Wavelet-based estimation for Gaussian time series and spatio-temporal processes

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Wenjun Zheng, B.S.

Graduate Program in Statistics

The Ohio State University

2014

Dissertation Committee:

Peter Craigmile, Advisor
Mark Berliner
Catherine Calder
Abstract

Modern statistical analyses often require the modeling non-Markov time series and spatio-temporal dependencies. Traditional likelihood methods are computationally demanding for these models, leading us to consider approximate likelihood methods that are computationally efficient, while not overly compromising on the efficiency of the parameter estimates. In this dissertation various wavelet-based Whittle approximations are investigated to model a certain class of nonstationary Gaussian time series and a class of Gaussian spatio-temporal processes. Wavelet transforms can help decorrelate processes across and within wavelet scales, allowing for the simplified modeling of time series and spatio-temporal processes. In addition to being computationally efficient, the proposed maximum wavelet-Whittle likelihood estimators of a Gaussian process are shown to be asymptotically normal. Asymptotic properties of the estimators are verified in simulation studies, demonstrating that the typical independence everywhere assumption assumed for wavelet-based estimation is not optimal. These methods are applied to the analyses of a Southern Oscillation Index climate series and the Irish wind speed data.

Keywords: Approximate likelihood, between and within-scale decorrelation, discrete wavelet transform, Markov approximations, Matérn processes, spatio-temporal processes.
Acknowledgments

First and foremost, I would like to express my grateful thanks to my advisor, Dr. Peter Craigmile for his patient guidance during my PhD study. I am fortunate to have a great advisor who is enthusiastic about research, and cares about my work. I am grateful to my committee members, Dr. Mark Berliner and Dr. Catherine Calder, for their assistance throughout my dissertation. I acknowledge the NSF grant, DMS-0906864, for support as a Graduate Research Assistant in 2012.

Finally, I am deeply thankful to my family and all my friends for their support and encouragement over the years. I am grateful to my husband, Shuchi Wu, for sacrificing his valuable time to proofread my dissertation. I also thank my son Dylan for being a good baby, making it possible for me to write and finish my dissertation in his first few months of life.
Vita

2008 ........................................... B.S. Statistics, Wuhan University

2008-present ............................. Graduate Teaching Associate, Department of Statistics, The Ohio State University.

Fields of Study

Major Field: Statistics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iii</td>
</tr>
<tr>
<td>Vita</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Approximation models</td>
<td>2</td>
</tr>
<tr>
<td>1.2 The wavelet Whittle method</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Extension to space-time processes</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Layout of the dissertation</td>
<td>6</td>
</tr>
<tr>
<td>2. Discrete wavelet transforms of time series processes</td>
<td>8</td>
</tr>
<tr>
<td>2.1 The Gaussian processes of interest</td>
<td>8</td>
</tr>
<tr>
<td>2.2 The discrete wavelet transforms of a time series</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Statistical properties of the wavelet coefficients</td>
<td>19</td>
</tr>
<tr>
<td>2.3.1 Expected value of the DWT coefficients</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2 Covariance properties</td>
<td>20</td>
</tr>
<tr>
<td>2.3.3 Spectral representations of the DWTs</td>
<td>23</td>
</tr>
<tr>
<td>2.4 An example</td>
<td>28</td>
</tr>
<tr>
<td>2.5 Conclusion</td>
<td>33</td>
</tr>
</tbody>
</table>
3. Asymptotic properties of the wavelet-Whittle estimators ............... 37
   3.1 Wavelet-based Whittle approximations .................................. 37
   3.2 A white noise model .......................................................... 40
   3.3 The proof of the white noise results .................................... 43
   3.4 An AR(1) model ..................................................................... 53
     3.4.1 The approximate likelihood .............................................. 53
     3.4.2 Discussion about the MLEs under the AR(1) model ............... 54
   3.5 Choices of $J_0$ ..................................................................... 58
   3.6 Monte Carlo studies ............................................................... 59
     3.6.1 The wavelet-based Whittle MLEs ....................................... 59
     3.6.2 Comparison to other approximate methods ......................... 63
     3.6.3 The limiting variance ....................................................... 65
   3.7 Southern Oscillation Index data .............................................. 67
   3.8 Conclusion ............................................................................. 73

4. The DWTs of space-time processes ............................................. 76
   4.1 A review of space-time processes ............................................ 77
     4.1.1 The statistical properties .................................................. 77
     4.1.2 The spectral representations ............................................. 80
   4.2 The DWT of a space-time process .......................................... 82
     4.2.1 Simplifying assumptions for the space-scale-time wavelet pro-
          cesses ............................................................................. 86
     4.2.2 Spectral representations of the DWT coefficients ................. 91
     4.2.3 Examples of the whitening effects ..................................... 95

5. Wavelet Whittle models for space-time processes ....................... 107
   5.1 Wavelet Whittle approximations for space-time processes ......... 107
     5.1.1 A white noise model ....................................................... 109
     5.1.2 An AR(1) model ............................................................ 110
   5.2 Asymptotic results of the wavelet-Whittle estimators ............... 112
   5.3 Proof of Theorem 5.1 and Theorem 5.2 .................................... 113
   5.4 Monte Carlo studies .............................................................. 130
   5.5 Application: Irish wind data ................................................... 133

6. Discussion and future work ....................................................... 141
   6.1 Conclusion and discussion ...................................................... 141
   6.2 Future work ........................................................................ 144
Appendix

A. Proof of the asymptotic normality of the approximate MLEs in the MODWT setting

Bibliography
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Matérn process: results for $\hat{\phi}$ using wavelet coefficients on Levels 3 to $J$</td>
<td>61</td>
</tr>
<tr>
<td>3.2 Matérn process: results for $\hat{\sigma}^2$ using wavelet coefficients on Levels 3 to $J$</td>
<td>61</td>
</tr>
<tr>
<td>3.3 Matérn process: results for $\hat{\phi}$ using wavelet coefficients on Levels 1 to $J$</td>
<td>62</td>
</tr>
<tr>
<td>3.4 Compare simulation time</td>
<td>62</td>
</tr>
<tr>
<td>3.5 Simulation results (FD process)</td>
<td>65</td>
</tr>
<tr>
<td>3.6 Limiting variance: all coefficients</td>
<td>66</td>
</tr>
<tr>
<td>3.7 Limiting variance: non-boundary coefficients</td>
<td>66</td>
</tr>
<tr>
<td>5.1 Simulation results of the temporal scaling parameter $a.$</td>
<td>132</td>
</tr>
<tr>
<td>5.2 Simulation results of the spatial scaling parameter $c.$</td>
<td>132</td>
</tr>
<tr>
<td>5.3 Simulation time for the space-time estimation (in seconds)</td>
<td>133</td>
</tr>
<tr>
<td>5.4 Estimation results for the wind speed covariance model</td>
<td>138</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>DWT filters</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>DWT squared gain functions</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>$LA(8)$ squared gain functions</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>DWT decorrelation example - SDF</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>DWT decorrelation example (within-scale) with the $LA(8)$ filter</td>
<td>31</td>
</tr>
<tr>
<td>2.6</td>
<td>DWT decorrelation example (cross-scale) with the $LA(8)$ filter</td>
<td>32</td>
</tr>
<tr>
<td>2.7</td>
<td>DWT decorrelation example (within-scale) with the $D(6)$ filter</td>
<td>33</td>
</tr>
<tr>
<td>2.8</td>
<td>DWT decorrelation example (cross-scale) with the $D(6)$ filter</td>
<td>34</td>
</tr>
<tr>
<td>2.9</td>
<td>DWT decorrelation example (within-scale) with the $D(4)$ filter</td>
<td>35</td>
</tr>
<tr>
<td>2.10</td>
<td>DWT decorrelation example (cross-scale) with the $D(4)$ filter</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Monthly SOI signal</td>
<td>68</td>
</tr>
<tr>
<td>3.2</td>
<td>SOI wavelet spectra</td>
<td>69</td>
</tr>
<tr>
<td>3.3</td>
<td>PACF of the standardized SOI</td>
<td>71</td>
</tr>
<tr>
<td>3.4</td>
<td>Fitted spectrum</td>
<td>74</td>
</tr>
<tr>
<td>3.5</td>
<td>Fitted spectrum - confidence bands</td>
<td>75</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>4.1</td>
<td>Space-time process between-scale covariances, with LA(8) filter and $</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Space-time process between-scale covariances, with LA(8) filter and $</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Space-time process between-scale covariances, with LA(8) filter and $</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>Space-time separable ACVS, LA(8) filter</td>
<td>101</td>
</tr>
<tr>
<td>4.5</td>
<td>Space-time separable correlation, LA(8) filter</td>
<td>102</td>
</tr>
<tr>
<td>4.6</td>
<td>Space-time nonseparable ACVS, LA(8) filter</td>
<td>103</td>
</tr>
<tr>
<td>4.7</td>
<td>Space-time nonseparable correlation, LA(8) filter</td>
<td>104</td>
</tr>
<tr>
<td>4.8</td>
<td>Space-time nonseparable ACVS with $\beta = 0.7$, LA(8) filter</td>
<td>105</td>
</tr>
<tr>
<td>4.9</td>
<td>Space-time process nonseparable ACVS with $\beta = 0.1$, LA(8) filter</td>
<td>106</td>
</tr>
<tr>
<td>5.1</td>
<td>Wind data wavelet spectra - 1</td>
<td>135</td>
</tr>
<tr>
<td>5.2</td>
<td>Wind data wavelet spectra - 2</td>
<td>136</td>
</tr>
<tr>
<td>5.3</td>
<td>Wind data wavelet spectra - 3</td>
<td>137</td>
</tr>
<tr>
<td>5.4</td>
<td>Wind data space-time correlation</td>
<td>139</td>
</tr>
<tr>
<td>5.5</td>
<td>Wind data cross (station) correlation</td>
<td>140</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Time series are collected in disciplines such as economics, environmental sciences, health sciences and geophysics. In time series analysis, researchers are interested in the statistical analysis of data observed over time. Modeling time series data helps people predict future outcomes and study the characteristics (statistical properties) of the process generating the data. One useful summary of a process, especially of a Gaussian process is its covariance function. Traditional statistical models fit to time series include autoregressive (AR) models, moving average (MA) models, and integrated models [e.g., Brockwell and Davis, 2002].

One challenge in analyzing time series data is to model lengthy processes with non-Markov dependencies. Typical examples of non-Markov are processes with long range dependence [e.g. Beran, 1994, Palma, 2007], certain Matérn covariance models [e.g. Cressie, 1993], and exponential processes [Bloomfield, 1973]. For non-Markov statistical models the computational complexity of the traditional exact likelihood method is challenging. Instead, for ease of computation, approximate methods often assume that the process follows a simpler approximate statistical model. Then, statistical inferences such as parameter estimation and forecasting are performed under this assumed model. The question we ask in this dissertation is can we obtain estimators of the statistical process model parameters from a wavelet-based approximate
model, but to obtain the asymptotic statistical properties of the estimator under the true model for the process.

1.1 Approximation models

Markov approximations, Whittle approximations, wavelet transforms and covariance tapering are popular approximation methods for analyzing time series and spatial processes. The Markov approximation (or autoregressive, AR approximation, e.g., Brockwell and Davis 2002) assumes the current outcome of a time series depends linearly on the previous \( p \) outcomes (an AR model of order \( p \)). Standard time series algorithms (e.g., Levinson-Durbin and Innovation algorithms, e.g., Brockwell and Davis 2002) speed up computation under a Markov approximation. However we may worry about using Markov models to approximate non-Markov dependence, because the important long-range dependencies are ignored by Markov approaches. For stationary processes, a spectral approach via the periodogram decomposing a time series into a series of sinusoids at different frequencies can be used to approximately estimate the time series parameters. Assuming independence of the periodogram ordinates leads to spectral-Whittle methods that are computationally efficient, but can be limited in their ability to analyze non-stationary processes without appropriate pre-processing of the data [Whittle, 1953, Dahlhaus, 1983].

More recently, covariance tapering [Furrer et al., 2006, Kaufman et al., 2008] has been introduced to improve the computational efficiency, yet preserve the asymptotic properties of the estimation and prediction results. In the basic tapering method we replace the covariance matrix in the likelihood calculation by a sparse matrix which is the direct product of the true covariance and a compactly-supported covariance
function called the taper. This method has been shown to simplify computation (with a computation cost of order $n$), and has been shown to be asymptotically unbiased for estimating certain parameters of spatial process of interest under an infill asymptotic. A modified two-taper method tapers both the model and sample covariance matrices and has an unbiased score function [Kaufman et al., 2008]. Both one-taper and two-taper methods are easier to compute because of the sparsity of the covariance matrix. It has been shown that the resulting approximate maximum likelihood estimators (MLEs) are consistent under certain conditions. We will, in this dissertation, implement the tapering method for estimation of time series parameters, and compare it with the proposed wavelet-Whittle method.

1.2 The wavelet Whittle method

The discrete wavelet transform (DWT) of a time series [Percival and Walden, 2000] produces a collection of wavelet and scaling coefficients on different temporal scales that are less correlated than the original time series process [e.g. Tewfik and Kim, 1992, Flandrin, 1992, Dijkerman and Mazumdar, 1994, Craigmile and Percival, 2005]. It can be shown that under certain conditions the wavelet coefficients are approximately uncorrelated between scales. An approximate likelihood is then obtained by assuming independence between wavelet coefficients on different time scales [e.g. McCoy and Walden, 1996, Craigmile and Percival, 2005, Moulines et al., 2008]. For certain long memory processes, the estimators of long memory parameters based on an approximate wavelet model has been shown to be consistent, and asymptotically normal [Moulines et al., 2008].
The work described in the following chapters uses DWT to model the class of Gaussian processes whose $d$th order backward differences are stationary. By taking the DWT of these processes, the resulting wavelet coefficients are decorrelated. The within scale coefficients and certain cross scale process [Moulines et al., 2008] are shown to be stationary if the length of the wavelet filter is long enough, and the spectrum of the original process is well defined. Because the DWT is an orthonormal transform, the exact likelihood of the Gaussian time series process can be written in terms of the wavelet coefficients and wavelet covariances. With simpler approximate models (such as white noise and AR(1)) for the wavelet coefficients on different scales, we can approximate the true likelihood. We are interested in jointly estimating all the parameters of the time series model, not just a subset. For example, in a model with long and short range dependencies, we estimate all the parameters [e.g. Jensen, 2000], rather than only estimating the long range parameters [e.g. Moulines et al., 2008]. Moreover, the existing theory of these ML estimators are derived under the assumption that the approximate model is the true model [e.g. Kaufman et al., 2008, Craigmile et al., 2005]. Our theory is, on the other hand, derived under the true process.

We obtain different wavelet-based methods of estimation by assuming different statistical models for the wavelet coefficients both within and between wavelet scales. While there has been some work on the development of wavelet-based estimation methods for certain long memory processes [e.g., McCoy and Walden, 1996, Craigmile et al., 2005, Jensen, 2000, Moulines et al., 2008, Vanucci and Corradi, 1999], in this dissertation we will study the still open question of how different wavelet-based models perform for the estimation of more general non-Markov models.
1.3 Extension to space-time processes

A natural extension of these approximate wavelet methods is to apply them on space-time processes (for a review of space-time processes see, e.g., Banerjee et al. 2004, Le and Zidek 2006, Cressie and Wikle 2011). Environmental data are often observed from a finite number of locations over a long period of time [e.g., Stein, 2005]. Researchers model space-time data to estimate the parameters of the process and to predict values at unobserved locations (kriging), which is computationally demanding because of the high dimensional covariance matrix. Approximation methods for space and space-time processes include Markov random fields, tapering, wavelet, and low rank approaches [e.g., Cressie, 1993, Gelfand et al., 2010]. In existing wavelet approaches, wavelet bases are used to represent the stationary or non-stationary structure of a spatial process. A threshold is set up to allow a limited proportion of nonzero element in the covariance matrix. Then a discrete wavelet transform approach are used to approximate the covariance function. Methods have been developed for the spatial processes observed on a grid, and for the ones that are irregular [Wikle et al., 2001, Nychka et al., 2003, Matsuo et al., 2011].

While these work are performing 2-D wavelet transforms in the spatial domain, our interest is in performing DWT in the temporal domain, which gives an asymptotic theory for estimators of parameters characterizing temporal dependence. We will define the DWT of a space-time process, and discuss its statistical properties. For space-time processes, common simplifying assumptions such as stationary, symmetric, and separable are made upon the covariance functions [e.g., Gneiting et al., 2007, Cressie and Wikle, 2011]. We will show in this dissertation that the DWT of a spatial-temporal process preserve certain properties from the original space-time process. 

In
addition, we will present the stationarity results of the within-scale and a cross-scale wavelet coefficients.

The spectral representation of the DWTs of a space-time process can be obtained from Bochner’s theorem, which provides a basic spectral representation of a space-time process [Cressie and Huang, 1999]. Stein [2005] and Cressie and Wikle [2011] discussed two different decompositions for the spectral density function (SDF) of a space-time process with a nonseparable covariance function. The decomposed spectral representations will be extended to the statistical processes for the DWT coefficients. With a thorough understanding of the statistical properties of the wavelet coefficients, one can implement the wavelet Whittle method on space-time processes and perform statistical inference. The exact likelihood of a space-time process can be approximated by the wavelet likelihood as for the univariate case. The decorrelation effects of the DWT enables us to apply a Whittle approximation on every wavelet scale. We will build wavelet Whittle-based white noise and AR(1) models, and will provide the asymptotic theory for the maximum likelihood estimators under the white noise approximate model.

1.4 Layout of the dissertation

In this dissertation, the wavelet-based Whittle approximations are presented to model Gaussian time series. In Chapter 2 we introduce the DWT of a time series and discuss the statistical properties of wavelet coefficients in both time domain and spectral domain. The idea of the wavelet Whittle likelihood approximation is discussed in the same chapter. In Chapter 3, the asymptotic properties of our wavelet-based approximate maximum likelihood estimators are presented and proved. Monte Carlo
studies are then used to explore the finite sample properties of the wavelet Whittle estimators, and to compare their performance with other estimators based on tapering and Markov approximations. A real data analysis on a Southern Oscillation Index dataset is used to illustrate the application of the time series methods. We define in Chapter 4 the DWT of a space-time process and derive the statistical properties of the resulting wavelet and scaling coefficients. Chapter 5 presents the wavelet Whittle approximate model, the asymptotic theory for the MLEs, and performs the Whittle estimation for space-time processes. We end with some conclusions and a discussion of future work in Chapter 6.
Chapter 2: Discrete wavelet transforms of time series processes

In this chapter we specify the general class of Gaussian time series processes of interest, which are those that have stationary backward differences. We briefly discuss the spectral representation of such processes. Then, the discrete wavelet transform (DWT) of a time series is defined and the wavelet coefficients’ statistical properties are presented. Two propositions on the important decorrelation features of the DWT are stated. At the end of this chapter, a Gaussian process with the Matérn covariance function is used to illustrate the decorrelation property.

2.1 The Gaussian processes of interest

Consider a real-valued univariate Gaussian process \( \{Y_t : t \in \mathbb{Z}\} \) with \( d \)th order stationary increments [Yaglom, 1958], defined as follows. Let \( \nabla \) denote the lag one differencing operator defined by \( \nabla Y_t = Y_t - Y_{t-1} \), with powers of \( \nabla \) defined recursively by \( \nabla^0 Y_t = Y_t \) and \( \nabla^k Y_t = \nabla^{k-1} [\nabla Y_t] \), for a positive integer \( k \). For a non-negative integer \( d \), we assume that the \( d \)th order difference of \( Y_t \), namely \( X_t = \nabla^d Y_t \), is stationary. Without loss of generality suppose that \( X_t \) has mean zero with an autocovariance sequence (ACVS) of \( \{\gamma_X(h) = \text{cov}(X_t, X_{t+h}) : h \in \mathbb{Z}\} \).
With the spectral representation of a stationary process, we can analyze the process in the frequency domain. Herglotz theorem defines the spectral distribution and the spectral density function (see, e.g., Brockwell and Davis [1991], page 117).

**Theorem 2.1** (Herglotz). Assume that the sampling interval $\Delta = 1$. A complex-valued function $\gamma(\cdot)$ defined on the integers is non-negative definite if and only if

$$
\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi fh} dS(f) \quad \text{for all } h = 0, \pm 1, \ldots,
$$

where $S(f) \in [-1/2, 1/2]$ and $S(0) = 0$. The function $S(f)$ is called the spectral distribution function of $\gamma(\cdot)$ and if $S(f) = \int_{-\lambda}^{\lambda} S(f)df, \ -1/2 \leq \lambda \leq 1/2$, then $S$ is called a spectral density of $\gamma(\cdot)$.

One important feature of the distribution function $S(f) \in [-1/2, 1/2]$ is that it can be uniquely defined by the function $\gamma(h)$ (usually an autocovariance function). When the spectral density exists, the autocovariances and the spectrum are a Fourier transform pair (written as $\{\gamma(h)\} \leftrightarrow S(\cdot)$; Percival and Walden [2000], page 23) with

$$
S(f) = \sum_h \gamma(h)e^{-i2\pi fh} \quad \text{and} \quad \gamma(h) = \int_{-1/2}^{1/2} S(f)e^{i2\pi fh}df.
$$

Therefore the general class of Gaussian process defined earlier has an associated spectral density function (SDF) of

$$
S_X(f) = \Delta \sum_{h \in \mathbb{Z}} e^{-i2\pi fh} \gamma_X(h), \quad f \in [-F, F].
$$

In the above equation $\Delta$ is the sampling interval for $Y_t$ and $F = 1/(2\Delta)$ is the Nyquist frequency. Then it can be shown [e.g., Yaglom, 1987] that the process $Y_t$ has a generalized SDF of

$$
S_Y(f) = \Delta |4\sin^2(\pi f)| S_X(f), \quad f \in [-F, F].
$$
Extending this definition so that \( d \) lies over the positive real line we obtain long range dependence and with \( d \) negative we obtain a class of antipersistent processes [e.g., Granger and Joyeux, 1980, Hosking, 1981, Beran, 1994]. Autoregressive fractionally integrated moving average (ARFIMA) models, and fractionally exponential processes are all included in this class of processes. Matérn processes are also included [e.g., Chang and Stein, 2013]. In this dissertation we suppose that the SDF \( S_Y(f) \) is parametrized by a parameter vector \( \theta \in \Theta \), where \( \Theta \) is a compact set in \( \mathbb{R}^p \) (which could include \( d \)). To make this clear we write the SDF as \( S_{Y,\theta}(f) \). Our interest is in estimating \( \theta \) based on approximate likelihoods calculated using a wavelet transform of \( Y_t \). Without loss of generality we assume that \( \Delta = 1 \).

### 2.2 The discrete wavelet transforms of a time series

The DWT of a time series is an orthonormal transform that maps the time series into the wavelet domain as wavelet and scaling coefficients on different time scales [Percival and Walden, 2000]. In the transformation, data are linearly filtered and downsampled by a factor of two, i.e., we only keep every second filtered value. Given a length \( L \) wavelet filter \( \{h_l : l = 0, \ldots, L - 1\} \) and a scaling filter \( \{g_l : l = 0, \ldots, L - 1\} \), we define the level 1 wavelet filters as \( h_{1,l} \equiv h_l \), \( g_{1,l} \equiv g_l = (-1)^{l+1}h_{L-1-l} \). The corresponding transfer functions (the Fourier transform of the filter) are

\[
H_1(f) \equiv \sum_{l=0}^{L-1} h_{1,l} e^{-i2\pi fl}, \quad f \in \left[ -\frac{1}{2}, \frac{1}{2} \right],
\]

and

\[
G_1(f) \equiv \sum_{l=0}^{L-1} g_{1,l} e^{-i2\pi fl}, \quad f \in \left[ -\frac{1}{2}, \frac{1}{2} \right].
\]
The level 1 transfer functions have the relationship: \( G_1(f) = e^{-i2\pi f(L-1)}H_1(\frac{1}{2} - f) \) [Percival and Walden, 2000, page 76]. For levels \( j = 2, 3, \ldots \), the wavelet and scaling filters \( \{h_{j,l}\} \) and \( \{g_{j,l}\} \) are obtained by the inverse discrete Fourier transform (DFT) of products of transfer functions [Percival and Walden, 2000, page 96-97]:

\[
H_j(f) \equiv H_1(2^j - 1) \prod_{l=0}^{j-2} G_1(2^l f),
\]

\[
G_j(f) \equiv \prod_{l=0}^{j-1} G_1(2^l f).
\]

The width of a level \( j \) wavelet/scaling filter is \( L_j = (2^j - 1)(L - 1) + 1 \). Alternatively, we can calculate the level \( j \) filter recursively from the level \( j-1 \) filters. Given \( \{g_{j-1,l}\} \), the next level filters \( \{h_{j,l}\} \), and \( \{g_{j,l}\} \) follow from

\[
h_{j,l} = \sum_{k=0}^{L_j-1} h_k g_{j-1,l-2^{j-1}k},
\]

\[
g_{j,l} = \sum_{k=0}^{L_j-1} g_k g_{j-1,l-2^{j-1}k},
\]

for \( l = 0, \ldots, L_j - 1 \) [Percival and Walden, 2000, page 102].

The wavelet filter \( \{h_{j,l}, l = 0, 1, \ldots, L_j - 1\} \) and scaling filter \( \{g_{j,l}, l = 0, 1, \ldots, L_j - 1\} \) for levels \( j = 1, \ldots, J \) have the following properties [Percival and Walden, 2000].

1. The level \( j \) wavelet filter \( \{h_{j,l}\} \) and the transfer function \( H_j(\cdot) \) constitute a Fourier transform pair [Percival and Walden, 2000, Chapter 2]: \( \{h_{j,l}\} \leftrightarrow H_j(\cdot) \).

   A similar result is also true for the level \( j \) scaling filter \( \{g_{j,l}\} \) and the corresponding transfer function \( G_j(\cdot) \): \( \{g_{j,l}\} \leftrightarrow G_j(\cdot) \).

2. The summation of the level \( j \) wavelet/scaling filters satisfy: \( \sum_{l=0}^{L_j-1} h_{j,l} = 0 \) and \( \sum_{l=0}^{L_j-1} g_{j,l} = 2^{j/2} \).
3. The level $j$ wavelet filters are orthogonal to even shifts: \( \sum_l h_{j,l} h_{j,l+2m} = 0 \), when \( m \) is a nonzero integer.

4. The level $j$ wavelet and scaling filters each have unit energy: \( \sum_{l=0}^{L_j-1} h_{j,l}^2 = 1 \) and \( \sum_{l=0}^{L_j-1} g_{j,l}^2 = 1 \).

5. In practice, the level $j$ wavelet filter \( \{h_{j,l}\} \) is approximately a band-pass filter with pass-band \( 1/2^{j+1} \leq |f| \leq 1/2^j \) (approximately passes only the frequencies within this range); the level $j$ scaling filter \( \{g_{j,l}\} \) is approximately a band-pass filter with pass-band \( 0 \leq |f| \leq 1/2^{j+1} \).

To further determine the filters, we also define the phase function and the squared gain function of a filter [Percival and Walden, 2000, Chapter 2]. In the polar representation of the transfer function, \( H_j(f) \),

\[
H_j(f) = |H_j(f)| e^{i\theta_j(f)}.
\]

In the above equation \( \theta_j(f) \) is called the phase function of the filter, and \( |H_j(f)| \) is the gain function of the filter. We let \( \mathcal{H}_j(f) \equiv |H_j(f)|^2 \) and \( \mathcal{G}_j(f) \equiv |G_j(f)|^2 \) denote the squared gain functions for respectively the level $j$ wavelet and the level $j$ scaling filters.

Two classes of filters are usually employed: the Daubechies extremal phase filter and the least asymmetric filter [Daubechies, 1992], which can produce DWTs whose wavelet coefficients can be interpreted as changes of averages of different scales. With a fixed \( L \), the extremal phase filter and the least asymmetric filter share the same squared gain functions. The squared gain function for the level 1 wavelet filter is

\[
\mathcal{H}_1^{(D)}(f) \equiv 2 \sin^2(\pi f) \sum_{l=0}^{L/2-1} \left( \frac{L}{2} - 1 + l \right) \cos^2(\pi f),
\]
and the squared gain function for the level 1 scaling filter is
\[ G_1^{(D)}(f) \equiv 2 \cos^L(\pi f) \sum_{l=0}^{L/2-1} \left( \frac{L}{2} - 1 + l \right) \sin^2(\pi f), \]

In the function \( H_1^{(D)}(f) \), \( \sin^L(\pi f) \) indicates that the filter is an order \( L/2 \) difference filter. (The difference filter \( \{-1, 1\} \) has the transfer function \( D(f) = 1 - e^{-i2\pi f} \), and the squared gain function \( D(f) = 4 \sin^2(\pi f) \), Craigmile and Percival [2005].) Given a squared gain function, different spectral factorizations [Percival and Walden, 2000, Chapter 11] are used to obtain different filters. The following conditions have been set up to uniquely determine the factorization, the transfer functions, and hence the filters under a given length \( L \). The minimum phase condition is
\[
\sum_{l=0}^{m} g_{1,l}^2 \leq \sum_{l=0}^{m} \left[ g_{1,l}^{(ep)} \right]^2, \quad m = 0, \ldots, L-1,
\]
where \( g_{1,l}^{(ep)} \) denote the extremal phase scaling filter, and \( g_{1,l}^{(ep)} \) are any other filters of length \( L \). The criterion can lead to the class of extremal phase scaling filters, which have the fastest energy build-up. These filters have length \( L = 4, 6, 8, \ldots \), and are called \( D(L) \) filters. The other class of filters, the least asymmetric (LA) scaling filters, are those filters that are closest to being symmetric (close to a linear phase filter). For the filters \( \{g_{1,l}\} \) with the same squared gain function \( G_1(\cdot) \), and for a given shift \( \tilde{\nu} \), the LA filter is the one that minimizes
\[
\rho_{\tilde{\nu}}(\{g_{1,l}\}) \equiv \max_{-1/2 \leq f \leq 1/2} |\theta_1^{(G)}(f) - 2\pi f \tilde{\nu}|,
\]
where \( \theta_1(f) \) is the phase function [Percival and Walden, 2000, Chapter 2]. Symmetry is important because a symmetric filter can help associate the wavelet coefficients with the process outcome at a certain time point. The length \( L \) of an LA filter takes values 8, 10, 12, \ldots. In this dissertation, we are interested in the two classes of
Daubechies filters with different lengths: \( D(4) \), \( D(6) \), \( D(8) \), and \( LA(8) \) filters. Figure 2.1 shows the values of level 1 Haar, \( D(4) \), \( D(6) \), \( D(8) \) and \( LA(8) \) filters. The squared gain functions of these filters are displayed in Figure 2.2. As stated earlier, when the length \( L \) increases the filter is closer to a band-pass filter. Based on the bottom four panels in Figure 2.2, the \( D(8) \) and the \( LA(8) \) filters have exactly the same squared gain functions (because they have the same length \( L = 8 \)). Figure 2.3 are plots of the \( LA(8) \) squared gain functions at wavelet levels 1 to 5. The approximate pass bands (indicated by dashed lines) of the level \( j \) wavelet filter \([−1/2^j, −1/2^{j+1}]\) and \([1/2^{j+1}, 1/2^j]\), and the level \( j \) scaling filter has pass band \([−1/2^{j+1}, 1/2^{j+1}]\).

For the process \( \{Y_t : t \in \mathbb{Z}\} \), the level \( j \) (\( j = 1, 2, \ldots \)) DWT wavelet coefficients \( \{W_{j,k} : k \in \mathbb{N}\} \) are defined by

\[
W_{j,k} = \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^j k - 1 - l}, \quad k = 1, 2, \cdots, \tag{2.1}
\]

and the level \( j \) scaling coefficients \( \{V_{j,k} : k \in \mathbb{N}\} \) are defined by

\[
V_{j,k} = \sum_{l=0}^{L_j-1} g_{j,l} Y_{2^j k - 1 - l}, \quad k = 1, 2, \cdots. \tag{2.2}
\]

The wavelet coefficients represent changes of averages over different (usually dyadic) time scales, and the scaling coefficients represent averages over the same time scales.

In practice, for a finite sample \( \{Y_0, Y_1, \ldots, Y_{N-1}\} \) with sample size \( N \) a multiple of powers of two, we can replace the indexing in equations (2.1) and (2.2) with their circularly filtered versions,

\[
W_{j,k} = \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^j k - 1 - l \mod N}, \tag{2.3}
\]

and

\[
V_{j,k} = \sum_{l=0}^{L_j-1} g_{j,l} Y_{2^j k - 1 - l \mod N}. \tag{2.4}
\]
with \( k = 1, \ldots, N_j \) and \( N_j = N/2^j \). It is usually more efficient to calculate the DWT using the cascade algorithm [Mallat, 1989, Percival and Walden, 2000], which requires that the sample size \( N \) is an integer multiple of \( 2^J \), for some integer \( J \). Here \( J \) denotes the maximum wavelet level that we can analyze the time series to. For the situation with a sample size that is not equal to powers of two, we can truncate the sample to a multiple of a power of two. Without loss of generality, in this dissertation we assume \( N = 2^J \).

