A WAVELETS BASED APPROACH FOR TIME SERIES MINING

Teză destinată obţinerii
titlului ştiinţific de doctor inginer
la
Universitatea "Politehnica" din Timişoara
în domeniul Inginerie Electronică şi Telecomunicaţii
de către

Ing. Cristina STOLOJESCU

Conducător ştiinţific:	prof.univ.dr.ing. Alexandru ISAR conf.univ.dr. Philippe LENCA
Referenţi ştiinţifici:	prof.univ.dr.ing. Monica BORDA prof.univ.dr. Lotfi SENHADJI prof.univ.dr. Ioan NAFORNITA prof.univ.dr. Thierry DHORNE prof.univ.dr.ing. Corneliu RUSU ing. Adrian IVAN
Invitat:	conf.univ.dr.ing. Sorin MOGA

Ziua susţinerii tezei: 13.01.2012

Seriile Teze de doctorat ale UPT sunt:

1. Automatică
2. Chimie
3. Energetică
4. Ingineria Chimică
5. Inginerie Civilă
6. Inginerie Electrică
7. Inginerie Electronică şi Telecomunicaţii
8. Inginerie Industrială
9. Inginerie Mecanică
10. Ştiinţa Calculatoarelor
11. Ştiinţa şi Ingineria Materialelor

Universitatea „Politehnica" din Timişoara a iniţiat seriile de mai sus în scopul diseminării expertizei, cunoştinţelor şi rezultatelor cercetărilor întreprinse în cadrul şcolii doctorale a universităţii. Seriile conţin, potrivit H.B.Ex.S Nr. 14 / 14.07.2006, tezele de doctorat susţinute în universitate începând cu 1 octombrie 2006.

România, 300159 Timişoara, Bd. Republicii 9,
tel. 0256 403823, fax. 0256 403221
e-mail: editura@edipol.upt.ro

Acknowledgements

I would like to use this opportunity to express my gratitude to a number of people who over the years have contributed in various ways to the completion of this work.

First, I would like to thank my scientific supervisor, Professor Alexandru Isar from "Politehnica" University of Timişoara, for all his help and contributions to this thesis. His encouragements, patience and guidance throughout my research were extremely important for me to get my work up to this point.

I would also like to thank my co-supervisors, Associate Professor Philippe Lenca and Associate Professor Sorin Moga from TELECOM Bretagne, for their useful adivices, guidance and support and for giving me the opportunity to work in the LUSSI Department, at TELECOM Bretagne, Brest, France.

I am extremely thankful to Professor Ioan Naforniţă from the Communications Department, Electronics and Telecommunications Faculty, "Politehnica" University of Timişoara who gave me the opportunity to be a member of the department and who continuously supported me since the beginning of my career.

The Communications Department of Electronics and Telecommunications Faculty, "Politehnica" University of Timişoara and LUSSI Department of TELECOM Bretagne, Brest, France, have provided me a wonderful working atmosphere. I spent many enjoyable hours with the members of these two departments and I thank them all for creating perhaps the most enjoyable academic environment.

This study received funding from the project no. 349/13.01.09, "Using Wavelets Theory for Decision Making" supported by the Romanian Research Council (CNCSIS). I am also greatful to Alcatel Lucent, Timişoara, for providing the WiMAX traffic database and for helpful discussions around WiMAX network.

Many thanks to the members of the thesis committee for accepting to be part of my thesis jury and for their valuable assistance.

Finally, I would like to thank my family for all the love, encouragements and interest in my career. This thesis is dedicated to them.

Timişoara, december 2011 Cristina STOLOJESCU

Familiei mele.

Stolojescu, Cristina

A Wavelets Based Approach for Time Series Mining

Teze de doctorat ale UPT, Seria 7, Nr. 43, Editura Politehnica, 2012, 108 pagini, 41 figuri, 12 tabele.

ISSN: 1842-7014

ISBN: 978-606-554-418-5

Keywords: time-series, wavelets, data mining, WiMAX.

Abstract:

This thesis has a practical objective. It consists in finding an answer to the following question:*"It is possible to identify the base stations (BS) which are bad positioned in a Worldwide Interoperability for Microwave Access (WiMAX) network topology by traffic analysis?"* The question is important for the planning and exploitation of WiMAX networks. This thesis is based on the research of time series analysis.

Time series analysis has been an area of considerable activity in recent years. The work of this thesis evaluates a set of time series conceived by monitoring the traffic developed in a WiMAX network, composed by 67 BSs, for a time interval of eight weeks, developed by Alcatel Lucent Timisoara, Romania. Taking into consideration the high volume of information contained in this database, a data-mining approach was preferred. Assuming that the traffic associated with a BS bad positioned is heavier than the traffic associated with a BS well positioned, two approaches for the appreciation of the heaviness of the traffic were developed. The first approach is based on the supposition that a BS with heavy traffic has a reduced risk of saturation. Hence, it is necessary to appreciate the risk of saturation of each BS. This is equivalent with the estimation of the moment when the BS will saturate. So, the first objective of this thesis is to propose an approach for predicting time series. This approach is based on a multiple resolution decomposition of the signal using the Stationary Wavelet Transform (SWT) and Autoregressive integrated moving average (ARIMA) model. The second approach for the appreciation of the heaviness of the traffic is based on Long Range Dependence (LRD) analysis. This is a relative new statistical concept in communication traffic analysis and can be implemented using wavelets as well. The estimation of LRD degree is realized through the estimation of the Hurst parameter of the time-series under analysis. This property has important implications on the performance, design and dimensioning of the network. Our objective is to highlight the particularities of WiMAX traffic from a LRD perspective and to analyze the positioning of BSs in the architecture of a WiMAX network. The results show which BSs have a good localization in the topology of the network and which BSs have a bad localization in the topology of the network and must be repositioned when the next session of network maintenance will take place.

The application of both data mining techniques (forecasting and LRD analysis) in the wavelets domain is decisive for their performance, improving the speed and the precision of the developed algorithms.

Contents

List of Figures

List of Tables

Abbreviations

- ACF - AutoCorrelation Function
- AIC - Akaike Information Criterion
- ANOVA - Analysis of Variance
- AR - AutoRegressive process
- ARMA - AutoRegressive Moving Average
- ARIMA - AutoRegressive Integrated Moving Average
- BIC - Bayesian Information Criterion
- BS - Base Station
- CRISP-DM - Cross Industry Standard Process for Data Mining
- CWT - Continuous Wavelet Transform
- DM - Data Mining
- DWT - Discrete Wavelet Transform
- fBm - fractional Brownian motion
- fGn - fractional Gaussian noise
- FPE - Final Prediction Error
- FT - Fourier Transform
- LRD - Long-Range Dependence
- LTE - Long Term Evolution
- MA - Moving Average process
- MLE - Maximum Likelihood Estimation
- MRA - Multiresolution Analysis
- MW - Mother Wavelets
- PACF - Partial AutoCorrelation Function
- QoS - Quality of Service
- STFT - Short-Time Fourier Transform
- SWT - Stationary Wavelet Transform
- UDWT - Undecimated Discrete Wavelet Transform
- WGN - White Gaussian Noise
- WT - Wavelet Transform
- WiMAX - Worldwide Interoperability for Microwave Access

Introduction

This thesis has a practical objective. It consists in finding an answer to the following question: "It is possible to identify the base stations (BS) which are bad positioned in a Worldwide Interoperability for Microwave Access (WiMAX) network?" The question is important for the planning and exploitation of WiMAX networks. The answer could be an explanation of the reasons for which the performance of WiMAX networks measured in practice is inferior to the performance estimated in the designing phase.

Our approach is based on the traffic analysis in a WiMAX network, composed by sixty-seven BSs, for a time interval of eight weeks. Taking into consideration the high volume of information, a data-mining approach was preferred. It is based on the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. This methodology is applied in the present thesis to extract information, to interpret it and to propose solutions. The selection of traffic as object of analysis is justified by the following reasons:

1. It can be measured,
2. Using planning and exploitation strategies, it allows the increasing of the performance of the network (especially in the case of wireless communications such as the WiMAX technology).

This selection oriented the thesis toward the research of time series analysis methods. Time series analysis has become a challenging issue for many researchers. Its origins can be found in mathematical research but today time series analysis is a multi-disciplinary field exploiting results obtained in mathematics, statistical signal processing, data mining, or engineering. This is the reason why the present thesis has a multi-disciplinary character as well, integrating the competences in informatics from the department LUSSI of Telecom Bretagne, Brest, France with the competences in communications from the Communications department of the Electronics and Telecommunications Faculty of "Politehnica" University from Timisoara, Romania.

One of the major difficulties of the analysis of time-series with long length (which correspond to large amount of data) is the big computational complexity involved. The computational complexity can be reduced by representing the data in a more favorable form. One of the phases of the CRISP-DM methodology, namely data preparation, supposes data representation in a more favorable form. Such a representation can be obtained using wavelets. A discrete wavelet transform (DWT) of the time-series is sparse and involves a reduced computational complexity. The Wavelet Transform has been used for time series analysis in many papers in recent years [2], [54], [56], [57], [65], [70]. One of the main properties of wavelets is that they are localized in time (or space) which makes them suitable for the analysis of non-stationary signals (signals containing transients and fractal structures).

The research framework associated with the present thesis has the following axes:

1. Wavelets - an introduction is presented in Chapter 1,
2. Statistical signal processing - basic tools are presented in Chapter 2,
3. Time-series analysis - performed in Chapter 3 and Chapter 4,
4. Data-mining - performed in Chapter 3 (where the CRISP-DM methodology is developed and highlighted) and Chapter 4,
5. WiMAX networks - described in Chapter 3 and analyzed in Chapter 3 and Chapter 4.

As it was already said, the goal of this thesis is to answer the question "It is possible to identify the BSs which are bad positioned in a WiMAX network topology by traffic analysis?" Assuming that the traffic associated with a BS bad positioned is heavier than the traffic associated with a BS well positioned, two approaches for the appreciation of the heaviness of the traffic were developed. The first approach is based on the supposition that a BS with heavy traffic has a reduced risk of saturation. Hence, it is necessary to appreciate the risk of saturation of each BS. This is equivalent with the estimation of the moment when the BS will saturate. So, the first objective of this thesis is to propose an approach for predicting time series. There are two types of prediction on short term and on long term. Both can be done in the wavelets domain with the aid of a multiple resolution decomposition of the signal using the Stationary Wavelet Transform (SWT). It is followed by an Autoregressive Integrated Moving Average (ARIMA) modeling in the case of long term prediction or by the utilization of Neural Networks (NN) in the case of short term prediction. These two types of prediction were compared in some companion papers which were elaborated in the department LUSSI from Telecom Bretagne, showing the superiority of NNs for short term prediction. Taking into consideration the fact that the moment of saturation could be situated far in the future we have preferred in this thesis the long term prediction approach based on ARIMA. Applied to all the traces from our database, this approach allowed a first classification of BSs from the heaviness of traffic point of view, presented at the end of Chapter 3.

The second approach for the appreciation of the heaviness of the traffic is based on Long Range Dependence (LRD) analysis. This is a relative new statistical concept in communication traffic analysis and can be implemented using wavelets as well. LRD is introduced in Chapter 2 in association with some of its estimators. The estimation of LRD degree is realized trough the estimation of the Hurst parameter of the time-series under analysis. The LRD analysis of WiMAX traffic is presented in Chapter 4. It can be assumed that a heavier traffic has a stronger LRD. This property of traffic has important implications on the performance, design and dimensioning of the network. By performing simulations and analysis, our results demonstrate that WiMAX traffic exhibits LRD behavior. The objective of Chapter 4 is to highlight the particularities of WiMAX traffic from a LRD perspective and to classify the BSs based on the heaviness of traffic. This second classification of the BSs, presented at the end of Chapter 4, is in agreement with the first classification presented at the end of Chapter 3, despite the fact that both classifications were performed by statistical estimations. For this reason, the response to the question generic for the present thesis is affirmative; the bad positioned BSs can be identified by traffic's analysis. The results show which BSs have a good localization in the topology of the network and which have not. These BSs must be repositioned when the next session of the network's maintenance will take place.

The results that we will present in this thesis are both of theoretical and practical nature. Between the theoretical results could be mentioned the following:

1. The second order statistical analysis of the wavelet coefficients presented in Chapter 2, which is original, allowed to give an elegant explanation of the Abry-Veitch estimator of the Hurst parameter. The estimation of the Hurst parameter is necessary to detect the presence of LRD in a time series. The same statistical analysis is at the basis of another theoretical contribution of this thesis, the Hurst parameter estimator based on wavelets which works for wide sense stationary input time-series proposed in Chapter 2. This is an original estimator which is very simple to be used but it has a limited applicability because the class of wide sense stationary random processes has a reduced degree of generality.
2. A new test of stationarity based on the reiteration of the Box-Jenkins methodology proposed in Chapter 3. Its utility is highlighted by comparisons with the classical stationarity tests based on the correlation or partial correlation functions. These theoretical results are not very general but could represent starting points for future research.

Between the practical results of the thesis we mention the following:

1. The adaptation of an algorithm previously published for the forecasting of wireless traffic time-series, presented in Chapter 3,
2. The identification of the best wavelet transform for traffic forecasting and of its best features in Chapter 3,
3. The selection of the best Hurst parameter estimators based on simulations presented in Chapter 4,
4. The comparative analysis of the results presented at the ends of Chapters 3 and 4.

The Matlab® codes required for the implementation of the estimation methods described in Chapters 3 and 4 represent personal contributions of the author of the thesis.

Chapter 1

Wavelet Transforms

The purpose of this chapter is to give a short introduction to the wavelet transforms and the wavelet families which will be used in the following sections. The transform of a signal is another form to represent it. It does not affect the information carried by the signal. In this context, a wave (see Figure 1.1, left) is an oscillating periodic function of time or space. In contrast, wavelets (see Figure 1.1, right) are localized waves. They have their energy concentrated in time or space and are suited to analysis of transient signals. While Fourier transform uses waves to analyze signals, the wavelet transform uses wavelets of finite energy.

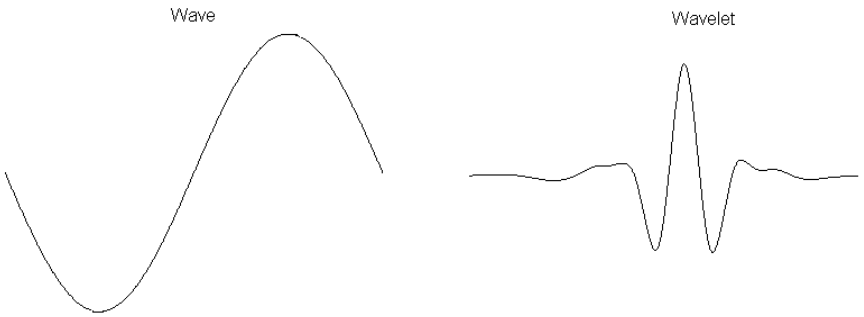

Figure 1.1: Plots of a wave and of a wavelet.

The wavelet theory deals with the properties of wavelets. It is a relatively new mathematical tool which appeared around 1980 when Grossman and Morlet [33], a physicist and an engineer, broadly defined wavelets in the context of quantum physics. Based on physical intuition, these two researchers provided a new way of thinking for wavelets based on physical intuition.

In 1985, Stephane Mallat [44] gave wavelets an additional jump-start through his work in digital signal processing. He discovered some relationships between quadrature mirror filters, pyramid algorithms, and orthonormal wavelet bases. An orthonormal wavelet basis of a given Hilbert space is an orthonormal basis of this space whose elements are obtained by translations with integers of a unique function named mother wavelets.

A couple of years later, in 1988, Ingrid Daubechies [18] used Mallat's work to construct a set of wavelet orthonormal basis functions that are perhaps the most

elegant, and have become the cornerstone of wavelet applications today.

Wavelet theory is used for analyzing various data studied in various domains such as mathematics [69], science [38], engineering [63], economics [28] and social studies: time series (as will be shown in the following sections of this thesis), radar signal [42], image [10], sound [13], video, mathematical functions, etc.

1.1 The Wavelet Transform

In the following we will present the main steps in the evolution of the wavelet transform (WT). As already said, the transform of a signal is nothing more than another form of representation of that signal. We will consider as starting point the Fourier transform. It is an alternative representation of a signal in the frequency domain. It has various forms: the Fourier series used for the representation of periodic signals, the Fourier transform in discrete time used for the representation of discrete in time signals, the short time Fourier transform which is a time-frequency representation and so on. Accordingly, there are different WTs, the wavelet series, the discrete WTs, the continuous WT.

1.1.1 Fourier Transform

Fourier series are named in honor of the french mathematician and physicist Joseph Fourier (1768-1830), who made important contributions to the study of trigonometric series. In 1807, Fourier presented a memoir [25] to the "Institut de France" in which he claimed that any periodic signal could be represented by a series of harmonically related sinusoids.

The concept can be extended to the Fourier transform (FT), which applies to aperiodic signals. The development of this representation for aperiodic signals in continuous time is one of Fourier's most important contribution, [51]. FT is a mathematical tool used to transform a signal from time-domain into frequency-domain. Being given a signal $x(t)$, the FT, or the spectrum of this signal, $X(\omega)$, is defined as:

$$X_{FT}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt, \tag{1.1}$$

while the inverse FT is given by:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega. \tag{1.2}$$

The difference between the FT and the Fourier series is the following: a Fourier series can only be applied to *periodic signals* and separates them into a number of discrete frequency components, while the FT can be used to break *aperiodic signals* into an infinite number of continuous frequency components using the integral, [7].

1.1.2 Short-Time Fourier Transform

The FT do not clearly indicate how the frequency content of a signal changes over time. Therefore, the Short-Time Fourier Transform (STFT), or windowed Fourier transform, was introduced. STFT extracts several frames of the signal which can be assumed to be stationary, to be analyzed with a window that moves with time, [4].

The STFT of a signal *x(t)* is defined as:

$$X_{STFT}(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t}dt, \tag{1.3}$$

where *w(t)* is the window function and $X(\tau,\omega)$ is the FT of $x(t)w(t-\tau)$, a complex function representing the phase and the magnitude of the signal over time and frequency.

The time resolution and frequency resolution of a STFT basis element is equal to those of the window. Narrow windows give good time resolution, but poor frequency resolution. Wide windows give good frequency resolution, but poor time resolution and may also violate the condition of stationarity, for signals which are stationary on portions. The effect of the selection of a window too long will be the smoothening of the analyzed signal and the information contained in its parts with rapid variations will be recovered with difficulty from its STFT. So, the window should be carefully chosen because it does not change during the period of analysis. Therefore, the time and frequency resolutions will remain unchanged on the entire duration of the analysis performed using the STFT, these resolutions being imposed by the window selected. A particular case of STFT is the Gabor Transform (1946) [27] which uses a Gaussian window.

1.1.3 Wavelet Transform

The Continuous Wavelet Transform (CWT), introduced by Grossman and Morlet, was developed as an alternative approach to the STFT, to overcome the problem of constant resolution. It is done in a similar way as the STFT, in the sense that the signal is multiplied with a function, the wavelet, similar to the window function in the STFT. The transform is computed separately for different segments of the time-domain. This transform is capable of providing the time and frequency information simultaneously, hence giving a time-frequency representation of the signal.

A wavelet is used to analyze a given function or continuous-time signal at a specified scale. This function plays the role of the window from the case of STFT, but it has a second parameter, additional to the position, the scale. It can be moved to various locations of the signal as shown in Figure 1.2. To highlight the influence of the additional parameter, in Figure 1.3 are represented three wavelets of the same type, having the same position but different scales. Each of these three wavelets allows the analysis of a signal at a different scale by translations across its waveform, obtaining three different representations which will be named in the following scale components. Usually one can assign a frequency range to each scale component. Each one can then be studied with a resolution that matches its scale.

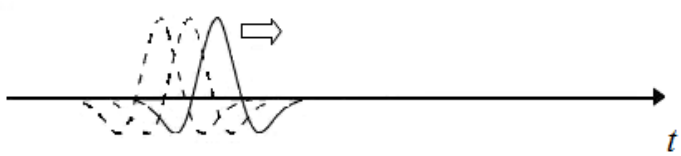

Figure 1.2: Location in time of a wavelet with a given scale.

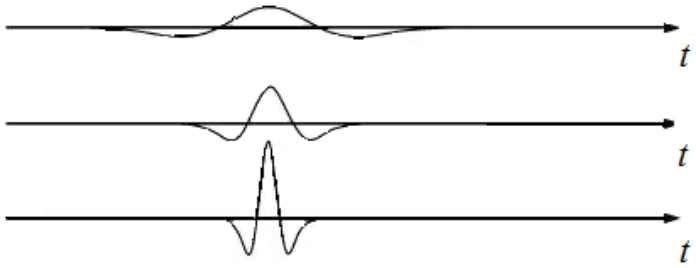

Figure 1.3: Same wavelet at a specified position and different scales.

To analyze signal structures of very different sizes, it is necessary to use time-frequency atoms with different time supports. A linear time-frequency transform correlates the signal with a family of waveforms that are well concentrated in time and in frequency. These waveforms are called time-frequency atoms [24].

The CWT decomposes signals over dilated and translated wavelets. A mother wavelets is a function, $\psi \in L^2(\Re)$, with a zero average:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0, \tag{1.4}$$

normalized ($||\psi|| = 1$), and centered in the neighborhood of $t = 0$.

A family of time-frequency atoms (wavelet functions) $\psi_{u,s}(t)$ are generated by translating and dilating the mother wavelets, ψ:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right), \tag{1.5}$$

that can form a basis. These atoms remain normalized: $||\psi_{u,s}|| = 1$.

1.1.4 Wavelet Transform versus Fourier Transform

Wavelet theory extends the ideas of the traditional Fourier theory. While the FT is useful for analyzing the spectral content of a stationary signal and for transforming difficult operations into very simple ones in the Fourier dual domain, it can not be used for the analysis of non-stationary signals or for real time applications. In this case are required time-frequency representations such as the STFT or the CWT. The CWT is a powerful time-frequency signal analysis tool which it is used in a wide variety of applications including biomedical signal processing, data mining, image compression, pattern recognition, etc. The CWT is one of the most important methods that are used to reduce the noise which perturbs non-stationary signals and to analyze the components of non-stationary signals, for which the traditional Fourier methods cannot be applied directly.

The wavelets have some properties: have good time-frequency (time-scale) localization, can represent data parsimoniously, can be implemented with very fast algorithms and are well suited for building mathematical models of data. The wavelet approach of signal analysis is also flexible in handling irregular data sets. Singularities

and irregular structures often carry essential information in a signal. So, the CWT has advantages over the STFT for representing functions that have discontinuities and sharp peaks, and for accurately decomposing and reconstructing finite, non-periodic and/or non-stationary signals.

The most interesting dissimilarity between these two kinds of transforms is that individual wavelet functions are localized in time. Fourier sine and cosine functions are not. This localization feature, along with wavelets localization in frequency, makes many functions and operators using wavelets "sparse" when transformed into the wavelet domain. This sparseness, in turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series.

Mathematically speaking, the CWT of a signal is a collection of scalar products which factors are the analyzed signal and a family of wavelets, defined in equation (1.5). All these wavelets are generated by translations (see the index u in (1.5)) and dilations (see the index s) of the mother wavelets, ψ . Hence the CWT is a bivariate function, having as variables u and s. One thing to remember is that the CWT has a large set of possible kernels (mother wavelets). Thus wavelet analysis provides immediate access to information that can be obscured by other time-frequency methods such as Fourier analysis.

There are also some similarities between the transforms obtained by the discretization of the CWT and STFT. The discrete transforms obtained by the discretization of continuous transforms are expressed by matrices. The mathematical properties of the matrices involved in the discrete transforms obtained by the discretization of the CWT and STFT are similar. The inverse transform matrix for both the Fast Fourier Transform (FFT) and the discrete WT is the transposed of the original. As a result, both transforms can be viewed as a rotation in functions space. For the FFT, this new domain contains basis functions that are sines and cosines. For the WT, this new domain contains more complicated basis functions called analyzing wavelets [44].

1.2 Time-frequency Representations

Fourier transform theory states that a given function of time can be characterized either in time or in frequency (spectral) domain. The transformation of a signal $x(t)$ between the time domain and the frequency domain can be done by computing the Fourier transform. Fourier transform is indispensable as data analysis tool for stationary signals. But if we deal with non-stationary signals the conventional Fourier transform becomes inadequate.

Time-frequency (time-scale) representation techniques overcome this problem as they are capable of representing a given function of time in both time and frequency domain simultaneously. These kind of representations aim to identify the parameters of a given signal: the starting/ending moments, the energy or the power, the instantaneous amplitude, the instantaneous frequency, the instantaneous frequency band, etc [36].

In Figure 1.4 is presented an ideal time-frequency representation of a given signal $x(t)$, composed by three non-overlapping sinusoids with frequencies in increasing order, each one truncated at its period. The representation is done in three dimensional space having as dimensions the time, the frequency and the amplitude.

This time-frequency representation realizes the perfect localization in time and frequency (the moments of time $t_1 - t_6$ and the frequencies $f_1 - f_3$ can be perfectly

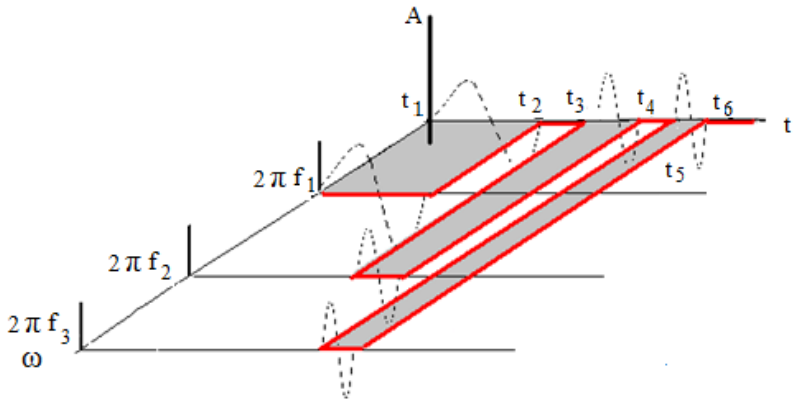

Figure 1.4: An ideal time-frequency representation of the signal $x(t)$.

localized in the time-frequency plane).

The projection of the time-frequency representation on the plane (A, t) represents the oscillogram of the signal $x(t)$ and allows the analysis of this signal in the time domain. The projection of the time-frequency representation on the plane (f, A) represents the ideal spectrum of $x(t)$ and permits us to analyze the signal $x(t)$ in the frequency domain, while the projection on the plane (f, t) represents the instantaneous frequency of $x(t)$ and allows the analysis of $x(t)$ in the modulation domain.

1.2.1 The Effective Duration and Effective Bandwidth

In [36] it is highlighted, based on the duality of the Fourier transform, that signals perfectly localized in time have an unlimited bandwidth, meaning that they are not localized in frequency. As well, band limited signals have an infinite duration. Therefore, to measure these quantities two concepts are used: the effective duration, σ_t, and the effective frequency band σ_ω. A measure of the time-frequency localization of a given signal can be obtained by the product $\sigma_\omega^2 \cdot \sigma_t^2$, [50].

Heisenberg uncertainty principle states that the following inequality is true:

$$\sigma_t^2 \cdot \sigma_\omega^2 \geq \frac{\pi}{2}. \tag{1.6}$$

The shorter is the effective duration of a signal, the wider is its effective frequency band.

In the case of the WT, both time and frequency localizations depend on the scale factor *s*, [36]. The CWT can be stated as a scalar product for every value of the scale factor *s*:

$$W_x(s,t) = \langle x(\tau), \psi_s(\tau - t) \rangle, \qquad \psi_s(\tau) = \sqrt{s\tau} \cdot \psi(s\tau). \tag{1.7}$$

If $\psi(\tau) \in \Re$, we will have:

$$W_x(u,s) = x(u) * \breve{\psi}_s(u), \qquad \breve{\psi}_s(t) = \psi_s(-t). \tag{1.8}$$

Therefore, for every *s*>0 the wavelet transform of a signal *x*(*t*) represents the response of a linear time invariant system at *x*(*t*), having the impulse response $\breve{\psi}_s(t)$.

The system has the frequency response:

$$F\{\breve{\psi}_s(t)\}(\omega) = F\{\psi_s(t)\}(-\omega) = F\{\sqrt{s}\psi(-st)\}(\omega) = \frac{1}{\sqrt{s}} F\{\psi(t)\}\left(\frac{-\omega}{s}\right). \tag{1.9}$$

So, the temporal "window" $\psi_s(t)$ is "responsible" for the temporal localization of the signal $x(t)$, while the frequency "window" $F\{\psi_s(t)\}(-\frac{\omega}{s})$ is "responsible" for the localization in frequency, at the scale s.

The effective duration and the effective bandwidth are:

$$_s\sigma_t^2 = \frac{_t\sigma^2}{s^2} \qquad \text{and} \qquad _s\sigma_\omega^2 = s^2 \cdot_\omega \sigma^2, \tag{1.10}$$

where $_t\sigma$ and $_\omega\sigma$ represent the duration of the temporal "window", respective the bandwidth of the frequency "window" associated to the mother wavelets. See [36] for more theoretical details.

It is noticed that the time localization is getting worse with the increasing of the factor s, while frequency localization improves with the increasing of s.

Also,

$$_s\sigma_t^2 \cdot _s\sigma_\omega^2 = _t\sigma_\omega^2 \cdot \sigma^2. \tag{1.11}$$

Regardless of the value of s, the time-frequency localization determined by $\psi_s(\tau)$ is identical with the one realized by the generating "window" $\psi(t)$.

In [50] is stated that the Haar functions (defined in equation (1.24) and represented in Figure 1.9 c) have good time localization, but they have an infinite effective bandwidth, meaning that they are not localized in frequency. Contrary, cardinal sinus functions have good frequency localization, but they have an infinite duration. These two examples represent extreme cases, but between them there are mother wavelets (for example the elements of the Daubechies family) for which the product gives finite values. These functions have poorer time localization than Haar functions and poorer frequency localization than the cardinal sinus, but they provide a better time-frequency "compromise" than Haar or cardinal sinus functions. Some conclusions can be drawn from [50]: the effective duration of Daubechies wavelets functions is stronger influenced by the number of vanishing moments (we will define this term in Section 1.6), than their effective bandwidth, meaning that it increases monotonically with the number of vanishing moments (an opposite evolution is observed for the effective bandwidth) and the time-frequency localization of wavelets from the Daubechies family monotonically increases with the number of vanishing moments.

1.2.2 Time-frequency Resolution Cell

We will present in the following a comparison between Fourier Series and the CWT in terms of time-frequency representations. Fourier Series, have a very good frequency localization but they have not a localization in time. Figure 1.5 presents the time-frequency localization of Fourier Series.

All the discrete frequencies which correspond to the harmonics of a periodic signal are perfectly localized but there is no time localization. All the harmonics already mentioned have infinite durations. The CWT, on the other hand, has a good frequency localization and poor time localization for low-frequencies, and poor frequency localization and good time localization for high-frequencies, as it can be seen in Figure 1.6.

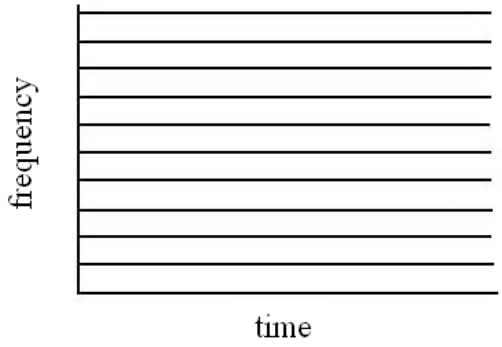

Figure 1.5: Time-frequency localization of the Fourier Series.

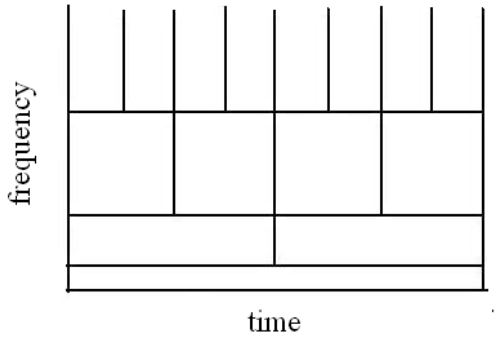

Figure 1.6: Time-frequency localization of the CWT.

1.3 Theoretical Aspects of Wavelet Transform

There are two main types of wavelet transforms - continuous and discrete.

1.3.1 Continuous Wavelet Transform

Any oscillating function with zero mean can be a mother wavelet. The wavelet transform of $f \in L^2(\Re)$ at time *u* and scale *s*, (1.6), is a convolution of the mother wavelet function $\psi \in L^2(\Re)$ with the function $f \in L^2(\Re)$:

$$Wf(u,s) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{(s)}}\psi^*(\frac{t-u}{s})dt = f * \bar{\psi}_s(u). \tag{1.12}$$

By applying Parseval formula, we can also write (1.12) as:

$$Wf(u,s) = \int_{-\infty}^{\infty} f(t)\psi^*_{u,s}(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)\hat{\psi}^*_{u,s}(\omega)d\omega. \tag{1.13}$$

The wavelet coefficients, *Wf*(*u,s*), depend on the signal *f(t)* and its spectrum $\hat{f}(\omega)$ in the time-frequency region where the energy of $\psi^*_{u,s}$ and $\hat{\psi}^*_{u,s}$ is concentrated. Since

it has a zero average, a wavelet coefficient *Wf*(*u,s*) measures the variation of *f* in the neighborhood of *u*, whose size is proportional to *s*.

