View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Texas A&M Repository

PREDICTION AND ESTIMATION OF RANDOM FIELDS

A Dissertation

by

PRIYA KOHLI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Statistics

PREDICTION AND ESTIMATION OF RANDOM FIELDS

A Dissertation

by

PRIYA KOHLI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Mohsen Pourahmadi
	Willa W. Chen
Committee Members,	Suhasini Subba Rao
	Gerald R. North
	Marc G. Genton
Head of Department,	Simon J. Sheather

August 2012

Major Subject: Statistics

ABSTRACT

Prediction and Estimation of Random Fields.

(August 2012)

Priya Kohli, B.S, Delhi University, New Delhi;

M.S., Indian Agricultural Statistics Research Institute, New Delhi;

M.S., Northern Illinois University, DeKalb

Co-Chairs of Advisory Committee: Dr. Mohsen Pourahmadi
Dr. Willa W. Chen

For a stationary two dimensional random field, we utilize the classical Kolmogorov-Wiener theory to develop prediction methodology which requires minimal assumptions on the dependence structure of the random field. We also provide solutions for several non-standard prediction problems which deals with the "modified past," in which a finite number of observations are added to the past. These non-standard prediction problems are motivated by the network site selection in the environmental and geostatistical applications. Unlike the time series situation, the prediction results for random fields seem to be expressible only in terms of the moving average parameters, and attempts to express them in terms of the autoregressive parameters lead to a new and mysterious projection operator which captures the nature of edge-effects. We put forward an approach for estimating the predictor coefficients by carrying out an extension of the exponential models. Through simulation studies and real data example, we demonstrate the impressive performance of our prediction method. To the best of our knowledge, the proposed method is the first to deliver a unified framework for forecasting random fields both in the time and spectral domain without making a subjective choice of the covariance structure.

Finally, we focus on the estimation of the hurst parameter for long range dependence

stationary random fields, which draws its motivation from applications in the environmental and atmospheric processes. Current methods for estimation of the Hurst parameter include parametric models like fractional autoregressive integrated moving average models, and semiparametric estimators which are either inefficient or inconsistent. We propose a novel semiparametric estimator based on the fractional exponential spectrum. We develop three data-driven methods which can automatically select the optimal model order for the fractional exponential models. Extensive simulation studies and analysis of Mercer and Hall's wheat data are used to illustrate the performance of the proposed estimator and model order selection criteria. The results show that our estimator outperforms existing estimators, including the GPH (Geweke and Porter-Hudak) estimator. We show that the proposed estimator is consistent, works for different definitions of long range dependent random fields, is computationally simple and is not susceptible to model misspecification nor poor efficiency.

To Lord Shiva and my Mom

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support of many people. First, I am greatly indebted to my advisor, Dr. Mohsen Pourahmadi for his invaluable advice, patience, guidance and cooperation during my years as a graduate student. I consider myself fortunate and honored to have had the opportunity to be his student; it has been a privilege to get a glimpse of the field of statistics through his eyes. He taught me how to ask questions and express my ideas. He showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. He patiently helped me become a more independent researcher.

A special thanks also goes to my co-advisor, Dr. Willa W. Chen whose constant encouragement and guidance have helped me in the completion of this dissertation. More importantly, she was always there to meet and talk about my ideas, to proofread and ask me good questions to help me think through my problems and to do qualitative research. Her affectionate and caring behavior was never limited to a guide; she gave all kind of moral support. She has modeled a lesson that I will gladly carry forward with me in my future work.

I am also grateful to the members of my dissertation committee, Dr. Gerald North, Dr. Suhasini Subba Rao and Dr. Marc Genton, for generously giving their time and insight to improve my work.

I owe a special note of gratitude to Dr. Michael Longnecker for his invaluable guidance, advice and constant encouragement throughout my period of study at Texas A&M University.

My special thanks and appreciation also go to all the Department of Statistics, Texas A&M University, especially Dr. Simon Sheather, Dr. Fred Dahm, Marilyn, Crystal and Carrie for their kind assistance during my study.

I would like to express my sincere thanks to the Division of Statistics, Northern Illinois University from where I started my journey as a graduate student. At this moment the motherly affection of Donna Lynn deserves a heartfelt thanks and appreciation. She always encouraged and supported me.

I would like to express my heartfelt thanks to my invaluable network of supportive, generous and loving friends, especially Yoyo, my classmates, seniors and juniors without whom I could not have survived the process.

This dissertation would never have been accomplished without the compassion, blessings and support of my mom, dad and my brother Kapil. They never failed to encourage and bless me without a second thought. I would like to express my love and respect to my family for providing moral support, for listening to my complaints and frustrations, and for believing in me to overcome my moments of doubt. I would never have made it up to this point if it had not been for the most special person in my life, my sweetheart mom Anita, whose love, unconditional support, encouragement and friendship provided constant comfort and enabled me to carry on when things did not look that bright. This work is dedicated to you, mom.

Diction is not enough to express my respect for my lovable friend, Deep and his family. His friendship, knowledge, hospitality, sacrifice and wisdom have enlightened me in every step. He gave me sufficient energy and enthusiasm to complete this work.

I feel very much indebted to all the above individuals and several colleagues who have assisted me one way or another.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS viii

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER

I INTRODUCTION 1

1.1 Dissertation Organization 2

II PREDICTION OF STATIONARY RANDOM FIELDS 5

2.1 Introduction 5
2.2 Prediction of Stationary Time Series 7
2.3 Prediction of Stationary Random Fields 11
2.3.1 MA Representation with Half-Plane Past 12
2.3.2 MA Representation with Quarter-Plane Past 14
2.3.3 Multi-Step Ahead Prediction 16
2.3.4 Recursive Formulas for the AR and MA Coefficients . 19
2.4 Prediction with Modified Quarter Plane Past 20
2.4.1 A Single Additional Observation 22
2.4.2 Several Additional Observations 23
2.5 Worth of Observation 25

III ESTIMATION OF STATIONARY RANDOM FIELDS 28

3.1 Introduction 28
3.2 Predictor Coefficients: Time Series 29
3.3 Bivariate Exponential Models 31
3.3.1 Regression Formulation for Data on a Lattice 32
3.4 Predictor Coefficients: Random Fields 34

CHAPTER Page

3.5 Simulations and Data Analysis 36
3.5.1 Simulation Study 37
3.5.2 Data Analysis 43

IV LONG RANGE DEPENDENT RANDOM FIELDS 46

4.1 Introduction 46
4.2 Existing Estimators 50
4.2.1 Parametric Approach 50
4.2.2 Semi-parametric Approach 51
4.3 Proposed Semi-parametric Estimator 52
4.3.1 FEXP Estimator 53
4.4 FEXP Model Order Selection 56
4.4.1 Model Selection 56
4.4.2 Subset Selection 57
4.5 Simulations 57
4.6 Mercer and Hall Wheat Data 67

V CONCLUSION 71

REFERENCES 74

APPENDIX A 82

APPENDIX B 90

APPENDIX C 91

VITA 93

LIST OF TABLES

TABLE Page

1 Mean and Variance of the (p, q) Values Selected Using AICC. 41

2 Predictions and Prediction Error Variances Obtained Using EXP Algorithm and Estimated ARMA Model. 42

3 Predictions and Prediction Error Variances for Wheat Yield Data Using EXP Algorithm and Kriging. (p, q) is the model order for the EXP model selected using AICC. 45

4 RMSE of FEXP Estimator for Separable AR(1) Models. 60

5 RMSE of FEXP and GPH Estimators. 60

6 RMSE of FEXP Estimator for Separable AR(1) Models Using Subset Selection. n_s(var) denotes the average (and variance) of the number of variables selected in 1000 replications. 61

7 RMSE of FEXP Estimator for Separable AR(1) Models Using Global and Local C_L and AICC. 62

8 RMSE of FEXP Estimator for Nonseparable AR(1) Models. 64

9 RMSE of FEXP Estimator for Nonseparable AR(1) Models Using Subset Selection. n_s(var) denotes the average (and variance) of the number of variables selected in 1000 replications. 65

10 RMSE of FEXP Estimator for Nonseparable AR(1) Models Using Global and Local C_L and AICC. 66

11 FEXP Estimators for Wheat Yield Data. 69

LIST OF FIGURES

FIGURE Page

1 Model 1: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA. 39

2 Model 2: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA. 40

3 Model 3: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA. 41

4 Wheat Yield Data: (a) Spatial Plot and (b) Histogram. Median Polish Residuals: (c) Spatial Plot and (d) Histogram. 44

5 Spectral Density for $\alpha = 0.6$ and AR(1) Model with $(\phi_{1,0}, \phi_{0,1})$ (a) $(0.6, 0.4)$, (b) $(0.6, -0.4)$ and (c) $(-0.6, -0.4)$. 59

6 Spectral Density for $\alpha = 0.6$ and AR(1) Model with (a) $(0.24, 0.24)$, (b) $(0.24, -0.24)$ and (c) $(-0.24, -0.24)$. 63

7 Log Periodogram of the (a) Wheat Data, (b) Row Series and (c) Column Series. 68

8 (a) Log Periodogram of Wheat Yield, and Log of Fitted Spectral Density for (b) Model (4.3), (c) Model (4.5) and (d) Model (4.7). 70

9 Prediction Errors in Quadrants II and IV. 84

CHAPTER I

INTRODUCTION

A stochastic process is a collection of random variables that are indexed by some set $D \subset \mathbb{R}^d$ containing spatial coordinates $\mathbf{s} = [s_1, s_2, \cdots, s_d]'$. When the dimension d of the index set is greater than one, the stochastic process is referred as the *random field*. A two dimensional (2-D) random field, $d = 2$ is often referred as spatial process because it contains information about the attribute of interest and its location. The location may be a set of coordinates, such as the latitude and longitude associated with an observed pollutant level, or it may be a small region such as a county associated with an observed disease rate. Let $X(\mathbf{s})$ be a random variable that can be measured at location $\mathbf{s} = (x, y)$ in the region $D \in \mathbb{R}^2$, then the random field is denoted by $\{X(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^2\}$ or for short as $\{X(\mathbf{s})\}$.

When the 2-D random field is indexed over a regular grid of points, that is when D is a finite (or countable) collection of spatial locations at which the random variable X is measured then it is called a *lattice process* with lattice D. The spatial sites in a lattice are typically identified using their longitude x and latitude y. These processes could be regular or irregular lattice type depending on the pattern of the locations. For instance, yield from an agricultural plot is usually observed on a regular lattice and the percentage of population below poverty line in five midwest states of Illinois, Indiana, Michigan, Ohio, and Wisconsin is observed on an irregular lattice.

We consider 2-D random fields which are observed on a regular lattice. These processes have seen a rapid rise in popularity due to demand from a wide range of fields. We

---

The format and style follow that of *Journal of the American Statistical Association*.

discuss two such areas of applications next.

1. **Remote Sensing Studies**

   In remote sensing studies, satellites and aircraft can be used to obtain images from which the spatial data is produced. In this case the data is in the form of pixels (or voxels), which are small rectangularly shaped regions. Some examples include inventory of natural resource, modeling weather patterns, monitoring the effects of forest clearing and erosion among various other application areas.

2. **Medical Imaging Studies**

   In this area the image is captured in a discrete, digitized format but the objects in an image vary continuously (colors, lines, etc.). Applications include images generated from magnetic resonance (MRI) imaging and positron emission tomography (PET).

The data from these two studies is measured on vastly different scales but the statistical problems posed are similar as they rely on the fact that the neighboring locations are dependent, see Cressie (1993, p.499). In this work, we propose a unified framework for the prediction and estimation of 2-D random fields observed on a regular lattice. For ease of notation, we will refer to the 2-D random fields observed on a regular lattice as random fields which are denoted by $\{X(\mathbf{s})\}$ from here on.

1.1 Dissertation Organization

In Chapter II, we present several results for prediction of random fields by extending the classical Kolmogorov-Wiener prediction theory of stationary (1-D) processes. Using these predictors, their prediction errors and variances, we also solve a number of non-standard prediction problems which deals with different modifications of the past for prediction. Most of the existing methods are based on the assumption that the dependence structure of the random field is known upto a finite number of unknown parameters. In a practical

situation, this means that for a given dataset one needs to specify a parametric model for covariance, which can be selected from a menu of available models in standard software packages. Our proposed method is first to deliver prediction theory for random fields based on a fixed past without making a subjective choice of the covariance structures.

In Chapter III, we answer the obvious question: How does one start to apply the prediction theory developed in Chapter II to a given dataset? The proposed methodology is implemented by first fitting an extended version of Bloomfield's (1973) exponential model to the spectrum and then using the recursive formulas which expresses the predictor coefficients in terms of the cepstral coefficients of the process. The extension of exponential models to random fields has been studied in the engineering literature in the contexts of texture and image analysis. However, the construction of predictors of stationary random fields with exponential spectrum has been lagging behind. The prediction theory combined with this exponential model based estimation thus, provides a unified framework for forecasting random fields both in the time and spectral domain. A simulation study that investigates the predictive performance of the proposed methodology is included. The proposed framework is then applied to a dataset of yields from an agricultural experiment. Both the simulation studies and application to real study show the satisfactory performance of the proposed prediction and estimation method.

Chapter IV focusses on long range dependent random fields. A common scientific objective for such processes is to estimate the parameter which controls the long range behavior of the process. We first review the existing methods of estimation and discuss their major pros and cons. Then we propose a novel semi-parametric estimator called fractional exponential (FEXP) estimator for the long range dependence parameter. There are many different ways to define a long range dependent random field and we show that the proposed estimator can be applied very easily to these different definitions. It turns out that the statistical properties of the proposed estimator depends on the order of the exponential

model. To alleviate this problem we generalize the existing model order selection techniques to introduce data dependent automatic selection of the appropriate model. Through extensive simulation experiments and an application to the wheat yield data, we conform that the proposed estimator not only leads to more reasonable estimates for the long range dependence parameter, but also avoids the inconsistencies resulting from other methods.

Finally, we conclude with the main findings of the work and some problems for further research in Chapter V.

CHAPTER II

PREDICTION OF STATIONARY RANDOM FIELDS

2.1 Introduction

One of the primary goals in geostatistics is prediction at unsampled locations based on the measurements available at known locations. These predictions provide a guide to make practical decisions, for instance spatial prediction can be done for variables like temperature, pressure, air pollution, home prices and disease concentration. Poor predictions will lead not only to poor decisions and planning but can also waste time, money and resources.

Kriging, the most commonly used method for spatial interpolation, deals with the prediction of $X(.)$ over all D when X is only observed at a finite number of points $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n\}$ in D. Matheron (1971) introduced kriging based on the work of Krige (1951). Kriging belongs to the family of linear least squares estimation methods, and estimates values at unobserved location as a weighted average of the neighboring observed values. The determination of unknown weights require specification of a parametric model for the covariance structure with few parameters. In practice, the model for the covariance is selected from a list of available models and then its parameters are estimated from the observed data $\{X(\mathbf{s})\}$. This presupposition that the data can be modelled by a specific covariance function is not universally accepted because a misspecified model can often lead to highly biased predictions. In addition, assuming a fixed dependence structure might produce too smooth image of the data at hand due to either underestimation of the extreme values or the poor reproduction of the short scale variations.

For stationary time series (or 1-D processes) the prediction theory began with the seminal work of Kolmogorov (1939, 1941) and Wiener (1949). Their prediction theory, hereafter referred to as the KW theory, makes no parametric model assumption for the

covariance structure and incorporates the dependence among the observations at different time points by using the Wold decomposition of the process.

Given the success of the KW theory, we expected its extension to play a key role in the prediction of 2-D stationary processes. The Wold decomposition is of basic importance in the KW theory, however, for random fields we have to assign different meaning to the "past" and "future" in different prediction problems. Hence, different definitions of "past" will lead to different Wold type decompositions. We develop the prediction theory for stationary random fields based on the Wold decomposition corresponding to quarter-plane past. Being an extension of the KW theory, the proposed prediction theory does not assume a specific model for the covariance structure, thus avoiding the pitfalls resulting from model misspecification.

Further, we study the solutions of several prediction problems for random fields by extending some nonstandard prediction problems for stationary time series based on the modification of the past. Their solutions lead to informative and explicit expressions involving the autoregressive (AR) and moving average (MA) parameters. These prediction problems provide useful information for assessing the worth of observations in the spatial setting and are closely related to the design issues or network site selection in the environmental, geostatistical and engineering applications, see Zimmerman (2006).

The rest of this chapter is organized as follows: In Section 2.2 we briefly describe the KW theory for prediction of time series along with a summary of the prediction results corresponding to the modified pasts. We review some basic results on stationary random fields, their Wold decompositions and Szegö's formula in Section 2.3. The multi-step ahead predictor based on the quarter plane past and its prediction error variance are also given. In Section 2.4, using the Wold decomposition of stationary random fields, their multi-step ahead prediction errors and variances, we provide solutions for various nonstandard prediction problems, when a number of observations are either added to the quarter-plane

past. Section 2.5 introduces the notion of worth of an observation in spatial prediction. We also describe the role of nonstandard prediction problems in finding the worth of the observations.

2.2 Prediction of Stationary Time Series

The prediction theory of a stationary time series $\{X(t), t \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers, is concerned with computing the linear least-squares predictors of the future values based on the knowledge of the infinite past and the spectral density function of $\{X(t)\}$. Kolmogorov (1939, 1941) and Wiener (1949) developed a comprehensive theory of prediction for a stationary process using its spectral factorization and the Wold decomposition. The Wold decomposition gives the fundamental unilateral representation of a stationary (nondeterministic) process $\{X(t)\}$ in terms of its innovations $\{\varepsilon(t)\}$ as:

$$X(t) = \sum_{k=0}^{\infty} b_k \varepsilon(t-k) + V(t),$$

where $\{b_k\}$ are the MA parameters with $b_0 = 1$, $\sum_{k=1}^{\infty} |b_k|^2 < \infty$, $b_k = 0, k < 0$ and $\{\varepsilon(t)\}$ is a white noise process with mean zero and variance σ^2 called the *innovation process* of $\{X(t)\}$. Here $\{V(t)\}$ is a deterministic process orthogonal to $\{\varepsilon(t)\}$. For *purely nondeterministic* (PND) process the Wold decomposition has no deterministic component such that

$$X(t) = \sum_{k=0}^{\infty} b_k \varepsilon(t-k). \tag{2.1}$$

It is well known that a necessary and sufficient condition for a process to be PND is given by

$$\int_{-\pi}^{\pi} \log f(\mathrm{e}^{i\lambda}) d\lambda > -\infty. \tag{2.2}$$

This condition is also necessary and sufficient for the spectral factorization theorem to hold. The factorization theorem provides an analytic function $\phi(z)$ of complex variable $z = \mathrm{e}^{i\lambda}$,

such that the spectral density function of $\{X(t)\}$ satisfies

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left|\phi(z)\right|^2. \tag{2.3}$$

This function $\phi(z)$ is called the *spectral factor* of $f(\lambda)$ and

$$\phi(z) = \sum_{j=0}^{\infty} b_j z^j,$$

with $\phi(0) = 1$. If $\phi(z) \neq 0$ in $|z| < 1$, then the inverse for $\phi(z)$ exists, see Pourahmadi (2001, §5.5) and Hannan (1970, Chap. 3). The Taylor series expansion of the inverse function, $\phi^{-1}(z) = 1/\phi(z)$ is

$$\phi^{-1}(z) = \sum_{j=0}^{\infty} a_j z^j, \ |z| < 1,$$

where $\{a_k\}$ are the AR(∞) parameters of $\{X(t)\}$ which gives the unilateral representation of $X(t)$ in terms of its past as:

$$X(t) = -\sum_{k=1}^{\infty} a_k X(t-k) + \varepsilon(t). \tag{2.4}$$

In KW theory the predictions are expressed in terms of the MA and AR parameters as given in the following result.

Lemma 1. Let $\{X(t)\}$ be a nondeterministic stationary process with spectral density function $f(\lambda)$ and spectral factor $\phi(z)$ with $\phi(z) \neq 0, \ |z| < 1$. Then

a.) the h-step ahead linear least-squares predictor in terms of the unobservable innovations is:

$$\widehat{X}(t+h) = \sum_{k=h}^{\infty} b_k \varepsilon(t+h-k),$$

where $\{b_k\}$ are the MA parameters or Fourier coefficients of $\phi(z)$.

b.) the corresponding prediction error is:

$$X(t+h) - \widehat{X}(t+h) = \sum_{k=0}^{h-1} b_k \varepsilon(t+h-k),$$

c.) the h-step ahead prediction error variance is:

$$\text{var}\left\{X(t+h) - \widehat{X}(t+h)\right\} = \sigma^2 \sum_{k=0}^{h-1} |b_k|^2,$$

d.) the h-step ahead linear least-squares predictor in terms of the observed past is:

$$\widehat{X}(t+h) = -\sum_{k=1}^{h-1} a_k \widehat{X}(t+h-k) - \sum_{k=h}^{\infty} a_k X(t+h-k),$$

where $\{a_k\}$ are the AR parameters or the Fourier coefficients of $\phi^{-1}(z)$. Here the first term corresponds to the unobserved future value between $X(t)$ and $X(t+h)$ whereas the second term consists of the infinite past of $X(t)$.

Solutions of some nonstandard prediction problems for a stationary time series $\{X(t)\}$ corresponding to the following modified pasts:

$$\begin{aligned} I_1 &= \{X(t); t \leq -1,\ t = h,\ t \neq 0\}, \\ I_2 &= \{X(t); t \leq h,\ t \neq 0\}, \\ I_3 &= \{X(t); t \leq -1,\ t \neq -h,\ t \neq 0\}, \end{aligned} \tag{2.5}$$

for any $h > 0$, are known to lead to the following informative and explicit expressions for the prediction error variance involving the AR and MA parameters of the process, see Nakazi (1984), Pourahmadi (1989), Miamee & Pourahmadi (1988), Cheng & Pourahmadi (1997), Bondon (2002, 2005), Pourahmadi, Inoue & Kasahara (2006) and Kasahara, Pourahmadi & Inoue (2009).

Lemma 2. Let $\{X(t); t \in \mathbb{Z}\}$ be a nondeterministic stationary process with the innovation process $\{\varepsilon(t); t \in \mathbb{Z}\}$, innovation variance σ^2, MA and AR parameters $\{b_k\}$ and $\{a_k\}$, respectively. Then, the prediction error variance of $X(0)$ based on

(a) the augmented past, I_1 is

$$\text{var}\left\{X(0) - \widehat{X}_{I_1}(0)\right\} = \sigma^2 \frac{1 + b_1^2 + b_2^2 + \cdots + b_{h-1}^2}{1 + b_1^2 + b_2^2 + \cdots + b_{h-1}^2 + b_h^2}, \tag{2.6}$$

(b) the augmented past, I_2 is

$$\text{var}\left\{X(0) - \widehat{X}_{I_2}(0)\right\} = \sigma^2 \frac{1}{1 + a_1^2 + a_2^2 + \cdots + a_h^2}, \tag{2.7}$$

(c) the incomplete past, I_3 is

$$\text{var}\left\{X(0) - \widehat{X}_{I_3}(0)\right\} = \sigma^2 \frac{1 + a_1^2 + a_2^2 + \cdots + a_{h-1}^2 + a_h^2}{1 + a_1^2 + a_2^2 + \cdots + a_{h-1}^2}. \tag{2.8}$$

These explicit expressions involving $\sum_{k=1}^{h} b_k^2$ and $\sum_{k=1}^{h} a_k^2$, which are reminiscent of the h-step ahead prediction error variance $\sigma^2 \sum_{k=0}^{h-1} b_k^2$, reveal the roles of the AR and MA parameters in assessing the effect of addition or deletion of observations from the past on the prediction error variance. For example, from (2.6) it is evident that for a stationary process with $b_h = 0$ adding the variable $X(h)$ to the past will not improve the prediction of $X(0)$. Similarly, from (2.8) one can see that deleting $X(-h)$ from the past will not deteriorate the prediction of $X(0)$ as long as $a_h = 0$.

The prediction error variance in (2.7) was obtained by Nakazi (1984) using spectral domain techniques and deep duality results in harmonic analysis. However, his method was too rigid to allow computing the predictor corresponding to the augmented past I_2. Nakazi's approach was modified and a time-domain (regression) method was developed by Pourahmadi (1989) to handle the other prediction problems in Lemma 2 without requiring the unnatural minimality condition on the process and where the linear predictor of $X(0)$ based on I_2 was found.

2.3 Prediction of Stationary Random Fields

Given the success of the KW prediction theory for a stationary time series, it is expected that the Wold decomposition and spectral factorization corresponding to a given past is the key for solving the prediction problem for stationary random fields. In this section, we review some of the basic results, and present some new results for the prediction of the 2-D random fields.