Since the DWT circularly filters the data, the first \( B_j = \min\{\lceil (L-2)(1-2^{-j}) \rceil, N_j \} \) coefficients use data from the start and the end of the series, and are called boundary coefficients. The remaining \( \{W_{j,k} : k = B_j + 1, \ldots, N_j \} \) are the \( M_j = N_j - B_j \) nonboundary coefficients [Percival and Walden, 2000]. When the filter width \( L \) is wider, there are more boundary coefficients. People prefer to use nonboundary coefficients for parameter estimation when a non-constant trend is presented [e.g., Craigmile et al., 2005]. But, it is believed that if a process has constant trend, it is better to include the boundary coefficients in the calculation because we are using more information. In our simulation study in Chapter 3, we apply the proposed models to both cases: all wavelet coefficients and non-boundary coefficients only, and we compare the estimation performance.

As an orthonormal transformation of \( \{Y_t\} \), the DWT coefficients has the matrix representation:

\[
W = (W_1^T, \ldots, W_J^T, V_J^T)^T = \mathcal{W}Y,
\]

where \( W_j = (W_{j,1}, \ldots, W_{j,N_j}) \) (or \( V_j \)) for \( j = 1, \ldots, J \) is the level \( j \) wavelet (or scaling) coefficients of length \( N_j \). In the above equation \( \mathcal{W} \) is an \( N \times N \) orthonormal transform matrix determined by the filters that satisfies \( \mathcal{W}^T\mathcal{W} = I_{N\times N} \). Because
Figure 2.1: Values of $\{h_{1,t}\}$ and $\{g_{1,t}\}$ for different filters.
Figure 2.2: The DWT squared gain functions of different filters (level 1).
Figure 2.3: $LA(8)$ squared gain functions (levels 1 to 5).
of the orthonormality, we will see that the likelihood of the time series data can be written as the likelihood of the DWT coefficients (see Chapter 3).

2.3 Statistical properties of the wavelet coefficients

2.3.1 Expected value of the DWT coefficients

For the class of Gaussian process \( \{Y_t\} \) defined at the beginning of this chapter, the expected value of its DWT coefficients have zero mean under the assumption of \( L \geq 2d_I \), where \( d_I \) denote the smallest integer so that the \( d_I \)th order differencing of \( \{Y_t\} \),

\[
X_t = \nabla^d Y_t = \sum \limits_{k=0}^{d_I} \binom{d_I}{m} (-1)^k Y_{t-k},
\]

is stationary. In terms of the wavelet filters, let \( b^{(d_I)}_{j,m} \) denote the \( d_I \)th order sum cumulative of the wavelet filter \( h_{j,l} \) [Craigmile et al., 2005, Lemma 1]; i.e.,

\[
\begin{align*}
b^{(0)}_{j,l} &= h_{j,l}, \\
b^{(r)}_{j,l} &= \sum \limits_{m=0}^{l} b^{(r-1)}_{j,m}, \quad l = 0, \ldots, L_j - r - 1, \quad r = 1, \ldots, d_I. \quad (2.5)
\end{align*}
\]

Then for \( d_I \geq 1 \), the wavelet coefficient can be shown to be equal to [Craigmile et al., 2005, page 1045]:

\[
W_{j,k} = \sum \limits_l h_{j,l} Y_{2^j k - 1 - l} \\
= \sum \limits_l \left[ \sum \limits_m \binom{d_I}{m} (-1)^m b^{(d_I)}_{j,l-m} \right] Y_{2^j k - 1 - l} \\
= \sum \limits_m b^{(d_I)}_{j,l-m} X_{2^j k - 1 - m},
\]

where \( \binom{d_I}{m} \) is defined to be 0 if \( m < 0 \) or \( m > d_I \). Because \( \sum_{m=0}^{L_j-d_I-1} b^{(d_I)}_{j,m} = 0 \) [Craigmile et al., 2005, page 1044], the expected value of the wavelet coefficient on a
fixed level $j$ is

$$E(W_{j,k}) = \sum_{m} b^{(d_{j})}_{j,m} E(X_{2^{j}k-1-m}) = 0,$$

for all $k$. For the special case of $d_{j} = 0$, i.e. the process $\{Y_{t}\}$ is stationary, then the level $j$ ($j = 1, \ldots, J$) scaling coefficients have expected value that does not depend on $k$,

$$E(V_{j,k}) = \sum_{l=0}^{L_{j}-1} g_{j,l} E(Y_{2^{j}k-1-l}) = \mu_{Y} \sum_{l=0}^{L_{j}-1} g_{j,l} = 2^{j/2} \mu_{Y},$$

because the level $j$ scaling filter satisfies $\sum_{l=0}^{L_{j}-1} g_{j,l} = 2^{j/2}$. We can further show some stationarity results of the DWT coefficients after the discussion about their covariance properties.

### 2.3.2 Covariance properties

The DWT of a stationary process preserve good properties of the time series process $\{Y_{t}\}$. We have already shown that wavelet coefficients have zero mean. We are now interested in the covariance structures of wavelet coefficients within and between scales.

The general expression for the covariance between wavelet coefficients on levels $j$ and $j'$ is

$$\text{cov}(W_{j,k}, W_{j',k+\tau}) = \text{cov}\left(\sum_{l=0}^{L_{j}-1} h_{j,l} Y_{2^{j}k-1-l}, \sum_{l'=0}^{L_{j'}-1} h_{j',l'} Y_{2^{j'}(k+\tau)-1-l'}\right) = \sum_{l=0}^{L_{j}-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} \text{cov}\left(Y_{2^{j}k-1-l}, Y_{2^{j'}(k+\tau)-1-l'}\right)$$

(2.6)
This covariance function depends on the levels \( j \) and \( j' \), time point \( k \), and time lag \( \tau \). In terms of the \( d_l \)th order difference of the process \( \{X_t\} \), the covariance can be written as:

\[
\text{cov}(W_{j,k}, W_{j',k+\tau}) = \text{cov}\left( \sum_{m=0}^{L_j-d_l-1} b^{(d_l)}_{j,m} X_{2^{2j}k-1-m}, \sum_{m'=0}^{L_{j'}-d_l-1} b^{(d_l)}_{j',m'} X_{2^{2j'}(k+\tau)1-m'} \right)
\]

\[
= \sum_{l=0}^{L_j-d_l-1} \sum_{l'=0}^{L_{j'}-d_l-1} b^{(d_l)}_{j,l} b^{(d_l)}_{j',l'} \gamma_X \left( 2^{2j'}l' - 2^j k + l - l' \right).
\]

(2.7)

**Lemma 2.2.** If the process \( \{Y_t, t \in \mathbb{Z}\} \) has stationary \( d_l \)th order difference, and the length of the wavelet filter \( L \) satisfies \( L \geq 2d_l \), then the wavelet coefficients at a fixed level \( j \) \((j = 1, \ldots, J)\), \( \{W_{j,k} : k \in \mathbb{Z}\} \), is a stationary process.

**Proof.** For \( L \geq 2d_l \geq 2d \) it has been shown that the expected value of level \( j \) wavelet coefficients \( \text{E}(W_{j,k}) = 0 \), and is free of \( k \). The within-scale wavelet autocovariance follows:

\[
\text{cov}(W_{j,k}, W_{j,k+l}) = \text{cov}\left( \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^{2j}k-1-l}, \sum_{l=0}^{L_{j'}-1} h_{j',l'} Y_{2^{2j'}(k+l)1-1-l} \right)
\]

\[
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} \gamma_X \left( Y_{2^{2j}k-1-l}, Y_{2^{2j'}(k+l)1-1-l} \right).
\]

By Equation (2.7), the above expression can be written in terms of the \( d_l \)th order sum cumulative of the wavelet filter \( h_{j,l} \):
\[
\text{cov}(W_{j,k}, W_{j,k+\tau}) = \text{cov}\left(Y_{2^j k - 1 - l}, Y_{2^j (k+\tau) - 1 - l'}\right) \\
= \sum_{l=0}^{L_j - d_I - 1} \sum_{l' = 0}^{L_j' - d_I' - 1} b_{j,l}^{(d_I)} b_{j,l'}^{(d_I')} \gamma_X (2^j \tau + l - l') \\
= \sum_{m = -(L_j - 1)}^{L_j - 1} \gamma_Y (2^j \tau + m) \sum_{l' = 0}^{L_j' - |m| - 1} h_{j,l'} h_{j,l' + |m|},
\]

because the \(d_I\)th order difference process \(\{X_t\}\) is stationary, and has an autocovariance function free of time point \(t\). The autocovariance function of the level \(j\) wavelet coefficients \(\gamma_{W,j}(\tau) \equiv \text{cov}(W_{j,k}, W_{j,k+\tau})\) is a function of the ACVS of \(\{X_t\}\), and does not depend on \(k\). Therefore \(\{W_{j,k} : k \in \mathbb{Z}\}\) is a stationary mean zero process at a given level \(j\). \(\square\)

When \(d_I = 0\), i.e., \(\{Y_t\}\) is stationary,
\[
\text{cov}(W_{j,k}, W_{j,k+\tau}) = \sum_{l=0}^{L_j - 1} \sum_{l' = 0}^{L_j - 1} h_{j,l} h_{j,l'} \gamma_Y (2^j \tau + l - l') \\
= \sum_{m = -(L_j - 1)}^{L_j - 1} \gamma_Y (2^j \tau + m) \sum_{l' = 0}^{L_j' - |m| - 1} h_{j,l'} h_{j,l' + |m|},
\]
where \(\sum_{l' = 0}^{L_j' - |m| - 1} h_{j,l'} h_{j,l' + |m|}\) is called the autocovariance of the level \(j\) wavelet filter [Percival and Walden, 2000]. A similar result holds for scaling coefficients:

**Lemma 2.3.** If the process \(\{Y_t\}\) is stationary (i.e., \(d_I = 0\)), the scaling coefficients at level \(j\) \((j = 1, \ldots, J)\), \(\{V_{j,k} : k \in \mathbb{Z}\}\), is stationary.

**Proof.** We showed previously
\[
E(V_{j,k}) = 2^{j/2} \mu_Y
\]
is free of \( k \). Also,

\[
\text{cov}(V_{j,k}, V_{j,k+\tau}) = \text{cov} \left( \sum_{l=0}^{L_j-1} g_{jl} Y_{2j,k-1-l}, \sum_{l=0}^{L_j-1} g_{jl} Y_{2j(k+\tau)-1-l} \right)
\]

\[
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} g_{jl} g_{jl'} \text{cov} \left( Y_{2j,k-1-l}, Y_{2j(k+\tau)-1-l'} \right)
\]

\[
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} g_{jl} g_{jl'} \gamma_Y \left( (2^j k - 1 - l) - (2^j (k + \tau) - 1 - l') \right)
\]

\[
= \sum_{m=-\left(L_j-1\right)}^{L_j-\left|m\right|-1} \gamma_Y \left( 2^j \tau + m \right) \sum_{l'=0}^{L_j-\left|m\right|-1} g_{jl} g_{jl'+|m|},
\]

where \( \sum_{l'=0}^{L_j-\left|m\right|-1} g_{jl} g_{jl'+|m|} \) is called the autocovariance of the level \( j \) scaling filter \cite{Percival2000}. The autocovariance function of scaling coefficients at a fixed level \( \gamma_{V,j}(\tau) \equiv \text{cov}(V_{j,k}, V_{j,k+\tau}) \) does not depend on \( k \). Thus at a given level \( j \), \( \{V_{j,k}\} \) is also a stationary process with mean \( 2^j/2 \mu_Y \).

While these results are restricted to the wavelet coefficients on the same scale, it is also interesting to study the wavelet covariance structures across scales. Based on Equation 2.6, the covariance function of \( W_{j,k} \) and \( W_{j',k+\tau} \) depends on scales \( j \) and \( j' \), time point \( k \), and lag \( \tau \). In the next part, we discuss the spectral representations of the DWT coefficients, the stationary within-scale wavelet coefficients, and the stationary cross scale wavelet process \cite{Moulines2007a}.

### 2.3.3 Spectral representations of the DWTs

The spectral representation defines the distribution of average power (variance) of the process versus frequency. With the spectral representation of a stationary process, we can analyze the process in the frequency domain. Let \( S_{W_j}(f) \) denote the
spectral density function (SDF) of the wavelet coefficients $\{W_{j,k} : k \in \mathbb{Z}\}$. The SDF of wavelet coefficients satisfies:

$$s_{j,\tau} = \int_{-1/2}^{1/2} e^{i2\pi f\tau} S_{W_j}(f) df.$$ 

By Percival and Walden [2000], Equation (267c), the ACVS and the SDF are a Fourier transform pair, and the ACVS is the inverse discrete Fourier transform (DFT) of the SDF. Write the level $j$ transfer function as $H_j(f) = D^{L/2}(f)A_j(f)$. Here $D(f) = 1 - e^{-12\pi f}$ is the transfer function of the (first order) difference filter, and let $D(f)$ be the squared gain function of the difference filter. We let $B_j^{(d_i)}(\cdot)$ denote the transfer function for $\{b_{j,l}^{(d_i)}\}$, then [Craigmile and Percival, 2005, page 1045]

$$H_j(f) = D^{d_i}(f)B_j^{(d_i)}(f),$$

and

$$B_j^{(d_i)}(f) = D^{L/2-d_i}(f)A_j(f).$$

The frequency equivalent of the wavelet covariance (Equation 2.7) can then be obtained [Craigmile and Percival, 2005, page 1045]:

24
\[
\begin{align*}
\text{cov}(W_{j,k}, W_{j',k+l}) &= \sum_{l=0}^{L_j-d_j-1} \sum_{l'=0}^{L_{j'}-d_{j'}-1} b_{j,l}^{(d_j)} b_{j',l'}^{(d_{j'})} \gamma_X \left(2^j k' - 2^j k + l - l'\right) \\
&= \sum_{l=0}^{L_j-d_j-1} \sum_{l'=0}^{L_{j'}-d_{j'}-1} b_{j,l}^{(d_j)} b_{j',l'}^{(d_{j'})} \int_{-1/2}^{1/2} e^{i2\pi f (2^j (k+l') - 2^j k + l - l')} S_X(f) df \\
&= \int_{-1/2}^{1/2} \left( \sum_{l=0}^{L_j-d_j-1} b_{j,l}^{(d_j)} e^{-i2\pi f l} \right) \left( \sum_{l'=0}^{L_{j'}-d_{j'}-1} b_{j',l'}^{(d_{j'})} e^{i2\pi f l'} \right)^* \cdot S_X(f) e^{i2\pi f (2^j (k+l') - 2^j k)} df \\
&= \int_{-1/2}^{1/2} \left( B_j^{(d_j)}(f) \right) \left( B_{j'}^{(d_{j'})}(f) \right)^* S_X(f) e^{i2\pi f (2^j (k+l') - 2^j k)} df \\
&= \int_{-1/2}^{1/2} \left( D_{L/2-d_j}^{L/2-d_{j'}}(f) A_j(f) \right) \left( D_{L/2-d_j}^{L/2-d_{j'}}(f) A_{j'}(f) \right)^* S_X(f) e^{i2\pi f (2^j (k+l') - 2^j k)} df \\
&= \int_{-1/2}^{1/2} H_j(f) H_{j'}^*(f) S_X(f) e^{i2\pi f (2^j (k+l') - 2^j k)} df \\
&= \int_{-1/2}^{1/2} H_j(f) H_{j'}^*(f) S_Y(f) e^{i2\pi f (2^j (k+l') - 2^j k)} df, \quad (2.8)
\end{align*}
\]

where \(H_j(f)\) is the transfer function for the \(j\)th level wavelet filter and \(H_{j'}^*(f)\) is the complex conjugate of the \(j'\)th level transfer function.

For the within-scale wavelet coefficients \((j = j')\), the above covariance expression reduces to:

\[
\text{cov}(W_{j,k}, W_{j,k+l}) = \int_{-1/2}^{1/2} e^{i2\pi 2^j f \tau} \mathcal{H}_j(f) S_Y(f) df,
\]

where \(\mathcal{H}_j(\cdot)\) is the squared gain function for the \(j\)th level wavelet filter \(\{h_{j,l} : l = 0, \ldots, L_j - 1\}\). An explicit expression of the wavelet spectral density can be obtained
from

\[ s_{j,\tau} = \cov(W_{j,k}, W_{j,k+\tau}) \]

\[ = \int_{-1/2}^{1/2} e^{i2\pi2^j f\tau} H_j(f) S_Y(f) df \]

\[ = \int_{-1/2}^{1/2} e^{i2\pi\zeta\tau} H_j \left( \frac{\zeta}{2^j} \right) S_Y \left( \frac{\zeta}{2^j} \right) d\zeta \left( \frac{1}{2^j} \right) \] (Let \( \zeta = 2^j f \), and change variable.)

\[ = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-2^j/2+m}^{-2^j/2+m+1} e^{i2\pi\zeta\tau} H_j \left( \frac{\zeta}{2^j} \right) S_Y \left( \frac{\zeta}{2^j} \right) d\zeta, \]

by breaking the integral into segments. Since \( H_j \left( \frac{\zeta}{2^j} \right) S_Y \left( \frac{\zeta}{2^j} \right) \) is periodic with period 1, we could further write the above equation as:

\[ \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-2^j/2+m}^{-2^j/2+m+1} e^{i2\pi\zeta\tau} H_j \left( \frac{\zeta}{2^j} \right) S_Y \left( \frac{\zeta}{2^j} \right) d\zeta = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-1/2}^{1/2} e^{i2\pi\left( \frac{\zeta + m}{2^j} \right)\tau} H_j \left( \frac{\zeta + m}{2^j} \right) S_Y \left( \frac{\zeta + m}{2^j} \right) d\zeta. \]

Thus, the level \( j \) wavelet SDF, as a function of \( S_Y(\cdot) \), is

\[ S_{W_j}(f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} H_j \left( \frac{f + m}{2^j} \right) S_Y \left( \frac{f + m}{2^j} \right). \]

In wavelet-based estimation it often helps to discriminate the within-level and the across-level properties of the wavelet coefficients. The following result for the wavelet coefficients within a given wavelet level is an extension of the results presented in Percival and Walden [2000] and Craigmile and Percival [2005] to our more general class of processes. Taking all the results presented before together, we obtain the following proposition.
Proposition 2.4. Suppose that $L \geq 2d$. Then the nonboundary wavelet coefficients at a given level $j$ are a portion of a zero mean stationary process with autocovariance

$$s_{j,\tau}(\theta) = \text{cov}(W_{j,k}, W_{j,k+\tau}) = \sum_{m=-(L_j-1)}^{L_j-1} \gamma_X(2^j \tau + m) \sum_{l'=0}^{L_j-|m|-1} b_{j,l'}(d_t) b_{j,l'+|m|}^{(d_t)}(2^j \tau + m)$$

(2.9)

and thus we identify the SDF of the level $j$ wavelet coefficients to be

$$S_{W_j,\theta}(f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \mathcal{H}_j\left(\frac{f + m}{2^j}\right) S_Y\left(\frac{f + m}{2^j}\right), \quad f \in [-1/2, 1/2], \quad (2.10)$$

where $\mathcal{H}_j(f) = H_j(f) H_j^*(f)$ is the squared gain function of the level $j$ wavelet filter $h_{j,l}$.

By the decorrelation properties of the DWT, the covariance between wavelet coefficients on different levels tend to be small. The width $L$ of the Daubechies wavelet filter used can control the level of decorrelation, because as we wrote earlier, as the length of the filter $L$ increases, the filter $\{h_{j,l}, l = 0, \ldots, L_j - 1\}$ will be an approximation of a bandpass filter with pass-band $[1/2^{j+1}, 1/2^j]$. It is showed the DWT covariance tends to zero when $\tau \neq 0$ and filter length $L$ is sufficiently large [Craigmille and Percival, 2005]. To understand the role of across-scale covariances upon our wavelet Whittle estimator we need a more general result first introduced by Moulines et al. [2007a], that uses the cross-level wavelet process. First, for each level $j$, let

$$W_{j,k}(u) = (W_{j-u,2^u k + r} : r = 0, \ldots, 2^u - 1)^T$$

be a $2^u$-dim vector, where $u$ is some integer such that $0 \leq u < j$. Then the following holds [adapted from Moulines et al., 2007a, Corollary 1].
Proposition 2.5. Let $L \geq 2d$. Then the cross-level wavelet process at level $j$ defined by $\{(W_{j,k}, W_{j,k}(u))^T \; : \; k \in \mathbb{Z}\}$, is stationary with covariance

$$\text{cov}(W_{j,k}, W_{j,k'}(u)) = \int_{-1/2}^{1/2} e^{i2\pi f(k-k')} D_{j,u}(f) df,$$

and a vector-valued cross-spectral density $D_{j,u}(f)$, given by

$$D_{j,u}(f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} 2^{u/2} e_u(f+m) H_j \left( \frac{f+m}{2^j} \right) H^*_j \left( \frac{f+m}{2^j} \right) S_Y \left( \frac{f+m}{2^j} \right),$$

where $e_u(\xi) = 2^{-u/2} \left[ 1, e^{-i2^{-u/2}2\pi\xi}, \ldots, e^{-i(2^u-1)2^{-u/2}2\pi\xi} \right]^T$ is a vector of unit norm for all $\xi \in \mathbb{R}$.

This result tells us that to obtain a stationary process across levels we must consider jointly the wavelet coefficients on previous scales that were used in the DWT to construct the coefficient $W_{j,k}$. We obtain Proposition 2.4 when $u = 0$.

The summation of all cross-scale coefficients can be decomposed as sums of the covariances of such stationary cross-scale coefficients. It can be shown that the squared cross-scale process $(W_{j,k}^2, W_{j,k}(u))^T = (W_{j,k}^2, W_{j-u,2^u k}, \ldots, W_{j-u,2^u k+2^u-1})^T$ is stationary with some vector-valued SDF $A_{j,u}(f)$ (to be defined later). These useful results are employed in our proof of the asymptotic normality of the wavelet Whittle estimators.

2.4 An example

A stationary Gaussian process with Matérn covariance function can be used to illustrate how the DWT decorrelates data. The Matérn covariance function with time lag $h$ is defined in Stein [1999] as

$$\gamma_Y(h) = \frac{1}{2\pi} \frac{\sqrt{\pi\sigma^2}}{2^{\nu-1}\Gamma(\nu+1/2)} \phi^{2\nu}(\phi|h|)^{\nu} K_\nu(\phi|h|).$$

(2.11)
The corresponding Matérn SDF is defined as
\[ S(f) = \frac{\Delta \sigma^2}{|\phi^2 + (2\pi \Delta f)^2|^{\nu+1/2}}, \]
where \( \Gamma \) is the gamma function, \( K_\nu \) is the modified Bessel function of the second kind with order \( \nu \); \( \phi \) is a scale parameter; \( \nu \) is a smoothness parameter. In the SDF, \( \Delta \) is the sampling interval, which is set to 1 in this work for simplicity. The smoothness parameter is related to the differentiability of the process. Specifically, when \( \nu = 0.5 \), the covariance is exponential; when \( \nu \to \infty \), the Matérn covariance converges to a squared exponential covariance function. The wavelet filters used in simulation studies are D(4), D(6), D(8), and LA(8), all of which satisfy the condition from Proposition 2.4 and Proposition 2.5 (\( L > 2d \)).

In Figure 2.4, compared to the SDF of the original process, the SDFs of the DWT of \( \{Y_t\} \) are much flatter, and closer to the SDF of a white noise process (a flat horizontal line). This means the wavelet transform decorrelates the original process. The plots also motivate us to consider using Whittle approximations to estimate process parameters. At a fixed level, the nonboundary wavelet coefficients are approximately a white noise process, especially at higher levels, such as SDFs at level 5 in Figure 2.4. We see decorrelation effects under different Matérn processes, but the decorrelation performance varies with different values of Matérn smoothness parameter. As shown in Figure 2.4, when the process is smoother (\( \nu \) is larger), it is harder to decorrelate.

We can also explore the whitening effects using autocovariance functions. In Percival and Walden [2000, Chapter 9], plots of cross-scale wavelet autocovariances showed how well the DWT decorrelates long memory. The cross-scale covariances (2.6) are close to 0 when time lag \( \tau \) is large. As the “distance” between levels \( j \) and \( j' \) becomes larger, the cross-scale covariance vanishes. Figure 2.5 is an example with
Figure 2.4: A Gaussian process with Matérn covariance function: \( \sigma^2 = 1 \) and \( \phi = 0.25 \). From top to bottom the Matérn smoothness parameter \( \nu \) takes value 0.25, 0.5, 1.5, respectively. From left to right the panels show SDF of the Gaussian process \( \{Y_t\} \) and the SDFs of the DWTs \( \{Y_t\} \), using various filters (D(4), D(6), and LA(8)). The wavelet spectrums at levels 1 to 5 are displayed using different colors and line types. The spectrums are in decibels, and the frequencies take values in the interval \([0, 0.5]\).

The cross-scale covariances are showed in Figure 2.6. The cross-scale covariances are very small, especially when time lag \( \tau \neq 0 \). The only concern is: with the D(4)
filter, $\text{cov}(W_{3,k}, W_{5,k})$ is about the value of -0.8, a large value; with the D(6) filter, the covariances of wavelet coefficients at level 2 and level 3 has relatively large values at the first two lags (-0.271 and 0.330). As expected, because the LA(8) filter has a wider length, it performs better in terms of decorrelation, while the D(4) filter performs poorly.

Figure 2.5: A Gaussian process with Matérn covariance function: $\sigma^2 = 1$, $\nu = 0.5$, and $\phi = 0.25$. The top-left panel shows the ACVS of the Gaussian process $\{Y_t\}$ at lags 0 to 15. The remaining panels show the within-scale ACVS of the LA(8) nonboundary wavelet coefficients.
Figure 2.6: The panels show the between-scale covariances of the LA(8) nonboundary wavelet coefficients.
Figure 2.7: A Gaussian process with Matérn covariance function: $\sigma^2 = 1$, $\nu = 0.5$, and $\phi = 0.25$. The top-left panel shows the ACVS of the Gaussian process $\{Y_t\}$ at lags 0 to 15. The remaining panels show the within-scale ACVS of the D(6) nonboundary wavelet coefficients.

### 2.5 Conclusion

In this chapter, we provide the basic definitions and statistical properties of the DWT, and focus on the decorrelation properties of the wavelet coefficients. Proposition 2.4 and Proposition 2.5 state the stationarity results of the within and between scale wavelet coefficients. Then the Matérn example and the figures in the last section show the motivation of using a wavelet based method and a spectral approximation approach to analyze a process with non-Markov dependencies. We will define the wavelet-Whittle method, and will study the asymptotic properties of the wavelet-Whittle estimators in Chapter 3.
Figure 2.8: The panels show the between-scale covariances of the D(6) nonboundary wavelet coefficients.
Figure 2.9: A Gaussian process with Matérn covariance function: $\sigma^2 = 1$, $\nu = 0.5$, and $\phi = 0.25$. The top-left panel shows the ACVS of the Gaussian process $\{Y_t\}$ at lags 0 to 15. The remaining panels show the within-scale ACVS of the D(4) nonboundary wavelet coefficients.
Figure 2.10: The panels show the between-scale covariances of the D(4) nonboundary wavelet coefficients.
3.1 Wavelet-based Whittle approximations

Following the setup in the previous chapter, \( \{Y_t : t \in \mathbb{Z}\} \) is a univariate real-valued Gaussian process with ACVS \( \gamma_Y(\cdot) \). Given a finite sample \( \{Y_t : t = 0, \ldots, N - 1\} \) with \( N = 2^J, J \in \mathbb{N}^+ \), we want to estimate the parameter vector \( \theta \in \Theta \), where \( \Theta \) is a compact set in \( \mathbb{R}^p \). The exact likelihood function is

\[
L_N(\theta) = f_Y(Y; \theta) = (2\pi)^{-N/2} (\det T_N(S_Y,\theta))^{-1/2} \exp \left( -\frac{1}{2} Y^T T_N(S_Y,\theta)^{-1} Y \right),
\]

and the log likelihood function is

\[
l_N(\theta) = \log L_N(\theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log (\det T_N(S_Y,\theta)) - \frac{1}{2} Y^T T_N(S_Y,\theta)^{-1} Y.
\]

For a mean zero Gaussian process \( \{Y_t\} \) with a (generalized) SDF given by \( S_{Y,\theta}(f) \), let \( T_N(S_{Y,\theta}) \) denote the \( N \times N \) covariance matrix. When \( Y_t \) is stationary, \( \gamma_Y(q - q'; \theta) = \gamma_Y(q' - q; \theta) \), so \( T_N(S_{Y,\theta}) \) has a symmetric Toeplitz structure with an \((q,q')\) element given by

\[
[T_N(S_{Y,\theta})]_{q,q'} = \gamma_Y(q - q'; \theta) = \int_{-1/2}^{1/2} e^{2\pi f(q-q')} S_{Y,\theta}(f) df.
\]
In the generalized SDF case when the process $Y_t$ is stationary after differencing $d_I$ times the matrix $T_N(S_Y, \theta)$ is generated using $d_I$th order cumulative sums of covariances of the form of Equation (3.1) where the SDF of $Y_t$ is replaced by the SDF of the differenced process (see, e.g., Equation (2.7)),

$$[T_N(S_Y, \theta)]_{q,q'} = \sum_{m=1}^{q} \sum_{m'=1}^{q'} b_m^{(d_I)} b_{m'}^{(d_I)} \gamma_X(m - m'; \theta),$$

(3.2)

where $b_m^{(0)} = 1$, and for $d \geq 1$,

$$b_m^{(d)} = \sum_{k=1}^{m} b_k^{(d-1)}, \quad m = 1, \cdots, N.$$

Taking the DWT of $Y$ to $J$ wavelet levels using the wavelet transform matrix $W$, the vector of DWT coefficients is $W = (W_1^T, \ldots, W_J^T, V_J^T)^T = WY$. The wavelet covariance matrix $\text{cov}(W) = W\text{cov}(Y)W^T = W T_N(S_Y, \theta) W^T$. From Chapter 2, DWT is an orthonormal transform, $W^TW = I_N$, so the likelihood of $Y$ is identical to the likelihood of $W$, i.e.,

$$L_N^W(\theta) = f_W(W; \theta)$$

$$= (2\pi)^{-\frac{N}{2}} (\det \text{cov}(W))^{-\frac{1}{2}} \exp \left( -\frac{1}{2} W^T \text{cov}(W)^{-1} W \right)$$

$$= (2\pi)^{-\frac{N}{2}} (\det(W T_N(S_Y, \theta) W^T))^{-\frac{1}{2}} \exp \left( -\frac{1}{2} W^T (W T_N(S_Y, \theta) W^T)^{-1} W \right),$$

with the log-likelihood

$$l_N^W(\theta) = \log L_N^W(\theta)$$

$$= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(\det(W T_N(S_Y, \theta) W^T)) - \frac{1}{2} W^T (W T_N(S_Y, \theta) W^T)^{-1} W.$$

The wavelet Whittle method [McCoy and Walden, 1996, Craigmile et al., 2005, Moulines et al., 2008] uses the sum of approximate log-likelihoods of wavelet coefficients at different levels to approximate the above true log-likelihood. The wavelet Whittle estimator, $\hat{\theta}_N$, is the value of $\theta$ that maximizes this approximate likelihood.
There are a number of different Whittle approximations possible. As discussed in Chapter 2, the discrete wavelet transform is a whitening transform. The DWT wavelet coefficients are approximately uncorrelated across time scales, i.e., \( W_1, \ldots, W_J, V_J \) are approximately uncorrelated. By assuming independence between scales, we obtain the first Whittle (log) likelihood:

\[
l_N^{W,1}(\theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} \left[ \log \det T_{N_j}(S_{W_j,\theta}) + W_j^T T_{N_j}^{-1}(S_{W_j,\theta}) W_j \right] - \frac{1}{2} \sum_{j=1}^{J} \left[ \log \det T_{N_j}(S_{V_j,\theta}) + V_j^T T_{N_j}^{-1}(S_{V_j,\theta}) V_J \right].
\]

A further approximation drops the scaling coefficients \( V_J \) and uses the sum of likelihoods of the wavelet coefficients \( W_j \) to approximate the true likelihood.

\[
l_N^{W,2}(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \left[ N_j \log(2\pi) + \log \det T_{N_j}(S_{W_j,\theta}) + W_j^T T_{N_j}^{-1}(S_{W_j,\theta}) W_j \right],
\]

where \( S_{W_j,\theta} \) is the level \( j \) wavelet SDF defined by (2.10). Proposition 2.4 states if \( L \geq 2d \), the DWT coefficients \( \{W_{j,}\} \) are stationary at a given level \( j \). Thus \( T_{N_j}(S_{W_j,\theta}) \) is a \( N_j \times N_j \) Toeplitz matrix with \((p,q)\)th element given by

\[
[T_{N_j}(S_{W_j,\theta})]_{p,q} = \text{cov}(W_{j,p}, W_{j,q}) = \int_{-1/2}^{1/2} e^{i2\pi f(p-q)} S_{W_j,\theta}(f) df.
\]

By further simplifying the covariance of the wavelet coefficients, one can get an approximate likelihood that is more (computationally) efficient than (3.4). Let \( \tilde{S}_{W_j,\theta} \) be an approximate SDF of \( S_{W_j,\theta} \). We call the following Equation (3.5) the wavelet Whittle likelihood function. With the additional assumption of white noise approximate model on every scale, we get the wavelet Whittle likelihood [Craigmiles, 2000].

\[
l_N^W(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \left( \log \det T_{N_j}(\tilde{S}_{W_j,\theta}) + W_j^T \left( T_{N_j}(\tilde{S}_{W_j,\theta}) \right)^{-1} W_j \right) - \frac{N - 1}{2} \log(2\pi).
\]

39
Common choices of the approximating SDF include white noise (WN) and AR(1), which have sparse covariance matrices $T_{N_j}({\tilde S}_{W_j,\theta})$. Assuming independence of the wavelet coefficients both across and within each wavelet level yields the white-noise (WN) Whittle likelihood [McCoy and Walden, 1996, Craigmile et al., 2005, Moulines et al., 2008]. In this work, we focus on two wavelet Whittle models: white noise and AR(1) [Craigmile et al., 2005].