The wavelet transform maps a raw data (observation of an underlying signal) into a collection of coefficients which provide the information on the behavior of the signal at certain point, during a certain time interval around that point. The coefficients tell us what the signal is doing and at what time. More precisely, it measures the change of the local average at a specific scale, around a specific moment.

The translation parameter *u* relates to the location of the wavelet function as it is shifted along the signal, while the scale parameter *s* is defined as the inverse of frequency.

The main disadvantage of the CWT is that it is computed for a large number of values both for the scale and for the translation, so it is a very redundant transform. Therefore, a discretization of the scale and translation variables was introduced.

1.3.2 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is obtained by the discretization of the CWT in the time-frequency plane [24] and is used to decompose discrete time signals. The result obtained at each decomposition level is composed by two types of coefficients: approximation coefficients and detail coefficients. The approximation coefficients are obtained by low-pass filtering the input sequence, followed by down-sampling. The detail coefficients are obtained by high-pass filtering the input sequence followed by down-sampling. The sequence of approximation coefficients constitutes the input for the next iteration. Each decomposition level corresponds to a specified resolution. The resolution decreases with the increasing of the number of decomposition levels. The DWT is invertible. Its inverse is named Inverse DWT (IDWT). At each resolution level, the approximation and the detail sequences are needed for the reconstruction of the approximation signal from the previous resolution level.

The Discrete Wavelet Transform has two features: the wavelet mother ψ and the number of decomposition levels. Discrete wavelets can be scaled and translated in discrete steps and a wavelet representation is the following:

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right), \tag{1.14}$$

where *j* is the scale factor and *n* is the translation index.

Classical DWT is not shift invariant meaning that the DWT of a translated version of a signal is not the same as the same translation of the DWT of the original signal. In order to achieve shift-invariance, several wavelet transforms have been proposed. One of them is presented in the following.

The Stationary Wavelet Transform (SWT) overcomes the absence of translation invariance of the DWT. The SWT, also known as the Undecimated Discrete Wavelet Transform (UDWT) is a time-redundant version of the standard DWT [64].

Unlike the DWT which down-samples the approximation coefficients and detail coefficients at each decomposition level [44], in the case of SWT no down-sampling is performed. This means that the approximation coefficients and the detail coefficients at each level have the same length as the original signal. This determines an increased number of coefficients at each scale and more accurate localization of signal features. Instead, the filters are up-sampled at each level.

The SWT has the translation-invariance, or shift-invariance, property. Thus, the SWT gives larger amount of information about the transformed signal as compared to the DWT. Larger amount of information is especially important when statistical approaches are used for analyzing the wavelet coefficients. The shift-invariant property is important in feature-extraction applications, denoising and detection.

The SWT can be implemented using the "à trous" algorithm, which will be detailed in a following section.

1.4 Multiresolution Analysis

The multiresolution analysis (MRA) was introduced in 1988 by Stephane Mallat and Yves Meyer [44] and uses the wavelet transform to decompose a data series in a cascade from the smallest scales to the largest ones. Adapting the signal resolution allows one to process only the relevant details for a particular task. The MRA is a method for analyzing a signal $x(t)$, that takes into account its representation at multiple time resolutions.

When the original signal $x(t)$ is involved, the maximal resolution is exploited. When a variant of the original signal (for example the signal $x(2t)$) is used, then a poorer resolution is exploited. Combining few analysis realized at different resolutions, a MRA is obtained.

The motivation of MRA is to use a sequence of embedded subspaces to approximate $L^2(\Re)$, allowing the selection of a proper subspace for a specific application task, to get a balance between accuracy and efficiency.

Mathematically, MRA represents a sequence of closed subspaces $V_j, j \in Z$ which approximate $L^2(\Re)$ and satisfy the following relations, [47]:

$$MR1 : ...V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2... \tag{1.15}$$

$$MR2 : \overline{\bigcup}_{j \in Z} V_j = L^2(\Re), \tag{1.16}$$

meaning that $L^2(\Re)$ space is the closure of the union of all subspaces $V_j, j \in Z$.

$$MR3 : \bigcap_{j \in Z} V_j = \{0\}, \tag{1.17}$$

that is, the intersection of all V_j contains a single element, the constant function equal with 0.

The multiresolution is reflected by the additional requirements:

$$MR4 : f \in V_j \longleftrightarrow f(2t) \in V_{j+1}, j \in Z, \tag{1.18}$$

$$f \in V_0 \longleftrightarrow f(t-k) \in V_0. \tag{1.19}$$

There exists a function, $\phi(t)$, such that its translates form an Riesz basis for V_0. Using the scale invariance condition, we see that $\{\phi(2t-k)\}$ is an Riesz basis for V_1.

Similarly, if we define $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$, then $\phi_{jk}(t)$ forms a Riesz basis for V_j. The function ϕ, which generates all the basis functions for all the spaces V_j, is called the scaling function of the multi-resolution analysis. Any Riesz basis [44] can be transformed into an orthonormal basis using the Gram-Schmidt orthogonalization

procedure [1]. Therefore, an orthonormal scaling functions basis corresponds to each scaling functions basis mentioned above.

Another important property of MRA is that, considering the subspace W_j, with $W_j \perp V_j$:

$$MR5 : V_{j+1} = V_j \bigoplus W_j. \tag{1.20}$$

The operator in the right hand side of equation (1.20) represents the direct sum of Hilbert spaces and the sequence of Hilbert spaces W_j.

A direct application of multi-resolution analysis is the fast discrete wavelet transform algorithm used to implement the DWT [44]. The fast discrete wavelet transform decomposes signals into low-pass and high-pass components sub-sampled by 2, while the inverse transform performs the reconstruction. Each mother wavelets ψ has a corresponding scaling function ϕ. The subspaces V_j are generated using bases obtained by the translations of a scaled variant of a scaling function. The subspaces W_j are generated using bases obtained by translations of a scaled version of the corresponding mother wavelets. In this case the subspaces W_j from (1.20) form an orthogonal decomposition of $L^2(\Re)$.

1.4.1 The Algorithm of Mallat

Generally, the MRAs are implemented based on the algorithm of Mallat [44]. It corresponds to the computation of the DWT, represented in Figure 1.7:

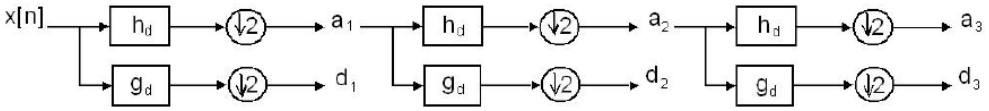

Figure 1.7: A three order Mallat decomposition tree.

The signal $x[n]$ is passed through a series of high pass filters with the impulse response (g_d), to analyze the high frequencies and it is passed through a series of low pass filters with the impulse response (h_d) to analyze the low frequencies. At each level, the high-pass filter produces after down sampling, the detail information d_k (k = 1, 2, 3 in this example), while the low-pass filter associated with scaling function produces, after down-sampling, coarse approximations, a_k (k = 1, 2, 3). The filtering operations determine the signal's resolution, meaning the quantity of detail information in the signal, while the scale is determined by up-sampling and sub-sampling operations.

There is a correspondence between the concepts of MRA and orthogonal decomposition mentioned above, and the diagram depicted in Figure 1.7. If $x[n]$ represents the decomposition coefficients of a signal $x(t)$ in the space V_0, then the sequence $a_1[n]$ represents the decomposition coefficients of $x(2t)$ in V_1 and the sequence $d_1[n]$ represents the decomposition coefficients in W_1 and so on.

The reconstruction operation is the reverse process of decomposition. The IDWT of the original signal is obtained by concatenating all the coefficients a_K and d_k, $k = 1...K$, starting from the last level of decomposition K. Due to successive sub-sampling by 2, the signal length must be a power of 2, or at least a multiple of power of 2 and it determines the number of levels that the signal can be decomposed into. The IDWT is implemented with the aid of up-samplers and finite impulse response (FIR) filters. The

sequence of approximation coefficients corresponding to a certain decomposition level is reconstructed starting from the sequences of approximation and detail coefficients corresponding to the next decomposition level. These approximation coefficients are up-sampled and the result is filtered with a low-pass filter. The detail coefficients are up-sampled and the result is filtered with a high-pass filter. The two new results are then added. The low-pass and high-pass filters used in the IDWT can be constructed starting with the corresponding filters used for the implementation of the DWT.

The disadvantage of the Mallat's algorithm is that the length of the coefficient sequences decreases with the increasing of the iteration index due to the decimators utilization. This fact produces translation variance, but the DWT is not redundant.

1.4.2 The Algorithm of Shensa

Another way to implement a MRA is the use of the algorithm "à trous" proposed by Shensa [64] which corresponds to the computation of the Stationary Wavelet Transform (SWT). The decomposition tree is represented in Figure 1.8.

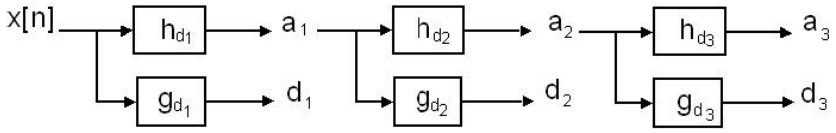

Figure 1.8: System for the computation of the SWT (3 levels).

In this case the use of decimators is avoided but at each iteration different low-pass (h_{d_1}, h_{d_2}, and h_{d_3}) and high-pass filters (g_{d_1}, g_{d_2} and g_{d_3}) are used. Each level filters are up-sampled versions of the previous ones.

So the differences between SWT and DWT are that the signal is never down-sampled, while the filters are up-sampled at each level in the case of SWT. The SWT is an inherently redundant scheme as each set of coefficients contains the same number of samples as the input - so for a decomposition of *N* levels there is a redundancy of 2*N*. Because no down-sampling is performed, a_1 and d_1 are of length *N* instead of *N*/2 as in the DWT case. At the next level of the SWT, a_1 is split into two using modified filters obtained by dyadically up-sampling the filters from the previous level. This process is continued recursively. The SWT is invertible and its inverse is named the Inverse SWT (ISWT). The implementation of the ISWT supposes to apply the inverse of the operations applied for the implementation of the SWT in inverse order. The SWT is translation invariant because all the filters composing the scheme in Figure 1.8 are linear time invariant systems.

1.5 Wavelet Families

There are several types of wavelet families whose qualities vary according to several criteria such as: the support of the mother wavelets, the symmetry, the number of vanishing moments, the regularity. These are associated with two properties that allow fast algorithm and space-saving coding: the existence of a scaling function and the orthogonality or the biorthogonality of the resulting analysis. They may also be

associated with these less important properties: the existence of an explicit expression, the ease of tabulating, or the familiarity with use. A possible classification of wavelets is into two classes: orthogonal and biorthogonal.

We have already mentioned that the set of functions obtained by translations and dilations of orthogonal mother wavelets forms an orthogonal basis and that the set of functions obtained by translations and dilations of biorthogonal mother wavelets forms a Riesz basis. Further details about the biorthogonal wavelets will be given in this section.

There is a variety of mother wavelets such as Daubechies, Symmlet, Haar or Coiflet, which generate orthogonal wavelet bases. An example of several mother wavelets waveforms, generated in Matlab®, is presented in Figure 1.9.

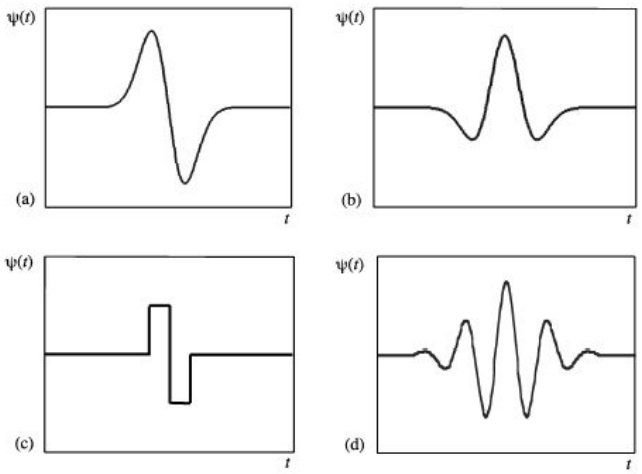

Figure 1.9: Several different mother wavelets: a) Gaussian wave; b) Mexican hat; c) Haar; d) Morlet.

The Haar mother wavelets is used for the computation of the Discrete Wavelet Transform, D_iWT, the other three mother wavelets showed in Figure 1.9 are used for the computation of the CWT.

Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling, it determines the characteristics of the resulting D_iWT. Therefore, the details of the particular application should be taken into account and the appropriate mother wavelets should be chosen in order to use the D_iWT effectively.

1.5.1 Vanishing Moments

The number of vanishing moments (or zero moments) is used to measure the local regularity of a signal [44]. According to [19] vanishing moments are a necessary condition for the smoothness of the wavelet function.

A wavelet $\psi(t)$ has *p* vanishing moments if:

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0, \tag{1.21}$$

with $0 \leq k < p$.

Substituting the mother wavelets in the integral from the left hand side of equation (1.21) with the probability density function, the integral becomes the moment of order k of the considered random variable. So, the equation (1.21) can be read as: the moment of order k of the random variable vanishes. This explains why p is named number of vanishing moments. The local regularity of mother wavelets is important because it can be chosen equal with the local regularity of the signal currently analyzed. This is an optimization technique for the procedure of selection of the mother wavelets. There are some features of mother wavelets which depend on its number of vanishing moments as the length of its support or its time, frequency or time-frequency localizations. The length of the support of a mother wavelets increases with the increasing of the number of vanishing moments. The time localization and the time-frequency localization of a mother wavelets decrease with the increasing of the number of vanishing moments. The frequency localization of a mother wavelets increases with the increasing of the number of vanishing moments.

Theorem 1 [44] associates the number of vanishing moments of ϕ with the number of vanishing derivatives of $\hat{\psi}(\omega)$ at $\omega = 0$, respective of $\hat{h}_d(\omega)$ at $\omega = \pi$.

Theorem 1

Let ψ and ϕ be a wavelet and the corresponding scaling function that generates an orthogonal basis. Suppose that $|\psi(t)| = O((1+t^2)^{-p/2-1})$ and $|\phi(t)| = O((1+t^2)^{-p/2-1})$. The following four statements are equivalent:

1. *The wavelet ψ has p vanishing moments;*
2. *$\hat{\psi}(\omega)$ and its first p-1 derivatives are* 0 *at $\omega = 0$;*
3. *$\hat{h}_d(\omega)$ and its first p-1 derivatives are* 0 *at $\omega = \pi$;*
4. *for any $0 \leq k < p$,*

$$q_k(t) = \sum_{n=-\infty}^{\infty} n^k \phi(t-n) \tag{1.22}$$

is a polynomial of degree k.

In this theorem $\hat{h}_d$ represents the Fourier transform of the conjugate mirror filter in Figure 1.7, while $\hat{\psi}$ represents the Fourier transform of ψ. A conjugate mirror filter represents a discrete filter that characterizes any scaling function, ϕ.

The proof of the theorem is presented in [44]. The Theorem 1 highlights the importance of the selection of the number of vanishing moments of mother wavelets. Condition 2 refers to the opportunity of the use of wavelets in spectral analysis. There are signals, as for example the long range dependent random signals (which will be studied in Chapter 4), whose spectral analysis is very difficult at low frequencies, because their Fourier transform tends to infinity when the frequency tends to zero. This spectral analysis can be successfully done with the aid of wavelets having an appropriate number of vanishing moments. Condition 3 in Theorem 1 gives indications about

the construction of the quadrature mirror filter associated with the mother wavelets. The construction of this filter is related to the length of the support of mother wavelets. The mother wavelet with the shortest support is the Haar mother wavelets. It has only one vanishing moment. Finally, Condition 4 specifies the degree of the polynomial which can be represented by linear combination of the corresponding scaling function. This degree depends on the number of vanishing moments as well.

1.5.2 Orthogonal Wavelet Families

In the case of orthogonal wavelets, vanishing moments, support, regularity and symmetry of the wavelet and scaling function are determined by the scaling filter. A scaling filter is a low-pass finite impulse response (FIR) filter of length $2N$ with the sum of coefficients of the impulse response equal with 1.

The coefficients of digital filters in Figures 1.7 and 1.8 are real numbers, the filters are of the same length and are not symmetric. The two filters *h* and *g* from a decomposition level are alternating flip of each other. This means that:

$$g[n] = (-1)^n h[M - n], \tag{1.23}$$

where *M* is an odd integer.

The alternating flip automatically gives double-shift orthogonality between the low-pass and high-pass filters. Perfect reconstruction is possible with alternating flip.

Orthogonal scaling functions and wavelets could have a high number of vanishing moments. This property is useful in many signal and image processing applications. They have regular structure which leads to easy implementation and scalable architecture.

An orthogonal wavelet has *p* vanishing moments if and only if its scaling function can generate polynomials of degree smaller than or equal to *p*.

If we refer to symmetry, it is well known that there is no symmetric compactly supported orthogonal mother wavelets, besides the wavelet of Haar.

Daubechies Wavelets

Daubechies wavelet family is named in the honor of its inventor, the Belgian physicist and mathematician Ingrid Daubechies and is one of the most widely used wavelet family. They represent a collection of orthogonal mother wavelets with compact support, characterized by a maximal number of vanishing moments for some given length of the support. Corresponding to each mother wavelets from this class, there is a scaling function (also called father wavelet) which generates an orthogonal MRA. The Daubechies mother wavelets are not symmetric.

A selection of Daubechies wavelets (left) and their scaling functions (right) is presented in Figure 1.10.

The elements of the Daubechies' family mostly used in practice are db1 - db20. The index refers to the number of vanishing moments. The number of vanishing moments is equal to half of the length of the digital filters length, N, in the case of Daubechies family of mother wavelets. For example, db1 (the Haar wavelet) has one vanishing moment, db2 has two vanishing moments and so on.

Haar wavelet (Daubechies wavelet of order 1) (represented in Figure 1.9, c) was the first mother wavelets proposed by Alfred Haar in 1909 [34] and has the shortest support among all orthogonal wavelets. The Haar mother wavelets generates, by

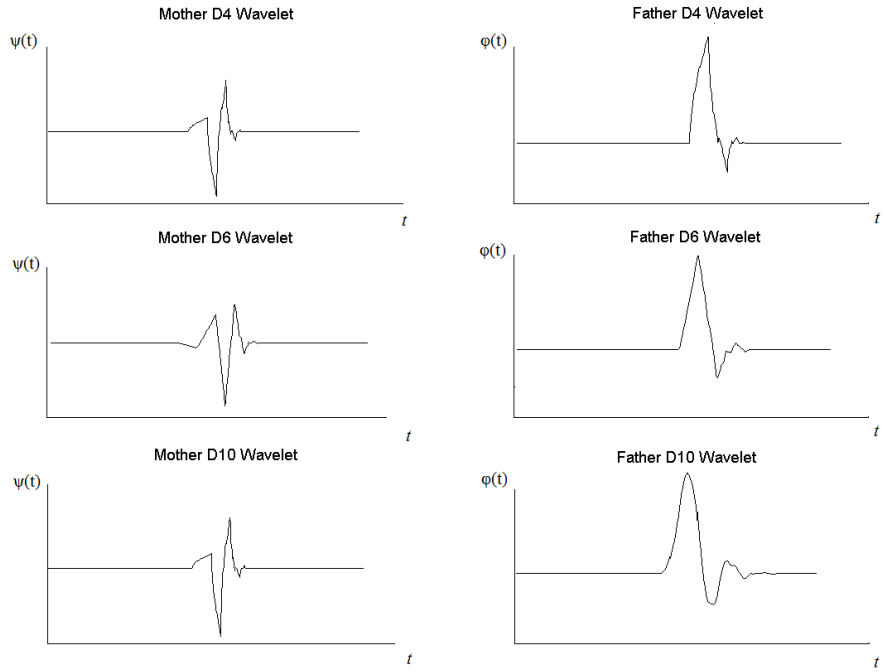

Figure 1.10: A selection of Daubechies wavelets (left) and their scaling functions (right): db4, db6 and db10.

translations and dilations, orthogonal wavelets. It is the single symmetric orthogonal mother wavelets. It is not well adapted for approximating smooth functions because it has only one vanishing moment.

Only Haar wavelets has an explicit expression, all other orders Daubechies wavelets are represented by wavelet coefficients and dilation equation. Haar mother wavelet function $\psi(t)$ has the expression:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{1.24}$$

The advantages of Haar wavelet transform are the following: it is conceptually simple and fast (the impulse response of its associated quadrature mirror filter has only two coefficients and the number of operations required by the implementation of the Haar transform is of the order of $2N$, where N represents the number of the samples of the input signal), it is memory efficient, and it is a good choice to detect time localized information. Because of these advantages we will use it with predilection in Chapters 3 and 4.

Symmlets

Daubechies wavelets are quite asymmetric. To improve symmetry Daubechies proposed Symmlets as a modification to her original wavelets [19].

Symmlets ($symN$, where N is the order), also known as Daubechies least asymmetric mother wavelets, are compact supported, orthogonal, continuous, but only nearly symmetric mother wavelets. The purpose was to create wavelets with the same size and same number of vanishing moments as Daubechies, but with near linear phase filters.

Symmlets have the highest number of vanishing moments for a given support width. Their construction is very similar to the construction of Daubechies wavelets, but the symmetry of Symmlets is stronger than the symmetry of Daubechies mother wavelets. Symmlets have $N/2$ vanishing moments, support length $N-1$ and filter length N.

Some examples of Symmlets (sym6 and sym8) and their associated scaling functions are presented in Figure 1.11.

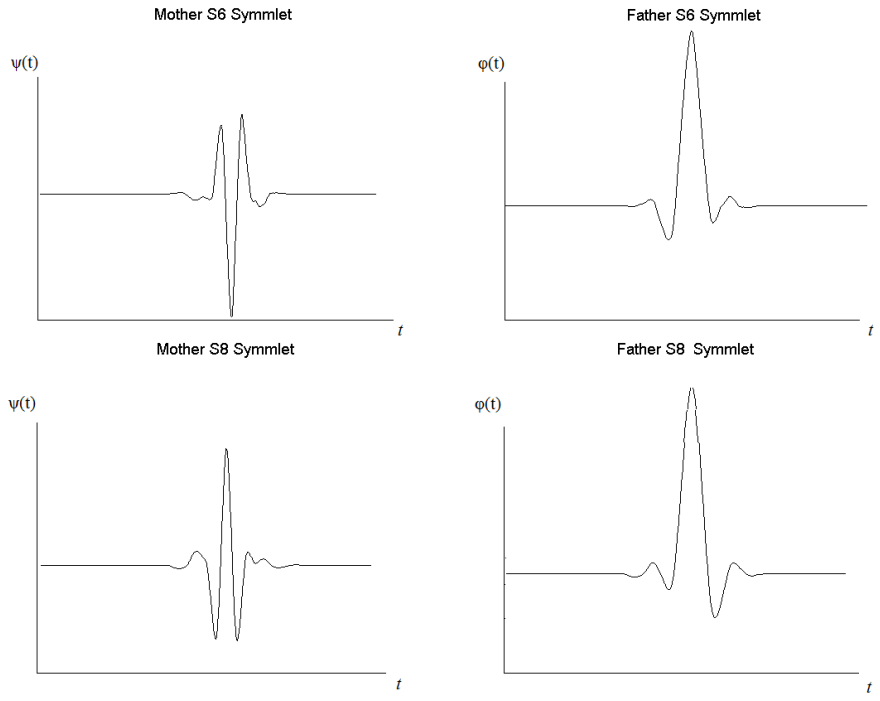

Figure 1.11: Symmlets (left) and their associated scaling functions (right): sym6 and sym8.

Coiflets

Coifman wavelets or "Coiflets" ($coifN$, where N is the order) are discrete wavelets designed by Ingrid Daubechies [19] and named in the honor of Ronald Coifman

(another researcher in the field of wavelets theory). Ronald Coifman suggested the construction of a orthonormal wavelets family with the same number of vanishing moments as the scaling functions they came from.

Coiflets are compactly supported wavelets and were designed to be more symmetrical than Daubechies mother wavelets to have a support of size $N-1$ and filter length N. The wavelet has $N/3$ vanishing moments, while the scaling function has $N/3-1$ vanishing moments. The number next to the wavelet's name represents the number of vanishing moments, related to the number of wavelet coefficients.

Two examples of Coiflets (coif3 and coif5) and their associated scaling functions are shown in Figure 1.12.

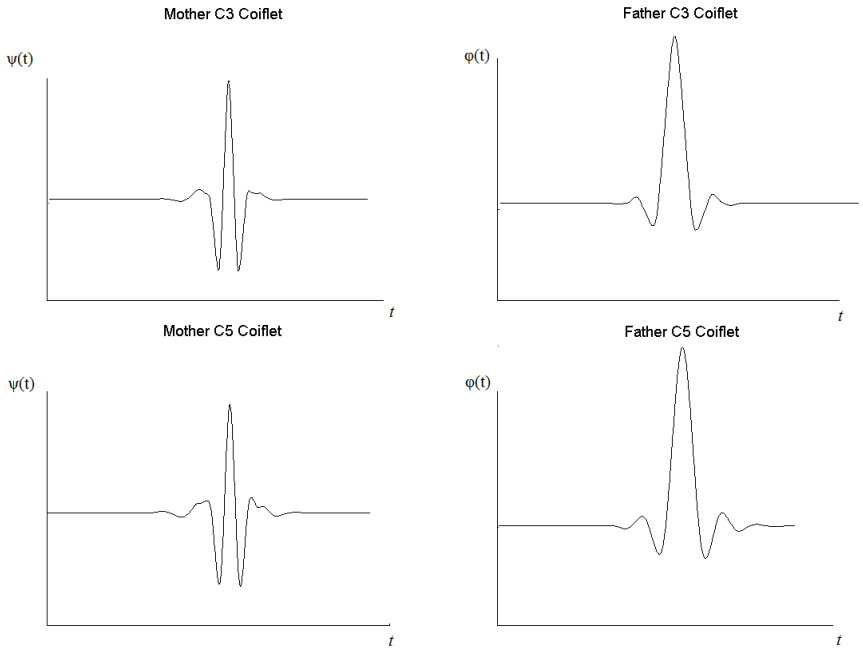

Figure 1.12: Coiflets (left) and their associated scaling functions (right): coif3, coif5.

1.5.3 Biorthogonal and Reverse Biorthogonal Wavelets

As already said the biorthogonal wavelets are elements of Riesz bases generating MRAs. In opposition with the orthogonal scaling functions which generate a single MRA, the biorthogonal scaling functions are associated in pairs which generate a pair of MRAs. The first element of the pair of biorthogonal scaling functions generates a MRA used for analysing the input signal of the associated forward WT. The second element generates a MRA used for the synthesis associated with the inverse WT. The elements of each MRA are orthogonal on the elements of a corresponding orthogonal decomposition. So, there are two orthogonal correspondences. They form a biorthogonal correspondence. More details about the concept of biorthogonality will be given

in the following.

Biorthogonal families include Biorthogonal and Reverse Biorthogonal wavelets. Generally, the biorthogonal scaling functions are selected from the family of spline functions. The Haar scaling function is a spline function of order zero. The spline function of first order is obtained by convolving the spline function of order zero with her self. The n^{th} order spline function is obtained by convolving the spline function of order $n-1$ with the spline function of order 0. Both families of wavelets, Biorthogonal and Reverse Biorthogonal, are composed by compactly supported wavelets associated with biorthogonal spline scaling functions implemented with FIR filters. Both symmetry and exact reconstruction are possible with FIR filters, [44].

Biorthogonal wavelets are families of compactly supported symmetric wavelets. Their construction can be made using an infinite cascade of perfect reconstruction filters which produce two scaling functions, ϕ and $\tilde{\phi}$ and two wavelets, ψ and $\tilde{\psi}$. For any $j \in Z$, $\phi_{j,n}$ and $\tilde{\phi_{j,n}}$ with $n \in Z$, generate bases of V_j and $\tilde{V}_j$ and the corresponding wavelets $\psi_{j,n}$ and $\tilde{\psi}_{j,n}$, with $n \in Z$, generate bases of two detail spaces W_j and $\tilde{W}_j$ such that:

$$V_{j+1} = V_j \bigoplus W_j, \qquad \tilde{V}_{j+1} = \tilde{V}_j \bigoplus \tilde{W}_j. \tag{1.25}$$

In Figure 1.13 is shown an example of biorthogonal wavelets with their associated scaling functions, for analysis and synthesis.

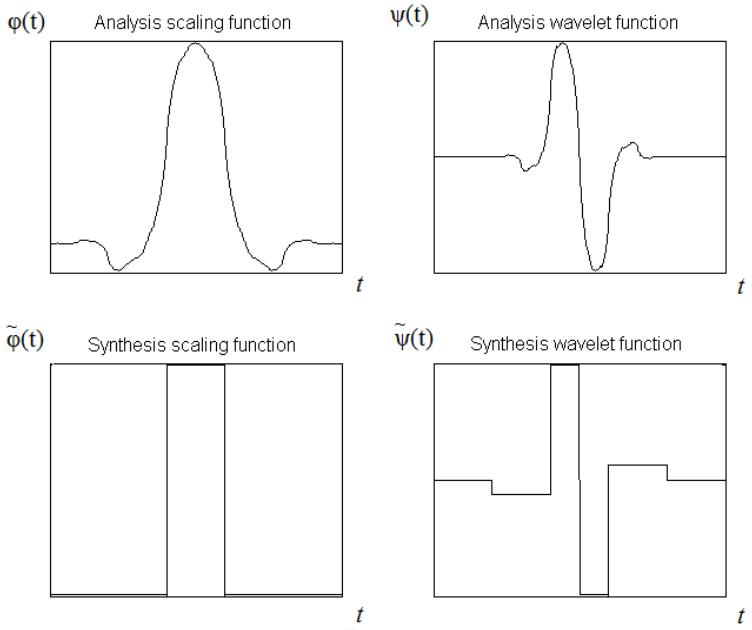

Figure 1.13: Biorthogonal wavelets, analysis and synthesis (right) and their associated scaling functions (left).

The digital filters associated with biorthogonal mother wavelets exhibit the property of linear phase which ensures the symmetry of the mother wavelets. Instead of a single orthogonal wavelet, in the case of biorthogonal wavelet transforms two

wavelets are used (one for decomposition and the other for reconstruction), as it can be seen in Figure 1.13. Designing biorthogonal wavelets allows additional degrees of freedom as compared to orthogonal wavelets, for example the possibility of constructing symmetric wavelet functions, [44]. In the case of the biorthogonal wavelet filters, the low pass and the high pass filters do not have the same length. The low pass filter is always symmetric, while the high pass filter could be either symmetric or asymmetric. The coefficients of the filters are either real numbers or integers. For perfect reconstruction, biorthogonal filter bank has all filters of odd length or even length. The two analysis filters can be symmetric with odd length or one symmetric and the other asymmetric with even length. Also, the two sets of analysis and synthesis filters must be dual. The linear phase biorthogonal filters are the most popular filters for data compression applications.

The biorthogonal wavelets are denoted as $biorNr.Nd$, where Nr is the order of the wavelet or the scaling functions used for reconstruction and Nd is the order of the functions used for decomposition. The reconstruction and decomposition functions have the support width equal to $2Nr+1$ and $2Nd+1$, respectively. The length of the associated filters is $max(2Nr, 2Nd)+2$.

Reverse biorthogonal ($rbioNr.Nd$, where Nr and Nd are the orders for the reconstruction and decomposition respectively) is obtained from biorthogonal wavelet pairs. This type of wavelets are compactly supported biorthogonal spline wavelets for which symmetry and exact reconstruction are possible with FIR filters. A comparison between the implementation of the D_iWT based on orthogonal and biorthogonal wavelets is presented in Figure 1.14.

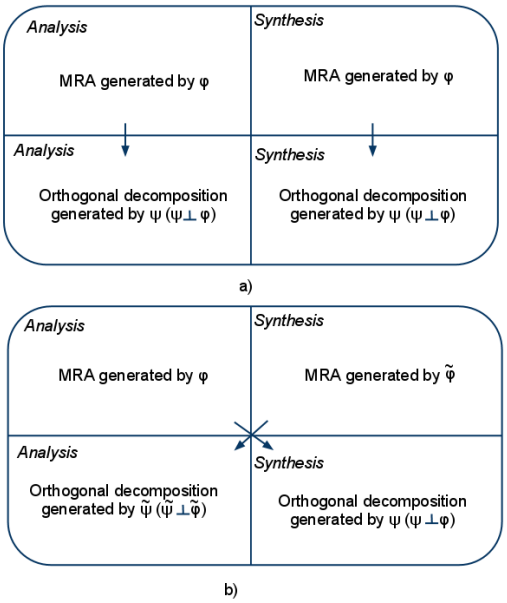

Figure 1.14: A comparison between D_iWT implementations based on orthogonal wavelet functions (a)) and biorthogonal wavelet functions(b)).