Time series models are unilateral in structure following a natural notion of past and future. Unlike the situation in time series, there is no unique definition of the past for a stationary random field with discrete time-index in the plane (2-D) or higher dimensions. Consequently, the prediction theory of stationary random fields $\{X(s_1, s_2); (s_1, s_2) \in \mathbb{Z}^2\}$ is very much dependent on the choice of a past like the half-plane, see Helson & Lowdenslager (1958) and the quarter-plane, Tjøstheim (1983), Kallianpur & Mandrekar (1983), Soltani (1984) and Rosenblatt (1985), among others. The typical examples of the half-plane and quarter-plane are:

$$S = \{(i, j) : i \leq -1, j \in \mathbb{Z}\} \cup \{(0, j) : j \leq -1\},$$
$$Q = \{(i, j) : i \leq 0, j \leq 0\} \backslash \{(0, 0)\}, \tag{2.9}$$

which correspond to the left half-plane and the third quadrant in $\mathbb{Z}^2$, respectively. We develop a framework for computing the best linear predictors and the prediction error variances for stationary random fields when Q (the third quadrant) is used as the past. The focus of the earlier work on prediction of random fields has been on *one-step ahead prediction* and the extension of the Szegö-Kolmogorov-Wiener formula for the innovation variance. Our focus, however, is on using the *multi-step ahead predictors* and their prediction error variances in solving several nonstandard prediction problems.

Let H be the Hilbert space of zero-mean, square-integrable random variables defined on a probability space. A sequence $\{X(s_1, s_2); (s_1, s_2) \in \mathbb{Z}^2\}$ with $X(s_1, s_2) \in H$ is called

a *stationary random field* if for all integers s_1, s_2, t_1 and t_2, the covariance of $X(s_1, s_2)$ and $X(t_1, t_2)$ depends only on the lags $(s_1 - t_1, s_2 - t_2)$, namely,

$$\text{cov}\,(X(s_1, s_2), X(t_1, t_2)) = \gamma(s_1 - t_1, s_2 - t_2).$$

Since $\gamma(.,.)$ is a positive-definite function on the group of lattice points $\mathbb{Z}^2$, by Bochner's Theorem there exists a unique distribution function $F(.,.)$ on the torus $(-\pi, \pi] \times (-\pi, \pi]$ such that

$$\gamma(s_1, s_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-i(s_1\lambda_1 + s_2\lambda_2)} dF(\lambda_1, \lambda_2);\ (s_1, s_2) \in \mathbb{Z}^2. \tag{2.10}$$

In what follows we assume that F is absolutely continuous with respect to the Lebesgue measure $d\lambda_1 d\lambda_2$ with the Radon-Nikodyn derivative $f(\lambda_1, \lambda_2)$ which is called the *spectral density function* of the random field.

Let $L^2(f)$ denote the Hilbert space of all functions on the torus which are square integrable with respect to the measure $d\lambda_1 d\lambda_2$. From (2.10), it is evident that the map $X(s_1, s_2) \to \mathrm{e}^{\{-i(s_1\lambda_1 + s_2\lambda_2)\}}$, extends to an isomorphism from $H_X =$ the closed linear subspace of H spanned by $\{X(s_1, s_2);\ (s_1, s_2) \in \mathbb{Z}^2\}$, onto $L^2(f)$. For any subset M of the lattice points in the plane, define H_X^M as the closed linear subspace spanned by $X(s_1, s_2)$ with $(s_1, s_2) \in M$, in the Hilbert space H. In the sequel, we use $H_X^{m\infty}, H_X^{\infty n}$ and H_X^{mn} corresponding to the indices from the sets $\{(s_1, s_2); s_1 \leq m, s_2 \in \mathbb{Z}\}$, $\{(s_1, s_2); s_1 \in \mathbb{Z}, s_2 \leq n\}$ and $\{(s_1, s_2); s_1 \leq m, s_2 \leq n\}$, respectively. For a given spectral density function $f(\lambda_1, \lambda_2)$, the subspaces $H_f^{m\infty}, H_f^{\infty n}$ and H_f^{mn} in $L^2(f)$ are defined as the closed linear span of exponentials $\mathrm{e}^{(s_1\lambda_1 + s_2\lambda_2)}$ with indices from the indicated sets. Finally, P_X^M stands for the orthogonal projection operator from H_X onto H_X^M.

2.3.1 MA Representation with Half-Plane Past

The general theory of prediction of stationary random fields based on half-plane as the past were developed in the seminal paper of Helson & Lowdenslager (1958). A half-plane $\mathcal{S}$

in the sense of Helson and Lowdenslager is a subset of lattice points located at the origin $\mathbf{0} = (0, 0)$, which satisfy

1. $\mathcal{S} \cup \{\mathbf{0}\}$ is an additive semigroup,
2. $\mathcal{S} \cup \{\mathbf{0}\} \cup (-\mathcal{S}) = \mathbb{Z}^2$,
3. $\mathcal{S} \cap (-\mathcal{S}) = \varnothing$,

where the last property ensures that $\mathbf{0} \notin \mathcal{S}$. The half-plane space $\mathcal{S}$ induces an order on $\mathbb{Z}^2$ which coincides with the lexicographic order, that is

$$(s_1, s_2) < (s_1', s_2') \text{ if } (s_1 - s_1', s_2 - s_2') \in \mathcal{S}.$$

Note that the half-plane S in (2.9) provides a specific example of such half-planes that we work with in this section.

For any $h = (h_1, h_2)$, S_h stands for its shifted version, that is,

$$S_h = \{(i + h_1, j + h_2); (i, j) \in S\}.$$

A random field $\{X(s_1, s_2); (s_1, s_2) \in \mathbb{Z}^2\}$ is said to be purely nondeterministic (PND) if

$$\bigcap_{h \in \mathbb{Z}^2} H_X^{S_h} = \{0\}.$$

It is known (Helson and Lowdenslager 1958, 1961) that a random field is PND, if and only if its spectral distribution $F(.,.)$ is absolutely continuous with respect to the Lebesgue measure $d\lambda_1 d\lambda_2$ and its spectral density function $f(\lambda_1, \lambda_2)$ satisfies the condition

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 > -\infty. \tag{2.11}$$

In this case, there exists a one-sided (unilateral) MA(∞) representation for $X(s, t)$:

$$X(s_1, s_2) = \varepsilon(s_1, s_2) + \sum_{(k,\ell) \in S} \sum b_{k,\ell} \varepsilon(s_1 - k, s_2 - \ell),$$

$$H_X^S = H_\varepsilon^S, \tag{2.12}$$

where $\{\varepsilon(s_1, s_2)\}$ is the *innovation process* of $\{X(s_1, s_2)\}$ with mean zero and variance σ^2, $\{b_{k,\ell}\}$ denote the MA parameters with $b_{0,0} = 1$, $b_{k,\ell} = 0$ when either $k < 0$ or $\ell < 0$ and $\sum\sum_{(k,\ell)\in S} b_{k,\ell}^2 < \infty$. Moreover, they also provided an extension of the Szegö's formula expressing the innovation variance $\sigma^2 = \sigma^2(S)$ as:

$$\sigma^2(S) = \exp\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \log f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2\right\}. \tag{2.13}$$

This formula for the innovation variance was used by Rosenblatt (1985) and Kallianpur, Miamee & Niemi (1990) to obtain similar formulas when infinite number of observations were added to or deleted from the half-plane S:

$$S^+ = \{(i,j) : i \leq 0, (i,j) \neq (0,0)\}, \;\; S^- = \{(i,j) : i \leq -1\}.$$

In particular, for S^+ it is known from Rosenblatt (1985, p. 225) and Kallianpur et al. (1990, Thm II.7) that

$$\sigma^2(S^+) = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda_1, \lambda_2) d\lambda_1\right\} d\lambda_2\right]^{-1},$$

where evidently $\sigma^2(S^+) \leq \sigma^2(S)$.

2.3.2 MA Representation with Quarter-Plane Past

In most of what follows we work with the quarter plane, denoted by Q from here on. With Q as the past, Kallianpur & Mandrekar (1983) obtained a four-fold Wold decomposition in the time-domain for a stationary random field. For the precise statement of the corresponding MA representation and the necessary and sufficient spectral conditions for it, we recall that a stationary random field $\{X(s_1, s_2)\}$ is said to have a one-sided (unilateral) MA(∞) representation on Q if there exists a white noise process $\{\varepsilon(s_1, s_2)\}$ such that

$$X(s_1, s_2) = \varepsilon(s_1, s_2) + \sum\sum_{(k,\ell)\in Q} b_{k,\ell}\varepsilon(s_1 - k, s_2 - \ell),$$
$$H_X^{mn} = H_\varepsilon^{mn} \quad \text{for all} \quad (m,n) \in \mathbb{Z}^2, \tag{2.14}$$

where the sequence $\{b_{k,\ell}\}$ consists of the MA parameters of the process, with $b_{0,0} = 1$, $b_{k,\ell} = 0$, when $k < 0$ or $\ell < 0$ and $\sum\sum_{(k,\ell)\in Q} b_{k,\ell}^2 < \infty$. In view of the equality of the two subspaces involving the pasts of the two process, we call $\{\varepsilon(s_1, s_2)\}$ the *innovation process* of $\{X(s_1, s_2)\}$ and the var $\{\varepsilon(s_1, s_2)\} \equiv \sigma^2 = \sigma^2(Q)$ is referred to as the *innovation variance*.

The following result by Soltani (1984) gives the spectral characterization of stationary random fields having a one-sided MA(∞) representation on Q.

Lemma 3. A stationary random field $\{X(s_1, s_2)\}$ with spectral density function $f(\lambda_1, \lambda_2)$ has a one-sided MA(∞) representation (2.14) on the quarter plane, if and only if

i.) $\log f \in L^1$, see (2.11),

ii.) the Fourier coefficients of $\log f$ vanish outside $Q \cup \{-Q\} \cup \{\mathbf{0}\}$,

iii.) $H_f^{00} = H_f^{0\infty} \bigcap H_f^{\infty 0}$,

where H_f^M is defined as before with M:

$$\begin{aligned} M &= \{(s_1, s_2); s_1 \le 0, s_2 \le 0\}, \text{for } H_f^{00}, \\ &= \{(s_1, s_2); s_1 \le 0, s_2 \in \mathbb{Z}\}, \text{for } H_f^{0\infty}, \\ &= \{(s_1, s_2); s_1 \in \mathbb{Z}, s_2 \le 0\}, \text{for } H_f^{\infty 0}. \end{aligned}$$

The spectral factorization of the spectral density function on Q in terms of an analytic function $\phi(z_1, z_2)$ of complex variables $z_1 = \mathrm{e}^{i\lambda_1}$ and $z_2 = \mathrm{e}^{i\lambda_2}$ follows from Lemma 3(ii). In fact, we have

$$f(\lambda_1, \lambda_2) = \frac{\sigma^2}{2\pi} \left|\phi(z_1, z_2)\right|^2,$$

where $\phi(0, 0) = 1$ and

$$\phi(z_1, z_2) = \exp\left(\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k,\ell} z_1^k z_2^\ell\right), \tag{2.15}$$

where $\{c_{k,\ell}\}$ are the *two-dimensional cepstral coefficients* of $f(\lambda_1, \lambda_2)$ defined as:

$$c_{k,\ell} = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log f(\lambda_1, \lambda_2) \mathrm{e}^{\{-i(k\lambda_1 + \ell\lambda_2)\}} d\lambda_1 d\lambda_2.$$

It is known that the Szegö's formula for $\sigma^2 = \sigma^2(Q)$ is the same as in (2.13), see Miamee (1986, Corollary, 4.3).

2.3.3 Multi-Step Ahead Prediction

Predicting future values other than $X(0,0)$ is important in the theory and applications of stationary random fields, a problem which has not been studied systematically in the literature. In this section, we derive the multi-step ahead predictor and its prediction error variance when Q is used as the past.

<u>Theorem</u> 1. Let $\{X(s_1, s_2)\}$ be a PND stationary random field with the spectral density function $f(\lambda_1, \lambda_2)$ satisfying the conditions of Lemma 3. Then for any $(h_1, h_2) \in Q^c$, where Q^c is the complement of Q, the (h_1, h_2)-step ahead linear least-squares predictor of $X(h_1, h_2)$ based on the past Q is given by

$$\widehat{X}(h_1, h_2) = \sum_{\substack{k=h_1 \\ (k,\ell)\neq(h_1,h_2)}}^{\infty} \sum_{\ell=h_2}^{\infty} b_{k,\ell} \varepsilon(h_1 - k, h_2 - \ell),$$

with the corresponding prediction error

$$\begin{aligned} X(h_1, h_2) - \widehat{X}(h_1, h_2) &= b_{h_1,h_2} \varepsilon(0,0) + \sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} b_{k,\ell} \varepsilon(h_1 - k, h_2 - \ell) \\ &+ \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} b_{k,\ell} \varepsilon(h_1 - k, h_2 - \ell) + \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} b_{k,\ell} \varepsilon(h_1 - k, h_2 - \ell), \end{aligned} \tag{2.16}$$

and the prediction error variance

$$\begin{aligned} \mathrm{var}\left\{X(h_1, h_2) - \widehat{X}(h_1, h_2)\right\} = \sigma^2 \Bigg(b_{h_1,h_2}^2 &+ \sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} b_{k,\ell}^2 + \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} b_{k,\ell}^2 \\ &+ \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} b_{k,\ell}^2 \Bigg), \end{aligned} \tag{2.17}$$

where $\sigma^2 = \sigma^2(Q)$ is the innovation variance. The (h_1, h_2)-step ahead linear least-squares predictor of $\{X(h_1, h_2)\}$ in terms of the observed past is:

$$\hat{X}(h_1,h_2) = -\Bigg(\sum_{\substack{k=h_1 \\ (k,\ell)\neq(h_1,h_2)}}^{\infty} \sum_{\ell=h_2}^{\infty} a_{k,\ell} X(h_1-k, h_2-\ell) + \sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} a_{k,\ell} \widehat{X}(h_1-k, h_2-\ell) + \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} a_{k,\ell} \widehat{X}(h_1-k, h_2-\ell) + \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} \widehat{X}(h_1-k, h_2-l) + a_{h_1,h_2} \widehat{X}(0,0) \Bigg), \quad (2.18)$$

where $\{a_{k,\ell}\}$ are the Fourier coefficients of ϕ^{-1}. Here the first term corresponds to the infinite quarter plane past of $X(0,0)$ and the remaining terms are the unobserved future value between $X(0,0)$ and $X(h_1,h_2)$. The proof of Theorem 1 is in Appendix A.

Remark 1. Depending on (h_1, h_2) some terms in (2.16) and (2.17) become zero for either of $h_i < 0, i = 1, 2$, but they always involve the infinite sums which makes them quite different from their counterparts in 1-D process in the sense that the latter are always finite sums, see Pourahmadi (2001, p.181).

Corollary 1. For a PND stationary random field $\{X(s_1, s_2)\}$ with the spectral density function $f(\lambda_1, \lambda_2)$ satisfying the conditions of Theorem 3 we have,

(a). the covariance between prediction errors based on the knowledge of Q, for any two different observations in Q^c is

$$\text{cov}\left\{X(h_1,h_2) - \widehat{X}(h_1,h_2), X(h_1^{'},h_2^{'}) - \widehat{X}(h_1^{'},h_2^{'})\right\} = \sigma^2 \Bigg(b_{h_1,h_2} b_{h_1^{'},h_2^{'}} + \sum_{k=0}^{M_1-1} \sum_{\ell=M_2}^{\infty} b_{k,\ell} b_{k+|h_1-h_1^{'}|,\ell+|h_2-h_2^{'}|} + \sum_{k=M_1}^{\infty} \sum_{\ell=0}^{M_2-1} b_{k,\ell} b_{k+|h_1-h_1^{'}|,\ell+|h_2-h_2^{'}|} + \sum_{k=0}^{M_1-1} \sum_{\ell=0}^{M_2-1} b_{k,\ell} b_{k+|h_1-h_1^{'}|,\ell+|h_2-h_2^{'}|} \Bigg), \quad (2.19)$$

where $M_1 = \min(h_1, h_1^{'})$ and $M_2 = \min(h_2, h_2^{'})$.

(b). The covariance between $X(0,0)$ and the prediction error of $X(h_1,h_2)$ based on Q is

$$\text{cov}\left\{X(h_1,h_2)-\widehat{X}(h_1,h_2),X(0,0)\right\}=\begin{cases}\sigma^2 b_{h_1,h_2}, & \text{if } (h_1,h_2)\geq(0,0),\\ 0, & \text{otherwise.}\end{cases} \tag{2.20}$$

The results in Theorem 1 and Corollary 1 are the counterparts of those for multi-step ahead prediction errors in time series, see Pourahmadi (1989). From Corollary 1(b) it is clear that there is no correlation between $X(0,0)$ and the prediction errors for observations with either $h_1<0$ or $h_2<0$.

The prediction error variance for the observations in the second and fourth quadrant are quite different in the sense discussed next. The following result (proof in Appendix A) establishes stationarity of the prediction error process along a fixed direction, a similar result for the half-plane past is given in Kallianpur et al. (1990, Thm II.1).

Corollary 2. Let $\{X(s_1,s_2)\}$ be a PND stationary random field with the spectral density function $f(\lambda_1,\lambda_2)$ satisfying the conditions of Lemma 3. Suppose $Y(h_1,h_2)=X(h_1,h_2)-\widehat{X}(h_1,h_2)$ denotes the prediction error of $X(h_1,h_2)$ based on Q.

(a). For (h_1,h_2) in the second quadrant and for a fixed h_2, the 1-D process $\{Y(s,h_2),s\leq -1\}$ is covariance stationary with the auto-covariance function

$$\gamma_{h_2}(h)\equiv\text{cov}(Y(s_1,h_2),Y(s_2,h_2))=\sigma^2\sum_{k=0}^{\infty}\sum_{\ell=0}^{h_2-1}b_{k,\ell}b_{k+|h|,\ell}, \tag{2.21}$$

which depends only on the lag $h=s_1-s_2$. For $s_1=s_2=s$ the variance of the prediction error is

$$\text{var}\{Y(s,h_2)\}=\sigma^2\sum_{k=0}^{\infty}\sum_{\ell=0}^{h_2-1}b_{k,\ell}^2,$$

which is independent of s and monotone in h_2. The spectral density function of $\{Y(s,h_2)\}$ is of the form:

$$f_{h_2}(\lambda)=\frac{\sigma^2}{2\pi}\sum_{\ell=0}^{h_2-1}\left|\phi_\ell(\mathrm{e}^{i\lambda})\right|^2, \tag{2.22}$$

where $\phi_\ell(.)$ is given by

$$\phi_\ell(\mathrm{e}^{i\lambda}) = \sum_{k=0}^{\infty} b_{k,\ell} \exp(ik\lambda).$$

(b). For (h_1, h_2) in the fourth quadrant and for fixed h_1, the 1-D process $\{Y(h_1, t), t \leq -1\}$ is covariance stationary with the auto-covariance function

$$\gamma_{h_1}(h) \equiv \mathrm{cov}(Y(h_1, t_1), Y(h_1, t_2)) = \sigma^2 \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{\infty} b_{k,\ell} b_{k,\ell+|h|}, \tag{2.23}$$

as a function of the lag $h = t_1 - t_2$. For $t_1 = t_2 = t$, the variance is

$$\mathrm{var}\,\{Y(h_1, t)\} = \sigma^2 \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{\infty} b_{k,\ell}^2,$$

which is independent of t and monotone in h_1. The spectral density function of $\{Y(h_1, t)\}$ is:

$$f_{h_1}(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=0}^{h_1-1} \left|\phi_k(\mathrm{e}^{i\lambda})\right|^2, \tag{2.24}$$

where $\phi_k(.)$ is given by:

$$\phi_k(\mathrm{e}^{i\lambda}) = \sum_{\ell=0}^{\infty} b_{k,\ell} \exp(ik\lambda).$$

2.3.4 Recursive Formulas for the AR and MA Coefficients

We provide recursive formulas to relate the MA and AR parameters in terms of the cepstral or Fourier coefficients of the logarithm of the spectral density function, see Pourahmadi (1984).

Consider the Taylor expansions of the optimal factor ϕ and its inverse:

$$\begin{aligned} \phi(z_1, z_2) &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} b_{k,\ell} z_1^k z_2^\ell, \\ \phi^{-1}(z_1, z_2) &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} a_{k,\ell} z_1^k z_2^\ell, \end{aligned} \tag{2.25}$$

from which it follows that the MA and AR parameters of the random field are related to each other via the recursions

$$b_{0,0} = a_{0,0} = 1,$$
$$b_{i,j} = \sum_{\substack{k=0 \\ (k,\ell)\neq(i,j)}}^{i} \sum_{\ell=0}^{j} b_{k,\ell} a_{i-k,j-\ell}. \tag{2.26}$$

2.4 Prediction with Modified Quarter Plane Past

In this section, we consider prediction of $X(0,0)$ when a finite number of observations are added to the quarter plane Q and the past is modified to:

$$I = Q \cup K,$$

where K represents a finite-dimensional space spanned by the additional observations. This is the first natural step in generalizing the 1-D results in Lemma 1 to the 2-D processes. It turns out that computing such predictors and prediction error variances are closely related to finding multi-step ahead predictions using the MA(∞) representation of a PND random field with Q as its past.

The orthogonal projection of $X(0,0)$ onto the linear subspace generated by the random variables with indices from the modified past $I = Q \cup K$ provides the best linear predictor of $X(0,0)$. However, due to non-orthogonality of Q and K computing this predictor is not easy even though it is known how to compute the projections of $X(0,0)$ onto Q and K separately. A natural way to alleviate this problem is to re-express I as the orthogonal sum of Q and a finite-dimensional subspace orthogonal to it. To this end, define

$$A = \text{sp}\{X(i,j) - \widehat{X}(i,j); (i,j) \in K\}, \tag{2.27}$$

where $\widehat{X}(i,j)$ is the orthogonal projection of $X(i,j)$ onto Q. The following results (outline of proof is given in Appendix A) provide the appropriate ingredients needed to compute

the predictors.

Lemma 4. Let $\{X(s_1, s_2)\}$ be a PND stationary random field with MA parameters $\{b_{k,\ell}\}$ and the modified past $I = Q \cup K$. Then,

(a). $X(0,0)$ is not in the modified set I, that is, $X(0,0) \notin I$.

(b). Q and A are orthogonal subspaces spanning I such that

$$I = Q \oplus A,$$

where A is as in (2.27).

(c). The orthogonal projection of $X(0,0)$ onto $I = Q \oplus A$ is given by

$$\widehat{X}_I(0,0) = P_I^{X(0,0)} = P_Q^{X(0,0)} + P_A^{X(0,0)} = \widehat{X}(0,0) + P_A^{X(0,0)},$$

where $P_Q^{X(0,0)}$ and $P_A^{X(0,0)}$ are the orthogonal projections of $X(0,0)$ onto Q and A, respectively, and $P_Q^{X(0,0)} = \widehat{X}(0,0)$ is obtained from Theorem 1 by substituting $(h_1, h_2) = (0,0)$.

(d). The orthogonal projection of $X(0,0)$ onto the finite-dimensional subspace A is given by

$$P_A^{X(0,0)} = \sum_{(i,j)\in K}\sum \beta_{i,j} \left\{ X(i,j) - \widehat{X}(i,j) \right\},$$

where the vector $\beta = \{\beta_{i,j};\ (i,j) \in K\}$, arranged using the lexicographic order of $\mathbb{Z}^2$, is given by

$$\beta = C^{-1} b, \tag{2.28}$$

and C is the variance-covariance matrix of the prediction errors in A with its elements given by (2.19), and $b = \{b_{i,j};\ (i,j) \in K\}$ is a vector of MA parameters with each $b_{i,j}$ given by (2.20).

Remark 2. The results in Lemma 4 are the counterparts of those for prediction problems in 1-D in stationary processes with modified past, see Pourahmadi (1989) and Pourahmadi et al. (2006). We now establish the prediction based on two different configurations of the set K in Lemma 5 and Theorems 2 & 3, and provide the details of the proof in Appendix A.

2.4.1 A Single Additional Observation

When the modified past has only the single additional observation $X(h_1, h_2)$, that is,

$$I_1 = Q \cup \{X(h_1, h_2)\} = Q \cup K, \tag{2.29}$$

with $h_i \geq 0, i = 1, 2$ and $(h_1, h_2) \neq (0, 0)$, then computing the prediction error variance for $\widehat{X}_{I_1}(0, 0)$ involves projecting onto a one-dimensional subspace. This is an analogue of the prediction of $X(0)$ when the past is modified to I_1 as in (2.5).