### 3.2 A white noise model

In this section, we provide the central limit theorem for wavelet-based Whittle MLEs under the white noise approximate likelihood presented below. Suppose the wavelet coefficients $\{W_{j,k} : k = 1, \cdots, N_j\}$ at level $j$ ($j = 1, \ldots, J$) follows the white noise model:

$$W_{j,k} \sim N(0, s_{j,0}(\theta)).$$

The wavelet whittle likelihood is

$$l^W_N(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \left( \log \det T_{N_j}({\tilde S}_{W_j,\theta}) + W_j^T (T_{N_j}({\tilde S}_{W_j,\theta}))^{-1} W_j \right) - \frac{N-1}{2} \log(2\pi)$$

$$= -\frac{1}{2} \sum_{j=1}^{J} \left( N_j \log(s_{j,0}(\theta)) + \sum_{k=1}^{N_j} \frac{W_{j,k}^2}{s_{j,0}(\theta)} \right) - \frac{N-1}{2} \log(2\pi). \quad (3.6)$$

The approximate covariance matrix $T_{N_j}({\tilde S}_{W_j,\theta})$ is a $J \times J$ diagonal matrix with wavelet variances $s_{j,0}(\theta)$ on its diagonals. The wavelet-based Whittle MLE, $\hat{\theta}_N$, is obtained by maximizing Equation (3.6).
It is important to study the asymptotic properties of the MLE of $\theta$. Our asymptotic result for the MLE relies on a theorem that provides the joint asymptotic distribution of the empirical wavelet variance

$$\hat{v}_j = \sum_{k=1}^{N_j} W_{j,k}^2 / N_j,$$

over a range of levels $j = 1, \ldots, J$, for some fixed positive integer $J$. From the results presented in Chapter 2, first note that when $L \geq 2d$, $\hat{v}_j$ is an unbiased estimator of the level $j$ wavelet variance $s_{j,0}(\theta)$.

$$E(\hat{v}_j) = E\left(\frac{\sum_{k=1}^{N_j} W_{j,k}^2}{N_j}\right) = E(W_{j,k}^2) = \text{var}(W_{j,k}) = s_{j,0}(\theta).$$

Then, Theorem 3.1 presents the limiting joint distribution of the empirical wavelet variances.

**Theorem 3.1.** Suppose that $L \geq 2d$. For each $j = 1, \ldots, J$ let $\rho_{j,1} = \sqrt{N_j/N_1}$. Then

$$\sqrt{N_1} [\rho_{1,1}(\hat{v}_1 - s_{1,0}), \ldots, \rho_{J,1}(\hat{v}_J - s_{J,0})] \xrightarrow{D} N_J(0, \Sigma),$$

as $N \to \infty$, where $\Sigma$ is the $J \times J$ covariance matrix with $(j, j')$th entry

$$2^{(j'-j)/2} \mathbf{1}^T A_{j,j'}(0).$$

In the theorem, $A_{j,u}$ for $0 \leq u < j$ is the vector-valued SDF of the squared cross-scale process $(W_{j,k}^2, W_{j,k}(u)^T)^T$, where $W_{j,k}(u) = (W_{j-u,2^u k+r}^2 : r = 0, \ldots, 2^{u-1} - 1)^T$ for each $j, k$, and $u$. By Proposition 2.5, the squared cross-scale process is covariance stationary when $L \geq 2d$. This result still holds if we replace the empirical wavelet variance by an estimator based only on the nonboundary wavelet coefficients.

We are now in a position to present the asymptotic distribution of $\tilde{\theta}_N$ based on the wavelet coefficients of the time series to a fixed level $J$, regardless of the length $N$. 

41
**Theorem 3.2.** Suppose that \( L \geq 2d \). Then the \( p \)-dimensional vector \( \hat{\theta}_N \) is a consistent estimator of the true parameter vector \( \hat{\theta}_0 \), and as \( N \to \infty \),

\[
\sqrt{N} \left[ \hat{\theta}_N - \theta_0 \right] \xrightarrow{D} N_p(0, \Omega(\theta_0)),
\]

where \( \Omega(\theta_0) \) is some \( p \times p \) covariance matrix that depends on \( \theta_0 \).

For each \( j \), let \( w_{j,q} = [\partial s_{j,0}(\theta)/\partial \theta_q]/s_{j,0}(\theta) \). Then from the proof of this theorem in Section 3.3 we have

\[
\Omega(\theta) = \Gamma^{-1}K\Gamma^{-1},
\]

where \( \Gamma \) is a \( p \times p \) matrix with \((q, q')\) element

\[
\Gamma_{q,q'} = \sum_{j=1}^{J} 2^{1-j} w_{j,q} w_{j,q'}, \tag{3.7}
\]

and \( K \) is a \( p \times p \) matrix with \((q, q')\) element

\[
K_{q,q'} = \sum_{j'=1}^{J} \sum_{j=1}^{J} 2^{1-j} w_{j,q} w_{j',q'} \frac{1^T A_{j, q-j'}}{s_{j,0}(\theta)s_{j',0}(\theta)}. \tag{3.8}
\]

Theorem 3.2 still applies when we estimate \( \theta \) using (3.6) restricting to wavelet coefficients on levels \( J_0, \ldots, J \), for \( 1 < J_0 < J \): in that case each sum in equations (3.7) and (3.8) starts at level \( J_0 \) and we replace \( 2^{1-j} \) by \( 2^{J_0-j} \).

The above two theorems hold for non-boundary coefficients and all-coefficients cases. We assume that the processes being considered all have zero mean. Then the estimations would be improved by including both the boundary and the non-boundary wavelet coefficients, because we are “using” more information. When there exists a trend component, the boundary coefficients will be influenced by the trend term (see, e.g., Craigmile et al. [2005]), so including the boundary coefficients will introduce bias in this case. In Section 3.6 simulation studies will be conducted to address the
performance of the estimators, their asymptotic properties and the benefit of including boundary coefficients for models without trend.

With the central limit result for the white noise approximation model, we can jointly estimate the vector of parameters $\theta$, and can assess the accuracy of the wavelet Whittle-based MLEs, $\hat{\theta}$, using approximate confidence intervals (or confidence regions). For example, under a white noise model, suppose we are interested in estimating the range parameter $\phi$ of a Matérn process with $\nu = 0.5$ and $\sigma^2 = 1$. Let $s_{j,0}(\phi) = \text{var}(W_{j,t})$, and let $d_{j,0}(\phi) = \frac{\partial s_{j,0}(\phi)}{\partial \phi}$. By Theorem 3.2, we can derive the limiting distribution of the wavelet-Whittle estimator $\hat{\phi}$,

$$\sqrt{N_{J_0}}(\hat{\phi} - \phi_0) \xrightarrow{D} N(0, \sigma^2_{\phi}), \text{ as } N_{J_0} \to \infty,$$

where $\sigma^2_{\phi} = \left( \sum_{j=J_0}^J 2^{J_0-j} \frac{d_{j,0}(\phi_0)}{s_{j,0}^2(\phi_0)} \right)^{-2} \sum_{j=J_0}^J \sum_{j'=J_0}^J \frac{\rho_{j,J_0} \rho_{j',J_0} d_{j,0}(\phi_0) d_{j',0}(\phi_0)}{s_{j,0}^2(\phi_0) s_{j',0}^2(\phi_0)} \Sigma_{jj'}$, and $\Sigma_{jj'}$ is the $(j,j')$th element of the limiting covariance matrix $\Sigma$ in Theorem 3.1. Then, an approximate $(1 - \alpha)100\%$ interval for $\phi$ is given by $\hat{\phi} \pm z_{\alpha/2} \sigma_{\phi}$.

### 3.3 The proof of the white noise results

In this section, we prove Theorem 3.1 and Theorem 3.2 with $J_0$ as the starting wavelet level. The following notations are used in the proofs,

$$R_j = \sum_k W_{j,k}^2, \quad \hat{\nu}_j = \frac{R_j}{N_j}, \quad \rho_{j,J_0} = \left( \frac{N_j}{N_{J_0}} \right)^{1/2},$$

where $\rho_{j,J_0} = \left( \frac{N_j}{N_{J_0}} \right)^{1/2}$ is the square root of the ratio of numbers of coefficients in levels $j$ and $J_0$. Let

$$\hat{\eta} = \left( \rho_{J_0,J_0} \hat{\nu}_{J_0} \right) \ldots \left( \rho_{J_0,J_0} \hat{\nu}_{J_0} \right) = \left( \rho_{J_0,J_0} \frac{R_{J_0}}{N_{J_0}} \right), \quad \eta = \left( \rho_{J_0,J_0} \text{var}(W_{J_0,k}) \right) \ldots \left( \rho_{J_0,J_0} \text{var}(W_{J_0,k}) \right),$$

43
where \( E(\hat{\eta}) = \eta \), \( \text{cov}(\hat{\eta}) = \frac{1}{N_{j_0}} \Sigma \). We first show the squared within-scale wavelet coefficients \( \{W_{j,k}^2 : k \in \mathbb{Z}\} \) is stationary.

**Proof.** For a fixed level \( j \), the expected value of \( W_{j,k}^2 \),

\[
E(W_{j,k}^2) = \text{var}(W_{j,k}) = s_{j,0}(\theta),
\]

does not depend on \( k \), and its covariance:

\[
\text{cov}(W_{j,k}^2, W_{j,k'}^2) = E(W_{j,k}^2 W_{j,k'}^2) - E(W_{j,k}^2)E(W_{j,k'}^2)
\]

(Since \( \{W_{j,k} : k \in \mathbb{Z}\} \) is Gaussian with mean 0, use Isserlis’ Theorem, Isserlis [1918])

\[
= E(W_{j,k}^2)E(W_{j,k'}^2) + E(W_{j,k} W_{j,k'})E(W_{j,k} W_{j,k'})
\]

\[
+ E(W_{j,k} W_{j,k'})E(W_{j,k} W_{j,k'}) - E(W_{j,k}^2)E(W_{j,k'}^2)
\]

\[
= 2 (E(W_{j,k} W_{j,k'}))^2
\]

\[
= 2 (\text{cov}(W_{j,k}, W_{j,k'}))^2
\]

\[
= 2s_{j,k-k'}^2(\theta),
\]

only depends on the time lag \( k-k' \). Thus the squared within-scale wavelet coefficients \( \{W_{j,k}^2 : k \in \mathbb{Z}\} \) is stationary. \( \square \)
With this stationarity result, the variance of $\sqrt{N_{j0}} \hat{\eta}_j$ can be shown to converge to the SDF of $W^2_{j,k}$ at zero frequency when $N_{j0} \to \infty$:

$$\text{var}(\sqrt{N_{j0}} \hat{\eta}_j) = \text{var} \left( \frac{1}{\sqrt{N_j}} R_j \right) = \frac{1}{N_j} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_j} \text{cov}(W^2_{j,k}, W^2_{j,k'})$$

(by the stationarity of $\{W^2_{j,k} : k \in \mathbb{Z}\}$)

$$= \sum_{h=-(N_j-1)}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right) \text{cov}(W^2_{j,1}, W^2_{j,h+1}).$$

In the above expression, $\sum_{h=-(N_j-1)}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right)$ is a Fejér’s kernel that behaves as a Dirac Delta function. It converges to a point mass at zero when $N_j \to \infty$ [Percival and Walden, 1993, page 199]. Thus we have

$$\text{var}(\sqrt{N_{j0}} \hat{\eta}_j) \xrightarrow{P} S_{W^2}(0), \text{ as } N_{j0} \to \infty,$$

which follows from Theorem 1 in Serroukh et al. [2000].

We use the following lemmas in Moulines et al. [2008] to prove Theorem 3.1, namely that

$$\sqrt{N_{j0}} (\hat{\eta} - \eta) \xrightarrow{P} N(0, \Sigma), \text{ as } N_{j0} \to \infty. \quad (3.10)$$

First define the **spectral radius** of a square matrix $\Gamma$: letting $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of a matrix $\Gamma_{n \times n}$, then its spectral radius is defined as

$$\rho(\Gamma) \overset{\text{def}}{=} \max_k (|\lambda_k|).$$

**Lemma 3.3.** Let $\{\xi_l, l \in \mathbb{Z}\}$ be a stationary process with spectral density $g$ and let $\Gamma_n$ be the covariance matrix of $[\xi_1, \ldots, \xi_n]$. Then, $\rho(\Gamma_n) \leq 2\pi ||g||_\infty$.  

45
Lemma 3.4. Let \( \{\xi_n, n \geq 1\} \) be a sequence of Gaussian vectors with zero mean and covariance \( \Gamma_n \). Let \( (A_n)_{n \geq 1} \) be a sequence of deterministic symmetric matrices such that \( \lim_{n \to \infty} \text{var}(\xi_n^T A_n \xi_n) = \sigma^2 \in [0, \infty) \). Assume that \( \lim_{n \to \infty} [\rho(A_n) \rho(\Gamma_n)] = 0 \). Then, \( \xi_n^T A_n \xi_n - E[\xi_n^T A_n \xi_n] \xrightarrow{D} N(0, \sigma^2) \).

We now prove Theorem 3.1, let \( \Gamma_N = \text{cov}(\xi_N) \), where \( \xi_N \) is a length \( N - 2^{J_0} \) vector \( (\rho^{-1/2} W_{j,k})_{j=J_0,\ldots,J; k=1,\ldots,N_j} \). With \( \{a_j\}_{j=J_0,\ldots,J} \in \mathbb{R} \), let \( A_N \) be an \( (N - 2^{J_0}) \times (N - 2^{J_0}) \) diagonal matrix with \( q \)th diagonal element \( a_j N_j^{-1/2} \), \( j = \max\{J_0, \lceil \log_2 q \rceil\} \). The spectral radius of \( A_N \) is \( \max_{J_0 \leq j \leq J} |a_j N_j^{-1/2}| \). Showing the asymptotic normality result in Equation (3.10) is equivalent to proving the quadratic form

\[
Q_N = \sum_{j=J_0}^J a_j N_j^{-1/2} \sum_{k=1}^{N_j} W_{j,k}^2 = \xi_N^T A_N \xi_N \xrightarrow{D} N(0, \sigma^2),
\]

where \( \sigma^2 \in [0, \infty) \). By Lemma 3.4, this proof proceeds in two steps. We first show the limit of \( \rho(A_N) \rho(\Gamma_N) \) is zero, then show that the limiting variance of \( Q_N \) is finite.

Lemma 3.5. [Moulines et al., 2007b] Let \( m \geq 2 \) be an integer and \( \Gamma \) be a \( m \times m \) covariance matrix. Let \( p \) be an integer between 1 and \( m - 1 \). Let \( \Gamma_1 \) be the top left submatrix with size \( p \times p \) and \( \Gamma_2 \) be the bottom right submatrix with size \( (m - p) \times (m - p) \),

\[
\Gamma_1 = [\Gamma_{i,j}]_{1 \leq i,j \leq p} \quad \text{and} \quad \Gamma_2 = [\Gamma_{i,j}]_{p+1 \leq i,j \leq m}.
\]

Then

\[
\rho(\Gamma) \leq \rho(\Gamma_1) + \rho(\Gamma_2).
\]

Since \( \Gamma_N \) is a covariance matrix, and \( \Gamma_j, j = J_0, \ldots, J \), are its diagonal blocks. By Lemma 3.5, \( \rho(\Gamma_N) \leq \sum_{j=J_0}^J \rho(\Gamma_j) \). Using the result in Lemma 3.3, \( \rho(\Gamma_j) \leq \ldots \)
\[ 2\pi \sup |S_{W_j}(f)| \text{ is bounded, so } \rho(\Gamma_N) \text{ is bounded. Because } \rho(A_N) = \max_j |a_j N_j^{-1/2}| \to 0 \text{ as } N_{j_0} \to \infty, \text{ we have } \lim_{N_{j_0} \to \infty} \rho(A_N)\rho(\Gamma_n) = 0. \]

We want to use the fact that the cross-scale process constructed in Moulines et al. [2007a] is covariance stationary, and decompose the variance of \( Q_N \):

\[
\text{var}(Q_N) = \sum_{j=J_0}^{J} \sum_{j'=J_0}^{J} \frac{a_j}{\sqrt{N_j}} \frac{a_{j'}}{\sqrt{N_{j'}}} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_{j'}-1} \text{cov}(W_{j,k}^2, W_{j',k'}^2)
\]

\[
= \frac{1}{N} \sum_{j=J_0}^{J} \sum_{j'=J_0}^{J} \frac{a_j}{2^{-j/2}} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_{j'}-1} \frac{a_{j'}}{2^{-j'/2}} \text{cov}(W_{j,k}^2, W_{j',k'}^2)
\]

\[
= \frac{1}{N} \sum_{j=J_0}^{J} \sum_{j'=J_0}^{J} \frac{a_j}{2^{-j/2}} \sum_{k=1}^{N_j} \text{cov}(W_{j,k}^2, W_{j,k'})
\]

\[
+ \sum_{j \neq j'} \sum_{k'=1}^{N_{j'}} \frac{a_{j'}}{2^{-j'/2}} \text{cov}(W_{j,k}^2, W_{j',k'})
\]

\[
\equiv A_N + B_N,
\]

where \( A_N = \frac{1}{N} \sum_{j=J_0}^{J} \sum_{k=1}^{N_j} \frac{a_j}{2^{-j/2}} \sum_{k'=1}^{N_{j'}} \text{cov}(W_{j,k}^2, W_{j,k'}) \) is a weighted sum of within-scale wavelet covariances, and

\( B_N = \frac{1}{N} \sum_{j=J_0}^{J} \sum_{k=1}^{N_j} \frac{a_j}{2^{-j/2}} \sum_{j \neq j'} \sum_{k'=1}^{N_{j'}} \frac{a_{j'}}{2^{-j'/2}} \text{cov}(W_{j,k}^2, W_{j',k'}) \) is a weighted sum of cross-scale wavelet covariances. Now

\[ A_N = \frac{1}{N} \sum_{j=J_0}^{J} \sum_{k=1}^{N_j} \frac{a_j}{2^{-j/2}} \sum_{k'=1}^{N_{j'}} \text{cov}(W_{j,k}^2, W_{j,k'})
\]

\[ = \sum_{j=J_0}^{J} \frac{a_j^2}{N_j} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \text{cov}(W_{j,k}^2, W_{j,k'})
\]

\[ = \sum_{j=J_0}^{J} \frac{a_j^2}{N_j} \sum_{h=-(N_j-1)}^{N_j-1} (N_j - |h|) \gamma_{W_j^2}(h)
\]

\[ = \sum_{j=J_0}^{J} \frac{a_j^2}{N_j} \sum_{h=-(N_j-1)}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right) \gamma_{W_j^2}(h).
\]
We previously showed that $\sum_{h=-(N_j-1)}^{N_j-1} \left(1 - \frac{|h|}{N_j}\right) \gamma_{W_j^2}(h) \to S_{W_j^2}(0)$, so $A_N \to \sum_{j=J_0}^{J} a_j^2 S_{W_j^2}(0)$ as $N_{J_0} \to \infty$. Next,

$$B_N = \frac{1}{N} \sum_{j=J_0}^{J} \sum_{k=1}^{N_j} \frac{a_j}{2-j/2} \sum_{j' \neq j}^{N_j-1} \sum_{k'=0}^{N_j-1} \frac{a_j a_j'}{2-j'/2} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_j-1} \text{cov}(W_{j,k}^2, W_{j',k'}^2) \left(1 - \frac{|h|}{N_j}\right) \gamma_{W_j^2}(h) \to S_{W_j^2}(0),$$

where $B_{j,J_0} = \sum_{j_0 \leq j' j < sj} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_j-1} \text{cov}(W_{j,k}^2, W_{j',k'}^2)$. Recall the result that $(W_{j,k}, W_{j,k}(u)^T)^T$ and the corresponding element-wise squared process $(W_{j,k}^2, W_{j,k}(u)^T)^T$ are stationary. Rewrite $B_{j,J_0}$ as:

$$B_{j,J_0} = \sum_{j_0 \leq j' j < sj} \frac{a_j}{2-j'/2} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_j-1} \text{cov}(W_{j,k}^2, W_{j',k'}^2) \left(1 - \frac{|h|}{N_j}\right) \gamma_{W_j^2}(h) \to S_{W_j^2}(0),$$

$$= \sum_{j_0 \leq j' j < sj} \frac{a_j}{2-j'/2} \sum_{k=1}^{N_j} \sum_{k'=0}^{N_j-1} \sum_{r=1}^{2-j'-1} \sum_{u=0}^{2-j'-1} \text{cov}(W_{j,k}^2, W_{j',2-j'-1+u}^2) \left[1^T \int_{-1/2}^{1/2} e^{i2\pi f(k-r)} A_{j,j'-j'}(f) df \right] \left[1^T \int_{-1/2}^{1/2} e^{i2\pi f(k-r)} A_{j,j'-j'}(f) df \right],$$

$$= \sum_{j_0 \leq j' j < sj} \frac{a_j}{2-j'/2} \sum_{h=-(N_j-1)}^{N_j-1} (N_j - |h|) \left[1^T \int_{-1/2}^{1/2} e^{i2\pi f(k-r)} A_{j,j'-j'}(f) df \right],$$

48
where $A_{j,u}(f)$ is the cross-scale SDF of the stationary squared process $(W_{j,k}^2, W_{j,k}^2(u)^T)^T$, as defined in Chapter 2. Thus we have

$$B_N = \frac{2}{N} \sum_{j=J_0}^J \frac{a_j}{2^{-j/2}} B_{j,J_0}$$

$$= 2 \sum_{j=J_0}^J a_j \sum_{j' < j} \frac{a_{j'}}{2^{j'/2-j/2}} \sum_{h=-N_j-1}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right) \left[ 1^T \int_{-1/2}^{1/2} e^{i2\pi f(t-r)} A_{j,j'}(f) df \right]$$

$$= 2 \sum_{j=J_0}^J \frac{a_{j-J_0}}{2^{j-J_0/2}} \sum_{u=0}^N a_{j-u} \sum_{h=-N_j-1}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right) \left[ 1^T \int_{-1/2}^{1/2} e^{i2\pi f(t-r)} A_{j,u}(f) df \right].$$

As discussed earlier, $\sum_{h=-N_j-1}^{N_j-1} \left( 1 - \frac{|h|}{N_j} \right)$ is a Fejér’s kernel that converges to a point mass at zero when $N_j \to \infty$, so $B_N \to 2 \sum_{j=J_0}^J \sum_{u=0}^N a_{j-u} 1^T A_{j,u}(0)$, where $A_{j,u}(0) = \sum_{h \in \mathbb{Z}} \text{cov}(W_{j,k}^2, W_{j,k+h}^2(u))$. Thus we have shown that

$$\lim_{N_{J_0} \to \infty} \text{var}(Q_N) = \sum_{j=J_0}^J a_j \left[ S_{W_j^2}(0) + 2 \sum_{u=0}^j \frac{a_{j-u}}{2^{j-u/2}} 1^T A_{j,u}(0) \right] \in [0, \infty).$$

This completes the proof of $Q_N \to N(0, \sigma^2)$ as $N_{J_0} \to \infty$. Therefore $\sqrt{N_{J_0}}(\rho_{J_0,j_0}(\hat{v}_{j_0} - s_{j_0,0}), \rho_{J_0+1,j_0}(\hat{v}_{j_0+1} - s_{j_0+1,0}), \ldots, \rho_{J,0}(\hat{v}_J - s_{J,0}))$ is jointly asymptotically normal.

We can now derive the explicit expressions for the entries in the limiting covariance matrix $\Sigma$. The diagonal elements are $\lim_{N_{J_0} \to \infty} \text{var}(\sqrt{N_{J_0}}\rho_{j,j}(\hat{v}_j - s_{j,0})) = \lim_{N_{J_0} \to \infty} \text{var}(\sqrt{N_{J_0}}\eta_j) = S_{W_j^2}(0)$ for $j = J_0, \ldots, J$, as we stated earlier.
Recall the wavelet Whittle likelihood under the white noise approximation.

Proof.

\[ l_N^W(\theta) = -\frac{1}{2} \sum_{j=J_0}^J \left( N_j \log(s_{j,0}(\theta)) + \sum_{k=1}^{N_j} W_{j,k}^2 \right) - \frac{1}{2} \sum_{j=J_0}^J N_j \log(2\pi). \]

Let \( d_{j,0}^{(q)}(\theta) = \frac{\partial s_{j,0}(\theta)}{\partial \theta_q}, \) \( e_{j,0}^{(q,q')}(\theta) = \frac{\partial^2 s_{j,0}(\theta)}{\partial \theta_q \partial \theta_{q'}}, q, q' = 1, \ldots, p. \) If \( q = q' \), we write \( e_{j,0}^{(q)}(\theta) \) for \( e_{j,0}^{(q,q')}(\theta). \)

\( e_{j,0}^{(q)}(\theta) \) is a consistent estimator of the true parameter vector \( \theta \), and as \( N \to \infty, \)

\[ \sqrt{N_1} \left[ \hat{\theta}_N - \theta_0 \right] \to N_p(0, \Omega(\theta_0)), \]

where \( \Omega(\theta_0) \) is some \( p \times p \) covariance matrix that depends on \( \theta_0. \)

Proof. Recall the wavelet Whittle likelihood under the white noise approximation model (Equation (3.6) with the starting level \( J_0) \) is

\[ \lim_{N,J_0 \to \infty} \text{cov} \left( \sqrt{N_{j,0}} \rho_{j,J_0} (\hat{s}_j - s_{j,0}), \sqrt{N_{j',0}} \rho_{j',J_0} (\hat{s}_{j'} - s_{j',0}) \right) \]

The above expression is obtained by the fact that the Fejér’s kernel

\[ e_{j,0}^{(q)}(\theta), e_{j,0}^{(q,q')}(\theta) = \frac{\partial^2 s_{j,0}(\theta)}{\partial \theta_q \partial \theta_{q'}}, q, q' = 1, \ldots, p. \]
Assume that the wavelet variance is identifiable on at least one wavelet level, the consistency of the MLE $\hat{\theta}_N$ can be shown following the proof of Theorem 7.1 in Craigmile et al. [2005], and the proof of Theorem 3.7 in Lehmann and Casella [1998], page 447.

We can write the first derivatives of $l_N^W(\theta)$, with respect to the parameter $\theta$, as a linear combination of $\sqrt{N_{J_0}(\hat{\eta}_{J_0} - \eta_{J_0}, \ldots, \hat{\eta}_J - \eta_J)^T}$:

$$
\frac{1}{\sqrt{N_{J_0}}} l_N^{W'}(\theta) = \frac{1}{\sqrt{N_{J_0}}} \left( \frac{1}{2} \sum_{j=J_0}^{J} N_j (\frac{R_j/N_j}{s_{j,0}(\theta)} - 1) \frac{d_{j,0}^{(1)}(\theta)}{s_{j,0}(\theta)} \right) = \frac{1}{2} \left( \begin{array}{cccc}
\frac{d_{J_0,0}^{(1)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(1)}(\theta)}{s_{J,0}(\theta)} \\
\vdots & \ddots & \vdots \\
\frac{d_{J_0,0}^{(p_j)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(p_j)}(\theta)}{s_{J,0}(\theta)}
\end{array} \right) \left( \frac{N_{J_0}}{\sqrt{N_{J_0}}} \frac{R_{J_0}/N_{J_0}}{s_{J_0,0}(\theta)} - 1 \right)
\right) = \frac{1}{2} \left( \begin{array}{cccc}
\frac{d_{J_0,0}^{(1)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(1)}(\theta)}{s_{J,0}(\theta)} \\
\vdots & \ddots & \vdots \\
\frac{d_{J_0,0}^{(p_j)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(p_j)}(\theta)}{s_{J,0}(\theta)}
\end{array} \right) \left( \begin{array}{cccc}
\frac{\rho_{J_0,J_0}}{s_{J_0,0}(\theta)} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \frac{\rho_{J,J_0}}{s_{J_0,0}(\theta)}
\end{array} \right) \sqrt{N_{J_0}} \left( \begin{array}{c}
\hat{\eta}_{J_0} - \eta_{J_0} \\
\vdots \\
\hat{\eta}_J - \eta_J
\end{array} \right).
$$

The linear association is established by a $p \times (J - J_0 + 1)$ matrix $B$,

$$
B = \left( \begin{array}{cccc}
\frac{d_{J_0,0}^{(1)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(1)}(\theta)}{s_{J,0}(\theta)} \\
\vdots & \ddots & \vdots \\
\frac{d_{J_0,0}^{(p_j)}(\theta)}{s_{J_0,0}(\theta)} & \cdots & \frac{d_{J,0}^{(p_j)}(\theta)}{s_{J,0}(\theta)}
\end{array} \right) \left( \begin{array}{cccc}
\frac{\rho_{J_0,J_0}}{s_{J_0,0}(\theta)} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \frac{\rho_{J,J_0}}{s_{J_0,0}(\theta)}
\end{array} \right) = \left( \frac{d_{J_0,0}^{(p_j)}(\theta)}{s_{J_0,0}(\theta)} \right)_{q,j} .
$$

Thus we have that

$$
\frac{1}{\sqrt{N_{J_0}}} l_N^{W'}(\theta) = \frac{1}{2} B \sqrt{N_{J_0}} \left( \begin{array}{c}
\hat{\eta}_{J_0} - \eta_{J_0} \\
\vdots \\
\hat{\eta}_J - \eta_J
\end{array} \right).
$$
As we previously proved in Theorem 3.1 that $\sqrt{N_j}(\hat{\eta} - \eta) \xrightarrow{D} N(0, \Sigma)$, we have the asymptotic result for the first derivative of the wavelet Whittle likelihood:

$$\frac{1}{\sqrt{N_{j_0}}} l_W'(\theta_0) \xrightarrow{D} N_p \left( 0, \frac{1}{4} B\Sigma B^T \right) \quad \text{as} \quad N_{j_0} \to \infty. \quad (3.12)$$

The Hessian matrix (the matrix of second-order partial derivatives) of $l_W^r(\theta)$ is

$$H(\theta) = \frac{\partial^2 l_W^r(\theta)}{\partial \theta^2} = \begin{pmatrix} \frac{\partial^2 l_W^r(\theta)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 l_W^r(\theta)}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l_W^r(\theta)}{\partial \theta_p \partial \theta_1} & \cdots & \frac{\partial^2 l_W^r(\theta)}{\partial \theta_p^2} \end{pmatrix},$$

with the $(q, q')$th element:

$$\frac{\partial^2 l_W^r(\theta)}{\partial \theta_q \partial \theta_{q'}} = \frac{1}{2} \sum_{j = j_0}^J N_j \left( \frac{R_j/N_j}{s_{j,0}(\theta)} - 1 \right) \frac{d_{q,j,0}(\theta)}{s_{j,0}(\theta)} - \frac{1}{2} \sum_{j = j_0}^J N_j \left( \frac{R_j/N_j}{s_{j,0}(\theta)} - 1 \right) \frac{d_{q,j,0}(\theta) d_{q',j,0}(\theta)}{s_{j,0}(\theta)}.$$

As we stated at the beginning of this section, since $\hat{\upsilon}_j$ is an unbiased and consistent estimator of $s_{j,0}(\theta)$, we have $\frac{R_j/N_j}{s_{j,0}(\theta)} \xrightarrow{P} 1$. It is easy to show that the $(q, q')$th element of the Hessian matrix $H(\theta)$ has the following result.

$$\frac{1}{N_{j_0}} \frac{\partial^2 l_W^r(\theta)}{\partial \theta_q \partial \theta_{q'}} \xrightarrow{P} - \sum_{j = j_0}^J 2^{J_0 - j - 1} \frac{d_{q,j,0}(\theta) d_{q',j,0}(\theta)}{s_{j,0}^2(\theta)}, \quad \text{as} \quad N_{j_0} \to \infty.$$

Let $\theta^* = \theta_0 + ch$, where $h = \hat{\theta} - \theta_0$ is small, and $c \in (0, 1)$. With the consistency of $\hat{\theta}$, we have $\theta^* \to \theta_0$ as $N_{j_0} \to \infty$, and the likelihood function has a local maximum at $\theta_0$. Thus $\frac{1}{N_{j_0}} H(\theta^*)$ converges in probability to a non-zero constant matrix $V = \left( - \sum_{j = j_0}^J 2^{J_0 - j - 1} \frac{d_{q,j,0}(\theta_0) d_{q',j,0}(\theta_0)}{s_{j,0}^2(\theta_0)} \right)_{q,q' = 1, \ldots, p}$.