1.6 Applications of Wavelet Transforms

Wavelet transforms are now used in many applications, replacing the traditional Fourier Transform. Wavelets are extensively used in Signal and Image Processing [23], Communications [49], Computer Graphics [10], Finance [28], Medicine [48], Biology [48], Geology [42] and many other fields.

Wavelets have been heavily utilized to find the edges in digital images, to digitally compress fingerprints, in the modeling of distant galaxies or in denoising noisy data. Musicologists used wavelets to reconstruct damaged recordings, [6].

Wavelet analysis is proving to be a very powerful tool for characterizing self-similar behavior, over a wide range of time scales, [31].

1.7 Conclusions

As it was shown in sub-sections 1.2.4, where the wavelet transform was compared with the Fourier transform and 1.4.1, where the CWT was introduced, the wavelet transforms are important tools for analysis and processing of non-stationary signals. The scaling functions associated to wavelets allow the implementation of MRAs as it was shown in the sub-section 1.5. This is an important concept because it allows the identification of the most appropriate resolution for the representation of a given signal in a specified application. The details of a signal which does not carry relevant information for the considered application can be neglected on the basis of MRA, speeding the implementation of the application. We will use the MRA concept in Chapter 3, for a traffic forecasting application. There are two algorithms for the implementation of a MRA, the algorithm of Mallat associated with the DWT, presented in sub-section 1.5.1 and the algorithm of Shensa associated with the SWT, presented in sub-section 1.5.2. We will use both algorithms in the following chapters of this thesis. The problem of WiMAX traffic forecasting is solved in Chapter 3 with the aid of the SWT. The problem of the long range dependence of the WiMAX traffic detection is solved in Chapter 4 with the aid of DWT. In both cases the use of wavelets speed-up considerably the application. The most used families of mother wavelets were presented in section 1.6. One of the goals of the present thesis is to find out the most appropriate mother wavelets for traffic forecasting and long range dependence detection. This purpose will be achieved by brute force search. Each element of the orthogonal and biorthogonal wavelet families presented in section 1.6 will be tested in both applications and the mother wavelets which will optimize the performance of each application will be retained. One of the most important parameters for the selection of mother wavelets is its number of vanishing moments, introduced in sub-section 1.6.1. Its importance will be highlighted in the future sections of this thesis in relation with the influence of non-stationarity of a random process on the detection of its long range dependence degree. Wavelets add literally another dimension to Digital Signal Processing. Instead of processing a signal in the time/frequency domain, we can simultaneously process the signal in time and frequency (scale). From the time-frequency methods currently available for high resolution decomposition in the time-frequency plane (including STFT or Wigner-Ville transform), the wavelet transform appeared to be the favorite tool for researchers due to its high flexibility and adaptability to a large set of applications. Another key advantage of wavelet transform is the variety of wavelet functions available, that allows us to choose the most appropriate for the signal under investigation.

Wavelet transform analysis has now been applied to a wide variety of applications including time series prediction. Generally, the prediction is done with the aid of statistical methods or with the aid of neural networks. Both types of prediction are speed-up if they are applied in the wavelets domain. This increasing of speed is due to the sparsity of the WTs. There are only few wavelet coefficients with big values, the majority of the wavelet coefficients have small values and can be neglected without loosing a large amount of information. We will refer in Chapter 3 to wavelet based prediction method for WiMAX traffic.

The performance of any signal processing method based on wavelets can be improved by the good selection of the wavelet transform used and of its features. For this reason we will investigate in section 3.4 the process of mother wavelets selection for the proposed traffic forecasting procedure, based on the quantities defined in section 1.3.1. We will also refer in Chapter 4 to a LRD detection method based on wavelets. Different wavelet transforms are used in the applications considered in Chapter 3 and Chapter 4. The reasons for these choices are indicated in the corresponding chapters.

Chapter 2

Statistical Tools

This chapter provides an introduction to some basic concepts in statistics and time series analysis. The aim of this section is to shortly present the theoretical bases of the statistical methods which will be used in the following two chapters of the thesis. The traffic forecasting methodology which represents the subject of Chapter 3 is based on an ARIMA model applied in the wavelets domain. The first goal of the present section is to define the ARIMA model, to show its utilization in estimation applications and to introduce some quality measures for this prediction. The long range dependence of the traffic is detected in Chapter 4 with the aid of Hurst parameter estimators. The description of those estimators represents the second goal of the present section. Let's present for the beginning some basic concepts in statistics.

2.1 Simple Statistical Measures

In the following we will define some statistical measures:

Definition 1. *Mean (μ): the mean of a random variable X can be defined as:*

$$\mu = E[X], \tag{2.1}$$

where E represents the statistical mean operator.

Definition 2. *Variance ($\sigma^2 = VAR(X)$): the variance of a random variable X is given by:*

$$\sigma^2 = VAR(X) = E[(X - \mu)^2]. \tag{2.2}$$

Definition 3. *Standard deviation (σ): the standard deviation of a random variable is the square root of the variance.*

Definition 4. *Autocovariance (γ): the autocovariance of a time-series X_t can be defined as:*

$$\gamma(i,j) = E[(X_i - \mu)(X_j - \mu)]. \tag{2.3}$$

Definition 5. *Autocorrelation function (ACF): the autocorrelation function of a time series X_t is given by:*

$$\rho(i,j) = \frac{\gamma(i,j)}{\sigma^2} = \frac{E[(X_i - \mu)(X_j - \mu)]}{\sigma^2}. \tag{2.4}$$

Definition 6. *Partial autocorrelation function (PACF): the partial autocorrelation function at the lag k is the autocorrelation between X_t and X_{t-k} that is not explained by all lower-order lags (1 to k-1, inclusive).*

$$\phi_{kk} = CORR(X_t - P(X_t|X_{t+1} + ... + X_{t+k-1}), X_{t+k} - P(X_{t+k}|X_{t+1} + ... + X_{t+k-1})). \quad (2.5)$$

where $P(X_t|X_{t+1}+...+X_{t+k-1})$ is the best linear projection of X_t over $X_{t+1}+...+X_{t+k-1}$, [26].

The PACF will vary between -1 and +1, with values near 1 indicating stronger correlation. PACF is a commonly used tool for model identification in Box-Jenkins methodology, identifying the order *p* of an AR model: the PACF for a AR(p) is nonzero if $k \leq p$ and zero for $k > p$, [8].

Definition 7. *Stationary process: A wide-sense stationary process X_t is a stochastic process characterized by the fact that its probability distribution, mean and variance do not change over time or position.*

An example of stationary process is the White Gaussian Noise (WGN). A Gaussian process is a stochastic process whose realizations consist in random values associated with every point in a range of time such that each random variable has a normal distribution:

$$P_G(x) = \frac{1}{\sqrt{2\pi}\sigma_G} e^{-\frac{(x-\mu_G)^2}{2\sigma_G^2}}, \quad (2.6)$$

where parameter μ_G is the mean and σ_G^2 is the variance. The distribution with $\mu_G = 0$ and $\sigma_G = 1$ is called standard normal.

A WGN process is a Gaussian process which has the following covariance function:

$$E\{WGN(t)WGN(s)\} = \sigma_G^2 \delta(t-s). \quad (2.7)$$

An example of non-stationary process is the fractional Brownian motion (fBm) [14]. It is a continuous-time Gaussian process $B_H(t)$ on [0,T], which starts at zero, has expectation zero for all $t \in [0, T]$ and has the following covariance function:

$$E\{B_H(t)B_H(s)\} = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \quad (2.8)$$

where H is a real number in the interval (0,1), called Hurst parameter [14]. The values of H determines what kind of process the fBm is:

- if $H = \frac{1}{2}$ the process is in fact a Brownian motion;
- if $H > \frac{1}{2}$ increments of the process are positively correlated;
- if $H < \frac{1}{2}$ increments of the process are negatively correlated.

The increment process, $X(t) = B_H(t+1) - B_H(t)$, is known as fractional Gaussian noise (fGn). The increment process is stationary:

$$B_H(t) - B_H(s) \sim B_H(t-s). \quad (2.9)$$

2.2 Basic Stochastic Model in Time-series Analysis

Autoregressive (AR) Process of Order *p*

An Autoregressive model of order *p* (AR(*p*)) is a weighted linear sum of the past *p* values [12] and it is defined by the following equation:

$$X_t = Z_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p}, \tag{2.10}$$

where X_t represent the time series which model must be established, $\phi_p(.)$ is a p^{th} degree polynomial and Z_t is a white noise time series.

Moving Average (MA) Process of Order *q*

Moving average (MA) process of order *q* is a weighted linear sum of the past *q* random shocks:

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q}. \tag{2.11}$$

where $\theta_q(.)$ is a q^{th} degree polynomial and Z_t is a white random process with constant variance and zero mean [12].

Autoregressive Moving Average Model (ARMA)

Given a time series of data X_t, the Autoregressive moving average model (ARMA) is a tool for understanding and predicting future values in this series. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The model is usually then referred to as the ARMA(*p*,*q*) model where *p* is the order of the autoregressive part and *q* is the order of the moving average part.

A time series X_t is an ARMA(*p*,*q*) process if X_t is stationary and if:

$$\phi(B)X_t = \theta(B)Z_t, \tag{2.12}$$

which can be expressed as:

$$\phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q}, \tag{2.13}$$

where $\phi_p(.)$ and $\theta_q(.)$ are p^{th} and q^{th} degree polynomials, and *B* is the backward shift operator ($B_j X_t = X_{t-j}$, $B_j Z_t = Z_{t-j}$, *j* = 0, ±1,...).

The ARMA model fitting procedure assumes the data to be stationary. If the time series exhibits variations that violate the stationary assumption, then there are specific approaches that could be used to render the time series stationary. As we will see in a following sub-section a stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are invariant in time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary (i.e., "stationarized") through the use of mathematical transformations. A stationarized series is relatively easy to predict: you simply predict that its statistical properties will be the same in the future as they have been in the past! The predictions for the stationarized series can then be "untransformed", by reversing whatever mathematical transformations were previously used, to obtain predictions for the original series. Thus, finding the sequence of transformations needed to stationarize a time series often provides important clues in the search for an appropriate

forecasting model. One of the operations which can be used for the stationarization of a time series is the differencing operation. The first difference of a time series is the series of changes from one period to the next. If $Y(t)$ denotes the value of the time series Y at period t, then the first difference of Y at period t is equal to $Y(t)-Y(t-1)$. If the first difference of Y is stationary and also not autocorrelated, then Y is described by a random walk model: each value is a random step away from the previous value. If the first difference of Y is stationary but autocorrelated, then a more sophisticated forecasting model such as exponential smoothing or ARIMA may be appropriate.

2.2.1 Autoregressive Integrated Moving Average Model (ARIMA)

Autoregressive integrated moving average (ARIMA) model is a generalization of an ARMA model. In statistics and signal processing, ARIMA models, sometimes called Box-Jenkins models after the iterative Box-Jenkins methodology usually applied to estimate them, are usually modeled for time series data.

ARIMA models are fitted to time series data either to better understand the data or to predict future points in the series. They are applied in some cases where data show evidence of non-stationarity, when some initial differencing steps must be applied to remove the non-stationarity.

The model is generally referred to as an ARIMA(*p*,*d*,*q*) model where *p*, *d*, and *q* are integers greater than or equal to zero and refer to the order of the autoregressive, integration (number of differencing steps needed to achieve stationarity), and moving average parts of the model respectively.

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t. \tag{2.14}$$

A generalization of standard ARIMA(*p*,*d*,*q*) processes is the Fractional ARIMA model referred to as FARIMA(*p*,*d*,*q*) [46]. The difference between ARIMA and FARIMA consist in the degree of differencing d, which for FARIMA models takes real values.

A time series X_t is a FARIMA(*p*,*d*,*q*) process if:

$$\phi(B)(\Delta)^d X_t = \theta(B)Z_t, \tag{2.15}$$

and $-0.5 < d < 0.5$.

2.2.2 Box-Jenkins Methodology

The Box-Jenkins methodology [8] applies to ARMA or ARIMA models to find the best fit of a time series to its past values, in order to make forecasts.

The original methodology uses an iterative three-stage modeling approach:

1. Model identification and model selection:
    - making sure that the time-series are stationary: Stationarity can be assessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay.
    - identifying seasonality in the dependent series. Seasonality (or periodicity) can usually be assessed from an autocorrelation plot, a seasonal sub-series plot, or a spectral plot.

2. Parameters estimation used to arrive at coefficients which best fit the selected ARIMA model. Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e., the p and q) of the autoregressive and moving average parts. The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known. Specifically, for an AR(1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components. The autocorrelation function of a MA(q) process becomes zero at lag q + 1 and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero. For models with p > 0 and q > 0, the ACF and PACF are difficult to recognize and have less value in order selection than in the cases where p or q equals 0. For these cases there are other selection criteria that will be discussed in the next sections.
3. Model checking by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals (elements of the time series that have no signification) should be as small as possible and should not follow a model. If the estimation is inadequate, we have to return to step one and attempt to build a better model.

Once the model has been selected, estimated and checked, the next step is to compute forecasts.

2.3 Parameter Estimation and Order Selection Criteria

The determination of an appropriate ARMA(p, q) model to represent an observed stationary time series involves the order p and q selection and estimation of the mean, the coefficients ϕ_p and θ_q, and the white noise variance σ^2.

When p and q are known, good estimators of ϕ and θ can be found by imagining the data to be observations of a stationary Gaussian time series and maximizing the likelihood with respect to the parameters ϕ_p, θ_q and σ^2. The estimators obtained using this procedure are known as maximum likelihood estimators.

2.3.1 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is the most popular method used for parameter estimation in statistics. The aim of this method is to determine the parameters that maximize the probability of observations. The likelihood function of data set represents the probability of obtaining that particular data set given that the probability density is known. A detailed theoretical approach regarding MLE is presented in [35].

In the following the problem of selecting appropriate values for the orders p and q will be discussed. Several criteria have been proposed in the literature, since the problem of model selection arises frequently in statistics, [9].

2.3.2 Final Prediction Error (FPE)

Developed by Akaike in 1969, FPE criterion is used to select the appropriate order of an AR process to fit to a time series $X_1, ..., X_n$. The most accurate model has the

smallest FPE. The FPE for an AR process of order *p* can be estimated according to the following equation:

$$FPE = \hat{\sigma}^2 \cdot \frac{n+p}{n-p}. \tag{2.16}$$

where *n* is the number of samples.

2.3.3 Akaike information criterion (AIC)

AIC is a measure of the goodness of fit of an estimated statistical model. In fact, AIC is the generalization of maximum likelihood principle. Given observations $X_1, ..., X_n$ of an ARMA process the AIC statistic is defined as:

$$AIC = -2 \cdot ln(L) + 2(p+q+1), \tag{2.17}$$

where *L* is the likelihood function.

The AICC is a bias-corrected version of the AIC, proposed by Hurvich, [35]. This criterion is applied as follows: choose *p*, *q*, ϕ_p, and θ_q to minimize:

$$AICC = -2 \cdot lnL + \frac{2(p+q+1)n}{n-p-q-2}. \tag{2.18}$$

where *n* is the number of samples.

2.3.4 Bayesian Information Criterion (BIC)

In the case of AICC and AIC statistics, for $n \to \infty$, the factors 2(*p* + *q* + 1)*n*/(*n*-*p*-*q*-2) respective 2(*p* + *q* + 1) are asymptotically equivalent. BIC is another criterion for model selection, that attempts to correct the overfitting nature of the AIC, [35]. For a zero-mean causal invertible ARMA(*p*, *q*) process, BIC is defined by the following equation:

$$BIC = (n-p-q) \cdot ln\left[\frac{n\hat{\sigma}^2}{n-p-q}\right] + n \cdot (1 + ln\sqrt{2\pi}) + (p+q) \cdot ln\left[\frac{(\sum_{t=1}^{n} X_t^2 - n\hat{\sigma^2})}{p+q}\right], \tag{2.19}$$

where $\hat{\sigma}^2$ is is the maximum likelihood estimator of σ^2 (the white noise variance of the AR(*p*) model).

2.4 Analysis of Variance

The Analysis of Variance (ANOVA) technique is a statistical method used to quantify the amount of variability accounted by each term in a multiple linear regression model. It can be used in the reduction of a multiple linear regression model process, identifying those terms in the original model that explain the most significant amount of variance.

We define the sum squared error (*SSE*):

$$SSE = \sum_{t=1}^{n} e(t)^2, \tag{2.20}$$

where *e(t)* represents the error of the model.

We denote the following sum with (*SSX*):

$$SSX = \sum_{t=1}^{n} y(t)^2, \tag{2.21}$$

where $y(t)$ is the observed response of the model.

The total sum of squares (*SST*) is defined as the uncertainty that would be present if one had to predict individual responses without any other information. The best one could do is to predict each observation to be equal to the sample mean. So, we compute *SST* as:

$$SST = \sum_{t=1}^{n} \left(y(t) - \overline{y(t)}\right)^2, \tag{2.22}$$

where $\overline{y(t)}$ represents the mean of $y(t)$.

The ANOVA methodology splits this variability into two parts. One component is accounted for by the model and it corresponds to the reduction in uncertainty that occurs when the regression model is used to predict the response. The remaining component is the uncertainty that remains even after the model is used. We define the regression sum of squares, *SSR*, as the difference between *SST* and *SSE*. This difference represents the sum of the squares explained by the regression.

The fraction of the variance that is explained by the regression determines the goodness of the regression and is called *the coefficient of determination*, R^2. The coefficient of determination can be expressed by the following formula:

$$R^2 = \frac{SSR}{SST}. \tag{2.23}$$

The model is considered to be statistically significant if it can account for a large fraction of the variability in the response, i.e. yields large values for R^2.

2.5 Measuring the Performance of a Forecasting Model

The performance of the forecasting model can be judged from its predictive ability in terms of the following well-known evaluation criteria:

1. *Symmetrical Mean Absolute Percentage Error* (SMAPE),
2. *Mean Absolute Percentage Error* (MAPE),
3. *Mean Absolute Error* (MAE).

SMAPE calculates the symmetric absolute error between the actuals *X* and the forecast *F* across all observations *t* of the test set of size *n*:

$$SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|X_t - F_t|}{(X_t + F_t)/2}. \tag{2.24}$$

MAPE is calculated as follows:

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \left|\frac{X_t - F_t}{X_t}\right| \times 100\%. \tag{2.25}$$

MAE measures how close forecasts or predictions are to the eventual outcomes and has the following expression:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|F_t - X_t|. \tag{2.26}$$

2.6 Second Order DWT Statistical Analysis

Let $x(t)$ be a wide sense stationary random signal. The DWT coefficients of its projection on a space V_0 are:

$$d_m[n] = \langle x(t), \psi_{m,n}(t)\rangle. \tag{2.27}$$

The autocorrelation of this sequence is:

$$\begin{aligned} R_{d_m}[k,l] &= E\left\{d_m[k]d_m^*[l]\right\} = E\left\{\langle x(t),\psi_{m,k}(t)\rangle, \langle x(t),\psi_{m,l}(t)\rangle^*\right\} = \\ &= E\left\{\int_{-\infty}^{\infty} x(t)\psi_{m,k}(t)^* dt \int_{-\infty}^{\infty} x(u)^*\psi_{m,l}(u)du\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left\{x(t)x^*(u)\right\}\psi_{m,k}^*(t)\psi_{m,l}(u)dtdu. \end{aligned} \tag{2.28}$$

Because x is stationary:

$$E\left\{x(t)x^*(u)\right\} = R_x(t-u), \tag{2.29}$$

so:

$$\begin{aligned} R_{d_m}[k,l] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_x(t-u)\psi_{m,k}^*(t)\psi_{m,l}(u)dtdu \\ &= \int_{-\infty}^{\infty}\left(R_x(t) * \psi_{m,l}(t)\right)\psi_{m,k}^*(t)dt, \end{aligned} \tag{2.30}$$

or, using Parseval theorem we obtain:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}\left\{R_x(t) * \psi_{m,l}(t)\right\}(\omega)\mathcal{F}^*\left\{\psi_{m,k}(t)\right\}(\omega)d\omega. \tag{2.31}$$

Taking into consideration that the wavelets are real functions:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}\left\{R_x(t)\right\}(\omega)\mathcal{F}\left\{\psi_{m,l}(t)\right\}(\omega)\mathcal{F}^*\left\{\psi_{m,k}(t)\right\}(\omega)d\omega. \tag{2.32}$$

But:

$$\mathcal{F}\left\{\psi_{m,l}(t)\right\}(\omega) = 2^{\frac{m}{2}}e^{-j\omega 2^m l}\mathcal{F}\left\{\psi\right\}(2^m\omega), \tag{2.33}$$

and:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}\left\{R_x(t)\right\}(\omega)2^{\frac{m}{2}}e^{-j\omega 2^m l}\mathcal{F}\left\{\psi\right\}(2^m\omega)2^{\frac{m}{2}}e^{-j\omega 2^m k}\mathcal{F}^*\left\{\psi\right\}(2^m\omega)d\omega. \tag{2.34}$$

or:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}\left\{R_x(t)\right\}(\omega)2^{\frac{m}{2}}e^{-j\omega 2^m (l-k)}\left|\mathcal{F}\left\{\psi\right\}(2^m\omega)\right|^2 d\omega. \tag{2.35}$$

After changing the variable $\nu = 2^m\omega$ we obtain:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}\{R_x(t)\}(2^{-m}\nu)e^{-j\nu(l-k)}|\mathcal{F}\{\psi\}(\nu)|^2 d\nu \tag{2.36}$$

The last integral can be decomposed into a series of integrals on intervals of length 2π:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\sum_{p=-\infty}^{\infty}\int_{(2p-1)\pi}^{(2p+1)\pi} \mathcal{F}\{R_x(t)\}(2^{-m}\nu)e^{-j\nu(l-k)}|\mathcal{F}\{\psi\}(\nu)|^2 d\nu. \tag{2.37}$$

After changing again the variable $w = \nu - 2p\pi$ we obtain:

$$R_{d_m}[k,l] = \frac{1}{2\pi}\sum_{p=-\infty}^{\infty}\int_{-\pi}^{\pi} \mathcal{F}\{R_x(t)\}\left(2^{-m}(w+2p-\pi)\right)e^{-jw(l-k)}|\mathcal{F}\{\psi\}(w+2p\pi)|^2 dw. \tag{2.38}$$

For $m \to \infty$ the previous equation becomes:

$$\begin{aligned} R_{d_\infty}[k,l] &= \frac{1}{2\pi}\sum_{p=-\infty}^{\infty}\int_{-\pi}^{\pi} \mathcal{F}\{R_x(t)\}(0)e^{-jw(l-k)}|\mathcal{F}\{\psi\}(w+2p\pi)|^2 dw = \\ &= \frac{1}{2\pi}\mathcal{F}\{R_x(t)\}(0)\int_{-\pi}^{\pi}\sum_{p=-\infty}^{\infty}|\mathcal{F}\{\psi\}(w+2p\pi)|^2 e^{-jw(l-k)}dw. \end{aligned} \tag{2.39}$$

If the DWT is computed using orthogonal wavelets, then:

$$\sum_{p=-\infty}^{\infty}|\mathcal{F}\{\psi\}(w+2p\pi)|^2 = 1 \tag{2.40}$$

and equation (2.38) becomes:

$$R_{d_\infty}[k,l] = \frac{1}{2\pi}\mathcal{F}\{R_x(t)\}(0)\int_{-\pi}^{\pi} e^{-jw(l-k)}dw. \tag{2.41}$$

The integral in equation (2.41) is proportional with the inverse discrete Fourier transform of the constant 1, which equals the sequence $\delta[k-l]$. So, equation (2.41) can be written:

$$R_{d_\infty}[k,l] = \mathcal{F}\{R_x(t)\}(0)\delta[k-l]. \tag{2.42}$$

Hence, the sequence $d_\infty[n]$ is not correlated. So, we just proved that asymptotically (when the number of decomposition levels tends to infinity), the DWT decorrelates the input random process.

The means and the variances of the DWT coefficients are computed in the following.

$$E\{d_m[k]\} = E\{\langle x(t), \psi_{m,k}(t)\rangle\} = E\left\{\int_{-\infty}^{\infty} x(t)\psi^*_{m,k}(t)dt\right\}. \tag{2.43}$$

After applying Fubini's theorem, the equation (2.43) becomes:

$$E\{d_m[k]\} = \int_{-\infty}^{\infty} E\{x(t)\}\psi^*_{m,k}(t)dt = \int_{-\infty}^{\infty}\mu_x\psi^*_{m,k}(t)dt = \mu_x\mathcal{F}\left\{\psi^*_{m,k}\right\}(0) \tag{2.44}$$

Using the equation (2.33), the expectation of the wavelet coefficients becomes:

$$E\left\{d_m[k]\right\} = \mu_x 2^{-\frac{m}{2}} \mathcal{F}\left\{\psi\right\}(0) = 0 \tag{2.45}$$

because $\mathcal{F}\left\{\psi\right\}(0) = 0$.

So, the expectation of each wavelet coefficients sequence is null. The variance of the detail wavelet coefficients can be computed with the aid of their autocorrelation function in (2.36) because they have null expectation:

$$E\left\{d_m^2[k]\right\} = R_{d_m}[0] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}\left\{R_x(t)\right\}(2^{-m}\nu)|\mathcal{F}\left\{\psi\right\}(\nu)|^2 d\nu \tag{2.46}$$

For $m \to \infty$ the previous equation becomes:

$$\begin{aligned} E\left\{d_\infty^2[k]\right\} &= \mathcal{F}\left\{R_x(t)\right\}(0)\frac{1}{2\pi}\int_{-\infty}^{\infty}|\mathcal{F}\left\{\psi\right\}(\nu)|^2 d\nu = \mathcal{F}\left\{R_x(t)\right\}(0)R_\psi(0) = \\ &= \mathcal{F}\left\{R_x(t)\right\}(0) \end{aligned} \tag{2.47}$$

for orthogonal wavelets.

Hence, asymptotically, the detail wavelet coefficients represent a zero mean white noise with the variance equal with the value of the power spectral density of the input signal computed in zero. This variance depends only on the input process being independent of the mother wavelets used in the implementation of the DWT.

In the following is analyzed an interesting particular case when the input process is a zero mean white noise with the variance equal to σ^2. In this case, $\mathcal{F}\left\{R_x(t)\right\}(\omega) = \sigma^2$ and the equation (2.38) becomes:

$$\begin{aligned} R_{d_m}[k,l] &= \frac{1}{2\pi}\sum_{p=-\infty}^{\infty}\int_{-\pi}^{\pi}\sigma^2 e^{-jw(l-k)}|\mathcal{F}\left\{\psi\right\}(w+2p\pi)|^2 dw = \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\sum_{p=-\infty}^{\infty}|\mathcal{F}\left\{\psi\right\}(w+2p\pi)|^2 e^{-jw(l-k)}d\omega = \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\sum_{p=-\infty}^{\infty}|\mathcal{F}\left\{\psi\right\}(w+2p\pi)|^2 e^{-jw(l-k)}dw. \end{aligned} \tag{2.48}$$

In the case of orthogonal wavelets the previous equation can be written:

$$R_{d_m}[k,l] = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} e^{-jw(l-k)}dw = \sigma^2\delta[k-l]. \tag{2.49}$$

So, if the input process is a white noise with zero mean and variance σ^2, then all the wavelet coefficients are zero mean white noises with the same variance as the input process.

The DWT does not correlate the white noise. This result could seem a paradox, taking into consideration the quadrature mirror filters used for the implementation of the DWT because any filter correlates the white noise. When the input signal of the DWT is a white noise, then the sequences of detail coefficients from any decomposition level of the DWT are white noises as well, having the same variance.

Another important particular case, when the DWT's input random process has long range dependence, will be treated in a following section, in connection with the problem of Hurst parameter estimation. We will propose in that section a new method for the estimation of the Hurst parameter for long range dependent stationary random processes.

2.7 Self-similarity and Long-Range Dependence

In this section we will introduce the concepts of self-similarity and long range dependence which will be applied to the analysis of communication networks traffic in Chapter 4. Self-similarity, or scale-invariance, is an important notion in the understanding of network traffic, [39]. A process is self-similar if its statistical behavior is independent of the time-scale, meaning that the statistical characteristics of the process may appear similar at different time scales, [59]. In the last years there have been made a series of empirical studies on traffic measurements from various communication networks. These studies have proved that the actual traffic is self--similar (fractal) or long-range dependent, [40], [17], [43], [39], [67], [3], [52], [30]. Thus, several models of long-range dependent processes have been introduced. The most well-known models of long-range dependent processes are fGn (thus second-order self-similarity) and FARIMA, [41]. There is a number of different definitions for self-similarity. Considering a continuous-time process Y = *Y*(*t*), we define its self--similarity in the sense of finite dimensional distributions:

Definition 8. *The process* Y(t), $t \geq 0$ *is self-similar with self-similarity parameter H (Hurst parameter) if:*

$$Y(t) = a^{-H}Y(at), \qquad \forall t \geq 0, \qquad \forall a > 0, \qquad 0 < H < 1. \tag{2.50}$$

The process *Y* can never be stationary because stationarity implies *Y*(*t*) = *Y*(*at*), but *Y* is assumed to have stationary increments.

Considering a discrete time stochastic process or time series (for example the traffic volume, measured in packets, bytes, or bits), *X*(*t*), with $t \in \mathbb{Z}$, we define the second-order self-similarity, respective the asymptotically second-order self-similarity, [39].

Definition 9. $X(t)$ *is exactly second-order self-similar with Hurst parameter* H *(0.5 <*H *<1) if:*

$$\gamma(k) = \frac{\sigma^2}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right), \qquad k \geq 1, \tag{2.51}$$

where $\gamma(k)$ *represents the autocovariance function of the m-aggregated process* $X^{(m)}$ *of X at aggregation level m.*

$X^{(m)}$ is defined as:

$$X^{(m)}(i) = \frac{1}{m}\sum_{t=m(i-1)+1}^{mi} X(t). \tag{2.52}$$

If *X* is the increment process of a self-similar process *Y* defined in (2.50), ($X(i) = Y(i+1) - Y(i)$), then for all integers *m*:

$$X = m^{1-H}X^{(m)}. \tag{2.53}$$

X is self-similar when it has the same ACF $\rho(k)$ as the series $X^{(m)}$ for all *m*, where $\rho(k)$ is defined as follows:

$$\rho(k) = E\left[(X_t - \mu)(X_{t+k} - \mu)\right]. \tag{2.54}$$

X is exactly self-similar if the relation (2.53) is satisfied for every *m*, [71].

A stationary sequence *X* is asymptotically self-similar if the relation (2.52) is satisfied for $m \to \infty$, [71].

From Definition 8 and 9 we can observe that a random process could be self-similar or exactly self-similar. The second concept, the exact self-similarity, imposes a certain structure for the autocorrelation of the corresponding random process (given in equation (2.51)). This constraint can be considered further in its exact form or only asymptotically.

Processes with LRD are often confused with self-similar processes, but they are different, meaning that some self-similar processes may exhibit LRD, but not all processes having LRD are self-similar.

The form $\gamma(k) = (k+1)^{2H} - 2k^{2H} + (k-1)^{2H}$ in equation (2.51) from Definition 9 is related to LRD. The autocorrelation function of a fractional Gaussian noise (fGn) satisfies the following equation, [52]:

$$\rho(k) = 2\gamma(k) = [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]\sigma^2. \tag{2.55}$$

For *k*=0 the equation (2.55) will be:

$$\rho(0) = [1^{2H} + (-1)^{2H}]\sigma^2 = 2\sigma^2. \tag{2.56}$$

We can observe that in this case the value in 0 of the ACF do not depend on *H*.

For *H*=0 the equation (2.55) will be:

$$\rho(k) = 0, \qquad \text{for} \qquad k \neq 0, \tag{2.57}$$

meaning that:

$$\rho(k) = \begin{cases} 2\sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{2.58}$$

which represent the autocorrelation of a white noise.

For *H* = 0 the fGn becomes a white noise, which does not manifest LRD.

In [15] the asymptotically behavior of the autocorrelation of a fGn is analyzed.

If Hurst parameter is between 0 and 0.5:

$$\rho(k) \sim H(2H-1)k^{2H-2}, \tag{2.59}$$

while if 0.5 <*H* <1, the autocorrelation has an asymptotically behavior, [15]:

$$\rho(k) \sim c_p k^{-\alpha}, \tag{2.60}$$

with $0 < \alpha < 1$ and c_p is a positive constant. At the border between the two asymptotically behaviors described in equations (2.59) and (2.60) lies the case : *H* = 0.5. In the case of equation (2.59):

$$0 < H < 0.5 \Longleftrightarrow 0 < 2H < 1 \Longleftrightarrow -2 < 2H - 2 < -1 \tag{2.61}$$

$$k^{2H-2} = \frac{1}{k^{2-2H}} < \frac{1}{k}. \tag{2.62}$$

The series with the general term k^{2H-2} described in (2.59) is convergent. In the case of (2.60):

$$k^{-\alpha} = \frac{1}{k^{\alpha}} > \frac{1}{k}. \tag{2.63}$$

In this case, the series with the general term the ACF $\rho(k)$ diverges:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty. \tag{2.64}$$

So, the convergence of the series with the general term the autocorrelation of the input signal depends on the values of the Hurst parameter. This is why analyzing the convergence of this series we can specify the interval in which the value of H is situated: (0, 0.5) if the series is convergent and (0.5, 1) if the series is divergent.