Theorem 2. Let $\{X(s_1, s_2)\}$ be a PND stationary random field with the innovation variance σ^2 and MA parameters $\{b_{k,\ell}\}$. Then, the best linear predictor of $X(0, 0)$ based on I_1 is given by

$$\widehat{X}_{I_1}(0, 0) = \widehat{X}(0, 0) + \beta_{h_1,h_2} \left(X(h_1, h_2) - \widehat{X}(h_1, h_2) \right), \tag{2.30}$$

where

$$\beta_{h_1,h_2} = \frac{b_{h_1,h_2}}{b^2_{h_1,h_2} + \sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} b^2_{k,\ell} + \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell} + \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell}}.$$

The corresponding prediction error variance is

$$\sigma^2(I_1) = \sigma^2 \frac{\sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} b^2_{k,\ell} + \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell} + \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell}}{b^2_{h_1,h_2} + \sum_{k=0}^{h_1-1} \sum_{\ell=h_2}^{\infty} b^2_{k,\ell} + \sum_{k=h_1}^{\infty} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell} + \sum_{k=0}^{h_1-1} \sum_{\ell=0}^{h_2-1} b^2_{k,\ell}}. \tag{2.31}$$

Remark 3. Note that if for a stationary random field $\{X(s_1, s_2)\}$ the MA parameter $b_{h_1,h_2} = 0$, then adding $X(h_1, h_2)$ to Q will have no effect on the prediction error variance of $X(0, 0)$ as in (2.6) for 1-D processes. However, the prediction error variance for 2-D processes in (2.31) involves infinite number of terms.

Remark 4. When the additional observation is in the second or fourth quadrant, the MA parameter $b_{h_1,h_2} = 0$ since $h_i < 0$ for $i = 1$ or 2. Therefore, augmenting the information set with an observation from second or fourth quadrant will not effect the prediction error variance of $X(0,0)$.

2.4.2 Several Additional Observations

Next we turn our attention to the prediction of $X(0,0)$ based on the knowledge of

$$I_2 = Q \cup \{X(i,j); 0 \le i \le h_1, 0 \le j \le h_2, (i,j) \ne (0,0)\} = Q \cup K, \tag{2.32}$$

where K is a finite set of n "future" observations in the first quadrant. This is the simplest form of the problem of finding the best linear predictor of $X(0,0)$ and its prediction error variance when a finite number of observations are added to Q. It is analogous to the *interpolation problem* in stationary 1-D processes to predict $X(0)$ based on the knowledge of I_2 from (2.5), see Nakazi (1984) and Pourahmadi (1989). In the next lemma, the vector of prediction errors with entries from the set A in (2.27) is written as a linear transformation of the innovation process.

Lemma 5. Let $\{X(s_1,s_2)\}$ be a PND stationary random field with innovations $\{\varepsilon(k,\ell)\}$, MA and AR parameters $\{b_{k,\ell}\}$ and $\{a_{k,\ell}\}$, respectively. Then,

(a). the vector of prediction errors for the observations in K based on the knowledge of Q can be expressed as:

$$X_K - \widehat{X}_K = b_K \varepsilon(0,0) + T' \varepsilon_K, \tag{2.33}$$

where $b_K = \{b_{k,\ell};\ (k,\ell) \in K\}$ is a vector of MA parameters corresponding to observations in the set K, ε_K consists of all the innovations $\{\varepsilon(k,\ell)\}$ involved in the prediction error of $X(h_1,h_2)$ based on Q. The terms in both b_K and ε_K are arranged lexicographically, and T is a rectangular matrix with n columns of MA parameters

for the observations in the set K. The MA parameters in each column are those involved in the prediction error of the observation in K.

(b). The variance-covariance matrix of the vector of prediction errors $X_K - \widehat{X}_K$ is

$$C = \sigma^2 \left(T'T + b_K b_K'\right). \tag{2.34}$$

(c). Let a_K be a vector of AR parameters with indices arranged as in ε_K. Then,

$$b_K' = T' a_K. \tag{2.35}$$

Remark 5. For 1-D stationary processes the matrix T involved in the prediction errors for observations in K, is a lower triangular, square and Toeplitz matrix of MA parameters, see Pourahmadi (2001, p.271).

Theorem 3. Let $\{X(s_1, s_2)\}$ be a PND stationary random field with the innovation variance σ^2, the MA and AR parameters $\{b_{k,\ell}\}$ and $\{a_{k,\ell}\}$, respectively. Then, the best linear predictor of $X(0,0)$ based on I_2 is

$$\widehat{X}_{I_2}(0,0) = \widehat{X}(0,0) + \beta'(X_K - \widehat{X}_K), \tag{2.36}$$

where $\widehat{X}(0,0)$ is the orthogonal projection of $X(0,0)$ onto Q and

$$\beta = G^{-1} b_K \left(1 + b_K' G^{-1} b_K'\right)^{-1}, \tag{2.37}$$

where $G = T'T$. The corresponding prediction error variance in terms of the MA parameters is

$$\sigma^2(I_2) = \sigma^2 \left(1 + b_K' G^{-1} b_K\right)^{-1}. \tag{2.38}$$

The prediction error variance can also be expressed in terms of the AR parameters as:

$$\sigma^2(I_2) = \sigma^2 \left(1 + a_K' P a_K\right)^{-1}, \tag{2.39}$$

where $P = T(T'T)^{-1}T'$ is a projection matrix.

Remark 6. Unlike the 1-D stationary process in (2.6), the prediction error variance in (2.39) based on the augmented set I_2 for random fields seems to be expressible only in terms of the MA parameters, and an attempt to express it in terms of the AR parameters runs into a mysterious projection operator P, which captures the nature of the "edge-effects" encountered in the estimation of random fields. Therefore, the results from 1-D and 2-D differ considerably because of the presence of the matrix P.

Corollary 3. If set I_2 in (2.32) is modified to include all the observations in the band between $X(0,0)$ and $X(h_1,h_2)$:

$$I_2 = \{X(i,j); i \leq h_1, j \leq h_2, (i,j) \neq (0,0), (h_1,h_2)\} = Q \cup K, \tag{2.40}$$

where both Q and K are infinite subspaces. Then the results in Theorem 3 for the prediction of $X(0,0)$ still hold true due to Corollary 1.

2.5 Worth of Observation

Suppose we are interested in quantifying the prediction worth of a generic information set denoted by $\mathcal{I}$, in terms of a meaningful measure of worth denoted by $W_{\mathcal{I}}$. Pourahmadi & Soofi (2000) introduced a measure of worth for observations in a stationary process when the purpose is prediction. We generalize this measure for the two-dimensional stationary random fields to propose an index of worth which quantifies the predictive worth of the observations in random fields.

To make this general set up more specific, imagine a situation in which two geologists A and B are interested in estimating $X(0,0)$ based on the knowledge of quarter plane past $Q = \{X(i,j); i \leq 0, j \leq 0 \text{ and} (i,j) \neq (0,0)\}$. Further, suppose that A has some additional information in the form of future observations;

$$\mathcal{I} = \{X(i,j); (r_1,r_2) \leq (i,j) \leq (0,0) \text{ and } (i,j) \neq (0,0)\},$$

for any $(r_1, r_2) > (0,0)$, then A is clearly at advantage as compared to B. In this case, A is actually predicting $X(0,0)$ based on the modified (amended) past $Q^+ = Q \cup \mathcal{I}$, while B is still computing the one-step ahead predictor for $X(0,0)$. The worth of observations in $\mathcal{I}$ gives a quantitative measure of how advantaged A is compared to B in presence of additional information.

Let $\widehat{X}_{Q^+}(0,0)$ denote the predictor of $X(0,0)$ based on the modified pasts Q^+ and let the prediction error variance be var $\left(X(0,0) - \widehat{X}_{Q^+}(0,0)\right) = \sigma^2\left(Q^+\right)$. Then we define the index of worth for set $\mathcal{I}$ as:

$$\mathcal{W}_{\mathcal{I}} = 1 - \frac{\sigma^2\left(Q^+\right)}{\sigma^2}, \text{ for } \mathcal{I} \subset Q^c.$$

The computation of this measure of worth requires $\sigma^2\left(Q^+\right)$, for modified sets Q^+. The results for the prediction error variance based on the modified pasts from Section 2.4 can be used here. From Theorem 2, where the interest was to predict $X(0,0)$ based on the augmented set $I_1 = Q \cup X(r_1, r_2)$, the set $\mathcal{I}$ consists of a single future observation $X(r_1, r_2)$. Then we can define the predictive worth of $\mathcal{I} = X(r_1, r_2)$ as:

$$\begin{aligned}\mathcal{W}_{\mathcal{I}} &= 1 - \frac{\sigma^2\left(I_1\right)}{\sigma^2} \\ &= \frac{b^2_{r_1,r_2}}{b^2_{r_1,r_2} + \sum_{k=r_1}^{\infty}\sum_{l=0}^{r_2-1} b^2_{k,l} + \sum_{k=0}^{r_1-1}\sum_{l=r_2}^{\infty} b^2_{k,l} + \sum_{k=0}^{r_1-1}\sum_{l=0}^{r_2-1} b^2_{k,l}}. \end{aligned} \tag{2.41}$$

Similarly, when a finite number of future observations are available then,

$$\mathcal{I} = \{X(i,j); (0,0) \leq (i,j) \leq (r_1, r_2), (i,j) \neq (0,0)\}, \tag{2.42}$$

as in Theorem 3 where a finite set K of future observations is augmented to Q such that the information set is $I_2 = Q \cup K$. Then we can find the worth of set $\mathcal{I} = K$ using the prediction error variance of the modified past I_2 as:

$$\begin{aligned}\mathcal{W}_{\mathcal{I}} &= 1 - \frac{\sigma^2\left(I_2\right)}{\sigma^2} \\ &= \frac{b'_K G^{-1} b_K}{1 + b'_K G^{-1} b_K}. \end{aligned} \tag{2.43}$$

Therefore, the change in prediction error variance due to addition of the observations from Q provides a way to measure the worth of observations in prediction. This measure of worth depends only on the MA parameters of the random field.

CHAPTER III

ESTIMATION OF STATIONARY RANDOM FIELDS

3.1 Introduction

In time series analysis the autoregressive moving average (ARMA) models:

$$X(t) = \sum_{r=1}^{p} \varphi_r X(t-r) + \varepsilon(t) + \sum_{r=1}^{q} \psi_r \varepsilon(t-r), \tag{3.1}$$

are used both in the time-domain for the purpose of forecasting and in the spectral-domain for parametric spectral density function estimation, see Brockwell & Davis (1991, Chap. 5, §4.4, §10.6). An alternative parametric model for the spectral density based on Bloomfield's (1973) exponential model of order p, EXP(p)

$$f(\lambda) = \frac{\sigma^2}{2\pi} \exp\left\{ 2 \sum_{r=1}^{p} \theta_r \cos(r\lambda) \right\}; \ \lambda \in (-\pi, \pi), \tag{3.2}$$

has a number of desirable statistical properties:

1. the parameters θ_i's have physical interpretation as the *cepstral* coefficients (Bogert, Healy &Tukey 1963, Chap. 15), and the estimated spectral density is guaranteed to be positive,

2. when estimating the parameters, the Hessian reduces to the identity matrix due to the orthogonality properties of sine and cosine functions which makes the maximization of the so-called Whittle Gaussian likelihood for the model relatively simple. Moreover, the parameter estimates are asymptotically uncorrelated in the sense that their Fisher information matrix is diagonal.

However, a drawback is that the construction of predictors for model (3.2) requires the optimal factorization of the spectral density function $f(\lambda)$ or finding the parameters of the moving average (MA) representation of the underlying process, see Section 2.2.

It seems Solo (1986) was the first to propose an extension of Bloomfield's (1973) EXP model for the two-dimensional stationary random fields as:

$$f(\lambda_1, \lambda_2) \approx \exp\left\{\sum_{r=-p}^{p}\sum_{s=-q}^{q}\theta_{rs}\cos(r\lambda_1 + s\lambda_2)\right\}; (\lambda_1, \lambda_2) \in (-\pi, \pi)^2. \tag{3.3}$$

The computational and statistical simplicity of models (3.2) and (3.3) are expected to be similar, so that the latter can enjoy the two properties listed earlier. However, to the best of our knowledge, this model has not been used effectively in the prediction of random fields observed on a lattice.

Based on the MA(∞) representation of the random field discussed in Section 2.3, we show that an important role is played by expressing the predictor coefficients in terms of the cepstrum of the random fields which amounts to extending the results in Pourahmadi (1983, 1984). Since the bivariate EXP model satisfy property (1), it provides a natural way to estimate the cepstral coefficients.

The outline of this chapter is as follows. In Section 3.2 we review the role of exponential models for prediction of stationary 1-D processes. In Section 3.3 we introduce the bivariate EXP model for the spectral density of a stationary spatial process and connect it to the classical linear regression. Section 3.4 provides an estimation methodology for the predictor coefficients by fitting exponential models. For a given data, we also put forward an algorithm to demonstrate the implementation of the proposed method. The procedure is illustrated using a simulation study and application to real data in Section 3.5.

3.2 Predictor Coefficients: Time Series

For a stationary time series $\{X(t), t \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers, the results in Lemma 1 provide linear least-squares predictors and prediction error variances of the future values based on the knowledge of infinite past and the spectral density of $\{X(t)\}$. Pourahmadi (1984) depicted the important role played by the formulas expressing the AR

and MA parameters in terms of the Fourier coefficients of $\log f(\lambda)$. These coefficients, known as *cepstral correlations/coefficients* of $f(\lambda)$ (Bogert et al. 1963), are given as:

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda)\exp(-ik\lambda)d\lambda. \tag{3.4}$$

The well-known, Szegö-Kolmogorov-Wiener formula for the innovation variance is related to c_0 as:

$$\sigma^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left\{f(\lambda)\right\} d\lambda\right].$$

The following recursive formulas for the MA and AR parameters in terms of the cepstral coefficients $\{c_k\}$ were derived in Pourahmadi (1983, 1984).

Lemma 6. Suppose that $\{X(t)\}$ is a purely nondeterministic, weakly stationary stochastic process with spectral density $f(\lambda)$, cepstral coefficients $\{c_k\}$, and spectral factor $\phi(z)$. Then, the Fourier coefficients of $\phi(z)$ and $\phi^{-1}(z)$ are given in terms of $\{c_k\}$ as:

(a) $b_k = \frac{1}{k}\sum_{j=0}^{k-1}(k-j)\,c_{k-j}b_j$,

(b) $a_k = \frac{-1}{k}\sum_{j=0}^{k-1}(k-j)\,c_{k-j}a_j$,

for $k = 0, 1, 2, \cdots$ with $b_0 = \exp(c_0/2)$ and $a_0 = \exp(-c_0/2)$.

Recursive formulas obtained later by Kaderli & Kayhan (2000) for the MA parameters using the cepstral coefficients of an ARMA process and by Hurvich (2002) for the coefficients $\{b_k\}$ and $\{a_k\}$ of the fractional exponential models, can be viewed as special cases of the Lemma 6. Bhansali (1974) used numerical factorization of the windowed estimates of $f(\lambda)$ to estimate the Kolmogorov-Wiener predictor coefficients when only a finite segment of the past of $\{X(t)\}$ is known. The performance of such estimators and their asymptotic properties have been studied by Bhansali (1973a, 1974, 1977) through Monte Carlo simulations.

It is clear from Lemma 6 that if cepstral coefficients of the spectral density are known, then the predictor coefficients can be computed easily. Bloomfield (1973) presented the least squares and maximum likelihood estimation of the cepstral coefficients by fitting EXP model.

3.3 Bivariate Exponential Models

The EXP model for the spectral density is motivated by observing that the log of the estimated spectral density is generally a smooth function, and thus can be approximated by a trigonometric polynomial (Bloomfield 1973). Consider a real-valued, zero mean, stationary random field $\{X(t_1, t_2), (t_1, t_2) \in \mathbb{Z}^2\}$ with summable autocovariance function $\{\gamma_{kl}\}$. Then its spectral density is,

$$f(\lambda_1, \lambda_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \gamma_{kl} \exp\{i(k\lambda_1 + l\lambda_2)\}.$$

If $\log f(\lambda_1, \lambda_2)$ is integrable then its Fourier series expansion is

$$\log f(\lambda_1, \lambda_2) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \theta_{rs} \exp\{i(r\lambda_1 + s\lambda_2)\}.$$

Truncating the above series at (p, q) we get,

$$\log f(\lambda_1, \lambda_2) \approx \sum_{r=-p}^{p} \sum_{s=-q}^{q} \theta_{rs} \exp\{i(r\lambda_1 + s\lambda_2)\}. \tag{3.5}$$

Using the symmetry of γ_{kl} and hence θ_{kl} we can write (3.5) as

$$\log f(\lambda_1, \lambda_2) \approx \theta_{00} + 2\sum_{r=1}^{p} \theta_{r0} \cos(r\lambda_1) + 2\sum_{s=1}^{q} \theta_{0s} \cos(s\lambda_2)$$
$$+2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(+)} \cos(r\lambda_1 + s\lambda_2) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(-)} \cos(r\lambda_1 - s\lambda_2), \tag{3.6}$$

with $R = 2pq + p + q + 1$ unknown parameters. Following Solo (1986), we express (3.6) as a classical linear regression model after replacing $f(\lambda_1, \lambda_2)$ by a suitable estimator. Consequently, all computational and inferential methods available for linear regression models can be used to fit EXP models to the spectral density.

3.3.1 Regression Formulation for Data on a Lattice

Let $\{X(t_1, t_2)\}$ be a stationary random field observed on a $n_1 \times n_2$ regular lattice with sufficiently smooth spectral density $f(.,.)$ which can be approximated by a bivariate EXP model as:

$$\log f(\lambda_{1j}, \lambda_{2k}) = \theta_{00} + 2\sum_{r=1}^{p} \theta_{r0} \cos(r\lambda_{1j}) + 2\sum_{s=1}^{q} \theta_{0s} \cos(s\lambda_{2k})$$
$$+2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(+)} \cos(r\lambda_{1j} + s\lambda_{2k}) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(-)} \cos(r\lambda_{1j} - s\lambda_{2k}), \tag{3.7}$$

where Fourier frequencies $\lambda_{1j} = (2\pi j/n_1)\,; 0 \leq j \leq (n_1 - 1)$ and $\lambda_{2k} = (2\pi k/n_2)\,; 0 \leq k \leq (n_2 - 1)$. Here $(j,k) \in A = \{(j,k) : 1 \leq j \leq m_1, -m_2 \leq k \leq m_2; j = 0, 1 \leq k \leq m_2\}$ and $m_i = [(n_i - 1)/2]$ for $i = 1, 2$, where $[.]$ is the greatest integer function.

A naive estimator for $f(.,.)$ is the periodogram of the random field. The periodogram for $\{X(t_1, t_2)\}$ is defined as:

$$I(\lambda_1, \lambda_2) = \frac{1}{(2\pi)^2 n_1 n_2} \left| \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} X(t_1, t_2) \exp\{-i(t_1\lambda_1 + t_2\lambda_2)\} \right|^2. \tag{3.8}$$

Brillinger (1974) provided a generalization for the asymptotic properties of the Fourier transforms of a spatial series. Under the assumption that all cumulants of the process $\{X(t_1, t_2)\}$ are bounded, he proved the asymptotic normality of the Fourier transform such that

$$\frac{I(\lambda_1, \lambda_2)}{f(\lambda_1, \lambda_2)} \sim \chi_2^2, \text{ as } \min(n_1, n_2) \to \infty, \tag{3.9}$$

with

$$E\left\{\frac{I(\lambda_{1j}, \lambda_{2k})}{f(\lambda_{1j}, \lambda_{2k})}\right\} = \gamma \text{ and var}\left\{\frac{I(\lambda_{1j}, \lambda_{2k})}{f(\lambda_{1j}, \lambda_{2k})}\right\} = \frac{\pi^2}{6}, \tag{3.10}$$

where $\gamma = -0.57721$ is the Euler's constant. Replacing $\log f$ in (3.7) with $I(.,.)$ from (3.8)

and using the property (3.10) we get,

$$\log I(\lambda_{1j}, \lambda_{2k}) + \gamma = \theta_{00} + 2\sum_{r=1}^{p} \theta_{r0} \cos(r\lambda_{1j}) + 2\sum_{s=1}^{q} \theta_{0s} \cos(s\lambda_{2k})$$
$$+2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(+)} \cos(r\lambda_{1j} + s\lambda_{2k}) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(-)} \cos(r\lambda_{1j} - s\lambda_{2k}) + \varepsilon_{jk}. \quad (3.11)$$

Representing (3.11) as a classical linear regression model of the form:

$$Y = X\beta + \varepsilon, \quad (3.12)$$

where for $(j,k) \in A$ the $N \times 1$ vector, $Y = \{y_{jk}\} = \{\log I(\lambda_{1j}, \lambda_{2k}) + \gamma\}$ with $N = 2m_1m_2 + m_1 + m_2 + 1$ columns and $N \times R$ design matrix, $X = \{x_{jk}^{'}\}$ has N rows, R columns with its (jk)-th row given as:

$$x_{jk} = [1, 2\cos(r\lambda_{1j}), 2\cos(s\lambda_{2k}), \cos(r\lambda_{1j} + s\lambda_{2k}), \cos(r\lambda_{1j} - s\lambda_{2k})]', \quad (3.13)$$

where $r = (1, 2, \cdots, p)'$ and $s = (1, 2, \cdots, q)'$. The $R \times 1$ vector of parameters, β is

$$\beta = [\theta_{00}, \theta_{10}, \cdots, \theta_{p0}, \theta_{01}, \cdots, \theta_{0q}, \theta_{11}^{+}, \cdots, \theta_{pq}^{+}, \theta_{11}^{-}, \cdots, \theta_{pq}^{-}]', \quad (3.14)$$

and $\varepsilon = \{\varepsilon_{jk}\}$ is the $N \times 1$ error vector. From regression model (3.12), $\{\theta_{rs}\}$ is estimated using the least squares as, $\hat{\beta} = (X'X)^{-1}X'Y$ with variance-covariance matrix for the estimates given as var$(\hat{\beta}) = \sigma^2(X'X)^{-1}$.

An important issue closely related to the parameter estimation for spatial models is the problem of *edge effects* caused by the data points on the boundary since the neighborhood is not completely observed at these locations. In time series data there are only two boundary points, corresponding to $t = 0$ and n, whereas in random fields the number of boundary points increases with the dimension. An alternative to overcome the bias introduced in the estimation due to edge effects is to use the periodogram based on the tapered data (Dalhaus & Künsch 1987). A tapered periodogram for random field $\{X(t_1, t_2)\}$ is defined as

$$I_T(\lambda_1, \lambda_2) = |H|^{-1} \left| \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} h_{t_1,t_2} X(t_1, t_2) \exp(-i(t_1\lambda_1 + t_2\lambda_2)) \right|^2, \quad (3.15)$$

where h_{t_1,t_2}, is a two-dimensional taper function and $H = \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} h_{t_1,t_2}^2$. We note that fitting an EXP model to estimate spectral density also removes edge effects to provide a smooth estimate. This is illustrated in Section 3.5.

It turns out that the performance of the proposed estimators depend on the model order of the EXP model fitted to $f(.,.)$. The choice of (p,q) is in accordance with the order selection in an ARMA(p,q) model, see Brockwell & Davis (1991, p. 301). Different ways to select optimal (p,q) are discussed in Section 4.4. Here, we use the generalized form of the corrected Akaike's information criteria (AICC) proposed by Hurvich & Tsai (1989) for model selection in the context of regression and time series models. For regression equation (3.12) with EXP model of order (p,q), the AICC is defined as:

$$AICC(p,q) = |A|\log\left(\frac{RSS_{p,q}}{|A|}\right) + \frac{2R|A|}{|A|-R-1}, \tag{3.16}$$

where $RSS_{p,q}$ is the residual sum of squares, R is the number of parameters and $|A|$ is the cardinality of set A. From all possible pairs (p,q) the pair which minimizes the AICC is selected. For more details and other model order selection methods, see Section 4.4

3.4 Predictor Coefficients: Random Fields

In Theorem 1, the linear least squares predictor for observations in Q^c and its prediction error variance are expressed in terms of the AR and MA parameters, respectively. Hence it is required to estimate these parameters to get hold of the predictions and prediction error variance. In this section, we present an extended version of Lemma 6 for stationary random fields which provides recursive relation between the AR and MA parameters with the cepstral coefficients of the random field.

From (2.15) and (2.25) we have

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell} z_1^k z_2^\ell = \exp\left(\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k,\ell} z_1^k z_2^\ell\right), \text{and} \tag{3.17}$$

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} a_{k,\ell} z_1^k z_2^\ell = \exp\left(-\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} c_{k,\ell} z_1^k z_2^\ell\right), \tag{3.18}$$

which are used to get the following result.