We can prove the non-negative definiteness of the matrix $-V$ by writing it as the product of a matrix $Q$ and its transpose:

$$-V = \left( \sum_{j = j_0}^J 2^{J_0 - j - 1} \frac{d_{q,j,0}(\theta_0) d_{q',j,0}(\theta_0)}{s_{j,0}^2(\theta_0)} \right)_{q,q' = 1, \ldots, p} = Q^T Q,$$
with
\[
Q = \begin{pmatrix}
    \frac{d_{j,0}^{(1)}(\theta_0)}{s_{j,0}(\theta_0)} & \cdots & \frac{d_{j,0}^{(1)}(\theta_0)}{s_{j,0}(\theta_0)} \\
    \vdots & \ddots & \vdots \\
    \frac{d_{J,0}^{(p)}(\theta_0)}{s_{J,0}(\theta_0)} & \cdots & \frac{d_{J,0}^{(p)}(\theta_0)}{s_{J,0}(\theta_0)}
\end{pmatrix}_{p \times (J-J_0+1)} \begin{pmatrix}
    \frac{N_{J_0}}{2N_{J_0}} & \cdots & 0 \\
    \vdots & \ddots & \vdots \\
    0 & \cdots & \frac{N_{J_0}}{2N_{J_0}}
\end{pmatrix}_{(J-J_0+1) \times (J-J_0+1)}^{1/2}.
\]

By a Taylor series expansion (applied element-wise to \(l_N^{W'}(\theta)\), e.g., Tsay, 2013),
\[
l_N^{W'}(\hat{\theta}) = l_N^{W'}(\theta_0) + l_N^{W''}(\theta^*)(\hat{\theta} - \theta_0).
\]

In this expression \(\theta^*\) is defined earlier as \(\theta_0 + c h\), where \(h = \hat{\theta} - \theta_0\) and \(c \in (0, 1)\).

Since \(\hat{\theta}\) is the maximum likelihood estimator: \(l_N^{W'}(\hat{\theta}) = 0\), we have
\[
\sqrt{N_{J_0}}(\hat{\theta} - \theta_0) = -\left(\frac{1}{N_{J_0}} l_N^{W''}(\theta^*)\right)^{-1} \left(\frac{1}{\sqrt{N_{J_0}}} l_N^{W'}(\theta_0)\right) = -\left(\frac{1}{N_{J_0}} H(\theta^*)\right)^{-1} \left(\frac{1}{\sqrt{N_{J_0}}} l_N^{W'}(\theta_0)\right).
\]

We previously showed \(\frac{1}{\sqrt{N_{J_0}}} l_N^{W'}(\theta_0) \xrightarrow{p} N_p(0, \frac{1}{4} \mathbf{B} \Sigma \mathbf{B}^T)\), as \(N_{J_0} \to \infty\), and we have
\[
\lim_{N_{J_0} \to \infty} \frac{1}{N_{J_0}} H(\theta^*) = \mathbf{V}.
\]

Denoting the covariance matrix \(\mathbf{B} \Sigma \mathbf{B}^T = K\), and \(2\mathbf{V} = \Gamma\).

By Slutsky’s theorem [e.g., Resnick, 1999, page 268], the wavelet-based MLEs are asymptotically joint normal,
\[
\sqrt{N_{J_0}}(\hat{\theta} - \theta_0) \xrightarrow{p} N_p(0, \Omega(\theta_0)),
\]
where the limiting covariance matrix \(\Omega(\theta_0) = \frac{1}{4} \mathbf{V}^{-1} \mathbf{B} \Sigma \mathbf{B}^T (\mathbf{V}^{-1})^T = \Gamma^{-1} K \Gamma^{-1}^T\).

\[\square\]

### 3.4 An AR(1) model

#### 3.4.1 The approximate likelihood

We can apply AR(1) approximations to the DWT coefficients at every level. Assume for level \(j, j = 1, \ldots, J\), the wavelet coefficients follow
\[
W_{j,k} = \phi_j(\theta) W_{j,k-1} + Z_{j,k-1}.
\]
Let $s_{j,0}(\theta) = \text{var}(W_{j,k})$, $s_{j,1}(\theta) = \text{cov}(W_{j,k}, W_{j,k-1})$, then
\[
\phi_j(\theta) = \frac{s_{j,1}(\theta)}{s_{j,0}(\theta)}.
\]
The white noise term $Z_{j,k-1}$ has variance $s_{j,0}(\theta)(1 - \phi_j(\theta))$. Under this AR(1) approximation, the covariance matrix $T_N(S_{j,\phi})$ at level $j$ reduce to a tridiagonal matrix.

The corresponding wavelet Whittle log-likelihood function can be written as a function of the parameter $\theta$, sums of squared wavelet coefficients, and the sums of lag 1 wavelet cross products:
\[
\begin{align*}
W_N(\theta) &= \sum_{j=J_0}^{J} \left( \log f(W_{j,1}; \theta) + \log f(W_{j,2}|W_{j,1}; \theta) + \cdots + \log f(W_{j,N_j}|W_{j,N_j-1}; \theta) \right) \\
&= -\frac{1}{2} \sum_{j=1}^{J} \left[ N_j \log s_{j,0}(\theta) + (N_j - 1) \log(1 - \phi_j^2) + \frac{1}{s_{j,0}(\theta)(1 - \phi_j^2)} \right. \\
&\quad \times \left( W_{j,1}^2 + W_{j,N_j}^2 + (1 + \phi_j^2) \sum_{k=2}^{N_j-1} W_{j,k}^2 - 2\phi_j \sum_{k=2}^{N_j} W_{j,k} W_{j,k-1} \right) \\
&\quad - \frac{N - 1}{2} \log 2\pi.
\end{align*}
\]
(3.14)
The maximizer of Equation (3.14) is the wavelet-based Whittle estimator under the AR(1) approximation model.

3.4.2 Discussion about the MLEs under the AR(1) model

We believe that the MLEs obtained under the AR(1) approximation model in this section also have good asymptotic properties. In the Monte Carlo studies presented in the next section, we implement the AR(1) model and it shows that the resulting MLEs performs better than the WN model. Follow the same steps as for WN model, the asymptotic results of the MLEs relies on a central limit result of the sums of
squared wavelet coefficients, and sums of lag-1 products at the same level. Let

\[ D_j(W, \phi_j(\theta)) = W_{j,1}^2 + W_{j,N_j}^2 + (1 + \phi_j^2) \sum_{k=2}^{N_j-1} W_{j,k}^2 - 2\phi_j \sum_{k=2}^{N_j} W_{j,k} W_{j,k-1}, \]

\[ d_{j,0}^{(q)} = \frac{\partial}{\partial \theta_q} s_{j,0}(\theta), \quad \text{and} \quad e_{j,0}^{(q,q')} = \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} s_{j,0}(\theta), \]

for each \( j = J_0, \ldots, J \), \( J_0 \) is a fixed integer and takes value between 1 and \( J \) (including 1). When \( q = q' \), we write \( e_{j,0}^{(q)} \) for \( e_{j,0}^{(q,q')} \). We also let \( a_j^{(q)} = \frac{\partial}{\partial \theta_q} \phi_j(\theta) \), and \( b_j^{(q,q')} = \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} \phi_j(\theta) \), for \( q, q' = 1, \ldots, p \).

We need the first and second derivatives of the log-likelihood function with respect to \( \theta \). The expressions of the first derivative is

\[
\frac{\partial}{\partial \theta_q} l_N^W(\theta) = -\frac{1}{2} \sum_{j=J_0}^{J} N_j \frac{d_{j,0}^{(q)}}{s_{j,0}(\theta)} + \sum_{j=J_0}^{J} (N_j - 1) \frac{\phi_j a_j^{(q)}}{1 - \phi_j^2} \\
+ \frac{1}{2} \sum_{j=J_0}^{J} \frac{d_{j,0}^{(q)}(1 - \phi_j^2)}{s_{j,0}^2(\theta)(1 - \phi_j^2)^2} - 2s_{j,0}(\theta) \phi_j a_j^{(q)} \frac{D_j(W, \phi_j(\theta))}{2} \\
- \frac{1}{2} \sum_{j=J_0}^{J} s_{j,0}(\theta) \frac{1}{(1 - \phi_j^2)} \left[ 2\phi_j a_j^{(q)} \sum_{k=2}^{N_j-1} W_{j,k}^2 - 2\phi_j \sum_{k=2}^{N_j} W_{j,k} W_{j,k-1} \right].
\]

The second derivatives \( 1/\sqrt{N_{j_0}} l_N^W(\theta) \) converge to some expressions depending on \( \theta \) when \( N_{j_0} \) is large. However, the expression is too complicated that we can not show analytically the Hessian matrix converges to a non-positive definite matrix with rate \( N_{j_0} \). Element-wise,

\[
E \left( \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} l_N^W(\theta) \right) \xrightarrow{p} -\sum_{j=J_0}^{J} \frac{N_j}{N_{j_0}} \left[ \frac{1 + \phi_j^2}{(1 - \phi_j^2)^2} a_j^{(q)} a_j^{(q')} + \frac{d_{j,0}^{(q)} d_{j,0}^{(q')}}{2s_{j,0}^2} \right] \\
- \frac{2\phi_j}{s_{j,0}(1 - \phi_j^2)} \left( a_j^{(q)} d_{j,0}^{(q')} + a_j^{(q')} d_{j,0}^{(q)} \right).
\]

If we compare the log-likelihood and its derivatives with the results in white noise model, \( D_j(W, \phi_j(\theta)) \) seems to play the role of \( R_j \). It is necessary to study the
properties of $D_j/N_j$ for large $N_j$:

$$E\left(\frac{D_j(W, \phi_j(\theta))}{N_j}\right) \xrightarrow{P} s_{j,0}(\theta)(1 - \phi_j^2),$$  \hspace{1cm} (3.15)

and

$$\text{var}\left(\frac{D_j(W, \phi_j(\theta))}{N_j}\right) = \frac{1}{N_j^2} \text{var}\left(\sum_{k=1}^{N_j} Z_{j,k}^2\right) \xrightarrow{P} S_{Z_j}(0).$$  \hspace{1cm} (3.16)

Here $S_{Z_j}(0)$ is the spectral density function of the noise term $Z_j^2$ in (3.13) at zero frequency. With the above results, it is possible to get $E\left(\frac{1}{\sqrt{N_j}} l_{j0}^W(\theta)\right)$ converges to 0 as $N_{j_0} \to \infty$. If one can show $(D_{j_0}/N_{j_0}, \ldots, D_j/N_j)^T$ converges jointly to a multivariate normal distribution, the limiting result for the wavelet-based MLEs can be derived.

Although this result is not currently available for DWT coefficients, we can show it with the maximum overlap discrete wavelet transform (MODWT) coefficients. By replacing the DWT coefficients in the likelihood function with the MODWT coefficients, the MLEs are consistent and joint asymptotically normal. We will introduce the MODWT based estimation and the asymptotic results next.

First, define the MODWT of a process $\{Y_t\}$ [Percival and Walden, 2000]:

$$\tilde{W}_{j,k} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{k-l} \quad \text{and} \quad \tilde{V}_{j,k} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} Y_{k-l},$$  \hspace{1cm} (3.17)

for $k = 1, 2, \ldots$, and level $j$ a positive integer. In the above equations, $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ are the level $j$ MODWT wavelet and scaling filter, respectively. The filters are obtained via the DWT filters:

$$\tilde{h}_{j,l} = \frac{h_{j,l}}{2^j} \quad \text{and} \quad \tilde{g}_{j,l} = \frac{g_{j,l}}{2^j}, \quad l = 0, \ldots, L_j,$$

with the same filter length $L_j = (2^j - 1)(L - 1) + 1$ as for the DWT. Compared to the DWT, the MODWT does not have the step of downsampling, and is using a...
rescaled version of the wavelet and scaling filters. Unlike the DWT, the MODWT is not an orthonormal transformation, but it still can be used to decompose the process variance to the wavelet variances over different scales [Percival and Walden, 2000].

When we have a finite sample, the MODWT is circularly filtering the data. The MODWT wavelet coefficients are given by

\[ \tilde{W}_{j,k} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{k-l} \pmod{N} \]

and

\[ \tilde{V}_{j,k} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} Y_{k-l} \pmod{N}, \]

for \( k = 1, \cdots, N \). One advantage of the MODWT is that it can be applied to any sample size \( N \). Another advantage is that the MODWT is time invariant: if the MODWT of \( Y \) is \( \tilde{W}_1, \cdots, \tilde{W}_J, \tilde{V}_J \), then MODWT of the (time) shifted data \( T^m Y \) is \( T^m \tilde{W}_1, \cdots, T^m \tilde{W}_J, T^m \tilde{V}_J \).

Mondal and Percival [2010] has shown the MODWT version of Theorem 3.1 with the MODWT non-boundary coefficients. The level \( j \) empirical wavelet variance \( \hat{v}_j \) in the theorem is replaced by \( \sum_k \tilde{W}_{j,k}^2 / M_j \), where \( M_j = N - L_j \) is the number of the level \( j \) non-boundary wavelet coefficients. We restate their result: the averaged sum of squared within-scale non-boundary MODWT coefficients, \( \left( \frac{1}{\sqrt{M_{j_0}}} \sum_k \tilde{W}_{j_0,k}^2, \cdots, \frac{1}{\sqrt{M_{j}}} \sum_k \tilde{W}_{j,k}^2 \right) \), is jointly asymptotically normal as \( M_{j_0} \to \infty \). To show that the wavelet Whittle MLEs of a AR(1) approximation model is consistent and asymptotically normal, we need a more general result.

**Theorem 3.6.** Suppose that \( L \geq 2d \). The averaged sum of squared non-boundary MODWT coefficients and their averaged lag-1 cross-products, \( \frac{1}{\sqrt{M_{j_0}}} \sum_k \tilde{W}_{j_0,k}^2, \cdots, \frac{1}{\sqrt{M_j}} \sum_k \tilde{W}_{j,k}^2, \frac{1}{\sqrt{M_{j_0}}} \sum_k \tilde{W}_{j_0,k} \tilde{W}_{j_0,k-1}, \cdots, \frac{1}{\sqrt{M_j}} \sum_k \tilde{W}_{j,k} \tilde{W}_{j,k-1} \), is jointly asymptotically normal with the mean \( (s_{j_0,0}, \cdots, s_{j_0,1}, \cdots, s_{j,1}) \), as \( M_{j_0} \to \infty \).
The proof of the above theorem is in Appendix A. Let $\hat{\theta}_{MODWT}$ be the wavelet Whittle MLE based on MODWT coefficients. Proposition 3.7 presents the asymptotic result for $\hat{\theta}_{MODWT}$ under the wavelet (MODWT) Whittle AR(1) approximation model.

**Proposition 3.7.** Suppose that $L \geq 2d$. Then the $p$-dimensional vector $\hat{\theta}_{MODWT}$ is a consistent estimator of the true parameter vector $\hat{\theta}_0$, and as $N \to \infty$,

$$\sqrt{N_1} \left[ \hat{\theta}_{MODWT} - \theta_0 \right] \xrightarrow{D} N_p(0, \Sigma(\theta_0)),$$

where $\Sigma(\theta_0)$ is some $p \times p$ covariance matrix that depends on $\theta_0$.

Here we briefly discuss the sketch of the proof for Proposition 3.7. Because $(D_{j_0}/\sqrt{M_{j_0}}, \ldots, D_J/\sqrt{M_J})^T$ is a linear combination of the vector $(\frac{1}{\sqrt{M_{j_0}}} \sum_k \tilde{W}_{j_0,k}^2, \ldots, \frac{1}{\sqrt{M_{j_0}}} \sum_k \tilde{W}_{j_0,k} \tilde{W}_{j_0,k-1}, \ldots, \frac{1}{\sqrt{M_J}} \sum_k \tilde{W}_{J,k} \tilde{W}_{J,k-1})$, which has the asymptotic result stated in Theorem 3.6, the vector of $D_j/\sqrt{M_j}$ thus converges in distribution to a multivariate normal distribution. Follow the previous discussion about the wavelet Whittle MLEs under an AR(1) approximation, this means the first derivatives of the wavelet Whittle likelihood are jointly normal. By a Taylor series expansion and Slutsky’s theorem, the wavelet based MLE $\hat{\theta}_{MODWT}$ under an AR(1) approximation model is also consistent and asymptotically normal with rate $\sqrt{M_{j_0}}$.

### 3.5 Choices of $J_0$

The proofs in Section 3.3 shows the wavelet Whittle MLEs obtained from the wavelet coefficients on levels $J_0, \ldots, J$ $(1 < J_0 < J)$ still follow the central limit result in Theorem 3.2. While the value of $J$ is determined by the sample size $N$, the choices of $J_0$ is related to the process, and need to be carefully made. One example of using
$J_0 > 1$ is based on a smooth Matérn process with $\nu = 1.5$. As shown in Figure 2.4, the level 1 wavelet spectrums (the black solid lines) and level 2 wavelet spectrums (the red dashed lines) increase steadily at low frequencies, then approach a climax and stop increasing in the bottom right three panels ($\nu = 1.5$ case). Neither the AR(1) nor WN spectra are good approximation to the lower level wavelet spectra. In this situation, whether or not to include the level 1 and level 2 wavelet coefficients in the approximation model is an open question. We will answer this question in the next section with a simulation study.

### 3.6 Monte Carlo studies

We now conduct some simulation studies to investigate the performance of the wavelet Whittle MLEs with finite samples. Two families of processes are studied in this section: the FD processes with $\sigma^2 = 1$ and Gaussian processes with a Matérn covariance (defined in Equation (2.11)) with parameters $\phi = 0.25$, $\sigma^2 = 1$ and smoothness parameter $\nu$ that takes values 0.25, 0.5, and 1.5. The FD processes are simulated using the Davies-Harte algorithm [Craigmile, 2003], and the Matérn data are simulated using the Levinson-Durbin algorithm [e.g., Brockwell and Davis, 1991]. We will use the statistical software package R to implement the wavelet Whittle models, and estimate the process parameters ($\phi$, $\sigma^2$ for the Matérn, and $d$ for the FD process).

#### 3.6.1 The wavelet-based Whittle MLEs

Suppose the sample series has length $N = 1024 (2^{10})$, for the Matérn process. Our purpose is to obtain the wavelet-Whittle estimates $\hat{\phi}$ and $\hat{\sigma}^2$. Uncertainty of the estimates is measured by standard deviations (SD) and root mean square errors (RMSE) in 10000 replications. Combinations of two approximation models, white
noise and AR(1), with various wavelet filters are applied to the simulated data. Table 3.1 and Table 3.2 show simulation results of the process with smoothness parameter $\nu = 1.5$. The wavelet Whittle MLEs are based on wavelet coefficients from level 3 to level $J = 10$. We do not use the first two levels because neither the white noise nor the AR(1) SDF is a good approximation of the lower level wavelet SDFs of a smooth Matérn process ($\nu = 1.5$). In Figure 2.4 (Chapter 2, page 30) the wavelet SDFs of level 1 and level 2 first increase then stay roughly constant, which are not close to the shape of WN or AR(1) spectrum. Additionally, when we use all-level-coefficients and apply the WN and the AR approximation models, the estimation results are not good in terms of bias, and the advantage of incorporating within-scale dependence (use AR model) is not obvious (see Table 3.3). But, it is demonstrated in the simulation (Table 3.1 and Table 3.3) that by dropping the first two levels wavelet coefficients and using only the level 3 to $J$ coefficients for the estimation, the AR(1) approximations perform better than WN. Both $\hat{\phi}$ and $\hat{\sigma}^2$ have significantly smaller SD and RMSE values under AR(1) approximation models (compared to using all levels coefficients, i.e., $J_0 = 1$). This answers the question we asked in the previous section: in the situation that the WN and the AR(1) are not good approximations to certain lower level spectrums, including the lower level wavelet coefficients in the wavelet Whittle model does not help improve the performance of the estimators, but brings in larger bias.

Comparing across filters, as the filter length $L$ increases, the estimation result gets better: they have significantly smaller bias, and small SDs and RMSEs (see also, Craigmile et al. [2005]). We find this pattern appearing repeatedly when we vary the parameter values of the Matérn process, i.e. $\nu = 0.25$ and 0.5, and when we
simulate and model processes other than Matérn. Particularly, with the same length $L = 8$, $LA(8)$ performs better than $D(8)$: in Table 3.1 and Table 3.2, the bias, SD and RMSEs are the smallest if with the $LA(8)$ filter. In conclusion, the combination of $LA(8)$ filter and AR(1) approximation is the best based on the results in Table 3.1 and Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EST</td>
<td>SE</td>
<td>EST</td>
</tr>
<tr>
<td>WN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D(4)$</td>
<td>0.0044</td>
<td>0.0003</td>
<td>0.0276</td>
</tr>
<tr>
<td>$D(6)$</td>
<td>0.0050</td>
<td>0.0003</td>
<td>0.0259</td>
</tr>
<tr>
<td>$D(8)$</td>
<td>0.0043</td>
<td>0.0002</td>
<td>0.0255</td>
</tr>
<tr>
<td>$LA(8)$</td>
<td>0.0028</td>
<td>0.0002</td>
<td>0.0252</td>
</tr>
<tr>
<td>AR(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D(4)$</td>
<td>0.0054</td>
<td>0.0002</td>
<td>0.0247</td>
</tr>
<tr>
<td>$D(6)$</td>
<td>0.0056</td>
<td>0.0003</td>
<td>0.0224</td>
</tr>
<tr>
<td>$D(8)$</td>
<td>0.0038</td>
<td>0.0002</td>
<td>0.0208</td>
</tr>
<tr>
<td>$LA(8)$</td>
<td>0.0029</td>
<td>0.0002</td>
<td>0.0208</td>
</tr>
</tbody>
</table>

Table 3.1: Matérn process: results for $\hat{\phi}$ using wavelet coefficients on Levels 3 to $J$

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EST</td>
<td>SE</td>
<td>EST</td>
</tr>
<tr>
<td>WN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D(4)$</td>
<td>0.0610</td>
<td>0.0027</td>
<td>0.287</td>
</tr>
<tr>
<td>$D(6)$</td>
<td>0.0612</td>
<td>0.0027</td>
<td>0.2611</td>
</tr>
<tr>
<td>$D(8)$</td>
<td>0.0545</td>
<td>0.0025</td>
<td>0.2509</td>
</tr>
<tr>
<td>$LA(8)$</td>
<td>0.0363</td>
<td>0.0026</td>
<td>0.2458</td>
</tr>
<tr>
<td>AR(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D(4)$</td>
<td>0.0676</td>
<td>0.0026</td>
<td>0.245</td>
</tr>
<tr>
<td>$D(6)$</td>
<td>0.0637</td>
<td>0.0021</td>
<td>0.2112</td>
</tr>
<tr>
<td>$D(8)$</td>
<td>0.0394</td>
<td>0.0019</td>
<td>0.1833</td>
</tr>
<tr>
<td>$LA(8)$</td>
<td>0.0268</td>
<td>0.0017</td>
<td>0.1788</td>
</tr>
</tbody>
</table>

Table 3.2: Matérn process: results for $\hat{\sigma}^2$ using wavelet coefficients on Levels 3 to $J$
Table 3.3: Matérn process: results for $\hat{\phi}$ using wavelet coefficients on Levels 1 to $J$

Simulation time per one replicate (Table 3.4) is also recorded to compare computation efficiency of different wavelet-Whittle models with various filters. The same models are applied to analyze three different processes: Matérn with $\nu = 0.5$, AR(1), ARMA(1, 1). The simulation time showed in the table are calculated by running 10 replicates using mcapply() function in R, with 10 cores, then calculating the average time used per replicate. As we expect, a simpler model (white noise) with shorter

<table>
<thead>
<tr>
<th>Process</th>
<th>Matérn</th>
<th>AR(1)</th>
<th>ARMA(1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D(4)</td>
<td>1.772</td>
<td>1.732</td>
</tr>
<tr>
<td>WN</td>
<td>D(6)</td>
<td>3.149</td>
<td>3.088</td>
</tr>
<tr>
<td></td>
<td>D(8)</td>
<td>4.949</td>
<td>4.697</td>
</tr>
<tr>
<td></td>
<td>LA(8)</td>
<td>4.656</td>
<td>4.651</td>
</tr>
<tr>
<td>AR(1)</td>
<td>D(4)</td>
<td>8.399</td>
<td>7.962</td>
</tr>
<tr>
<td></td>
<td>D(8)</td>
<td>20.375</td>
<td>19.316</td>
</tr>
<tr>
<td></td>
<td>LA(8)</td>
<td>20.615</td>
<td>19.221</td>
</tr>
</tbody>
</table>

Table 3.4: Compare simulation time (in seconds).
filter takes the shortest simulation time. As length of the filter becomes larger, simulation time gets longer. AR(1) models are about 4 times slower than WN models in analyzing the same process using the same filter. There is no big difference in time between the three types of process we used. The simulation time may be larger if we try a more complicated process, such as fractionally exponential process or Matérn process with \( \nu = 1.5 \).

### 3.6.2 Comparison to other approximate methods

As introduced in Chapter 1, there are other approximation methods for parameter estimations. In order to compare across these methods, a fractionally differenced (FD) process with \( d = 0.4 \) and \( \sigma^2 = 1 \) is simulated and analyzed to study the difference between our wavelet Whittle approximation model and other models (the exact likelihood method, Markov approximation, and the one-taper method), in terms of estimation performance.

We first define the fractionally differenced (FD) process [Percival and Walden, 2000, page 281]: for \(-1/2 < d < 1/2\), the process \( \{Y_t : t \in \mathbb{Z}\} \) is a fractionally differenced process if

\[
(1 - \nabla)^d Y_t = Z_t, \tag{3.18}
\]

where \( \nabla \) is the lag one differencing operator (as defined in Chapter 2), and \( Z_t \) is a white noise process with mean zero and variance \( \sigma^2 \). The spectral density of \( Y_t \) is

\[
S_Y(f) = \frac{\sigma^2}{[4 \sin^2(\pi f)]^d}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \tag{3.19}
\]

When \( 0 < d < 1/2 \) the FD process is stationary, so under our setting of \( d = 0.4 \), \( \{Y_t\} \) is a stationary long memory process. Table 3.5 summarizes the simulation results.
The eight wavelet Whittle models (WN and AR(1) approximations with 4 different filters) all have significantly smaller bias than Markov model with order 1 and the tapering model. Some wavelet models, the WN model with $D(4)$ filter, the AR(1) models with $D(8)$ and with $LA(8)$ filters, can beat the exact likelihood results in terms of bias. For the tapering method, we use Wendland_1 taper as suggested in Furrer et al. [2006],

$$
\left(1 - \frac{h}{\theta}\right)^4 \left(1 + 4\frac{h}{\theta}\right), \text{ where } x_+ = \max\{0, x\},
$$

with different values of $\theta$. The range parameter $\theta$ equals the number of nonzero tapered autocovariances. The results of tapering models in Table 3.5 have significantly larger bias and RMSE than the other methods. And a drawback of tapering method is that there is no clear guide in choosing $\theta$. A larger $\theta$ does not necessarily give a better estimate since tapering is also introducing bias to the nonzero autocovariances. As we change the process, the optimal $\theta$ value changes. For some time series processes, it is not appropriate to use the tapering approximation. For example, when we apply the tapering method to estimate the parameter $\phi$ of a Matérn process with $\nu = 1.5$, the bias is around 0.75 while the true $\phi$ is 0.25. The results of the two-taper method [Kaufman et al., 2008] is worse than the one-taper method (with huge bias and RMSEs).

The values in the last column of Table 3.5 are the simulation time for 10000 replicates using 6 cores in mcapply() function in R. The time for exact likelihood method and Markov method are much smaller than the other two methods since it is using C code. Both the wavelet Whittle and tapering code are written in R. Among all tapering models in this simulation study, when $\theta$ in the Wendland taper is 5, it takes the shortest simulation time (1959.739 seconds). However, using wavelet
<table>
<thead>
<tr>
<th>Method</th>
<th>Bias EST</th>
<th>Bias SD</th>
<th>Bias RMSE EST</th>
<th>Bias RMSE SE</th>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>-0.0021</td>
<td>0.0002</td>
<td>0.0233</td>
<td>0.0002</td>
<td>39.096</td>
</tr>
<tr>
<td>Markov</td>
<td>-0.0125</td>
<td>0.0003</td>
<td>0.0334</td>
<td>0.0002</td>
<td>6.680</td>
</tr>
<tr>
<td>Tapering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta = 2$</td>
<td>-0.0260</td>
<td>0.0002</td>
<td>0.0230</td>
<td>0.0002</td>
<td>2180.538</td>
</tr>
<tr>
<td>$\theta = 5$</td>
<td>-0.0200</td>
<td>0.0003</td>
<td>0.0282</td>
<td>0.0002</td>
<td>1959.739</td>
</tr>
<tr>
<td>$\theta = 10$</td>
<td>-0.0212</td>
<td>0.0003</td>
<td>0.0287</td>
<td>0.0002</td>
<td>2588.371</td>
</tr>
<tr>
<td>$\theta = 30$</td>
<td>-0.0225</td>
<td>0.0003</td>
<td>0.0285</td>
<td>0.0002</td>
<td>2825.827</td>
</tr>
<tr>
<td>$\theta = 60$</td>
<td>-0.0200</td>
<td>0.0003</td>
<td>0.0275</td>
<td>0.0002</td>
<td>5805.882</td>
</tr>
<tr>
<td>$\theta = 100$</td>
<td>-0.0184</td>
<td>0.0003</td>
<td>0.0259</td>
<td>0.0002</td>
<td>7207.526</td>
</tr>
<tr>
<td>Wavelet WN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D(4)</td>
<td>0.0007</td>
<td>0.0003</td>
<td>0.0273</td>
<td>0.0002</td>
<td>946.479</td>
</tr>
<tr>
<td>D(6)</td>
<td>0.0011</td>
<td>0.0003</td>
<td>0.0272</td>
<td>0.0002</td>
<td>1583.593</td>
</tr>
<tr>
<td>D(8)</td>
<td>0.0014</td>
<td>0.0003</td>
<td>0.0265</td>
<td>0.0002</td>
<td>2222.319</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.0014</td>
<td>0.0003</td>
<td>0.0266</td>
<td>0.0002</td>
<td>2216.814</td>
</tr>
<tr>
<td>Wavelet AR(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D(4)</td>
<td>0.0012</td>
<td>0.0003</td>
<td>0.0262</td>
<td>0.0002</td>
<td>1108.934</td>
</tr>
<tr>
<td>D(6)</td>
<td>0.0015</td>
<td>0.0003</td>
<td>0.0262</td>
<td>0.0002</td>
<td>1663.136</td>
</tr>
<tr>
<td>D(8)</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0258</td>
<td>0.0002</td>
<td>2303.738</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.0009</td>
<td>0.0003</td>
<td>0.0263</td>
<td>0.0002</td>
<td>2331.613</td>
</tr>
</tbody>
</table>

Table 3.5: FD process with $d = 0.4$. $N = 1024$, replication = 10000, cores = 6.

Whittle method with D(4)( or D(6)) filter only takes 50% (or 75%) of the time. Our model is computationally more efficient than tapering model.

