Spatio-Temporal Models by Wavelets

Yangyang Chen

THESIS SUBMITTED

TO

INSTITUTE OF MATHEMATICS AND STATISTICS

OF

UNIVERSITY OF SÃO PAULO

TO

OBTAIN THE TITLE

OF

DOCTORATE IN SCIENCE

Program: Statistics

Advisor: Prof. Dr. Pedro Alberto Morettin

Prof. Dr. Ronaldo Dias

SÃO PAULO, 2023

Spatio-Temporal Models by Wavelets

This version of the thesis includes the corrections and modifications suggested by the Examining Committee during the defense of the original version of the work, which took place on April 26, 2023. A copy of the original version is available at the Institute of Mathematics and Statistics of the University of São Paulo.

Examining Committee:

- Prof. Dr. Pedro Alberto Morettin (orientador) - IME-USP
- Profa. Dra. Chang Chiann - IME-USP
- Prof. Dr. João Ricardo Sato - UFABC
- Prof. Dr. Reinaldo Castro Souza - PUC-RJ
- Profa. Dra. Telma Safadi - UFLA

Acknowledgements

First of all, I would like to give my heartfelt thanks to my advisors, Prof. Pedro and Prof. Ronaldo, for their patience, support and kindness. Besides my advisors, I must express my sincere thanks to Prof. Chang, not only for professional help, but also for the friendship. Also, I am very thankful to all the people who have ever helped me in this thesis.

I am grateful to my parents and Tudou, for supporting me spiritually and encouragement. I also would like to thank my friends who gave me their help and time to listening to me.

I would like to thank the professors and staff of the Department of Statistic, Institute of Mathematics and Statistics, University of São Paulo, for their guidance and support during my university years. Finally, my acknowledgement go to FAPESP, for financial support through grant 2019/05917-6.

Abstract

The space-time autoregressive moving average model is one of the models that is frequently used in several studies of multivariate time series data. In time series analysis, the assumption of stationarity is important, but it is not always guaranteed in practice and one way to proceed is to consider the locally stationary process. In this thesis we propose a time-varying spatio-temporal model based on the local stationarity assumption. The time-varying parameters are expanded as a linear combination of the wavelet bases and some estimation procedures are used to estimate the coefficients. Some simulations were realized to study the performance of the algorithm and the effects of different types of the spatial weights matrices. And then, an application to historical daily precipitation records of Midwestern states of the USA is illustrated.

For the non stationary case, a procedure for estimating the non stationary spatial covariance function for spatio-temporal deformation was proposed. The procedure is based on a monotonic function approach and the functions are expanded using wavelet bases. The deformation proposed guarantees a injective transformation. That is, two distinct locations in the geographic plane are not mapped into the same point in the deformation plane. Finally, some simulations and an application to historical daily maximum temperature records are illustrated.

Keywords: Kalman filter, Locally stationary processes, Spatial covariance function, Spatio-temporal, Non stationary processes, Time-varying, Wavelets.

Resumo

O modelo autoregressivo e média móvel espaço-temporal é um dos modelos frequentemente utilizados em diversos estudos de séries temporais multivariadas. Nesta análise, a suposição de estacionariedade é importante, mas nem sempre é garantida na prática e uma forma de proceder é considerar o processo localmente estacionário. Nesta tese propomos um modelo espaço-temporal variando no tempo, baseado na suposição de estacionariedade local. Os parâmetros variando no tempo são expandidos como uma combinação linear de ondaletas e alguns procedimentos de estimação são usados para estimar os coeficientes. Simulações são realizadas para estudar o desempenho do algoritmo e os efeitos dos diferentes tipos de matrizes de pesos espaciais. Em seguida, é ilustrada uma aplicação aos registros históricos diários de precipitação dos estados do meio-oeste dos EUA.

Para o caso não estacionário, propomos um procedimento para estimar a função de covariância espacial não estacionária e estudamos o problema de deformação no espaço e tempo. O procedimento é baseado em uma abordagem de função monótona e as funções são expandidas usando bases de ondaletas. A deformação proposta garante uma transformação injetiva. Ou seja, duas localizações distintas no plano geográfico não são mapeadas no mesmo ponto no plano deformado. Por fim, simulações e uma aplicação aos registros históricos diários de temperatura máxima são ilustradas.

Palavras-chave: Filtro de Kalman, Processos localmente estacionários, Função de covariância espacial, Espaço-temporal, Processos não estacionários, Variante no tempo, Ondaletas.

Contents

List of Abbreviations

SARIMA	Seasonal Autoregressive Integrated Moving Average
VARMA	Vector Autoregressive and Moving Average
STARMA	Space-Time Autoregressive and Moving Average
tvVAR	time-varying Vector Autoregressive
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
TVP	Time-Varying Parameter
tvSTARMA	time-varying Space-Time Autoregressive and Moving Average
G plane	Geographic Domain
D plane	Deformation Domain
SOS	Second-order Stationarity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive and Moving Average
AR	Autoregressive
MA	Moving Average
STAR	Space-Time Autoregressive
STMA	Space-Time Moving Average
STARIMA	Space-Time Autoregressive Integrated and Moving Average
STACF	Space-Time Autocorrelation Function
STPACF	Space-Time Partial Autocorrelation Function
FDA	Functional Data Analysis
SSE	Sum of Squared of Errors
FT	Fourier Transform
ECG	Electrocardiography
MRA	Multiresolution Analysis
MSE	Mean Square Error
tvSTAR	time-varying Space-Time Autoregressive
GHCN	Global Historical Climatology Network
NOAA	National Oceanic and Atmospheric Administration

List of Symbols

$\mathbb{Z}$	The set of integer numbers
$\mathbb{R}$	The set of real numbers
$Z_i(t)$	A spatio-temporal data at location i and discrete time t
$\mu(\cdot)$	Mean function
$C(\cdot)$	Covariance function
$E[\cdot]$	Expected value
B	Backward shift operator
$\psi(B)$	Autoregressive operator
$\theta(B)$	Moving average operator
Δ	Differencing operator
$\Psi(B^k)$	Seasonal autoregressive operator with seasonal period k
$\Theta(B^k)$	Seasonal moving average operator with seasonal period k
Δ_k	seasonal differencing operator with seasonal period k
$L^{(l)}$	Spatial lag operator with spatial order l
$W^{(l)}$	Spatial weight matrix of order l
d_{ij}	Distance between regions i and j
w_{ij}	spatial relationship between regions i and j
D^m	The mth derivative operator
D^{-1}	The integration operator
L^p	L^p space
$\phi(t)$	Scaling function
$\psi(t)$	Wavelet function
$\phi_{j,k}(t)$	Scaling function under binary dilation and dyadic translation
$\psi_{j,k}(t)$	Wavelet function under binary dilation and dyadic translation
$\mathcal{N}$	Normal distribution
$L(\cdot)$	Likelihood function
C^k	Class C^k
I_n	Identity matrix of size n

List of Figures

List of Tables

Chapter 1

Introduction

Since Box and Jenkins (1970) introduced the class of seasonal autoregressive integrated moving average (SARIMA) model as a procedure for identifying, estimating, and checking models for a specific time series dataset, the method became standard for analyzing stationary and homogeneous non stationary time series with constant coefficients and variance. The models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). In the case of multivariate time series that are stationary, models of the vector autoregressive and moving average (VARMA) family are often used.

The space-time autoregressive and moving average (STARMA) model is a special case of the VARMA model. Cliff and Ord (1975) and Martin and Oeppen (1975) were the first to use models of the class STARMA and then, several methods were developed by Pfeifer and Deutsch (1980a, 1980b, 1981a, 1981b, 1981c) at the early eighties. The models are characterized by linear dependence lagged in both space and time, that is, the modeling processes are characterized by a random variable observed at n geographic locations and, at each location, T observations over time. In this case, in addition to recent past values have more influence, close locations also influence more than distant locations, through the specification of spatial weight matrices which give the highest weights to the nearest neighbors. The STARMA model has already been widely for different types of spatio-temporal data for example, real estate price (Pace et al. (2000)), traffic flow data (Kamarianakis and Prastacos (2005)), damage detection (Hu et al. (2011)), regional bank deposits (Kurt and Tunay (2015)), wind power (Zou et al. (2018)), etc.

As the models of class SARIMA, the STARMA models also have the assumption of stationarity, that is, the distribution does not change when shifted in the origin of the index set, however, in practice, this assumption is not warranted. There are many well known techniques that convert non stationary time series into stationary ones, such as differencing, log transformation, detrending, etc. An attractive alternative is the idea of locally stationary process which the process is approximately stationary over small periods of time, but whose characteristics (covariances, parameters, etc.) are

gradually changing throughout the time period. Many estimators and asymptotic results were developed by Dahlhaus (1996a, 1996b, 1996c, 1997, 2000) and Dahlhaus et al. (1999) and an overview of locally stationary process can be found in Dalhaus (2012).

Time-varying models based on the locally stationary process introduced by Dahlhaus have been used over the years in many different cases, for example, Chiann and Morettin (1999, 2005) investigated the estimation of time varying coefficients of a linear system, Sato et al. (2007) proposed an estimation procedure for time-varying vector autoregressive (tvVAR) models and applied to functional magnetic resonance imaging dates, Rohan and Ramanathan (2013) introduced a nonparametric estimation to time-varying Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and Yousuf and Ng (2021) proposed two algorithm for estimating high dimensional linear time-varying parameter (TVP) model. In this paper, we propose a time-varying STARMA (tvSTARMA) modelling, based on the wavelet expansion of coefficients.

On the other hand, the estimation of spatial covariance structures is an important problem in many fields. In addition to stationarity, isotropicity is a common assumption, which means that the process is invariant under rotations around the origin. But in practice, the assumption of stationarity and isotropicity are often difficult to hold in real applications; see for instance Guttorp et al. (1994), Schmidt and O'Hagan (2003) and Finley (2011).

If a function f that maps the sampling locations at a geographic domain (G plane) into space representations of the deformation domain (D plane) is built, the spatial correlation can be considered isotropic in the D plane. The injectivity of transformation is one of the most important requirements to guarantee that two distinct locations in the G plane are not mapped into the same point in the D plane. A sufficient condition for f being injective is that the Jacobian determinant of f be non-zero.

To guarantee the injectivity of the mapping function, Choi and Lee (2000) suggested the box constraints for uniform cubic *B-spline* deformation coefficients. Musse et al. (2001) enforced Jacobian positivity by a novel constrained hierarchical parametric model. Chun and Fessler (2009) provided sufficient conditions that restrict *B-splines* based deformation coefficients to ensure that the Jacobian of such transformation is positive which extended the conditions of Kim (2004).

Although these methods ensure the positivity of Jacobian, deterministic models were used. Sampson and Guttorp (1992) introduced the topic with a stochastic model but no guarantee that the transformation based on thin plate splines is injective. Damian et al. (2001) suggested a solution to guarantee the injectivity of the transformation in stochastic model using a Bayesian approach.

1.1 Objective

In this thesis, we propose a time-varying space-time autoregressive and moving average modeling based on the multivariate locally stationary process, the time-varying parameters are expanded as a

linear combination of wavelet bases. And then, a method for nonstationary and isotropicity covariance function modeling, the deformation guarantees the injectivity of the transformation and is based on the monotonic function approach. Note that the wavelet expansion is also used for the deformation.

The organization of this thesis is as follows. Chapter 2 provides some basis backgrounds on spatio-temporal data, STARMA model, locally stationary processes, functional data analysis and wavelet analysis. In Chapter 3, the tvSTARMA model proposed with some simulations and an application to historical daily precipitation records are presented. The deformation proposed that guarantees the injectivity of transformation is in Chapter 4 and several simulations and an application to historical daily maximum temperature records are performed in the same Chapter. Conclusions and some further comments are presented in Chapter 5.

Chapter 2

Preliminaries

In this chapter, some basic concepts related to this thesis are presented. They are organized as follows: spatio-temporal process (2.1), STARMA models (2.2), locally stationary processes (2.3), functional data analysis (2.4) and wavelet analysis (2.5).

2.1 Spatio-Temporal Data

As the name suggests, spatio-temporal data are collected across space and time. An example would be that of the daily maximum temperature at 138 weather stations in the central USA recorded between the years 1990 and 1993 (inclusive), where the spatial property is the location where the object (maximum temperature) was collected, and the temporal property is the timestamp for which the spatial object is valid. Many fields including geoscience, meteorology, neuroscience, and climate science generate data that have both spatial and temporal components.

Let $\{Z(\mathbf{x},t) : \mathbf{x} \in \mathbb{R}^2,\ t \in \mathbb{R}\}$ a space-time random field. In most cases, the time index t is considered to be a subset of $\mathbb{Z}^+$, the set of positive integers and the spatial locations are typically either a subset of $\mathbb{Z}^2$, the two-dimensional integer lattice, or of $\mathbb{R}^2$. In general, spatio-temporal data is denoted by $Z_i(t) = Z(\mathbf{x}_i,t)$ for $i = 1,2,..,n,\ t = 1,...,T$, and usually, there is a large number of temporal observations, with relatively few spatial locations.

Stationary process do not change with time and space, while a nonstationary process is inconsistent with these components. Let $\mu(\mathbf{x},t)$ the mean of $Z(\mathbf{x},t)$ and the covariance between two space–time variables is

$$\text{Cov}[Z(\mathbf{x},t), Z(\mathbf{y},v)] = E\{[Z(\mathbf{x},t) - \mu(\mathbf{x},t)][Z(\mathbf{y},v) - \mu(\mathbf{y},v)]\}. \tag{2.1.1}$$

Process $Z(\mathbf{x},t)$ is called Second-order stationarity (SOS) if and only if for any locations $\mathbf{x}$ and times t

i. E[Z($\mathbf{x}$,t)] = μ, where μ is a constant;

ii. $\text{Cov}[\text{Z}(\mathbf{x}+\mathbf{h}, t+\tau), \text{Z}(\mathbf{x}, t)] = \text{Cov}[\text{Z}(\mathbf{h},\tau), \text{Z}(\mathbf{0}, 0)]$ can be expressed in terms of $\mathbf{h}$ and τ for all spatial shifts $\mathbf{h}$, and temporal shifts τ.

Sometimes, the assumption of isotropic on the spatial correlation is made out of convenience. This means that the correlation between any two observations depends merely on the distance between those locations while anisotropic is directional dependence.

2.1.1 Exploratory Analysis of Spatio-Temporal Data

Examining the empirical means and empirical covariances can be helpful to explore spatio-temporal data. The empirical spatial mean for location $\mathbf{x}_i$, $\hat{\mu}(\mathbf{x}_i)$, is calculated with the following formula:

$$\hat{\mu}(\mathbf{x}_i) = \frac{1}{T}\sum_{t=1}^{T} Z(\mathbf{x}_i, t),\ i = 1, ..., n. \tag{2.1.2}$$

The spatial mean of all locations is an n-dimensional vector, $\hat{\boldsymbol{\mu}}_{\mathbf{x}} = [\hat{\mu}(\mathbf{x}_1), ..., \hat{\mu}(\mathbf{x}_n)]'$. Similarly, the empirical temporal mean for time $t, \hat{\mu}(t)$, is given by

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n} Z(\mathbf{x}_i, t). \tag{2.1.3}$$

The empirical τ lag covariance between locations $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined by

$$\widehat{\text{Cov}}(\mathbf{x}_i, \mathbf{x}_j, \tau) = \text{E}[[Z(\mathbf{x}_i, t) - \hat{\mu}(\mathbf{x}_i)][Z(\mathbf{x}_j, t-\tau) - \hat{\mu}(\mathbf{x}_j)]], \tag{2.1.4}$$

where $\tau = 0, 1, ..., T-1$. Then, the $n \times n$ empirical τ lag covariance matrix is obtained, denoted by $\widehat{\mathbf{Cov}}(\tau)$, where the (i, j)th element is provided by (2.1.4). Alternatively, it can be calculated directly by

$$\widehat{\mathbf{Cov}}(\tau) = \text{E}[(\mathbf{Z}_t - \hat{\boldsymbol{\mu}}_{\mathbf{x}})(\mathbf{Z}_{t-\tau} - \hat{\boldsymbol{\mu}}_{\mathbf{x}})'], \tag{2.1.5}$$

where $\tau = 0, 1, ..., T-1$ and $\mathbf{Z}_t = [Z(x_1, t), Z(x_2, t), ..., Z(x_n, t)]'$.

According to Wikle et al. (2019), since locations in a two-dimensional space do not have any natural sort order, it can be difficult to obtain any intuitive information from these matrices. In some situations, splitting the domain into "strips" can be helpful, for example, world map can be split in some longitudinal strips and then calculate empirical covariance matrices for those strips.

Compared with the estimated spatial covariance above (2.1.4), covariograms and semivariograms are useful for measure the joint spatio-temporal dependence, then the covariability of the spatio-temporal data can be explored.

The empirical covariogram is a function of specific lags in time and in space, for time lag τ and spatial lag $\mathbf{h}$, it is given by

$$\hat{C}(\mathbf{h}, \tau) = \frac{1}{|N_{\mathbf{x}}(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in N_{\mathbf{x}}(\mathbf{h})}\sum_{(t_k,t_l)\in N_t(\tau)}[Z(\mathbf{x}_i, t_k) - \hat{\mu}(\mathbf{x}_i)][Z(\mathbf{x}_j, t_l) - \hat{\mu}(\mathbf{x}_j)], \tag{2.1.6}$$

where

$$N_{\mathbf{x}}(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j) : |\mathbf{x}_i, \mathbf{x}_j| = \mathbf{h},\ i, j = 1, ..., n\}$$

and

$$N_t(\tau) = \{(t_k, t_l) : t_k - t_l = \pm\tau,\ k, l = 1, ..., T\}.$$

Under isotropy, the spatial lag as a function of distance is usually considered, $h = ||\mathbf{h}||$, where $||\cdot||$ is the Euclidean norm.

And the semivariogram is defined as

$$\hat{\gamma}(\mathbf{x}_i, \mathbf{x}_k, t_j, t_l) = \frac{1}{2}\text{Var}[Z(\mathbf{x}_i, t_j) - Z(\mathbf{x}_k, t_l)]. \tag{2.1.7}$$

Note that the empirical covariogram of $Z(\mathbf{x}_i, t_k)$ and $Z(\mathbf{x}_j, t_l)$ can be expressed by $2\hat{\gamma}(\mathbf{x}_i, \mathbf{x}_k, t_j, t_l)$.

2.1.2 Spatio-Temporal Covariance Models

The spatio-temporal covariance function plays an important role in analyzing data, it describes the second-order dependence of random processes. Many studies have been developed for spatio–temporal covariance functions as models for dependence, such as Kyriakidis and Journel (1999), Jun and Stein (2008), Reich et al. (2011), Gneiting and Schlather (2013), Ma (2003), Guinness and Fuentes (2016) and De Iaco et al. (2019).

Due to the large size of spatio-temporal data, it is usually computationally expensive and sometimes infeasible to implement traditional techniques. Simplifying structures are often used, such as stationarity, spatial isotropy, full symmetry and separability. Stationarity and spatial isotropy have been defined in the previous section, denoting covariance function (2.1.1) in $C(\mathbf{h}, \tau)$, we can say that for all $\mathbf{h} \in \mathbb{R}^d$ and $\tau \in \mathbb{R}$, a covariance function C is called fully symmetric if

$$C(\mathbf{h}, \tau) = C(\mathbf{h}, -\tau) \tag{2.1.8}$$

and is called separable if

$$C(\mathbf{h}, \tau) = C_s(\mathbf{h})C_t(\tau), \tag{2.1.9}$$

where C_s and C_t are purely spatial and purely temporal covariance functions, respectively. The relationships among several spatio-temporal covariance structures can be found in Gneiting et al. (2006).

The model proposed by Gneiting (2002) is one example of stationary and fully symmetric covariance models and it is given by

$$C(\mathbf{h}, \tau) = \frac{\sigma^2}{\beta(|\tau|^2)^{d/2}}\omega\left\{\frac{||\mathbf{h}||^2}{\beta(|\tau|^2)}\right\},\ (\mathbf{h}; \tau) \in \mathbb{R}^d \times \mathbb{R}, \tag{2.1.10}$$

where $\sigma^2 = C(\mathbf{0}, 0)$, $\omega(\cdot) \geq 0$ is a completely monotone function and $\beta(\cdot)$ is a Bernstein function, that is, it has a completely monotone derivative.

Under separability, an easy way to build a spatio–temporal covariance function is through the product of a spatial and a temporal covariance function. Separable spatio-temporal covariance function with the well known exponential model is given by

$$C(\mathbf{h}, \tau) = \sigma^2 \exp(-\nu_s ||\mathbf{h}||) \exp(-\nu_t |\tau|), \tag{2.1.11}$$

where $\sigma^2 = C(\mathbf{0}, 0)$ is scale parameter, $\nu_s > 0$ is decay parameter and the temporal correlation is from an autoregressive process of order 1 with parameter $\rho = \exp(-\nu_t)$.

2.2 STARMA Models

A time series is a sequence of data points generated in successive order over some period of time. To describe and forecast several time series simultaneously, vector autoregressive and moving average (VARMA) models are often used.

In this section, we briefly review the models of ARIMA class and introduce STARMA models, which are special cases of the VARMA models, with their procedure of identification, estimation, diagnostic and forecast.

2.2.1 ARIMA Models

In time series analysis, stationarity is a common assumption. Let $\{Z_t : t \in \mathscr{T}\}$ a stochastic process, such that $\mathscr{T} = [0, \infty)$ and $\{t_1, t_2, ..., t_n\}$ is a collection of values of $\mathscr{T}$. Z_t is called strictly stationary if

$$P(Z_{t_1} \leq z_1, ...Z_{t_n} \leq z_n) = P(Z_{t_1+\tau} \leq z_1, ...Z_{t_n+\tau} \leq z_n) \tag{2.2.1}$$

for all $n = 1, 2, ...$, time points $t_1, ..., t_n$, numbers $z_1, ..., z_n$ and $\tau \in \mathscr{T}$. That is, all distributions are invariant under time translations.

However, it is very difficult to use the strict stationary definition, as we do not know these distribution functions in practice, thus, a milder version that imposes conditions only on the first two moments of the series is considered. Z_t is called weakly stationary or SOS if and only if

- $E[Z_t]$ is constant for all $t \in \mathscr{T}$;
- $E[Z_t^2] < \infty$ for all $t \in \mathscr{T}$;
- $\text{Cov}[Z_{t_1}, Z_{t_2}]$ depends on $|t_1 - t_2|$ only.

2.2.1.1 The Models

The Autoregressive and moving average (ARMA) models are characterized by two polynomials, one for the autoregressive (AR) and the second for the moving average (MA) that depend on their own past values and past error terms, respectively. An ARMA process with order p in the AR part and order q in the MA part, denoted by ARMA(p, q), is given by

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{2.2.2}$$

where $\phi = (\phi_1, ...\phi_p)$ are parameters of autoregressive term, $\theta = (\theta_1, ..., \theta_q)$ are parameters of moving average term and a_t is white noise. Using the backward shift operator B which defined as $BZ_t = Z_{t-1}$, $B^k Z_t = Z_{t-k}$, (2.2.2) can be written as

$$\begin{aligned} (1 - \phi_1 B - \cdots - \phi_p B^p) Z_t &= (1 - \theta_1 B - \cdots - \theta_q B^q) a_t & (2.2.3) \\ \phi(B) Z_t &= \theta(B) a_t, & (2.2.4) \end{aligned}$$

where $\phi(B)$ and $\theta(B)$ are the autoregressive operator and the moving average operator of orders p and q, respectively.

Within the ARMA (p, q) model, there are two particular cases, when $p = 0$ or $q = 0$. When $q = 0$, it has the AR(p) model with the following form:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t. \tag{2.2.5}$$

And when $p = 0$, the MA(q) model is defined by

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \tag{2.2.6}$$

Generalization of the ARMA models, the ARIMA models are applied in some cases where the data show evidence of non stationarity, where an initial differencing step can be applied one or more times to eliminate the non stationarity. Thus, the ARIMA(p, d, q) model can be expressed as

$$(1 - \phi_1 B - \cdots - \phi_p B^p) \Delta^d Z_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t, \tag{2.2.7}$$

where $\Delta = 1 - B$ is the differencing operator and d is a non-negative integer.

2.2.1.2 SARIMA Models

In practice, many time series exhibit cyclical and periodic behavior, the seasonal autoregressive integrated moving average (SARIMA) model can be used to capture the seasonality of a series. Incorporating seasonal AR and/or MA polynomials into an ARIMA model multiplicatively, a SARIMA$(p, d, q) \times (P, D, Q)_k$ model is obtained and written as

$$\phi(B)\Phi(B^k)\Delta^d \Delta_k^D Z_t = \theta(B)\Theta(B^k) a_t, \tag{2.2.8}$$

where

$$\begin{aligned}\phi(B) &= 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p,\\ \Phi(B^k) &= 1-\Phi_1 B^k-\Phi_2 B^{2k}-\cdots-\Phi_P B^{Pk},\\ \theta(B) &= 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q,\\ \Theta(B^k) &= 1-\Theta_1 B^k-\Theta_2 B^{2k}-\cdots-\Theta_Q B^{Qk},\end{aligned}$$

$\Phi(B^k)$ and $\Theta(B^k)$ are the k period seasonal autoregressive operator and the seasonal moving average operator of orders P and Q, respectively, $\Delta_k^D = (1-B^k)^D$ is seasonal differencing operator with seasonal period k and non-negative integer D.

2.2.2 Spatial Weight Matrix

In the construction of the STARMA models, the definition of the spatial lag operator is necessary. Let $Z_i(t)$ be a random variable at location i and time t and $L^{(l)}$, the spatial lag operator with spatial order l is defined as

$$L^{(l)}Z_i(t) = \begin{cases} Z_i(t), & \text{if } l=0,\\ \sum_{j=1}^n w_{ij}^{(l)} Z_j(t), & \text{if } l>0,\end{cases} \tag{2.2.9}$$

where $w_{ij}^{(l)}$ are weights of order l with

$$\sum_{j=1}^{n} \omega_{ij}^{(l)} = 1. \tag{2.2.10}$$

Each element of the matrix reflects the spatial relationship between two regions, i and j and $w_{ij}^{(l)} = 0$ when $i = j$, that is, the matrix has zeros on its main diagonal and the other elements will consist of positive numbers. Let $\mathbf{Z}(t) = [Z_1(t), ..., Z_n(t)]'$ a vector of n observations,

$$L^{(l)}\mathbf{Z}(t) = \begin{cases} W^{(0)}\mathbf{Z}(t) = \mathbf{I}_n\mathbf{Z}(t), & \text{if } l=0,\\ W^{(l)}\mathbf{Z}(t), & \text{if } l>0,\end{cases} \tag{2.2.11}$$

where $W^{(l)}$ is an $n \times n$ square matrix consist of weights $w_{ij}^{(l)}$ and the sum of items in each row is 1.

According to Pfeifer and Deutsch (1980a), the weights must depict a hierarchical ordering of spatial neighbors, that is, first-order neighbors are locations closest to the location of interest; second-order neighbors are neighbors that are further away than first-order neighbors and closer than third-order neighbors and so on. Figure 2.1 shows the first four spatial order neighbors of a location X in bidimensional grid.

There are many way to define the weights, according to a priori information, the specification of $w_{ij}^{(l)}$ is chosen by the researcher. A simple way proposed by Rao and Antunes (2004) can be written as

$$w_{ij}^{(l)} = \begin{cases} \frac{1}{n_i^{(l)}}, & \text{if } i \text{ and } j \text{ are neighbors of order } l,\\ 0, & \text{otherwise,}\end{cases} \tag{2.2.12}$$

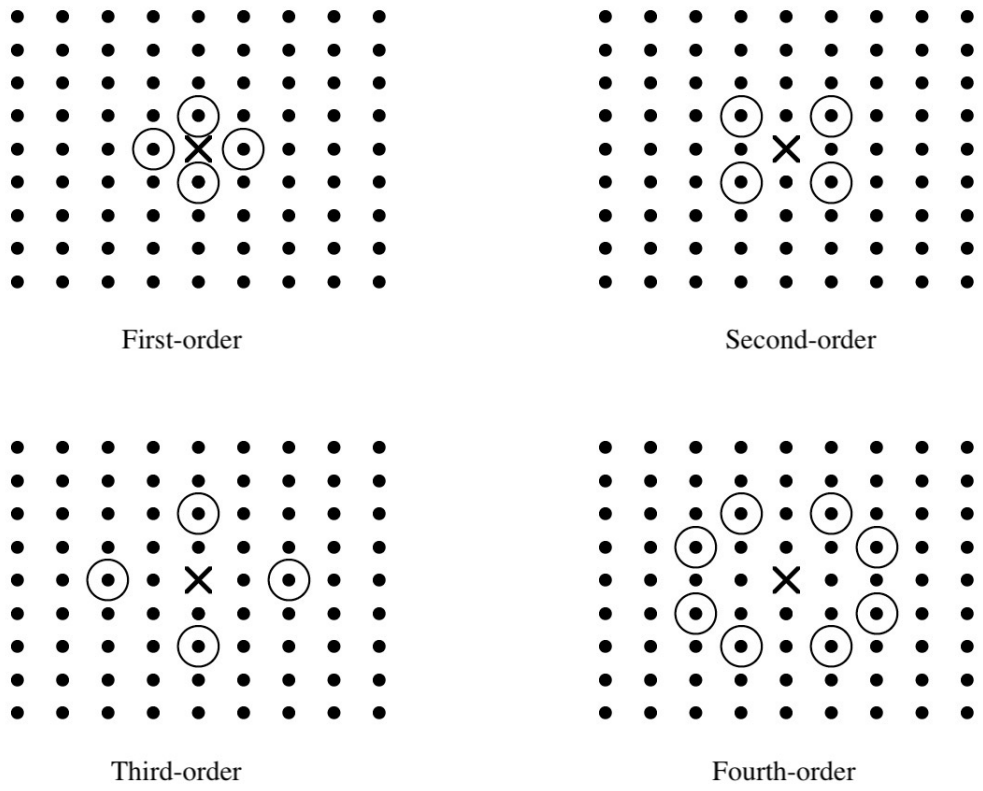


Figure 2.1: Spatial order in bidimensional grid.

where $n_i^{(l)}$ is the number of neighbors of order l of location i. Another alternative to calculate the matrix proposed by Rao and Antunes (2004) is to use the inverse of the distance between the locations, as the spatial effect decreases with distance, only one spatial weight matrix is used, different from the weights calculated by the method presented above (2.2.12), where each spatial order is represented by a matrix. The distance inverse weights are defined as

$$w_{ij} = \begin{cases} \frac{d_{ij}^{-\alpha}}{\sum_{k \neq i} d_{ik}^{-\alpha}}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{2.2.13}$$

where $\alpha > 0$ is the changing rate of weights over distance d_{ij} and d_{ij} is the orthodromic distance (great-circle or spherical distance) between two locals $\mathbf{x}_i$ and $\mathbf{x}_j$. It is the shortest distance between two points on the surface of a sphere. Let $\mathbf{x}_i = (\xi_i, \kappa_i)$ and $\mathbf{x}_j = (\xi_j, \kappa_j)$ be the geographical locations, where ξ_η and $\kappa_\eta, \eta = i, j$, are the latitude and longitude, respectively. The distance d_{ij} is given by

$$d_{ij} = R\sigma, \tag{2.2.14}$$

where R is radius of the Earth and

$$\sigma = \arccos[\sin \xi_i \sin \xi_j + \cos \xi_i \cos \xi_j \cos(|\kappa_i - \kappa_j|)]. \tag{2.2.15}$$

Zhou and Lin (2008) proposed a matrix in which the weights are negative exponential distance

functions and are given by

$$w_{ij} = \begin{cases} \frac{\exp\{-\alpha d_{ij}\}}{\sum_{k\neq i}\exp\{-\alpha d_{ik}\}}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{2.2.16}$$

where d_{ij} can be euclidean distance or orthodromic distance and α is coefficient of friction, generally, $\alpha = 1$ or 2.

2.2.3 STARMA Model

The STARMA($p_{\lambda_1,...,\lambda_p}, q_{m_1,...,m_q}$) model can be defined as

$$Z_i(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}L^{(l)}Z_i(t-k) - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}L^{(l)}\epsilon_i(t-k) + \epsilon_i(t),\ i = 1,...,n,\ t = 1,...,T, \tag{2.2.17}$$

where p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the kth AR term and m_k is the spatial order of the kth MA term. ϕ_{kl} and θ_{kl} are parameters at time lag k and space lag l for the AR and MA parameters, respectively. $\epsilon_i(t)$ is random error follows a normal distribution with mean zero and covariance σ^2.

Rewriting (2.2.17) in matrix form, we have

$$\mathbf{Z}(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}\mathbf{Z}(t-k) - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\boldsymbol{\epsilon}(t-k) + \boldsymbol{\epsilon}(t), \tag{2.2.18}$$

where $\boldsymbol{\epsilon}(t)$ are random errors follow a multivariate normal distribution with $E[\boldsymbol{\epsilon}(t)] = \mathbf{0}$ and

$$E[\epsilon(t)\epsilon(t+s)'] = \begin{cases} \mathbf{I}_n\sigma^2, & \text{if } s = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.19}$$

There are two particular cases of STARMA model, when $p = 0$ or $q = 0$. When $q = 0$, only autoregressive terms remain, there is a space-time autoregressive (STAR) model. The STAR($p_{\lambda_1,...,\lambda_p}$) model is given by

$$\mathbf{Z}(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}\mathbf{Z}(t-k) + \boldsymbol{\epsilon}(t) \tag{2.2.20}$$

and when $p = 0$, it called space-time moving average (STMA) model, the STMA($q_{m_1,...,m_q}$) is written as

$$\mathbf{Z}(t) = \boldsymbol{\epsilon}(t) - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\boldsymbol{\epsilon}(t-k). \tag{2.2.21}$$

The condition to ensure stationarity of STARMA model is that every root (x_u) that solve

$$\det\left[x_u^p\mathbf{I} - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}x_u^{p-k}\right] = 0 \tag{2.2.22}$$

must lie inside the unit circle, equivalently, $|x_u| < 1$, this requirement is to determine a region of possible values of ϕ_{kl} that will result in a stationary process.

The invertibility condition is independent of the sationarity condition, it is applied in MA terms, then the STARMA process is called invertible if all possible roots of

$$\det\left[x_u^q\mathbf{I}-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}x_u^{q-k}\right]=0, \tag{2.2.23}$$

lie inside the unit circle ($|x_u|<1$). Note that pure STAR models are always invertible (since they contain no MA terms) and pure STMA models are always stationary (since they contain no AR terms).

Similar to the models of ARIMA class, when the stationarity condition is not met due to having a trend, it is enough to take the necessary order of differencing in the series and later incorporate it to the STARMA model. Thus, the space-time autoregressive integrated moving average (STARIMA) model is obtained as

$$\boldsymbol{\Delta}^d\mathbf{Z}(t)=\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}\boldsymbol{\Delta}^d\mathbf{Z}(t-k)-\sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\boldsymbol{\epsilon}(t-k)+\boldsymbol{\epsilon}(t), \tag{2.2.24}$$

where $\boldsymbol{\Delta}^d$ are differencing operators and defined by $\Delta^d Z(t)=(1-B)^d Z(t)$ with d non-negative. Analogous to STARIMA model, there is also the seasonal STARMA model that incorporates the seasonal component, see Biz (2014) for details.

The methodology introduced by Box and Jenkins (1970) is widely used in the analysis of parametric models. The method is an iterative three-stage modeling approach. According to Morettin and Toloi (2018), the process of model construction is based on an interactive cycle, consists of a sequence of procedures:

- Identification: select a class of models that is considered for analysis and determine one or more candidate models through the analysis of the autorelation and partial autocorrelation functions;
- Estimation: estimate the parameters of the identified models;
- Verification or diagnosis: through residual analysis and forecasting to check if the fitted model is suitable.

If the model is not suitable, the cycle is repeated from the identification phase. Basic reference books are Box et al. (2015), Brockwell and Davis (2016), Shumway and Stoffer (2017) and Morettin and Toloi (2018).

We mention below a brief summary of construction of STARMA models.

2.2.3.1 Identification

Different from ARIMA models, the STARMA models have a space-time autocovariance function, the covariance between points lagged both in space and time by combining the covariance between

all possible pairs of locations. The space-time autocovariance between lth and kth neighbors at time lag s is defined by

$$\gamma_{lk}(s) = E\left\{\sum_{i=1}^{n} \frac{L^{(l)}Z_i(t)L^{(k)}Z_i(t+s)}{n}\right\}. \tag{2.2.25}$$

Rewriting (2.2.25) in matrix form, we have

$$\begin{aligned}\gamma_{lk}(s) &= E\left\{\frac{[W^{(l)}\mathbf{Z}(t)]'[W^{(k)}\mathbf{Z}(t+s)]}{n}\right\} \\ &= \mathrm{tr}\left\{\frac{W^{(k)'}W^{(l)}\Gamma(s)}{n}\right\},\end{aligned} \tag{2.2.26}$$

where $\mathrm{tr}[\mathbf{A}]$ is the trace of $\mathbf{A}$ and $\Gamma(s) = E[\mathbf{Z}(t)\mathbf{Z}(t+s)']$. The estimator of $\Gamma(s)$ is

$$\hat{\Gamma}(s) = \sum_{t=1}^{T-s} \frac{\mathbf{z}(t)\mathbf{z}(t+s)'}{T-s}, \tag{2.2.27}$$

where $\mathbf{z}$ are the observations of $\mathbf{Z}$, replacing (2.2.26) by (2.2.27), the estimator of $\gamma_{lk}(s)$,

$$\hat{\gamma}_{lk}(s) = \frac{1}{n}\mathrm{tr}\left\{W^{(k)'}W^{(l)}\sum_{t=1}^{T-s}\frac{\mathbf{z}(t)\mathbf{z}(t+s)'}{T-s}\right\}, \tag{2.2.28}$$

is obtained.

In particular, when $l = k = 0$,

$$\gamma_{00}(s) = \frac{1}{n}\mathrm{tr}[\Gamma(s)] \tag{2.2.29}$$

is the average of the τ lag autocovariance for all n locations. And when $l = 0,\ k = 1$,

$$\gamma_{01}(s) = \frac{1}{n}\mathrm{tr}[W^{(1)}\Gamma(s)] \tag{2.2.30}$$

presents the average over all locations of the s lag covariance between each location and their spatial weight matrix of order 1.

Note that a important property of the space-time autocovariance is

$$\gamma_{lk}(s) = \gamma_{lk}(-s), \tag{2.2.31}$$

which can be proved using $\mathrm{tr}[\mathrm{AB}] = \mathrm{tr}[\mathrm{BA}]$ and $\Gamma(s) = \Gamma(-s)'$.