Theorem 4. Suppose $\{X(t_1,t_2)\}$ is a real-valued, stationary process on the quarter-plane Q with spectral density function $f(\lambda_1,\lambda_2)$, cepstral coefficients $\{c_{k,\ell}\}$, and optimal spectral factor $\phi(z_1,z_2)$ such that the conditions of Lemma 3 hold. Then, the MA and AR parameters of $\{X(t_1,t_2)\}$ are given in terms of its cepstral coefficients as:

(a). $b_{k,0} = \frac{1}{k}\sum_{j_1=0}^{k-1}(k-j_1)c_{k-j_1,0}b_{j_1,0}$,

(b). $b_{0,\ell} = \frac{1}{\ell}\sum_{j_2=0}^{\ell-1}(\ell-j_2)c_{0,\ell-j_2}b_{0,j_2}$,

(c). $b_{k,\ell} = \frac{1}{k\ell}\sum_{j_1=0}^{k-1}\sum_{j_2=0}^{\ell-1}\Big\{(k-j_1)(\ell-j_2)c_{k-j_1,\ell-j_2}b_{j_1,j_2}+(\ell-j_2)(j_2+1)c_{k-j_1,\ell-(j_2+1)}b_{j_1,j_2+1}\Big\}$,

(d). $a_{k,0} = \frac{-1}{k}\sum_{j_1=0}^{k-1}(k-j_1)c_{k-j_1,0}a_{j_1,0}$,

(e). $a_{0,\ell} = \frac{-1}{\ell}\sum_{j_2=0}^{\ell-1}(\ell-j_2)c_{0,\ell-j_2}a_{0,j_2}$,

(f). $a_{k,\ell} = \frac{-1}{k\ell}\sum_{j_1=0}^{k-1}\sum_{j_2=0}^{\ell-1}\Big\{(k-j_1)(\ell-j_2)c_{k-j_1,\ell-j_2}a_{j_1,j_2}+(\ell-j_2)(j_2+1)c_{k-j_1,\ell-(j_2+1)}a_{j_1,j_2+1}\Big\}$,

for $k,\ell = 1,2,\cdots$, $b_{0,0} = \sigma$ and $a_{0,0} = 1/\sigma$. The proof of Theorem 4 is derived in Appendix B.

These recursive relations provide a way to estimate the MA and AR parameters using least squares estimate for the cepstral coefficients as explained in the Section 3.3.1.

Remark 7. The MA and AR parameters of a causal and invertible (stationary) random field decay at an exponential rate, see Yao & Brockwell (2006).

The steps involved in the computation of the estimates for the linear least predictors and their prediction error variance are summarized in the following algorithm:

Exponential Model Based Prediction Algorithm

For a given data observed on a $n_1 \times n_2$ lattice,

1. Fit a bivariate exponential model, EXP(p, q) for the spectral density of the data, where the model order (p, q) is selected using AICC given in (3.16).

2. Using the fitted EXP(p, q) model compute the least squares estimates for the cepstral coefficients, $\{\widehat{c}_{k,\ell}\}$ as explained in Section 3.3.1.

3. An estimate for σ^2 is computed from $\widehat{c}_{0,0}$ as:

$$\widehat{\sigma}^2 = 4\pi^2 \exp(c_{0,0}).$$

4. Using estimates $\{\widehat{c}_{k,\ell}\}$ from step 2 and the recursive relations from Theorem 4, estimate MA and AR parameters. We compute $\{\widehat{b}_{k,\ell}\}$ and $\{\widehat{a}_{k,\ell}\}$ for $k = 1, \cdots, L_1$ and $\ell = 1, \cdots, L_2$, where mostly large enough L_1 and L_2 are sufficient for the purpose of computation since only first few coefficients are significant, see Remark 7.

5. Substitute $\{\widehat{b}_{k,\ell}\}$ and $\{\widehat{a}_{k,\ell}\}$ in Theorem 1 to obtain predictor and prediction error variance for $X(t_1 + h_1, t_2 + h_2)$. Note that the double infinite sums in the Theorem 1 are truncated at (L_1, L_2) for computation.

3.5 Simulations and Data Analysis

In this section, we illustrate the performance of the proposed methodology for estimation of the predictor coefficients through simulation study and real data analysis.

3.5.1 Simulation Study

Data is generated from a spatial ARMA model of the form:

$$X(t_1,t_2)+\sum_{\substack{i=0\\(i,j)\neq(0,0)}}^{p_1}\sum_{j=0}^{p_2}\varphi_{i,j}X(t_1-i,t_2-j)=\sum_{k=0}^{q_1}\sum_{l=0}^{q_2}\psi_{k,l}\varepsilon(t_1-k,t_2-l), \tag{3.19}$$

where $\varphi_{i,j}$ and $\psi_{k,l}$ are AR and MA coefficients, respectively and $\varepsilon(t_1,t_2)$ is Gaussian white noise process with mean 0 and variance σ^2. The exponential model based algorithm is used to obtain $\widehat{X}(t_1+h_1,t_2+h_2)$ and its prediction error variance. We repeat steps (1)-(5) of the algorithm for 200 replications and report the mean and variance of the prediction errors, $(X-\widehat{X})$ obtained from these replications. The prediction error variance is also computed for each replication.

Kizilkaya & Kayran (2005) proposed an estimation method for the AR and MA coefficients of ARMA models using the cepstral coefficients . Substituting these estimates in model (3.19) we compute forecast for $X(t_1+h_1,t_2+h_2)$ denoted by $\tilde{X}(t_1+h_1,t_2+h_2)$ and its prediction error, $\left(X-\tilde{X}\right)$ for 200 replications. We compare the mean and variance of these prediction errors with those obtained from our algorithm. An important point to note here is that the estimates provided by Kizilkaya & Kayran (2005) are based on the assumption that the true ARMA model order is known, which results in nearly unbiased estimates for the AR and MA coefficients. Hence, the values obtained from estimated ARMA model are very close to those from the true model and the prediction errors are very small.

We consider three different ARMA models to compare the performance of EXP and estimated ARMA models based predictions. We fixed $\sigma^2=1$, $(n_1,n_2)=(64,64)$, $L_1=L_2=100$ and computed prediction errors for
$(h_1,h_2)=\{(1,1),(1,2),(1,3),(2,1),(2,2),(3,3),(3,1),(3,2),(3,3)\}$.

Model 1: ARMA(1,1,1,1)

$$X(t_1,t_2) - 0.13X(t_1,t_2-1) - 0.05X(t_1-1,t_2) - 0.06X(t_1,t_2-1) =$$
$$Z(t_1,t_2) + 0.32Z(t_1,t_2-1) + 0.15Z(t_1-1,t_2) + 0.305Z(t_1-1,t_2-1).$$

Figure 1(a) shows the (true) spectral density of the above model, and Figure 1(b)-(d) exhibits periodogram, fitted spectral density obtained using the EXP model and estimated ARMA model, averaged over 200 realizations. Plot (b) shows that the periodogram is not estimating density very well, whereas (c) and (d) are close to the true density. Note that the estimate based on the EXP model is less constraint as it makes no assumption regarding the ARMA model order and selects (p,q) using the data-driven approach (AICC here).

Model 2: ARMA(1,1,1,1)

$$X(t_1,t_2) - 0.48X(t_1,t_2-1) - 0.285X(t_1-1,t_2) - 0.18X(t_1,t_2-1) =$$
$$Z(t_1,t_2) + 0.42Z(t_1,t_2-1) + 0.25Z(t_1-1,t_2) + 0.136Z(t_1-1,t_2-1).$$

Figure 2(a) shows that the (true) spectral density of the above model has a significant peak at $(\lambda_1,\lambda_2) = (0,0)$, (b) shows that the periodogram provides a poor estimate, and the fitted spectral density using EXP and estimated ARMA model in (c) and (d) provide reasonable estimates. The roots of this ARMA model are close to the unit circle and to circumvent this Kizilkaya & Kayran (2005) proposed to use the MA residuals in place of the original data. The MA residuals, $\{Z(t_1,t_2)\}$ are obtained by applying the inverse filtering to the data by using estimated AR coefficients as $Z(t_1,t_2) = X(t_1,t_2) + \sum_{\substack{k=0\\(k,l)\neq(0,0)}}^{1}\sum_{l=0}^{1}\widehat{a}_{k,l}X(t_1-k,t_2-l)$.

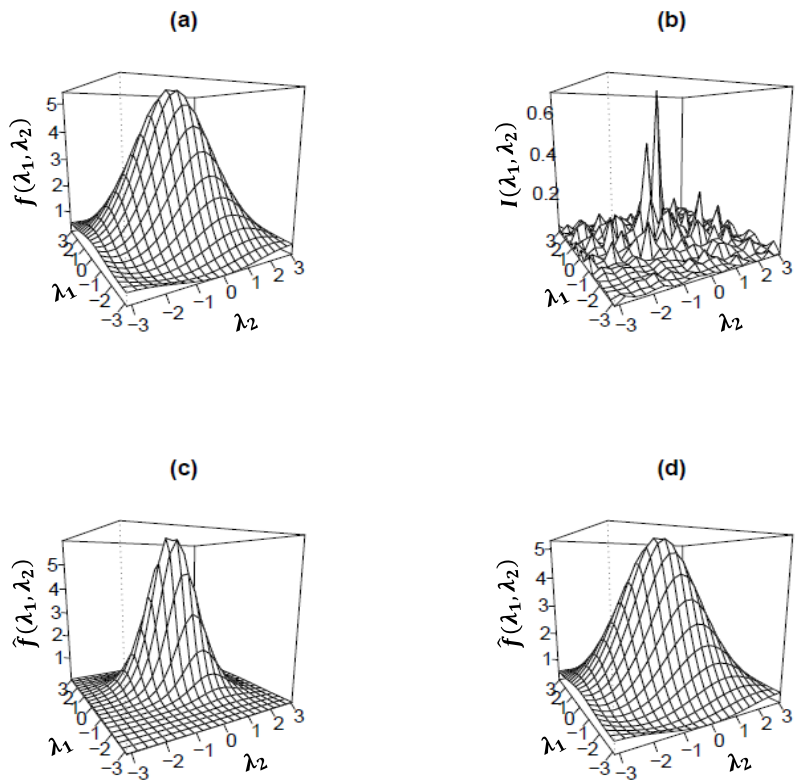

Figure 1: Model 1: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA.

Model 3: ARMA(2,2,2,2)

$$X(t_1,t_2)+0.15X(t_1,t_2-1)+0.06X(t_1,t_2-2)+0.085X(t_1-1,t_2)-0.1X(t_1-1,t_2-1)+0.053X(t_1-1,t_2-2)-0.05X(t_1-2,t_2)+0.13X(t_1-2,t_1-1)+0.115X(t_1-2,t_2-2)=Z(t_1,t_2)+0.20Z(t_1,t_2-1)+0.23Z(t_1,t_2-2)0.15Z(t_1-1,t_2)+0.18Z(t_1-1,t_2-1)+0.16Z(t_1-1,t_2-2)+0.175Z(t_1-2,t_2)+0.24Z(t_1-2,t_1-1)+0.30Z(t_1-2,t_2-2).$$

The true spectral density, periodogram, fitted density using EXP and estimated ARMA model given in Figure 3(a)-(d) show that the estimate obtained using EXP model is reason-

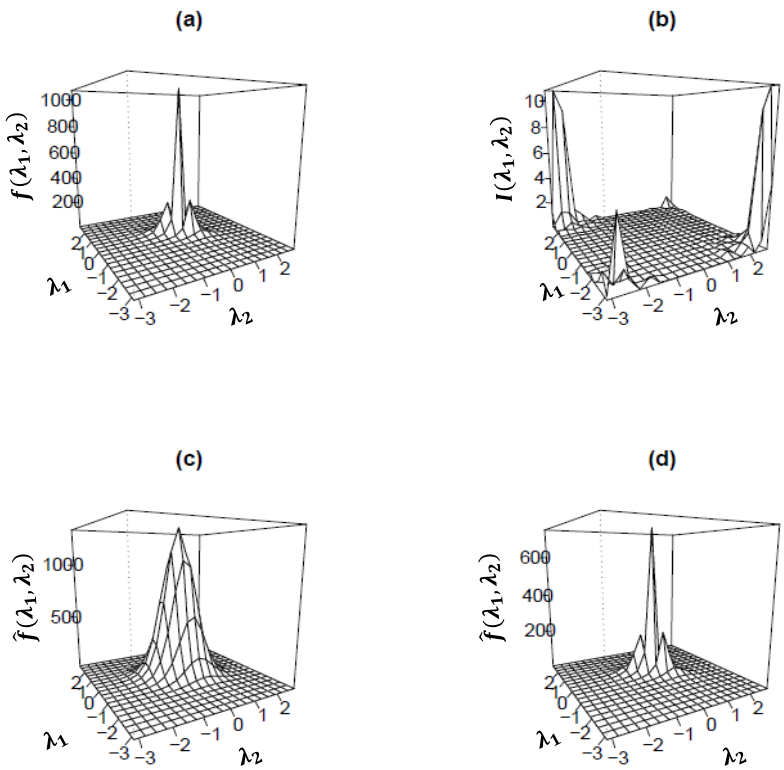

Figure 2: Model 2: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA.

able.

Kizilkaya & Kayran (2005) have also used these three ARMA models. For each case the model order for EXP(p, q) is selected using AICC from all possible pairs in $\{(1,1) \leq (p,q) \leq (5,5)\}$. The mean and variance of p and q from 200 replications reported in Table 1 show that a small value of 2 and 3 is sufficient for both p and q for all models.

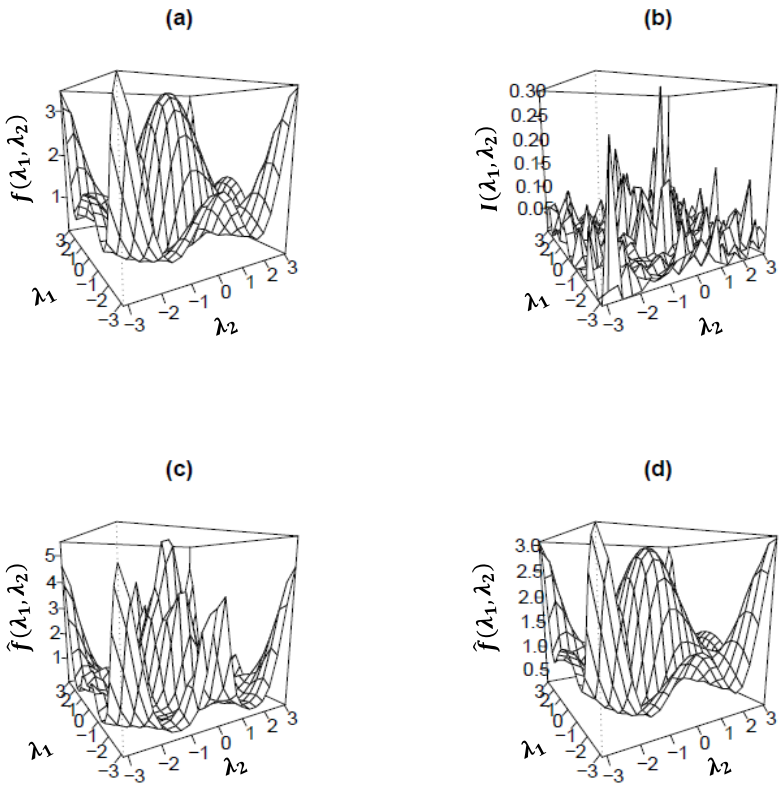

Figure 3: Model 3: Plot for (a) True Spectral Density (b) Periodogram, (c) Exponential Model, (d) Estimated ARMA.

Table 1: Mean and Variance of the (p, q) Values Selected Using AICC.

ARMA Model	p(var)	q(var)
1	2.01(0.03)	2.06(0.07)
2	2.73(0.25)	3.40(0.37)
3	3.18(0.19)	3.10(0.24)

The prediction results are given in Table 2 which includes the mean and variance of the prediction errors obtained using EXP model and estimated ARMA model based

predictions. Since the estimated ARMA model is similar to the true model, so its prediction errors and variances are expected to be small as seen in Table 2. These results show that the prediction errors and variances from the EXP model are comparable to those from the ARMA models. For Model 2 the prediction errors obtained from EXP model were quite poor as compared to ARMA model based errors (values not reported here) if $\{X(t_1, t_2)\}$ was used, however the prediction results obtained by using the MA residuals $\{Z(t_1, t_2)\}$ are reasonable as reported in Table 2. We also computed predictions using MA residuals for other models 1 and 3 also, since the results were not too different as compared to those from the original data, therefore, those values are not reported here. These results also show that the (empirical) variance for the prediction errors increases with (h_1, h_2).

Table 2: Predictions and Prediction Error Variances Obtained Using EXP Algorithm and Estimated ARMA Model.

		Model 1				**Model 2**				**Model 3**			
		EXP(p,q)		ARMA		EXP(p,q)		ARMA		EXP(p,q)		ARMA	
h_1	h_2	$(X-\widehat{X})$	PEV	$(X-\tilde{X})$	PEV	$(X-\widehat{X})$	PEV	$(X-\tilde{X})$	PEV	$(X-\widehat{X})$	PEV	$(X-\tilde{X})$	PEV
1	1	-0.03	1.28	-0.01	0.97	-0.36	1.52	-0.05	1.13	-0.002	1.46	-0.005	1.09
1	2	-0.06	1.48	-0.05	1.43	-0.13	1.79	-0.08	1.40	-0.03	1.63	-0.009	1.45
1	3	-0.09	1.67	-0.04	1.55	-0.19	1.82	-0.04	1.47	0.04	1.53	-0.02	1.36
2	1	-0.07	1.21	-0.05	1.35	-0.72	1.74	-0.12	1.22	-0.05	1.84	-0.04	1.18
2	2	-0.12	1.71	0.05	1.72	-0.32	1.83	-0.02	1.58	-0.09	1.68	-0.06	1.44
2	3	-0.15	1.98	-0.03	1.44	-0.51	2.09	-0.03	1.52	-0.11	1.97	-0.07	1.29
3	1	-0.10	1.86	-0.06	1.42	-0.19	1.80	-0.11	1.34	-0.06	1.78	-0.06	1.38
3	2	-0.16	1.97	-0.04	1.41	-0.62	2.04	-0.07	1.29	-0.13	2.07	-0.08	1.46
3	3	-0.18	2.34	-0.11	1.34	-0.92	2.42	-0.07	1.14	-0.17	2.19	0.14	1.17

3.5.2 Data Analysis

In this section we rely on the classical Mercer & Hall (1911, p.232) wheat yield data to compare the performance of the EXP model based predictions with those obtained using Kriging. A wheat yield experiment was conducted at Rothamsted Experimental Station in Great Britain. The experiment, a uniformity trial, consisted of giving a 20 by 25 lattice of plots the same treatment with approximately 1 acre area under each plot. Yield of grains were measured in pounds. On the 20 by 25 layout each of the 20 rows runs in the east-west direction and each of the 25 columns runs in the north-south direction. The exact size of the plots from the original data set seems to be unknown, although some researchers have used 3.30 meters east to west, and 2.51 meters north to south. This dataset has been broadly studied by many authors Whittle (1954), Cressie (1993, p.248) and Young & Young (1998). The dataset is available in the R library `(spdep)` using simple command `data(wheat)`. The data is shown in Figure 4(a) and its histogram in Figure 4(b) show the familiar bell-shaped curve, indicating the nearly normal distribution for the 500 wheat yield measurements. The mean (3.95), median (3.94) and mode (3.97) are nearly equal, and the skewness (.036) and kurtosis (-.254) are close to zero, all providing evidence that the distribution is close to normal. Cressie (1993, pp. 284-259) conducted exploratory data analysis of this data in which he confirmed the presence of an irregular east-west trend in the data. The trend effect is removed by applying median polishing, an exploratory data analysis technique proposed by Tukey (1977), to the data. The residuals of the median polished data are presented in Figure 4(c) along with its histogram in (d). Using the empirical variogram of these residuals Cressie (1993) showed that an exponential model is suitable for modeling the covariance.

We also use these residuals for prediction. To compute $X(t_1 + h_1, t_2 + h_2)$, first divide the data into two parts: observed and unobserved. The observed part consists of

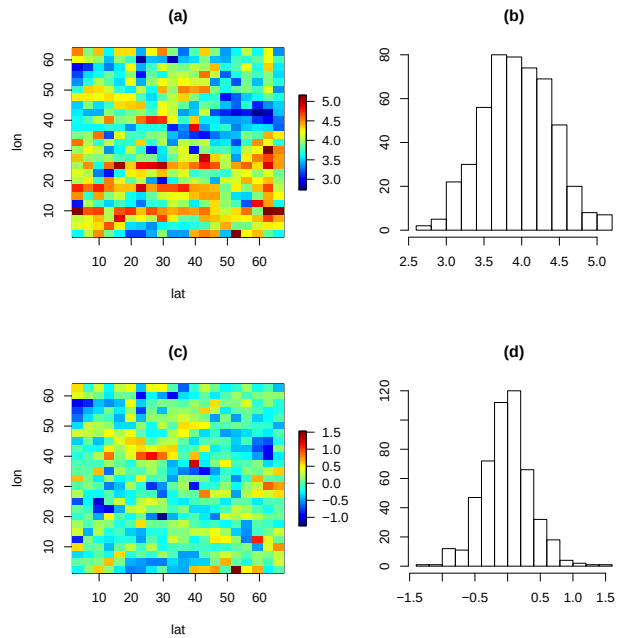

Figure 4: Wheat Yield Data: (a) Spatial Plot and (b) Histogram. Median Polish Residuals: (c) Spatial Plot and (d) Histogram.

$(20 - h_1) \times (25 - h_2)$ out of 20×25 and the remaining rows and columns constitute the unobserved part. Following steps (1)-(5) of the EXP based algorithm we compute $\widehat{X}(t + 1 + h_1, t_2 + h_2)$, and the prediction error variance for different (h_1, h_2). For the original data, Solo (1986) suggested $(p, q) = (2, 10)$ as a suitable EXP model in order to ensure the separability of the field. Using AICC we obtain $(p, q) = (4, 1)$ for the original data and $(p, q) = (2, 1)$ for the residuals. The (p, q) values selected for $(20-h_1)\times(25-h_2)$ residuals are reported in Table 3 for different (h_1, h_2) values.

To perform prediction using Kriging on the residuals we used exponential covariance structure as suggested by Cressie (1993). In Table 3 we report the residual value at $(20, 25)$ and its estimates using EXP model and Kriging along with the corresponding prediction error variances. These results show that for small (h_1, h_2), that is, when we are predict-

ing not too far from the observed values, the EXP model based method outperforms the Kriging. Whereas as (h_1, h_2) increase, that is prediction is for a far away location, both the prediction error and its variance increase and Kriging has smaller bias and variance as compared to the EXP model based predictions.

Table 3: Predictions and Prediction Error Variances for Wheat Yield Data Using EXP Algorithm and Kriging. (p, q) is the model order for the EXP model selected using AICC.

			EXP			Kriging	
h_1	h_2	$X(20,25)$	$\widehat{X}(20,25)$	PEV	(p,q)	$\tilde{X}(20,25)$	PEV
1	1	-0.386	-0.387	0.274	(1,1)	-0.411	0.338
1	2	-0.386	-0.378	0.298	(1,1)	-0.379	0.362
1	3	-0.386	-0.395	0.317	(1,1)	-0.214	0.400
2	1	-0.386	-0.382	0.306	(1,1)	-0.359	0.374
2	2	-0.386	-0.394	0.323	(2,1)	-0.373	0.377
2	3	-0.386	-0.398	0.378	(2,1)	-0.311	0.397
3	1	-0.386	-0.398	0.403	(1,1)	-0.401	0.445
3	2	-0.386	-0.411	0.421	(1,1)	-0.514	0.436
3	3	-0.386	-0.414	0.453	(1,1)	-0.358	0.449

CHAPTER IV

LONG RANGE DEPENDENT RANDOM FIELDS

4.1 Introduction

Long range dependent random fields are characterized by spectral density function which is unbounded at certain frequencies including zero frequency. Alternatively, a random field whose covariance function tends to zero like a power function and so slowly that their sums diverge also exhibits long range dependence. These two notions of long range dependence are closely related but not equivalent, and we will focus on the spectral domain approach here. The motivation to study long range dependence in random fields comes from many areas of applications like agricultural field trials in which there is empirical evidence of slow decay of correlations between yield, see Whittle (1956, 1962), Pearce (1976) and Martin (1986). This led to the study of power law correlation functions by Whittle (1962) and Besag (1981). More applications include spatial variability of soil properties, air ozone concentration of ocean temperature, see Anh, Lam, Leung & Tieng (2000), Akkaya & Yucemen (2002) and Lim, Kim & Lee (2002).

