 = 1.5$, $\nu_2 = 1.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9902	0.9998	0.9962	0.9972
421	0.8961	0.9884	0.9238	0.9295
761	0.7186	0.9283	0.7463	0.7518
1201	0.5957	0.8614	0.6131	0.6159
1741	0.3975	0.6532	0.4026	0.4026
2381	0.2921	0.5004	0.2936	0.2934
3121	0.2189	0.3754	0.2193	0.2192
3961	0.1664	0.2796	0.1666	0.1666
4901	0.1509	0.2421	0.1510	0.1511

[1] NUM.LOCS is total number of locations
[2] MAT.MSPE is the MSPE using the true covariance
[3] ASK.MSPE is the MSPE for Askey taper
[4] WEND.1.MSPE is the MSPE for Wendland type one taper
[5] WEND.2.MSPE is the MSPE for Wendland type two taper
[6] Refer to Figure 2.7 for graph of results

Table 2.8: *Time (in sec.) Results for Matérn :* $\nu_1 = 1.5$, $\nu_2 = 1.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0000	0.0000	0.0000
181	0.1900	0.0000	0.0000	0.0100
421	0.6000	0.0100	0.0200	0.0100
761	3.5100	0.0600	0.0500	0.0600
1201	13.8500	0.2200	0.2200	0.2200
1741	41.9100	0.6800	0.6600	0.8000
2381	107.6300	1.5700	1.6100	1.7200
3121	241.7400	3.5200	3.4300	3.7600
3961	494.1900	7.8000	7.8900	7.9200
4901	936.4500	12.8100	13.1400	12.9500

[1] NUM.LOCS is total number of locations
[2] MAT.TIME time to calculate cholesky decomposition for Matérn
[3] ASK.TIME time to calculate cholesky decomposition for Askey
[4] WEND.1.TIME time to calculate cholesky decomposition for Wendland 1
[5] WEND.2.TIME time to calculate cholesky decomposition for Wendland 2
[6] Refer to Figure 2.7 for graph of results

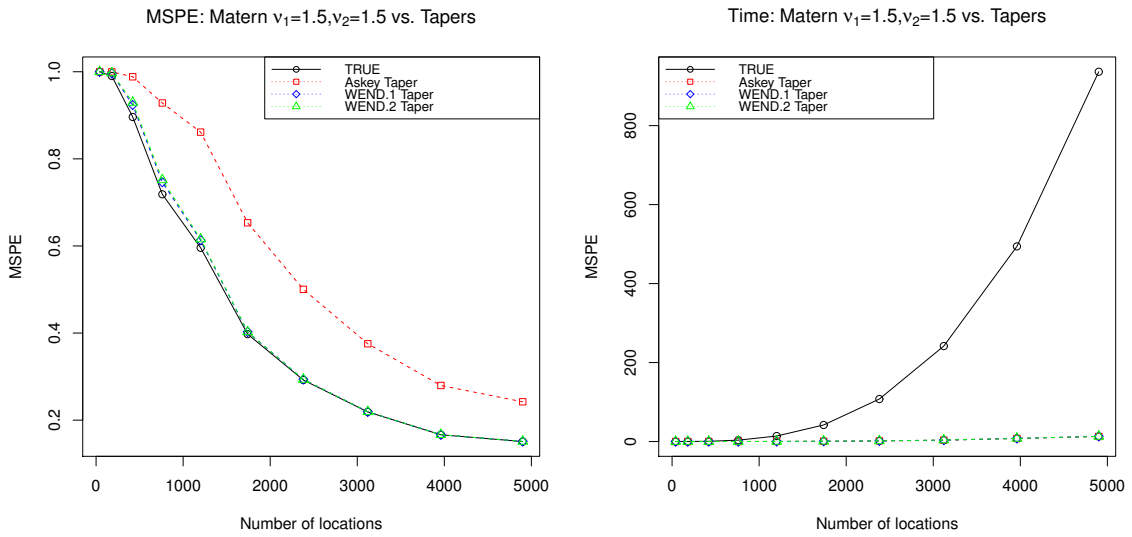

Figure 2.7: Graphical results of comparing the Matérn : $\nu_1 = 1.5$, $\nu_2 = 1.5$ and Tapers

Table 2.9: *MSPE Results for Matérn :* $\nu_1 = 2.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9551	0.9993	0.9832	0.9875
421	0.7161	0.9680	0.7878	0.8039
761	0.4432	0.8469	0.4838	0.4927
1201	0.3011	0.7214	0.3177	0.3198
1741	0.1500	0.4593	0.1536	0.1532
2381	0.0881	0.3019	0.0891	0.0887
3121	0.0540	0.1955	0.0543	0.0541
3961	0.0341	0.1268	0.0342	0.0342
4901	0.0273	0.0971	0.0273	0.0273

[1] NUM.LOCS is total number of locations
[2] MAT.MSPE is the MSPE using the true covariance
[3] ASK.MSPE is the MSPE for Askey taper
[4] WEND.1.MSPE is the MSPE for Wendland type one taper
[5] WEND.2.MSPE is the MSPE for Wendland type two taper
[6] Refer to Figure 2.8 for graph of results

Table 2.10: *Time (in sec.) Results for Matérn :* $\nu_1 = 2.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0100	0.0000	0.0000
181	0.1800	0.0000	0.0100	0.0000
421	0.5900	0.0100	0.0100	0.0100
761	3.5200	0.0600	0.0600	0.0600
1201	13.7800	0.2100	0.2200	0.2200
1741	42.5300	0.6700	0.6500	0.7900
2381	107.5000	1.5300	1.5500	1.6400
3121	241.6000	3.6500	3.4800	3.5300
3961	495.2000	7.8400	7.8600	7.7700
4901	938.1000	12.7000	13.0700	13.2800

[1] NUM.LOCS is total number of locations
[2] MAT.TIME time to calculate cholesky decomposition for Matérn
[3] ASK.TIME time to calculate cholesky decomposition for Askey
[4] WEND.1.TIME time to calculate cholesky decomposition for Wendland 1
[5] WEND.2.TIME time to calculate cholesky decomposition for Wendland 2
[6] Refer to Figure 2.8 for graph of results

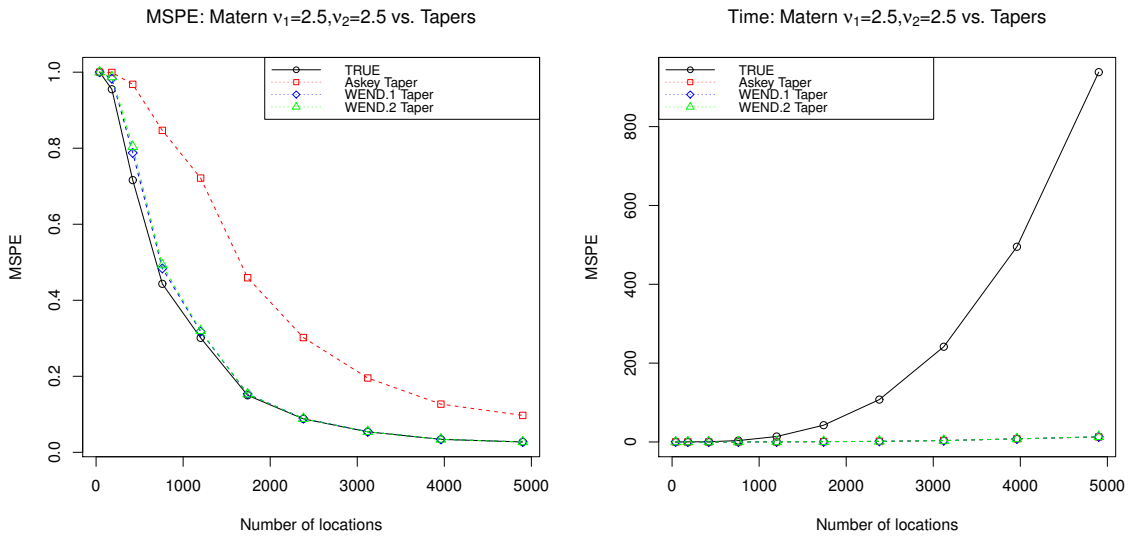

Figure 2.8: Graphical results of comparing the Matérn : $\nu_1 = 2.5$, $\nu_2 = 2.5$ and Tapers

2.5 USA Multivariate Climate Data

To show the application of multivariate tapering we examine precipitation and temperature data from the National Climatic Data Center for the years from 1895 to 1997. In this analysis we consider yearly total precipitation which is standardized over the long-run mean and standard deviation for each station. The average yearly temperature is also standardized in the same way.

The question here is, can yearly average temperature improve the spatial prediction of yearly total precipitation? So we co-krigged yearly total precipitation as the primary variable of interest with temperature being the secondary. To illustrate the tapering methods we used the year 1969, because this year had one of the most complete data records for both precipitation and temperature, with 5,182 stations and did not show any non-stationarity. The data can be found at *http://www.image.ucar.edu/public/Data*. The complexity here comes in the fact we are dealing with a covariance matrix that has dimension 10364×10364 since we have two variables of interest. The goal is to perform multivariate cokriging to show how tapering can give accurate predictions in the multivariate case as well as allow for time savings in highly complex calculations. We used the Multivariate Matérn Model discussed in Gneiting et al. (2010) Theorem 1 and using the RandomFields package in R, (Schlather (2011)) we found MLEs for the following parameters, (a subscript of P is for precipitation and T is for temperature): $\sigma_P = 0.747$, $\sigma_T = 0.238$, $\nu_P = 0.186$, $\nu_T = 0.799$, $\nu_{PT} = 0.493$, $\frac{1}{a_P} = 162$, $\frac{1}{a_T} = 162$, $\frac{1}{a_{PT}} = 162$, $\rho_{PT} = -0.0819$, $\tau_P = 0$, $\tau_T = 0.113$. Based on simulations in Sections 2.3 and 2.4, and the works of Kaufman et al. (2008) and Furrer et al. (2006) we use a tapering range of 50 miles (80.40 km) and a tapering $\alpha = 0.5$. To account of the smoothness of the multivariate Matérn , the Askey taper used a tapering parameter $\nu = 7$. Parameters for tapers where chosen based on previous simulation work and the lowest MSPE. For results we provide MSPE, predictions, and times to complete the process of calculating results for 100,000 points which create a equally spaced grid across the US. For matrix calculations, we used the Matrix package in R (Bates and Maechler (2011)), All

calculations were carried out on the Beocat of Kansas State University with a node having 16 GB of RAM and a 4x Quad-Core AMD Opteron 8350 Processor. .

Table 2.11 give the results for the time of constructing the covariance matrix, solving for the inverse and calculating all predictions for the 100,000 new points. Note that the average time it take to calculate one prediction of a new location for the Matérn , Askey, Wendland 1, and Wendland 2 are; 0.52606 sec., 0.51958 sec., 0.53198 sec., and 0.52192 sec. The time taken to calculate a prediction in all four cases is relatively the same. As in works of Kaufman et al. (2008) and Furrer et al. (2006) the time savings comes at the construction step and in the calculation of the inverse. Note that there is a huge improvement of time savings for calculating the the inverse from approximately 64 minutes for the true covariance to 6 minutes for the tapered covariances.

Table 2.11: *US Precipitation Time Results for Tapering*

Task	Matérn	Askey	Wendland 1	Wendland 2
Matrix Construction	86.01	53.17	55.72	73.66
Inverse Calculation	3852.49	416.09	448.33	419.07
Prediction Calculations	52601.91	51954.32	53194.61	52188.26
Total	56540	52424	53699	52681

[1] Time given in seconds

Table 2.12 shows the average MSPE and standard deviation of the MSPEs over the 100,000 predicted points. Notice that Wendland 1 has the best results which correspond to results found in the simulation studies, when choosing the proper taper.

Table 2.12: *US Precipitation MSPE Results for Tapering*

Taper Range	MATERN	Askey	Wendland 1	Wendland 2
Avg. MSPE	0.29288	0.36847	0.30523	0.30872
Std. MSPE	0.043293	0.081602	0.052982	0.055655

[1] Average MSPE and Std. taken over the 100,000 predicted points.

The following are images of the 100,000 predicted points using the tapering methods

and the original Matérn covariance matrices. Based on the simulations and kriging we found that the tapering alpha parameter does not effect predictions greatly. However as alpha increases to 1 the Askey taper improves slightly and as alpha decreases to 0 both the Wendland Type 1 and Wendland Type 2 improve slightly. Notice that the Wendland Type 1 kriging produces the best image when compared to the Matérn in Figure 2.9 further backing up the simulation study results.

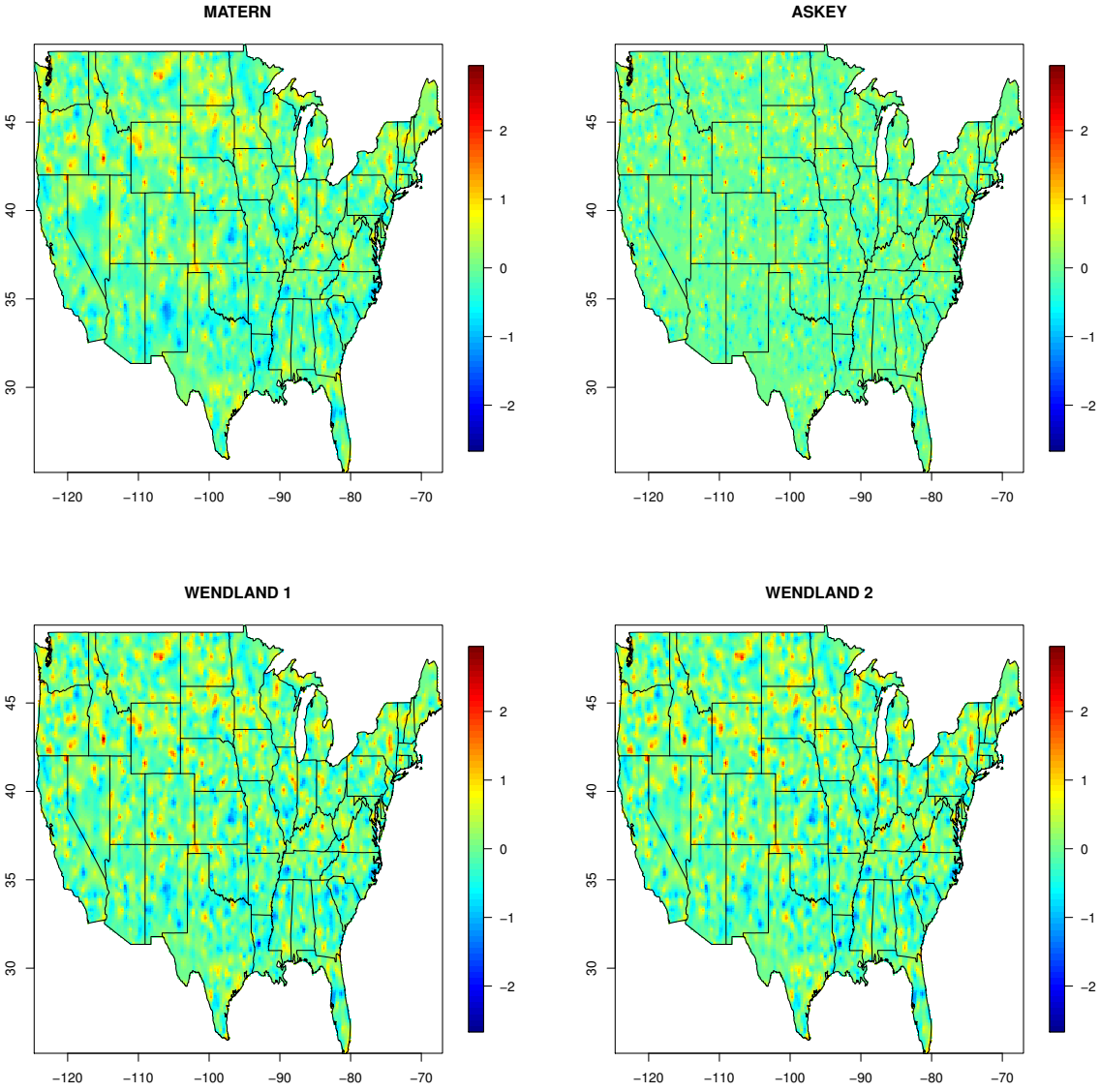

Figure 2.9: Total yearly precipitation kriging using Matérn , Askey, Wendland 1, and Wendland 2