According to Martin and Oeppen (1975), the definition of space-time autocorrelation is different from the autocorrelation of univariate models, due to the fact that space-time autocorrelation has several combinations that can be used. In that case, an autocorrelation function that has constant variance for all spatial lags is a good choice to be defined. Then, a space-time autocorrelation function (STACF) between lth and kth order neighbors in time lag s is given by

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{\sqrt{\gamma_{ll}(0)\gamma_{kk}(0)}}. \tag{2.2.32}$$

And an estimator of this correlation is obtained using (2.2.28).

Another important function is the space-time partial autocorrelation function (STPACF) which finds the correlation between two values of the process, eliminating the intermediate values. Martin and Oeppen (1975) proposed an extension of the Yule-walker equations of univariate models to calculate STPACF. Multiplying both sides of the STAR $(k_{0,...,\lambda})$ by $[W^{(h)}\mathbf{Z}(t-s)]'$,

$$\mathbf{Z}(t-s)'W^{(h)'}\mathbf{Z}(t) = \sum_{j=1}^{k}\sum_{l=0}^{\lambda}\phi_{jl}\mathbf{Z}(t-s)'W^{(h)'}W^{(l)}\mathbf{Z}(t-j) + \mathbf{Z}(t-s)'W^{(h)'}\boldsymbol{\epsilon}(t). \tag{2.2.33}$$

Calculating the expected value and dividing both sides by N, since $E[\mathbf{Z}(t-s)'\boldsymbol{\epsilon}(t)] = 0$, for $s > 0$,

$$\gamma_{h0}(s) = \sum_{j=1}^{k}\sum_{l=0}^{\lambda}\phi_{jl}\gamma_{hl}(s-j). \tag{2.2.34}$$

Hence, the system of equations can be written as the Yule-Walker equations shown in Figure 2.2 for $s = 1, ..., k$ and $h = 0, 1, ..., \lambda$. The last coefficient, ϕ_{kl}, obtained by solving the system of equations as $l = 1, ..., k$ for $k = 1, .., \lambda$ is called STPACF of spatial order λ, note that the value of λ is determined by researcher.

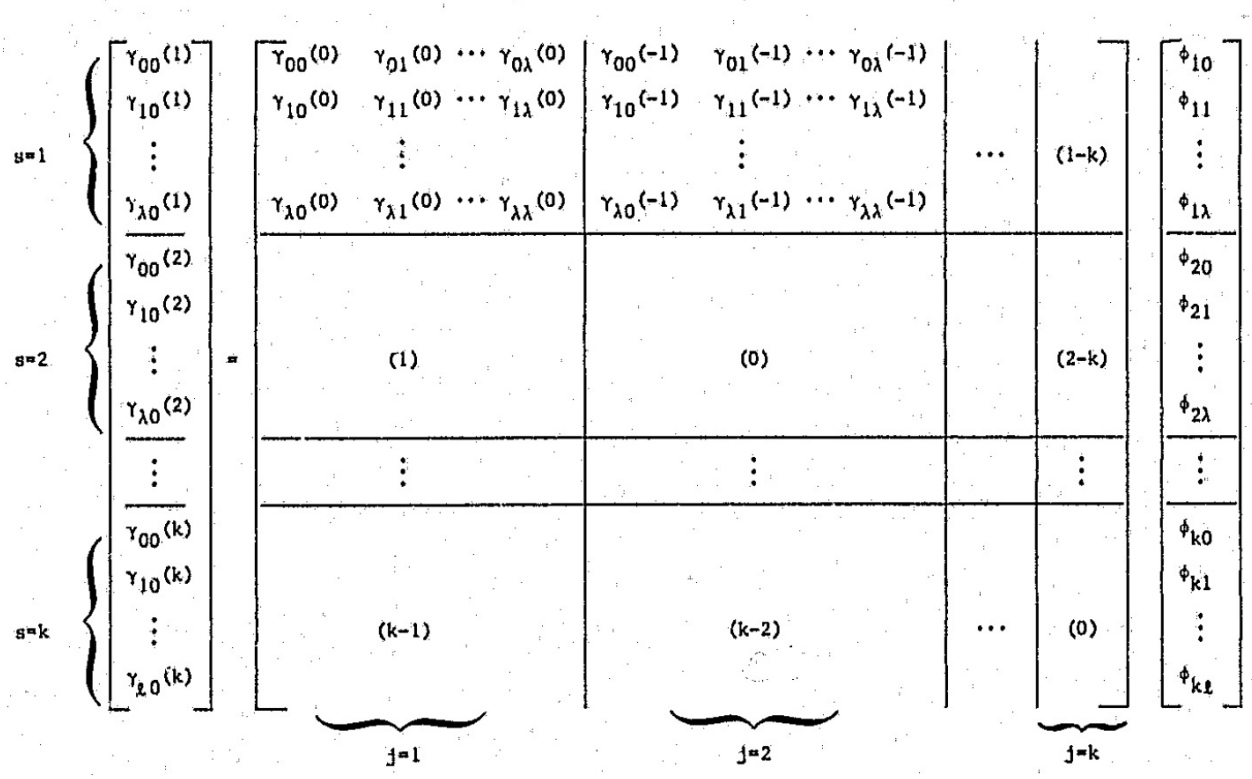


Figure 2.2: System of Yule-Walker, Pfeifer and Deutsch (1980a).

Since the $\gamma_{lk}(s)$ can be estimated by solving Yule-Walker equations using a recursive procedure for $l = 0, 1, ..., \lambda$, for each $k = 1, 2, ...,$ STACF and STFACP are estimated. Analogous to univariate time series models, STACF and STFACP plots of series can be use to determine the order of AR and/or MA terms of STARMA models. Table 2.1 shows a general behavior of the STACF and STPACF for STAR, STMA and STARMA models. Note that the functions decay with both time and space.

Table 2.1: General behavior of the STACF and STPACF.

Modelo	STFAC	STFACP
STAR($p_{\lambda_1,...,\lambda_p}$)	gradual decreasing	$\phi_{kl} = 0$, for $k > p$ and $l > \lambda_p$
STMA($q_{m_1,...,m_q}$)	$\rho_{l0}(s) = 0$, for $s > q$ and $l > m_q$	gradual decreasing
STARMA($p_{\lambda_1,...,\lambda_p}, q_{m_1,...,m_q}$)	gradual decreasing	gradual decreasing

2.2.3.2 Estimation

When the errors $\boldsymbol{\epsilon}(t)$ are white noise, that is, follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_{nT}\sigma^2$, the maximum likelihood estimation can be used to estimate the parameters $\boldsymbol{\Phi} = (\phi_{10}, \phi_{11}, ..., \phi_{1\lambda_1}, ..., \phi_{p0}, ..., \phi_{p\lambda_p})$, $\boldsymbol{\Theta} = (\theta_{10}, ..., \theta_{1m_1}, ...\theta_{q0}, \theta_{qm_q})$ and σ^2. Then, the conditional joint probability density function is written as

$$\begin{aligned} f(\boldsymbol{\epsilon}|\boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma^2) &= (2\pi)^{-\frac{NT}{2}} |\mathbf{I}_{nT}\sigma^2|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}\boldsymbol{\epsilon}'\mathbf{I}\boldsymbol{\epsilon}\} \\ &= (2\pi)^{-\frac{nT}{2}} (\sigma^2)^{-\frac{nT}{2}} \exp\{-\frac{S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta})}{2\sigma^2}\}, \end{aligned}$$

where $S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta}) = \boldsymbol{\epsilon}'\mathbf{I}\boldsymbol{\epsilon} = \sum_{i=1}^{n}\sum_{t=1}^{T}\epsilon_i^2(t)$. Note that the errors are unknown and one way to obtain them is to calculate by a recursive method using the STARMA model from the observations $\mathbf{z}(t)$, that is,

$$\boldsymbol{\epsilon}(t) = \mathbf{z}(t) - \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}\mathbf{z}(t-k) - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}\boldsymbol{\epsilon}(t-k) \tag{2.2.35}$$

with the initial values of $\mathbf{z}(t)$ and $\boldsymbol{\epsilon}(t)$ equal to zero for $t = 1, ..., T$.

The conditional likelihood function is given by

$$L(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma^2|\mathbf{z}) = (2\pi)^{-\frac{nT}{2}} (\sigma^2)^{-\frac{nT}{2}} \exp\{-\frac{S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta})}{2\sigma^2}\}, \tag{2.2.36}$$

where $S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta}) = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$ and $\hat{\boldsymbol{\epsilon}}$ are obtained by equation (2.2.35). Then, the conditional maximum likelihood estimators of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma^2)$ are $(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Theta}})$ that minimize $S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta})$ and $\hat{\sigma}^2 = \frac{S_*(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Theta}})}{nT}$.

The Kalman filter is a good alternative to estimate the parameters of STARMA model. Consider the system

$$\mathbf{x}_{t+1} = F_t\mathbf{x}_t + \Gamma_t\mathbf{w}_{t+1}, \tag{2.2.37}$$

$$\mathbf{z}_t = M_t\mathbf{x}_t + \mathbf{v}_t, \tag{2.2.38}$$

where (2.2.37) is the state equation and (2.2.38) is the observation equation of the system. Here, $\mathbf{x}_t$ is the state variable at time t, $\mathbf{z}_t$ is the observation at t, F_t, Γ_t and M_t are the state transition matrix, control input matrix and observation matrix, respectively. The vectors $\mathbf{w}_t$ and $\mathbf{v}_t$ have mean zero and covariance matrices Q_t and R_t, respectively, that fulfill

$$\text{Cov}(\mathbf{w}_s, \mathbf{w}_t) = 0, \ \text{Cov}(\mathbf{v}_s, \mathbf{v}_t) = 0, \ \text{for } s \neq t,$$

$$\text{Cov}(\mathbf{w}_s, \mathbf{v}_t) = 0.$$

The algorithm provides recursive estimation formulas to estimate the state of a process, in the sense of minimizing the squared error. The filter can be written in the form

$$\hat{\mathbf{x}}_{t+1} = F_t\hat{\mathbf{x}}_t + K_{t+1}(\mathbf{z}_{t+1} - M_{t+1}F_t\hat{\mathbf{x}}_t), \tag{2.2.39}$$

$$P_{t+1} = (I - K_{t+1}M_{t+1})(F_tP_tF_t' + \Gamma_tQ_{t+1}\Gamma_t'), \tag{2.2.40}$$

$$K_{t+1} = P_{t+1}M_{t+1}'R_{t+1}^{-1}. \tag{2.2.41}$$

For the STARMA model, (2.2.18) can be rewritten in linear form:

$$\begin{aligned}\mathbf{z}_t &= \sum_{l=0}^{\lambda_1}\phi_{1l}W^{(l)}\mathbf{z}_{t-1} + ... + \sum_{l=0}^{\lambda_p}\phi_{pl}W^{(l)}\mathbf{z}_{t-p} - \sum_{l=0}^{m_1}\theta_{1l}W^{(l)}\boldsymbol{\epsilon}_{t-1} - ... - \sum_{l=0}^{m_q}\theta_{ql}W^{(l)}\boldsymbol{\epsilon}_{t-q} + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\Phi}_1\mathbf{z}_{t-1} + ... + \boldsymbol{\Phi}_p\mathbf{z}_{t-p} + \boldsymbol{\Theta}_1\boldsymbol{\epsilon}_{t-1} + ... + \boldsymbol{\Theta}_q\boldsymbol{\epsilon}_{t-q} + \boldsymbol{\epsilon}_t \\ &= \mathbf{x}_t\mathbf{b} + \boldsymbol{\epsilon}_t,\end{aligned}$$

where

$$\mathbf{b}_{n^2(p+q)\times 1} = \begin{bmatrix}\text{vec } \boldsymbol{\Phi}_1 \\ \vdots \\ \text{vec } \boldsymbol{\Phi}_p \\ \text{vec } \boldsymbol{\Theta}_1 \\ \vdots \\ \text{vec } \boldsymbol{\Theta}_q\end{bmatrix},$$

$$\mathbf{x}_{t\ n\times n^2(p+q)} = [\mathbf{y}_t\ \mathbf{e}_t],$$

$$\mathbf{y}_{t\ n\times n^2p} = [\text{diag}_n \otimes \mathbf{z}_{t-1}'\ \ \ldots\ \ \text{diag}_n \otimes \mathbf{z}_{t-p}'],$$

$$\mathbf{e}_{t\ n\times n^2q} = [\text{diag}_n \otimes \boldsymbol{\epsilon}_{t-1}'\ \ \ldots\ \ \text{diag}_n \otimes \boldsymbol{\epsilon}_{t-q}'].$$

According to Cipra and Motykova (1987), the state space representation (2.2.37) and (2.2.38) for the estimation of the parameters $\mathbf{b}$ can be written as

$$\mathbf{b}_{t+1} = \mathbf{b}_t, \tag{2.2.42}$$

$$\mathbf{z}_t = \mathbf{x}_t\mathbf{b}_t + \boldsymbol{\epsilon}_t, \tag{2.2.43}$$

where $F_t = \mathbf{I}_{n^2(p+q)}$, $\Gamma_t = \mathbf{0}_{n^2(p+q)\times n^2(p+q)}$, $M_t = \mathbf{x}_t$, $\mathbf{v}_t = \boldsymbol{\epsilon}_t$. And the recursive equations have the following form:

$$\hat{\mathbf{b}}_{t+1} = \hat{\mathbf{b}}_t + \mathbf{P}_{t+1}\mathbf{z}_{t+1}'\hat{\boldsymbol{\Sigma}}_t^{-1}(\mathbf{z}_{t+1} - \mathbf{z}_{t+1}\hat{\mathbf{b}}_t), \tag{2.2.44}$$

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \mathbf{P}_t\mathbf{z}_{t+1}'(\mathbf{z}_{t+1}\mathbf{P}_t\mathbf{z}_{t+1}' + \hat{\boldsymbol{\Sigma}}_t)^{-1}\mathbf{z}_{t+1}'\mathbf{P}_t, \tag{2.2.45}$$

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \frac{1}{t+1-n^2(p+q)}\left\{[t-n^2(p+q)]\hat{\boldsymbol{\Sigma}}_t + (\mathbf{z}_{t+1} - \mathbf{z}_{t+1}\hat{\mathbf{b}}_{t+1})(\mathbf{z}_{t+1} - \mathbf{z}_{t+1}\hat{\mathbf{b}}_{t+1})'\right\} \tag{2.2.46}$$

$$\hat{\boldsymbol{\epsilon}}_{t+1} = \mathbf{z}_{t+1} - \mathbf{z}_{t+1}\hat{\mathbf{b}}_{t+1}. \tag{2.2.47}$$

With the linear form, the confidence intervals can be obtained from the result of linear regression theory. Another alternative is calculate approximate confidence regions, using the approximate likelihood function of $S_*(\boldsymbol{\Phi}, \boldsymbol{\Theta})$, see Pfeifer and Deutsch (1980a) for details.

2.2.3.3 Diagnostic

In the diagnostic checking, the object is to verify if the selected models are adequate. For this, a residual analysis will be carried out and observe if the residuals are in accordance with the assumptions of the model. In the case of the STARMA model, the residuals of the fitted model should be white noise, that is, they are multivariate normal distribution with zero mean, covariance matrix σ^2 and all autocovariances at non-zero lags equal to 0.

For the residual analysis, the usual way is to calculate the STACF and STPACF of the residuals and construct a confidence interval using theoretically variance

$$\mathrm{Var}(\hat{\rho}_{l0}(s)) \approx \frac{1}{n(T-s)}, \tag{2.2.48}$$

where $\hat{\rho}_{l0}(s)$ is sample STACF of the residuals of the fitted model. If the STACF and the STPACF are inside the confidence interval, the residuals are approximately white noise, otherwise, the dependent term is identified and the model is updated.

2.2.3.4 Forecasting

For STARMA models, the forecasting is performed in a similar way as SARIMA models. According to Biz (2014), the predictions from time t and horizon h for STARMA$(p_{\lambda_1,\ldots,\lambda_p}, q_{m_1,\ldots,m_q})$ are given by

$$\hat{\mathbf{Z}}_t(h) = E\left\{\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\hat{\phi}_{kl}W^{(l)}\mathbf{Z}(t+h-k) - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\hat{\theta}_{kl}W^{(l)}\boldsymbol{\epsilon}(t+h-k) + \boldsymbol{\epsilon}(t+h)\right\} \tag{2.2.49}$$

whereas

$$\begin{aligned}
E[\mathbf{Z}(t+k)] &= \hat{\mathbf{Z}}_t(k),\ k>0,\\
E[\mathbf{Z}(t+k)] &= \mathbf{Z}(t+k),\ k\leq 0,\\
E[\boldsymbol{\epsilon}(t+k)] &= \mathbf{0},\ k>0,\\
E[\boldsymbol{\epsilon}(t+k)] &= \boldsymbol{\epsilon}(t+k),\ k\leq 0.
\end{aligned}$$

2.3 Locally Stationary Processes

As mentioned in Section 1, stationarity is basis assumption for time series analysis, however, many phenomena in the practical applications show a non stationary behavior. The disadvantage of

non-stationary data is that for most of the time series models, the model assumptions are violated when non-stationary data are used. This leads to the estimators no longer having the nice properties such as asymptotic normality and sometimes even consistency and hence poor forecasts. One way to proceed is to assume that the processes involved are locally stationary processes.

The idea of having locally approximately a stationary process was considered by Priestlety (1965), who generalized the usual definition of spectra for stationary processes with the concept of evolutionary spectra. According to Priestley (1965), the process X_t with $\text{var}(X_t) < \infty$ for each t has a time varying spectral representation:

$$X_t = \int_{-\pi}^{\pi} \exp\{i\pi t\} A_t(\lambda) d\xi(\lambda), \ t \in \mathbb{Z}, \tag{2.3.1}$$

where $A_t(\lambda) = \int_{-\infty}^{\infty} \exp\{it\theta\} dH_\lambda(\theta)$ with $|H_\lambda(\theta)|$, which has an absolute maximum at $\theta = 0$ is a time varying transfer function and $\xi(\lambda)$ an orthogonal increment process with $E[d\xi(\lambda)]^2 = d\mu(\lambda)$, the measure $d\mu(\lambda)$ is analogous to the spectrum in the case of stationary processes.

There is a lack of asymptotic considerations in the approach introduced by Priestley (1965). Many estimators and asymptotic results were developed by Dahlhaus and his coworkers, thus, the locally stationary processes used in this thesis is in the sense of the Dahlhaus's theory of locally stationary processes which is based on the infill asymptotic approach.

2.3.1 The General Definition

Let

$$X_t + \alpha_t X_{t-1} = \sigma_t \epsilon_t \tag{2.3.2}$$

a time varying AR(1) process where ϵ_t is an independent and identically distributed standard normal. Infill asymptotics is applied by rescaling the parameter curves α_t and σ_t to the unit interval. That is, them are replaced by $\alpha(\frac{t}{T})$ and $\sigma(\frac{t}{T})$ with curves $\alpha(\cdot) : [0,1] \to (-1,1)$ and $\sigma(\cdot) : [0,1] \to (0,\infty)$, respectively. And then, X_t is replaced by a triangular array of observations, $X_{t,T}, t = 1, ..., T$, where T is the sample size.

Under the model (2.3.2), there are many way to construct estimators for the parameter curves, but it is nearly impossible to derive the finite sample properties of theses estimators. This problem can be overcomed by rescaling the parameter curves as described above. As $T \to \infty$, more and more observations of each local structure are available and many results of asymptotic analysis for non-stationary processes can be retained.

Formally, the class of locally stationary process is defined as follows.

Definition 2.3.1 (Dahlhaus(1996a)) A sequence of stochastic processes $X_{t,T}, t = 1, ..., T$, is called locally stationary with transfer function A^0 and trend μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^0_{t,T}(\lambda) d\xi(\lambda), \tag{2.3.3}$$

where the following hold:

(i) $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\lambda)} = \xi(-\lambda)$ and

$$\text{cum}\{d\xi(\lambda_1), ..., d\xi(\lambda_k)\} = \eta\left(\sum_{j=1}^{k} \lambda_j\right) g_k(\lambda_1, ..., \lambda_{k-1}) d\lambda_1 ... d\lambda_k, \tag{2.3.4}$$

where cum$(\cdot)$ denotes the cumulant of kth order, $g_1 = 0$, $g_2(\lambda) = 1$, $|g_k(\lambda_1, ..., \lambda_{k-1})| \leq \text{cont}_k$ for all k and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

(ii) There exists a constant K and a 2π-periodic function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ with $A(\mu, -\lambda) = \overline{A(\mu, \lambda)}$ and

$$\sup_{t,\lambda} \left| A^0_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leq KT^{-1} \tag{2.3.5}$$

for all T; $A(\mu, \lambda)$ and $\mu(u)$ are assumed to be continuous in μ.

The smoothness of A in μ guarantees that the process has locally stationary behavior. In Dahlhaus and Polonik (2006), locally stationary process was introduced by using a time-varying MA(∞) representation, the definition is more general since the parameter curves are assumed to be of bounded variation instead of continuity in the time direction. Let

$$V(g) = \sup\left\{ \sum_{k=1}^{m} |g(x_k) - g(x_{k-1})| : 0 \leq x_0 < ... < x_m \leq 1, m \in \mathbb{N} \right\} \tag{2.3.6}$$

be the total variation of a function g on [0, 1], and let

$$l(j) = \begin{cases} 1, & \text{if } |j| \leq 1, \\ |j| \log^{1+k} |j|, & \text{if } |j| > 1 \end{cases} \tag{2.3.7}$$

for some $k > 0$.

Definition 2.3.2 (Dahlahus and Polonik (2006)) The sequence $X_{t,T}$, $t = 1, ..., T$, is a locally stationary process if it has the representation

$$X_{t,T} = \sum_{j=-\infty}^{\infty} a_{t,T}(j) \epsilon_{t-j}, \tag{2.3.8}$$

where the ϵ_t are identically distributed with $E(\epsilon_t) = 0$, $E(\epsilon_s \epsilon_t) = 0$ for $s \neq t$, $E(\epsilon_t^2) = 1$ and where the following condition holds:

$$\sup_t |a_{t,T}(j)| \leq \frac{K}{l(j)} \quad \text{(with } K \text{ not depending on } T) \tag{2.3.9}$$

and there exist functions $a(\cdot, j) : (0, 1] \to \mathbb{R}$ with

$$\sup_u |a(u, j)| \leq \frac{K}{l(j)}, \tag{2.3.10}$$

$$\sup_j \sum_{t=1}^{T} \left| a_{t,T}(j) - a\left(\frac{t}{T}, j\right) \right| \leq K, \tag{2.3.11}$$

$$V(a(\cdot, j)) \leq \frac{K}{l(j)}. \tag{2.3.12}$$

Dahlhaus and Polonik (2009) have shown that time-varying ARMA model fulfills the above conditions (2.3.10-2.3.12). Dahlahus (2012) mentioned that the above conditions are weak since the definition only require bounded variation. For local results some further conditions have to be imposed, namely

$$\sup_u \left| \frac{\partial^i \mu(u)}{\partial \mu^i} \right| \leq K, \text{ for some } i, \tag{2.3.13}$$

$$\sup_u \left| \frac{\partial^i a(u,j)}{\partial u^i} \right| \leq \frac{K}{l(j)} \text{ for } j = 0, 1, ..., \tag{2.3.14}$$

$$\sup_{t,T} \left| a_{t,T}(j) - a\left(\frac{t}{T}, j\right) \right| \leq \frac{K}{Tl(j)}. \tag{2.3.15}$$

The process $X_{t,T}$ can be approximated by the stationary processes

$$\tilde{X}_t(u) = \mu(u) + \sum_{j=-\infty}^{\infty} a(u,j)\epsilon_{t-j} \tag{2.3.16}$$

and the derivative processes is given by

$$\frac{\partial^i \tilde{X}_t(u)}{\partial u^i} = \frac{\partial^i \mu(u)}{\partial u^i} + \sum_{j=-\infty}^{\infty} \frac{\partial^i a(u,j)}{\partial u^i} \epsilon_{t-j}. \tag{2.3.17}$$

The time varying spectral density and the time-varying covariance of lag k at rescaled time u of the stationary approximation $\tilde{X}_t(u)$ are defined by

$$f(u,\lambda) = \frac{1}{2\pi} |A(u,\lambda)|^2 \tag{2.3.18}$$

where

$$A(u,\lambda) = \sum_{j=-\infty}^{\infty} a(u,j) \exp(-i\lambda j) \tag{2.3.19}$$

and

$$c(u,k) = \int_{-\pi}^{\pi} f(u,\lambda) \exp(i\lambda k) d\lambda = \sum_{j=-\infty}^{\infty} a(u,k+j)a(u,j), \tag{2.3.20}$$

respectively.

Under the Definition 2.3.2 and assumption (2.3.15), it can be shown that

$$\text{Cov}(X_{[uT],T} X_{[ut+k],T}) = C(u,k) + O(T^{-1}) \tag{2.3.21}$$

uniformly in u and k, thus, $C(u,k)$ also is the time-varying covariance of the processes $X_{t,T}$ and $f(u,k)$ is the uniquely defined time-varying spectral density of $X_{t,T}$ under the same conditions, the proof can be found in Dahlahus (1996a).

There are many results such as local estimation for the parameter curves and covariance, Kullback-Leibler information divergence, Parametric Whittle-type estimates, Gaussian likelihood theory, empirical spectral processes, etc. These can be found in Dahlhaus (2012) which provides an overview of the locally stationary processes.

2.3.2 Multivariate Locally Stationary Processes

For multivariate processes, the definition is analogous to the univariate case, formally, the definition of multivariate locally stationary processes is as follows.

Definition 2.3.3 (Dahlhaus (2000)) A sequence of Gaussian multivariate stochastic processes $X_{t,T} = (X_{t,T}^{(1)}, ..., X_{t,T}^{(d)})', t = 1, ..., T$, is called locally stationary with transfer function matrix $\mathbf{A}^0$ and mean function vector μ if there exists a representation

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-\pi}^{\pi} \exp(i\lambda t)\mathbf{A}_{t,T}^0(\lambda)d\xi(\lambda) \tag{2.3.22}$$

with the following properties:

(i) $\xi(\lambda)$ is a complex valued Gaussian vector process on $[-\pi, \pi]$ with $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$, $E[\xi_a(\lambda)] = 0$ and

$$E[d\xi_a(\lambda)d\xi_b(\lambda)] = \delta_{ab}\eta(\lambda + \mu)d\lambda d\mu, \tag{2.3.23}$$

where $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

(ii) There exists a constant K and a 2π-periodic matrix valued function $\mathbf{A}$:$[0, 1] \times \mathbb{R} \to \mathbb{C}^{d\times d}$ with $\overline{\mathbf{A}(u, \lambda)} = \mathbf{A}(u, -\lambda)$ and

$$\sup_{t,\lambda} \left| \mathbf{A}_{t,T}^0(\lambda)_{ab} - \mathbf{A}\left(\frac{t}{T}, \lambda\right)_{ab} \right| \leq \frac{K}{T} \tag{2.3.24}$$

for all $a, b = 1, ..., d$ and $T \in \mathbb{N}$. $\mathbf{A}(u, \lambda)$ and $\mu(u)$ are assumed to be continuous in μ.

$f(u, \lambda) := \mathbf{A}(u, \lambda)\overline{\mathbf{A}(u, \lambda)}'$ is the time varying spectral density matrix of the process.

Suppose $X_{t,T}$ is a time-varying VARMA model, the difference equations is defined by

$$\sum_{j=0}^{p} \Phi_j\left(\frac{t}{T}\right)\left[X_{t-j,T} - \mu\left(\frac{t-j}{T}\right)\right] = \sum_{j=0}^{q} \Psi_j\left(\frac{t}{T}\right)\boldsymbol{\epsilon}_{t-j}, \tag{2.3.25}$$

where $\boldsymbol{\epsilon}_t$ are independent, identically distributed with mean zero and covariance matrix $\mathbf{I}_d$ and $\Phi_0(u) \equiv \Psi_0(u) \equiv \mathbf{I}_d$. For $z \in \mathbb{C}$, let $\Phi(u, z) = \sum_{j=0}^{p} \Phi_j(u)z^j$ and $\Psi(u, z) = \sum_{j=0}^{q} \Psi_j(u)z^j$. If $\det[\Phi(u, z)] \neq 0$ for all $|z| \leq 1 + c$ with $c > 0$ uniformly in u and all entries of $\Phi_j(u)$ and $\Psi_j(u)$ are continuous in u, then the solution of these difference equations has an infinite time-varying MA presentation, that is, the solution is locally stationary of the form (2.3.22). The time-varying spectral density of the process is

$$f(u, \lambda) = \frac{1}{2\pi}\frac{\Psi(u, e^{i\lambda})\Psi(u, e^{-i\lambda})'}{\Phi(u, e^{i\lambda})\Phi(u, e^{-i\lambda})'}. \tag{2.3.26}$$

Since STARMA models are special cases of VARMA models, the time-varying STARMA model proposed in this thesis is a locally stationary process and for simplicity, in the sequel we assume that $\mu(\cdot) = 0$.

2.4 Functional Data Analysis

While the term "Functional Data Analysis (FDA)" was introduced by Ramsay (1982), the history of this area is much older. FDA is a topic in growing development as it has sparked great interest from the global statistical community. Most of the initial techniques have been developed for FDA were introduced by Ramsay and Dalzell (1991) and Ramsay and Silverman (1997). Generally, FDA deals with data that are in the form of functions, that is, the ith observation is a real-valued function, $f_i(t)$, $i = 1, ..., n,\ t \in T$, where T is a real interval. Functional data are being observed and investigated with increasing frequency in several areas.

In practice, the functional observation f consists of n pairs (t_i, y_i), where y_i is an observation of $f(t_i)$, that is,

$$y_i = f(t_i) + \epsilon_i, \tag{2.4.1}$$

where ϵ_i denotes measurement error.

A function f can be expressed as a weighted sum or linear combination of elementary functional building blocks called basis functions. Therefore, the conversion of the data to functional form requires two steps: 1) Choosing and defining a set of basis functions. 2) Computing the best linear combination. If the observation has an observational error, some techniques will be used to filter out this noise as efficiently as possible. In some cases, the alternative strategy of leaving the noise in the estimated function is used.

2.4.1 Basis Expansions

A common smoothing method is employed to represent the function as a linear combination of a sufficiently large number K of known basis functions $\phi_k, k = 1, ..., K$, that are mathematically independent of each other and have the property that any function can be approximated arbitrarily well. A function f by a linear combination of known basis functions is defined by

$$f(t) = \sum_{k=1}^{K} c_k \phi_k(t), \tag{2.4.2}$$

where $\mathbf{c} = (c_1, ..., c_K)'$ is the coefficients vector and it is determined by minimizing the sum of squared of errors (SSE). The criterion in matrix terms is expressed as

$$\text{SSE} = [\mathbf{y} - \mathbf{\Phi c}]' \mathbf{W} [\mathbf{y} - \mathbf{\Phi c}],$$

where $\mathbf{y} = (y_1, ..., y_n)'$, $\mathbf{\Phi} = [\phi_k(t_i)]$ is a $n \times K$ matrix and $\mathbf{W}$ is a weight matrix. Then, the weighted least squares estimate of the coefficient vector $\mathbf{c}$ is

$$\hat{\mathbf{c}} = (\mathbf{\Phi' W \Phi})^{-1} \mathbf{\Phi' W y}. \tag{2.4.3}$$

It is important that the basis functions have characteristics similar to those of the functions to be estimates. For periodic data, the Fourier series is well known, function f can be expressed as a linear combination of sine and cosine functions, that is,

$$f(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \dots \tag{2.4.4}$$

This basis is periodic and defined by the basis $\phi_0(t) = 1,\ \phi_{2r-1}(t) = \sin r\omega t$ and $\phi_{2r}(t) = \cos r\omega t$, for $r \geq 1$. The parameter ω determines the period $2\pi/\omega$ which is equal to the length of interval T. As the number of observations is finite, the first K basis functions of the expansion above (2.4.4) can be considered, thereby, the smooth estimate of $f(t)$ is given by

$$\hat{f}(t) = \sum_{k=1}^{K} \hat{c}_k \phi_k(t). \tag{2.4.5}$$

For non-periodic data, although there are many ways that such systems can be constructed, the *B-splines* basis system developed by de Boor (2001) is the most popular, due to its advantage of flexibility and computational speed. The *B-splines* are piecewise polynomials functions joined at certain values called knots. The notation $B_{i,m}(t)$ indicates the value at t of the ith *B-spline* basis function of degree m defined by the knot sequence $\tau = \{\tau_i\}$. By the de Boor's algorithm, *B-splines* of any degree can be computed from *B-splines* of lower degree, the recursion formula is:

$$B_{i,m}(t) = \frac{t - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1} + \frac{\tau_{i+m} - t}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}, \tag{2.4.6}$$

where

$$B_{i,1}(t) = \begin{cases} 1, & \text{if } \tau_i \leq t \leq \tau_{i+1}, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the smoothed estimate of $f(t)$ is given by

$$\hat{f}(t) = \sum_{i=1}^{K} \hat{c}_i B_{i,m}(t). \tag{2.4.7}$$

Note that the number of *B-spline* basis functions is equal the number of interior knots plus the order of the polynomial segments.

There are other basis which are well known as well, they are:

- Exponential basis systems, consist of exponential functions $e^{\lambda_k t},\ k = 1, ..., K$.
- Power bases, $t^{\lambda_k},\ k = 1, ..., K$.
- Polynomial bases, the bases are constructed by the monomial basis $\phi_k(t) = (t - \omega)^k,\ k = 0, 1, ..., K$.
- Wavelets which consist of functions $\psi_{j,k}(t)$ and/or $\phi_{j,k}(t), j, k \in \mathbb{Z}$.

2.4.2 Roughness Penalty

Similar to the least squares methods of the previous section, roughness penalty methods are based on optimizing a fitting criterion that we define a smooth of the data in ways that are appropriate to our problems.

The square of the second derivative of a function at t measures the curvature in f at t, thus, the penalty that measures function's roughness is defined as

$$\text{PEN}_2(f) = \int [D^2 f(s)]^2 ds, \tag{2.4.8}$$

where D^2 refer to the second derivative.

The roughness penalty above can be generalized by allowing a derivative $D^m x$ of arbitrary order, then, the penalty is given by

$$\text{PEN}_m(f) = \int [D^m f(s)]^2 ds. \tag{2.4.9}$$

Adding the roughness penalty to the last squares criterion of section previous, the penalized least squares is defined by

$$\text{PENSSE}_\lambda = [\mathbf{y} - f(\mathbf{t})]' \mathbf{W} [\mathbf{y} - f(\mathbf{t})] + \lambda \text{PEN}_2(f), \tag{2.4.10}$$

where λ is a smoothing parameter that controls the trade-off between fidelity to the data and roughness of the function estimate. As the smoothing parameter varies from zero to infinity, functions vary from being rough to being very smooth. The smoothing parameter λ can be obtained by a procedure such as cross-validation method.

2.4.3 Strictly Monotonic Functions

In practical applications, the estimated function has to satisfy some condition, for example, be non-decreasing and the use of a basis expansion may not be suitable. In this section we consider the case where a monotonic function is of interest.

A monotonic function is a function which is either entirely non-increasing or non-decreasing. An important property of this kind of function is that if a function is a strictly monotonic function, then it is injective on its domain. There are many techniques developed. In this thesis, we use the monotonic function introduced by Ramsay (1998). It is an arbitrary twice differentiable strictly monotonic function defined on an interval closed on the left.

The strictly monotonic function f satisfies the conditions: $\ln(Df)$ is differentiable and $D\{\ln(Df)\} = D^2 f/Df$ is squared Lebesgue integrable and D refers to the operation of taking the derivative and D^{-1} means the integration operator. These conditions guarantee that the function's first derivative is smooth and bounded almost everywhere. By the following theorem, we can see that f is a general solution of differential equation.

Theorem 2.4.1 (Ramsay (1998)) Every monotonic function f is representable as either

$$f(x) = C_0 + C_1 D^{-1}\{\exp\big(D^{-1}w(x)\big)\} \tag{2.4.11}$$

or as a solution of the homogeneous linear differential equation

$$D^2 f(x) = w(x)Df(x), \tag{2.4.12}$$

where $w(x)$ is a Lebesgue square integrable function and C_0 and C_1 are arbitrary constants.

Note that to guarantee that the function f is strictly monotone increasing, $Df = e^W$, where $W = D^{-1}w + \log C_1$ was assumed, for more details see Ramsay and Silverman (2006).

2.5 Wavelets Analysis

The Fourier transform (FT) decomposes a function into simple sines and cosines, that is, any function can be represented as a sum of sine and cosine functions of different amplitudes and frequencies. The FT is suitable to analyze stationary processes but is not appropriate for studying the local behavior of a signal, for example electrocardiography (ECG) signal, where signals have short intervals of characteristic oscillation. Then, the wavelet transform is a good alternative.

According to Morettin (2014), wavelets are used in many areas of statistics, like estimation of a density function, non-parametric regression, estimation of the spectrum of a stationary or non-stationary process etc. In this thesis, we consider wavelets as an alternative of basis functions for the representation of a function of interest, in general, time-varying.

In the case of wavelets, a function $f \in L^2(\mathbb{R})$ can be approximated by a linear combination of binary dilations 2^j and dyadic translations $k2^{-j}$ of a function $\phi(t)$, called scaling function or father wavelet (which is used for capturing the smooth and the low-frequency of the data) and/or of a function $\psi(t)$, called mother wavelet (which is used for capturing the details and the high-frequency of the data). Note that $\phi(t)$ and $\psi(t)$ have to satisfy the following conditions:

- $\int_{-\infty}^{\infty} t^j \phi(t)dt = 1$;
- $\int_{-\infty}^{\infty} \psi(t)dt = 0$;
- $\int_{-\infty}^{\infty} |\psi(t)|dt < \infty$;
- $\frac{|\Psi(\omega)|^2}{|\omega|}d\omega < \infty$, where $\Psi(\omega)$ is the FT of $\psi(t)$;
- $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$ or $\int_{-\infty}^{\infty} |\Psi(\omega)|d\omega = 2\pi$;
- $\int_{-\infty}^{\infty} t^j \psi(t)dt = 0, j = 0, 1, ..., r-1$, for at least a $r \geq 1$ and $\int_{-\infty}^{\infty} |t^r \psi(t)|dt < \infty$.

