WAVELET TECHNIQUES IN TIME SERIES ANALYSIS
WITH AN APPLICATION TO SPACE PHYSICS

by

Agnieszka Jach

A dissertation submitted in partial fulfillment
of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
Mathematical Sciences

UTAH STATE UNIVERSITY
Logan, Utah
2006
ABSTRACT

Wavelet Techniques in Time Series Analysis
with an Application to Space Physics

by

Agnieszka Jach, Doctor of Philosophy
Utah State University, 2006

Major Professor: Dr. Piotr Kokoszka
Department: Mathematics and Statistics

Several wavelet techniques in the analysis of time series are developed and applied to real data sets.

Methods for long-memory models include wavelet-based confidence intervals for the self-similarity parameter in potentially heavy-tailed observations. Empirical coverage probabilities are used to assess the procedures by applying them to Linear Fractional Stable Motion with many choices of parameters. Asymptotic confidence intervals provide empirical coverage often much lower than nominal and it is recommended to use subsampling confidence intervals. A procedure for monitoring the constancy of the self-similarity parameter is proposed and applied to Ethernet data sets.

A test to distinguish a weakly dependent time series with a trend component, from a long-memory process, possibly with a trend, is proposed. The test uses a generalized likelihood ratio statistic based on wavelet domain likelihoods. The test is robust to trends that are piecewise polynomials. The empirical size and power are good and do not depend on specific choices of wavelet functions and models for the wavelet coefficients. The test is applied to annual minima of the water levels of the Nile River and confirms the presence of long-range dependence in this time series.
A wavelet-based method of computing an index of geomagnetic storm activity is put forward. The new index can be computed automatically using statistical procedures and does not require operator's intervention in selecting quiet days and removal of the secular component by polynomial fitting. This one-minute index is designed to facilitate the study of the fine structure of geomagnetic storm events and requires only the most recent magnetogram records, e.g., the two months including the storm event of interest. It can thus be computed over a moving window as soon as new magnetogram records become available. Averaged over one-hour periods, it is practically indistinguishable from the traditional Dst index.
ACKNOWLEDGMENTS

I would like to thank my advisor, Piotr Kokoszka, for the guidance over the last four years, for showing me a down-to-earth approach to research and life, and for his sense of humor.

I would also like to thank my committee members, Joe Koebbe, Jürgen Symanzik, Stephen Yeung, and Lie Zhu, for valuable interactions during my program.

Many thanks to my family back in Poland for long-distance support.

This research was partially supported by NATO grant PST.EAP.CLG 980599 and NSF grants INT-0223262 and DMS-0413653.

Agnieszka Jach
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>viii</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>WAVELET-BASED CONFIDENCE INTERVALS FOR THE SELF-SIMILARITY PARAMETER</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Estimation of the Self-Similarity Parameter of LFSM</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Confidence Intervals for the Parameter $H$</td>
<td>8</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Asymptotic approach</td>
<td>9</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Resampling approach</td>
<td>10</td>
</tr>
<tr>
<td>2.4</td>
<td>Simulation Study</td>
<td>12</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Generation of LFSM and estimation of $H$</td>
<td>13</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Asymptotic confidence intervals for $H$</td>
<td>15</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Resampling confidence intervals</td>
<td>16</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Summary and conclusions</td>
<td>19</td>
</tr>
<tr>
<td>2.5</td>
<td>Application to Ethernet Traffic</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>WAVELET DOMAIN TEST FOR LONG-RANGE DEPENDENCE IN THE PRESENCE OF A TREND</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>33</td>
</tr>
<tr>
<td>3.2</td>
<td>Discrete Wavelet Transform of the ARFIMA Process</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>The Test Procedure</td>
<td>39</td>
</tr>
<tr>
<td>3.4</td>
<td>Extension to AR(1) Model for Wavelet Coefficients</td>
<td>40</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulation Study</td>
<td>42</td>
</tr>
<tr>
<td>3.6</td>
<td>Application to the Annual Minima of the Nile</td>
<td>44</td>
</tr>
<tr>
<td>3.7</td>
<td>Summary and Discussion</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>WAVELET-BASED INDEX OF MAGNETIC STORM ACTIVITY</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>MODWT of the Magnetometer Data</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>Noise Removal via Wavelet Thresholding</td>
<td>59</td>
</tr>
<tr>
<td>4.4</td>
<td>Removal of the Periodic Component</td>
<td>61</td>
</tr>
<tr>
<td>4.5</td>
<td>Index Algorithm</td>
<td>63</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparison to Dst</td>
<td>67</td>
</tr>
<tr>
<td>4.7</td>
<td>Conclusions</td>
<td>70</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Methods of constructing wavelet-based confidence intervals for the self-similarity parameter $H$.</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>P-values for the GLR test applied to $N = 512$ annual Nile minima.</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>Empirical size of the GLR test based on $R = 1000$ replications of $X_t = Y_t + m_t$ of length $N = 512$, where ${Y_t}$ follows AR(1) model with parameters $\phi = 0.5820$ and $\sigma^2 = 4333$ and $m_t$ is given by (3.13).</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>Empirical size of the GLR test based on $R = 1000$ replications of $X_t = Y_t + m_t$ of length $N = 512$, where ${Y_t}$ follows AR(1) model with parameters $\phi = 0.5628$ and $\sigma^2 = 4394$ and $m_t$ is given by (3.14).</td>
<td>47</td>
</tr>
<tr>
<td>3.4</td>
<td>Empirical power of the GLR test based on $R = 1000$ replications of ARFIMA$(1,\delta,0)$ process of length $N = 512$ with parameters $\phi = 0.0660$, $\delta = 0.4013$ and $\sigma^2 = 0.0028$.</td>
<td>47</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The 95% confidence intervals (2.8) for the bias of the asymptotic estimator as a function of $j_{min}$ based on realizations of length $N = 2^{13}$. Estimators $\hat{H}$ were calculated using periodic (●) and reflection (○) boundary rules, and excluding boundary coefficients (plain bars). All estimators were calculated using scales $2^{j_{min}}, \ldots, 2^{10}$.</td>
</tr>
<tr>
<td>2.2</td>
<td>Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{13}$. Estimators were calculated using indicated range of scales.</td>
</tr>
<tr>
<td>2.3</td>
<td>Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{12}$. Estimators were calculated using indicated range of scales.</td>
</tr>
<tr>
<td>2.4</td>
<td>Example of the 95% confidence intervals (B, AP, SN, BB) based on $R = 50$ series of length $N = 2^{13}$ with subtracted $H$; scales $2^{3}, \ldots, 2^{9}$.</td>
</tr>
<tr>
<td>2.5</td>
<td>The time series of bytes per 12 milliseconds based on “pAug” trace.</td>
</tr>
<tr>
<td>2.6</td>
<td>The time series of bytes per 10 milliseconds based on “pOct” trace.</td>
</tr>
<tr>
<td>2.7</td>
<td>The time series of bytes per 1000 milliseconds based on “OctExt” trace.</td>
</tr>
<tr>
<td>2.8</td>
<td>The time series of bytes per 1000 milliseconds based on “OctExt4” trace.</td>
</tr>
<tr>
<td>2.9</td>
<td>The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 26 adjacent blocks obtained from “pAug” time series; scales: $2^{3}, \ldots, 2^{9}$.</td>
</tr>
<tr>
<td>2.10</td>
<td>The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 17 adjacent blocks obtained from “pOct” time series; scales: $2^{3}, \ldots, 2^{9}$.</td>
</tr>
<tr>
<td>2.11</td>
<td>The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 12 adjacent blocks obtained from “OctExt” time series; scales: $2^{3}, \ldots, 2^{9}$.</td>
</tr>
<tr>
<td>2.12</td>
<td>The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 7 adjacent blocks obtained from “OctExt4” time series; scales: $2^{3}, \ldots, 2^{9}$.</td>
</tr>
</tbody>
</table>
3.1 Top: Nile River annual minimum water levels for years 773 to 1284 with fitted second order polynomial trend and piecewise regression line. Middle: Simulated realization of \( X_t = m_t + Y_t \), where \( m_t \) is the fitted polynomial and \( Y_t \) follows an AR(1) model fitted to the residuals. Bottom: Simulated realization of \( X_t = m_t + Y_t \), where \( m_t \) is the fitted piecewise regression line and \( Y_t \) follows an AR(1) model fitted to the residuals.

3.2 Empirical size of the GLR test based on \( R = 500 \) replications of \( X_t = Y_t + m_t \) of length \( N = 512 \), where \( \{Y_t\} \) follows AR(1) model with given \( \phi \) and \( m_t = 0.25t^2/N \). Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.

3.3 Empirical size of the GLR test based on \( R = 500 \) replications of \( X_t = Y_t + m_t \) of length \( N = 1024 \), where \( \{Y_t\} \) follows AR(1) model with given \( \phi \) and \( m_t = 0.25t^2/N \). Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.

3.4 Empirical power of the GLR test based on \( R = 500 \) replications of \( X_t \) of length \( N = 512 \), where \( \{X_t\} \) follows ARFIMA\((0,\delta,0)\) model with given \( \delta \). Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.

3.5 Empirical power of the GLR test based on \( R = 500 \) replications of \( X_t \) of length \( N = 1024 \), where \( \{X_t\} \) follows ARFIMA\((0,\delta,0)\) model with given \( \delta \). Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.
4.1 Squared gain functions $\tilde{H}_{j,L}(f)$ for LA(8) filter, levels $j = 8, 9, 10$. Nominal pass-bands are indicated by solid vertical lines. Periodic frequencies $1/1440$, $2/1440$, $3/1440$ are marked by dashed, dotted and dashed-dotted lines, respectively. ................................. 71

4.2 Details from levels $j = 8, 9, 10$ for station Kakioka, March-April, 2001. Bar in the bottom left-hand side corner of the bottom panel spans one day. ... 71

4.3 Hourly values of WISA and Dst, January-December, 2001. .................. 72

4.4 Differences between Dst (+18 nT) and WISA, January-December, 2001. 72

4.5 Histograms of differences between Dst (+18 nT) and WISA for “quiet” and “disturbed” periods, January-December, 2001. ................................. 73

4.6 Dst, WISA and 95% confidence intervals for hourly WISA for the first week of 2001. ................................................................. 73

4.7 Hourly values of WISA and Dst, March-April, 2001. ......................... 74

4.8 Differences between Dst (+31 nT) and WISA, March-April, 2001. .... 74

4.9 Histograms of differences between Dst (+31 nT) and WISA for “quiet” and “disturbed” periods, March-April, 2001. ..................................... 75

4.10 Dst, WISA and 95% confidence intervals for hourly WISA for the first week of March. ......................................................... 75
CHAPTER 1
INTRODUCTION

Wavelet-based methods have been widely used in time series analysis for the last twenty years or so, in areas such as seismology ([34]), image and signal processing ([32]), and more recently in computer network traffic ([43]) and space physics ([53]). Popularity of wavelet-related techniques stems from the nature and statistical properties of the wavelet coefficients such as decorrelation, scaling, multi-resolution decomposition, among others.

This dissertation consists of three main parts. Each of them represents an individual manuscript that illustrates the usefulness of wavelet methods in time series analysis. Chapters 1 and 2 are concerned with self-similar or/and long-range dependent processes (LRD), whilst Chapter 3 introduces wavelet-based application to space physics.

Decorrelation and scaling properties of the discrete wavelet transform (DWT) serve as a basis in point estimation of long-range dependent (long-memory) or/and self-similar time series. Long-memory and self-similar time series have become an established modeling tool in many areas of science and technology, including geosciences ([42]), medical sciences ([39]), telecommunication networks ([38]) and to some extend financial econometrics ([22]). Intuitively, self-similar phenomena display structural similarities that encompass a wide range of time scales. In the network traffic literature ([36]), the term “burstiness” is often used in this context. Roughly speaking, self-similarity means that the pattern of traffic in a 10-minute interval looks the same as the pattern in a 1-minute subinterval, which in turn looks the same as the traffic in its 6-second subinterval, provided that the “size” of the traffic is appropriately rescaled. The key parameter describing a stochastic process used to model such phenomena is the self-similarity (Hurst) parameter $H$. Long-memory in a stationary time series occurs when data are found to be strongly dependent over large time lags, more precisely, when the covariances tend to zero in a power-law like manner and their sum diverges. The relationship between long-range dependence and self-similarity is very close and the two terms are used interchangeably. For example, if we consider a self-similar process $X(t)$, $H > 1/2$, with stationary increments (e.g., aggregated traffic data), then the
increment process \( Y(t) = X(t) - X(t-1) \) \((Y(t)\) corresponds to the total traffic volume reported between times \( t - 1 \) and \( t \)) is LRD, and vice versa.

An example of a self-similar time series that we study in Chapter 2 is a linear fractional stable motion (LFSM). It is often regarded as the simplest extension of the fractional Brownian motion to the infinite variance case. Heavy-tailed LFSM with stable increments is a good model for Internet traffic, because the latter exhibits lots of burstiness. Heavy-tailed distribution can yield a very wide range of different values, including “very large” values with nonnegligible probabilities. In addition, for some values of the Hurst parameter, the LFSM has strongly dependent increments, and thus it can be used to model long-memory observed in the network traffic.

Estimation of the self-similarity parameter \( H \) is of great importance in modeling the physical phenomena and the wavelet decomposition facilitates the statistical inference. DWT allows to reduce LRD in the time domain to weak-dependence in the wavelet domain, whilst replacing nonstationarity by stationarity. At the same time, due to scaling, wavelet coefficients reproduce the self-similarity of the process. As a result, it is feasible to use a standard technique such as linear regression to construct a wavelet-based estimator of the self-similarity parameter. This estimator is asymptotically unbiased and normal, however the asymptotic variance depends in a complex way on unknown parameters. Although useful approximations to the asymptotic variance have been established, they are based on the assumption of the Gaussianity, which in practice is not always justifiable. Therefore, in Chapter 2, various resampling methods (bootstrap, block bootstrap, subsampling) of constructing confidence intervals for the self-similarity parameter are proposed. Based on Monte-Carlo simulations, the empirical coverage probabilities are computed, and assessment of the methods are made. Practical recommendation are provided allowing to apply the optimal method(s) to real traffic measurements. In addition, a procedure for monitoring the constancy of the self-similarity parameter is proposed and applied to real Ethernet data sets. We can summarize this approach in the following points: 1) the goal is to construct confidence intervals for the self-similarity parameter \( H \) of a highly correlated time series; 2)
apply the wavelet transform and work with weakly correlated wavelet coefficients instead of the original observations in the time domain; 3) a wavelet-domain estimator for $H$ is well defined, but its variance depends on unknown parameters, hence it is difficult to construct confidence intervals for $H$; 4) resample wavelet coefficients, compute many copies of the estimator of $H$ and construct resampling-based confidence intervals.

Another example of LRD time series is the fractionally integrated autoregressive moving average (ARFIMA) process derived as a natural extension of the weakly-dependent ARMA process to a LRD setting. The ARFIMA process is used as a long-memory model in Chapter 3 in a test aiming at distinguishing between a long-range dependent process, possibly with a trend, and a weakly dependent process with a trend.

It is of importance to be able to discriminate between short and long memory as it has been recently realized that practically all statistical procedures intended to detect and estimate long memory give spurious results if a time series without long memory is perturbed by nonstationarities, like trends or structural breaks (change-points). The wavelet-based test that we propose in Chapter 3 relies on yet another feature of the DWT, i.e., its invariance to the polynomial trends of particular orders and certain piece-wise continuous functions. The idea of the test is as follows: 1) select a parametric model for which spectral densities under the null and the alternative hypotheses must be specified in such a way that weak dependence corresponds to a fixed value of a memory parameter, $d = 0$, and long-range dependence to a range of values, $d > 0$; 2) under the assumptions that the DWT coefficients are approximately uncorrelated and unaffected by additive trends, write down the likelihood function based on the coefficients under both the null and alternative hypotheses, and construct the generalized likelihood ratio (GLR) statistic; 3) the asymptotic distribution of the GLR statistic is known and thus it is feasible to either reject or fail to reject the null hypothesis. The empirical size and power of the test analyzed by means of simulations are good, but show that the test is somewhat conservative. The size and the power do not depend on specific choices of wavelet functions and models for the wavelet coefficients. We assume two reasonable representations for the wavelet coefficients, white
noise and autoregressive of the first order. The test is applied to annual minima of the water levels of the Nile River and confirms the presence of long-range dependence in this time series.