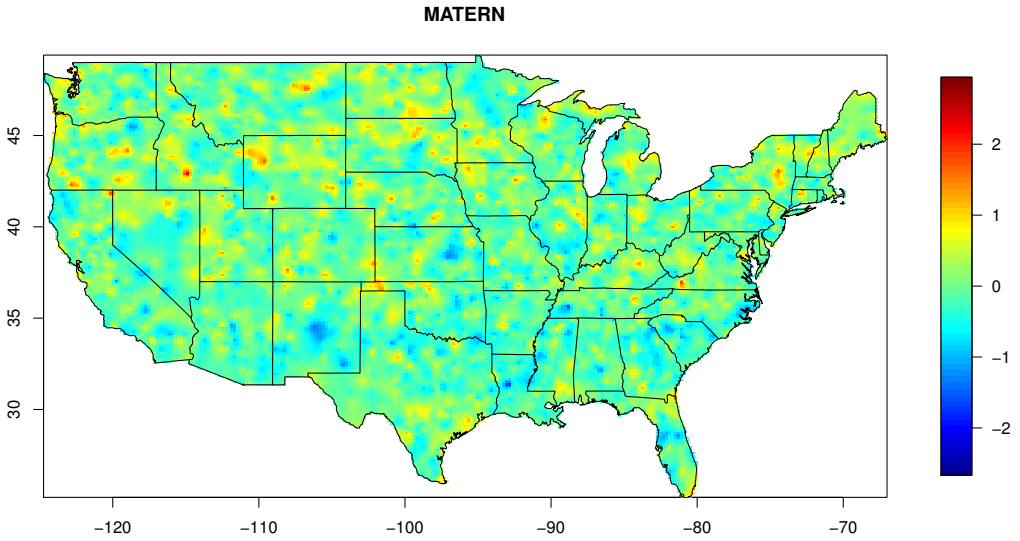

Figure 2.10: Total yearly precipitation prediction for USA in 1969 Using original Matérn

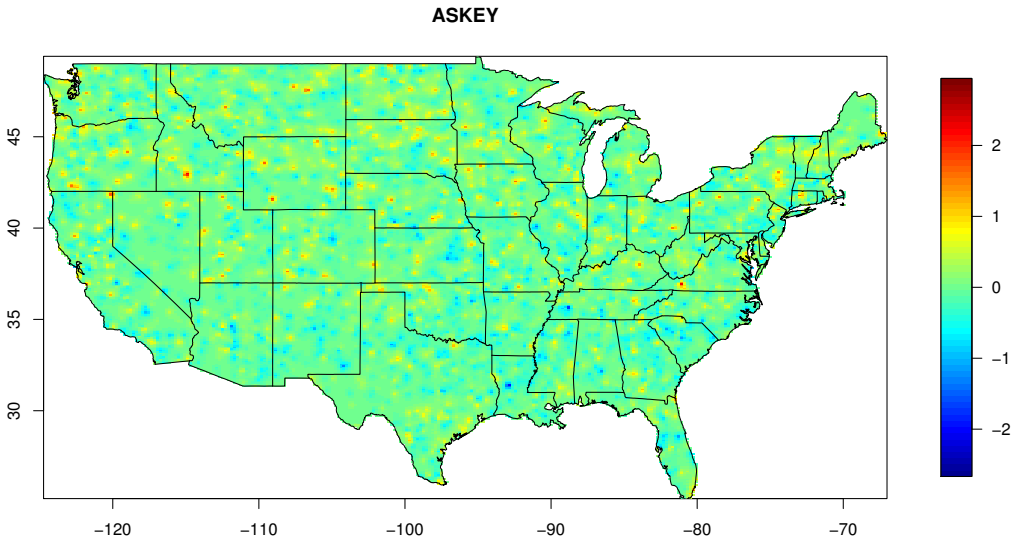

Figure 2.11: Total yearly precipitation prediction for USA in 1969 Using Askey Taper

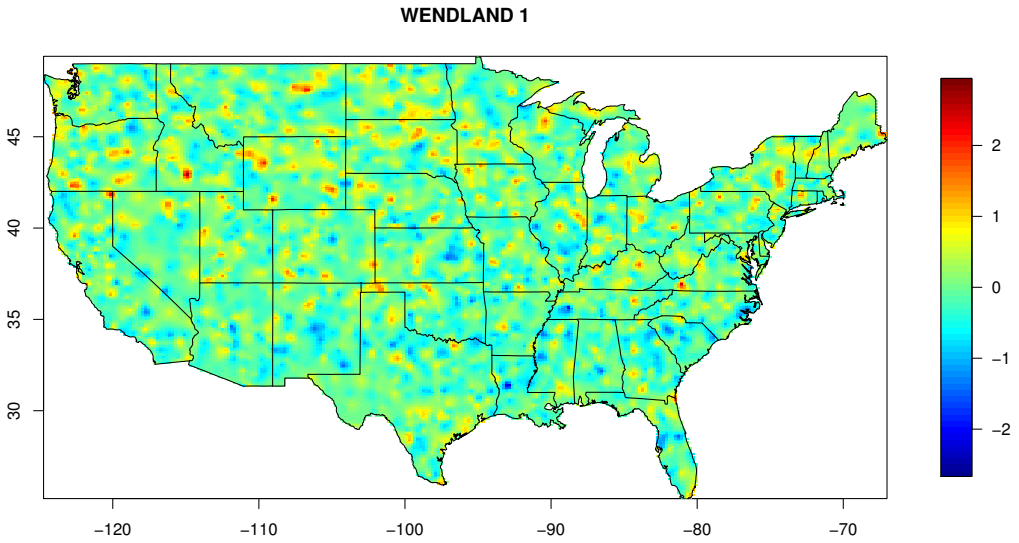

Figure 2.12: Total yearly precipitation prediction for USA in 1969 Using Wendland 1 Taper

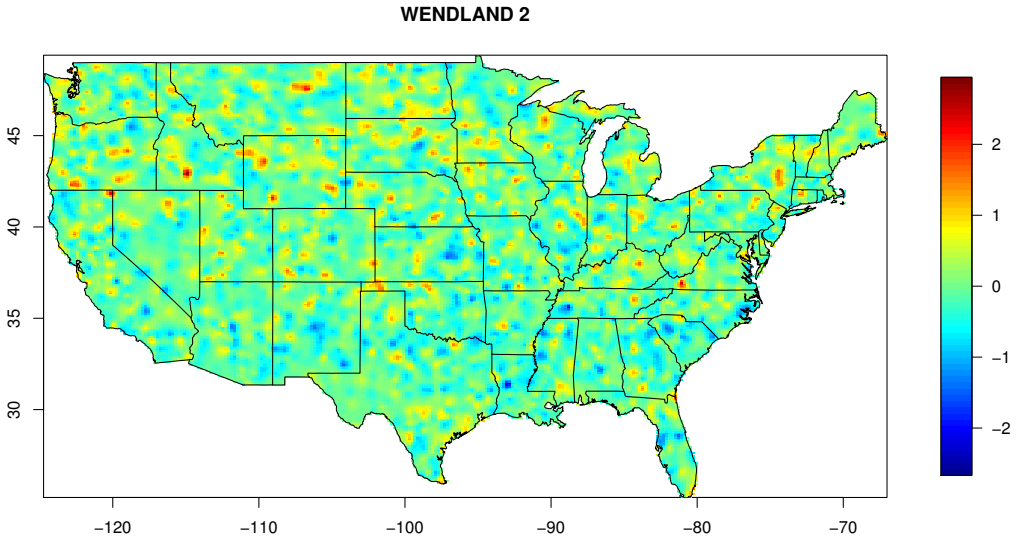

Figure 2.13: Total yearly precipitation prediction for USA in 1969 Using Wendland 2 Taper

2.6 Chapter 2 Appendix

2.6.1 Proof of Theorem 2.2.1

We evaluate the Fourier transform matrix of 2.10, which is positively proportional to

$$\begin{pmatrix} (\theta_{11}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & (\theta_{12}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & \dots & (\theta_{1m}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} \\ (\theta_{21}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & (\theta_{22}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & \dots & (\theta_{2m}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ (\theta_{m1}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & (\theta_{m2}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} & \dots & (\theta_{mm}+\|\boldsymbol{\omega}\|^2)^{-\frac{d+1}{2}} \end{pmatrix}, \tag{2.24}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$, by using Theorem 1.14 of Stein and Weiss (1971). This matrix is positive definite for each fixed $\boldsymbol{\omega} \in \mathbb{R}^d$ since its entries can be rewritten as

$$(\theta_{ij}+\|\boldsymbol{\omega}\|^2)^{-1} = \int_0^\infty \exp(-\|\boldsymbol{\omega}\|^2 u)\exp(-\theta_{ij}u)du,\ i,j=1,\dots,m,$$

and the matrix with entries $\exp(-\theta_{ij}u)$ is positive definite due to the assumption that Θ is conditionally negative definite (see Theorem 1 in Du and Ma (2011)). According to the Cramér-Kolmogorov Theorem (Cramár (1940)), (2.10) is a covariance matrix function. ■

2.6.2 Proof of Theorem 2.2.2

The Fourier transform matrix of 2.12 is positively proportional to

$$\begin{pmatrix} \prod_{k=1}^d(\theta_{11}+\omega_k^2)^{-1} & \prod_{k=1}^d(\theta_{12}+\omega_k^2)^{-1} & \dots & \prod_{k=1}^d(\theta_{1m}+\omega_k^2)^{-1} \\ \prod_{k=1}^d(\theta_{21}+\omega_k^2)^{-1} & \prod_{k=1}^d(\theta_{22}+\omega_k^2)^{-1} & \dots & \prod_{k=1}^d(\theta_{2m}+\omega_k^2)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{k=1}^d(\theta_{m1}+\omega_k^2)^{-1} & \prod_{k=1}^d(\theta_{m2}+\omega_k^2)^{-1} & \dots & \prod_{k=1}^d(\theta_{mm}+\omega_k^2)^{-1} \end{pmatrix}, \tag{2.25}$$

and is positive definite for each fixed $\boldsymbol{\omega} = (\omega_1,\dots,\omega_d)' \in \mathbb{R}^d$ since its entries can be rewritten as

$$\prod_{k=1}^d(\theta_{ij}+\omega_k^2)^{-1} = \prod_{k=1}^d \int_0^\infty \exp(-\omega_k^2 u)\exp(-\theta_{ij}u)du,\ i,j=1,\dots,m,$$

and, by assumption, Θ is conditionally negative definite. According to the Cramér-Kolmogorov Theorem, Cramár (1940), (2.12) is a covariance matrix function. ■

2.6.3 Proof of Theorem 2.2.5

Clearly, an equivalent form of (2.22) is

$$C_{ij}(\mathbf{x}) = \int_0^1 K(\frac{\|\mathbf{x}\|}{u}) g_{ij}(u) du, \ \mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \ldots, m.$$

The Hardmard product of given matrix function and a positive definite matrix with entries $g_{ij}(u)$ results in a valid covariance matrix function by Lemma 2.2.1. Consequently this theorem is proved by applying modified Lemma 2.2.2.

2.6.4 Proof of Theorem 2.2.6

In Theorem 2.2.5 take $K(\cdot) = u^k I^k \phi_{\nu,0}(\cdot)$. Note that $u^k I^k \phi_{\nu,0}(\cdot)$ is a univariate covariance function because $I^k \phi_{\nu,0}(\cdot)$ is, when $\nu \geq \frac{d+1}{2}, k = 0, 1, 2, \ldots$. ■

2.6.5 Proof of Corollary 2.2.3

By Theorem 2.2.6, together with the univariate Wendland tapering function (2.19), (2.23) becomes

$$C_{w,ij}(\mathbf{x}) = \begin{cases} \int_0^1 (u - \|\mathbf{x}\|)_+^4 (u + 4\|\mathbf{x}\|) g_{ij}(u) du, \ \|\mathbf{x}\| \leq 1, \\ 0, \ \|\mathbf{x}\| > 1, \end{cases} \tag{2.26}$$

$\mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \ldots, m$. Completing the integration completes the proof. ■

2.6.6 Proof of Corollary 2.2.4

By Theorem 2.2.6, together with the univariate Wendland tapering function (2.20), , (2.23) becomes

$$C_{w,ij}(\mathbf{x}) = \begin{cases} \int_0^1 (u - \|\mathbf{x}\|)_+^6 \left(u^2 + 6u\|\mathbf{x}\| + \frac{35}{3}x\right) g_{ij}(u) du, \ \|\mathbf{x}\| \leq 1, \\ 0, \ \|\mathbf{x}\| > 1, \end{cases} \tag{2.27}$$

$\mathbf{x} \in \mathbb{R}^d, \ i, j = 1, \ldots, m.$

Completing the integration completes the proof. ■

2.6.7 Additional Bivariate Exponential Results

The following are more results for the bivariate exponential case when $\gamma = .5, .3$.

Table 2.13: *Results for Exponential and Askey-Taper with* $\gamma = .5$

NUM.LOCS	EXP.MSPE	EXP.TIME	Askey.TIME	Askey.MSPE
400	0.0288870	0.30	0.30	0.0295190
625	0.0142470	0.71	0.71	0.0143110
900	0.0107040	2.12	1.75	0.0107220
1225	0.0160650	5.35	4.33	0.0160900
1600	0.0117850	12.04	9.48	0.0117940
2025	0.0024041	24.09	19.79	0.0024050
2500	0.0104930	45.88	38.51	0.0104940
3025	0.0096929	88.50	63.78	0.0096927
3600	0.0026003	181.06	105.31	0.0025995
4225	0.0068860	525.65	200.30	0.0068912
4900	0.0079591	727.54	287.45	0.0079560
5625	0.0046308	1207.70	418.64	0.0046212
6400	0.0039480	1882.70	606.73	0.0039294
7225	0.0065214	2934.00	920.14	0.0065031
8100	0.0051621	5185.10	1456.30	0.0051650

[1] Refer to Figure 2.14 for graph of results

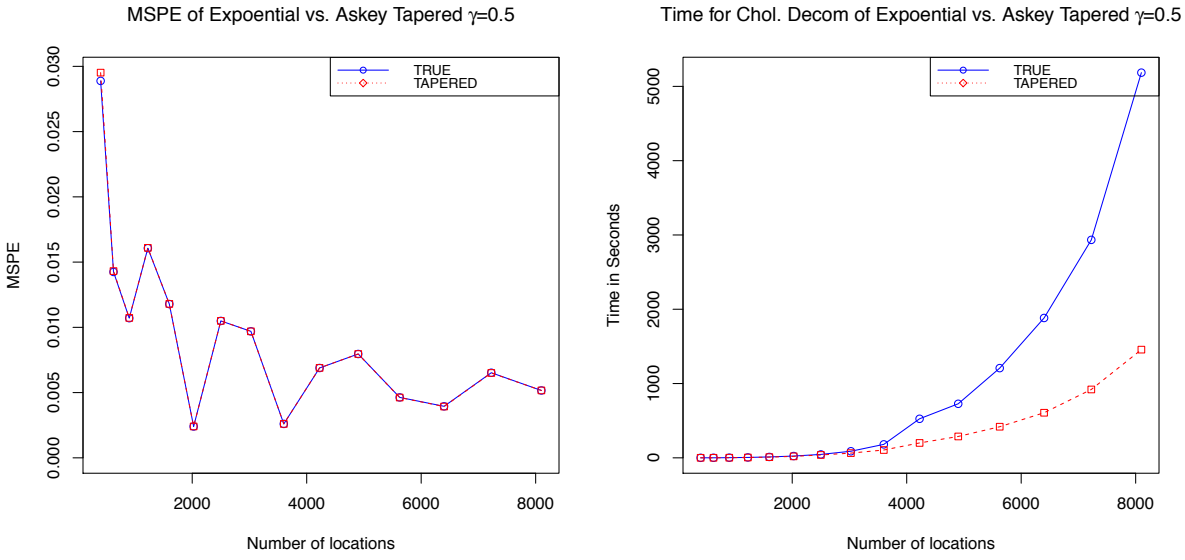

Figure 2.14: Results Graph: Exponential and Askey taper, tapered with $\gamma = 0.5$

Table 2.14: *Results for Exponential and Askey-Taper with* $\gamma = 0.3$

NUM.LOCS	EXP.MSPE	EXP.TIME	Askey.TIME	Askey.MSPE
400	0.0288870	0.32	0.20	0.0369410
625	0.0142470	0.72	0.36	0.0151380
900	0.0107040	2.14	0.74	0.0109690
1225	0.0160650	5.37	1.60	0.0164190
1600	0.0117850	11.94	3.44	0.0118980
2025	0.0024041	24.28	6.22	0.0024076
2500	0.0104930	45.97	12.36	0.0105300
3025	0.0096929	88.98	21.20	0.0097140
3600	0.0026003	178.38	37.15	0.0026005
4225	0.0068860	588.64	62.17	0.0068968
4900	0.0079591	773.91	100.78	0.0079618
5625	0.0046308	1318.50	142.92	0.0046226
6400	0.0039480	2094.10	213.86	0.0039302
7225	0.0065214	2781.20	312.71	0.0065050
8100	0.0051621	4306.90	445.93	0.0051659

[1] Refer to Figure 2.15 for graph of results

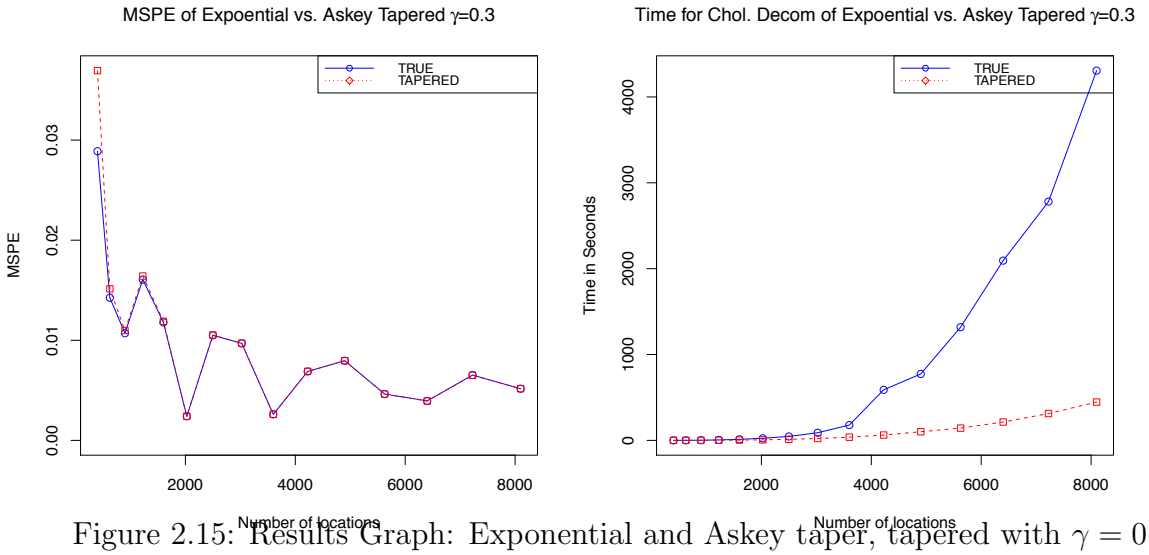

Figure 2.15: Results Graph: Exponential and Askey taper, tapered with $\gamma = 0.3$

2.6.8 Additional Bivariate Matérn Results

The following are more results for the bivariate Matérn case.