Thus, a wavelet basis is composed of functions $\{\phi_{j,k}(t) \cup \psi_{j,k}(t), j, k \in \mathbb{Z}\} \in L^2(\mathbb{R})$, where

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \tag{2.5.1}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k). \tag{2.5.2}$$

And the function $f(t) \in L^2(\mathbb{R})$ can be written as

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j \geq j_0}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(t), \tag{2.5.3}$$

where

$$c_{j_0,k} = \int_{-\infty}^{\infty} f(t)\phi_{j_0,k}(t)dt, \tag{2.5.4}$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt \tag{2.5.5}$$

for some coarse scale j_0, see Section 2.5.3.

Sometimes, the function can be expanded only through mother wavelet, that is,

$$f(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(t), \tag{2.5.6}$$

where the wavelet coefficients $c_{j,k}$ can be obtained by the following scalar product

$$c_{j,k} = < g, \psi_{j,k} > = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt. \tag{2.5.7}$$

The scaling function $\phi(t)$ can be obtained by solving the equation

$$\phi(t) = \sqrt{2}\sum_k l_k \phi(2t - k), \tag{2.5.8}$$

and $\psi(t)$ is obtained from $\phi(t)$ by

$$\psi(t) = \sqrt{2}\sum_k h_k \phi(2t - k), \tag{2.5.9}$$

where

$$h_k = (-1)^k l_{1-k}. \tag{2.5.10}$$

In fact, l_k and h_k are the low-pass and high-pass filter coefficients, respectively, the coefficients are given by

$$l_k = \sqrt{2}\int_{-\infty}^{\infty} \phi(t)\phi(2t - k)dt, \tag{2.5.11}$$

$$h_k = \sqrt{2}\int_{-\infty}^{\infty} \psi(t)\phi(2t - k)dt. \tag{2.5.12}$$

2.5.1 Wavelet Families

There are many different families of wavelets in the literature, see Ogden (1997), Misiti et al. (2007), Stéphane (2009), Morettin (2014), etc. Some type of wavelet families are presented in this section.

The oldest and the simplest possible wavelet, Haar wavelet, was proposed by Haar (1910). The Haar wavelet is not continuous, and therefore not differentiable, it is usually used to the analysis of signals with sudden sharp transitions. The scaling and wavelet functions can be described as

$$\psi(t) = \begin{cases} 1, & \text{if } 0 \leq t < \frac{1}{2}, \\ -1, & \text{if } \frac{1}{2} \leq t < 1, \\ 0, & \text{otherwise} \end{cases} \tag{2.5.13}$$

and

$$\phi(t) = \begin{cases} 1, & \text{if } 0 \leq t < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.5.14}$$

Under binary dilations and dyadic translations, we obtain

$$\psi_{j,k}(t) = \begin{cases} 2^{\frac{j}{2}}, & \text{if } 2^{-j}k \leq t < 2^{-j}(k+\frac{1}{2}), \\ -2^{\frac{j}{2}}, & \text{if } 2^{-j}(k+\frac{1}{2}) \leq t < 2^{-j}(k+1), \\ 0, & \text{otherwise} \end{cases} \tag{2.5.15}$$

and

$$\phi_{j,k}(t) = \begin{cases} 2^{\frac{j}{2}}, & \text{if } 2^{-j}k \leq t < 2^{-j}(k+1), \\ 0, & \text{otherwise,} \end{cases} \tag{2.5.16}$$

respectively.

Daublets and Symmlets are wavelets from Daubechies' family (Daubechies (1988) and Daubechies (1992)). If a Daubechies wavelet has N vanishing moments, the support of the corresponding scaling function is $[0, 2N-1]$ and is $[-N+1, N]$ for wavelet function. As a special case of the Daubechies wavelet, the Haar wavelet is also known as Daublet wavelet with 1 vanishing moment.

Different from the wavelets described above, the Mexican hat and Shannon wavelets are continuous wavelets with closed form expression. We have

- Mexican hat wavelet:

$$\psi(t) = \frac{2}{\sqrt{3}\pi^{1/4}}(1-t^2)e^{-t^2/2} \tag{2.5.17}$$

  and in bivariate case:

$$\psi(t_1, t_2) = \frac{1}{\pi}(1 - \frac{t_1^2 + t_2^2}{2})e^{-\frac{t_1^2+t_2^2}{2}}. \tag{2.5.18}$$

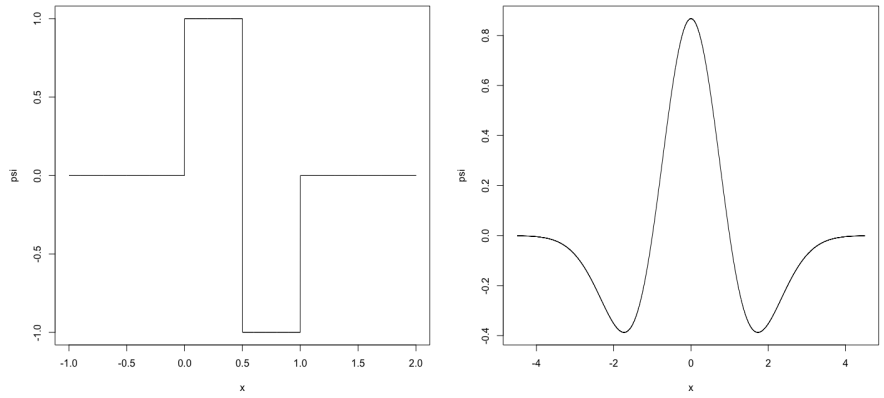
- Shannon wavelet:

$$\psi(t) = \text{sinc}(\frac{t}{2})\cos\left(\frac{3\pi t}{2}\right) = 2\text{sinc}(2t) - \text{sinc}(t), \tag{2.5.19}$$

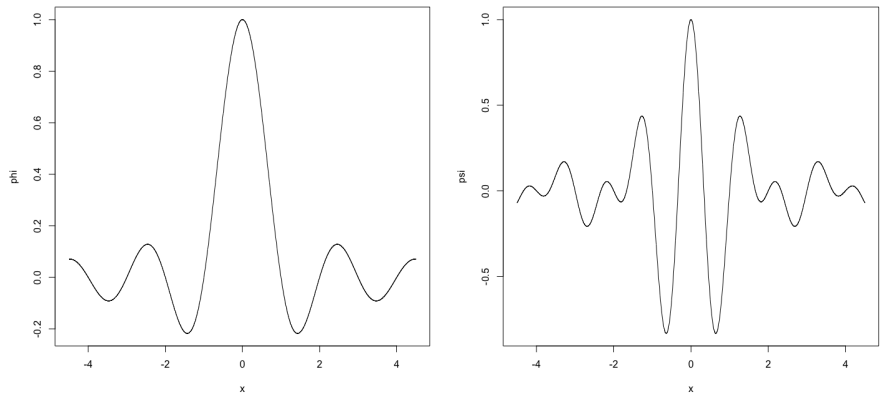
where $\text{sinc}(t) = \frac{\sin \pi t}{\pi t}$. And scaling function is

$$\phi(t) = \text{sinc}(t). \tag{2.5.20}$$

Figures 2.3-2.5 present the scaling ($\phi(t)$) and/or wavelet ($\psi(t)$) functions of Haar wavelet, Daubechies wavelets with different vanishing moments, Mexican hat wavelet and Shannon wavelet.



(a) Haar wavelet (b) Mexican hat wavelet



(c) Shannon wavelet (d) Shannon wavelet

Figure 2.3: Haar, Mexican hat and Shannon wavelets.

2.5.2 Daubechies-Lagarias Algorithm

The wavelets with compact support only have a finite amount of their non-zero filter coefficients and for all compactly supported orthonormal families of wavelet, except of the Haar wavelet, the scaling and wavelet functions have no explicit expression. However, it is always necessary to find their values at given points.

Base on Daubechies and Lagarias (1991 and 1992) local pyramidal algorithm, the functions $\phi(t)$ and $\psi(t)$ can be calculated directly at a point with preassigned precision. Suppose the support of ϕ is

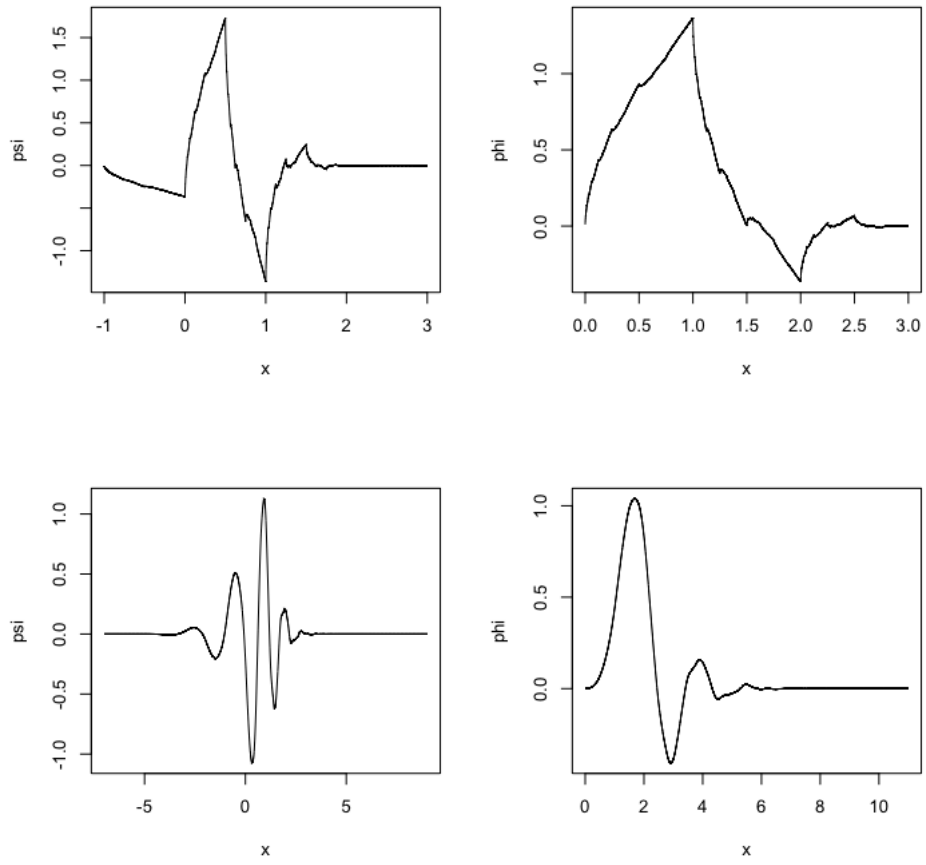


Figure 2.4: Scaling functions (left) and wavelet functions (right) from Daublet with vanishing moments N = 2 (top), 6 (bottom).

[0, 2N-1], let $x \in (0,1)$ and dyad$(x) = d_1, d_2, ..., d_n, ...$ be the set of 0/1 digits in dyadic representation of x, that is, $x = \sum_{j=1}^{\infty} d_j 2^{-j}$. The subset of the first n digits from dyad(x) is denoted by dyad(x, n).

Let $h = (h_0, h_1, ..., h_{2N-1})$ be the wavelet filter coefficients, two $(2N-1) \times (2N-1)$ matrices are defined as

$$T_0 = (\sqrt{2}h_{2i-j-1})_{1 \le i,j \le 2N-1}, \tag{2.5.21}$$

$$T_1 = (\sqrt{2}h_{2i-j})_{1 \le i,j \le 2N-1}. \tag{2.5.22}$$

The algorithm can be obtained from

$$\lim_{n\to\infty} T_{d_1} \cdot T_{d_2} \dots T_{d_n} = \begin{bmatrix} \phi(x) & \phi(x) & \dots & \phi(x) \\ \phi(x+1) & \phi(x+1) & \dots & \phi(x+1) \\ \vdots & \vdots & \dots & \vdots \\ \phi(x+2N-2) & \phi(x+2N-2) & \dots & \phi(x+2N-2) \end{bmatrix}. \tag{2.5.23}$$

The Daubechies-Lagarias algorithm gives only the values of the scaling function, for the wavelet function, Pinheiro and Vidakovic (1997) proposed the following theorem

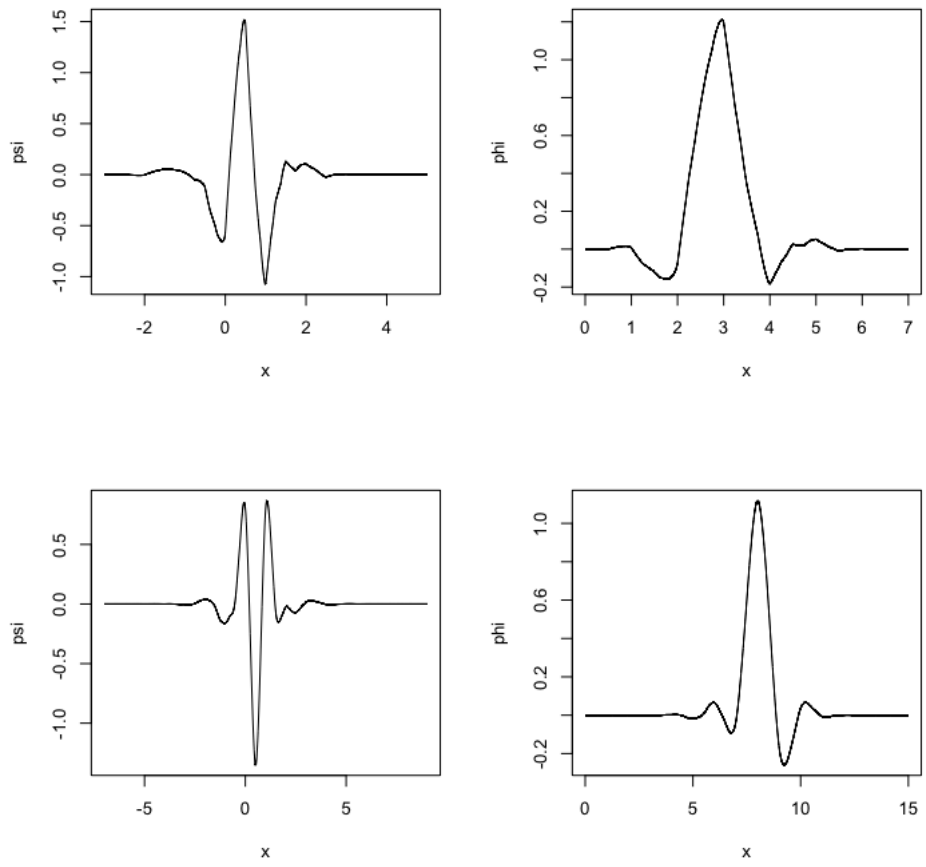


Figure 2.5: Scaling functions (left) and wavelet functions (right) from Symmlet with vanishing moments N, for N = 4 (top), 8 (bottom).

Theorem 2.5.1 (Pinheiro and Vidakovic (1997)) Let x be an arbitrary real number, let the wavelet be given by its filter coefficients, and let u with $2N$-1 be a vector defined as

$$u(x) = \{(-1)^{1-[2x]} h_{i+1-[2x]}, i = 0, ..., 2N-2\}. \tag{2.5.24}$$

If for some i the index $i+1-[2x]$ is negative or larger than $2N-1$, then the corresponding component of u is equal to 0.

Let the vector v be

$$v(x,n) = \frac{1}{2N-1}\mathbf{1}' \prod_{i \in \text{dyad}(2x,n)} T_i, \tag{2.5.25}$$

where **1**' is the row-vector of ones. Then

$$\psi(x) = \lim_{n\to\infty} u(x)' v(x,n), \tag{2.5.26}$$

and the limit is constructive.

An alternative to calculate $\psi(t)$ is using equation (2.5.9), with $\phi(t)$ obtained by equation (2.5.23).

2.5.3 A Multiresolution Analysis

A Multiresolution analysis (MRA) allows us to analysis the data at different levels of resolution. The data with coarse resolution contain information about lower-frequency components and retain the main features of the original signal and the data with finer resolution retain information about the higher-frequency components.

Formally, a MRA can be viewed as a sequence of approximations of a given function $f(t)$ of $L^2(\mathbb{R})$ at different resolutions. Let V_j and W_j be closed subspaces generated by $\{\phi_{j,k}, k = 0, ..., 2^j - 1\}$ and $\{\psi_{j,k}, k = 0, ..., 2^j - 1\}$, respectively, then the MRA have the following properties:

- $... \subset V_{-1} \subset V_0 \subset V_1 \subset ...$
- $L^2(\mathbb{R}) = \bigcup_j \overline{V_j}$
- $\bigcap_j V_j = \{0\}$
- $x(t) \in V_j \Leftrightarrow x(2t) \in V_{j+1},\ \forall j$
- $V_{j+1} = V_j \oplus W_j,\ W_j \perp V_j.$

The above properties imply $W_j = V_{j+1} \ominus V_j$, and then,

$$L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{\infty} W_j. \tag{2.5.27}$$

Let $f \in L^2(\mathbb{R})$, there is J that $f_J \in V_J$ approximates f. If $g_i \in W_i,\ f_i \in V_i$, by the last property, we have

$$f_J = f_{J-1} + g_{J-1}, \tag{2.5.28}$$

repeating the argument,

$$f \simeq f_J = g_{J-1} + g_{J-2} + ... + g_{J-M} + f_{J-M},\ M = 1, ..., J-1. \tag{2.5.29}$$

Observe that f_{J-M} is a linear combination of $\phi_{J-M,k}$ and g_j are linear combinations of $\psi_{j,k},\ j = J-M, ..., J-1$. And,

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j = V_0 \oplus \bigoplus_{j\geq 0} W_j = V_{j_0} \oplus \bigoplus_{j\geq j_0} W_j, \tag{2.5.30}$$

for some coarse scale j_0, any $f(t) \in L^2(\mathbb{R})$ can be expressed as

$$\begin{aligned} f(t) &= \sum_{j,k} d_{j,k}\psi_{j,k}(t) \\ &= \sum_k c_{J_0,k}\phi_{j_0,k}(t) + \sum_{j\geq j_0}\sum_k d_{j,k}\psi_{j,k}(t) \\ &= \sum_k c_{0,k}\phi_{0,k}(t) + \sum_{j\geq 0}\sum_k d_{j,k}\psi_{j,k}(t) \end{aligned} \tag{2.5.31}$$

where $j_0 = 0$ and

$$c_{j,k} = \int_{-\infty}^{\infty} f(t)\phi_{j,k}(t)dt,$$
$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt.$$

Note that in wavelet analysis, the number of observations n has to be power of 2 and the number of basis functions depends on the value of J and/or J_0, where $n = 2^J,\ J > 0$ integer and $J_0 = 0, 1, ..., J-1$ represent the level associated with the scale 2^{J_0}, the finest scale corresponds to the $J-1$ level and the most coarse scale corresponds to the zero level.

Chapter 3

Time-varying Space-time Autoregressive and Moving Average Models

3.1 Wavelet Based Time-varying STAR Model

Let $\mathbf{Z}(\frac{t}{T}) = [Z_1(\frac{t}{T}), ..., Z_n(\frac{t}{T})]'$ be a n-dimensional time series with T observations. The time-varying STAR (tvSTAR) model is defined by

$$\mathbf{Z}\left(\frac{t}{T}\right) = \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\phi_{sl}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right),\ t = 1, ..., T, \tag{3.1.1}$$

where $\boldsymbol{\epsilon}(\frac{t}{T})$ is an independent, identically distributed gaussian vector with mean zero, $\phi_{sl}(\frac{t}{T})$ is time-varying parameter at time lag s and space lag l and $W^{(l)}$ is spatial weight matrix of the order l.

Wavelet bases, despite having irregular shapes, are able to perfectly reconstruct functions with linear and higher-order polynomial shapes. The idea is to expand the time-varying parameters $\phi_{sl}(\frac{t}{T})$ in wavelet expansions as

$$\phi_{sl}\left(\frac{t}{T}\right) = \sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right). \tag{3.1.2}$$

Notice that, for simplification, we define $\psi_{-1,0}(\frac{t}{T}) = \phi_{0,0}(\frac{t}{T})$, the scaling function with $j, k = 0$.

Then, the wavelet based tvSTAR model is given by

$$\begin{aligned}
\mathbf{Z}\left(\frac{t}{T}\right) &= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right) \\
&= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) \\
&+ \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right)
\end{aligned}$$

$$= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)+\boldsymbol{\nu}\left(\frac{t}{T}\right), \tag{3.1.3}$$

where

$$\boldsymbol{\nu}\left(\frac{t}{T}\right)=\sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)+\boldsymbol{\epsilon}\left(\frac{t}{T}\right) \tag{3.1.4}$$

and $W^{(l)}$ is $n\times n$ matrix of weights with each row having sum one. Note that $J-1$ is the the finest resolution level such that $2^{J-1}\leq\sqrt{T}\leq 2^J$, and then, $\sqrt{n}2^J/2\leq\sqrt{nT}\leq\sqrt{n}2^J$ with $n\geq 2$.

The linear form of (3.1.3) can be written as

$$\mathbf{Z}=\boldsymbol{\Psi}\boldsymbol{\beta}+\boldsymbol{\nu}, \tag{3.1.5}$$

where

$$\mathbf{Z} = \left[\text{vec}\left(\mathbf{Z}\left(\frac{p+1}{T}\right)\right)',...,\text{vec}\left(\mathbf{Z}\left(\frac{T}{T}\right)\right)'\right]', \tag{3.1.6}$$

$$\boldsymbol{\nu} = \left[\text{vec}\left(\boldsymbol{\nu}\left(\frac{p+1}{T}\right)\right)',...,\text{vec}\left(\boldsymbol{\nu}\left(\frac{T}{T}\right)\right)'\right]', \tag{3.1.7}$$

$$\boldsymbol{\Psi} = \left[\Psi_{-1,0},\Psi_{0,0},...,\Psi_{J-1,2^{J-1}-1}\right] \tag{3.1.8}$$

and

$$\Psi_{j,k}=\begin{bmatrix}\psi_{j,k}\left(\frac{p+1}{T}\right)\sum_{m=1}^{n}w_{1,m}^{(0)}Z_m\left(\frac{p}{T}\right) & \dots & \psi_{j,k}\left(\frac{p+1}{T}\right)\sum_{m=1}^{n}w_{1,m}^{(\lambda_p)}Z_m\left(\frac{1}{T}\right)\\ \vdots & & \vdots\\ \psi_{j,k}\left(\frac{T}{T}\right)\sum_{m=1}^{n}w_{1,m}^{(0)}Z_m\left(\frac{T-1}{T}\right) & \dots & \psi_{j,k}\left(\frac{T}{T}\right)\sum_{m=1}^{n}w_{1,m}^{(\lambda_p)}Z_m\left(\frac{T-p}{T}\right)\\ \psi_{j,k}\left(\frac{p+1}{T}\right)\sum_{m=1}^{n}w_{2,m}^{(0)}Z_m\left(\frac{p}{T}\right) & \dots & \psi_{j,k}\left(\frac{p+1}{T}\right)\sum_{m=1}^{n}w_{2,m}^{(\lambda_p)}Z_m\left(\frac{1}{T}\right)\\ \vdots & & \vdots\\ \psi_{j,k}\left(\frac{T}{T}\right)\sum_{m=1}^{n}w_{n,m}^{(0)}Z_m\left(\frac{T-1}{T}\right) & \dots & \psi_{j,k}\left(\frac{T}{T}\right)\sum_{m=1}^{n}w_{n,m}^{(\lambda_p)}Z_m\left(\frac{T-p}{T}\right)\end{bmatrix}, \tag{3.1.9}$$

for $j=-1,...,J-1,\ k=0,...,2^j-1$.

$\boldsymbol{\beta}$ is a $s_a(2^J)\times 1$ vector containing the wavelet coefficients, where $s_a=\sum_{k=1}^{p}(1+\lambda_k)$.

3.1.1 Estimation

The number of parameters depends on the spatial order, λ_s, for $s=1,..,p$ and the resolution level J. For really large datasets, there are many parameters to be estimated, the method of least squares is a computationally efficiency estimation method in this case.

Assume that the matrix $\boldsymbol{\Psi}'\boldsymbol{\Psi}$ is positive definite, the least squares estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{Z}. \tag{3.1.10}$$

Consider the following assumptions:

A.1 (Dahlhaus et al. (1999)) The functions ϕ_{kl} are real, bounded and belong to the following set of functions:

$$\mathscr{F} = \left\{ f(x) : f(x) = \sum_{j=-1}^{\infty} \sum_{k=0}^{\infty} \beta_{j,k} \psi_{j,k}(x) \,\middle|\, \|\beta_{..}\|_{d,p,q} < \infty \right\}, \tag{3.1.11}$$

where

$$\|\beta_{..}\|_{d,p,q} = \left[\sum_{j \geq -1} \left(2^{jup} \sum_{k=0}^{2^j - 1} |\beta_{j,k}|^p \right)^{q/p} \right]^{1/q}, \tag{3.1.12}$$

$u = d + 1/2 - 1/\tilde{p} > 1$ with $\tilde{p} = \min\{p, 2\}$. Note that d is the degree of smoothness and $1 < p, q \leq \infty$ specify the norm in which smoothness is measured.

A.2 The functions $\phi(t)$ and $\psi(t)$ belong to $C^r[0,1]$, with degree of regularity $r > d$ and they have compact support. In addition, $\int \phi(t)dt = 1$ and $\int \psi(t)^k dt = 0$ for $0 \leq k \leq r$.

A.3 There exists some $\gamma \geq 0$ with $|\text{cum}_n(\nu_t)| \leq A^n (n!)^{1+\gamma}$ for all n, t, where A is a positive constant and $\nu_t = \nu(\frac{t}{T})$.

A.4 $\mathbf{Z}(\frac{t}{T})$ are locally stationary processes.

Proposition 1 Suppose the assumptions A.1-A.4 hold, then

(i.) $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + O((nT)^{-1/2})$.

(ii.) $E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = O((nT)^{-1})$.

Proof. See Appendix.

Proposition 2 Suppose the assumptions A.1-A.4 hold, then we have

$$\sqrt{N}\mathbf{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathscr{N}_K(\mathbf{0}, \boldsymbol{\Gamma}), \tag{3.1.13}$$

where D means convergence in distribution and covariance matrix

$$\boldsymbol{\Gamma} = \lim_{N \to \infty} N\mathbf{H}E[\boldsymbol{\Psi}'\boldsymbol{\Psi}]^{-1}\mathbf{H}', \tag{3.1.14}$$

where $N = n(T - p)$ and $\mathbf{H}$ is a matrix with K rows.

Proof. See Appendix.

3.2 Wavelet Based Time-varying STARMA Model

3.2.1 The Model

Let $\mathbf{Z}\left(\frac{t}{T}\right) = [Z_1(\frac{t}{T}), Z_2(\frac{t}{T}), ..., Z_n(\frac{t}{T})]'$ be a multivariate locally stationary of n-dimensional time-varying STARMA (tvSTARMA) model, defined by

$$\mathbf{Z}\left(\frac{t}{T}\right) = \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\phi_{sl}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) - \sum_{s=1}^{q}\sum_{l=0}^{m_s}\theta_{sl}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right),\ t = 1, ..., T, \tag{3.2.1}$$

where $\boldsymbol{\epsilon}(\frac{t}{T})$ is independent, identically distributed gaussian vector with mean zero, $\phi_{sl}(\frac{t}{T})$ and $\theta_{sl}(\frac{t}{T})$ are time-varying parameters at time lag s and space lag l and $W^{(l)}$ is a spatial weight matrix of the order l.

The time-varying parameters $\phi_{sl}(\frac{t}{T})$ and $\theta_{sl}(\frac{t}{T})$ can be written in wavelet expansions as

$$\phi_{sl}\left(\frac{t}{T}\right) = \sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right), \tag{3.2.2}$$

$$\theta_{sl}\left(\frac{t}{T}\right) = \sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right). \tag{3.2.3}$$

Replacing equations (3.2.2), (3.2.3) into (3.2.1), the wavelet based tvSTARMA model is given by

$$\begin{aligned}
\mathbf{Z}\left(\frac{t}{T}\right) &= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)\\
&- \sum_{s=1}^{q}\sum_{l=0}^{m_s}\sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1}b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right)\\
&= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)\\
&- \sum_{s=1}^{q}\sum_{l=0}^{m_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right)\\
&+ \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1}a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)\\
&- \sum_{s=1}^{q}\sum_{l=0}^{m_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1}b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right)\\
&= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right)\\
&- \sum_{s=1}^{q}\sum_{l=0}^{m_s}\sum_{j=-1}^{J-1}\sum_{k=0}^{2^j-1}b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right) + \boldsymbol{\nu}\left(\frac{t}{T}\right),
\end{aligned} \tag{3.2.4}$$

where

$$\begin{aligned}\boldsymbol{\nu}\left(\frac{t}{T}\right) &= \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1} a_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) \\ &- \sum_{s=1}^{q}\sum_{l=0}^{m_s}\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1} b_{j,k}^{sl}\psi_{j,k}\left(\frac{t}{T}\right)W^{(l)}\boldsymbol{\epsilon}\left(\frac{t-s}{T}\right)+\boldsymbol{\epsilon}\left(\frac{t}{T}\right). \qquad (3.2.5)\end{aligned}$$

3.2.2 Estimation

As in the case of STARMA model, linear and non-linear estimators can be used to tvSTARMA model's estimation. Maximum likelihood estimation is a widely used technique, but it is computationally expensive when there is a large number of parameters to be estimated and can be sensitive to the choice of starting values and then the optimization algorithms may converge to a local minimum or even not converge.

An alternative is the Kalman filter (Kalman (1960)), associated with the state space model

$$\mathbf{x}_{t+1} = F_t\mathbf{x}_t + \Gamma_t\mathbf{w}_{t+1}, \qquad (3.2.6)$$

$$\mathbf{z}_t = M_t\mathbf{x}_t + \mathbf{v}_t, \qquad (3.2.7)$$

where (3.2.6) is the state equation and (3.2.7) is the observation equation of the system, $\mathbf{x}_t$ is the state variable at time t, $\mathbf{z}_t$ is the observation at t, F_t, Γ_t and M_t are state transition matrix, control input matrix and observation matrix, respectively. The vectors $\mathbf{v}_t$ and $\mathbf{w}_t$ are errors assumed to be Gaussian. The algorithm provides recursive estimation formulas to estimate the state of a process, in the sense of minimizing the squared error.

For adaptive parameter estimation in tvSTARMA model by Kalman filter, first we rewrite (3.2.4) in linear form

$$\mathbf{Z}\left(\frac{t}{T}\right) = \mathbf{Y}\left(\frac{t}{T}\right)\mathbf{c} + \boldsymbol{\nu}\left(\frac{t}{T}\right), \qquad (3.2.8)$$

where $\boldsymbol{\nu}\left(\frac{t}{T}\right)$ is assume do have a multivariate normal distribution with mean zero and covariance matrix Σ,

$$\begin{aligned} s_a &= \sum_{s=1}^{p}(1+\lambda_s), \\ s_m &= \sum_{s=1}^{q}(1+m_s), \\ \mathbf{c}_{2^J[s_a+s_m]\times 1} &= \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}, \\ \mathbf{a}'_{2^J[s_a]\times 1} &= [a^{10}_{-1,0}\ a^{10}_{0,0}\ \ldots\ a^{10}_{J-1,2^{J-1}-1}\ \ldots\ a^{p\lambda_p}_{J-1,2^{J-1}-1}], \\ \mathbf{b}'_{2^J[s_m]\times 1} &= [b^{10}_{-1,0}\ b^{10}_{0,0}\ \ldots\ b^{10}_{J-1,2^{J-1}-1}\ \ldots\ b^{qm_q}_{J-1,2^{J-1}-1}], \end{aligned}$$

$$\mathbf{Y}\left(\frac{t}{T}\right)_{n\times 2^J[s_a+s_m]} = \left[\text{diag}_n \otimes \mathbf{1}_{1\times[s_a+s_m]}\mathbf{WD}\left(\frac{t}{T}\right)\right] \otimes \Psi\left(\frac{t}{T}\right), \text{ where,}$$

$$\Psi\left(\frac{t}{T}\right)'_{2^J\times 1} = \left[\psi_{-1,0}\left(\frac{t}{T}\right)\ \psi_{0,0}\left(\frac{t}{T}\right)\ \ldots\ \psi_{J-1,2^{J-1}-1}\left(\frac{t}{T}\right)\right],$$

$$\mathbf{W}_{n[s_a+s_m]\times n[s_a+s_m]} = \begin{bmatrix} \mathbf{W}^{ar} & \\ & \mathbf{W}^{ma} \end{bmatrix},$$

$$\mathbf{W}^{ar}_{n\times n[s_a]} = \begin{bmatrix} W^{(0)} & & & & & & & \\ & \ddots & & & & & & \\ & & W^{(\lambda_1)} & & & & & \\ & & & W^{(0)} & & & & \\ & & & & \ddots & & & \\ & & & & & W^{(\lambda_2)} & & \\ & & & & & & \ddots & \\ & & & & & & & W^{(\lambda_p)} \end{bmatrix},$$

$$\mathbf{W}^{ma}_{n\times n[s_m]} = \begin{bmatrix} W^{(0)} & & & & & & & \\ & \ddots & & & & & & \\ & & W^{(m_1)} & & & & & \\ & & & W^{(0)} & & & & \\ & & & & \ddots & & & \\ & & & & & W^{(m_2)} & & \\ & & & & & & \ddots & \\ & & & & & & & W^{(m_q)} \end{bmatrix},$$

$$\mathbf{D}\left(\frac{t}{T}\right)_{n[s_a+s_m]\times[s_a+s_m]} = \begin{bmatrix} \mathbf{X}\left(\frac{t}{T}\right) & \\ & \mathbf{E}\left(\frac{t}{T}\right) \end{bmatrix},$$

$$\mathbf{X}\left(\frac{t}{T}\right)_{n[s_a]\times s_a} = \begin{bmatrix} \text{diag}_{(1+\lambda_1)} \otimes \mathbf{Z}(\frac{t-1}{T}) & & & \\ & \text{diag}_{(1+\lambda_2)} \otimes \mathbf{Z}(\frac{t-2}{T}) & & \\ & & \ddots & \\ & & & \text{diag}_{(1+\lambda_p)} \otimes \mathbf{Z}(\frac{t-p}{T}) \end{bmatrix},$$

$$\mathbf{E}\left(\frac{t}{T}\right)_{n[s_m]\times s_m} = \begin{bmatrix} \text{diag}_{(1+m_1)} \otimes \boldsymbol{\epsilon}(\frac{t-1}{T}) & & & \\ & \text{diag}_{(1+m_2)} \otimes \boldsymbol{\epsilon}(\frac{t-2}{T}) & & \\ & & \ddots & \\ & & & \text{diag}_{(1+m_q)} \otimes \boldsymbol{\epsilon}(\frac{t-q}{T}) \end{bmatrix}.$$

According to Cipra and Motykova (1987), the state space representation (3.2.6) and (3.2.7) for

the estimation of the parameters $\mathbf{c}$ can be written as

$$\mathbf{c}_{t+1} = \mathbf{c}_t, \tag{3.2.9}$$

$$\mathbf{z}_t = \mathbf{y}_t\mathbf{c}_t + \boldsymbol{\epsilon}_t, \tag{3.2.10}$$

where $F_t = \mathbf{I}_{2^J[s_a+s_m]}$, $\Gamma_t = \mathbf{0}_{2^J[s_a+s_m]}$, $\mathbf{z}_t = \mathbf{Z}(\frac{t}{T})$, $M_t = \mathbf{y}_t = \mathbf{Y}(\frac{t}{T})$, $\mathbf{v}_t = \boldsymbol{\nu}_t = \boldsymbol{\nu}(\frac{t}{T})$. Then, the recursive equations have the following form:

$$\hat{\mathbf{c}}_{t+1} = \hat{\mathbf{c}}_t + \mathbf{P}_{t+1}\mathbf{y}'_{t+1}\hat{\boldsymbol{\Sigma}}_t^{-1}(\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\hat{\mathbf{c}}_t), \tag{3.2.11}$$

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \mathbf{P}_t\mathbf{y}'_{t+1}(\mathbf{y}_{t+1}\mathbf{P}_t\mathbf{y}'_{t+1} + \hat{\boldsymbol{\Sigma}}_t)^{-1}\mathbf{y}'_{t+1}\mathbf{P}_t, \tag{3.2.12}$$

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \frac{1}{t+1-(s_a+s_m)}\{[t-(s_a+s_m)]\hat{\boldsymbol{\Sigma}}_t + (\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\hat{\mathbf{c}}_{t+1})(\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\hat{\mathbf{c}}_{t+1})'\}, \tag{3.2.13}$$

$$\hat{\boldsymbol{\nu}}_{t+1} = \mathbf{z}_{t+1} - \mathbf{y}_{t+1}\hat{\mathbf{c}}_{t+1}. \tag{3.2.14}$$

Note that if there is no a priori information on the parameters, the initial values of the estimates at time t_0 can be chosen as

$$\hat{\mathbf{c}}_{t_0} = \mathbf{0}_{2^J[s_a+s_m]\times 1},\ \mathbf{P}_{t_0} = \mathbf{I}_{2^J[s_a+s_m]\times 1},\ \hat{\boldsymbol{\Sigma}}_{t_0} = h\mathbf{I}_{n\times n}, \tag{3.2.15}$$

where h is a small positive constant.

3.3 Simulations

This section presents some simulation examples in order to evaluate the performance of the proposed estimation procedure.

3.3.1 Simulation Procedure

The simulations consist of the following steps:

[1.] Let $n = 15$ sample locations generated (as shown in Figure 3.1), and then, $M = 1000$ experiments of n time series with length $T = 1024$ are simulated.

[2.] Each sample data $\mathbf{z} = [\mathbf{z}(\frac{1}{T}), ..., \mathbf{z}(\frac{T}{T})]$ is simulated from a tvSTAR(1_1) process

$$\mathbf{z}\left(\frac{t}{T}\right) = \phi_{10}(t)W^{(0)}\mathbf{z}\left(\frac{t-1}{T}\right) + \phi_{11}(t)W^{(1)}\mathbf{z}\left(\frac{t-1}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right) \tag{3.3.1}$$

or a tvSTARMA($1_1, 1_1$) process

$$\begin{aligned}\mathbf{z}\left(\frac{t}{T}\right) &= \phi_{10}(t)W^{(0)}\mathbf{z}\left(\frac{t-1}{T}\right) + \phi_{11}(t)W^{(1)}\mathbf{z}\left(\frac{t-1}{T}\right) \\ &+ \theta_{10}(t)W^{(0)}\boldsymbol{\epsilon}\left(\frac{t-1}{T}\right) + \theta_{11}(t)W^{(1)}\boldsymbol{\epsilon}\left(\frac{t-1}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right),\end{aligned} \tag{3.3.2}$$

where $\boldsymbol{\epsilon}(t)$ is a multivariate normal process with mean zero and covariance matrix $\mathbf{I}_{nT}\sigma^2$, with $\sigma^2 = 1$. The elements of the spatial weight matrix W are defined by

$$w_{ij} = \frac{d_{ij}^{-0.5}}{\sum_{k \neq i} d_{ik}^{-0.5}}, \tag{3.3.3}$$

d_{ij} is great-circle distance between locals $\mathbf{x}_i$ and $\mathbf{x}_j$ and it is the same as (2.2.14).