An application of wavelet techniques to space physics is described in Chapter 4. In the task of constructing an index of magnetic storm activity we do not assume any specific parametric model for the underlying time series. The need of creating an alternative index to the traditionally used Dst index is justified by the main difficulty in producing the latter, i.e., the removal of the daily, so-called Sq, variation, which requires subjective human input ([37], [44], [49]). Ground-based magnetogram records from the low-latitude magnetic observatories are used to isolate the signatures of the ring current that becomes prominent during magnetic storms. These records are multi-scaled, impulsive, and asynchronous with non-stationary frequency spectra. Because of the nature of the underlying time series, multiresolution decomposition based on wavelets is well suited for the task of creating an automated procedure for and index of magnetic storm activity. It allows to decompose the signal (observed magnetogram record) into the high- and low-frequency components, which are localized in time. Thresholding of the high-frequency components removes that part of the signal that cannot be attributable to the ring current activity. Modification of the time scales representing the periodic (daily and half-daily) oscillations in magneotogram record allows to estimate and consequently extract the Sq variation. The resulting wavelet-based procedure is fully automated. A comparison of the traditional and wavelet-based indices of magnetic storm activity for the year 2001 is presented and shows that the two indices are practically indistinguishable.
CHAPTER 2
WAVELET-BASED CONFIDENCE INTERVALS FOR THE
SELF-SIMILARITY PARAMETER\(^1\)

2.1 Introduction

Empirical studies have shown ([30], [36], [38]) that many network traffic traces exhibit\(^1\) self-similarity or are long-range dependent (LRD). It is of importance in modeling the flow of information through a network to estimate the self-similarity parameter. Significant contribution in this direction have been made by several authors, see, e.g., [2], [3] for a review and references. The existing theoretical and numerical studies are chiefly concerned with point estimation of the self-similarity parameter. As discussed later in this paper, it is known that the appropriate estimators are asymptotically unbiased and normal, but the asymptotic variance depends in a very complex way on unknown parameters. Therefore, useful approximations to the asymptotic variance have been derived, but these are based on the assumption of the Gaussianity of observations. As will be seen in real data examples discussed in Section 2.5, the latter assumption is not always justified and often the data exhibit heavy tails characterized by the tail index \(\alpha\). Suitable approximations to the asymptotic variance have also been derived in the case of heavy-tailed observations. One should hope that these approximations would yield useful confidence intervals for the self-similarity parameter, but our research has shown that this is typically not the case, even if the tail index \(\alpha\) is correctly specified. This index is very difficult to estimate in practice, and a misspecification of \(\alpha\) worsens the empirical coverage probability of the asymptotic confidence intervals. Boundary effects are an important source of bias in finite samples. However, even after removing the boundary wavelet coefficients, it is not unusual for an asymptotic confidence interval with nominal coverage probability of 95% to have an empirical coverage probability lower than 75%. A first-order bias corrected estimator has been derived in the case of Gaussian observations by [50] but this modification relies on some very specific properties of the normal distribution and is not directly applicable to heavy-tailed

\(^1\)Coauthored by A. Jach and P. Kokoszka.
In this paper we propose a number of alternative methods of constructing confidence intervals for the self-similarity parameter. We compare five different methods and provide practical recommendations. We also propose a procedure for monitoring the constancy of the self-similarity parameter and apply it to Ethernet traffic measurements.

In our simulation study we focus on self-similar (motion-type) processes and use the Linear Fractional Stable Motion as an archetype. Recall that a stochastic process \( X = \{X_t\}_{t \in \mathbb{R}} \) is self-similar with self-similarity parameter \( H > 0 \) if for any positive constant \( c \)

\[
\{X_{ct}\}_{t \in \mathbb{R}} =_d \{c^H X_t\}_{t \in \mathbb{R}},
\]

where \( =_d \) denotes the equality of the finite-dimensional distributions. We say that the process \( X = \{X_t\}_{t \in \mathbb{R}} \) is a Linear Fractional Stable Motion (LFSM) if \( X_t \) is defined by the integral

\[
X_t = \int_{\mathbb{R}} ((t-s)_+)^d - ((-s)_+)^d |M_\alpha(ds)|,
\]

where \( \alpha \in (0, 2] \), and \( d = H - 1/\alpha \), for some \( H \in (0, 1) \), \( H \neq 1/\alpha \), \( M_\alpha \) is an \( \alpha \)-stable symmetric random measure on \( \mathbb{R} \) with Lebesgue control measure, see Section 3.3 of [45], \( x_+ = \max\{x, 0\} \). This process has stationary increments and is self-similar with parameter \( H \). When \( \alpha = 2 \), i.e., when the process is Gaussian, and when \( H > 1/2 \) the increments of \( X \) are LRD. In the case of \( \alpha < 2 \), the process defined by (2.1) has heavy tails and even though there is no universally agreed upon definition of long-range dependence when the variance is infinite, the increment process of \( X \), called a Linear Fractional Stable Noise, is said to be LRD if \( H > 1/\alpha \). For further details about LFSM and heavy-tailed self-similar and LRD processes, see Chapter 7 of [45]. The LFSM has been used for modeling large network traffic fluctuations ([25], [55]).

The objective of this paper is to develop and compare several wavelet-based methods of constructing confidence intervals for the parameter \( H \) in the LFSM. Wavelets are known to be an efficient tool in the context of long-range dependence and self-similarity, see [3], [4], Chapter 9 of [41], [47], among others. When a discrete realization of a process is
available, Mallat’s algorithm can be used to produce the set of discrete wavelet transform (DWT) coefficients making these procedures applicable to very long time series obtained, for example, from network traffic measurements.

Two features of the DWT coefficients of a long-range dependent process, decorrelation and scaling, are utilized in the estimation of $H$ and the construction of confidence intervals through the asymptotic and resampling approaches. We assess the procedures under consideration using empirical coverage probabilities. Our overall assessment of the relative performance of the methods enables us to provide useful guidance for their practical application.

The paper is organized as follows: We first describe in Section 2.2 a wavelet-based estimator of the self-similarity parameter of the LFSM. This estimator forms the basis for the construction of the confidence intervals. In Section 2.3, we introduce several methods of constructing confidence intervals. We then give in Section 2.4 a detailed description of the simulation study and discuss the results. In Section 2.5, an application of our techniques to Ethernet traces is presented.

2.2 Estimation of the Self-Similarity Parameter of LFSM

In this section we present a wavelet-based method of estimating the self-similarity parameter $H$ of a LFSM, which serves as a cornerstone for confidence interval construction. This method was derived in [1], see also [3].

The wavelet transform coefficients of a continuous time process $X = \{X_t\}_{t \in \mathbb{R}}$ is a collection of the quantities

\begin{equation}
    d_{j,k} = \int_{\mathbb{R}} X_t \psi_{j,k}(t) dt, \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z},
\end{equation}

where $\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k), \ t \in \mathbb{R}$. The function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, called the mother wavelet, satisfies some regularity assumptions (see [15] for the details). In this setting, $2^j$ and $k$ are called scale and location, respectively. In our paper, we work Discrete Wavelet Transform (DWT) coefficients which are obtained as the output of Mallat’s algorithm ap-
plied to discrete time observations $X_0, X_1, \ldots, X_{N-1}$. This scheme produces a set of DWT coefficients $\tilde{d}_{j,k}$; $N_j$ of them are available at scale $n_j = 2^j$, $j = 1, 2, \ldots, J$, where $N_j = 2^{J-j}$, $J = \lfloor \log_2(N) \rfloor$ and $\lfloor \cdot \rfloor$ denotes the integer part. The coefficients $\tilde{d}_{j,k}$ can be viewed as approximations to the $d_{j,k}$, see Chapter 11 of [41].

The regression-based approach for estimating the self-similarity parameter $H$ in LFSM involves the statistic

$$Y(N_j) = \frac{1}{N_j} \sum_{k=0}^{N_j-1} \log_2 |\tilde{D}_{j,k}|,$$

where

$$\tilde{D}_{j,k} = 2^{-j/2} \tilde{d}_{j,k}.$$

The generalized least squares regression estimator for $H$ is defined by

$$\hat{H} = \sum_{j=j_{\min}}^{j_{\max}} w_j Y(N_j),$$

where $j_{\min}$ and $j_{\max}$ dictate a range of scales upon which the estimator is constructed. [47] refer to (2.5) as the “log” estimator. In most applications, due to bias-variance trade-off, the number of scales used in (2.5) is smaller than the total number of available scales (see Section 2.4). The weights $w_j$ in equation (2.5) satisfy the following conditions

$$\sum_{j=j_{\min}}^{j_{\max}} w_j = 0, \quad \sum_{j=j_{\min}}^{j_{\max}} j w_j = 1.$$

The specific choice of the $w_j$ is discussed in Section 2.4.1.

Detailed explanation of the estimation technique along with the asymptotic results are given in [47].

2.3 Confidence Intervals for the Parameter $H$

In this section we present several methods of constructing confidence intervals for the self-similarity parameter $H$ in LFSM. These methods can be grouped into two broad categories: methods based on an asymptotic approximation, and resampling methods. We describe them in Sections 2.3.1 and 2.3.2, respectively.
2.3.1 Asymptotic approach

The decorrelation property of wavelet coefficients allows us to approximate the variance of the self-similarity estimator $\hat{H}$. [4] derived the following approximate formula

$$\text{Var}\{\hat{H}\} = \sigma^2(\alpha) \approx \frac{\log_2(e) \pi^2}{12} \left(1 + \frac{2}{\alpha^2}\right) \left(\sum_{j = j_{\text{min}}}^{j_{\text{max}}} \frac{w_j^2}{N_j}\right),$$

where the summation extends over the octaves used to construct the estimator $\hat{H}$. Consequently, the $100(1 - \beta)\%$ asymptotic confidence interval for $H$ is defined as

$$\left(\hat{H} + q_Z(\beta/2)\sigma(\alpha), \hat{H} + q_Z(1 - \beta/2)\sigma(\alpha)\right),$$

where $q_Z(\beta)$ denotes the $\beta$-th quantile of the standard normal distribution. Note that the definition of the asymptotic confidence intervals for $H$ requires the knowledge of the stability index $\alpha$ and thus we assume that this parameter is known. We comment on this issue when we discuss our conclusions and recommendations in Section 2.4.4.

Our study of the estimator $\hat{H}$ showed a significant effect of the boundary wavelet coefficients on the bias of this estimator (see Section 2.4.2). To study this bias, we define the $100(1 - \beta)\%$ confidence interval for the bias as

$$\left(\overline{\hat{H}} - H\right) + q_Z(\beta/2)s/\sqrt{R}, \left(\overline{\hat{H}} - H\right) + q_Z(1 - \beta/2)s/\sqrt{R},$$

where $\overline{\hat{H}}$ and $s$ are, respectively, the sample mean and the sample standard deviation of $R$ independent replications of $\hat{H}$.

Recall that if a wavelet filter of length $L$ is used, the first $K_j = \min([\frac{(L-2)(1-2^{-j})}{2}], N_j)$ wavelet coefficients at octave $j$ are affected by circularly filtering the data and the remaining $M_j = N_j - K_j$ coefficients,

$$\tilde{d}_{j,k} \equiv \tilde{d}_{j,K_j + k}, \quad j = 1, 2, \ldots, J, \quad k = 0, 1, \ldots, M_j - 1,$$

called the non-boundary wavelet coefficients, are not influenced (for more details see Comments and Extensions to Section 4.11 of [41]). The corresponding estimator of $H$ based on the non-boundary wavelet coefficients is therefore defined as

$$\hat{H}^{\text{nb}} = \sum_{j = j_{\text{min}}}^{j_{\text{max}}} w_j Y(M_j),$$
where

\( Y(M_j) = \frac{1}{M_j} \sum_{k=0}^{M_j-1} \log_2 |\tilde{D}_{j,k}| \)  

and

\( \tilde{D}_{j,k} = 2^{-j/2} \tilde{d}_{j,k} \).

Consequently, the 100(1 - \beta)\% asymptotic confidence interval using the non-boundary wavelet coefficients is defined as

\[
(\hat{H}^{nb} + q_Z(\beta/2)\sigma(\alpha), \hat{H}^{nb} + q_Z(1 - \beta/2)\sigma(\alpha)).
\]

The use of the non-boundary wavelet coefficients \( \tilde{d}_{j,k} \) combined with the appropriate selection of \( j_{\min} \) and \( j_{\max} \) provide confidence intervals for bias which contain 0. Such confidence intervals can be calculated using formula (2.8) with \( \hat{H} \) replaced by \( \hat{H}^{nb} \) (see Figure 2.1).

2.3.2 Resampling approach

In this section we focus on the resampling methods of constructing confidence intervals. The resampling procedures are applied to DWT coefficients within given scale. These methods are heuristically justified by the approximate decorrelation property of these coefficients.

2.3.2.1 Subsampling confidence intervals with non-overlapping blocks

Consider the coefficients \( \tilde{D}_{j,k} \) defined by (2.12) at some fixed scales \( n_1 = 2^{j_{\min}}, n_2 = 2^{j_{\min}+1}, \ldots, n_m = 2^{j_{\max}}, m = j_{\max} - j_{\min} + 1 \), and \( k = 0, 1, \ldots, M_j - 1 \) at scale \( n_j \). For each octave \( j, j = j_{\min}, j_{\min} + 1, \ldots, j_{\max}, \) we choose a number of blocks \( B_j \) of lengths \( l_0 = l_1 = \ldots = l_{B_j-2} = \lfloor M_j/B_j \rfloor, l_{B_j-1} = \lfloor M_j/B_j \rfloor + (M_j \mod B_j) \) at scale \( n_j \). In other words, if \( M_j \) is not divisible by \( B_j \), the first \( B_j - 1 \) blocks are of the same length, \( l_0 \), and the last one is longer by \( (M_j \mod B_j) \). At each octave \( j \), we consider the following blocks of the rescaled non-boundary wavelet coefficients

\[
\{\tilde{D}_{j,0}, \ldots, \tilde{D}_{j,l_0-1}\}, \quad \{\tilde{D}_{j,l_0}, \ldots, \tilde{D}_{j,l_0+l_1-1}\}, \quad \ldots, \quad \{\tilde{D}_{j,\sum_{i=0}^{B_j-2} l_i}, \ldots, \tilde{D}_{j,M_j-1}\}.
\]
Based on each of these blocks, we compute

$$Y(M_j^m) = \frac{1}{B_j} \sum_{r=0}^{m-1} \sum_{k=0}^{l_r-1} \log_2 |\tilde{D}_{j,k}|, \quad m = 0, 1, \ldots, B_j - 1.$$  

The index $m$ indicates the position of the block, its range depends on the octave $j$. Next we select $J_{\text{max}} - J_{\text{min}} + 1$ blocks, one on each scale, and compute the estimator

$$\hat{H}_{SN} = \hat{H}_{SN}(m(J_{\text{min}}), \ldots, m(J_{\text{max}})) = \sum_{j=J_{\text{min}}}^{J_{\text{max}}} w_j Y(M_j^{m(j)}),$$  

where $m(j)$ is the index of the block selected on scale $j$. There are $S = \prod_{j=J_{\text{min}}}^{J_{\text{max}}} B_j$ estimators (2.14) based on one realization of LFSM. Consequently, the $100(1 - \beta)\%$ subsampling confidence interval for $H$ is

$$\left(q_s(\beta/2), q_s(1 - \beta/2)\right),$$  

where $q_s(\beta)$ denotes the $\beta$-th empirical quantile of the empirical distribution of the $S$ estimators $\hat{H}_{SN}$. The subscript $SN$ in (2.14) stands for “Subsampling with Non-overlapping blocks”.

### 2.3.2.2 Naive block bootstrap and bootstrap confidence intervals

We first describe the naive block bootstrap method subsequently called for brevity block bootstrap. We use the notation introduced in Section 2.3.2.1. For each octave $j$ we choose a block length $l_j$. It follows that there are $M_j - l_j + 1$ (overlapping) blocks of the form

$$\{\tilde{D}_{j,0}, \ldots, \tilde{D}_{j,l_j-1}\}, \quad \{\tilde{D}_{j,1}, \ldots, \tilde{D}_{j,l_j}\}, \quad \ldots, \quad \{\tilde{D}_{j,M_j-l_j}, \ldots, \tilde{D}_{j,M_j-1}\}.$$  

Next, by drawing with replacement from the set $\{0, 1, \ldots, M_j - l_j\}$, we select $B_j = \left[\frac{M_j}{l_j}\right]$ integers and denote them by $\{s_0, s_1, \ldots, s_{B_j-1}\}$. Based on this selection, we construct the bootstrap series of coefficients

$$\{\tilde{D}_{j,s_0}, \ldots, \tilde{D}_{j,s_0+l_j-1}\}, \quad \{\tilde{D}_{j,s_1}, \ldots, \tilde{D}_{j,s_1+l_j-1}\}, \quad \ldots, \quad \{\tilde{D}_{j,s_{B_j-1}}, \ldots, \tilde{D}_{j,s_{B_j-1}+l_j-1}\}.$$  

If $(M_j \mod l_j) \neq 0$ the length of this series exceeds the number of the non-boundary coefficients at scale $j$, therefore we truncate it to obtain a series of length $M_j$, i.e., of the
same length as the original series of the non-boundary coefficients at scale $2^j$. Denoting these coefficients by $\tilde{D}_{j,k}$ we can compute the estimator

$$\hat{H}_{BB}^* = \sum_{j=j_{\text{min}}}^{j_{\text{max}}} w_j Y^*(M_j),$$

where

$$Y^*(M_j) = \frac{1}{M_j} \sum_{k=0}^{M_j-1} \log_2 |\tilde{D}_{j,k}^*|.$$

If we draw one resample at each octave $j$, we obtain one estimator (2.16). Resampling $B$ times (one resample on each scale) yields $B$ block bootstrap estimators $\hat{H}_{BB}^*$. Consequently, the $100(1 - \beta)\%$ block bootstrap confidence interval for $H$ is

$$\left( q_{B}(\beta/2), q_{B}(1 - \beta/2) \right),$$

where $q_{B}(\beta)$ denotes the $\beta$-th empirical quantile of the empirical distribution of the $B$ estimators $\hat{H}_{BB}^*$.

Note that when the blocks of length 1 are used, the block bootstrap method is equivalent to the DWT-based bootstrap method described in [40]. We refer to confidence intervals obtained in this way as bootstrap confidence intervals even though it is a special case of the naive block bootstrap procedure.