Table 2.15: *MSPE Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 0.25$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASKEY.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9996	1.0000	0.9999	0.9999
421	0.9923	0.9991	0.9944	0.9948
761	0.9668	0.9922	0.9714	0.9721
1201	0.9412	0.9838	0.9470	0.9478
1741	0.8666	0.9357	0.8710	0.8714
2381	0.8061	0.8855	0.8092	0.8094
3121	0.7483	0.8285	0.7504	0.7504
3961	0.6941	0.7686	0.6954	0.6955
4901	0.6779	0.7484	0.6790	0.6790

[1] Refer to Figure 2.16 for graph of results

Table 2.16: *Time (in sec.) Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 0.25$ *and Tapers*

NUM.LOCS	MAT.TIME	ASKEY.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0000	0.0100	0.0000
181	0.1800	0.0100	0.0000	0.0000
421	0.6000	0.0100	0.0100	0.0100
761	3.5400	0.0600	0.0600	0.0600
1201	13.8100	0.2200	0.2100	0.2100
1741	41.9100	0.7000	0.6900	0.8300
2381	107.2400	1.5200	1.5400	1.6300
3121	241.7700	3.5500	3.3500	3.7000
3961	494.4600	7.6100	7.6800	7.7300
4901	936.7200	12.7900	12.8400	12.7600

[1] Refer to Figure 2.16 for graph of results

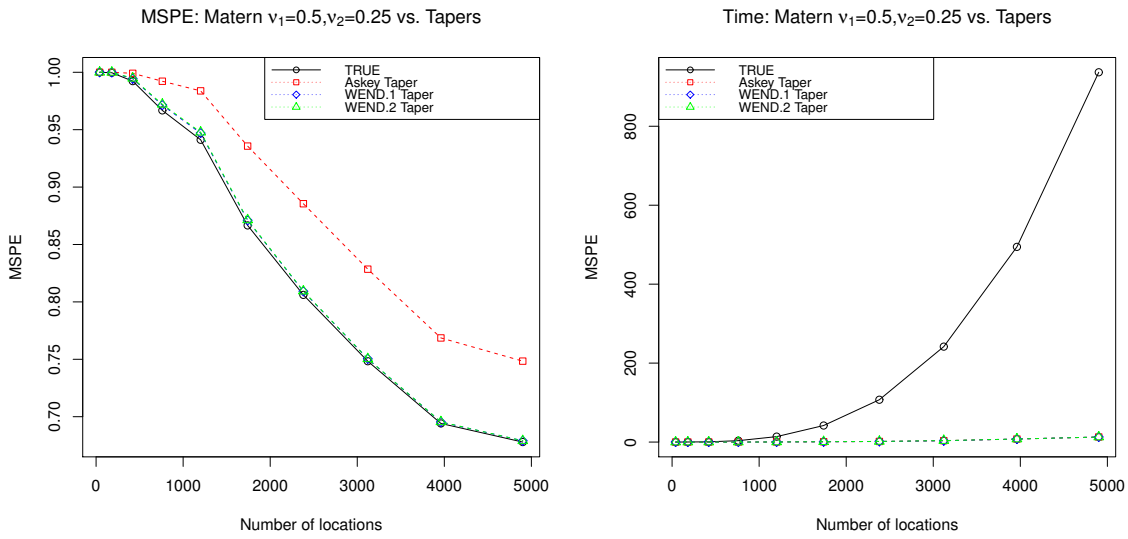

Figure 2.16: Graphical results of comparing the Matérn : $\nu_1 = 0.5$, $\nu_2 = 0.25$ and Tapers

Table 2.17: *MSPE Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 1.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9993	1.0000	0.9997	0.9998
421	0.9882	0.9988	0.9920	0.9926
761	0.9569	0.9899	0.9632	0.9641
1201	0.9290	0.9794	0.9355	0.9364
1741	0.8515	0.9252	0.8564	0.8568
2381	0.7914	0.8718	0.7953	0.7955
3121	0.7346	0.8134	0.7379	0.7381
3961	0.6816	0.7537	0.6846	0.6848
4901	0.6669	0.7344	0.6695	0.6697

[1] Refer to Figure 2.17 for graph of results

Table 2.18: *Time (in sec.) Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 1.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0000	0.0100	0.0000
181	0.1700	0.0100	0.0000	0.0000
421	0.6000	0.0100	0.0100	0.0100
761	3.5400	0.0600	0.0500	0.0600
1201	13.7800	0.2300	0.2100	0.2200
1741	41.9400	0.6700	0.6600	0.8300
2381	107.2000	1.5400	1.5900	1.7600
3121	241.4700	3.4900	3.3400	3.3100
3961	493.2400	7.8000	7.8100	7.7900
4901	937.0200	12.9700	13.2400	12.7100

[1] Refer to Figure 2.17 for graph of results

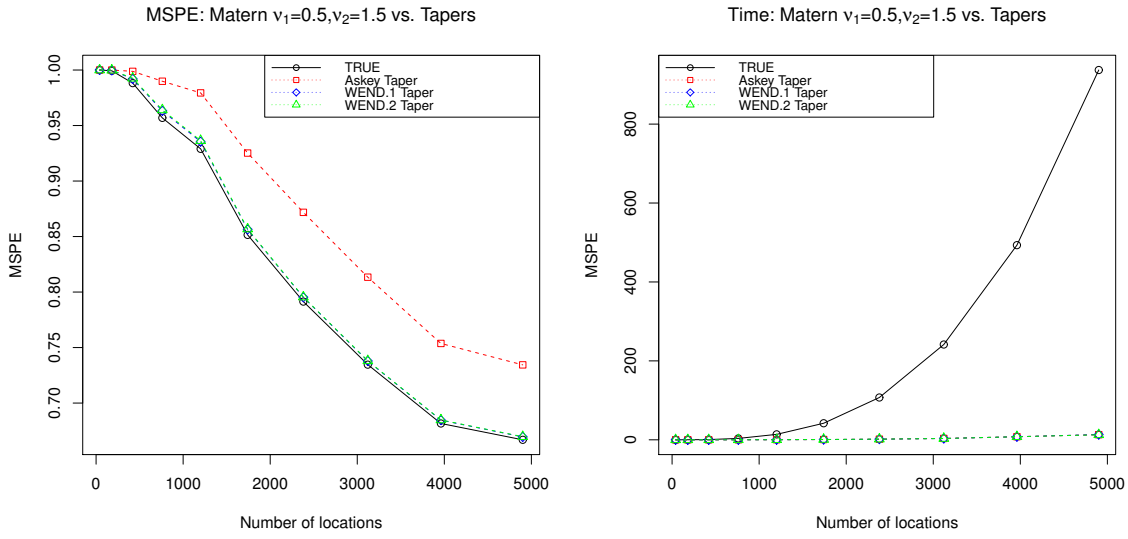

Figure 2.17: Graphical results of comparing the Matérn : $\nu_1 = 0.5$, $\nu_2 = 1.5$ and Tapers

Table 2.19: *MSPE Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9975	1.0000	0.9993	0.9995
421	0.9754	0.9977	0.9845	0.9857
761	0.9321	0.9845	0.9441	0.9455
1201	0.9007	0.9698	0.9123	0.9133
1741	0.8201	0.9087	0.8319	0.8323
2381	0.7614	0.8536	0.7734	0.7737
3121	0.7070	0.7965	0.7193	0.7195
3961	0.6564	0.7395	0.6689	0.6690
4901	0.6434	0.7217	0.6554	0.6555

[1] Refer to Figure 2.18 for graph of results

Table 2.20: *Time (in sec.) Results for Matérn :* $\nu_1 = 0.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0100	0.0000	0.0000
181	0.1700	0.0000	0.0100	0.0000
421	0.6000	0.0100	0.0100	0.0100
761	3.5200	0.0600	0.0500	0.0500
1201	13.8400	0.2200	0.2200	0.2300
1741	41.9500	0.6700	0.7000	0.8100
2381	107.4100	1.5500	1.5500	1.6400
3121	241.3400	3.6400	3.4800	3.6300
3961	493.9600	7.6800	7.9000	7.8300
4901	935.7000	12.9100	13.0200	13.1000

[1] Refer to Figure 2.18 for graph of results

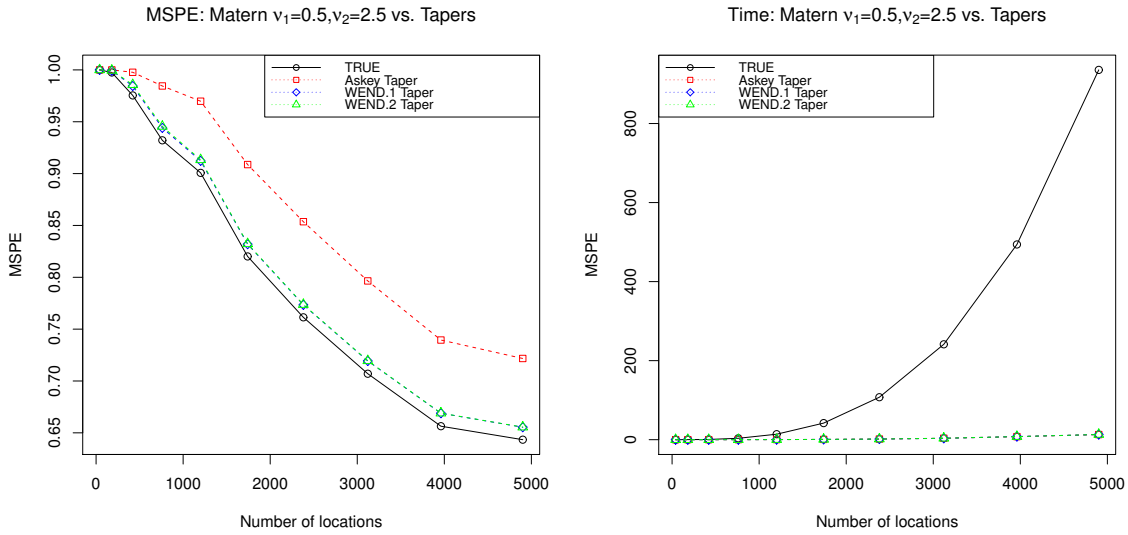

Figure 2.18: Graphical results of comparing the Matérn : $\nu_1 = 0.5$, $\nu_2 = 2.5$ and Tapers

Table 2.21: *MSPE Results for Matérn : $\nu_1 = 1.5$, $\nu_2 = 0.5$ and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9895	0.9998	0.9961	0.9971
421	0.8915	0.9882	0.9222	0.9281
761	0.7111	0.9278	0.7426	0.7483
1201	0.5885	0.8606	0.6089	0.6118
1741	0.3918	0.6527	0.3985	0.3985
2381	0.2878	0.5005	0.2903	0.2899
3121	0.2156	0.3761	0.2167	0.2164
3961	0.1640	0.2807	0.1645	0.1644
4901	0.1488	0.2434	0.1491	0.1491

[1] Refer to Figure 2.19 for graph of results

Table 2.22: *Time (in sec.) Results for Matérn :* $\nu_1 = 1.5$, $\nu_2 = 0.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0100	0.0000	0.0000
181	0.1900	0.0100	0.0100	0.0000
421	0.5900	0.0100	0.0100	0.0200
761	3.5200	0.0500	0.0600	0.0600
1201	13.8300	0.2200	0.2000	0.2200
1741	42.0100	0.7200	0.6800	0.8300
2381	107.2000	1.6000	1.5900	1.6500
3121	241.5600	3.6200	3.3500	3.5500
3961	493.8200	7.7500	7.6700	7.8100
4901	935.8400	12.8800	12.9200	12.9100

[1] Refer to Figure 2.19 for graph of results

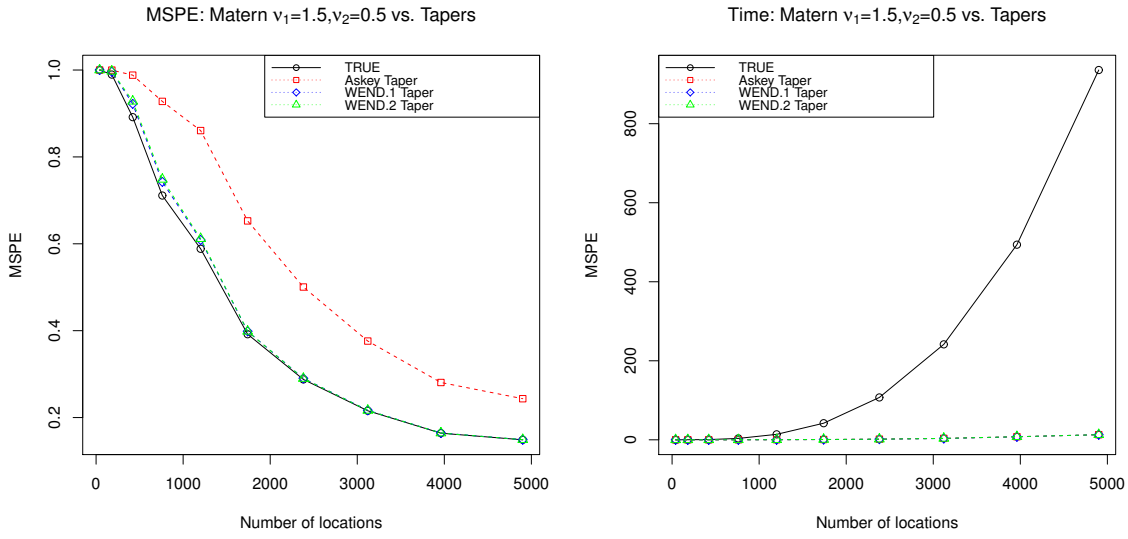

Figure 2.19: Graphical results of comparing the Matérn : $\nu_1 = 1.5$, $\nu_2 = 0.5$ and Tapers

Table 2.23: *MSPE Results for Matérn :* $\nu_1 = 1.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.MSPE	ASK.MSPE	WEND.1.MSPE	WEND.2.MSPE
41	1.0000	1.0000	1.0000	1.0000
181	0.9892	0.9998	0.9959	0.9969
421	0.8923	0.9879	0.9207	0.9266
761	0.7146	0.9265	0.7416	0.7471
1201	0.5926	0.8584	0.6092	0.6119
1741	0.3956	0.6496	0.4005	0.4005
2381	0.2908	0.4972	0.2924	0.2922
3121	0.2179	0.3731	0.2186	0.2185
3961	0.1658	0.2781	0.1661	0.1661
4901	0.1503	0.2409	0.1506	0.1507

[1] Refer to Figure 2.20 for graph of results

Table 2.24: *Time (in sec.) Results for Matérn :* $\nu_1 = 1.5$, $\nu_2 = 2.5$ *and Tapers*

NUM.LOCS	MAT.TIME	ASK.TIME	WEND.1.TIME	WEND.2.TIME
41	0.0000	0.0000	0.0000	0.0000
181	0.2000	0.0100	0.0000	0.0000
421	0.6000	0.0100	0.0100	0.0100
761	3.5400	0.0600	0.0600	0.0500
1201	13.8100	0.2100	0.2300	0.2200
1741	42.0400	0.6900	0.6600	0.8200
2381	107.3100	1.5200	1.5100	1.6700
3121	241.8300	3.5100	3.3900	3.5700
3961	494.1800	7.6500	7.7800	7.7300
4901	937.1500	13.0100	13.2500	12.8500