### 3.6.3 The limiting variance

We also check the performance of MLEs under different sample sizes. In the simulation study, RMSEs of $\hat{\phi}$ were calculated already. The empirical variance of estimator $\sqrt{N_{J_0}}\hat{\phi}$ can be calculated using $N_{J_0}RMSE^2$ (or $M_{J_0}RMSE^2$ if using non-boundary coefficients). The theoretical limiting variance is given by the Equation (3.9) in this chapter:

$$
\sigma^2_{\phi} = \left( \sum_{j=J_0}^{J} 2^{J_0-j} \frac{d^2_j(\phi_0)}{s^2_j(\phi_0)} \right)^{-2} \cdot \sum_{j=J_0}^{J} \sum_{j'=J_0}^{J} \frac{\rho_{j,J_0} \rho_{j',J_0} d_j(\phi_0) d_{j'}(\phi_0)}{s^2_j(\phi_0) s^2_{j'}(\phi_0)} \Sigma_{j,j'},
$$

where $\Sigma_{j,j'}$ is the $(j, j')$th element of $\Sigma$ in Theorem 3.1. We can compare this limiting variance with the empirical value to check the variance of the estimator under finite
sample sizes. Values in Table 3.6 and Table 3.7 are results from WN approximation models.

### Table 3.6: Limiting variance: all coefficients

<table>
<thead>
<tr>
<th>$N_J \text{RMSE}^2$ (empirical variance)</th>
<th>LA(8)</th>
<th>D(8)</th>
<th>D(6)</th>
<th>D(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limiting variance from equation (3.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 1024$</td>
<td>0.2367</td>
<td>0.2433</td>
<td>0.2411</td>
<td>0.2456</td>
</tr>
<tr>
<td>$N = 2048$</td>
<td>0.2329</td>
<td>0.2331</td>
<td>0.2346</td>
<td>0.2394</td>
</tr>
<tr>
<td>$N = 4096$</td>
<td>0.2322</td>
<td>0.2300</td>
<td>0.2333</td>
<td>0.2383</td>
</tr>
</tbody>
</table>

### Table 3.7: Limiting variance: non-boundary coefficients

<table>
<thead>
<tr>
<th>$M_J \text{RMSE}^2$ (empirical variance)</th>
<th>LA(8)</th>
<th>D(8)</th>
<th>D(6)</th>
<th>D(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limiting variance from equation (3.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 1024$</td>
<td>0.2353</td>
<td>0.2419</td>
<td>0.2397</td>
<td>0.244</td>
</tr>
<tr>
<td>$N = 2048$</td>
<td>0.3424</td>
<td>0.3518</td>
<td>0.3140</td>
<td>0.2902</td>
</tr>
<tr>
<td>$N = 4096$</td>
<td>0.2918</td>
<td>0.2903</td>
<td>0.2748</td>
<td>0.2670</td>
</tr>
</tbody>
</table>

If we use all coefficients (both boundary and non-boundary ones), the empirical variances are smaller than theoretical values with different sample sizes. As the length $N$ of the process increases the theoretical limiting variance is getting closer to the empirical variance $N_J \text{RMSE}^2$. In Table 3.7, the limiting variances based on non-boundary wavelet coefficients are close to the empirical variance if the length of the series $N$ is large, but much slower. This is because we are only using nonboundary wavelet coefficients, and we lose some information from the data. Thus it requires a larger length $N$ for the variance to “converge”. Compare across filters, the variance
tends to be smaller if length of the filter is larger when using all coefficients: with LA(8) filter and \( N = 4096 \), the variance is 0.2318, while with D(4) filter and same value of \( N \), the variance is 0.2359. However, this is not true if we consider only non-boundary wavelet coefficients (Table 3.7), which may still due to the loss of information.

### 3.7 Southern Oscillation Index data

Southern Oscillation was first brought up by Gilbert Walker in early 20th century. Researchers use Southern Oscillation Index (SOI) to record the fluctuation in sea level atmospheric pressure between eastern and western Pacific. The index is calculated from the air pressure difference between Tahiti and Darwin. Prolonged negative SOI values usually indicate that there will be El Nino, while the positive values are related to La Nina. Studying SOI data can help assess the relationship between southern oscillation, and temperature, pressure, precipitation changes in tropical Pacific. The association between SOI values, El Nino, and global climate changes is also of great interest for meteorologists.

The dataset is the SOI signal from National Center for Atmospheric Research (NCAR), and shown in Figure 3.1. We truncate the data to \( N = 1664 = 13 \times 2^7 \) monthly SOI values (from October 1874 to May, 2013), so that the sample size is a multiple of a power of two. Let \( Y_t \) denote the monthly SOI, the time index \( t \) take values 1874.75, \ldots, 2013.33, where \( t = 1875.00 \) represents 1895 January, and \( t = 2013\frac{1}{12} = 2012.75 \) represents May 2013. There is no missing data, but it is suggested by NCAR website that one should be cautious when using the data prior
to 1935, because there are some problems about the quality of the data from Tahiti prior to 1935.

Figure 3.1: Monthly SOI signals from October 1874 to May, 2013; the SOI signals are standardized by the monthly standard deviations.

In the exploratory analysis, we first fit an ANOVA model with a monthly effect, but none of the terms are significant. We standardize the SOI values by their monthly standard deviation. In Figure 3.1, most of the standardized SOI values lie between -4 and 4. There are some very low values (smaller than -6) around 1982 (one of the strongest El Nino of the century) and 2004. Then we perform the DWT with the LA(8) filter, and the empirical spectra of the wavelet and scaling coefficients are in
Figure 3.2. The wavelet coefficients are highly decorrelated and have spectra close to white noise spectra (are very flat), which motivates us to implement the wavelet Whittle approximation to perform analysis.

Figure 3.2: The empirical spectrum of the DWT coefficients (with the LA(8) filter, \( J = 7 \)).

The candidate models considered for the standardized data are Matérn (defined in Equation (2.11)), FD (defined in Equation (3.18)), autoregressive (AR) models of order \( p \), ARFIMA, and fractionally exponential models. Now we define AR\((p)\) models, ARFIMA models, and fractionally exponential models. If \( \{Y_t : t \in \mathbb{Z}\} \) is an
AR($p$) process, then

\[ Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + Z_t, \]

where $Z_t$ is a mean zero white noise process with variance $\sigma^2$, and $\phi_k$'s are the AR coefficients [Percival and Walden, 2000, page 268]. When the sampling interval $\Delta = 1$, the SDF is given by

\[ S_Y(f) = \frac{\sigma^2}{|1 - \sum_{k=1}^{p} \phi_k e^{-i2\pi fk}|^2}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \]

An ARFIMA($p, d, q$) model is defined by

\[ (1 - \sum_{k=1}^{p} \phi_k B^k)(1 - \nabla)^d Y_t = (1 + \sum_{k'=1}^{q} \theta_{k'} B^{k'}) Z_t, \]

where $\phi_k$ and $\theta_{k'}$ are the autoregressive and moving average coefficients, respectively, $Z_t$ is a mean zero white noise process, $B$ is the backward shift operator defined by $BY_t = Y_{t-1}$, $\nabla$ is the lag one differencing operator. Compared with FD models, ARFIMA models can model the small scale properties of a process [Percival and Walden, 2000, page 366]. A fractional exponential model is defined Beran [1993] through the spectral density function:

\[ S_Y(f) = \sigma^2 \exp \left\{ 2 \sum_{r=0}^{p} \theta_r \cos(2\pi rf) \right\}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \]

This model is in the form of truncated Fourier series, and can be used to approximate various processes. In this application we use the wavelet Whittle approximation method (via the likelihood in Equation (3.5)) to estimate the parameters of the Matérn and the FD model. Then we use the least square methods to estimate the parameters of the AR($p$), the ARFIMA, and the fractional exponential model.

First we try the traditional AR($p$) model. A plot of the sample partial autocorrelation sequence (Figure 3.3) shows a cut-off at lag 5, so the order of the AR model
is $p = 5$. The estimated AR coefficients are (0.4590, 0.1613, 0.0641, 0.0292, 0.0436), and $\hat{\sigma}^2 = 0.5755$. The fitted SDF is shown in Figure 3.4.

Figure 3.3: PACF of the standardized SOI

Now assume the process is stationary with a Matérn covariance. The wavelet Whittle approximation approach stated in Chapter 3 is using WN or AR(1) approximation. In this application of the wavelet Whittle method, the LA(8) filter is implemented because the Monte Carlo studies in the previous section conclude that

71
longer wavelet filter performs better. The wavelet Whittle MLEs (using WN approximation) for the Matérn process are: \( \hat{\nu}_{WN} = 0.20, \hat{\phi}_{WN} = 0.082, \hat{\sigma}^2_{WN} = 0.391, \)
\( l^W_{Whittle-WN} = -7579.704; \) the estimation results obtained under the AR(1) approximation are almost the same as the WN approximation, with differences in estimates that are less than 0.001.

Then assume that the SOI data is an FD process, the wavelet Whittle estimation with WN approximation gives \( \hat{d}_{WN} = 0.5000, \hat{\sigma}^2_{WN} = 0.5910, l^W_{Whittle-WN} = -7599.138; \) and the wavelet Whittle AR(1) approximation has \( \hat{d}_{AR} = 0.5000, \hat{\sigma}^2_{AR} = 0.5909, l^W_{Whittle-AR} = -7544.316. \)

The approximate likelihood are also computed for AR(5), ARFIMA, and fractional exponential models. Comparing across these log-likelihood values and AIC values, the Matérn process under the AR(1) approximation model has the smallest AIC. However, we can not simply draw the conclusion that this model is the best because it is based on the assumption that the SOI data is stationary with Matérn (\( \nu = 0.2 \)) covariance. Some modified model selection criteria that include a penalty on model misspecification should be considered instead.

A plot of the empirical spectrum and the fitted spectrum (Figure 3.4) is used to graphically assess the fit of the models. We also calculate the empirical point-wise 95% confidence bands for the fitted SDFs (Figure 3.5). The orange line in Figure 3.4 is the SDF of the AR(5) model. It captures the overall behaviors of the empirical spectrum. The blue solid and dashed lines are assuming a Matérn model with \( \nu = 0.2, \)
and using the wavelet Whittle WN and AR(1) approximations, respectively. These two fitted lines almost overlap, and they follow the overall behavior of the empirical spectrum. If we assume an FD model using the wavelet Whittle approximation, the
fitted spectrum (the red lines) overlap with the AR(5) spectrum, except at the origin. The green line representing ARFIMA with AR order of 1 also falls above the average empirical values. It is obviously not as good as the models we discussed previously. The spectrum of the fractional exponential model follows the empirical one in high frequencies well, but it seems to underestimate the spectrum when the frequency is small: the purple line is below the gray line for \( f < 0.25 \). From this plot of fitted spectra, we can conclude the Matérn model with \( \hat{\nu} = 0.2 \) and the FD model with \( \hat{d} = 0.5 \) fit the best.

With these estimates, one can predict the monthly SOI values, which is critical in predicting El Nino and La Nina. We calculate the one-step ahead predictions using the Levinson-Durbin algorithm and use the RMSE and median absolute deviation (MAD) to evaluate the prediction performance. The Matérn model has RMSE=2.147 and MAD=1.694, while the FD model has smaller MSE and MAD values: 1.965 and 1.564 respectively. In terms of prediction, the FD model is better under both criteria.

### 3.8 Conclusion

In this chapter we define the wavelet Whittle (log) likelihood of a time series process and study the statistical properties of the approximate MLEs. For a general class of Gaussian processes with stationary \( d \)th backward differences, we show that the empirical wavelet variances converge jointly to a multivariate normal distribution (Theorem 3.1). We further present and show that the wavelet Whittle MLEs under a white noise approximation are jointly asymptotically normal (Theorem 3.2). The wavelet Whittle method is implemented in some Monte Carlo studies and in the analysis of an SOI climate data. Comparing to the white noise approximation, we
get better estimates in the simulation studies by using the AR(1) approximation with wider filters such as D(8) and LA(8).
Figure 3.5: Fitted spectrum - confidence bands
Chapter 4: The DWTs of space-time processes

In this chapter, we first review the statistical properties of space-time processes and the common simplifying assumptions made upon the covariance functions, which include weakly stationary, fully symmetric, separable and compactly supported. We also summarize the spectral representations of a space-time process: Bochner’s theorem, the Cressie-Huang representation [Cressie and Huang, 1999], and Stein’s representation [Stein, 2005]. Then we define the DWTs of a space-time process over the time domain, which is equivalent to performing the DWT to the process observed at every spatial location. We show that the DWT coefficients preserve the simplifying assumptions assumed for the original process. We also prove that the multidimensional wavelet coefficients of a general class of Gaussian processes are stationary on each wavelet level. Extending a result due to Moulines et al. [2007a] to the spatially varying wavelet process, we show a particular between-scale process is stationary. Similar to the whitening feature of the DWTs in time series analysis, a space-time process can be decorrelated by a wavelet transform. This decorrelation property is discussed and some examples are provided at the end of the chapter.
4.1 A review of space-time processes

4.1.1 The statistical properties

Suppose \( Y(s, t) \) is a space time process with discrete time index \( t \in \mathbb{Z} \), and \( s \in D \subseteq \mathbb{R}^d \) for some positive integer \( d \). The process has a mean function

\[
E(Y(s, t)) = \mu(s, t), \quad s \in D, t \in \mathbb{Z},
\]

and a nonnegative definite covariance function

\[
\text{cov}(Y(s_1, t_1), Y(s_2, t_2)) = C_Y(s_1, s_2, t_1, t_2), \quad s_1, s_2 \in D, t_1, t_2 \in \mathbb{Z}.
\] (4.1)

In this dissertation, we assume the process has a constant mean for all values of \( t \) and \( s \), and assume that the second moments of \( Y(s, t) \) exist and are finite. Some common simplifying assumptions made upon the covariance function of the space-time process are defined and discussed in this section [Gneiting et al., 2007, Cressie and Wikle, 2011].

A valid covariance function needs to be nonnegative definite. Below is the definition of nonnegative definiteness of a function defined on \( D^* \times D^* \), where \( D^* \) is a subset of \( \mathbb{R}^d \times \mathbb{R} \).

**Definition 4.1 (Nonnegative-definite).** A function \( \{ f((s, t), (s', t')) : (s, t) \) and \( (s', t') \in D^* \} \) defined on \( D^* \times D^* \) is said to be nonnegative-definite, if for any complex numbers \( \{a_k : k = 1, \ldots, m\} \), any \( \{(s_k, t_k) : k = 1, \ldots, m\} \) in \( D^* \), and any integer \( m \), we have

\[
\sum_{k=1}^{m} \sum_{k'=1}^{m} a_k a_k^{*} f((s_k, t_k), (s_{k'}, t_{k'})) \geq 0,
\]

where recall that \( a^{*} \) denotes the complex conjugate of \( a \).
The nonnegative-definite covariance function $C_Y$ of a space-time process is **spatially stationary** if $\text{cov}(Y(s_1,t_1), Y(s_2,t_2))$ depends on the observation sites $s_1, s_2$ only through the spatial separation vector, $h = s_1 - s_2 \in \mathbb{R}^d$. The process has **temporal stationary** covariance if $\text{cov}(Y(s_1,t_1), Y(s_2,t_2))$ depends on the observation times $t_1$ and $t_2$ only through the temporal lag $u = t_1 - t_2 \in \mathbb{Z}$. If the covariance function can be written as $C_Y(h,u)$, it is a stationary spatio-temporal covariance function. Thus a weakly stationary space time process is defined as:

**Definition 4.2** (Weakly stationary). *If a space-time process $\{Y(s,t)\}$ has constant expectation, and a stationary covariance function $C_Y(h,u)$, then it is said to be second-order (or weakly) stationary.*

To incorporate the concept of isotropy (the covariance depends on the distance between points but not the direction) into the space-time covariance, we say that the process has **spatial isotropy** [Cressie and Wikle, 2011] if

$$\text{cov}(Y(s_1,t_1), Y(s_2,t_2)) = C_Y(||s_1 - s_2||, t_1, t_2), \quad s_1, s_2 \in D, t_1, t_2 \in \mathbb{Z}.$$  

The covariance function adapted from Gneiting [2002] is spatial isotropy (i.e., it depends on $||h||$ but not $h$, see Equation (4.17)), and is used later in this chapter to visualize the decorrelation effects of the DWTs.

**Definition 4.3** (Fully symmetric). *The space-time process $Z$ has fully symmetric covariance if*

$$\text{cov}(Y(s_1,t_1), Y(s_2,t_2)) = \text{cov}(Y(s_1,t_2), Y(s_2,t_1)), \quad s_1, s_2 \in D, t_1, t_2 \in \mathbb{Z}. \quad (4.2)$$

If a process has a fully symmetric covariance function, then the corresponding covariance matrix is symmetric. But this assumption does not apply if the processes
are influenced by prevailing wind or water flows, which often happens in areas like environmental science and geophysics. A special case of fully symmetric is that of a separable process.

**Definition 4.4** (Separable). A random process \( \{Y(s, t)\} \) is said to have separable covariance if there exist purely spatial and purely temporal covariance functions \( C_S \) and \( C_T \), respectively, such that

\[
\text{cov}(Y(s_1, t_1), Y(s_2, t_2)) = C_S(s_1, s_2)C_T(t_1, t_2), \quad s_1, s_2 \in D, t_1, t_2 \in \mathbb{Z},
\]

for all space-time coordinates \((s_1, t_1)\) and \((s_2, t_2)\) in \( D \times \mathbb{Z} \).

The separable assumption on the covariance function is, in most cases, not true in practice, because there is always some space-time interactions in the covariance [Genton, 2007]. But with a separable covariance, the computation burden is dramatically reduced because the covariance matrix of the space time process is the Kronecker product of the temporal covariance matrix and the spatial covariance matrix. Mixtures of separable covariance functions have also been used to construct nonseparable space-time covariance functions [e.g., Ma, 2002, De Iaco et al., 2002].

**Definition 4.5** (Compactly supported). A spatio-temporal process \( \{Y(s, t)\} \) has compactly supported covariance if

\[
\text{cov}(Y(s_1, t_1), Y(s_2, t_2)) = 0 \quad \text{when} \quad \|s_1 - s_2\| \text{ and/or } |t_1 - t_2| \text{ are sufficiently large.}
\]

The compactness of the covariance is also related to computational efficiency. When the time and/or space “distance” is greater than a certain threshold, the corresponding covariance is zero, which results in a sparse covariance matrix. However, this is not a reasonable assumption for a long memory process.
4.1.2 The spectral representations

Gneiting et al. [2007] and Section 6.1.6 of Cressie and Wikle [2011] discuss the spectral representations for space-time process. Bochner’s theorem [Cressie and Huang, 1999, Gneiting et al., 2007] gives the spectral representation of a space-time process in both the spatial and temporal frequency domain.

**Theorem 4.6** (Bochner). *Suppose that $C_Y$ is a continuous and symmetric function on $\mathbb{R}^d \times \mathbb{R}$. Then $C_Y$ is a covariance function if and only if it is of the form*

$$C_Y(h, u) = \int \int e^{i2\pi(h^T \omega + u\tau)} dS(f)(\omega, f), \quad (h, u) \in \mathbb{R}^d \times \mathbb{R},$$  \hspace{1cm} (4.4)

*where $S(f)$ is a finite, nonnegative and symmetric measure on $\mathbb{R}^d \times \mathbb{R}$.*

We call $S(f)$ in Equation (4.4) the *spectral measure*. If $C_Y$ is integrable, the spectral measure is absolutely continuous with Lebesgue density, and we obtain the spectral density function

$$S(\omega, \tau) = \int \int e^{-i2\pi(h^T \omega + u\tau)} C_Y(h, u) dhdud, \quad (\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}.$$  

If the spectral density exists, the representation (4.4) in Bochner's theorem reduces to

$$C_Y(h, u) = \int \int e^{i2\pi(h^T \omega + u\tau)} S(\omega, \tau) d\omega d\tau, \quad (h, u) \in \mathbb{R}^d \times \mathbb{R},$$  

and $C_Y$ and $S$ can be obtained from each other via a Fourier transform. Both functions describe weakly stationary space-time dependence [Cressie and Wikle, 2011].

With the additional assumption on the integrability of $C_Y$, we have the Cressie-Huang representation [Cressie and Huang, 1999]:

80
Theorem 4.7 (Cressie and Huang). Suppose that $C_Y$ is a continuous, bounded, integrable and symmetric function on $\mathbb{R}^d \times \mathbb{R}$. Then $C_Y$ is a stationary covariance if and only if
\[ \rho(\omega, u) = \int e^{-i2\pi h' \omega} C_Y(h, u) \, dh, \quad u \in \mathbb{R}, \] (4.5)
is positive definite for almost all $\omega \in \mathbb{R}^d$.

Here $\rho(\omega, u)$ is a function of time $u$ and spatial frequency $\omega$. For each $\omega$, $\rho(\omega, u)$ is a valid autocorrelation function in $u$. The function
\[ C_Y(h, u) = \int e^{i2\pi \omega} \rho(\omega, u) \, d\omega, \quad (h, u) \in \mathbb{R}^d \times \mathbb{R}, \]is a space-time covariance function constructed by the inverse Fourier transform of $\rho(\cdot, u)$.

Stein [2005] uses the same idea but studies the process in the frequency domain in time. The space-time covariance function can be written as
\[ C_Y(h, u) = \int_{-1/2}^{1/2} S_Y(h, f) e^{iu2\pi f} \, df, \quad h \in \mathbb{R}^d, u \in \mathbb{R}, \] (4.6)
where $S_Y(h, f) = S(f)\rho(h, f)$. The function $S(f)$ is assumed to be even, nonnegative, and integrable on $\mathbb{R}$, and the function $\rho(h, f)$ is a correlation function on $\mathbb{R}$ for each $f$. A special form of the function $\rho(h, f)$ is used in Stein [2005]:

Theorem 4.8. For $C_Y$ a measurable real-valued isotropic autocorrelation function on $\mathbb{R}^d$, $\gamma$ a non-negative even function on $\mathbb{R}$, $\theta$ an odd measurable function on $\mathbb{R}$ and $S$ an even non-negative integrable function on $\mathbb{R}$, the function
\[ C_Y(h, u) = \int_{-1/2}^{1/2} S(f) c(|h|\gamma(f)) e^{i2\pi \omega(f)w'h} e^{iu2\pi f} \, df \] (4.7)
is a real-valued positive definite function on $\mathbb{R}^d \times \mathbb{R}$. 81
In the above covariance function, \( w \) is a unit vector in \( \mathbb{R}^d \). Stein’s representation will be very useful in our discussion because we are interested in the discrete wavelet transform of a space-time process in the time domain. This spectral representation help us to look at the whitening effect of the DWT in the frequency domain in time, and allow us to study the possible whitening effect on the function \( \rho(h, f) \) in spatial and temporal frequency.

4.2 The DWT of a space-time process

In this section, we will introduce the DWT of a space-time process, where the wavelet transforms are performed in the time domain. Unlike the 2-D DWT in space [Wikle et al., 2001, Nychka et al., 2003, Matsuo et al., 2011], the DWT introduced here emulates most of the results we have for the DWTs of a time series. The expected values, covariance structures and properties of the wavelet and scaling coefficients are derived. Then we presented two propositions on the stationarity of the within and between-scale wavelet coefficients. The section ends up with a discussion on the decorrelation effects of the DWTs. Space-time covariance functions from Stein [2005] and Gneiting [2002] are used to illustrate these statistical results.

The wavelet and scaling coefficient of the space-time process \( \{Y(s, t) : s \in D \subseteq \mathbb{R}^d, t \in \mathbb{Z}\} \) are defined by

\[
W_{j,k}(s) = \sum_{l=0}^{L_j-1} h_{j,l} Y(s, 2^j k - 1 - l),
\]

and

\[
V_{j,k}(s) = \sum_{l=0}^{L_j-1} g_{j,l} Y(s, 2^j k - 1 - l),
\]

for \( k \in \mathbb{Z} \) and levels \( j = 1, \ldots, J \), with \( J \) a positive integer.
Assume that the process $Y(s, t)$ is (weakly) stationary in space and has $d$th order stationary increments in the time domain. The process $Y(s, t)$ has the covariance function $C_Y(h, u)$, and the mean $\mu_Y$. Suppose that the space-time covariance function is parameterized by a parameter vector $\theta$ on a compact set $\Theta$ in $\mathbb{R}^p$, and we use the notation $C_{Y, \theta}(h, u)$ to denote the covariance function of $Y(s, t)$. As defined in Chapter 2, with $\nabla$ being the differencing operator in the time domain, we have that $\nabla Y(s, t) = Y(s, t) - Y(s, t - 1)$. Now let $X(s, t) = \nabla^d Y(s, t)$ be the $d$th order differencing process, and assume it is stationary in time with zero mean and an autocovariance function $C_X(h, u) = \text{cov}(X(s + h, t + u), X(s, t))$. With $b_j^{(d)}$ being the $d$th order sum cumulative of the wavelet filter $h_{j,l}$ (Equation (2.5)), the wavelet coefficients has the expressions in terms of $X(s, t)$:

$$W_{j,k}(s) = \sum_l h_{j,l} Y(s, 2^j k - 1 - l)$$

$$= \sum_l \left[ \sum_m \binom{d}{m} (-1)^m b_j^{(d)} \delta_{j-l, m} \right] Y(s, 2^j k - 1 - l)$$

$$= \sum_m b_j^{(d)} X(s, 2^j k - 1 - m),$$

where $\binom{d}{m}$ is defined to be 0 if $m < 0$ or $m > d$. It is showed in [Craigmile et al., 2005] that $\sum_{m=0}^{L_j-d-1} b_j^{(d)} = 0$. Then the expected value of the wavelet coefficient on a given level $j$ is

$$E(W_{j,k}(s)) = \sum_m b_j^{(d)} E(X(s, 2^j k - 1 - m)) = 0.$$
When the process \( \{Y(s, t)\} \) is stationary in space and time, the level \( j \) \( (j = 1, \ldots, J) \) scaling coefficients have expected value that does not depend on \( k \),

\[
E(V_{j,k}(s)) = \sum_{l=0}^{L_j-1} g_{j,l} E(Y(s, 2^j k - 1 - l))
= \mu_Y \sum_{l=0}^{L_j-1} g_{j,l}
= 2^{j/2} \mu_Y,
\]

The wavelet coefficients have the wavelet covariance:

\[
C_{W_j,W_{j'},\theta}(s - s', k, k') = \text{cov}(W_{j,k}(s), W_{j',k'}(s')) = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_{Y,\theta}(s - s', 2^j k - 2^{j'} k' + l' - l), (4.10)
\]

and the scaling covariance:

\[
C_{V_j,V_{j'},\theta}(s - s', k, k') = \text{cov}(V_{j,k}(s), V_{j',k'}(s')) = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} g_{j,l} g_{j',l'} C_{Y,\theta}(s - s', 2^j k - 2^{j'} k' + l' - l). (4.11)
\]

The covariance functions of wavelet coefficients and scaling coefficients both depend on \( s - s' \), levels \( j, j' \), and “time” points \( k, k' \). The coefficients \( \{W_{j,k}(s)\} \) and \( \{V_{j,k}(s)\} \) are stationary in space. If we fix the level \( j \), the within-scale autocovariances are obtained from the above two equations by letting \( j = j' \):

\[
C_{W_j,\theta}(s - s', k - k') = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_{Y,\theta}(s - s', 2^j(k - k') + l' - l),
\]

and

\[
C_{V_j,\theta}(s - s', k - k') = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} g_{j,l} g_{j',l'} C_{Y,\theta}(s - s', 2^j(k - k') + l' - l).
\]
The resulting covariance functions only depend on the spatial distance \( s - s' \), the difference in time \( k - k' \) and the level \( j \). We will show later the coefficients \( W_{j,k}(s) \) and \( V_{j,k}(s) \), are stationary in space and time on a given level \( j \) under some given conditions.

Suppose the weakly stationary space-time process \( \{Y(s, t) : t \in \mathbb{Z}, s \in D\} \) is observed at time points \( t = 0, \ldots, N - 1 \) and at locations \( D = \{s_1, s_2, \ldots, s_m\} \). Let \( Y_t \) denote the \( m \)-dim vector \((Y(s_1, t), \ldots, Y(s_m, t))^T\). The covariance matrix of \( Y(s, t) \) is

\[
\begin{align*}
\Gamma_{Y, \tau} &= \text{cov}(Y_{t+\tau}, Y_t) \\
&= \begin{bmatrix}
\text{cov}(Y(s_1, t + \tau), Y(s_1, t)) & \ldots & \text{cov}(Y(s_1, t + \tau), Y(s_m, t)) \\
\vdots & \ddots & \vdots \\
\text{cov}(Y(s_m, t + \tau), Y(s_1, t)) & \ldots & \text{cov}(Y(s_m, t + \tau), Y(s_m, t)) \\
\end{bmatrix} \\
&= \begin{bmatrix}
C_{Y, \theta}(0, \tau) & \ldots & C_{Y, \theta}(s_1 - s_m, \tau) \\
\vdots & \ddots & \vdots \\
C_{Y, \theta}(s_m - s_1, \tau) & \ldots & C_{Y, \theta}(0, \tau) \\
\end{bmatrix}. \quad (4.12)
\end{align*}
\]

Notice that the entries in the above matrix satisfy \( C_{Y, \theta}(s_q - s_q', \tau) = \text{cov}(Y(s_q, t + \tau), Y(s_q', t)) = \text{cov}(Y(s_q', t), Y(s_q, t + \tau)) = C_{Y, \theta}(s_q' - s_q, -\tau), \ q = 1, \ldots, m \), i.e., \( \Gamma_{Y, \tau} = \Gamma_{Y, -\tau}^T \). Letting \( N \) be a multiple of a power of two, the DWT wavelet coefficients \( W_{j,k} = (W_{j,k}(s_1), \ldots, W_{j,k}(s_m))^T \) are obtained by circularly filtering the data:

\[
W_{j,k}(s_i) = \sum_{l=0}^{L_j-1} h_{j,l} Y(s_i, 2^j k - 1 - l \ mod \ N), \quad i = 1, \ldots, m.
\]

And the scaling coefficients follow in a similar way:

\[
V_{j,k}(s_i) = \sum_{l=0}^{L_j-1} g_{j,l} Y(s_i, 2^j k - 1 - l \ mod \ N), \quad i = 1, \ldots, m,
\]

where \( j = 1, \ldots, J \). The wavelet and scaling covariance are

\[
C_{W_{j,j'}, \theta}(s - s', k, k') = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_{Y, \theta}(s - s', 2^j k - 2^{j'} k' + l' - l \ mod \ N),
\]

85
and
\[ C_{V_j, V_{j'}, \theta}(s - s', k, k') = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} g_{j,l}g_{j',l'}C_Y(s - s', 2^j k - 2^{j'} k' + l' - l \mod N), \]
respectively. Then we have the expression of the covariance matrix for nonboundary wavelet coefficients,

\[
\text{cov}(W_{j,k}, W_{j',k}) = \begin{bmatrix}
\text{cov}(W_{j,k}, W_{j',k}(s_1), W_{j',k}(s_1)) & \ldots & \text{cov}(W_{j,k}, W_{j',k}(s_m), W_{j',k}(s_m)) \\
\vdots & \ddots & \vdots \\
\text{cov}(W_{j,k}, W_{j',k}(s_m), W_{j',k}(s_1)) & \ldots & \text{cov}(W_{j,k}, W_{j',k}(s_m), W_{j',k}(s_m)) \\
C_{W_{j},W_{j'},\theta}(0, k + \tau, k) & \ldots & C_{W_{j},W_{j'},\theta}(s_1 - s_m, k + \tau, k) \\
\vdots & \ddots & \vdots \\
C_{W_{j},W_{j'},\theta}(s_m - s_1, k + \tau, k) & \ldots & C_{W_{j},W_{j'},\theta}(0, k + \tau, k)
\end{bmatrix}.
\]

When \( j = j' \),

\[
\Gamma_{W_{j},\tau} = \text{cov}(W_{j,k+\tau}, W_{j,k}) = \begin{bmatrix}
C_{W_{j},\theta}(0, \tau) & \ldots & C_{W_{j},\theta}(s_1 - s_m, \tau) \\
\vdots & \ddots & \vdots \\
C_{W_{j},\theta}(s_m - s_1, \tau) & \ldots & C_{W_{j},\theta}(0, \tau)
\end{bmatrix}.
\]

### 4.2.1 Simplifying assumptions for the space-scale-time wavelet processes

At the beginning of this section we discussed various simplifying assumptions made on the mean and covariance of the original space-time process. It can be shown that the wavelet transforms of \( Y(s, t) \) preserves the assumptions of weakly stationary, separability, fully symmetric, and being compactly supported.