[3.] Calculate the estimates of $\phi_{10}(t), \phi_{11}(t)$ and/or $\theta_{10}(t), \theta_{11}(t)$ using the estimation procedure described in Sections 3.1.2 and 3.2.2.

[4.] Calculate mean squared error (MSE) of the predictors $\hat{\mathbf{z}}$,

$$\text{MSE} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} [z_i \left(\frac{t}{T}\right) - \hat{z}_i \left(\frac{t}{T}\right)]^2, \tag{3.3.4}$$

where $\hat{z}_i \left(\frac{t}{T}\right)$ is the estimator of $z_i \left(\frac{t}{T}\right)$, the observation at location i and time t.

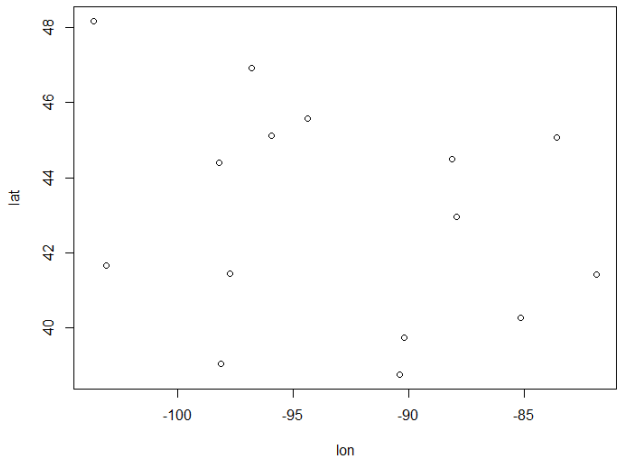


Figure 3.1: Simulated locations.

3.3.2 Formulation of Time-varying Parameter

Four groups of parameters will be used to generate the sample data, they are

- Group 1:

$$\phi_{10}(t) = 0.4\mathrm{I}_{(\frac{t}{T} \leq 0.5)} - 0.8\mathrm{I}_{(\frac{t}{T} > 0.5)}, \tag{3.3.5}$$

$$\phi_{11}(t) = -0.6\mathrm{I}_{(\frac{t}{T} \leq 0.5)} + 0.2\mathrm{I}_{(\frac{t}{T} > 0.5)}; \tag{3.3.6}$$

- Group 2:

$$\phi_{10}(t) = 0.138 + \left(0.316 + 0.982\frac{t}{T}\right) \exp\left(-3.89 \left(\frac{t}{T}\right)^2\right), \tag{3.3.7}$$

$$\phi_{11}(t) = -0.437 - \left(0.659 + 1.26\frac{t}{T}\right) \exp\left(-3.89 \left(\frac{t}{T}\right)^2\right); \tag{3.3.8}$$

- Group 3:

$$\phi_{10}(t) = 0.5 - \frac{\sin\left(\frac{2\pi t}{T}\right)}{4}, \tag{3.3.9}$$

$$\phi_{11}(t) = -0.5 - \frac{\cos\left(\frac{2\pi t}{T}\right)}{4}; \tag{3.3.10}$$

- Group 4:

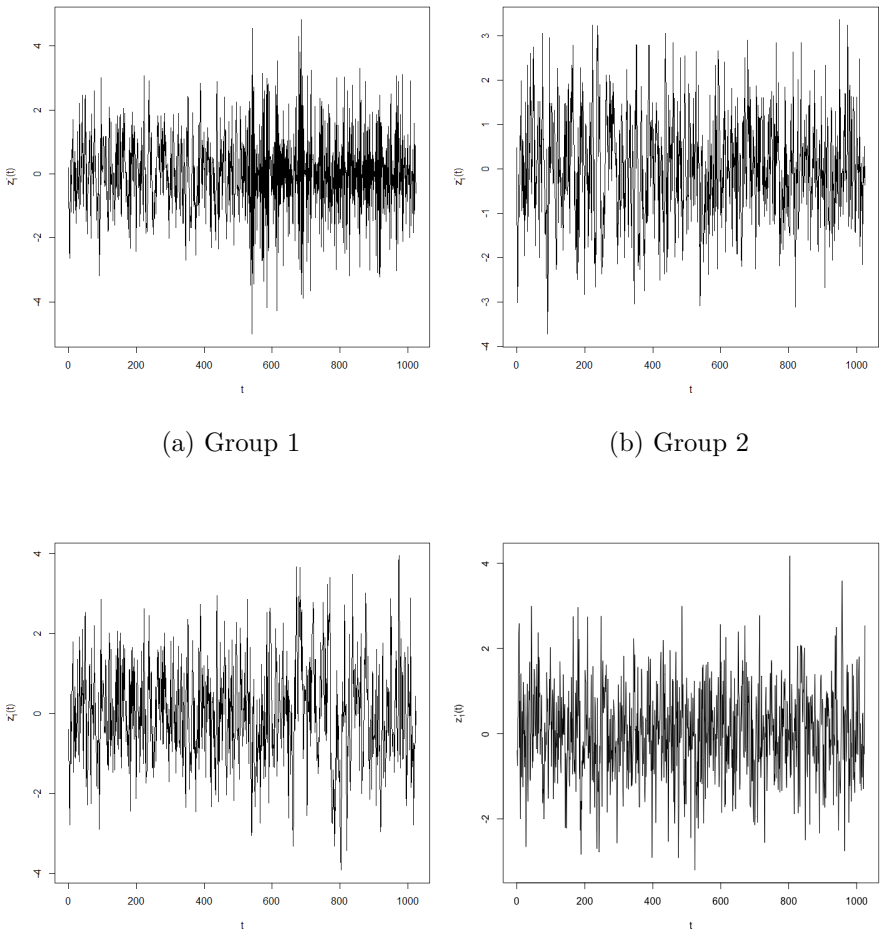
$$\phi_{10}(t) = 0.5\left(1 - \frac{t}{T}\right)^2, \tag{3.3.11}$$

$$\phi_{11}(t) = -0.5\left(1 - \frac{t}{T}\right)^2, \tag{3.3.12}$$

$$\theta_{10}(t) = 0.5\left(\frac{t}{T}\right)^2, \tag{3.3.13}$$

$$\theta_{11}(t) = -0.5\left(\frac{t}{T}\right)^2. \tag{3.3.14}$$

The series of one location of the sampling tvSTAR(1_1) and tvSTARMA($1_1, 1_1$) processes simulated are presented in Figure 3.2.



(a) Group 1 (b) Group 2

(c) Group 3 (d) Group 4

Figure 3.2: Simulated series for one location.

3.3.3 Results

Since the true parameters of Group 1 are discontinuous functions, Haar wavelet was used. In other groups, the true parameters are smooth functions, then, Mexican hat wavelet was chosen. Note that all estimates were calculated with $J = 2$.

Figures 3.3-3.4 show the boxplots of MSEs of the estimates of $\mathbf{z}$, we can see that the MSEs of Group 1 are a little larger than others which the parameters were calculated using the Mexican hat wavelets. Figures 3.5, 3.6, 3.7 and 3.8 show the comparison of the true parameters versus the average of estimates over 1000 experiments of Groups 1, 2, 3 and 4, respectively. The comparisons present a high similarity between the parameters estimated and the true parameters, that is, all groups show a satisfactory performance of the estimation proposed.

Figures B.1-B.6 of Appendix B present the histograms of estimated coefficients of ϕ_{10} and ϕ_{11} of Groups 1-3 which were obtained by the least squares estimation. Independent of the wavelet basis selected, the estimates look like asymptotic normality. Although the coefficients of Group 4 are estimated by Kalman filter, the histograms of Figures B.7-B.10 of Appendix B show the same conclusion.

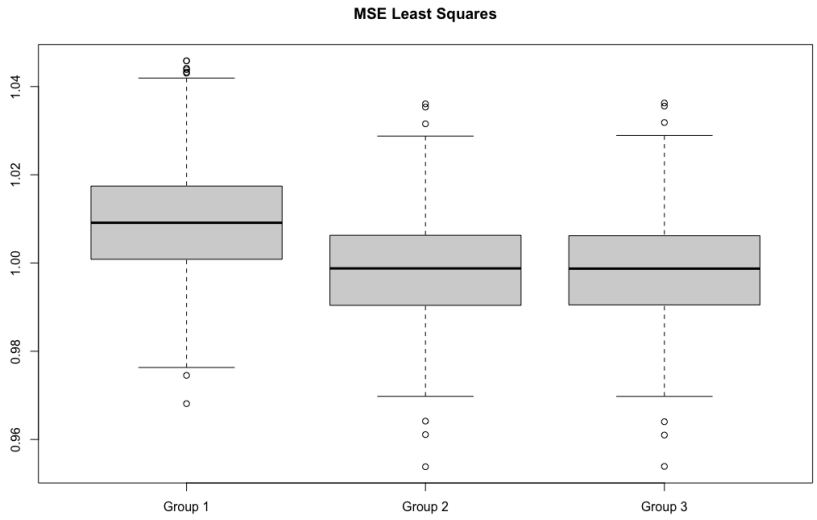


Figure 3.3: Boxplots of MSEs of 1000 experiment for Groups 1-3.

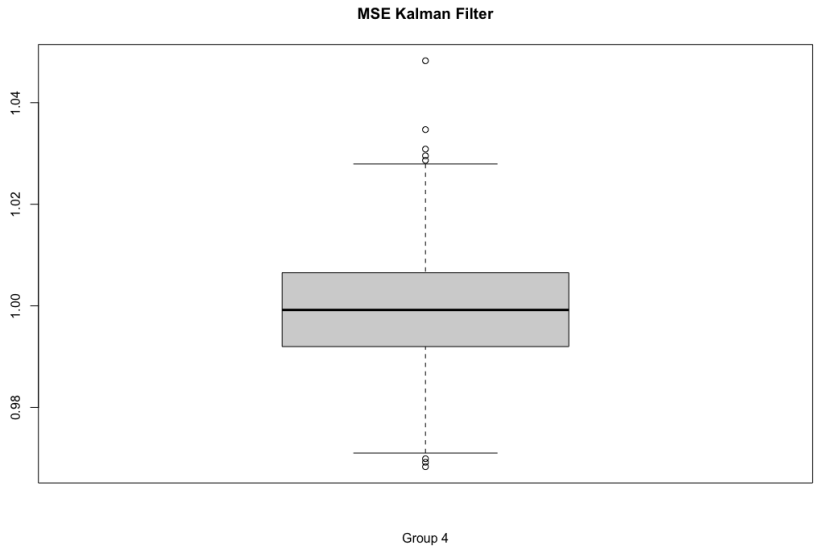


Figure 3.4: Boxplot of MSEs of each experiment for Groups 4.

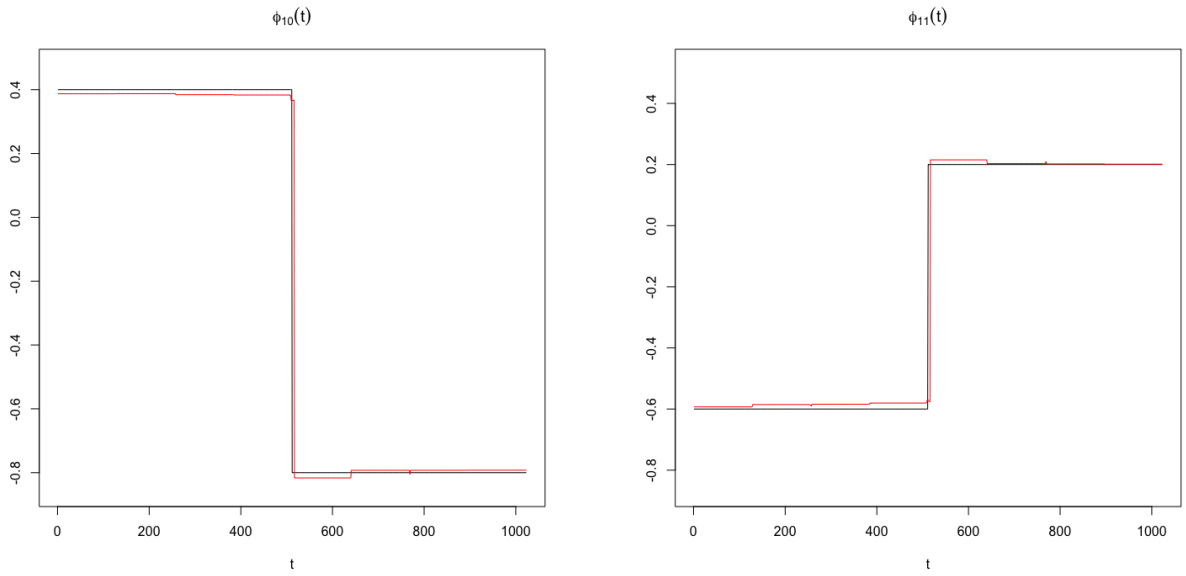


Figure 3.5: Comparison of the true parameters (in black) versus the averages of estimates obtained by Haar wavelet (in red) of Group 1.

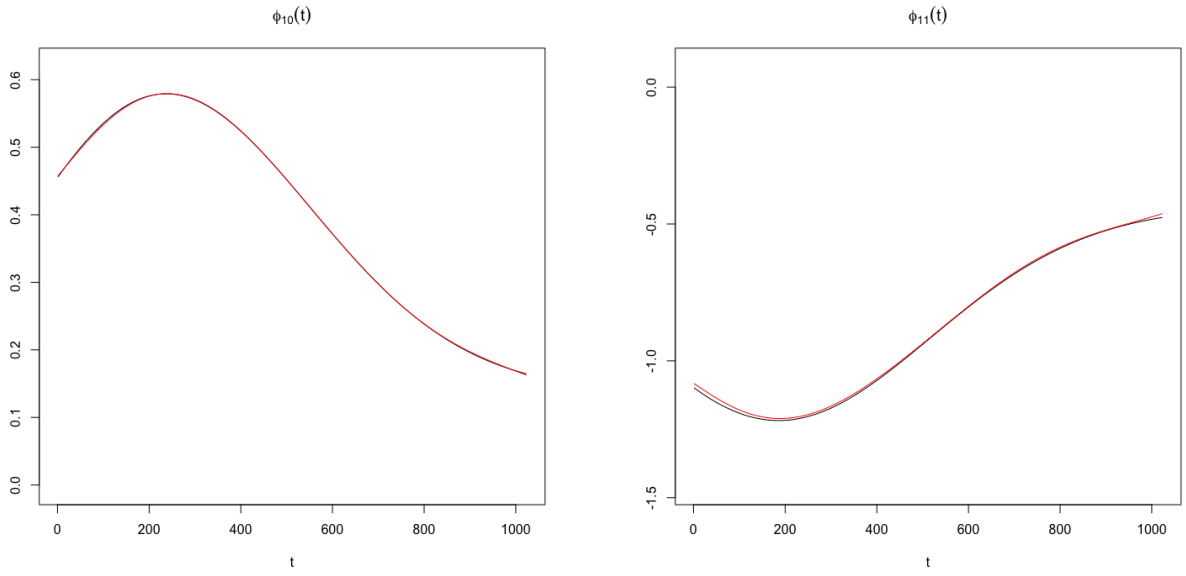


Figure 3.6: Comparison of the true parameters (in black) versus the averages of estimates obtained by Mexican hat wavelet (in red) of Group 2.

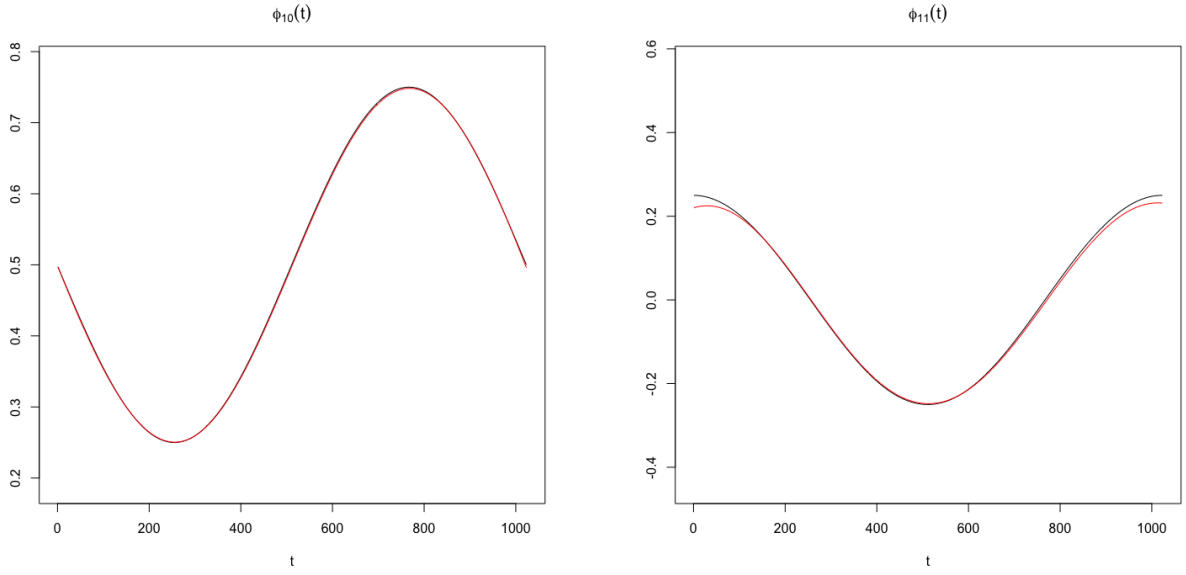


Figure 3.7: Comparison of the true parameters (in black) versus the averages of estimates obtained by Mexican hat wavelet (in red) of Group 3.

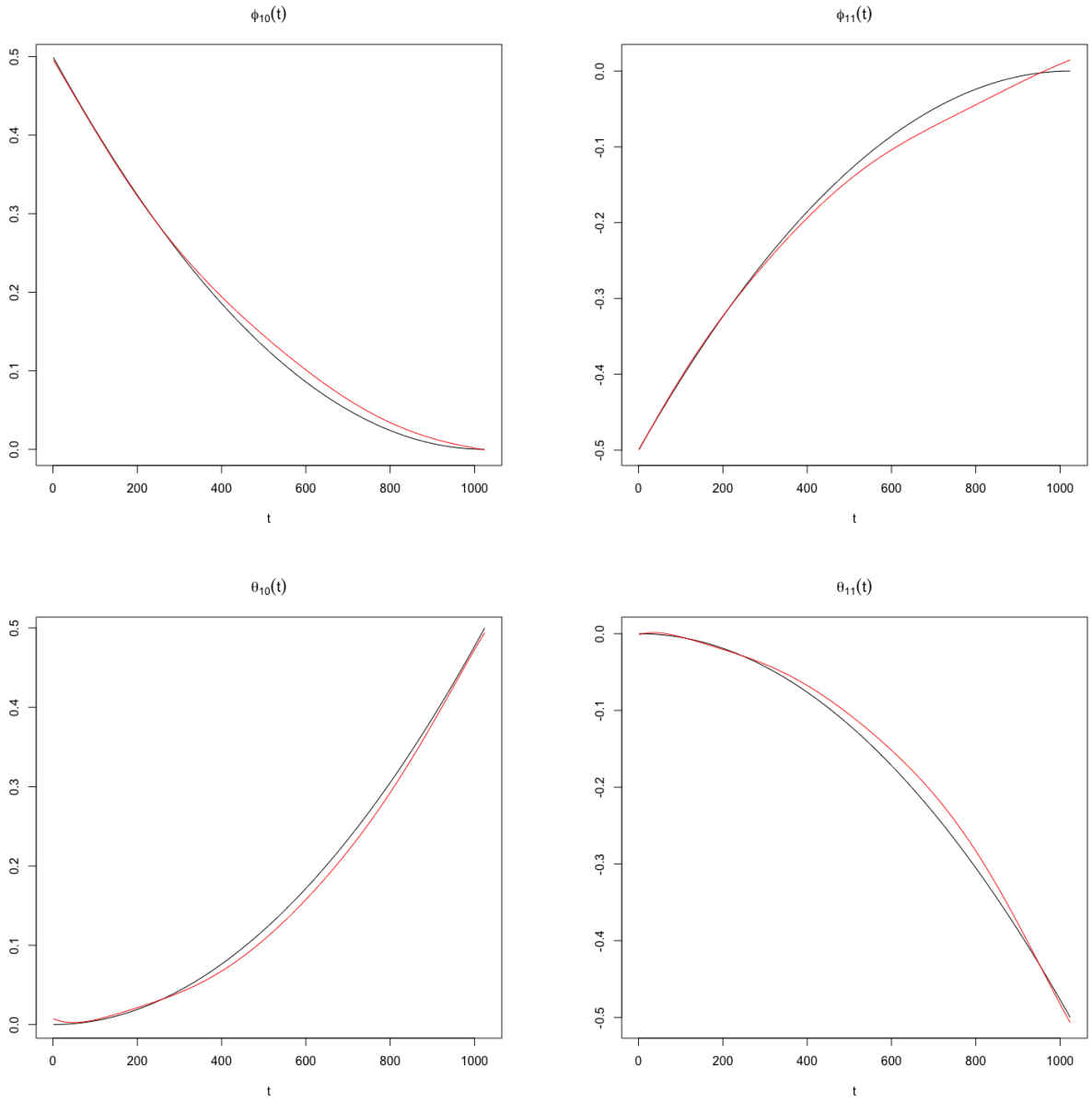


Figure 3.8: Comparison of the true parameters (in black) versus the averages of estimates obtained by Mexican hat wavelet (in red) of Group 4.

3.4 Comparisons of Spatial Weight Matrix via Simulation

Inspired by Jin (2017), this section presents some simulation examples in order to compare different types of spatial weights matrices.

3.4.1 Spatial Weight Matrix

As mentioned in Section 2.2.2, the main characteristic of the STARMA model is the spatial weight matrix,

$$W^{(l)} = (w_{ij}^{(l)}),\ i, j = 1, ..., n, \tag{3.4.1}$$

where

$$\sum_{j=1}^{n} w_{ij}^{(l)} = 1 \tag{3.4.2}$$

and l is the spatial order. Each element of the matrix reflects the spatial relationship between two regions, $\mathbf{x}_i$ and $\mathbf{x}_j$ and $w_{ij}^{(l)} = 0$ when $i = j$, that is, the matrix has zeros on its main diagonal and the other elements will consist of positive numbers.

In the tvSTARMA models, let $\mathbf{Z}(t) = [Z_1(t), ..., Z_n(t)]$,

$$W^{(l)}\mathbf{Z}(t) = \begin{cases} \mathbf{I}_n\mathbf{Z}(t), & \text{if } l = 0, \\ L^{(l)}\mathbf{Z}(t), & \text{if } l > 0. \end{cases} \tag{3.4.3}$$

The spatial lag operator $L^{(l)}$ is defined by

$$L^{(l)}Z_i(t) = \begin{cases} Z_i(t), & \text{if } l = 0, \\ \sum_{j=1}^{n} w_{ij}^{(l)} Z_j(t), & \text{if } l > 0. \end{cases} \tag{3.4.4}$$

There are many choices to the matrix W, such as inverse distance, negative exponential model, k-nearest neighbors, etc. The distance inverse weight and negative exponential weight are selected in our simulations. They are defined by

$$w_{ij}^{di} = \begin{cases} \frac{d_{ij}^{-\alpha}}{\sum_{k \neq i} d_{ik}^{-\alpha}}, & i \neq j, \\ 0, & i = j \end{cases} \tag{3.4.5}$$

and

$$w_{ij}^{ne} = \begin{cases} \frac{\exp\{-\alpha d_{ij}\}}{\sum_{k \neq i} \exp\{-\alpha d_{ik}\}}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{3.4.6}$$

where d_{ij} is great-circle distance and the changing rate of weights over distance d_{ij} is determined by the parameter α, a positive number.

3.4.2 Simulation Procedure

The simulations consist of the following steps:

[1.] Let $n = 15$ sample locations generated (the same as Figure 3.1), and then, $M = 500$ experiments of n time series with length $T = 1024$ are simulated.

[2.] The sample data $\mathbf{z} = [\mathbf{z}(\frac{1}{T}), ..., \mathbf{z}(\frac{T}{T})]$ are simulated from a Gaussian random field with mean function $\mu(\mathbf{z}) = \mathbf{0}$ and covariance function

$$\text{Cov}(z_{\mathbf{x}+\mathbf{h}}(t+\tau), z_{\mathbf{x}}(t)) = \frac{\sigma^2}{\beta(|\tau|)^{d/2}}\omega\left\{\frac{||\mathbf{h}||}{\sqrt{\beta(|\tau|)}}\right\} + \sigma_\epsilon^2 \mathbb{I}_{(z_{\mathbf{x}+\mathbf{h}}(t+\tau)=z_{\mathbf{x}}(t))}, (\mathbf{h};\tau) \in R^d \times R, \quad (3.4.7)$$

where $\mathbf{h}$ and τ are spatial and temporal shifts, respectively, $\sigma^2 = \text{Var}(z_{\mathbf{x}}(t))$, $d = 2$ and $\sigma_\epsilon^2 = 0.05$; $\omega(\cdot) \geq 0$ is a completely monotone function and $\beta(\cdot)$ is a positive function with a completely monotone derivative. And then,

$$\beta(f) = (f^\zeta + 1)^{\delta/\zeta}, \quad (3.4.8)$$

$$\omega(f) = \exp\{-f/\gamma\}, \quad (3.4.9)$$

where $\zeta = 1$, $\delta = 0.5, 1, 1.5$ and $\gamma = 0.25, 0.5, 1$ are selected.

[3.] Fit tvSTAR(1_1) and tvSTARMA($1_1, 1_1$) models using the estimation procedure described in Sections 3.1.2 and 3.2.2 with different spatial weight matrices described in Section 3.4.1, we set $\alpha = 0.5$, 1 for distance inverse weight and $\alpha = 0.5$, 1, 2 for negative exponential weight.

[4.] Calculate mean squared error (MSE) of the predictors $\hat{\mathbf{z}}$,

$$\text{MSE} = \frac{1}{MnT}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{t=1}^{T}[z_i^m\left(\frac{t}{T}\right) - \hat{z}_i^m\left(\frac{t}{T}\right)]^2, \quad (3.4.10)$$

where $\hat{z}_i^m\left(\frac{t}{T}\right)$ is the estimator of $z_i^m\left(\frac{t}{T}\right)$, the observation at location i and time t of the mth experiment.

3.4.3 Results

From tables 3.1-3.6, we can say that under same spatial weight matrix, Haar wavelet is better and MSEs decrease as J increases. The MSEs of tvSTARMA($1_1, 1_1$) models are smaller for almost every fitted model. For both Haar and Mexican hat wavelets, the MSEs are better when the distance inverse weight with $\alpha = 1$ was used and the negative exponential weight with $\alpha = 0.5$ always obtains the largest MSE.

Note that only in the case tvSTAR(1_1) model with Haar wavelet, tvSTAR(1_1) model with Mexican hat wavelet and $J = 2$ and tvSTARMA($1_1, 1_1$) model with Haar wavelet and $J = 2$, the negative

exponential weight with $\alpha = 2$ has the lowest MSE and it presents the second best MSE in other cases.

Figures B.11-B.37 of Appendix B show the boxplots of MSEs of different simulated datasets obtained by several models. Similar as the results of the tables, the effects of spatial weights matrices don't change in models used, that is, for both tvSTAR(1_1) and tvSTARMA($1_1, 1_1$) models, Haar and Mexican hat wavelets, $J = 2, 3$, MSEs of the models with the distance inverse weight with $\alpha = 1$ are always better. Note that there are some outliers when tvSTARMA($1_1, 1_1$) models with Mexican hat wavelet were used with the negative exponential weight.

Table 3.1: MSEs of different datasets from fitted tvSTAR(1_1) model with Haar wavelet and $J = 2$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5718754	0.5718671	0.5718751	0.5718726	**0.5718640**
$\gamma = 0.25$ $\delta = 1$	0.8098079	0.8097984	0.8098102	0.8098068	**0.8098007**
$\gamma = 0.25$ $\delta = 1.5$	0.9288534	**0.9288364**	0.9288597	0.9288551	0.9288474
$\gamma = 0.5$ $\delta = 0.5$	0.5718170	**0.5717461**	0.5718260	0.5718180	0.5717843
$\gamma = 0.5$ $\delta = 1$	0.8096359	**0.8094605**	0.8096958	0.8096638	0.8095814
$\gamma = 0.5$ $\delta = 1.5$	0.9285955	**0.9283395**	0.9286984	0.9286436	0.9285171
$\gamma = 1$ $\delta = 0.5$	0.5711949	**0.5707646**	0.5713536	0.5712403	0.5709691
$\gamma = 1$ $\delta = 1$	0.8082413	**0.8073662**	0.8086525	0.8083505	0.8077541
$\gamma = 1$ $\delta = 1.5$	0.9266655	**0.9256044**	0.9272141	0.9267818	0.9260058

Table 3.2: MSEs of different datasets from fitted tvSTAR(1_1) model with Haar wavelet and $J = 3$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5707591	0.5707378	0.5707555	0.5707495	**0.5707273**
$\gamma = 0.25$ $\delta = 1$	0.8086490	0.8086369	0.8086449	0.8086422	**0.8086318**
$\gamma = 0.25$ $\delta = 1.5$	0.9277130	**0.9276963**	0.9277128	0.9277093	0.9277001
$\gamma = 0.5$ $\delta = 0.5$	0.5707055	**0.5706292**	0.5706956	0.5706897	0.5706473
$\gamma = 0.5$ $\delta = 1$	0.8084639	**0.8082959**	0.8085023	0.8084771	0.8083968
$\gamma = 0.5$ $\delta = 1.5$	0.9274310	**0.9271860**	0.9275153	0.9274675	0.9273463
$\gamma = 1$ $\delta = 0.5$	0.5700613	**0.5696411**	0.5701736	0.5700780	0.5698094
$\gamma = 1$ $\delta = 1$	0.8069982	**0.8061549**	0.8073586	0.8070843	0.8065077
$\gamma = 1$ $\delta = 1.5$	0.9254121	**0.9243868**	0.9259187	0.9255126	0.9247610

Table 3.3: MSEs of different datasets from fitted tvSTAR(1_1) model with Mexican hat wavelet and $J = 2$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5720284	0.5720198	0.5720292	0.5720263	**0.5720187**
$\gamma = 0.25$ $\delta = 1$	0.8099595	**0.8099531**	0.8099612	0.8099605	0.8099591
$\gamma = 0.25$ $\delta = 1.5$	0.9289999	**0.9289853**	0.9290054	0.9290036	0.9290000
$\gamma = 0.5$ $\delta = 0.5$	0.5719711	**0.5718976**	0.5719838	0.5719739	0.5719395
$\gamma = 0.5$ $\delta = 1$	0.8097927	**0.8096183**	0.8098543	0.8098237	0.8097445
$\gamma = 0.5$ $\delta = 1.5$	0.9287466	**0.9284917**	0.9288502	0.9287975	0.9286744
$\gamma = 1$ $\delta = 0.5$	0.5713549	**0.5709173**	0.5715235	0.5714033	0.5711274
$\gamma = 1$ $\delta = 1$	0.8084149	**0.8075349**	0.8088334	0.8085280	0.8079301
$\gamma = 1$ $\delta = 1.5$	0.9267936	**0.9256779**	0.9272953	0.9268617	0.9260850

Table 3.4: MSEs of different datasets from fitted tvSTAR(1_1) model with Mexican hat wavelet and $J = 3$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5713301	0.5713054	0.5713330	0.5713241	**0.5713024**
$\gamma = 0.25$ $\delta = 1$	0.8092726	**0.8092524**	0.8092760	0.8092697	0.8092562
$\gamma = 0.25$ $\delta = 1.5$	0.9283331	**0.9283079**	0.9283398	0.9283333	0.9283203
$\gamma = 0.5$ $\delta = 0.5$	0.5712721	**0.5711897**	0.5712766	0.5712649	0.5712213
$\gamma = 0.5$ $\delta = 1$	0.8090920	**0.8089110**	0.8091462	0.8091135	0.8090265
$\gamma = 0.5$ $\delta = 1.5$	0.9280607	**0.9278012**	0.9281587	0.9281040	0.9279749
$\gamma = 1$ $\delta = 0.5$	0.5706313	**0.5702010**	0.5707714	0.5706620	0.5703890
$\gamma = 1$ $\delta = 1$	0.8076571	**0.8067883**	0.8080496	0.8077554	0.8071624
$\gamma = 1$ $\delta = 1.5$	0.9259958	**0.9249419**	0.9265289	0.9261045	0.9253332

Table 3.5: MSEs of different datasets from fitted tvSTARMA($1_1, 1_1$) model with Haar wavelet and $J = 2$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5490047	0.5489930	0.5490053	0.5490008	**0.5489899**
$\gamma = 0.25$ $\delta = 1$	0.7955578	**0.7955132**	0.7955640	0.7955502	0.7955169
$\gamma = 0.25$ $\delta = 1.5$	0.9214984	**0.9214105**	0.9215132	0.9214881	0.9214192
$\gamma = 0.5$ $\delta = 0.5$	0.5489881	**0.5489317**	0.5489881	0.5489827	0.5489529
$\gamma = 0.5$ $\delta = 1$	0.7954177	**0.7952247**	0.7954742	0.7954330	0.7953230
$\gamma = 0.5$ $\delta = 1.5$	0.9212156	**0.9208820**	0.9213314	0.9212477	0.9210446
$\gamma = 1$ $\delta = 0.5$	0.5486359	**0.5483734**	0.5487187	0.5486533	0.5484805
$\gamma = 1$ $\delta = 1$	0.7942841	**0.7935768**	0.7945893	0.7943409	0.7938358
$\gamma = 1$ $\delta = 1.5$	0.9193444	**0.9183398**	0.9198217	0.9194045	0.9186393

Table 3.6: MSEs of different datasets from fitted tvSTARMA$(1_1, 1_1)$ model with Mexican hat wavelet and $J = 2$.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$\gamma = 0.25$ $\delta = 0.5$	0.5493941	**0.5493536**	0.5494022	0.5493865	0.5493567
$\gamma = 0.25$ $\delta = 1$	0.7960634	**0.7960057**	0.7960567	0.7960479	0.7960121
$\gamma = 0.25$ $\delta = 1.5$	0.9221163	**0.9220534**	0.9221938	0.9223765	0.9221010
$\gamma = 0.5$ $\delta = 0.5$	0.5493746	**0.5492904**	0.5493905	0.5493671	0.5493187
$\gamma = 0.5$ $\delta = 1$	0.7960146	**0.7957764**	0.7963313	0.7960042	0.7958850
$\gamma = 0.5$ $\delta = 1.5$	0.9217563	**0.9214193**	0.9218807	0.9217912	0.9215954
$\gamma = 1$ $\delta = 0.5$	0.5490224	**0.5487275**	0.5491266	0.5490394	0.5488448
$\gamma = 1$ $\delta = 1$	0.7948874	**0.7941313**	0.7952156	0.7949564	0.7944082
$\gamma = 1$ $\delta = 1.5$	0.9200001	**0.9189014**	0.9204669	0.9200260	0.9192314

3.5 Application

Precipitation is one of the most important variables for climate and hydro-meteorology. All plants need at least some water to survive, therefore precipitation, especially rain, is an important guiding standard for agricultural production. The annual amount of precipitation can be conducted as an important index degree for drought which presents more differences. There are many studies of precipitation in different regions using time series models, such as Dalezios and Adamowski (1995), Wang et al. (2013) and Wu et al. (2021). In this paper, we use historical daily precipitation records, the data can be obtained directly from GHCN (Global Historical Climatology Network)-Daily, an integrated public database of NOAA (National Oceanic and Atmospheric Administration) using R package **rnoaa**.

The data selected are daily precipitation records (in tenths of millimeters) from Midwestern states of the USA. The region consists of 11 states: North Dakota, South Dakota, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, Ohio and Wisconsin. We use only climate monitoring stations which contains no missing data during the period between 1990-01-01 and 2009-12-30 (inclusive). The 30 stations selected are showed in Figure 3.9 and Figure 3.10 shown precipitation recorded of the 2 weather stations.

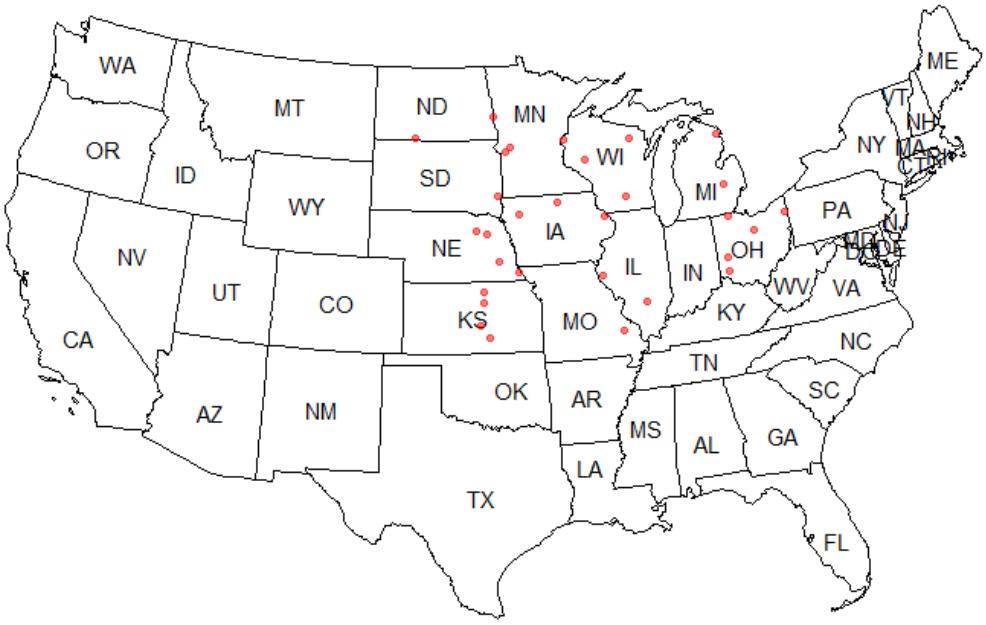


Figure 3.9: Locations of the stations selected from Midwestern states of the USA.