[1] Refer to Figure 2.20 for graph of results

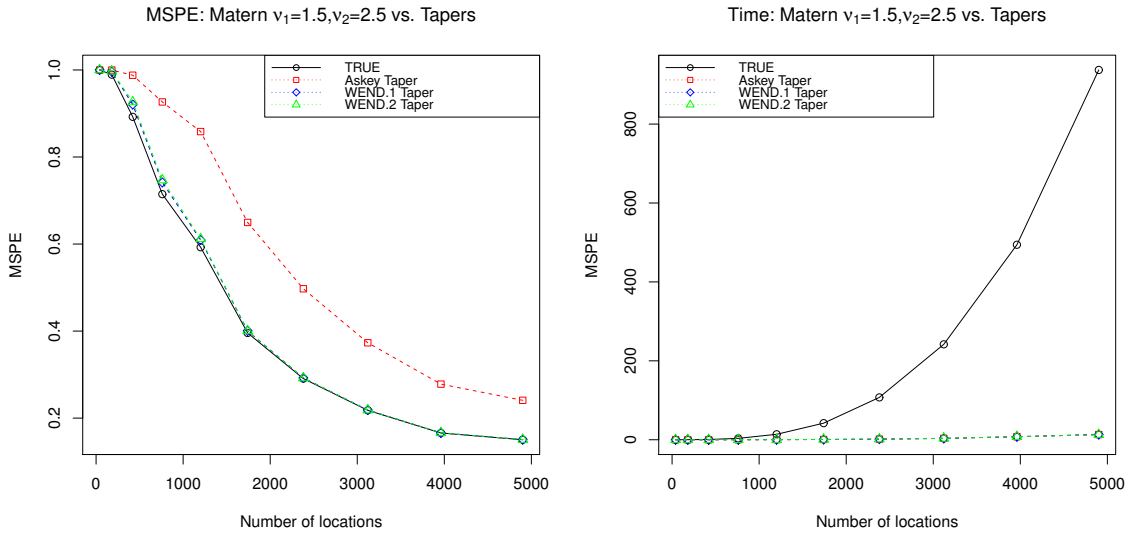

Figure 2.20: Graphical results of comparing the Matérn : $\nu_1 = 1.5$, $\nu_2 = 2.5$ and Tapers

Chapter 3

Functional space-time modeling

3.1 Introduction and Background

The development of a crop during its life cycle is heavily influenced by weather variables such as precipitation, maximum temperature, and minimum temperature. In agriculture the behavior of these weather variables can be used to construct a model that explains weather impact on corp production, and also predict crop yield. Usually the data collected on these weather elements can be seen as a function taken over time on a yearly, daily, or hourly scale. Throughout the life span of a crop there are important time periods that weather can positively or negatively influence the resulting crop yield. It would be beneficial if there was a model that would identify these important time points and give a measure of how influential each weather element is on crop yield. Staggenborg et al. (2008) did an exploratory analysis of minimum temperature, maximum temperature, and precipitation to see what relationship each weather element has with crop production. They explain that the growing season for Kansas corn is form April to September and that the vital time for a corn plant's life is from June 20th to July 10th. Within this short time frame they find guidelines for what weather conditions produce a good crop. For example, during this time the corn needs lots of rain, hot days, and cool nights according to Staggenborg et al. (2008). Our goal is to analytical explore as well as pin point the time periods in a growing season where these weather variables have either a negative or positive impact on the resulting corn harvest.

In this chapter we will review functional linear modeling and how it can be used to better examine weather effects on crop yield. We will also explore a spatial dependence covariance structure and add it to the functional linear model. Using these ideas of a functional linear model with spatial dependence we will model the effects of weather on Kansas corn yield from the years 1990-2011.

The Kansas weather data that are analyzed are gathered from 1123 weather stations in Kansas provided by the National Climatic Data Center (NCDC), and the corn yield data was gathered from U.S. Department of Agriculture (USDA) for each of the 105 counties. Figure 3.1 shows the locations of all 1123 stations represented by the black dots. Maximum

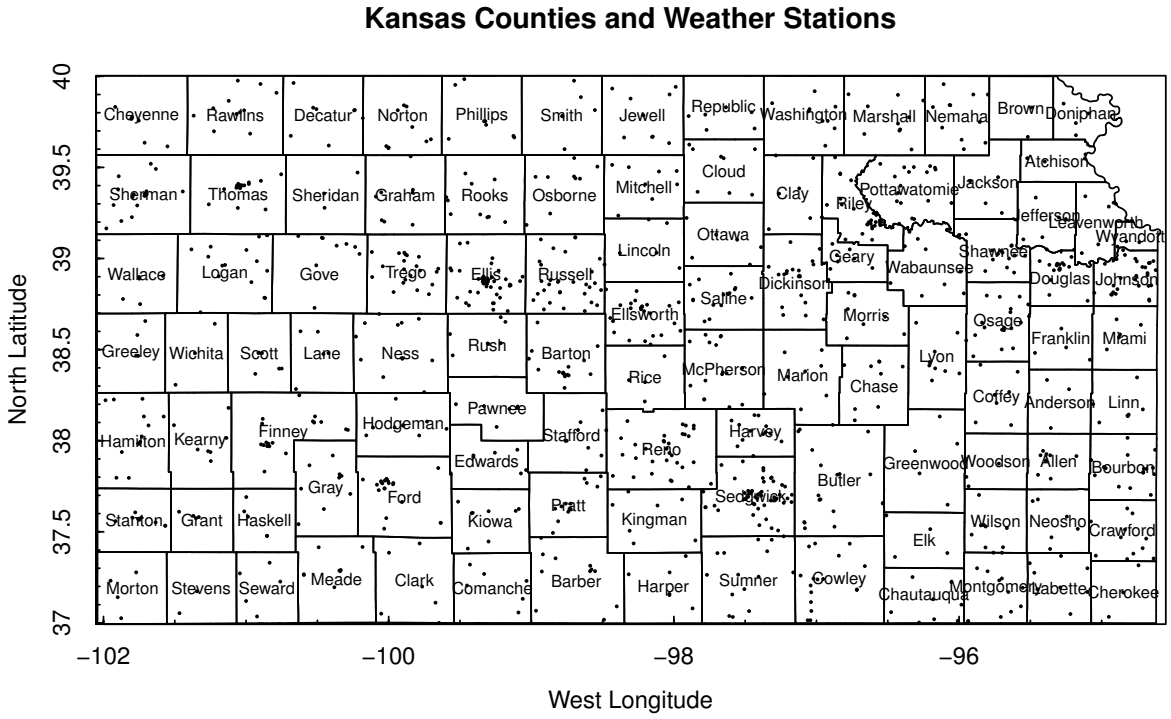

Figure 3.1: Weather stations of Kansas: black dots are each station.

and minimum daily temperatures are measured in degrees Celsius and daily precipitation is recorded in millimeters. Take a look at an example of raw daily weather data in Figure 3.2,

here the red line is maximum temperature, blue is minimum temperature, and black is precipitation. Notice how the daily weather readings can be thought of as a function of time. The idea that daily weather readings can be seen as a function gives us the motivation to use functional data analysis. One of the biggest challenges when dealing with raw data is

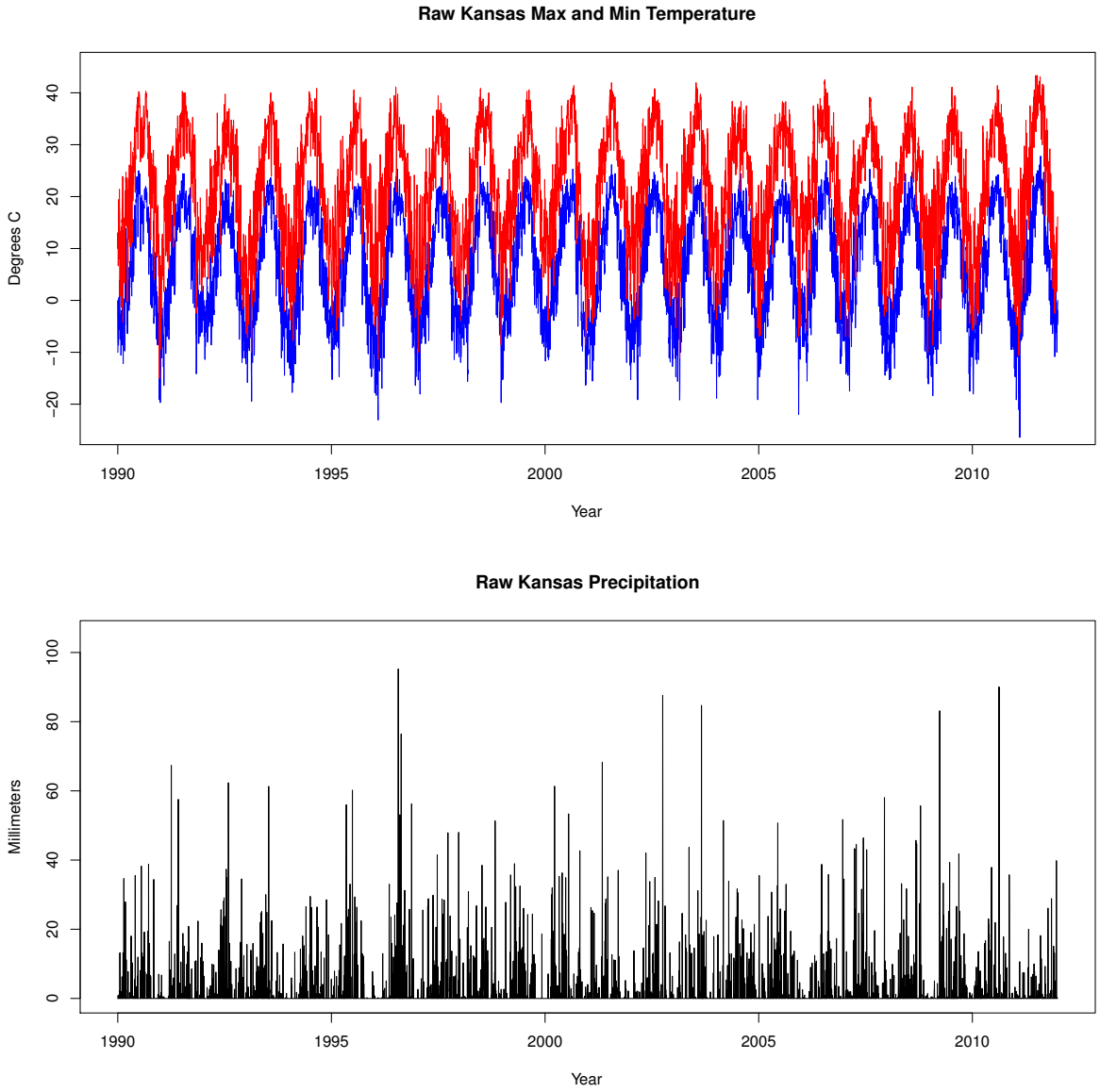

Figure 3.2: Finney County weather elements, red is Max Temp, blue is Min Temp, and black is Precipitation. Temperature is measured in degrees C and precipitation is measured in millimeters.

missing data. Throughout the Kansas weather data missing information frequently occurs

from station to station. Some stations have years with missing months which causes a problem, example Thomas County. Because of this, a significant portion of the data needs to be cleaned and in some case interpolated. To solve this missing data problem covariance models discussed in Chapter 1 and time series techniques are used to interpolate the data. To interpolate the precipitation data we use similar techniques discussed in Chapter 1 by applying a modified version of the spatio-temporal covariance function with a MA(1) temporal margin. Time series and spatial statistics are used to predict the missing temperature values over both space and time.

The response variable of interest is the yield data, which is measured in average bushels of corn per acre harvested each year. The yield data was collected form 1990 to 2011, an example of the yearly crop yield for three counties is given in Figure 3.3; Finney county in blue, Pottawatomie in red, and Thomas in green. For the yield data there is really no

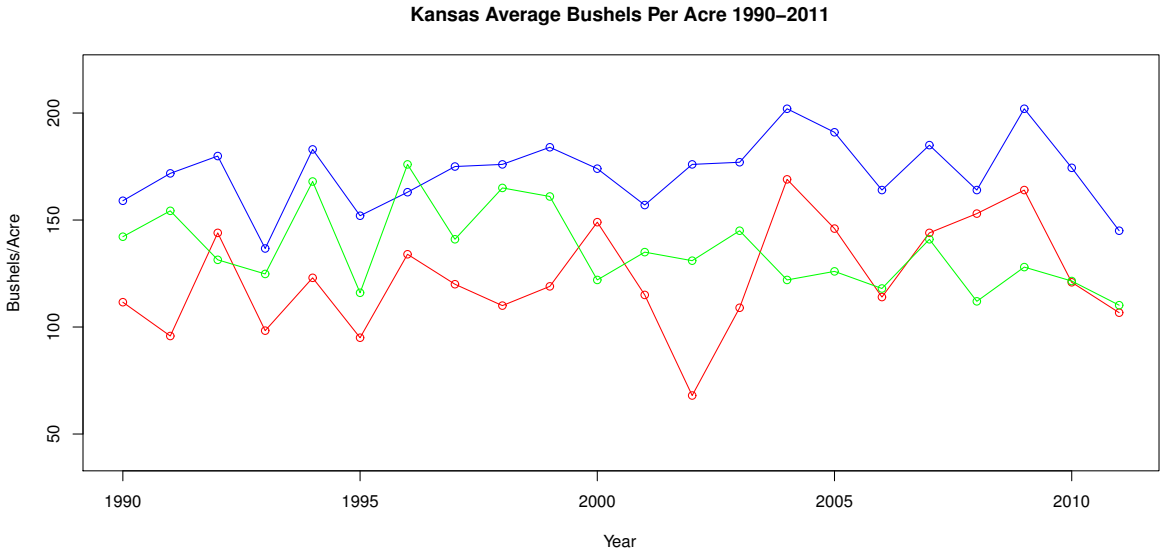

Figure 3.3: Plot of Corn Yield for Dickinson, Finney, and Sherman Counties. Blue is Dickinson County, green is Finney County, and red is Sherman County

missing data problems and minimal cleaning was done. One very interesting part of this modeling is the different scales of time. Weather data is recorded on a daily basis were yield is only collected once a year after harvest. This again is another reason functional

data analysis can provide insight to how weather plays a role in the development of crops. Spatial because yield is only recorded at the county level we have to aggregate the data. To do this we find an average daily weather reading for each county based on the weather stations in that county. This results in having a time series for each weather component for each county, a total of 315 total time series. With the back ground of the data now in place it is time to introduce the model. To begin we review the B-Spline basis functions in Section 3.2.1 that are used to estimate the functional parameter of the functional linear model in Section 3.2.2. In Section 3.2.3 we introduce a functional linear model that allows the scalar responses to have a spatial dependence structure.

3.2 Methodology

3.2.1 B-Splines

Functional data usually have a framework that can be represented by a smooth curve. In order to approximate this curve we can use many techniques, one in particular is splining. A spline is a smooth polynomial piecewise function that has controllable smoothness at the places where the polynomial pieces connect known as knots. Splining is commonly used for interpolating a function, smoothing a function, or in approximating the behavior of a function. Historical the word "spline" came from the ship building industry which meant a thin piece wood to create curve. To create the curve in the wood weights were placed on the wood in order to bend the wood into the desired shape. This is the same process to create a spline curve. A spline curve is a sequence of curve segments that are connected together to form a single continuous curve. The connecting points of the spline are called knots. The knots are like the weights in the ship yard. Each segment between the knots can be modeled with some polynomial, then the polynomials are splined together to create a single smooth curve. One such type of spline curve for representing complicated smooth curves is a B-spline.

B-splines, short for basis splines, are commonly used as basis functions to fit smoothing

curves to large datasets. A B-spline of order p is a piecewise order p Bezier curve and is $C^{(p-1)}$-continuous. B-splines are generally chosen because they are relatively easy to manipulate having local control points, the knots, which only affect a small interval of the entire spline. This ease of manipulation provides an excellent way of approximating a functions behavior over the intervals between the knots. Using a B-spline allows us to create small intervals over the grow season of a crop in order to approximate the behavior of weather elements and the influence they have on crop production. Given a sequence of increasing numbers $\{x_j\}_{j=1}^{N}$, called knots, between an interval $[a,\ b]$, and extend it to the following sequence,

$$x_{-(p)} = \cdots = x_0 = a < x_1 < x_2 < \cdots < x_{N-1} < x_N < b = x_{N+1} = \cdots = x_{N+p+1}.$$

Now define $b_{j,i}(x) = I_{[x_j,\ x_{j+1}]}(x),\ j = 0, 1, 2, \ldots, N$, then B-splines are defined by the Cox-de Boor recursion formula:

$$\begin{cases} b_{j,(p+1)}(x) = \frac{x - x_j}{x_{j+(p+1)-1} - x_j} b_{j,(p+1)-1}(x) + \frac{x_{j+(p+1)} - x}{x_{j+(p+1)} - x_{j+1}} b_{j+1,(p+1)-1}(x), \\ j = -(p), -(p-1), \ldots, -1, 0, 1, 2, \ldots, N-1, N;\ (p+1) > 1., \end{cases}$$

If the knots are equidistant from each other the B-spline is said to be uniform otherwise it is called non-uniform. The knots are vital to the construction of the B-spline, they control their shape, smoothness, and ability to fit the data. Find the correct number of knots is a challenge in its own right. One goal B-splines, is to pick a set of knots such that over parameterization does not occur and that the B-spline accurately fit the data. Another goal of B-splines is to approximate the behavior of the a curve using a linear combination of B-spline basis functions and weights. The second goal will be used in Section 3.3 to approximate the behavior of weather elements precipitation, maximum temperature, and minimum temperature to see how each plays a role in the development of a crop. To utilize the B-splines for this purpose we create a functional linear model.