**Lemma 4.9.** Suppose the process \( \{Y(s, t)\} \) is weakly stationary in space and has \( d \)th order stationary increments in time domain. Assume \( L \geq 2d \), then \( W_{j,k}(s) \) is weakly stationary in space and time at a given level \( j \).
Proof. We have previously showed that if $L \geq 2d$ then

$$E(W_{j,k}(s)) = 0.$$  

The covariance of the level $j$ wavelet coefficients is

$$\text{cov}(W_{j,k}(s), W_{j,k'}(s'))$$

$$= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}} h_{j,l} h_{j,l'} \text{cov} \left( Y(s, 2^j k - 1 - l), Y(s', 2^j k' - 1 - l') \right)$$

$$= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}} h_{j,l} h_{j,l'} C_{Y,\theta} \left( s - s', 2^j k - 1 - l - (2^j k' + 1 - l') \right)$$

$$= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}} h_{j,l} h_{j,l'} C_{Y,\theta} \left( s - s', 2^j (k - k') - l + l' \right)$$

$$= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} b_{j,l}^{(d)} b_{j,l'}^{(d)} C_X \left( s - s', 2^j (k - k') + l - l' \right)$$

$$= \sum_{m=-(L_j-1)}^{L_j-1} C_X \left( s - s', 2^j (k - k') + m \right) \sum_{l'=0}^{L_{j'}-1} b_{j,l'}^{(d)} b_{j,l'+|m|}^{(d)}.$$

The covariance function $\text{cov}(W_{j,k}(s), W_{j,k'}(s'))$ only depends on the spatial lag $s - s'$ and the temporal lag $k - k'$. Thus at any fixed level, the wavelet coefficients $\{W_{j,k}(s) : k \in \mathbb{Z}\}$ is weakly stationary. \qed

Lemma 4.10. If the process $Y(s, t)$ has separable covariance, then the wavelet covariance $\text{cov}(W_{j,k}(s), W_{j',k'}(s'))$ is also separable.

Proof. The space-time process $Y(s, t)$ has separable covariance, thus

$$\text{cov}(Y(s, t), Y(s', t')) = C_S(s, s') C_T(t, t'), \ s, s' \in D, \ t, t' \in \mathbb{Z}.$$  

We have the wavelet

87
the wavelet covariance:

\[
\text{cov}(W_{j,k}(s), W_{j',k'}(s')) \\
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} \text{cov} \left( Y(s, 2^j k - 1 - l), Y(s', 2^{j'} k' - 1 - l') \right) \\
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_S(s, s') C_T \left( 2^j k - 1 - l, 2^{j'} k' - 1 - l' \right) \\
= C_S(s, s') \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_T \left( 2^j k - 1 - l, 2^{j'} k' - 1 - l' \right) \\
= C_S(s, s') C_{T,j,j'}(k, k'),
\]

where \( C_{T,j,j'}(k, k') = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} C_T \left( 2^j k - 1 - l, 2^{j'} k' - 1 - l' \right) \) for each \( j, j' = 1, \ldots, J \). Thus \( \text{cov}(W_{j,k}(s), W_{j',k'}(s')) \) is separable.

**Lemma 4.11.** If the process \( Y(s, t) \) has a fully symmetric covariance function, then the wavelet covariance, \( \text{cov}(W_{j,k}(s), W_{j',k'}(s')) \), is fully symmetric.

**Proof.** The space-time process \( Y(s, t) \) has fully symmetric covariance if

\[
\text{cov}(Y(s, t), Y(s', t')) = \text{cov}(Y(s, t'), Y(s', t)), \text{ for all } s, s' \in D, t, t' \in \mathbb{Z}. \text{ Then }
\]

\[
\text{cov}(W_{j,k}(s), W_{j',k'}(s')) \\
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} \text{cov} \left( Y(s, 2^j k - 1 - l), Y(s', 2^{j'} k' - 1 - l') \right) \\
= \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_{j'}-1} h_{j,l} h_{j',l'} \text{cov} \left( Y(s, 2^j k - 1 - l'), Y(s', 2^{j'} k - 1 - l) \right) \\
= \text{cov} \left( \sum_{l'=0}^{L_{j'}-1} h_{j',l'} Y_{2^{j'} k' - 1 - l'}(s'), \sum_{l=0}^{L_j-1} h_{j,l} Y_{2^j k - 1 - l}(s) \right) \\
= \text{cov}(W_{j',k'}(s), W_{j,k}(s')).
\]

Thus the covariance function \( \text{cov}(W_{j,k}(s), W_{j',k'}(s')) \) is fully symmetric. \( \square \)
Lemma 4.12. If the process $Y(s, t)$ has a compactly supported covariance, then the wavelet covariance function, $\text{cov}(W_{j,k}(s), W_{j',k'}(s'))$, is also compactly supported.

Proof. When $Y(s, t)$ has compactly supported covariance in space, $\exists \, d > 0$ s.t. if $\|s - s'\| > d$, $\text{cov}(Y(s, t), Y(s', t')) = 0$.

Suppose that $\|s - s'\| > d$, then

$$\text{cov}(W_{j,k}(s), W_{j',k'}(s')) = \sum_{l=0}^{L_j - 1} \sum_{l'=0}^{L_{j'} - 1} h_{j,l}h_{j',l'}\text{cov}\left(Y(s, 2^j k - 1 - l), Y(s', 2^j k' - 1 - l')\right) = 0.$$ 

Thus $\{W_{j,k}(s)\}$ has a covariance function that is compactly supported in space.

When $Y(s, t)$ is compactly supported in time, i.e., $\exists \, q > 0$ s.t. if $|t - t'| > q$, $\text{cov}(Y(s, t), Y(s', t')) = 0$. Then at any fixed levels $j, j'$,

$$\text{cov}(W_{j,k}(s), W_{j',k'}(s')) = \sum_{l=0}^{L_j - 1} \sum_{l'=0}^{L_{j'} - 1} h_{j,l}h_{j',l'}\text{cov}\left(Y(s, 2^j k - 1 - l), Y(s', 2^j k' - 1 - l')\right).$$

We can check the value of the time lag $2^j k - 1 - l - (2^j k' - 1 - l')$. Without loss of generality let $j \geq j'$, then $L_j = 2^j (L - 1) + 1 > L_{j'}$, and

$$2^j k - 1 - l - (2^j k' - 1 - l') = 2^j k - 2^j k' + (l' - l),$$

where $l = 0, \ldots, L_j - 1$ and $l' = 0, \ldots, L_{j'} - 1$. Thus $\min\{l' - l\} = -(L_j - 1)$, and $\max\{l' - l\} = L_{j'} - 1$ are known when levels $j, j'$ are given. The minimum value of the time lag is

$$2^j k - 2^j k' - (L_j - 1) = 2^j k - 2^j k' - 2^j (L - 1) = 2^j (k - L + 1) - 2^j k',$$
and the maximum value is

\[ 2^j k - 2^j k' + (L^j - 1) = 2^j k - 2^j k' + 2^j (L - 1) = 2^j k - 2^j (k' - L + 1). \]

If the wavelet covariance is compactly supported, \( \text{cov}(W_{j,k}(s), W_{j',k'}(s')) = 0 \) when \( k \) and \( k' \) are far apart, which means all possible values of the time lag \( 2^j k - 1 - l - (2^j k' - 1 - l') \) must have absolute values greater than \( q \) (the covariance of \( Y(s, 2^j k - 1 - l) \) and \( Y(s', 2^j k' - 1 - l') \) is zero).

There are two possible situations: all time lags are greater than \( q \), or all of them smaller than \(-q\). For the first situation, it is equivalent to show the minimum time lag is greater than \( q \):

\[ 2^j (k - L + 1) - 2^j k' > q \quad \Leftrightarrow k - 2^j k' > 2^j \left(q + 2^j (L - 1)\right). \]

The second situation is equivalent to show the maximum time lag is smaller than \(-q\):

\[ 2^j k - 2^j (k' - L + 1) < -q \quad \Leftrightarrow k - 2^j k' < -2^j \left(q + 2^j (L - 1)\right). \]

In terms of the absolute value, when \( |k - 2^j k'| > \max\{2^{-j} (q + 2^j (L - 1)), 2^{-j} (q + 2^j (L - 1))\} \), \( 2^{-j} (q + 2^j (L - 1)) \}, the covariance \( \text{cov}(W_{j,k}(s), W_{j',k'}(s')) = 0 \). Thus the wavelet covariance function is compactly supported in time.

Similarly if the original process \( Y(s,t) \) is compact in space and time, the wavelet coefficients \( \{W_{j,k}(s)\} \) of a given level \( j \) have covariance function compactly supported in both space and time.

\[ \square \]
Because a compactly supported function is not necessarily nonnegative definite, it is not an easy task to construct a compact function that is a valid covariance function. If the compactness property can be preserved by the DWT coefficients from the original process \(Y(s, t)\), the covariance function is guaranteed to be well defined.

### 4.2.2 Spectral representations of the DWT coefficients

Recalling the spectral representations of a space-time process, we define the spectral density function of the DWT coefficients of a space-time process. Let \(C_{W_j, \theta}(s + h, s, u)\) denotes the level \(j\) wavelet covariance of \(W_{j,k+u}(s + h)\) and \(W_{j,k}(s)\). If \(C_{W_j, \theta}(s + h, s, u)\) is integrable with respect to \(u\), the corresponding spectral density of the wavelet coefficient is

\[
S_{W_j, \theta}(s + h, s, f) = \int_{-1/2}^{1/2} e^{-i2\pi uf} C_{W_j, \theta}(s + h, s, u) du.
\]

If the space-time process is spatial and temporal stationary, then the wavelet covariance function is \(C_{W_j, \theta}(h, u)\), and the above expression reduces to

\[
S_{W_j, \theta}(h, f) = \int_{-1/2}^{1/2} e^{-i2\pi uf} C_{W_j, \theta}(h, u) du,
\]
where \( h \) is the space lag. Then we can derive the expression of the wavelet SDF at a fixed level \( j \). We have

\[
\text{cov} \left( W_{j,k+u}(s+h), W_{j,k}(s) \right) \\
= \int_{-1/2}^{1/2} e^{i2\pi f u} S_{W_{j,\theta}}(h, f) df \\
= \int_{-1/2}^{1/2} e^{i2\pi f u} \mathcal{H}_j(f) S_{Y,\theta}(h, f) df \quad \text{(Let } \zeta = 2^j f, \text{ and change variable.)} \\
= \int_{-1/2}^{1/2} e^{i2\pi \zeta u} \mathcal{H}_j \left( \frac{\zeta}{2^j} \right) S_{Y,\theta} \left( h, \frac{\zeta}{2^j} \right) d\zeta \\
= \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-2^j/2+m}^{-2^j/2+m+1} e^{i2\pi \zeta u} \mathcal{H}_j \left( \frac{\zeta}{2^j} \right) S_{Y,\theta} \left( h, \frac{\zeta}{2^j} \right) d\zeta \\
= \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-1/2}^{1/2} e^{i2\pi \zeta u} \mathcal{H}_j \left( \frac{\zeta + m}{2^j} \right) S_{Y,\theta} \left( h, \frac{\zeta + m}{2^j} \right) d\zeta \\
= \int_{-1/2}^{1/2} e^{i2\pi f u} S_{W_{j,\theta}}(h, f) df,
\]

where the level \( j \) SDF is defined to be

\[
S_{W_{j,\theta}}(h, f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \mathcal{H}_j \left( \frac{f + m}{2^j} \right) S_{Y,\theta} \left( h, \frac{f + m}{2^j} \right). \quad (4.13)
\]

This result is obtained by breaking the integral into segments, and by the fact that

\[
\mathcal{H}_j \left( \frac{\zeta}{2^j} \right) S_{Y,\theta} \left( h, \frac{\zeta}{2^j} \right)
\]

is periodic with period 1 in the temporal frequency domain for each \( j = 1, \ldots, J \).

For any two levels \( j \) and \( j' \), a general expression of the wavelet spectral density is derived below:
\begin{align*}
\text{cov} \left( W_{j,k+u}(s+\mathbf{h}), W_{j',k}(s) \right) \\
= \int_{-1/2}^{1/2} e^{i2\pi f(2^j(k+u)-2^{j'}k)} S_{W_j,W_{j'},\theta}(\mathbf{h}, f) df \\
= \int_{-1/2}^{1/2} e^{i2\pi f(2^j(k+u)-2^{j'}k)} H_j(f) H_{j'}^*(f) S_{Y,\theta}(\mathbf{h}, f) df \\
= \int_{-1/2}^{1/2} e^{i2\pi \zeta(2^j(k+u)-2^{j'}k)} H_j \left( \frac{\zeta}{2^j} \right) H_{j'}^* \left( \frac{\zeta}{2^{j'}} \right) S_{Y,\theta} \left( \mathbf{h}, \frac{\zeta}{2^j} \right) d \left( \frac{\zeta}{2^j} \right) \\
(\text{Break the integral into segments.}) \\
= \frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-2^{j}/2+m}^{-2^{j}/2+m+1} e^{i2\pi \zeta(2^j(k+u)-2^{j'}k)} H_j \left( \frac{\zeta + m}{2^j} \right) H_{j'}^* \left( \frac{\zeta + m}{2^{j'}} \right) S_{Y,\theta} \left( \mathbf{h}, \frac{\zeta + m}{2^j} \right) d \zeta.
\end{align*}

Because \( H_j \left( \frac{\zeta}{2^j} \right) H_{j'}^* \left( \frac{\zeta}{2^{j'}} \right) S_{Y,\theta} \left( \mathbf{h}, \frac{\zeta}{2^j} \right) \) is periodic with period 1, the above equation equals

\begin{align*}
\frac{1}{2^j} \sum_{m=0}^{2^j-1} \int_{-1/2}^{1/2} e^{i2\pi \zeta(2^j(k+u)-2^{j'}k)} H_j \left( \frac{\zeta + m}{2^j} \right) H_{j'}^* \left( \frac{\zeta + m}{2^{j'}} \right) S_{Y,\theta} \left( \mathbf{h}, \frac{\zeta + m}{2^j} \right) d \zeta \\
= \int_{-1/2}^{1/2} e^{i2\pi f(2^j(k+u)-2^{j'}k)} S_{W_j,W_{j'},\theta}(\mathbf{h}, f) df,
\end{align*}

where the cross-level wavelet spectral density is defined to be

\begin{equation}
S_{W_j,W_{j'},\theta}(\mathbf{h}, f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} H_j \left( \frac{f + m}{2^j} \right) H_{j'}^* \left( \frac{f + m}{2^{j'}} \right) S_{Y,\theta} \left( \mathbf{h}, \frac{f + m}{2^j} \right). \tag{4.14}
\end{equation}

Recall the stationarity result of the within-scale wavelet coefficients for the univariate time series, the following proposition is an extended result to space-time processes.
Proposition 4.13. Suppose that \( L \geq 2d \). Then the nonboundary wavelet coefficients at a given level \( j \) are a portion of a zero mean stationary process with autocovariance

\[
C_{W_j,\theta}(h, \tau) = \text{cov}(W_{j,k+r}(s + h), W_{j,k}(s)) = \sum_{m=-(L_j-1)}^{L_j-1} C_X(h, 2^j \tau + m) \sum_{t'=0}^{L_j-|m|-1} b_{j,t}^{(dt)} b_{j,t'+|m|}^{(dt)}
\]

\[
= \int_{-1/2}^{1/2} e^{i2\pi f \tau} S_{W_j,\theta}(h, f) df,
\]

and thus we identify the SDF of the level \( j \) wavelet coefficients to be

\[
S_{W_j,\theta}(h, f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \mathcal{H}_j \left( \frac{f + m}{2^j} \right) S_Y \left( h, \frac{f + m}{2^j} \right), \quad f \in [-1/2, 1/2],
\]

where \( \mathcal{H}_j(f) = H_j(f) H_j^*(f) \) is the squared gain function of the level \( j \) wavelet filter \( h_{j,l} \).

We also have the cross-scale covariance result extended from [Moulines et al., 2007a, Corollary 1]. First define a cross-level wavelet space-time process,

\[
W_{j,k}(s, v) = (W_{j,v+2^j k+r}(s) : r = 0, \ldots, 2^v - 1)^T,
\]

for each level \( j \). It is a \( 2^v \)-dim vector, where \( v \) is some integer such that \( 0 \leq v < j \).

Then the following holds.

Proposition 4.14. Let \( L \geq 2d \). Then the cross-level wavelet process at level \( j \) defined by \( \{ (W_{j,k}(s), W_{j,k}(s, v))^T : k \in \mathbb{Z} \} \), is stationary with covariance

\[
\text{cov}(W_{j,k}(s + h), W_{j,k}(s, v)) = \int_{-1/2}^{1/2} e^{i2\pi f(k-k')} D_{j,v}(h, f) df,
\]

and a vector-valued cross-spectral density \( D_{j,v}(h, f) \), given by

\[
D_{j,v}(h, f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} 2^{v/2} e_v(f + m) H_j \left( \frac{f + m}{2^j} \right) H_j^* \left( \frac{f + m}{2^j} \right) S_Y \left( h, \frac{f + m}{2^j} \right),
\]
where \( e_v(\xi) = 2^{-v/2} [1, e^{-i2^{-v}2\pi\xi}, \ldots, e^{-i(2^{v-1})2^{-v}2\pi\xi}]^T \) is a vector of unit norm for all \( \xi \in \mathbb{R} \).

The above propositions hold for a space-time process because the DWT is carried out in time at each spatial location. The stationarity with respect to time applies here as for the time series process, and the stationarity in space can be preserved from the original process \( Y(s, t) \). These propositions are employed to prove the central limit result of the wavelet Whittle MLEs of the parameters of a space-time process.

### 4.2.3 Examples of the whitening effects

Using Stein’s spectral representation, the SDF of the space-time process \( Y(s, t) \) has the expression \( S_{Y,\theta}(h, f) = S(f)\rho(h, f) \), then Equation (4.13) can be rewritten as:

\[
S_{W,\theta}(h, f) = \frac{1}{2^j} \sum_{m=0}^{2^j-1} \mathcal{H}_j \left( \frac{f + m}{2^j} \right) S \left( \frac{f + m}{2^j} \right) \rho \left( h, \frac{f + m}{2^j} \right). \tag{4.16}
\]

Similar to the DWT whitening result for one-dimensional (time series) processes, the DWT defined in this chapter decorrelates a space-time process in time. One interesting question is: what is the effect of DWT on the function \( \rho(h, f) \)? Since the function \( \rho \) depends on both the spatial lag \( h \) and the frequency \( f \) in time, we expect the space-time wavelet covariance to decay faster than the process covariance in the spatial domain as well. In this section, some examples are used to visualize the whitening effect of the wavelet transform.

We consider the space-time covariance functions constructed by Gneiting [2002]:

\[
C(h; u) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \left( \frac{c ||h||}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right)^\nu \\
\times K_\nu \left( \frac{c ||h||}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right), \quad (h; u) \in \mathbb{R}^d \times \mathbb{R}. \tag{4.17}
\]
In this expression, $a \geq 0$ is the temporal scaling parameter, $c \geq 0$ is the spatial scaling parameter; $\alpha \in (0, 1]$ is the temporal smoothness parameters, and $\nu > 0$ is the spatial smoothness parameters; $\beta \in [0, 1]$ measures the separability of the covariance; $\delta \geq 0$, $\sigma^2 > 0$; $K_\nu$ is the modified Bessel function of the second kind. When $\beta = 0$ the space-time covariance (4.17) is a separable function, and when $\beta$ is close to 1, there is a very strong space-time interaction. We plot the process and wavelet autocovariances of a separable and a nonseparable covariance functions, with $a = 1$, $c = 1$, $\alpha = 1/2$, $\nu = 1/2$, and $d = 2$. Let $\beta = 0$ and $\delta = 1$, we obtain the separable covariance function:

$$C(h, u) = (|u| + 1)^{-1} \exp (-||h||).$$

And for the nonseparable case, we let $\beta = 1$ and $\delta = 0$:

$$C(h, u) = \frac{1}{(|u| + 1)} \exp \left( - \frac{||h||}{(|u| + 1)} \right).$$

Figure 4.4 and Figure 4.5 are the within-scale autocovariances and correlations, respectively. The process and the wavelet autocovariance sequences in time with a given spatial lag $||h||$ are proportional to each other, and the correlation plots of the process and of the wavelet coefficients are exactly the same for different values of $||h||$. The autocovariances in Figure 4.6 and the correlations in Figure 4.7 behave differently under different $||h||$, there is a very clear space-time interaction.

Figures 4.1 - 4.3 display the between-scale covariances of the space-time process. These covariances are small in values, and are decreasing when the space lag $||h||$ increases (Figure 4.3 has almost all covariances close to zero). The covariances between level 3 and level 5 are relatively large, but this may because the original process $Y(s, t)$ has large variation at the corresponding scales ($2^3$ and $2^5$).
Figure 4.8 and Figure 4.9 show the value of ACVS versus spatial lag $||h||$. In these figures, we let the spatial locations be equally spaced between 0 and 10. The distance between neighboring locations is 0.5. When the space-time interaction is large, e.g., $\beta = 0.7$ as in Figure 4.8, the wavelet coefficients are less correlated compared to the original process. In Figure 4.8 the space-time interaction is small, $\beta = 0.1$, the DWT in the time domain greatly reduces the wavelet covariances in space as well as in the time domain. This is because the covariance function is close to a separable space-time covariance, which is be approximately proportional to the wavelet covariance in time, thus will be small. In general, regardless of spatial distance both separable and nonseparable covariance functions have highly decorrelated wavelet coefficients in time on different levels after we perform the DWTs in time domain. And this wavelet decorrelations also have some effects on the spatial dependence. This motivates us to use a similar wavelet Whittle approach as for the time series to approximate the likelihood of a space-time process, and to estimate parameters of the process covariance function.
Figure 4.1: Space-time process between-scale covariances, with LA(8) filter and $||h|| = 0$
Figure 4.2: Space-time process between-scale covariances, with LA(8) filter and $||h|| = 1$
Figure 4.3: Space-time process between-scale covariances, with LA(8) filter and $||h|| = 2$
Figure 4.4: Space-time separable ACVS, LA(8) filter. (The covariance function is adapted from [Gneiting, 2002, Equation (15)], with $\beta = 0$).
Figure 4.5: Space-time separable correlation, LA(8) filter.
Figure 4.6: Space-time nonseparable ACVS, LA(8) filter. (The covariance function is adapted from [Gneiting, 2002, Equation (15)], with $\beta = 1$.)
Figure 4.7: Space-time nonseparable correlation, LA(8) filter.
Figure 4.8: Space-time nonseparable ACVS with $\beta = 0.7$, LA(8) filter.
Figure 4.9: Space-time process nonseparable ACVS with $\beta = 0.1$, LA(8) filter.
In this chapter we present the wavelet Whittle methods for estimating the parameters of a space-time model. We first provide the wavelet Whittle approximate likelihood, which is extended from the univariate time series case that we introduced in Chapter 2. White noise and order 1 autoregressive wavelet-based approximation models are then discussed in detail, as well as a central limit result for the estimators under the white noise approximation model. At the end of this chapter, we analyze the Irish wind data [Haslett and Raftery, 1989] to illustrate an application of the proposed wavelet Whittle approximation model.

5.1 Wavelet Whittle approximations for space-time processes

Suppose a space-time process \( \{Y(s, t) : s \in D \subset \mathbb{R}^2, t \in \mathbb{Z}\} \) is observed at \( m \) sites, \( s_1, \ldots, s_m \) and \( N \) equally spaced time points \( \{Y(s_1, t), \ldots, Y(s_m, t) : t = 0, \ldots, N - 1\} \). Letting the process belong to the general class of Gaussian space-time process with stationary \( d \)th order difference, we want to estimate the parameter vector \( \theta \) of its covariance function \( C_Y(h, u) \) using maximum likelihood methods. The computational cost of likelihood methods is already expensive for univariate time series with non-Markov dependencies, and it will be more computational demanding
for space-time processes. Approximating the likelihood is essential in performing analysis under a reasonable computational cost.

As defined in Chapter 4, let \( \Gamma_{Y,\tau} \) be the covariance matrix of the \( m \)-dimensional vector \( Y_t = (Y(s_1, t), \ldots, Y(s_m, t))^T \) with time lag \( \tau \). The exact likelihood of \( Y = (Y_1^T, \ldots, Y_N^T)^T \) is

\[
L(\theta; Y) = (2\pi)^{-Nm/2} \det(\text{cov}(Y))^{-1/2} \exp \left\{ -\frac{1}{2} Y^T \text{cov}(Y)^{-1} Y \right\}, \tag{5.1}
\]

with the log likelihood function

\[
l(\theta; Y) = -\frac{Nm}{2} \log(2\pi) - \frac{1}{2} \log \det(\text{cov}(Y)) - \frac{1}{2} Y^T \text{cov}(Y)^{-1} Y. \tag{5.2}\]

Here \( Y \) is a \( Nm \)-dimensional data vector and \( \text{cov}(Y) \) is a \( Nm \times Nm \)-dimensional covariance matrix. A direct calculation of this likelihood has \( O([Nm]^3) \) computational complexity, so there is a need for likelihood approximation for long time series \( N \) or a large number of sites \( m \). Based on the DWT of a space-time process defined in Chapter 4 and its decorrelation property, a wavelet Whittle approximation method is proposed.

As discussed in Chapter 4, the DWT decorrelates the space-time data: the coefficients are approximately uncorrelated between different wavelet scales. Assuming that the DWT coefficients are independent between scales, the following wavelet-based likelihood is used to approximate the exact likelihood. Before we define the wavelet likelihood, let \( W_j = (W_{j,1}, \ldots, W_{j,N_j})^T \), a \( N_jm \)-dim vector of the level \( j \) wavelet coefficients. And let \( T_{N_j}(S_{W_j,\theta}) \) denote the covariance matrix, \( \text{cov}(W_j) \), then

\[
T_{N_j}(S_{W_j,\theta}) = \begin{bmatrix}
T_{N_j}(S_{W_j,1,1,\theta}) & T_{N_j}(S_{W_j,1,2,\theta}) & \cdots & T_{N_j}(S_{W_j,1,N_j,\theta}) \\
\vdots & \vdots & \ddots & \vdots \\
T_{N_j}(S_{W_j,N_j,1,\theta}) & T_{N_j}(S_{W_j,N_j,2,\theta}) & \cdots & T_{N_j}(S_{W_j,N_j,N_j,\theta})
\end{bmatrix}_{mN_j \times mN_j},
\]
where $T_{N_j}(S_{W_j,k,k',\theta})$ is a $m \times m$ matrix defined by

$$
[T_{N_j}(S_{W_j,k,k',\theta})]_{q,q'} = \text{cov} (W_{j,k}(s_q), W_{j,k'}(s_{q'})) \\
= \int_{-1/2}^{1/2} e^{i2\pi f(k-k')} S_{W_j,\theta}(s_q - s_{q'}, f) df.
$$

The SDF $S_{W_j,\theta}(h,f)$ is defined by Equation (4.13). We have the following wavelet-based approximate likelihood:

$$
l^{W,1}(\theta) = -\sum_{j=1}^{J} \frac{N_j m}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} \log \det (T_{N_j}(S_{W_j,\theta})) \\
- \frac{1}{2} \sum_{j=1}^{J} W_j^T \text{cov} (T_{N_j}(S_{W_j,\theta}))^{-1} W_j.
$$

A further approximation is obtained by using the approximate SDF $\tilde{S}_{W_j,\theta}(h,f)$.

$$
l^{W,2}(\theta) = -\sum_{j=1}^{J} \frac{N_j m}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} \log \det \left(T_{N_j}(\tilde{S}_{W_j,\theta})\right) \\
- \frac{1}{2} \sum_{j=1}^{J} W_j^T \left(T_{N_j}(\tilde{S}_{W_j,\theta})\right)^{-1} W_j.
$$

In this chapter, we will consider the WN and AR(1) approximations for $\tilde{S}_{W_j,\theta}$.

### 5.1.1 A white noise model

Assuming that a white noise (WN) model applies at every level, then the covariance matrix reduces to a block diagonal matrix with a $m \times m$ symmetric covariance matrix $\Gamma_{W_j,0}$ on the diagonals:

$$
\Sigma_j(\theta) = \text{cov}(W_j) = \begin{bmatrix} \text{cov}(W_{j,1}) & 0 & \cdots & 0 \\
0 & \text{cov}(W_{j,2}) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \text{cov}(W_{j,N_j}) \end{bmatrix}_{N_j m \times N_j m} \\
= \begin{bmatrix} \Gamma_{W_j,0} & 0 & \cdots & 0 \\
0 & \Gamma_{W_j,0} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Gamma_{W_j,0} \end{bmatrix}_{N_j m \times N_j m},
$$
To calculate the determinant of a block diagonal matrix, the following result is useful:

$$\det \Sigma_j(\theta) = \det \Gamma_{W,j,0} \times \cdots \times \det \Gamma_{W,j,0}.$$ 

Also the inverse of a block diagonal matrix is

$$\Sigma_j(\theta)^{-1} = \text{diag}(\Gamma_{W,j,0}^{-1}, \ldots, \Gamma_{W,j,0}^{-1}).$$

The white noise wavelet Whittle (log) likelihood is then:

$$l^W(\theta) = -\frac{(N - 1)m}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \log(\det \Gamma_{W,j,0})$$

$$-\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} W_{j,k}^T \Gamma_{W,j,0}^{-1} W_{j,k}.$$  (5.3)

The white noise wavelet Whittle estimator $\hat{\theta}$ is the maximizer of the above approximate (log) likelihood function. In terms of computational complexity, the calculation requires a much smaller order of operation because it only needs to invert and calculate the determinant of a $m \times m$ symmetric matrix, with the order of calculation $m^3$.

### 5.1.2 An AR(1) model

To obtain the autoregressive of order one approximate Whittle likelihood we assume a multivariate AR(1) model on every scale $j = 1, \ldots, J$:

$$W_{j,k} = A_j(\theta)W_{j,k-1} + Z_{j,k-1},$$  (5.4)

where $A_j(\theta)$ is the propagation matrix depending on $\theta$, and $\{Z_{j,k}\}$ is a multivariate white noise process with covariance function $\Gamma_{Z,j}$. Let $\Gamma_{W,j,0} = \text{cov}(W_{j,k})$ and $\Gamma_{W,j,1} = \text{cov}(W_{j,k}, W_{j,k-1})$. In the calculation of this AR(1) approximate likelihood, we also need to obtain the matrices $A_j(\theta)$, $\Gamma_{W,j,0}$, and $\Gamma_{W,j,1}$ for each level $j$ using
multivariate Levinson-Durbin algorithm [e.g., Brockwell and Davis, 1991, page 432].

The covariance matrices $\Gamma_{W_j,0}$, $\Gamma_{W_j,1}$, $\Gamma_{Z,j}$ depend on parameter $\theta$ and using a few steps of the Levinson-Durbin algorithm we can get that

$$\Gamma_{W_j,1} = A_j(\theta)\Gamma_{W_j,0},$$

$$A_j(\theta) = \Gamma_{W_j,1}\Gamma_{W_j,0}^{-1},$$

$$\Gamma_{W_j,0} = A_j(\theta)\Gamma_{W_j,0}A_j^T(\theta) + \text{cov}(Z_{j,t}), \text{ and}$$

$$\Gamma_{Z,j} = \text{cov}(Z_{j,k}) = \Gamma_{W_j,0} - A_j(\theta)\Gamma_{W_j,0}A_j^T(\theta).$$

The (approximate) autoregressive of order one Whittle (log) likelihood is then:

$$l^{(W)}(\theta) = \sum_{j=1}^{J} \left[ \log(f(W_{j,1})) + \sum_{k=2}^{N_j} \log f(W_{j,k}|W_{j,k-1}) \right]$$

$$= -\frac{(N - 1)m}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{J} \left[ W_{j,1}^T \text{cov}(W_{j,1})^{-1} W_{j,1} + \log \det \text{cov}(W_{j,1}) ight.$$

$$+ \sum_{k=2}^{N_j} \log \det \text{cov}(Z_{j,k}) + \sum_{k=2}^{N_j} (W_{j,k} - A_j(\theta)W_{j,k-1})^T \text{cov}(Z_{j,k})^{-1} (W_{j,k} \right.$$

$$- A_j(\theta)W_{j,k-1}) \right]$$

$$= -\frac{(N - 1)m}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{J} \left[ \log \det \Gamma_{W_j,0} + (N_j - 1) \log \det \Gamma_{Z,j} \right.$$

$$- \frac{1}{2} \sum_{j=1}^{J} \left[ W_{j,1}^T \Gamma_{W_j,0}^{-1} W_{j,1} + \sum_{k=2}^{N_j} (W_{j,k} - A_j(\theta)W_{j,k-1})^T \Gamma_{Z,j}^{-1} (W_{j,k} \right.$$

$$- A_j(\theta)W_{j,k-1}) \right].$$

Obtaining the MLE using this AR(1) approximation requires an efficient multivariate Levinson-Durbin algorithm. We will be using the statistical software package R [R...
Core Team, 2014] to implement Levinson-Durbin algorithms. Therefore the computational time could be large for AR(1) approximation (compared with the white noise approximation). A further discussion about this issue is in Chapter 6.