For ease of reference, we list in Table 2.1 the methods investigated in this paper and their abbreviations as well as plotting characters used for marking the empirical coverage probabilities in Section 2.4. In case of subsampling method and block bootstrap we tried several choices of the blocks and thus we have more than one plotting character (see Sections 2.4.3.1-2.4.3.2 for more details).

### 2.4 Simulation Study

This section contains the main results of the paper. In Section 2.4.1 we describe in detail the procedures for generating realizations of LFSM and for estimating $H$. Sections 2.4.2 and 2.4.3 focus on the performance of the asymptotic and resampling methods, respectively. Conclusions are discussed in Section 2.4.4. We present only the results that best illustrate
Table 2.1. Methods of constructing wavelet-based confidence intervals for the self-similarity parameter $H$.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>ABBR.</th>
<th>CHARACTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic (correct $\alpha$)</td>
<td>AP</td>
<td>×</td>
</tr>
<tr>
<td>Asymptotic ($\alpha = 1.5$)</td>
<td>A</td>
<td>◦</td>
</tr>
<tr>
<td>Subsampling with Non-overlapping Blocks</td>
<td>SN</td>
<td>◦, ▽</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>B</td>
<td>△</td>
</tr>
<tr>
<td>Block Bootstrap</td>
<td>BB</td>
<td>+, ●, ★</td>
</tr>
</tbody>
</table>

our findings. Tables with numerical values of the empirical coverage probabilities and a large number of additional graphs are presented in an extended version of this paper available from the authors.

2.4.1 Generation of LFSM and estimation of $H$

Consider a discrete realization of LFSM $X_0, X_1, \ldots, X_N$, ($X_0 = 0$). The random variables $X_t$, $t = 1, 2, \ldots, N$ can be expressed as $X_t = \sum_{k=1}^t U_k$, where

$$U_k = \int_{\mathcal{R}} [((k - s)_+)^d - ((k - 1 - s)_+)^d] M_\alpha(ds).$$

As in [45], Section 7.11, the random variable $U_k$ is approximated as

$$U_k \approx \sum_{i=1, i \neq m}^{M/m} [((i/m)_+)^d - ((i/m - 1)_+)^d] m^{-1/\alpha} Z_\alpha(i - mk),$$

for some choice of discretization parameters $m$ and $M$ which correspond to the mesh size and the truncation level in integral (2.19), respectively. It is possible to implement an algorithm for calculating the last summation directly (see Section 7.11 of [45]) or indirectly via the Fast Fourier Transform (see [48]). However since our goal is to compare methods of constructing confidence intervals, the computational issues concerning generating paths of LFSM are beyond the scope of this paper. We refer to [48] for more details. In our study we used the indirect implementation with discretization parameters $m = 64$ and $M = 2^{14}$. According to [48], the choice of $m = 64$ when $\alpha = 1.5$ provides good results for several values of $M = 60, 600, 6000$, and for various choices of $H$. In our study $M$ equals the length of the generated series.
We used nine pairs \((\alpha, H)\) of the parameters obtained as combination of the following sets

\[
\alpha \in \{1.2, 1.6, 2.0\} \quad \text{and} \quad H \in \{0.6, 0.75, 0.9\}.
\]

For each pair, we generated \(R = 300\) independent realizations of LFSM of length \(2^{14}\) via the indirect method and truncated them to our target length \(N = 10000\) for further computations. The generated series were originally of length \(2^{14}\) and not 10000, in order to utilize the Fast Fourier transform. Same choice for \(M\), i.e., \(M = 2^{14}\) combined with \(m = 64\) yielded \(m\) and \(M + \text{length of generated series}\) as powers of 2.

To compute wavelet-based estimators of the self-similarity parameter \(H\) and to construct confidence intervals, we computed discrete wavelet transform coefficients using Mallat’s algorithm with a wavelet filter of width \(L = 6\), corresponding to Daubechies wavelet with 3 vanishing moments. The definition and the role of vanishing moments is explained in Section 11.9 of [41].

Practical selection of scales upon which the estimators are built depends on the length of the series, as well as the bias-variance trade-off ([47]). When the series length is 10000, there are at most \(J = 13\) dyadic scales available, however not all of them should be included in the estimation. Based on the analysis of the bias of the asymptotic estimator (2.8) presented in Section 2.4.2 and rationale behind the resampling methods we made the following selection. We used two sets of scales \(n_1 = 2^3, n_2 = 2^4, \ldots, n_6 = 2^8\) and \(n_1 = 2^3, n_2 = 2^4, \ldots, n_7 = 2^9\), to see which was the better option. In addition, we considered series of length \(N = 2^{12}\), to compare methods of constructing confidence intervals for \(H\) with respect to the length of the series (longer series, \(N = 2^{13}\) and shorter series, \(N = 2^{12}\)). For shorter series, we computed the estimators using two collections of scales \(n_1 = 2^3, n_2 = 2^4, \ldots, n_5 = 2^7\) and \(n_1 = 2^3, n_2 = 2^4, \ldots, n_6 = 2^8\). The weights \(w_1, w_2, \ldots, w_m\) (\(m\) is the number of octaves involved in the estimation) can be determined by an \(m \times m\) strictly positive definite matrix \(G\). The connection between the weights and the matrix \(G\) is described in [47]. We used \(G = \text{diag}\{n_1, \ldots, n_m\}\), where \(n_1 = 2^{j_{\min}}, n_2 = 2^{j_{\min}+1}, \ldots, n_m = 2^{j_{\max}}\), which is the only practically available choice of \(G\) when the autocovariances of the observations are unknown.
2.4.2 Asymptotic confidence intervals for $H$

In this section we discuss the results of simulations for the asymptotic methods.

Based on confidence intervals for the bias of the asymptotic estimator, (2.8), calculated for two boundary rules, periodic and reflection (see [40] for more details), and as a function of $j_{\text{min}} = 1, 2, \ldots, 6$, we conclude that; firstly, it is necessary to drop first few scales to avoid bias ($j_{\text{min}} = 3$ is a reasonable choice), but not too many to keep the variance low; secondly, the bias caused by the periodic rule is greater than that for the reflection rule; thirdly, the exclusion of the boundary wavelet coefficients yields the desirable bias containing 0, located between the biases introduced by the two boundary rules (see Figure 2.1 for the longer series, similar results were obtained for the shorter series).

The choice of the upper cut-off is not as critical and the behavior of the confidence intervals when all wavelet coefficients are included is very similar for three choices of $j_{\text{max}} = 10, 11, 12$ we considered. We present the comparison of the confidence intervals for bias (2.8) of the asymptotic estimator computed with and without the boundary coefficients, for $j_{\text{max}} = 10$, since for $J = 13$ and wavelet filter of length $L = 6$ this is the largest scale containing the non-boundary coefficients (for $J = 12$ the largest scale is 9). On the other hand, since the number of the non-boundary coefficients at this scale is low, it is reasonable to use $j_{\text{max}} < 10$, especially for the resampling methods.

However, even after excluding the boundary coefficients and carefully selecting $j_{\text{min}}$ and $j_{\text{max}}$, the asymptotic method does not provide satisfactory results. The empirical coverage probabilities for the two choices of $J$, 13 and 12, are very similar with values in the 40%-75% range, and increasing with $H$, for all values of $\alpha$ considered (see Figures 2.2-2.3, ×). The variability among the confidence intervals based on different replications but for the same $(\alpha, H)$ is large when $\alpha = 1.2$, and small when $\alpha = 2.0$, and overall smaller for the longer series.

A practical implementation of the asymptotic method requires the knowledge of the tail index $\alpha$. Our simulations have shown, however, that $\alpha$ need not be known precisely, as long as it is not too close to zero. The variance (2.6) as a function of $\alpha$ changes very slowly
for $\alpha > 0.5$. The empirical coverage probabilities with the variance (2.6) computed with a possibly misspecified $\alpha = 1.5$ behave very similarly to those based on the correct $\alpha$ (see Figures 2.2-2.3, $o$).

We summarize our conclusions for the asymptotic approach in the following points:

- Asymptotic intervals (2.7) are strongly influenced by the boundary coefficients.
- Asymptotic intervals (2.13) based on the non-boundary wavelet coefficients have coverage in the range 40%-75%.
- Asymptotic intervals (2.13) based on the variance (2.6) computed for incorrect $\alpha = 1.5$ provide similar coverage to those based on the correct $\alpha$.
- Asymptotic intervals (2.13) computed for different replications but fixed $(\alpha, H)$ vary a lot among each other when the series length is $2^{12}$ and $\alpha$ small.

### 2.4.3 Resampling confidence intervals

#### 2.4.3.1 Subsampling confidence intervals with non-overlapping blocks

We continue the discussion of the simulation results with the resampling methods, focusing first on the subsampling procedure with non-overlapping blocks. Choosing the lengths of the blocks in procedures of this type is always a difficult task. We tried two strategies. First, we partitioned different scales into blocks with lengths roughly the same across all scales. At the same time we aimed for the total number of estimators, $S$, which is not too high. Then, we used blocks of length equal to approximately 25% of the coefficients at the lowest scales, and to all the coefficients at the remaining scales, this time making sure that $S$ was sufficiently large to construct the percentile-type confidence intervals.

Using the first strategy, for the shorter series ($N = 2^{12}$) the number of blocks were as follows:

\begin{equation}
B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^7,
\end{equation}

\begin{equation}
B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8.
\end{equation}
For the longer series \((N = 2^{13})\), we used

\[(2.21)\]
\[B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8,\]
\[B_3 = 16, B_4 = 8, B_5 = 4, B_6 = 2, B_7 = B_8 = B_9 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^9.\]

In each case the number of estimators \((2.14)\) was \(S = 1024\).

For both lengths, and for both choices of scales, the empirical coverage probabilities are very similar and very close to 100\% (see Figures 2.2-2.3, o). The intervals are much longer than those based on the asymptotic method, especially for the shorter series, and thus not very informative. This strategy is thus not recommended.

Using the second strategy, for the shorter series \((N = 2^{12})\) the number of blocks were as follows:

\[(2.22)\]
\[B_3 = B_4 = \ldots = B_6 = 4, B_7 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^7,\]
\[B_3 = B_4 = \ldots = B_6 = 4, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^9.\]

For the longer series \((N = 2^{13})\), we used

\[(2.23)\]
\[B_3 = B_4 = \ldots = B_6 = 4, B_7 = B_8 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^8,\]
\[B_3 = B_4 = \ldots = B_6 = 4, B_7 = B_8 = B_9 = 1, \quad \text{at scales } 2^3, 2^4, \ldots, 2^9.\]

In each case the number of estimators \((2.14)\) was \(S = 256\). Such a choice of the number of blocks, not only yields shorter confidence intervals, but also provides coverage probabilities very close to nominal (see Figures 2.2-2.3, ∨).

We observe an increase in the coverage probability for the longer series with \(j_{\text{min}} = 3\) and \(j_{\text{max}} = 9\) as a function of \(H\) for all \(\alpha\), with the values approaching 95\% as \(H\) increases. This slight undercoverage can be improved by choosing \(j_{\text{min}} = 3\) and \(j_{\text{max}} = 8\). The empirical coverage is then very close to the nominal value. The empirical coverage probability for the shorter series (both sets of scales) is more stable and is close to 95\%, however intervals are longer and so less informative.

The results for the subsampling method can be summarized as follows:

- The subsampling method yields longer confidence intervals than the asymptotic method and higher coverage.

- When the subsamples are too short, coverage is close to 100\%, the intervals are too long.
• Sampling about 25% of the coefficients at the lowest scales provides coverage of 90%-95% (with a target value of 95%).

• Confidence intervals for the longer series are shorter. The choice of \( j_{\text{min}} = 3 \) and \( j_{\text{max}} = 8 \) for both lengths is optimal.

A referee of this paper suggested to study hybrid “asymptotic-subsampling” 100(1-\( \alpha \))% confidence intervals defined as

\[
(\hat{H}^{\text{nb}} + q_Z(\beta/2)s_{SN}, \hat{H}^{\text{nb}} + q_Z(1 - \beta/2)s_{SN}),
\]

where \( s_{SN} \) is the sample standard deviation based on the 256 estimators \( \hat{H}_{SN} \) with the number of blocks determined by (2.22) and (2.23). This method gives results very similar, the difference in empirical coverage probabilities is around 1-2% in either direction, to the percentile based subsampling method discussed in this section. All conclusions listed above remain valid for this modification.

2.4.3.2 Block bootstrap and bootstrap confidence intervals

In this section we focus on confidence intervals based on block bootstrap. In the special case of block size equal to 1, following [40], we refer to them as bootstrap intervals.

For comparison with the subsampling method, we used the lengths of blocks implied by (2.20)-(2.23). We also used much shorter blocks: \( l_j = 4 \) and \( l_j = 2, j = j_{\text{min}}, \ldots, j_{\text{max}}, \) for both series lengths and both sets of scales.

Using \( l_j \) dictated by (2.20)-(2.23) leads to undercoverage. For \( J = 13 \) and blocks of lengths \( l_3 = l_4 = l_5 = 63, l_6 = 62, l_7 = 60, l_8 = 28, l_9 = 12 \) and \( l_3 = l_4 = l_5 = 63, l_6 = 62, l_7 = 60, l_8 = 28, \) the coverage is in the range 60%-75%, while for the shorter series with \( l_3 = l_4 = l_5 = 31, l_6 = 30, l_7 = 28, l_8 = 12 \) and \( l_3 = l_4 = l_5 = 31, l_6 = 30, l_7 = 28, \) the coverage falls in 65%-80% interval (see Figures 2.2-2.3, +). These results can be improved dramatically if shorter blocks are used (see Figures 2.2-2.3, •). By taking \( l_j = 4 \) at all scales, the coverage for \( N = 2^{13} \) and \( N = 2^{12}, \) for both sets of scales, increases to 80%-95%.
Very similar results are obtained for \( l_j = 2 \) (see Figures 2.2-2.3, \( * \)). The bootstrap method yields similar coverage of about 75%-95%, (see Figures 2.2-2.3, \( \Delta \)). It has the advantage of not relying on a block size selection.

The results of simulations for the block bootstrap and bootstrap methods can be summarized as follows:

- Block bootstrap with short blocks and bootstrap methods generally provide good coverage which is slightly below the nominal value of 95%, often by 5% and no more than 20%.

- When longer blocks are used, the undercoverage is more serious.

2.4.4 Summary and conclusions

Based on all our experiments the following overall conclusions can be drawn. Recall that the target coverage is 95%.

(1) The asymptotic method provides quite low empirical coverage probabilities, in the range 40%-75%, even if only non-boundary wavelet coefficients are used. The use of all wavelet coefficients is not recommended as they introduce large bias, whose direction depends on the boundary rule applied. An approximate value of \( \alpha \) must be known to determine the length of the interval, however a reasonable choice of this value, for example \( \alpha = 1.5 \), leads to confidence intervals yielding similar coverage probabilities to those based on correct \( \alpha \). The asymptotic method, even with correctly specified \( \alpha \), is outperformed by all resampling methods considered.

(2) Bootstrap method gives coverage between 75% and 95%.

(3) Block bootstrap and subsampling with non-overlapping blocks yield best results (coverage of 80%-95%) for appropriate choice of the block lengths. If the blocks in the block bootstrap method are too long, the coverage is too low (60%-80%). If the blocks in the subsampling method are too short, the coverage is too high (98%-100%).
(4) Resampling confidence intervals are longer than the asymptotic, especially for the shorter series.

Subsampling with appropriate choice of blocks yields coverage probabilities typically some 1%-5% below the nominal coverage. Block bootstrap gives shorter, more informative, confidence intervals, but at the expense of lower empirical coverage, 1%-15% below the nominal coverage. The bootstrap method is an attractive alternative leading to only slightly greater undercoverage than the block bootstrap, but not requiring block size selection. These three methods provide coverage reasonably close to nominal and are much better than the asymptotic method which has been used so far. To achieve better coverage it is necessary to use longer confidence intervals than the asymptotic ones.

We conclude this section with a heuristic explanation of the conclusions stated in point (3) above. The block bootstrap method reconstructs the series of DWT coefficients at octave $j$ from blocks of these coefficients. If these blocks are too long, the reconstructed series look too much like the original series, not enough variability is introduced, the estimators of $H$ are too close to the estimator computed from the original sample, consequently, the confidence intervals based on the empirical distribution of these estimators are too short. By contrast, the subsampling method treats blocks in the same way as the original series of coefficients. These blocks are not put together, so in order for them to "imitate" the original coefficients, they must be long enough. If they are too short, too much variability is introduced, resulting in confidence intervals that are too long.

2.5 Application to Ethernet Traffic

In this section we present an application of our techniques to four Ethernet data sets. Each set contains a million packet arrival times together with the packet sizes in bytes recorded at an Ethernet link at the Bellcore Morristown Research and Engineering Facility. In these two-column data sets, the first column gives arrival time in seconds since the start of the trace and the second gives the corresponding Ethernet packet size in bytes (for more information about these traces see [30] or http://ita.ee.lbl.gov/html/contrib/BC.html). To
construct the discrete time series we used in our application, we computed the total number of bytes transmitted during consecutive time intervals of constant lengths 12, 10, 1000, and 1000 milliseconds, thus obtaining four traces "pAug", "pOct", "OctExt", and "OctExt4", respectively. More specifically, let $\delta$ denote the length of the time interval and $Z_t$ the size of a packet arriving at time $t$; the index $t$ can take any value in the interval $[t_0, T]$, where $t_0$ and $T$ are the first and the last values in the time column, respectively. The discrete time process $\{Z^{(\delta)}_n\}$ based on a given trace is obtained by putting

$$Z^{(\delta)}_n = \sum_{\{t: t-t_0 \in [(n-1)\delta, n\delta]\}} Z_t, \quad n = 1, 2, \ldots, N, \quad N = [(T - t_0)/\delta].$$

The four time series are plotted in Figures 2.5-2.8.