Lavancier (2005) gave many possible definitions for long range dependent random fields. A definition parallel to the long range dependent time series (1-D processes) characterizes a stationary 2-D random field as isotropic long range dependent if its spectral density admits the form:

$$f(\boldsymbol{\lambda}) \sim |\boldsymbol{\lambda}|^{-2\alpha} L\left(\frac{1}{|\boldsymbol{\lambda}|}\right) b\left(\frac{\boldsymbol{\lambda}}{|\boldsymbol{\lambda}|}\right); \;\; 0 < \alpha < 1, \tag{4.1}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$, $L(.)$ is a slowly varying function at infinity and b is continuous on the unit sphere in $\mathbb{R}^2$. Here α controls the behavior of $f(.,.)$ near the zero frequency and is termed as the *long range dependence parameter*. In this case, the behavior of the covariance

function at infinity is known and is given as

$$\gamma(h) \sim |h|^{-2\alpha} L(|h|)\, b\left(\frac{h}{|h|}\right), \;\; 0 < \alpha < 1, \;\; h \to \infty, \tag{4.2}$$

where L and b are same as above. Many authors (see Section 4.2) considered models which satisfy the isotropic long range dependence property (4.1) and their spectral density is characterized by singularity at zero frequency as:

$$f(\boldsymbol{\lambda}) = \left|1 - e^{-i\boldsymbol{\lambda}}\right|^{-2\alpha} f^*(\boldsymbol{\lambda}), \;\; 0 < \alpha < 1, \tag{4.3}$$

where $f^*(\boldsymbol{\lambda})$ is a even, positive, continuous function on the torus $[-\pi, \pi]^2$, bounded above and away from zero.

Two ways to construct long range dependent time series include filtering the white noise process through unbounded filters and aggregation of autoregressive moving average (ARMA) processes, see Brockwell & Davis (1991). Lavancier (2005) generalized these methods for d-dimensional random fields with $d > 1$. He showed that the covariance function of the resulting processes have a rather closed form. We provide examples for long range dependent random fields based on filtering and aggregation, for details see Lavancier (2005, 2011).

Filtering

An extension of the stationary fractional autoregressive integrated moving average (FARIMA) process for any $k \in \mathbb{Z}^+$ is defined as:

$$\left(1 - B_1 B_2^k\right)^{\alpha} X(s_1, s_2) = \varepsilon(s_1, s_2), \;\; 0 < \alpha < 1/2,$$

where B_1 and B_2 are the backward shift operators in the horizontal and vertical direction, respectively such that $B_1 X(s_1, s_2) = X(s_1 - 1, s_2)$ and $B_2 X(s_1, s_2) = X(s_1, s_2 - 1)$. The spectral density of $X(s_1, s_2)$ is

$$f(\lambda_1, \lambda_2) = \frac{\sigma^2}{4\pi^2} \left|1 - \mathrm{e}^{-i(\lambda_1 + k\lambda_2)}\right|^{-2\alpha}, \tag{4.4}$$

which is unbounded all over the line $\lambda_1 + k\lambda_2 = 0$. The function fails to satisfy (4.1) and exhibits non-isotropic long range dependence. The result given in Brockwell & Davis (1991) for time series easily generalize to give the correlation function for $X(s_1, s_2)$ as:

$$\begin{cases} \rho(h, kh) = \prod_{0<j\leq h} \dfrac{j-1+\alpha}{j-\alpha}; h = \pm 1, \pm 2, \cdots \\ \rho(k, l) = 0, l \neq kh. \end{cases}$$

The correlation function ρ is non-summable when $l = kh$ due to its proportionality to $h^{2\alpha-1}$. As an example, consider the spectral density of the form

$$f(\lambda_1, \lambda_2) = \left|1 - \mathrm{e}^{-i(\lambda_1 + k\lambda_2)}\right|^{-2\alpha} f^*(\lambda_1, \lambda_2), \tag{4.5}$$

where the function $f^*(.,.)$ is bounded above and away from zero.

Aggregation

Another definition based on the aggregation leads to a long range dependent random field with spectral density unbounded when either $\lambda_1 = 0$ or $\lambda_2 = 0$ or both are zero. The spectral density is the tensorial product of two 1-D spectral densities given as:

$$f(\lambda_1, \lambda_2) \sim c\,|\lambda_1|^{-2\alpha_1}\,|\lambda_2|^{-2\alpha_2}\,;\ \text{when } (\lambda_1, \lambda_2) \to (0, 0), \tag{4.6}$$

where c is a positive constant and $0 < \alpha_1, \alpha_2 < 1/2$. Notice that long range behavior depends on the direction including parameter α_1 for the "horizontal" direction and α_2 for the "vertical" direction. For instance, a spectral density of the form

$$f(\lambda_1, \lambda_2) = \left|1 - \mathrm{e}^{-i\lambda_1}\right|^{-2\alpha_1} \left|1 - \mathrm{e}^{-i\lambda_2}\right|^{-2\alpha_2} f^*(\lambda_1, \lambda_2), \tag{4.7}$$

with $f^*(.,.)$ as the short run behavior function which is bounded above and away from zero. Some authors have considered (4.7) and modelled $f^*(.,.)$ using fully parametric models like AR or ARMA, as discussed in Section 4.2.

In long range dependent random fields, a common scientific goal is to estimate the parameter which controls the behavior of the spectral density at unbounded frequencies,

subjected to only a single long range dependence parameter, α in models like (4.3) and (4.5) or two different long range dependence parameters, α_1 and α_2, one for each direction as in model (4.7). Current methods can be categorized into two broad classes based on the different ways of dealing with the short range dependence component $f^*(.,.)$. The first is *parametric* approach which assumes a finite dimensional parametric model for $f^*(.,.)$ and then employ either least squares or the Gaussian/Whittle maximum likelihood method to estimate the long range dependence and the model parameters. Sometimes, the data under study suggests a suitable parametric function which offers meaningful interpretation of the study. Still, for many situations there may not be sufficient knowledge to warrant such a parametric function and concerns of model misspecification naturally arise. To alleviate those issues more flexible *semi-parametric* modelling and estimation of the spectral density becomes essential. In semi-parametric approach $f^*(.,.)$ is not required to obey any parametric model or the model can be misspecified. Wang (2009) and Guo, Lim & Meerschaert (2009) used a semi-parametric approach to estimate α in model (4.3) by treating $f^*(.,.)$ as a constant and considering frequencies only in the neighborhood of zero. However, this assumption of constant behavior at frequencies far from zero may not be satisfied for some real data and might lead to inconsistent estimates.

Due to shortcomings of the existing estimators, we take a semi-parametric approach and cast this problem in a *broadband* framework which includes all the nonzero frequencies. Given a specific definition for long range dependence, we propose to use a bivariate exponential model (Solo 1986) for $f^*(.,.)$. The proposed semi-parametric estimator is efficient, robust to model misspecification and can be implemented for different definitions of long range dependent random fields. It turns out that the statistical properties of the proposed estimator depends on the model order of the exponential model used for $f^*(.,.)$. In this work, we discuss data-driven choice of the model order which minimizes the root mean squared error (RMSE) of the estimator for the long range dependence parameter.

We generalize two model order selection criteria proposed for 1-D processes: Mallow's C_L (Hurvich 2001) and corrected Akaike's information criteria (AICC) (Hurvich & Tsai 1989). We also consider a subset selection approach based on Akaike's information criteria (AIC) (Akaike 1974) to select only a subset of significant variables in exponential model. These techniques can be used to automatically select the appropriate model for $f^*(.,.)$ without any additional modelling assumptions. We demonstrate the performance of the proposed semi-parametric estimator through simulation studies and an application to the real data.

4.2 Existing Estimators

4.2.1 Parametric Approach

Boissy, Bhattacharyya, Li & Richardson (2005) considered spatial autoregressive (AR) model (Martin 1979) for $f^*(.,.)$ in (4.7) such that the spectral density is

$$f(\lambda_1,\lambda_2) = \frac{\sigma^2}{4\pi^2}\left|1-\mathrm{e}^{-i\lambda_1}\right|^{-2\alpha_1}\left|1-\mathrm{e}^{-i\lambda_2}\right|^{-2\alpha_2}\left|\phi(\mathrm{e}^{-i\lambda_1},\mathrm{e}^{-i\lambda_2},\varphi_1,\varphi_2)\right|^{-2}, \qquad (4.8)$$

where $-1/2 < \alpha_1, \alpha_2 < 1/2$ and $\phi(z_1, z_2, \varphi_1, \varphi_2) = (1-\varphi_1 z_1)(1-\varphi_2 z_2)$ is the polynomial in terms of the AR coefficients $\varphi_i, i = 1, 2$. The unknown parameter vector is $\theta = (\varphi_1, \varphi_2, \alpha_1, \alpha_2)'$, $|\varphi_1| < 1$, $|\varphi_2| < 1$. For data observed on a square grid, they generalized the Whittle estimation technique used by Fox and Taqqu (1986, 1987) and derived a central limit theorem for the random fields.

For model (4.7), Beran, Ghosh & Dieter (2009) assumed that $f^*(.,.)$ follows a stationary and invertible ARMA model such that the spectral density is given as:

$$f(\lambda_1,\lambda_2) = \frac{\sigma^2}{4\pi^2}\left|1-\mathrm{e}^{-i\lambda_1}\right|^{-2\alpha_1}\left|1-\mathrm{e}^{-i\lambda_2}\right|^{-2\alpha_2}\left|\frac{\psi_1(\mathrm{e}^{-i\lambda_1})}{\varphi_1(\mathrm{e}^{-i\lambda_1})}\right|^2\left|\frac{\psi_2(\mathrm{e}^{-i\lambda_2})}{\varphi_2(\mathrm{e}^{-i\lambda_2})}\right|^2, \qquad (4.9)$$

where $\varphi_i(.)$ and $\psi_i(.)$ are the polynomials of AR and MA coefficients, respectively. This function is the product of the spectral densities of two FARIMA models, such that

$$f(\lambda_1,\lambda_2) = \sigma^2 f_1(\lambda_1) f_2(\lambda_2).$$

The process is stationary and invertible when $0 < \alpha_1, \alpha_2 < 1/2$ and roots of the polynomials φ_i and ψ_i are outside the unit circle. Using an AR(∞) representation of the two-way FARIMA model, they considered least squares estimation of the parameter vector $\theta = (\alpha_1, \alpha_2, \varphi_1', \varphi_2', \psi_1', \psi_2')$.

It is well known that the parametric methods provide efficient estimators only when the model is correctly specified. However, for a given data it might be that the model (AR or ARMA) assumed for $f^*(.,.)$ is not correct or even if the model holds exactly, there is the possibility that the model order is misspecified and as a result, the maximum likelihood or Whittle estimate for the long range dependence parameter(s) will be asymptotically biased. Since the primary interest is in the estimation of long range dependence parameter only, therefore, it would be preferred to have a method of estimation even if we are not able to specify a fully parametric model for the short range dependence component $f^*(.,.)$. For a rectangular $n_1 \times n_2$ grid, the mean squared errors of the parametric estimators of long range dependence parameters are of the order $O\left([n_1 n_2]^{-1}\right)$ if the model is correctly specified, but the estimator may be inconsistent if the model is misspecified. This motivates *semi-parametric approach* in which $f^*(.,.)$ is not required to obey any parametric model.

4.2.2 Semi-parametric Approach

Anh & Lunney (1995) considered model (4.3) to propose an extension of the Gaussian semi-parametric (or local Whittle) estimator (GSE) of α proposed by Künsch (1987) in the time series context and its statistical properties were analysed by Robinson (1995). This estimator maximizes an approximate form of the Whittle likelihood only at frequencies in the neighborhood of the zero. Frias, Alonso, Ruiz-Medina & Angulo (2008) extended the averaged periodogram estimator of Robinson (1994) which models spectral density using (4.3). The discrete averaging is carried out over the neighborhood of the zero frequency only. They proved weak consistency of the proposed estimator under certain regularity

assumptions on the spectral density of the process at low frequencies. Wang (2009) also considered model (4.3) to extend the GPH estimator proposed by Geweke & Porter-Hudak (1983). Assuming $f^*(.,.)$ as a constant and regressing log of the periodogram on $-2\log|1-\mathrm{e}^{-i\boldsymbol{\lambda}}|$ only at frequencies in the neighborhood of zero, he proposed the semi-parametric GPH estimator for α. He also obtained weak consistency of the proposed estimator under some regularity assumptions. Since all of the above methods aim at constructing estimators of α without any strict restrictions on $f^*(.,.)$ away from zero frequency, they belong to the class of *narrowband* or *local* estimators. The main advantage of these estimators is that they are consistent under certain regularity conditions without the need to specify the parametric model correctly for the spectral density. However, a major drawback of these estimators is that irrespective of how smooth the true spectral density is, generally the mean squared error is of the order $O\left([n_1 n_2]^{-m}\right)$ for $m \leq 4/5$. Therefore, if the actual spectral density is smooth enough then a potentially misspecified parametric model can easily outperform these estimators.

4.3 Proposed Semi-parametric Estimator

For time series (1-D processes), Moulines & Soulier (1999) and Hurvich & Brodsky (2001) considered an estimator based on fitting of the potentially misspecified parametric models. The proposed estimator assumes that $f^*(.,.)$ is sufficiently smooth such that it can be approximated by Bloomfield's (1973) exponential model with a finite number of parameters. This estimator is known as fractional exponential (FEXP) estimator since the fitted models are in the FEXP class, see Beran (1993, 1994). In contrast to the restricted narrowband estimators, the FEXP estimator includes all the nonzero frequencies, therefore, it is called *broadband* or *global* estimator by Hurvich & Brodsky (2001). We consider extension of the FEXP estimator to estimate the long range dependence parameter(s) in random fields (2-D processes).

4.3.1 FEXP Estimator

Suppose that $f^*(\lambda_1, \lambda_2)$ is a sufficiently smooth function so that $\log f^*$ can be approximated by truncated bivariate Fourier series as:

$$\log f^*(\lambda_1, \lambda_2) \approx \sum_{r=-p}^{p} \sum_{s=-q}^{q} \theta_{rs} \exp\{i(\lambda_1 r + \lambda_2 s)\} = \theta_{00} + 2\sum_{r=1}^{p} \theta_{r0} \cos(r\lambda_1)$$

$$+2\sum_{s=1}^{q} \theta_{0s} \cos(s\lambda_2) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(+)} \cos(r\lambda_1 + s\lambda_2) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(-)} \cos(r\lambda_1 - s\lambda_2), \quad (4.10)$$

for real-valued coefficients $\{\theta_{rs}\}$. This model is called the bivariate exponential model of order (p, q) and is denoted by EXP(p, q). The model order p and q are such that

$$1/p + p/n_1 \to 0 \text{ and } 1/q + q/n_2 \to 0.$$

Let $\{X(t_1, t_2), 1 \le t_1 \le n_1, 1 \le t_2 \le n_2\}$ be a stationary random field with spectral density $f(.,.)$ and periodogram given as:

$$I(\lambda_1, \lambda_2) = \frac{1}{(2\pi)^2 n_1 n_2} \left| \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} X(t_1, t_2) \exp\{-i(t_1\lambda_1 + t_2\lambda_2)\} \right|^2, \quad (4.11)$$

where λ_1 and λ_2 are the Fourier frequencies. Brillinger (1974) provided a generalization for the asymptotic properties of the Fourier transforms of a spatial series. Under the assumption that all cumulants of the process $\{X(t_1, t_2)\}$ are bounded he proved the asymptotic normality of the Fourier transform such that the periodogram satisfies:

$$\frac{I(\lambda_1, \lambda_2)}{f(\lambda_1, \lambda_2)} \sim \chi_2^2, \text{ as } \min(n_1, n_2) \to \infty,$$

for Fourier frequencies $\lambda_1 = 2\pi j/n_1$ and $\lambda_2 = 2\pi k/n_2$ with $(j, k) \in A$, where set A is

$$A = \{(j, k) : 1 \le j \le m_1, -m_2 \le m_2;\ j = 1, 1 \le k \le m_2\}, \quad (4.12)$$

with $m_i = [(n_i - 1)/2]\,, i = 1, 2$. We can write

$$\log I(\lambda_1, \lambda_2) = \log f(\lambda_1, \lambda_2) - \gamma + \varepsilon, \quad (4.13)$$

where $\gamma = 0.57721$ is the Euler's constant and $\varepsilon = \log\{I(\lambda_1, \lambda_2)/f(\lambda_1, \lambda_2)\} + \gamma$, has zero mean and variance$= \pi^2/6$.

For long range dependent random field, we can write (4.13) as

$$\log I(\lambda_1, \lambda_2) = \log L(\alpha_1, \alpha_2) + \log f^*(\lambda_1, \lambda_2) - \gamma + \varepsilon, \tag{4.14}$$

where $L(\alpha_1, \alpha_2)$ is a function of the long range dependence parameters of the spectral density. Using the EXP(p, q) model in (4.10) to replace $f^*(.,.)$ in (4.14), we have

$$\log I(\lambda_1, \lambda_2) + \gamma \approx \log L(\alpha_1, \alpha_2) + \theta_{00} + 2\sum_{r=1}^{p} \theta_{r0} \cos(r\lambda_1) + 2\sum_{s=1}^{q} \theta_{0s} \cos(s\lambda_2) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(+)} \cos(r\lambda_1 + s\lambda_2) + 2\sum_{r=1}^{p}\sum_{s=1}^{q} \theta_{rs}^{(-)} \cos(r\lambda_1 - s\lambda_2) + \varepsilon \tag{4.15}$$

Representing (4.15) as a classical linear regression model:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{4.16}$$

where for $(j, k) \in A$, the $N \times 1$ vector $Y = \{y_{jk}\} = \{\log I(\lambda_{1j}, \lambda_{2k}) + \gamma\}$ with $N = 2m_1m_2 + m_1 + m_2 + 1$, $N \times v$ matrix $X_1 = \{x_1(j, k)\}$ consists of the terms associated with the long range dependence component $L(\alpha_1, \alpha_2)$ and β_1 is vector of these v long range dependence parameters. The $N \times R$ matrix $X_2 = \{x_2(j, k)\}$ corresponds to the short range dependence component which is approximated by an EXP(p, q) model with $R = 2pq + p + q + 1$ parameters. The (jk)-th row of X_2 is given as:

$$x_2(j, k) = [1, 2\cos(r\lambda_{1j}), 2\cos(s\lambda_{2k})\cos(r\lambda_{1j} + s\lambda_{2k}), \cos(r\lambda_{1j} - s\lambda_{2k})], \tag{4.17}$$

where $r = (1, 2, \cdots, p)'$ and $s = (1, 2, \cdots, q)'$. The $R \times 1$ vector β_2 is

$$\beta_2' = [\theta_{00}, \theta_{10}, \cdots, \theta_{p0}, \theta_{01}, \cdots, \theta_{0q}, \theta_{11}^{+}, \cdots, \theta_{pq}^{+}, \theta_{11}^{-}, \cdots, \theta_{pq}^{-}], \tag{4.18}$$

and $\varepsilon = \{\varepsilon_{jk}\}$ is $N \times 1$ error vector. Note that X_2 and β_2 will remain same across different long range dependence models whereas X_1 and β_1 will vary. From regression model (4.16),

the least squares estimate of parameters in β_1 give the semi-parametric FEXP estimator(s). As a byproduct we also get estimates for $\{\theta_{rs}\}$, the Fourier coefficients of $\log f$ which provides an estimator for the spectral density of the process as well. The standard errors for these $(R+v)$ parameters in (4.16) are given by the square root of the diagonal element of the variance-covariance matrix, that is

$$\text{cov}\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \sigma^2 \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}.$$

We show some simplifications and details for the three spectral density models discussed in Section 4.2.

For model (4.3), the long range dependence component is given as

$$\begin{aligned} L(\alpha) &= \left|1 - e^{-i(\lambda_1,\lambda_2)'}\right|^{-2\alpha} = \left|1 - e^{-i\lambda_1}\right|^{-2\alpha} + \left|1 - e^{-i\lambda_2}\right|^{-2\alpha} \\ &= \left\{4\sin^2\left(\frac{\lambda_1}{2}\right) + 4\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-\alpha}, \end{aligned} \tag{4.19}$$

with only one unknown α. The (jk)-th row of $N \times 1$ matrix X_1 is

$$x_1(j,k) = \log\left\{4\sin^2\left(\frac{\lambda_{1j}}{2}\right) + 4\sin^2\left(\frac{\lambda_{2k}}{2}\right)\right\},$$

and $\beta_1 = \alpha$.

For spectral density in (4.5),

$$L(\alpha) = \left|1 - e^{-i(\lambda_1 + k\lambda_2)}\right|^{-2\alpha} = \left\{4\sin^2\left(\frac{\lambda_1 + k\lambda_2}{2}\right)\right\}^{-\alpha},$$

with only one unknown α. The (jk)-th row of $N \times 1$ matrix X_1 is

$$x_1(j,k) = \log\left\{16\sin^2\left(\frac{\lambda_{1j}}{2}\right)\sin^2\left(\frac{\lambda_{2k}}{2}\right)\right\},$$

and $\beta_1 = \alpha$.

The situation is slightly different for model (4.7) in which the long range dependence component is

$$L(\alpha_1,\alpha_2) = \left|1 - \mathrm{e}^{-i\lambda_1}\right|^{-2\alpha_1}\left|1 - \mathrm{e}^{-i\lambda_2}\right|^{-2\alpha_2} = \left\{4\sin^2\left(\frac{\lambda_1}{2}\right)\right\}^{-\alpha_1}\left\{4\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-\alpha_2},$$

with $\beta_1 = [\alpha_1\ \alpha_2]'$ and X_1 is a $N \times 2$ matrix with (jk)-th row given as

$$x_1(j,k) = \left[\log\left\{4\sin^2\left(\frac{\lambda_{1j}}{2}\right)\right\}\ \log\left\{4\sin^2\left(\frac{\lambda_{2k}}{2}\right)\right\}\right].$$

The FEXP estimator can thus be applied directly to different definitions of the long range dependent random fields. An important issue closely related to the model order selection in ARMA type models or variable selection in regression setup is the choice of order (p,q) for the exponential model. We present some selection criteria in the next section.

4.4 FEXP Model Order Selection

4.4.1 Model Selection

We set the notations first. For the regression model (4.16) with EXP model of order (p,q), let $\beta' = [\beta_1\ \beta_2]$, $X = [X_1\ X_2]$, $\widehat{Y} = \{y_{j,k}\} = X\widehat{\beta}$ be the least squares estimate of Y, $RSS_{p,q} = \sum\sum_{(j,k)\in A}(y_{jk} - \widehat{y}_{jk})^2$ be the corresponding residual sum of squares and $H = X(X'X)^{-1}X'$ be the hat matrix. The set A with all nonzero Fourier frequencies is truncated to include frequencies in the neighborhood of zero as:

$$A^* = \{(j,k) : 1 \le m_1^*, -m_2^* \le k \le m_2^*\},$$

where $m^* \in \mathbb{Z}^+$ such that $1/m^* + m^*/n \to 0$ as $n \to \infty$. We generalize Mallow's C_L criterion proposed by Hurvich (2001) in the long range dependent time series context, and Corrected Akaike's Information Criteria introduced by Hurvich & Tsai (1989) for model selection in regression and time series models.

Mallow's C_L Criterion is modified to:

$$C_L(p,q) = RSS^*_{p,q} + \frac{2\pi^2}{6}\sum_{j=1}^{|A^*|} h_{j,j}, \tag{4.20}$$

where $|A^*|$ is cardinality of set A^*, $RSS^*_{p,q} = \sum\sum_{(j,k)\in A^*}(y_{jk} - \hat{y}_{jk})^2$ and $h_{j,j}$ is the j-th diagonal element of the hat matrix H. For $A = A^*$, that is, when all nonzero Fourier frequencies are

included, we get global C_L denoted by C_L^*. The optimal (p, q) minimizes C_L, that is,

$$(p, q) = \underset{(p,q)}{\operatorname{argmin}}\, C_L(p, q).$$

Corrected Akaike's Information Criteria (AICC) is defined as

$$AICC(p, q) = |A^*| \log\left(\frac{RSS^*_{p,q}}{|A^*|}\right) + \frac{2R|A^*|}{|A^*| - R - 1}, \tag{4.21}$$

where $R = 2pq + p + q + 1$ and all other terms are as defined above. Again optimal (p, q) is such that

$$(p, q) = \underset{(p,q)}{\operatorname{argmin}}\, AICC(p, q).$$

4.4.2 Subset Selection

For the regression model (4.16), the variable selection technique choose only a subset of the relevant variables from all R parameters in the EXP(p, q) model. This removes irrelevant variables from the exponential model, thus reducing the number of unknown parameters in the model. Several different optimality criteria can be used for the variable selection, we use Akaike's information criterion (AIC). For more details on subset selection in regression see Miller (2002).