5.2 Asymptotic results of the wavelet-Whittle estimators

For each \( j = 1, \ldots, J \) let \( \hat{\Gamma}_{W_j,0} = \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k} W_{j,k}^T \). Taking expectations of \( \hat{\Gamma}_{W_j,0} \) (see the top of p.115) we have the result that \( \hat{\Sigma}_j \) is an unbiased estimator of the wavelet variance \( \Gamma_{W_j,0} \). The following theorem states the joint asymptotic results of \( \hat{\Gamma}_{W_j,0} \).

**Theorem 5.1.** Suppose that \( L \geq 2d \). For each \( j = 1, \ldots, J \) let \( \rho_{j,1} = \sqrt{N_j/N_1} \). Then

\[
\sqrt{N_1} \left( \rho_{1,1} \text{vec}(\hat{\Gamma}_{W_1,0} - \Gamma_{W_1,0})^T, \ldots, \rho_{J,1} \text{vec}(\hat{\Gamma}_{W_J,0} - \Gamma_{W_J,0})^T \right)^T \xrightarrow{D} N_{m^2J}(0, \Sigma),
\]

as \( N_1 \to \infty \), where \( \Sigma \) is the \( m^2J \times m^2J \) covariance matrix with the block structure:

\[
\Sigma = \begin{bmatrix}
\Sigma_{1,1} & \cdots & \Sigma_{1,J} \\
\vdots & \ddots & \vdots \\
\Sigma_{J,1} & \cdots & \Sigma_{J,J}
\end{bmatrix},
\]

where

\[
\Sigma_{j,j'} = \left[ \text{cov}(W_{j,k}(s)W_{j',k}(s')) W_{j',k}(s^*W_{j',k}(s^*)) \right]_{s,s',s^*,s^*'} \in \{s_1, \ldots, s_m\}
\]

\[
= \left[ 2^{j'/2-j/2} \left( 1^T A_{j,v}(s, s', s^*, s^*') \right) \right]_{s,s',s^*,s^*'} \in \{s_1, \ldots, s_m\}.
\]

In the theorem \( A_{j,v}(s, s', s^*, s^*') \) follows the vector-valued cross-scale SDF defined later by Equation (5.5).

Letting \( \hat{\theta}_N \) be the wavelet-Whittle MLE obtained under the white noise approximation model, we have the following asymptotic result.
**Theorem 5.2.** Suppose that \( L \geq 2d \). Then the \( p \)-dimensional vector \( \hat{\theta}_N \) is a consistent estimator of the true parameter vector \( \hat{\theta}_0 \) of the space-time model, and as \( N \to \infty \),

\[
\sqrt{N} \left[ \hat{\theta}_N - \theta_0 \right] \overset{D}{\to} N_p(0, \Omega(\theta_0)),
\]

where \( \Omega(\theta_0) \) is some \( p \times p \) covariance matrix that depends on \( \theta_0 \).

In the above theorem, the covariance matrix \( \Omega(\theta_0) = KB\Sigma B^TK^T \), where

\[
K = \sum_{j=1}^J \frac{1}{2} \begin{bmatrix}
\text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_1} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_1} \right) & \cdots & \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \right) \\
\vdots & \ddots & \vdots \\
\text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \right) & \cdots & \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \right)
\end{bmatrix}
\]

is a \( p \times p \) matrix, and

\[
B = -\frac{1}{2} \begin{bmatrix}
\rho_{1,1} \text{vec} \left( \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_1} \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_1} \right)^T & \cdots & \rho_{J,1} \text{vec} \left( \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \right)^T \\
\vdots & \ddots & \vdots \\
\rho_{1,1} \text{vec} \left( \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \right)^T & \cdots & \rho_{J,1} \text{vec} \left( \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \Gamma_{W,0}^{-1} \frac{\partial \Gamma_{W,0}}{\partial \theta_p} \right)^T
\end{bmatrix}
\]

is a \( p \times m^2J \) matrix. The \( m^2J \times m^2J \) covariance matrix \( \Sigma \) is defined in Theorem 5.1.

### 5.3 Proof of Theorem 5.1 and Theorem 5.2

Theorem 5.1 and Theorem 5.2 can be proved using the same approach as in Chapter 3 for the univariate time series results. Lemma 3.10 and Lemma 3.3 from Moulines et al. [2008] are essential in the proof. We will also need the stationarity results stated below.

**Lemma 5.3.** Assume that the condition in Proposition 4.13 is satisfied. Then the process \( \{W_{j,k}(s)W_{j,k}(s') : k \in \mathbb{Z}\} \) with \( s, s' \in D \subset \mathbb{R}^2 \), is also stationary in time.
Proof. By Proposition 4.13, at a fixed level \( j \), the wavelet process \( \{W_{j,k}(s) : k \in \mathbb{Z} \} \) is stationary in time and with mean zero. (It is also stationary in space if the original process \( Y(s,t) \) is.) Thus the expected value is

\[
E(W_{j,k}(s)W_{j,k}(s')) = \text{cov}(W_{j,k}(s), W_{j,k}(s')) - E(W_{j,k}(s))E(W_{j,k}(s'))
\]

\[
= C_{W_j}(s-s', 0) - 0
\]

\[
= C_{W_j}(s-s', 0),
\]

is free of \( k \). And the covariance

\[
\text{cov}(W_{j,k}(s)W_{j,k}(s'), W_{j,k'}(s^*)W_{j,k'}(s'^*))
\]

\[
= E(W_{j,k}(s)W_{j,k}(s'))E(W_{j,k'}(s^*)W_{j,k'}(s'^*)) + E(W_{j,k}(s)W_{j,k'}(s^*))
\]

\[
\times E(W_{j,k'}(s^*)W_{j,k'}(s'^*)) + E(W_{j,k}(s)W_{j,k'}(s'^*))E(W_{j,k'}(s^*))
\]

\[
- E(W_{j,k}(s), W_{j,k}(s'))E(W_{j,k'}(s^*)W_{j,k'}(s'^*))
\]

\[
= E(W_{j,k}(s)W_{j,k'}(s^*))E(W_{j,k}(s')W_{j,k'}(s'^*)) + E(W_{j,k}(s)W_{j,k'}(s'^*))
\]

\[
\times E(W_{j,k'}(W_{j,k}(s')s^*))
\]

\[
= C_{W_j}(s-s^*, k-k')C_{W_j}(s'-s'^*, k-k')
\]

\[
+ C_{W_j}(s-s'^*, k-k')C_{W_j}(s'-s^*, k-k'),
\]

only depends on time lag \( k - k' \), and relative distances \( s - s^* \), \( s' - s'^* \), \( s - s'^* \), and \( s' - s^* \). Thus the within-scale wavelet “square” process \( \{W_{j,k}(s)W_{j,k}(s') : k \in \mathbb{Z} \} \) is stationary in time. \( \square \)
Proposition 4.14 states that the cross-level wavelet process
\[ \left\{ (W_{j,k}(s), W_{j,k}(s, v))^T : k \in \mathbb{Z} \right\} \]
is covariance stationary, and with the vector-valued cross spectral density \( D_{j,v}(h, f) \). We define the “squared” process
\[
W_{j,k}^2(s, s', v)^T = \left( W_{j,k}^2 - v, W_{j,k}^2 + r(s) : r = 0, \ldots, 2^v - 1 \right)^T.
\]

It can be shown that the following result holds.

**Lemma 5.4.** Assume that \( L \geq 2d \), then the element-wise “squared” process
\[
\left\{ (W_{j,k}(s)W_{j,k}(s'), W_{j,k}^2(s, s', v))^T : k \in \mathbb{Z} \right\},
\]
is covariance stationary.

**Proof.** With the results in Proposition 4.14, for each pair of the cross-scale process the covariance is
\[
\text{cov}(W_{j,k}(s)W_{j,k}(s'), W_{j-v,2^v k+r}(s^*)W_{j-v,2^v k+r}(s'^*))
\]
\[
= E(W_{j,k}(s)W_{j,k}(s')W_{j-v,2^v k+r}(s^*)W_{j-v,2^v k+r}(s'^*))
\]
\[-E(W_{j,k}(s)W_{j,k}(s')W_{j-v,2^v k+r}(s^*)W_{j-v,2^v k+r}(s'^*))
\]
\[
+ E(W_{j,k}(s)W_{j-v,2^v k+r}(s^*)) E(W_{j,k}(s')W_{j-v,2^v k+r}(s'^*))
\]
\[
+ E(W_{j,k}(s)W_{j-v,2^v k+r}(s^*)) E(W_{j,k}(s')W_{j-v,2^v k+r}(s^*))
\]
\[
= C_{W_{j,v}, W_{j-v}}(s - s^*, v, r) C_{W_{j,v}, W_{j-v}}(s' - s'^*, v, r)
\]
\[
+ C_{W_{j,v}, W_{j-v}}(s - s^*, v, r) C_{W_{j,v}, W_{j-v}}(s' - s^*, v, r),
\]
which does not depend on \( k \). So the squared process is covariance stationary in time. \( \square \)
For this squared process, we define the vector-valued cross spectral density,
\[ A_{j,v}(s, s', s^*, s'^*) \]
between \( \{W_{j,k}(s) : k \in \mathbb{Z}\} \) and \( \{W_{j,k}^2(s, s', v)^T : k \in \mathbb{Z}\} \):
\[
A_{j,v}(s, s', s^*, s'^*) = \sum_{\tau \in \mathbb{Z}} \text{cov} \left( W_{j,k}(s)W_{j,k}(s'), W_{j,k+\tau}(s, s', v)^T \right)e^{-i2\pi f\tau}, \tag{5.5}
\]
for \( 0 \leq v < j \). Then,
\[
\text{cov}(W_{j,k}(s)W_{j,k}(s'), W_{j-v,2^v+k+r}(s^*)W_{j-v,2^v+k+r}(s'^*))
\]
\[
= \int_{-1/2}^{1/2} e^{i2\pi f2^{-v}} [A_{j,v}(s, s', s^*, s'^*), r, r']_{r} df.
\]

To show Theorem 5.1, it is equivalent to show for \( \forall a_j \in \mathbb{R}, j = 1, \ldots, J \),
\[
\sum_{j=1}^{J} a_j \sqrt{N_j} \text{vec} \left( \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}W_{j,k}^T - \Gamma_{W,j,0} \right)
\]
converges to a mean zero, \( m^2 \)-dimensional multivariate normal distribution as \( N \to \infty \). The mean zero result is easy to show. For the \((r, r')\)th element, \( r, r' = 1, \ldots, m \), we have
\[
E \left( \left[ \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}(s_r)W_{j,k}(s_{r'}) \right]_{r,r'} \right) - [\Gamma_{W,j,0}]_{r,r'}
\]
\[
= E \left( \left[ \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}(s_r)W_{j,k}(s_{r'}) \right] \right) - [\Gamma_{W,j,0}]_{r,r'}
\]
\[
= E \left( W_{j,k}(s_r)W_{j,k}(s_{r'}) \right) - [\Gamma_{W,j,0}]_{r,r'}
\]
\[
= \text{cov} \left( W_{j,k}(s_r), W_{j,k}(s_{r'}) \right) - \text{cov}(W_{j,k}(s_r), W_{j,k}(s_{r'}))
\]
\[
= 0.
\]

Showing the asymptotic normality is the same as showing for \( \forall a_{j,r,r'} \in \mathbb{R} \):
\[
\sum_{j=1}^{J} \sum_{r=1}^{m} \sum_{r'=1}^{m} a_{j,r,r'} \frac{1}{\sqrt{N_j}} \sum_{k=1}^{N_j} (W_{j,k}(s_r)W_{j,k}(s_{r'}))
\]
116
converges to a normal distribution as $N \to \infty$.

We construct a $(N - 1)m$-dim vector $\xi_N$ and a $(N - 1)m \times (N - 1)m$ dimensional square matrix $A_n$ to rewrite the above expression in the quadratic form: $Q_N = \xi_N^T A_N \xi_N = \sum_{j=1}^{J} \sum_{r=1}^{m} \sum_{r'=1}^{m} \frac{1}{\sqrt{N_j}} \sum_{k=1}^{N_j} W_{j,k}(s_r) W_{j,k}(s_{r'})$. The $(N - 1)m$-dim vector $\xi_N$ has the form:

$$
\xi_N^T = [\xi_{1,s_1}^T, \xi_{1,s_2}^T, \ldots, \xi_{1,s_m}^T, \ldots, \xi_{J,s_1}^T, \ldots, \xi_{J,s_m}^T]^T,
$$

(5.6)

with

$$
\xi_{j,s_q} = \rho_{j,1}^{-1} [W_{j,1}(s_q), \ldots, W_{j,N_j}(s_q)]^T,
$$

for $j = 1, \ldots, J$, and the $(N - 1)m \times (N - 1)m$ matrix $A_N$ follows:

$$
A_N = \frac{1}{\sqrt{N_1}} \begin{bmatrix}
A_{N,1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & A_{N,J}
\end{bmatrix},
$$

(5.7)

where

$$
A_{N,j} = \begin{bmatrix}
a_{j,1,1} & \cdots & 0 & a_{j,1,2} & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & a_{j,1,1} & 0 & \cdots & a_{j,1,2} & \cdots & a_{j,1,m} \\
a_{j,2,1} & \cdots & 0 & a_{j,2,2} & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & a_{j,2,1} & 0 & \cdots & a_{j,2,2} & \cdots & a_{j,2,m} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \ddots & \vdots \\
0 & \cdots & a_{j,m,1} & 0 & \cdots & a_{j,m,2} & \cdots & a_{j,m,m}
\end{bmatrix}_{N_j \times N_j}
$$

Since $a_{j,r,r'}$ and $a_{j,r',r}$ are the coefficients of the same terms $W_{j,k}(s_r) W_{j,k}(s_{r'})$, $k = 1, \ldots, N_j$, for arbitrary real values of $a_{j,r,r'}$ and $a_{j,r',r}$, we can assign the average of the two as their new values to make them equal. This means $a_{j,r,r'} = a_{j,r',r}$, and the matrix $A_{N,j}$ becomes a symmetric matrix.

As we did in Theorem 3.1, Theorem 5.1 will be proved in two steps:
• Show that \( \lim_{N \to \infty} \rho(A_N)\rho(\Gamma_\xi) = 0 \), where \( \rho(\cdot) \) is the spectral radius, and \( \Gamma_\xi = \text{cov}(\xi_N) \).

• Show that \( \lim_{N \to \infty} \text{var}(Q_N) = \sigma^2 \in [0, \infty) \).

Lemma 5.5.

\[
\lim_{N \to \infty} \rho(A_N)\rho(\Gamma_\xi) = 0,
\]
where the matrix \( A_N \) is defined by Equation 5.7, and \( \Gamma_\xi = \text{cov}(\xi_N) \) is the covariance matrix of the vector \( \xi_N \) defined in Equation 5.6.

Proof. By permuting rows and columns of \( A_{N,j} \), we can get a block diagonal matrix with the same submatrices on the diagonals:

\[
A_{N,j}^p = \begin{bmatrix}
  a_{j,1,1} & \cdots & a_{j,1,m} & \cdots & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  a_{j,m,1} & \cdots & a_{j,m,m} & \cdots & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  0 & \cdots & 0 & \cdots & a_{j,1,1} & \cdots & a_{j,1,m} \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  0 & \cdots & 0 & \cdots & a_{j,m,1} & \cdots & a_{j,m,m}
\end{bmatrix}_{m \times N_j, m}.
\]

Thus we can permute the rows and columns of \( A_N \) to make it a block diagonal matrix with repeated block diagonals \( A'_{N,j}, j = 1, \ldots, J \):

\[
A'_{N,j} = \begin{bmatrix}
  a_{j,1,1} & \cdots & a_{j,1,m} \\
  \vdots & \ddots & \vdots \\
  a_{j,m,1} & \cdots & a_{j,m,m}
\end{bmatrix}, \quad j = 1, \ldots, J.
\]

Because eigenvalues of a block diagonal matrix are the eigenvalues of the blocks combined [Co, 2013, page 113], the eigenvalues of \( A_N \) are the combined eigenvalues of the matrices \( A'_{N,1}, \ldots, A'_{N,J} \). We mentioned previously that we can assign the same
value to $a_{j,r,r'}$ and $a_{j,r',r}$, thus the matrix $A'_{N,j}$ is symmetric. The maximum eigenvalue of $A'_{N,j}$, among all $j$, is a real value. Thus $\lim_{N_1 \to \infty} \rho(A'_{N,j})/\sqrt{N_1} = 0$ for all $j$, and $\lim_{N_1 \to \infty} \rho(A_N) = \max_{1 \leq j \leq J} \left\{ \frac{1}{\sqrt{N_1}} \rho(A_{N,j}) \right\} \to 0$.

Because $\Gamma_\xi$ is a covariance matrix, then the submatrices

$$
\Gamma_{j,r} = \text{cov}(W_{j,1}(s_r), W_{j,2}(s_r), \cdots, W_{j,N_j}(s_r)), \quad j = 1, \ldots, J, \quad r = 1, \ldots, m
$$

are on the diagonals of $\Gamma_\xi$. By Lemma 3.5, the following holds:

$$
\rho(\Gamma_\xi) \leq \sum_{j=1}^{J} \sum_{r=1}^{m} \rho(\Gamma_{j,r}).
$$

According to Lemma 3.3, since $\{W_{j,1}(s_r), W_{j,2}(s_r), \cdots, W_{j,N_j}(s_r)\}$ is stationary in time for each $j$ in $1, \ldots, J$, $\rho(\Gamma_j)$ is bounded by $2\pi ||g||_\infty$ with $g$ being the spectral density of this stationary process (as defined in Chapter 4). $\square$

**Lemma 5.6.** The variance of $Q_N$ converges to a non-negative real value as $N_1 \to \infty$. 

---

119
Proof. The quadratic form $Q_N = \xi_N^T A_N \xi_N$, so we have:

$$\text{var}(Q_N) = \text{var}\left( \sum_{j=1}^{J} \sum_{r=1}^{m} \sum_{r'=1}^{m} a_{j,r,r'} \frac{1}{\sqrt{N_j}} \sum_{k=1}^{N_j} W_{j,k}(s_r) W_{j,k}(s_{r'}) \right)$$

$$= \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \frac{1}{\sqrt{N_j} N_{j'}} \sum_{r=1}^{m} \sum_{r'=1}^{m} \sum_{r''=1}^{m} a_{j,r,r'} a_{j',r',r''} \times \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \text{cov} (W_{j,k}(s_r) W_{j,k}(s_{r'}), W_{j',k'}(s_{r'}) W_{j',k'}(s_{r''}))$$

$$= \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \frac{1}{\sqrt{N_j} N_{j'}} \sum_{r=1}^{m} \sum_{r'=1}^{m} \sum_{r''=1}^{m} a_{j,r,r'} a_{j',r',r''} \times \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \text{cov} (W_{j,k}(s_r) W_{j,k}(s_{r'}), W_{j',k'}(s_{r'}) W_{j',k'}(s_{r''}))$$

$$+ \sum_{j \neq j'} \sum_{k \neq k'} \frac{1}{\sqrt{N_j} N_{j'}} \sum_{r=1}^{m} \sum_{r'=1}^{m} \sum_{r''=1}^{m} a_{j,r,r'} a_{j',r',r''} \times \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \text{cov} (W_{j,k}(s_r) W_{j,k}(s_{r'}), W_{j',k'}(s_{r'}) W_{j',k'}(s_{r''})) \Bigg\}.$$
Let \( S_{W^2_j,s,s^*,s^*} \) be the SDF of the within-scale wavelet product process,
\[
\text{cov} (W_{j,k}(s)W_{j,k}(s^*), W_{j,k'}(s^*)_j) = \int_{-1/2}^{1/2} S_{W^2_j,s,s^*,s^*}e^{2\pi f(k-k')} df.
\]

Then the first term in the expression of \( \text{var}(Q_N) \), \( K_N \), can be written as:
\[
K_N = \sum_{r,r',r^*,r^*' \atop j=1}^{N_j-1} \sum_{u=-(N_j-1)}^{N_j-1} a_{j,r,r'}a_{j,r^*,r^*'} \sum_{u=-(N_j-1)}^{N_j-1} \left( 1 - \frac{|u|}{N_j} \right) \\
\times \text{cov} (W_{j,k}(s_r)W_{j,k}(s_{r'}), W_{j,k+u}(s_{r'})W_{j,k+u}(s_{r^*}))
\]
\[
= \sum_{r,r',r^*,r^*' \atop j=1}^{J} \sum_{u=-(N_j-1)}^{N_j-1} \left( a_{j,r,r'}a_{j,r^*,r^*'} \right) \\
\times \int_{-1/2}^{1/2} \sum_{u=-(N_j-1)}^{N_j-1} \left( 1 - \frac{|u|}{N_j} \right) e^{2\pi fu} S_{W^2_j,s,s^*,s^*,s^*,s^*} df.
\]

As discussed in Chapter 3 the Fejér's kernel \( \sum_{h=-(N_j-1)}^{N_j-1} (1 - |h|/N_j) \) behaves like a Dirac Delta function (with a point mass at zero frequency) when \( N_j \to \infty \), so we have
\[
\lim_{N \to \infty} K_N = \sum_{r,r',r^*,r^*' \atop j=1}^{J} \sum_{u=-(N_j-1)}^{N_j-1} a_{j,r,r'}a_{j,r^*,r^*'} S_{W^2_j,s,s^*,s^*,s^*,s^*}(0).
\]

Now we will show the second part of \( \text{var}(Q_N) \) has a limit when \( N \to \infty \). Let
\[
B_N = \sum_{j=1}^{J} \sum_{k=1}^{N_j} \frac{1}{\sqrt{N_j}} \sum_{j' \neq k'}^{N_j} \sum_{k'=1}^{N_j} \frac{1}{\sqrt{N_{j'}}} \sum_{r,r',r^*,r^*' \atop j'=1}^{J} a_{j,r,r'}a_{j',r^*,r^*'} \sum_{u=-(N_{j'}-1)}^{N_{j'}} \text{cov} (W_{j,k}(s_r)W_{j,k}(s_{r'}), W_{j',k'}(s_{r'})W_{j',k'}(s_{r^*}))
\]
\[
= 2 \sum_{j=1}^{J} \sum_{r,r',r^*,r^*' \atop j'=1}^{J} B_{N,j},
\]
where
\[
B_{N,j} = \sum_{0<j'<j} \frac{a_{j',r^*,r^*'}}{\sqrt{N_{j'}}} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \text{cov} (W_{j,k}(s_r)W_{j,k}(s_{r'}), W_{j',k'}(s_{r'})W_{j',k'}(s_{r^*}))
\]
\[
= \sum_{0<j'<j} \frac{a_{j',r^*,r^*'}}{\sqrt{N_{j'}}} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \sum_{u=0}^{2^{j'-j'-1}} \text{cov} (W_{j,k}(s_r)W_{j,k}(s_{r'}), W_{j',2j'-j'k'+u}(s_{r'})W_{j',2j'-j'k'+u}(s_{r^*})).
\]

121
By the covariance stationary result in Lemma 5.4 and the definition of be the vector-valued SDF \( A_{j,v}(s, s', s^*, s'^*, f) \) (Equation (5.5)) we have

\[
B_{N,j} = \sum_{0 < j' < j} \frac{a_{j',r,r^*,r'^*}}{N_{j'}} \sum_{k=1}^{N_j} \sum_{k'=1}^{N_{j'}} \left[ 1^T \int_{-1/2}^{1/2} e^{i2\pi f(k-k^*)} A_{j,v}(s, s', s^*, s'^*, f) df \right]
\]

\[
= \sum_{1 \leq j' < j} \frac{a_{j',r,r^*,r'^*}}{N_{j'}} \sum_{u=-(N_j-1)}^{N_j-1} (N_j - |u|) \left[ 1^T \int_{-1/2}^{1/2} e^{i2\pi fu} A_{j,v}(s, s', s^*, s'^*, f) df \right].
\]

Thus we have

\[
B_N = 2 \sum_{j=1}^{J} \sum_{r,r',r^*,r'^*} B_{N,j}
\]

\[
= 2 \sum_{j=1}^{J} \sum_{r,r',r^*,r'^*} \sum_{1 \leq j' < j} \frac{a_{j',r,r^*,r'^*}}{2j/2-j'/2} \left[ 1^T \int_{-1/2}^{1/2} \sum_{u=-(N_j-1)}^{N_j-1} (1 - |u|) \frac{N_j}{N_j} e^{i2\pi fu} A_{j,v}(s, s', s^*, s'^*, f) df \right].
\]

In the above equation, the Fejér’s kernel \( \sum_{u=-(N_j-1)}^{N_j-1} (1 - |u|) \) again has a point mass at zero when \( N_j \to \infty \). Therefore,

\[
\lim_{N \to \infty} B_N = 2 \sum_{j=1}^{J} \sum_{r,r',r^*,r'^*} \sum_{1 \leq j' < j} \frac{a_{j',r,r^*,r'^*}}{2j/2-j'/2} \left[ 1^T A_{j,v}(s, s', s^*, s'^*, 0) \right].
\]

The limiting variance of \( Q_N \) is

\[
\lim_{N \to \infty} \text{var}(Q_N) = \lim_{N \to \infty} (K_N + B_N)
\]

\[
= \sum_{r,r',r^*,r'^*} \sum_{j=1}^{J} a_{j,r,r^*,r'^*} a_{j',r,r^*,r'^*} S_{W_j^2,s_r,s_{r'},s_{r^*},s_{r'^*}}(0)
\]

\[
+ 2 \sum_{j=1}^{J} \sum_{r,r',r^*,r'^*} \sum_{1 \leq j' < j} \frac{a_{j',r,r^*,r'^*}}{2j/2-j'/2} \left[ 1^T A_{j,v}(s, s', s^*, s'^*, 0) \right]
\]

\[
\equiv \sigma_{Q_N}^2 \in [0, \infty).
\]
Proof of Theorem 5.1. Because \( \lim_{N \to \infty} \rho(A_N) \rho(\Gamma \epsilon) = 0 \) and \( \lim_{N \to \infty} \text{var}(Q_N) = \sigma^2_{Q_n} \) is nonnegative, based on Lemma 3.4, the quadratic form \( Q_N \) converges to a normal distribution. Thus for \( \forall a_j \in \mathbb{R}, j = 1, \ldots, J \),

\[
\sum_{j=1}^{J} a_j \sqrt{N_1} \text{vec} \left( \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k} W_{j,k}^T - \Gamma_{W_j,0} \right)
\]

converges to a mean zero, \( m^2 \)-dimensional multivariate normal distribution as \( N_1 \to \infty \). By Cramér-Wold device [e.g., Ash and Doléans-Dade, 2000, page 343],

\[
\sqrt{N_1} \left( \rho_{1,1} \text{vec}(\hat{\Gamma}_{W_1,0} - \Gamma_{W_1,0})^T, \ldots, \rho_{J,1} \text{vec}(\hat{\Gamma}_{W_J,0} - \Gamma_{W_J,0})^T \right)^T \overset{D}{\to} N_{m^2J}(0, \Sigma),
\]

as \( N_1 \to \infty \). The limiting covariance matrix \( \Sigma \) is a block matrix

\[
\Sigma = \begin{bmatrix}
\Sigma_{1,1} & \cdots & \Sigma_{1,J} \\
\vdots & \ddots & \vdots \\
\Sigma_{J,1} & \cdots & \Sigma_{J,J}
\end{bmatrix},
\]

where

\[
\Sigma_{j,j'} = \text{cov}(W_{j,k}(s)W_{j',k}(s'))W_{j,k}(s^*W_{j',k}(s^*')), s, s', s^*, s'^* \in \{s_1, \ldots, s_m\}.
\]

This completes the proof of Theorem 5.1. \( \square \)

In order to prove Theorem 5.2, we first show the results stated in Lemma 5.7 and Lemma 5.8.

Lemma 5.7. Assume that the wavelet Whittle (log) likelihood is differentiable with respect to the parameter vector \( \theta \), and assume that \( L \geq 2d \). Then the first derivative of the wavelet Whittle (log) likelihood, \( \frac{\partial}{\partial \theta} l_W(\theta) \), is jointly asymptotically normal as \( N_1 \to \infty \).
Proof. Recall that the wavelet Whittle likelihood under the white noise approximation model (Equation 5.3):

\[
l_W(\theta) = -\frac{(N - 1)m}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^{N_j} \log(\det \Gamma_{W_{j,0}}) \\
- \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^{N_j} W_{j,k}^T \Gamma_{W_{j,0}}^{-1} W_{j,k},
\]

with \( \Gamma_{W_{j,0}} = \text{cov}(W_{j,k}) \), for all \( k = 1, \ldots, N_j \).

Now we need the first and the second derivatives of the approximate log likelihood with respect to each parameter. For \( q = 1, \ldots, p \),

\[
\frac{\partial}{\partial \theta_q} \log(\det(\Gamma_{W_{j,0}})) = \frac{1}{\det \Gamma_{W_{j,0}}} \frac{\partial}{\partial \theta_q} (\det \Gamma_{W_{j,0}}) = \frac{1}{\det \Gamma_{W_{j,0}}} \det \Gamma_{W_{j,0}} \text{tr} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \right) = \text{tr} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \right),
\]

and

\[
\frac{\partial}{\partial \theta_q} \left( W_{j,k}^T \Gamma_{W_{j,0}}^{-1} W_{j,k} \right) = \frac{\partial}{\partial \theta_q} \left( \sum_{r=1}^m \sum_{r'=1}^m W_{j,k}(s_r) \left[ \Gamma_{W_{j,0}}^{-1} \right]_{r,r'} W_{j,k}(s'_{r'}) \right) \\
= \sum_{r=1}^m \sum_{r'=1}^m W_{j,k}(s_r) W_{j,k}(s'_{r'}) \frac{\partial}{\partial \theta_q} \left[ \Gamma_{W_{j,0}}^{-1} \right]_{r,r'} \\
= W_{j,k}^T \frac{\partial}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} W_{j,k} \\
= W_{j,k}^T \left( -\Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} \right) W_{j,k} \\
= \text{tr} \left( -\Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} W_{j,k} W_{j,k}^T \right).
\]
So the first derivative of the approximate likelihood is

\[
\frac{\partial}{\partial \theta} l^W(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \frac{\partial}{\partial \theta} \log(\det \Gamma_{W_j,0}) - \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \frac{\partial}{\partial \theta} W_{j,k} \Gamma_{W_j,0}^{-1} W_{j,k}
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} N_j \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \right) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left( -\Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right)
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \left\{ \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \right) - \text{tr} \left( -\Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right\}
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right)
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta} \left( \Gamma_{W_j,0} - W_{j,k} W_{j,k}^T \right) \right).
\]
Because only the term $W_{j,k}W_{j,k}^T$ depends on the index $k$, we rewrite the above equation. For $q = 1, \cdots, p$, we have:

$$
\frac{\partial}{\partial \theta_q} f_W(\theta)
$$

$$
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} (\Gamma_{W_{j,0}} - W_{j,k}W_{j,k}^T) \right)
$$

$$
= -\frac{1}{2} \sum_{j=1}^{J} \text{tr} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} \sum_{k=1}^{N_j} (\Gamma_{W_{j,0}} - W_{j,k}W_{j,k}^T) \right)
$$

$$
= -\frac{1}{2} \sum_{j=1}^{J} N_j \text{tr} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} \left( \Gamma_{W_{j,0}} - \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}W_{j,k}^T \right) \right)
$$

$$
= -\frac{1}{2} \sum_{j=1}^{J} N_j \sum_{r=1}^{m} \sum_{r'=1}^{m} \left[ \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} \right]_{r,r'} \left[ \Gamma_{W_{j,0}} - \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}W_{j,k}^T \right]_{r',r'}
$$

$$
= -\frac{1}{2} \sum_{j=1}^{J} N_j \text{vec} \left( \Gamma_{W_{j,0}}^{-1} \frac{\partial \Gamma_{W_{j,0}}}{\partial \theta_q} \Gamma_{W_{j,0}}^{-1} \right)^T \text{vec} \left( \Gamma_{W_{j,0}} - \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k}W_{j,k}^T \right).
$$
Thus in a vector form, the first derivative of the likelihood scaled by $1/\sqrt{N_1}$ is

$$
\frac{1}{\sqrt{N_1}} \frac{\partial}{\partial \theta} l^W(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \frac{N_j}{\sqrt{N_1}} \begin{bmatrix}
\text{vec} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_1} \Gamma_{W_j,0}^{-1} \right) & \cdots & \text{vec} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_p} \Gamma_{W_j,0}^{-1} \right)
\end{bmatrix}^T \begin{bmatrix}
\Gamma_{W_j,0} - \frac{1}{N_j} \sum_{k=1}^{N_j} W_{j,k} W_{j,k}^T
\end{bmatrix}.
$$

The above result shows that the first derivative of the likelihood, $\frac{\partial}{\partial \theta} l^W(\theta)$, is a linear combination of the jointly asymptotically normal vector

$$
\sqrt{N_1} \begin{bmatrix}
\rho_{1,1} \text{vec}(\hat{\Gamma}_{W_1,0} - \Gamma_{W_1,0})^T, \ldots, \rho_{J,1} \text{vec}(\hat{\Gamma}_{W_J,0} - \Gamma_{W_J,0})^T
\end{bmatrix}.
$$

The linear association is established by the following $p \times m^2 J$ matrix,

$$
B = -\frac{1}{2} \begin{bmatrix}
\rho_{1,1} \text{vec} \left( \Gamma_{W_1,0}^{-1} \frac{\partial \Gamma_{W_1,0}}{\partial \theta_1} \Gamma_{W_1,0}^{-1} \right) & \cdots & \rho_{J,1} \text{vec} \left( \Gamma_{W_J,0}^{-1} \frac{\partial \Gamma_{W_J,0}}{\partial \theta_p} \Gamma_{W_J,0}^{-1} \right)
\end{bmatrix}.
$$

Thus $\frac{\partial}{\partial \theta} l^W(\theta)$ is also jointly asymptotically normal with mean zero and covariance matrix $B \Sigma B^T$. \qed
Lemma 5.8. Assume that the wavelet Whittle (log) likelihood is twice differentiable with respect to the parameter vector \( \theta \), and assume that \( L \geq 2d \). Then the Hessian matrix

\[
H(\theta^*) = \frac{\partial^2}{\partial \theta^2} l^W(\theta) \bigg|_{\theta = \theta^*}
\]

converges to a nonnegative definite matrix that depends on \( \theta_0 \). In the expression \( \theta^* = \theta_0 + ch \), where \( h = \hat{\theta} - \theta_0 \) is small, and \( c \in (0, 1) \).