3.2.2 Functional linear models for scalar responses

According to Ramsay and Silverman (2005), a functional linear model can be formulated in the following ways: the dependent variable Y with argument t is a function of t, one or more

of the independent variables or covariates Z is a function of t, or both Y and Z are functions of t. We will discuss the case when the response Y is scalar and the independent variable(s) Z are a function of t. The main difference from ordinary linear models is that instead of the β_is being a constant value they will now be functions $\beta_i(t)$. Functional data is similar to time series data, but the data are usually collected on a finer time scale such as daily or hourly weather readings. The goal of this type of functional data analysis is to approximate the $\beta_i(t)$ functions which use the explanatory variables $X(t)$to explain the response variable Y (Cardot et al. (1999)). Here instead of obtaining regression coefficients we find regression coefficient functions $\beta_i(t)$ that show the influence the independent variables have on the response (Ramsay and Silverman (2005)).

To start let $\mathbf{y}$ be a vector of responses , $\mathbf{Z}$ is a design matrix, $\boldsymbol{\beta}$ is a vector of coefficients, and $\boldsymbol{\epsilon}$ is the error vector then the multiple general linear model is given by.

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{3.1}$$

Instead of $\mathbf{Z}$ being a typical design matrix, define $\mathbf{Z}$ as a vector of functions with entries $Z_i(t)$. Now the functional extension of (3.1) is where the scalar values y_i will be predicted by functions Z_i. So in the functional form (3.1) becomes

$$y_i = \alpha + \int_0^T Z_i(t)\beta(t)dt + \epsilon. \tag{3.2}$$

Notice that the summation in the matrix product $\mathbf{Z}\boldsymbol{\beta}$ in (3.1) is replaced by the integration over a continuous index t in the interval $[0,\ T]$ of (3.2).

One way to solve this problem is through discretizing the covariate function, however Ramsay and Silverman (2005) state that this is a naive approach and results in "infinitely many sets of solutions". To solve the problems of the naive approach we turn to basis functions. We can expand β in terms of a set of basis functions $B_j(s)$. Let $B_j(t)$ be a vector of basis functions of length J_β, where $J_\beta = N + p + 1$ is the number of knots that have been selected to create the basis, so that

$$\beta(t) = \sum_{j=1}^{J_\beta} b_j B_j(t) \tag{3.3}$$

The number of control points should be chosen so there is no significant loss of information, but small enough so that we can reasonably interpret $\beta(t)$. With the representation of $\beta(t)$ in (3.3) the model (3.2) can now be expressed as

$$y_i = \alpha + \int_0^T Z_i(t)\beta(t)dt + \varepsilon_i = \alpha + \sum_{j=1}^{J_\beta} \int_0^T Z_i(t) b_j B_j(t) dt + \varepsilon_i. \tag{3.4}$$

Now we write (3.4) in matrix form, define matrix $\mathbf{W}$ to be a $n \times J_\beta$ matrix with entires,

$$w_{i,j} = \int_0^\infty Z_i(t) B_j(t) dt, \ 1 \leq i \leq n; 1 \leq j \leq J_\beta, \tag{3.5}$$

We can further simplify this notation by defining the $(J_\beta + 1) \times 1$ vector $\boldsymbol{\theta} = (\alpha, \mathbf{b}')' = (\alpha, b_1, \ldots, b_{J_\beta})'$ and define the coefficient matrix $\mathbf{Z}$ to be the $N \times (J_\beta + 1)$ matrix $\mathbf{Z} = [\mathbf{1}\ \mathbf{W}]$, and $\mathbf{1}$ is a $N \times 1$ vector of ones. Then (3.4) becomes

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{3.6}$$

and the least squares estimate of the parameter vector $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \tag{3.7}$$

Now the problem is just a standard multiple regression problem.

Now the are many choices of the basis functions Ramsay and Silverman (2005) gives an example of predicting total annual precipitation for Canadian weather stations with daily temperature observations from 35 different weather stations using Fourier basis functions. Ramsay and Silverman (2005) also mention that B-spline basis functions could be used as well. Cardot et al. (2003) show how B-spline basis functions can be used and also introduce a penalized version of the B-spline to account for the roughness of the spline. The penalty involves a adding a constant determined by the derivatives of the B-spline basis functions. Usually the squared first derivatives are used. Note here that we have yet to consider when the responses have some type of dependence structure. Giraldo et al. (2009) showed that space time kriging was possible when the responses are functional in both space and

time. They do this by using classical multivariable geostatistics techniques and extending them to the functional data context. Horváth and Kokoszka (2012) give an in depth review of spatially distributed functional data when the response is functional. They develop functional spatial-temporal covariance functions as well as techniques for estimating them using basis functions. In the cases of Giraldo et al. (2009) and Horváth and Kokoszka (2012) both the response variable and dependent variables have the same time scale. The difference here in this work is the time scale for the yield is in years and the weather data is measured in days. This leads us to propose a similar model to that of (3.2) but with a spatial twist resulting in a weighted least square procedure. This weighted least square procedure will allow for the scalar responses to have a spatial/spatio-temporal dependence structure over years. The introduction of this new functional linear model is driven the desire to explain weather impact on crop yield in the state of Kansas. In Section 3.3 we will perform a data analysis on how weather impacts Kansas corn yield using this modified functional linear model.

3.2.3 The spatial functional linear model

Our proposed model for predicting Kansas corn yield based on multiple weather elements has the from

$$y_{k,i} = \mu_k + \sum_{w=1}^{3} \int_0^T W_{k,i,w}(t)\beta_{w_i}(t)\ dt + \varepsilon_{k,i}. \tag{3.8}$$

Here y_{ki} is the scalar response bushels per acre for the kth year ($k = 1, \ldots, K$) and ith county ($i = 1, \ldots, n$). Each weather component is represented by $W_{k,i,w}(t)$ and $\beta_w(t)$, for example, $W_{k,i,1}(t)$ is the precipitation process for year k and county i. The function $\beta_1(t)$ is the weight function for precipitation indicating at what times and to what degree precipitation influences corn production. Just to be clear k indicates year, i is the index for county, $t \in [0, T]$ is a time point within the growing season and is common for k and i. The μ_k is spatial yield trend per year over the whole state and $\varepsilon_{k,i}$ is the yearly spatial error term over counties that will later be modeled using spatial statistic techniques.

Note that each $\beta_w(t)$ is a function of time that will describe when and how important each weather element is important to the prediction of corn yield. Just as in the review of the functional linear model the goal is to estimate the $\beta_w(t)$s. In order to estimate each $\beta_w(t)$ we use B-spline basis functions for each weather element with order p given by $B_{w,q}(t),\ w = 1,2,3,\ q = 1,2,\ldots,N^*$ so that $\beta_w(t) = \sum_{q=1}^{N^*} b_{w,q}B_{w,q}(t)$ is the smoothing approximation so that

$$y_{k,i} = \mu_k + \sum_{w=1}^{3}\sum_{q=1}^{N^*} b_{w,q}\int_0^T W_{k,i,w,}(t)B_{w,q}(t)\ dt. \tag{3.9}$$

For our proposed weighted least squares procedure to solve for $\boldsymbol{\beta}(t) = [\beta_1(t),\ \beta_2(t),\ \beta_3(t)]$ we do the following

$$\hat{\beta}_w(t) = \underset{b_{w,q}}{\arg\min} \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(y_{k,i} - \mu_k - \sum_{w=1}^{3}\sum_{q=1}^{N^*} b_{w,q}\int W_{k,i,w}(t)B_{w,q}(t)\right)\right. \tag{3.10}$$
$$\left.\times\sigma^{-(ijk)} \times \left(y_{k,j} - \mu_k - \sum_{w=1}^{3}\sum_{q=1}^{N^*} b_{w,q}\int W_{k,i,w}(t)B_{w,q}(t)\right)\right],$$

where $\sigma^{-(ijk)}$ is the $(i,\ j)$th element of Σ_k^{-1}, the spatial covariance matrix for yearly yield among counties.

Similar to the discussion in Section 3.2.2 we can represent (3.9) and (3.10) in matrix form. Let $\mathbf{Y} = [Y_1', Y_2' \ldots, Y_k']'$ is a vector of length $kn \times 1$ and each component $Y_k = [y_{k,1}, \ldots, y_{k,n}]'$, here $y_{k,i}$ represents the yearly yield observation from year k and county i. The vector $\boldsymbol{\mu} = [\mathbf{1}_n'\mu_1, \mathbf{1}_n'\mu_2, \ldots, \mathbf{1}_n'\mu_k]'$ is $kn \times 1$ vector with each element μ_k is the average crop yield over all counties for the kth year and $\mathbf{1}_n$ is a $n \times 1$ vector of ones. Define the matrix $\mathbf{W}_w, w = 1,2,3$ as a $kn \times N^*$, $N^* = N + p + 1$, matrix that consists of the integrations over the B-spline basis functions and data components maximum temperature, minimum temperature, or precipitation. $N^* = N + p + 1$ is the number of total knots for order $p = 3$

B-spline basis functions with N interior knots so that

$$\mathbf{W}_w = \left(\begin{array}{cccc} \int W_{1,1,w}(t)B_1(t)dt & \int W_{1,1,w}(t)B_2(t)dt & \dots & \int W_{1,1,w}(t)B_{N^*}(t)dt \\ \int W_{1,2,w}(t)B_1(t)dt & \int W_{1,2,w}(t)B_2(t)dt & \dots & \int W_{1,2,w}(t)B_{N^*}(t)dt \\ \vdots & \vdots & \vdots & \vdots \\ \int W_{1,n,w}(t)B_1(t)dt & \int W_{1,n,w}(t)B_2(t)dt & \dots & \int W_{1,n,w}(t)B_{N^*}(t)dt \\ \hline \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hline \int W_{k,1,w}(t)B_1(t)dt & \int W_{k,1,w}(t)B_2(t)dt & \dots & \int W_{k,1,w}(t)B_{N^*}(t)dt \\ \int W_{k,2,w}(t)B_1(t)dt & \int W_{k,2,w}(t)B_2(t)dt & \dots & \int W_{k,2,w}(t)B_{N^*}(t)dt \\ \vdots & \vdots & \vdots & \vdots \\ \int W_{k,n,w}(t)B_1(t)dt & \int W_{k,n,w}(t)B_2(t)dt & \dots & \int W_{k,n,w}(t)B_{N^*}(t)dt \end{array} \right) \quad w = 1, 2, 3$$

Here $W_{k,i,w}(t)$ are the functions of maximum temperature, minimum temperature, or precipitation ate each year and county. B_J is the Jth b-spline $J = 1, 2, \dots, N^*$ with integration being over the time interval from $[0,\ T]$. Since we have three weather elements there will be three $\mathbf{W}$ matrices which can be represented as $[\mathbf{W}_1\ \mathbf{W}_2\ \mathbf{W}_3]$. Now the weight vector for the basis functions with elements $b_{w,J},\ J = 1, \dots . N^*$ will also be in three pieces one for each weather element defining $\mathbf{b} = [b_1'\ b_2'\ b_3']'$, where $b_w = [b_{w,1}, \dots, b_{w,N^*}]',\ w = 1, 2, 3$. So now let $\mathbf{Z} = [\mathbf{M}\ \mathbf{W}_1\ \mathbf{W}_2\ \mathbf{W}_3]$ where $\mathbf{M} = \mathbf{I} \otimes \mathbf{1}_n$, $\mathbf{I}$ is a $K \times K$ identity matrix and $\mathbf{1}_n$ is a $n \times 1$ vector of ones. Also define the parameter vector $\mathbf{\Omega} = [\boldsymbol{\mu}', \mathbf{b}']'$, where $\mathbf{b} = [b_1', b_2', b_3']'$.

Allowing the spatial covariance to change from year to year we get a block diagonal $kn \times kn$ matrix defined as

$$\mathbf{\Sigma}_* = \begin{pmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_K \end{pmatrix}.$$

and Σ_k is the $n \times n$ spatial covariance matrix for yield across counties for the kth year. Notice that the off diagonals are zero meaning there is no yearly space time covariance, but if you data where better modeled by a more complex covariance structure the only change would be the off diagonals.

Combining all the notation we obtain the matrix form of (3.9) is

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Omega}$$

and a weighted least squares estimate function of (3.10)

$$\hat{\boldsymbol{\Omega}} = \left[\mathbf{Z}'\boldsymbol{\Sigma}_*^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'\boldsymbol{\Sigma}_*^{-1}\mathbf{Y}.$$

By solving for Ω we obtain estimates for the yearly mean yield over the state of Kansas and obtain three functional parameters one for each weather element. Looking a the graphs of these functional parameters we can determine if a weather element as a positive or negative effect on crop yield and how strong it is. Taking the first derivative of these functions determines what time points are vital to the growth of the crop in terms of each weather element. We now fit our model and discuss the findings.

3.3 Kansas corn and weather data analysis

As mentioned in the introduction there are $n = 105$ counties in Kansas on which we collect data over $K = 22$ years from 1990-2011. The growing season for Kansas is from April to September, however, to avoid complications of having boundary knots during the growing season we add March and October to allow the B-spline to have a more freedom at the end points of the growing season. This totals to $T = 245$ days, 183 days which make up the corn growing season, so the time interval for the B-spline basis functions is $[1, 245]$. The B-spline basis functions were of order $p = 3$ and a total of $N^* = 19$ knots, four boundary knots and $N = 15$ interior knots were equally spaced over the growing season creating a two week interval between the knots. This setup captured the behavior of $\beta_w(t)$s the best. To fit the model and estimate our $\beta(t)$ functions requires a two step process; first fitting the mean structure then analyzing the residuals to develop the covariance structure then refitting it with the yearly spatial covariance structure. After the initial fit, we noticed that the residuals have an approximate normal distribution centered at zero which allowed us to

use Gaussian random field theory to fit the covariance structure of the residuals. To check for isotopy we looked at directional variograms and noticed that all directions have the same pattern. We also looked at the space-time empirical correlations and noticed that from year to year corn yield is linearly uncorrelated. Since the weather does not have year to year correlation this effects the corn yield by only having spatial correlation. Using the yearly empirical variograms we fit an exponential covariance function and parameter estimation for the range, scale, and nugget was done by maximum likelihood. Figure 3.4 shows a sample of the empirical and fitted variograms from the first nine years. Now that we have the

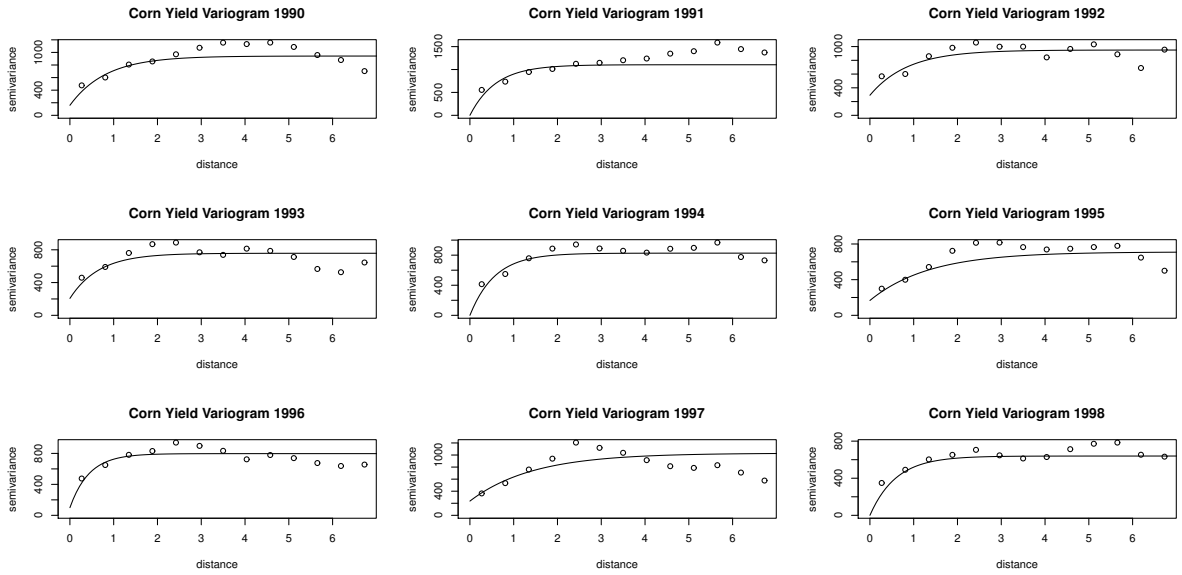

Figure 3.4: Variograms for the first nine years. Distance scale in 100 kilometers

estimate for our $\mathbf{\Sigma}_*$ we refit the model, and the estimation of $\mathbf{\Omega}$ in (3.2.3) can be done to find estimates for μ_ks and functional parameters β_ws.