Finally, for *H* = 0, the fGn becomes a white noise and:

$$\rho(k) = 2\sigma^2\delta(k), \qquad \text{and} \tag{2.65}$$

$$\sum_{k=-\infty}^{\infty} \rho(k) = 2\sigma^2 \sum_{k=-\infty}^{\infty} \delta(k) = 2\sigma^2. \tag{2.66}$$

So, in this case the series with the general term the ACF, $\rho(k)$, converges as well.

When the ACF $\rho(k)$ slowly decays and the equation (2.62) is satisfied, we call the corresponding stationary process *X*(*t*) long-range dependent, [39]. On the contrary, short-range dependence is characterized by quickly decaying correlations. These two regions are separated by the case *H*=0.5. In this case, the series with the general term $\rho(k)$ is divergent. For 0 <*H* <0.5 the series with the general term $\rho(k)$ is convergent.

The LRD of a stationary process X_t can be also defined in terms of power spectral density [14], taking into account the Winer-Hincin theorem.

Definition 10. *A stationary process X_t presents LRD if its power spectral density satisfies the following relation:*

$$f(\lambda) \sim c_f|\lambda|^{-\beta}, \tag{2.67}$$

with $\lambda \to 0$, $\beta \in (0,1)$ and c_f is a constant.

The term $f(\lambda)$ represents the power spectral density of the stationary process and can be computed applying the Wiener-Hincin theorem:

$$f(\lambda) = \sum_{k=-\infty}^{\infty} \rho(k)e^{-ik\lambda}. \tag{2.68}$$

In this case, the relation between the Hurst parameter *H* and β is the following: $H = (1+\beta)/2$.

Definition 11. *A stationary process X_t presents LRD (seasonal long memory) with a pole at λ_0 if the power spectral density satisfies the following relation:*

$$f(\lambda) \sim c_f(cos\lambda - cos\lambda_0)^{-\beta}, \tag{2.69}$$

with $\lambda \to \lambda_0$, $\lambda_0 \in [0,\pi]$, $\beta \in (0,1)$ and c_f is a constant.

The expressions of the power spectral densities from equations (2.67) and (2.68) enable the estimation of Hurst parameter by spectral analysis. The major difficulty of this kind of Hurst parameter estimation lays in the fact that these power spectral densities are divergent around 0 for the equation (2.67) and around λ_0 for the equation (2.68). The standard spectral estimation techniques, based on Fourier transform, fail due to these divergences, but they can be successfully substituted by spectral estimation techniques based on wavelets, as we shall see in the following sub-sections.

2.8 The Estimation of Hurst Parameter

The Hurst parameter (*H*) characterizes a process in terms of the degree of self--similarity and LRD (the degree of persistence of the statistical phenomenon).

The degree of self-similarity and LRD increases as $H \to 1$, [59]. Theoretically, *H* must be between 0 and 1. A value equal to 0.5 indicates the lack of self-similarity or the presence of short-range dependence (SRD), [3]. A value of *H* smaller than 0.5 indicates that between the samples there is a SRD (the autocorrelation function is absolutely summable). A value greater than 0.5 indicates the existence of LRD (the ACF is not absolutely summable).

It is very important to know that Hurst parameter can not be calculated, it can only be estimated [40], because there are not mathematical methods to calculate the H in equation (2.55) for which the series with the general term the ACF is convergent. There are various statistical techniques to estimate Hurst parameter (*H*). By definition, the LRD phenomenon is related to the power-law behavior of certain second-order statistics (variance, covariance,...) of the process with respect to the durations of observation [3]. Many estimators of H are therefore based on the idea of measuring the slope of a linear fit in a log-log plot.

The Hurst parameter estimators can be classified into two categories: operating in the time domain and operating in the frequency domain.

- the estimators operating in the time domain are:
  - Rescaled Adjusted Range (R/S) Method,
  - Aggregated Variance Method,
  - Absolute Value method,
- the estimators operating in the frequency domain are the following:
  - Periodogram,
  - Whittle estimator,
  - Wavelet based LRD estimators.

2.8.1 Time Domain Estimators

The so-called variogram or R/S estimators are famous examples of the idea of measuring the slope of a linear fit in a log-log plot.

Rescaled Adjusted Range (R/S) Method

Proposed by Hurst in 1951, the R/S statistic is one of the oldest and better known methods for estimating the Hurst parameter, *H*, in a time series which presents LRD. For a selection of subsets of the time series A_i, starting at t_i and of size *n*+1, R/S statistic is defined as presented in the following equation:

$$\frac{R}{S}(t_i, n) = \frac{R(t_i, n)}{S(t_i, n)}, \tag{2.70}$$

where R represents the adjusted range of the considered series, A_i, and *S* is the sample variance of A_i.

The adjusted range $R(t_i, n)$ has the following physical interpretation. We suppose the time series A_i represents the amounts of water per time unit flowing into a reservoir. Furthermore, water flows out of the reservoir with a constant rate, such that the reservoir contains the same amount of water at the t_{i+n}^{th} time unit as at the t_i^{th} time unit. Then $R(t_i, n)$ is the minimum capacity of the reservoir such that it will not overflow in the period t_i to t_{i+n} inclusive.

The calculation of $R(t_i, n)$ proceeds as follows. Given t_i and *n*, and the mean:

$$\mu(t_i, n) = \frac{1}{n+1} \sum_{l=t_i}^{t_i+n} A_l, \tag{2.71}$$

the standard deviation can be expressed as follows:

$$S(t_i, n) = \sqrt{\frac{1}{n+1} \sum_{l=t_i}^{t_i+n} (A_l - \mu(t_i, n))^2}, \tag{2.72}$$

while $R(t_i, n)$ will be:

$$R(t_i, n) = \max_{0 \le l \le n} \left(\sum_{j=t_i}^{t_i+l} A_j - (l+1)\mu(t_i, n) \right) - \min_{0 \le l \le n} \left(\sum_{j=t_i}^{t_i+l} A_j - (l+1)\mu(t_i, n) \right). \tag{2.73}$$

The rescaled adjusted range is then just $R(t_i, n)/S(t_i, n)$. A single such calculation results in one point on a graph of $\log_{10} R(t_i, n)/S(t_i, n)$ against $\log_{10} n$. By varying t_i and *n* we obtain a plot of R/S. The size *n* is varied from 10 to about *N*/8 (*N* is the total sample size) in 5,000 logarithmically-spaced steps (except for small *n*, where several calculations of R/S are made for the same *n* and different t_i), [20]. The starting value t_i is chosen randomly in the range 1 to *N*-*n*. Finally, linear regression is used to fit a straight line to the R/S plot, the slope of this line being an estimate of *H*, [43].

Aggregated Variance Method

The variance-time plot method is one of the easiest methods used to estimate Hurst parameter. Being given a time series X_t with $t \in (1, N)$, it is divided into blocks of length m and aggregated as follows [14]:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \qquad k = 1, 2, ...N/m. \tag{2.74}$$

The variance is given by:

$$\widehat{var(X^{(m)})} = \frac{1}{N/m - 1} \sum_{k=1}^{N/m} (X^{(m)}(k) - \bar{X})^2. \tag{2.75}$$

The plots are obtained by computing $log(var(X^{(m)}))$ against $log(m)$ ("time") and by fitting a simple least squares line through the resulting points in the plane, ignoring the small values for *m*. If the estimate β of the asymptotic slope has values between -1 and 0 it means LRD, and an estimate for the degree of LRD is given by $H = 1 + \beta/2$, [43].

Absolute Value Method

Absolute value method uses different block sizes *m* for defining an aggregated series $X^{(m)}$. The absolute moment of a discrete time series X_t is defined as:

$$AM^m = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \bar{X}|, \tag{2.76}$$

where $X^{(m)}(k)$ is the aggregate series of level m [53].

The log-log plot of AM^m versus m, for varying levels m, should result in a straight line with slope of *H*-1, if the data are LRD, [40]. The slope is computed using a least squares regression on the points.

2.8.2 Frequency Domain Estimators

LRD determines the spectrum of a process to behave as a power law for frequencies close to 0. Therefore, it is normal to think of using spectral estimation to measure H parameter.

Periodogram

Periodogram method plots the logarithm of the spectral density of a time series *X*, versus the logarithm of the frequency. In the case in which X is a long-range dependent random process, this plot becomes a line. The slope provides an estimate of *H*, [40]. The periodogram is given by:

$$I(\nu) = \frac{1}{2\pi N} |\sum_{j=1}^{N} X(j) e^{ij\nu}|^2, \tag{2.77}$$

and represents an estimator for the spectral density of X, where ν is the frequency and *N* is the length of the time series *X*.

The periodogram is a standard estimator for the power spectral density. Taking into account the limitations of the Fourier transform already presented in Chapter 1, this estimator can be improved by substituting the Fourier transform with the Short Time Fourier Transform (STFT). A new spectral estimator is obtained by averaging smoothed periodograms computed with STFT on different sequences of data:

$$E(\omega) = \sum_{k=1}^{P} |x(t - kL) \cdot w_L(t) e^{j\omega t} dt|^2, \tag{2.78}$$

where P is the number of data pieces, L their length and w_L a weighting window.

When applied to long-range dependent data, such a spectral estimator results in an estimator of H based on a linear fit in a $log(\omega)$ versus $logE(\omega)$ plot, which is strongly biased. This happens because the constant-bandwidth spectral analysis performed does not match with the structure of the power spectral density of a long-range dependent process. The wavelet based spectral estimators, which will be presented in the following, perform a constant relative bandwidth spectral analysis that matches with the structure of the power spectral density of a long-range dependent process, [3].

Whittle Estimator

To diminish the drawbacks of the H estimator based on periodogram already mentioned, Whittle proposed the minimization of a likelihood function, which is applied to the periodogram of the time series. It involves to find a function $f(\nu;\eta)$ which minimize the expression:

$$Q(\eta) = \int_{-\infty}^{\infty} \frac{I(\nu)}{f(\nu;\eta)} d\nu + \int_{-\infty}^{\infty} \log f(\nu;\eta) d\eta, \tag{2.79}$$

where η is the vector of unknown parameters and $I(\nu)$ is the periodogram. The minimization is performed for the variable η. By normalizing $f(\nu;\eta)$, the term $\int_{-\infty}^{\infty} \log f(\nu;\eta) d\eta$ becomes equal to 0. Finding the value of η which minimizes Q, the unknown parameters and the function f are identified. Substituting the expression of the periodogram (which depends on H), computed with one of the methods already presented, in the first term of the right hand side of equation (2.79), the integral (which will depend on H as well) could be computed and the minimum value of Q (which depends on H as well) will be found. This is the reason why the Whittle estimator can be used to estimate H. The Whittle estimator is defined as the value of η that minimize *Q*, [66].

The Discrete Whittle estimator

The MLE gives a coherent approach to estimator design, which is capable of producing an unbiased, asymptotically efficient estimator for H. Proposed by Whittle in 1953, the Whittle estimator consists of two analytic approximations to the exact Gaussian MLE, in order to avoid the huge computational complexity of the exact algorithm, [3].

The first approximation basically replaces the covariance matrix by an integral of a function of the spectrum. Because the computational difficulties remain after this approximation, a second approximation is performed. It consists in the discretization of the frequency-domain integration rewritten in terms of the periodogram. Performing the two approximations, the Discrete-Whittle (D-Whittle) estimator is obtained.

The D-Whittle estimator relies on the periodogram, inheriting the structure of the Whittle estimator (which relies on the periodogram as it was already said) and the periodogram has a low computational cost. However, a minimization procedure is involved. This procedure requires many iterative evaluations, resulting a higher overall cost. Furthermore, problems of convergence to local minima may be found.

The use of periodogram makes the D-Whittle estimator asymptotically unbiased only. This asymptotic comportment is not enough for a good estimator, which must be unbiased, robust and efficient.

Wavelet based LRD estimators

Abry and Veitch [3] proposed a Hurst parameter estimator at each scale of the wavelet decomposition of the random process which must be analyzed. The m^{th} scale estimation is realized computing the expectation of the random variable:

$$e_m = \frac{1}{n_m} \sum_{k=1}^{n_m} |d_m[k]|^2, \tag{2.80}$$

where n_m represents the length of the sequence of wavelet coefficients obtained at the m^{th} decomposition level.

For the beginning, let us consider that the input random process is wide sense stationary. The Abry-Veich estimator takes the form:

$$E\{e_m\} = \frac{1}{n_m}\sum_{k=1}^{n_m} E\left\{d_m^2[k]\right\}. \tag{2.81}$$

For stationary input random processes, the expectation of the wavelet coefficients square represents their variance, which is constant. Denoting this constant as σ_m^2, the previous equation becomes:

$$E\{e_m\} = \frac{1}{n_m} n_m \sigma_m^2 = \sigma_m^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}\{R_x(t)\}(2^{-m}\nu)|\mathcal{F}\{\psi\}(\nu)|^2 d\nu. \tag{2.82}$$

with the aid of equation (2.46).

For a random input process, continuous in time and with LRD:

$$\mathcal{F}\{R_x(\tau)\}(\nu) = c_f|\nu|^{1-2H}, \tag{2.83}$$

which substituted in (2.81) will give:

$$\begin{aligned} E\{e_m\} = \sigma_m^2 &= \frac{c_f}{2\pi}\int_{-\infty}^{\infty} |\nu 2^{-m}|^{1-2H}|\mathcal{F}\{\psi\}(\nu)|^2 d\nu = \\ &= \frac{c_f}{2\pi} 2^{m(2H-1)}\int_{-\infty}^{\infty} |\nu|^{1-2H}|\mathcal{F}\{\psi\}(\nu)|^2 d\nu. \end{aligned} \tag{2.84}$$

This estimator can be used in practice if the integral from the right side is convergent.

If the mother wavelet (MW) ψ is selected from the Daubechies family of MWs, then it has a finite effective bandwidth ω_m:

$$\int_{-\infty}^{\infty} u^2|\mathcal{F}\{\psi\}|^2 du < \infty \tag{2.85}$$

Hence, the integral in equation (2.84) can be written as:

$$\int_{|u|<\omega_m} |u|^{1-2H}|\mathcal{F}\{\psi\}(u)|^2 du. \tag{2.86}$$

The convergence problem can appear only around the point $u = 0$, if:

$$\lim_{u\to 0}|u|^{1-2H}|\mathcal{F}\{\psi\}(u)|^2 = \infty. \tag{2.87}$$

Fortunately, Daubechies MWs have a positive number of vanishing moments, N_ν, which means that:

$$\forall k = 0, 1, ..., N_\nu - 1, \qquad \int_{-\infty}^{\infty} t^k \psi(t) dt \equiv 0. \tag{2.88}$$

Taking in consideration the derivation in the frequency domain property of the Fourier transform, the previous equation can be written as:

$$\forall k = 0, 1, ..., N_\nu - 1, \qquad \frac{1}{j^k}\frac{d^k\mathcal{F}\{\psi(t)\}(\omega)}{d\omega^k}\Big|_{\omega=0} = 0. \tag{2.89}$$

Hence, the Fourier transform and its $N_\nu - 1$ derivatives vanish in $u = 0$.

It can be decomposed around $u = 0$:

$$\mathcal{F}\{\psi(t)\}(u) = \sum_{k=N_\nu}^{\infty} \frac{u^k}{k!} \frac{d^k \mathcal{F}\{\psi(t)\}(u)}{du^k}\Big|_{u=0}. \tag{2.90}$$

obtaining a polynomial with the degree higher or equal to N_ν.

Hence, the power spectral density of MW, $|\mathcal{F}\{\psi(t)\}(u)|^2$ behaves around the point $\nu = 0$ as a polynomial with the degree higher or equal to $2N_\nu$. So, the condition (2.87) is not accomplished if:

$$2N_\nu > 2H - 1 \Longleftrightarrow N_\nu > H - \frac{1}{2}. \tag{2.91}$$

The LRD detection supposes that:

$$\frac{1}{2} < H < 1. \tag{2.92}$$

So, the integral in equation (2.84) is convergent for any MW from Daubechies family, because for this family:

$$\min_{\nu}\{N_\nu\} = 2. \tag{2.93}$$

By making the substitution $m \to m - 1$ in equation (2.84) it results:

$$E\{e_{m-1}\} = \sigma^2_{m-1} = \frac{c_f 2^{m(2H-1)} \int_{-\infty}^{\infty} |u|^{1-2H} |\mathcal{F}\{\psi\}(u)|^2 du}{2^{2H-1}} = \frac{\sigma^2_m}{2^{2H-1}}, \tag{2.94}$$

or:

$$\frac{\sigma^2_m}{\sigma^2_{m-1}} = 2^{2H-1}. \tag{2.95}$$

Applying the logarithm in the previous equation, an estimator of the Hurst parameter is obtained:

$$2\hat{H} - 1 = log_2\hat{\sigma}^2_m - log_2\hat{\sigma}^2_{m-1} \Longleftrightarrow \hat{H} = \frac{1 + log_2\hat{\sigma}^2_m - log_2\hat{\sigma}^2_{m-1}}{2}. \tag{2.96}$$

This is a very simple H estimator, which requires only the wavelet decomposition of the input process and the estimation of the wavelets coefficient variances at two successive scales. For non-stationary input random processes this estimator is improper, because the local variance of wavelet coefficients is not longer constant. In this case, the random variable from equation (2.80) must be used. Abry and Veitch [3] proposed the following estimator:

$$log_2\left(\frac{1}{n_m}\sum_{k=1}^{n_m} E\left\{d^2_m[k]\right\}\right) = (2\hat{H_{AV}} - 1)m + \hat{c} \tag{2.97}$$

where $\hat{c}$ estimates the constant $log_2(c_f \int_{-\infty}^{\infty} |\nu|^{1-2H} |\mathcal{F}\{\psi(t)\}(\nu)|^2 d\nu)$.

Performing a weighted least squares fit between the scales m_1 and m_2 yields the following explicit formula:

$$\hat{H}_{AV}(m_1, m_2) = \frac{1}{2}\left[\frac{\sum_{m=m_1}^{m_2} s_m m\eta_m - \sum_{m=m_1}^{m_2} s_m m \sum_{m=m_1}^{m_2} s_m\eta_m}{\sum_{m=m_1}^{m_2} s_m \sum_{m=m_1}^{m_2} s_m m^2 - (\sum_{m=m_1}^{m_2} s_m m)^2} + 1\right], \tag{2.98}$$

where

$$\eta_m = log_2(\frac{1}{n_m}\sum_{k=1}^{n_m} d^2_m[k]), \tag{2.99}$$

and the weight $s_m = (n \cdot ln^2 2)/2^{m+1}$ is the inverse of the theoretical asymptotic variance of n_m and n represents the length of the entire sequence of wavelet coefficients [3].

Stationarity hypothesis testing is difficult in the presence of LRD, where many classical statistical approaches cease to hold.

The estimator (2.98) was treated in [3] for non-stationary random processes with LRD which have stationary increments as the fBm processes. This kind of non stationarity appears frequently in practice. For example in the context of Ethernet traffic the first thing to note is that data is not stationary and this is caused by the hidden periodicities or of the diurnal cycle, lunch breaks etc. On the other hand, it is reasonable to expect that for smaller timescales where network conditions are relatively stable, stationarity will be a natural and useful assumption. So, Ethernet traffic is stationary at same scales and not stationary at other scales. For this reason were considered the scales m_1 and m_2 in the design of Abry-Veitch estimator. Finally, the input random process could have overall trends. These trends represent the last source of non-stationarity for the input random process.

The performance of Abry-Veitch estimator is analyzed in [3]. It is an unbiased estimator, robust and efficient which requires less computational resources than other H estimators, because it uses the DWT which can be computed very fast (multiple of $\mathcal{O}(N)$), for example Whittle estimator (which requires more computational complexity due to its recursive nature). Its efficiency comes from the fact that it attains the Cramer-Rao bound.

In the design of the Abry-Veitch estimator is assumed that a continuous-time random process, $x(t)$, is available. There are numerous cases where only discrete time observations of the input process are available. In the following these observations will be denoted as $x[1], x[2], ..., x[N]$. For continuous-time random processes, the wavelet coefficients are computed using the equation:

$$d_m[k] = \langle x(t), \psi_{m,k}(t) \rangle = \int_{-\infty}^{\infty} x(t)\psi_{m,k}(t)dt. \tag{2.100}$$

The integral from the previous equation can not be computed if only discrete observations of the process $x(t)$ are available. So, this integral must be discretized:

$$\begin{aligned} d_x[m,k] &= \sum_l x[l]\psi_{m,k}[l] = \sum_l x[l]2^{-\frac{m}{2}}\psi(2^{-m}l - k) = \\ &= 2^{-\frac{m}{2}} \sum_l x[l]\psi(\frac{l}{2^m} - k) = \frac{1}{\sqrt{n_m}} \sum_{l=1}^{N} x[l]\psi(\frac{l}{n_m} - k). \end{aligned} \tag{2.101}$$

With the change of variable $l = n_m(p + k)$, the last equation becomes:

$$d_x[m,k] = \frac{1}{\sqrt{n_m}} \sum_{p=n_m^{-1}-k}^{Nn_m^{-1}-k} x[n_m(p+k)]\psi(p) = \sum_{p=n_m^{-1}-k}^{Nn_m^{-1}-k} a_m[p]x[n_m(p+k)], \tag{2.102}$$

where $a_m[p]$ is a collection of discrete filters coefficients determined by the MW used [37].

The wavelets coefficients $d_m[k]$ and the coefficients $d_x[m,k]$ have similar properties. One can replace $d_m[k]$ with $d_x[m,k]$ and use the Abry-Veitch estimator, [2]. The resulting quantities:

$$e_N(n_m) = \frac{1}{N_m} \sum_{l=1}^{N_m} d_x^2[n,k], \tag{2.103}$$

are called generalized quadrature variations of the process $x[l]$ and their substitution in the expression of the Abry-Veitch estimator give a new H estimator, which can be called discrete Abry-Veitch estimator, or generalized quadrature variations Hurst parameter estimator based on wavelets, because one of its parameters is the mother wavelets selected for its implementation. The performance of those Hurst estimators is analyzed in [2]. They are consistent and have asymptotic normality. The complexity of the corresponding algorithms is $\mathcal{O}(N)$, the same as the complexity of the Abry-Veitch estimator.

This class of estimators is robust against non-stationarity. They were tested for fBm and linear fractional stable processes in [2]. This is the class of Hurst parameter estimators which are the most appropriate the solve the problem of WiMAX traffic which represent the subject of Chapter 4 of this thesis. WiMAX traffic data, which represent the subject of this thesis, are discrete observations of a continuous-time random process, which is non-stationary, because it has overall trend as it will be proved in Chapter 3. So, for LRD analysis of WiMAX traffic, the best class of Hurst parameter estimators based on wavelets, seems to be the generalized quadrature variations estimators class.

2.9 Conclusions

In this chapter we aimed to present some elementary statistics and we introduced some basic ideas of time series analysis and forecasting that will be used in the following chapters. We introduced an important parametric family of stationary time series, ARMA processes which are frequently used in the modeling of time series, due to their high generality. A generalization of this class, which incorporates a wide range of non-stationary series, is provided by the ARIMA processes. ARIMA models are flexible and can be applied to a wide spectrum of time series analysis. They are used for: financial, meteorological or derived from man made activities time series. Finding an appropriate model implies the order selection and parameters estimation.

The estimation methods presented in this chapter enable us to find, for given values of *p* and *q*, an ARMA(*p*,*q*) model to fit a given series of data. A number of different procedures can be employed to test whether the selected model is really a statistically sufficient description of the time series.

In the last few years measurements of various types of network traffic proved that the traffic exhibits LRD and self-similarity. A key parameter characterizing self-similar processes is the Hurst parameter H. Thus, an overview of the theory and methods developed to deal with long-range dependent data were presented in this chapter.

One of the theoretical contributions of this thesis is the second order DWT statistical analysis presented in section 2.6, in equations (2.27) to (2.49). It permits to understand how does the sequences of wavelet coefficients obtained by computing the DWT of a wide sense stationary random process look like. The mean, variance and autocorrelation of those sequences are computed in general and for the particular case of input white Gaussian noise. An asymptotic analysis, proving the decorrelation effect of the DWT is also reported. A further research direction is the generalization of this statistical analysis for non-stationary random signals.

A very simple Hurst parameter's estimation method, based on the previously mentioned second order DWT statistical analysis, is proposed in equation (2.95). This estimation method works for second order wide sense stationary random processes. It was simply generalized to the Abry-Veitch Hurst parameter's estimator which works

for non-stationary continuous in time random processes. Next this estimator was discretized obtaining the generalized quadrature variations Hurst parameter's estimator based on wavelets. This is another theoretical contribution of this thesis.

The selection of the most appropriate mother wavelets for the computation of the DWT implied in the Hurst parameter estimation method must be made based on equation (2.93). The superiority of Hurst parameter's estimation method based on wavelets against other methods will be proved by simulation in Chapter 4.

Chapter 3

Time-series Mining. Application to Forecasting

Time-series forecasting is an important area of forecasting where the historical values are collected and analyzed in order to develop a model describing the behavior of the series. There are examples of communication signals which represent time-series. A typical example is the traffic developed in a communication network. To support the growth of demands, communication companies are investing in new technologies to improve their services. However, in the case of permanent growing of the demands, in order to assure the users supply, a network capacity planning tool should be used. For capacity planning purposes, one only needs to know the traffic baseline in the future along with possible fluctuations of the traffic around this particular baseline. The communication services providers should anticipate future demands and should know where and when the upgrades must be done in the network. This requirement is even more important in wireless networks. There are three modern technologies for the wireless networks: Wi-Fi, WiMAX and LTE.

Worldwide interoperability for Microwave Access (WiMAX) is a telecommunication technology based on IEEE 802.16 standard, capable of delivering advanced IP applications, such as voice, video, and data over the microwave spectrum RF to stationary and moving users.

A fundamental question about this technology is: How do the technologies compare in terms of prioritizing traffic and controlling quality? A partial response to the previous question can be given by studying the traffic forecasting methodology for the WiMAX technology. The amount of traffic through a BS can not be higher than the capacity of that BS. If the amount of traffic approaches the capacity of the BS, then the BS saturates and a lot o messages are lost. So, it is necessary to know the capacity of the BS.

By assuming a deployment scenario –e.g., available bandwidth in MHz per cell, distribution of various user types, and application breakdown - it is then possible to calculate the total traffic volume of a BS.

Actual data rates, namely the throughput provided by the BSs throughout the cell and experienced by users, depend on several factors including user distribution and propagation conditions and pilot distribution and will need to be taken into account. More, different users can use different modulation techniques in the same interval of time. In addition, the QoS is not constant with the amount of the BS capacity used. Lately, there is a significant increase in the need for delivering multimedia-based services to home residences and business premises.

One of the most important attributes of a WiMAX network is its ad-hoc nature.

Any user localized into a cell of the network must obtain the access any time. So, the number of users is not a priori known, neither the amount of traffic in a cell.

3.1 Related Work

Time series forecasting has always been a challenging issue for many researchers. Recently, many approaches involving time series models have been used for traffic forecasting such as pure statistical or based on neural networks [55].

For more than two decades, Box Jenkins ARIMA technique has been widely used for time series forecasting. This class of models is used to build the time series model in a sequence of steps which are repeated until the optimum model is achieved. The Box-Jenkins models can be used to represent processes that are stationary or non-stationary.

As it was shown in Chapter 2, there are a lot of classical linear predictive models: Auto Regressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) which deal with stationary series or Autoregressive Integrated Moving Average (ARIMA) which deals with non-stationary series. All these models were already used for communications traffic prediction. For example, in [54], [5], [11] and [32] ARIMA models are used for traffic prediction. In [62] the authors propose a prediction algorithm based on the Auto-Regressive Moving Average (ARMA) model and the Markov-Modulated Poisson Process (MMPP). Fractional ARIMA models are used to predict traffic in [22]. The communication traffic forecasting could have different goals as for example: the anticipation of the following pick of traffic or the estimation of the moment when one of the features of the traffic will allow a given condition. If in the first case a short range prediction is required, in the second case a long range prediction seems to be more appropriate. So, the selection of the linear model must be made in accordance with the application. Another solution for the traffic forecasting is the use of neural networks.

The authors of [16] propose a Neural Network (NN) approach to predict TCP/IP traffic for all links of a backbone network. The data collected from the United Kingdom Education and Research Network (UKERNA) was recorded into two datasets (every 10 minutes and every hour), between 12 AM of 14th June 2006 and 12 AM of 23th July 2006. The data was analyzed using two forecasting types (or scales): real-time (every 10 minutes) and short-term (hourly values). The equipments of the networks considered in the examples already presented were connected through cables. Some papers published recently, present cases of wireless traffic forecasting. Neural networks are also used in [30] where the wireless network traffic is predicted for short time scale. Methods based on the use of Artificial Neural Networks (ANN) for traffic forecasting are also presented in [57], [60] and [59].

According to the results presented in these papers we can conclude that ANN performs better then the other forecasting techniques for small future time intervals, several weeks at most. But if the goal of the forecasting method is to predict the moment when a feature of the traffic (as for example its overall tendency) will allow a specific condition (as for example the saturation of a base station), meaning prediction for several months, than pure statistical models are the ones that should be taken into consideration, because the performance of NNs deteriorates in the absence of training. Both forecasting methods (based on linear predictive models or based NNs) can be accelerated if they are applied in the wavelet domain, taking into consideration the sparsity of the wavelet coefficients.

The wavelet transform has been frequently used for time series analysis and forecasting in recent years [54]. Wavelets can localize data in time-scale space. At high scales, wavelets have a small time support and can "catch" discontinuities or singularities, while at low scales the wavelets have a larger time support and can identify periodicities. Wavelets are able to characterize the physical properties of the data. At low scales, the wavelets identify the long-term trend of the data. By increasing the scale, the wavelets begin to reveal the details of data, zooming in on its behavior at a moment of time.

A paper in which the authors proposed to model the traffic evolution in a IP backbone network at large time scales is [54]. The authors combined the wavelet analysis and statistical data processing and developed models for long-term forecasting for capacity planning purposes. A combination between wavelet analysis and traffic foresting is made also in [70].

Inspired by [54] this chapter proposes a methodology to build forecasting models for WiMAX traffic. The goal of the forecasting methodology proposed in [54] was to predict the moment when a part of a communication network, in which the equipments are connected by cables, will saturate. The forecasting methodology proposed in [54] supposes the utilization of ARIMA models in the wavelet domain to estimate two features (the overall tendency and the variability) of the time-series belonging to a network traffic data base. In Figure 3.1 are shown the main steps followed in [54].

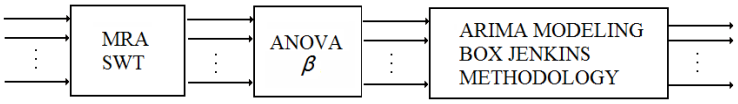

Figure 3.1: The forecasting methodology proposed in [54].

Our purpose is to adapt the forecasting algorithm proposed in [54] to the case of a WiMAX network. In Figure 3.2 are presented the main steps followed in the case of WiMAX traffic. A series of modification have been made in order to adapt the methodology proposed in [54]. These modifications will be highlighted in the following sections of this chapter.

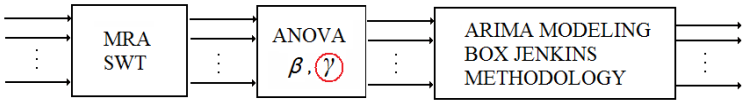

Figure 3.2: The forecasting methodology in the case of WiMAX traffic.

A prediction of the moment when a BS will saturate can be realized using both estimated features for the traffic which corresponds to that BS. So, the traffic forecasting methodology, which will be applied in this chapter, analyzes the elements of the database (for knowledge discovery purposes) and extracts two features of those time-series. The data features extraction represents a common operation in the data mining field. This methodology is based on statistical data processing in the field of wavelets and follows CRISP-DM [21] phases, as will be shown in the following section.

3.2 Phases of a Data Mining Project

Knowledge Discovery is a domain that searches new knowledge about an application domain. One of its branches is Data mining which is an analytic process designed to explore and to extract useful information from large volume of data.

According to CRISP-DM the process has several steps, [21]:

1. Business understanding,
2. Data understanding,
3. Data preparation,
4. Modeling,
5. Evaluation,
6. Deployment.

The succession of those phases and their interdependence are represented in Figure 3.3.

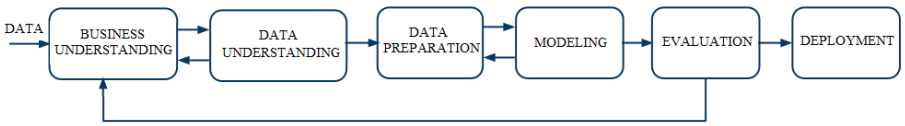

Figure 3.3: Phases of a data mining project.