Proof. Let \( H(\theta) = \frac{\partial^2}{\partial \theta^2} l^W(\theta) \) denote the Hessian matrix of the approximate likelihood. We can derive the \((q, q')\)th entry of the matrix \( H(\theta) \):

\[
\frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} l^W(\theta) = \frac{\partial}{\partial \theta_q} \left\{ -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left( \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_q} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right) \right\} \right.
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left\{ \frac{\partial \Gamma_{W_j,0}}{\partial \theta_q} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_{q'}} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right\}
\]

\[
\left. + \Gamma_{W_j,0}^{-1} \frac{\partial}{\partial \theta_q} \left( \frac{\partial \Gamma_{W_j,0}}{\partial \theta_{q'}} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right) \right\}
\]

\[
= -\frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{N_j} \text{tr} \left\{ -\Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_q} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_{q'}} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right. \right.
\]

\[
\left. + \Gamma_{W_j,0}^{-1} \frac{\partial^2 \Gamma_{W_j,0}}{\partial \theta_q \partial \theta_{q'}} \left( I - \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right) \right. \right.
\]

\[
\left. + \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_{q'}} \Gamma_{W_j,0}^{-1} \frac{\partial \Gamma_{W_j,0}}{\partial \theta_q} \Gamma_{W_j,0}^{-1} W_{j,k} W_{j,k}^T \right\}. \quad (5.9)
\]
As we previously shown \( \left( \sum_{k=1}^{N_j} W_{j,k} W_{j,k}^T \right) / N_j \) is an unbiased estimator of \( \Gamma_{Wj,0} \), we have

\[
\frac{1}{N_1} \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} l^W(\theta) = -\frac{1}{2} \sum_{j=1}^{J} \left[ \frac{1}{2j-1} \sum_{k=1}^{N_j} \right] \text{tr} \left\{ -\Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_q} \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_{q'}} \left( I - \Gamma_{Wj,0}^{-1} \frac{1}{N_j} W_{j,k} W_{j,k}^T \right) \right. \\
+ \Gamma_{Wj,0}^{-1} \frac{\partial^2 \Gamma_{Wj,0}}{\partial \theta_q \partial \theta_{q'}} \left( I - \Gamma_{Wj,0}^{-1} \frac{1}{N_j} W_{j,k} W_{j,k}^T \right) \\
+ \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_{q'}} \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_q} \left( I - \Gamma_{Wj,0}^{-1} \frac{1}{N_j} W_{j,k} W_{j,k}^T \right) \}.
\]

With the consistency assumption on \( \hat{\theta} \), we have \( \theta^* \to \theta_0 \) as \( N_1 \to \infty \), and

\[
\lim_{N_1 \to \infty} \frac{1}{N_1} \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} l^W(\theta^*) = -\sum_{j=1}^{J} \frac{1}{2j} \text{tr} \left( \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_q} \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_{q'}} \right).
\]

Thus the limit of \( -H(\theta^*)/N_1 \) is

\[
K = \sum_{j=1}^{J} \frac{1}{2j} \left[ \text{tr} \left( \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_1} \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_1} \right) \cdots \text{tr} \left( \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_p} \Gamma_{Wj,0}^{-1} \frac{\partial \Gamma_{Wj,0}}{\partial \theta_p} \right) \right],
\]

\[
(5.10)
\]

which is a Gram matrix (or Gramian matrix) with trace as the inner product (Frobenius inner product). Because Gram matrices are nonnegative definite [Schwerdtfeger, 1950], we have the result that \( -H(\theta^*)/N_1 \) converges in probability to a nonnegative definite matrix \( K \), where \( K \) depends on \( \theta_0 \).

\[
\square
\]

Proof of Theorem 5.2. By the Taylor expansion of the wavelet-Whittle likelihood at \( \theta_0 \),

\[
l_N^{W'}(\hat{\theta}) = l_N^{W'}(\theta_0) + H(\theta^*)(\hat{\theta} - \theta_0).
\]

\[
(5.11)
\]
In the above expression \( \theta^* \) is defined earlier as \( \theta_0 + ch \), where \( h = \hat{\theta} - \theta_0 \) and \( c \in (0, 1) \).

Since \( \hat{\theta} \) is a maximum likelihood estimator \( l^W_N(\hat{\theta}) = 0 \), we have

\[
\sqrt{N_1}(\hat{\theta} - \theta_0) = -\left( \frac{1}{N_1} l^W''(\theta^*) \right)^{-1} \left( \frac{1}{N_1} l^W_N(\theta_0) \right)
= -\left( \frac{1}{N_1} H(\theta^*) \right)^{-1} \left( \frac{1}{N_1} l^W_N(\theta_0) \right).
\]

Because Lemma 5.7 states that \( \left( \frac{1}{N_1} l^W_N(\theta_0) \right) \) is asymptotically joint normal with zero mean, and Lemma 5.8 states that the limit of \( -1/N_1H(\theta^*) \) is a nonnegative definite matrix. Using Slutsky’s theorem [e.g., Resnick, 1999, page 268], we have \( \sqrt{N_1}(\hat{\theta} - \theta_0) \) converges to a \( p \)-dimensional normal distribution with mean zero, and covariance matrix \( \Omega = KB\Sigma B^T K^T \), where the matrix \( K \) is defined by Equation (5.10), the matrix \( B \) is defined by Equation (5.8), and \( \Sigma \) as defined in Theorem 5.1. All three matrices depend on \( \theta_0 \). \( \square \)

5.4 Monte Carlo studies

We design and conduct Monte Carlo studies to investigate the performance of the wavelet Whittle estimators of the parameters in a space-time model. Consider Gneiting’s nonseparable covariance model (Equation (4.17)),

\[
C(h; u) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)(a|u|^{2\alpha} + 1)^{\delta + \beta d/2}} \left( \frac{c||h||}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right)^
u \times K_\nu \left( \frac{c||h||}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right), \quad (h; u) \in \mathbb{R}^d \times \mathbb{R}.
\]

The parameters \( a, c \) and \( \delta \) are nonnegative real numbers, and \( \sigma^2 > 0 \). The ranges of the smoothness parameters are \( \alpha \in (0, 1) \) and \( \nu \in (0, +\infty) \). The parameter \( \beta \in [0, 1] \) here represents the space-time interaction of the model: when \( \beta = 0 \) the covariance is separable, while \( \beta = 1 \) means the covariance function is nonseparable and has
the highest space-time interaction [Gneiting, 2002]. We assume that the covariance model has the spatial smoothness parameter $\nu = 1/2$ (the spatial covariance model is exponential), and let the smoothness parameter in time be $\alpha = 1/2$. For the other parameters we set $\delta = 0$, $d = 2$, $\sigma^2 = 1$, and $\beta = 1$. Thus the space-time covariance model in this simulation study has the expression:

$$
C(\mathbf{h}; u) = \frac{1}{(a |u| + 1)} \exp \left( - \frac{c ||\mathbf{h}||}{(a |u| + 1)^{1/2}} \right), \quad (\mathbf{h}; u) \in \mathbb{R}^2 \times \mathbb{Z}. \tag{5.12}
$$

We are interested in estimating $a$ and $c$, the scaling parameters in space and in time, respectively.

The simulated spatio-temporal data are from $M = 20$ equally spaced spatial locations on a line between 0 and 1 (including 0 and 1), and are observed at $N = 2^8 = 256$ time points. With the statistical software package R [R Core Team, 2014], a sequence of $N$ autocovariance matrices $\Gamma_{Y,\tau}$ (Equation 4.12) are calculated by Equation (5.12), then the data are generated using the multivariate Levinson-Durbin algorithm [e.g., Brockwell and Davis, 1991, page 432]. The wavelet-Whittle white noise approximation (Equation (5.3)) is implemented to estimate the two parameters simultaneously. For comparison purpose, the exact likelihood method is implemented as well. The simulation results are based on 5000 replications, and the estimation bias, SD, and RMSE are summarized in Tables 5.1 – 5.2. As expected, the biases of the estimates are smaller with the exact method than with the wavelet Whittle method. There is a situation that D(6) filter outperforms the LA(8) filter with a significantly smaller bias, but in general different wavelet filters have similar performances in terms of bias. The measurements of uncertainties, SD and RMSE, are very close for both exact and the wavelet Whittle approaches. Unlike the univariate case (Section 3.6), there is no significant different between the wavelet filters in terms of SD and RMSE.
For this simulation study it is not necessary that longer filters lead to estimates of $a$ that are better than with short wavelet filters.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
 & Bias & SD & RMSE \\
 & EST & SE & EST & SE & EST & SE \\
\hline
Exact & 0.0020 & 0.0007 & 0.0536 & 0.0006 & 0.0537 & 0.0006 \\
WN & D(4) & -0.0054 & 0.0008 & 0.0575 & 0.0006 & 0.0577 & 0.0006 \\
 & D(6) & -0.0028 & 0.0008 & 0.0567 & 0.0005 & 0.0568 & 0.0005 \\
 & D(8) & -0.0033 & 0.0008 & 0.0562 & 0.0006 & 0.0563 & 0.0006 \\
 & LA(8) & -0.0055 & 0.0008 & 0.0561 & 0.0006 & 0.0563 & 0.0006 \\
\hline
\end{tabular}
\caption{Simulation results of the temporal scaling parameter $a$.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
 & Bias & SD & RMSE \\
 & EST & SE & EST & SE & EST & SE \\
\hline
Exact & 0.0001 & 0.0003 & 0.0237 & 0.0002 & 0.0237 & 0.0002 \\
WN & D(4) & 0.0047 & 0.0003 & 0.0240 & 0.0003 & 0.0245 & 0.0003 \\
 & D(6) & 0.0054 & 0.0003 & 0.0242 & 0.0002 & 0.0248 & 0.0002 \\
 & D(8) & 0.0060 & 0.0003 & 0.0240 & 0.0002 & 0.0247 & 0.0003 \\
 & LA(8) & 0.0055 & 0.0003 & 0.0242 & 0.0002 & 0.0248 & 0.0003 \\
\hline
\end{tabular}
\caption{Simulation results of the spatial scaling parameter $c$.}
\end{table}

The computation time of one replication is listed in Table 5.3. For each replication we include both the time spent simulating the space-time data, and the time for estimation. included. The wavelet Whittle method with some filters, e.g., D(4), is about 2.5 times faster than the exact likelihood method. Longer filters such as LA(8) costs more time, but is still slightly faster than the exact method. The computation time could be faster if one is using other languages (e.g., C) instead of R. Considering the fact there are only 256 time points and 20 sites, the wavelet Whittle estimation of
a finite sample could be highly improved if the method is applied to analyze a longer sequence of space-time data observed at a larger number of sites.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Wavelet Whittle approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation time (per replicate)</td>
<td>Exact</td>
</tr>
<tr>
<td></td>
<td>D(4)</td>
</tr>
</tbody>
</table>

Table 5.3: Simulation time for the space-time estimation (in seconds).

5.5 Application: Irish wind data

In this section, the wavelet Whittle method is applied to analyze the Irish wind data [Haslett and Raftery, 1989]. The Irish wind power data was obtained from the R package “gstat”, and it consists of daily average wind speeds collected at 12 meteorological stations between 1961 and 1978. Statistical analyses of this dataset usually exclude the station at Rosslare because the data have very low correlation with the rest of the stations. The daily wind speeds are stabilized using a square root transformation and then deseasonalized [see, e.g., Haslett and Raftery, 1989, Gneiting, 2002]. In order to compare our estimation results with the results of other methods, we follow these pre-processing procedures, and call the processed data, “velocity measures”.

We perform the DWT of the velocity measures in the time domain with the LA(8) filter. Figures 5.1 – 5.3 show plots of the velocity empirical spectrum and level 1 to level 5 wavelet spectrums of each station. The whitening effects are very obvious in these plots: the wavelet spectrums on each level are very flat (“white”) compared to the process spectrums, especially at higher wavelet levels. This is a good
motivation for using wavelet-based methods to analyze these wind data. Since the wavelet coefficients are approximately uncorrelated, as the wavelet-Whittle method assumed in this chapter, we can use some simpler spectrums (WN or AR(1)) to approximate the wavelet spectrums within each scale.

Assume the data follows the space-time model (Equation (4.17)) with $d = 2$, $\nu = 1/2$, because this model has previously been shown to fit well to these data [Gneiting, 2002]. We want to estimate the scaling parameters $a$ and $c$, the temporal smoothness parameter $\alpha$, the parameter that measures the space-time separability, $\beta$, as well as $\delta$ and $\sigma^2$. The space-time covariance function belongs to the family:

$$C(h; u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{-(\beta+\delta)}} \exp\left(-\frac{c||h||}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right), \quad (h; u) \in \mathbb{R}^2 \times \mathbb{Z}. \quad (5.13)$$

Because the DWT requires a multiple of a power of two as the sample size, the length of the velocity measures is truncated to $51 \times 2^7 = 6528$ time points. Based on the white noise and AR(1) wavelet models fit to $J = 7$ levels of wavelet coefficients of the wind speed dataset, the parameter estimates are summarized in Table 5.4. Comparing our estimates to those in Gneiting [2002], the values for scaling parameters are close, e.g., we have $\hat{c}_{WN} = 0.00141$ and their value is 0.00134. But the estimated $\beta$ is very different: their estimated value of 0.61 is obtained by a weighted-least-squares method [Cressie, 1993, page 96], but we have $\hat{\beta}_{WN} = 0.9194$, which indicates a stronger interaction in space and time. This may due to the fact that they are not estimating the parameters simultaneously. They first fit the temporal correlation to get estimates for $a$ and $\alpha$, then consider the spatial correlation parameters.

To assess the fit of the model, space-time correlations and cross (stations) correlations are drawn in Figure 5.4 and Figure 5.5, respectively. Figure 5.4 shows the spatial correlation at time lag 0 (the left panel), and at time lag 1 (the right panel).
Figure 5.1: The empirical spectrum of the process and the DWT coefficients (with LA(8) filter, $J = 7$) at stations Valentia, Belmullet, Claremorris and Shannon.
Figure 5.2: The empirical spectrum of the DWT coefficients (with LA(8) filter, $J = 7$) at stations Roche’s Point, Birr, Mullingar and Malin Head.
Figure 5.3: The empirical spectrum of the DWT coefficients (with LA(8) filter, $J = 7$) at stations Kilkenny, Clones and Dublin.
Table 5.4: Estimation results for the wind speed covariance model.

<table>
<thead>
<tr>
<th></th>
<th>Wavelet Whittle estimation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{a}$</td>
<td>$\hat{c}$</td>
</tr>
<tr>
<td>WN approximation</td>
<td>1.1993</td>
<td>0.00141</td>
</tr>
<tr>
<td>AR(1) approximation</td>
<td>1.1611</td>
<td>0.00146</td>
</tr>
</tbody>
</table>

The fitted lines (the green solid line and the purple dashed line) given by the wavelet Whittle method and using the Gneiting’s results (the red dotted line) are close for both panels. The differences in fit can be seen more clearly in Figure 5.5. In the figure we plot the cross-correlations of four pairs of stations (Valentia v.s. Belmullet, Claremorris, Shannon, and Roche’s Point). In general, the fitted cross-correlation given by the wavelet Whittle approximation is closer to the empirical cross-correlation than Gneiting’s results. Gneiting’s model slightly over estimate the correlations than our model. It seems that the AR(1) approximation result fit better than the WN approximation result when the time lag is between 3 and 25, but we also notice that none of these fitted correlations can capture the behavior of the process at large time lags (e.g., the bump in the correlation around lag 27). This is because the covariance model (Equation (4.17)) can not model such behavior, unless a more general and valid nonseparable space-time covariance is proposed.

In this application of analyzing the wind speed data, we estimate the parameters of the covariance model simultaneously, and the correlations fit better than Gneiting’s estimates do, in which his estimates are obtained separately. If needed, we note that confidence region for the parameters can be calculated using the asymptotic theory presented in Section 5.2.
Figure 5.4: Wind data space-time correlation for time lag 0 and time lag 1.
'VAL' and 'BEL'

'VAL' and 'CLA'

'VAL' and 'SHA'

'VAL' and 'RPT'

Figure 5.5: Wind data cross-correlation between two stations.
6.1 Conclusion and discussion

In this dissertation we proposed a wavelet Whittle approximation method for estimating the parameters of a general class of Gaussian process whose $d$th order backwards difference is stationary. In analyzing time series processes with complicated non-Markov dependencies, the traditional exact likelihood method is computationally intensive. As one of the popular decorrelation approaches, discrete wavelet transforms are used to decorrelate the data. The DWT can be used to filter a sequence of data, and gives wavelet coefficients (representing changes of averages) on different scales, and scaling coefficients (representing averages) on a smooth scale. Previous work has shown that the between-scale wavelet coefficients are approximately uncorrelated [e.g., Percival and Walden, 2000, Craigmile and Percival, 2005]. With the between-scale independence assumption on the wavelet coefficients, the wavelet likelihood can be used to approximate the exact likelihood. In addition, approximating the wavelet spectrum on every level using a simpler SDF leads to the wavelet Whittle approximate likelihood.

The statistical properties of the DWTs coefficients of a time series process were summarized in Chapter 2. Two propositions regarding the stationarity of the wavelet
coefficients were presented for the general class of Gaussian process. Under the assumption that the filter length $L \geq 2d$, the within-scale wavelet and scaling coefficients are stationary [Craigmile and Percival, 2005]. With the same assumption, a between scale process as constructed by Moulines et al. [2008] is covariance-stationary. We used a time series with smoothness $\nu = 1.5$ in a Matérn covariance setting to illustrate the decorrelation property of the DWTs.

Motivated by the decorrelation feature of the DWTs, we presented the wavelet Whittle method (Chapter 3), the WN and the AR(1) approximate likelihood, and an asymptotic normality result for the wavelet Whittle MLEs under a WN approximation (Theorem 3.2). Unlike other existing results with limiting theories assuming the approximate model is true [Craigmile et al., 2005, Furrer et al., 2006, Kaufman et al., 2008], our theory is obtained under the original process. Another contribution of our work is that the current wavelet-based estimation methods are mostly estimating one parameter (the long range dependence parameter, $d$), while the proposed estimation method in this dissertation can jointly estimate all the parameters and obtain their confidence regions.

The estimation performance of the wavelet Whittle method was compared with the Markov approximation with order 1 and the tapering method (one-taper). It turned out our method performed better than the Markov and the tapering method in terms of estimation bias, SD and RMSE. Long wavelet filters (e.g., LA(8) and D(8)) with an AR approximation in the wavelet Whittle method outperformed the exact likelihood method. In terms of computation time, the wavelet Whittle method was much faster than the tapering method.
In Chapter 4 we extended the DWT and wavelet Whittle methods to analyzing space-time processes. We defined the statistical properties of wavelet coefficients calculated at each location of the space-time process. Analogues of the stationarity propositions in Chapter 2 for the space-time process were provided. Then, we studied the whitening effect of the DWT in the space-time scenario using Gneiting’s (2002) nonseparable covariance model.

We introduced in Chapter 5 the wavelet Whittle method to approximate the likelihood of a space-time process. The asymptotic theory for these estimators was developed. The Monte Carlo studies in Section 5.4 compared the wavelet Whittle estimation results with the exact likelihood method. The estimates are not as good as the exact method in terms of bias because we are ignoring the covariances of wavelet coefficients within and between scales. But there is a big improvement in computation time. The approximation method with D(4) filter is about 2.5 times faster than the exact. We also made a comparison among different wavelet filters. Unlike the univariate case, there is no significant difference in space-time estimation performance among filters, but shorter filter can be much faster than longer ones, e.g., D(4) is 2.5 times faster than LA(8). Thus one can just use a short wavelet filter to achieve gain in computation time while keep a relatively good estimation performance.

In summary, contributions of this dissertation are listed below:

- While wavelet-based likelihood approximation methods are often used to estimate the long range dependence (LRD) parameter, our method can jointly estimate all parameters of a statistical model.
• We provided the asymptotic results of the wavelet Whittle MLEs under the true process, rather than under the assumed approximate model.

• The DWT of a space-time process is defined and detailed. We studied the expected value, the covariance structure of the DWT coefficients of a space-time process, as well as the whitening effect of the wavelet transform. The stationary results (extended from the univariate case) on the within and between scale coefficients are presented.

• Wavelet whittle estimation method for space time processes is developed, with the asymptotic theory for the MLEs (under white noise approximation).

6.2 Future work

In the discussions of the wavelet Whittle method for time series data, we obtained the approximate likelihood with an AR(1) model on each wavelet scale. This AR(1) approximation is also implemented in the Monte Carlo studies in Chapter 3. Generally, the AR(1) approximation performs much better than the WN approximation with slightly longer computation time. We have shown that the MLE based on the MODWT coefficients is jointly asymptotic normal, and it is easy to show that element-wise, the (DWT) wavelet Whittle MLE is asymptotically normal. However, the derivation of the joint limiting distribution of the MLEs requires a stronger stationarity results on the between-scale wavelet coefficients. One possible future direction is to further investigate the between-scale covariance structure. Open questions include: can we construct a between-scale covariance-stationary process similar to the process in Moulines et al. [2008]? What is the loss in parameter estimation when assuming independence between scales?
In the wavelet Whittle approximate likelihood, Equation (3.5), other approximating SDFs can be used. Higher orders autoregressive spectrums, MA(1), and ARMA(1,1) SDFs are possible approximation candidates.

For space-time data, although our theory assumes spatial stationary as well, non-stationarity in space should not affect the result because the method and asymptotic results mainly depend on the stationarity or the backwards stationarity in the time domain. One can drop the stationary assumption in space and extend the explore the application of the wavelet Whittle method to more general space-time processes.

Due to the time constrains and lack of a more efficient process simulation algorithm, the simulated space-time data in the Monte Carlo studies in Chapter 5 has a length of 256 in time, and is assumed to be observed at 20 equally spaced sites. The length of the data in time can be increased, and the spatial setting can be generalized to irregularly spaced sites to further study the performance of the estimators under different situations.

In Chapter 5 we only summarized the results of the WN approximation. We ran some simulation studies for the AR(1) approximation as well. There is no significant improvement in estimation performance, however the computation time is about twice as much as the WN approximation. This might be because the R code is not intensively optimized. We can not ignore the fact that incorporating some dependencies of the wavelet coefficients within a scale may not be worthwhile. The trade-off between computational and statistical efficiency is an interesting question to work on.

Since researchers have been using the 2-D DWTs in space to analyze space-time data, and we have defined the 1-D DWT in time in this dissertation, an extension is to apply the DWTs both in space and in time. This approach should be able to
largely decorrelate a process in both space and time. One limitation is that the process needs to be observed on a complete regular grid to apply the 2-D DWT [Nychka et al., 2003]. Also, the theory for Whittle estimations using the 2-D DWT still needs to be developed, with which one can give a clear definition of the DWT in both space and time, and can study the statistical properties of the “space-scale-time-scale” DWT coefficients.
Appendix A: Proof of the asymptotic normality of the approximate MLEs in the MODWT setting

To show the asymptotic results of the MODWT based MLEs in the AR(1) approximation, we can follow the proof procedure given by Mondal and Percival [2010]. The following lemmas are provided in their paper and can be used to prove a more general result as stated in Theorem 3.6 (compared with Theorem 4.2 in Mondal and Percival, 2010). Proposition 3.7 can then be proved using a similar approach as for the MLEs obtained in the DWT white noise model. In this appendix, we will show briefly the steps of the proof for Theorem 3.6.

Lemma A.1. Assume that \( L \geq 2d \). Let \( U_{p,k} = Y_{k-1-l \mod N}Y_{k-1-l' \mod N} \) and \( E U_{p,k} = \phi_p \), in which \( p = (l, l') \). Then for \( n \geq 3 \) and fixed \( p_1, \ldots, p_n \),

\[
\sum_{k_1, \ldots, k_n} |\text{cum}(U_{p_1,k_1} - \phi_{p_1}, \ldots, U_{p_n,k_n} - \phi_{p_n})| = o(M^{n/2}),
\]

where each \( k_i \) ranges from 1 to \( M \) (\( M \) is shorthand for \( M_j \) in the main text).

Lemma A.2. Assume that \( L \geq 2d \). Let \( U_{p,k} = -\frac{1}{2}(Y_{k-1-l \mod N} - Y_{k-1-l' \mod N})^2 \) and \( E U_{p,k} = \phi_p \), in which \( p = (l, l') \). Then for \( n \geq 3 \) and fixed \( p_1, \ldots, p_n \),

\[
\sum_{k_1, \ldots, k_n} |\text{cum}(U_{p_1,k_1} - \phi_{p_1}, \ldots, U_{p_n,k_n} - \phi_{p_n})| = o(M^{n/2}),
\]

where each \( k_i \) ranges from 0 to \( M \) (\( M \) is shorthand for \( M_j \) in the main text).
Lemma A.3. Let $U_{p,k}$ be either as in Lemma A.1 or as in Lemma A.2 and assume $\kappa_n(p_1, \ldots, p_n, k_1, \ldots, k_n) = \text{cum}(U_{p_1,k_1} - \phi_{p_1}, \ldots, U_{p_n,k_n} - \phi_{p_n})$. Define for $i = 1, 2, \ldots, n - 1$

$$\kappa_n(p_1, \ldots, p_n, k_1, \ldots, k_i) = \sum_{k_{i+1}, \ldots, k_n} M^{-\frac{1}{2}(n-i+1)} \kappa_n(p_1, \ldots, p_n, k_1, \ldots, k_n),$$

where the summation in $k_i$ ranges from 1 to $M$. Then $\kappa_n(p_1, \ldots, p_n, k_1, \ldots, k_i)$ is bounded and satisfies

$$\sum_{k_1, \ldots, k_i} \kappa_n(p_1, \ldots, p_n, k_1, \ldots, k_i) = o(M^{\frac{1}{2}(i+1)}), \quad i = 1, 2, \ldots, n. \quad (A.1)$$

Proof of Theorem 3.6. Let

$$Q'_k = \sum_{j=J_0}^{J} \beta_j \left( \tilde{W}_{j,k}^2 - \nu_j \right) + \sum_{j=J_0}^{2(J-J_0+1)} \beta_j^* \left( \tilde{W}_{j,k} \tilde{W}_{j,d-1} - \gamma_{\tilde{W}_j}(1) \right)$$

$$= \sum_{j'=J_0}^{2(J-J_0+1)} \beta_j'(Z_{j',k} - z_{j',k}),$$

where $\gamma_{\tilde{W}_j}(1)$ is the lag 1 wavelet covariance at level $j$, and $E(Q'_k) = 0$. Let

$$Z_{j',k} = \begin{cases} \tilde{W}_{j',k}^2 & \text{if } j' = J_0, \ldots, J \\ \tilde{W}_{j'-J,k} \tilde{W}_{j-J,k-1} & \text{if } j' = J + 1, \ldots, 2(J-J_0+1) \end{cases},$$

and let

$$z_{j',k} = \begin{cases} \nu_{j'} & \text{if } j' = J_0, \ldots, J \\ \gamma_{\tilde{W}_{j'-J}(1)} & \text{if } j' = J + 1, \ldots, 2(J-J_0+1) \end{cases}.$$
When \( j' = J + 1, \ldots, 2(J - J_0 + 1) \),

\[
Z_{j',k} = \tilde{W}_{j'-J,k} \tilde{W}_{j',k-1} \\
= \left( \sum_{l=0}^{L_{j'-J-1}} h_{j'-J,l} Y_{k-1-l} \mod N \right) \left( \sum_{l'=0}^{L_{j'-J-1}} h_{j'-J,l'} Y_{(k-1)-1-l'} \mod N \right) \\
= \sum_{l=0}^{L_{j'-J-1}} \sum_{l'=0}^{L_{j'-J-1}} h_{j'-J,l} h_{j'-J,l'} Y_{k-1-l} \mod N Y_{(k-1)-1-l'} \mod N \\
= -\frac{1}{2} \sum_{l=0}^{L_{j'-J-1}} \sum_{l'=0}^{L_{j'-J-1}} h_{j'-J,l} h_{j'-J,l'} \left( Y_{k-1-l} \mod N - Y_{(k-1)-1-l'} \mod N \right)^2.
\]

We can construct a \( U'_{l,l',k} \), which plays the role of \( U_{p,l} \) in the proof of the Theorem 4.2 in Mondal and Percival [2010]:

\[
U'_{l,l',k} = \begin{cases} 
\frac{1}{2} (Y_{k-1-l} \mod N - Y_{k-1-l'} \mod N)^2, & \text{if } j' = J_0, \ldots, J; \\
\frac{1}{2} (Y_{k-1-l} \mod N - Y_{(k-1)-1-l'} \mod N)^2, & \text{if } j' = J + 1, \ldots, 2(J - J_0 + 1).
\end{cases}
\]

It has the expected value:

\[
\theta'_{l,l'} = E(U'_{l,l',k}) = \begin{cases} 
s_{l',l} & \text{if } j' = J_0, \ldots, J \\
s_{y,2l'-J+l'-l} \mod N & \text{if } j' = J + 1, \ldots, 2(J - J_0 + 1)
\end{cases}.
\]

Follow the Lemma A.5 in Mondal and Percival [2010], \( Q'_k \) is second order stationary. Lemma A.1 through Lemma A.3 still hold just with a change in time index in the proof. Thus the CLT of \( R' = \frac{1}{M} \sum_k Q'_k \) can be shown using its log characteristic function [Mondal and Percival, 2010]:

\[
\log F(z) = \sum_{n=1}^{\infty} \frac{z^n}{n!} \sum_{k_1, \ldots, k_n} \frac{B_n(k_1, \ldots, k_n)}{M^{n/2}},
\]

where \( B_n(k_1, \ldots, k_n) \) is the nth cumulant of \( Q'_k \). We also have \( B_1(t_1) = 0 \) because \( E(Q_t) = 0 \). Based on the results in lemmas A.1–A.3, Mondal and Percival [2010]
have shown that when $M_{J_0}$ is large the third and higher order cumulants of $R'$ satisfy

$$
\sum_{t_1, \ldots, t_n} M^{-n/2} B_n(t_1, \ldots, t_n) = o(M^{n/2}),
$$

and they converge in probability to zero. Thus $R'$ converges in distribution to a normal distribution. By Cramér Wold device [e.g., Ash and Doléans-Dade, 2000, page 343] we have:

$$
\left( \frac{\sum_k \tilde{W}_{J_0,k}^2}{\sqrt{M_{J_0}}} , \ldots , \frac{\sum_k \tilde{W}_{J,k}^2}{\sqrt{M_{J}}} , \frac{\sum_k \tilde{W}_{J_0,k} \tilde{W}_{J_0,k-1}}{\sqrt{M_{J_0}}} , \ldots , \frac{\sum_k \tilde{W}_{J,k} \tilde{W}_{J,k-1}}{\sqrt{M_{J}}} \right)
$$

is jointly asymptotically normal as $M_{J_0} \to \infty$. \qed