Figure 3.5 shows the estimate $\hat{\beta}_1(t)$ and its derivative for precipitation. When $\frac{d}{dt}\hat{\beta}(t) = 0$ for some t we obtain an important time point in the growth cycle of corn when precipitation has either a positive or negative influence. Recall that the important dates determined by Staggenborg et al. (2008) are June 20th to July 10th when using complex computer simulations. Based on the raw data during the early part of April rain is vital since ground moisture is needed for the seed to sprout. Next, notice the declining behavior of precipitation

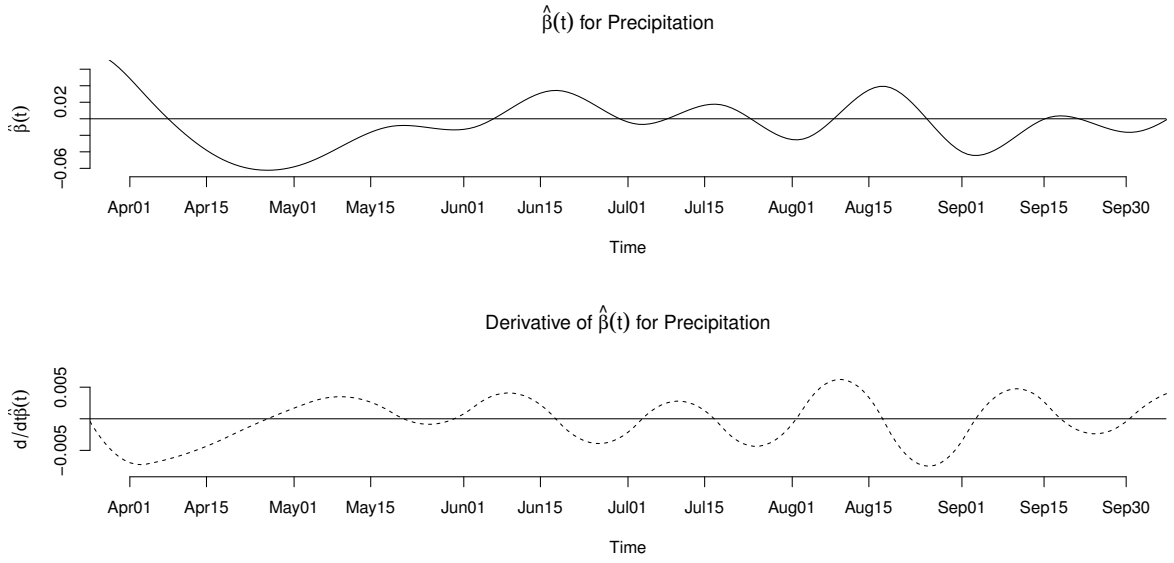

Figure 3.5: Precipitation $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$

producing a negative effect. After the seed sprouts if there is too much moisture, the plant has the possibility of drowning resulting in a replant and a shorter growing season. This shorter growing season can prevent the crop from reaching the maximum yield potential. Around April 26th precipitation begins to increase maxing out around June 18th and another positive spike around July 16th. Notice that these days coincide with the days set forth by Staggenborg et al. (2008), meaning that at the beginning and end of the vital period June 20th to July 10th precipitation has a major positive influence on corn yield. There is one last major positive effect coming in the middle of August which is mostly for cooling the plants from the summer heat. Precipitation towards the end of the season has a slight negative effect because farmers need the corn to dry in order to harvest and moisture prevents this process.

Figure 3.6 shows the estimate $\hat{\beta}_2(t)$ and its derivative for maximum temperature. From the plots for maximum daily temperature notice hot days in May are vital because a freeze would cause the crop to die. Again, the agronomy time frame holds true with the raw data in that days surrounding June 26th-27th maximum temperature has a big positive influence on corn yield. Also note the sharp decline to a negative effect at the very end of

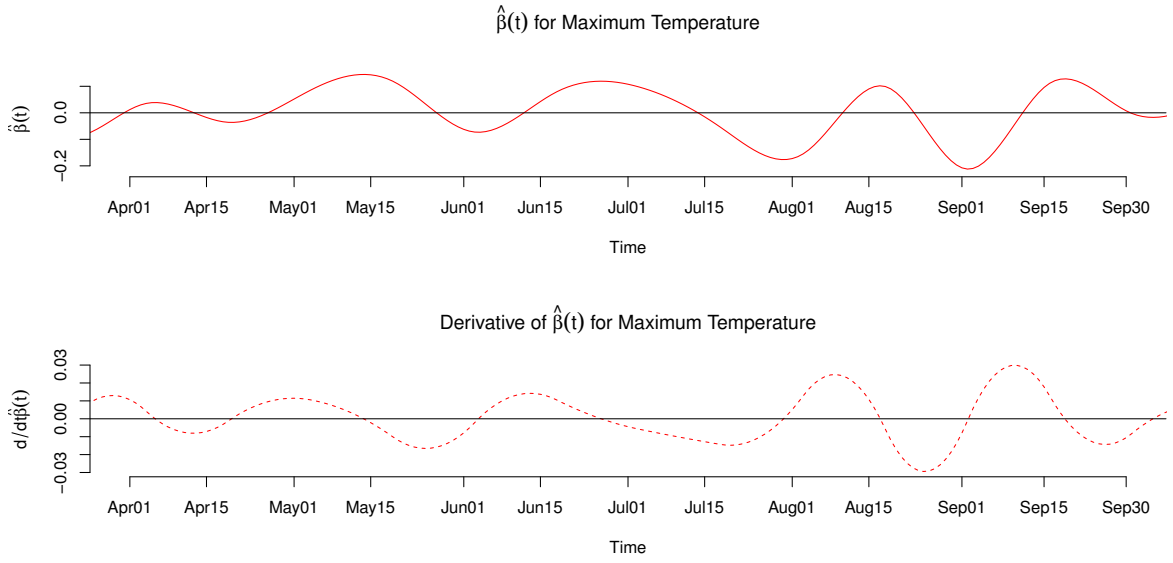

Figure 3.6: Maximum Temperature $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$

July beginning of August on through to September. This is showing how the hot summer months, especially without precipitation, can cause a drought which can damage or destroy the corn plants deceasing yield. At the end of the growing season there is one last positive effect for drying the corn around September 20th.

Figure 3.7 shows the estimate $\hat{\beta}_2(t)$ and its derivative for minimum temperature. Through the beginning of a crops life freezing can in most cases cause serious damage to yield. Notice that minimum temperature has a mostly negative effect from April 1st to May 6th this is because the young plant can drastically be effected by freezing conditions. During the in vital June 20th to July 10th time period minimum temperature does not have major effect on yield, but going into the hot summer months of August and into September cool weather has a positive effect on yield. Around August 3rd to September 4th having cool weather to battle against the hot days is very beneficial for a good corn crop at the end of the season.

Looking at the big picture weather plays a vital role for good crop yield. To sum up what weather conditions produce a bunker crop; in the beginning of the season a farmer would want warm days and nights with a little rain, during the crucial June 20th - July 10th hot days with good amounts of rain, and at the end of the season some rain and cool conditions

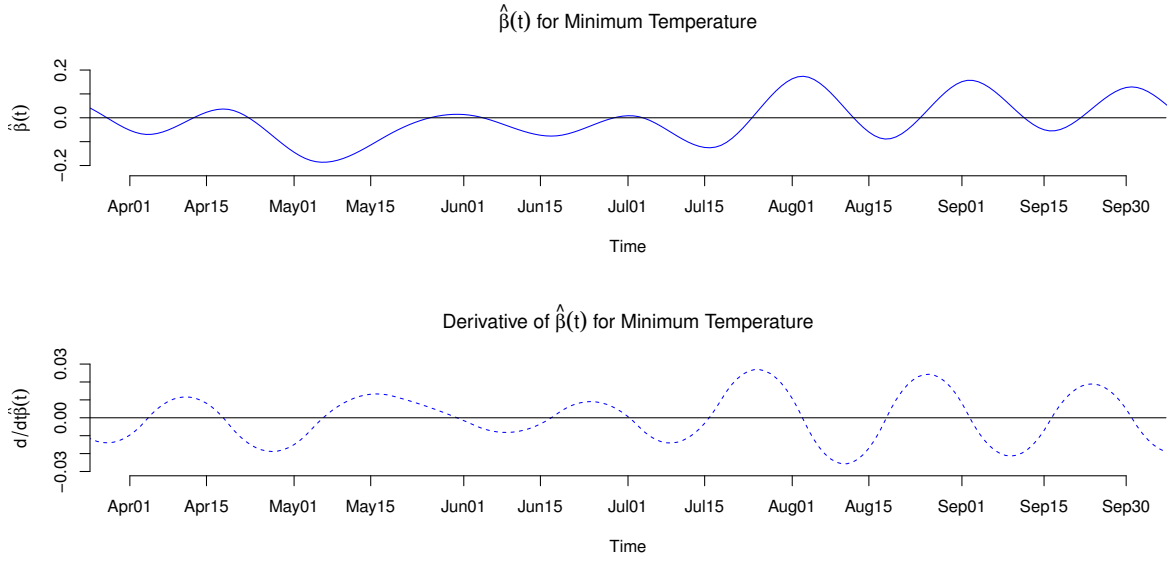

Figure 3.7: Minimum Temperature $\hat{\beta}(t)$ and $\frac{d}{dt}\hat{\beta}(t)$

to handle the summer heat of August and September. Overall the proposed model fits the data nicely to give agronomist, farmers, and statisticians a way to analyze raw weather to predict yearly yield outcomes. Weather does not change much year to year, but what about over a ten or twenty year time period? Next we investigate how corn yield is affected by ten year weather patterns.

3.4 Ten year weather patterns

Through the analyses of Chaper 1 and 3 we notice that weather patterns are almost independent from year to year, but there may be something changing over larger periods of time. To look at this relationship we will look at how agriculture districts change over a ten year period with respect to yield and each weather element using functional boxplots presented by Sun and Genton (2011). Figure 3.8 shows the agriculture districts of Kansas. Kansas is almost perfectly divided from the northwest corner to the southeast corner with the west side having more crops than the east side. We will use the functional boxplot over two areas. The heavily farmed or rural area will be made up by the agriculture districts;

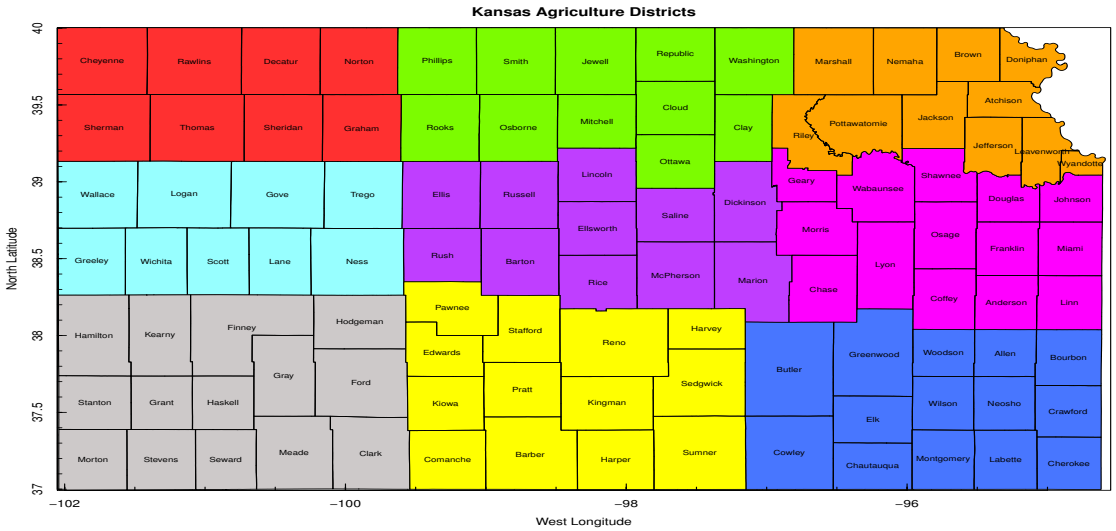

Figure 3.8: Kansas Agriculture Districts: Red-Northwest, Green-North Central, Orange-Northeast, Cyan-East Central, Purple-Central, Pink-West Central, Gray-Southwest, Yellow-South Central, Blue-Southeast.

Northwest, West Central, Southwest, South Central. The other five districts will make up the urban area or light farming. Figure 3.9 this split of the rural and urban areas of Kansas as described above. The functional data plots that will be compared are from 1990-2000

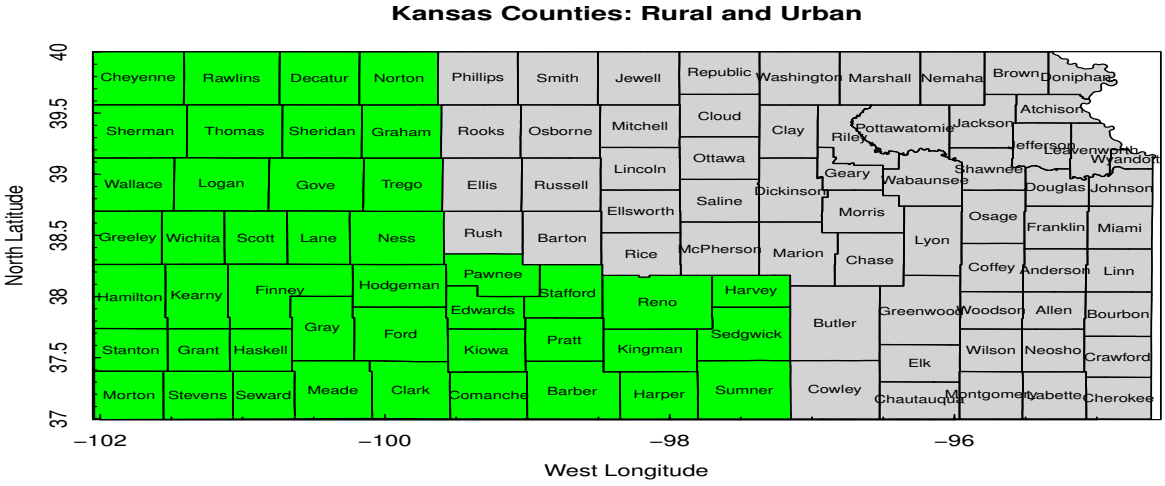

Figure 3.9: Kansas Rural and Urban Areas.

and 2001-2011. Here we are looking for change over ten years in crop yield, precipitation,

maximum temperature, and minimum temperature.

When looking at the following functional boxplots the upper line represents the maximum. The colored band is the 50% central region measured outward from the median curve given by the middle black line. The lowest blue band is the minimum. For the 10 year change in yield for both rural and urban areas the overall yield has a very small increase over a ten year period, see Figures 3.10 and 3.11. However, notice that in the years 2001 to 2011 the functional boxplot in Figures 3.10 has a colored band that is slightly wider for the rural area than that of the rural area in years 1990 to 2000. There could be a number of reasons for this increase in variability such as weather changes, expansion of urban areas, or data collection is inaccurate.