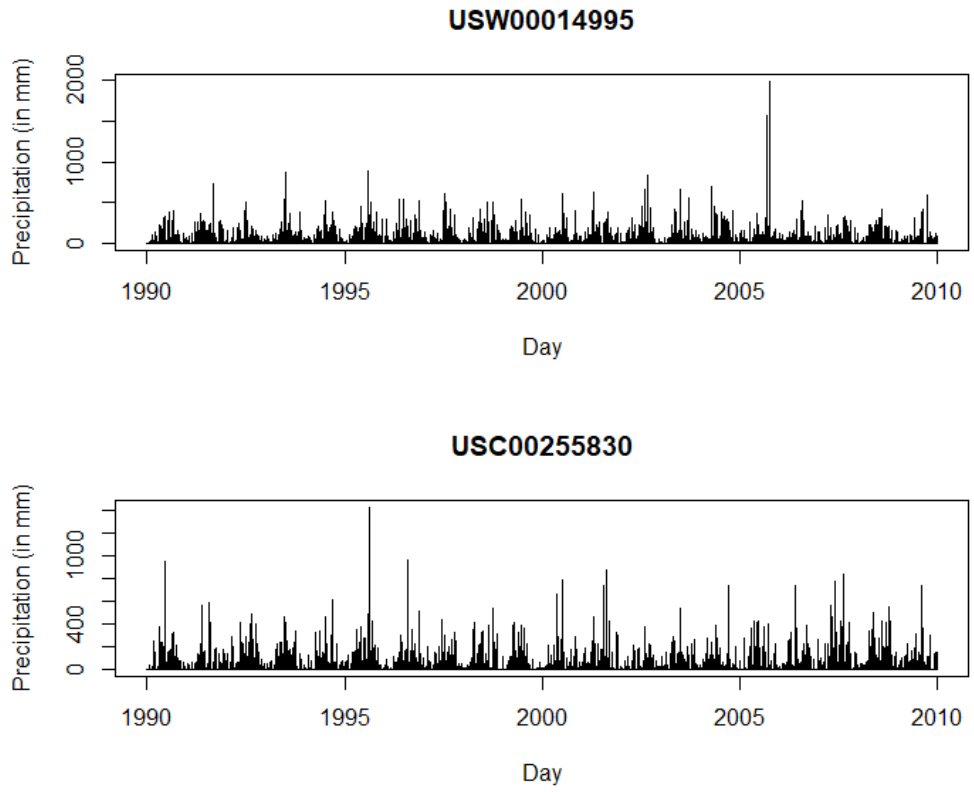


Figure 3.10: Precipitation recorded of 2 weather stations.

As precipitation data may have heavy tail distribution due to very large values with finite probabilities, we apply the transformation

$$Z_i\left(\frac{t}{T}\right) = \log_{10}\left(Y_i\left(\frac{t}{T}\right) + 1\right), \tag{3.5.1}$$

where $Y_i\left(\frac{t}{T}\right)$ is the original records at location $\mathbf{x}_i$ and time t before applying the estimation proposed to remove the effects.

To define the weights, the distance inverse weight and negative exponential weight are selected in this application. The weights are defined as (3.4.5) and (3.4.6). In this application, we set $\alpha = 0.5, 1$ and 0.5, 1, 2 for distance inverse weight and negative exponential weight, respectively.

The results were obtained from the estimation described in Sections 3.1.2 and 3.2.2 using tvSTAR(1_1) and tvSTARMA($1_1, 1_1$) models. The wavelet expansion of time-varying parameters is built using Haar and Mexican hat wavelets with $J = 2, 3, 4$ (tvSTAR models only).

Tables 3.7 and 3.8 show the MSEs of the estimates of different models with different selected wavelets. We can see that MSEs are better for both Haar and Mexican hat wavelets when the spatial weight matrix is distance inverse weight with $\alpha = 1$. Under same weight matrix, MSEs decrease as the number of J increases, however the MSEs are similar, therefore $J = 2$ may be satisfactory to fit the data since it has less coefficients. Note that Mexican hat is better only in case tvSTAR(1_1) and tvSTARMA($1_1, 1_1$) with $J = 2$.

Table 3.7: MSEs of different models where Haar wavelet were used.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
tvSTAR(1_1) $J = 2$	0.54997	**0.54881**	0.55414	0.55057	0.54972
tvSTAR(1_1) $J = 3$	0.54974	**0.54855**	0.55392	0.55036	0.54949
tvSTAR(1_1) $J = 4$	0.54938	**0.54822**	0.55354	0.54999	0.54913
tvSTARMA$(1_1, 1_1)$ $J = 2$	0.54905	**0.54809**	0.55348	0.54967	0.54878
tvSTARMA$(1_1, 1_1)$ $J = 3$	0.54847	**0.54757**	0.55288	0.54907	0.54824

Table 3.8: MSEs of different models where Mexican hat wavelet were used.

Spatial weight matrix	w_{ij}^{di}		w_{ij}^{ne}		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
tvSTAR(1_1) $J = 2$	0.54996	**0.54880**	0.55414	0.55057	0.54972
tvSTAR(1_1) $J = 3$	0.54983	**0.54866**	0.55401	0.55044	0.54959
tvSTAR(1_1) $J = 4$	0.54966	**0.54851**	0.55383	0.55028	0.54944
tvSTARMA$(1_1, 1_1)$ $J = 2$	0.54909	**0.54816**	0.55355	0.54972	0.54886
tvSTARMA$(1_1, 1_1)$ $J = 3$	0.54901	**0.54789**	0.55349	0.54963	0.54861

Chapter 4

Deformation Based on Monotonic Functions

4.1 Introduction

By the Theorem 2.4.1, instead of estimating the constrained function g, the problem becomes computing the unconstrained function w. Since w can be positive or negative, we expand it as a linear combination of a set of wavelet basis functions.

As previously mentioned, two types of wavelets will be used:

- Mexican hat, given by

$$\psi^{Mex}(x) = \frac{2}{\sqrt{3}\pi^{1/4}}(1-x^2)e^{-x^2/2}. \tag{4.1.1}$$

  Then we can expand ω as

$$w_{Mex}(x) = \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c_{j,k}\psi_{j,k}^{Mex}(x). \tag{4.1.2}$$

- Shannon wavelet, given by

$$\psi^{Shan}(x) = \text{sinc}(\frac{x}{2})\cos\left(\frac{3\pi x}{2}\right) = 2\text{sinc}(2x) - \text{sinc}(x), \tag{4.1.3}$$

$$\phi^{Shan}(x) = \text{sinc}(x), \tag{4.1.4}$$

  where $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$. And ω can be expressed by

$$w_{Shan}(x) = c_0\phi^{Shan}(x) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c_{j,k}\psi_{j,k}^{Shan}(x). \tag{4.1.5}$$

4.2 Deformation Based on Monotonic Functions

Let $\mathbf{x}_i = (x_{i1}, x_{i2})$, location i in a G plane and $\mathbf{y}_i = (y_{i1}, y_{i2})$, its deformed location in a D plane, $i = 1, ..., n$. As there is no natural sort order in $\mathbb{R}^2$, it's very difficult to get a bidimensional

monotonic function. Adapting the generalized additive model introduced by Hastie e Tibshirani (1990) to estimate the deformation, we can consider each coordinate of the representation of the D plane as the response variable and the coordinates of the location in the G plane as predictor variables and thus we have an additive model as

$$y_{il} = \beta_0 + \sum_{j=1}^{2} g_j^l(x_{ij}) + \epsilon_{il},\ l = 1, 2, \tag{4.2.1}$$

where β_0 represent intercept and ϵ_{il} indicates random effect.

Suppose that $\beta_0 = 0$ and random effect is null. Therefore, the representation in the D plane, $\mathbf{y}_i = (y_{i1}, y_{i2})$, can be written as

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} = \begin{bmatrix} g_1^1(x_{i1}) + g_2^1(x_{i2}) \\ g_1^2(x_{i1}) + g_2^2(x_{i2}) \end{bmatrix}, \tag{4.2.2}$$

where $g_1^l(x_{i1})$ and $g_2^l(x_{i2})$ are monotonic functions given by (2.4.11). Since $g_1^l(x_{i1})$ and $g_2^l(x_{i2})$ are strictly monotonic on the range $[0, \infty)$, y_{ik} is also strictly monotonic in this range. Then, y_{il} is a injective function. And we can write

$$y_{il} = C_{10}^l + C_{11}^l D^{-1} \exp\{D^{-1}\omega^{l1}(x_{i1})\} + C_{20}^l + C_{21}^l D^{-1} \exp\{D^{-1}\omega^{l2}(x_{i2})\}. \tag{4.2.3}$$

According to Theorem 2.4.1, C_0 and C_1 are arbitrary constants, suppose that $C_{10}^l = C_{20}^l = 0$ and $C_{11}^l = C_{21}^l = 1$. Therefore, (4.2.3) can be simplified to

$$y_{il} = D^{-1} \exp\{D^{-1}\omega^{l1}(x_{i1})\} + D^{-1} \exp\{D^{-1}\omega^{l2}(x_{i2})\}. \tag{4.2.4}$$

Note that $\omega^{l1}(x_{i1})$ and $\omega^{l2}(x_{i2})$ can be written in the form (4.1.2) or (4.1.5). Thus, the estimated deformations using Mexican hat and Shannon wavelets can be written as

$$\hat{y}_{il}^{Mex} = D^{-1} \exp\{\omega_{Mex}^{k1}(x_{i1})\} + D^{-1} \exp\{\omega_{Mex}^{l2}(x_{i2})\}, \tag{4.2.5}$$

$$\hat{y}_{il}^{Shan} = D^{-1} \exp\{\omega_{Shan}^{k1}(x_{i1})\} + D^{-1} \exp\{\omega_{Shan}^{l2}(x_{i2})\}, \tag{4.2.6}$$

for $i = 1, ..., n,\ l = 1, 2$.

4.3 Process Optimization

Let $\mathbf{x}_i = (x_{i1}, x_{i2}) \in G \subset \mathbb{R}^2$ coordinates of location i in a G plane and suppose that a collection of n locations $[(x_{11}, x_{12}), ..., (x_{n1}, x_{n2})]$ was obtained. The optimization procedure is used to estimate the deformation $\mathbf{y}$, the spatial variance $\nu = \nu(\mathbf{x})$ and the parameters $\boldsymbol{\theta}$ of the correlation function ρ, maximizing the likelihood function of the samples $z_{it} = Z(\mathbf{x}_i, t)$, $i = 1, .., n$, $t = 1, .., T$.

Assuming $\mathbf{z}_t = (z_{1t}, ..., z_{nt})' \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma_{\boldsymbol{\eta}})$, the likelihood function is

$$L(\boldsymbol{\mu}, \Sigma_{\boldsymbol{\eta}}|\mathbf{z}) \quad = \quad (2\pi)^{-nT/2} \times \det(\Sigma_{\boldsymbol{\eta}})^{-T/2} \times \exp\{-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{z}_t - \boldsymbol{\mu})^T \Sigma_{\boldsymbol{\eta}}^{-1}(\mathbf{z}_t - \boldsymbol{\mu})\}$$

$$= (2\pi)^{-nT/2} \times \det(\Sigma_{\boldsymbol{\eta}})^{-T/2} \times \exp\{-\frac{1}{2}\mathrm{tr}[\Sigma_{\boldsymbol{\eta}}^{-1}(\sum_{t=1}^{T}(\mathbf{z}_t - \bar{\mathbf{z}})(\mathbf{z}_t - \bar{\mathbf{z}})^T + T(\bar{\mathbf{z}} - \boldsymbol{\mu})(\bar{\mathbf{z}} - \boldsymbol{\mu})^T)]\}$$

$$= (2\pi)^{-nT/2} \times \det(\Sigma_{\boldsymbol{\eta}})^{-T/2} \times \exp\{-\frac{T}{2}\mathrm{tr}[\Sigma_{\boldsymbol{\eta}}^{-1}S] - \frac{T}{2}(\bar{\mathbf{z}} - \boldsymbol{\mu})^T\Sigma_{\boldsymbol{\eta}}^{-1}(\bar{\mathbf{z}} - \boldsymbol{\mu})\}, \quad (4.3.1)$$

where $\bar{\mathbf{z}}$ is the vector of means at each location, S is sample spatial covariance matrix whose element $S_{ij} = \sum_{t=1}^{T}(z_{it} - \bar{z}_{i.})(z_{jt} - \bar{z}_{j.})$ and the covariance matrix

$$\Sigma_{\boldsymbol{\eta}=(\mathbf{c},\nu,\boldsymbol{\theta})} = (\sigma_{ij}), \quad (4.3.2)$$

where

$$\sigma_{ij} = \sqrt{v_i v_j}\rho_{\boldsymbol{\theta}}(|\mathbf{y}_i - \mathbf{y}_j|), \quad (4.3.3)$$

with the coordinates $\mathbf{y}_i$ and $\mathbf{y}_j$ of the representation of the D plane, the spatial variance ν_i and ν_j at locations i and j, respectively, and the parameters $\boldsymbol{\theta}$ of the correlation function ρ.

Without loss of generality, set $\boldsymbol{\mu} = \mathbf{0}$, the optimization problem becomes

$$\max_{\boldsymbol{\eta}} L(\Sigma_{\boldsymbol{\eta}}|\mathbf{z}). \quad (4.3.4)$$

The optimization procedure to estimate the covariance matrix (4.3.2) consists of the following steps:

1. Let $\mathbf{c}^0$ be the initial values of $\mathbf{c}$, coefficients of ω and calculate $\boldsymbol{\gamma}^0 = (\nu^0, \boldsymbol{\theta}^0)$ such that

$$\boldsymbol{\gamma}^0 = \arg\max\, l(\Sigma_{\nu,\boldsymbol{\theta}}|\mathbf{z}, \mathbf{c}^0), \quad (4.3.5)$$

   where $l(\Sigma_{\nu,\boldsymbol{\theta}}|\mathbf{z}, \mathbf{c}^0) = \log L(\Sigma_{\nu,\boldsymbol{\theta}}|\mathbf{z}, \mathbf{c}^0)$.

2. Given $\boldsymbol{\gamma}^0$ obtained in Step 1, calculate

$$\mathbf{c}^1 = \arg\max\, l(\mathbf{c}|\mathbf{z}, \Sigma_{\nu^0,\boldsymbol{\theta}^0}), \quad (4.3.6)$$

   where $l(\mathbf{c}|\mathbf{z}, \Sigma_{\nu^0,\boldsymbol{\theta}^0}) = \log L(\mathbf{c}|\mathbf{z}, \Sigma_{\nu^0,\boldsymbol{\theta}^0})$.

3. Replace $\mathbf{c}^0$ by $\mathbf{c}^1$, return to Step 1.

4. Repeat Step (2) - (3) until convergence.

4.4 Simulations

This section presents some simulations in order to assess the performance of the algorithm. These simulations are carried out in the cases that the deformations are generated by functions linear, quadratic, non-linear and wavelet.

4.4.1 Formulation of Deformation

The deformed coordinates $\mathbf{y}_i$'s, i.e. representations of D plane at location i, $i = 1, .., n$, are generated as follows:

- Linear case:

$$y_{i1} = 0.75x_{i1} + x_{i2}, \tag{4.4.1}$$

$$y_{i2} = x_{i1} + 0.25x_{i2}. \tag{4.4.2}$$

- Quadratic case:

$$y_{i1} = -0.5(x_{i1} - 0.5)^2 + (x_{i2} - 0.5) + 0.6, \tag{4.4.3}$$

$$y_{i2} = (x_{i1} - 0.5) - 0.5(x_{i2} - 0.5)^2 + 0.6. \tag{4.4.4}$$

- Non-linear case:

$$y_{i1} = \cos(\text{angle})(x_{i1} - 0.5) + \sin(\text{angle})(x_{i2} - 0.5) + 0.5, \tag{4.4.5}$$

$$y_{i2} = -\sin(\text{angle})(x_{i1} - 0.5) + \cos(\text{angle})(x_{i2} - 0.5) + 0.5, \tag{4.4.6}$$

  where angle $= 2.5\exp\{-(x_{i1} - 0.5)^2 - (x_{i2} - 0.5)^2\} + 3\pi/2$.

- Wavelet case:

$$y_{i1} = D^{-1}\exp\{\sum_{j=0}^{1}\sum_{k=0}^{2^j-1} c_{j,k}^{11}\psi_{j,k}(x_{i1})\} + D^{-1}\exp\{\sum_{j=0}^{1}\sum_{k=0}^{2^j-1} c_{j,k}^{12}\psi_{j,k}(x_{i2})\}, \tag{4.4.7}$$

$$y_{i2} = D^{-1}\exp\{\sum_{j=0}^{1}\sum_{k=0}^{2^j-1} c_{j,k}^{21}\psi_{j,k}(x_{i1})\} + D^{-1}\exp\{\sum_{j=0}^{1}\sum_{k=0}^{2^j-1} c_{j,k}^{22}\psi_{j,k}(x_{i2})\}, \tag{4.4.8}$$

  where $\psi(t)$ are Mexican hat wavelets and $c_{0,0}^{11} = 0.25$, $c_{1,0}^{11} = 0.01$, $c_{1,1}^{11} = -0.036$, $c_{0,0}^{12} = -0.37$, $c_{1,0}^{12} = 0.065$, $c_{1,1}^{12} = -1.2$, $c_{0,0}^{21} = -0.032$, $c_{1,0}^{21} = -0.043$, $c_{1,1}^{21} = -1$, $c_{0,0}^{22} = -0.031$, $c_{1,0}^{22} = 0.11$, $c_{1,1}^{22} = 0.19$.

4.4.2 Simulation Procedure

The simulations consists of the following steps:

1. Let $n = 50$ sample locations generated in the geographical domain G $= [0, 1] \times [0, 1]$, the coordinates (longitude and latitude) are generated following a uniform dispersion in (0, 1). Figure 4.1 shows the sampling locations and the deformed locations are presented in Figure 4.2.

2. The sample data ($\mathbf{z} = (\mathbf{z}_1, ..., \mathbf{z}_T)$) are simulated from a Gaussian random field with mean function $\mu(\mathbf{z}) = \mathbf{0}$ and covariance function

$$\text{Cov}(\mathbf{z}_i, \mathbf{z}_j) = \nu\rho_{\boldsymbol{\theta}}(|\mathbf{y}_i - \mathbf{y}_j|) + \sigma_\epsilon^2 \mathbb{I}_{(\mathbf{z}_i = \mathbf{z}_j)} = \nu \exp\{-|\mathbf{y}_i - \mathbf{y}_j|/\theta\} + \sigma_\epsilon^2 \mathbb{I}_{(\mathbf{z}_i = \mathbf{z}_j)} \tag{4.4.9}$$

with parameters $\nu = 1,\ \theta = 0.25$ and $\sigma_\epsilon^2 = 0.05$. The length of each series was fixed at $T = 2048$.

3. Calculate the estimates of $\mathbf{y}_i$, $\hat{\mathbf{y}}_i = (\hat{y}_{i1}, \hat{y}_{i2})$ for $i = 1, ..., n = 50$ and the parameters of the covariance function using the optimization procedure described in Section 4.3 with Mexican hat and Shannon wavelets.

4. Calculate the MSE of the estimates of the correlation matrix,

$$\text{MSE} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\text{corr}_{ij} - \exp\{-|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j|/\hat{\theta}\})^2, \tag{4.4.10}$$

where corr_{ij} is the element of the ith row and jth column in the correlation matrix of $\mathbf{z}$, i.e. is the correlation between $\mathbf{z}_i$ and $\mathbf{z}_j$.

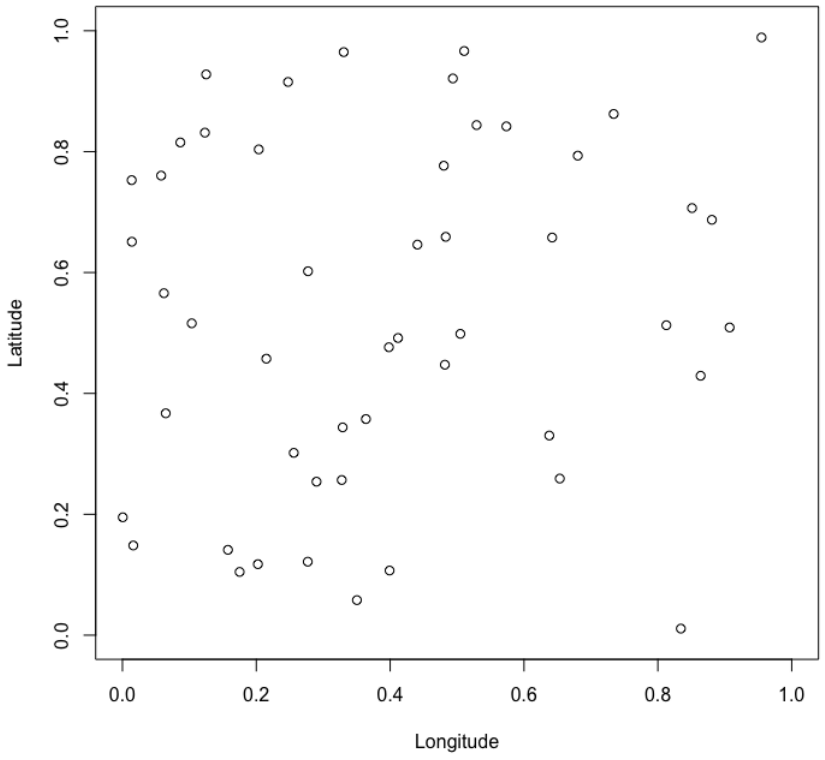
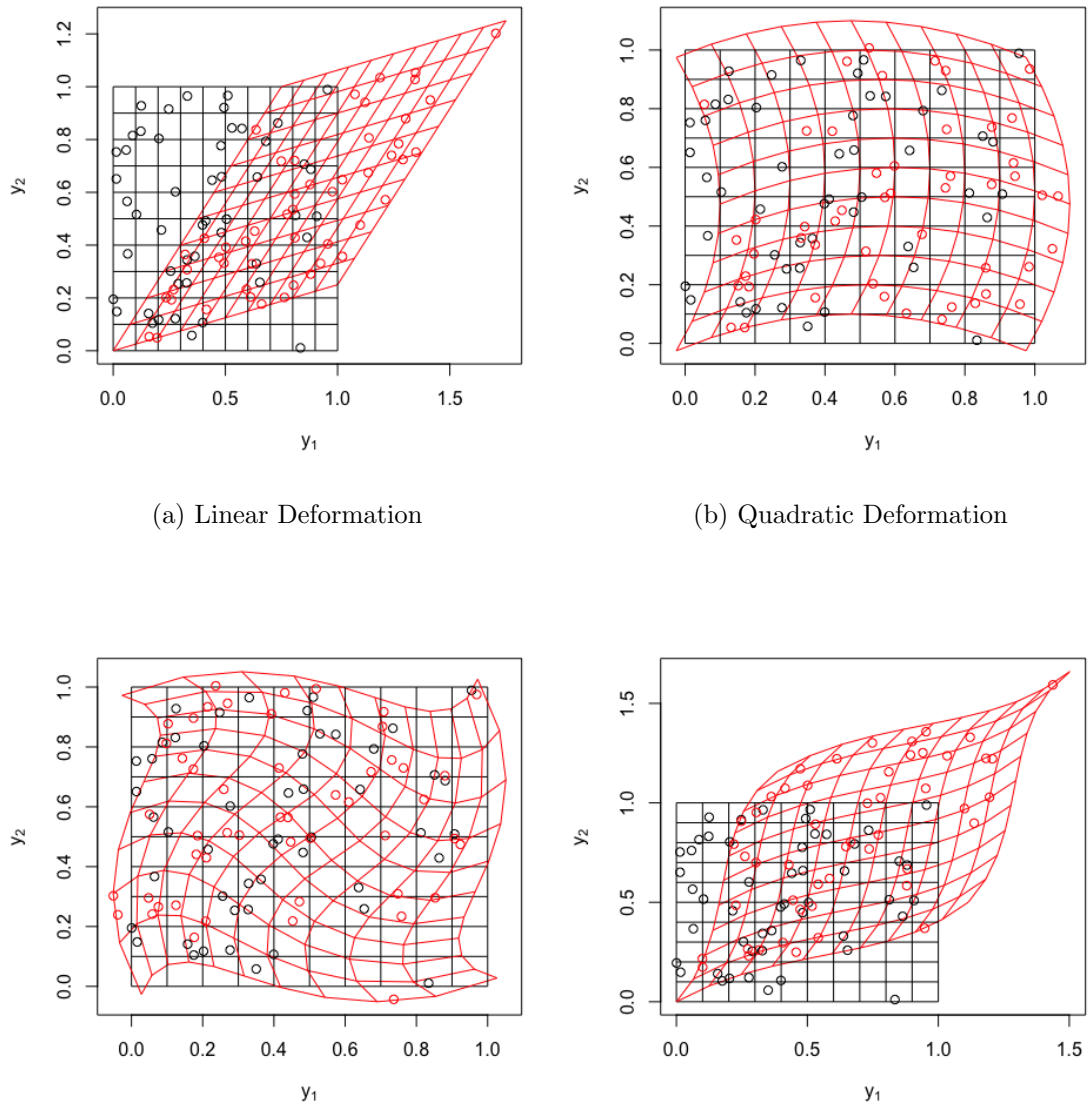


Figure 4.1: Sampling locations.



(a) Linear Deformation

(b) Quadratic Deformation

(c) Non-linear Deformation

(d) Wavelet Deformation

Figure 4.2: Sampling locations (and regular grid) in G plane, in black, and deformed locations (and deformed grid) in plane D, in red.

4.4.3 Results

Table 4.1 shows the estimated parameters of the covariance function and the MSEs of the estimates of the correlation matrix in several fits. Observe that in the linear, non-linear and wavelet case, the estimates that used $J = 3$ and Shannon wavelet were closer to the true values of the parameters, whereas in the quadratic case, they were closer when using Mexican hat wavelet. Note that in three cases the lowest MSE was obtained using the Mexican hat wavelet, however with $J = 2$ in the linear case and $J = 3$ in the non-linear and wavelet case. Only in the quadratic case the smallest MSE was obtained using Shannon wavelet and $J = 4$.

Figures 4.3-4.10 show the estimated deformation and the scatter plots of the upper-diagonal entries of the estimated correlation matrices for the sample data, versus the true correlation matrices. The comparison of estimated correlation matrices provides great convenience to evaluate the estimation results. According to the MSEs in Table 4.1, the scatter plot is more accurate for smaller MSE. Note that in the linear and wavelet case, the estimated deformations are very close to the true one.

Table 4.1: Estimated parameters of the covariance function and MSEs of the correlation matrix of different fits.

True value of the parameters: $\nu = 1 \quad \theta = 0.25 \quad \sigma_\epsilon^2 = 0.05$

	Mexican hat			Shannon		
	Linear Deformation					
	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$
ν	1.03540	1.04229	1.03757	1.03493	1.02579	1.03192
θ	0.19197	0.18178	0.20000	0.20204	0.23975	0.21140
σ_ϵ^2	0.04117	0.03905	0.03771	0.04380	0.06739	0.04512
MSE	0.00173	0.00505	0.00290	0.00196	0.00409	0.00708
	Quadratic Deformation					
	$J=3$	$J=4$	$J=5$	$J=3$	$J=4$	$J=5$
ν	1.01631	1.03535	1.01537	1.04410	1.04491	1.00799
θ	0.25529	0.26601	0.25003	0.26182	0.25878	0.25973
σ_ϵ^2	0.04272	0.04371	0.04082	0.04140	0.04032	0.04690
MSE	0.00309	0.00165	0.00195	0.00220	0.00124	0.00205
	Non-linear Deformation					
	$J=3$	$J=4$	$J=5$	$J=3$	$J=4$	$J=5$
ν	1.04400	1.04378	1.02342	1.02903	0.99906	1.04291
θ	0.22920	0.21801	0.23132	0.25214	0.24048	0.23826
σ_ϵ^2	0.04117	0.03905	0.03771	0.04183	0.03259	0.03297
MSE	0.00387	0.00539	0.00485	0.00409	0.00436	0.00451
	Wavelet Deformation					
	$J=3$	$J=4$	$J=5$	$J=3$	$J=4$	$J=5$
ν	1.00827	1.02275	1.01732	1.02485	1.00019	1.00496
θ	0.19449	0.19605	0.20986	0.22260	0.20877	0.21827
σ_ϵ^2	0.05629	0.04681	0.05375	0.06022	0.06107	0.06728
MSE	0.00179	0.00279	0.00218	0.00296	0.00222	0.00226

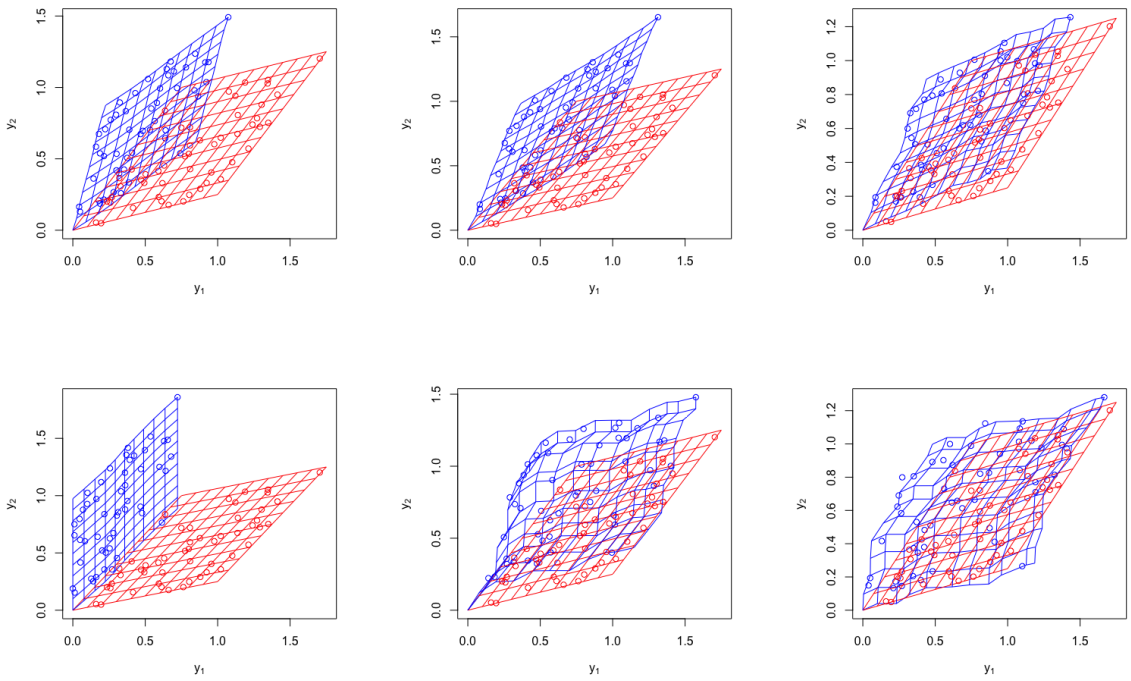


Figure 4.3: Estimated deformation (in blue) when $J = 2, 3, 4$ (from left to right) for linear deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

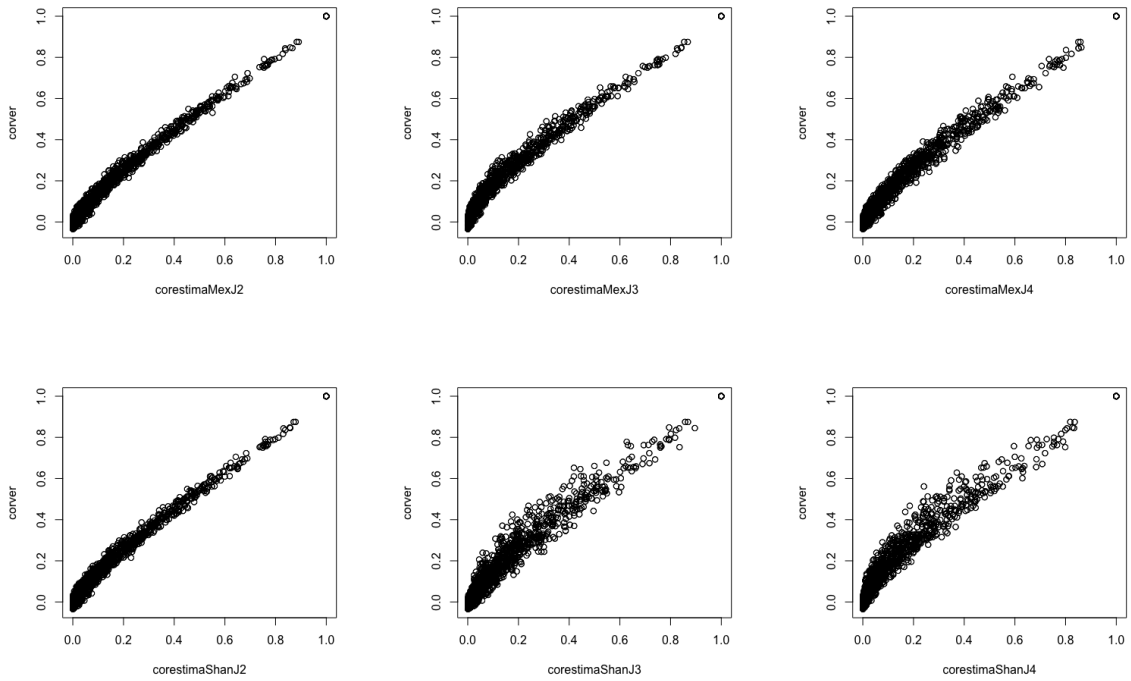


Figure 4.4: Comparison of the estimated correlation matrix versus the true correlation matrix when $J = 2, 3, 4$ (from left to right) for linear deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

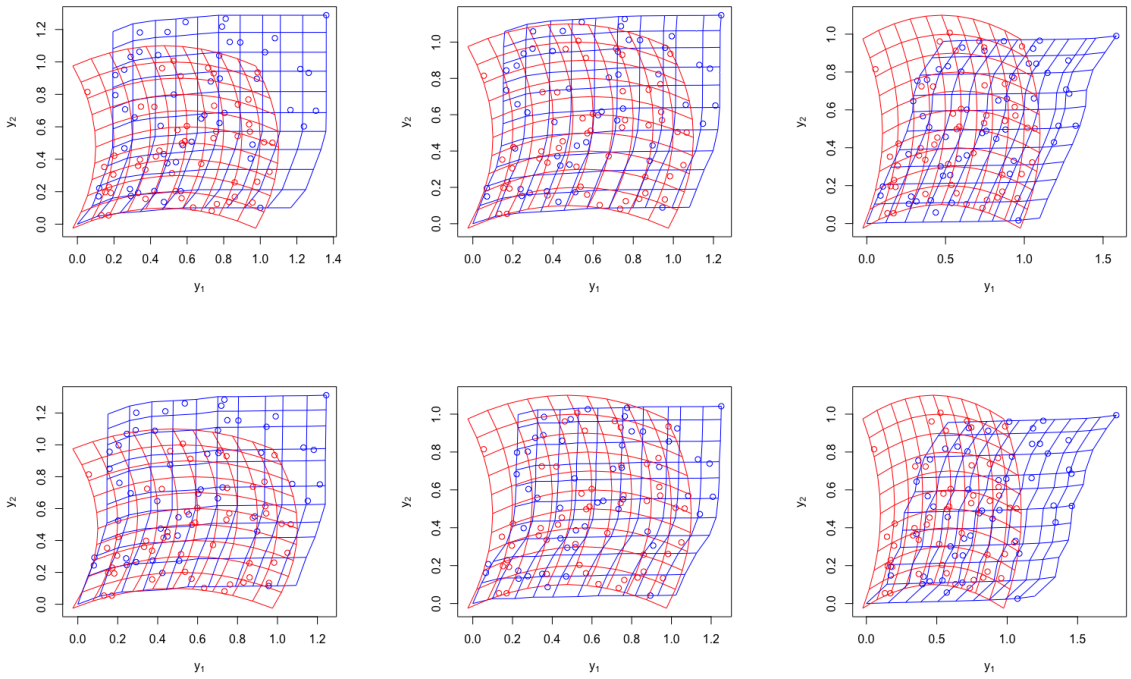


Figure 4.5: Estimated deformation (in blue) when $J = 3, 4, 5$ (from left to right) for quadratic deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

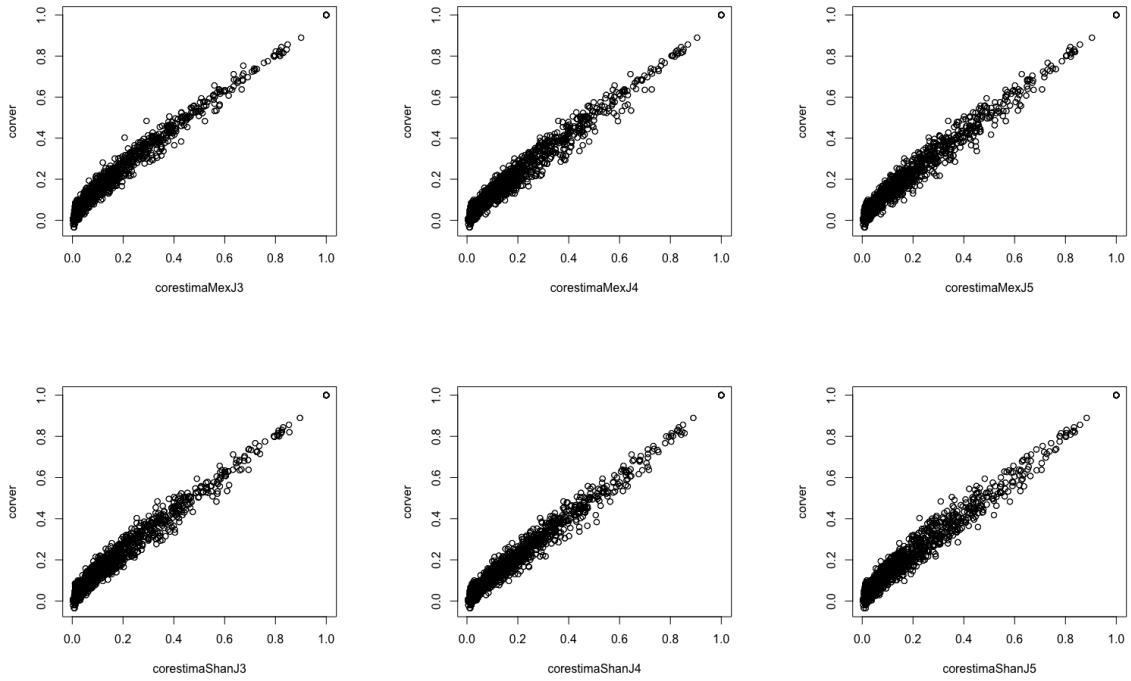


Figure 4.6: Comparison of the estimated correlation matrix versus the true correlation matrix when $J = 3, 4, 5$ (from left to right) for quadratic deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