The first step of the data mining project is to understand the application in which it is involved (business understanding). This goal is implemented iteratively, by collaboration with other phases as data understanding or evaluation. The second step of the data mining project is data understanding. It is based on the first step and has an iterative implementation as well. The data understanding goal can be seen as a feedback for the business understanding goal. A first business understanding allows a preliminary data understanding. With this acquired knowledge, the process of business understanding is improved and as a consequence, the data understanding process is improved as well. The third phase of the process is data preparation. Generally, the raw data are affected by inaccuracies of sensors and acquisition systems, for example the traces from data bases contain missing data. For this reason, the phase of data preparation is required in a data mining project. This phase consist in putting the data in the most appropriate form for the subsequent phase of the data mining project, which is the modeling phase. As in the case of the pair of phases composed by the business understanding phase and the data understanding phase, the data preparation and the modeling phases are interdependent. The modeling step is one of the most important for the data mining project, because it allows the representation of data in a form favorable for the extraction of some features useful for the application considered. The evaluation phase is very important as well. It permits the appreciation of the quality of the model selected. It has also a regulatory function in the data mining project, influencing its first phase, the business understanding. The last step of the CRISP-DM project consists in deployment. In the following are presented details about these phases and how are they implemented for WiMAX traffic

prediction based on the forecasting algorithm proposed in [54]. This algorithm was conceived for the prediction of the moment when the amount of traffic will produce the saturation of a router from a given node of a wired network. We have adapted this algorithm for the case of WiMAX traffic. The phases of the new algorithm are: MRA of traffic traces, selection of the most important resolutions for the extraction of the overall tendency and of the variability of the WiMAX traffic, ARIMA modeling of those two features, models validation using the Box-Jenkins methodology and extraction of an estimation of the moment when the corresponding BS will saturate. In the following are presented details about the correspondence between the phases of the proposed algorithm and the phases of a data-mining project, already mentioned.

3.2.1 Business Understanding

The first phase of a data mining project involves: understanding the objectives and the requirements of the project, the problem definition and designing a preliminary plan to achieve the objectives. The objective of the proposed algorithm is to predict when upgrades of a given BS have to take place. We compute an aggregate demand for each BS and we look at its evolution at time scales larger than one hour. The requirement of the project is to do this prediction fast and precise. We have chosen the forecasting methodology proposed in [54] and our preliminary plan was to adapt this methodology to the case of a WiMAX network.

3.2.2 Data Understanding

Data understanding phase implies collecting initial data, describing and exploring data.

In our case, the data was obtained by monitoring the traffic from sixty-seven BSs composing a WiMAX network. The duration of collection is of eight weeks, from March 17th till May 11th, 2008. Our database is formed by numerical values representing the total number of packets/bytes from uplink and downlink channels, for each of the 67 BSs. The values were recorded every 15 minutes, so it can be easily deduced that for a given BS we have 96 samples/day, 672 samples/week, and a total number of 5376 samples. We observed that for the 34th BS from the 7th week (April 28-May 5) there is a significant loss of data due to an error at the BS's level so we decided to remove this BS from our analysis. Hence, we deal with sixty-six traces in uplink and sixty-six traces in downlink, eight weeks long.

The traces from the database are accessible in two formats corresponding to two measures of the traffic, in bytes and in packets. For our simulations we will analyze the traffic measured in packets (because it is simpler to handle time-series with smaller values of samples), corresponding to uplink and downlink channel. For the application under consideration, which consists in estimating the moment when each BS's traffic becomes comparable with the BSs' capacity, it is more important the downlink channel (where the traffic has a higher volume). Therefore, the results presented in the following correspond to downlink channel. The risk of saturation of the BS in uplink is considerably smaller.

3.2.3 Data Preparation

This phase includes selecting data to be used for analysis, data clearing, such as identification of the potential errors in data sets, handling missing values, and removal of noises or other unexpected results that could appear during the acquisition process.

The incomplete or missing data constitute a problem. Despite the efforts made to reduce their occurrence, in most cases missing values cannot be avoided. If the number of missing values is big, the results are nor relevant. It is therefore essential to know how to minimize the amount of missing values and which strategy to select in order to handle missing data.

There are several strategies of handling missing data, for example delete all instances where there is at least one missing value, replacing missing attribute values by the attribute mean or to estimate each of the missing values using the values that are present in the dataset (interpolation) [58]. There are many different interpolation methods such as linear, polynomial, cubic or nearest neighbor interpolation. We choose the cubic interpolation because for some BSs the missing values are situated on the first/last position of the vector and this fact forbids us to use, for example, the linear interpolation.

At this stage, the input data is also analyzed in order to find if it contains periodicities. The simple plot of the traffic curves proved the existence of periodicities in the traffic. In Figure 3.4 is presented a signal representing the traffic evolution during one week for a given BS randomly selected.

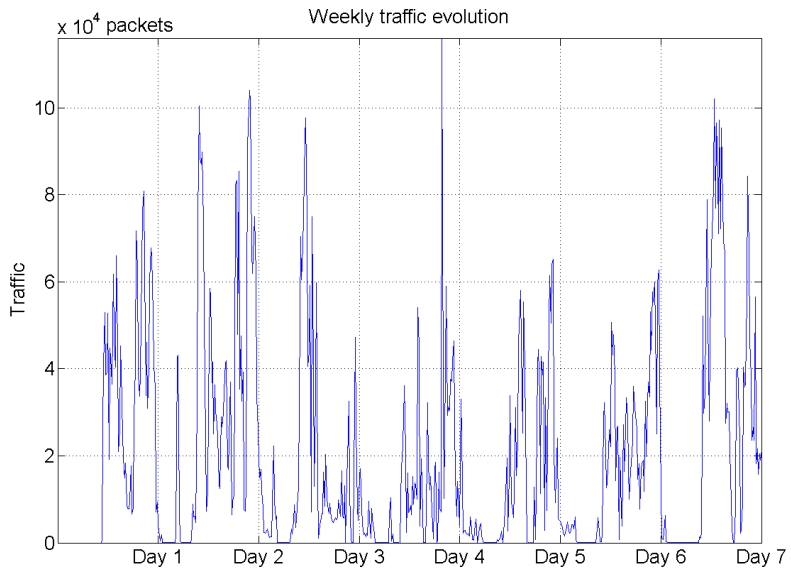

Figure 3.4: A curve describing the weekly traffic evolution for a BS arbitrarily selected.

In order to verify the existence of periodicities we calculate the Fourier transform of the signal and we analyze the power spectral density in Figure 3.5.

Looking at Figure 3.5 we can remark the eight harmonic. The sampling step used

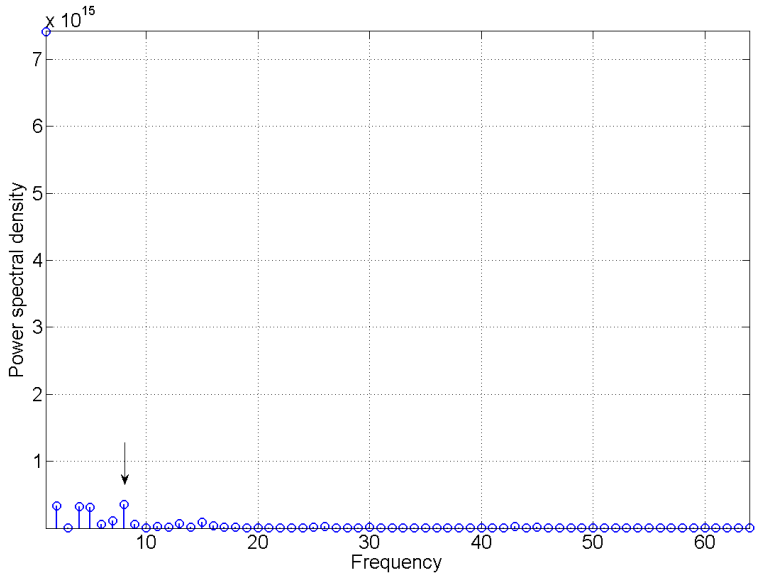

Figure 3.5: The power spectral density of the signal from Figure 3.4.

has a value of 15 minutes. It corresponds to a sampling frequency of 1.1 mHz. So, the maximal frequency contained in the analyzed power spectral densities equals half of the sampling frequency, 0.55 mHz.

The representation contains 670 values. Hence the fundamental frequency of the representation equals 0.55 mHz/335 (the value 355 was obtained by dividing the total number of samples with 2). The frequency of the eighth harmonic equals 0.013 mHz. The corresponding period is equal with 76923.07 s, or 1282.05 minutes or 21.36 hours (near 24 hours). The period corresponding to the ninth harmonic equals 3,69 hours. Hence, we can associate the eight harmonic with a period of 24 hours, because this value is closer of 21.36 hours than 3,69 hours. So, the results in previous figure indicate that one of the most dominant periods across all traces is the 24 hours one. The trace in Figure 3.5 was arbitrarily chosen. There are other traces in the data base for which the eight harmonic is dominant, being several times bigger than the other harmonics and proving the periodicity with the period of 24 hours of the traffic. However, depending on trace, the periodicity with the period of 24 hours can be also hidden. This is for example the case of the BS 2 showed in Figure 3.6. This periodicity has social reasons reflecting a pattern of diurnal comportment of the network. Such seasonal behavior is commonly observed in practical time-series. In the rest of this chapter we will explain the phases of the forecasting algorithm using an example of a particular trace (corresponding to BS1).

Next, we consider a traffic curve recorded during eight weeks represented in Figure 3.7 with blue. The long-term trend (red line) and the deviations from the long-term trend (the green and the black lines) are also shown in Figure 3.7.

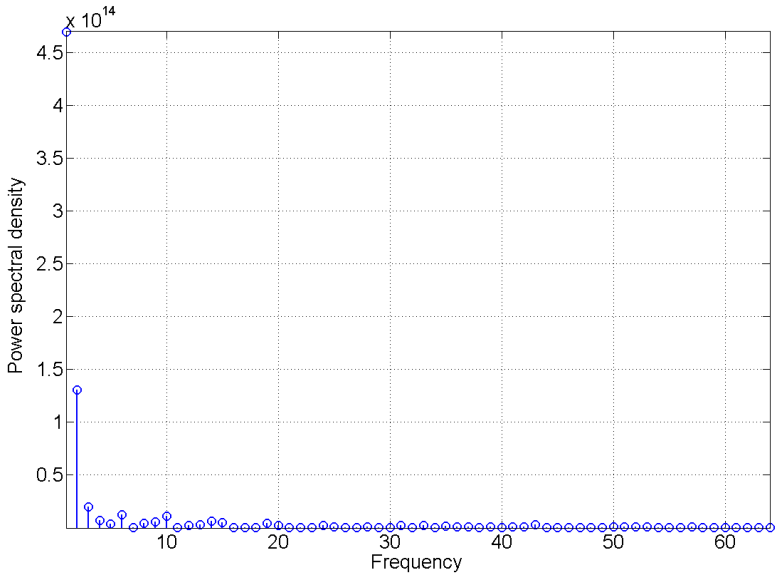

Figure 3.6: The power spectral density of the traffic trace corresponding to BS2.

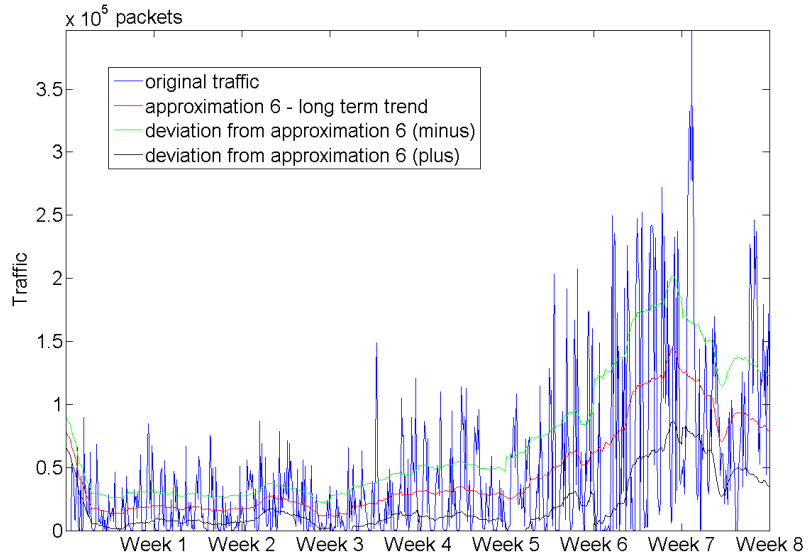

Figure 3.7: A traffic curve recorded during 8 weeks, its long term trend (approximation 6) and the deviations from sixth approximation.

The curve contain specific underlying overall trends, represented in red. The curve in blue describes the traffic evolution measured in number of packets/s for a BS1, during eight weeks. The other two curves show the deviation, plus (in green)/minus (in black), from the signal approximation. It can be observed that a large part of the traffic is contained between the green and black lines. The red line indicates an increasing of the traffic in time suggesting the possibility of saturation of the corresponding BS1.

Next, we propose a multi-timescale analysis. We used the SWT to decompose the original signal into a range of frequency bands. The level of decomposition (n), depends on the length of the original signal. For a discrete signal, in order to be able to apply the SWT, if the decomposition at level n is needed, 2^n must divide evenly the length of the signal. The n^{th} level of decomposition, gives us $n + 1$ signals for processing: one approximation signal corresponding to the current level and n detail sequences corresponding to each of the n decomposition levels. The n approximation sequences compose a multiresolution analysis (MRA). The value of n gives the maximal number of resolutions which can be used in the MRA. It corresponds to the poorer time resolution.

There is shown that WiMAX traffic exhibits some periodicities which are better noticed if we modify the sampling interval from 15 minutes to 90 minutes. So, by temporal decimation with a factor of six, these time series can be transformed in signals at a temporal resolution of 1.5 hours. This represents the highest time resolution which is used in the proposed MRA. Further on these temporal series will be denoted by $ca_{sd}(t)$. The derived temporal series $ca_{sd}(2^p t)$ have a temporal resolution of $2^p * 1,5$ hours. To extract the overall trends of the traffic time series, the MRA of the temporal series $ca_{sd}(t)$ using temporal resolutions between 1.5 and 96 hours is done.

We used Shensa's algorithm (which corresponds to the computation of the SWT with six levels of decomposition). In this case the utilization of decimators (required for the computation of DWT) is avoided but at each iteration different low-pass and high-pass filters are used. The impulse responses of the filters from the second iteration are obtained by sub-sampling the impulse responses of the filters from the first iteration and so on. At each temporal resolution two categories of coefficients are obtained: approximation coefficients and detail coefficients. In Figure 3.8 are shown approximation coefficients for six level of decomposition. It can be observed that with the increasing of the level of decomposition, the sequence of approximation coefficients becomes more smoothed. The first sequence of approximation (approximation 1) contains very rapid and high oscillations. The sequence corresponding to the sixth approximation is much smoothed and does not contain any rapid oscillation. Preliminary simulations presented in a research rapport afferent to a contract developed by our department for Alcatel-Lucent Timisoara prove that the overall trend of the traffic time series is better highlighted by the approximation coefficients obtained at the time resolution of 96 hours (corresponding to the sixth decomposition level), c_6. In the data-mining context, the separation of the last sequence of approximation coefficients obtained based on a MRA can be regarded as a data preparation operation, because the form of this sequence is appropriate for modeling the overall tendency of the traffic with the aid of linear predictive models.

The goal of using the MRA in our work is to extract the overall trend of the temporal series that describes the traffic under analysis with the aid of the approximation coefficients, and to extract the variability around the overall trend with the aid of some detail coefficients.

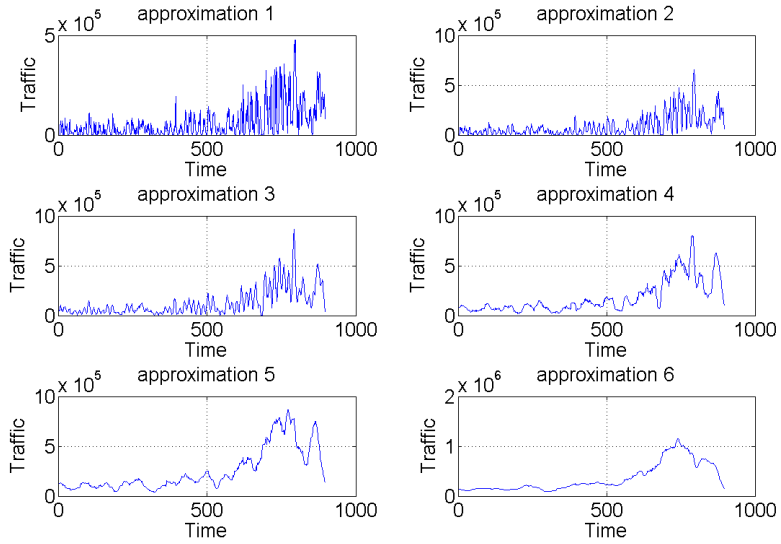

Figure 3.8: The approximation coefficients.

The six detail coefficients obtained after decomposition are depicted in Figure 3.9.

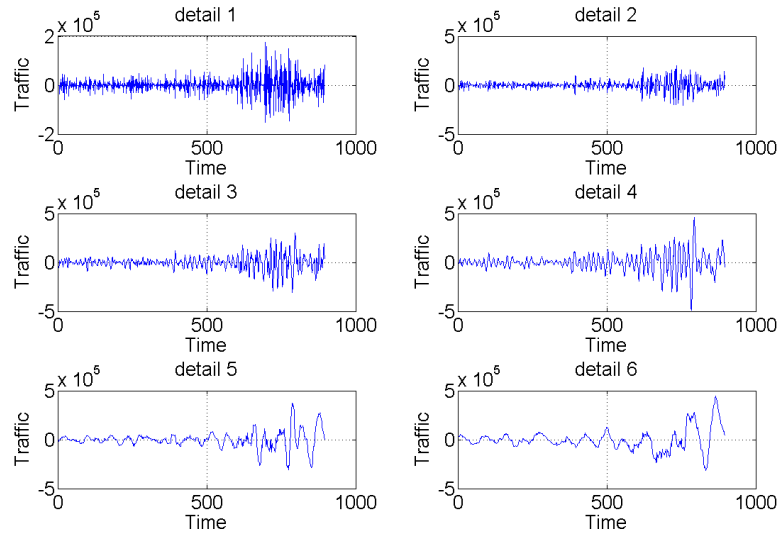

Figure 3.9: The detail coefficients.

Another data preparation operation contained in the forecasting algorithm proposed in [54] refers to detail coefficients illustrated in Figure 3.9. These sequences reflect the variability of the traffic and have different energies. In the following the detail sequences corresponding to time resolutions between 1.5 hours and 96 hours will be denoted by $d_1 - d_6$.

The equation describing the proposed multi-timescale analysis is:

$$ca_{sd}(t) = c_6(t) + \sum_{p=1}^{6} d_p(t). \tag{3.1}$$

Computing the energies of the detail sequences corresponding to our example the higher energy corresponds to d_3 (the time resolution of 12 hours). The next detail energy value in decreasing order corresponds to a time resolution of 24 hours (the detail d_4), where the higher periodicities of the time series were observed. The energy of the coefficients c_6, d_3 and d_4 represents a great quantity of the overall energy of the analyzed time series. The total energy contained in $ca_{sd}(t)$ is defined as:

$$E = \int_{-\infty}^{\infty} |ca_{sd}(t)|^2 dt = ||ca_{sd}(t)||^2. \tag{3.2}$$

Hence, we have decided to ignore the detail sequences with small energies to reduce the amount of computation and to keep in our multi-timescale analysis only the details d_3 and d_4, which explain the deviation of the time series around its overall trend :

$$ca_{sd}(t) = c_6(t) + \beta d_3(t) + \gamma d_4(t). \tag{3.3}$$

The model in (3.3) represents the new statistical model for the traffic time-series which we want to forecast. It reduces the multiple linear regression model in (3.1) at two components only: the overall trend of the traffic (described by c_6) and the variability (described by the detail coefficients d_3 and d_4).

In order to use the new statistical model, the weights β and γ must be identified (see Figure 3.10 and 3.11). First, for the identification of the weight β, the contribution of the coefficients d_4 is neglected. So, the new statistical model will be expressed by:

$$ca_{sd}(t) = c_6(t) + \beta d_3(t) + e(t), \tag{3.4}$$

where $e(t)$ represents the error of the new statistical model.

The parameter β can be found by minimizing the mean square of *e*(*t*), (see Figure 3.10),

$$\beta_{opt} = \underset{\beta}{argmin} \|ca_{sd}(t) - c_6(t) - \beta d_3(t)\|^2. \tag{3.5}$$

The already mentioned search procedure can be also used for the computation of γ. This time, the contribution of the coefficients d_4 is taken into account. The new statistical model will be expressed by:

$$ca_{sd}(t) = c_6(t) + \beta_{opt} d_3(t) + \gamma d_4(t) + e(t). \tag{3.6}$$

The parameter γ can be found by minimizing the new mean square of *e*(*t*), (see Figure 3.11),

$$\gamma_{opt} = \underset{\gamma}{argmin} \|ca_{sd}(t) - c_6(t) - \beta_{opt} d_3(t)\|^2. \tag{3.7}$$

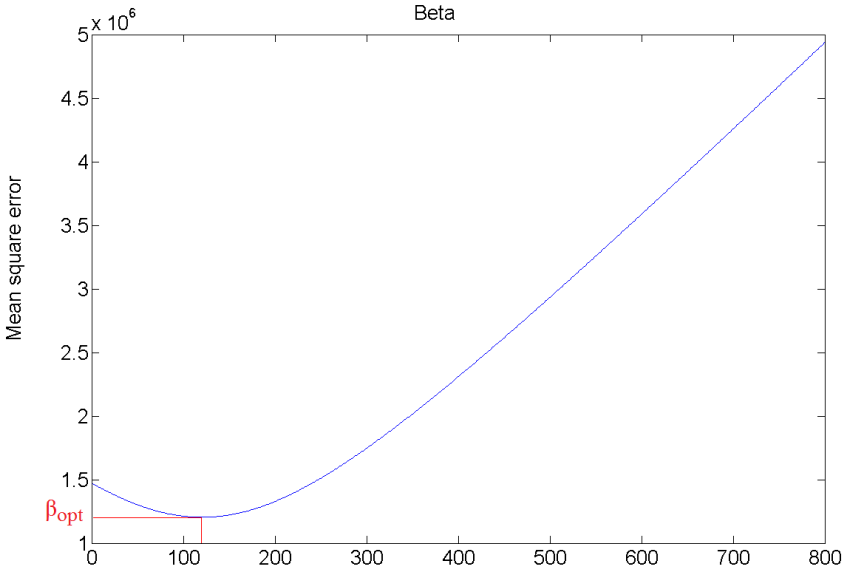

Figure 3.10: The search of the best value of β.

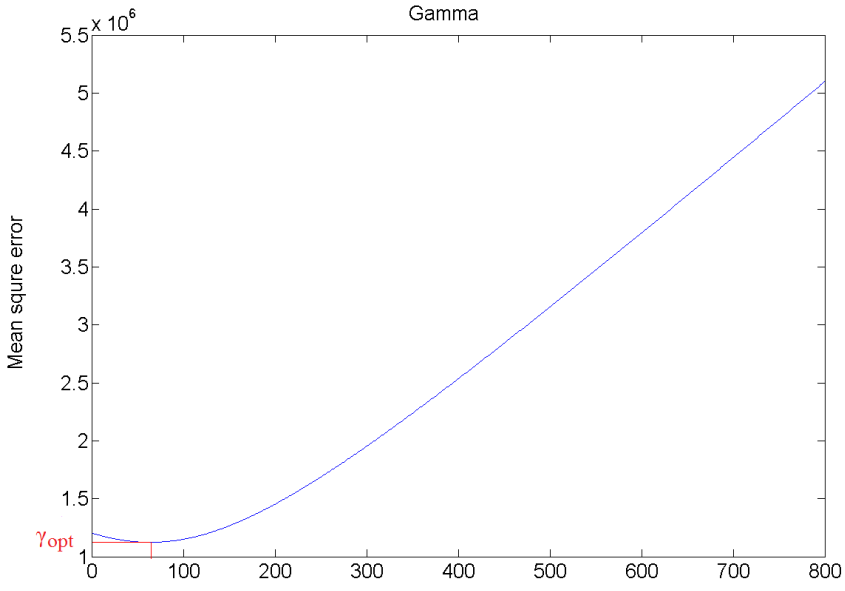

Figure 3.11: The search of the best value of and γ.

In Figure 3.12 is presented the reconstruction of the original traffic (first plot) using the estimation of the overall trend (realized using c_6) and the estimation of the variability (realized using βd_3-second plot and $\beta d_3 + \gamma d_4$-third plot). The approximation errors are higher in the second plot than in the third plot. This remark justifies the utilization of both weights β and γ. The approximation in the second plot is smoother than the approximation in the third plot. So, the utilization of the weight γ diminishes the high frequency components of the errors.

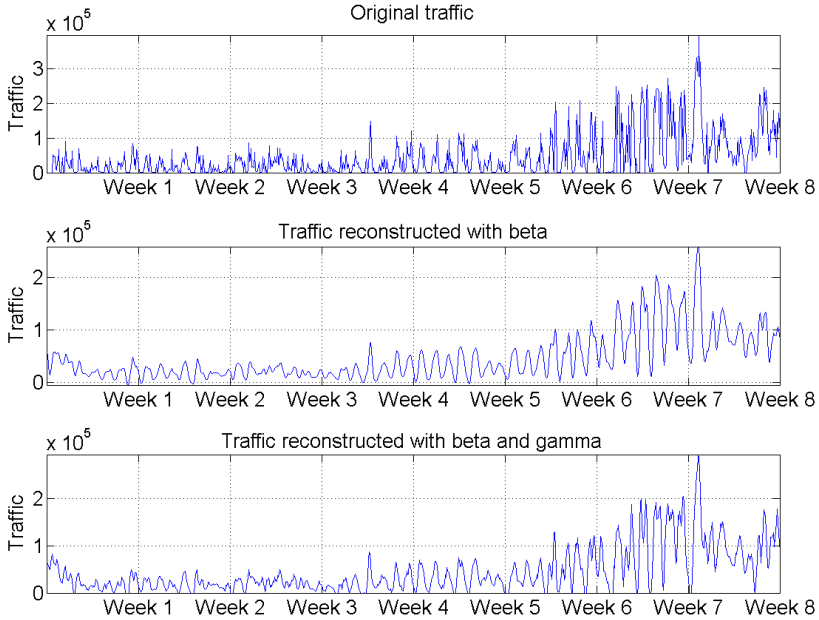

Figure 3.12: The reconstruction of the original traffic using the estimation of the overall trend and the estimation of the variability.

For capacity planning purposes, one only needs to know the traffic baseline in the future along with possible fluctuations of the traffic around this particular baseline. Since our goal is not to forecast the exact amount of traffic on a particular day in the future, we calculate the weekly standard deviation as the average of the seven values computed within each week. Such a metric represents the fluctuations of the traffic around the long term trend from day to day within each particular week.

Given that the 6^{th} approximation signal is a very smooth approximation of the original signal, we calculate its average across each week, and create a new time series capturing the long term trend from one week to the next. The resulting signal is presented in Figure 3.13, in red. It can be observed that this signal represents a good approximation of the overall tendency of the traffic.

Approximating the original signal using weekly average values for the overall long term trend, and the daily standard deviation results in a model which accurately captures the desired behavior. So, our data are prepared now for the modeling of the overall tendency of the traffic and of the variability around this tendency.

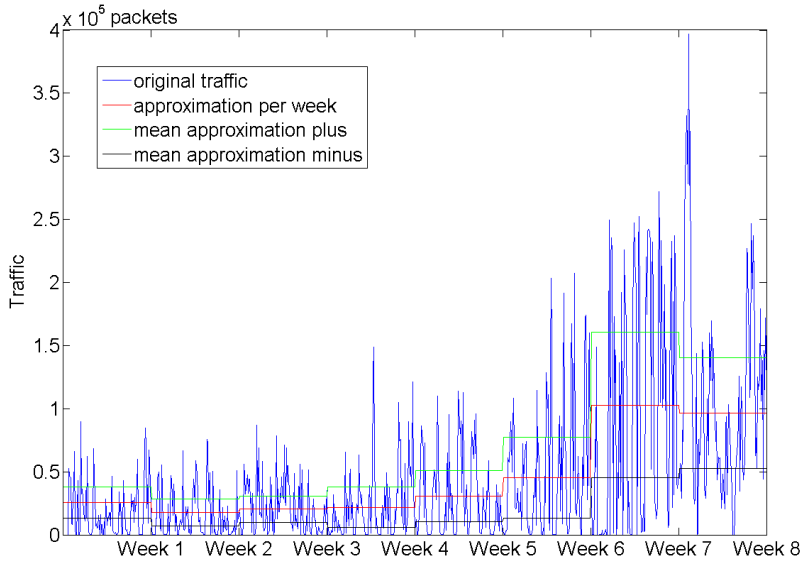

Figure 3.13: Approximation of the signal using the average weekly long term trend and the average daily standard deviation within a week.

3.2.4 Modeling

Modeling phase involves the selection of modeling technique and the estimation of model's parameters. Our goal is to model the tendency and the variability of the traffic using linear time series models.

Let us denote the terms describing the variance with:

$$dt_3(t) = \beta_{opt} d_3(t) + \gamma_{opt} d_4(t). \tag{3.8}$$

We used the Box-Jenkins methodology [8] to fit linear time series models, separately for the overall trend and for the variability, starting with the estimations in Figure 3.12. The estimations "mean approximation plus" and "mean approximation minus" are used for modeling the variability while the estimation "approximation per week" is used for modeling the overall trend. Such a procedure involves the following steps: determine the number of differencing operations needed to render the time series stationary, determine the values of *p* and *q*, estimate the polynomials ϕ, and θ.

In Figure 3.14 is presented the algorithm that lies at the basis of our Matlab® implementation of the Box-Jenkins methodology.

The goal of the Box-Jenkins methodology is to find an appropriate model so that the residuals are as small as possible and exhibit no pattern, [8]. The residuals represent all the influences on the time series which are not explained by other of its components (trend, seasonal component, trade cycle).

The steps involved to build the model are repeated, in order to find a specific multiple times formula that copies the patterns in the series as closely as possible

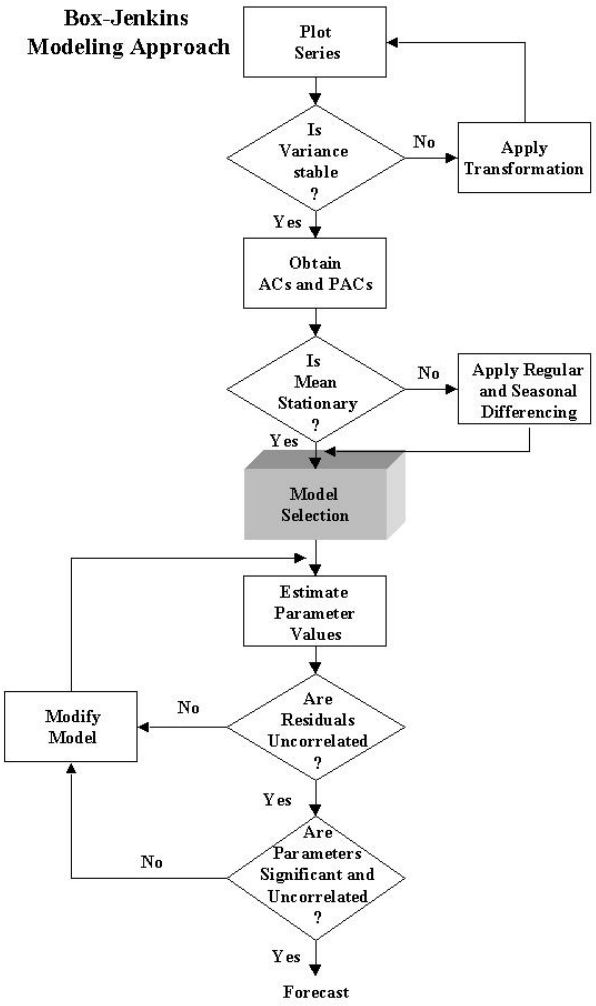

Figure 3.14: The Box-Jenkins methodology algorithm.

and produces accurate forecasts. The input data must be adjusted first to form a stationary series and next, a basic model can be identified [8].

ACF and PAC are used to analyze the stationarity of a time series and to estimate the orders p and q. The ACF and PAC plots are compared to the theoretical behavior of these plots when the order is known. For AR(p) processes the sample ACF should have an exponentially decreasing appearance for AR(1), while higher-order AR processes needs to be supplemented with a PAC plot because they are often a mixture of exponentially decreasing and damped sinusoidal components. The PAC of an AR(p) process becomes zero at lag $p + 1$ and greater. In the case of MA(q) processes the ACF becomes zero at lag $q + 1$ and greater, while the sample PAC function is generally not helpful for identifying the order of the MA(q) process.

In the following we will give an example to show how is applied the Box-Jenkins methodology for WiMAX traffic prediction. In Figure 3.15 are presented the approximation coefficients (the signal c_6) and their first and second differences.