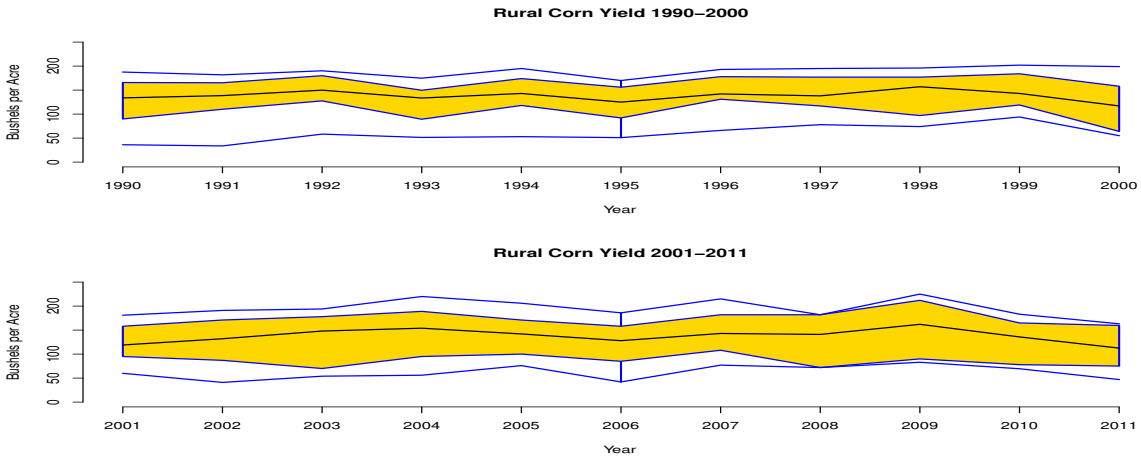

Figure 3.10: Rural Functional Boxplot of Yield; Top: years 1990-2000, Bottom: years 2001-2011

Figures 3.12 and 3.13 show the ten year change functional boxplots for precipitation. Recall from Figure 3.5 that at the middle of June and August are very important times for a corn plant in terms of precipitation. In Figure 3.12 notice that the years 2001 to 2011 have more precipitation during these critical times which could explain the wider quartile band in Figure 3.10. The Urban precipitation during 2001 to 2011 has a loss of precipitation around August 15th which can explain the decline in yield for the Urban area for those years. Overall in the rural area the precipitation pattern did not change over the ten year

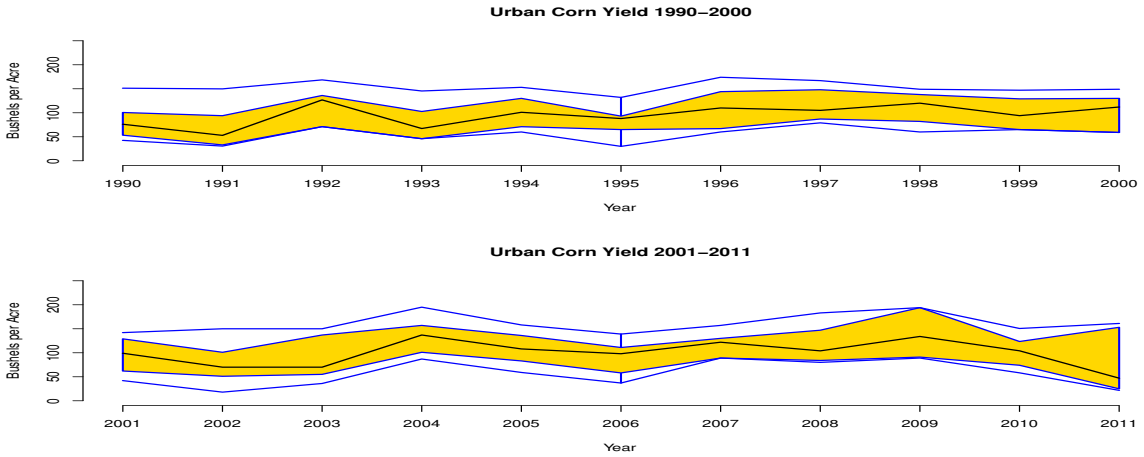

Figure 3.11: Urban Functional Boxplot of Yield; Top: years 1990-2000, Bottom: years 2001-2011

gap, however, in the urban area precipitation has declined during the August 15th time frame. This would explain the sharp decline of crop yield in the the urban area during the years 2009-2001.

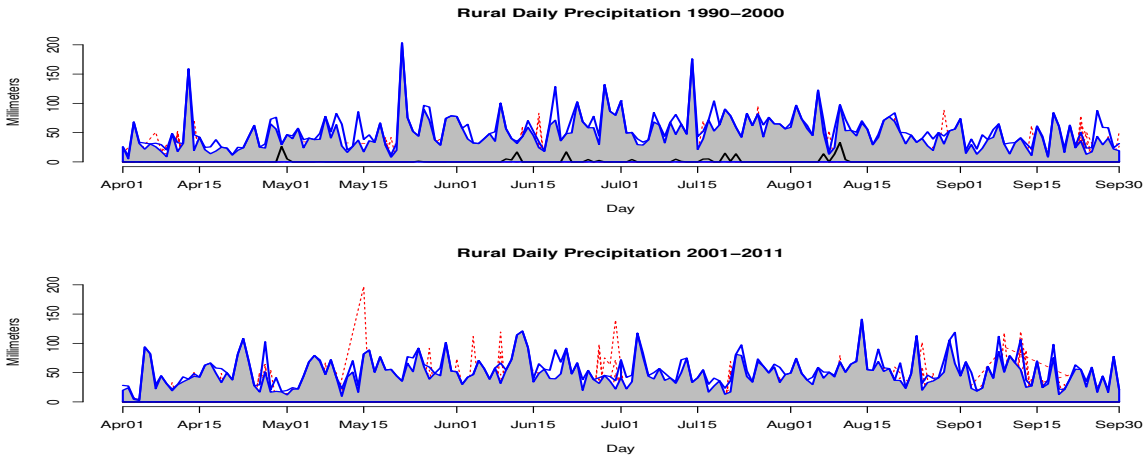

Figure 3.12: Rural Functional Boxplot of Precipitation; Top: years 1990-2000, Bottom: years 2001-2011

Figures 3.14, 3.15, 3.16, and 3.17, show the change over 10 years for maximum and minimum temperatures. In all the cases there is not much change, but in all cases note that in the later 2001 to 2011 years the quartile band is tighter. With the addition of newer technology to collect data and more stations to collect data from, this should be expected

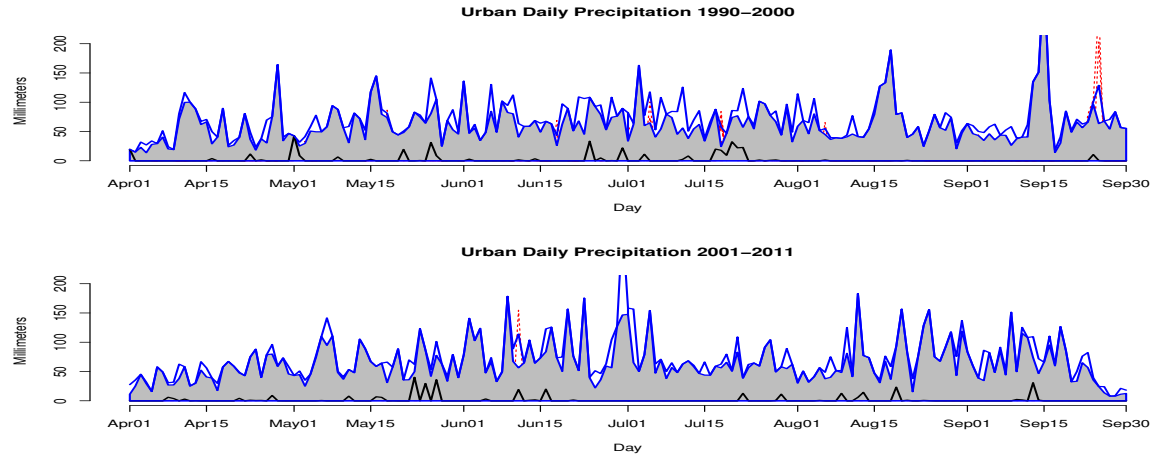

Figure 3.13: Urban Functional Boxplot of Precipitation; Top: years 1990-2000, Bottom: years 2001-2011

signifying more accurate data collection.

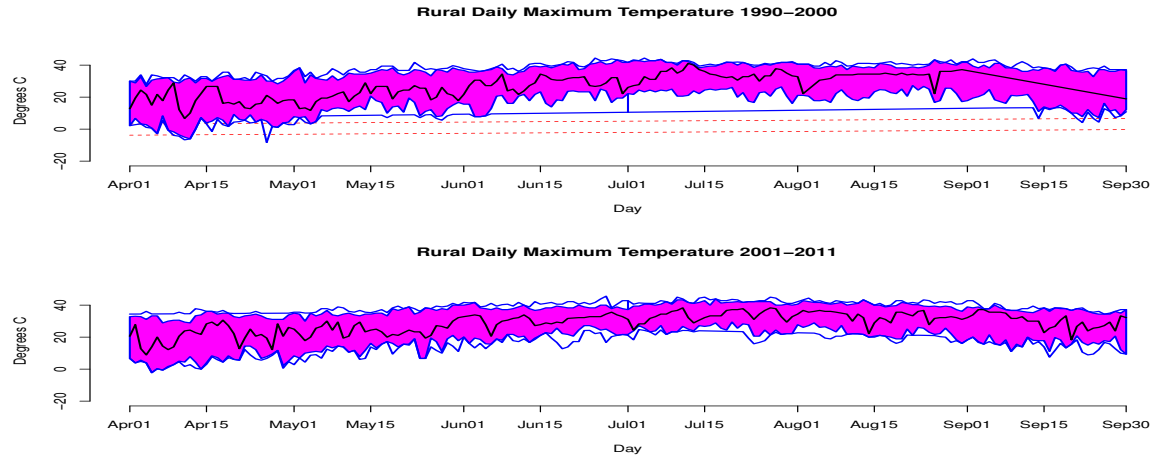

Figure 3.14: Rural Functional Boxplot of Maximum Temperature; Top: years 1990-2000, Bottom: years 2001-2011

Taking a look spatially in Figure 3.18 we plot a three dimensional functional boxplot over the state of Kansas. Again we see the same conclusions, the southwest portion of Kansas maintains a higher corn yield than the rest of the state and the urban areas have a lower yield. Figure 3.18 shows that the minimum and maximum corn yield surface do not really change over a ten year time span. The interquartile surfaces do have slight changes with yield increasing in the northeast and southwest parts of Kansas. Examining Figure

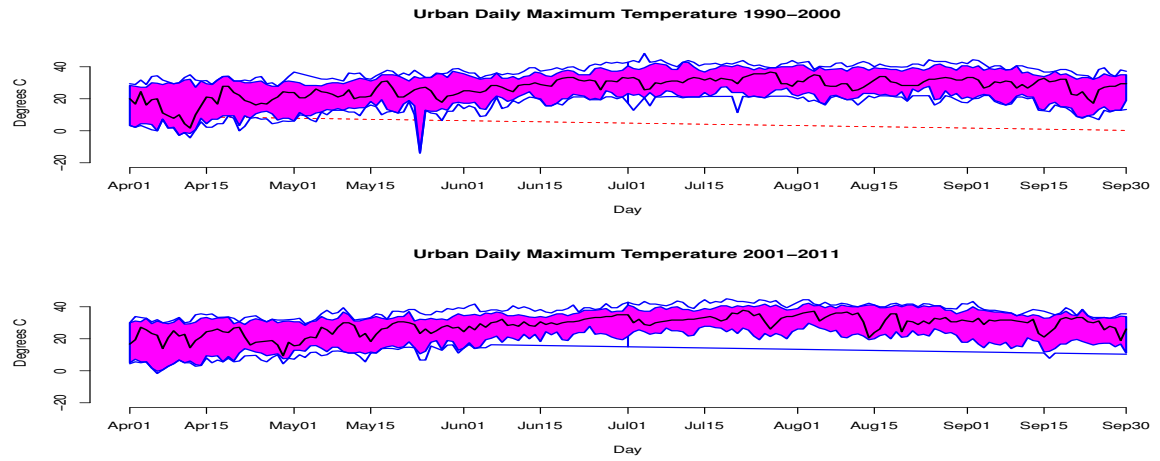

Figure 3.15: Urban Functional Boxplot of Maximum Temperature; Top: years 1990-2000, Bottom: years 2001-2011

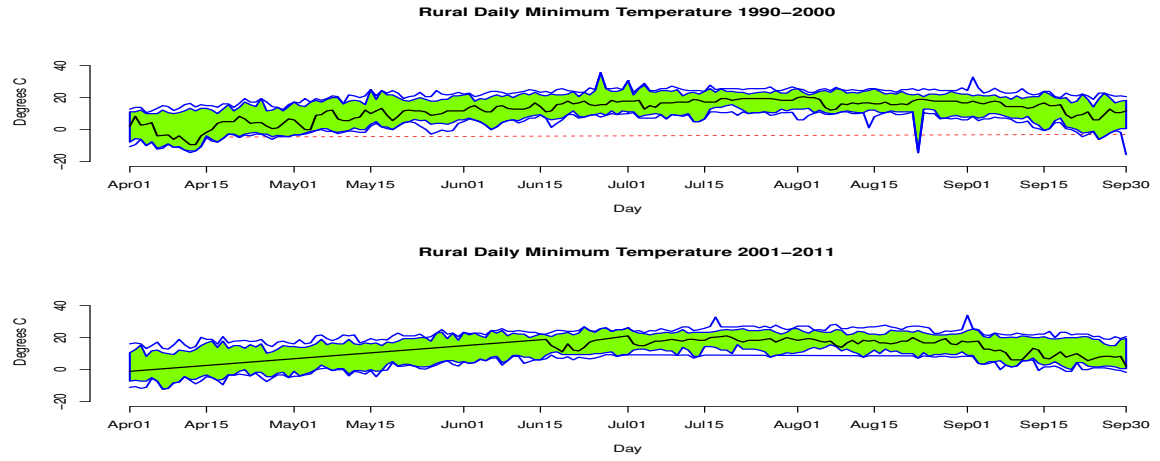

Figure 3.16: Rural Functional Boxplot of Minimum Temperature; Top: years 1990-2000, Bottom: years 2001-2011

3.19 clearly there is a lot of activity going on with precipitation during the ten year time span. Here the total precipitation was taken over the time period from June 15th to June 21st which is the first critical time span for corn plants to obtain moisture. The maximum surface in Figure 3.19 shows drastic changes for different areas of Kansas. For example, during 1990 to 2000 the southwest corner of Kansas did not have near the maximum rainfall as it did in the 2001 to 2011. This explains the raise of the surfaces in that area in Figure 3.18. On the negative side looking at the southeast part of the surfaces notice the decline

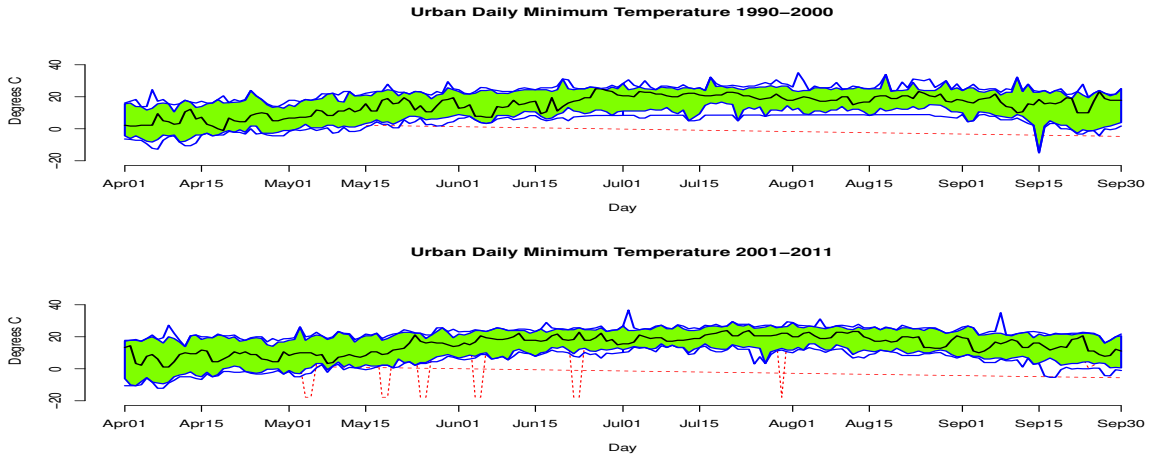

Figure 3.17: Urban Functional Boxplot of Minimum Temperature; Top: years 1990-2000, Bottom: years 2001-2011

in precipitation over the ten year gap. This coincides with the low surface in that area of Figure 3.18. Maximum and minimum surface plots were also examined, but they did not show any change over the ten year time span. Overall it is clear that changes in weather patterns over a ten year time span can have a huge impact on crop production.