4.5 Simulations

In this section we compare the performance of our FEXP estimator with Wang's (2009) GPH estimator through Monte Carlo simulations. Lee & Berman (1997) proposed a method to approximate 2-D random fields with known spectral density function. We use their approach to simulate data on a $n_1 \times n_2$ regular lattice when its spectral density is of the form (4.3). This method of simulation is a 2-D generalization of the algorithm proposed by Davies & Harte (1987) for stationary Gaussian processes (1-D) with known autocovariance. The details of the algorithm along with the modifications and a proof to show that the

method is exact is given in Appendix C. The criterion used for comparison is the root mean squared error (RMSE) of these estimators. For $f^*(.,.)$ in model (4.3) different AR(1) models are used. The long range dependence parameter α is set equal to $0.3, 0.6$ and 0.9. All three α values resulted in mostly similar conclusions, therefore, we report results for $\alpha = 0.6$ only. Two different grid sizes $(n_1, n_2) = \{(50, 50), (50, 100)\}$ are used and for each case 1000 realizations of the process are generated. For the GPH estimator, we use $m_i = n_i^{4/5}, i = 1, 2$ as suggested by Wang (2009) .

To see the performance of different model order selection criteria discussed in Section 4.3 we compare the average and variance of the (p, q) values selected using different criteria from 1000 realizations. The RMSE of the FEXP estimator corresponding to selected (p, q) values is also given. The global and local Mallow's C_L are denoted by C_L^* and C_L, respectively. For C_L and AICC, $(m_1, m_2) = (n_1^a, n_2^a)$ with $a = \{1/2, 2/3, 3/4, 4/5\}$ are used. For all four criteria: C_L^*, C_L, AICC and subset selection, all possible pairs in $\{(1, 1) \leq (p, q) \leq (5, 5)\}$ are used.

Consider a stationary, invertible and separable AR(1) model with spectral density given as

$$f^*(\lambda_1, \lambda_2) = \frac{\sigma^2}{(2\pi)^2} \left|(1 - \phi_{1,0}e^{-i\lambda_1})(1 - \phi_{0,1}e^{-i\lambda_2})\right|^{-2}, \tag{4.22}$$

where $|\phi_{1,0}| < 1$ and $|\phi_{0,1}| < 1$. We use $(\phi_{1,0}, \phi_{0,1}) = \{(0.6, 0.4), (0.6, -0.4), (-0.6, -0.4)\}$. Figure 5(a)-(c) represents the log of spectral densities for these three processes when $\alpha = 0.6$, all these plots show that there is a significant peak at $(0, 0)$, small peaks at higher frequencies in (a), peaks along $\lambda_1 = 0$ in (b), peaks along $\lambda_1 = -\pi$ and $\lambda_1 = \pi$ in (c).

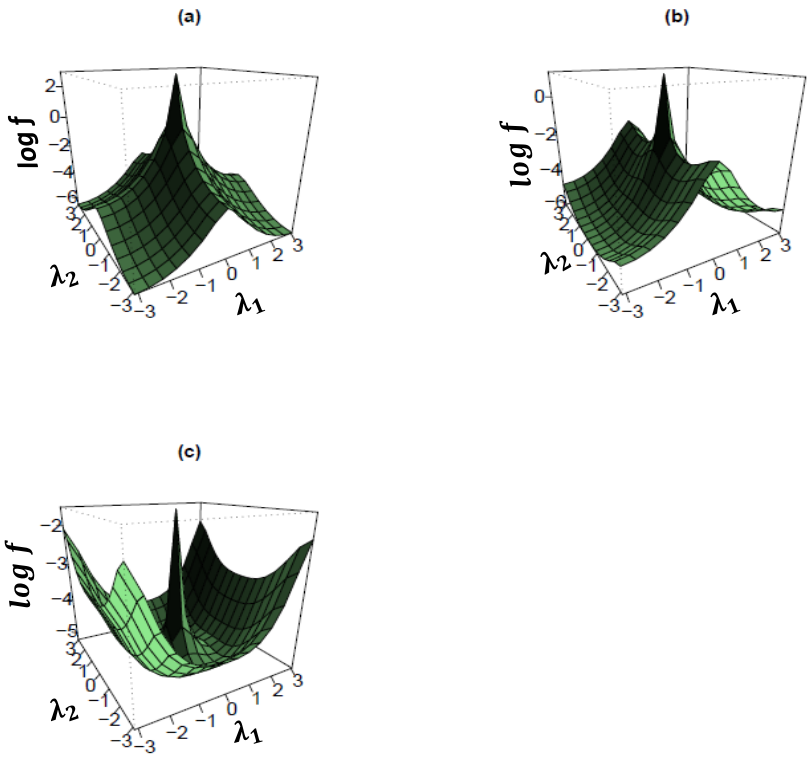

Figure 5: Spectral Density for $\alpha = 0.6$ and AR(1) Model with $(\phi_{1,0}, \phi_{0,1})$ (a) $(0.6, 0.4)$, (b) $(0.6, -0.4)$ and (c) $(-0.6, -0.4)$.

The RMSE of FEXP estimator for all possible pairs of (p, q) is given in Table 4. The value of RMSE decreases for each (n_1, n_2) as p increases and q decreases. For $(\phi_{1,0}, \phi_{0,1}) = (0.6, 0.4), (0.6, -0.4)$ and $(-0.6, -0.4)$, the smallest RMSE value is obtained when $(p, q) = (4, 2), (3, 2)$ and $(2, 1)$, respectively for both grid sizes. The FEXP estimator is consistent since its RMSE (and the variance) decreases as the sample size increases for all models.

Table 4: RMSE of FEXP Estimator for Separable AR(1) Models.

	$(0.6, 0.4)$					$(0.6, -0.4)$					$(-0.6, -0.4)$				
$q\downarrow$ $p\rightarrow$															
						$(n_1, n_2) = (50, 50)$									
.572	.374	.333	.328	.326		.531	.327	.286	.280	.278	.397	.176	246	.240	.246
.608	.344	.287	.286	.290		.552	.269	.242	.258	.267	.385	.258	.243	.269	.274
.639	.382	.336	.357	.383		.623	.372	.333	.357	.385	.443	.284	.314	.349	.376
.652	.405	.377	.427	.488		.636	.394	.370	.422	.483	.452	.319	.357	.425	.476
.657	.419	.402	.481	.582		.643	.412	.401	.483	.587	.457	.329	.383	.486	.579
						$(n_1, n_2) = (50, 100)$									
.534	.349	.310	.304	.302		.503	.304	.263	.256	.254	.375	.149	.223	.216	.223
.553	.279	.209	.198	.199		.503	.202	.156	.164	.172	.343	.206	.154	.177	.176
.578	.297	.227	.233	.245		.569	.291	.225	.232	.245	.398	.220	.201	.225	.239
.590	.314	.251	.275	.303		.581	.306	.247	.273	.302	.407	.252	.229	.276	.301
.600	.326	.266	.300	.341		.588	.321	.265	.300	.343	.412	.262	.246	.302	.338

The RMSE of the GPH estimator and lowest RMSE of the FEXP estimator is given in Table 5. These values show that for all three models, FEXP estimator has smaller RMSE than the GPH estimator.

Table 5: RMSE of FEXP and GPH Estimators.

	$(n_1, n_2) = (50, 50)$			$(n_1, n_2) = (50, 100)$		
$(\phi_{1,0}, \phi_{0,1})$	(p, q)	FEXP	GPH	(p, q)	FEXP	GPH
(0.6,0.4)	(4,2)	0.286	1.055	(4,2)	0.198	1.015
(0.6,-0.4)	(3,2)	0.242	0.316	(3,2)	0.156	0.416
(-0.6,-0.4)	(2,2)	0.176	0.797	(2,2)	0.149	0.718
(.24,.24)	(4,4)	0.605	1.670	(5,5)	0.414	1.396
(.24,-.24)	(1,2)	0.164	0.121	(1,2)	0.110	0.039
(-.24,-.24)	(2,1)	0.212	0.919	(1,2)	0.154	0.762

We use the subset selection in which only the variables selected using AIC are kept in vector β_2 of the regression model (4.16). The results reported in Table 6 suggest that only 5 variables (including α) out of $R + v = 73$ variables are selected, clearly there is a significant reduction in the number of parameters of the EXP(p, q) model. The RMSE of FEXP estimator is comparable to those from three selection criteria used above.

Table 6: RMSE of FEXP Estimator for Separable AR(1) Models Using Subset Selection. n_s(var) denotes the average (and variance) of the number of variables selected in 1000 replications.

	(0.6,0.4)		(0.6,-0.4)		(-0.6,-0.4)	
(n_1, n_2)	n_s(var)	RMSE	n_s(var)	RMSE	n_s(var)	RMSE
(50,50)	4.67(0.79)	0.357	4.65(0.95)	0.344	4.48(0.72)	0.307
(50,100)	5.31(0.30)	0.123	5.25(0.430)	0.137	5.09(0.21)	0.099

The results for global and local C_L, AICC are reported in Table 7 in which the average and variance of the (p, q) from 1000 realizations are given. For all three AR(1) models, the average and variance of p and q selected using the global and local C_L are close to each other whereas those selected using AICC are always smaller. The RMSE is lowest for global C_L in all cases. For both local C_L and AICC, the RMSE decreases as a increases with lowest value achieved for $a = 4/5$. Therefore, the bandwidth $m_i = n_i^{4/5}$ gives minimum RMSE, this is in correspondence with the choice of optimum bandwidth given by Hurvich, Deo & Brodsky (1998) for long range dependent time series.

Table 7: RMSE of FEXP Estimator for Separable AR(1) Models Using Global and Local C_L and AICC.

		$(n_1, n_2) = (50, 50)$			$(n_1, n_2) = (50, 100)$		
Criterion	a	p(var)	q(var)	RMSE($\widehat{\alpha}$)	p(var)	q(var)	RMSE($\widehat{\alpha}$)
				(0.6,0.4)			
CL^*	-	2.81(0.66)	1.82(0.64)	0.326	3.15(0.58)	2.18(0.53)	0.230
CL	1/2	2.86(1.32)	2.21(1.73)	0.390	3.44(1.28)	2.42(1.59)	0.282
	2/3	2.86(1.01)	1.98(1.17)	0.345	3.26(1.02)	2.18(1.21)	0.267
	3/4	2.87(0.88)	1.95(0.95)	0.334	3.10(0.81)	2.26(0.90)	0.249
	4/5	2.86(0.78)	1.76(0.70)	0.329	3.13(0.67)	2.25(0.74)	0.237
$AICC$	1/2	1.25(0.19)	1.01(0.01)	0.525	1.67(0.23)	1.01(0.01)	0.419
	2/3	1.83(0.35)	1.10(0.09)	0.421	2.12(0.26)	1.07(0.06)	0.350
	3/4	2.34(0.38)	1.28(0.24)	0.349	2.43(0.36)	1.45(0.33)	0.296
	4/5	2.62(0.63)	1.51(0.44)	0.335	2.79(0.50)	1.84(0.42)	0.259
				(0.6,-0.4)			
CL^*	-	2.75(.61)	1.73(0.50)	0.289	3.13(.57)	2.06(0.33)	0.192
CL	1/2	2.86(1.35)	2.22(1.61)	0.373	3.29(1.33)	2.28(1.57)	0.265
	2/3	2.84(1.05)	2.02(1.13)	0.326	3.22(0.99)	2.07(1.09)	0.243
	3/4	2.77(0.74)	1.86(0.81)	0.305	3.07(0.80)	2.08(0.75)	0.220
	4/5	2.78(0.76)	1.71(0.56)	0.295	3.13(0.66)	2.10(0.58)	0.203
$AICC$	1/2	1.26(0.19)	1.01(0.01)	0.480	1.67(0.22)	1.02(0.02)	0.377
	2/3	1.80(0.31)	1.13(0.13)	0.382	2.09(0.24)	1.07(0.07)	0.306
	3/4	2.31(0.37)	1.29(0.27)	0.303	2.39(0.35)	1.45(0.28)	0.247
	4/5	2.56(0.59)	1.49(0.37)	0.291	2.77(0.48)	1.76(0.27)	0.210
				(-0.6,-0.4)			
CL^*	-	2.87(.67)	1.56(0.45)	0.270	3.14(.53)	1.92(0.42)	0.195
CL	1/2	2.95(1.32)	2.09(1.63)	0.332	3.35(1.34)	2.30(1.68)	0.253
	2/3	2.87(1.01)	1.83(1.01)	0.281	3.26(0.92)	2.00(1.14)	0.221
	3/4	2.89(0.88)	1.70(0.73)	0.271	3.13(0.78)	1.96(0.75)	0.208
	4/5	2.86(0.78)	1.59(0.54)	0.268	3.16(0.66)	1.93(0.56)	0.198
$AICC$	1/2	1.35(0.23)	1.01(0.01)	0.332	1.64(0.25)	1.00(0.005)	0.257
	2/3	1.92(0.26)	1.09(0.09)	0.247	2.27(0.34)	1.10(0.09)	0.207
	3/4	2.35(0.39)	1.23(0.20)	0.245	2.55(0.38)	1.34(0.25)	0.214
	4/5	2.62(0.58)	1.39(0.32)	0.259	2.84(0.51)	1.61(0.35)	0.207

Next we consider a noncausal, nonseparable AR(1) model with spectral density given as

$$f^*(\lambda_1, \lambda_2) = \frac{\sigma^2}{(2\pi)^2}|1 - 2\phi_{1,0}\cos(\lambda_1) - 2\phi_{0,1}\cos(\lambda_2)|^{-2}; \;\; |\phi_{1,0}| + |\phi_{0,1}| < 1/2, \quad (4.23)$$

to simulate data using $(\phi_{1,0}, \phi_{0,1}) = \{(0.24, 0.24), (0.24, -0.24), (-0.24, -0.24)\}$. Figure 6(a)-(c) presents log of the spectral densities for these three models when $\alpha = 0.6$. In (a) there is a high peak at $(0, 0)$ and small ones at high frequencies, whereas in (b) there are high peaks at $(0, \pi)$ and $(0, -\pi)$ and only a small peak at $(0, 0)$ and in (c) there is a small peak at $(0, 0)$, and high peaks at four edges $(-\pi, -\pi)$, (π, π), $(\pi, -\pi)$ and $(-\pi, \pi)$.

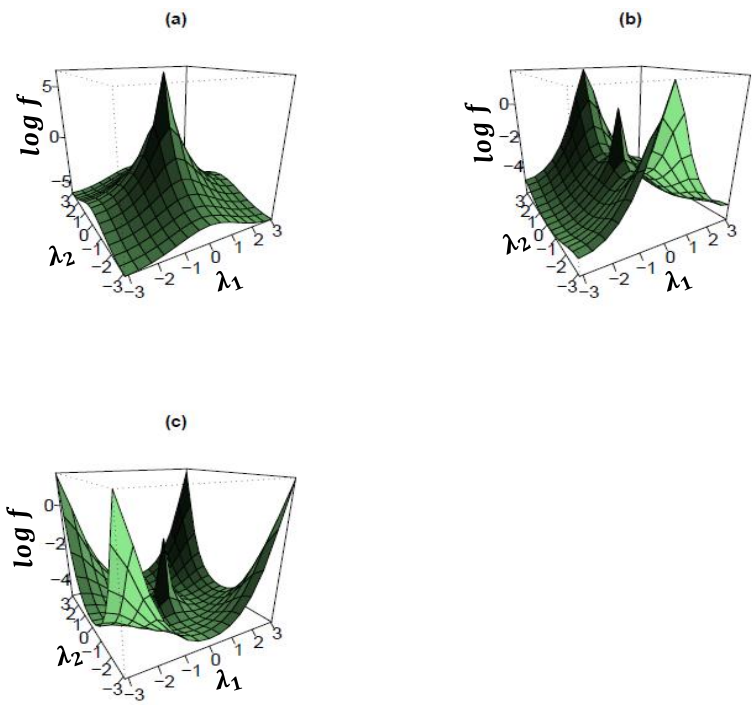

Figure 6: Spectral Density for $\alpha = 0.6$ and AR(1) Model with (a) $(0.24, 0.24)$, (b) $(0.24, -0.24)$ and (c) $(-0.24, -0.24)$.

The RMSE of the FEXP estimator in Table 8 shows no clear (increasing/decreasing) as p and q change from 1 to 5. For $(\phi_{1,0}, \phi_{0,1}) = (0.24, 0.24), (0.24, -0.24)$ and $(-0.24, -0.24)$,

the minimum RMSE value is obtained when $(p, q) = (4, 4), (1, 2)$ and $(2, 1)$, respectively for $(n_1, n_2) = (50, 50)$ and $(p, q) = (5, 5), (1, 2)$ and $(1, 2)$, respectively for $(n_1, n_2) = (50, 100)$. The FEXP estimator is consistent since the RMSE decreases as the sample size increases for all models.

Table 8: RMSE of FEXP Estimator for Nonseparable AR(1) Models.

	$(0.24, 0.24)$					$(0.24, -0.24)$					$(-0.24, -0.24)$				
$q\downarrow$	$p\rightarrow$														
						$(n_1, n_2) = (50, 50)$									
1.316	1.201	1.189	1.190	1.193		.469	.572	.633	.655	.663	.240	**.212**	.238	.240	.244
1.196	0.917	0.840	0.834	0.842		**.164**	.286	.393	.447	.472	.218	.235	.247	.270	.277
1.181	0.835	0.681	0.644	0.649		.245	.255	.326	.377	.416	.246	.249	.310	.344	.374
1.182	0.832	0.649	**0.605**	0.618		.234	.259	.357	.436	.507	.250	.277	.361	.429	.496
1.184	0.839	0.653	0.607	0.638		.242	.267	.374	.473	.578	.254	.282	.384	.475	.583
						$(n_1, n_2) = (50, 100)$									
1.252	1.177	1.180	1.189	1.195		.455	.574	.640	.665	.675	.200	.192	.212	.218	.221
1.083	0.833	0.769	0.770	0.781		**.110**	.239	.330	.373	.393	**.154**	.166	.172	.188	.190
1.053	0.719	0.569	0.533	0.534		.190	.176	.232	.273	.299	.176	.163	.208	.235	.250
1.050	0.697	0.506	0.446	0.441		.173	.171	.240	.292	.325	.176	.183	.232	.281	.310
1.049	0.695	0.489	0.419	**0.414**		.179	.174	.247	.311	.354	.178	.184	.248	.311	.354

Table 5 also gives RMSE of the GPH and FEXP estimator. These values show that for $(\phi_{1,0}, \phi_{0,1}) = (0.24, 0.24)$ and $(-0.24, 0.24)$, the FEXP estimator has smaller RMSE than the GPH estimator for all cases whereas when $(\phi_{1,0}, \phi_{0,1}) = (0.24, -0.24)$ the RMSE of GPH estimator is smaller.

Using subset selection around 8 to 10 variables (including α) are significant as reported in Table 9, thus leading to reduction in the number of parameters from $R = 72$ to 10. The RMSE of FEXP is smaller than those from three selection criteria for model with $(\phi_{1,0}, \phi_{0,1}) = (0.24, 0.24)$.

Table 9: RMSE of FEXP Estimator for Nonseparable AR(1) Models Using Subset Selection. n_s(var) denotes the average (and variance) of the number of variables selected in 1000 replications.

	$(0.24, 0.24)$		$(0.24, -0.24)$		$(-0.24, -0.24)$	
(n_1, n_2)	n_s(var)	RMSE	n_s(var)	RMSE	n_s(var)	RMSE
(50,50)	7.32(3.09)	0.573	7.54(1.29)	0.628	7.49(0.98)	0.551
(50,100)	8.89(3.08)	0.552	10.48(1.83)	0.430	9.59(2.20)	0.425

The results for global and local C_L and AICC are reported in Table 10 in which the average and variance of the (p, q) from 1000 realizations are given. For all three nonseparable AR(1) models, the average and variance of p and q selected using AICC is smaller than the values obtained using global and local C_L criteria. However, the values selected using global and local C_L are close. The RMSE of FEXP is lowest for global C_L when $(\phi_{1,0}, \phi_{0,1}) = (0.24, 0.24)$, whereas for $(\phi_{1,0}, \phi_{0,1}) = (0.24, -0.24)$ both local C_L and AICC outperformed the global C_L and the RMSE is lowest for local C_L and AICC when $(n_1, n_2) = (50, 50)$ and $(50, 100)$, respectively. For the third model with $(\phi_{1,0}, \phi_{0,1}) = (-0.24, -0.24)$, AICC outperforms the other two criteria for all cases. The RMSE using local C_L mostly decreases as a increases with lowest value achieved for $a = 4/5$, whereas for AICC the smallest RMSE is not achieved at $a = 4/5$ always.

Table 10: RMSE of FEXP Estimator for Nonseparable AR(1) Models Using Global and Local C_L and AICC.

		$(n_1, n_2) = (50, 50)$			$(n_1, n_2) = (50, 100)$		
Criterion	a	p(var)	q(var)	RMSE($\widehat{\alpha}$)	p(var)	q(var)	RMSE($\widehat{\alpha}$)
				(0.24,0.24)			
CL^*	-	3.61(1.51)	3.67(1.46)	0.689	3.52(1.45)	4.46(0.59)	0.505
CL	1/2	3.16(1.48)	3.11(1.48)	0.757	4.05(1.22)	3.64(1.27)	0.538
	2/3	3.20(1.63)	3.25(1.53)	0.770	3.84(1.33)	4.09(0.95)	0.512
	3/4	3.42(1.54)	3.43(1.51)	0.724	3.66(1.45)	4.36(0.74)	0.517
	4/5	3.52(1.72)	3.57(1.68)	0.728	3.63(1.41)	4.44(0.65)	0.507
$AICC$	1/2	1.29(0.21)	1.28(0.20)	1.214	1.45(0.42)	1.59(0.26)	1.077
	2/3	1.84(0.88)	1.84(0.90)	1.081	2.31(1.63)	2.31(1.31)	0.928
	3/4	2.63(1.53)	2.69(1.51)	0.889	2.60(1.64)	3.63(1.31)	0.721
	4/5	3.20(1.87)	3.19(1.88)	0.794	3.09(1.63)	4.12(0.95)	0.603
				(0.24,-0.24)			
CL^*	-	2.62(0.58)	2.66(0.67)	0.370	2.93(0.52)	3.17(0.68)	0.270
CL	1/2	2.46(1.34)	2.87(1.52)	0.402	2.59(1.53)	3.11(1.44)	0.278
	2/3	2.36(0.99)	2.60(1.26)	0.390	2.40(1.04)	3.14(1.08)	0.255
	3/4	2.43(0.67)	2.51(0.85)	0.375	2.63(0.67)	2.99(0.92)	0.263
	4/5	2.49(0.58)	2.50(0.74)	0.367	2.90(0.64)	2.94(0.84)	0.284
$AICC$	1/2	1.20(0.09)	1.23(0.18)	0.427	1.09(0.08)	1.53(0.25)	0.334
	2/3	1.45(0.28)	1.53(0.33)	0.414	1.33(0.23)	2.03(0.16)	0.213
	3/4	1.96(0.33)	1.88(0.43)	0.398	2.06(0.18)	2.16(0.26)	0.266
	4/5	2.29(0.37)	2.23(0.50)	0.369	2.55(0.42)	2.47(0.45)	0.287
				(-0.24,-0.24)			
CL^*	-	2.56(0.52)	2.58(0.54)	0.279	2.89(0.50)	3.01(0.55)	0.207
CL	1/2	3.11(1.22)	3.16(1.12)	0.357	3.46(1.22)	3.42(1.21)	0.256
	2/3	2.80(0.90)	2.81(0.91)	0.311	2.99(1.04)	3.23(0.92)	0.223
	3/4	2.59(0.70)	2.62(0.71)	0.276	2.74(0.75)	2.98(0.75)	0.198
	4/5	2.52(0.60)	2.52(0.59)	0.258	2.86(0.65)	2.80(0.70)	0.199
$AICC$	1/2	1.50(0.26)	1.52(0.26)	0.224	1.66(0.24)	1.60(0.25)	0.172
	2/3	1.83(0.22)	1.84(0.23)	0.234	1.90(0.32)	1.92(0.34)	0.173
	3/4	2.06(0.27)	2.05(0.25)	0.242	2.13(0.17)	2.22(0.26)	0.169
	4/5	2.26(0.34)	2.31(0.38)	0.250	2.47(0.37)	2.41(0.36)	0.176

4.6 Mercer and Hall Wheat Data

We base our analysis on the Mercer & Hall's (1911) wheat yield data. Their main objective was to find the optimal size and number of field plots needed to reduce the error of estimating yield to an acceptable level. This data is also used in Section 3.5.2 to predict the values at different field locations. Whittle (1954) computed correlations between the wheat yield up to lag 3, and found that the correlations are stronger along the north-south as compared to the east-west direction. He speculated that there are some 'waves of fertility' of the kind often remarked in a ploughed field. From 2-D spectral analysis, McBratney & Webster (1981) plotted the correlogram and an estimate of the spectrum obtained by using the Bartlett window of 10 lags. They found an obvious peak in the east-west direction with a three plot long period and also reported evidence of an earlier ridge and furrows pattern of plowing on the field under study. Ripley (1981) found a similar pattern in the periodogram plot and attributed this to the variation in the soil fertility caused by layers in the outcropping rocks. Cressie (1993) considered the median polishing of this data before modelling the dependence. He showed that the estimated additive effect of the median polished row, column and overall effects have a striking variability in the east-west direction.