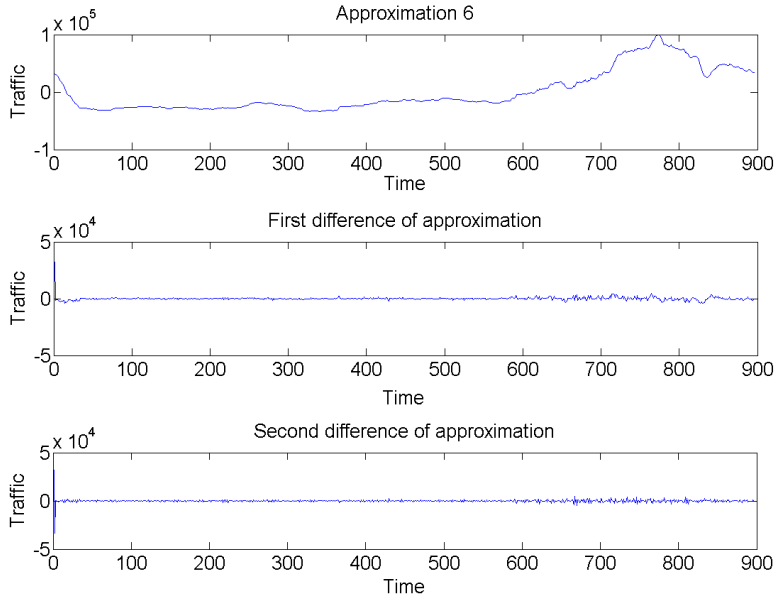

Figure 3.15: The approximation coefficients (first line) and their first (second line) and second (third line) differences.

The first step is to study which of these sequences are stationary in order to establish the value of the parameter d (the number of differencing operations required to obtain a stationary time-series) of the ARIMA model. There are few tests for different type of stationarity: stationarity in mean, stationarity in variance, or wide sense (second order) stationarity. The first test is to verify if the variance is constant, so we must compute and compare the partial variances (defined on two disjoint intervals) of each of the three time series (approximation, its first and its second difference). When both partial variances corresponding to the same time-series has the same value, we can decide that the considered series is stationary in variance. The second test is to verify if the time-series is stationary in mean. To test this type of stationarity we have to compute and to compare the partial means (defined on two disjoint intervals) of each of the three time-series. When both partial means corresponding to the same time-series has the same value we can decide that the considered series is stationary in mean. The third category of tests is dedicated to the wide sense stationarity. The first form of this category of tests uses the ACFs of the three sequences. These correlations are represented in Figure 3.16. The correlation of a stationary sequence must vanish after few samples. We can observe, analyzing Figure 3.16, that the third sequence (from up to bottom) has the higher decreasing speed. It has a peak in its middle. The sample values decreases rapidly at the left and the right of this pick becoming close to zero. This decreasing is most rapidly than the corresponding decreasing observed in the middle plot from Figure 3.16 or in the up plot.

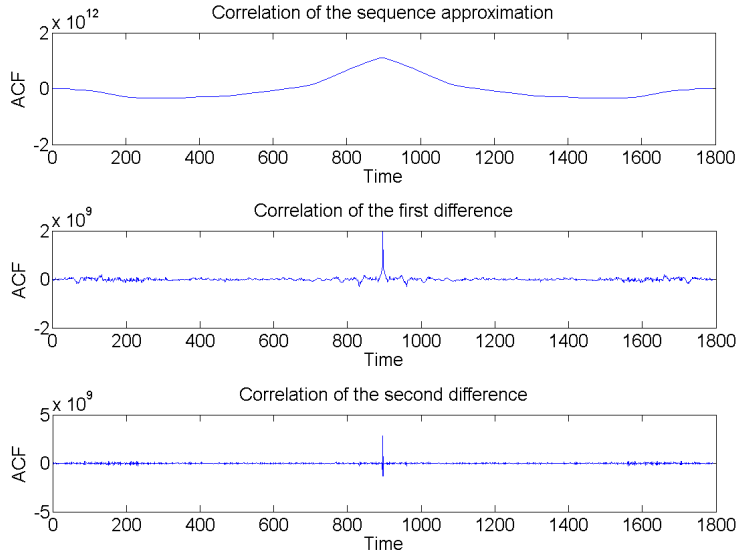

Figure 3.16: The autocorrelations of the three sequences approximaion (first line), their first (second line) and second differences (third line).

The second form of this category of tests uses the PACs of the three sequences. These functions are represented in Figure 3.17.

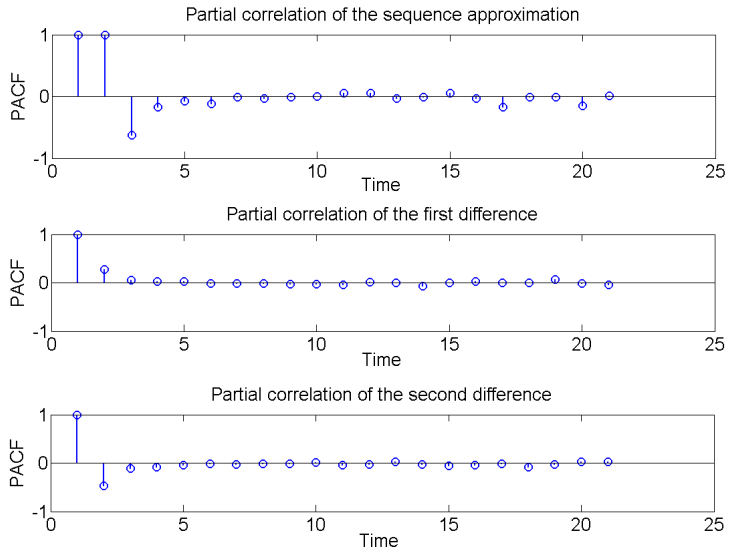

Figure 3.17: The partial correlations of the three sequences approximation (first line), their first (second line) and second differences (third line).

They are also useful for the estimation of orders p and q.

Analyzing Figure 3.17, we obtain the same conclusion as in the case of Figure 3.16, namely that the sequence obtained by computing the second difference of the sequence of approximations is more stationary than the sequence obtained by computing the first difference of approximations or that the sequence of approximations itself.

Finally, we propose a last functional test to appreciate which of the three time series is stationary. The idea of this test is to verify the repeatability of the Box-Jenkins procedure. We will present in the following the implementation of this test in parallel with the presentation of the implementation in Matlab® of the Box-Jenkins methodology. In fact, in [8] is devised to apply the Box-Jenkins methodology two times.

First is established an initial model that is optimized in the second run. To initialize the Matlab® Box-Jenkins methodology (the function $bj.m$) some information is required: the data to be modeled (one of the three sequences: the approximation c_6, its first difference or its second difference in our case) and the initialization of the model (the values p and q and the initial coefficients ϕ of order p and θ of order q). The results of the function $bj.m$ are: the optimal values of coefficients θ and ϕ (which permit the mathematical description of the model), the degree in which the model fits the data (it must be as small as possible) and the value of FPE which must be as small as possible. The orders p and q of the polynomials ϕ and θ can be identified based on of their coefficients but the value of the parameter d from the ARIMA model can not be identified using the function $bj.m$. For this reason it is identified on the basis of stationarity tests. We have applied five times the same Box-Jenkins methodology to each of the three sequences using the same initialization model and we have appreciated the repeatability of this procedure, comparing the individual results obtained. The repeatability of the Box-Jenkins methodology is very poor for the sequence of approximation c_6, so it can be concluded that this time-series is not stationary and that the stationarity of its first derivative must be tested.

Applying the same test to the first difference of the approximation, c_6, we have obtained that the repeatability of the Box-Jenkins methodology is very good for this sequence, so it can be concluded that this time-series is stationary. So, for the time-series considered in this example, $d = 1$, as it was indicated by the other stationarity tests previously presented as well. We have obtained the results presented in Table 3.1.

Run index	Initial model fit	Final model fit
1	20,85%	1,14%
2	0,5%	7,8%
3	8,4%	7,9%
4	2,6%	3,8%
5	0%	8,3%

Table 3.1: Results obtained running five times the Box-Jenkins methodology for the first difference of the approximation c_6.

The results corresponding to the first line of the Table 3.1 are represented in Figure 3.18 and those corresponding to the last line are represented in Figure 3.19.

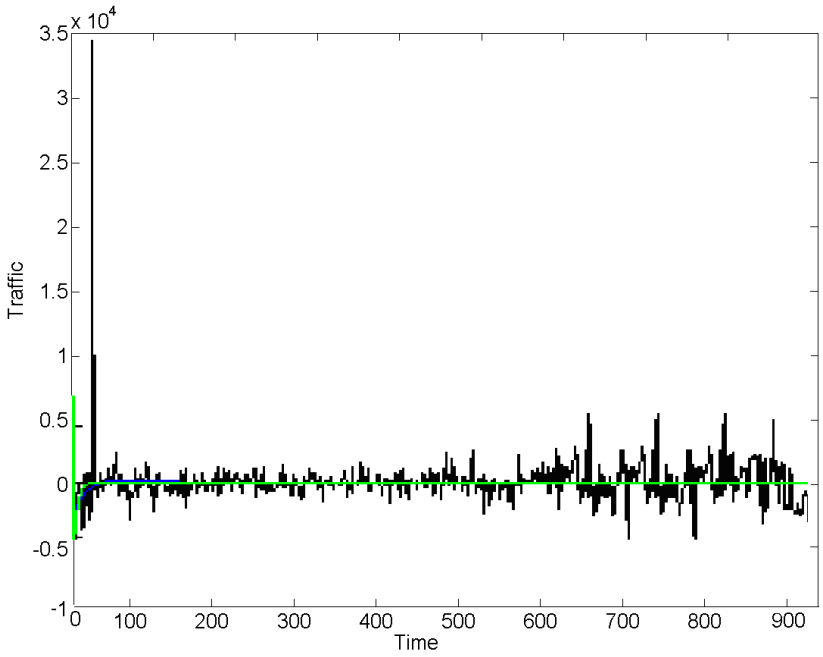

Figure 3.18: Results obtained applying first time the Box-Jenkins methodology on the first difference of the approximation c_6.

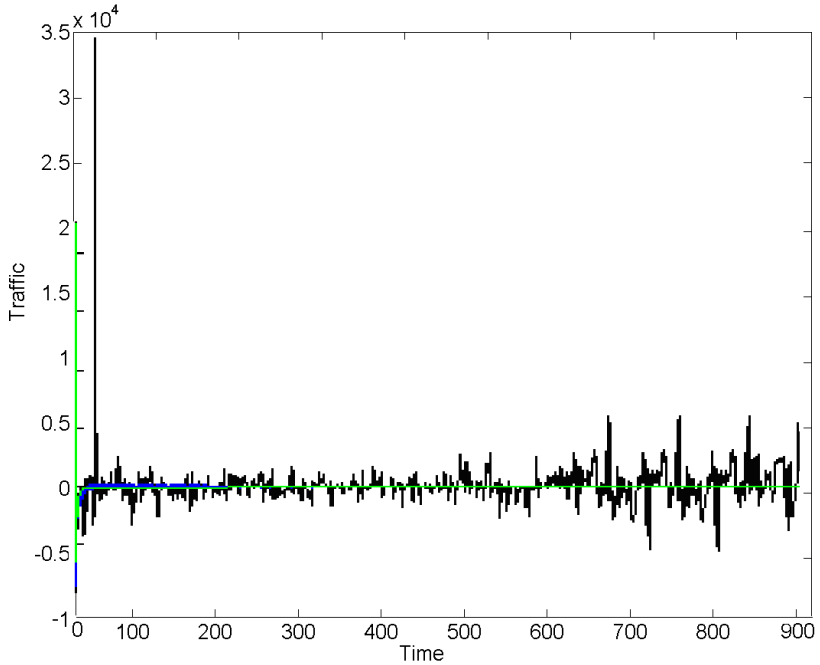

Figure 3.19: Results obtained applying fifth time the Box-Jenkins methodology on the first difference of the approximation c_6.

In both Figure 3.18 and Figure 3.19 the original signal is represented in black, the output of the initial model is represented in green and the output of the final model is represented in blue. Applying the same test to the second difference of the approximation c_6 we have obtained results (which are not presented here for seek of concision) proving that the repeatability of those results is inferior to the repeatability of the results obtained for the previous time series. So, in the following we will use the sequence formed by the first difference of the approximation c_6.

The initial model can be selected using Matlab® function *idpoly*, and specifying the polynomials B, C, D and F from the following equation:

$$y(t) = [B(q)/F(q)]u(t) + [C(q)/D(q)]e(t) \tag{3.9}$$

Applying once again the Box-Jenkins methodology, the final model (represented in Figure 3.20 for the current example) is obtained. In Figure 3.20 is presented a comparison of the first difference of the approximation c_6 with the simulation in Matlab® of its ARIMA model given by the Box-Jenkins methodology.

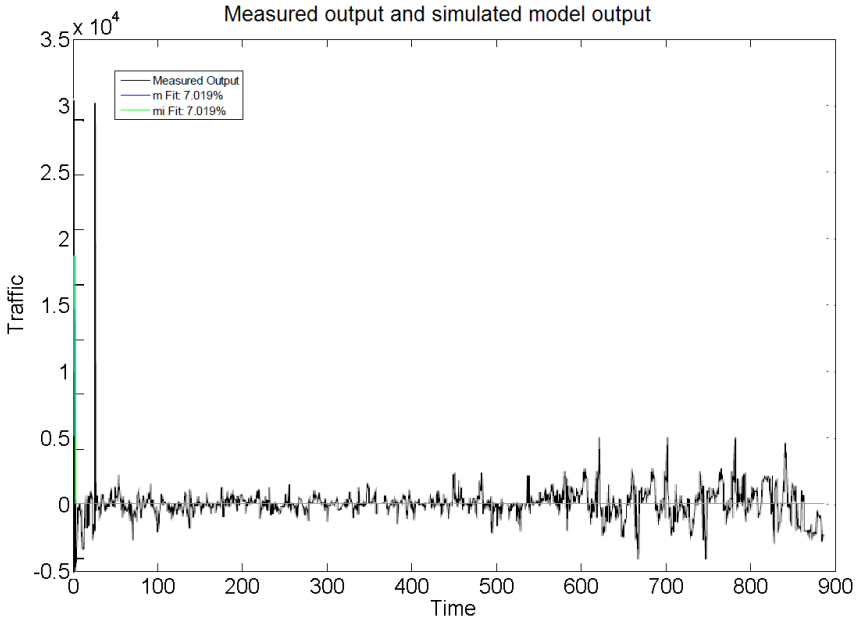

Figure 3.20: First difference of traffic overall tendency before and after ARIMA modeling.

A comparison of the approximation c_6 with the simulation in Matlab® of its ARIMA model is presented in Figure 3.21.

In this figure we represented the original time series with red, while the simulated model is represented in blue. A good match of the original time series with the model can be observed analyzing this figure.

The models computed for the long term trend of all downlink traces from the database are used for the next phase of our data mining project which consists in

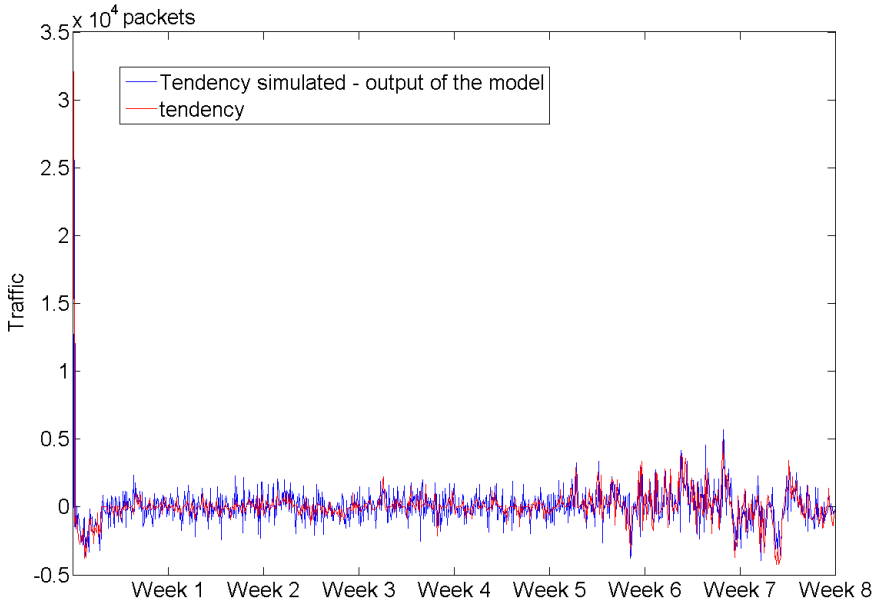

Figure 3.21: Traffic overall tendency before and after ARIMA modeling.

evaluation. These models indicate that the first difference of those time-series is consistent with a simple MA model ($p = 0$) with one or two terms (q=1 and d=1 or q=2 and d=1) plus a constant value μ_{ot}. This conclusion ends the modeling phase for the feature long-range dependence.

A similar modeling phase is implemented for the other feature of the WiMAX downlink traffic, the variability. As the approximation coefficients c_6 are used for the modeling of traffic long-range dependence, the detail coefficients dt_3 are used to appreciate the variability of the traffic. They are treated following a similar procedure based on the Box-Jenkins methodology. A comparison of the original time series (represented in red) with the data obtained simulating the model obtained applying two times the Box-Jenkins methodology (represented in blue) in the case of the traffic's variability is shown in Figure 3.22. A good match of the original time series with the model is observed in this case as well.

Applying the Box-Jenkins methodology on the deviation measurements (that reflect the variability of the traffic), we found that the deviation dt_3 can be expressed with simple MA ($p = 0$) processes after one differencing operation. This conclusion ends the modeling part of our data mining project.

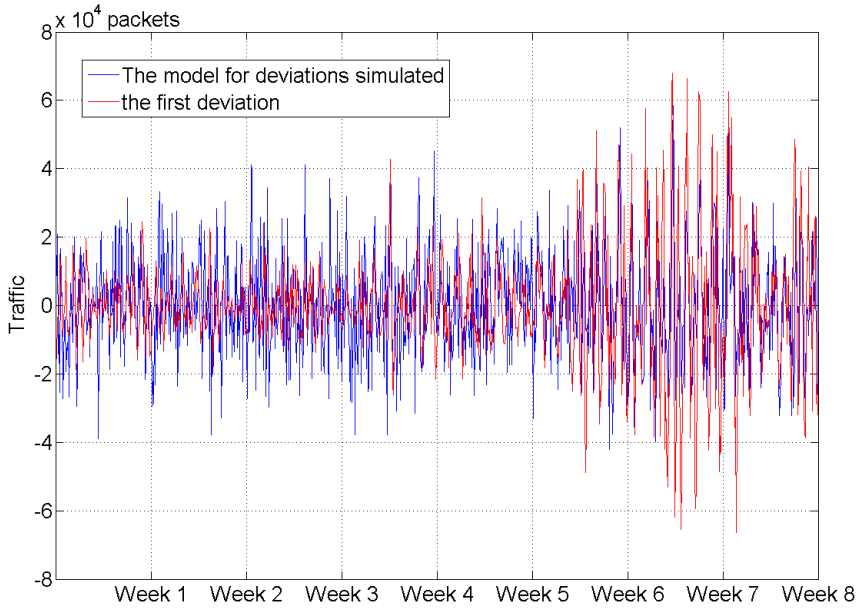

Figure 3.22: Modeling the variability of the traffic.

3.2.5 Evaluation

In this phase, both models (for overall tendency and variability of the traffic) are evaluated and all the precedent steps are reviewed. In order to see if the new statistical model in (3.3) is representative, we used ANOVA and we computed the coefficient of determination defined in Chapter 2. The model is considered to be statistically significant if it can account for a large fraction of the variability in the response, i.e. yields large values for the coefficient of determination.

We have applied the forecasting algorithm to all the downlink traces, from the data base and we have obtained statistically significant ARIMA models for each traffic overall tendency and variability. We have identified the model parameters (p and q) using MLE. The best model was chosen as the one that provides the smallest AICC, BIC and FPE measures, while offering the smallest mean square prediction error for a number of weeks ahead.

3.2.6 Deployment

The final stage, deployment, involves the application of the model to new data in order to generate predictions. The moment when the saturation of the BS takes place can be predicted comparing the trajectory of the overall traffic forecast with the BS's saturation threshold as shown in Figure 3.23.

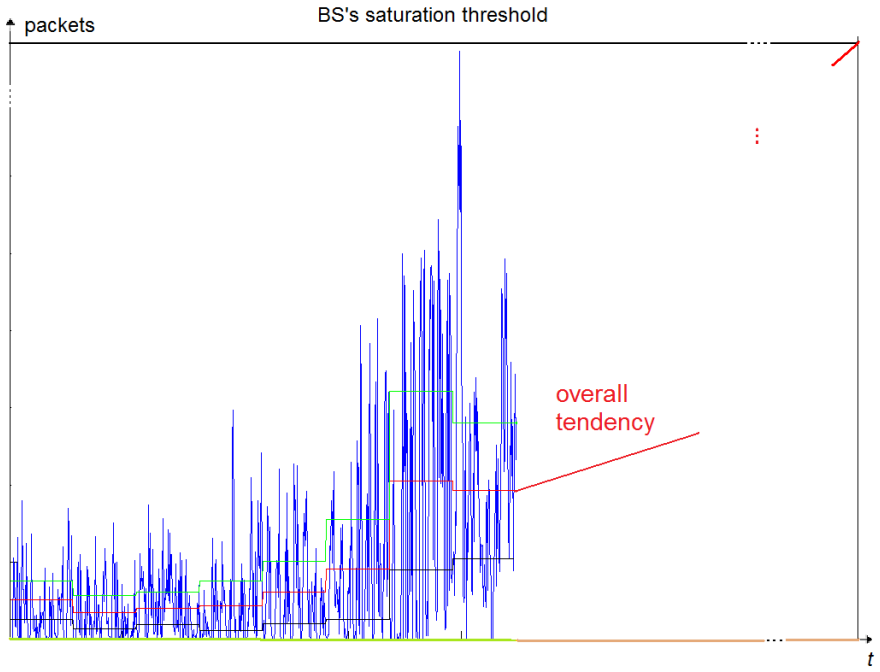

Figure 3.23: The trajectory for the long-term forecasts.

The need for one differencing operation at lag 1, and the existence of term μ_{ot} across the model indicate that the long-term trend of the downlink WiMAX traffic is a simple exponential smoothing with growth. The trajectory for the long-term forecasts will be a sloping line, whose slope is equal to μ_{ot}. Similar conclusion can be formulated for the variability of the downlink WiMAX traffic. The single difference between the long-term trend and the variability is that the slope of the variability is much smaller than the slope of the long-term trend. The trajectory for the variability forecast is a sloping line as well but it has a much smaller slope. The sum of these sloping lines is a third line, parallel with the trajectory of the long term forecast, which represents the trajectory of the overall forecast. Hence, the risk of saturation of a BS is direct proportional with the slope of its overall tendency. Given the estimates of μ_{ot} across all models, corresponding to all BSs, we can conclude based on the positive values of those slopes that all traces exhibit upward trends, but grow at different rates.

In Table 3.2 is presented a classification of BSs in terms of the saturation risk. For BS50 the value of μ_{ot} was estimated as equal to 2.0610^{16} which is much higher than all the other slopes, so we decided to exclude it from the classification. BSs showed on the first column in Table 3.2 have high values of μ_{ot}. This means that these BSs have a higher risk of saturation than the other BSs. The BSs with higher risk of saturation are the following: BS63, BS60, BS3, BS49, BS61.

We cannot come up with a single WiMAX network-wide forecasting model for the aggregate demand. Different parts of the network grow at different rates (long-term trend), and experience different types of variation (deviation from the long-term trend).

BS	μ_{ot} (Mb/s)	BS	μ_{ot} (Mb/s)	BS	μ_{ot} (Mb/s)	BS	μ_{ot} (Mb/s)
63	239.860	48	114.810	13	68.311	1	45.068
60	185.470	52	110.250	53	66.329	2	44.729
3	177.680	8	109.040	6	65.579	9	43.102
49	176.070	7	105.240	5	63.415	42	42.878
61	164.030	56	104.720	26	59.885	33	41.441
57	157.260	55	99.920	12	58.708	30	41.395
62	146.310	65	99.174	39	57.789	28	39.973
67	144.630	20	97.943	38	57.675	41	38.129
54	143.880	29	97.655	35	54.498	40	33.587
18	138.230	46	93.711	37	53.458	36	32.224
64	134.220	10	91.557	23	52.729	25	30.601
16	131.730	19	83.567	45	51.019	15	29.400
59	130.530	43	79.215	22	50.872	11	27.622
58	130.350	44	78.572	24	49.404	31	26.144
51	123.960	66	74.149	27	46.704	17	25.052
4	118.100	14	71.564	47	45.879	21	24.614
32	15.921						

Table 3.2: BSs risk of saturation.

Our methodology extracts those trends from historical measurements and can identify those BSs in the network that exhibit higher growth rates and thus may require additional capacity in the future.

Our technique is capable, based on MRA and ARIMA modeling, of isolating the overall long term trend and identifying those components that significantly contribute to its variability. Predictions based on approximations of those components provide accurate estimates with a minimal computational overhead.

3.3 Selection of Mother Wavelets

Another important goal in our work is to compare the influence of different wavelets families on the prediction accuracy. We propose a comparison between the following wavelets families:

- Daubechies: db_1, db_2, db_3, db_4, db_5.
- Coiflets: $coif_1$, $coif_2$, $coif_3$, $coif_4$, $coif_5$.
- Symlets: sym_2, sym_3, sym_4.
- Biorthogonal: $bior_{1.1}$, $bior_{2.2}$, $bior_{3.1}$, $bior_{4.4}$, $bior_{5.5}$.
- Reverse Biorthogonal: $rbio_{1.1}$, $rbio_{2.2}$, $rbio_{3.3}$, $rbio_{4.4}$, $rbio_{5.5}$.

These families of mother wavelets were already introduced in Chapter 1 and their particularities as: length of support, number of vanishing moments, time, frequency and time-frequency localizations. There are some features of the forecasting algorithm which can be optimized by the mother wavelets selection such as: the results of the MRA, the values of the weights β and γ given by the ANOVA procedure or the accuracy of the traffic prediction.

We have applied the forecasting algorithm choosing each mother wavelets already mentioned and for each trace corresponding to each BS from the database. In Figure 3.24 are shown the main steps followed in our work.

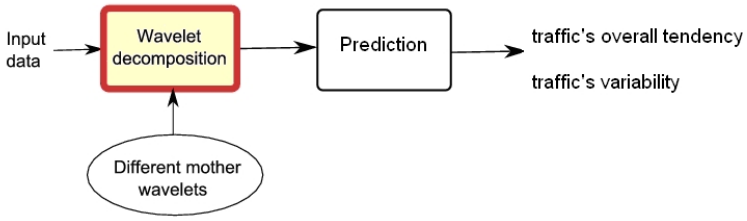

Figure 3.24: Main steps followed in our algorithm.

We have divided each data sequence into two parts, each corresponding to a specific interval of time. Data from the first interval (of seven weeks) were considered as historical and were used for prediction, while data from the second interval (the last week) were used to evaluate the quality of prediction. The results obtained are presented in Table 3.3.

Wavelet	SMAPE	MAPE	MAE
coif 1	0.890	0.0020	0.9599
coif 2	0.837	0.0019	**0.7191**
db 1	**0.812**	**0.0016**	0.7327
db 2	0.855	0.0019	0.7768
db 3	0.857	0.0018	0.7678
db 4	0.834	0.0018	0.7563
db 5	0.823	0.0019	0.7730
bior 3.1	0.860	0.0018	**0.7071**
rbio 1.1	**0.820**	**0.0017**	0.8947
rbio 2.2	0.891	0.0018	0.7690
rbio 3.3	0.907	0.0022	1.0690
sym 2	0.895	0.0019	0.7412

Table 3.3: Comparison between wavelets, WiMAX traffic.

These results consist in values of the three quality measurements, SMAPE, MAPE and MAE. These measurements were computed for all BSs, using each of MWs mentioned. In Table 3.3 are shown the mean values obtained for all the BSs.

SMAPE, MAPE and MAE, are calculated between the actual and the predicted tendency, because for linear models the trajectory of the forecasts is represented through sloping line which represents the weekly increase.

According to the results presented in Table 3.3., the 1^{st} order Daubechies wavelet, db_1, which is the simplest of the Daubechies family, gives the best prediction performance. So, in the case of communication traffic time series, the time localization is more important than the frequency localization. Good forecasting accuracy was obtained using mother wavelets with good time-frequency localization as well, which have a reduced number of vanishing moments, like $rbio_{1.1}$.

3.4 Extension to Financial Domain

The prediction method already described can be applied not only in communications domain but also in finance.

The financial time-series analyzed in this section is composed by numerical values representing the total number of EUR-USD currency exchanges realized in a time interval of fifteen weeks. The values are recorded every 15 minutes. The overall tendency of this time-series was estimated on the basis of a MRA, followed by an ARIMA modeling of the sequence of approximation coefficients. We have divided the data sequence into two parts: training part of fourteen weeks and a testing part of one week. We followed the same steps shown in Figure 3.24 as in the case of WiMAX traffic. The results we obtained are presented in Table 3.4.

Wavelet	SMAPE	MAPE	MAE
coif 1	0.516	0.0821	0.4240
coif 2	**0.453**	0.0732	0.3799
db 1	0.454	**0.0713**	**0.3690**
db 2	0.497	0.0812	0.4199
db 3	0.531	0.0864	0.4461
db 4	0.527	0.0863	0.4459
db 5	0.546	0.0912	0.4710
bior 3.1	0.433	**0.0712**	**0.3681**
rbio 1.1	0.457	0.0716	0.3704
rbio 2.2	0.491	0.0790	0.4081
rbio 3.3	0.455	0.0716	0.3705
sym 2	**0.453**	0.0736	0.3813

Table 3.4: Comparison between wavelets on financial data.

The best forecasting performance is obtained using the mother wavelets $db_1(Haar)$ and $bior_{3.1}$, which have good time localization.

The financial data (the EUR-USD currency exchanges) exhibit an almost constant tendency, while WiMAX traffic presents a strong variability and its tendency (long term trend) represents a sloping line. However, our algorithm is applicable to both types of data and the obtained predictions are accurate.

3.5 Conclusions

In this chapter, we have proved by extensive simulations that the traffic forecasting methodology proposed in [54] can be adapted in the case of WiMAX traffic. We have adapted the methodology proposed in [54] to our case, by taking into consideration the particularities of wireless networks as their ad-hoc nature or the non-stationary behavior of their traffic. To do this adaptation we have modified the number of detail sequences retained after the ANOVA analysis, performed using the MRA and required for the estimation of the variability of the traffic, from one (number used in [54]) to two. We have proposed a methodology to select the two weights required, β and γ and we have added to the methodology proposed in [54] a new test of stationarity (original in our knowledge) based on the reiteration of the Box-Jenkins methodology.

Our technique is capable of isolating the overall long term trend and identifying those components that significantly contribute to its variability.

The statistical models for the overall tendency and for the variability of the traffic can be found neglecting some resolutions from the corresponding MRA. Selecting only approximation coefficients at the sixth decomposition level of the SWT, c_6, we can predict the overall tendency of WiMAX traffic. Hence, the overall tendency of the traffic is a very low frequency signal, requiring mother wavelets with good frequency localization (which have a big number of vanishing moments). The traffic variability can be predicted by selecting only detail coefficients at the third and fourth decomposition levels, d_3 and d_4. Hence, the traffic's variability is a relative high frequency signal, requiring mother wavelets with good time localization (as db_1 or $rbio_1.1$, which have a reduced number of vanishing moments).

In consequence, the proposed forecasting methodology requires mother wavelets with both time and frequency good localization. So, we consider that the best results will be obtained using mother wavelets with a good time-frequency localization, which corresponds to a reduced number of vanishing moments.

Considering that the prediction accuracy of the traffic variability is more important than the prediction accuracy of the traffic overall tendency, it results that time localization is more important than the frequency localization. So, the best mother wavelet seems to be the Haar wavelet (db_1) in our case. With the increase of the number of vanishing moments, the performance of the traffic's prediction deteriorates. In [65], [56] and [57] we proposed an approach for predicting traffic time series based on the association of the SWT with Artificial Neural Networks (ANN). According to the results, we can conclude that ANN performs better then the other forecasting techniques for small future time intervals, several weeks at most. The superiority of short-term forecasting methods based on ANNs can be explained by the higher computational complexity of these methods. They require a supplementary training phase and are applied to all wavelet coefficients ($d_1 - d_6$), without computing weekly averages. But, if we are interested in the tendency of the traffic, meaning prediction for several months, than pure statistical models are the ones that should be taken into consideration.

The non-stationarity of the traces from the database was highlighted. It is explained by the non-constant overall tendency extracted from each trace. This behavior makes more difficult the estimation of the Hurst parameter presented in the following chapter.

Chapter 4

Knowledge Discovery in WiMAX Traffic

Communication equipment generates and stores large amounts of data. In the last years the topics of self-similarity and LRD in communication networks has become a very popular research domain [17], [67], [3]. The analysis of real data challenged the engineers as well as the researchers, so LRD has become more and more used in data analysis [52], [41]. Recent analysis of traffic measurements from various communication networks has revealed that the traffic is long-range dependent or fractal (self-similar). These findings revolutionized the understanding of network traffic, giving an explanation of the difference which appears between the theoretical estimated efficiency and the efficiency measured in practice, so time-series analysis and modeling in terms of LRD have become more and more used in data analysis.