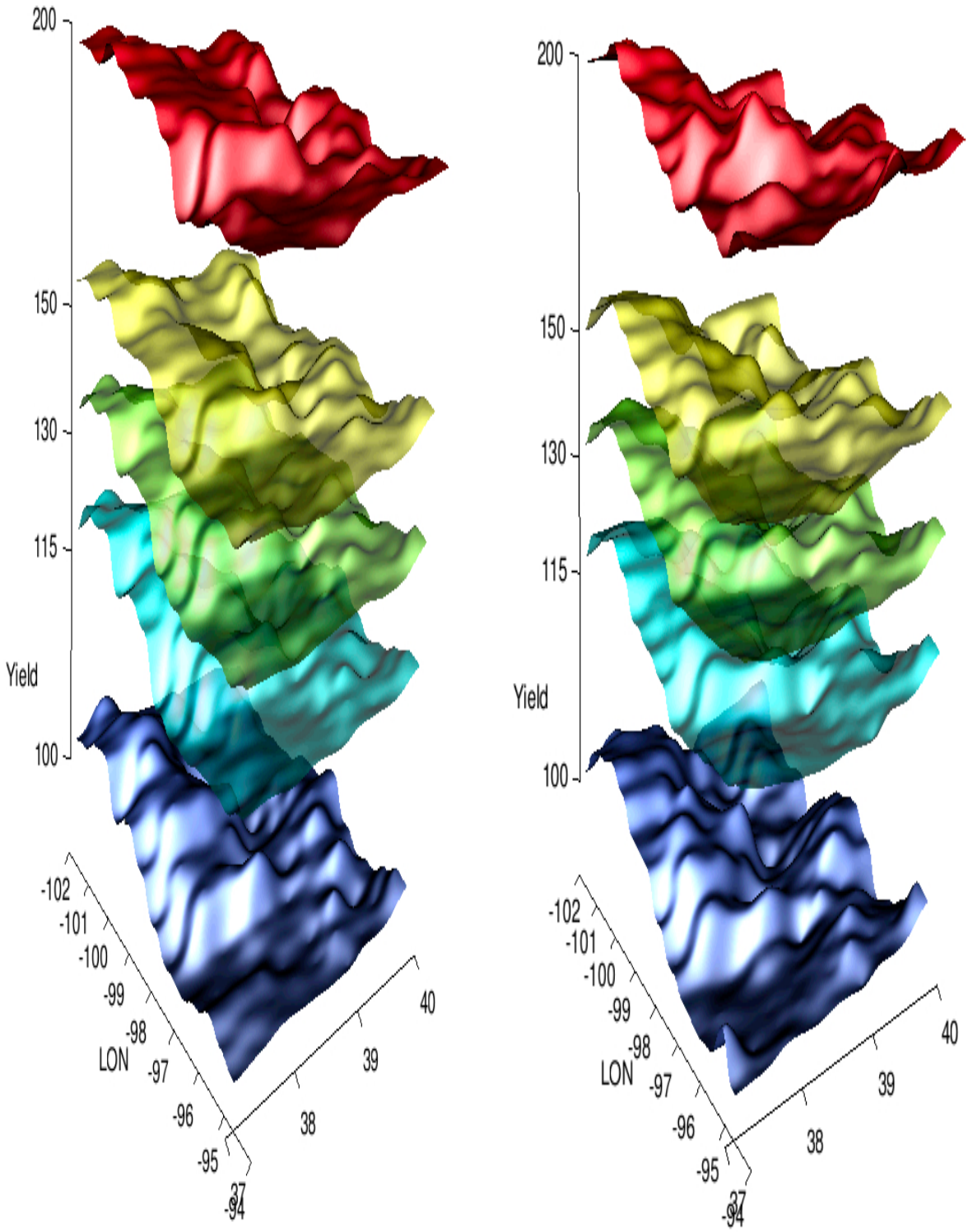

Figure 3.18: 3D Functional Surface Boxplot of Yield. Left: years 1990-2000, Right: years 2001-2011

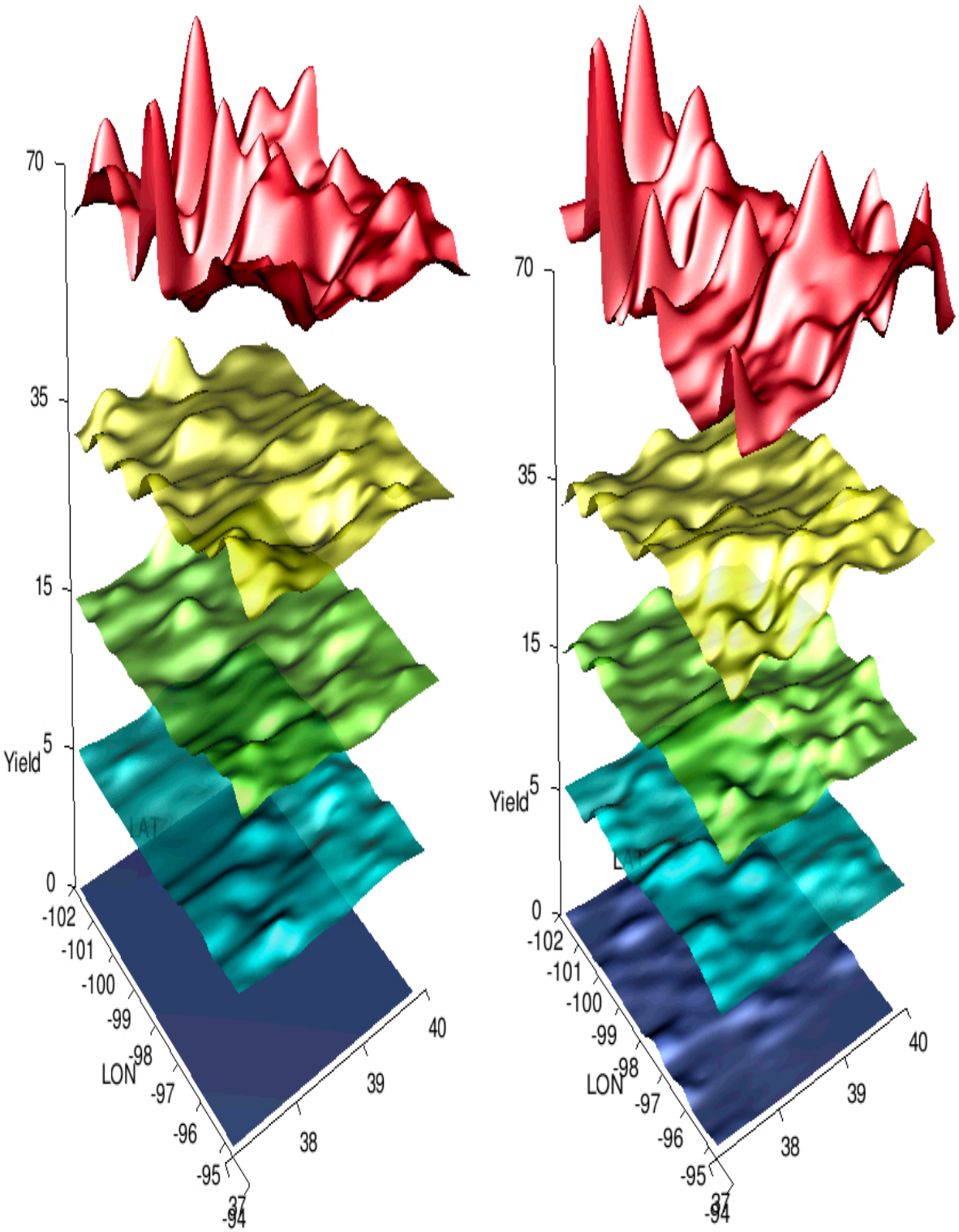

Figure 3.19: 3D Functional Surface Boxplot of Precipitaton June 15th to June 21st. Left: years 1990-2000, Right: years 2001-2011

Conclusion and Future Work

Throughout this work there has been a central theme of explaining correlation or correlated structure in space/space-time processes. In Chapter 1 we found that it is beneficial to consider the discreteness of the time points in spatio-temporal modeling in terms of ease of model identification and plausibility. It is seen that the proposed covariance functions generally out perform Gneiting's separable and non separable covariance functions in the Kanas Precipitation data analysis and competitive with those methods in the Irish Wind data analysis. However, in the asymmetric case Gneiting's model provides a slightly better fit, but the proposed asymmetric model is still very comparable. Applying the asymmetric framework to the proposed model causes the temporal margin to be no longer autoregressive in nature. To fix this it could be possible to choose a slightly different Lagrangian function to maintain the autoregression structure. Another improvement of both the models proposed and Gneiting's models is the ability to use maximum likelihood estimation. The problem here is trying to maximize over a large parameter vector, but efficiency of MLE might be beneficial in final model estimation. One solution would be to use profile likelihoods to estimate certain parameters so that the overall parameter space dimension becomes smaller. It would also be relevant to extend the proposed models (1.8) and (1.22) to include more types of time series margins and the constrains for each. The spatio-temporal models presented in Chapter 1 have successfully started to provide an effective approach that can be used in an intuitive manner as well as easier to apply to space-time data with some discrete temporal margins.

In Chapter 2 we introduced some multivariate covariance matrix functions and extended tapering techniques to the multivariate case. Several classes of compactly supported correlation matrix functions are developed to play as tapering matrix functions, especially the multivariate Wendland type tapering matrix functions with desired flexibility of smoothness. We established a guideline for choosing a tapering function based on the original covariance

function. Simulation study shows that as the number of locations in a space increases the MSPE of the tapered covariance tends to zero with a great deal of time savings in inverting the large covariance matrix. Future work would consist of establishing theoretical conditions on the tapering matrix function to assure there is no loss of optimality of prediction and efficiency of estimation.

Chapter 3 introduced the spatial functional linear model, which was used to examine how weather variables affect Kansas corn yield. Overall the model performed well giving insight into what important time points are within the growing season and how weather has either a positive or negative impact on corn yield. In the end we were successful in creating a monitor-like system that can predict the yield given daily weather observations. There are many extensions that can be studied in the future. First our study focuses on the state of Kansas only, if national information is concerned, it would be interesting to explore large scale influence of weather change on yield. Second, we only studied the corn yield in this work, while multiple crops are harvested at each state, such as sorghum, bean, etc. Then multivariate spatial function regression can be considered, where we could also take advantage of the multivariate covariance functions and tapering techniques introduced in Chapter 2. Third, for the data analysis in this chapter we could also let the model include over covariates such as soil type, humidity, and other variables that play a significant role in the production of crop yield. Fourth, not apparent in the Kansas application, but in other applications the Σ_* matrix could have a spatio-temporal structure. For example, if the independent variables had a scale of minutes and the response variable had a scale in hours, there might be space-time correlation in the response variable, which could be modeled with the models of Chapter 1. Furthermore, it would be beneficial to construct a functional confidence intervals for the parameter functions $\beta_w(t)$. Research on the asymptotics of the spatial functional linear model would also be a worthwhile venture in future work. With functional data analysis being a new technique for spatial statistics there are many more possible avenues to be exploited.

Bibliography

Askey, R. (1973), "Radial characteristic functions. Tech. Report No. 1262," Tech. rep., Math. Research Center, University of Wisconsin-Madison.

Bates, D. and Maechler, M. (2011), *Matrix: Sparse and Dense Matrix Classes and Methods*, r package version 0.999375-50.

Cambanis, S., Huang, S., and Simons, G. (1981), "On the theory of elliptically contoured distributions." *Journal of Multivariate Analysis*, 11, 368–385.

Cardot, H., Ferraty, F., and Sarda, P. (1999), "Functional Linear Model." *Statistics and Probability Letters*, 45, 11–22.

— (2003), "Spline estimators for the functional linear model." *Statistica Sinica*, 13, 571–591.

Cramár, H. (1940), "On the theory of stationary random processes." *Annals of Mathematics*, 41, 215–230.

Cressie, N. and Huang, H. C. (1999), "Classes of non-separable, spatio-temporal stationary covariance functions." *Journal of the American Statistical Association*, 94, 1330–1340.

Davis, T. A. (2006), *Direct Methods for Spare Linear Systems.*, Philadelphia: Society for Industrial and Applied Mathematics.

de Luna, X. and Genton, M. G. (2002), "Predictive spatio-temporal models for spatially sparse environmental data." *Statistica Sinica*, 15, 547–568.

Demel, S. S. and Du, J. (2011), "Spatio-temporal covariance modeling with some ARMA temporal margins," in *Conference for Applied Statistics in Agriculture Proceedings*, vol. 23.

Du, J., Leonenko, N., Ma, C., and Shu, H. (2012), "Hyperbolic vector random fields with hyperbolic direct and cross covariance functions." *Stochastic Analysis and Applications*, 30, 662–674.

Du, J. and Ma, C. (2011), "Spherically invariant vector random fields in space and time." *IEEE Trans. Signal Proc.*, 59, 5921–5929.

— (2012), "Variogram matrix functions for vector random fields with second-order increments." *Mathematical Geoscience*, 44, 411–425.

Du, J., Zhang, H., and Mandrekar, V. S. (2009), "Fixed-domain asymptotic properties of tapered maximum likelihood estimators." *Annals of Statistics*, 37, 3330–3361.

Fuentes, M. (2007), "Approximate likelihood for large irregularly spaced spatial data." *Journal of the American Statistical Association*, 102, 321–331.

Furrer, R. and Genton, M. (2011), "Aggregation-cokriging tapering for interpolation of large spatial datasets." *Biometrika*, 98, 615–631.

Furrer, R., Genton, M. G., and Nychka, D. (2006), "Covariance tapering for interpolation of large spatial datasets." *Journal of Computational and Graphical Statistics*, 15, 502–523.

Gilbert, J. R., Moler, C., and Schreiber, R. (1992), "Sparse matrices in MATLAB: Design and implementation." *SIAM Journal on Matrix Analysis and Applications*, 15, 333–356.

Giraldo, R., Delicado, P., and Mateu, J. (2009), "Continuous time-varying kriging for spatial prediction of functional data: An environmental application," *Journal of Agricultural Biological and Environmental Statistics*, 15, 66–82.

Gneiting, T. (2002a), "Compactly supported correlation functions." *Journal of Multivariate Analysis*, 83, 493–508.

— (2002b), "Nonsepable, stationary covariance functions for space-time data." *Journal of the American Statistical Association*, 97, 590–600.

Gneiting, T., Genton, M., and Guttorp, P. (2007), *Statistical Methods for Spatio-Temporal Systems.*, Boca Raton: Chapman and Hall/CRC, chap. Geostatistical space-time models, stationarity, separability, and full symmetry., pp. 151–174.

Gneiting, T., Kleiber, W., and Schlather, M. (2010), "Matéern cross-covariance functions for multivariate random fields." *Journal of the American Statistical Association*, 105, 1167–1177.

Gradshteyn, I. S. and Ryzhik, I. M. (2000), *Table of Integrals, Series, and Products*, Academic Press, San Diego, 6th ed.

Haslett, J. and Raftery, A. (1989), "Space-Time Modeling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37, 1–50.

Horváth, L. and Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer New York, chap. Spatially distributed functional data., pp. 343–374.

Kaufman, C., Schervish, M., and Nychka, D. (2008), "Covariance tapering for likelihood-based estimation in large spatial datasets." *Journal of American Statistical Association*, 103, 1545–1555.

Le, N. D. and Zidek, J. V. (2006), *Statistical Analysis of Environmental Space-time Processes*, New York: Springer.

Letac, G. and Rahman, Q. I. (1986), "A factorization of the Askey's characteristic function $(1 + \|t\|_{2n+1})_+^{n+1}$." *Ann. Inst. H. Poincaré Probab. Statist.*, 22, 169–174.

Ma, C. (2003), "Families of spatio-temporal stationary covariance models." *Journal of Statistical Planning and Inference*, 116, 489–501.

— (2005), "Semi-parametric spatio-temporal covariance models with the ARMA temporal margin." *Annals of the Institute of Statistical Mathematics*, 57, 221–233.

— (2011), "Vector random fields with second-order moments or second order increments." *Stochastic Analysis and Applications.*, 29, 197–215.

Matérn, B. (1960), "Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations." *Meddelanden fran Statens Skogsforskningsinstitut*, 49, 144.

Pearce, I. G., Chaplain, M. A. J., Schofield, P. G., Anderson, A. R. A., and Hubbard, S. F. (2006), "Modeling the spatio-temporal dynamics of multi-species host-parasitoid interactions: heterogeneous patterns and ecological implications." *Journal of Theoretical Biology*, 241, 876–886.

Pissanetzky, S. (1984), *Spare Matrix Technology.*, Academic Press.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Ramsay, J. and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer, 2nd ed.

Rudin, W. (1963), "The extension problem for positive-definite functions." *Illinois Journal of Mathematics*, 7, 532–539.

Schlather, M. (2011), *RandomFields: Simulation and Analysis of Random Fields*, r package version 2.0.45.

Staggenborg, S. A., Dhuyvetter, K. C., and Gordon, W. B. (2008), "Grain sorghum and corn comparisons: Yield, economic, and environmental responses." *Agronomy Journal*, 100, 1600–1604.

Stein, E. M. and Weiss, G. (1971), *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, New Jersey.

Stein, M. L. (2005a), "Space-time covariance functions." *Journal of the American Statistical Association*, 100, 310–321.

— (2005b), "Statistical methods for regular monitoring data." *Journal of the Royal Statistical Society. Series B*, 67, 667–687.

Stein, M. L., Chi, Z., and Welty, L. J. (2004), "Approximating likelihoods for large spatial datasets." *Journal of the Royal Statistical Society. Series B*, 66, 275–296.

Storvik, G., Frigessi, A., and Hirst, D. (2002), "Stationary space-time Gaussian fields and their time autoregressive representation." *Statistical Modeling*, 2, 139–161.

Sun, Y. and Genton, M. G. (2011), "Functional Boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334.

Sun, Y., Li, B., and Genton, M. G. (2012), *Advances and Challenges in Space-time Modeling of Natural Events.*, Springer Berlin Heidelberg, chap. Geostatistics for Large Datasets., pp. 55–77.

von Kármán, T. (1948), "Progress in the statistical theory of turbulence." *Proceedings of the National Academy of Sciences U.S.A.*, 34, 530–539.

Wendland, H. (1995), "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree." *Advances in Computational Mathematics*, 4, 389–396.

Whittle, P. (1954), "On stationary processes in the plane." *Biometrika*, 41, 434–449.

Wikle, C. K. and Royle, J. A. (2005), "Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data." *Environmetrics*, 16, 507–522.

Xu, K., Wikle, C. K., and Fox, N. I. (2005), "A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities." *Journal of the American Statistical Association*, 100, 1133–1144.

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions, Vol. I, Basic Results.*, Springer, New York.

Zhang, H. and Du, J. (2008), *Positive Definite Functions: From Schoenberg to Space-time Challenges*, Gráficas Casta, Spain, chap. Covariance tapering in spatial statistics.