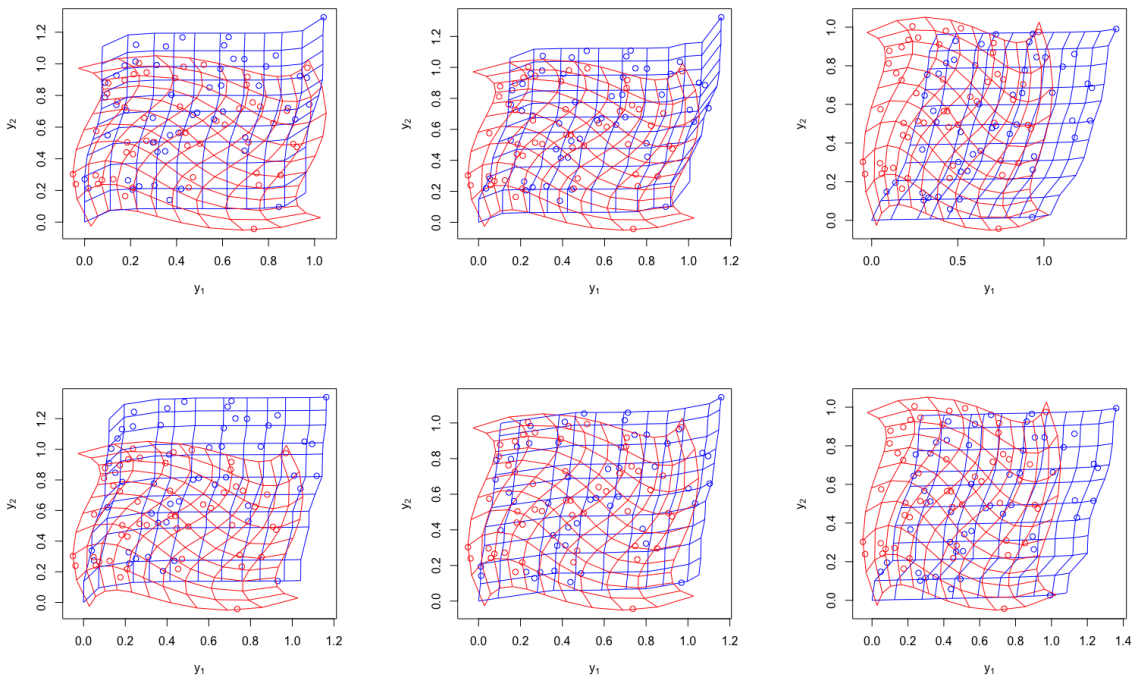


Figure 4.7: Estimated deformation (in blue) when $J = 3, 4, 5$ (from left to right) for non-linear deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

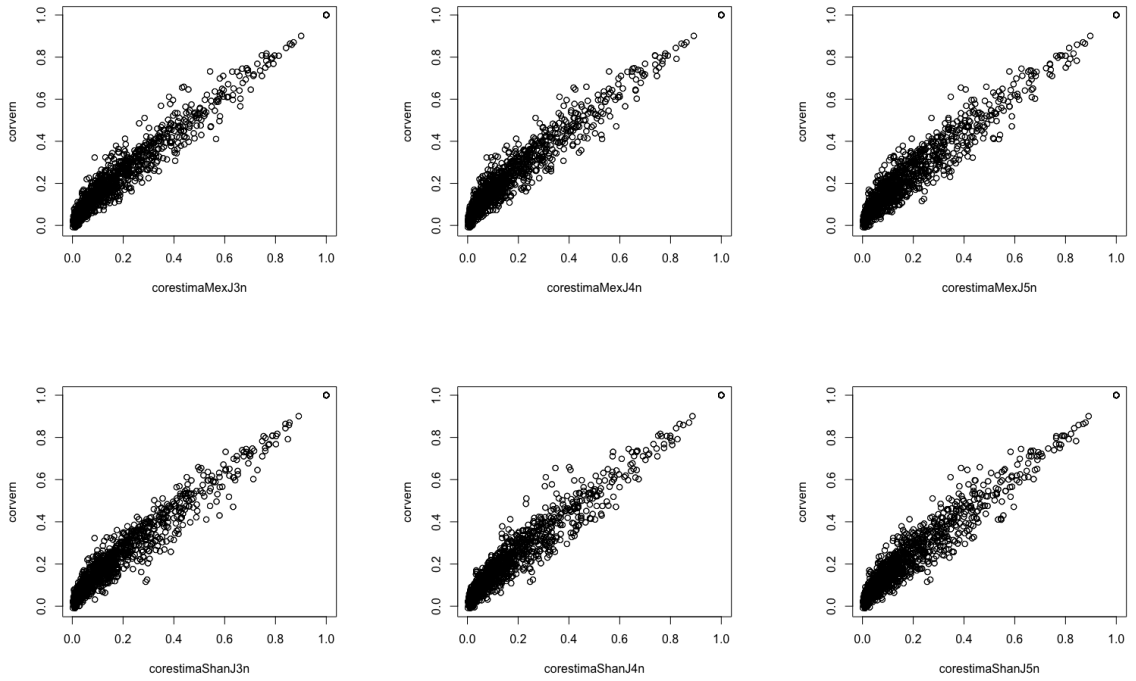


Figure 4.8: Comparison of the estimated correlation matrix versus the true correlation matrix when $J = 3, 4, 5$ (from left to right) for non-linear deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

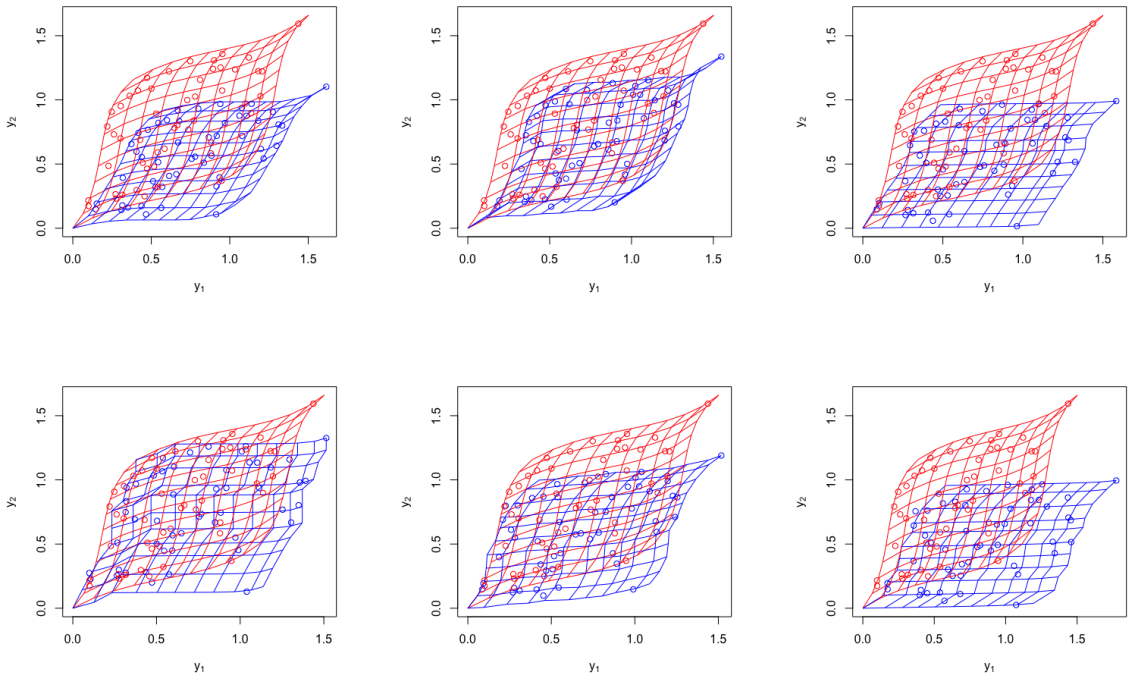


Figure 4.9: Estimated deformation (in blue) when $J = 3, 4, 5$ (from left to right) for wavelet deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

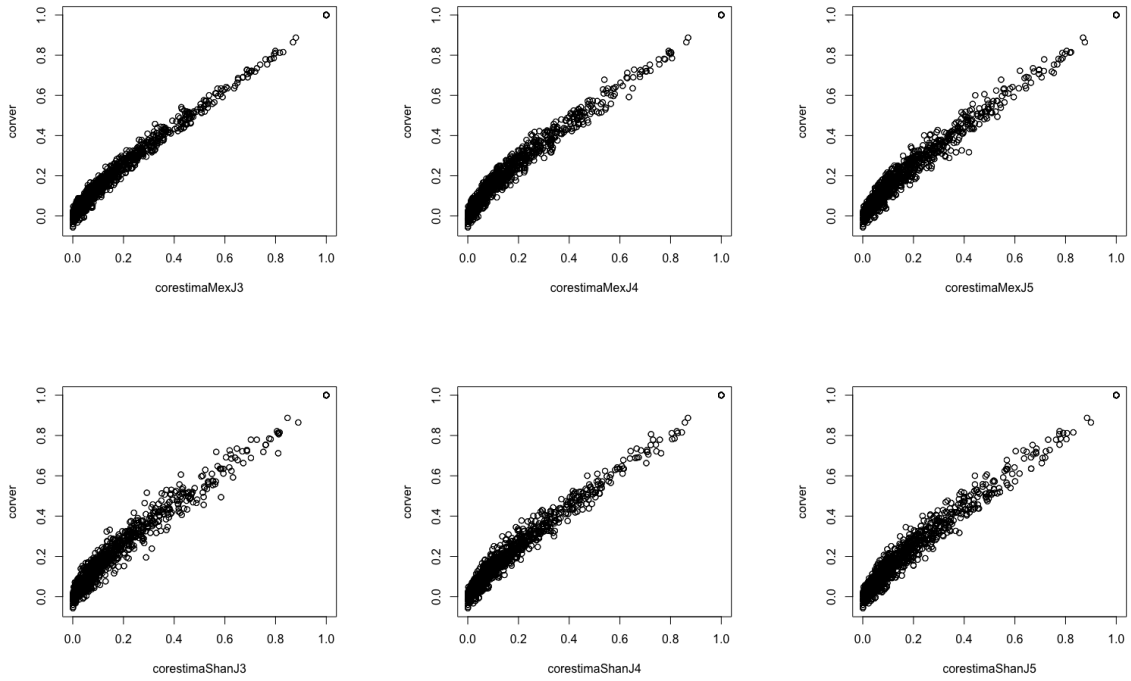


Figure 4.10: Comparison of the estimated correlation matrix versus the true correlation matrix when $J = 3, 4, 5$ (from left to right) for wavelet deformation case using Mexican hat (upper) and Shannon (bottom) wavelets.

4.5 Application

The dataset we use to illustrate the optimization procedure described in Section 4.3 is composed of historical daily maximum temperature records. The data can be obtained directly from GHCN (Global Historical Climatology Network)-Daily, an integrated public database of NOAA (National Oceanic and Atmospheric Administration) using R package **rnoaa**.

The data selected are daily maximum temperature records (in tenths of degrees Celsius) from Midwestern states of the USA. The region consists of 12 states: North Dakota, South Dakota, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, Ohio and Wisconsin. We use only climate monitoring stations which contains no missing data during the period between 1980 and 1999 (inclusive). The 51 stations selected are showed in Figure 4.11 and Figure 4.12 shown maximum temperature recorded at the 4 sampling stations.

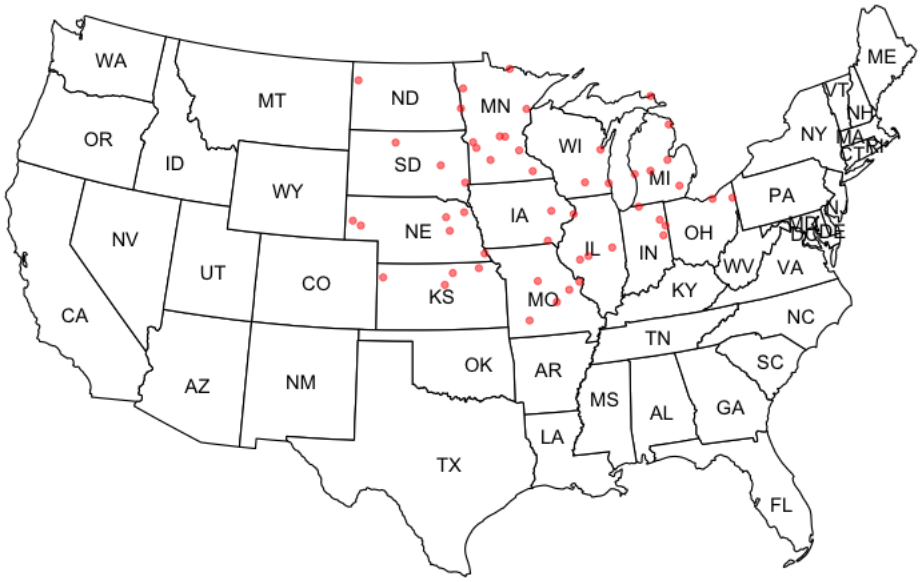


Figure 4.11: Locations of the stations selected from Midwestern states of the USA.

The results were obtained from the optimization procedure described in Section 4.3 and deformation (4.2.5), (4.2.6) with $J = 2, 3, 4$ using Mexican hat and Shannon wavelets. And the covariance function used has the formula of equation (4.4.9). Figure 4.13 presents the estimated deformation and Table 4.2 shows the MSEs of the correlation matrix for the several fits. We can see that MSEs are better for both Mexican hat and Shannon when $J = 3$. Same conclusions can be seen in the scatter plots (Figure 4.14). Note that the correlations are strong, maybe because the stations are from the same region.

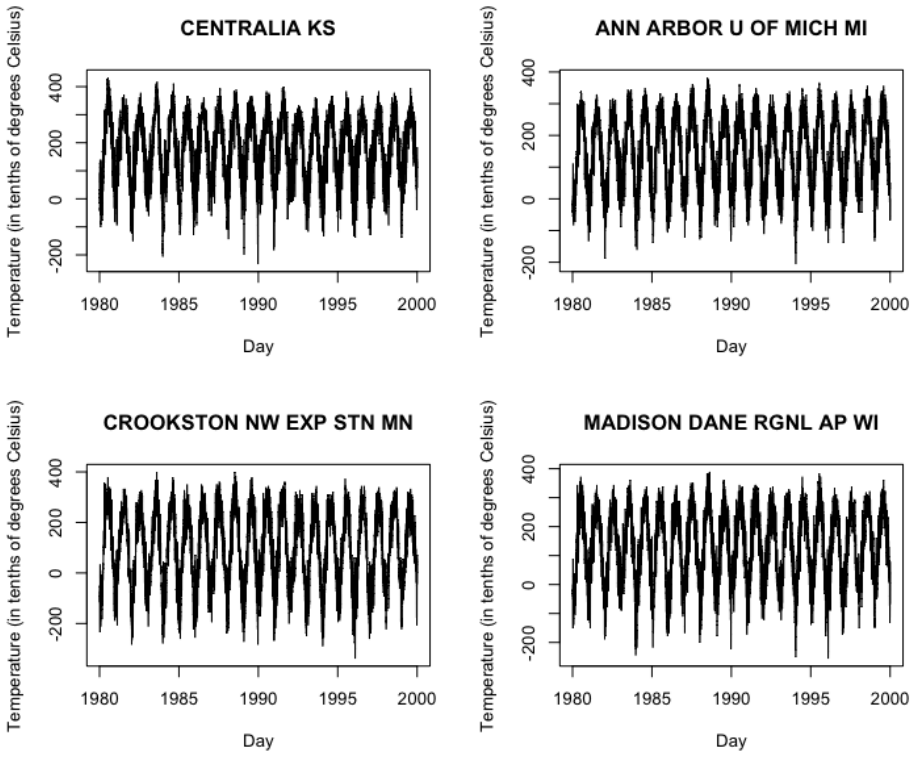


Figure 4.12: Maximum temperature recorded at the 4 sampling stations.

Table 4.2: Estimated parameters and MSEs of the correlation matrix for the several fits.

	Mexican hat			Shannon		
	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$
ν	0,06438	0,06299	0,07400	0,06826	0,06871	0,06746
θ	5,11259	4,24264	5,86402	4,04539	4,11567	4,98565
σ_ϵ^2	0,00024	0,00016	0,00022	0,00015	0,00011	0,00019
MSE	0,00106	0,00103	0,00174	0,00146	0,00109	0,00122

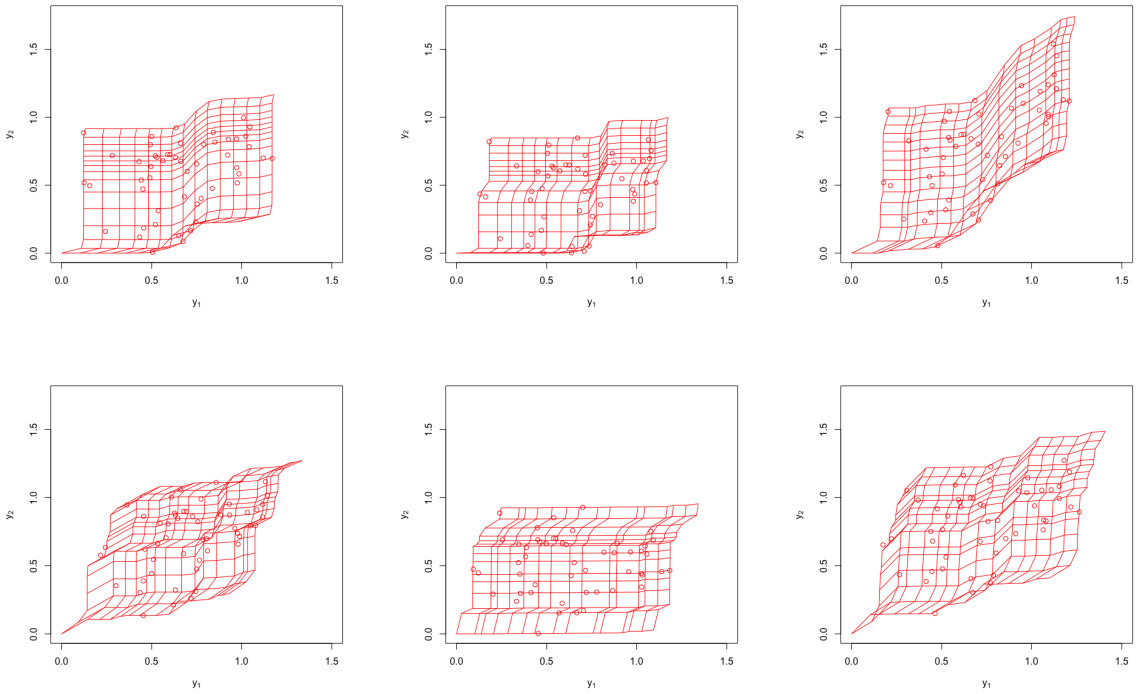


Figure 4.13: Estimated deformation (in red) when $J = 2, 3, 4$ (from left to right) using Mexican Hat (upper) and Shannon (bottom) wavelets.

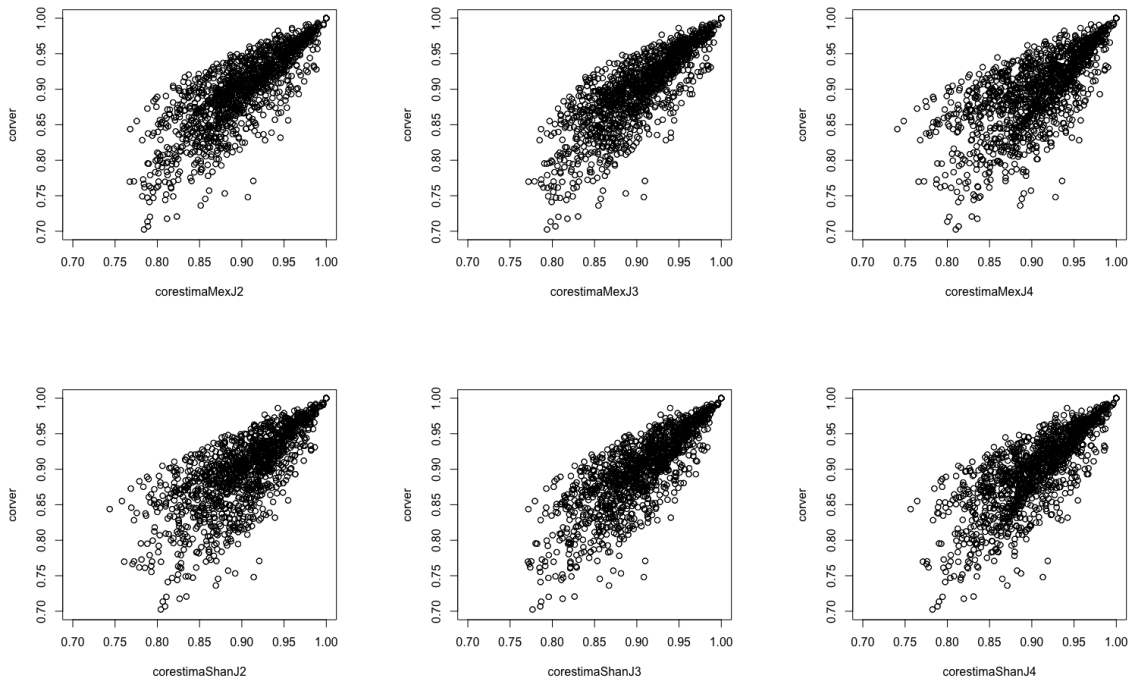


Figure 4.14: Comparison of the estimated correlation matrix versus the true correlation matrix when $J = 2, 3, 4$ (from left to right) using Mexican Hat (upper) and Shannon (bottom) wavelets.

Chapter 5

Conclusions

In this thesis, we studied two topics about spatio-temporal data analysis. First, a time-varying STARMA models based on locally stationary process in the sense of Dahlhaus was proposed. The STARMA models have been used in different fields of studies, but stationarity is a basic assumption for the models of STARMA class and is difficult to guarantee it in practice. Then, a multivariate locally stationary process is considered and we defined the time-varying STARMA model, where the time-varying parameters are expanded in wavelets. Two estimation methods were used, the least squares method when the tvSTAR models were selected and the Kalman filter when the terms of MA part were included. Some simulations examples in order to evaluate the performance of the proposed estimation procedure and an application were realized using Haar and Mexican hat wavelets. Since the models are characterized by the spatial weight matrix W and there are many choices to the matrix W, some simulation were done to compare the effects of different types of the matrices in the datasets which have non-separable covariance function.

And then, we have presented a method for non-stationary covariance function modeling, based on the monotonic functions. The mapping function f that maps the sampling locations in a geostatistical domain into space representations at a deformation domain guarantee the isotropicity of the spatial correlation of the deformation plane. Furthermore, the function f should be injective, then we used the strictly monotonic function with wavelet expansion to construct f. The maximum likelihood estimation was used for the process optimization and the results of the simulations showed that the performance of the algorithm is satisfactory when the deformation is simple. Note that the Mexican hat and Shannon wavelets were used in the simulations and application.

For the time-varying STARMA models, there are many interesting studies that can be explored in the future research, for examples:

- incorporate seasonal factors;
- explore innovation covariance function heterogeneity;

- consider the case when the spatial weight matrix is time-varying;
- compare the proposed model and other spatio-temporal models, such as spatial dynamic models and hierarchical spatio-temporal models.

By the inverse function theorem, if the Jacobian determinant of f, denoted by $\det(J(f))$, is non-zero, the f is an injective function. The deformation proposed in this thesis satisfies the condition that the Jacobian determinant is invertible. When $\det(J(f)) > 0$, the f is considered as a bijective function, that is, the orientation preserving of the function f is guaranteed. Then, some methods that guarantee the bijectivity of the transformation under the stochastic model considered could be explored in future studies.

Appendix A

Proofs

A.1 Proof of Proposition 1

Equation (3.1.4) can be written as

$$\boldsymbol{\nu}\left(\frac{t}{T}\right) = \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s} R_{sl}\left(\frac{t}{T}\right) W^{(l)}\mathbf{Z}\left(\frac{t-s}{T}\right) + \boldsymbol{\epsilon}\left(\frac{t}{T}\right), \tag{A.1.1}$$

where

$$R_{sl}\left(\frac{t}{T}\right) = \sum_{j\geq J}^{\infty}\sum_{m=0}^{2^j-1} \beta_{j,m}^{sl}\psi_{j,m}\left(\frac{t}{T}\right). \tag{A.1.2}$$

Define

$$S = \left[\sum_{s=1}^{p}\sum_{l=0}^{\lambda_s} R_{sl}\left(\frac{p+1}{T}\right) W^{(l)}\mathbf{Z}\left(\frac{p+1-s}{T}\right), ..., \sum_{s=1}^{p}\sum_{l=0}^{\lambda_s} R_{sl}\left(\frac{T-p}{T}\right) W^{(l)}\mathbf{Z}\left(\frac{T-p-s}{T}\right)\right]' \tag{A.1.3}$$

and

$$\boldsymbol{\epsilon} = \left[\text{vec}\left(\boldsymbol{\epsilon}\left(\frac{p+1}{T}\right)\right)', ..., \text{vec}\left(\boldsymbol{\epsilon}\left(\frac{T-p}{T}\right)\right)'\right]', \tag{A.1.4}$$

equation (3.1.10) can be decomposed as

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'(\boldsymbol{\Psi}\boldsymbol{\beta} + S + \boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{\epsilon} + [(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1} - (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}]\boldsymbol{\Psi}'\boldsymbol{\epsilon} + (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'S \\
&= \boldsymbol{\beta} + T_1 + T_2 + T_3.
\end{aligned} \tag{A.1.5}$$

Since $T_1 = (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{\epsilon}$, we have $E[T_1] = 0$ and by Taylor expansion of the matrix $(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}$, $T_2 = T_{21} + T_{22}$, where

$$T_{21} = (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}(E\boldsymbol{\Psi}'\boldsymbol{\Psi} - \boldsymbol{\Psi}'\boldsymbol{\Psi})(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{\epsilon} \tag{A.1.6}$$

and

$$T_{22} = (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}(E\boldsymbol{\Psi}'\boldsymbol{\Psi} - \boldsymbol{\Psi}'\boldsymbol{\Psi})(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}(E\boldsymbol{\Psi}'\boldsymbol{\Psi} - \boldsymbol{\Psi}'\boldsymbol{\Psi})(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{\epsilon}. \tag{A.1.7}$$

Analogously to Chang and Morettin (2005), we obtain

$$\begin{aligned}\|T_{21}\|_2 &\leq \left\|(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\right\|_2^2 \left\|\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\Psi})\right\|_2^{1/2} \left\|\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\epsilon})\right\|_2^{1/2} \\ &= O((nT)^{-2})O(2^{2J}(nT)^{1/2})O((nT)^{1/2}) \\ &= O(2^J(nT)^{-1}) \end{aligned} \tag{A.1.8}$$

and

$$\begin{aligned}\|T_{22}\|_2 &\leq \left\|(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\right\|_2^3 \left\|\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\Psi})\right\|_2^2 \left\|\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\epsilon})\right\|_2 \\ &= O((nT)^{-3})O(2^{2J}nT)O((nT)^{1/2}) \\ &= O(2^{2J}(nT)^{-3/2}). \end{aligned} \tag{A.1.9}$$

From Donoho et al. (1995), we have

$$\sup\left\{\sum_{j\geq J}^{\infty}\sum_{k=0}^{2^j-1}|\beta_{j,k}^{sl}|^2\right\} = O(2^{-2Ju}), \tag{A.1.10}$$

where $u = d + 1/2 - 1/\tilde{p}$ with $\tilde{p} = \min\{p, 2\}$ and similarly to Dahlhaus et al. (1999),

$$\begin{aligned}\|T_3\|_2 &\leq (\left\|(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\right\|_2 + \left\|(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1} - (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\right\|_2)\left\|\boldsymbol{\Psi}'S\right\|_2 \\ &= (O((nT)^{-1}) + O(2^J(nT)^{-3/2}) + O(2^{2J}(nT)^{-2}))\left\|\boldsymbol{\Psi}'S\right\|_2 \\ &= O((nT)^{-1})O(nT(2^{-Ju} + (nT)^{-1/2}2^{-J(d-1/2-1/(2\tilde{p}))})\sqrt{\log(nT)})) \\ &= O((2^{-Ju} + (nT)^{-1/2}2^{-J(d-1/2-1/(2\tilde{p}))})\sqrt{\log(nT)})) \\ &= O((nT)^{-1/2-\tau}), \end{aligned} \tag{A.1.11}$$

for some $\tau > 0$. Then, $\|T_3\|_2 \leq O((nT)^{-1/2})$.

The result $(i.)$ follows.

For $(ii.)$, we have

$$\begin{aligned}E[T_1T_1'] &= (E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\epsilon})(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1} \\ &\leq \left\|(E\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\right\|_2^2 \left\|\mathrm{Cov}(\boldsymbol{\Psi}'\boldsymbol{\epsilon})\right\|_2 \\ &= O((nT)^2)O(nT) \\ &= O((nT)^{-1}). \end{aligned} \tag{A.1.12}$$

Analogously,

$$E[T_2T_2'] = O(2^{2J}(nT)^{-2}) + O(2^{4J}(nT)^{-3}), \tag{A.1.13}$$

$$E[T_3T_3'] = O((nT)^{-1}), \tag{A.1.14}$$

$$E[T_1T_2'] = E[T_1'T_2] = E[T_2T_3'] = E[T_2'T_3] = O(2^J(nT)^{-3/2}) + O(2^{2J}(nT)^{-2}), \tag{A.1.15}$$

$$E[T_1T_3'] = E[T_1'T_3] = O((nT)^{-1}), \tag{A.1.16}$$

and then the result $(ii.)$ follows.

A.2 Proof of Proposition 2

Lemma 1 Suppose the assumptions A.1-A.4 hold, then

$$N\mathbf{H}(\mathbf{\Psi}'\mathbf{\Psi})^{-1}\mathbf{H}' \xrightarrow{P} \mathbf{\Gamma}, \tag{A.2.1}$$

where $\mathbf{\Psi}$ is as in (3.1.8) and

$$\mathbf{\Gamma} = \lim_{N\to\infty} N\mathbf{H}E[\mathbf{\Psi}'\mathbf{\Psi}]^{-1}\mathbf{H}'. \tag{A.2.2}$$

Proof. Let ψ_{gl} the gth row and lth column cell of $\mathbf{\Psi}$, we have

$$\psi_{gl} = \sum_{i=1}^{n} f_{gl}^{i} Z_{gl}^{i}, \tag{A.2.3}$$

where f_{gl}^{i} is obtained by the product of a wavelet function and a weight function up to one and Z_{gl}^{i} is a locally stationary process, both are elements of (g,l)th cell of $\mathbf{\Psi}$ at location i.

Let $(\psi\psi)_{lm}$ the lth row and mth column cell of any subpartition of $\mathbf{\Psi}'\mathbf{\Psi}$, $(\psi\psi)_{lm}$ can be expressed as

$$\begin{aligned}
(\psi\psi)_{lm} &= \sum_{g=1}^{N} \psi_{gl}\psi_{gm} \\
&= \sum_{g=1}^{N} \left[\sum_{i=1}^{n} (f_{gl}^{i} Z_{gl}^{i}) \sum_{i=1}^{n} (f_{gm}^{i} Z_{gm}^{i}) \right] \\
&= \sum_{g=1}^{N} f_{gl}^{1} Z_{gl}^{1} f_{gm}^{1} Z_{gm}^{1} + \sum_{g=1}^{N} f_{gl}^{1} Z_{gl}^{1} f_{gm}^{2} Z_{gm}^{2} + \cdots + \sum_{g=1}^{N} f_{gl}^{n} Z_{gl}^{n} f_{gm}^{n} Z_{gm}^{n}.
\end{aligned}$$

The proof of $(\psi'\psi)_{lm}$ is similar to that of Sato et al. (2007) and the result follows from the Weak Law for $L^1 - Maxingales$ of Andrews (1988).

Lemma 2 Suppose the assumptions A.1-A.4 hold, then

$$\frac{1}{\sqrt{N}}\mathbf{G}(\mathbf{\Psi}'\boldsymbol{\epsilon}) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, \mathbf{\Gamma}^{-1}), \tag{A.2.4}$$

where $\boldsymbol{\epsilon}$ is as in (A.1.4) and covariance matrix

$$\mathbf{\Gamma}^{-1} = \lim_{N\to\infty} \mathbf{G}\frac{E[\mathbf{\Psi}'\mathbf{\Psi}]}{N}\mathbf{G}', \tag{A.2.5}$$

where $N = n(T-p)$, $\mathbf{G}$ is a matrix with the same number of columns of $\mathbf{\Psi}$ and K rows.

Proof. Let $(\psi e)_l$ the lth element of $\mathbf{\Psi}'\boldsymbol{\epsilon}$,

$$\begin{aligned}
(\psi e)_l &= \sum_{g=1}^{N} (\psi e)_{gl} \\
&= \sum_{g=1}^{N} f_{gl}\tilde{Z}_{gl}\epsilon_g,
\end{aligned}$$

where f_{gl} is a wavelet function and $\tilde{Z}_{gl}$ is a weighted arithmetic mean of n locally stationary processes with weights sum to one, both are of (g,l)th cell of $\boldsymbol{\Psi}$ and ϵ_g, gth element of $\boldsymbol{\epsilon}$, is independent Gaussian process.

Let $\boldsymbol{\Psi}'\boldsymbol{\epsilon} = \sum_{g=1}^{N} b_g$, where b_g is a vector consist of $(\psi e)_{gl}$ and define $\mathscr{F}_g$ as a σ-algebra containing all the information up to moment g. Since

$$E[b_g \mid \mathscr{F}_{g-1}] = 0, \tag{A.2.6}$$

$\{b_g\}_{g=1}^{\infty}$ is a martingale difference sequence. Then

$$\begin{aligned} E\left[\frac{\sum_{g=1}^{N} b_g b_g'}{N}\right] &= E\left[\frac{\boldsymbol{\Psi}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\boldsymbol{\Psi}}{N}\right] \\ &= E\left[E\left[\frac{\boldsymbol{\Psi}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\boldsymbol{\Psi}}{N}\,\middle|\, \boldsymbol{\Psi}\right]\right] \\ &= E\left[\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{N}\right]. \end{aligned}$$

Analogously,

$$\mathbf{G}E\left[\frac{\sum_{g=1}^{N} b_g b_g'}{N}\mathbf{G}' = \mathbf{G}E\left[\frac{\boldsymbol{\Psi}'\boldsymbol{\Psi}}{N}\right]\right]\mathbf{G}' \to \boldsymbol{\Gamma}^{-1}, \tag{A.2.7}$$

where

$$\boldsymbol{\Gamma}^{-1} = \lim_{N\to\infty} \mathbf{G}\frac{E[\boldsymbol{\Psi}'\boldsymbol{\Psi}]}{N}\mathbf{G}' \tag{A.2.8}$$

and similar to Lemma 1,

$$\mathbf{G}\frac{\sum_{g=1}^{N} b_g b_g'}{N}\mathbf{G}' \xrightarrow{P} \boldsymbol{\Gamma}^{-1}. \tag{A.2.9}$$

The result follows the central limit theorem for martingale difference sequence of White (2000).

From equation (A.1.11), $\sqrt{N}\|T_3\|_2 = O((nT)^{-\tau}) = o_p(1)$, applying Lemma 1, Lemma 2 and Slutsky theorem, we have

$$\begin{aligned} \sqrt{N}\mathbf{H}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{N}\mathbf{H}(T_1 + T_2 + T_3) && \text{(A.2.10)} \\ &= \sqrt{N}\mathbf{H}(T_1 + T_2) + o_p(1) && \text{(A.2.11)} \\ &= \sqrt{N}\mathbf{H}(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\boldsymbol{\epsilon} + o_p(1) && \text{(A.2.12)} \\ &= \mathscr{N} + o_p(1), && \text{(A.2.13)} \end{aligned}$$

where $\mathscr{N}$ has a K-dimensional normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}$ as in (A.2.2). The result follows the Slutsky's theorem.

Appendix B

Figures of Simulations

B.1 Histograms of Estimated Coefficients

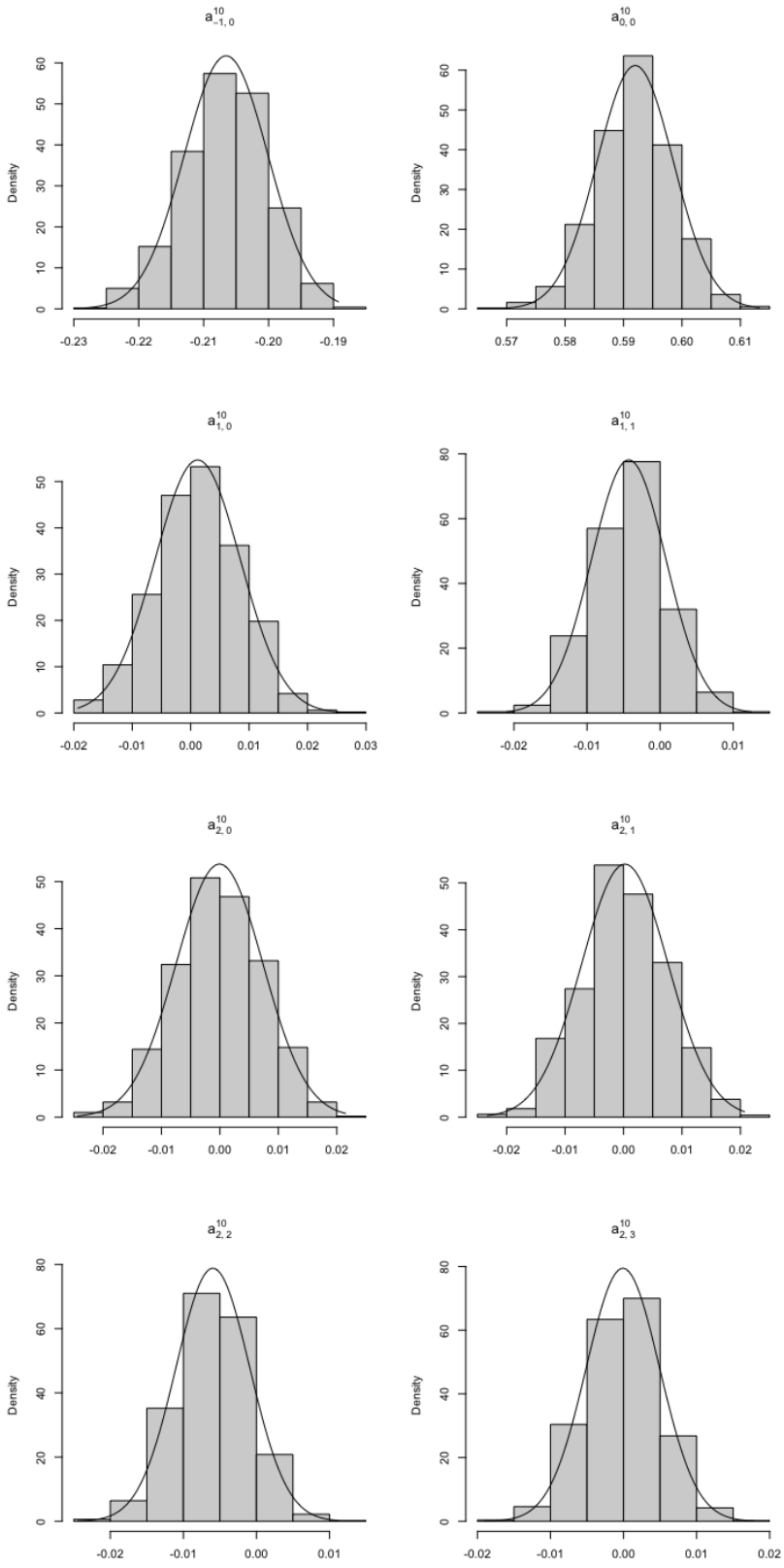


Figure B.1: Histograms of estimated coefficients of ϕ_{10} of Group 1 obtained by Haar wavelet and least squares estimation.