Veitch and Abry ([51], [52]) developed a test for the time constancy of scaling exponents in self-similar or LRD Gaussian time series and applied it to Ethernet sequences similar to the sequences $\{Z^{(\delta)}_n\}$ defined above. If the observations can be modeled as increments of a fractional Brownian motion with self-similarity parameter $H$, then the scaling exponent $\gamma$ is defined as $\gamma = 2H - 1$.

Their procedure can be summarized as follows:

1. Choose an appropriate $m > 1$ and divide the time series into $m$ adjacent blocks.

2. Use a common range of octaves $[j_1(m), j_2(m)]$ to compute the estimates $\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_m$, which can be considered as uncorrelated Gaussian variables with unknown means $\gamma_i$ and known variances $\sigma^2_i, \quad i = 1, 2, \ldots, m$.

3. For a given significance level $\beta$, reject

$$H_0: \text{ all scaling exponents are equal } (\gamma_i = \gamma^0, i = 1, 2, \ldots, m)$$

in favor of

$$H_1: \text{ some scaling exponents are different } (\gamma_i \neq \gamma_j, \text{ for some } i \text{ and } j),$$

if $V_m > C_m(\beta)$. Under the null hypothesis, the distribution of the test statistic

$$V_m = \sum_{i=1}^{m} \frac{1}{\sigma^2_i} \left( \hat{\gamma}_i - \frac{\sum \hat{\gamma}_i/\sigma^2_i}{\sum 1/\sigma^2_i} \right)^2$$


is the chi-squared distribution with \( m - 1 \) degrees of freedom.

The procedure of [51] is essentially a one-way analysis of variance and relies on the fact that the estimators \( \hat{\gamma}_i \), \( i = 1, 2, \ldots, m \), are approximately independent normal random variables with identical known variance which is approximated assuming that the observations themselves are normal.

The procedure we propose is applicable when the data can be modeled as realizations of a LFSM and does not require the assumptions that the observations are normal. Its practical implementation is based on one of the resampling methods of constructing confidence intervals for \( H \). Similar procedures have been proposed in other contexts, e.g., [26].

Suppose then that \( X_0, X_1, \ldots, X_N \) is a realization of a self-similar (motion-type) process with stationary increments sampled at equi-spaced time points. Our approach can be summarized as follows:

1. Divide a time series into \( m \) adjacent blocks of the same length.

2. For a given significance level \( \beta \), use a common to all blocks range of octaves \([j_{\min}, j_{\max}]\) to construct \( 100(1 - \beta)\% \) bootstrap (or block bootstrap, or subsampling) confidence intervals for the self-similarity parameter \( H_i, i = 1, 2, \ldots, m \) in each block; denote this interval by \((l_i, r_i)\).

3. Denote by \( k \) the largest number of intervals \((l_i, r_i)\) with non-empty intersection.

4. At level of significance \( \beta \) reject

\[
H_0 : H_i = H^0, i = 1, 2, \ldots, m
\]

in favor of

\[
H_1 : H_i \neq H_j, \text{ for some } i \text{ and } j,
\]

if \( [100((m - k)/m)]\% > \beta \) and conclude that \( H \) is not constant. Otherwise accept \( H_0 \) and conclude that there is no evidence that \( H \) is not constant.
For example, when $\beta = 0.05$, we fail to reject the null hypothesis if at least 95% of the confidence intervals (with a confidence level of 95%) overlap. Note that since the bootstrap and block bootstrap confidence intervals have tendency to undercover the true $H$ (Section 2.4.4), the acceptance of the null hypothesis will yield a very strong evidence supporting the claim that there is no change in the self-similarity parameter. Rejection, on the other hand, has to be treated with caution.

We applied our test with B, BB and SN methods to the cumulative sums of the four time series plotted in Figures 2.5-2.8. After subtracting a linear trend, these cumulative sums can be modeled as realizations of a self-similar process with stationary increments. Recall that the wavelet coefficients, and hence our procedure, are not affected by a linear trend. We split the four series into 26,17,12, and 7 blocks, respectively, of lengths approximately equal to 10000. We used scales $2^3, \ldots, 2^9$, the same conclusions for all series and all types of intervals (with slightly different intersections) are drawn when smaller set of scales $2^3, \ldots, 2^8$ is used.

Nominal B, BB, SN, 95% confidence intervals based on the first two series, “pAug” and “pOct”, are presented in Figures 2.9 and 2.10. For the series “pAug”, all 26 SN confidence intervals cover the same range (0.8330, 0.8690) and therefore we accept the null hypothesis supporting the claim about constancy in $H$. Less conservative tests based on B and BB methods reject this claim. Slightly different conclusions hold for the series “pOct”. At least $100 - |100(1/17)|\%$ of BB and SN confidence intervals overlap, yielding intersections (0.7706, 0.7819) and (0.7422, 0.7468), respectively. Our results for the series “pAug” and “pOct” based on the SN method thus accord with those arrived at by [51].

The 95% confidence intervals for the series “OctExt” and “OctExt4” are plotted in Figures 2.11 and 2.12. The intersection of SN confidence intervals for the series “OctExt” is (0.9829, 1.1855) and for the series “OctExt4” is (0.9675, 1.2308). For the former series the constancy in $H$ is not confirmed by the tests based on B and BB methods, however for the “OctExt4” series, B and BB-based tests also indicate that $H$ is constant, yielding intersections (1.0404, 1.1072) and (1.0163, 1.1365), respectively. If we were to accept the
constancy in $H$ for both series, the estimated value of $H$ would fall into the range including values greater than 1. These series thus cannot be viewed as increments of a LFSM, for which $0 < H < 1$, but might well be assumed to be increments of a different self-similar process with $0 < \alpha < 1$. Both series, especially “OctExt4”, are seen to have very heavy tails. Recall, see Corollary 7.1.11 of [45], that if $0 < \alpha < 1$ the upper bound on $H$ is $1/\alpha$ and not 1. Recall also that the test of [51] used an approximation to the variance of $\hat{\gamma}_i$, which was based on the assumptions that the observations are approximately normal. Such an assumption is questionable for the series “OctExt” and “OctExt4”, and the rejection reported by [51] might be spurious. We note however that our numerical experiments discussed in Section 2.4 considered only LFSM with $1 < \alpha \leq 2$, so our conclusion of the constancy of $H$ in the series “OctExt” and “OctExt4” must be treated with caution.
Figure 2.1. The 95% confidence intervals (2.8) for the bias of the asymptotic estimator as a function of $j_{\text{min}}$ based on realizations of length $N = 2^{13}$. Estimators $H$ were calculated using periodic (●) and reflection (○) boundary rules, and excluding boundary coefficients (plain bars). All estimators were calculated using scales $2^{j_{\text{min}}}, \ldots, 2^{10}$. 
Figure 2.2. Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{13}$. Estimators were calculated using indicated range of scales.
Figure 2.3. Empirical coverage probabilities for 95% confidence intervals based on realizations of length $N = 2^{12}$. Estimators were calculated using indicated range of scales.
Figure 2.4. Example of the 95% confidence intervals (B, AP, SN, BB) based on $R = 50$ series of length $N = 2^{13}$ with subtracted $H$: scales $2^3, \ldots, 2^9$. 
Figure 2.5. The time series of bytes per 12 milliseconds based on “pAug” trace.

Figure 2.6. The time series of bytes per 10 milliseconds based on “pOct” trace.
Figure 2.7. The time series of bytes per 1000 milliseconds based on “OctExt” trace.

Figure 2.8. The time series of bytes per 1000 milliseconds based on “OctExt4” trace.
Figure 2.9. The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 26 adjacent blocks obtained from “pAug” time series; scales: $2^3, \ldots, 2^9$.

Figure 2.10. The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 17 adjacent blocks obtained from “pOct” time series; scales: $2^3, \ldots, 2^9$. 
Figure 2.11. The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 12 adjacent blocks obtained from “OctExt” time series; scales: $2^3, \ldots, 2^9$.

Figure 2.12. The 95% confidence intervals for $H$ (B, BB, SN, respectively) from 7 adjacent blocks obtained from “OctExt4” time series; scales: $2^3, \ldots, 2^9$. 
CHAPTER 3
WAVELET DOMAIN TEST FOR LONG-RANGE DEPENDENCE IN
THE PRESENCE OF A TREND

3.1 Introduction

Long-range dependent (LRD) processes, also known as long-memory processes, have been extensively used and studied in the past few decades. The relevant literature that has accumulated is too extensive to attempt even a limited review here, so we instead refer the reader to the collection [17] which covers not only the most recent developments in the theory and applications of LRD processes, but also contains a number of review chapters tracing the historical development of important facets of these processes.

The goal of the present paper is to propose and explore a test aimed at distinguishing between a LRD process, possibly with a trend, and a weakly dependent process with a trend. Roughly speaking, under the null hypothesis the observed time series \( X_0, X_1, \ldots, X_{N-1} \) follows the model \( X_t = Y_t + m_t \), where the process \( \{Y_t\} \) is stationary and weakly dependent and \( m_t \) is a deterministic function, and under the alternative the \( X_t \) follow an LRD model. The testing problem is formulated precisely later in this section where the relevant background and research are also reviewed. The test is constructed in the wavelet domain and is motivated by the recent work of [14]. The main reason why the wavelet domain is suitable for such a test is that (non-boundary) wavelet coefficients are invariant with respect to an additive polynomial trend, i.e., the wavelet coefficients of \( \{Y_t + m_t\} \) and \( \{Y_t\} \) are the same, provided that \( m_t \) is a polynomial of sufficiently low order. Since wavelet coefficients are localized in time, only very few of them will be influenced by discontinuities in the function \( m_t \) or its derivatives, provided the number of discontinuities is small relative to the length of the realization. Thus, in the wavelet domain, the process \( \{Y_t + m_t\} \) looks very much the same as the process \( \{Y_t\} \), so the deterministic trend is effectively eliminated from the testing problem. The next step is to find a test statistic based on the wavelet coefficients which has a known distribution, at least in an asymptotic sense, if the underlying process

\footnote{Coauthored by A. Jach and P. Kokoszka.}
is weakly dependent, and which diverges, if it is LRD. In this paper, we focus on the generalized likelihood ratio statistic which can be easily computed because of special properties of the wavelet coefficients, which will be discussed in the following, and whose asymptotic distribution is known to be chi-squared.

Realizations of stationary LRD processes exhibit long, non-periodic cycles which, in finite samples, resemble non-stationary behavior with trends and level shifts. It has been argued for some time that the observed manifestations of long-range dependence can be explained assuming that the observations are weakly dependent, but follow a slightly non-stationary model, for example the model \( X_t = Y_t + m_t \), introduced above. [7] used mathematical arguments to show that the so-called Hurst effect, which motivated Mandelbrot and his collaborators to advocate the use of self-similar LRD processes, can also be explained if the observations are weakly dependent with a trend satisfying certain assumptions. That research was elaborated on by [20] who proved that several statistics akin to the modified R/S statistic of [31] diverge to infinity under either long-range dependence or weak dependence with trend or change-points. In a similar spirit, [16] argued that the appearance of long-memory can be explained by models whose parameters change or evolve with time. [24] demonstrated that many estimators of the memory parameter can be “fooled” in the presence of periodicity or a trend. There is by now ample evidence that, in finite samples, standard tools like ACF plots and periodogram-based spectral estimates behave in a very similar way for LRD processes and for certain types of nonstationarities. Most long-memory tests reject in the presence of a trend or change-points. There is often a controversy which of the two modeling approaches is more appropriate for a specific time series.

There has however not been much research that produced effective tools for distinguishing between long-range dependence and a trend with weakly dependent noise. [29] developed theoretical foundations for a periodogram-based procedure to discriminate between a LRD process and the process \( \{X_t = Y_t + m_t\} \) with a “small” monotonic function \( m_t \). [21] showed that procedures for detecting long-memory which are based on a smoothed periodogram are robust in the presence of “small” trends. These ideas were recently developed by [46]
who proposed a test based on a difference between the estimator of [19] and its version based on the tapered periodogram. In the latter test, the observations are LRD under the null, so it is not comparable with the test proposed in this paper. [6] proposed a test for discriminating between long-range dependence and weak dependence with a change-point in mean, which is a time domain procedure based on a CUSUM statistic for the partial sums of observations. The test of [6] is, however, not suitable if the mean changes smoothly under the null.

We now formulate precisely the testing problem and describe the testing procedure in greater detail.

We assume that the observations follow a Gaussian process both under the null and the alternative. If the weakly dependent process \( \{Y_t\} \) has absolutely summable autocovariance function and its spectral density is positive at every frequency, then it admits both autoregressive and moving average representation of infinite order with absolutely summable coefficients, see e.g., [10], p. 78. Such a process can thus be approximated in mean square by a causal and invertible ARMA\((p,q)\) process. We therefore postulate that under the null \( \{Y_t\} \) is an ARMA\((p,q)\) process and \( m_t \) is a polynomial. As we will see in Section 3.6, we may in practice assume that \( m_t \) is a piecewise polynomial, but the theoretical arguments are available only if \( m_t \) is a polynomial. As a model for the LRD process under the alternative we use the fractional ARIMA model with the differencing parameter \( \delta > 0 \) and the same order \( p,q \) as for the ARMA process \( \{Y_t\} \). We denote this model as ARFIMA\((p,\delta,q)\). The testing problem is thus formulated as follows:

**Null Hypothesis.** The observations \( X_0, X_1, \ldots, X_{N-1} \) follow the model

\[
X_t = Y_t + m_t, \quad 0 \leq t \leq N - 1,
\]

where \( \{Y_t\} \) is a causal and invertible Gaussian ARMA\((p,q)\) process and \( m_t \) is a polynomial.

**Alternative Hypothesis.** The observations \( X_0, X_1, \ldots, X_{N-1} \) follow model (3.1), where \( \{Y_t\} \) is a Gaussian ARFIMA\((p,\delta,q)\) process with \( \delta > 0 \).

The test is based on the approximate decorrelation property of the discrete wavelet transform (DWT) which asserts that the DWT coefficients, especially of an LRD process,
exhibit very small correlations within each level and between levels. The decorrelation property has been established through simulation and theoretical studies, see e.g., [2], [3], Section 9.1 of [41], and references therein. The decorrelation property holds to a particularly good approximation for the non-boundary DWT (NBDWT) coefficients which are also not influenced by a polynomial of a sufficiently low order, see Section 3.2 for further details. Modeling the NBDWT coefficients within each level as either white noise or an AR(1) process and assuming that the coefficients at different levels are uncorrelated, as proposed in [14], we can write down the likelihood function under both the null and alternative hypotheses and construct the generalized likelihood ratio (GLR) statistic. Since one more parameter, $\delta$, is estimated under the alternative, the $-2\log$ GLR has asymptotic $\chi^2(1)$ distribution.

The paper is organized as follows: We first review in Section 3.2 properties of the NBDWT coefficients of an ARFIMA process and the so-called white noise model for these coefficients. In Section 3.3 we introduce the GLR test procedure based on the white noise model. We extend it to the so-called AR(1) model for the NBDWT coefficients in Section 3.4. Section 3.5 contains a simulation study. We further investigate our procedure by applying it to a time series of Nile River yearly minimum water levels in Section 3.6. In Section 3.7, we summarize our findings and provide a broader perspective on the proposed procedure.

3.2 Discrete Wavelet Transform of the ARFIMA Process

In this section we discuss the relevant properties of the NBDWT coefficients of the ARFIMA process. We use the notation and terminology introduced in [41].

Stationary causal and invertible ARFIMA($p, \delta, q$) process \{$X_t$\} is defined by the difference equation

$$(1 - B)^{\delta}\Phi(B)X_t = \Theta(B)Z_t, \quad |\delta| < 0.5,$$

where

$$\Phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p, \quad \Theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$$
satisfy $\Phi(z) \neq 0$ and $\Theta(z) \neq 0$ for all $|z| \leq 1$, $B$ is the backward shift operator, and $\{Z_t\}$ is a white noise (WN) sequence with mean 0 and variance $\sigma^2$. If $\delta = 0$, the $X_t$ follow the ARMA$(p, q)$ model.

The spectral density of $\{X_t\}$ is

$$S_X(f) = \sigma^2 |\Theta(e^{-i2\pi f})|^2 \left|2\sin(\pi f)\right|^{-2\delta}, \quad |f| \leq 1/2.$$  

If $\delta > 0$, the spectral density diverges at the origin and the process is seen to be LRD.

The GLR test described in Section 3.3 is based on the wavelet domain maximum likelihood estimation of the parameter vector $\beta$ of the ARFIMA$(p, \delta, q)$ process defined as

$$\beta = (\delta, \phi, \theta, \sigma^2), \quad \phi = (\phi_1, \phi_2, \ldots, \phi_p), \quad \theta = (\theta_1, \theta_2, \ldots, \theta_q).$$

The estimation is based on the autocovariance sequences of the NBDWT coefficients introduced below. Similar approach for the ARFIMA$(0, \delta, 0)$ process is described in Sections 9.1-9.4 of [41] as well as in [14].