We present the plot of log periodogram in Figure 7(a) which shows a peak at the origin and along the line $\lambda_1 = 0$ for different values of λ_2, thus showing evidence for long range dependence. There are also prominent parallel ridges in the north-south direction. The periodogram plot obtained from the row and column series (in log-log coordinates) is shown in Figure (b) and (c), respectively. These plots show the presence of long range dependence along both directions.

We fit the FEXP models (4.3), (4.5) and (4.7) to this data. For $f^*(.,.)$ in these models we use (p, q) selected using AICC, Mallow's C_L and subset selection from all possible values in $\{(0,0) \leq (p,q) \leq (7,7)\}$. Since AICC and C_L selected same model order so we

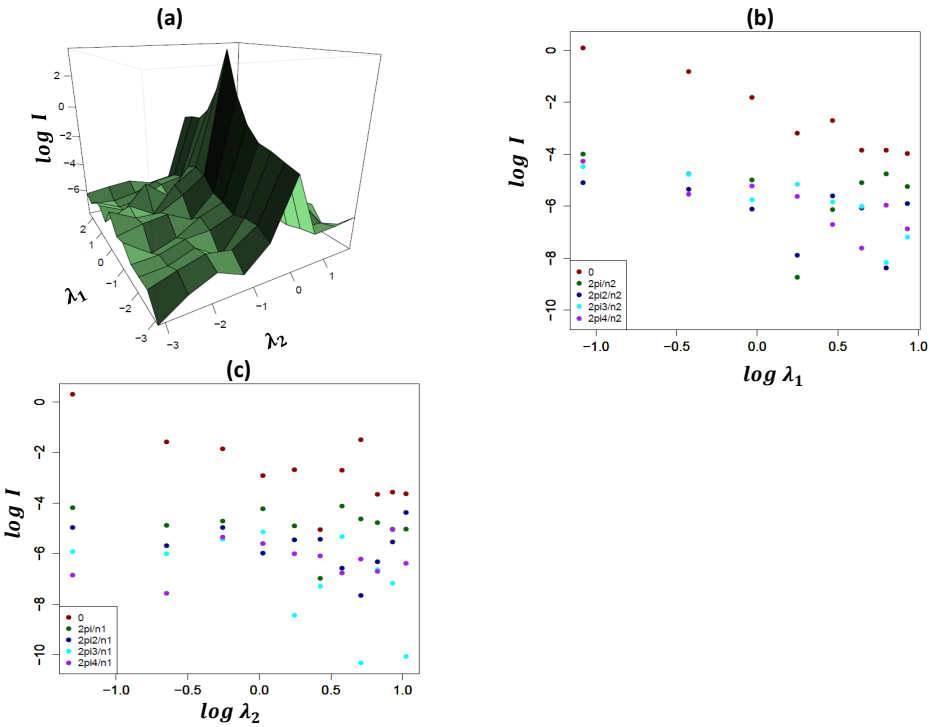

Figure 7: Log Periodogram of the (a) Wheat Data, (b) Row Series and (c) Column Series.

report results for C_L only. The results for the FEXP estimators from different models and their 95% confidence intervals are provided in Table 11.

For model (4.3), C_L selected $(p, q) = (4, 1)$ and for subset selection only α is included in the model. The results in Table 11 show that the estimates are above 0.5, thus it provides strong evidence for long range dependence. Note that the confidence interval based on C_L are much wider (nearly four times) than the subset selection. The log of fitted spectral densities for $(p, q) = (4, 1)$ is shown in Figure 8(b) along with the log periodogram in (a). The fitted density represents the data fairly well.

For model (4.5), C_L selected $(p, q) = (1, 1)$ and for subset selection only α_1 and α_2 are included in the model. The results in Table 11 provide evidence for long range dependence in both directions with stronger effect in the east-west direction. The confidence intervals

obtained using C_L are wider than those from the subset selection. The log of fitted spectral densities for $(p, q) = (1, 1)$ in Figure 8(c) which represents the log periodogram in (a) much better than the fitted model in (b).

For model (4.7) with $k = 0$, C_L selected $(p, q) = (1, 1)$ and for subset selection only α is included in the model. The results in Table 11 show evidence for long range dependence. The confidence intervals obtained using C_L are wider than those from the subset selection. The log of fitted spectral densities for $(p, q) = (1, 1)$ in Figure 8(d) is not a good representation of the data.

These results provide evidence that the long range dependence parameters differ with stronger dependence in the east-west direction. Model (4.5) provides the most appropriate fit for the wheat data, and explains the presence of long range dependence and periodicity quite well. Our results also support the claim of McBratney & Webster (1981).

Table 11: FEXP Estimators for Wheat Yield Data.

	Model (4.3)		Model (4.5)				Model (4.7)	
Criteria	$\widehat{\alpha}$	95% CI	$\widehat{\alpha}_1$	95% CI	$\widehat{\alpha}_2$	95% CI	$\widehat{\alpha}$	95% CI
Local C_L	0.796	(0.02,1.57)	0.260	(0.14,0.38)	0.137	(0.02,0.26)	0.255	(0.14,0.37)
Subset Selection	0.814	(0.61,1.02)	0.315	(0.24,0.39)	0.167	(0.09,0.25)	0.309	(0.23,0.39)

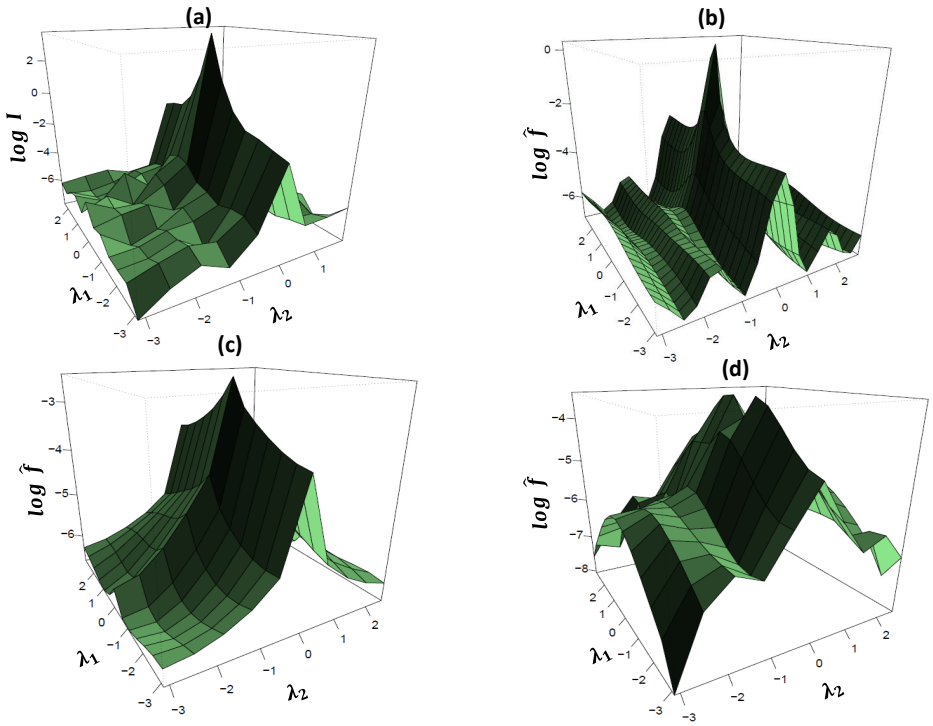

Figure 8: (a) Log Periodogram of Wheat Yield, and Log of Fitted Spectral Density for (b) Model (4.3), (c) Model (4.5) and (d) Model (4.7).

CHAPTER V

CONCLUSION

Random fields have applications in a large diversity of disciplines, such as geography, geology, biology, environmental sciences, and agriculture. Motivated by these applications random fields have been the subject of research. In this dissertation, we have developed methods for prediction and estimation of two dimensional random fields which are defined on a regular lattice.

Kriging, the most commonly used method for spatial prediction, requires specification of a parametric model for the dependence structure. In Chapter II, we develop prediction methodology which requires minimal assumptions on the dependence structure of the process. Using the unilateral moving average representation of a stationary random field, we solve the multi-step ahead prediction problem when the quarter-plane (third quadrant) is chosen and fixed as the past. The multi-step ahead prediction errors and their variances play a central role in solving a number of non-standard prediction problems, which deals with augmentation of the quarter-plane past with either a single or a group of observations. These non-standard prediction problems are closely related to the design problem and motivated by the network site selection in the environmental and geostatistical applications. We obtain informative and explicit prediction error variance formulas in terms of either the autoregressive or moving average parameters of the random fields which links the worth of an additional observation to its spatial location via the size of these parameters. The prediction error variances for random fields seem to be expressible only in terms of the moving average parameters, and attempts to express them in terms of the autoregressive parameters lead to a new and mysterious projection operator which captures the nature of the edge-effects encountered in the estimation of the spectral density function of stationary

random fields. This is in sharp contrast with the solution of prediction problems in the time series setup. The role of projection operator in prediction of random fields remains an open problem for future research. Another interesting problem for further research is to extend the prediction theory when a half-plane is used as the past. This extension would examine the role of past in predicting values at unsampled locations and also in finding the worth of observations for prediction.

In Chapter III, we show that an extension of the exponential models to the stationary random fields play a crucial role in the implementation of the prediction methodology developed in Chapter II. The predictor coefficients for stationary random fields are expressed in terms of the cepstrum coefficients. The developments in Chapters II and III provide a unified framework for forecasting stationary random fields both in the time and spectral domain. Through various simulation studies which account for different random field models, we demonstrate that the proposed methodology delivers impressive prediction results with reasonable prediction error variances. A problem for further research is to generalize this prediction framework to include the long range dependent or fractional exponential processes. This involves using the definitions of fractional exponential models given in Chapter IV to incorporate the long range dependence in the model.

Additionally, in Chapter IV, we develop a new broadband semiparametric estimator for the long range dependence parameter of random fields, which is based on the fractional exponential spectrum. The proposed estimator is consistent, easily adaptable to different long range dependence definitions, is computationally simple, and avoids the pitfalls of model misspecification. We also propose three different criteria to automatically select the appropriate model order for the fractional exponential models. Several Monte Carlo simulations show that our estimator works better than the other estimators, including GPH estimator. Through analysis of the Mercer and Hall's wheat data we illustrate the use of the proposed estimator for various different long range dependent spectral density models.

In many applications, like total column ozone amounts, the observed process have spatial as well as temporal component. Future work involves developing fractional exponential estimators for spatio-temporal processes with long range dependence.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Trans. Automatic Control*, 19, 716–723.

Akkaya, A. D., & Yucemen, M. S. (2002), "Stochastic Modeling of Earthquake Occurrences and Estimation of Seismic Hazard: A Random Field Approach," *Probab. Eng. Mech.*, 17, 1–13.

Anh, V. V., Lam, K. C., Leung, Y., & Tieng, Q. (2000), "Multifractal Analysis of Hong Kong Air Quality Data," *Environmetrics*, 10, 139–149.

Anh, V. V., & Lunney, K. E. (1995), "Prediction Estimation of Random Fields With Long-range Dependence," *Mathematical Computational Modelling*, 21, 67–77.

Beran, J. (1993), "Fitting Long-memory Models by Generalized Linear Regression," *Biometrika*, 80, 817–822.

Beran, J. (1994), *Statistics for Long-memory Processes*, New York: Chapman and Hall.

Beran, J., Ghosh, S., & Dieter, S. (2009), "On Least Squares Estimation for Long-memory Lattice Processes," *Journal of Multivariate Analysis*, 100, 2178–2194.

Besag, J. (1981), "On a System of Two-dimensional Recurrence Equations," *Journal of Royal Statistical Society, Series B*, 43, 302–309.

Bhansali, R. (1973a), "A Simulation Study of the Wiener-Kolmogorov Predictor," *Sankhyā: The Indian Journal of Statistics, Series A*, 35(3), 357–376.

Bhansali, R. (1974), "Asymptotic Properties of the Wiener-Kolmogorov Predictor. I," *Journal of Royal Statistical Society, Series B*, 36(1), 61–73.

Bhansali, R. (1977), "Asymptotic Properties of the Wiener-Kolmogorov Predictor. II," *Journal of Royal Statistical Society, Series B*, 39(1), 66–72.

Bloomfield, P. (1973), "An Exponential Model for the Spectrum of a Scalar Time Series," *Biometrika*, 60, 217–226.

Bogert, B. P., Healy, M. P. R., & Tukey, J. W. (1963), *The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking. Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed)*, New York: Wiley.

Boissy, Y., Bhattacharyya, B. B., Li, X., & Richardson, G. D. (2005), "Parameter Estimates for Fractional Autoregressive Spatial Processes," *The Annals of Statistics*, 33, 2553–2567.

Bondon, P. (2002), "Prediction With Incomplete Past of a Stationary Process," *Stochastic Processes and their Applications*, 98(1), 67–76.

Bondon, P. (2005), "Influence of Missing Values on the Prediction of a Stationary Time Series," *Journal of Time Series Analysis*, 26, 519–525.

Brillinger, D. R. (1974), "Fourier Analysis of Stationary Processes," *Proceedings of IEEE*, 62, 628–642.

Brockwell, P. J., & Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer.

Cheng, R., & Pourahmadi, M. (1997), "Prediction With Incomplete Past and Interpolation of Missing Values," *Statistics and Probability Letters*, 33, 341–346.

Cressie, N. A. (1993), *Statistics for Spatial Data*, New York: Wiley.

Dalhaus, R., & Künsch, H. (1987), "Edge Effects and Efficient Parameter Estimation for Stationary Random Fields," *Biometrika*, 74, 877–882.

Davies, R. B., & Harte, D. S. (1987), "Tests for Hurst Effects," *Biometrika*, 74, 95–101.

Fox, R., & Taqqu, M. S. (1986), "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *The Annals of Statistics*, 14, 517–532.

Fox, R., & Taqqu, M. S. (1987), "Central Limit Theorems for Quadratic Forms in Random Variables having Long-range Dependence," *Probab. Theory Related Fields*, 74, 213–240.

Frias, M. P., Alonso, F. J., Ruiz-Medina, M. D., & Angulo, J. M. (2008), "Semiparametric Estimation of Spatial Long-range Dependence," *Journal of Statistical Planning and Inference*, 138, 1479–1495.

Geweke, J., & Porter-Hudak (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Guo, H., Lim, C. Y., & Meerschaert, M. M. (2009), "Local Whittle Estimator for Anisotropic Random Fields," *Journal of Multivariate Analysis*, 100, 993–1028.

Hannan, E. J. (1970), *Multiple Time Series*, New York: Wiley.

Helson, H., & Lowdenslager, D. (1958), "Prediction Theory and Fourier Series in Several Variables," *Acta Math.*, 99, 165–202.

Helson, H., & Lowdenslager, D. (1961), "Prediction Theory and Fourier Series in Several Variables II," *Acta Math.*, 106, 175–213.

Hurvich, C. M. (2001), "Model Selection for Broadband Semiparametric Estimation of Long Memory in Time Series," *Journal of Time Series Analysis*, 22, 679–709.

Hurvich, C. M. (2002), "Multistep Forecasting of Long Memory Series Using Fractional Exponential Models," *International Journal of Forecasting*, 18, 167–179.

Hurvich, C. M., & Brodsky, J. (2001), "Broadband Semiparametric Estimation of the Memory Parameter of a Long-memory Time Series Using Fractional Exponential Models," *Journal of Time Series Analysis*, 22, 221–249.

Hurvich, C. M., Deo, R., & Brodsky, J. (1998), "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series," *Journal of Time Series Analysis*, 19, 19–46.

Hurvich, C. M., & Tsai, C. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297–307.

Kaderli, A., & Kayhan, A. S. (2000), "Spectral Estimation of ARMA Processes Using ARMA-cepstrum Recursion," *IEEE Sigma Processing Letters*, 7(9), 259–261.

Kallianpur, G., & Mandrekar, V. (1983), "Nondeterministic Random Fields and Wold and Halmos Decomposition for Commuting Isometries," *Prediction Theory and Harmonic Analysis, A Pesi Masani Volume, V. Mandrekar and H. Salehi, (Eds)*, pp. 165–190.

Kallianpur, G., Miamee, A. G., & Niemi, H. (1990), "On the Prediction Theory of Two Parameter Stationary Random Fields," *Journal of Multivariate Analysis*, 32, 120–149.

Kasahara, Y., Pourahmadi, M., & Inoue, A. (2009), "Duals of Random Vectors and Processes With Applications to Prediction Problems With Missing Values," *Statistics and Probability Letters*, 79(14), 1637–1646.

Kizilkaya, A., & Kayran, A. H. (2005), "ARMA-cepstrum Recursion Algorithm for the Estimation of the MA Parameters of 2-D ARMA Models," *Multidimensional Systems and Signal Processing*, 16, 397–415.

Kolmogorov, A. N. (1939), "Sur L'interpolation et Extrapolation des Suites Stationnaires," *C. R. Acad. Sci. Paris*, 208, 2043–2045.

Kolmogorov, A. N. (1941), "Stationary Sequences in a Hilbert Space," *Bulletin of Moscow State University*, 2, 1–40.

Krige, D. G. (1951), "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Chem. Metal. and Mining Society of South Africa*, 52, 119139.

Künsch, H. R. (1987), "Statistical Aspects of Self Similar Processes," *In Proceedings of the First World Congress of the Bernoulli Society, Edited by Prohorov, Y. and Sazanov, V. V.*, 1, 67–74.

Lavancier, F. (2005), "Long Memory Random Fields," *Dependence in Probability and Statistics, Springer, New York, Lecture Notes in Statistics*, 187, 195–220.

Lavancier, F. (2011), "Aggregation of Isotropic Autoregressive Fields," *Journal of Statistical Planning and Inference*, 141, 3862–3866.

Lee, C. M., & Berman, M. (1997), "Nonparametric Estimation and Simulation of Two-dimensional Gaussian Image Texture," *Graphical Models and Image Processing*, 59, 434–445.

Lim, Y. K., Kim, K. Y., & Lee, H. S. (2002), "Temporal and Spatial Evolution of the Asian Summer Monsoon in the Seasonal Cycle of Synoptic Fields," *Journal of Climate*, 15, 3630–3644.

Martin, R. J. (1986), "On the Design of Experiments Under Spatial Correlation," *Biometrika*, 73, 247–277.

Matheron, G. (1971), "The Theory of Regionalized Variables," *Les Cahiers du Centre de Morphologie Mathematique de Fountainebleau*, 5.

McBratney, A. B., & Webster, R. (1981), "Detection of Ridge and Furrow Pattern by Spectral Analysis of Crop Yield," *International Statistical Review*, 49, 45–52.

Mercer, W. B., & Hall, A. D. (1911), "The Experimental Error of Field Trials," *Journal of Agricultural Science (Cambridge)*, 4, 107–132.

Miamee, A. G. (1986), "Extension of Three Theorems for Fourier Series on the Disc to the Torus," *Bulletein of Australian Mathematical Society*, 33, 335–350.

Miamee, A. G., & Pourahmadi, M. (1988), "Best Approximation in $\mathrm{L}^p(d\mu)$ and Prediction Problems of Szegö, Kolmogorov, Yaglom, and Nakazi," *Journal of London Mathematical Society*, 38(2), 133–145.

Miller, A. J. (2002), *Subset Selection in Regression*, London: Chapman and Hall/CRC Press.

Moulines, E., & Soulier, P. (1999), "Broadband Log-periodogram Regression of Time Series With Long-range Dependence," *Annals of Statistics*, 27, 1415–1439.

Nakazi, T. (1984), "Two Problems in Prediction Theory," *Studia Math.*, 78, 7–14.

Pearce, S. C. (1976), "An Examination of Fairfield Smiths Law of Environmental Variation," *Journal of Agricultural Science*, 87, 21–24.

Pourahmadi, M. (1983), "Exact Factorization of the Spectral Density and its Application to Forecasing and Time Series Analysis," *Communication in Statistics Theory and Methods*, 12, 2085–2094.

Pourahmadi, M. (1984), "Taylor Expansion of $\exp(\sum_{k=1}^{\infty} a_k z^k)$ and Some Applications.," *The American Mathematical Monthly*, 91(5), 303–307.

Pourahmadi, M. (1989), "Estimation and Interpolation of Missing Values of a Stationary Time Series," *Journal of Time Series Analysis*, 10, 149–169.

Pourahmadi, M. (2001), *Foundation of Time Series Analysis and Prediction Theory*, New York: John Wiley and Sons.

Pourahmadi, M., Inoue, A., & Kasahara, Y. (2006), "A Prediction Problem in $\mathrm{L}^2(w)$," *Proceedings of American Mathematical Society*, 135(4), 1233–1239.

Pourahmadi, M., & Soofi, E. S. (2000), "Prediction Variance and Information Worth of Observations in Time Series," *Journal of Time Series Analysis*, 21, 413–434.

Ripley, B. D. (1981), *Spatial Statistics*, New York: Wiley.

Robinson, P. M. (1994), "Semiparametric Analysis of Long Memory Time Series," *The Annals of Statistics*, 22, 515–539.

Robinson, P. M. (1995), "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23, 1630–1661.

Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*, Boston: Birkhauser.

Solo, V. (1986), "Modeling of Two-dimensional Random Fields by Parametric Cepstrum," *IEEE Information Theory*, 32, 743–750.

Soltani, A. R. (1984), "Extrapolation and Moving Average Representation for Stationary Random Fields and Beurling's Theorem," *Annals of Probability*, 12, 120–132.

Tjøstheim, D. (1983), "Statistical Spatial Series Modelling II: Some Further Results in Unilateral Processes," *Advances in Applied Probability*, 15, 562–584.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Wang, L. (2009), "Memory Parameter Estimation for Long Range Dependent Random Fields," *Statistics and Probability Letters*, 79, 2297–2306.

Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.

Whittle, P. (1956), "On the Variation of Yield Variance With Plot Size," *Biometrika*, 43, 337–343.

Whittle, P. (1962), "Topographic Correlation, Power-Law Covariance Functions, and Diffusion," *Biometrika*, 49, 305–314.

Wiener, N. (1949), *Time Series*, Cambridge: MIT Press.

Yao, Q., & Brockwell, P. J. (2006), "Gaussian Maximum Likelihood Estimation for ARMA Models II: Spatial Processes," *Bernoulli*, 12(3), 403–429.

Young, L. J., & Young, J. H. (1998), *Statistical Ecology: A Population Perspective*, Boston: Kluwer Academic Publishers.

Zimmerman, D. L. (2006), "Optimal Network Design for Spatial Prediction, Covariance Parameter Estimation, and Empirical Prediction," *Environmetrics*, 17, 635–652.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER II

Proof of Theorem 1

Using the MA(∞) representation of $\{X(s_1, s_2)\}$ in (2.14), orthogonality of the innovations and suitably partitioning the infinite sums in terms of the indices belonging to Q and its complement Q^c, we obtain

$$X(h_1,h_2) = b_{h_1,h_2}\varepsilon(0,0) + \sum_{\substack{k=h_1 \\ (k,\ell)\neq(h_1,h_2)}}^{\infty}\sum_{\ell=h_2}^{\infty} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell) + \sum_{k=0}^{h_1-1}\sum_{\ell=h_2}^{\infty} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell)$$
$$+ \sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell) + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell).$$

Using linearity of the orthogonal projection and the second equality in (2.14), we find that the (h_1, h_2)-step ahead predictor is

$$\widehat{X}(h_1,h_2) = \sum_{\substack{k=h_1 \\ (k,\ell)\neq(h_1,h_2)}}^{\infty}\sum_{\ell=h_2}^{\infty} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell).$$

The corresponding prediction error is

$$X(h_1,h_2) - \widehat{X}(h_1,h_2) = b_{h_1,h_2}\varepsilon(0,0) + \sum_{k=0}^{h_1-1}\sum_{\ell=h_2}^{\infty} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell)$$
$$+ \sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell) + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b_{k,\ell}\varepsilon(t_1+h_1-k,t_2+h_2-\ell),$$

with variance

$$\text{var}\left\{X(h_1,h_2)-\widehat{X}(h_1,h_2)\right\} = b^2_{h_1,h_2}\text{var}\left\{\varepsilon(0,0)\right\} + \sum_{k=0}^{h_1-1}\sum_{\ell=h_2}^{\infty} b^2_{k,\ell}\text{var}\{\varepsilon(h_1-k,h_2-\ell)\}$$
$$+\sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell}\text{var}\{\varepsilon(h_1-k,h_2-\ell)\} + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell}\text{var}\{\varepsilon(h_1-k,h_2-\ell)\}$$
$$= \sigma^2\left(b^2_{h_1,h_2} + \sum_{k=0}^{h_1-1}\sum_{\ell=h_2}^{\infty} b^2_{k,\ell} + \sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell} + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell}\right).$$

For quarter plane past, Tjøstheim (1983) pointed out that a wide range of homogeneous lattice processes can be approximated in terms of unilateral AR models as:

$$X(s_1,s_2) = -\sum_{\substack{k\\(k,\ell)\in Q}}\sum_{\ell} a_{k,\ell}X(s_1-k,s_2-\ell) + \varepsilon(s_1,s_2), \tag{A.1}$$

where $\{a_{k,\ell}\}$ are the AR parameters of X with $a_{0,0} = 1$, $a_{k,\ell} = 0$, when either $k < 0$ or $l < 0$, and the double infinite sum converges in the mean-square sense. The AR parameters are same as the coefficients in Taylor expansion of inverse of $\phi(z_1,z_2)$ in (2.15). Partitioning the sums in (A.1) into two parts corresponding to the indices belonging to Q and its complement Q^c we obtain (2.18).