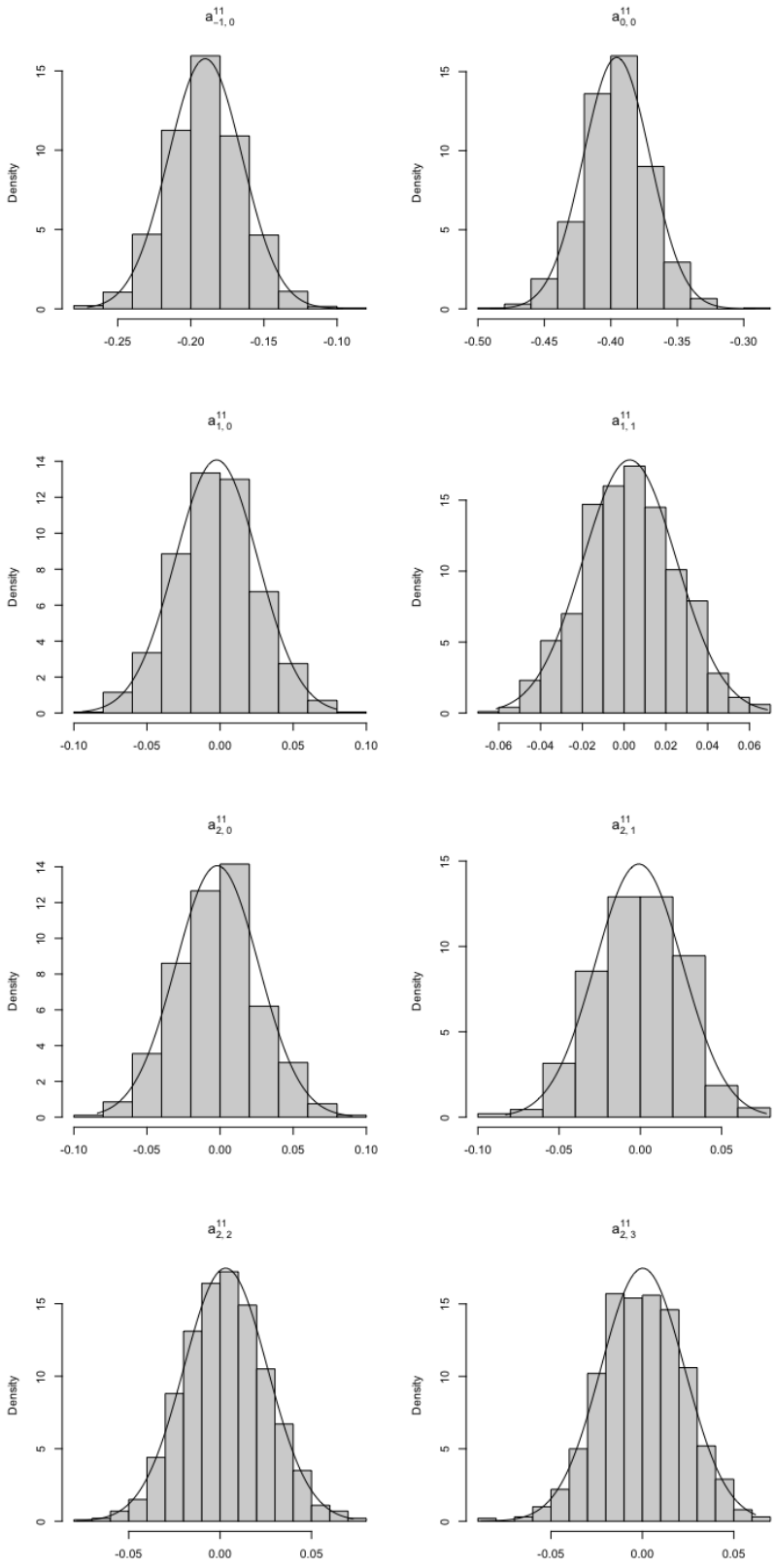


Figure B.2: Histograms of estimated coefficients of ϕ_{11} of Group 1 obtained by Haar wavelet and least squares estimation.

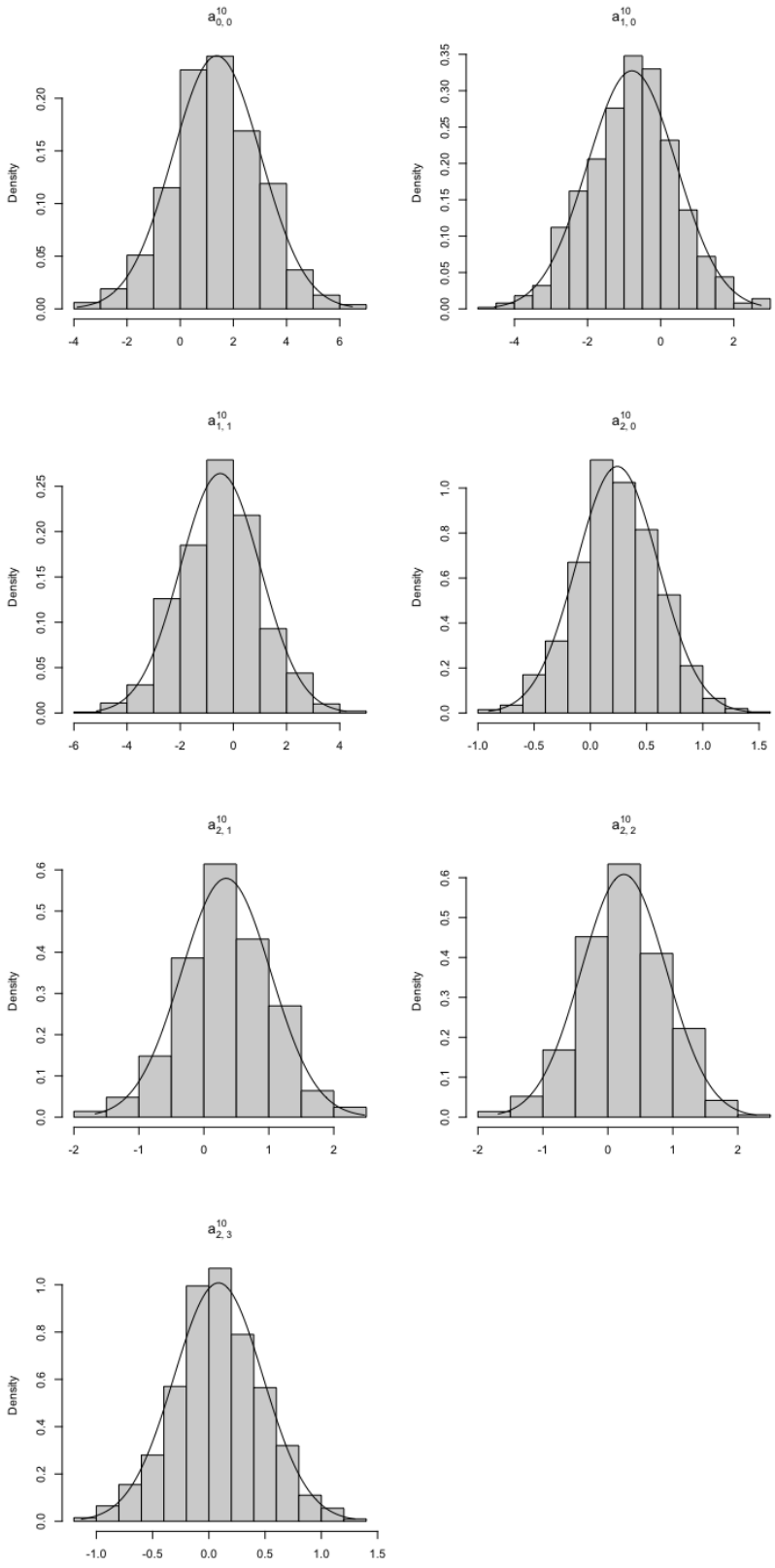


Figure B.3: Histograms of estimated coefficients of ϕ_{10} of Group 2 obtained by Mexican hat wavelet and least squares estimation.

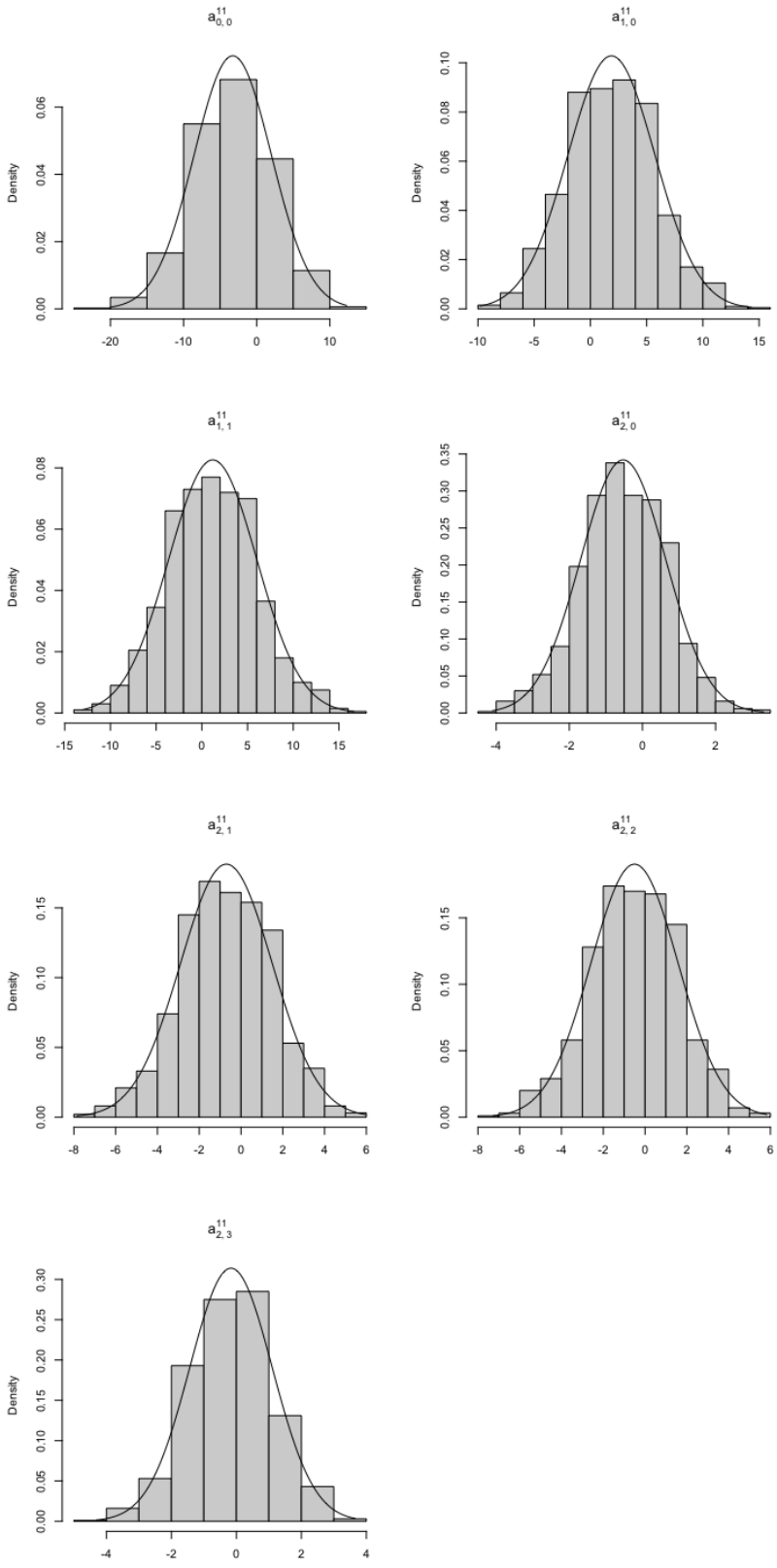


Figure B.4: Histograms of estimated coefficients of ϕ_{11} of Group 2 obtained by Mexican hat wavelet and least squares estimation.

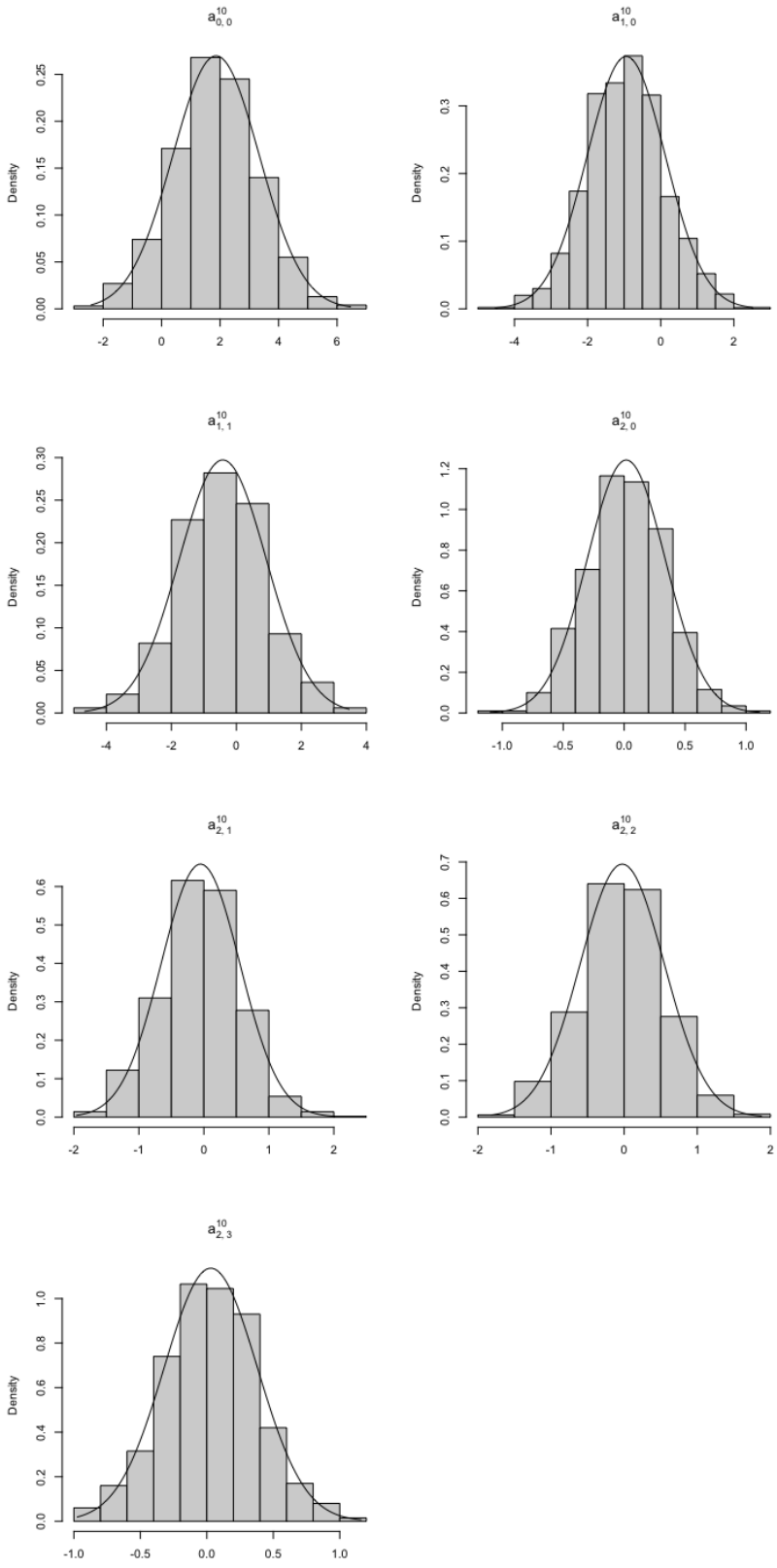


Figure B.5: Histograms of estimated coefficients of ϕ_{10} of Group 3 obtained by Mexican hat wavelet and least squares estimation.

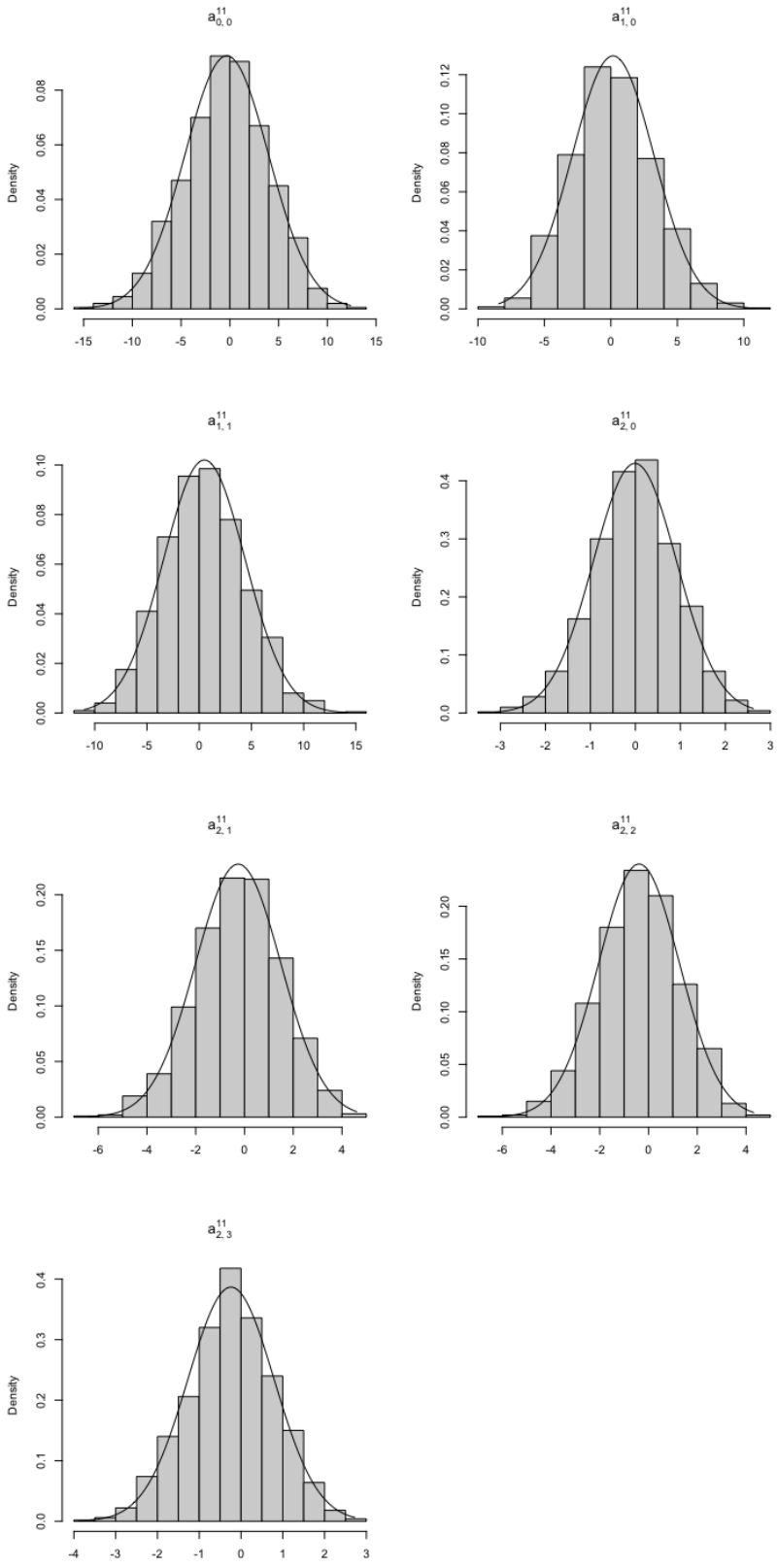


Figure B.6: Histograms of estimated coefficients of ϕ_{11} of Group 3 obtained by Mexican hat wavelet and least squares estimation.

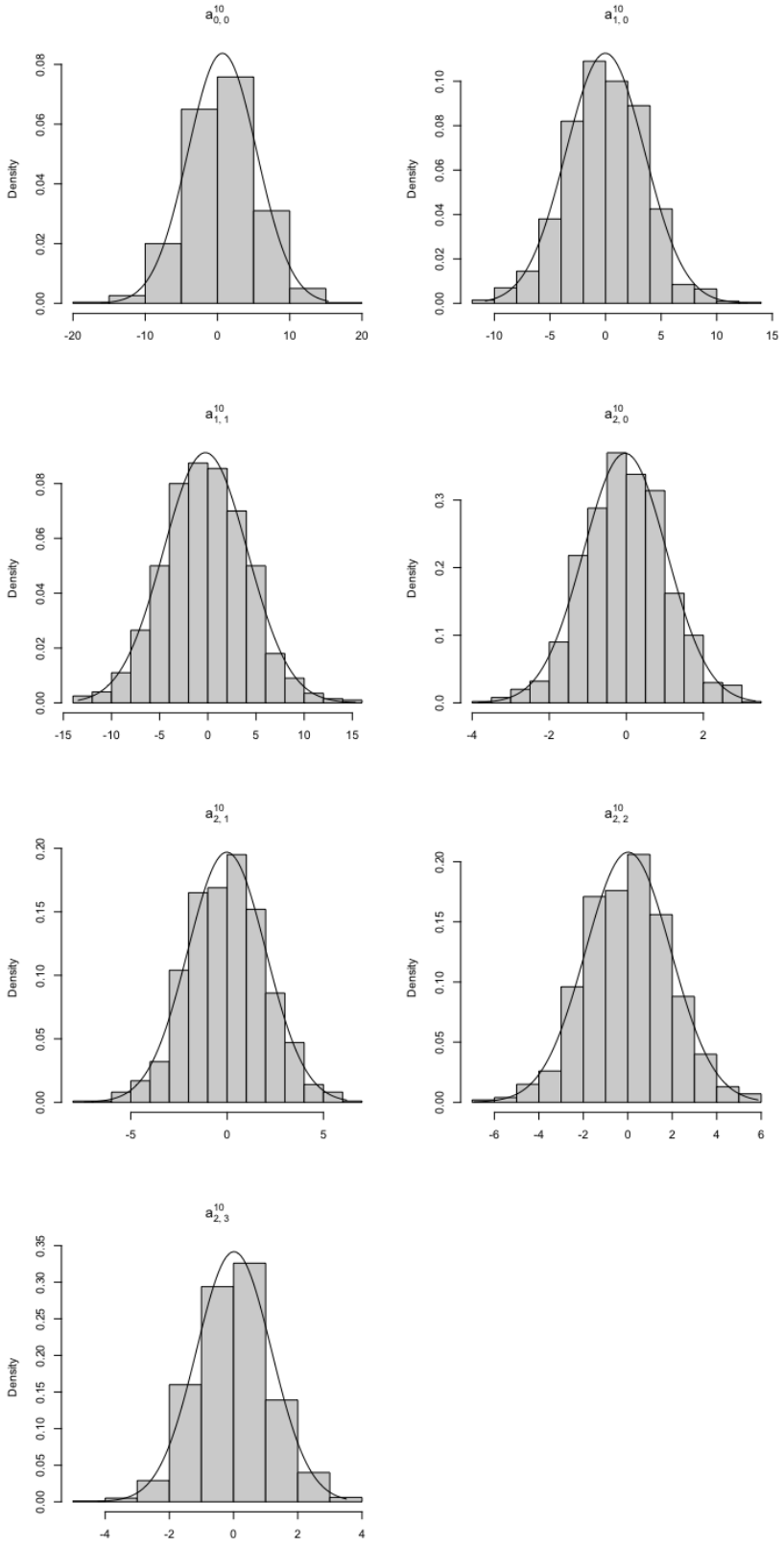


Figure B.7: Histograms of estimated coefficients of ϕ_{10} of Group 4 obtained by Mexican hat wavelet and Kalman filter estimation.

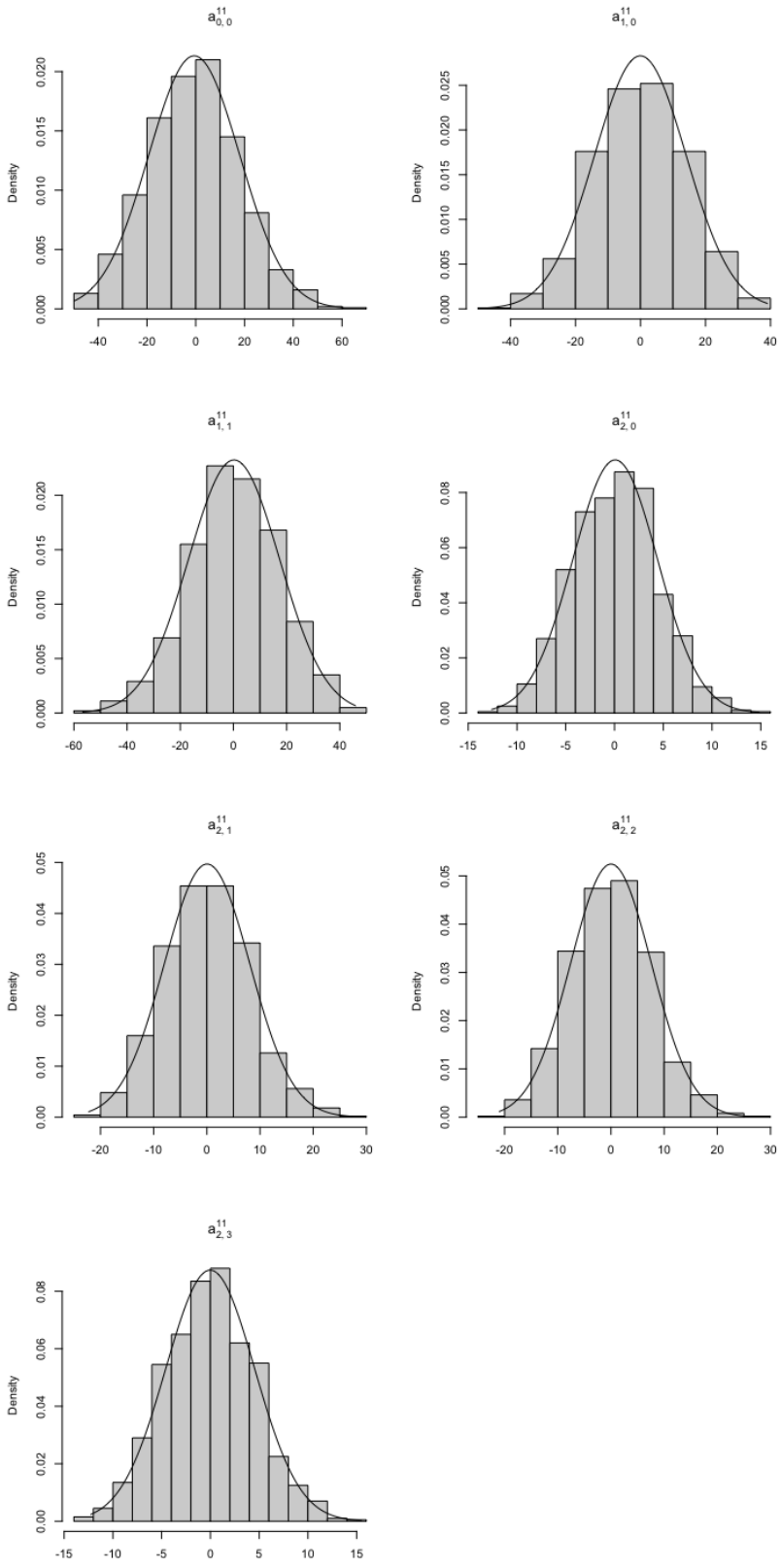


Figure B.8: Histograms of estimated coefficients of ϕ_{11} of Group 4 obtained by Mexican hat wavelet and Kalman filter estimation.

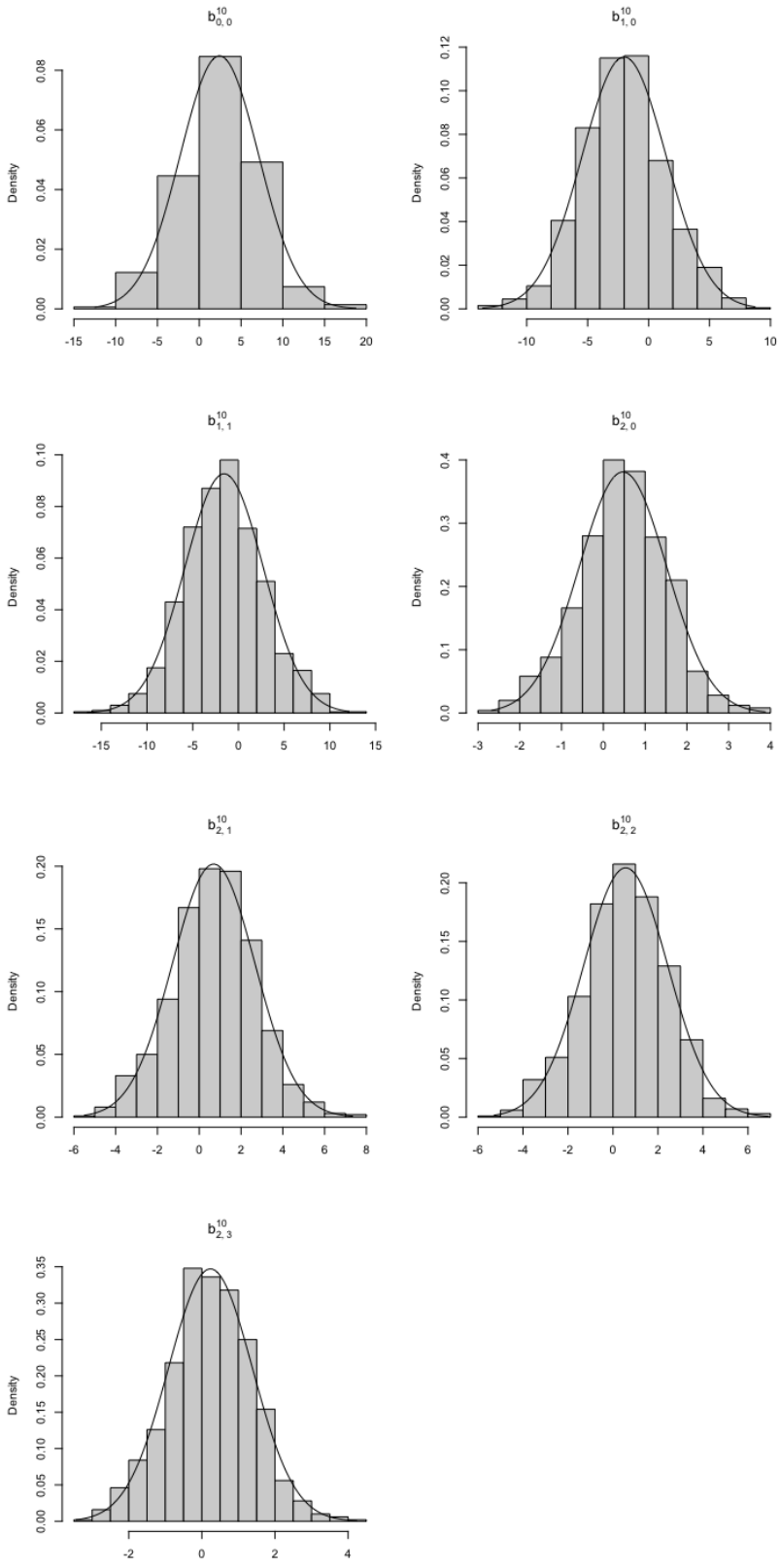


Figure B.9: Histograms of estimated coefficients of θ_{10} of Group 4 obtained by Mexican hat wavelet and Kalman filter estimation.

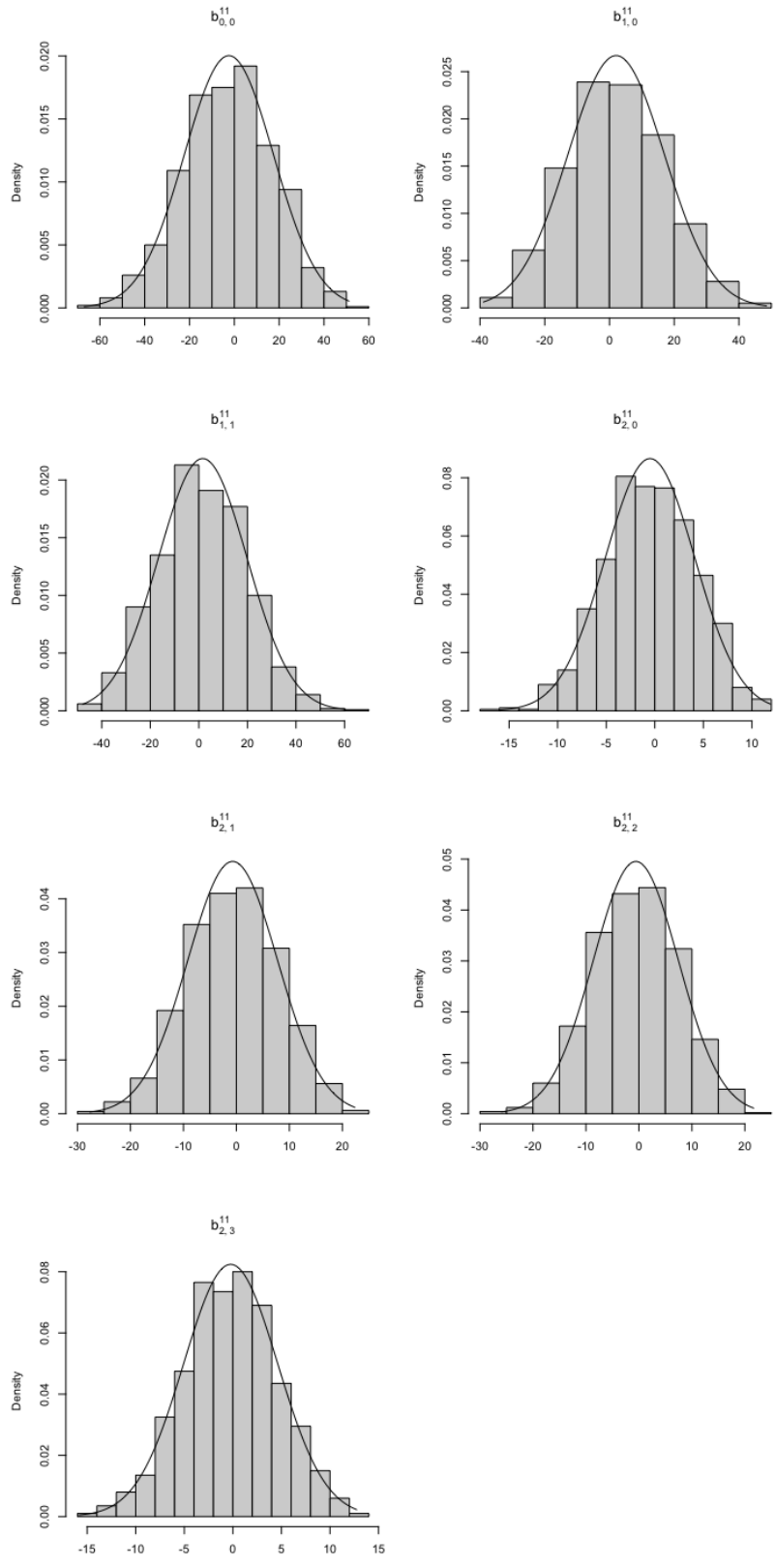


Figure B.10: Histograms of estimated coefficients of θ_{11} of Group 4 obtained by Mexican hat wavelet and Kalman filter estimation.

B.2 Boxplots of MSEs

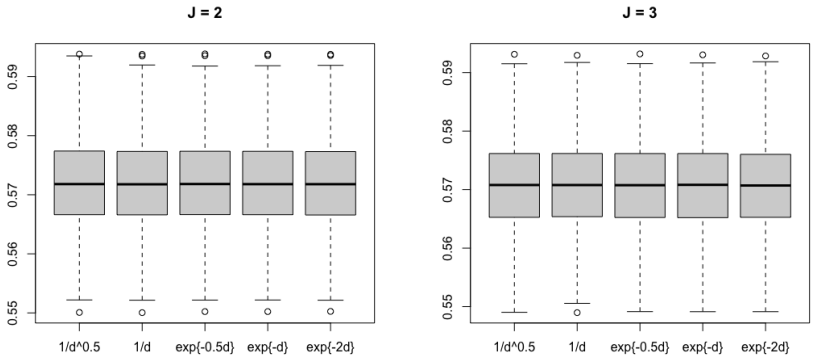


Figure B.11: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 0.5$.