Given a realization of a time series $X_0, X_1, \ldots, X_{N-1}, (N = 2^J, J$ - positive integer), the $N_j = 2^{J-j}$ DWT coefficients for the $j$-th level are obtained (theoretically) by filtering the data with a level $j$ wavelet filter $\{h_{j,l}: l = 0, 1, \ldots, L_j - 1\}$, where $L_j = (2^j - 1)(L - 1) + 1$ and $L$ denotes the length of the corresponding wavelet filter (for example, Daubechies $D(L)$ or least asymmetric $L(L)$ filter, see Section 4.8 of [41]). The transfer function for the level $j$ filter is

$$H_{j,L}(f) = e^{-i2\pi(L_j-1)f} H_{1,L}(2^j f) \prod_{k=0}^{j-2} H_{1,L}(1/2 - 2^k f)$$

and its squared gain function is $H_{j,L}(f) = |H_{j,L}(f)|^2$. Here $H_{1,L}(\cdot)$ is the Fourier transform of the wavelet filter.

Recall that $L'_j = \min([(L - 2)(1 - 2^{-j})], N_j)$ DWT coefficients at level $j$ are influenced by circular filtering (for more details see Comments and Extensions to Section 4.11 of [41]). Let

$$d_{j,k}, \quad j = 1, 2, \ldots, J, \quad k = 0, 1, \ldots, M_j - 1, \quad M_j = N_j - L'_j$$

denote the NBDWT coefficients. The exclusion of the boundary coefficients has two important consequences. Firstly, the NBDWT are “blind” to polynomials of order $K, K \leq L/2 - 1$.  

Secondly, it allows us to view the NBDWT coefficients as a sequence following approximately a white noise WN model, i.e., to a good approximation we may assume that the $d_{j,k}$ are uncorrelated. For a fixed $j$, the autocovariance sequence of the $d_{j,k}$ is, in fact, given by

$$s_{j,r}(\delta, \phi, \theta) = \int_{-1/2}^{1/2} e^{i2\pi r f} S_j(f) df,$$

where

$$S_j(f) = 2^{-j} \sum_{k=0}^{2^j-1} \mathcal{H}_{j,L}(2^{-j}(f+k))S_X(2^{-j}(f+k)),$$

and where $S_X(\cdot)$ is defined by (3.2). Under the WN model all autocovariances except at lag $r = 0$ are assumed to vanish. In the sequel, it is convenient to work with the quantities $c_{j,r}(\delta, \phi, \theta)$ defined by

$$s_{j,r}(\delta, \phi, \theta) = \sigma^2 c_{j,r}(\delta, \phi, \theta).$$

Thus, under the WN model

$$d_{j,k} \sim \text{i.i.d. } N(0, c_{j,0}(\delta, \phi, \theta)\sigma^2).$$

The quantities $c_{j,r}(\delta, \phi, \theta)$ are explicitly given as

$$c_{j,r}(\delta, \phi, \theta) = \int_{-1/2}^{1/2} e^{i2\pi f 2^{-j}} \sum_{k=0}^{2^j-1} \mathcal{H}_{j,L}(2^{-j}(f+k)) \frac{\Theta(e^{-i2\pi 2^{-j}(f+k)})}{|\Theta(e^{-i2\pi 2^{-j}(f+k)})|^2} \frac{|2\sin(\pi 2^{-j}(f+k))|^{-2\delta}}{2^{2j+1}} df.$$

To speed up the calculations, what is particularly important for a simulation study, an approximation for $c_{j,r}(\delta, \phi, \theta)$ is needed. As [14], we obtained very good results using the so-called bandpass approximation. This method involves replacing $\mathcal{H}_{j,L}(f)$ in (3.6) by the squared gain function for the exact bandpass filter with passband $[1/2^{j+1}, 1/2^j]$ and yields

$$c_{j,r}(\delta, \phi, \theta) \approx 2^{j+1} \int_{1/2^{j+1}}^{1/2^j} \cos(2\pi f 2^j r) \frac{\Theta(e^{-i2\pi f})}{|\Theta(e^{-i2\pi f})|^2} \frac{|2\sin(\pi f)|^{-2\delta}}{2^{2j+1}} df.$$
3.3 The Test Procedure

Since the NBDWT coefficients do not depend on the polynomial $m_\ell$ in (3.1), any test based on a statistic which is a function of the NBDWT coefficients can be reformulated as testing

(3.8) $H_0 : \delta = 0$ against $H_A : \delta > 0$.

In fact, we could consider a broader class of alternatives in which the observations follow an ARFIMA model with a polynomial trend. Thus, a rejection of $H_0$ means that the data exhibit long-range dependence, possibly with a polynomial trend, but they are not weakly dependent with a polynomial trend.

Our task is thus to propose a test statistic corresponding to the testing problem (3.8). To do so, we combine the techniques described by [14] and [41] with the generalized likelihood ratio (GLR) test, see Section 18.1 of [5] for an elementary introduction to the GLR test.

Assuming the WN model for the NBDWT coefficients, the likelihood function takes the form

(3.9) $f(d; \delta, \phi, \theta, \sigma^2) = \prod_{j=1}^{J} \prod_{k=0}^{M_j-1} (2\pi c_{j,0}(\delta, \phi, \theta)\sigma^2)^{-1/2} \exp \left( -\frac{d_{j,k}^2}{2c_{j,0}(\delta, \phi, \theta)\sigma^2} \right),$

where the array

$d = \{d_{j,k} : j = 1, 2, \ldots, J, k = 0, 1, \ldots, M_j - 1\}$

plays the role analogous to the vector of observations in the classical maximum likelihood estimation.

Twice the negative log likelihood, $-2 \log f(d; \delta, \phi, \theta, \sigma^2)$ (log(·) denotes the natural logarithm), is given as

$$-2 \log f(d; \delta, \phi, \theta, \sigma^2) = M \log(2\pi \sigma^2) + \sum_{j=1}^{J} M_j \log(c_{j,0}(\delta, \phi, \theta)) + \sum_{j=1}^{J} \sum_{k=0}^{M_j-1} \frac{d_{j,k}^2}{c_{j,0}(\delta, \phi, \theta)\sigma^2}.$$

$$= M \log(2\pi \sigma^2) + \sum_{j=1}^{J} M_j \log(c_{j,0}(\delta, \phi, \theta)) + \sum_{j=1}^{J} \frac{R_j}{c_{j,0}(\delta, \phi, \theta)\sigma^2},$$
where

\[ R_j = \sum_{k=0}^{M_j-1} d_{j,k}^2 \quad \text{and} \quad M = \sum_{j=1}^{J} M_j. \]

Minimizing this expression with respect to \( \sigma^2 \) yields the maximum likelihood estimate of \( \sigma^2 \) as a function of the remaining parameters:

\[ 3.10 \]

\[ \hat{\sigma}^2(\delta, \phi, \theta) = \frac{1}{M} \sum_{j=1}^{J} \frac{R_j}{c_{j,0}(\delta, \phi, \theta)}. \]

Replacing \( \sigma^2 \) by \( \hat{\sigma}^2(\delta, \phi, \theta) \) in \( -2 \log f(d; \delta, \phi, \theta, \sigma^2) \) yields a function of \( \delta, \phi, \) and \( \theta \) only, namely,

\[ -2 \log f(d; \delta, \phi, \theta, \hat{\sigma}^2(\delta, \phi, \theta)) \]

\[ = M(\log(2\pi) + 1) + M \log(\hat{\sigma}^2(\delta, \phi, \theta)) + \sum_{j=1}^{J} M_j \log(c_{j,0}(\delta, \phi, \theta)). \]

Recall now the definition (3.3) of the parameter vector \( \beta \) and introduce the parameter spaces

\[ \Omega_0 = \{ \beta : \delta = 0 \}, \quad \Omega = \{ \beta : \delta \geq 0 \}. \]

Minimizing \( -2 \log f(d; \delta, \phi, \theta, \sigma^2) \) over \( \Omega_0 \) leads to the estimators \( \hat{\phi}_0 \) and \( \hat{\theta}_0 \) whereas minimizing over \( \Omega \) yields estimators \( \hat{\delta}, \hat{\phi}, \) and \( \hat{\theta} \). We thus obtain two estimators of the parameter vector \( \beta \):

\[ \hat{\beta}_0 = (0, \hat{\phi}_0, \hat{\theta}_0, \hat{\sigma}^2(0, \hat{\phi}_0, \hat{\theta}_0)), \quad \hat{\beta} = (\hat{\delta}, \hat{\phi}, \hat{\theta}, \hat{\sigma}^2(\hat{\delta}, \hat{\phi}, \hat{\theta})), \]

from which we can construct the GLR statistic

\[ 3.11 \]

\[ \lambda(d) = \frac{\max\{ f(d; \beta) : \beta \in \Omega_0 \}}{\max\{ f(d; \beta) : \beta \in \Omega \}} = \frac{f(d; \hat{\beta}_0)}{f(d; \hat{\beta})}. \]

Under \( H_0 : \delta = 0 \), \( -2 \log \lambda(d) \) converges to the \( \chi^2(1) \) distribution, see, e.g., Theorem 6.3.2 of [8]. Therefore, the size \( \alpha \) asymptotic GLR test rejects \( H_0 \) if \( -2 \log \lambda(d) > \chi^2_{1-\alpha}(1) \), where \( \chi^2_{q}(r) \) denotes \( q \)-th quantile of the chi-square distribution with \( r \) degrees of freedom.

3.4 Extension to AR(1) Model for Wavelet Coefficients

In this section we explain how the WN model for the NBDWT coefficients and the resulting GLR test can be extended to the AR(1) model for the the NBDWT coefficients.
Such an extension was proposed by [14] to better fit the autocorrelation structure of the wavelet coefficients within each scale.

Recall that the exact autocovariances within each level are given by (3.4). In the WN model it is assumed that all but lag 0 autocovariances vanish. In the AR(1) model we assume that only lag 0 and lag 1 autocovariances are non-zero. The covariances between different levels are still assumed to vanish. These assumptions also lead to a tractable wavelet domain likelihood, as explained below.

We thus say that the NBDWT coefficients \( d_{j,k} \) follow the AR(1) model if

\[
d_{j,k} = r_j(\delta, \phi, \theta)d_{j,k-1} + Z_{j,k}, \quad j = 1, 2, \ldots, J, \quad k = 0, 1, \ldots, M_j - 1,
\]

where \( Z_{j,k} \sim N(0, \kappa_j(\delta, \phi, \theta)\sigma^2) \) is an array of independent random variables. Yule-Walker equations provide formulas for \( r_j \) and \( \kappa_j \):

\[
  r_j(\delta, \phi, \theta) = c_{j,1}(\delta, \phi, \theta)/c_{j,0}(\delta, \phi, \theta),
\]

\[
  \kappa_j(\delta, \phi, \theta) = c_{j,0}(\delta, \phi, \theta)(1 - r_j^2(\delta, \phi, \theta)),
\]

with \( c_{j,r}(\delta, \phi, \theta) \) defined by (3.6). Similar considerations as in Section 3.3 show that the maximum likelihood estimates are obtained through the minimization of

\[
-2 \log f(d; \delta, \phi, \theta, \hat{\sigma}^2(\delta, \phi, \theta))
= M(\log(2\pi \hat{\sigma}^2(\delta, \phi, \theta)) + 1) + \sum_{j=1}^{J} (M_j \log(\kappa_j(\delta, \phi, \theta)) - \log(1 - r_j^2(\delta, \phi, \theta))),
\]

with respect to \( \delta, \phi, \) and \( \theta \), where the estimate of \( \sigma^2 \) is given by

\[
\hat{\sigma}^2(\delta, \phi, \theta) = \frac{1}{M} \sum_{j=1}^{J} \frac{1}{\kappa_j(\delta, \phi, \theta)} \left( d_{j,0}^2(1 - r_j^2(\delta, \phi, \theta)) + \sum_{k=1}^{M_j-1} (d_{j,k} - r_j(\delta, \phi, \theta)d_{j,k-1})^2 \right).
\]

The null hypothesis is rejected if

\[
-2 \log \lambda(d) = -2[\log f(d; 0, \hat{\phi}_0, \hat{\theta}_0, \hat{\sigma}^2(0, \hat{\phi}_0, \hat{\theta}_0)) - \log f(d; \hat{\delta}, \hat{\phi}, \hat{\theta}, \hat{\sigma}^2(\hat{\delta}, \hat{\phi}, \hat{\theta}))]
\]

\[
> \chi^2_{1-\alpha}(1).
\]
3.5 Simulation Study

Design of the study. For our simulation study, we implemented the test procedure assuming the ARMA(1, 0) model under the null and the ARFIMA(1, δ, 0) under the alternative. We assessed its finite sample performance based on at least $R = 500$ replications of data generating processes of lengths $N = 512$ and $N = 1024$. We analyzed the simulated realizations with wavelet filters $D(6)$ and $LA(8)$, sufficiently long to provide good bandpass approximation (3.7), which we used. We focused on nominal sizes $\alpha = 0.05$ and $\alpha = 0.10$ and considered both WN and AR(1) models for the NBDWT coefficients. According to the theory explained in Sections 3.2, 3.3 and 3.4, the size and power of the test do not depend on a polynomial trend, provided the order of the polynomial does not exceed $L/2 - 1$, where $L$ is the length of the wavelet filter. This was confirmed in several test cases we considered: by adding different polynomials of degree 0, 1 or 2, we obtained essentially the same sizes, with differences only slightly greater than machine precision. We report here sizes obtained with the polynomial

$$m_r = 0.25t^2/N$$

under the null hypothesis and no polynomial trend under the alternative.

When evaluating the empirical size, the process $\{Y_t\}$ in (3.1) was generated according to the AR(1) model with nine choices of $\phi$, $\phi \in \{0.1, 0.2, \ldots, 0.9\}$ and common variance $\sigma^2 = 1$. Only positive values of the autocorrelation coefficient were considered because realizations with $\phi \leq 0$ (and possibly a trend) do not resemble realizations of LRD processes and can be told apart from them by eye. To compute the empirical power of the test, we used realizations of ARFIMA(0, δ, 0) processes, also with nine different values of the parameter $\delta$, $\delta \in \{0.10, 0.15, \ldots, 0.50\}$, and common variance $\sigma^2 = 1$. For fixed $\phi, \delta$ and $N$, the same replications were used to better assess the effect of the model assumed for the NBDWT coefficients and the type of the wavelet used.

All numerical experiments reported here were performed in R. We also implemented the procedure in Matlab and obtained very similar results, but the Matlab implementation was slower. R’s optimization routines, “optimize” (minimization with respect to one variable)
and "optim" (minimization with respect to two variables), were used to estimate the parameters. Both of them allow to specify the range of minimization. Due to the constraints on \( \delta \) dictated by the testing problem, this flexibility plays an important role in the practical implementation of the test procedure. We minimized over \( \delta \in [0, 0.99] \) under the alternative and over \( \phi \in [-0.99, 0.99] \) under both null and the alternative, c.f. (3.11).

As will be seen in the following, the test is somewhat conservative for \( N = 512 \). We therefore also investigated the performance of the test with calibrated quantiles, namely, we used 17-th and 8-th upper percentiles of the \( \chi^2(1) \) distribution as the critical values for the test with nominal size of 10 and 5 percent, respectively.

**Discussion of the results.** Empirical sizes and powers, for different choices of \( N \), wavelet models, wavelet filters, and nominal sizes are presented in Figures 3.2-3.5. Figures corresponding to the empirical sizes include the 95 percent asymptotic confidence bounds \( \pm 1.96\sqrt{\hat{\alpha}(1 - \hat{\alpha})/R} \), where \( \hat{\alpha} \) is the empirical size. Results for the calibrated test and all numerical values are available upon request.

We begin with the discussion of the empirical sizes. The method works well when the parameter \( \phi \) of is between 0.1 and 0.8. The rejection probabilities become too large as \( \phi \) approaches unity. This is not surprising, as for large values of \( \phi \) the spectrum of an AR(1) process looks similar to the spectrum of an LRD process. In the following, we therefore focus our discussion on the cases with \( 0.1 \leq \phi \leq 0.8 \).

By comparing the empirical sizes obtained under the assumption of the WN model for the NBDWT coefficients to those computed under the AR(1) model assumption (left panel against the right in Figures 3.2 and 3.3), we conclude that the former provides more accurate results. The empirical sizes calculated under the assumption of the AR(1) model exhibit slightly greater undercoverage than the sizes obtained under the WN model.

Empirical sizes plotted in Figures 3.2 and 3.3 indicate that our test is conservative, especially for \( N = 512 \). The undercoverage corresponding to the nominal value of \( \alpha = 0.05 \) is slightly smaller than that for \( \alpha = 0.10 \), for both choices of \( N \) and for both models for the NBDWT coefficients. For \( N = 512 \), the 5 percent test has empirical size of about 3.5
percent. The empirical sizes are closer to the nominal sizes for $N = 1024$ and for both values of $N$ generally increase with $\phi$. For $0.3 \leq \phi \leq 0.8$, the range most commonly encountered in practice, the empirical sizes are very good for $N = 1024$ and the LA(8) filter.

The choice of the wavelet filter, D(6) versus LA(8), does not substantially affect the results (compare top two panels to bottom two in Figures 3.2 and 3.3), which suggests that either of the two filters can be used. The LA(8) filter should be used if the presence of a cubic polynomial trend is suspected ($K = 8/2 - 1 = 3$).