In this chapter, we propose to analyze the uplink and downlink traffic in a WiMAX network in terms of LRD. The aim is to obtain the estimated values of *H* parameter, for each uplink and downlink trace corresponding to all the BSs that compose the considered network. We prove that some particularities of the network can be established analyzing these values. Rules for the optimization of the network's exploitation can be derived analyzing these particularities.

4.1 Related Work

Previous work, [40], [39], [67], [3], [52], [30], [41], proved the utility of *H* for the analysis of the Internet traffic. In [40] is detected self-similarity in world wide web (www) traffic and are presented some possible causes for this comportment. In [39] is observed that the self similarity of the traffic influences the performance of the corresponding network. Some network performance evaluation methods which take into account the self-similarity of the traffic are proposed. In [30] is observed that the prediction of wireless network traffic is influenced by its LRD. In 1993 was identified the presence of LRD in data sets captured on Ethernet Local Area Network (LAN) traffic, [14]. In the case of Ethernet LAN traffic, LRD is manifested in the absence of a natural length of a "burst"; at every time scale ranging from a few milliseconds to minutes and hours, bursts consist of bursty sub-periods separated by less bursty sub-periods. So, a cause of LRD is the hidden periodicities which are present in the time series analyzed. It is also shown that the value of the Hurst parameter typically depends on the utilization level of the Ethernet and can be used to measure "burstiness" of LAN traffic. In [43] it is shown that the H parameter is a function of the usage of Ethernet (higher usage meaning a higher value of H). The reason for the

considerable interest in this subject is the fact that the engineering implications of LRD on queuing performance can be considerable. If Internet traffic is not well modeled using independent or short-range dependent (SRD) models, then traditional queuing theory based on the assumption of Poisson processes is no longer appropriate. Traffic which is long-range dependent by nature can have a queuing performance which is significantly worse than Poisson traffic [3]. It has been found [3] that a higher H parameter often increases the delays in a network, the probability of packet loss and affects a number of measures of engineering importance. The majority of the papers already mentioned refer to networks on wires. The exception is the reference [30] which refers to wireless networks.

4.2 Sources of LRD

In the literature, [14], [45], [29], [68], [17] several possible origins for LRD in networks as: the hidden periodicities in traffic, its variable rate, the existence of heavy tail data streams, the feedback mechanisms in the TCP protocol or the bad positioning of BSs, are commonly cited.

One of them is the hidden periodicities which are present in the time series [14]. These hidden periodicities are revealed by analyzing the power spectral densities of the time series.

In [14] it is proved that the existence of video traffic coded with Variable-bit rate (VBR) induces the LRD. In this case, the LRD arises from the encoding mechanism whereby video is encoded as a series of differences between frames with occasional full updates.

Aggregate traffic is made up of many connections which arrive randomly. Each connection is characterized by its "size" representing the number of packets and by the "rate" of transmitted packets. As showed in [17], the distributions of connections have very long tails.

A random variable *X* is heavy tailed if for all $\alpha > 0$ it satisfies:

$$P(X > x)e^{\alpha x} \to \infty, \qquad x \to \infty. \tag{4.1}$$

In [45] the authors show that LRD results by the aggregation of heavy-tailed data streams.

Another potential cause of LRD is given by the feedback mechanisms in the Transmission Control Protocol (TCP). Let's consider the transmission of a packet between a sender-receiver pair on a network. The data is sent usually according to a reliable transport protocol like TCP. The release of packets to the network is decided by the flow and congestion control mechanism. Using a Markov model to simulate the behavior of TCP traffic, the authors of [29] concluded that the multiple timescale nature of traffic generation, coupled with transport protocol issues, make the appearance of LRD-like behavior inevitable, while in [68] it is said that "TCP congestion control creates self similar traffic (...) showing both short-range and long-range dependence depending on system parameters".

LRD arises from network topology or routing algorithms as well [14]. If a BS is bad positioned into a wireless network, its traffic is more difficult than the traffic of the other BSs, which are well positioned, and it can behave LRD.

Determining the origin of LRD remains an important research area and it is essential to find out which is the real origin of LRD. The presence of LRD can be controlled

only its origins are known. A possibility remains that it is a mixture of some or all of the sources already mentioned.

The goal of this chapter is to analyze the WiMAX traffic recorded in the data base, which was already used and described in previous chapter, in order to establish if the corresponding traces exhibit LRD or not. As it was already mentioned, one of the sources of LRD for a BS is its bad positioning. The goal of LRD analysis reported in this chapter is to identify the BSs which could be bad positioned in the architecture of the WiMAX network which corresponds to the data base. To do this, we intend to separate the different sources of LRD and to eliminate the effects of all LRD sources with the exception of the bad positioning. To attain this goal we have identified some software products which can be used to estimate the Hurst parameter for each uplink or downlink trace in the database, such as SELFIS which will be presented in the next sub-section or some Matlab® functions as $HEST$, which will be used in a subsequent sub-section.

SELFIS (SELF-similarity analysIS) is a java-based software tool for self-similarity and LRD analysis, developed by T. Karagiannis and M. Faloutsos, [41] at University of California. It implements the following estimators of H: Aggregate Variance, Periodogram, Variance of Residuals, Whittle Estimator, R/S, Absolute Moments and the Abry-Veitch Estimator which were already presented in Chapter 2.

Using the H parameters estimated using the R/S estimator from SELFIS, we will identify in the following some hidden periodicities in the WiMAX traffic and we will find as solution to reduce the effects of this first source of LRD, the segmentation of the time-series. As it was already said, the other sources of LRD in data communication traffic are: its variable rate, the existence of heavy tail data streams, the feedback mechanisms in the TCP protocol and the bad positioning of the BSs. These sources of LRD have different effects in downlink and in uplink. The principal difference between these two phases of wireless communications is given by the access to Internet. The majority of users make more frequently downloads that uploads on different Internet sites. Generally, the messages transmitted in uplink are shorter than the messages transmitted in downlink. The variability of the traffic rate is produced by mechanisms such as data streaming (required by multimedia applications) which are more specific for the downlink than for the uplink. Same mechanism, the data streaming, produces time-series with heavy tail distributions, so this kind of distributions appear more frequently in downlink than in uplink. The feedback mechanism in TCP protocol depends on the length of the message which is currently transmitted. This mechanism works heavy in the case of long messages. So, the feedback mechanism produces stronger LRD in downlink than in uplink. For these reasons, we can consider as principal sources of LRD in uplink traffic the hidden periodicities and the bad positioning of BSs. In consequence, after the reduction of LRD produced by hidden periodicities, we can consider that the single LRD source remained in the uplink traffic is the bad positioning of BSs. After the reduction of the effect of the hidden periodicities we will compare the LRD comportment in uplink and downlink of each BS, we will identify the normal behavior of a BS from the LRD point of view and we will isolate the BSs which deviate from this normal behavior. These BSs could be considered as bad positioned.

4.3 Evaluation of *H* Using R/S Method

By initial tests performed using the WiMAX data base (which are not presented here to keep a decent length of the thesis), we have observed that SELFIS makes an

acceptable estimation of H for time series which have a long enough length, when the R/S estimator is used. As we have stated in Chapter 2, the quality of the R/S estimation decrease with the decreasing of the analyzed sequence length, due to the increasing of the polarization of that estimator. For this reason, we will use the R/S estimator to prove that the traffic from our database contains some hidden periodicities. We have already observed a periodicity of 24 hours in the previous chapter. To make simpler the analysis of hidden periodicities, we have separated the uplink and the downlink traffic.

4.3.1 Downlink Traffic

This section presents the evaluation of H with the aid of the R/S method in the case of WiMAX downlink traffic. Its goal is to highlight the particularities of WiMAX traffic from a LRD perspective. As it was already said, one of the sources of LRD in time-series is represented by the hidden periodicities. To identify the possible periodicities, few LRD analysis, for different lengths of the time-series are helpful. First, the H parameter of the entire time-series must be estimated. Second, the time-series must be segmented and the H parameter of each segment must be estimated. If the H values of all segments are smaller than the H value of the entire time-series, it can be deduced that the time-series contains a hidden periodicity with a period belonging to a time interval having as inferior limit the length of the segments and as superior limit the length of the entire time-series. To identify other hidden periodicities, with shorter periods, the method already described can be repeated, using shorter segments.

The length of the entire time-series in our database is of eight weeks. As a first experiment we calculated the value of H using the R/S estimator for the sixty-six time series, corresponding to all BSs. The results are presented in Table 1 in Appendix. We can observe that the values of *H* are between 0.57 and 0.756, so H belongs to the theoretical interval that proves the presence of LRD ($H \in [0.5, 1]$). So, the downlink WiMAX traffic exhibits LRD, observation which can explain why the real performance of a real wireless network is inferior to its value estimated theoretically.

Next, we searched the hidden periodicities. We have split the time series into weeks. We have obtained eight new time series for each BSs and for each of these series we have calculated the new values of H. The results are presented in Table 2 in Appendix. The majority of the values of H are smaller than the H value of the entire series (composed by the eight weeks). Hence, the downlink WiMAX series contain hidden periodicities with values between a week and eight weeks. This source of LRD can be eliminated by performing the LRD analysis on segments having the length of one day.

In the following we will present a similar LRD analysis for the uplink traffic. Contrary to the method presented in Chapter 3, where was considered only the downlink traffic, we are interested now in the uplink traffic as well. The problem discussed in Chapter 3 was connected with the estimation of the moment when a given BS will saturate. The saturation appears only in the presence of a wide traffic, and we have observed that the traffic in downlink is more intense than the traffic in uplink. For this reason we have considered only the downlink traffic in Chapter 3. The problem proposed in this section is different. We try to separate the normal and the exceptional behaviors of the traffic. To identify these behaviors we must know the comportment of a BS in both downlink and uplink. For the moment we are interested if the uplink traffic contains hidden periodicities as well.

4.3.2 Uplink Traffic

A simplified LRD analysis can be done for the uplink traffic, when some sources of LRD are not present. Indeed video traffic does not exist in uplink and the feedback mechanisms in the TCP protocol do not manifest in uplink. So, the only two sources of LRD in uplink are the hidden periodicities and the bad topology of the network. For this analysis we use the same method for the estimation of H as in downlink, namely the R/S method.

First we use the R/S method to estimate the H parameter corresponding to the sixty-six time series (66 BSs). The results are presented in Table 3 in Appendix. As in downlink, the Hurst parameter belongs to the theoretical interval that proves the presence of LRD ($H \in [0.5, 1]$). The values of H are between 0.56 and 0.754.

Next, we have split the time series into weeks. We have obtained eight new time series, for each BS and for each of these series we have calculated the new values of H. The results are presented in Table 4 in Appendix. The conclusion is that the entire series (of eight weeks) exhibits stronger LRD than each weekly series. Hence, the uplink WiMAX traffic contains hidden periodicities with periods between one week and eight weeks as well. So, the normal comportment of the WiMAX traffic supposes the presence of hidden periodicities. It is plausible that one of those hidden periodicities to correspond to a period of four weeks, taking into account social reasons connected with the organization of the work in enterprises which supposes more deliveries at the end of the month and more production at the beginning of the month. Another hidden periodicity could correspond to a period of a week, taking into account the reduction of activity on the duration of weekends. The effect of those hidden periodicities can be reduced by the segmentation of the time series.

To continue the identification of hidden periodicities in WiMAX traffic we must make another segmentation of the time-series. The segments will have shorter lengths, of one day. To make the LRD analysis of those segments we need a better estimator for H than the R/S estimator. As it was specified in Chapter 2, based on the theoretical arguments, the generalized quadrature variations estimator based on wavelets is the best estimator for the Hurst parameter in the case of discrete observations, as the traces in our database are. To verify experimentally this assertion, we will compare in the following sub-section some SELFIS estimators with the generalized quadrature variations Hurst parameter estimator based on wavelets.

4.4 A Comparison of Some Estimators of the Hurst Parameter Based on Simulation

The following estimators of the Hurst parameter are considered in the next simulations: Aggregate variance, R/S, Periodogram, and Absolute moments, implemented in SELFIS, and generalized quadrature variations estimator (discrete Abry-Veitch), implemented in Matlab® - function HEST. All these estimators were already defined and analyzed in Chapter 2. We will use in the following simulations two types of random processes with known values of H and we will check the estimated values given by different estimation methods.

For the first simulations we considered a White Gaussian Noise (WGN) as input process. The corresponding value of H must be 0.

The simulation results are presented in Table 4.1.

No. of samples	Aggregate variance	R/S	Periodogram	Abs. moments	HEST
100 000	0.39	0.52	0.5	0.37	0.0028
10 000	0.51	0.53	0.47	0.49	-0.0035
4096	0.51	0.59	0.53	0.48	-0.034
1024	0.46	0.59	0.40	0.43	0.036

Table 4.1: WGN input process.

Analyzing the results in Table 4.1, it can be observed the increasing of the bias of R/S estimations with the decreasing of the length of the input sequences. The superiority of the wavelet based estimator (HEST) is obvious.

Next, we considered a fBm input process containing 10.000 samples, with the following values of H: 0.5, 0.6, 0.7, 0.8, and 0.9. The results are shown in Table 4.2.

Input process	Aggregate variance	R/S	Periodogram	Abs. moments	HEST
fBm05	1.015	0.463	1.497	0.865	0.483
fBm06	0.986	0.387	1.597	0.817	0.583
fBm07	0.991	0.272	1.639	0.795	0.685
fBm08	0.994	0.162	1.614	0.774	0.789
fBm09	0.996	0.06	1.551	0.751	0.892

Table 4.2: fBm input process.

The results presented in Table 4.2 prove that the precision of Hurst parameter estimators from SELFIS depends on its values. For small values of H, the R/S estimator seams to be the best. For intermediate values of H, the Absolute moments estimator is the best one. For high values of H, the best estimator used by SELFIS is the Aggregate variance. Once again, the results obtained by applying the generalized quadrature variations estimator based on wavelets are better than the results obtained by using estimators implemented in SELFIS.

The two types of simulations already analyzed recommend the use of the wavelet based generalized quadrature variations estimator for Hurst parameter estimation. These results are in agreement with the theoretical analysis made in sections 2.8.1 and 2.8.2, showing that the generalized quadrature variations estimator based on wavelets is not polarized, robust and efficient and recommends the use of the generalized quadrature variations estimator based on wavelets for the estimation of daily WiMAX traffic's H parameter. The appropriateness of this estimator for the LRD analysis of traces from our database can be explained adding two supplementary reasons.

The first reason is the discrete nature of the traces from our data base (the number of packets of data is acquired every 15 minutes) because the generalized quadrature variations estimator based on wavelets was specially conceived for discrete events (see Chapter 2).

The second reason is the non-stationary nature of the traces from our database (already observed in Chapter 3 as a consequence of the evolution in time of the overall tendency of the traffic) which does not affect the precision of the estimation made by the generalized quadrature variations estimator based on wavelets as it was shown in Chapter 2. The traces in the data base are discrete in time non-stationary random processes. The source of non-stationarity is the traffic's overall tendency.

4.5 Estimation of H Parameter Using a Wavelet Based Method

In the following we will continue the LRD analysis of WiMAX traffic by segmenting the time-series having the length equal with a week, into segments with the length equal with a day. In this section, we will apply the generalized quadrature variations Hurst parameter estimator based on wavelets to the daily WiMAX traffic. The wavelet transform with its natural scale invariance and low computational cost is suitable for analyzing of LRD process.

4.5.1 Downlink Traffic

This section presents the evaluation of *H* in the case of WiMAX daily downlink traffic. We have already identified in Chapter 3 a hidden periodicity of the downlink WiMAX traffic, with the period of a day (24 hours), which can be also explained by social reasons, taking into account the cycle day-night. We believe that it is the hidden periodicity in the WiMAX traffic with the shorter period. The goal of this sub-section is to identify the normal behavior of the downlink traffic. The idea is to separate the BSs with normal behavior from the LRD perspective. In the following, we will not be interested in the intrinsic values of the Hurst parameter. The goal will be to identify the days with long-range dependent traffic. We will consider that a day has LRD if the corresponding value of H is bigger than 0.5. The results are shown in Figure 1 in Appendix. The days with long-range dependent traffic are represented in black. For the majority of BSs, the number of black rectangles is bigger than the number of white rectangles. Indeed, downlink daily traffic exhibits LRD. Based on these results we realized the BSs classification in terms of the number of days for which the downlink traffic exhibits LRD (number of values H greater than 0.5). The results are shown in Table 4.3.

We can observe the extreme cases: BS63 with only 14 days with LRD traffic (the best case) and BS32 with 42 days with LRD traffic (the worse case).

4.5.2 Uplink Traffic

In the following we will analyze the *H* parameter for WiMAX uplink daily traffic. Taking into account the fact that the bad positioning of the BS, as source of LRD, affects both uplink and downlink traffic and the fact that there are more sources of LRD in downlink than in uplink, it is preferable to analyze the uplink traffic to identify the bad positioned BSs. We have analyzed all the sixty-six traces of uplink traffic using the generalized quadrature variations estimator based on wavelets, for each of the days of the eight weeks, and we have obtained the results presented in Figure 2 in Appendix. The uplink daily traffic exhibits LRD as well. For the uplink traffic some sources of LRD can be eliminated, but the bad BSs positioning is a common source of LRD for both uplink and downlink traffic.

In Table 4.4 we realized the classification of the sixty-six BSs in terms of the number of days for which the uplink traffic exhibits LRD (number of values H greater than 0.5).

The worse case is given by BS32 for which the number of days with LRD traffic is greater than the number of days without LRD traffic (42 days with LRD traffic). The BSs with the smallest number of days for which the traffic manifest LRD are BS49 and BS62 (17 days with LRD traffic).

Number of values H>0.5	BS
42	BS32
40	BS15
36	BS12, BS13, BS33, BS40
35	BS6, BS25, BS29, BS35, BS38
34	BS21, BS23, BS24, BS50
33	BS36, BS37, BS47
32	BS18, BS19, BS28, BS30, BS31
31	BS44
30	BS2, BS66
29	BS10, BS11, BS14, BS52
28	BS5, BS22, BS26, BS43, BS56
27	BS20, BS39, BS42, BS53, BS67
26	BS17, BS41, BS45, BS62
25	BS1, BS16, BS27, BS55, BS57, BS60
24	BS8
23	BS64
22	BS54, BS59
21	BS4, BS9
20	BS7, BS51
19	BS61, BS 65
18	BS48, BS58
17	BS3, BS46
16	BS49
14	BS63

Table 4.3: BSs classification in downlink.

Number of values H>0.5	BS
42	BS32
35	BS15
34	BS23
33	BS1, BS13, BS24, BS25, BS29, BS33, BS35, BS38,
32	BS21, BS50
30	BS31, BS37
29	BS6, BS26, BS30, BS36, BS41, BS44
28	BS10, BS12, BS66
27	BS20, BS40
26	BS5, BS11, BS22, BS42, BS47
25	BS14, BS19, BS27, BS52, BS54, BS60, BS67
24	BS8, BS18, BS28, BS45, BS55, BS56
23	BS2, BS16, BS39, BS64
22	BS43, BS57, BS62
21	BS17, BS53
19	BS9, BS65
18	BS46, BS59
17	BS58, BS61
16	BS4, BS7, BS51
15	BS48
14	BS3
12	BS49, BS62

Table 4.4: BSs classification in uplink.

4.5.3 BSs localization analysis in uplink and downlink

The goal of this section is the comparison between the daily uplink and downlink traffic for each BS. If the value estimated for Hurst parameter, corresponding to a

given day, was greater than 0.5 we decided that the traffic of that day exhibits LRD.

The existence of LRD in the downlink channel is marked in Figure 3, in Appendix, with red and the existence of LRD in the uplink channel is marked with green. If in a day both channels are characterized by LRD then we have a black rectangle in Figure 3 in Appendix.

The number of days without LRD is greater than the number of days with LRD for a number of twenty-two BSs. These BSs are: BS7, BS49, BS61, BS3, BS4, BS8, BS9, BS16, BS17, BS46, BS48, BS51, BS53, BS55, BS58, BS59, BS60, BS62, BS63, BS64, BS65 and BS67. The normal behavior of those BSs is without LRD. Hence, they are well positioned.

The downlink traffic contains more days with LRD than the uplink traffic for fifty-three BSs. So, the normal behavior of one BS supposes more LRD in downlink than in uplink. Taking into account the fact that the bad positioning of the BS affects both uplink and downlink traffic, it is preferable to analyze the uplink traffic to identify the bad positioned BSs. So, we can use the classification in Table 4.4.

For other twenty-five BSs, the number of days with LRD is smaller than the number of days without LRD in uplink. These BSs are: BS19, BS52, BS2, BS5, BS10, BS11, BS12, BS14, BS18, BS20, BS21, BS22, BS27, BS28, BS30, BS39, BS40, BS42, BS43, BS45, BS47, BS54, BS56, BS57 and BS66. Because the normal behavior of those BSs is without LRD in uplink, we can consider that they are well positioned as well.

The traffic of other six BSs has an atypically behavior, the number of days with LRD in uplink being greater than the number of days with LRD in downlink. These BSs are: BS23, BS26, BS32, BS35, BS41 and especially BS1. We consider that these BSs could be repositioned at the next network release.

Finally, for thirteen BSs the LRD analysis proposed in this paper is not relevant because there are more days with LRD in uplink than days without LRD. We will come back to these BSs in the following sub-section. These BSs are: BS6, BS13, BS15, BS24, BS25, BS29, BS31, BS33, BS36, BS37, BS38, BS44 and BS50.

4.6 Conclusions

The aim of this section was to identify the BSs bad positioned in a WiMAX network, by traffic analysis. The identification of the BSs bad positioned is very important for the planning of a wireless network. The LRD traffic analysis can be developed further by the investigation of different kinds of traffic which are separated at the input of each BS as for example, best effort traffic or multimedia traffic or voice traffic. We preferred to analyze the traffic in its most general form. We have proved that WiMAX traffic exhibits LRD. This LRD is influenced by the presence of some hidden periodicities in the time series. We have identified the hidden periodicities with the following periods: one month, one week and one day. All of them are produced by social mechanisms. They can be removed by the segmentation of traces. We consider that the analysis of daily traffic could solve the considered problem. The bad position of the BS is a common LRD source for both downlink and uplink channels.

There are more LRD sources for the downlink channel than for the uplink channel. So, for the identification of the BSs bad positioned it is simpler to analyze the uplink daily traffic.

We have presented and compared some Hurst parameter estimators highlighting the superiority of the generalized quadrature variations estimator based on wavelets

for the problem of the identification of the BSs which are bad positioned. This estimator is unbiased, effective and robust, if the mother wavelets used for its implementation has a number of vanishing moments equal or greater than two, as it was proved in Chapter 2. It was conceived for discrete random processes, as are the traces in the considered database are. Its performance is not affected by the non stationarity of the input random processes, as the traces in the considered database are, whose non-stationarity is given by their overall tendencies. These overall tendencies were identified in Chapter 3. The superiority of generalized quadrature variations estimator based on wavelets against the other estimators of Hurst parameter presented in section 2.8 was proved through simulations in Table 4.1 and Table 4.2 and justifies once again the opportunity of using wavelets in communications, giving an unitary character to this thesis.

We were interested only in the presence of LRD, considering that a value of H greater or equal with 0.5 proves the presence of LRD and that a value of H smaller than 0.5 proves the absence of LRD. Based on this assumption we have introduced a new type of network's representation in Figures 1, 2 and 3 showed in Appendix. This is a very simple and suggestive representation of a network, enabling its analysis in uplink and downlink simultaneously. It does not allow to appreciate the LRD degree, but this information seems to be redundant for the BSs positioning analysis.

We realized the classification of the positioning of BSs for the daily traffic in terms of the number of days for which the uplink traffic shows LRD. This classification is presented in Table 4.4. It is very interesting to compare this table with the Table 3.2 which presents the BSs' risk of saturation. The BSs with the bigger number of days with LRD in Table 4.4 are: BS32, BS15, BS23, BS1, BS13, BS24, BS25, BS33, BS35 and BS38. All these BSs have a reduced risk of saturation in accordance with Table 3.2. The BS32 can be found on the last position of Table 3.2, the BS15 is on the 60^{th} position, the BS23 on the 43^{th} position, the BS1 on the 49^{th} position and so on. So, the presence of LRD proves that the corresponding traffic is heavy, which reduces the efficiency of the considered BS. A good network must have an uniform risk of saturation and an uniform efficiency for all its BSs. The last observation permits the reciprocal validation of the results obtained in Chapter 3 (Table 3.2) and Chapter 4 (Table 4.3) despite the fact that both are estimation results.

Applying two estimation techniques, one for the risk of saturation and the second one for the Hurst parameter, we have obtained results which are in agreement. Indeed, a BS with heavy traffic (appreciated with the high frequency of apparition of daily LRD) will have a reduced risk of saturation. This remark can be verified analyzing comparatively Table 3.2 and Table 4.3. The first BSs in Table 4.3 could be found on the last column of Table 3.2 and vice versa. Taking into consideration this new rule, we can refine now the list of BSs for which only the LRD analysis is not relevant. This list is composed by the following BSs: BS6, BS13, BS15, BS24, BS25, BS29, BS31, BS33, BS36, BS37, BS38, BS44 and BS50. Some of them can be found on the last column of Table 3.2. These BSs are: BS6, BS15, BS24, BS25, BS31, BS33, BS36, BS37 and BS38. Taking into consideration their reduced risk of saturation we believe that those BSs must be also repositioned. BS13, BS29 and BS44 are well positioned because they have a high risk of saturation too. Finally, for BS50 we can not come with a conclusion because for this BS the value of μ_{ot} in Table 3.2 has an aberrant value.

Chapter 5

Conclusions and Perspectives

The aim of this thesis is finding an answer to the following question: "It is possible to identify the BSs which are bad positioned in a WiMAX network using traffic analysis? Taking into consideration the big volume of information contained in the database which represented the object of investigation for this thesis, data mining was preferred as working tool. Generally, data mining techniques require high computational complexity. One of the phases of a data mining project is data preparation. A modality to reduce the computational complexity is to use an alternative representation of data in this phase of data mining project. For this reason the association of data mining techniques with the wavelet theory was assumed in this thesis. The pretext of this thesis is a database containing uplink and downlink traffic traces for 66 BSs composing a WiMAX network. Two data mining techniques, forecasting and LRD analysis, were applied in the wavelet domain. The SWT was used in the first case and the DWT was used in the second one. The well time-frequency behavior of the WTs permitted the fast and appropriate treatment of these non stationary signals. Both data mining techniques were applied in statistical form.

The most general form of traffic was chosen for the experimental part of this thesis, considering all types of packets collected at the input of a BS. This choice was made to obtain the results in their most general form. This is a non-parametric strategy and provides very robust results.

5.1 Contributions

In Chapter 3, we tested an algorithm for time series prediction proposed in [54] for wired networks, in the case of wireless networks. This method is based on the SWT and statistical time series analysis techniques, but can be viewed as an implementation of the CRISP-DM methodology. We analyzed historical information, continuously collected during a period of eight weeks, at the level of each BS composing the WiMAX network. The main contributions of the thesis contained in Chapter 3 can be summarized as follows:

The utility of the algorithm proposed in [54] was validated in the case of wireless networks. The genuine algorithm was adapted for wireless traffic. In the phase of data preparation were considered two sequences of detail wavelet coefficients as result of the MRA, and the ANOVA procedure was modified to find the weights of those sequences, which minimize the mean square error of the approximation of traffic

variability. In the phase of modeling, the application of the Box-Jenkins methodology was modified also, applying a new test of stationarity.

Predictions provided accurate estimates with a minimal computational cost (all the forecasts were obtained in seconds). The BSs with higher risk of saturation were identified.

A strategy for the selection of MWs, based on their time-frequency localization was proposed. It was verified by simulations that in the case of traffic forecasting, the time localization is the most important feature of the MW used to compute the SWT. The best forecasting results are obtained using the Haar MW. The SWT represents the best choice of wavelet transform for wireless traffic forecasting. This is due to its translation invariance.

The new forecasting algorithm proposed in this thesis is flexible enough to work with many different datasets such as network traffic, financial data or transportation data, without requiring important modifications.

We have compared the proposed prediction algorithm with other algorithms, developed in our research team and published in companion papers, based on neural networks, and we proved its utility for long term predictions. The proposed forecasting algorithm is faster than other forecasting algorithms due to the use of wavelets (the wavelet transforms are fast), due to the use of MRA (we used only three sequences of data: c_6, d_3 and d_4) and due to the use of weekly averages. It does not require any training phase.

In Chapter 4 we analyzed the traffic data in a WiMAX network, in order to identify its particularities. The strategy chosen for this purpose is based on the LRD of traffic. The presence of LRD in network traffic has significant impact on the network performance. The performance of wireless communication networks depends on an efficient architecture (good positioning of base stations). The thesis contributions in Chapter 4 can be summarized as follows:

The uplink and downlink traffic of a wireless network was analyzed in terms of LRD. It was observed that WiMAX network traffic exhibits LRD. A cause for LRD appearance, which is typical for wireless networks, was highlighted: the periodicities of one month, one week and one day.

Some Hurst parameter's estimators were compared and the superiority of the estimator based on wavelets was proved by simulations. Using the second order DWT statistical analysis presented in section 2.6, in equations (2.27) to (2.49), we have proposed a new very simple estimation method of the Hurst parameter in equation (2.95), which works for second order wide sense stationary random processes. It was simply generalized to the Abry-Veitch Hurst parameter's estimator which works for non-stationary continuous in time random processes. Next, this estimator was discretized obtaining the generalized quadrature variations Hurst parameter's estimator based on wavelets, which was applied to the traffic traces from the data base.

Using the R/S estimator of the Hurst parameter, it was observed that LRD can be reduced by splitting the time series corresponding to each BS into daily series. We observed that normally, the daily traffic through a BS should not manifest LRD. During the days with LRD traffic, some traffic anomalies appeared.

The positioning of BSs in the topology of the WiMAX network was analyzed. The BSs for which the number of days with LRD traffic both in uplink and downlink is high, are incorrect positioned. This time the generalized quadrature variations Hurst parameter's estimator based on wavelets was used due to its better performance in comparison with the R/S estimator. This fact was proved in Table 4.1 and Table 4.2.

The results show which BSs have a good localization in the topology of the network and which have not. The BSs which have a bad localization in the topology of the network must be repositioned in the future.

The BSs with bad localization have a reduced risk of saturation as well due to their heavy traffic. This remark permits to make a reciprocal validation of the results of estimation from Chapter 3 with the aid of the estimation results presented in Chapter 4. From a total number of sixty-six BSs the results of the positioning analysis made are not conclusive only for one BS.

5.2 Perspectives

The appearance of LRD could be the result of some anomalies that occur during some days. Anomaly detection refers to the problem of finding patterns in data that do not conform to the expected behavior. A PhD thesis recently presented in the Communications Department of Electronics and Telecommunications Faculty of "Politehnica" University of Timisoara [61] treated the problem of anomaly detection in wired networks and proposed an anomaly detector based on the association of the hyperanalytic SWT with the forth order cumulant of the traffic trace. A method to identify anomalies in the wireless traffic would be very interesting in further research.

Another future continuation of this research work consists in the statistical analysis of DWT coefficients of a non-stationary random process. This research could lead to find new and better Hurst estimators among other results.

Appendix

BS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
H	0.693	0.628	0.658	0.682	0.691	0.676	0.606	0.665	0.665	0.643	0.657
BS	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
H	0.689	0.656	0.692	0.645	0.641	0.706	0.641	0.618	0.657	0.657	0.600
BS	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
H	0.723	0.706	0.717	0.679	0.740	0.656	0.665	0.619	0.637	0.719	0.653
BS	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
H	0.678	0.729	0.667	0.626	0.719	0.697	0.698	0.756	0.622	0.660	0.681
BS	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**
H	0.641	0.560	0.608	0.618	0.704	0.667	0.629	0.703	0.628	0.654	0.636
BS	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**
H	0.603	0.591	0.669	0.648	0.664	0.661	0.581	0.631	0.657	0.727	0.628

Table 1: H values for the time series corresponding to all 66 BSs in downlink.