Q.E.D

Proof of Corollary 2

(a). Figure 9 depicts the 1-D process $Y(s,h_2) = \{\cdots, Y(-3,h_2), Y(-2,h_2), Y(-1,h_2)\}$ of prediction errors in the second quadrant, along the rows $h_2 = 1,2,3,\cdots$. From (2.16), the prediction error of $X(h_1,h_2)$ with $h_1 < 0$ and $h_2 > 0$ is:

$$Y(h_1,h_2) = X(h_1,h_2) - \widehat{X}(h_1,h_2) = \sum_{k=0}^{\infty}\sum_{\ell=0}^{h_2-1} b_{k,\ell}\varepsilon(h_1-k,h_2-\ell).$$

Using (2.19) for $s_i < 0, i = 1,2$ the covariance between $Y(s_1,h_2)$ and $Y(s_2,h_2)$ is as in (2.21). Substituting $s_1 = s_2 = s$ in (2.21), the variance is trivial. By definition, for a

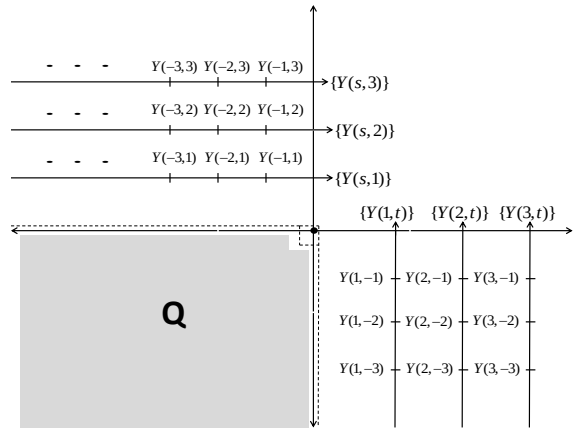

Figure 9: Prediction Errors in Quadrants II and IV.

stationary 1-D process $\{Y(s, h_2)\}$ with auto-covariance function $\gamma(.)$, the spectral density function is

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-ih\lambda), \tag{A.2}$$

substituting $\gamma_{h_2}(h)$ from (2.21) into (A.2) we get,

$$f_{h_2}(\lambda) = \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \sum_{\ell=0}^{h_2-1} b_{k,\ell} b_{k+|h|,\ell} \right) \exp(-ih\lambda).$$

Let $t = k + h$, such that, $h = t - k$ and $0 < t < \infty$, then

$$\begin{aligned}
f_{h_2}(\lambda) &= \frac{\sigma^2}{2\pi} \sum_{t=0}^{\infty} \sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{h_2-1} b_{k,\ell} b_{t,\ell} \right) \exp(-i(t-k)\lambda) \\
&= \frac{\sigma^2}{2\pi} \sum_{t=0}^{\infty} \sum_{k=0}^{\infty} (b_{k,0} b_{t,0} + b_{k,1} b_{t,1} + \cdots + b_{k,h_2-1} b_{t,h_2-1}) \exp(-i(t-k)\lambda) \\
&= \frac{\sigma^2}{2\pi} \left\{ \left| \sum_{k=0}^{\infty} b_{k,0} \exp(ik\lambda) \right|^2 + \left| \sum_{k=0}^{\infty} b_{k,1} \exp(ik\lambda) \right|^2 + \cdots + \left| \sum_{k=0}^{\infty} b_{k,h_2-1} \exp(ik\lambda) \right|^2 \right\} \\
&= \frac{\sigma^2}{2\pi} \sum_{\ell=0}^{h_2-1} \left| \sum_{k=0}^{\infty} b_{k,\ell} \exp(ik\lambda) \right|^2,
\end{aligned}$$

which is (2.22). Proof of (b) is similar, and hence omitted.

Outline of the proof of Lemma 4

(a) holds since $\{X(s_1, s_2)\}$ is PND. For (b),

$$Q \oplus A \subseteq I,$$

it suffices to note that any $X \in I$ can be written as

$$X = Y + \sum_{(k,\ell)\in K}\sum d_{k,\ell} X(k,\ell),$$

for some $Y \in Q$ and scalars $\{d_{k,\ell}\}$. Adding and subtracting $\sum_{(k,\ell)\in K}\sum d_{k,\ell}\widehat{X}(k,\ell)$ gives

$$X = Y + \sum_{(k,\ell)\in K}\sum d_{k,\ell}\widehat{X}(k,\ell) + \sum_{(k,\ell)\in K}\sum d_{k,\ell}\left(X(k,\ell) - \widehat{X}(k,\ell)\right) \in Q \oplus K.$$

Proof of Theorem 2

Since A is one-dimensional, from Lemma 4(d) we have,

$$P_A^{X(0,0)} = \beta_{h_1,h_2}\left(X(h_1,h_2) - \widehat{X}(h_1,h_2)\right),$$

where β_{h_1,h_2} is obtained from Corollary 1(a)-(b) as:

$$\begin{aligned}\beta_{h_1,h_2} &= \left\{\text{var}\left(X(h_1,h_2) - \widehat{X}(h_1,h_2)\right)\right\}^{-1} \text{cov}\left(X(0,0), X(h_1,h_2) - \widehat{X}(h_1,h_2)\right) \\ &= \left\{\sigma^2\left(b_{h_1,h_2}^2 + \sum_{k=0}^{h_2-1}\sum_{\ell=h_2}^{\infty} b_{k,\ell}^2 + \sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b_{k,\ell}^2 + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b_{k,\ell}^2\right)\right\}^{-1} \left(\sigma^2 b_{h_1,h_2}\right).\end{aligned} \tag{A.3}$$

The prediction error of $\widehat{X}_{I_1}(0,0)$ is

$$\begin{aligned}X(0,0) - \widehat{X}_{I_1}(0,0) &= X(0,0) - \widehat{X}(0,0) - \beta_{h_1,h_2}\left(X(h_1,h_2) - \widehat{X}(h_1,h_2)\right), \\ &= \varepsilon(0,0) - \beta_{h_1,h_2}\left(X(h_1,h_2) - \widehat{X}(h_1,h_2)\right).\end{aligned}$$

Computing the variance of the above prediction error gives

$$\begin{aligned}\sigma^2(I_1) &= \text{var}\left(X(0,0) - \widehat{X}_{I_1}(0,0)\right)\\ &= \text{var}\left\{\varepsilon(0,0) - \beta_{h_1,h_2}\left(X(h_1,h_2) - \widehat{X}(h_1,h_2)\right)\right\}\\ &= \sigma^2\left\{1 - \beta_{h_1,h_2}\text{cov}\left(X(0,0), X(h_1,h_2) - \widehat{X}(h_1,h_2)\right)\right\}. \qquad \text{(A.4)}\end{aligned}$$

Substituting β_{h_1,h_2} from (A.3) into (A.4) and using Corollary 1(b), it follows that

$$\sigma^2(I_1) = \sigma^2\left(1 - \frac{b^2_{h_1,h_2}}{b^2_{h_1,h_2} + \sum_{k=0}^{h_1-1}\sum_{\ell=h_2}^{\infty} b^2_{k,\ell} + \sum_{k=h_1}^{\infty}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell} + \sum_{k=0}^{h_1-1}\sum_{\ell=0}^{h_2-1} b^2_{k,\ell}}\right).$$

Q.E.D

Proof of Lemma 5

Representing the three double sums in the prediction error of $X(h_1,h_2)$ given in (2.16) in a matrix form, we can write

$$X(h_1,h_2) - \widehat{X}(h_1,h_2) = b_{h_1,h_2}\varepsilon(0,0) + [T'_{n-1}, T'_{n-2}, T'_{n-3}, \cdots, T'_0]\varepsilon_K,$$

where T_{i-1} is a vector of MA parameters for the i-th observation in set K with $i = 1, \cdots, n$. For instance, if $i = n$ then the additional observation is $X(h_1,h_2)$ and

$$T'_{n-1} = \{b_{0,0}, b_{1,0}, b_{2,0}, \cdots, b_{0,1}, b_{0,2}, \cdots\}$$

$$\vdots$$

$$T'_0 = \{b_{h_1-1,h_2-1}, b_{h_1,h_2-1}, b_{h_1+1,h_2-1}, \cdots, b_{h_1-1,h_2}, b_{h_1-1,h_2+1}, \cdots\}.$$

Likewise, writing the prediction error for each observation in the set K we get

$$X_K - \widehat{X}_K = \begin{pmatrix} X(h_1,h_2) - \widehat{X}(h_1,h_2)\\ X(h_1-1,h_2) - \widehat{X}(h_1-1,h_2)\\ X(h_1-2,h_2) - \widehat{X}(h_1-2,h_2)\\ \cdot\\ \cdot\\ X(1,1) - \widehat{X}(1,1)\\ X(1,0) - \widehat{X}(1,0)\\ X(0,1) - \widehat{X}(0,1)\end{pmatrix} = \begin{pmatrix} b_{h_1,h_2} & T'_{n-1} & T'_{n-2} & \cdots & T'_0\\ b_{h_1-1,h_2} & T'_{n-2} & T'_{n-3} & \cdots & 0'\\ b_{h_1-2,h_2} & T'_{n-3} & T'_{n-4} & \cdots & 0'\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ b_{1,1} & T'_0 & 0' & \cdots & 0'\\ b_{1,0} & b'_{10} & 0' & \cdots & 0'\\ b_{0,1} & 0' & b'_{01} & \cdots & 0' \end{pmatrix}\begin{pmatrix}\varepsilon(0,0)\\ \varepsilon_K\end{pmatrix},$$

which is

$$X_K - \widehat{X}_K = \begin{pmatrix} b_K & T' \end{pmatrix} \begin{pmatrix} \varepsilon(0,0) \\ \varepsilon_K \end{pmatrix} = b_K \varepsilon(0,0) + T'\varepsilon_K.$$

Part (b) follows from (a) and orthogonality of the innovations as:

$$\begin{aligned} C = \text{cov}\left(X_K - \widehat{X}_K\right) &= \text{cov}\left\{b_K \varepsilon(0,0) + T'\varepsilon_K\right\} \\ &= \sigma^2 b_K b_K' + T'(\sigma^2 I)T \\ &= \sigma^2 (b_K b_K' + T'T). \end{aligned}$$

Part (c) is immediate from the recursive relation (2.26) between MA and AR parameters.

Proof of Theorem 3

From Lemma 4(c)-(d), the best linear predictor of $X(0,0)$ based on I_2 is:

$$\begin{aligned} \widehat{X}_{I_2}(0,0) &= \widehat{X}(0,0) + \sum_{(i,j)\in K}\sum \beta_{i,j}\left\{X(i,j) - \widehat{X}(i,j)\right\} \\ &= \widehat{X}(0,0) + \beta'(X_K - \widehat{X}_K), \end{aligned}$$

where

$$\begin{aligned} \beta &= \left\{\text{var}\left(X_K - \widehat{X}_K\right)\right\}^{-1} \text{cov}\left(X_K - \widehat{X}_K, X(0,0) - \widehat{X}(0,0)\right) \\ &= \left\{\sigma^2(T'T + b_K b_K')\right\}^{-1}(\sigma^2 b_K) \\ &= (T'T + b_K b_K')^{-1} b_K. \end{aligned} \tag{A.5}$$

The inverse of $(T'T + b_K b_K')$ is found formally by using a matrix inversion identity Pourahmadi (2001, p. 155) for a rank-1 perturbation of a nonsingular matrix G:

$$(G + UV')^{-1} = G^{-1} - (1 + V'G^{-1}U)^{-1}(G^{-1}U)(V'G^{-1}), \tag{A.6}$$

where U and V are column vectors. Using $G = T'T$ and $U = V = b_K$ in (A.6) and substituting the inverse into (A.5), we get

$$\begin{aligned}
\beta &= \left\{G^{-1} - (1 + b_K' G^{-1} b_K)^{-1} (G^{-1} b_K)(b_K' G^{-1})\right\} b_K \\
&= G^{-1} b_K - (1 + b_K' G^{-1} b_K)^{-1} (G^{-1} b_K)(b_K' G^{-1} b_K) \\
&= G^{-1} b_K \left\{1 - (1 + b_K' G^{-1} b_K)^{-1} (b_K' G^{-1} b_K)\right\} \\
&= G^{-1} b_K \left(1 + b_K' G^{-1} b'\right)^{-1}.
\end{aligned}$$

The prediction error for $\widehat{X}_{I_2}(0,0)$ is obtained by using Lemma 5(a) for $X_K - \widehat{X}_K$ as:

$$\begin{aligned}
X(0,0) - \widehat{X}_{I_2}(0,0) &= X(0,0) - \widehat{X}(0,0) - \beta' \left(X_K - \widehat{X}_K\right) \\
&= \varepsilon(0,0) - \beta' \left(b_K \varepsilon(0,0) + T' \varepsilon_K\right). \\
&= \varepsilon(0,0)(1 - \beta' b_K) - (T\beta)' \varepsilon_K.
\end{aligned}$$

Using the orthogonality of innovations, the prediction error variance is

$$\begin{aligned}
\sigma^2(I_2) &= \text{var}\left\{X(0,0) - \widehat{X}_{I_2}(0,0)\right\} \\
&= \text{var}\left\{\varepsilon(0,0)(1 - \beta' b_K) - (T\beta)' \varepsilon_K\right\} \\
&= \sigma^2 \left\{(1 - \beta' b_K)(1 - \beta' b_K)' + \beta'(T'T)\beta\right\}. \qquad \text{(A.7)}
\end{aligned}$$

Working on simplification of the terms

$$\begin{aligned}
1 - \beta' b_K &= 1 - b_K' C^{-1} b_K \\
&= 1 - b_K' \left\{G^{-1} - (1 + b_K' G^{-1} b_K)^{-1} (G^{-1} b_K)(b_K' G^{-1})\right\} b_K \\
&= 1 - \left\{(b_K' G^{-1} b_K)(1 + b_K' G^{-1} b_K)^{-1}\right\} \\
&= \left(1 + b_K' G^{-1} b_K\right)^{-1}, \qquad \text{(A.8)}
\end{aligned}$$

and

$$\begin{aligned}\beta'(T'T)\beta &= \beta' G \beta \\ &= \left\{\left(1 + b_K' G^{-1} b_K'\right)^{-1} b_K' G^{-1}\right\} G \left\{G^{-1} b_K \left(1 + b_K' G^{-1} b_K'\right)^{-1}\right\} \\ &= \left(1 + b_K' G^{-1} b_K'\right)^{-2} b_K' G^{-1} b_K. \end{aligned} \tag{A.9}$$

Substituting (A.8) and (A.9) into (A.7) we get

$$\begin{aligned}\sigma^2(I_2) &= \sigma^2 \left\{\left(1 + b_K' G^{-1} b_K'\right)^{-2} + \left(1 + b_K' G^{-1} b_K'\right)^{-2} \left(b_K' G^{-1} b_K\right)\right\} \\ &= \sigma^2 \left(1 + b_K' G^{-1} b_K'\right)^{-2} \left(1 + b_K' G^{-1} b_K'\right) \\ &= \sigma^2 \left(1 + b_K' G^{-1} b_K'\right)^{-1}\end{aligned}$$

Using (2.35) and $P = T(T'T)^{-1}T'$, we get

$$\begin{aligned}\sigma^2(I_2) &= \sigma^2 \left\{1 + (T'a_K)' G^{-1} (T'a_K)\right\}^{-1} \\ &= \sigma^2 \left\{1 + a_K' T (T'T)^{-1} T' a_K\right\}^{-1} \\ &= \sigma^2 \left(1 + a_K' P a_K\right)^{-1}.\end{aligned}$$

Q.E.D

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER III

Proof of Theorem 4

Differentiating both sides of (3.17) with respect to z_1 we get;

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)b_{k+1,\ell}z_1^k z_2^\ell = \left(\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell}z_1^k z_2^\ell\right)\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)c_{k+1,\ell}z_1^k z_2^\ell, \quad \text{(B.10)}$$

or equivalently,

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)b_{k+1,\ell}z_1^k z_2^\ell = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\left\{\sum_{j_1=0}^{k}\sum_{j_2=0}^{\ell}(k+1-j_1)c_{k+1-j_1,\ell-j_2}b_{j_1,j_2}\right\}z_1^k z_2^\ell. \quad \text{(B.11)}$$

Similarly, differentiating both sides of (3.18) with respect to z_2 we get;

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(\ell+1)b_{k,\ell+1}z_1^k z_2^\ell = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\left\{\sum_{j_1=0}^{k}\sum_{j_2=0}^{\ell}(\ell+1-j_2)c_{k-j_1,\ell+1-j_2}b_{j_1,j_2}\right\}z_1^k z_2^\ell. \quad \text{(B.12)}$$

Equating the coefficients of like powers of z_1 and z_2 on both sides of (B.11) and (B.12) leads to expressions Theorem 4(a) and (b) for $\{b_{k,\ell}\}$, respectively.

Now differentiating both sides of (B.10) with respect to z_2 we get;

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)(\ell+1)b_{k+1,\ell+1}z_1^k z_2^\ell = \left\{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} b_{k,\ell}z_1^k z_2^\ell \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)(\ell+1)c_{k+1,\ell+1}z_1^k z_2^\ell\right\}$$
$$+\left\{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(\ell+1)b_{k,\ell+1}z_1^k z_2^\ell \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)c_{k+1,\ell}z_1^k z_2^\ell\right\},$$

or equivalently,

$$\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(k+1)(\ell+1)b_{k+1,\ell+1}z_1^k z_2^\ell = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\Big\{\sum_{j_1=0}^{k}\sum_{j_2=0}^{\ell}(k+1-j_1)(\ell+1-j_2)c_{k+1-j_1,\ell+1-j_2}b_{j_1,j_2} +$$
$$(k+1-j_1)(j_2+1)c_{k+1-j_1,\ell-j_2}b_{j_1,j_2+1}\Big\}z_1^k z_2^\ell.$$

Derivation of (d)-(f) for $\{a_{k,\ell}\}$ is similar, and hence omitted.

Q.E.D

APPENDIX C

SUPPLEMENTARY MATERIAL FOR CHAPTER IV

Lee and Berman's (1997) Algorithm

To simulate data from a Gaussian random field for given spectral density $f(\lambda_1, \lambda_2)$. Suppose the required process is $\{X(t_1, t_2)\}$, $t_1 = 1, \cdots, n_2$, $t_2 = 1, \cdots, n_2$ on a regular lattice. Then, the algorithm proceeds as follows:

1 Compute $g(\lambda_1, \lambda_2) = \sqrt{f(\lambda_1, \lambda_2)}$.

2 Let $\varepsilon(t_1, t_2)$ be a white noise process with zero mean and unit variance. Compute the discrete Fourier transform of Z as:

$$\tilde{\varepsilon}(\lambda_{1j}, \lambda_{2k}) = \frac{1}{\sqrt{n_1 n_2}} \sum_{s_1=0}^{n_1-1} \sum_{s_2=0}^{n_2-1} \varepsilon(t_1, t_2) \exp\{-i(s_1\lambda_{1j} + s_2\lambda_{2k})\}.$$

Let u and v be the real and imaginary parts respectively of $\tilde{\varepsilon}(\lambda_{1j}, \lambda_{2k})$.

3 Calculate the inverse Fourier transform of $h = (gu + igv)$. The real part of this inverse Fourier transform gives a realization of the Gaussian random field with spectral density $f(\lambda_1, \lambda_2)$ such that,

$$X(t_1, t_2) = \frac{2\pi}{n_1 n_2} \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} \sqrt{f(\lambda_{1j}, \lambda_{2k})} \left[\sum_{s_1=0}^{n_1-1} \sum_{s_2=0}^{n_2-1} \varepsilon(s_1, s_2) \exp\{-i(s_1\lambda_{1j}) + s_2(\lambda_{2k})\} \right] \exp\{i(j\lambda_{1t_1} + k\lambda_{2t_2})\}.$$

The above algorithm is used to simulate data from a Gaussian random field first on a $N_1 \times N_2$ rectangular lattice where $N_i > n_i, i = 1, 2$ and N_i is not a multiple of n_i. Then we select the required process $X(t_1, t_2)$ as a subset of the simulated process. A slight modification to avoid the zero frequencies at which the spectral density function takes very high values

is as follows:

$$X(t_1,t_2) = \frac{2\pi}{N_1N_2}\sum_{j=0}^{N_1-1}\sum_{k=0}^{N_2-1}\sqrt{f(\lambda_{1\tilde{j}},\lambda_{2\tilde{k}})}\left[\sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1}\varepsilon(s_1,s_2)\exp\{-i(s_1\lambda_{1j})+s_2(\lambda_{2k})\}\right]$$
$$\exp\{i(j\lambda_{1\tilde{t_1}}+k\lambda_{2\tilde{t_2}})\},$$

where $\tilde{j}=j+1/2$, $\tilde{k}=k+1/2$ so that $\lambda_{1\tilde{j}}=2\pi(j+1/2)/N_1$ and $\lambda_{2\tilde{k}}=2\pi(k+1/2)/N_2$.

Proof: Above algorithm produces process with given spectral density function

For s stationary random field $\{X(t_1,t_2), t_1=1,\cdots,n_1, t_2=1,\cdots,n_2\}$ with spectral density $f(.,.)$:

$$X(t_1,t_2) = \frac{2\pi}{n_1n_2}\sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1}\sqrt{f(\lambda_{j1},\lambda_{j2})}\left[\sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1}\varepsilon(s_1,s_2)\exp\left\{-i\left(\lambda_{j1}s_1+\lambda_{j2}s_2\right)\right\}\right]$$
$$\exp\left\{i\left(\lambda_{t1}j1+\lambda_{t2}j_2\right)\right\}.$$

Using the definition of the autocovariance and simplifying we get:

$$E\left[X(t_1,t_2)X(t_1+h_1,t_2+h_2)\right] = \left(\frac{2\pi}{n_1n_2}\right)^2$$
$$\sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1}\sum_{k_1=0}^{n_1-1}\sum_{k_2=0}^{n_2-1}\sqrt{f(\lambda_{j_1},\lambda_{j_2})f(\lambda_{k_1},\lambda_{k_2})}\left[\sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1}\exp\left\{-i\left(\lambda_{j_1+k_1}s_1+\lambda_{j_2+k_2}s_2\right)\right\}\right]$$
$$\times\exp\left\{i\left(\lambda_{t_1}j_1+\lambda_{t_2}j_2\right)\right\}\exp\left\{i\left(\lambda_{t_1+h_1}k_1+\lambda_{t_2+h_2}k_2\right)\right\}$$

$$= \left(\frac{2\pi}{n_1n_2}\right)^2 n_1n_2\sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1}f(\lambda_{j_1},\lambda_{j_2})\exp\left\{-i\left(\lambda_{h_1}j_1+\lambda_{h_2}j_2\right)\right\}$$
$$= \frac{4\pi^2}{n_1n_2}\sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1}f(\lambda_{j_1},\lambda_{j_2})\exp\left\{-i\left(\lambda_{j_1}h_1+\lambda_{j_2}h_2\right)\right\}.$$

VITA

Priya Kohli was born in New Delhi, India. She received a Bachelor of Science Honors degree in Physics in 2003 from Delhi University, India. She continued on to obtain a Master of Science degree in Agricultural Statistics in 2006 from Indian Agricultural Statistics Research Institute, and a second Master of Science degree in Applied Probability and Statistics from the Northern Illinois University, Illinois in 2009. In January 2009, she started the doctoral program at Texas A&M University, and in August 2012, she received her Ph.D. degree in statistics under the advisement of Profs. Mohsen Pourahmadi and Willa W. Chen. Her mailing address is:

3143 TAMU,
Department of Statistics,
Texas A&M University,
College Station, TX 77845-3143.