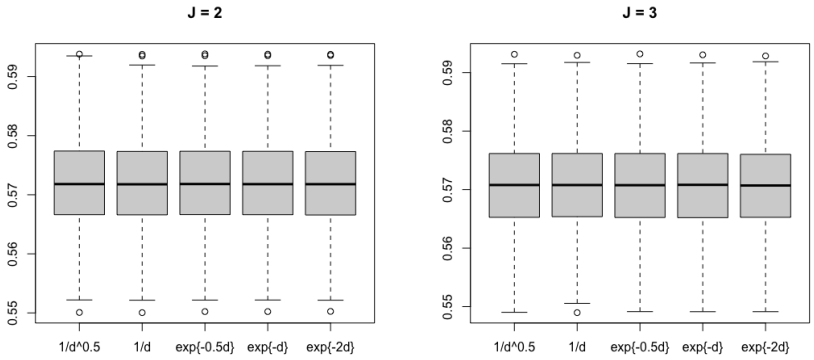


Figure B.12: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 0.5$.

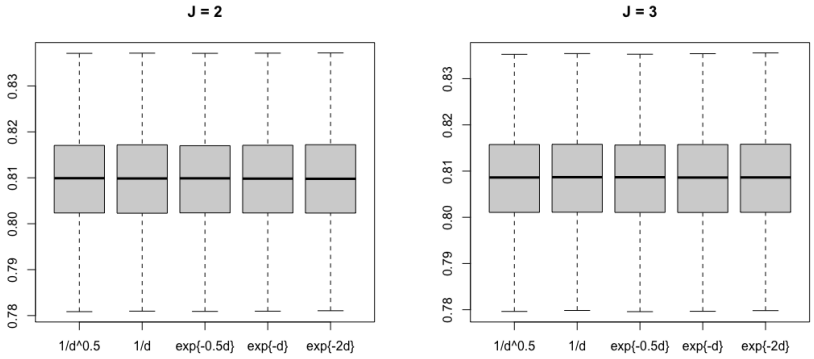


Figure B.13: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1$.

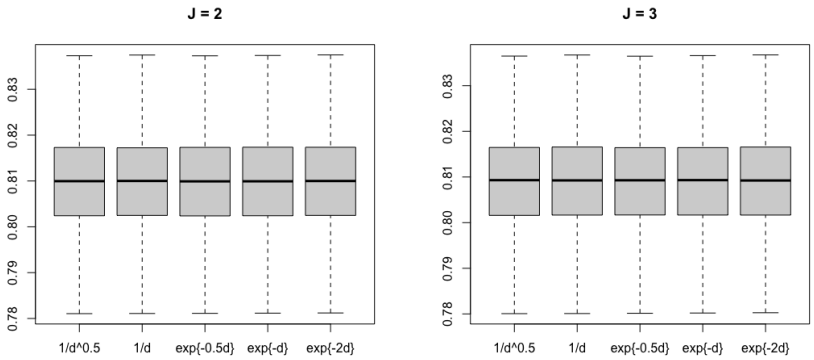


Figure B.14: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1$.

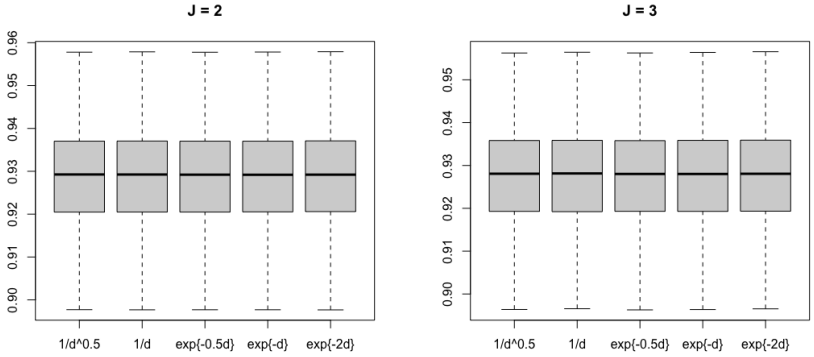


Figure B.15: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1.5$.

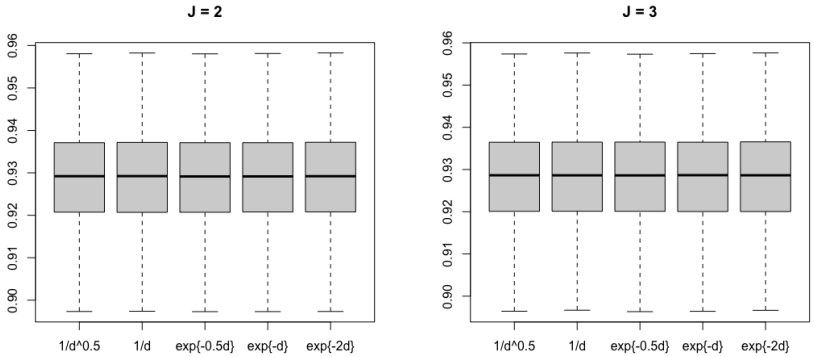


Figure B.16: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1.5$.

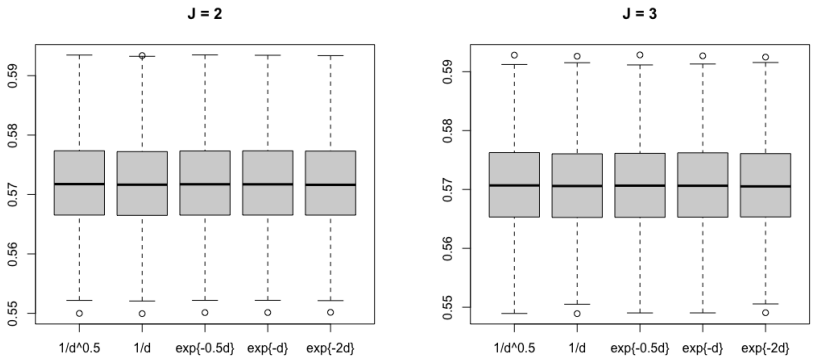


Figure B.17: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 0.5$.

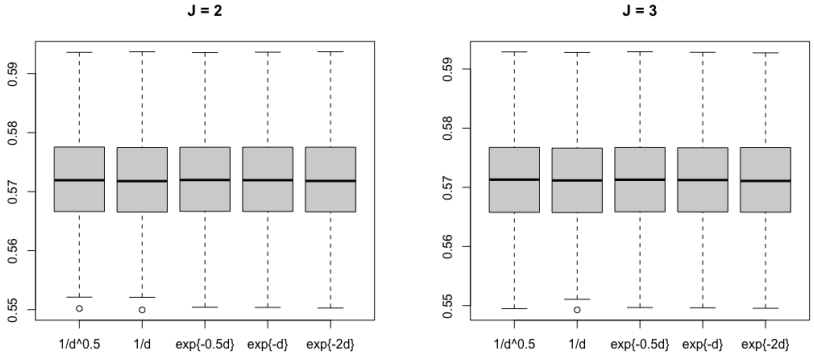


Figure B.18: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 0.5$.

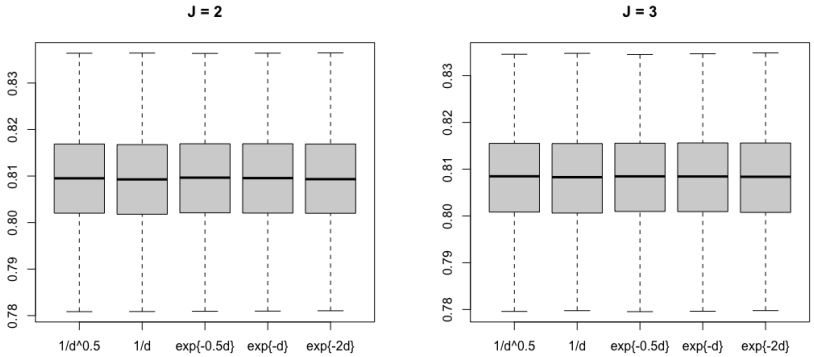


Figure B.19: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1$.

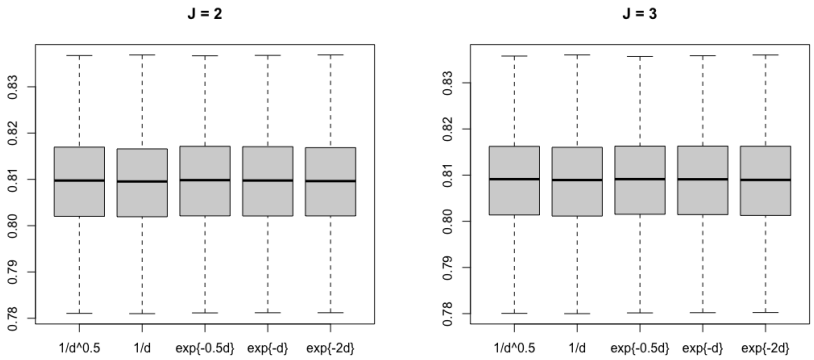


Figure B.20: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1$.

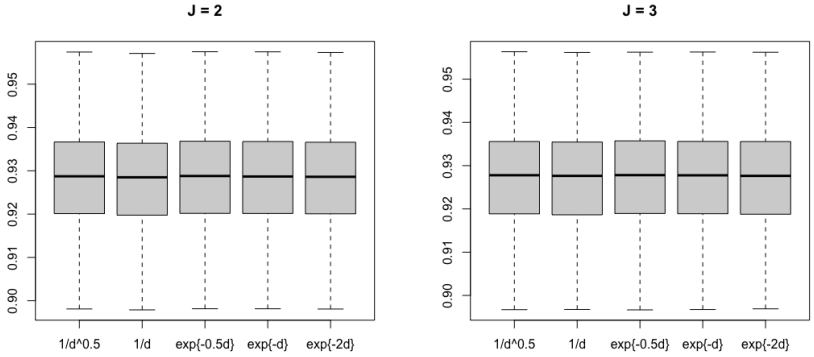


Figure B.21: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1.5$.

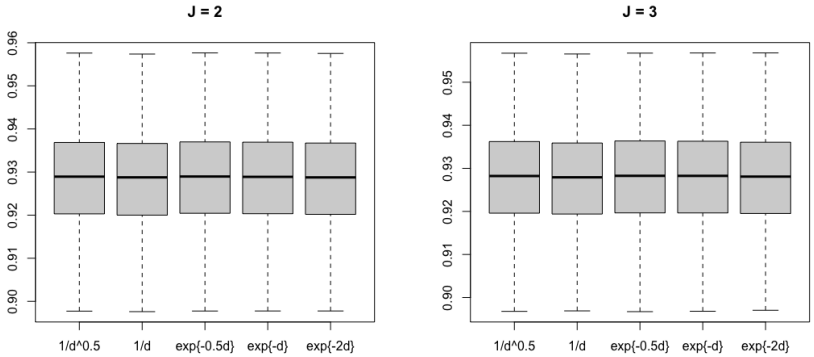


Figure B.22: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1.5$.

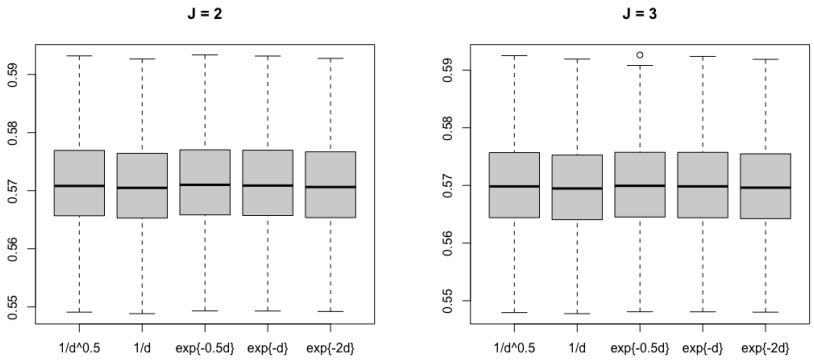


Figure B.23: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 0.5$.

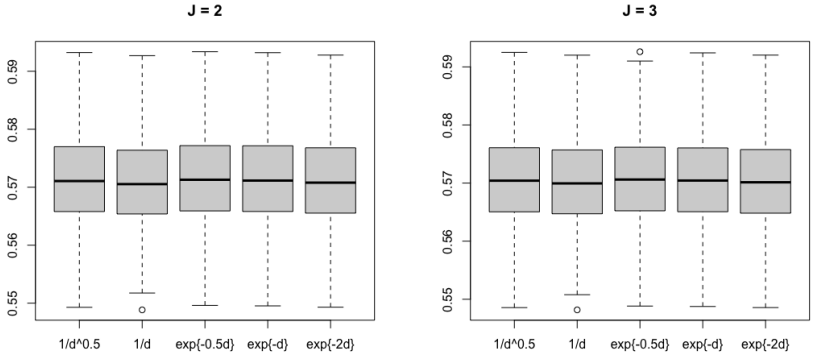


Figure B.24: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 0.5$.

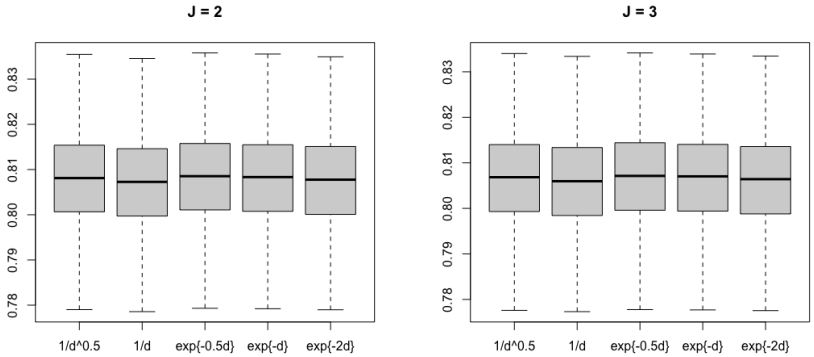


Figure B.25: Boxplots of MSEs obtained by tvSTAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1$.

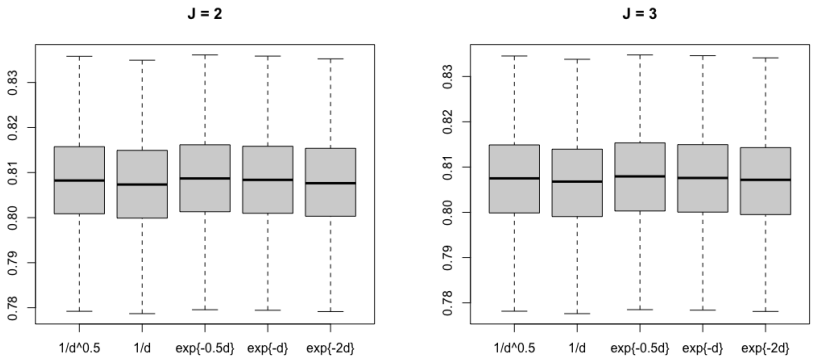


Figure B.26: Boxplots of MSEs obtained by tvSTAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1$.

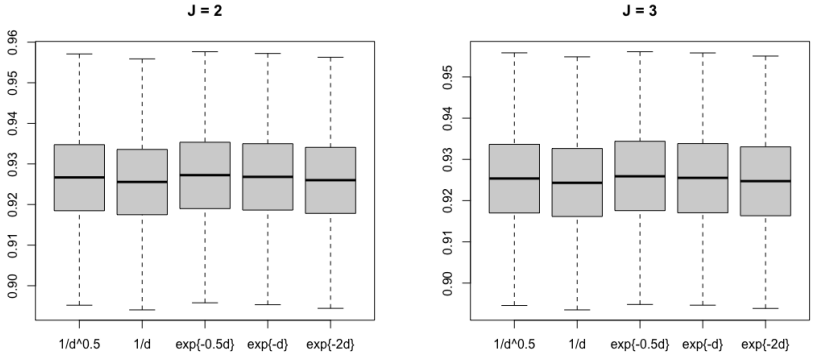


Figure B.27: Boxplots of MSEs obtained by STAR(1_1) model with Haar wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1.5$.

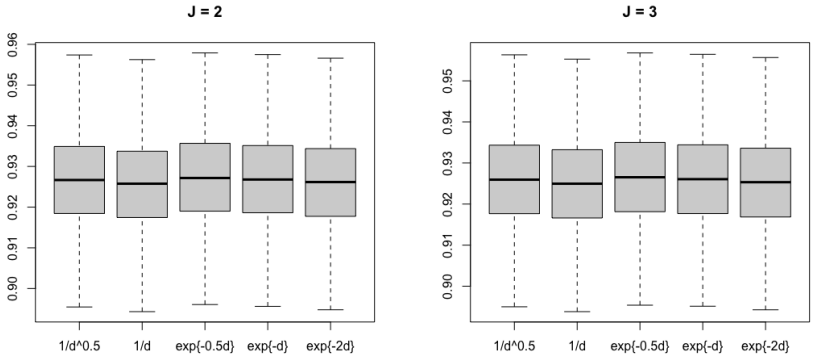


Figure B.28: Boxplots of MSEs obtained by STAR(1_1) model with Mexican hat wavelet and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1.5$.

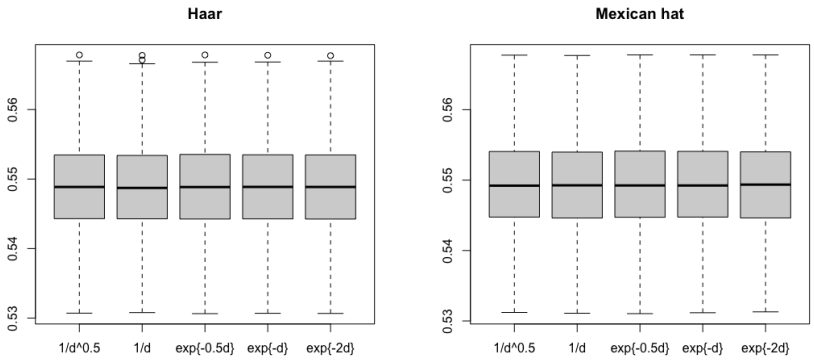


Figure B.29: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 0.5$.

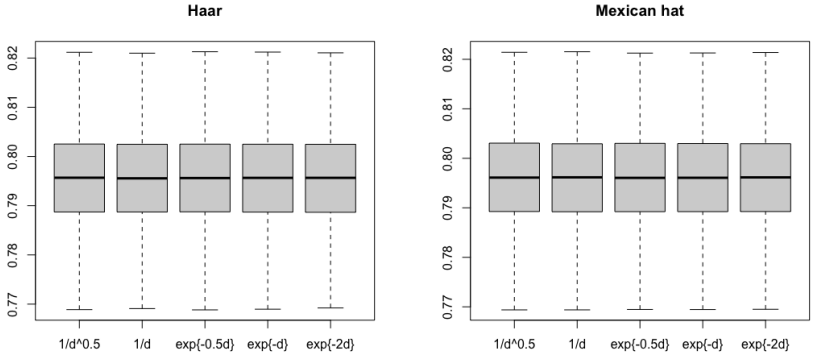


Figure B.30: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1$.

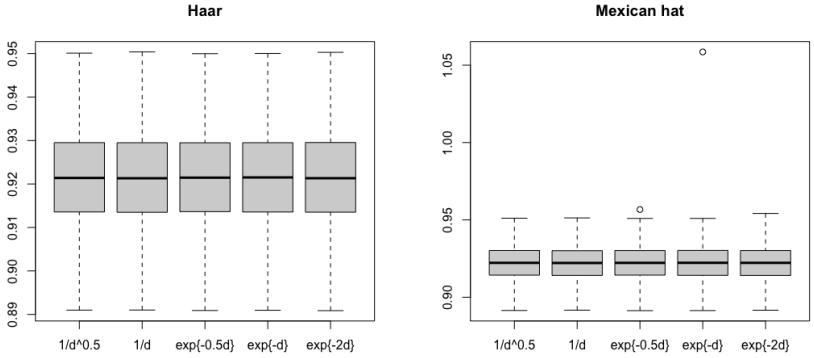


Figure B.31: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.25$ and $\delta = 1.5$.

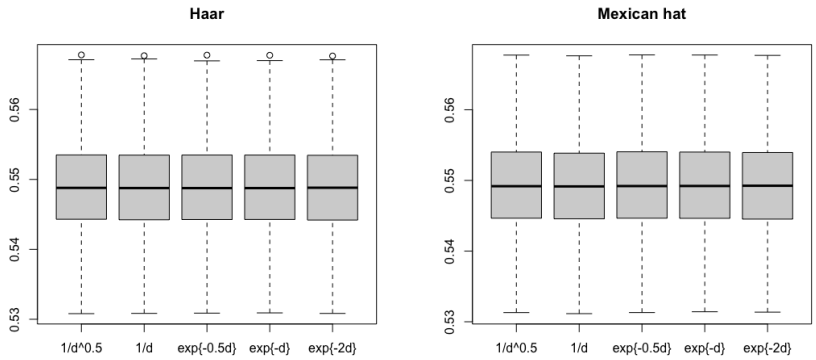


Figure B.32: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 0.5$.

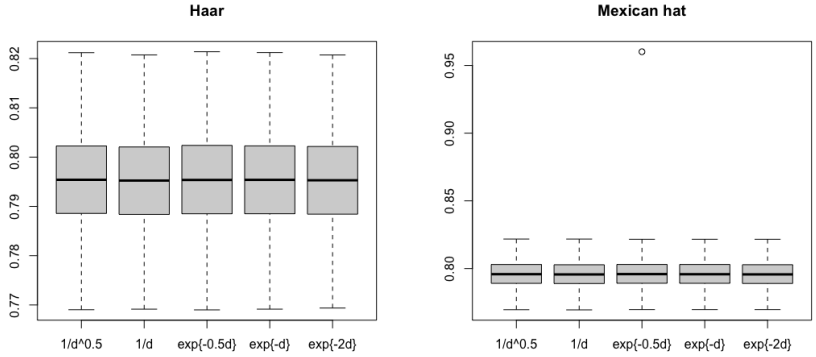


Figure B.33: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1$.

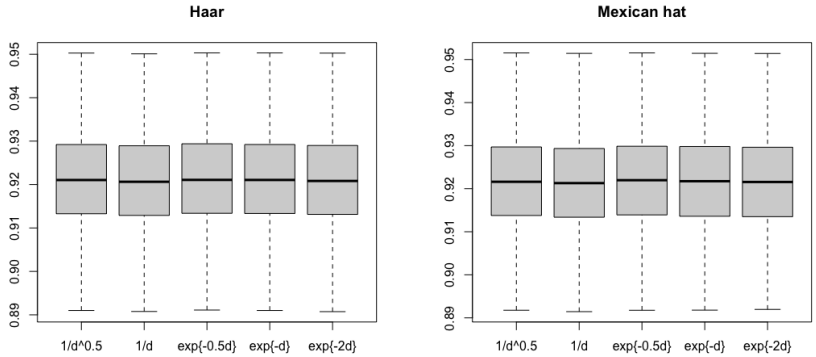


Figure B.34: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 0.5$ and $\delta = 1.5$.

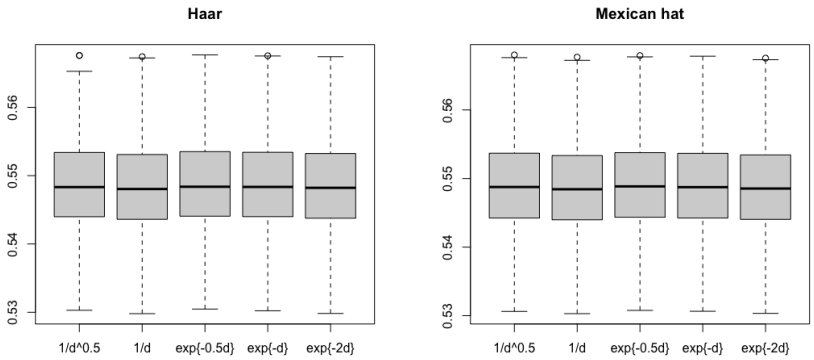


Figure B.35: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 0.5$.

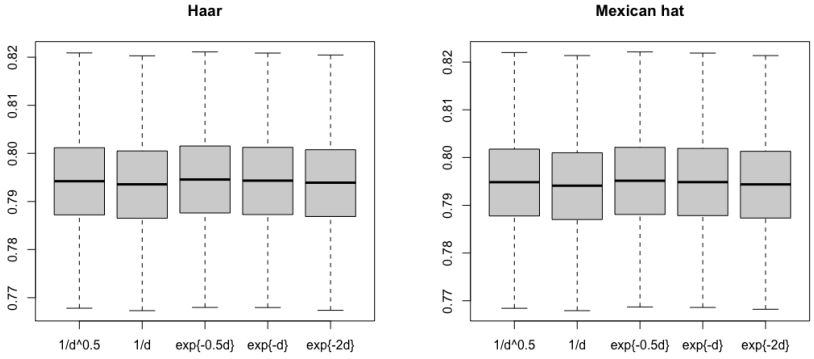


Figure B.36: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1$.

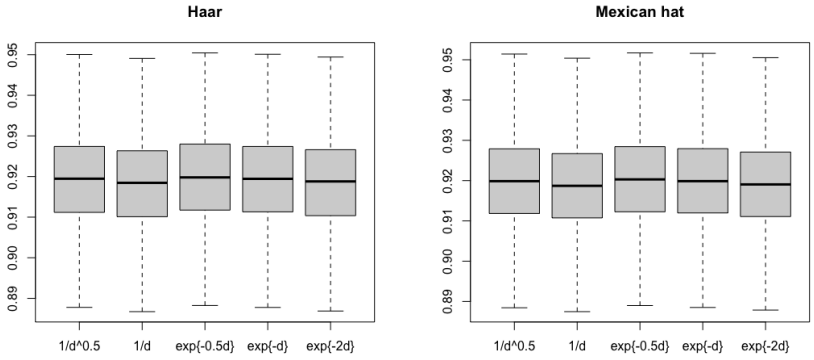


Figure B.37: Boxplots of MSEs obtained by tvSTARMA$(1_1, 1_1)$ model with $J = 2$ and different spatial weights matrices, which the dataset was simulated using $\gamma = 1$ and $\delta = 1.5$.

References

Andrews, D. W. K. (1988) Laws of large number for dependent non-identically distributed random variables, *Econometric Theory*, **4**(3), 458-467.

Biz, G. (2014) Simulações de pesos espaciais para o modelo STARMA e aplicações, tese de doutorado, Universidade de São Paulo Escola Superior de Agiricultura Luiz de Queiroz, Piracicaba, São Paulo.

Box, G. E. P., Jenkins, G. M, Reinsel, G. C. and Ljung, G. M. (2015) *Time Series Analysis: Forecasting and Control*, 5th Edition, Wiley.

Box, G. E. P. and Jenkins, G. M. (1970) *Time series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Brockwell, P. J. and Davis, R. A. (2016) *Introduction to Time Series and Forecasting*, Springer.

Bruce, A. and Gao, H.-Y. (1996) *Applied wavelet analysis with S-plus*, New York: Springer-Verlag.

Chiann, C. and Morettin, P. A. (1999) Estimation of time-varying linear systems, *Statistical Inference for Stochastic Processes* **2**, 253–285.

Chiann, C. and Morettin, P. A. (2005) Time domain nonlinear estimation of time varying linear systems, *Journal of Nonparametric Statistics*, **17**(3), 365–383.

Choi, Y. and Lee, S. (2000) Injectivity conditions of 2D and 3D uniform cubic B-spline functions, *Graphical Models*, **62**(6), 411–427.

Chun, S. Y. and Fessler, J. A. (2009) A simple regularizer for B-spline nonrigid image restration that encourages local invertibility, *IEEE journal of selected topics in signal processing* **3**(1), 159-169.

Cipra, T. and Motykova, I. (1987) Study on Kalman filter in time series analysis, *Commentationes Mathematicae Universitatis Carolinae*, **28**(3):549-563.

Cliff, A. D. and Ord, J. K. (1975) Space-Time Modeling with an Application to Regional Forecasting, *Transactions of the Institute of British Geographers*, No.64, 119-128.

Dahlhaus, R. (1996a) Maximum likelihood estimation and model selection for locally stationary processes, *Journal of Nonparametric Statistics*, **6**(2-3), 171-191.

Dahlhaus, R. (1996b) On the kullback-leibler information divergence of locally stationary processes, *Stochastic Processes and their Applications*, **62**(1), 139–168.

Dahlhaus, R. (1996c) Asymptotic statistical inference for nonstationary processes with evolutionary spectra, In: Robinson, P. M., Rosenblatt, M. (eds) *Athens Conference on Applied Probability and Time Series Analysis. Volume II: Time Series Analysis In Memory of E.J. Hannan*, **115**, 145–159. Springer, New York, NY.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes, *The Annals of Statistics*, **25**(1), 1–37.

Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes, *The Annals of Statistics*, **28**(6), 1762–1794.

Dahlahus, R. (2012) Locally stationary processes, *Time Series Analysis: Methods and Applications*, **30**, 351-413.

Dahlhaus, R., and Polonik, W. (2006) Nonparametric Quasi-Maximum Likelihood Estimation for Gaussian Locally Stationary Processes, *The Annals of Statistics*, **34**(6), 2790–2824.

Dahlahus, R., and Polonik, W. (2009) Empirical spectral processes for locally stationary time series, *Bernoulli*, **15**(1), 1–39.

Dahlhaus, R., Neumann, M. H. and von Sachs, R. (1999) Nonlinear wavelet estimation of time-varying autoregressive processes, *Bernoulli* **5**(5), 873–906.

Dalezios, N. R and Adamowski, K. (1995) Spatio-temporal precipitation modelling in rural watersheds, *Hydrological Sciences Journal*, **40**(5), 553-568.

Damian, D., Sampson, P. and Guttorp, P. (2001) Bayesian estimation of semiparametric nonstationary spatial covariance structures, *Environmentrics* **12**(2), 161-178.

Daubechies, I. (1992) *Ten Lectures on Wavelets*, SIAM, Philadelphia.

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41**(7), 909-996.

Daubechies, I. and Lagarias, J. C. (1991) Two-scale difference equations I: existence and global regularity of solutions, *SIAM Journal on Mathematical Analysis*, **22**, 1388–1410.

Daubechies, I. and Lagarias, J. C. (1992) Two-scale difference equations II: local regularity, infinite products of matrices and fractals, *SIAM Journal on Mathematical Analysis*, **23**, 1031–1079.

De Boor, C. (2001) *A Practical Guide to Splines*, Revised Edition, New York: Springer-Verlag.

De Iaco, S., Posa, D., Cappello, C. and Maggio, S. (2019) Isotropy, symmetry, separability and strict positive definiteness for covariance functions: A critical review, *Spatial Statistics*, **29**, 89–108.

Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D. (1995) Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society*, Series B (Methodological), **57**(2), 301-337.

Finley, A. O. (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence, *Methods in Ecology and Evolution*, **2**, 143-154.

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

Gneiting, T. and Schlather, M. (2013) Space-time covariance models, *In Encyclopedia of Environmetrics*, John Wiley & Sons.

Gneiting, T., Genton, M. G., Guttorp, P. (2006) Geostatistical space-time models, stationarity, separability, and full symmetry, *In Statistics of Spatio-Temporal Systems*, In: Finkenstaedt, B., Held, L., Isham, V. (eds), Chapman & Hall/CRC, New York.

Guinness, J. and Fuentes, M. (2016) Isotropic covariance functions on spheres: Some properties and modeling considerations, *Journal of Multivariate Analysis*, **143**, 143–152.

Guttorp, P., Meiring, W. and Sampson, P. D. (1994) A space-time analysis of ground-level ozone data, *Environmetrics* **5**(3), 241-254.

Haar, A. (1910) Zur Theorie der orthogonalen Funktionensysteme, *Mathematische Annalen*, **69**(3), 331–371.

Hu, X., Qin, Z. and Chu, F. (2011) Damage detection in plate structures based on spacetime autoregressive moving average processes, *Journal of Physics: Conference Series*, **305**(1), 012119.

Jin, E. Y. (2017) Estrutura de vizinhanças espaciais nos modelos autorregressivos e de médias móveis espaço-temporais STARMA, dissertação de mestrado, Universidade de São Paulo Instituto de Matemática e Estatística, São Paulo.

Jun, M. and Stein, M. L. (2008) Nonstationary covariance models for global data, *The Annals of Applied Statistics*, **2**(4), 1271–1289.

Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized Additive Models*, Chapman and Hall, New York.

Kalman, R. (1960) A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, **82**(1), 35-45.

Kamarianakis, Y. and Prastacos, P. (2005) Space–time modelling of traffic flow, *Computers & Geosciences*, **31**(2), 119-133.

Kim, J. (2004) Intensity Based Image Registration Using Robust Similarity Measure and Constrained Optimization: Applications for Radiation Therapy, Ph.D. dissertation, University of Michigan, Ann Arbor, MI.

Kurt, S. and Tunay, K. B. (2015) STARMA Models Estimation with Kalman Filter: The Case of Regional Bank Deposits, *Procedia - Social and Behavioral Sciences*, **195**, 2537-2547.

Kyriakidis, P. C. and Journel, A. G. (1999) Geostatistical space–time models: a review, *Mathematical Geology*, **31**, 651–84.

Ma, C. (2003) Spatio-temporal stationary covariance models *Journal of Multivariate Analysis*, **86**(1), 97–107.

Martin, R. L. and Oeppen, J. E. (1975) The identification of regional forecasting models using space: time correlation functions, *Transactions of the Institute of British Geographers*, No.66, 95-118.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.-M. (2007) *Wavelets and their Applications*, Wiley.

Morettin, P. A. e Toloi, C. (2018) *Análise de Séreis Temporais*, Terceira Edição, Blucher Editora.

Morettin, P. A. (2014) *Ondas E Ondaletas: Da Análise de Fourier à Análise de Ondaletas de Séries Temporais*, EDUSP.

Musse, O. Heitz, F. and Armspach, J. (2001) Topology preserving deformable image matching using constrained hierarchical parametric models, *IEEE Transactions on Image Processing*, **10**(7), 1081–1093.

Ogden, R. D. (1997) *Essential wavelets for statistical applications and data analysis*, Springer.

Pace, R. K., Barry, R., Gilley, O. W. and Sirmans, C. F. (2000) A method for spatial-temporal forecasting with an application to real estate prices, *International Journal of Forecasting*, **16**(2), 229-246.

Pfeifer, P. E. and Deutsch, S. J. (1980a) A three-stage iterative procedure for space-time modeling, *Technometrics*, **22**(1), 35-47.

Pfeifer, P. E. and Deutsch, S. J. (1980b) Identification and Interpretation of First-Order Space-Time ARMA Models, *Technometrics*, **22**(3), 397-408.

Pfeifer, P. E. and Deutsch, S. J. (1981a) Variance of the Sample-Time Autocorrelation Function of Contemporaneously Correlated Variables, *SIAM Journal of Applied Mathematics*, **40**(1), 133-136.

Pfeifer, P. E. and Deutsch, S. J. (1981b) Seasonal Space-Time ARIMA modeling, *Geographical Analysis*, **13**(2), 117-133.

Pfeifer, P. E. and Deutsch, S. J. (1981c) Space-Time ARMA Modeling with contemporaneously correlated innovations, *Technometrics*, **23**(4), 410-409.

Pinheiro, A. and Vidakovic, B. (1997) Estimating the square root of a density via compactly supported wavelets, *Computational Statistics & Data Analysis*, **25**(4), 399-415.

Priestley, M. B. (1965) Evolutionary Spectra and Non-Stationary Processes, *Journal of the Royal Statistical Society*, Series B (Methodological), **27**(2), 204–237.

Ramsay, J. O. (1982) When the data are functions, *Psychometrika*, **47**, 379–396.

Ramsay, J. O. and Dalzell, C. J. (1991) Some tools for functional data analysis, *Journal of Royal Statistical Society*, Series B, **53**(3), 539-572.

Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data Analysis*, New York: Springer.

Ramsay, J. O. and Silverman, B. W. (2006) *Functional Data Analysis*, 2nd Edition, New York: Springer.

Ramsay, J. O. (1998) Estimating smooth monotone functions. *Journal of Royal Statistical Society*, Series B, **60**(2), 365-375.

Rao, T. S. and Antunes, A. M. C. (2004) Spatio-temporal Modelling of Temperature Time Series: A Comparative Study. In: Brillinger, D. R., Robinson, E. A., Schoenberg, F. P. (eds), *Time Series Analysis and Applications to Geophysical Systems The IMA Volumes in Mathematics and its Applications*, **139**, Springer, New York, NY.

Reich, B. J., Eidsvik, J., Guindani, M., Nail, A. J. and Schmidt, A. M. (2011) A class of covariate-dependent spatiotemporal covariance functions, *The Annals of Applied Statistics*, **5**(4), 2265–2687.

Rohan, N. and Ramanathan T. V. (2013) Nonparametric estimation of a time-varying GARCH model, *Journal of Nonparametric Statistics*, **25**(1), 33-52.

Sampson, P. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure, *Jounal of the American Statistical Association*, **87**(417), 108-119.

Sato, J. R., Morettin, P. A., Arantes, P. R. and Amaro Jr, E. (2007) Wavelet based time-varying vector autoregressive modelling, *Computational Statistics & Data Analysis*, **51**(12), 5847-5866.

Schmidt, A. M. and O'Hagan, A. (2003) Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society*, Series B, **65**(3), 743-758.

Shumway, R. H. and Stoffer, D. S. (2017) *Time Series Analysis and Its Applications With R Examples*, Springer.

Stéphane, M. (2009) *A Wavelet Tour of Signal Processing*, Third Edition, Academic Press.

Wang, S., Feng, J. and Liu, G. (2013) Application of seasonal time series model in the precipitation forecast, *Mathematical and Computer Modelling*, **58**(3–4), 677-683.

White, H., (2000) Asymptotic Theory for Econometricians, revised edition, Academic Press, New York.

Wikle, C. K., Zammit-Mangion, A and Cressie, N. (2019) *Spatio-Temporal Statistics With R*, Chapman & Hall/CRC, Boca Raton, FL.

Wu, X., Zhou, J., Yu, H., Liu, D., Xie, K., Chen, Y., Hu, J., Sun, H. and Xing, F. (2021) The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis, *Atmosphere*, **12**(1):74.

Yousuf, K. and Ng, S. (2021) Boosting high dimensional predictive regressions with time varying parameters, *Journal of Econometrics*, **224**(1), 60-87.

Zhou, X. and Lin, H. (2008) Spatial Weights Matrix, In: Shekhar, S., Xiong, H. (eds), *Encyclopedia of GIS*, Springer, Boston, MA.

Zou, J., Zhu, J., Xie, P., Xuan, P. and Lai, X. (2018) A STARMA Model for Wind Power Space-Time Series, *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 1-5.