The sample size has the most pronounced effect. Our test provides more accurate results for $N = 1024$ than for $N = 512$ (compare Figure 3.2 to Figure 3.3). This is in agreement with the asymptotic nature of the test, and is likely due to the larger number of the wavelet coefficients available for $N = 1024$, especially those unaffected by the circular filtering.

Calibrated quantiles improve the performance of the test for the smaller sample size, $N = 512$, but should not be used for $N = 1024$ where they lead to an overcoverage.

The above discussion was based on results obtained from $R = 500$ replications. For the LA(8) filter we also obtained empirical sizes based on $R = 2000$ replications which confirm the findings discussed above and show that the empirical size increases slowly with $\phi$.

We conclude this section with a brief discussion of the empirical power. The general shape of the power curve is very similar for the two significance levels considered and for fixed $N$ (see Figures 3.4 and 3.5). The power is high and exceeds 80% for $N = 1024$ and $\delta \geq 0.2$. It is slightly higher under the AR(1) model for the NBDWT coefficients (compare left and right panels in Figures 3.4 and 3.5). The power converges to 1 much faster for $N = 1024$ than for $N = 512$ (compare Figure 3.4 to Figure 3.5). The choice of the filter does not matter much. Calibration of the quantiles does not noticeably affect the results.

3.6 Application to the Annual Minima of the Nile

We applied the test to Nile River yearly minimum water levels. The whole data set covers years 622 to 1284 and has been extensively studied in the long-memory literature. Here we focus on the last 512 observations plotted in the top panel of Figure 3.1. Visual
Figure 3.1. Top: Nile River annual minimum water levels for years 773 to 1284 with fitted second order polynomial trend and piecewise regression line. Middle: Simulated realization of $X_t = m_t + Y_t$, where $m_t$ is the fitted polynomial and $Y_t$ follows an AR(1) model fitted to the residuals. Bottom: Simulated realization of $X_t = m_t + Y_t$, where $m_t$ is the fitted piecewise regression line and $Y_t$ follows an AR(1) model fitted to the residuals.

examination of this time series suggests that rather than considering a long-memory model one might use an AR(1) model with a positive AR coefficient and with a smooth or discontinuous trend. Using the least squares method, we fitted a second order polynomial and a piecewise linear function to the Nile minima. The break points of the piecewise linear function were chosen somewhat arbitrarily to reflect the apparent level and slope shifts in the observations. We then estimated the AR(1) model on the two sets of the residuals. For comparison, we simulated realizations from the resulting two models and added them to the corresponding trends. Visual comparison of the two lower panels of Figure 3.1 with the top panel shows that an AR(1) model with a trend, especially with a piecewise linear trend, might be a reasonable alternative to an LRD model.

We applied our test procedure to the 512 observations assuming the WN and AR(1) models for the NBDWT coefficients and using both D(6) and LA(8) wavelet filters. The
Table 3.1. P-values for the GLR test applied to $N = 512$ annual Nile minima.

<table>
<thead>
<tr>
<th>Wavelet filter</th>
<th>WN model</th>
<th>AR(1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>D(6)</td>
<td>0.0033</td>
<td>0.0026</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.0155</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

Table 3.2. Empirical size of the GLR test based on $R = 1000$ replications of $X_t = Y_t + m_t$ of length $N = 512$, where $\{Y_t\}$ follows AR(1) model with parameters $\phi = 0.5820$ and $\sigma^2 = 4333$ and $m_t$ is given by (3.13).

<table>
<thead>
<tr>
<th>Filter $\alpha$</th>
<th>WN model</th>
<th>AR(1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>D(6)</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.077</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>0.026</td>
</tr>
</tbody>
</table>

p-values presented in Table 3.1 are small, so we reject the null hypothesis and conclude that an AR(1) model with a trend is not suitable. We now provide a more detailed validation of our conclusion. The polynomial fitted to the data is

$$m_t = -0.0009t^2 + 2.0796t + 16.7069.$$  

Fitting an AR(1) model to the residuals, we obtained the autoregressive coefficient $\phi = 0.5820$ and the WN variance $\sigma^2 = 4394$. The simulations in Section 3.5 show that for $\phi = 0.6$ our test has about the correct size. This is confirmed by additional simulation results presented in Table 3.2. For example, when the test with the WN model and the D(6) filter is applied to the estimated model, the empirical sizes of 10 and 5 percent level tests are 10.8 and 5.5 percent, respectively. The theory and simulations presented earlier in the paper do not apply to piecewise polynomial trends. It can however be expected that if there are relatively few break points compared to the length of the series, only few DWT coefficients will be affected by these breaks and the test will continue to have correct size. This is indeed confirmed by our simulations. The piecewise linear function we considered has constant slope over the following three periods: 773-1009, 1010-1099, 1100-1284. If $t = 0$ corresponds to year 773, it can be written as

$$m_t = \begin{cases} 
0.2t + 934.9, & t = 0, 1, \ldots, 236, \\
0.8t + 299.4, & t = 237, 238, \ldots, 326, \\
-0.8t + 2088.8, & t = 327, 328, \ldots, 511.
\end{cases}$$  

(3.14)
Table 3.3. Empirical size of the GLR test based on $R = 1000$ replications of $X_t = Y_t + m_t$ of length $N = 512$, where $\{Y_t\}$ follows AR(1) model with parameters $\phi = 0.5628$ and $\sigma^2 = 4394$ and $m_t$ is given by (3.14).

<table>
<thead>
<tr>
<th>Filter \ $\alpha$</th>
<th>WN model</th>
<th>AR(1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>D(6)</td>
<td>0.078</td>
<td>0.048</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.083</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Table 3.4. Empirical power of the GLR test based on $R = 1000$ replications of ARFIMA$(1,\delta,0)$ process of length $N = 512$ with parameters $\phi = 0.0660$, $\delta = 0.4013$ and $\sigma^2 = 0.0028$.

<table>
<thead>
<tr>
<th>Filter \ $\alpha$</th>
<th>WN model</th>
<th>AR(1) model</th>
</tr>
</thead>
<tbody>
<tr>
<td>D(6)</td>
<td>0.838</td>
<td>0.773</td>
</tr>
<tr>
<td>LA(8)</td>
<td>0.769</td>
<td>0.693</td>
</tr>
</tbody>
</table>

The AR(1) model for the residuals has parameters $\phi = 0.5628$ and $\sigma^2 = 4333$. The empirical sizes are shown in Table 3.3. For the test with the WN model and the D(6) filter, the empirical sizes of 10 and 5 percent level tests are 7.8 and 4.8 percent, respectively. The ARFIMA$(1,\delta,0)$ model estimated on the observations has parameters $\phi = 0.0660$, $\delta = 0.4013$, and the WN variance $\sigma^2 = 0.0028$. The empirical power of the test for this model is presented in Table 3.4. For all but one combinations of nominal size, wavelet filter, and wavelet model, the power exceeds 70 percent and is about 80 percent for the D(6) filter and 5 percent nominal significance level.

3.7 Summary and Discussion

Motivated by the work of [14], we developed and investigated a wavelet domain test in which under the null the time series is weakly dependent with a polynomial trend and under the alternative it is LRD, possibly with a polynomial trend. We assumed that the $Y_t$ in (3.1) follow and ARMA$(p,q)$ model under the null and ARFIMA$(p,\delta,q)$, $\delta > 0$ under the alternative. We investigated the finite sample performance of the test for $p = 1$ and $q = 0$. Our findings can be summarized as follows:

1. The test has about correct size for moderate weak dependence which can be quantified
by the condition \(0.1 \leq \phi \leq 0.8\). It performs noticeably better for \(N = 1024\) than for \(N = 512\). For \(N = 512\) the test is somewhat conservative for \(0.1 \leq \phi \leq 0.6\).

2. The test has very good power.

3. The WN model for the NBDWT coefficients is slightly better than the AR(1) model, as far as the empirical size is concerned.

4. Both wavelet filters, D(6) and LA(8), yield similar results.

5. The test can be applied if a piecewise polynomial trend is suspected, provided there are few break points relative to the length of the series.

The approach proposed here can be extended to different settings. The key requirement is that the spectral densities under the null and the alternative must be specified by a parametric model such that weak dependence corresponds to a fixed value of a memory parameter and long-range dependence to a range of values. A very natural alternative to the ARFIMA specification considered here is the fractional exponential model (FEXP) recently studied by [33], among others. The spectral density of the FEXP\((r,\delta)\) model is

\[
S(f) = \sigma^2 |2 \sin(\pi f)|^{-2\delta} \exp \left\{ \sum_{j=1}^{r} c_j \cos(2\pi j f) \right\},
\]

where the memory parameter \(\delta\) has the same interpretation as in the ARFIMA model. If \(\delta = 0\), the FEXP model becomes the exponential model of [9].

The choice of the order \(p, q\) may require further investigations. From the practical point of view, AR(1) and ARMA(1,1) models can be used as good approximations to the autocovariance structure of a weakly dependent linear process. Order selection in the wavelet domain, however, offers itself as an interesting and challenging problem.

The investigation of these and similar modifications is, however, beyond the intended scope of the present paper.
Figure 3.2. Empirical size of the GLR test based on $R = 500$ replications of $X_t = Y_t + m_t$ of length $N = 512$, where $\{Y_t\}$ follows AR(1) model with given $\phi$ and $m_t = 0.25\sigma^2/N$. Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.
Figure 3.3. Empirical size of the GLR test based on $R = 500$ replications of $X_t = Y_t + m_t$ of length $N = 1024$, where $\{Y_t\}$ follows AR(1) model with given $\phi$ and $m_t = 0.25t^2/N$. Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.
Figure 3.4. Empirical power of the GLR test based on $R = 500$ replications of $X_t$ of length $N = 512$, where $\{X_t\}$ follows ARFIMA$(0,\delta,0)$ model with given $\delta$. Wavelet filter: $D(6)$, $LA(8)$; model for the wavelet coefficients: $WN$ (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.
Figure 3.5. Empirical power of the GLR test based on $R = 500$ replications of $X_t$ of length $N = 1024$, where $\{X_t\}$ follows ARFIMA(0,δ,0) model with given $\delta$. Wavelet filter: D(6), LA(8); model for the wavelet coefficients: WN (left panel) and AR(1) (right panel); nominal size indicated by the solid horizontal line.
CHAPTER 4
WAVELET-BASED INDEX OF MAGNETIC STORM ACTIVITY

4.1 Introduction

The currents flowing in the magnetosphere-ionosphere (M-I) form a complicated multi-scale geosystem that contains the temporal scales from seconds to days. Ground-based magnetometers have long been an important tool to observe the M-I current system and a number of indices based on magnetometer data have been introduced to characterize the variations of specific current components. Due to the nature of this current system, the magnetometer data are multi-scale, impulsive, and asynchronous with non-stationary frequency spectra. Based on the assumption that magnetometers in certain latitude bands are most sensitive to specific currents, a traditional way to separate the magnetic effects of different currents is to use magnetometer data from a specific region and combine them into an index characterizing the variation of a specific current. References to review papers on geomagnetic indices are given on p. 409 of [27] and a concise account of the main indices is presented in Appendix 13B on p. 451 of that monograph.

In this paper we propose an automatic wavelet-based statistical procedure designed to develop an index of storm activity associated with the intensification of the ring current. The Dst index, see [49], has long been used to characterize this variation and is produced from the magnetometer data recorded in the equatorial region. The Dst index was originally designed to describe the variation of the symmetric ring current. But the procedure of producing the Dst index can not eliminate the magnetic effects from many local-time dependent currents in the ionosphere and magnetosphere, including the partial ring current, auroral currents, magnetotail current, etc. Therefore, what the Dst index describes is actually the overall magnetic effect of storm activity at the low- and mid-latitude regions (see [12], [18], [23], [44]). A main difficulty in producing this index lies in subtracting the quiet day variation from a magnetometer data at a given location. After this quiet day component and the long term component have been subtracted, the remainder is believed to describe the storm

---

related magnetic activity. In the procedure of the standard Dst, constructing the quiet day component involves somewhat subjectively choosing several days in a month, during which the storm activity is believed to be absent, and averaging them. We show that using a wavelet technique it is possible to produce a one-minute index which, after averaging over one hour periods, is very close to the standard Dst. The unique strength of the technique we propose is that it is fully automatic and, in particular, does not involve selecting quiet days. In this technique, the daily and long-term variation, as well as inconsequential noise, are removed in the wavelet domain by means of statistical filtering. The procedure requires only the most recent stretch of data, and can be used to quickly produce a signature of a storm event.

We now introduce a wavelet-based representation for magnetometer data and formulate a time series model needed to explain the central ideas of the proposed approach. More details are presented in Section 4.2. For the purpose of this study, the specific magnetometer data to be used is the station’s H component, i.e., the magnitude of the horizontal magnetic field. This is the same component as used in the standard Dst calculation, see [49].

Suppose \(X_s = \{X_{s,t} : t = 0, 1, \ldots, N-1\}\) is a magnetometer data at station \(s\), the sample size \(N\) is the length of the record in minutes. The magnetogram \(X_s\) can be decomposed as

\[
X_s = \sum_{j=1}^{J} D_{s,j} + S_{s,j},
\]

where

\[
D_{s,j} = \{D_{s,j,0}, \ldots, D_{s,j,N-1}\}, \quad S_{s,j} = \{S_{s,j,0}, \ldots, S_{s,j,N-1}\}.
\]

Decomposition (4.1) is known as the Multi Resolution Analysis (MRA). The details \(D_{s,j,t}\) correspond to the component of the record at time of approximately \(2^j\) minutes and to frequencies in the range of \(2^{-j-1}\) to \(2^{-j}\) cycles per minute. This range corresponds to physical scales between \(2^j/60\) and \(2^{j+1}/60\) hours. The smooths \(S_{s,j,t}\) correspond to time of approximately \(2^j\) minutes and to frequencies lower than \(2^{-j-1}\) cycles per minute, or, alternatively, to broadly understood averages over intervals of \(2^{j+1}/60\) hours.

As will be seen in the following, the different components of the magnetogram are most
prominent at specific levels $j$ of decomposition (4.1). By applying suitable statistical filters to each level $j$, we are able to isolate and remove the components which do not reflect the storm activity. It should be noted that the following procedure still cannot separate and distinguish the magnetic effects associated with several local time dependent currents (for example, parital-ring current and auroral currents). Therefore, resulting wavelet-based index basically describes the variation of the same physical processes as the Dst index does, that is the storm time activity. However the important new features of the wavelet index, which are not encountered in the conventional Dst, are automation (with no human input) and flexibility on the data stretch.

To justify our methodology, we assume the following approximate model for the time series $\{X_{s,t}\}$:

\[
X_{s,t} = E_{s,t} + A_{s,t} + P_{s,t} + L_{s,t} + N_{s,t},
\]

with the terms on the right-hand side defined as follows:

- $B_s$: Internal magnetic field of the Earth measured at a low latitude station $s$ (assumed to be time independent).
- $A_{s,t}$: The disturbance component attributable to storm activity.
- $P_{s,t}$: Periodic component.
- $L_{s,t}$: Slowly varying trend (assumed to be a low degree polynomial).
- $N_{s,t}$: Noise component.

The components $A_{s,t}$ and $N_{s,t}$ should be thought of as random; $A_{s,t}$ is a “large” random component which becomes pronounced during a storm; $N_{s,t}$ is a “small” random component reflecting all kinds of random disturbances from the measurement error to irregular random disturbances of the M-I system. Small random variations in deterministic components $B_s$, $P_{s,t}$ and $L_{s,t}$ are thought of as being moved to $N_{s,t}$, so, for example, $P_{s,t}$ is considered a deterministic periodic component to be statistically estimated. By contrast, the component
As, which we want to isolate, is random because we do not assume anything about the timing or signature of a storm event.

Admittedly, model (4.2) is a simplification needed to develop a usable statistical methodology. For example, \( L_{s,t} \) evolves in a complex manner with the solar cycle and treating it as a polynomial is only an approximation. The periodicity of \( P_{s,t} \) is only a mathematical assumption; \( P_{s,t} \) should not be identified with the simple diurnal variability, a loose association is however helpful to understand the procedure. Without these assumptions, the components on the right-hand side would not be identifiable, as only the \( X_{s,t} \) are observable and five components cannot be uniquely identified from their sum. The mathematical assumptions thus play a role of additional equations. A further caveat is that the present analysis does not make any corrections for the ground induction effects. The temporary variations of various components of induction currents are similar to their source currents. Our omission of this correction is in keeping with the existing procedures for calculating the Dst index.

The paper is organized as follows. In Section 4.2 we present some background the wavelet analysis. Sections 4.3 and 4.4 focus on the removal of the noise and periodic components, respectively. In Section 4.5, we formulate the procedure for computing our wavelet based index, and in Section 4.6 we compare it to the standard Dst index. We conclude with final remarks in Section 4.7.

4.2 MODWT of the Magnetometer Data

In this section we describe the Maximum Overlap Discrete Wavelet Transform (MODWT) and explain how it can be applied to a single magnetogram component (scalar data) from a single station. We will also point out the advantages of the MODWT over the Discrete Wavelet Transform (DWT). Finally, we will discuss the choice of the parameters of the wavelet analysis. Our choices are motivated by the task of isolating storm signatures in the presence of a strong periodic component.