Week	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	BS10	BS11	BS12	BS13	BS14
1	0.732	0.534	0.680	0.594	0.715	0.631	0.524	0.692	0.617	0.636	0.624	0.597	0.622	0.747
2	0.724	0.153	0.656	0.544	0.651	0.651	0.512	0.534	0.662	0.541	0.596	0.426	0.550	0.574
3	0.662	0.334	0.527	0.446	0.622	0.636	0.570	0.593	0.615	0.550	0.599	0.580	0.567	0.636
4	0.736	0.580	0.575	0.480	0.535	0.528	0.604	0.500	0.647	0.458	0.645	0.586	0.600	0.593
5	0.737	0.568	0.507	0.547	0.635	0.597	0.563	0.605	0.618	0.565	0.698	0.569	0.508	0.643
6	0.724	0.662	0.424	0.617	0.599	0.548	0.515	0.570	0.626	0.523	0.640	0.550	0.559	0.635
7	0.693	0.545	0.398	0.698	0.575	0.511	0.622	0.591	0.730	0.504	0.781	0.610	0.616	0.691
8	0.634	0.597	0.435	0.541	0.651	0.495	0.538	0.573	0.628	0.574	0.601	0.548	0.661	0.512

Week	BS15	BS16	BS17	BS18	BS19	BS20	BS21	BS22	BS23	BS24	BS25	BS26	BS27	BS28
1	0.745	0.607	0.624	0.561	0.470	0.170	0.788	0.549	0.689	0.681	0.711	0.707	0.814	0.530
2	0.640	0.545	0.715	0.487	0.661	0.538	0.677	0.516	0.639	0.649	0.645	0.669	0.642	0.724
3	0.645	0.528	0.726	0.554	0.449	0.616	0.594	0.725	0.656	0.637	0.682	0.644	0.590	0.627
4	0.643	0.590	0.651	0.547	0.568	0.560	0.547	0.633	0.506	0.611	0.622	0.633	0.584	0.633
5	0.613	0.524	0.753	0.451	0.530	0.595	0.596	0.592	0.696	0.594	0.709	0.616	0.663	0.681
6	0.534	0.562	0.685	0.424	0.603	0.623	0.642	0.573	0.626	0.656	0.572	0.633	0.726	0.594
7	0.726	0.553	0.694	0.505	0.541	0.519	0.568	0.630	0.593	0.630	0.685	0.629	0.543	0.469
8	0.662	0.543	0.453	0.603	0.663	0.653	0.659	0.488	0.551	0.515	0.548	0.565	0.632	0.762

Week	BS29	BS30	BS31	BS32	BS33	BS35	BS36	BS37	BS38	BS39	BS40	BS41	BS42	BS43
1	0.595	0.434	0.434	0.727	0.467	0.605	0.739	0.727	0.661	0.702	0.675	0.711	0.629	0.637
2	0.470	0.490	0.490	0.679	0.521	0.657	0.609	0.537	0.580	0.611	0.624	0.658	0.742	0.587
3	0.547	0.542	0.542	0.707	0.640	0.514	0.590	0.515	0.557	0.630	0.633	0.633	0.683	0.610
4	0.522	0.535	0.535	0.733	0.617	0.637	0.612	0.668	0.552	0.688	0.681	0.679	0.634	0.539
5	0.594	0.573	0.573	0.344	0.636	0.505	0.709	0.580	0.535	0.669	0.724	0.730	0.687	0.568
6	0.538	0.619	0.619	0.511	0.656	0.597	0.667	0.612	0.520	0.544	0.708	0.620	0.615	0.524
7	0.552	0.637	0.637	0.695	0.654	0.714	0.690	0.589	0.617	0.534	0.651	0.626	0.572	0.592
8	0.578	0.689	0.689	0.481	0.602	0.554	0.697	0.618	0.662	0.705	0.479	0.653	0.595	0.572

Week	BS44	BS45	BS46	BS47	BS48	BS49	BS50	BS51	BS52	BS53	BS54	BS55	BS56	BS57
1	0.565	0.634	0.734	0.093	0.574	0.714	0.704	0.629	0.542	0.654	0.632	0.623	0.497	0.622
2	0.582	0.542	0.630	0.618	0.590	0.375	0.466	0.568	0.540	0.714	0.562	0.627	0.615	0.628
3	0.466	0.571	0.707	0.698	0.542	0.473	0.484	0.584	0.502	0.616	0.544	0.652	0.546	0.602
4	0.489	0.541	0.641	0.611	0.526	0.441	0.535	0.548	0.580	0.661	0.554	0.593	0.588	0.614
5	0.581	0.597	0.572	0.550	0.520	0.450	0.469	0.628	0.622	0.724	0.524	0.623	0.569	0.561
6	0.559	0.419	0.587	0.593	0.518	0.435	0.507	0.576	0.662	0.571	0.498	0.635	0.475	0.584
7	0.702	0.587	0.532	0.549	0.550	0.563	0.725	0.596	0.440	0.586	0.582	0.550	0.589	0.517
8	0.601	0.583	0.618	0.573	0.448	0.574	0.484	0.600	0.504	0.589	0.456	0.504	0.533	0.583

Week	BS58	BS59	BS60	BS61	BS62	BS63	BS64	BS65	BS66	BS67				
1	0.552	0.653	0.642	0.649	0.651	0.527	0.687	0.574	0.770	0.608				
2	0.517	0.590	0.625	0.568	0.625	0.510	0.618	0.636	0.668	0.606				
3	0.583	0.611	0.539	0.525	0.576	0.459	0.549	0.595	0.683	0.534				
4	0.501	0.549	0.501	0.638	0.510	0.519	0.491	0.620	0.719	0.539				
5	0.519	0.552	0.546	0.580	0.586	0.528	0.512	0.645	0.626	0.530				
6	0.553	0.555	0.498	0.489	0.616	0.514	0.552	0.591	0.587	0.582				
7	0.499	0.544	0.524	0.472	0.577	0.525	0.457	0.565	0.490	0.451				
8	0.577	0.484	0.454	0.532	0.518	0.514	0.483	0.599	0.587	0.565				

Table 2: Weekly values of H, corresponding to 66 BSs in downlink.

BS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
H	0.718	0.746	0.646	0.673	0.695	0.652	0.599	0.656	0.666	0.636	0.656
BS	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
H	0.674	0.652	0.692	0.657	0.607	0.648	0.613	0.600	0.637	0.737	0.603
BS	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
H	0.754	0.707	0.730	0.690	0.721	0.658	0.657	0.638	0.651	0.723	0.648
BS	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
H	0.702	0.720	0.667	0.629	0.718	0.676	0.714	0.743	0.628	0.658	0.695
BS	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**
H	0.631	0.637	0.597	0.598	0.697	0.654	0.631	0.713	0.614	0.655	0.614
BS	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**
H	0.594	0.585	0.650	0.649	0.653	0.661	0.567	0.622	0.722	0.619	

Table 3: H values for the time series corresponding to all 66 BSs in uplink.

Week	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	BS10	BS11	BS12	BS13	BS14
1	0.735	0.664	0.665	0.562	0.694	0.614	0.512	0.680	0.581	0.615	0.624	0.587	0.606	0.730
2	0.736	0.605	0.656	0.529	0.648	0.623	0.485	0.503	0.629	0.513	0.596	0.372	0.536	0.558
3	0.680	0.653	0.510	0.410	0.611	0.617	0.646	0.583	0.609	0.511	0.599	0.513	0.511	0.555
4	0.766	0.595	0.558	0.472	0.531	0.485	0.606	0.450	0.629	0.472	0.645	0.561	0.567	0.589
5	0.750	0.551	0.459	0.513	0.602	0.589	0.563	0.566	0.608	0.526	0.698	0.543	0.487	0.612
6	0.717	0.665	0.393	0.630	0.572	0.517	0.566	0.511	0.629	0.485	0.640	0.538	0.534	0.559
7	0.676	0.544	0.401	0.670	0.550	0.497	0.585	0.571	0.717	0.473	0.781	0.604	0.592	0.655
8	0.612	0.570	0.479	0.531	0.647	0.417	0.519	0.541	0.621	0.552	0.601	0.502	0.680	0.472

Week	BS15	BS16	BS17	BS18	BS19	BS20	BS21	BS22	BS23	BS24	BS25	BS26	BS27	BS28
1	0.745	0.585	0.556	0.530	0.462	0.710	0.800	0.553	0.683	0.680	0.695	0.720	0.820	0.633
2	0.640	0.534	0.706	0.470	0.625	0.505	0.680	0.526	0.608	0.650	0.663	0.658	0.558	0.724
3	0.645	0.491	0.684	0.514	0.500	0.603	0.629	0.732	0.666	0.643	0.663	0.634	0.558	0.599
4	0.643	0.553	0.598	0.543	0.550	0.560	0.497	0.616	0.458	0.592	0.636	0.646	0.545	0.632
5	0.613	0.519	0.653	0.415	0.517	0.574	0.610	0.569	0.665	0.543	0.708	0.583	0.648	0.639
6	0.534	0.482	0.623	0.439	0.574	0.605	0.652	0.525	0.639	0.649	0.654	0.598	0.701	0.559
7	0.726	0.554	0.670	0.490	0.522	0.495	0.591	0.595	0.530	0.643	0.694	0.608	0.507	0.499
8	0.662	0.513	0.423	0.584	0.603	0.631	0.687	0.460	0.541	0.470	0.471	0.576	0.625	0.769

Week	BS29	BS30	BS31	BS32	BS33	BS35	BS36	BS37	BS38	BS39	BS40	BS41	BS42	BS43
1	0.581	0.530	0.653	0.763	0.426	0.622	0.739	0.718	0.665	0.710	0.665	0.716	0.652	0.647
2	0.446	0.557	0.676	0.696	0.526	0.662	0.571	0.524	0.625	0.622	0.626	0.662	0.716	0.540
3	0.516	0.556	0.693	0.719	0.644	0.486	0.507	0.518	0.521	0.611	0.641	0.661	0.666	0.583
4	0.475	0.557	0.738	0.800	0.602	0.617	0.592	0.663	0.554	0.658	0.592	0.691	0.609	0.474
5	0.581	0.565	0.634	0.366	0.629	0.506	0.711	0.565	0.540	0.680	0.691	0.711	0.680	0.513
6	0.527	0.622	0.575	0.482	0.658	0.566	0.672	0.618	0.490	0.527	0.688	0.626	0.614	0.483
7	0.530	0.614	0.616	0.647	0.644	0.726	1.219	0.546	0.596	0.529	0.662	0.624	0.626	0.619
8	0.585	0.436	0.710	0.502	0.580	0.550	0.682	0.612	0.656	0.658	0.525	0.626	0.574	0.539

Week	BS44	BS45	BS46	BS47	BS48	BS49	BS50	BS51	BS52	BS53	BS54	BS55	BS56	BS57
1	0.539	0.668	0.735	0.307	0.580	0.696	0.695	0.611	0.516	0.674	0.603	0.602	0.499	0.594
2	0.562	0.601	0.638	0.591	0.541	0.349	0.460	0.556	0.502	0.711	0.505	0.612	0.567	0.614
3	0.425	0.612	0.662	0.762	0.518	0.426	0.449	0.567	0.467	0.619	0.496	0.589	0.494	0.582
4	0.463	0.555	0.615	0.646	0.517	0.401	0.536	0.545	0.537	0.651	0.504	0.558	0.556	0.566
5	0.549	0.592	0.531	0.589	0.504	0.392	0.422	0.602	0.598	0.709	0.499	0.599	0.521	0.586
6	0.535	0.521	0.550	0.679	0.512	0.366	0.486	0.589	0.649	0.496	0.476	0.630	0.411	0.580
7	0.658	0.582	0.503	0.521	0.587	0.509	0.713	0.597	0.418	0.551	0.575	0.548	0.569	0.489
8	0.625	0.590	0.567	0.579	0.437	0.556	0.482	0.552	0.478	0.544	0.455	0.490	0.505	0.553

Week	BS58	BS59	BS60	BS61	BS62	BS63	BS64	BS65	BS66	BS67				
1	0.533	0.646	0.617	0.637	0.654	0.491	0.685	0.573	0.770	0.611				
2	0.482	0.558	0.577	0.526	0.591	0.500	0.578	0.600	0.645	0.544				
3	0.539	0.609	0.518	0.496	0.566	0.440	0.486	0.599	0.677	0.502				
4	0.486	0.548	0.497	0.635	0.514	0.497	0.447	0.593	0.699	0.520				
5	0.471	0.519	0.524	0.553	0.590	0.502	0.487	0.639	0.627	0.491				
6	0.525	0.531	0.491	0.458	0.569	0.501	0.515	0.552	0.538	0.540				
7	0.468	0.530	0.518	0.449	0.550	0.462	0.461	0.542	0.448	0.429				
8	0.560	0.460	0.436	0.504	0.494	0.531	0.438	0.619	0.582	0.550				

Table 4: Weekly values of H, corresponding to 66 BSs in uplink.

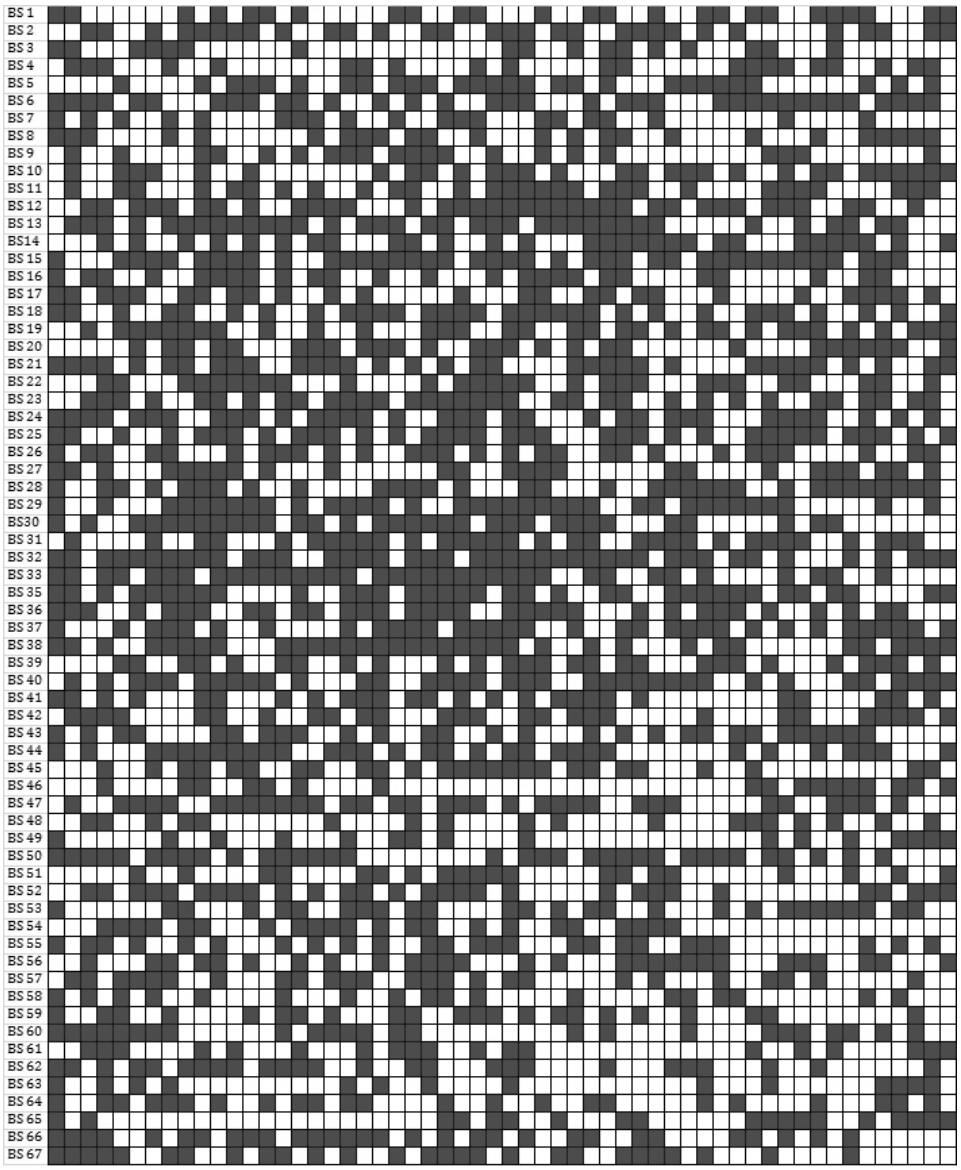

Figure 1: The values of H corresponding to all 66 BSs, daily series in downlink. The Hurst parameter's values bigger than 0.5 are representsed in black.

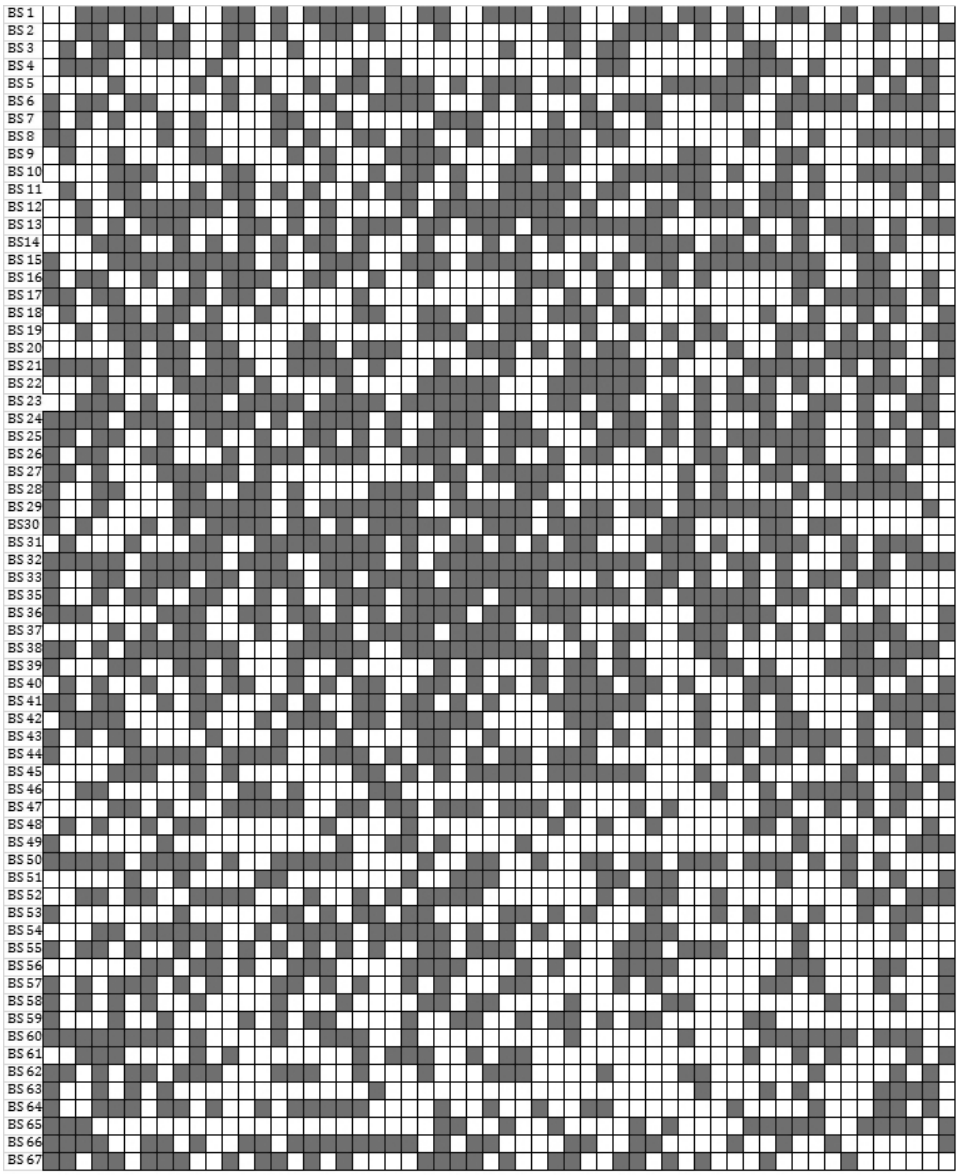

Figure 2: The values of H corresponding to all 66 BSs, daily series in uplink.

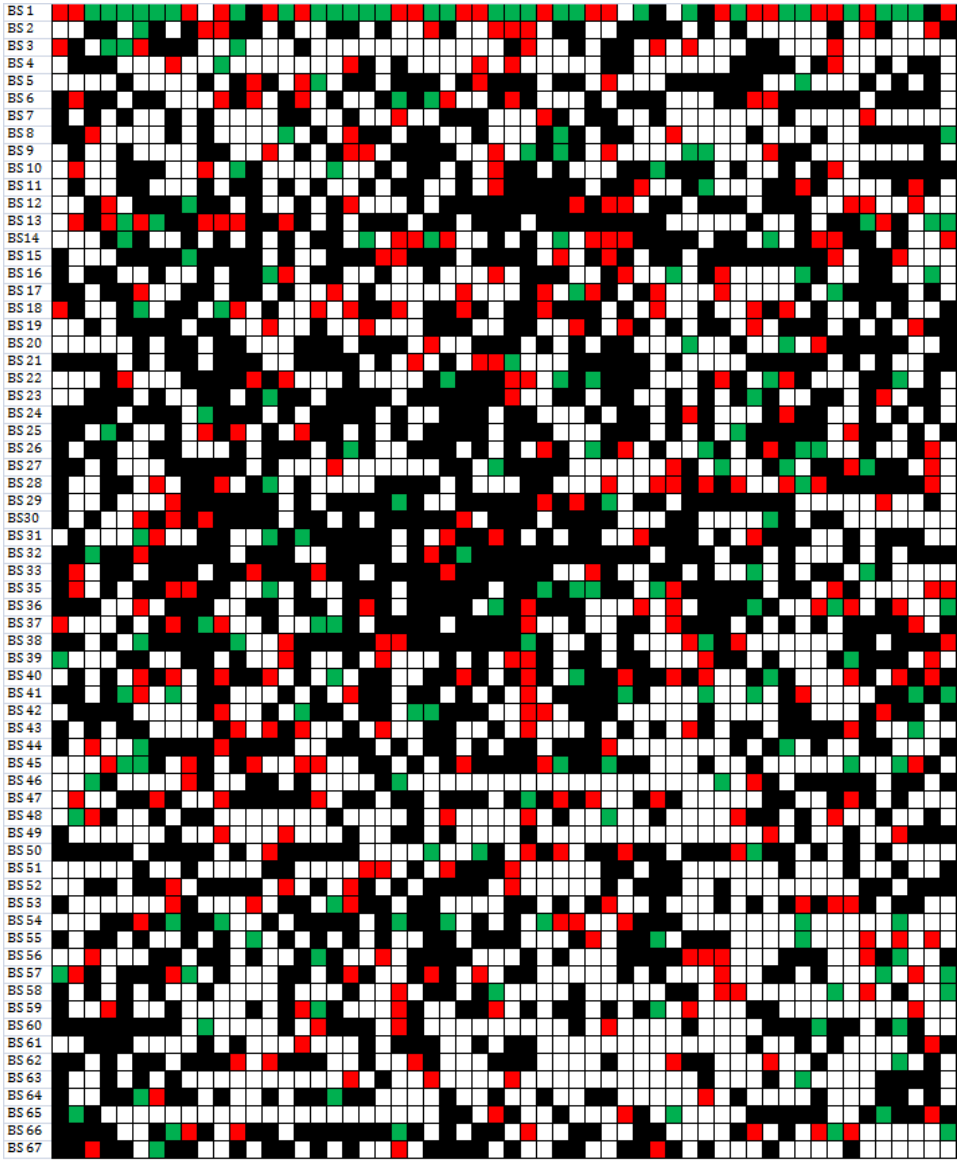

Figure 3: The LRD comportment of the considered WiMAX network.

Bibliography

[1] http://en.wikipedia.org/wiki/Gram-Schmidt_process.

[2] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch. *Self-Similarity and Long-Range Dependence Through the Wavelet Lens*, chapter Theory and Applications of Long-Range Dependence, pages 527–556. P. Doukhan, G. Oppenheim and M.S. Taqqu, eds., 2003.

[3] P. Abry and D. Veitch. Wavelet analysis of long range dependent traffic. *IEEE Trans. Information Theory, 1998, vol. , pp.2-15.*, 44(1):2–15, 1998.

[4] N. Addison. *The Illustrated Wavelet Transform Handbook*. Taylor and Francis; 1st edition, 2002.

[5] S. Basu and A. Mukherjee. Time series models for internet traffic. In *24th Conference on Local Computer Networks*, pages 164–171, 1999.

[6] C. Bénéteau and P. Van Fleet. An introduction to wavelets and their application to digital imaging. In *Workshop SACNAS'09, Dallas, TX*, 2009.

[7] B. Boulet. *Fundamentals of Signals and Systems*. Da Vinci Engineering Press, 1st edition, 2005.

[8] G. E. P. Box, G. M. Jenkins, and G. Reinsel. *Time-series analysis: forecasting and control*. Prentice Hall; 3rd edition, 1994.

[9] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2002.

[10] T. Chan and J. Shen. Image processing and analysis: Variational, pde, wavelet and stochastic methods. *SIAM*, 2005.

[11] K. Chandra, C. You, G. Olowoyeye, and C. Thompson. Non-linear time-series models of ethernet traffic. Technical report, Technical Report, CACT, 1998.

[12] C. Chatfield. *Time-series forecasting*. Chapman and Hall/CRC Press, 2001.

[13] C. Cheng. Wavelet signal processing of digital audio with applications in electro-accoustic music. Master's thesis, Hanover, New Hampshire, 1996.

[14] R. G. Clegg. *The Statistics of Dynamic Networks*. PhD thesis, University of York, Department of Mathematics, 2004.

[15] R. G. Clegg. A practical guide to measuring the hurst parameter. In *Proceeding of 21st UK Performance Engineering Workshop, School of Computing Science Technical Report Series, CS-TR-916, Univ. of Newcastle*, 2005.

[16] P. Cortez, M. Rio, Pedro Sousa, and M. Rocha. Topology aware internet traffic forecasting using neural networks. *Springer Berlin / Heidelberg, Lecture Notes in Computer Science*, 4669/2007., 2007.

[17] M. E. Crovella and A. Bestavros. Self-similarity in world wide web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5:835–846, 1997.

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

[19] I. Daubechies. Ten lectures on wavelets. *SIAM*, 1992.

[20] H. B. Deane, C. Smythe, D. J. Jefferies, and Guildford Gu Xh. Self-similarity in a deterministic model of data transfer. *Journal of Electronics (80),*, pages 677–691, 1996.

[21] P. Chapman et al. Crisp-dm 1.0 step-by-step data mining guide, 2000.

[22] Y. Shu et al. Traffic prediction using farima models. In *Proceedings of IEEE International Conference on Communications,*, volume 1, pages 891–895, 1999.

[23] I. Firoiu. *Complex Wavelet Transform. Application to Denoising*. PhD thesis, Universitatea Politehnica, Timisoara, 2010.

[24] P. Flandrin. *Representation temps-fréquence*. Hermes, 1993.

[25] J. Fourier. Mémoire sur la propagation de la chaleur dans les corps solides. *Nouveau Bulletin des sciences par la Société philomatique de Paris. I. Paris: Bernard.*, pages 112–116, 1808.

[26] J. Franke, W. K. Hardle, and C. M. Hafner. Statistics of financial markets: An introduction, second edition. *Springer-Verlag Berlin Heidelberg*, 2008.

[27] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, 93:429–457, 1946.

[28] R. Gencay, F. Selcuk, and B. Whitcher. An introduction to wavelets and other filtering methods in finance and economics. *Elsevier*, 2001.

[29] Wei-Bo Gong, Y. Liu, V. Misra, and D. Towsley. Self-similarity and long range dependence on the internet: A second look at the evidence, origins and implications. *Computer Networks*, 48:377–399, 2005.

[30] S. Gowrishankar and P. S. Satyanarayana. A time series modeling and prediction of wireless network traffic. *International Journal of Interactive Mobile Technologies*, 3 (1), 2009.

[31] A. Graps. An introduction to wavelets. *IEEE Computational Sciences and Engineering*, 2:50–61, 1995.

[32] N. K. Groschwitz and G. C. Polyzos. A time series model of long-term nsfnet backbone traffic. In *Proceeding of the IEEE International conference on Communications*, 1994.

[33] A. Grossman and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15 (4):723–736, 1984.

[34] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69 (3):331–371, 1910.

[35] C. M. Hurvich and C. L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989.

[36] A. Isar and I. Nafornita. *Reprezentari timp-frecventa*. Editura Politehnica, Timisoara, 1998.

[37] J. Istas and G. Lang. Quadratic variations and estimation of the local holder index of a gaussian process. *Annales Institut Henri Poincaré*, 33 (4):407–436, 1987.

[38] S. Jaffard, Y. Meyer, and R. Dean Ryan. Wavelets: tools for science and technology. *SIAM*, 2001.

[39] W. Willinger K. Park. Self-similar network traffic: An overview. `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.88.475`, 1999.

[40] T. Karagiannis, M. Faloutsos, and R. Riedi. Long-range dependence: Now you see it, now you don't! In *Global Internet Symposium (in IEEE GLOBECOM), Taipei, Taiwan*, 2002.

[41] T. Karagiannis, M. Molle, and M. Faloutsos. Long-range dependence: Ten years of internet traffic modeling. *IEEE Internet Computing. Special Issue - Measuring the Internet,*, 2004.

[42] N. Z. Kolev. *Sonar Systems*. InTech, 2011.

[43] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. In *Proceedings of ACM SIGCOMM '93*, pages 183–193, 1993.

[44] S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition, Wavelet Analysis and Its Applications*. Academic Press, 1999.

[45] V. Misra. Measurement, modeling, and analysis of the internet: Part ii. In *IMA Workshop, Probability and Statistics in Complex Systems: Genomics, Networks, and Financial Engineering, University of Minnesota, USA.*, 2003.

[46] H. Zare Moayedi and M.A. Masnadi-Shirazi. Arima model for network traffic prediction and anomaly detection. In *International Symposium on Information Technology, ITSim2008*, pages 1–6, 2008.

[47] P. A. Morettin. Wavelets in statistics. In *Tutorial in The Third International Conference on Statistical Data Analysis Based on The L1 Norm and Related Methods*, 1997.

[48] J. Olkkonen. *Discrete Wavelet Transforms-Theory and Applications*. InTech, 2011.

[49] M. Oltean. *Radio transmission improvement with wavelets*. PhD thesis, Politehnica University of Timisoara, 2010.

[50] M. Oltean and A. Isar. On the time-frequency localization of the wavelet signals, with application to orthogonal modulations. In *Proceeding of the International Symposium on Signals, Circuits and Systems, ISSCS 2009*, pages 1–4, 2009.

[51] A. V. Oppenheim and Alan S. Willsky. *Signals and Systems*. Prentice Hall; 2 edition, 1996.

[52] S. Ostring and H. Sirisena. The influence of long-range dependence on traffic prediction. In *IEEE International Conference on Communications, ICC2001*, volume 4, pages 1000–1005, 2001.

[53] J. C. Ramirez Pacheco and D. Torres Roman. Accuracy of time-domain algorithms for self-similarity: An empirical study. In *Proceedings of the 15th International Conference on Computing (CIC'06)*, pages 379–385, 2006.

[54] K. Papagiannaki, N. Taft, Z. Zhang, and C. Diot. Long-term forecasting of internet backbone traffic: Observations and initial models. In *Proceedings of IEEE Infocom, San Francisco*, 2003.

[55] C. R. Qiu and Z. Rong. A case based reasoning system for individual demand forecasting. In *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM'08*, pages 1–6, 2008.

[56] I. Railean, S. Moga, M. Borda, and C. Stolojescu. Neural networks vs genetically optimized neural networks in time series prediction. In *Proceedings of SMTDA2010, Chania, Crete*, 2010.

[57] I. Railean, C. Stolojescu, S. Moga, and P. Lenca. Wimax traffic forecasting based on neural networks in wavelet domain. In *Proceedings of the 4th International Conference on Research Challenges in Information Science, RCIS 2010, Nice, France*, 2010.

[58] L. Rokach and O. Maimon. *The Data Mining and Knowledge Discovery Handbook: A Complete Guide for Researchers and Practitioners*. Springer, 2005.

[59] G. Rutka. Neural network models for internet traffic prediction. *Proceedings of Electronics and Electrical Engineering, Lithuania*, 4(68):55–58, 2006.

[60] G. Rutka and G. Lauks. Study on internet traffic prediction models. *Proceedings of Electronics and Electrical Engineering, Lithuania, Kaunas, 21-23. May, 2007*, 6(78):47–50, 2007.

[61] M. Salagean. *Utilizarea reprezentarilor timp-frecventa in analiza fenomenelor nestationare*. PhD thesis, "Politehnica" University of Timisoara, 2011.

[62] A. Sang and S. Li. A predictability analysis of network traffic. In *Proceedings of IEEE INFOCOM'2000, Tel Aviv, Israel*, 2000.

[63] T. K. Sarkar, M. Salzar-Palma, and M. C. Wicks. *Wavelet Applications in Engineering Electromagnetics*. Artech House, 2002.

[64] M. J. Shensa. Discrete wavelet transform. wedding the a trous and mallat algorithms. *IEEE Transactions and Signal Processing*, 4:2464–2482, 1992.

[65] C. Stolojescu, S. Moga, P. Lenca, and A. Isar. A wavelet based prediction model for time-series. In *Proceedings of SMTDA2010, Chania, Crete*, 2010.

[66] M. S. Taqqu and V. Teverovsky. *On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. A Practical Guide to Heavy Tails: Statistical Techniques and Applications,*. 1998.

[67] P. Ulanovs and E. Petersons. Modeling methods of self-similar traffic for network performance evaluation. *Scientific Proceedings of RTU. Series 7, Telecommunications and Electronics*, 2002.

[68] A. Veres and M. Boda. The chaotic nature of tcp congestion control. In *Proceedings of IEEE INFOCOM*, 2000.

[69] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice Hall, 1995, 1995.

[70] X. Wang and X. Shan. A wavelet-based method to predict internet traffic. In *Proceedings of IEEE Communications, Circuits and Systems and West Sino Expositions*, pages 690–694, 2002.

[71] W. Willinger, V. Paxson, and M. S. Taqqu. Self-similarity and heavy tails: Structural modeling of network traffic. `http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.30.6033`, 1998.