Throughout this section, we follow closely the exposition and notation of [41]. We focus
on the aspects of the MODWT which are most relevant to our task, and cannot present all
details. An interested reader is referred to [41], especially Chapter 5 of that monograph.

The MODWT is a non-orthogonal modification of the DWT which addresses some
shortcomings of the latter, such as sample size restriction and sensitivity to the starting point
of the series. Like the DWT, the MODWT produces a set of wavelet and scaling coefficients
obtained by linear filtering of the signal (this is done by an efficient pyramid algorithm,
although with a somewhat higher computational burden than that for the DWT). Unlike
the DWT, there are \( N \) coefficients at each scale, where \( N \) is the number of observations.
The MODWT details and smooths are associated with zero phase filters and, unlike details
and smooths of the DWT, do not require any shifting to align time events.

Suppose \( X_s = \{X_{s,t} : t = 0, 1, \ldots, N - 1\} \) is a scalar component of a magnetometer
recorded at station \( s \). For any integer \( 1 \leq j \leq \log_2(N) \), the MODWT wavelet and scaling
coefficients at level \( j \) are defined as

\[
W_{s,j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{s,t-l \mod N}
\]

and

\[
V_{s,j,t} = \sum_{l=0}^{L_j-1} g_{j,l} X_{s,t-l \mod N}
\]

where \( t = 0, 1, \ldots, N - 1 \), and \( \{h_{j,l}\} \) and \( \{g_{j,l}\} \) are the \( j \)th level MODWT wavelet and scaling
filters, both of length \( L_j = (2^j - 1)(L - 1) + 1 \); \( L \) denotes the length of the underlying wavelet
filter, for example, the Daubechies \( D(L) \) or the least asymmetric \( LA(L) \) filter, see Section
4.8 of [41]. The “mof \( N \)” indicates circular convolution, see Section 2.5 of [41]. In practice,
due to the application of the pyramid algorithm, these filters need not be computed.

The transfer function \( \tilde{H}_{j,L}(f) \) for the level \( j \) filter can be regarded as an approximation to
a transfer function of a perfect band-pass filter with pass-band \([1/2^j+1, 1/2^j]\). The squared
gain functions, defined as \( \tilde{H}_{j,L}(f) = |\tilde{H}_{j,L}(f)|^2 \), are shown in Figure 4.1, which will be
referred to later. Level \( j \) wavelet coefficients are thus associated with the portion of Discrete
Fourier Transform (DFT) of \( X_s \) with frequencies in the interval \([1/2^j+1, 1/2^j]\).

The MODWT yields decomposition (4.1) with level \( j \) detail and smooth sequences \( D_{s,j} \)
and $S_{s,j}$ defined by

\begin{align}
D_{s,j,t} &= \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} W_{s,j,t+l \mod N}, \quad t = 0, 1, \ldots, N - 1 \\
S_{s,j,t} &= \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} V_{s,j,t+l \mod N}, \quad t = 0, 1, \ldots, N - 1.
\end{align}

The circular (mod $N$) filtering in (4.3)-(4.4) and (4.5)-(4.6) generates the so-called boundary MODWT coefficients which do not have a physical interpretation, but are due to the fact that the observations are not part of an infinite sequence with period $N$. For long records, there are relatively few boundary coefficients, especially for small $j$. Several methods of reducing the impact of circular filtering have been devised. In this paper we use the reflection rule, see [41], pp. 140-141.

The application of the MODWT to the magnetometer data requires additional specifications, such as the choice of a wavelet filter and the maximum level $J$. With the goals of the analysis in mind, we discuss these issues below.

Wavelet filter: In our study, we use the LA(8) filter which has been extensively applied in several recent quantitative analyzes of geophysical data, see [13], [35], [54], among others. Using Daubechies filter D(8) produces the same results. LA filters are parametrized by even widths $L$. For a specific analysis, a width must be selected which would provide sufficient concentration in the octave pass-band and reduce the overlapping of frequencies from outside of it. This can be achieved by selecting a relatively long filter, but choosing $L$ too large will result in many boundary coefficients, especially at higher levels $j$. Choosing $L = 8$ has been found to offer a reasonable balance between these two competing requirements. This choice also guarantees that the MODWT coefficients are uninfluenced by a polynomial of order $K \leq L/2 - 1 = 3$. Thus, a slowly varying trend (secular variation) in the magnetic field, which is typically modeled by a low order polynomial, see [27], p. 457, is practically not reflected in the details of the MODWT of magnetometer data. On the other hand, for one-year-long magnetometer data, this trend can be captured by a smooth $S_{s,j}$ on the highest available level, and thus an appropriate adjustment of $S_{s,j}$ will automatically remove it.
**Number of levels J:** The maximum level $J$ in (4.1) is typically chosen to be smaller than its largest admissible value $\lfloor \log_2(N) \rfloor$. However, for one-year-long data, we need to account for the secular variation. Decomposing the signal up to level $J = \lfloor \log_2(N) \rfloor = 19$ and replacing $S_{19}$ by its average (this average is equal to the average of the magnetogram) allow to eliminate the annual trend. For shorter records, for example, two months long, the smooth corresponding to the highest available level is affected by the storm events and cannot be regarded as a good approximation to the annual trend. We assume that the long-term component is not visible in the records extending over the period of two months. This assumption is validated by comparing our index computed from two months of data to the standard Dst index. For such records we propose to set $J = 10$, because $J$ should be large enough to allow the removal of the noise and periodic components. Since the latter is visible at higher scales, $J$ will be dictated by the frequencies contributing to the periodic (Sq) variation. As we will see in Section 4.4, the lowest frequency corresponding to the periodic portion of the magnetogram is the daily frequency, $1/1440$, that falls in $[2^{-j-1}, 2^{-j}]$, $j = 10$, but there are also some higher frequencies. Taking $J = 10$ ensures that these frequencies can be effectively analyzed.

### 4.3 Noise Removal via Wavelet Thresholding

Wavelet thresholding is a nonparametric method of estimating a signal in the presence of additive noise, see Chapter 10 of [41] for a review, and has been developed both theoretically and practically over the last decade. The underlying assumption of this methodology is that the noise is "small" compared to the signal. We use thresholding to remove the component $N_{s,t}$ in decomposition (4.2). To justify this approach, note that by (4.2) and (4.3)

\begin{equation}
W_{s,j,t} = W_{s,j,t}^{(A)} + W_{s,j,t}^{(P)} + W_{s,j,t}^{(N)},
\end{equation}

where the $W_{s,j,t}^{(A)}$ are the MODWT coefficients of the storm component $A_{s,t}$, $W_{s,j,t}^{(P)}$ and $W_{s,j,t}^{(N)}$ being defined analogously. The wavelet coefficients of the components $B_s$ and $L_{s,t}$ vanish because these components are assumed to be a constant and a polynomial of degree not exceeding three, respectively. At low levels $j \leq J_0$, where the value of $J_0$ is specified at the
end of this section, the periodic coefficients $W_{s,j,t}^{(P)}$ are negligible, see Section 4.4. Thus we may in fact assume that

$$W_{s,j,t} = W_{s,j,t}^{(A)} + W_{s,j,t}^{(N)}.$$ \tag{4.8}

Moreover, for $j \leq J_0$, the $W_{s,j,t}$ are mostly "small noise" with large coefficients during storm events. The goal of thresholding is to remove the $W_{s,j,t}^{(N)}$ which constitute the vast majority of the coefficients at low scales, and to estimate $W_{s,j,t}^{(A)}$. This can be done only in some statistical sense because the $W_{s,j,t}^{(N)}$ are not observable. We assume that the absolute values of the $W_{s,j,t}^{(N)}$ approximately do not exceed a level $\delta_j$. The observed coefficients $W_{s,j,t}$ are thus replaced by zero if their absolute value is smaller than $\delta_j$ and their length is reduced by $\delta_j$ if their absolute value is greater than $\delta_j$. The resulting coefficients are denoted by $\bar{W}_{s,j,t}$ and are thus given by the formula

$$\bar{W}_{s,j,t} = \text{sign}(W_{s,j,t})(|W_{s,j,t}| - \delta_j)_+, \quad t = 0, 1, \ldots, N - 1 \quad (j \leq J_0).$$ \tag{4.9}

The $\bar{W}_{s,j,t}$ are statistical estimates of the $W_{s,j,t}^{(A)}$ in the absence of the $W_{s,j,t}^{(P)}$.

Formula (4.9) defines the so-called soft level-dependent thresholding. In hard thresholding, coefficients $W_{s,j,t}$ whose absolute value exceeds $\delta_j$ are not modified. Soft thresholding seems more appropriate in our context because formula (4.8) suggests that $W_{s,j,t}^{(N)}$ should be removed from all coefficients. Moreover, our exploratory analysis has shown that hard thresholding introduces occasional long spikes to the proposed index of storm activity which have no reasonable physical interpretation. In level-independent thresholding $\delta_j = \delta$ is the same for all levels and is typically determined by the statistical behavior of the coefficients at the finest level $j = 1$. This approach is suitable for the task of removing noise from a signal, but is not appropriate in our context because for magnetometer data the magnitude of the $W_{s,j,t}^{(N)}$ visibly changes with level $j$.

A natural and convenient way to select $\delta_j$ is to define it as the $p$th quantile of the distribution of the absolute values of the $W_{s,j,t}$. Thus, we define $\delta_j = \delta_{j,s}(p)$ by the formula

$$P(|W_{s,j,t}| \leq \delta_{j,s}(p)) = p.$$ \tag{4.10}
The probability $P$ on the left-hand side of (4.10) is the empirical probability, i.e., the proportion of the $|W_{s,j,t}|, t = 0, 1, \ldots, N - 1$ which exceed $\delta_{j,s}(p)$. Our exploratory analysis has shown that $p$ should be taken relatively large (close to 1). For example, if $p = 0.98$, all coefficients, except the largest (in absolute terms) 2% of are set to zero. Note that this statistical approach guarantees that the thresholding can be automatically applied to every station no matter what the typical magnitude for the magnetic field is.

Recall that $J_0$ is the largest scale on which the effect of the periodic component is negligible. The periodic component is associated with the Sq variation which has a pronounced daily frequency (1/1440 cycles per minute), half-daily frequency (2/1440) and whose spectrum also contains higher frequencies, even though these are much less pronounced. Note that the half-daily frequency corresponds to the scale of $1440/2 = 770$ minutes, which is contained in the scale range $[512, 1024]$ corresponding to $j = 9$. Because of the presence of frequencies higher than $2/1440$ in the spectrum of the Sq variation, the periodic component is also visible at level $j = 8$. Visual inspection reveals that at levels $j \leq 7$, the periodic component is not visible and the MODWT coefficients at these levels form an approximately stationary process. We therefore choose $J_0 = 7$ based both on theoretical grounds discussed above and on an exploratory analysis of the wavelet decompositions of the magnetometer data.

4.4 Removal of the Periodic Component

This section proposes a statistical method of removing the periodic component $P_{s,t}$ in decomposition (4.2). We loosely identify the periodic component with the daily variation caused by the rotation of the Earth. As the position of the station relative to the M-I current system changes, various currents, including the Sq currents on the dayside of the ionosphere, the magnetopause and tail currents, leave approximately periodic signatures in the magnetometer data.

The method will be applied to selected details $D_{s,j}$ because, unlike the MODWT coefficients $W_{s,j,t}$, times $t$ in the details $D_{s,j,t}$, are aligned with times $t$ in the observations.
such a time alignment is not necessary in the removal of the noise which does not have a regular pattern over time. Instead, the wavelet domain equation (4.8) was exploited.

Exploratory analysis of the details from four Dst stations, showed that the periodicity is clearly visible in the $D_{s,j}$ for $j = 8, 9, 10$ and not visible by eye at other scales. This finding agrees with the known properties of the daily variation and the spectral properties of the LA(8) wavelet filter. The daily variation, see Section 4.4 of [37], has the spectrum dominated by peaks at frequencies of 1, 2 and possibly 3 cycles per day. These frequencies correspond to wavelet levels $j = 10, 9, 8$, respectively, as already discussed in Section 4.3. Figure 4.1 shows that the the squared gain functions $H_{j,t}(f)$, $j = 8, 9, 10$, of the LA(8) filter together completely cover the interval $[0.0005, 0.0040]$ which contains practically all frequencies present in the spectrum of the daily component, including the daily frequency of $1/1440 \approx 0.0007$ and the half-daily frequency of about 0.0014 (all frequencies are in cycles per minute).

As an example, Figure 4.2 shows the details $D_j$ for levels 8 through 10 for station Kakioka for March-April of 2001. As we can see, the largest contribution to the periodic component comes from the “slow-oscillating” details capturing the daily frequency (bottom panel) and the smallest from the “fast-oscillating” $D_8$.

Recall that our goal is to extract a periodic daily component from the data. We propose to use a robust median-based filter which is analogous to the usual method of removing the periodic component, see e.g., Section 1.5 of [11], but uses the median instead of the average. Unlike the average, the median is not sensitive to unusually large or small observations. Moreover, the wavelet coefficients of magnetometer data have heavy tails, see [28]. Discussing this property here would distract us from the focus of this paper, so we merely note that for heavy-tailed observations, median-based procedures generally work better than procedures based on averaging.

In our context, using the median rather than the average produces an index which is closer to the Dst index. To describe our method, suppose $Y_0, Y_1, \ldots, Y_{N-1}$ is a time series
without a trend. Here, it should be thought of as the time series of details at a given level \( j \) from which we wish to remove the periodic component. Recall from Section 4.2 that the MODWT coefficients and the details do not contain a trend because the slowly varying trend in the data is modeled as a polynomial of order not exceeding 3, and we use wavelet filter of length 8.

Given a series \( Y_0, Y_1, \ldots, Y_{N-1} \) from which a periodic component with period \( d \) is to be removed, we follow these steps:

1: Construct \( R = \lfloor N/d \rfloor \) sequences

\[ Y_t, Y_{t+1d}, Y_{t+2d}, \ldots, Y_{t+(R-1)d}, \quad t = 0, 1, \ldots, d - 1. \]

(\( R \) is the number of consecutive, non-overlapping sequences of length \( d \) that "fit into" the sequence of length \( N \).)

2: For each such sequence compute the median \( m_t, t = 0, 1, \ldots, d - 1 \).

3: For \( t > d - 1 \), write \( t = Kd + u \), for some integers \( K \geq 1 \) and \( 0 \leq u \leq d - 1 \), and set \( m_t = m_u \). (In this step the sequence \( m_t, t = 0, 1, \ldots, d - 1 \), is extended periodically to a sequence of length \( N \).)

4: Compute the mean \( \mu = N^{-1} \sum_{t=0}^{N-1} m_t \) and return sequence \( \{Y_t - (m_t - \mu)\} \). (This step ensures that the estimated periodic component \( m_t - \mu, t = 0, 1, \ldots, N - 1 \), has mean zero.)

In the estimation of the periodic component we include the boundary coefficients, but since the above procedure is robust to atypical observations, it is not affected by these coefficients.

In Section 4.5, to remove the daily variation, we apply the above procedure to the details \( D_{s,8}, D_{s,9}, D_{s,10} \) with periods \( d = 480, 770, 1440 \), respectively.

4.5 Index Algorithm

We now describe the complete algorithm for obtaining an index of storm activity which incorporates the statistical procedures described in the previous sections. In light of the
motivation and objectives outlined in Section 4.1, our primary focus is on an algorithm which uses two months worth of data, but we also show how to construct an index over a period of one year.

Recall that for $s = 1, 2, \ldots, S$, $X_s = \{X_{s,t} : t = 0, 1, \ldots, N - 1\}$ is a magnetometer data at station $s$.

1) If $N$ corresponds to two months set $J = 10$, if $N$ corresponds to one year, set $J = 19$.
In both cases, set $J_0 = 7$. Compute the MODWT coefficients $W_{s,j,t}$, $j = 1, 2, \ldots, J$, and $V_{s,j,t}$;

2) Threshold the wavelet coefficients $W_{s,j,t}$, $j = 1, 2, \ldots, J_0 = 7$, to obtain $\overline{W}_{s,j,t}$ according to Section 4.3, and using equation (4.5) compute the details $\overline{D}_{s,j,t}$ by replacing $W_{s,j,t}$ with $\overline{W}_{s,j,t}$;

3) Compute the details $D_{s,j}$, $j = 8, \ldots, J$ and the smooth $S_{s,j}$. To the details $D_{s,8}$, $D_{s,9}$, $D_{s,10}$, apply median-based filter of Section 4.4 with periods $d = 480, 770, 1440$, respectively, and denote the resulting details without the periodic components by $\overline{D}_{s,j}$, $j = 8, 9, 10$;

4a) If $N$ corresponds to two months, compute the filtered magnetogram from which the periodic and noise components have been removed:
$$\hat{A}_{s,t} + \hat{B}_s = \sum_{j=1}^{7} \overline{D}_{s,j,t} + \sum_{j=8}^{10} \overline{D}_{s,j,t} + S_{s,10,t}$$
$$= X_{s,t} - \hat{P}_{s,t} - \hat{L}_{s,t} - \hat{N}_{s,t};$$

4b) If $N$ corresponds to one year ($J = 19$), replace the smooth $S_{s,j}$, which reflects the long-term (secular) component, by its average $N^{-1} \sum_{t=0}^{N-1} S_{s,j,t} = N^{-1} \sum_{t=0}^{N-1} X_{s,t}$ and compute the filtered magnetogram from which the trend, periodic and noise components have been removed:
$$\hat{A}_{s,t} + \hat{B}_s = \sum_{j=1}^{7} \overline{D}_{s,j,t} + \sum_{j=8}^{10} \overline{D}_{s,j,t} + \sum_{j=11}^{19} \overline{D}_{s,j,t} + \frac{1}{N} \sum_{t=0}^{N-1} X_{s,t}$$
$$= X_{s,t} - \hat{P}_{s,t} - \hat{L}_{s,t} - \hat{N}_{s,t};$$
5) Compute the average of the filtered magnetogram

\[ \tilde{M}_s = \frac{1}{N} \sum_{t=0}^{N-1} (\hat{A}_{s,t} + \hat{B}_s) = \left\{ \frac{1}{N} \sum_{t=0}^{N-1} \hat{A}_{s,t} \right\} + \hat{B}_s \]

and center the filtered data to have mean zero:

\[ \hat{A}^c_{s,t} = (\hat{A}_{s,t} + \hat{B}_s) - \tilde{M}_s = \hat{A}_{s,t} - \frac{1}{N} \sum_{t=0}^{N-1} \hat{A}_{s,t}. \]

(This step eliminates, in particular, the constant field \( B_s \).)

6) Adjust for the magnetic latitude:

\[ \hat{A}^{c*}_{s,t} = \hat{A}^c_{s,t} / \cos(\theta_s) \]

(\( \theta_s \) is the magnetic latitude of station \( s \).)

7) Average over all stations to obtain a global index:

\[ I_t = \frac{1}{S} \sum_{s=1}^{S} \hat{A}^{c*}_{s,t}, \quad t = 0, 1, \ldots, N - 1. \]

To facilitate the discussion below, we refer to the index \( I_t \) as the *Wavelet Index of Storm Activity* or WISA.

*Comments.*

1) The last two steps of the algorithm are the same as for the standard Dst index, see [27], p. 457. (In the construction of the Dst index \( \hat{A}^c_{s,t} \) are the data with \( B_s, L_s, t \) and \( P_s, t \) manually removed.)

2) In order to obtain the WISA over a given period of time, only the data from that period are used. For example, to construct an index over a period of two months, no observations outside that period are needed. This feature offers the flexibility of constructing the index only from new data or of using a moving window of suitably chosen length.

3) The WISA reflects only the dynamic range of storm activity over a given period of time and does not provide any reference level relative to the past. If the WISA is
constructed for a period of one year, its mean over that year is zero, but the means over any subperiods of that year need not be zero. If the index is constructed over two months, its mean over these two months is zero.

4) The WISA can be constructed practically over any period of time, but to obtain values similar to the standard Dst, at least two months worth of data are needed. The periods of two months and one year were chosen for illustration only. Note that if $I_t^{1y}$ is an index constructed from one year worth of data and $I_t^{2m}$ is an index constructed only from, e.g., March-April data, then over the months of March and April the indices $I_t^{1y}$ and $I_t^{2m}$ will differ slightly.

5) Once the length of the period of time over which the WISA is to be computed is set, the procedure is fully automatic and requires merely one minute measurements of the horizontal intensity as input.

6) Even quality one-minute data from INTERMAGNET CDs which we used contain some missing values. We used linear interpolation to estimate these missing values (magnetometer data from Honolulu: 2% in 2001 and 0.1% in March-April of 2001; San Juan: 2% in 2001 and 0.3% in March-April of 2001). Linear interpolation gives slightly different values of the WISA, the difference is however not perceptible by eye on a graph showing both indices over a period of two months.

The data we worked with had only occasional stretches of a few missing values. The presence of long stretches of missing values may be a serious problem. This issue and the use of data from non-Dst stations will be explored in future work.

7) The WISA is a one-minute index, so in order to compare it to the standard Dst, some averaging is needed. This issue is taken up in greater detail in Section 4.6.

8) The algorithm described in this section can be modified in several ways while preserving the general idea. For example, different thresholding levels could be used and the removal of the periodic component could be extended to level $j = 7$. Values of
in the range 0.90 to 0.99 give only a negligibly different index. At levels \( j = 8, 9 \), a component with period of 1 day could be removed rather than components with periods of 8 and 12 hours, respectively. All these modifications would produce only a slightly different index. The algorithm we settled on seems most logical.

9) Naive removal of the details containing the noise and periodic component rather than statistical filtering produces highly oversmoothed storm events because the information about the storms is contained in all levels and must be extracted with care.

10) For one year data, the long-term trend at a given station is estimated by the smooth \( S_{a,j} \), see step 4b) of the algorithm. Since our goal is to produce an index of storm activity, \( S_{a,j} \) was replaced by its average.

11) It is well-established that the pattern of the daily variation changes with season and undergoes other long term changes. We therefore see the primary application of the proposed method in computing the index over a window of two months. Assuming daily periodicity of \( P_{s,t} \) over the period of one year is more questionable, but the WISA computed over one year is still remarkably close to the standard Dst. A detailed comparison is presented in Section 4.6.

12) The complexity of MODWT is \( O(n \log_2(n)) \). On Sun V2-40, with two 1.28 GHz processors and 8192 MB of RAM, it takes slightly under an hour to obtain WISA for \( N = \) one year, and slightly under 3 min for \( N = 2 \) months.

4.6 Comparison to Dst

The objective of this section is to compare the new WISA index to the Dst index. There are several sources of differences:

1) WISA and Dst are calculated in fundamentally different way.

2) Dst is an hourly index and WISA is a one-minute index. To obtain hourly values of WISA we used one-minute averages from INTERMAGNET's CDs (which are them-
selves averages of raw data recorded with 5 s frequency) and then averaged the sixty
values of one-minute WISA following the time for which Dst is reported.

3) Even the best quality records contain missing values. We interpolated the missing
via linear interpolation. A different interpolation technique would produce slightly
different WISA values. We do not know how missing values are handled to produce
Dst.

Ideally, we would like to compare WISA to Dst with respect to differences arising from the
different methodology, i.e., from source 1). This source cannot, however, be separated from
sources 2) and 3) and keeping this in mind we will argue that the contribution of source 1)
is negligible, i.e., over a period of one year our technique can automatically reproduce Dst
up to the accuracy determined by the preprocessing of the raw data.

In the comparison below we use the final one-hour Dst index and WISA computed from
one-minute values of the H-component for Hermanus, Honolulu, Kakioka, and San Juan
(Dst stations). We consider the time period January-December, 2001, which contains a
few very strong storms with the dynamic range approaching 500 nT. By construction, the
average of the WISA is zero, cf. comment 3) in Section 4.5. The Dst is computed over longer
periods of time and its average is not zero for any subperiod (but is close to zero). To make
the comparison possible, we subtracted the average value of the Dst for the periods over
which we compare it to WISA. Specifically, when using the Dst over the whole year 2001,
we added 18 nT to it to make its average for 2001 zero. For the subperiods of two months,
different constants had to be added. For example, for the period of March-April, which
contains a very strong storm, we added 31 nT to the Dst to make its average over this two
month period zero.

Figure 4.3 shows the Dst index (bottom panel) and the hourly averages of WISA (top
panel). The two indices look very similar. In Figure 4.4, the differences between Dst and
WISA are plotted and are seen to be generally smaller than 10 nT. These differences tend
to be slightly larger during disturbed period (Dst < -30 nT) than during quiet periods.
(Dst ≥ -30 nT). This point is illustrated in Figure 4.5. The physical interpretation of the larger spread of differences for disturbed conditions requires additional study with larger data sets, but it can be conjectured that our approximation of the diurnal component is more accurate during quiet periods; during disturbed periods this component appears less stable. The differences for both the disturbed and calm periods have an error-type distribution, indicating that they may be attributable to chance errors rather than a systematic bias.

As explained above, the averaging of WISA to obtain an hourly index introduces some arbitrary variability. To assess this variability, we computed the differences

\[ \Delta I_t = I_t - I_{t-1}, \quad t = 1, 2, \ldots, \lfloor N/60 \rfloor - 1. \]

The sample standard deviation of these differences, denoted \( s_{\Delta I} \), is a measure of variability due to averaging over one hour intervals. Since the differences are approximately normally distributed, the \( 100(1 - \alpha)\% \) confidence interval for the hourly WISA can be defined as

\[ (I_t - z_{1-\alpha/2}s_{\Delta I}, I_t + z_{1-\alpha/2}s_{\Delta I}), \]

where \( z_\beta \) denotes \( \beta \)th quantile of the standard normal distribution. For the whole of 2001, \( s_{\Delta I} = 5.20 \) nT, and 97.72\% of the Dst values are within the 95\% confidence limits. Figure 4.6 shows the 95\% confidence intervals together with the two indices for the first week of January. The graphs have a similar appearance for the remaining weeks of 2001.

We also compared hourly WISA computed over 2 month periods in 2001 to the Dst. Graphs for the most stormy period, March-April, are presented in Figures 4.7-4.10. For this period, all Dst values are within the 95\% confidence limits. The increasing spread of differences visible in Figure 4.8 is due to a very strong storm in the middle of that period. As for the WISA computed from one year worth of data, the variability due to the averaging of the one-minute WISA is statistically greater than the differences between the hourly WISA and the Dst. Perhaps more importantly, the difference between the WISA and the Dst, which is typically about 5 nT, is negligible compared to the dynamic range of a storm which is several hundred nT.
4.7 Conclusions

We propose an automatic procedure for calculating an index of geomagnetic storm activity which uses scale time-dependent decomposition provided by the Maximum Overlap Discrete Wavelet Transform and statistical filtering techniques. We use the wavelet thresholding to remove the noise and a median-based filter to remove the periodic component. The procedure produces a one-minute index and requires only a recent short stretch of data. It replaces the Dst determination of quiet days involving subjective human selection by robust statistical estimation of the diurnal component. When the H-components of the four Dst stations are provided as input, the hourly averages of the new index are statistically indistinguishable from the Dst values.

The procedure can take magnetometer data from different stations as input thus allowing the study of the effect of the choice of the stations on the index of storm activity. Such a study is, however, beyond the scope of this work. The proposed procedure is flexible and amenable to various modifications and can potentially serve as a useful tool in the study of the magnetic storms. Its full automation has a substantial potential for operational purposes.

Research is ongoing to investigate both the short-term, storm-related wavelet decompositions and the long-term seasonal and solar cycle dependencies of the other components in equation (4.2).
Figure 4.1. Squared gain functions $\tilde{H}_{j,L}(f)$ for LA(8) filter, levels $j = 8, 9, 10$. Nominal pass-bands are indicated by solid vertical lines. Periodic frequencies $1/1440, 2/1440, 3/1440$ are marked by dashed, dotted and dashed-dotted lines, respectively.

Figure 4.2. Details from levels $j = 8, 9, 10$ for station Kakioka, March-April, 2001. Bar in the bottom left-hand side corner of the bottom panel spans one day.
Figure 4.3. Hourly values of WISA and Dst, January-December, 2001.

Figure 4.4. Differences between Dst (+18 nT) and WISA, January-December, 2001.
Figure 4.5. Histograms of differences between Dst (+18 nT) and WISA for "quiet" and "disturbed" periods, January-December, 2001.

Figure 4.6. Dst, WISA and 95% confidence intervals for hourly WISA for the first week of 2001.
Figure 4.7. Hourly values of WISA and Dst, March-April, 2001.

Figure 4.8. Differences between Dst (+31 nT) and WISA, March-April, 2001.
Figure 4.9. Histograms of differences between Dst (+31 nT) and WISA for "quiet" and "disturbed" periods, March-April, 2001.

Figure 4.10. Dst, WISA and 95% confidence intervals for hourly WISA for the first week of March.
CHAPTER 5
SUMMARY AND FUTURE RESEARCH

The unique strength of wavelets follows from many appealing properties that can be utilized in the context of highly-correlated time series as well as semi- and nonparametric estimation.

In Chapter 2, we showed how to construct useful wavelet-based confidence intervals for the self-similarity parameter $H$. We focused on the Linear Fractional Stable Motion, which is used as a modeling tool in the high-resolution traffic measurements from modern communication networks. Statistical inference about the self-similarity parameter may help to detect unusual network traffic flow patterns. We examined methods of constructing confidence intervals for $H$, asymptotic versus resampling. The superiority of the latter approach lies in the fact that it does not require (typically unavailable) knowledge about the tail index $\alpha$. Although parameters such as subsample or block size in resampling methods must be specified, we provide practical recommendations for their values. It was shown that the asymptotic method with variance based on the correct and incorrect, but reasonably chosen $\alpha$, provides quite low empirical coverage probabilities. This method is outperformed by all resampling methods considered. When blocks are appropriately chosen, subsampling method yields coverage probabilities about 1%-5% below the nominal coverage. Block bootstrap method with short blocks provides confidence intervals that are shorter, but undercoverage increases to about 1%-15% below the nominal coverage. Similar empirical coverage can be achieved by the bootstrap method, however without the problem of the block size selection. These three methods provide coverage very close to the nominal values and are much better than the asymptotic method which has been used so far, and do not require the knowledge of the tail index $\alpha$. Application of the resampling confidence intervals, via a procedure for monitoring constancy of $H$, to a important network traffic traces provides new insights about the constancy and level of self-similarity. On contrary to a previous study, which concluded the rejection of constancy in the self-similarity parameter, we show that there is not enough evidence to reject constancy in $H$. Interestingly, the
estimated value of this parameter exceeds the theoretical range of (0, 1) suggesting that we may be dealing with a different type of self-similar process than LFSM, so the conclusions of all studies must be treated with caution.

In Chapter 3, an important and difficult task of discriminating between long-range dependence and weak dependence with nonstationarities was investigated. A wavelet-domain test in which under the null hypothesis the time series is weakly dependent with a polynomial trend and under the alternative hypothesis it is LRD, possibly with a polynomial trend, was developed and studied. It was assumed that the time series $Y_t$ follows an ARMA($p, q$) model under the null and ARFIMA($p, \delta, q$), $\delta > 0$, model under the alternative. A simulation study established that the test has roughly the correct size for moderate weak dependence ($0.1 \leq \phi \leq 0.8$) and performs noticeably better for $N = 1024$ than for $N = 512$. For $N = 512$ the test is conservative for $0.1 \leq \phi \leq 0.6$. Overall, the test has very good power.

The test can be applied if a piecewise polynomial trend is suspected, provided there are few break points relative to the length of the series. Application of the test to annual minima of the Nile River confirmed the presence of long-range dependence in this time series.

Flexibility in decomposing the signal, across various frequencies, but without the loss of location in time, via wavelets was demonstrated in Chapter 4, where a new index of magnetic storm activity was developed. In the procedure, the extraction of the daily $Sq$ variation is achieved by deseasonalizing the details associated with the daily and half-daily frequencies, and is free of human input. Deseasonalizing is robust due to the use of median-rather than mean-based filter. Specifications of the wavelet analysis are chosen to address the issues associated with the nature of the magnetogram data. Magnetogram readings from 2001 were used to compare the traditional Dst index with the new wavelet-based index. The comparison showed that the differences are negligible compared to the dynamic range of a storm.

A possible extension of the topics covered by this dissertation would be an investigation of the estimation techniques from Chapter 3 in case of an ARFIMA process with stable innovations. Chapter 3 poses many theoretical problems related to wavelet domain approx-
imation of time domain likelihoods. In Chapter 4, it would be interesting to see if it is possible to modify the wavelet-based procedure to estimate the essential contributor of the magnetic storm activity, i.e., the ring current. In other words, whether it would be feasible to remove the effects of auroral current systems, that intensify during the magnetic storms, and extract the so-called symmetric ring current.
REFERENCES


March 30th 2006
Agnieszka Jach
Department of Mathematics and Statistics
3900 Old Main Hill
Logan, UT 84322-3900
Phone: (435)-764-0591

Dear Jan Sojka:

I am in the process of preparing my dissertation in the Department of Mathematics and Statistics at Utah State University. I hope to complete my degree in the spring of 2006. I am requesting your permission to include the attached material as shown. I will include acknowledgements and/or appropriate citations to your work as shown and copyright and reprint rights information in a special appendix. The bibliographical citation will appear at the end of the manuscript as shown. Please advise me of any changes you require.

Please indicate your approval of this request by signing in the space provided, attaching any other form or instruction necessary to confirm permission. If you charge a reprint fee for use of your material, please indicate this as well. If you have any questions, please call me at the number above.

I hope you will be able to reply immediately. If you are not the copyright holder, please forward my request to the appropriate person or institution.

Thank you for your cooperation,

I hereby give permission to Agnieszka Jach to reprint the following material in her dissertation.

Chapter 4 of this dissertation based on manuscript by A. Jach, P. Kokoszka, J. Sojka, L. Zhu, Wavelet-based index of magnetic storm activity.

Signed
CURRICULUM VITAE

Agnieszka Jach

EDUCATION:

MS (Mathematics, Financial and Actuarial Mathematics) 1996-2001, Wroclaw University of Technology, Poland (Aleksander Janicki)
PhD (Mathematical Sciences, Statistics) 2002-2006, Utah State University, USA (Piotr Kokoszka)

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, WAMI (Wavelet Analysis of Magnetosphere-Ionosphere) Project, 2004-2006, Utah State University
Graduate Teaching Assistant, 2002-2005, Utah State University
Graduate Tutor and Grader, 2002, Utah State University
Graduate Teaching Assistant, 2001, Wroclaw University of Technology

PUBLICATIONS: