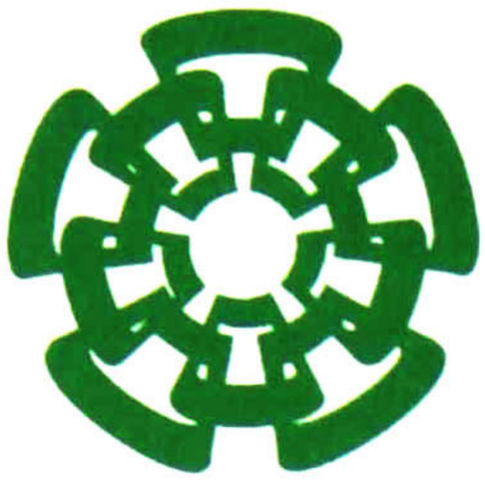
CT-888.551
DOU: 2015



Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional
Unidad Guadalajara

Análisis y modelado de procesos auto-similares con aplicación en comunicaciones de VoIP

Tesis que presenta:

Leopoldo Estrada Vargas

para obtener el grado de:
Doctor en Ciencias

en la especialidad de:
Ingeniería Eléctrica

Director de Tesis
Dr. Deni Librado Torres Román

CINVESTAV del IPN Unidad Guadalajara, Guadalajara, Jalisco, Febrero de 2015.

Análisis y modelado de procesos auto-similares con aplicación en comunicaciones de VoIP

Tesis de Doctorado en Ciencias
Ingeniería Eléctrica

Por:

Leopoldo Estrada Vargas

M.C. en la especialidad de Ingeniería Eléctrica

Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional 2004-2006

Becario de CONACYT, expediente no. 166532

Director de Tesis

Dr. Deni Librado Torres Román

CINVESTAV del IPN Unidad Guadalajara, Febrero de 2015.

Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional
Unidad Guadalajara

Self-similar time series: analysis and modeling with applications to VoIP

Tesis que presenta:

Leopoldo Estrada Vargas

para obtener el grado de:
Doctor en Ciencias

en la especialidad de:
Ingeniería Eléctrica

Director de Tesis
Dr. Deni Librado Torres Román

CINVESTAV del IPN Unidad Guadalajara, Guadalajara, Jalisco, Febrero de 2015.

Self-similar time series: analysis and modeling with applications to VoIP

Tesis de Doctorado en Ciencias
Ingeniería Eléctrica

Por:

Leopoldo Estrada Vargas

M.C. en la especialidad de Ingeniería Eléctrica

Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional 2004-2006

Becario de CONACYT, expediente no. 166532

Director de Tesis

Dr. Deni Librado Torres Román

CINVESTAV del IPN Unidad Guadalajara, Febrero de 2015.

Centro de Investigación y de Estudios Avanzados

del Instituto Politécnico Nacional

Unidad Guadalajara

Análisis y modelado de procesos auto-similares con aplicación en comunicaciones de VoIP

Autor: Leopoldo Estrada Vargas

Director de tesis: Dr. Deni Librado Torres Román

Resumen:

Este trabajo se divide en 2 partes. La primera es el análisis matemático de series de tiempo auto-similares. Los estimadores clásicos para la media y varianza de una muestra, o sus estadísticas, comúnmente llevan a cálculos polarizacos cuando las observaciones de dicha muestra están correlacionadas. Para proponer estimadores no polarizados para estas estadísticas, se propone un conjunto de funciones ortonormales (de tipo *wavelet*). A través del análisis de muestras mediante ese conjunto de funciones se obtienen estimaciones no polarizadas de la media, varianza y el parámetro de Hurst (H) de series de tiempo. Uno de los principales resultados de este trabajo consiste en la clarificación acerca de estimador basado en la varianza de las series agregadas: varios autores afirman que el estimador es polarizado y recomiendan no usarlo más que para obtener una estimación burda el índice de auto-similitud (H). En este trabajo se muestra que esto es debido a una mala interpretación de un problema mal planteado, es decir, para estimar el parámetro H es necesario obtener las respectivas varianzas de las series agregadas, pero éstas a su vez dependen del valor de H. Se plantea una alternativa a este problema, de la cual se deriva un estimador no polarizado, con varianza mínima, basado en wavelets. La segunda parte de este trabajo consiste en el desarrollo de una metodología para el análisis estadístico y el modelado de las características de una comunicación VoIP (*Voice over Internet Protocol*), tales como el retardo de los paquetes y su pérdida en la red. Para este propósito, las ecuaciones de análisis presentadas en la primera parte se usan en combinación con otros modelos estadísticos como los procesos discretos de Markov (para capturar el comportamiento en ráfaga de las pérdidas de paquetes), el factor R del modelo E de la ITU-T, y la distribución de Cauchy (para describir la natura del *jitter*). La referencia de máxima calidad de servicio (QoS) es la voz PCM 64*kbps* 8bits/muestra, tal como en la red telefónica pública conmutada (RTPC). Los resultados de este trabajo sirven para incrementar la calidad de servicio percibida con presencia de deteriorios, mediante el ajuste adaptivo de algunos parámetros como el tiempo entre salida de paquetes, la redundancia y el tamaño del buffer de recepción. Los análisis y metodologías propuestos se verifican mediante simulaciones.

Centro de Investigación y de Estudios Avanzados

del Instituto Politécnico Nacional

Unidad Guadalajara

Self-similar time series: analysis and modeling with applications to VoIP communications

Author: Leopoldo Estrada Vargas

Thesis advisor: Dr. Deni Librado Torres Román

Summary:

This work is divided in two parts. The first one consists of the mathematical analysis of self-similar time series. For that, a set of orthonormal functions (wavelet-type) is proposed. Such analysis allows the unbiased estimation of the mean, variance, and Hurst index (H) of those series, differently from the classic estimators, which are biased always except for the case where there is no correlation. One of the main results is the clarification about the variance-based H estimator. Several works affirm that this estimator is biased and they recommend not to use it, but in this work it is explained that this is a misinterpretation of an ill conditioned problem: to estimate H it is needed to calculate the variances of the aggregated series, but these also depend on H. A solution to this ill conditioned problem is proposed, leading to an unbiased, minimal variance waveled-based method. The second part of this work consists of the development of a methodology for 1) the statistical modeling of the characteristics of a VoIP communication, such as packet delay and loss, and for 2) the optimization of the Quality of Service (QoS) by setting adaptively the adjustable parameters, such as: the packet inter-departure time, the level and type of redundancy and the de-jitter buffer size. The reference of "maximum possible QoS" is the 64kbps 8-bit/sample PCM voice codification (either mu-law or A-law) with zero delay and loss, similarly to the Public Switched Telephone Network service. The proposed methodology estimates the probabilities of loss and the delay from the last n packets and sets the mentioned adjustable parameters in order to achieve the maximum QoS, measured using the E-model's R factor, for the reception of the following packet (e.g., packet n+1). The proposed estimator considers that packet loss can be modeled using two-state discrete-time Markov chains, and that the packet delay is correlated so that it can be predicted using a linear predictor. An estimation of the impairment of the Quality of Service when using two different codifications, i.e., one type for the normal information and other type for the redundant data, is proposed. The proposed methodology is verified by simulation and measurements.

Agradecimientos

A mi querido Dios, por amarme incondicionalmente y por cuidarme en cada paso de mi vida.

A mis padres, Arturo y Graciela, por su amor y su generosidad sin límite. Dios los bendiga.

A mi asesor, Dr. Deni Torres, por su paciencia y por todo el conocimiento que he obtenido con su guía.

A CONACYT, por el apoyo económico.

Y a Lupita, mi esposa amada, por inspirarme a ser una mejor persona cada día.

Acknowledgements

To my dear God, for loving me unconditionally and for taking care of me at every step through my life.

To my parents, Arturo and Graciela, for their endless love and generosity. God bless them.

To my advisor, Dr. Deni Torres, for his patience and for all the knowledge I have got with his guidance.

To CONACYT, for the financial support.

And, to my lovely wife, Lupita, for inspiring me to be a better person everyday.

Index

Abstract

This work is divided in two parts. The first one consists of the mathematical analysis of self-similar time series. The classical estimators for the sample mean and variance, or their statistics, commonly lead to biased calculations when the observations of the sample under study are correlated. In order to propose unbiased estimators for these statistics, a set of orthonormal functions (wavelet-type) is proposed. Through the analysis of samples by means of the proposed orthonormal basis, unbiased estimation of the mean, variance, and Hurst index (H) of those series is achieved. One of the main results of this work is the clarification about the variance-based H estimator: several authors affirm that this estimator is biased and they recommend not to use it but to provide a rough estimation of the self-similarity index. In this work it is explained that this is a misinterpretation of an ill-conditioned problem, i.e., to estimate H it is needed to calculate the variances of the aggregated series, but it is shown that these also depend on H. A solution to this problem is proposed, leading to an unbiased, minimal variance waveled-based method. The second part of this work consists of the development of a methodology for the statistical analysis and modeling of the characteristics of a *Voice over Internet Protocol* (VoIP) communication, such as packet delay and loss. For that purpose, the analysis equations in the first part of this work are used in combination with other statistical models such as the discrete Markov processes (to capture the *bursty* behavior of packet losses), the E-model's R factor (to estimate the perceived Quality of Service, or QoS, of a VoIP call), and the Cauchy distribution function (to describe the nature of the packet delay jitter). The reference corresponding to the defined maximum possible QoS is the $64kbps$ 8bit/sample Pulse Code Modulation (PCM) voice codification (either μ-law or A-law) with zero delay and loss, similarly to the Public Switched Telephone Network service. The results of this work can be used to increase the perceived QoS, in the presence of impairments, by setting adaptively some adjustable parameters, such as: the packet inter-departure time, the level and type of redundancy, and the de-jitter buffer size. The proposed analyses and methodologies are verified by simulations and measurements.

Introduction

In the past decades, the self-similar processes and long range dependence (LRD), also named long memory, have been applied to the study and modeling of many natural and man-made complex phenomena. These kinds of processes have been particularly attractive in the pursuit of optimal design and configuration of network communications.

The published work of Leland et al in 1993 demonstrated that Ethernet traffic is statistically self-similar and that the commonly used models are unable to capture that fractal behavior; highlighting that a *burstiness* and LRD are present when $H > 0.5$ [1], [2]. Since then, researchers have been studying extensively long memory processes and their impact on network performance, e.g., Karagiannis et al stated that the identification of LRD is not trivial and that not all scenarios in modern networks present LRD characteristics, e.g., traffic in the Internet backbone is more likely to be Poisson type instead of LRD [3].

Many researchers have also addressed their studies to determine if network traffic is sufficiently modeled by self-similar processes or a more general model is needed, e.g., one that considers *multiscaling* or *multifractality* [4], [5], [6]. The advantage of the capability to model complex systems with self-similar processes is that the correlation structure is defined by a single parameter: the Hurst index (H).

Unlike other statistics, the Hurst index, although it is mathematically well defined, cannot be estimated unambiguously from real world samples. Several methods have been developed then in order to estimate it. Examples of classical estimators are those based on R/S statistic [7] (and its unbiased version [8]), detrended fluctuation analysis (DFA) [8], [9], maximum likelihood (ML) [10], aggregated variance (VAR) [8], wavelet analysis [11], [12], etc. In [13], Clegg developed an empirical comparison of estimators for data in raw form and corrupted. An important observation is that the estimation of the Hurst index may differ from one estimator to another, and the selection of the most adequate estimator is a difficult task. This selection depends greatly on how well the data sample meets the assumptions the estimator is based on. However, through analytical and empirical studies it has been discovered that the estimators that have the best performance in bias and standard deviation, and, consequently, in mean squared error (MSE), are Whittle ML and the wavelet-based estimator proposed by Veitch and Abry in [11]. From these two estimators, the wavelet-based is computationally simpler and faster [11], [7].

In addition to the Hurst index, other statistical characteristics are needed to describe the phenomenon under study. The most common are the first and second order statistics, i.e., mean, variance, and correlation. The classical estimators of these characteristics have been proposed

decades ago, e.g., Kenney and Keeping demonstrated in 1939 that the classical variance estimator is unbiased for independent and identically distributed Gaussian observations [14], [15]. Confidence interval is also given for these estimations, e.g., $P\left[\bar{X} \in \left(-3/\sqrt{N}, 3/\sqrt{N}\right)\right] \approx 99\%$ (where $\bar{X}$ is the sample mean estimated from a sample of size N) for a standardized white noise process. This confidence interval is narrower as the sample size increases.

It has been claimed that most processes satisfy those common assumptions [16]. As it has been expressed, however, other authors (Leland et al [1], [2], Taqqu et al [5], Tsybakov et al [17], Veitch and Abry [11], and many others) conclude that traffic characteristics present correlation and that the estimation of these statistics (including confidence interval) with the classical estimators (which do not consider correlation) may lead to estimation errors and, consequently, to wrong decisions or inaccurate models, especially when data presents accentuated LRD.

Several estimators of the Hurst index have been proposed, but many of them do not consider the effect of correlation on the estimation of first and second order statistics, thus applying incorrectly the classical formulae. Particularly, it has been claimed that the aggregated variance method can only be used as a heuristic method, and that the *Variance-Plot* (see definition in Section 1.1.4), also named *Variance-time Plot*, can only be used to check whether the time series is self-similar or not and, if so, to obtain a crude guess for the Hurst index [18 pp. 44]. Section 3.6 clarifies this point, demonstrating that the *Variance-Plot* can be estimated efficiently and that the estimation of the Hurst index from it is actually unbiased and has minimum variance (similarly to the wavelet-based estimator).

The importance of self-similar stochastic processes in communication networks is due to its random and complex nature. It is the result of the convergence of information and media transmission (data, voice, and video) through the same communication channel. As there are a very high (and increasing) number of nodes (i.e., devices) connected to the network, and these are being added in a random, decentralized manner, the network is asymmetric and also practically random in both its topology and its usage. Additionally, the service provided by the Internet is generally a "best effort" type, which means that the nodes, with some exceptions, e.g. [19], [20], do not differentiate between traffic types and there is neither resource reservation nor prioritization.

Congestion, due to the high demand of network resources, is a cause of the impairment of the Quality of Service, which consists of delay problems (i.e., the delay and its variation, namely the delay *jitter*, are higher) and packet loss. For time-critical communications, such as *Voice over Internet Protocol* (VoIP), end-to-end delay and packet loss can have high impact on quality of service [21]. A set of techniques are implemented in order to reduce the impairment of the time-critical communications. *Multiple packet transmission* (MPT) and *forward error correction* (FEC) are used to reduce packet losses, at the expense of bandwidth and delay. Many codification

schemes of the redundant information on later packets have been proposed in the literature [22]. The *automatic repeat request* (ARQ) technique, the correction scheme of the *transmission control protocol* (TCP), is generally not suitable for real-time applications, which have a tighter delay tolerance than usual. Also, the presence of a de-*jitter* buffer at the receiver helps compensating the packet delay variation, and even packet reordering, at the expense of additional delay and cost. After all, a de-jitter buffer is a memory buffer and its size depends, among other factors, on the delay variation.

This document is organized as follows: theoretical basis is presented In Chapter 0, including the definitions related to probability, self-similar processes, discrete Markov processes, and quality of VoIP. Chapter 2 describes wavelet-based analysis and synthesis of self-similar time series, and formulae regarding the variance of the aggregated series and the wavelet components, which are used throughout the next chapters. Chapter 3 explains how the mean and variance of discrete self-similar processes are distributed. It clarifies also a spread misinterpretation of the popular *Variance-Plot*, settling that it does not underestimates the Hurst index by itself, but that it can be used to estimate it without bias and with minimal variance. In Chapter 4 the E-model and the R-factor are described [23], as well as the improvements achieved by using FEC and two different voice codecs in the same flow. Also, algorithms to generate artificial series of loss and delay are presented. In Chapter 5, the performance of the estimators of the Hurst index, the mean and variance of second-order self-similar processes as well as the synthesis of Gaussian- and Cauchy-distributed time series are evaluated. Artificial and real world time series are used for that purpose. Chapter 0 summarizes the work.

Objectives

This work is motivated by the mentioned importance of the self-similar processes in many areas, especially in the analysis and modeling of Internet traffic, and by the fact that there are still some misunderstandings and bad practices that must be overcome. The main objectives are summarized as follows:

- To provide a mathematical basis for the analysis of time series, particularly those corresponding to discrete self-similar processes and discrete Markov processes.
- Based on the mathematical definitions, to derive additional formulae regarding the statistics of those series, such as the Hurst index, the mean, variance, correlation, and distribution.
- To describe a VoIP communication channel, and to identify several applications of the presented mathematical concepts to improve the quality of the communication.
- To evaluate the performance of the proposed estimators and improvements using artificial and real world time series.

1 Theoretical Basis

1.1 Self-similarity

Self-similarity describes the phenomenon where certain properties are preserved irrespective of scaling in space or time. Deterministic self-similarity is clearly exemplified by popular figures as Sierpinski's triangle or Koch's snow flake. This form of self-similarity is named scale invariance, and makes different scales of the same object undistinguishable. Stochastic self-similarity is not that obvious; it refers to how statistical properties of a stochastic process are preserved under time expansion. Stochastic self-similarity is defined for continuous and discrete time stochastic processes.

Self-similarity (either continuous or discrete time) is tightly related to short- and long-range dependencies (SRD and LRD, respectively). The degree of self-similarity is defined and measured through the so named Hurst index H ($0 < H < 1$). It is known that processes with $H < 0.5$ are SRD, and processes with $H > 0.5$ are LRD. If $H = 0.5$, neither SRD nor LRD are present. E.g., the commonly used white Gaussian noise (WGN) has always $H = 0.5$ and does not present any dependency.

Processes with LRD are also named long-memory, as current and future realizations of these are strongly correlated. The dividing line between SRD and LRD processes is not ambiguous: for LRD processes the auto-covariance function is not absolutely convergent (i.e., the sum is not finite), while it is for SRD processes. This work refers only to stochastic discrete time self-similarity.

1.1.1 Internet traffic is self-similar

Although created by man and machine, the complexity of Internet traffic is such that in many ways it requires treatment as a natural phenomenon. Even if each atomic component of the network is well understood, the whole is so complex that it must be measured and its emergent properties "discovered" [1]. For more than a decade, network researchers have been heavily using self-similar and LRD models, as these seem to capture the bursty behavior of many network characteristics [1], [24]. Despite of its widespread use, LRD analysis is hindered by our difficulty in actually identifying dependence and estimating its parameters unambiguously [3]. Authors have been researching about the estimation of the statistical parameters characterizing self-

similarity and LRD, for both off-line and real-time estimation, and various estimators have been proposed in the pursuit of higher accuracy and efficiency [11], [13], [25], [26], [27].

The discovery of LRD in packet data was followed by detailing evidence for multi-fractal behavior of TCP/IP traffic in WANs. However, in terms of networking physical mechanisms for such behavior have never been convincingly demonstrated. It has been shown that the "evidence" for multi-fractal behavior of aggregated traffic is actually weak in many ways. Veitch et al state that "pseudo scaling" is often confused with true scaling, due to shortcomings in the statistical tools [28].

Self-similarity describes the phenomena where certain properties are preserved irrespective of scaling in space or time. It can be defined as follows:

Definition 1: A real valued continuous time stochastic process $\{Y(t), -\infty < t < \infty\}$ is said to be self-similar if for any constant $a > 0$, there exists $0 < H < 1$, called index of self-similarity or Hurst parameter, such that $Y(at) \sim a^H Y(t); \forall a > 0; t \in R$, where the symbol $\sim$ means equality in the sense of finite-dimensional distributions.

1.1.2 Self-similarity and Long Range Dependence

A common definition of LRD [11] is the slow, power-like decrease at large lag of the *auto-covariance function* (ACV) of a stochastic process X_t, given by:

$$\gamma_X(k) \sim c_\gamma |k|^{-(1-\alpha)};\ \alpha \in (0,1) \tag{1}$$

Equivalently, it can be defined as the power-law divergence at the origin of its spectrum:

$$f_X(\upsilon) \sim c_f |\upsilon|^{-\alpha};\ \upsilon \to 0 \tag{2}$$

Each of these definitions includes two parameters: (α, c_γ) or (α, c_f) respectively, which are related as:

$$c_f = 2(2\pi)^{-\alpha} c_\gamma \Gamma(\alpha) \sin[(1-\alpha)\pi/2] \tag{3}$$

where Γ is the Euler function.

The parameter α is related to the Hurst exponent as:

$$H = \frac{1+\alpha}{2} \tag{4}$$

1.1.3 Discrete Time Self-similarity

When considering discrete stochastic time series the definition of self-similarity is given in terms of the aggregated processes. Let $\{X_t; t \in \mathbb{N}\}$ be a discrete time series derived from a self-similar

process with stationary increments (H-SSSI); then, others series can be obtained by aggregation. The new aggregated time series is a sequence given by (5):

$$X(m) = \{X_k^{(m)}; k \in \mathbb{N}\} \tag{5}$$

where each term X_k^m is defined as:

$$X_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i \, ; k \in \mathbb{N} \tag{6}$$

and where m represents the aggregation level. That is, each new time series is obtained by partitioning the original time series into non-overlapping blocks of size m and then averaging each block to obtain its respective values.

Let X_t be a covariance stationary discrete time series with mean $\mu_X = 0$, variance σ_X^2 and ACV $\gamma_X(k)$, and $X_k^{(m)}$ its aggregated series. Then it is said that X_t is self-similar, H-ss, if the following (7) holds [29]:

$$X_k^{(m)} \sim m^{H-1} \cdot X_t \tag{7}$$

where $\sim$ means equality in distribution.

The definition of discrete self-similarity (7) has important implications; some of them are [5]:

i. Zero mean

$$E(X_t) = E\left(X_k^{(m)}\right) = 0 \tag{8}$$

ii. Power law of the q^{th} order moments:

$$E\left[\left(X_k^{(m)}\right)^q\right] = m^{q(H-1)} \cdot E[(X_t)^q] \tag{9}$$

iii. Power law of the q^{th} order absolute moments:

$$E\left(\left|X_k^{(m)}\right|^q\right) = m^{q(H-1)} \cdot E(|X_t|^q) \tag{10}$$

Many estimators are based on these properties, per example, the classical variance estimator is based on (10) for $q = 2$.

1.1.4 Second-order Discrete Time Self-similarity

A second-order definition of self-similarity is derived from (9) with $q = 2$. The variance and covariance of the aggregated time series are defined, respectively, by (11) and (12) [17], i.e.:

$$var\left(X_k^{(m)}\right) = m^{2H-2} \cdot var(X_t) \tag{11}$$

and

$$\gamma_X^{(m)}(k) = \frac{\sigma_X^2}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]; k \geq 0 \tag{12}$$

The normalized ACV, also named correlation coefficient, is then:

$$\rho_X(k) = \frac{\gamma_X^{(m)}(k)}{\sigma_X^2} = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}];\ k \geq 0 \tag{13}$$

If a discrete time series X_t satisfies these conditions, it is called second-order self-similar with Hurst index H (H-SOSS).

It is widely known that H-SOSS processes with $H > 0.5$ present LRD and, in this case, the sum of their ACV function diverges (see Section 2.4). Note also that the mean of an H-SOSS process is not necessarily zero.

The plot $\log\left[var\left(X_k^{(m)}\right)\right]$ vs. $\log(m)$ is usually known as *Variance-Plot*. It is a straight line of slope $2H - 2$ for self-similar processes. This plot is the basis of the variance-based estimator of the Hurst index. It has been "shown" in the literature that the variance-based estimator underestimates the Hurst index and that the variance-based estimator throws a coarse estimation of the true Hurst index. Section 3.6 demonstrates that this is a consequence of inadequate implementations of this estimator, i.e., the aggregated variance is estimated with the classical formula [25], which is not correct if there exists correlation. An apparent solution is to use the proposed unbiased estimator, but that leads to an ill-conditioned problem: the Hurst index is needed to estimate the variance and vice versa. The solution for this situation is also described.

1.1.5 Wavelet Decomposition and the *Logscale Diagram*

The wavelet decomposition transforms a signal X_t into a sum of orthogonal components as follows:

$$X_t = \sum_{j=i}^{J} \sum_{k=1}^{2^j} d_X(j,k)\psi_{j,k}(t) \tag{14}$$

where each function $\psi_{j,k}(t)$ is derived from a basis function $\psi_0(t)$, namely the mother wavelet, by scaling and displacement, i.e.,

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi_0(2^{-j}t - k) \tag{15}$$

and coefficients $d_X(j,k)$ is the value at time k of scale j, computed as a inner product between the signal X_t and the wavelet function $\psi_{j,k}(t)$:

$$d_X(j,k) = \langle X(t), \psi_{j,k}(t)\rangle \tag{16}$$

The statistic $S_2(j)$ is then defined from these coefficients as:

$$S_2(j) = E|d_X(j,\cdot)|^2 \tag{17}$$

which, for an H-SOSS process, is related to the Hurst index as:

$$S_2(j) = c_f C 2^{j(2H-1)} \tag{18}$$

where the quantity $c_f C$, related to the power of the process, is considered a constant.

The plot $log_2 S_2(j)$ vs. j forms the widely known *Logscale Diagram* (*LD-Diagram*) described by Veitch et al [11]. The *LD-Diagram* of an H-SOSS process is a straight line of slope $2H - 1$. To obtain an unbiased estimation of the Hurst index from a real world time series based on the *LD-Diagram*, it is also necessary to subtract the bias that results from averaging the logarithms of the variances, i.e.,

$$E\{\log_2[S_2(j)]\} \approx E\{\log_2[\hat{S}_2(j)]\} - \hat{g}_j \tag{19}$$

where $\hat{g}_j$ is the estimated bias [11]:

$$\hat{g}_j = \Psi(n_j/2)/\ln(2) - log_2(n_j/2) \tag{20}$$

and n_j is the number of coefficients available at each octave.

The *LD-Diagram* of an ideal second-order self-similar time series is a straight line; for real world traces, in general, it is not. Then a linear regression, per example, can be applied in order to estimate the Hurst index. If the *LD-Diagram* cannot be adequately modeled with a linear model, then either the time series is not self-similar, i.e., the Hurst parameter is not an adequate statistic, or the scaling behavior should be described with more than one scaling parameter, i.e., a single Hurst parameter is insufficient across the spectrum, and a multi-fractal phenomenon occurs [26] [18].

Note that even when the time series under study is not self-similar, the *LD-Diagram* can show whether or not it presents LRD.

1.1.6 Classical Hurst Index Estimators

In this sections, the mathematical basis for several widely used Hurst index estimator is described.

1.1.6.1 R/S Estimator

Let Y_t be the partial sum of X_t: $Y_t = \sum_{i=1}^{t} X_i$. Then, the R/S statistic is defined as:

$$\frac{R}{S}(l) = \frac{1}{S(l)}\left[\max_{0\le t\le l}\left(Y_t - \frac{t}{l}Y_l\right) - \min_{0\le t\le l}\left(Y_t - \frac{t}{l}Y_l\right)\right] \tag{21}$$

where $S^2(l)$ is the variance of X_t and l is called *lag*.

If X_t is self-similar, then the R/S statistic is related to the Hurst parameter as expressed by (22):

$$E\left\{\frac{R}{S}(l)\right\} \sim cl^H; l \to \infty \tag{22}$$

The slope of the line that best fits the log-log plot of $E\left\{\frac{R}{S}(l)\right\}$ vs. l is the estimation of H.

1.1.6.2 AM Estimator

The absolute moment of aggregation level m is defined as:

$$AM^{(m)} = \frac{M}{N}\sum_{k=1}^{N/M}\left|X_k^{(m)} - E\left\{X_k^{(m)}\right\}\right| \tag{23}$$

The definition of discrete self-similarity (7) implies that

$$AM^{(m)} \sim m^{H-1}; m \to \infty \tag{24}$$

Then, the slope ($\hat{s}$) of the line that best fits the log-log plot of $AM^{(m)}$ vs. m and the Hurst parameter estimation ($\widehat{H}$) are related as (25) expresses:

$$\widehat{H} = \hat{s} + 1 \tag{25}$$

1.1.6.3 Variance-Based Estimator

The definition of second-order self-similarity (11) is the basis of the aggregated variance estimator (VAR), which consists of calculating the variance at many aggregation levels, i.e. a set of values for m, then applying a linear regression to the $log\left(var\{X^{(m)}\}\right)$ vs. $log(m)$ plot.

It is known that the classical variance-based estimator underestimates H, particularly for those traces with theoretical H greater than 0.5. Then, it has been claimed that this estimator is useful only to obtain a fast-coarse approximation of the real Hurst parameter and, possibly, to check whether or not the time series under study is H-SOSS [18].

1.1.6.4 Periodogram-based Estimator

The periodogram-based estimator (PER) considers the power-law behavior of the spectral density function of self-similar processes, as expressed by (26). An estimation of H can be obtained by computing a linear model from the log-log plot of the spectral density vs. frequency with the lowest frequencies.

The estimation of the Hurst parameter ($\widehat{H}$) and the slope of the linear model ($\hat{s}$) are related as expressed by:

$$\widehat{H} = \frac{1 - \hat{s}}{2} \tag{26}$$

1.1.6.5 MAVAR Estimator

The modified Allan's variance is defined by the ITU-T [27] as:

$$\sigma_X^{2'}(\tau) = \frac{1}{2n^4\tau_0^2(N-3n+1)} S_n \tag{27}$$

where S_n is:

$$S_n = \sum_{j=1}^{N-3n+1} \left[\sum_{i=1}^{n+j-1} X_{i+2n} - 2X_{i+n} + X_i \right]^2 \tag{28}$$

$n = 1,2,3,\ldots,\left\lfloor \frac{N}{3} \right\rfloor$, $\tau = n\tau_0$ and τ_0 is the sampling period.

The modified variance is related to the Hurst parameter as expressed by (29):

$$\sigma_X^{2'}(\tau) \sim k\tau^{2H-4} \tag{29}$$

Then, the modified Allan variance-based (MAVAR) estimation can be calculated by applying a linear regression to the log-log plot of $\sigma_X^{2'}(\tau)$ vs τ. H is related to the slope $\hat{s}$ as expressed by (30):

$$\widehat{H} = \frac{\hat{s}}{2} + 2 \tag{30}$$

1.1.6.6 Local Whittle Estimator

Whittle estimator has its basis on (31):

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\nu)}{f(\nu,\eta)} d\nu + \int_{-\pi}^{\pi} \log(f(\nu,\eta)) d\nu \tag{31}$$

where η is a vector of parameters, $I(\nu)$ is the periodogram and $f(\nu,\eta)$ is the spectral density function.

A more simple way of estimating H is known as local Whittle (LWHI) estimator, whose basis is:

$$f(\nu) \sim G(H)\nu^{1-2H}; \nu \to 0 \tag{32}$$

For this spectral density, the analog of (31) is:

$$Q(G,H) = \frac{1}{M} \sum_{j=1}^{M} \left(\frac{I(\nu_j)}{G\nu_j^{1-2H}} + logG\nu_j^{1-2H} \right) \tag{33}$$

M must satisfy $M < \frac{N}{2}$ and $\frac{1}{M} + \frac{M}{N} \to 0$ as $N \to \infty$.

An estimation of G function is:

$$\hat{G} = M^{-1} \sum_{j=1}^{M} \frac{I(\nu_j)}{\nu_j^{1-2H}} \tag{34}$$

Then, the estimation $\hat{H}$ is the value that minimizes the function $R(H)$ defined as:

$$R(H) = log(\hat{G}) - (2H-1)M^{-1} \sum_{j=1}^{M} log\nu_j \tag{35}$$

1.1.6.7 Wavelet-based Estimator

This estimator (WAV) is based on the estimation of the wavelet spectrum: the *LD-Diagram* described in section 2.1, i.e., $log_2[E|d_X(j,\cdot)|^2]$ vs. j. The statistic $E|d_X(j,\cdot)|^2$ is an exponential function of α [11]:

$$E|d_X(j,\cdot)|^2 = 2^{j\alpha} c_f C \tag{36}$$

The parameter α is estimated then as the slope of the *LD-Diagram* and it is related to the Hurst index as expressed by (4).

The statistic $E|d_X(j,\cdot)|^2$ is estimated as:

$$E|d_X(j,\cdot)|^2 = 2^j var\left(C_{X,t}^{n,i}\right) - g_j \tag{37}$$

where the time series $C_{X,t}^{n,i}$ are estimated from X_t as defined by (79) and its variance is estimated as (127). g_j is the bias described by (20) [11].

1.2 Inverse Cumulative Distribution Function Transformation

The inverse cumulative distribution function (ICDF) transformation produces a time series Y_t with cumulative distribution function (CDF) $F_Y(y)$ from a random time series X_t with CDF $F_X(x)$ by applying the following sample-to-sample formula:

$$y_t = F_Y^{-1}\left(F_X(x_t)\right) \tag{38}$$

If X_t is uniformly distributed between 0 and 1, as $F_X(x_t) = x_t$, then (38) simplifies to:

$$y_t = F_Y^{-1}(x_t) \tag{39}$$

1.2.1 Gaussian and Cauchy Distributions

The CDF function of a Gaussian RV is defined by the mean (μ) and variance (σ^2) of the distribution:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda \tag{40}$$

In the absence of a closed form for (40), numerical approximations are used to estimate it.

A Cauchy RV has the probability density function (PDF):

$$f_X(x) = \frac{1}{\pi s\left[1 + \left(\frac{x-c}{s}\right)^2\right]} \tag{41}$$

where c, the location parameter, defines the location of the peak of the distribution and s, the scale parameter, specifies the half-width at half-maximum. A standardized Cauchy RV has $c = 0$ and $s = 1$. An approximated standardization for a Cauchy RV is:

$$X_{(0,1)} \sim s^{-1}\left(X_{(s,c)} - c\right) \tag{42}$$

The inverse of (42) produces a Cauchy distributed sample of certain location and scale parameters from a standardized Cauchy sample, i.e.,

$$X_{(c,s)} \sim sX_{(0,1)} + c \tag{43}$$

Unlike a Gaussian RV, there exist closed forms for the CDF and ICDF of a Cauchy RV, which are:

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-c}{s}\right) \tag{44}$$

and

$$F_X^{-1}(p) = c + s \cdot tan\left[\pi\left(p - \frac{1}{2}\right)\right] \tag{45}$$

1.3 Finite-state Discrete Markov Processes

Finite-state discrete Markov processes (also called discrete Markov chains, commonly) are used for the modeling of packets receptions and losses. Consecutive received packets form a gap and consecutive packet losses form a burst. According to the definition of burst used in this work, there are not lost packets within a gap and there are not received packets within a burst. A second (and slightly different) definition is used by other authors [30], which considers gaps as the longest sequence of packets beginning and ending with a loss or the start or end of reception, where the sequences of consecutive received packets are greater than or equal to certain value (G_{min}) in length (measured in packets). Similarly, these authors define the burst as the longest

sequence of packets beginning and ending with a loss, where the sequences of consecutive received packets are less than certain value (G_{min}) in length. Other two quantities need to be defined to describe the packet loss rate (PLR): the gap and burst densities, defined as the percent of received/loss packets within a gap/burst.

The first definition is used in this work in order to compare the two studied models (two- and four-state) easily. Also, this definition helps avoiding the usage of gap and burst densities. Note that the second definition converges to the first by setting $G_{min} = 1$ and, consequently, both densities are equal to 100%. Both definitions include also the burst and gap durations, defined as the timestamp of the end packet minus the timestamp of the beginning packet plus the duration of the end packet (expressed commonly in milliseconds) of the current gap or burst. These durations can be computed from the gap length and burst length distributions, but they are not used directly in the estimation of the E-model's *R* factor.

Previous works on error and loss modeling date from about five decades ago. These works address the distribution of bit errors on telephone channels [30]. These models are also applied on loss modeling for Internet Protocol (IP) networks. Although the losses in IP networks are more common to occur at packet level, instead of bit level, these are correlated due to the time-correlated occupancy of the network. Consecutive packet receptions and losses (respectively named *gap* and *burst*) occur in such way that their respective length follow geometric-type distribution [31] [32]. Bernoulli models are not well suited to this modeling because although they can produce geometric distributions for gap and burst lengths, losses in real networks seem to be correlated, instead of independent. I.e., the probability that a packet is lost (or received) depends on whether or not the previous packet was lost.

At small time scales, i.e. a few seconds or minutes, a two-state Markov process can reproduce the geometric type phenomenon for the homogeneous case, i.e., the PLR is the same (or almost) through certain short time interval, but a non-homogeneous behavior becomes noticeable at larger scales and, in this case, the two-state Markov process is insufficient, thus a more general model is necessary. The four-state Markov process seems to capture or simulate better this widely known non-homogeneous behavior of the characteristics of network traffic. The four-state model approach allows us to represent and simulate those periods with low and high PLR that alternate in sequence according to certain probability.

1.3.1 Matrix Representation of the Steady State

Let $S = S_1, S_2, \ldots, S_m$ be the m states of an m-state Markov process and let p_{ij} be the probability of the process to pass from the state S_i to the state S_j, i.e., $p_{ij} = P(X_i = x_i | X_{i-1} = x_{i-1})$. Having the Markov property means that, given the present state, future states are independent of the past

states, i.e., $P(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \dots) = P(X_{n+1} = x_{n+1}|X_n = x_n)$. The Markov processes used in this work also are time-homogeneous, which means that the probabilities of transition between states are constant over time, i.e., $P(X_{n+1} = x_{n+1}|X_n = x_n) = P(X_n = x_n|X_{n-1} = x_{n-1})$.

All states communicate (are reachable from) each other, which makes the states diagram irreducible. Also, the states diagram is aperiodic, i.e., state S_i can be reached from itself in any number of steps ($n = 1,2,3, \dots$).

The probabilities of transitions between states can be represented by a *transition matrix*. The elements of the one-step $m \times m$ transition matrix $\underline{T}$ are $T_{ij} = p_{ij}$. To obtain the n-step transition matrix it is necessary to multiply the matrix itself n times [33], i.e.,

$$\underline{T_n} = \underline{T}^n \tag{46}$$

As the number of steps (n) increases, the probability of the matrix to be in the state S_i from an initial state depends less on this one. i.e., as n tends to ∞, the matrix $\underline{T_n}$ converges to a matrix with the next form:

$$\underline{T_\infty} = \lim_{n\to\infty} \underline{T_n} = \begin{bmatrix} s_1 & s_2 & \cdots & s_m \\ s_1 & s_2 & \cdots & s_m \\ \vdots & \vdots & & \vdots \\ s_1 & s_2 & \cdots & s_m \end{bmatrix} \tag{47}$$

such that

$$s_1 + s_2 + \cdots + s_m = 1 \tag{48}$$

In (47) and (48), s_i represents the named *steady* probability of state S_i. The steady-state transition matrix $\underline{T_\infty}$ can be obtained then by solving (48) and (49) [34]:

$$\overline{S}\underline{T} = \overline{S} \tag{49}$$

where $\overline{S} = [s_1 \quad s_2 \quad \dots \quad s_m]$.

Assuming that the states diagram is irreducible and aperiodic, the matrix $\underline{T_\infty}$ is well defined and unique.

1.3.1.1 Numerical Approximation

Obtaining analytical expressions for the elements of $\underline{T_\infty}$ (i.e., s_1, $s_2 \dots$) can be difficult when the number of states is large. In this case, a numerical approximation is more suitable, which is described as follows:

Let $\underline{T}$ be a $m \times m$ transition matrix, which has a unique steady-state solution, and let $\{(\lambda_i, \bar{v}_i); i = 1, \dots, m\}$ be its pairs of eigenvalues and eigenvectors (i.e., $\underline{T}\bar{v}_i = \lambda_i \bar{v}_i$), such that $\lambda_i > \lambda_j$ for $i < j$. This matrix $\underline{T}$ can be decomposed into the special form

$$\underline{T} = \underline{PDP}^{-1} \tag{50}$$

where $\underline{P}$ is a matrix composed of the eigenvectors of $\underline{T}$, $\underline{D}$ is the diagonal matrix constructed from the corresponding eigenvalues and $\underline{P}^{-1}$ is the inverse of $\underline{P}$. Then $\underline{T_n}$ can be calculated easily as

$$\underline{T_n} = \underline{PD^n P}^{-1} \tag{51}$$

As all elements of the diagonal of the matrix $\underline{D}$ are lower than 1 except $D_{1,1}$ (which is equal to 1), then

$$\underline{T_\infty} = \underline{PD^\infty P}^{-1} = \underline{PD'P}^{-1} \tag{52}$$

where the only non-zero element of $\underline{D'}$ is $D'_{1,1} = 1$.

This method is also useful when obtaining short-term approximations, i.e., $\underline{T_n} = \underline{PD^n P}^{-1}$ for small n.

1.3.2 Two-state Discrete Markov Process

The two-state Markov process states diagram is shown in Figure 1. State S_1 represents packet loss and S_2, packet reception. Two substitutions ($p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$) are made in order to represent the states diagram with the lowest number of parameters. The steady-state probability of state S_1, i.e., the probability for this process to be in the state S_1 in a random point in time, namely the PLR, is given by (53) [35]:

$$s_1 = \frac{p_{21}}{p_{12} + p_{21}} \tag{53}$$

and clearly $s_2 = 1 - s_1$.

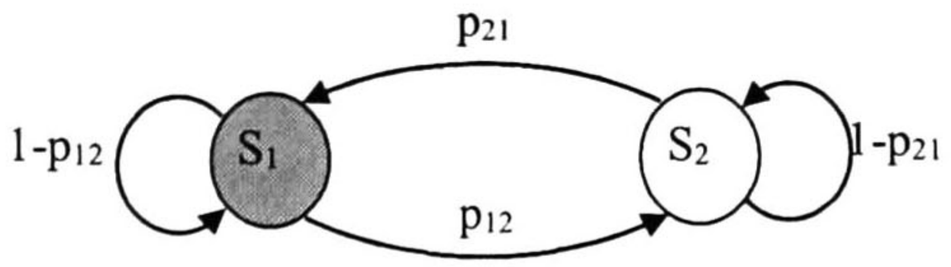


Figure 1. Two-state Markov process states diagram. White and shady circles represent correct and erroneous states, respectively.

The burst and gap length distributions ($f_b(k)$ and $f_g(k)$, respectively) can be expressed in terms of p_{12} and p_{21}, as expressed by (54) and (55):

$$f_b(k) = p_{12}(1 - p_{12})^{k-1} \tag{54}$$

$$f_g(k) = p_{21}(1 - p_{21})^{k-1} \tag{55}$$

which have also respective means $E\{f_b(k)\} = 1/p_{12}$ and $E\{f_g(k)\} = 1/p_{21}$. It is easy to proof (54), as $\sum_{k=1}^{\infty} f_b(k) = 1$ and $f_b(k+1) = f_b(k) \cdot (1 - p_{21})$; and similarly for (55).

1.3.3 Four-state Discrete Markov Process

The four-state Markov process states diagram is shown in Figure 2. Missing arrows indicate zero probability. States S_1 and S_3 (shady circles) represent packet losses (erroneous); S_2 and S_4 (white circles), packet reception (correct).

Six parameters ($p_{21}, p_{12}, p_{43}, p_{34}, p_{23}, p_{32} \in (0,1)$) are necessary to define all the transition probabilities. Without loss of generality, probabilities of transitions between correct states, as well as transitions between erroneous ones, have been set to zero.

The four steady-state probabilities of this process are:

$$s_1 = \frac{1}{1 + \frac{p_{12}}{p_{21}} + \frac{p_{12}p_{23}}{p_{21}p_{32}} + \frac{p_{12}p_{23}p_{34}}{p_{21}p_{32}p_{43}}} \tag{56}$$

$$s_2 = \frac{1}{1 + \frac{p_{21}}{p_{12}} + \frac{p_{23}}{p_{32}} + \frac{p_{23}p_{34}}{p_{32}p_{43}}} \tag{57}$$

$$s_3 = \frac{1}{1 + \frac{p_{34}}{p_{43}} + \frac{p_{32}}{p_{23}} + \frac{p_{21}p_{32}}{p_{12}p_{23}}} \tag{58}$$

$$s_4 = \frac{1}{1 + \frac{p_{43}}{p_{34}} + \frac{p_{32}p_{43}}{p_{23}p_{34}} + \frac{p_{21}p_{32}p_{43}}{p_{12}p_{23}p_{34}}} \tag{59}$$

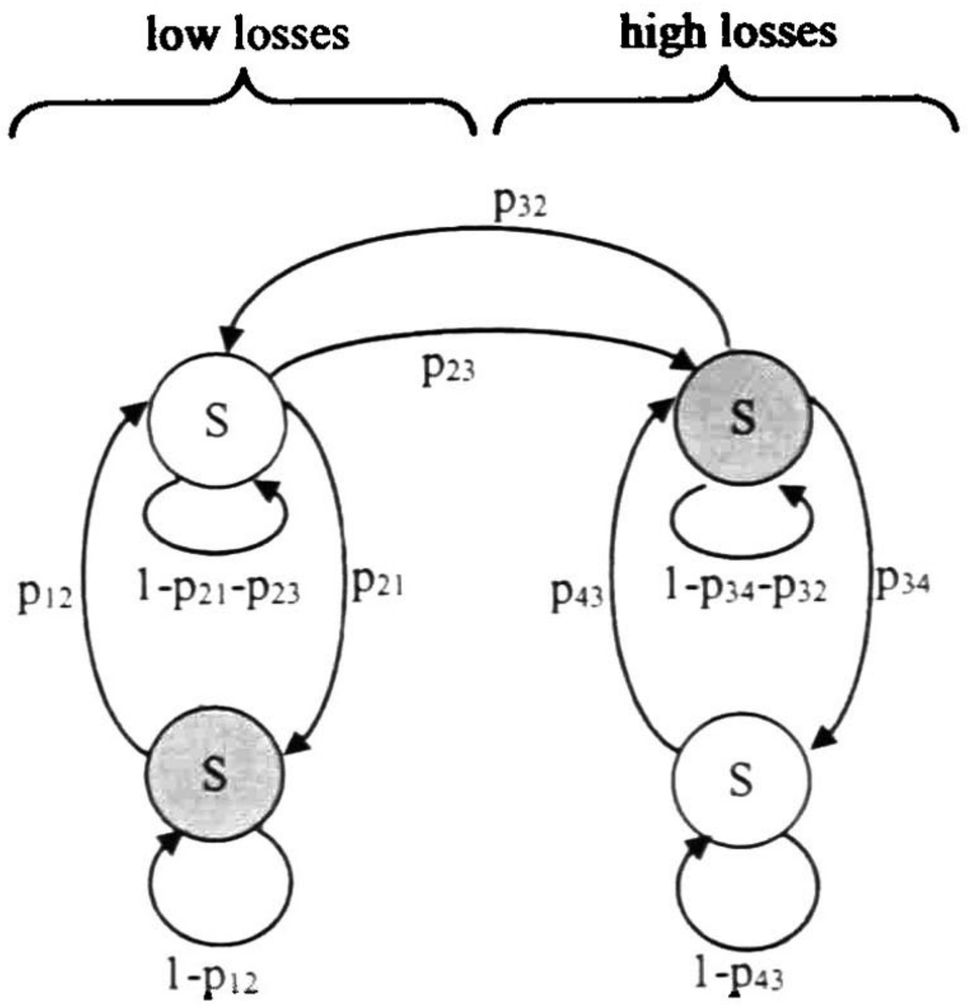


Figure 2. Four-state Markov process states diagram. Only two types of transitions between different states are allowed: from correct to erroneous and from erroneous to correct.

The probability of the process to be either in S_1 or in S_3, i.e., the PLR, is then:

$$r = s_1 + s_3 \tag{60}$$

The average burst length ($\bar{b}$) is calculated as the quotient of the probability of loss and the probability of transition from a lossless state to a loss state (61), that is:

$$\bar{b} = \frac{s_1 + s_3}{s_2(p_{21} + p_{23}) + s_4(p_{43})} \tag{61}$$

Similarly, the average gap length is:

$$\bar{g} = \frac{s_2 + s_4}{s_1(p_{12}) + s_3(p_{34} + p_{32})} \tag{62}$$

It is also easy to proof that the transitions from error state to correct state and vice versa have equal probability, i.e. $s_2(p_{21} + p_{23}) + s_4(p_{43}) = s_1(p_{12}) + s_3(p_{34} + p_{32})$.

The distribution of the burst length can be derived the following the next procedure:

Let $f_b(k)$ denote the probability that the burst length is k; $C_1(k)$, the probability that the burst length is k or greater and the k^{th} transmission is from state S_1 and $C_3(b)$, the probability that the burst length is k or greater and k^{th} transmission is from state S_1 and $C_b(k)$, the probability that the burst length is k or greater such that $C_b(k) = C_1(k) + C_3(k)$ and $f_b(k) = C_b(k) - C_b(k + 1)$. Clearly $C_b(k) = \sum_{i=k}^{\infty} f_b(i)$. Also, as transitions between states S_1 and S_3 have zero probability, $C_1(k+1) = C_1(k)(1 - p_{12}) = C_1(1)(1 - p_{12})^k$ and $C_3(k+1) = C_3(k)(1 - p_{34} - p_{32}) = C_3(1)(1 - p_{34} - p_{32})^k$ Then to calculate $f_b(k)$ it is necessary to obtain $C_1(1)$ and $C_3(1)$, whose respective values are $C_1(1) = s_2p_{21}/[s_2(p_{21} + p_{23}) + s_4p_{43}]$ and $C_3(1) = (s_2p_{23} + s_4p_{43})/[s_2(p_{21} + p_{23}) + s_4p_{43}]$.

As the minimum burst length is 1, $C_b(1) = C_1(1) + C_3(1) = 1$. Then, the distribution of the burst length is:

$$f_b(k) = C_1(1)Q_1(k) + C_3(1)Q_3(k) \tag{63}$$

where $Q_1(k) = (1 - p_{12})^{k-1} - (1 - p_{12})^k = p_{12}(1 - p_{12})^{k-1}$ and $Q_3(k) = (1 - p_{34} - p_{32})^{k-1} - (1 - p_{34} - p_{32})^k = (p_{34} + p_{32})(1 - p_{34} - p_{32})^{k-1}$. As expressed by (63), $f_b(k)$ is the sum of two geometric series with respective rates $1 - p_{12}$ and $1 - p_{34} - p_{32}$; this implies that $f_b(k)$ is a decreasing function of k, i.e., bursts of greater length have lower probabilities than shorter ones.

A similar procedure can be followed to obtain the gap length distribution ($f_g(k)$), which is:

$$f_g(k) = C_2(1)Q_2(k) + C_4(1)Q_4(k) \tag{64}$$

where $C_2(1) = (s_1p_{12} + s_3p_{32})/[s_1p_{12} + s_3(p_{32} + p_{34})]$, $C_4(1) = (s_3p_{34})/[s_1p_{12} + s_3(p_{32} + p_{34})]$, $Q_2(k) = (1 - p_{21} - p_{23})^{k-1} - (1 - p_{21} - p_{23})^k = (p_{21} + p_{23})(1 - p_{21} - p_{23})^{k-1}$ and $Q_4(k) = (1 - p_{43})^{k-1} - (1 - p_{43})^k = p_{43}(1 - p_{43})^{k-1}$. Also note that $C_2(1) + C_4(1) = 1$.

Note that, although the resulting equations correspond to the four-state model of Figure 2, this procedure can be applied for any finite-state Markov process, which consists of finding firstly the cumulative density functions (CDF), i.e., $C_b(k)$ and $C_g(k)$.

1.4 Network Performance Metrics

The analysis and statistical study of time series, that represent network characteristics, are often used for the design of communications systems. These studies are used for the design and testing of communication improvements.

Also, the Quality of Service (QoS) of network communication can be estimated or predicted from these measurements. There are two main issues that must be conducted in order to obtain data samples: 1) configuration of the measurement setting, e.g., phones, gatekeepers, traffic monitors and network interfaces, and 2) the design and realization of the measurement protocols. Additionally to all this work, there is also certain amount of time necessary to capture a data sample of certain size, e.g., the duration of an Internet call. Artificial data generators are then used in order to gather a large volume of data without conducting the mentioned issues and saving a lot of effort and time. The artificial time series must produce time series that are representative, in the statistical sense, of the characteristics of the communication systems, e.g., their distribution or correlation structure.

1.5 Real-time Transport Protocol

The R*eal-time Transport Protocol* (RTP) header format is shown in Table 1 [36].

bit offset	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9	1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	1 9	2 0	2 1	2 2	2 3	2 4	2 5	2 6	2 7	2 8	2 9	3 0	3 1	3 2
0	V		P	X	CC				M	PT							Sequence number															
32	Timestamp																															
64	SSRC ID																															
96	CSRC IDs																															
	...																															
96+32(CC)	Profile-specific extension header ID																Extension header length															
128+32(CC)	...																															

Table 1: RTP header format.

The fields in this header are as follows :

- Version (V): 2-bits. This field identifies the version of RTP. The version defined by this specification is two (2). (The value 1 is used by the first draft version of RTP and the value 0 is used by the protocol initially implemented in the "vat" audio tool.)
- Padding (P): 1 bit. If the padding bit is set, the packet contains one or more additional padding octets at the end which are not part of the payload. The last octet of the padding contains a count of how many padding octets should be ignored (itself included). Padding may be needed by some encryption algorithms with fixed block sizes or for carrying several RTP packets in a lower-layer protocol data unit.
- Extension (X): 1 bit. If the extension bit is set, the fixed header MUST be followed by exactly one header extension.
- Contributors count (CC): 4 bits. The CSRC count contains the number of CSRC identifiers that follow the fixed header.
- Marker (M): 1 bit. Used at the application level and defined by a profile. If it is set, it means that the current data has some special relevance for the application.
- Payload type (PT): 7 bits. Indicates the format of the payload and determines its interpretation by the application.
- Sequence number: 16 bits. The sequence number increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence.
- Timestamp: 32 bits. The timestamp reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations.
- Synchronization source (SSRC): 32 bits.

1.6 VoIP System Components

Figure 3 represents the *source terminal* diagram from a typical VoIP system [30]. The shown blocks have a direct effect on perceived speech and voiceband application quality.

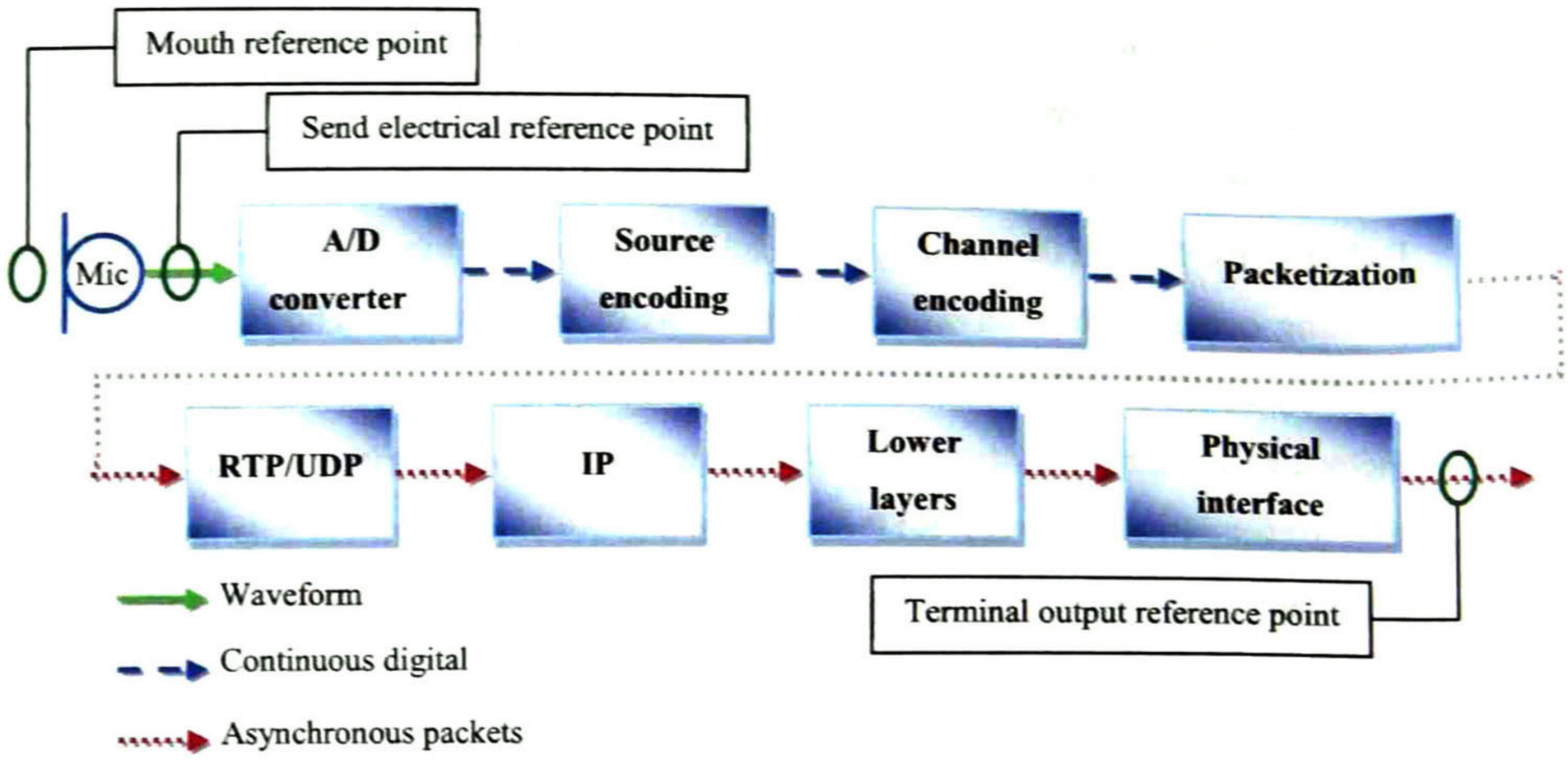


Figure 3: Source terminal diagram and reference points

The *A/D converter* of the source terminal transforms the voice waveform into its digital version. In order to reduce the quantization error, a non-uniform quantification (either *A*-law or μ-law) is applied in its internal quantization block. The source encoding consists of compressing the digital signal in order to reduce redundancy and improve bandwidth utilization. The channel coding is a viable method to reduce information rate by increasing reliability. This goal is achieved by adding redundancy to the information so that it is more distinguishable at the output of the channel (receiver).

The voice packetization protocols use a sequence number field in the transmit packet stream to maintain temporal integrity of voice, allowing the receiver to detect lost packets and to properly reproduce silence intervals during playout. The packetized voice stream traverse to lower layers, i.e., RTP, UDP, etc., and it is sent asynchronously into the network.

The main impairment due to the source terminal is named *source terminal delay*, which is defined as the sum of all delays of the system components belonging to the source, i.e., those between the mouth reference point and the terminal output reference point, more specifically, it is defined as the interval defined by the time that a signal enters the mouth reference point and the time that the first bit of the correspondence encoded, packetized signal exits the terminal output reference point. The source terminal delay variation (ΔS_n) is defined as the time reference between the first bit of a packet emission at the terminal output reference point and the ideal periodic reference time. For the first packet in a flow, the ideal periodic reference time is set equal to the emission time. Subsequent packet emissions are compared to this periodic reference as:

$$\Delta S_n = \hat{t}_n - t_n \tag{65}$$

where t_n is the emission time of the n^{th} packet of the ideal periodic reference stream. Note that this delay variation ΔS_n tends to be more significant as time passes due to the undesired effects

of source frequency offset. This variation should be noted and removed as a measurement error, when possible.

Encoders and decoders process sequences of speech samples known as *frames*. The encoder does not start processing the speech frame until all corresponding samples have been collected. I.e., there is an unavoidable delay before the encoder produce an encoded output, equal to the frame length, before the encoding process can begin. Another delay, defined in [37] is named *look-ahead*, and it corresponds to the time for the encoder to take a preprocessing into the speech frame, in order to determine how to carry out an efficient compression. For an efficient use of processor resources, the time required to process an input frame is assumed to be the same as the frame size (f, in time units). The delay through an encoder/decoder pair ($D_{e/d}$) is sometimes assumed to be then [37]:

$$D_{e/d} = 2f + l \tag{66}$$

In IP networks, the additional delay due to the packetized frame traversing to the IP layer should not exceed the frame size. Otherwise, an output buffer of variable length would be necessary at the output of the encoder and some frames could be also discarded, impairing the quality of the communication unnecessarily. Then, the maximum delay attributable to encoder/decoder processing in IP-based systems operating in real-time must be [37]:

$$D_{e/d} = 3f + l \tag{67}$$

If multiple frames are packetized into a single IP packet, e.g., an IP packet carrying $40ms$ of voice containing four $10ms$-G.729 frames, further delay is added to the speech signal. This delay will be at least the duration of one extra voice frame at the encoder. Then the minimum delay attributable to encoder processing is [37]:

$$D_{e/d} = (K + 1)f + l \tag{68}$$

where K is the number of frames per IP packet.

Similarly to the time necessary to start encoding a speech frame, a packet cannot be delivered to the IP layer until all speech frames, corresponding to that packet, are encoded. And the delivering time for each frame should not exceed its own duration (again, to avoid discarding packets and unnecessary impairments). Then, the maximum delay attributable to encoder/decoder processing in IP-based systems operating in real-time with multiple packets per frame is [37]:

$$D_{e/d} = (2K + 1)f + l \tag{69}$$

Figure 4 exemplifies the delay due to the collection, encoding and delivering of a 3-frames packet into the IP layer. The maximum attributable delay is shown at the bottom of this figure. Note that the look-ahead and processing of the k^{th} frame are carried out simultaneously with the collection of speech samples for the $(k + 1)^{th}$ frame.

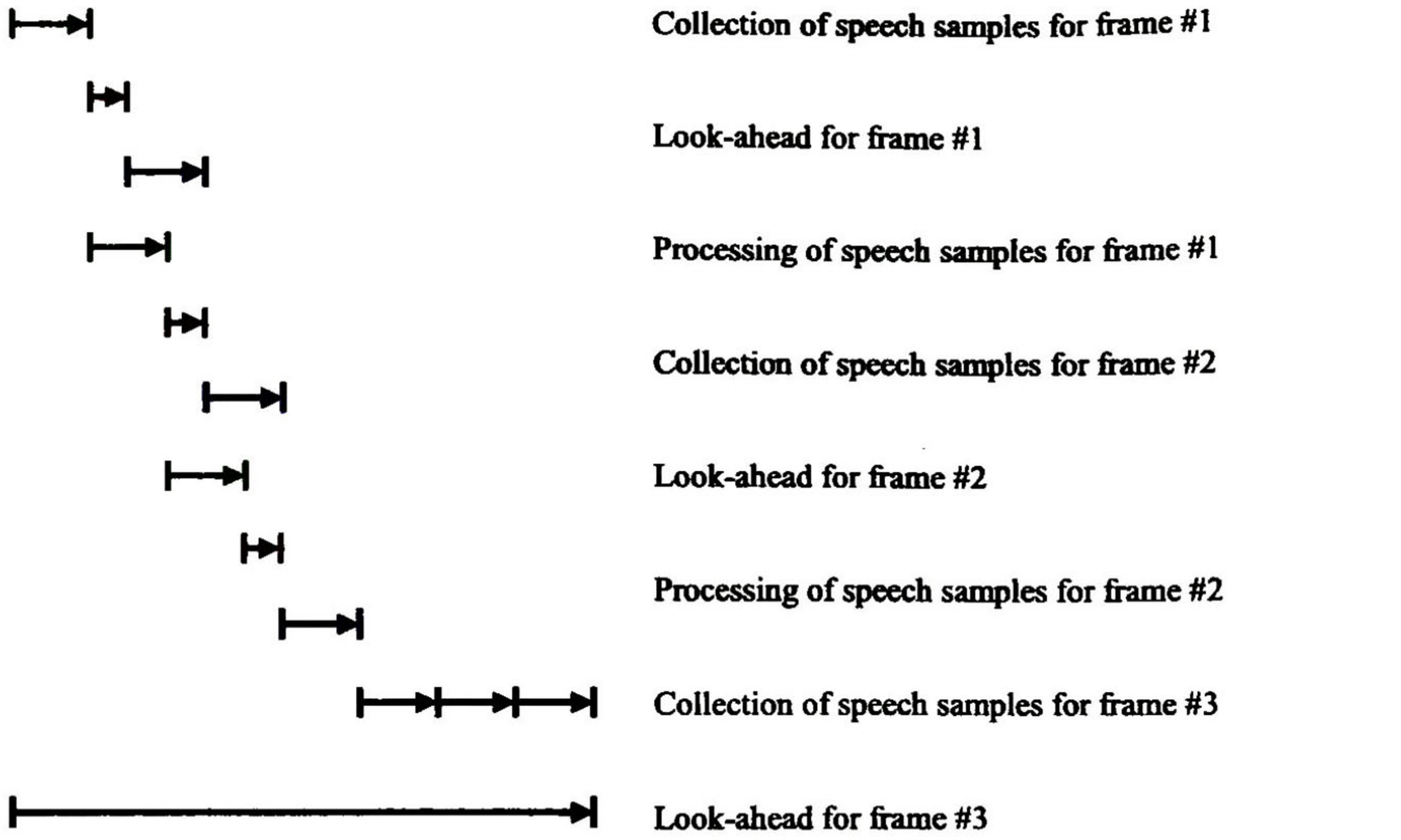


Figure 4: Example of encoder-related delay of three frames delivered to IP layer

Figure 5 represents the *destination terminal* diagram of a typical VoIP system [30]. The components and the positions of measurement points are indicated. The input bit stream, that contains the asynchronous voice packets, arrives at the terminal input reference point. Then it passes through the lower layers and then through the IP and RTP/UDP blocks. After that, the sequence of voice datagrams is passed to the de-jitter buffer, which performs very important tasks, in the sense of *QoS*, in order to reconstruct the digitized voice waveform. At this point, voice stream is already impaired because of the packet delay and loss due to the traversed network. Additional impairments occur by the additional delay of the remaining blocks and the additional loss caused by the discarding of highly delayed packets, compared to de-jitter buffer timeout, which is desirable optimized [38].

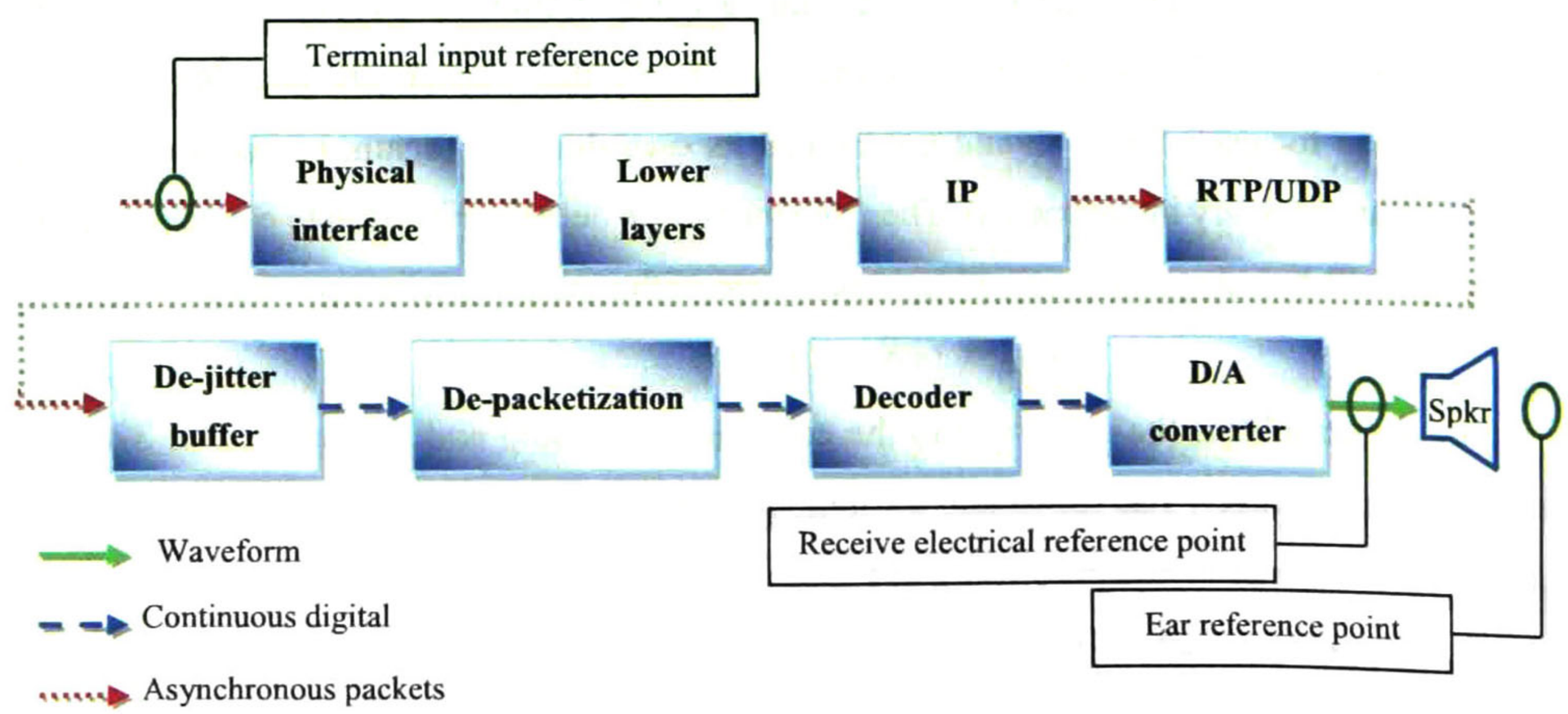


Figure 5: Destination terminal diagram and reference points

Figure 6 shows the process through which IP packet impairments and parameters (transfer delay, jitter and packet loss and errors) can be mapped to application layer performance in terms of overall loss and delay.

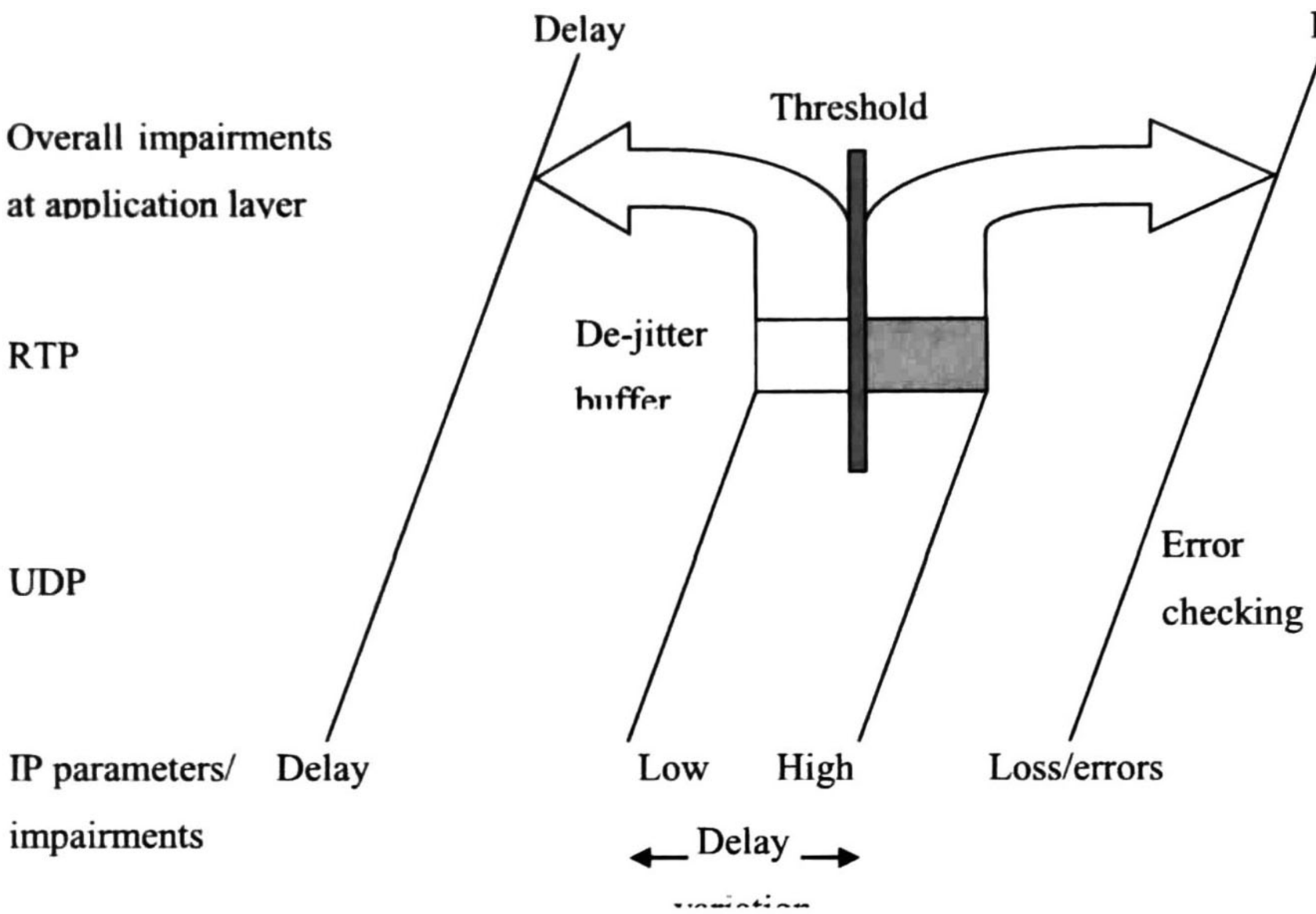


Figure 6: Mapping IP packet performance to application layer.

The input digital signal that carries the speech frames arrives at the receiver from the physical and lower layers. As this packet traverse to higher layers, the speech information is processed and impairments are removed, if possible. Note that some forms of impairments map into other, e.g., delay jitter is reduced at the expense of an increase in delay and highly delayed packets translate into packet loss. At the bottom of the figure, packets arrive with various impairments due to source terminal and network, or never arrive (lost). The arriving packets are processed and impairments are removed as possible as they move up to upper protocols. Some forms of impairment map into other [30].

Note that only those packets within the range between lowest allowed OWD and the threshold are delivered to application layer. This threshold consists of a trade-off between application level delay and loss.

Although it is not mandatory, it is strongly recommended that malformed packets due to bit errors that arrive at the receiver are discarded and not reproduced, because the resultant impairment due to voice waveform distortion may be worse to that produced by loss, especially for low bit rate codecs. These discarded packets are additional losses that, coupled with those due

to network congestion, must be evaluated by the application layer in order to apply a good estimation of the voice quality using a performance tool as the E-model's R factor. E.g., it could have more sense to estimate the QoS locally instead of globally, although the second estimation can always be obtained.

1.6.1 Packet One Way Delay and Delay Jitter Impairment for VoIP Applications

The one way delay (OWD) of the k^{th} packet is defined as the difference of its reception (R_k) and sending (S_k) times, i.e.,

$$O_k = R_k - S_k; k = 1, \dots, N \tag{70}$$

It is recommended that OWD should not exceed $400ms$, regardless of the type of application. This OWD is referred sometimes as *mouth-to-ear delay* (when transmitting speech), and the ITU-T recommendation G.107 [37] provides a tool to estimate the effect of this delay on the speech transmission quality (the mathematical expression is explained in more detail in Section 1.7).

The OWD can be represented as the sum of all delays that occur from the mouth reference point at the sender to the ear reference point at the receiver.

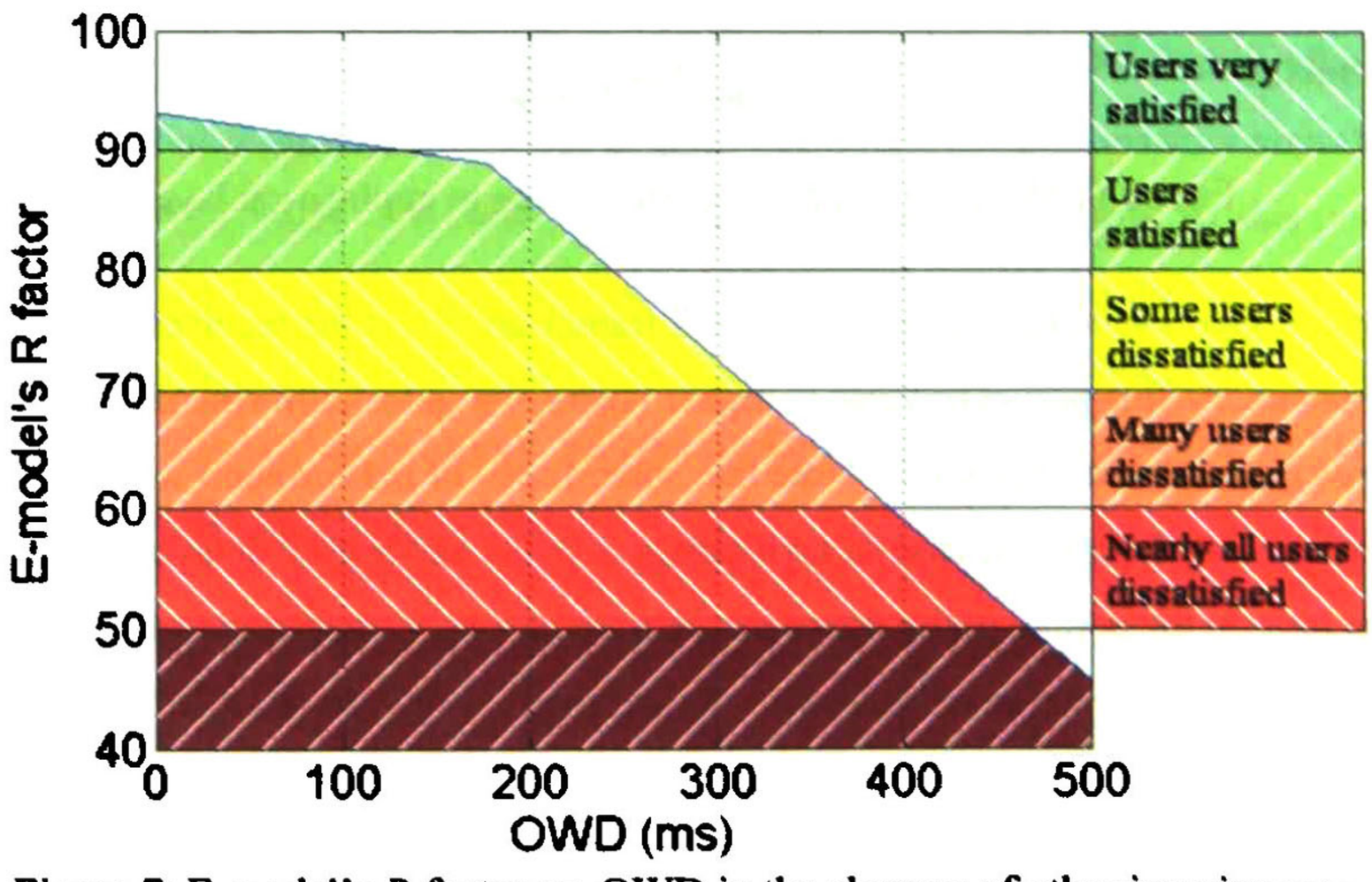


Figure 7: E-model's R factor vs. OWD in the absence of other impairments.

Figure 7 shows the effect of OWD on the E-model's R factor without considering other impairments (e.g., echo, noise and other distortions). To the right of this figure are shown the speech quality categories of ITU-T recommendation G.109 [39], which translate the values of R to users' acceptance levels. These are also explained in Table 2. Impairments caused by delays over $500ms$ are not fully validated in the recommendation G.109. It is recommended to avoid the transmission of speech through connections for which the estimated R factor is lower than 50.

It is also important to clarify that the R value is an estimation of the quality perception to be expected by the average user under the connection under consideration. Then, this assignment of range boundaries from best to poor quality cannot be exactly delimited. Rather, the quality should be viewed as a continuous, as shown in Figure 8.

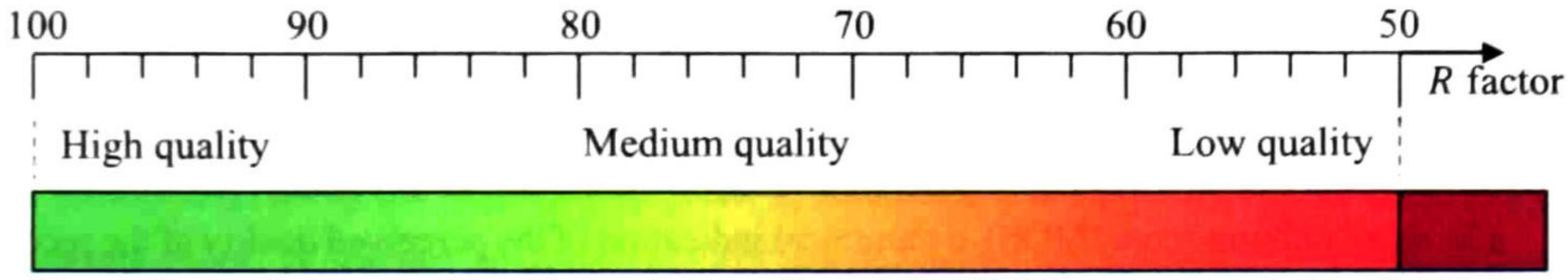


Figure 8: Quality perception of speech according to the ITU-T recommendation G.109.

The *delay jitter* of two packets is defined as the difference of their respective OWD, i.e.,

$$J_{k+1,k} = OWD_{k+1} - OWD_k; k = 1, \dots, N-1. \quad (71)$$

1.7 Quality Impairments and The E-Model's *R* Factor

The International Telecommunication Union (ITU) defined the E-model in the ITU-T Recommendation G.107 [23], as a tool for quality measurement for planning purposes. The E-model provides a prediction of the expected quality as perceived by the end user. This model is based on impairment factors, as expressed by (E-model's R factor):

$$R = R_0 - I_s - I_d - I_e + A \quad (72)$$

where:

- R_0 is the signal-to-noise ratio,
- I_s represents all impairments which occur more or less simultaneously with the voice signal,
- I_d sums all delay impairments due to delay and echo effects,
- I_e represents the impairments that are caused by low-bit rate codecs and A represents and advantage factor which certain systems provide in comparison to conventional systems.

A simplified version of (72), that represents the impairment as a function of the packet delay and PLR, is:

$$R = 93.2 - I_d - I_e \quad (73)$$

where I_d, the delay impairment, is defined as [40]:

$$I_d = 0.024d + 0.11(d - 177.3)H(d) \quad (74)$$

where d is the mouth-to-ear delay in ms and

$$H(d) = \begin{cases} 0; & d < 177.3 \\ 1; & d \geq 177.3 \end{cases} \quad (75)$$

The quantity I_e is the impairment caused by low bit rate codecs [23], and its general expression is:

$$I_e(r) = \gamma_1 + \gamma_2 \ln(1 + \gamma_3 r) \tag{76}$$

where r is the packet loss probability and the values of γ_1, γ_2 and γ_3 are constants that depend on the type of codec used [41], e.g.,

$$I_{e,G.711}(r) = 0 + 30 \ln(1 + 15r) \tag{77}$$

and

$$I_{e,G.729}(r) = 11 + 40 \ln(1 + 10r) \tag{78}$$

The mean opinion score (MOS), a numerical indication of the perceived quality of the received media after compression and/or transmission, can be estimated from the R factor as: $MOS = 1$ for $R < 0$, $MOS = 1 + 0.035R + R(R - 60)(100 - R)7 \cdot 10^{-6}$ for $0 \leq R < 100$ and $MOS = 4.5$ for $R > 100$ [23]. In order to achieve the best perceived quality as possible, the R factor, and consequently the MOS, must be maximized.

Table 2: Definition of speech transmission quality categories.

R-value range	MOS	Speech transmission quality category	User satisfaction
$90 \leq R < 100$	$4.33 \leq MOS < 4.5$	Best	Very satisfied
$80 \leq R < 90$	$4.02 \leq MOS < 4.33$	High	Satisfied
$70 \leq R < 80$	$3.59 \leq MOS < 4.02$	Medium	Some users satisfied
$60 \leq R < 70$	$3.10 \leq MOS < 3.59$	Low	Many users satisfied
$50 \leq R < 60$	$2.57 \leq MOS < 3.10$	Poor	Nearly all users satisfied
Note: Transmissions with $R < 50$ are not recommended.			

In order to do this, the impairments factors I_d and I_e must be adjusted. I_d depends on many parameters, many of them cannot be controlled by the end user directly (e.g., constant and variable network delays) but others can, e.g. the size (in ms) of the de-jitter buffer timeout. In its turn, I_e depends on the codec used and the PLR. It can be adjusted by selecting the type of codec used and by applying an error correction technique, e.g. N-packet FEC described in section 4.1.

There are two main sources of packet losses: network congestion (from A to B) and de-jitter buffer (from C to D). The PLR due to network congestion is reduced by the N-packet FEC. The network also adds variable delays, i.e., delay jitters, that are eliminated by the de-jitter buffer, but at the expense of an additional delay for all (or most) packets and, possibly, a slight increase in packet loss. This occurs since long delayed packets, although they successfully arrive to the receiver, are discarded and, consequently, lost from the point of view of the listener. Furthermore, bit-level errors that may be present in received but corrupted packets are an important source of errors, especially for wireless communications.

In order to quantify the voice quality by means of the E-model's *R* factor, the performance of the *N*-packet FEC block and the de-jitter buffer must be described analytically, as presented in Chapter 0.

1.8 Chapter Summary

This chapter presents mathematical definitions and other concepts that are used throughout this work. Particularly, about the following subjects: 1) stochastic self-similarity for the continuous and discrete cases, 2) wavelet analysis 3) ICFD transformation, 4) discrete Markov processes (including 2-state and 4-state formulae), and 5) network communications, particularly about VoIP systems.

Two definitions are given for the discrete self-similarity: the exact self-similarity and the second order self-similarity. The second definition is more likely to be satisfied by real world time series.

The degree of self-similarity and, consequently, some statistical properties, are completely determined by the Hurst index (*H*). This index of self-similarity, although it is mathematically well defined, is difficult to estimate from real-world traces. A set of the most commonly used Hurst index estimators is described in this section.

Wavelet analysis and the *LD-Diagram*, which are important for data series analysis, area also described, as well as their relationship with the Hurst index.

Respective mathematical descriptions of the discrete two-state and four-state Markov processes are presented. These models are applied to network packet losses. The probabilities of packet reception and loss, as well as the distribution of gaps and bursts, are calculated in terms of the probabilities of transition between the states for each model.

Finally, the VoIP system components and the E-model's R-factor are also described.

2 Analysis and Synthesis of Second-order Self-similar Time Series

2.1 A Wavelet-Based Decomposition of Time-Series

Any time series X_t can be decomposed into a set of time series, each one defined as expressed by (79).

$$C_{X,t}^{n,l} = X_t^{(n^{l-1}E)} - X_t^{(n^lE)}; n, i \in \mathbb{N} \tag{79}$$

where $X_t^{(n^lE)}$ is the time series X_t after two operations, which are:

- Aggregation at level n^l, as defined by (5) and (6). i.e. $m = n^l$.
- Expansion of level n^l, which consists in repeating each element of a time series n^l times.

  i.e. $X_j^{(n^lE)} = X_k^{(n^l)}$ for $k = 1 + \left\lfloor \frac{j-1}{n^l} \right\rfloor$ and $j \in \mathbb{N}$.

These zero mean components $C_{X,t}^{n,l}$ have three important properties:

i. They synthesize the original time series without loss, i.e.

$$X_t = \sum_l C_{X,t}^{n,l} \tag{80}$$

ii. They are pair wise orthogonal:

$$\langle C_{X,t}^{n,l}, C_{X,t}^{n,j} \rangle = 0; i \neq j \tag{81}$$

iii. If X_t is exactly self-similar or at least satisfies (9) for $q = 2$, then the variance of the components satisfy:

$$var\left(C_{X,t}^{n,l}\right) = r \cdot var(C_{X,t}^{n,l-1}) \tag{82}$$

where

$$r = n^{2H-2} \tag{83}$$

Properties i, ii and iii imply (84) and (85):

$$\sigma_X^2 = \sum_l var(C_{X,t}^{n,l}) \tag{84}$$

$$\sigma_X^2 = \frac{1}{1 - n^{2H-2}} \cdot var(C_{X,t}^{n,1}) \tag{85}$$

Then, the variance of the i^{th} component is related to the variance of X_t as (86) expresses:

$$var\left(C_{X,t}^{n,l}\right) = (1-r)\cdot r^{l-1}\cdot \sigma_X^2 \tag{86}$$

where

$$r = n^{2H-2} \tag{87}$$

Another useful property relates the variance of the component to the variances of the aggregated series, i.e.,

$$var\left(C_{X,t}^{n,j}\right) = var\left[X_t^{(n^{j-1}E)}\right] - var\left[X_t^{(n^{j}E)}\right] \tag{88}$$

It is easy to proof (88): from (79) it turns that $var\left[C_{X,t}^{n,j}\right] = var\left[X_t^{(n^{j-1}E)}\right] + var\left[X_t^{(n^{j}E)}\right] - 2cov\left[X_t^{(n^{j-1}E)}, X_t^{(n^{j}E)}\right]$, but as $X_t^{(n^{j}E)}$ is itself an aggregation of $X_t^{(n^{j-1}E)}$ it turns that $cov\left[X_t^{(n^{j-1}E)}, X_t^{(n^{j}E)}\right] = var\left[X_t^{(n^{j}E)}\right]$, and (88) comes after a substitution.

Properties i, ii, and iii imply:

$$\sigma_X^2 = \sum_i var(C_{X,t}^{n,j}) \tag{89}$$

$$\sigma_X^2 = \frac{1}{1-r} var(C_{X,t}^{n,1}) \tag{90}$$

Then, the variance of the j^{th} component is related to the variance of X_t as:

$$var\left(C_{X,t}^{n,j}\right) = (1-r)r^{j-1}\sigma_X^2 \tag{91}$$

It is easy to proof the following relation:

$$var\left(C_{X,t}^{n,j}\right) = n^{-j}S_2(j) \tag{92}$$

An immediate consequence of (92) is that the plot $j + \log_n[var(C_{X,t}^{n,j})]$ vs. j is equivalent to the *LD-Diagram*, it is a straight line for H-SOSS and its slope is related to the Hurst index as $= 2H-1$, i.e.:

$$var\left(C_{X,t}^{n,i}\right) = \frac{E|d_X(j,\cdot)|^2}{2^j} \tag{93}$$

Also, for $n = 2$ it coincides with the Haar wavelet case ([42] [43]), that is:

$$\psi_{j,k}(t) = \psi_0(2^{-j}t - k) \tag{94}$$

and

$$\psi_0(t) = \begin{cases} +1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & otherwise \end{cases} \tag{95}$$

For a finite-length time series with L octaves, the number of octave (j) of the *LD-Diagram* is related to index i of (79) according to (96):

$$j = i \tag{96}$$

To estimate the *LD-Diagram*, it is necessary to subtract the bias as expressed by (19), resulting: where n_j is the number of coefficients available at octave, i.e.:

$$E\{\log_2[var(C_{X,t}^{n,i})]\} \sim E\{\log_2[var(\hat{C}_{X,t}^{n,i})]\} - g_j \tag{97}$$

Figure 1 shows the components obtained from an H-SOSS sample of size 32 and $H = 0.9$. The squared form of the components is due to the expansion of the aggregated series and disappears after the downsampling.

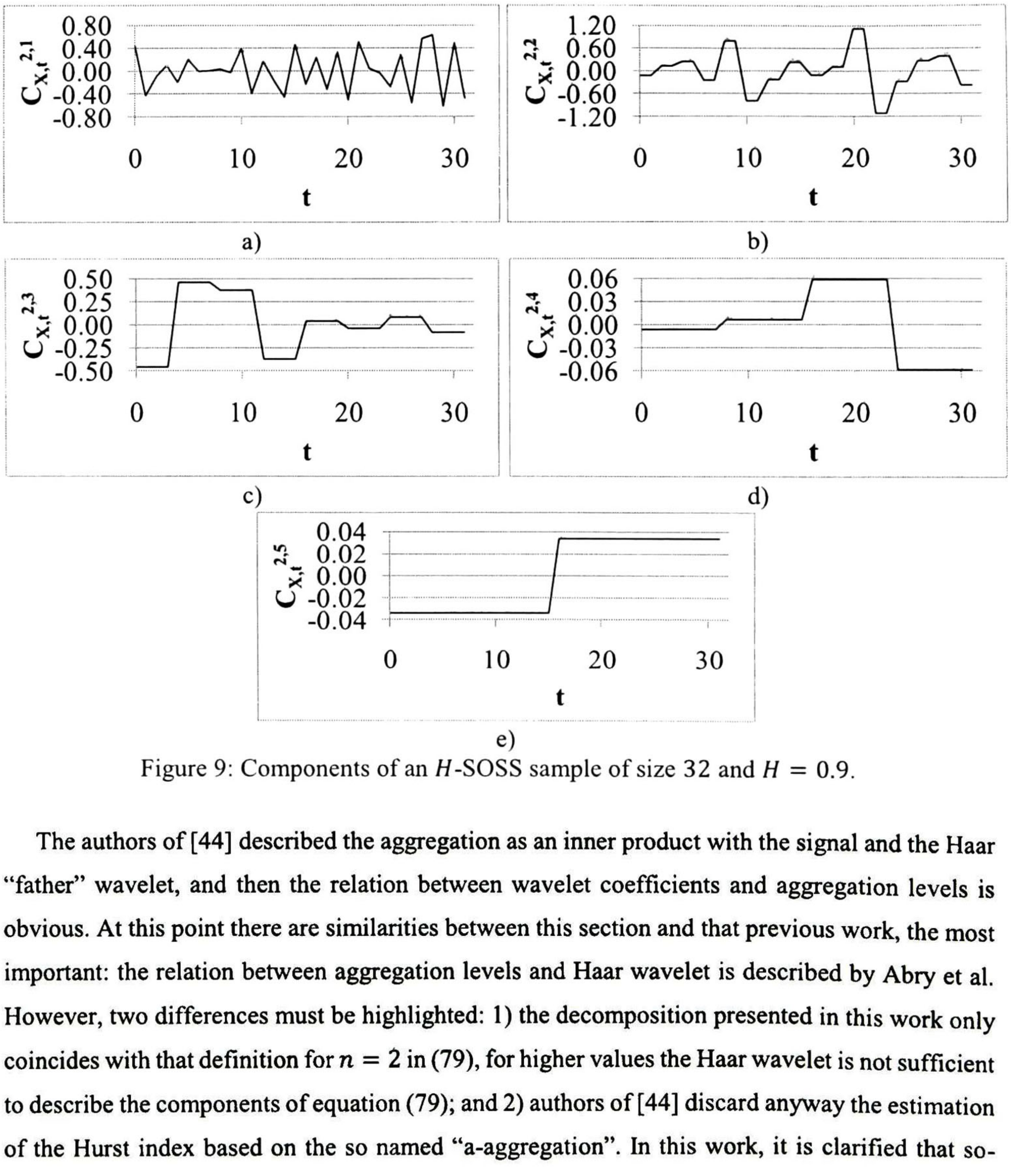


Figure 9: Components of an H-SOSS sample of size 32 and $H = 0.9$.

The authors of [44] described the aggregation as an inner product with the signal and the Haar "father" wavelet, and then the relation between wavelet coefficients and aggregation levels is obvious. At this point there are similarities between this section and that previous work, the most important: the relation between aggregation levels and Haar wavelet is described by Abry et al. However, two differences must be highlighted: 1) the decomposition presented in this work only coincides with that definition for $n = 2$ in (79), for higher values the Haar wavelet is not sufficient to describe the components of equation (79); and 2) authors of [44] discard anyway the estimation of the Hurst index based on the so named "a-aggregation". In this work, it is clarified that so-

called "a-aggregation" is misunderstood leading to incorrect implementations (i.e., the "classical" variance estimator) and a generalization of the "d-aggregation" is proposed.

2.2 General Waveforms of the Basis Functions

The orthogonal decomposition defined by (79) can be expressed in terms of an inner product between the signal and a set of orthogonal wavelet-type functions. Let us describe the waveform for the general case.

The wavelet-based function is:

$$\psi_n(t) = \begin{cases} 1 - \frac{1}{n} & 0 \le t < 1/n \\ -\frac{1}{n} & \frac{1}{n} \le t < 1 \\ 0 & otherwise \end{cases} \tag{98}$$

and the wavelet functions derived from $\psi_n(t)$ are obtained by three operations: scaling, displacement (similarly, but not exactly equal, to equation (15)) and a phase shift, i.e.:

$$\begin{aligned} \psi_{n,j,k,\theta}(t) = \{1 - u[n^j(k-1)]\}\psi_{n,j,k}(t - \theta n^{j-1}) \\ + u[n^j(k-1)]\psi_{n,j,k}(t + n^j - \theta n^{j-1}) \end{aligned} \tag{99}$$

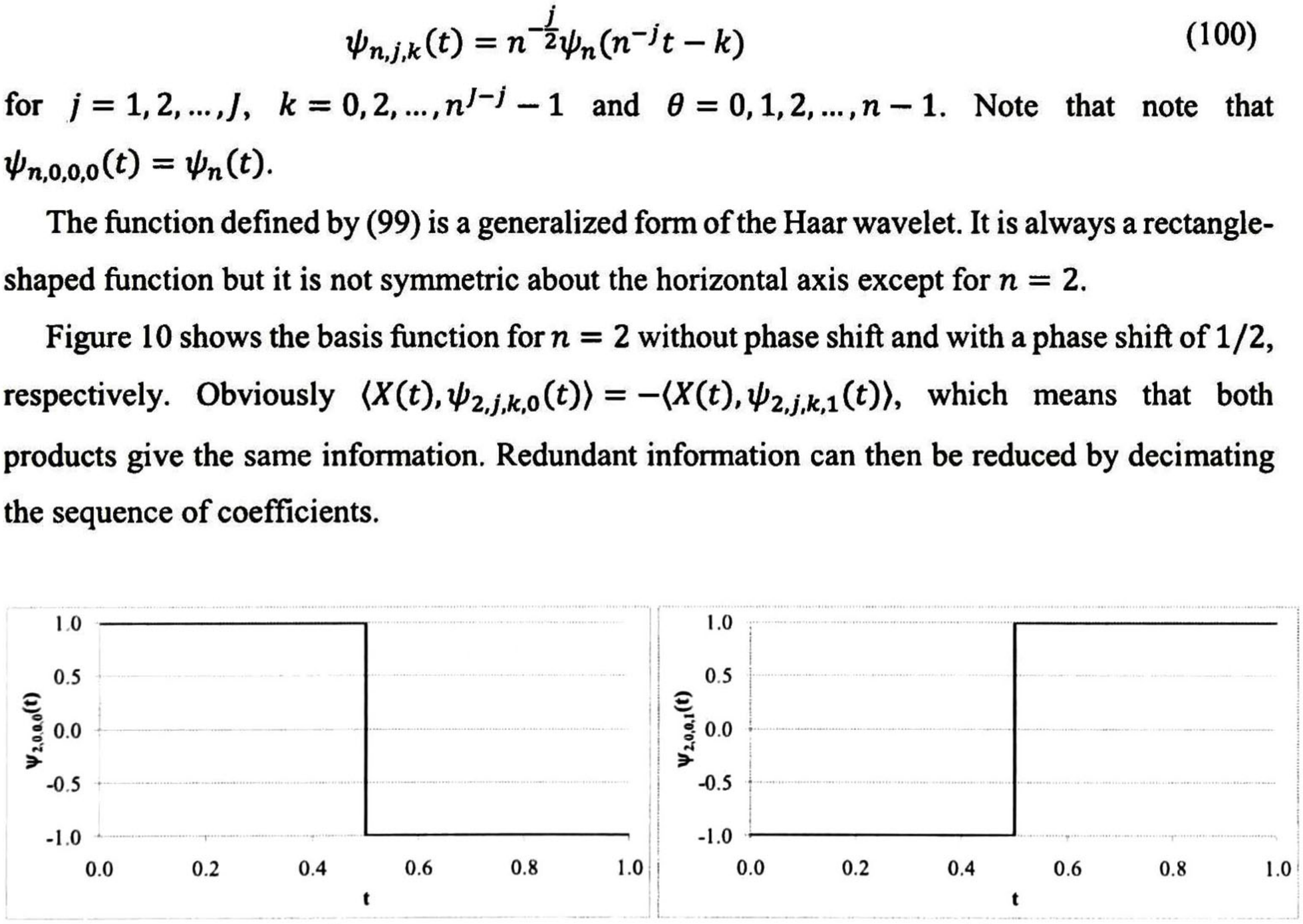
and

$$\psi_{n,j,k}(t) = n^{-\frac{j}{2}}\psi_n(n^{-j}t - k) \tag{100}$$

for $j = 1, 2, \ldots, J$, $k = 0, 2, \ldots, n^{J-j} - 1$ and $\theta = 0, 1, 2, \ldots, n-1$. Note that note that $\psi_{n,0,0,0}(t) = \psi_n(t)$.

The function defined by (99) is a generalized form of the Haar wavelet. It is always a rectangle-shaped function but it is not symmetric about the horizontal axis except for $n = 2$.

Figure 10 shows the basis function for $n = 2$ without phase shift and with a phase shift of $1/2$, respectively. Obviously $\langle X(t), \psi_{2,j,k,0}(t)\rangle = -\langle X(t), \psi_{2,j,k,1}(t)\rangle$, which means that both products give the same information. Redundant information can then be reduced by decimating the sequence of coefficients.



a) b)

Figure 10: Basis functions: a) $\psi_{2,0,0,0}(t)$ (no phase shift) and b) $\psi_{2,0,0,1}(t)$ (equal to $\psi_{2,0,0,0}(t)$ with a phase shift of $1/2$).

Figure 11 shows the basis function for $n = 3$. Note that the phase shift moves the rectangle of height $2/3$ from one third to another. In this case there exists also redundant information, as $\psi_{3,0,0,2}(t) = -\psi_{3,0,0,0}(t) - \psi_{3,0,0,1}(t)$ and in the general case $\psi_{n,j,k,\theta}(t) = -\sum_{i\neq\theta}\psi_{n,j,k,i}(t)$, which means that one of the coefficients, e.g., the one obtained with the last phase shift, can be discarded. This means that the sequence of coefficients can be downsampled without loss of information. For a sample of length N, it is easy to verify that the number of observations that remains in all components (sequences of coefficients) after the downsampling is $N - 1$, which can be complemented with the sample mean, in the case that this is not zero.

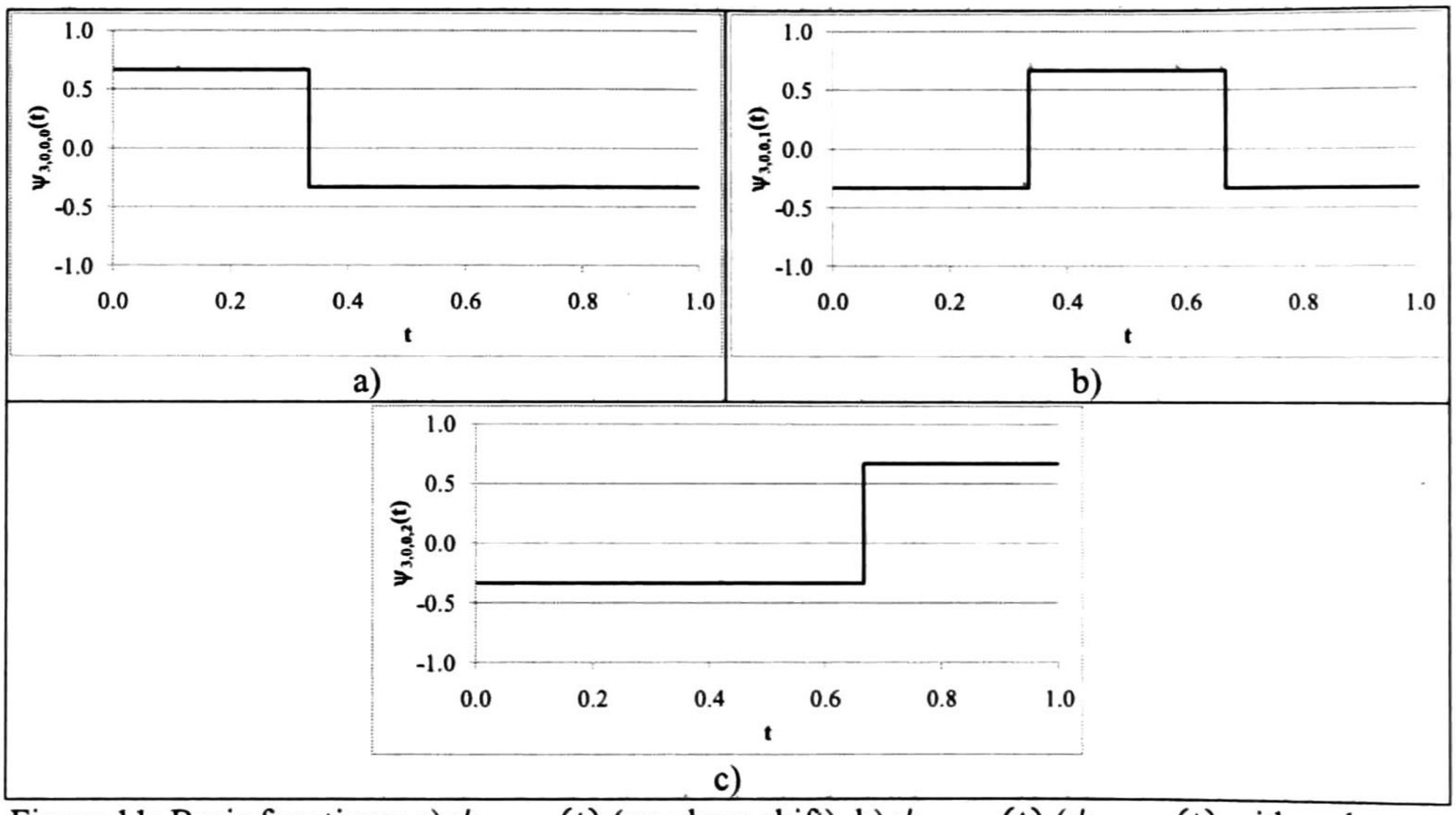


Figure 11: Basis functions: a) $\psi_{3,0,0,0}(t)$ (no phase shift), b) $\psi_{3,0,0,1}(t)$ ($\psi_{3,0,0,0}(t)$ with a phase shift of $1/3$) and c) $\psi_{3,0,0,2}(t)$ ($\psi_{3,0,0,0}(t)$ with a phase shift of $2/3$).

2.3 Wavelet-based Synthesis of *H*-SOSS Time Series

A method to synthesize H_1-SOSS from practically any time series, regardless of whether it is or not self-similar or its marginal distribution, is proposed. This method consists of adjust the sum expressed by (14) with a set of weights, i.e.,

$$X_{H_1}(t) = \sum_{j=i}^{J} w_j \sum_{k=1}^{2^j} d_X(j,k)\psi_{j,k}(t) \tag{101}$$

where the weights w_j are defined as:

$$w_j = \sqrt{\frac{\widehat{c_f C} \cdot 2^{J(2H_1-1)}}{\hat{S}_2(j)}} \tag{102}$$

and where $\hat{S}_2(j)$ and $\widehat{c_f C}$ are the respective estimations of $S_2(j)$ and $c_f C$ (the associated power parameter [1]) from X_t and H_1 is the desired Hurst index of the new synthetic series. It is necessary that $\hat{S}_2(j) > 0 \forall j = 1, \ldots, J$.

The weighted sum (101) can also be expressed in terms of the orthogonal components described in Section 3 and defined by (11) as follows:

$$X_{H_1}(t) = \sum_{j=1}^{J} w_j \hat{C}_{X,t}^{n,j} \tag{103}$$

where the weights w_i are estimated as:

$$w_j = \sqrt{r_1^{j-1} \cdot \frac{1 - r_1^J}{1 - r_1} \cdot \frac{\hat{\sigma}_X^2}{var\{\hat{C}_{X,t}^{n,j}\}}} \tag{104}$$

where r_1, similarly to (83), is defined as:

$$r_1 = n^{2H_1 - 2} \tag{105}$$

To proof derive (104), consider that a finite length time series $X_{H_1}(t)$ whose Hurst index estimation is already H_1 and r_1 is defined as in (105). Then, knowing that 1) the variance of $X_{H_1}(t)$ is the sum of the respective variances of its components, as expressed by (84), 2) that there are a finite number, say L, of non-zero components, and 3) that those variances are related as (82), it yields:

$$\hat{\sigma}_{X_{H_1}}^2 = var\left\{C_{X_{H_1},t}^{n,1}\right\} \cdot \sum_{j=0}^{J-1} r_1^j = var\left\{\hat{C}_{X_{H_1},t}^{n,1}\right\} \cdot \frac{1 - r_1^J}{1 - r_1} \tag{106}$$

The variance of the j^{th} component ($\hat{C}_{X,t}^{n,j}$) is related to that of the first one ($\hat{C}_{X,t}^{n,1}$) by

$$var\left\{\hat{C}_{X_{H_1},t}^{n,j}\right\} = r_1^{j-1} var\left\{\hat{C}_{X_{H_1},t}^{n,1}\right\} \tag{107}$$

then

$$var\left\{\hat{C}_{X_{H_1},t}^{n,j}\right\} = \frac{\left(r_1^{j-1}\right) \cdot (1 - r_1)}{1 - r_1^L} \cdot \hat{\sigma}_{X_{H_1}}^2 \tag{108}$$

Finally, it is straightforward that the weight w_j for $i = 1, \ldots, J$ is the square root of the quotient of the desired component variance over the estimated one. E.g.,

$$w_j = \sqrt{\frac{var\left\{\hat{C}_{X_{H_1},t}^{n,j}\right\}}{var\{\hat{C}_{X,t}^{n,j}\}}}$$

yielding (104) after substituting $var\left\{\hat{C}^{n,j}_{X_{H_1},t}\right\}$ as in (108). Using this synthesis method, an H-SOSS self-similar series with equal variance, mean (i.e., zero), and length than the original is obtained, i.e.,

$$\hat{\sigma}^2_{X_{H_1}} = \hat{\sigma}^2_X \tag{109}$$

Note that the only restriction for (103) is that the estimated variance of $\hat{C}^{n,j}_{X,t}$ must be non-zero, similarly to $\hat{S}_2(j)$ for (101). An additional warning is that even though this synthesis does not depend either on the marginal distribution or the correlation structure of the input signal, it is preferable that this is already self-similar (e.g., FGN) and that its Hurst index is close to that of $X_{H_1}(t)$, i.e., H_1 is close to $\hat{H}$.

Pathological behavior can be produced in the output series for some critical conditions, e.g., noticeable steps, which may produce a non-stationary signal, result of the transforming a SRD or uncorrelated input signal to a LRD output with H close to 1. Impulses of very large magnitude (outliers) can be also be produced when, for some j, $var\left\{\hat{C}^{n,j}_{X_{H_1},t}\right\}$ or, equivalently, $\hat{S}_2(j)$, is close to zero.

A similar methodology was developed by Deléchelle *et al* [45], to synthesize fractional Gaussian noise by performing a weighted sum of the intrinsic mode functions (IMF) of a white noise process (see also [46]). The advantages of the proposed synthesis compared to that of Deléchelle *et al* are that the components defined by (79) are exactly orthogonal, the relation between the weights for the reconstruction sum are mathematically well defined and the Hurst index of the synthesized time series matches perfectly the wavelet estimator proposed by D. Veitch and P. Abry [11], i.e., the estimated H of the synthesized series is unbiased ($E(\hat{H}) = H$) and has zero variance ($var(\hat{H}) = 0$). The disadvantages are that the components are sequences of squared signals (because of the expansion described in Section 3) and not sinusoids, so noticeable steps arise when synthesizing a time series with high Hurst index, e.g., close to 1, from an input that is SRD or weakly correlated. A solution for this problem is to apply interpolation (as in EMD) instead of expanding the series in order to produce softer components (sinusoids or polynomial) instead of square-type, with the consequence that the Hurst index is no longer exact, but approximated.

2.4 Auto-covariance function of Discrete *H*-SOSS Processes with Finite Variance

The normalized ACV of a discrete *H*-SOSS process is defined by (13). To use properly this definition, it is necessary that the process under study has finite variance. Let us calculate the sum of the first terms of the normalized ACV, i.e.:

$$P_{X,1}(k) = \sum_{i=0}^{k} \rho_X(i) = \rho_X(0) + \rho_X(1) + \cdots + \rho_X(k) \tag{110}$$

It is easy to reduce the sum to:

$$P_{X,1}(k) = \frac{1 + (k+1)^{2H} - k^{2H}}{2} \tag{111}$$

And, the limit $k \to \infty$ is:

$$\lim_{k\to\infty} P_{X,1}(k) = \begin{cases} 0.5; & 0 < H < 0.5 \\ 1 & 0.5 \\ \infty & 0.5 < H < 1 \end{cases} \tag{112}$$

Note that the difference $(k+1)^{2H} - k^{2H}$ tends to ∞ as $k \to \infty$. This is the reason why the sum of the ACV function of *H*-SOSS processes with LRD diverges.

Expression (113) shown the ACV matrix of an *H*-SOSS process calculated up to certain lag *k*:

$$\underline{P} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(k-1) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(k-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k) & \rho(k-1) & \rho(k-2) & \cdots & 1 \end{bmatrix} \tag{113}$$

The sum of the upper part of the ACV matrix (including the diagonal) can be obtained as:

$$U(K) = \sum_{i=0}^{k} P_{X,1}(i) = P_{X,1}(0) + P_{X,1}(1) + \cdots + P_{X,1}(k) \tag{114}$$

which can be reduced as:

$$U_X(k) = \frac{(k+1)^{2H} + k + 1}{2} \tag{115}$$

And the sum of all elements of the matrix is twice the upper parte minus once the sum of the elements of the main diagonal (which has $k+1$ ones):

$$P_{X,2}(k) = 2U_X(k) - (k+1) \tag{116}$$

$$P_{X,2}(k) = (k+1)^{2H} \tag{117}$$

From (117), it can be seen that $P_{X,2}(k)$ diverges for all values of H in the range between 0 and 1 and that, the higher the Hurst index, the faster it diverges, e.g., it is a line with slope 1 for $H = 0.5$. Figure 12 shows the divergence of $P_{X,2}(k)$ for many values of H.

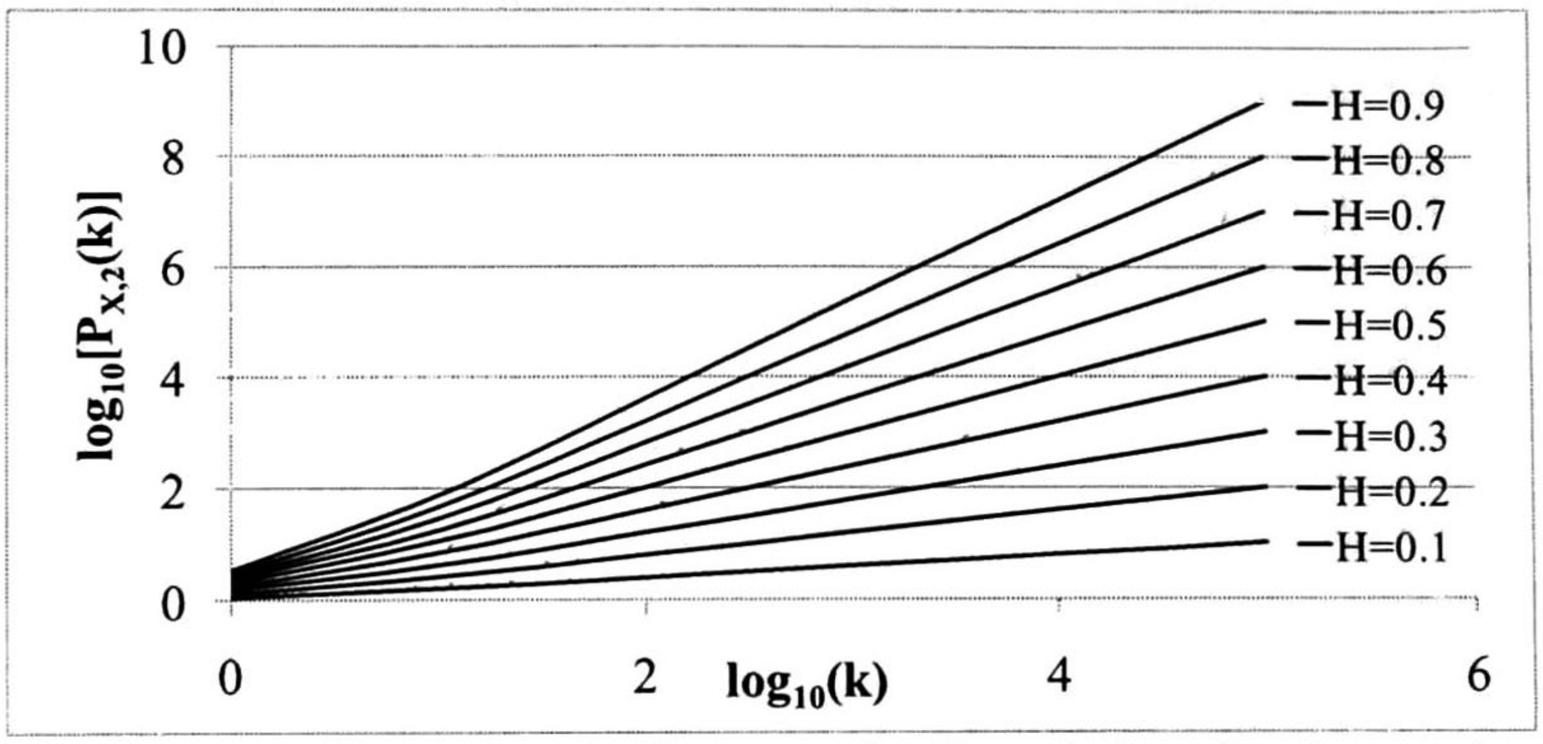


Figure 12: Log-log plot of $P_{X,2}(k)$ vs. k. The respective Hurst index of each curve is at its right. All curves diverge to $+\infty$ as $k \to \infty$.

2.5 Variance of the Aggregated Series

Using (87), (82) can be written as:

$$var\left(C_{X,t}^{n,i}\right) = r \cdot var(C_{X,t}^{n,i-1}) \tag{118}$$

where $r = n^{2H-2}$

Let $\hat{X}_t = \{\hat{X}_t; t = 1, \ldots, N\}$ be a finite-length self-similar time series such that $N = n^L$ ($L < \infty$) and $n \geq 2$ (i.e. N is a power of n) then, a set of non-zero L components ($\hat{C}_{X,t}^{n,i}; i = 1, \ldots, L$) can be obtained as expressed by the analysis equation (79). As the components are pair wise orthogonal, the variance of $\hat{X}_t$ ($\hat{\sigma}_X^2$) is the sum of a finite number of variances:

$$\hat{\sigma}_X^2 = \sum_{i=1}^{L} var(\hat{C}_{X,t}^{n,i}) \tag{119}$$

Then, the variance of the i^{th} component ($\hat{C}_{X,t}^{n,i}$), which has also finite length, is:

$$var\left(\hat{C}_{X,t}^{n,i}\right) = \frac{1-r}{1-r^L} \cdot r^{i-1} \cdot \hat{\sigma}_X^2 \tag{120}$$

and the variance of the aggregated series is:

$$var\left(\hat{X}_k^{(n^i)}\right) = \sum_{j=i+1}^{L} var\left(\hat{C}_{X,t}^{n,j}\right) \tag{121}$$

or, equivalently:

$$var\left(\hat{X}_k^{(n^i)}\right) = \frac{r^i - r^L}{1-r^L} \cdot var(\hat{X}_t) \tag{122}$$

Making $m = n^i$ in (11), it yields:

$$var\left(X_k^{(n^i)}\right) = r^i \cdot var(X_t) \tag{123}$$

Considering consider $var(\hat{X}_t) = \sigma_X^2$ (i.e. the sample variance is equal to the variance of the process), then the difference between (122) and (123) is:

$$\Delta_{n,i} = var\left(\hat{X}_k^{(n^i)}\right) - var\left(X_k^{(n^i)}\right) \tag{124}$$

$$\Delta_{n,i} = -r^L \cdot \frac{1-r^i}{1-r^L} \cdot \sigma_X^2 \tag{125}$$

This means that the definition of second order self-similar time series cannot be complied by finite-length time series (such as those obtained from measurements) i.e. the variance of the aggregated series, expressed by (122) is always smaller than the expected (theoretical) for $i > 1$. For a fixed variance (σ_X^2), the absolute error ($|\Delta_{n,i}|$) is greater as L is smaller (shorter traces), i is greater (larger aggregation levels) and r is greater (larger Hurst parameter).

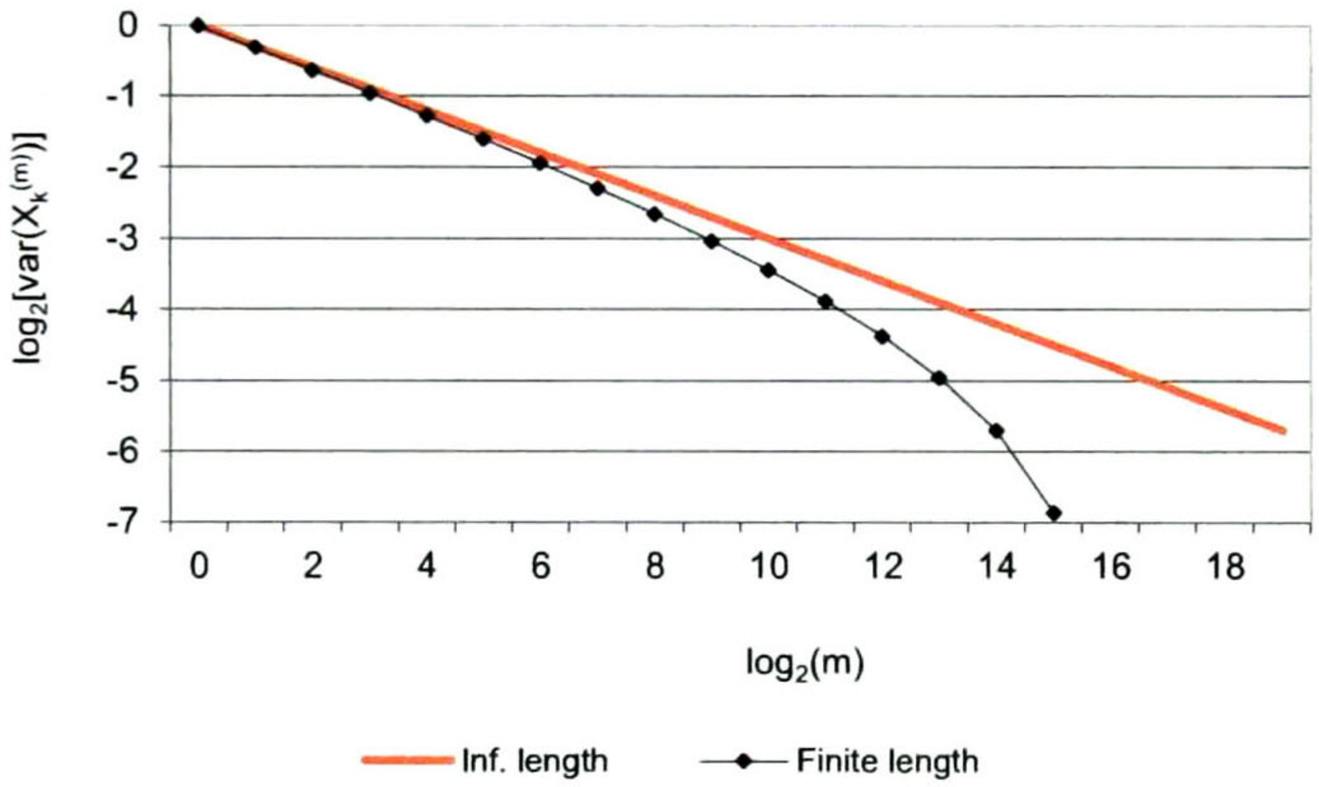


Figure 13. Variance of the aggregated series ($m = 2^i$) of two time series with $H = 0.85$

2.6 Variance of the Wavelet Components

It is easy to proof that the variance of the components $C_{X,t}^{n,i}$ (defined by (79)) can be calculated as follows (126):

$$var\left(C_{X,t}^{n,i}\right) = var\left(X_k^{(2^{i-1})}\right) - var\left(X_k^{(2^i)}\right) \tag{126}$$

But the variance of the components for finite-length time series also presents a theoretical error:

$$var\left(\hat{C}_{X,t}^{n,i}\right) = var\left(\hat{X}_k^{(2^{i-1})}\right) - var\left(\hat{X}_k^{(2^i)}\right) \tag{127}$$

$$var\left(\hat{C}_{X,t}^{n,i}\right) = var\left(C_{X,t}^{n,i}\right) + \left(\Delta_{n,i} - \Delta_{n,i-1}\right) \tag{128}$$

where

$$\Delta_{n,i} - \Delta_{n,i-1} = \frac{r^{L+i-1}(1-r)\sigma_X^2}{1-r^L} > 0 \tag{129}$$

The positive quantity expressed by (129) shows that the calculation of respective variance of the components $\hat{C}_{X,t}^{n,i}$ overestimates the real variance. Calculating then the quotient between $var\left(\hat{C}_{X,t}^{n,i}\right)$ and $var\left(\hat{C}_{X,t}^{n,i-1}\right)$ from (128) and (129), it yields:

$$\frac{var\left(\hat{C}_{X,t}^{n,i}\right)}{var\left(\hat{C}_{X,t}^{n,i-1}\right)} = \frac{var\left(C_{X,t}^{n,i}\right) + r^{i-1}}{var\left(C_{X,t}^{n,i-1}\right) + r^{i-2}} = r \tag{130}$$

The quotient (130) shows that, although the respective variance of the components obtained from a finite-length time series present a theoretical (positive) error, it does not affect the *LD-Diagram*, i.e. the ordinate at the origin of the linear model derived from the *LD-Diagram* is a little greater than the expected (theoretical) value but its slope is identical. It is also shown by (130) that both the theoretical variance of the components and their respective error form a geometrical series of rate $r = n^{2H-2}$

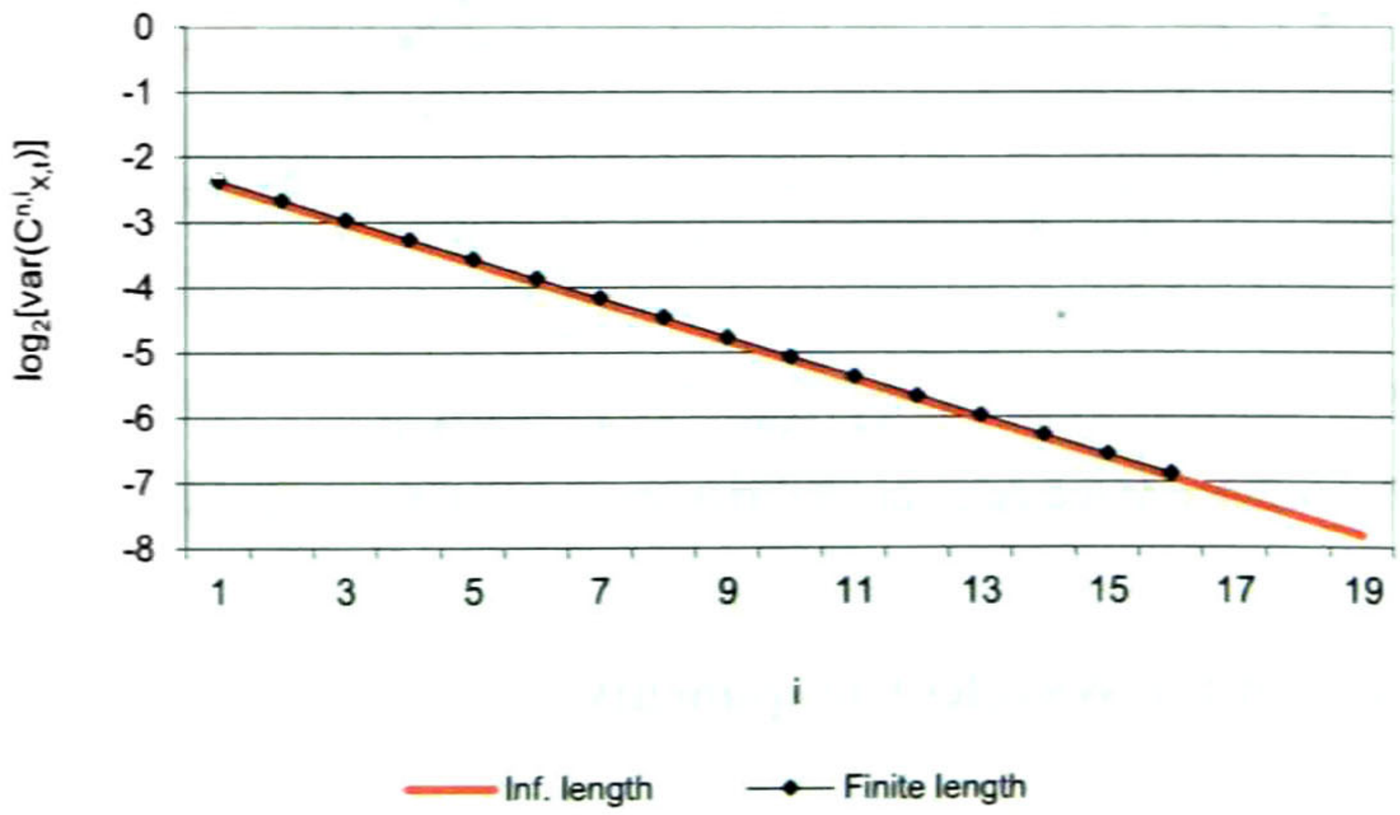


Figure 14. Respective *LD-Diagram* for two time series with $H = 0.85$

The variance of the aggregated series of a real world time series of finite length ($var(\tilde{X}_k^{(m)})$), assuming that it represents a self-similar process, can be different of the theoretical ($var(\hat{X}_k^{(m)})$).

For a set of real world time series, it can be assumed that $E\left[var\left(\tilde{X}_k^{(m)}\right)\right] = var(\hat{X}_k^{(m)})$, but this also implies that $E\{\log_2\left[var(\tilde{X}_k^{(m)})\right]\} \leq \log_2[var(\hat{X}_k^{(m)})]$.

2.7 Chapter Summary

In this chapter, a wavelet-based orthogonal decomposition of time series is presented. The analysis functions are scaled, displaced, and phase-shifted versions of the basis function. Note that wavelet functions are generally scaled and displaced only. The basis function is defined by equation (98), and the rest of the analysis functions are defined by (99) and (100).

After analyzing, i.e., decomposed, the original time series, the orthogonal components can be used to two important goals: to observe the frequency content by scales, i.e., through the *LD-Diagram* and to estimate the Hurst index. Also, those components can be added by means of a weighted sum in order to produce a new time series with different Hurst index and, consequently, different correlation. It is worthy to mention that abrupt changes in the Hurst index may lead to pathological series so it is convenient that the desired new Hurst index is close to the original.

Finally, the variance of the aggregated series and the components are described, for both the ideal and real world cases. I.e., theoretically the variance of the time series is the sum of an infinite number of components but as a real world time series is of finite size, the number of components is also finite and the formulae have to be adapted.

3 Estimation of Mean, Variance, and Hurst Index of Second-order Self-similar Time Series

3.1 Sample Mean

The sample mean of a self-similar process is unbiased: its expected value is the process mean, i.e., $E(\bar{X}) = \mu_X$, where $\bar{X} = 1/N \sum_{t=1}^{N} X_t$, regardless of the presence of correlation between observations. However, its variance does not depend only on the sample size (N), but also on the degree of self-similarity (H) of the process, as follows:

$$var(\bar{X}) = \sigma_X^2 N^{2H-2} \tag{131}$$

which becomes:

$$var(\bar{X}) = \frac{\sigma_X^2}{N} \tag{132}$$

(classical estimator) for $H = 0.5$ (uncorrelated observations).

To derive (131), consider that $\bar{X}$, estimated from a sample of size N, behaves exactly the same as the stationary aggregated process $X_k^{(N)}$, defined by (5), and its variance is determined by the definition of second-order self-similarity (11). Expression (131) can be also derived (for $H > 0.5$) from the auto-correlation coefficient $\rho(k) = 0.5[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]$ for $k \geq 1$ ($\rho(0) = 1$) and $var(\bar{X}) = \frac{\sigma_X^2}{N^2} \sum_{i,j=1}^{N} \rho(k)$.

Important implications of (131) about the uncertainty of the mean are: 1) that it increases with the Hurst index, e.g., $var(\bar{X})$ tends to σ_X^2 as H tends to 1, which makes the sample mean worth as a single observation, and 2) that it cannot be zero for any case when estimated from a finite-size sample.

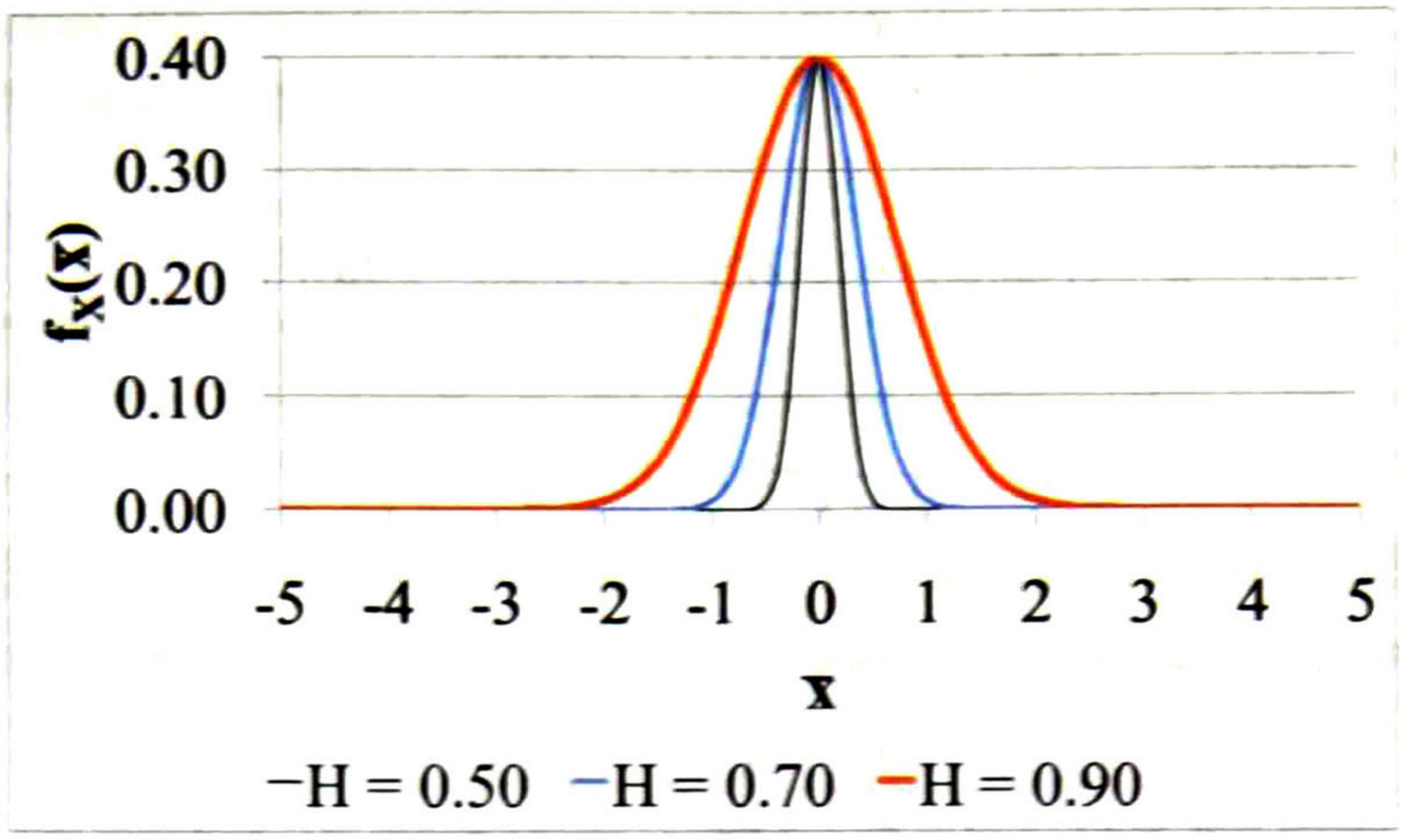


Figure 15. Distribution of the sample mean of standardized fractional Gaussian noise processes with $H = \{0.50, 0.70, 0.90\}$ and $N = 32$.

3.2 Sample Variance

For high number of observations, sample variance is usually calculated using the following:

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{t=1}^{N}(X_t - \hat{\mu}_X)^2 \tag{133}$$

which is widely known to be biased so it is more adequate to use:

$$\hat{\sigma}_X^2 = \frac{1}{N-1}\sum_{t=1}^{N}(X_t - \bar{X})^2 \tag{134}$$

especially for small samples.

Particularly, if the observations X_t are independent and come from a normal distribution, $\hat{\sigma}_X^2$ is distributed as $\hat{\sigma}_X^2 \sim \frac{\sigma_X^2}{N}\chi_{n-1}^2$, as stated by Cochran's theorem [47], [48]. Formula (134) is the most used estimator of the sample variance but, as Beran indicates in [49], it is needed to know which assumptions this estimator is based on in order to apply it correctly; otherwise it may be the source of errors that in practice cannot be negligible for all cases.

A self-similar process is uncorrelated only and only if the Hurst index is 0.5. In this particular case the classical estimator of sample variance, defined by (134), is unbiased.

The expected value of the sample variance defined by (134) is:

$$E(\hat{\sigma}_X^2) = E\left[\frac{1}{N-1}\sum_{t=1}^{N}(X_t - \hat{\mu})^2\right] \tag{135}$$

which can be expressed as:

$$E(\hat{\sigma}_X^2) = \frac{N}{N-1}[E(X_t^2) - E(\hat{\mu}^2)] \tag{136}$$

which is equivalent to:

$$E(\hat{\sigma}_X^2) = \frac{N}{N-1}\{var(X_t) + [E(X_t)]^2 - \{var(\hat{\mu}) + [E(\hat{\mu})]^2\}\} \tag{137}$$

$$E(\hat{\sigma}_X^2) = \frac{N}{N-1}\sigma_X^2(1 - N^{2H-2}) \tag{138}$$

Expression (138) proves that the classical estimator (134) is biased, i.e., $E(\hat{\sigma}_X^2) \neq \sigma_X^2$, for $H \neq 0.5$. It is straightforward that the unbiased variance estimator for self-similar processes is then:

$$\hat{\sigma}_X^2 = \frac{1}{N - N^{2H-1}}\sum_{t=1}^{N}(X_t - \bar{X})^2 \tag{139}$$

which obviously becomes (134) for $H = 0.5$.

A plot of $\log_{10}(\hat{\sigma}_X^2 - \sigma_X^2)$ vs. N is shown in Figure 16. Note that, for a fixed sample size, as H increases the estimation of the variance by means of the classical estimator (134) becomes less reliable.

In Figure 16 the following details are observed: as the sample size is greater, the magnitude of the bias of the classical variance estimator is lower; however, as the Hurst index of the process is greater, that magnitude also greater. As H approaches to 1, the variance is considerably underestimated, makes the classical variance estimator useless.

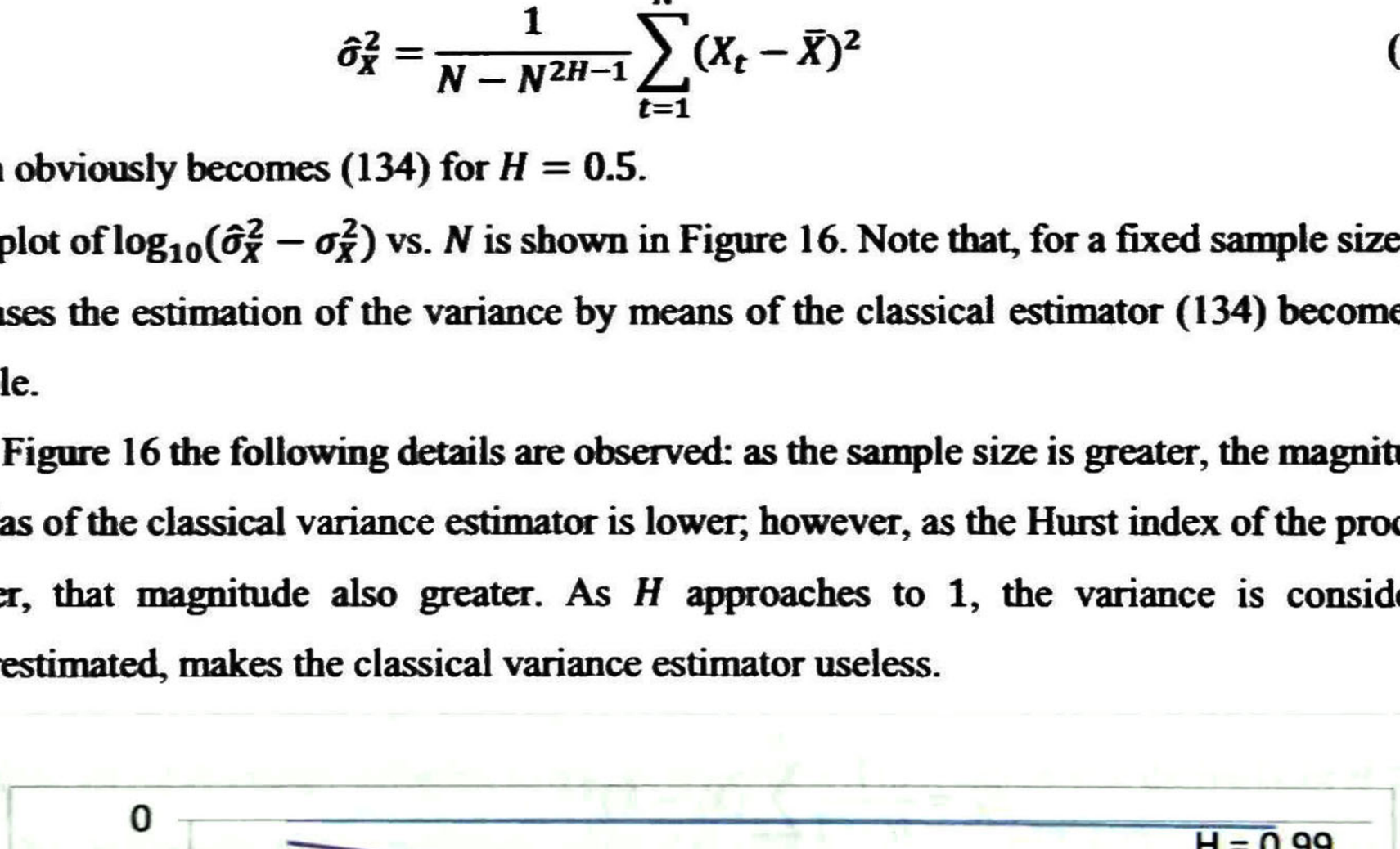


Figure 16. Logarithm of the variance bias for different sample sizes.

The variance of the estimated variance of a self-similar time series can be approximated by applying the formula proposed by Yunhua in [50] for $k = 0$, i.e.,

$$var(\hat{\sigma}_X^2) = \frac{(2N^{4H-3} + 8H - 7)(2H^2)(2H-1)^2}{N(4H-3)} + \frac{1}{N} \tag{140}$$

This approximation is close to the variance of $\hat{\sigma}_X^2$, with the disadvantage that it has a discontinuity in $H = 0.75$. A further work can be developed in order to verify this approximation and to quantify its error.

It is worthy to mention that, although the proposed estimator of the sample variance is unbiased, its performance relies on the estimation of the Hurst index. This dependence is very noticeable as H approaches to 1, as the statistic $1 - N^{2\hat{H}-2}$ is especially sensitive to the variation of $\hat{H}$ under that condition. The derivative $\frac{d\left(1-N^{2\hat{H}-2}\right)}{d\hat{H}} = -2N^{2\hat{H}-2}ln(N)$ vs. H is shown in Figure 17. Note that as the estimated H increases, the estimation of the mean of the sample variance is more variable.

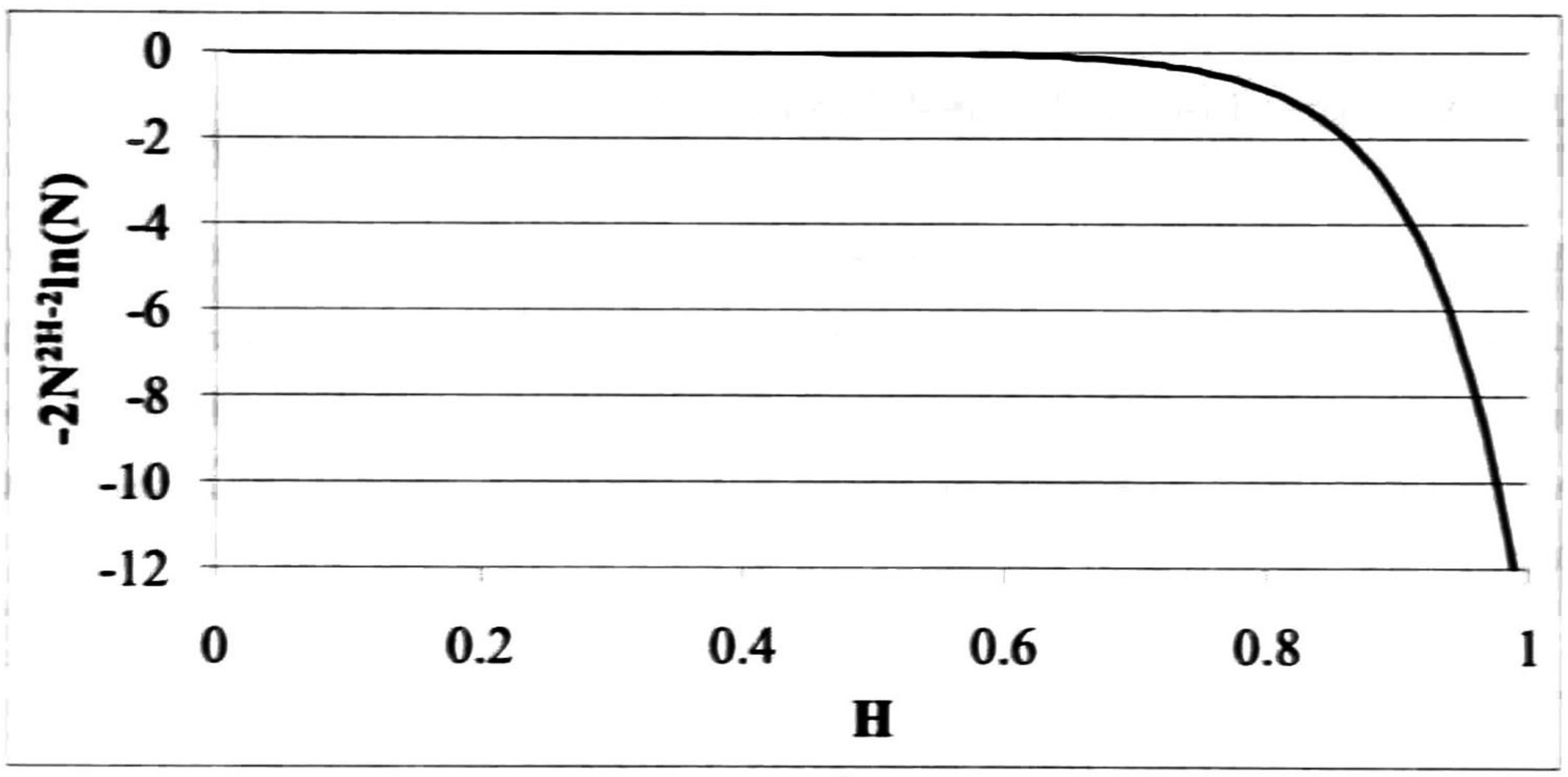


Figure 17. Plot $-2N^{2\hat{H}-2}ln(N)$ vs $\hat{H}$.

An immediate implication of this is that processes with Hurst index close to 1 must be carefully treated, as slight deviations of the Hurst index estimation derive in a non-negligible error in the estimation of the process variance.

3.3 Statistics of the Aggregated Process

The aggregated process $X_k^{(m)}$, as defined by (5), that is derived from an H-SS process is also H-SS (self-similar with the same Hurst index). It is also true for the case of H-SOSS processes. And

this aggregated process is, by definition, identically distributed to the sample mean obtained from a set of $N = m$ observations (m is the aggregation level, as in (5), i.e., $X_k^{(m)} \sim \hat{\mu}$. Also, the aggregated sample variance obtained with the classical estimator (134) is biased, as it is well known [51].

The variance of the aggregated series of an H-SOSS process must be then estimated with the unbiased formula (139) adapted to the number of observations in the sample, i.e.,

$$var\left(\hat{X}_k^{(m)}\right) = \frac{1}{N_i - N_i^{2H-1}} \sum_{k=1}^{N_i} \left(X_k^{(m)} - \overline{X_k^{(m)}}\right)^2 \tag{141}$$

where N_i is the size of the series after aggregation (i.e., $N_i = N/m$). Note that the estimation of the sample mean from the aggregated sample is also unbiased, i.e., $E\left[\overline{X_k^{(m)}}\right] = E(X_t)$ and it is more reliable than the mean estimated from a sample of the same size, i.e., $var\left[\overline{X_k^{(m)}}\right] = m^{2H-2} var(\bar{X})$ if the samples are of equal size.

3.4 Statistics of the Orthogonal Components

Let $\{\hat{X}_t; t = 1, \ldots, N\}$ be a finite-length self-similar time series such that $N = n^J$ ($J < \infty$) and $n \geq 2$ (i.e. N is a power of n), then a set of non-zero J components ($\hat{C}_{X,t}^{n,j}; j = 1, \ldots, J$) can be obtained as expressed by the analysis equation (79). As the components are pair wise orthogonal, the variance of $\hat{X}_t$ ($\hat{\sigma}_X^2$) is the sum of a finite number of variances, i.e., applying (89) to real world traces, that is:

$$\hat{\sigma}_X^2 = \sum_{j=1}^{J} var\left(\hat{C}_{X,t}^{n,j}\right) \tag{142}$$

Expression (142) and (82) imply that the variance of the j^{th} component $\hat{C}_{X,t}^{n,j}$ (which has finite length) and the variance of $\hat{X}_t$ are related as (108), i.e.:

$$var\left(\hat{C}_{X,t}^{n,j}\right) = \frac{1-r}{1-r^J} r^{j-1} \hat{\sigma}_X^2 \tag{143}$$

Applying (86) in (143), it is derived:

$$var\left(\hat{C}_{X,t}^{n,i}\right) = \frac{var\left(C_{X,t}^{n,j}\right)\left(1 - N^{2H-2}\right)}{1 - r^J} \tag{144}$$

And, finally, by replacing r^J by N^{2H-2}, it yields:

$$var\left(\hat{C}_{X,t}^{n,i}\right) = var\left(C_{X,t}^{n,j}\right) \tag{145}$$

Expression (145) implies that the estimation of the variance of components is unbiased (a desirable property) and, as a consequence, so it is the estimation of the statistic $S_2(j)$. Another implications of (145) is that the estimations of the Hurst index and the power parameter $c_f C$ from the *LD-Diagram* are unbiased, as have previously proven by the authors of [11].

3.5 Correlation of the Wavelet Coefficients

In this Section, the auto-covariance of the Haar wavelet coefficients is derived, i.e., the case for $n = 2$ in analysis equation (79). The structure of the j^{th} component ($C_{X,t}^{2,j}$), as a function of the elements of X_t is:

$$C_{X,\tau}^{2,j} = \frac{x_{2\tau-1}^{(2^{j-1})} - x_{2\tau}^{(2^{j-1})}}{2} \tag{146}$$

where $C_{X,\tau}^{2,j}$ is a downsampled version of $C_{X,t}^{2,j}$, i.e., only the first observation of each 2^j of $C_{X,t}^{2,j}$ remains. Then, two wavelet coefficients are correlated as:

$$\gamma_{C_j}(k) = E\left(C_{X,\tau}^{2,j}, C_{X,\tau+k}^{2,j}\right) = E\left(\frac{x_{2\tau-1}^{(2^{j-1})} - x_{2\tau}^{(2^{j-1})}}{2} \frac{x_{2(\tau+k)-1}^{(2^{j-1})} - x_{2(\tau+k)}^{(2^{j-1})}}{2}\right) \tag{147}$$

Assuming that X_t represents a zero-mean H-SOSS process and according to the definition (12), $E\left(x_{2\tau-1}^{(2^{j-1})} x_{2(\tau+k)-1}^{(2^{j-1})}\right) = E\left(x_{2\tau}^{(2^{j-1})} x_{2(\tau+k)}^{(2^{j-1})}\right) = \gamma_X(2k)$, $E\left(x_{2\tau-1}^{(2^{j-1})} x_{2(\tau+k)}^{(2^{j-1})}\right) = \gamma_X(2k+1)$ and $E\left(x_{2\tau}^{(2^{j-1})} x_{2(\tau+k)-1}^{(2^{j-1})}\right) = \gamma_X(2k-1)$. Then (147) is calculated as:

$$\gamma_{C_j}(k) = \frac{2^{2H-2}\sigma_X^2}{4}[2\rho_X(2k) - \rho_X(2k+1) - \rho_X(2k-1)] \tag{148}$$

and the correlation coefficient of the j^{th} component is then

$$\rho_{C_j}(k) = \rho_C(k) = \frac{[\rho_X(2k+1) - 2\rho_X(2k) + \rho_X(2k-1)]}{2^{2H} - 4} \tag{149}$$

As (149) shows, the correlation structure is the same for all components, i.e., $\rho_{C_j}(k)$ is independent of j. Another implication of (149) is that:

$$\sum_{k=0}^{\infty} \rho_{C_j}(k) = 1 + \frac{-\rho_X(1) - \rho_X(2k+1) + 2\sum_{i=1}^{2K+1}[(-1)^{i-1}\rho_X(i)]}{2^{2H} - 4} \tag{150}$$

which, in practice, can be approximated as:

$$\sum_{k=0}^{\infty} \rho_{C_j}(k) \approx -0.052H^2 - 0.311H + 1.168 < \infty \tag{151}$$

i.e., the coefficients are weakly correlated and the sum of correlations is finite [52]; furthermore, none of the components (sequences of wavelet coefficient) can be a self-similar time series except for $H = 0.5$, e.g., components of a white noise process are also white noise processes. A plot of $\rho_{C_j}(k)$ vs. k is shown in Figure 18 for $H = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The sum $\sum_{k=0}^{\infty} \rho_{C_j}(k)$ vs. H is shown in Figure 19.

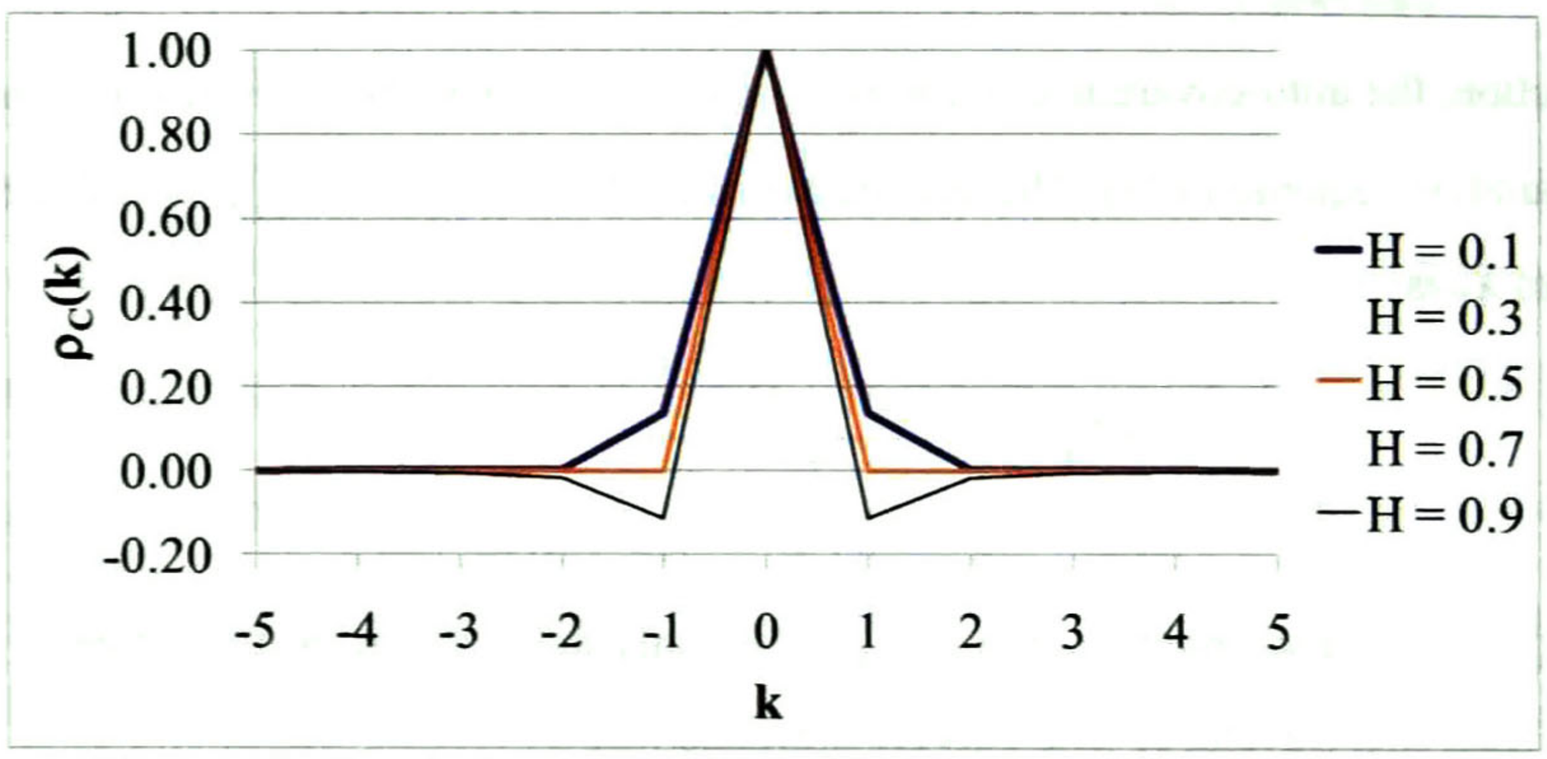


Figure 18. Correlation coefficient of the Haar wavelet components.

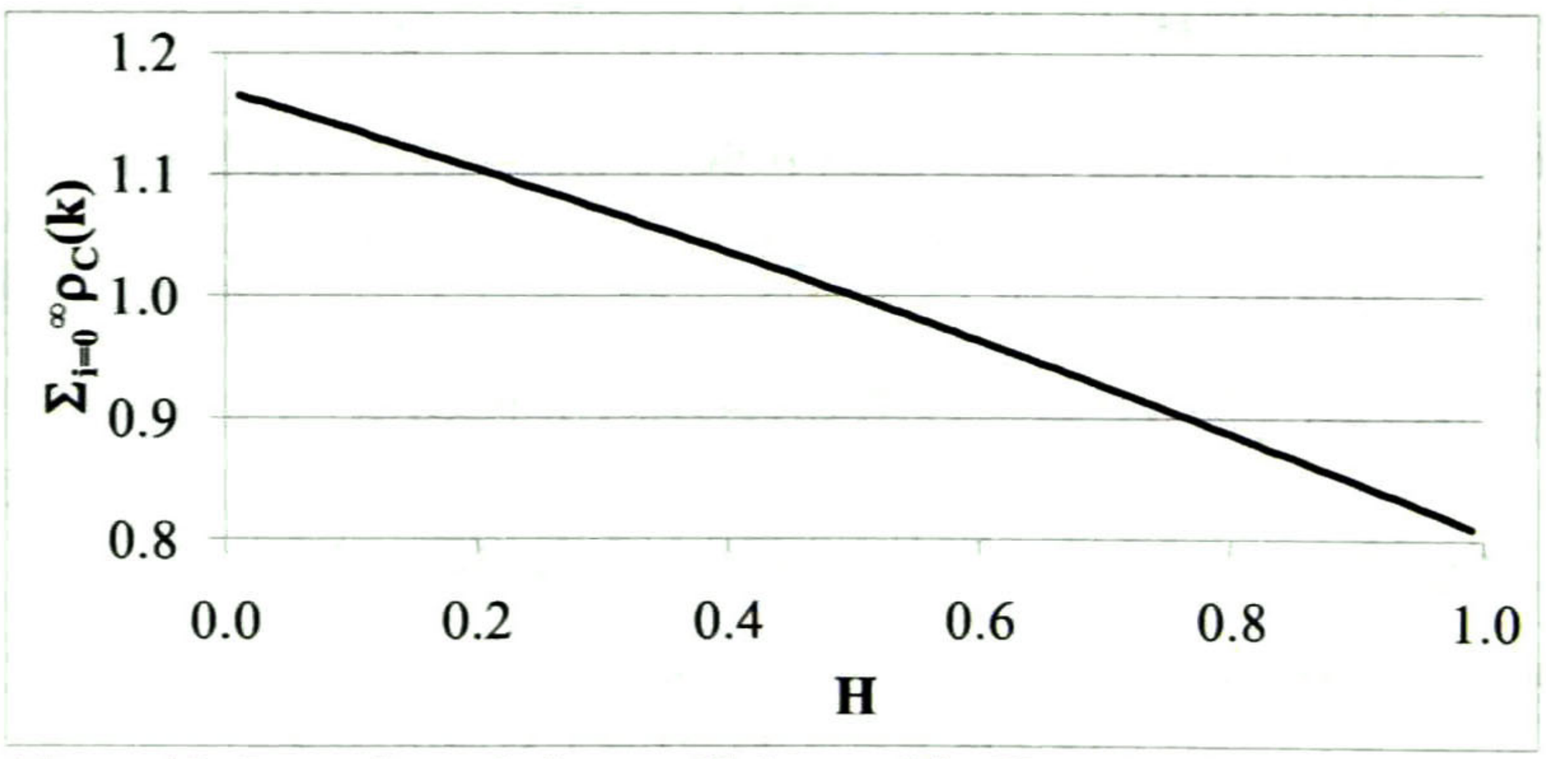


Figure 19. Sum of correlation coefficients of the Haar wavelet components.

As the maximum magnitude of $\gamma_{C1}(1)$ is $1/8$ (when $H \to 0$), it is said that these coefficients are *quasi-uncorrelated*. Note that $\gamma_{C1}(1) = 0$ for $H = 0.5$ and $H \to 1$.

Although the assumption that the estimation of the 1^{st} component variance ($var\left(\hat{C}_{X,t}^{n,1}\right)$) is unbiased is nearly accurate, it may not hold for components of greater order. As the wavelet coefficients $d_X(j,k)$ are almost uncorrelated, the estimation of the variance as $\sum_{t=1}^{N/n^i}\left(\hat{C}_{X,t}^{n,i}\right)^2 \approx \frac{N-n^i}{N} var\left(C_{X,t}^{n,i}\right)$.

3.6 *Variance-Plot*-based Estimation of the Hurst Index

As described in Section 1.1.4, the *Variance-Plot* is a straight line for self-similar time series but, as many authors have claimed, it underestimates the Hurst index when working with real-world data. This is already noted and documented in the literature, per example: Riedi et al name this phenomenon "the decay of the aggregate variance" [53]. Sheluhin et al state that the *Variance-Plot*, can only be used to check whether the time series is self-similar or not and, if so, to obtain a crude guess for the Hurst index [18]. Clegg developed a comparison of Hurst index estimators in the presence of additional noise (e.g., linear trends or sinusoidal), including the aggregated variance estimator, reporting unbiased estimations when no additional noise is added [13]. Krunz et al perform a comparison of the variance-based estimator of LRD with other three estimators. They conclude that the estimator is inherently biased, that can often lead to incorrect conclusions, and that the bias diminishes gradually with the data [51]. Jeong et al also perform a comparison of several estimators on simulated FGN, concluding that the variance-based estimator becomes increasingly negatively biased as H increases [7].

Of course, these conclusions and claims are valid in some sense. That is, applying the classical variance estimator to variance-based methods leads to biased and not reliable calculations. But the objective of this section is to clarify that these claims are a consequence of the inadequate estimation of the aggregated variance, caused by the application of the classical formula (134) regardless of whether the original process presents any type of correlation.

The solution would be then to apply the unbiased formula (141), but it leads to an ill-conditioned problem: the Hurst index need to be estimated and known at the same time. Then it is not that the *Variance-Plot* is not adequate to estimate the Hurst index, rather the flaw of those implementations is that the aggregated variance is underestimated.

Nevertheless, it is actually possible to estimate the Hurst index analytically from the *Variance-Plot*: the key is to choose the aggregation levels so that they form a geometric series, as explained in Section 5.1. Note that a numerical method can also be applied to estimate simultaneously the *Variance-Plot* and the Hurst index, but the proposed solution is computationally simpler and more efficient.

3.6.1 Analytical Solution to the Ill-conditioned Problem

Let $\{m_i; i = 1, \dots, M\}$ be the set of aggregation levels such that $m_i = am_{i-1} = a^{i-1}m_1, a, m_1 \in \mathbb{N}$ and $a > 1$, i.e., the levels of aggregation follow a geometric series, e.g., $\{m_i\} = \{2, 4, 8, \dots, 2^M\}$

or $\{m_i\} = \{10, 100, 1000, \ldots, 10^M\}$, and let $\hat{\sigma}^2_{X(m_i)}$ be the variance of the aggregated series $X_k^{(m_i)}$ estimated with formula (133). Obviously $\hat{\sigma}^2_{X(m_i)}$ is biased, as $\hat{\sigma}^2_{X(m_i)} = \sigma^2_{X(m_i)}\left[1 - \left(\frac{N}{m_i}\right)^{2H-2}\right] = \sigma_X^2 m_i^{2H-2}(1 - N^{2H-2})$, but then the difference between $\hat{\sigma}^2_{X(m_i)}$ and $\hat{\sigma}^2_{X(m_{i+1})}$ is calculated as:

$$\widehat{\Delta}_i = \hat{\sigma}^2_{X(m_i)} - \hat{\sigma}^2_{X(m_{i+1})} = \sigma_X^2 m_1^{2H-2} a^{(i-1)(2H-2)}(1 - a^{2H-2}); i = 1, \ldots, M-1 \quad (152)$$

and its logarithm

$$\log_a \Delta_i = \log_a[\sigma_X^2 m_1^{2H-2}(1 - a^{2H-2})] + (i-1)(2H-2) \quad (153)$$

Finally, the slope (s) of the plot $log_a \Delta_i$ vs. i is obtained (e.g., with a weighted least square regression) and, using it, H is estimated as $\widehat{H} = \frac{\hat{s}}{2} + 1$. It can be easily proven as $\hat{\sigma}^2_{X(m_i)} = X_t^{(n^{j-1}E)}$ and substituting it in (79) and (82).

The slope is computed by the following weighted formula:

$$\hat{s} = \frac{\sum_{i=1}^{M-1}(i\widehat{\Delta}_i W_i) - \sum_{i=1}^{M-1}(iW_i) \cdot \sum_{i=1}^{M-1}(\widehat{\Delta}_i W_i)}{\sum_{i=1}^{M-1}(i^2 W_i) - \left[\sum_{i=1}^{M-1}(iW_i)\right]^2} \quad (154)$$

where the weights are such that $\sum_{i=1}^{M-1}(W_i) = 1$ and they are adequate so that $\hat{s}$ has minimal variance, e.g., $W_i = W_{i-1}/m_1$.

Note that the *Variance-Plot* can be estimated without bias but the Hurt index is not estimated from it. Furthermore, if the aggregation levels are taken as $m_i = 2^i$, the estimator is exactly the same than one that uses Haar wavelet. The authors of [54] developed an empirical study of the estimation of the Hurst index from series with the presence of trends. They conclude that a method named differenced-variance (a variation of the variance-bases estimator) should not be used for estimating the Hurst index. The proposed solution is also a differenced-variance type method, but it can be used to estimate the Hurst index without bias and with optimal variance. Evidently, when working with real world traces, the *Variance-Plot* may differ from the straight line and an additional bias results from the logarithm as $E[log(\cdot)] \neq log[E(\cdot)]$. This bias can be corrected applying by estimating the statistic analogously to (97), i.e.,

$$log_a\, var\left[X_t^{(m_i)}\right] = log_a\, var\left[X_t^{(m_i)}\right] - g_i \quad (155)$$

where g_i is the bias described by (20).

3.7 Chapter Summary

In this chapter, the mean and variance of an H-SOSS self-similar time series are calculated in terms of the Hurst index. The estimation of the mean from a real world sample is unbiased,

regardless of its distribution and Hurst index. Only the confidence intervals for that estimation are different when correlation is present. I.e., they are narrower for samples with SRD ($H < 0.5$) and wider for samples with LRD ($H > 0.5$), as expression (131) indicates.

The classical sample variance estimator, defined by expression (134), is unbiased when $H = 0.5$ but biased otherwise. The disregarding of this fact may lead to incorrect conclusions. As the sample size is greater the magnitude of the bias is smaller, but for traces with strong LRD, i.e., $H > 0.90$, the classical variance estimator is practically useless. The unbiased estimator is defined by expression (139), which considers the existing correlation between samples by means of the Hurst index. It is necessary to highlight that, although the estimator is mathematically unbiased, its performance on real world samples relies on the accurate estimation of the Hurst index. And, as shown in Figure 17, a slight error on the estimation of the Hurst index may lead to a very different variance estimation for samples with H close to 1.

The use of the classical variance estimation to obtain the *Variance-Plot* from a self-similar sample has been a very common mistake in works related to this subject. Several works in the literature have claimed that the variance-based estimator is unbiased, and that the higher the Hurst index, the less reliable the estimation. In Section 3.6, it is clarified that this has been a misunderstanding, originated by the application of the classical variance formula to correlated samples. A possible solution leads to an ill-conditioned problem, i.e., to estimate the variance, it is necessary to estimate the Hurst index, and vice versa. In order to avoid this vicious circle, the a different plot is derived from the original *Variance-Plot* calculating the difference of the variance between consecutive aggregation levels. That produces a plot whose slope is not affected by the bias (although the intersect is) from which the Hurst index can be calculated. And even more, it is proven that the "new" plot is exactly the *LD-Diagram*, i.e., the variances' differences are actually the variances of the components obtained with the analysis functions defined by (98), (99), and (100).

This chapter is based on publication [55]. Also, a more recent study [56] expresses that the aggregated variance is underestimated by a factor of $1 - CM^{2H-2}$, where C is some constant and M is the number of aggregated sub-series. A previous study of the variance underestimation of self-similar time series in published in [57].

4 VoIP Performance Metrics

4.1 *N*-packet Forward Error Correction

N-packet FEC consists of that packet $n + 1$ contains information about the last N packets, so that if any of the packets $n, n-1, n-2, \ldots, n-N+1$ is lost, it can be approximately reconstructed from the associated information. Any packet n cannot be reconstructed if there is no redundant information, e.g., when packets $n+1, n+2, \ldots, n+N$ are also lost, but that is not very likely for high values of N, so generally $N = 1$ will suffice. An alternative to the usage of FEC is MPT, which consists of sending copies of packets when high losses occur. In order to maximize the probability of reception, these copies must be equally spaced in the time [58]. Although MPT has the advantage that it is very easy to implement, unless a low bit rate coded is used, it has the disadvantage of high bandwidth requirement consumption.

When using N-packet FEC, the last N packets of a burst can be reconstructed; that way, the perceived PLR of the end user is lower than the real PLR due to the network. Generally, the amount of redundancy is defined as a function of the PLR [35], e.g., it is not efficient to send redundant information if there are no missing packets. The FEC technique also reduces the *burstiness* of the perceived packet loss at the receiver, which affects the quality of a VoIP communication [59]. Many packet recovery schemes are documented in the literature [22].

The 1-packet FEC technique performance can be described as: it reduces the size of a burst of length k to $k-1$. The perceived PLR (r_1') is proportional to the perceived average burst length, which in this case decreases by 1 (packet), then it is equal to:

$$r_1' = \frac{(\bar{b}-1)r}{\bar{b}} \tag{156}$$

where $\bar{b}$, the average burst length, is:

$$\bar{b} = \sum_{k=1}^{x} k f_b(k) \,; k = 1, 2, \ldots \tag{157}$$

and $f_b(k)$ is the burst length distribution, i.e., $f_b(k) \geq 0 \; \forall \; k = 1, 2, \ldots$ and $\sum_{k=1}^{\infty} f_b(k) = 1$.

If the redundancy level extends to N packets, i.e. packet n has information about $n+1, n+2, \ldots, n+N$ packets, the length of all bursts decreases from k to $\max(0, k-N)$ packets, then the new burst length distribution $f_b'(k)$ is:

$$f_b'(k) = \begin{cases} \sum_{i=1}^{N} f_b(i); & k = 0 \\ f_b(k+N); & k > 0 \end{cases} \tag{158}$$

Note that (158) includes the case when $k = 0$, that is, it considers bursts of zero length. The interpretation of this is as follows: bursts do really occur in the network but, as they are corrected by a N-packet FEC technique, they are diminished (when $k > N$) or eliminated (when $k \leq N$) in the receiver. Then, $f_b'(k)$ is the new average burst length can be calculated as:

$$\bar{b}' = \sum_{k=1}^{\infty} [k f_b'(k)] \tag{159}$$

$$\bar{b}' = \bar{b} - N + \sum_{k=1}^{N} (N-k) f_b(k) \tag{160}$$

Consequently, the perceived PLR is:

$$r_N' = \frac{\left[\bar{b} - N + \sum_{k=1}^{N-1} (N-k) f_b(k)\right] r}{\bar{b}} \tag{161}$$

which is a generalized form of (156).

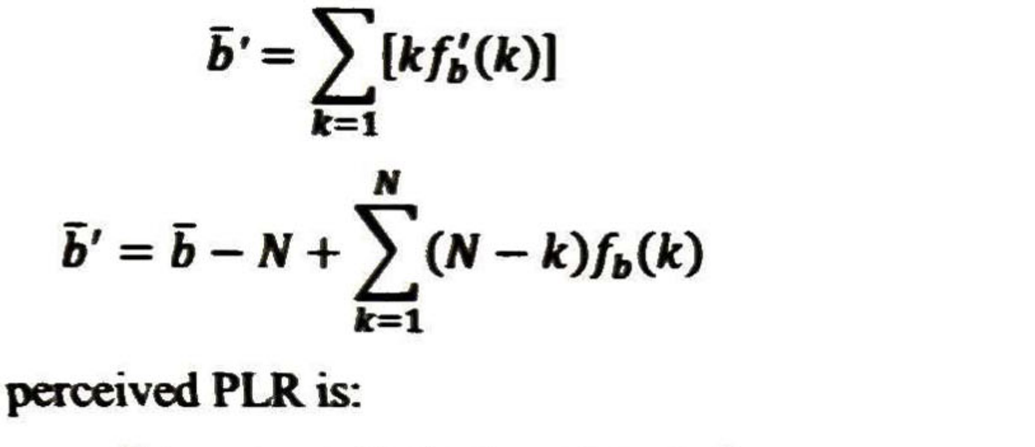


Figure 20. Example of the burst length distribution before and after 1-packet FEC. PLR is reduced from 4% to 0.55%, approximately.

Figure 20 shows an example of the burst length distribution, as expressed by (63) with $p_{21} = 0.001350$, $p_{12} = 1.000000$, $p_{43} = 0.054507$, $p_{34} = 0.845146$, $p_{23} = 0.001968$ and $p_{32} = 0.016989$, and how it is modified, from the point of view of the receiver, after applying 1-packet FEC. In this example, all bursts are reduced in 1 packet, as is defined by (158) and, as a consequence, the perceived PLR is reduced from 4.022% to 0.548%. Similarly, Figure 21 shows the comparison of the burst length distribution before and after 2-packet FEC. In this example the PLR is reduced to 0.075%.

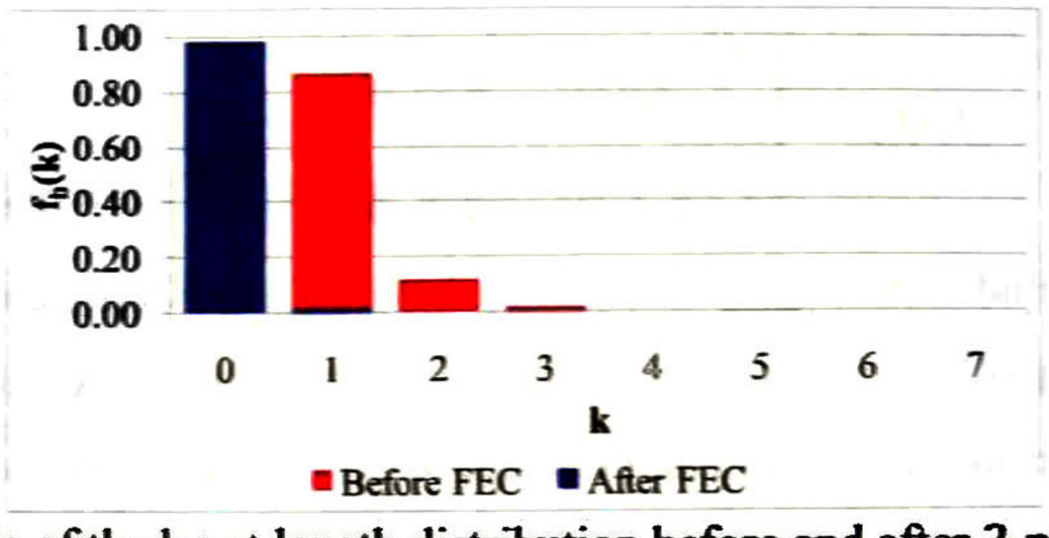


Figure 21. Example of the burst length distribution before and after 2-packet FEC. PLR is reduced from 4% to 0.075%, approximately.

Note that (161) expresses the perceived PLR of the receiver without considering other sources of losses, e.g., additional perceived losses occur if packets are delayed more than certain threshold (i.e., de-jitter buffer timeout).

4.1.1 Estimation of Burst Length Distribution Using Discrete Markov Processes

Packet losses due to network generally present a non-homogeneous behavior. This means that periods of low losses and high losses occur alternatively in time. Intuitively, it seems more appropriate to apply a non-homogeneous model to represent these losses. However, unlike the received packets, these losses generally occur in bursts of short length, e.g., the probability that a burst has a length greater than, per example, 10 packets or even less, is negligible (an example is shown in Figure 46). This quick decay of the burst length distribution allows us to apply a homogeneous model, e.g., the 2-state Markov process, without significant error.

From the perceived burst length distribution, expressed by (158), and the 2-state burst length distribution (54) is derived:

$$f_b'(k) = \begin{cases} \sum_{i=1}^{N} p_{12}(1-p_{12})^{i-1}; & k = 0 \\ p_{12}(1-p_{12})^{k+N-1}; & k > 0 \end{cases} \tag{162}$$

The perceived average burst length, defined by (159) is then:

$$\bar{b}' = \sum_{k=1}^{\infty} [k f_b'(k)] = \sum_{k=1}^{\infty} [k p_{12}(1-p_{12})^{k+N-1}] \tag{163}$$

$$\bar{b}' = \sum_{k=1}^{\infty} [k p_{12}(1-p_{12})^{k+N-1}] = (1-p_{12})^N \sum_{k=1}^{\infty} [k p_{12}(1-p_{12})^{k-1}] \tag{164}$$

$$\bar{b}' = \frac{(1-p_{12})^N}{p_{12}} \tag{165}$$

The perceived PLR, that is proportional to the average burst length, is then:

$$r'_N = r(1-p_{12})^N \tag{166}$$

Note that the perceived PLR, without considering discarded packets in the de-jitter buffer, decays geometrically as N increases.

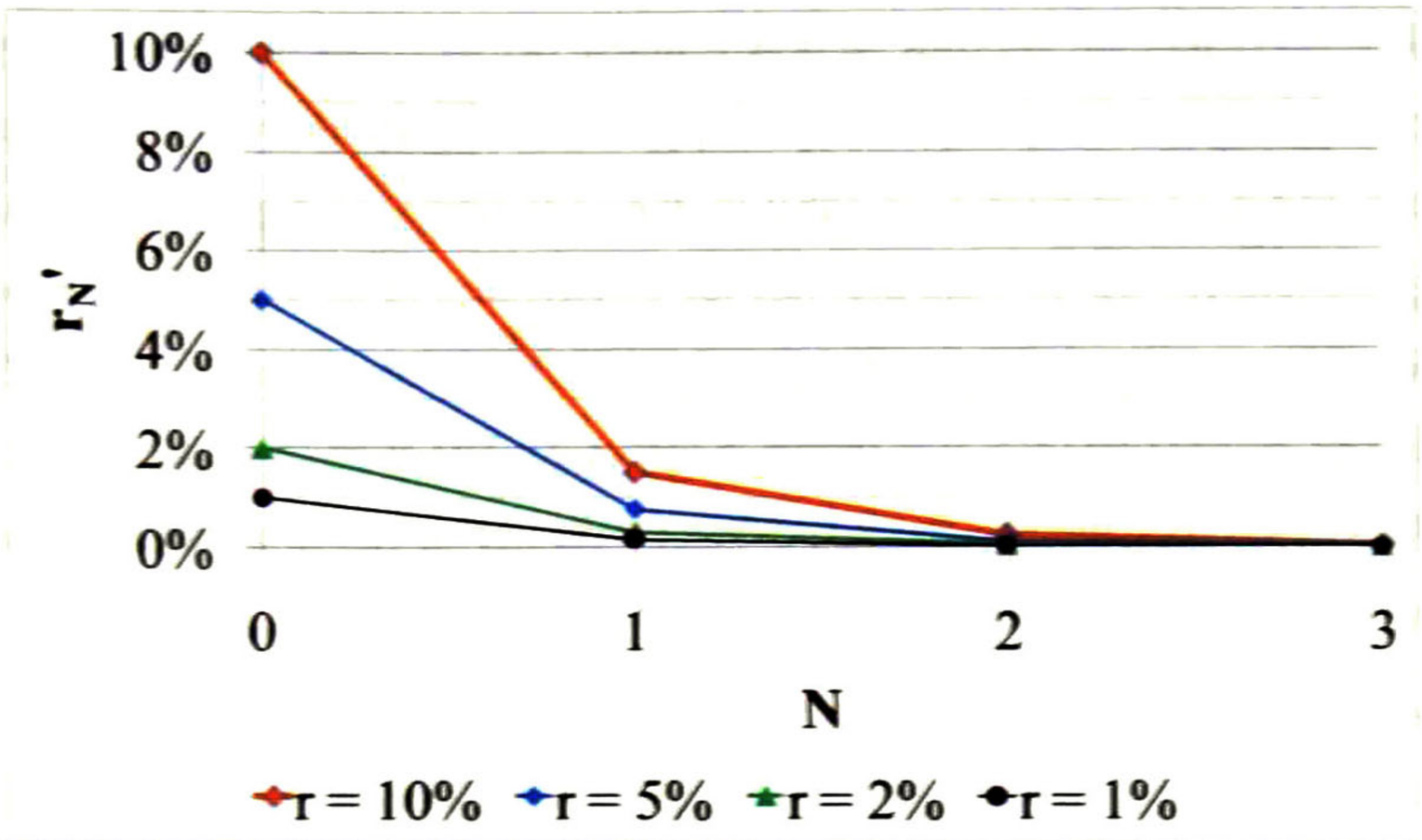


Figure 22: Plot r'_N vs. N for $r = \{10\%, 5\%, 2\%, 1\%\}$ and $p_{12} = 85\%$.

Figure 22 is an example of the perceived PLR after FEC reconstruction when $p_{12} = 85\%$ for four cases: $r = \{10\%, 5\%, 2\%, 1\%\}$. For the four cases in this example, the perceived PLR is reduced below 1% for $N = 2$. Expression (166) can also give an estimation of the level of redundancy needed to achieve certain perceived PLR. Also note that, generally one or two levels of redundancy are necessary to reduce the perceived PLR below 1%.

An important matter is that the modeling of the burst length distribution using a 2-state Markov process may present an error for non-homogeneous losses. The advantage of using this model is that there is only one parameter needed to estimate: the probability p_{12} of expressions (54) and (166). The error of the burst length distribution can be estimated by comparing the 2-state vs. the 4-state models, i.e., by obtaining the quantity:

$$\varepsilon(f_b) = \sum_{k=1}^{\infty} [\hat{f}_b(k) - f_b(k)]^2 \tag{167}$$

The estimation of p_{12} using the algorithm explained in Section 4.4.2 produces the following approximation:

$$\hat{p}_{12} \approx \frac{s_2(p_{21} + p_{23}) + s_4 p_{43}}{s_1 + s_3} \tag{168}$$

which is the multiplicative inverse of the average burst length (61). Note that the parameters of the right part of expression (167) correspond to the 4-state model. The estimated burst length distribution, according to (54), is then:

$$\hat{f}_b(k) \approx \frac{s_2(p_{21}+p_{23})+s_4p_{43}}{s_1+s_3}\left(1-\frac{s_2(p_{21}+p_{23})+s_4p_{43}}{s_1+s_3}\right)^{k-1} \quad (169)$$

which will be compared to $f_b(k)$, defined by (63) in order to obtain the estimation error (167).

Comparing expressions (63) and (169), it is noted that the reliability of this approximation of the burst length distribution depends on how different are the rates two geometric series with respective rates $1-p_{12}$ and $1-p_{34}-p_{32}$. As p_{32} can be considered very low against p_{34} (see average values for these parameters in Table 8), the performance of this approximation reduces to compare the probabilities p_{12} and p_{34}. I.e., the performance of the estimator will be worse as the absolute difference between p_{12} and p_{34} is greater. The expected error of the estimation of the burst length distribution ($\varepsilon(f_b)$) is shown in Figure 23. The average values shown in Table 8 are used for p_{21}, p_{23}, p_{32} and p_{43}. It is noted that, for the shown values of these parameters (both p_{12} and p_{43} between 0.6 and 1.0), the error is lower than 0.005.

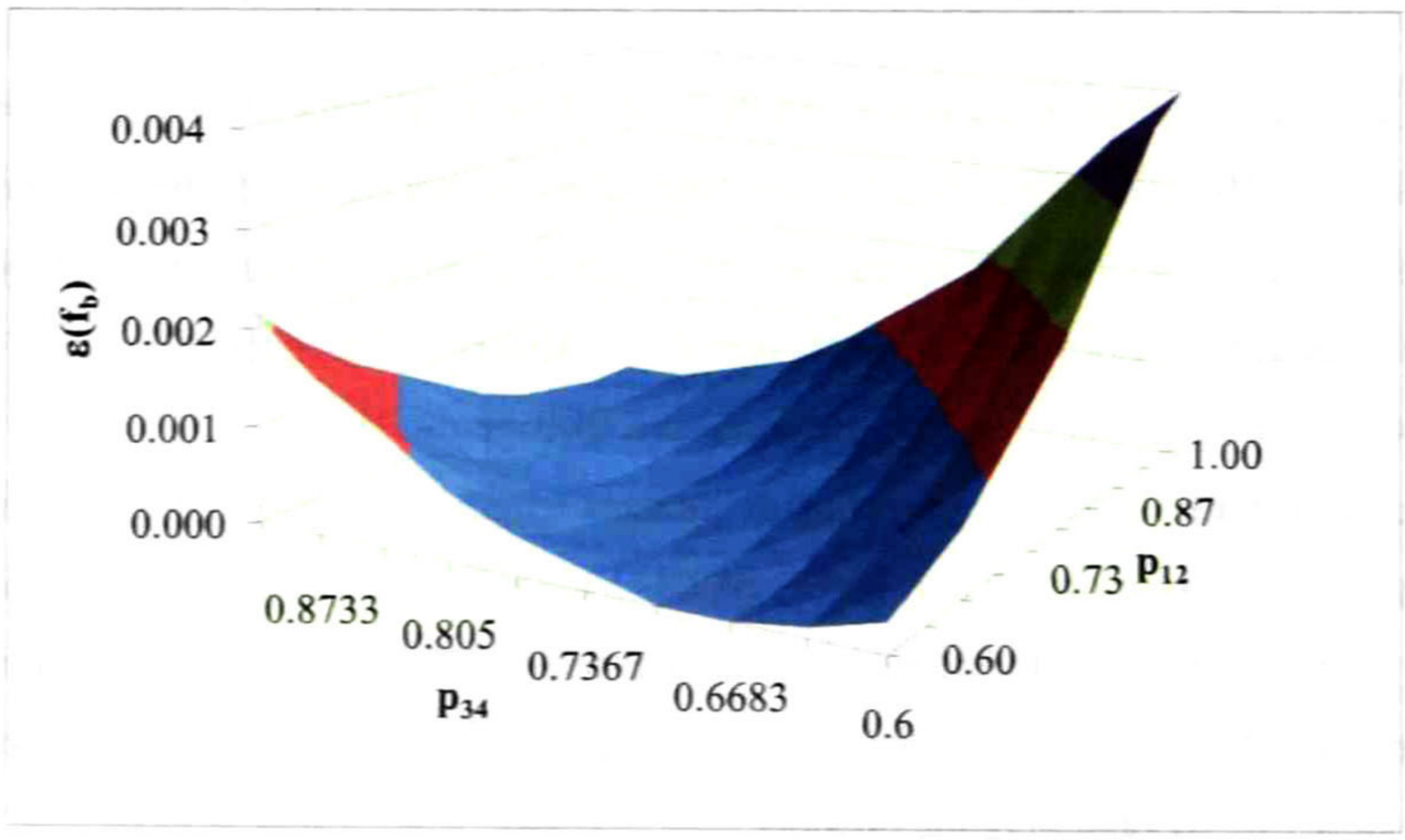


Figure 23: SMSE the estimated burst length distribution.

4.1.2 Estimation of Loss Impairment after FEC Reconstruction

As explained in Section 4.1, the *N*-packet FEC technique sends additional information of the immediate previous packets within the current sending packet. The amount of redundant information (I), e.g., in bytes, of one packet must be less (or at most equal) than N times the

packet size (P) without redundancy, i.e., $I \leq N(P)$. Let us estimate the codec impairment (I_e) for these two encoding schemes:

1. Normal packets coded using G.711 and redundant information coded using G.729, as shown in Figure 25a.
2. Normal packet and redundant information both coded using G.729, as shown in Figure 25b.

a) b)

Figure 24: FEC encoding schemes: The first one combines normal frames, encoded using G.711 (light rectangle) and redundant frames (dark rectangle) encoded using G.729, while the second uses G.729 for both frames.

For the first case, if the PLR due to the network is r and the perceived PLR after the FEC block, then the impairment after this block depends on the original loss rate (r) and the percent of reconstructed packets ($r - r'$). Note that this function $I_{e,G.711,G.729}(r, r')$ is neither equal to (77) nor to (78) and must satisfy the following conditions:

i. If $r' = r$ (i.e., no reconstructed packets):

$$I_{e,G.711,G.729}(r, r) = I_{e,G.711}(r) \tag{170}$$

ii. If $0 < r' < r$ (i.e., some packets are reconstructed):

$$I_{e,G.711,G.729}(r, r') > I_{e,G.711}(r') \tag{171}$$

iii. If $r' = 0$ (i.e., all packets are reconstructed):

$$I_{e,G.711,G.729}(r, 0) > I_{e,G.711}(0) \tag{172}$$

iv. For a fixed $r > 0$, it must be a non-decreasing function of r'.

According to (171), for a fixed value of r', the impairment increases as the percent of reconstructed packets is greater. Based on these properties, the estimation of the codec impairment is proposed as:

$$I_{e,G.711,G.729}(r, r') = I_{e,G.711}(r') + (r - r')\Delta_e \tag{173}$$

where Δ_e is:

$$\Delta_e = I_{e,G.729}(r') - I_{e,G.711}(r') \tag{174}$$

This estimation (173) satisfies the properties described previously.

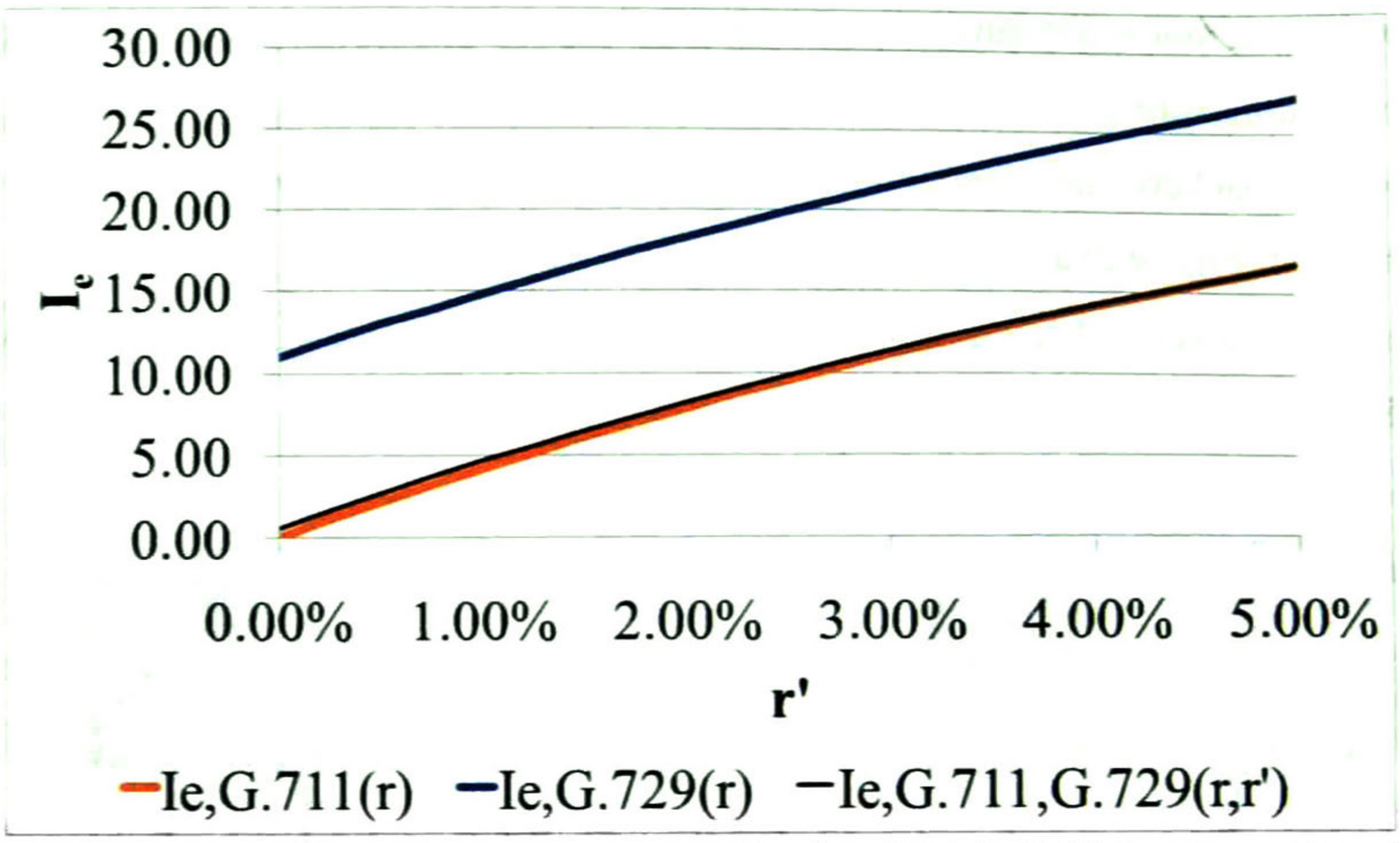


Figure 25. Estimation of the codec impairments (I_e) for G.711, G.729 and combined G.711-G.729. PLR before FEC reconstruction is 5%.

Figure 25 shows the proposed model (173), compared with the estimations for codecs G.711 and G.729. Note that the percent of reconstructed packets ($r - r'$) depends on the level of redundancy (N) and the burst length distribution ($f_b(k)$).

For the second codification scheme (i.e., using G.729 for normal packets and redundant information), the estimation of I_e is the same as (78), substituting r with r':

$$I_{e,G.729,G.729}(r, r') = I_{e,G.729}(r') \tag{175}$$

Note that this estimation is not including those additional losses caused by de-jitter buffer, which discard long delayed packets. To estimate the impairment after the de-jitter buffer block, the values of r and r' must be updated in the case of that any packet is discarded.

4.2 Impact of the De-jitter Buffer Size on the Quality of Service

The design of an adaptive optimal de-jitter buffer consists of determining a time window which is used to determine whether or not each packet is delivered to the player. This is, at each packet arrival the receive estimates its OWD, if this is within the time window, it is decoded and sent to the de-jitter buffer to wait there to be played; otherwise, it will be discarded. The de-jitter buffer is necessary to play the audio stream at constant rate but at the expense of an additional delay for all (or most) packets and, possibly, a slight increase in PLR.

Let $F_D(t)$ be the packet OWD distribution and d, the OWD for a certain packet. Then, the probability for this single packet to be discarded (p_d) is:

$$p_d = 1 - F_D(d) \tag{176}$$

Note that (176) considers the packet delay process as stationary; otherwise it would be time dependent.

Let r'' be the perceiver PLR after FEC and the receiver buffer, in this order. It is the sum of two probabilities: the probability of a packet to be lost due to network congestion and the probability that a packet successfully reaches the receiver and be discarded by the receiver buffer, then an approach to r'' is:

$$r'' = r' + (1 - r')[1 - F_D(w)] \tag{177}$$

where r' is the same than r_N' and defined by (161).

As the theoretical waiting time that minimizes the perceived PLR is ∞ (which implies that the packets would never be sent to the listener), the optimal value of w must consider the expected voice quality of the communication, e.g., by means of the E-Model's R-Factor and not considering PLR only.

4.3 Generation of Cauchy-distributed Time Series with Specific Hurst Index

A method to simulate Cauchy-type processes, that represent delay jitter series of a VoIP call, is proposed. This generator produces a random sequence of Cauchy-distributed observations whose correlation is determined by the self-similarity parameter (H). This proposed method shows that it is possible to synthesize artificial time series with both a specific distribution and Hurst index, even in the infinite variance case (a study of the finite-variance case is presented in [60].

The proposed algorithm for generating artificial Cauchy time series with defined size, location, scale and Hurst index consists of the following four stages:

1) Generate of a random sample of certain size, e.g., a uniformly distributed series or FGN.
2) Convert the random sample into a Cauchy-distributed series, with a specific location and scale, by means of the ICDF transformation (see Section 1.2).
3) Adjust the Hurst index of the series by means of the weighted synthesis described in Section 2.3.
4) Adjust the location and scale parameters by means of the linear transformations (42) and (43), which do not alter the estimation of the Hurst index.

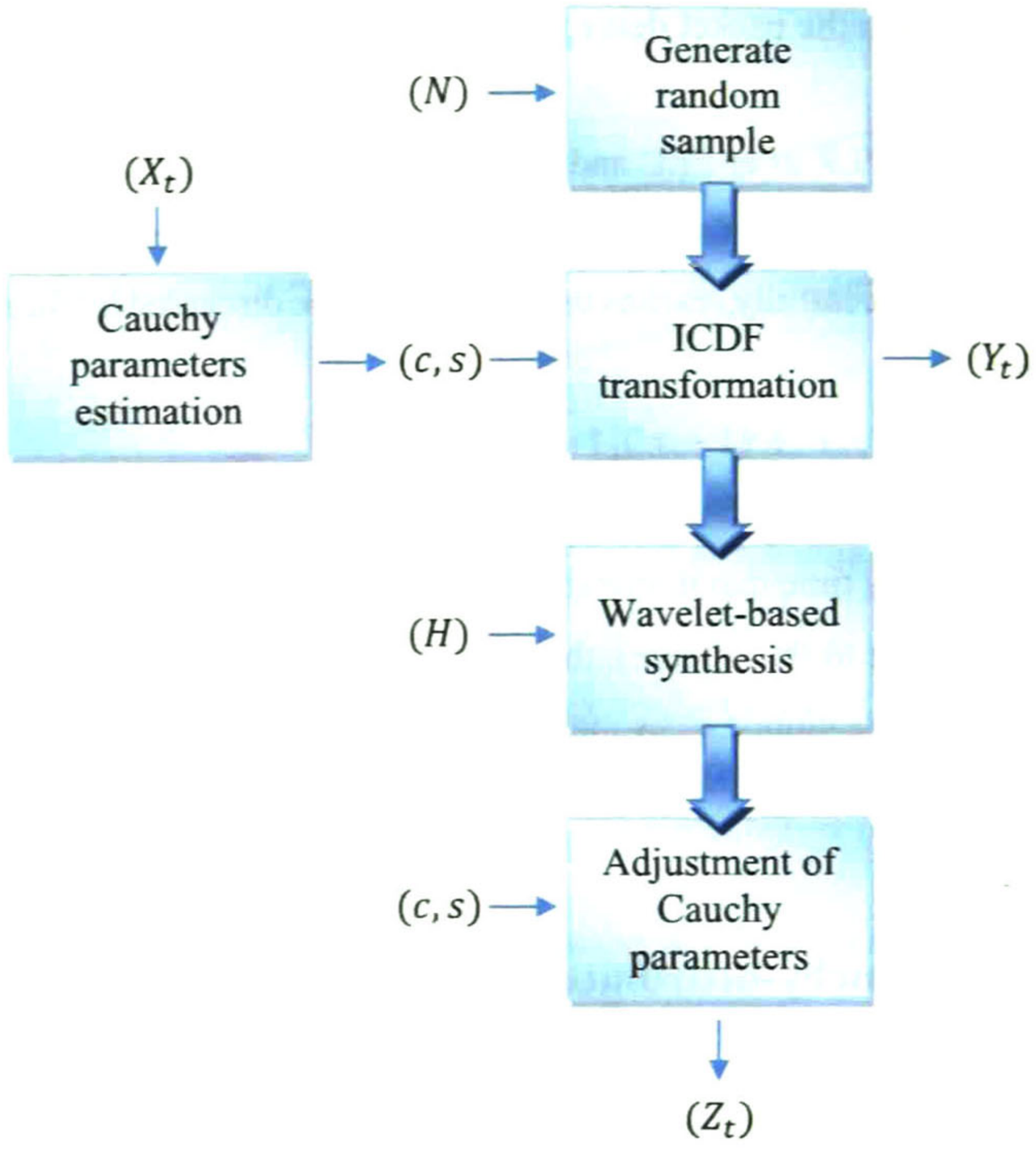


Figure 26: Algorithm to generate Cauchy-distributed time series with Hurst index *H*

Figure 26 shows the proposed algorithm, where X_t represents a measured jitter sample, e.g., from a monitored VoIP call, from which the location and scale parameter are estimated, and Z_t is the output. Y_t is the resulting series of the ICDF transformation alone.

Note that the wavelet-based synthesis is not a linear transformation (it is only for multiplication by a real constant). Consequently, the output (Z_t) of the generator may not be exactly Cauchy-distributed, but it is actually very close if the Hurst index adjustment is not abrupt.

4.4 Generation and Modeling of Packet Loss Sequences

Let us define the loss sequence as follows:

$$Y_n = \begin{cases} 0; & \text{if packet } n \text{ is received} \\ 1; & \text{if packet } n \text{ is lost} \end{cases} \quad (178)$$

From the loss sequence, the probabilities of transitions were also estimated using the algorithms explained in sections 4.4.2 and 4.4.3, for the two-state and the four-state models, respectively.

4.4.1 Artificial Loss Sequence Generation

A simply algorithm is used in this work to simulate packet losses modeled by discrete Markov processes. The algorithm for the two-state Markov process sequence is shown in Table 3. Algorithms to generate sequences based on processes with greater number of states are just extensions. For this, let X_t be a generated uniformly distributed sequence such that $X_t \in [0,1]$ for $t = 1, \ldots, N$. The values of p_{12} and p_{21} are set. The output sequence Y_t represents packet receipt and losses by zeros and ones, respectively for $t = 1, \ldots, T$.

```
state = 0;
FOR t = 1 TO T,
   IF (state),
      IF (X[t] < p12)
         state = 0;
         Y[t] = 0;
      ELSE
         Y[t] = 1;
      END IF;
   ELSE
      IF (X[t] < p21)
         Y[t] = 1;
         state = 1;
      ELSE
         Y[i] = 0;
      END IF;
   END IF;
END FOR;
```

Table 3: Algorithm for two-state Markov process sequence generation

4.4.2 Two-state Parameters Estimation

The estimations of p_{12} and p_{21} are: $p_{12} = t_{c\to e}/n_1$ and $p_{21} = t_{e\to c}/n_0$, where $t_{c\to e}$ and $t_{e\to c}$ are the respective number of transitions from correct states to error states (i.e., when $Y_k = 0$ and $Y_{k+1} = 1$) and from error states to correct states (i.e., when $Y_k = 1$ and $Y_{k+1} = 0$), and n_0 and n_1 are the respective number of received and lost packets (i.e., the respective numbers of zeros and ones of Y_k).

4.4.3 Four-state Parameters Estimation

In the *4*-state case, the values of the sequence Y_t are divided into two regions: the first one with lower loss rate (whose first and last values are zeros) and the second one with higher loss rate (whose first and last values are ones) than certain threshold, e.g. 1%. Then, from the first region, p_{12} and p_{21} are estimated as explained in section 4.4.2. Similarly, p_{43} and p_{34} are estimated from the second region. Finally, let $t_{1st\to 2nd}$ be the number of transitions from the first region to the second; $t_{2nd\to 1st}$, the number of transitions from the second to the first; n_{1st}, the number of received packets in the first region (zeros) and n_{2nd}, the number of lost packets in the second region (ones), then $p_{23} = t_{1st\to 2nd}/n_{1st}$ and $p_{32} = t_{2nd\to 1st}/n_{2nd}$.

4.5 Parameter-optimizable Quality of the VoIP Communication

The maximum value for the MOS, which indicates the maximum quality of the communication, is achieved when the *R* factor is also maximized. The strategy is then to set the adjust parameters to their respective optimal values, e.g., the redundancy level (N), the de-jitter buffer size (w), the type of codec (in this case, G.711 or G.729) and the voice data length (inter-departure time or IDT). Many of these adjust parameters are easier to optimize than the others, as they are independent (or almost) of the others. E.g., the level of redundancy depends only on the PLR; it is increased as the PLR become greater than certain thresholds to decrease it to acceptable levels.

The E-model's *R* factor can be estimated then in terms of these adjust parameters as follows: As it is defined by (73), it is the sum of the impairments due to coded-PLR (I_e) and those due to mouth-to-ear delay (I_d). The first one depends on the types of codec used, as described in section 1.7, and the type and level of redundancy, as the perceived PLR decreases by packet reconstruction. The second one depends of the mouth-to-ear delay, which can be expressed as the sum of two delays, i.e.,

$$d = w + \delta \tag{179}$$

where w is the delay caused by the de-jitter buffer (which is equal to its size) and δ is the sum of all other delays, e.g., packet transmission, queuing, coding/decoding, etc. The delay impairment is then estimated as expressed by (74) for $d = w + \delta$. The value of d for received packets must be a constant. Then, w is a random variable that depends on δ (i.e., $w = d - \delta$).

Finally, the E-model's *R* factor is estimated as a function of the following parameters:

a) Communication-intrinsic parameters: the network PLR (r), the burst length distribution ($f_b(k)$) and the delay due to network and the devices except the de-jitter buffer.

b) Adjust parameters: the codec used, the level of redundancy (N), and the de-jitter buffer size w.

A methodology to obtain the maximum quality is to estimate the communication-intrinsic parameters and then to find the optimum adjust parameters that maximize the estimation of the E-model's R factor, and consequently, the MOS. This analysis can be extended if other improvement techniques are implemented, e.g., bit-level error correction [61].

4.6 Chapter Summary

In this chapter, modeling and characterization of packet loss and delay for a VoIP communication is presented.

The performance of the communication is measured by means of the E-model's R factor, which can be converted to MOS. Packet reception and loss is modeled by finite-state Markov processes. An innovative contribution of this study corresponds to the models based on two-state and four-state Markov processes: the equations for theoretical gap and burst length distributions, as a function of the probabilities of transitions for both models are proposed. The strategy used to obtain the gap and burst length distributions for the four-state model presented in Section 1.3.3 exemplifies the generalized methodology for a m-state Markov process model, which consists of finding firstly their respective CDF, i.e., $C_b(k)$ and $C_g(k)$.

Algorithms for reconstructions, i.e., estimation of the probabilities of transitions between states for two-state and four-state models, are also described.

It is shown, through an evaluation based on SMSE, that both two-state (at least for most cases) and four-state models can capture the geometric-type decay of the distribution of the burst length, but the two-state model fails to model the gap length distribution when non-homogeneous losses are present. I.e., the gap length distribution is the sum of two geometrical series with different decaying ratio, as defined by (64), not only one, as defined by (55).

An analysis of the N-packet FEC scheme is also presented. The expected perceived PLR obtained with this correction scheme is quantified, as expresses (161). This resulting general formula applies for the N-packet FEC scheme, regardless of the shape of the burst length distribution.

Through an example using real world samples, it is shown that 1-packet FEC is generally sufficient to improve the quality of the communication to an acceptable level, that is, with a PLR lower than 1%.

The equations for the estimation of the impairments for combined codification, i.e., when normal packets and redundant information may be coded differently, are proposed. These

estimations satisfy certain properties that result of the combination of two different codification schemes, as described in Section 4.1.2.

This chapter is based on publications [62], [63].

5 Simulations and Performance Evaluation

5.1 Comparison of the Hurst Index Estimators with Synthetic Traces

5.1.1 Evaluation Metrics and Methodology

A criterion to evaluate the performance of the H estimators is needed. The conclusions are based on the SMSE, which depends on the bias and the standard deviation of the estimations (see Definition 5).

Definition 2: Average estimated H: $\hat{\bar{H}} = E\{\hat{H}\}$

Definition 3: Bias: $\hat{\delta} = \hat{\bar{H}} - H$

Definition 4: Standard deviation: $\hat{\sigma}_H = \sqrt{E\left\{\left(\hat{H} - \hat{\bar{H}}\right)^2\right\}}$

Definition 5: Square root of mean squared error (SMSE): $\hat{\varepsilon} = \sqrt{E\left\{\left(\hat{H} - H\right)^2\right\}} = \sqrt{\hat{\delta}^2 + \hat{\sigma}_H{}^2}$

The experiment 1 consisted of:

1. The generation of a set of 2^{22-L} traces of *fractional Gaussian noise* (FGN) of length 2^L for $L = 10, \ldots, 16$ and Hurst parameter $H = 0.50, 0.60, 0.70, 0.80, 0.90$ by using Paxson's algorithm, see [64].
2. Estimation of H for each generated trace using the estimators described in Section 1.1.6; these are labeled, respectively, R/S, AM, VAR, PER, MAVAR, LWHI, and WAV. For each pair of values of L and H (theoretical), the average H, standard deviation and SMSE were calculated. The implementation is that used in the tool developed by J. C. Ramirez et al [65] [66].

Important note: VAR is the classical variance-based estimator, i.e., the variance of the aggregated series is calculated using formula (134), and not (139). It is expected to be negatively biased.

5.1.2 Results

The results of the calculations are shown from Figure 27 to Figure 33.

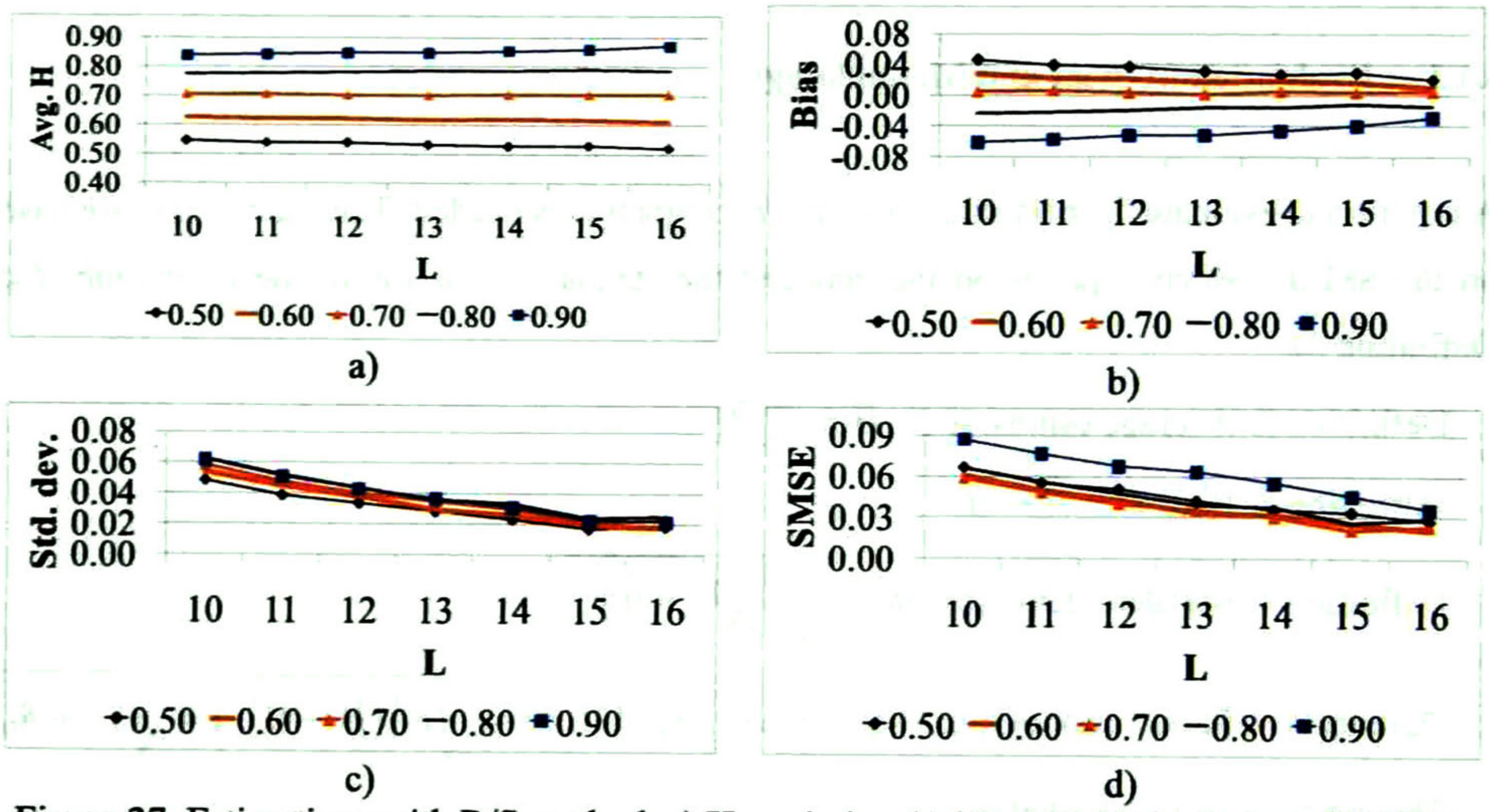


Figure 27. Estimations with R/S method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.

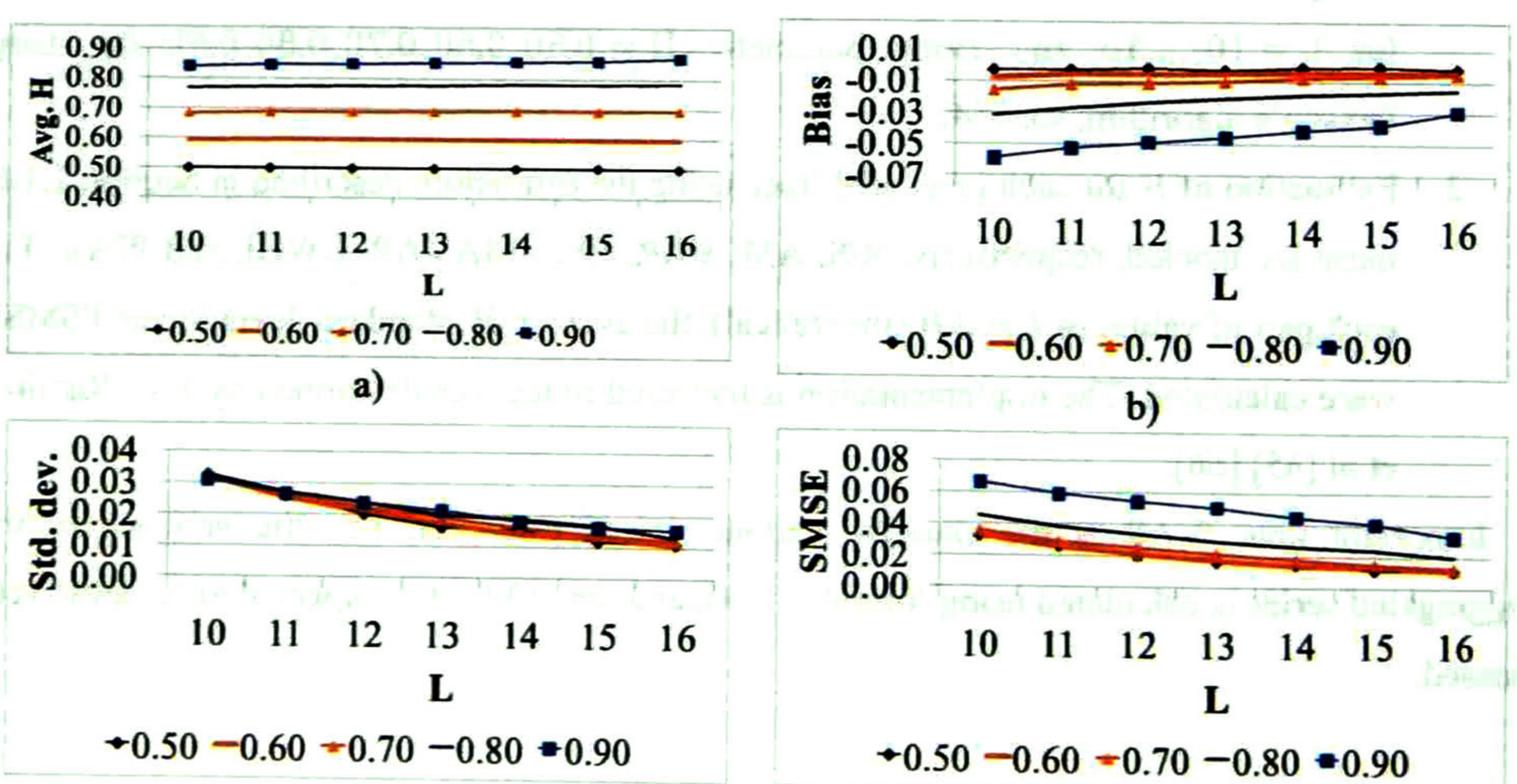


Figure 28. Estimations with AM method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.

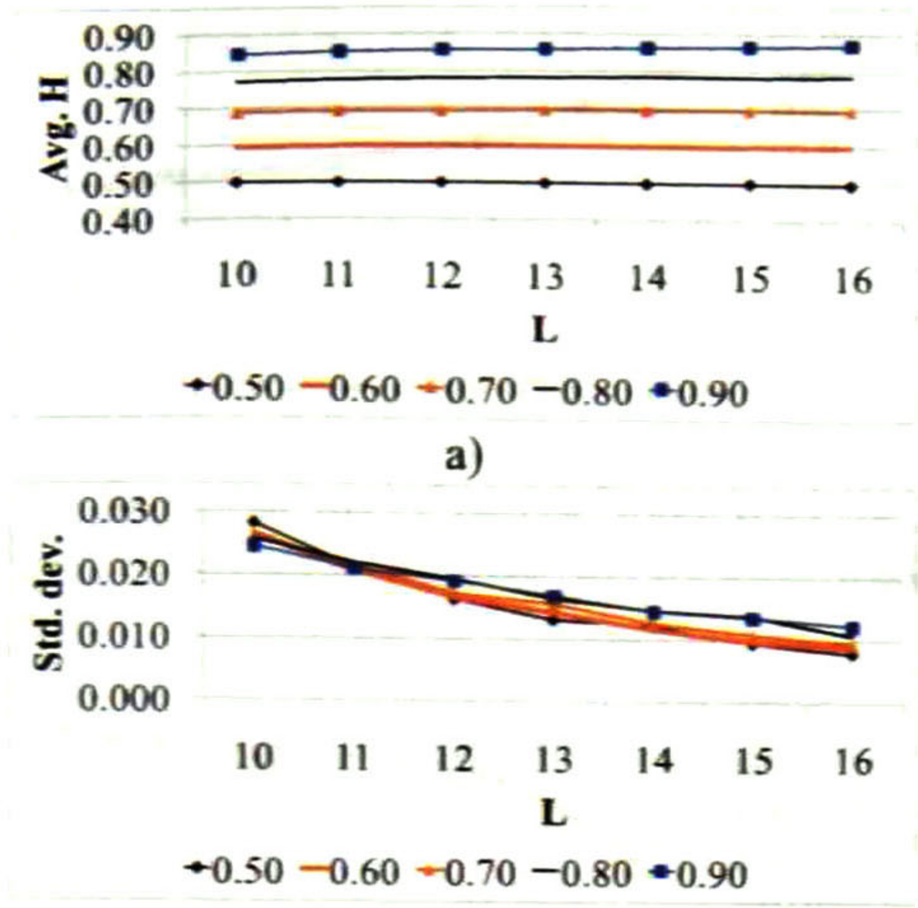


Figure 29. Estimations with VAR method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.

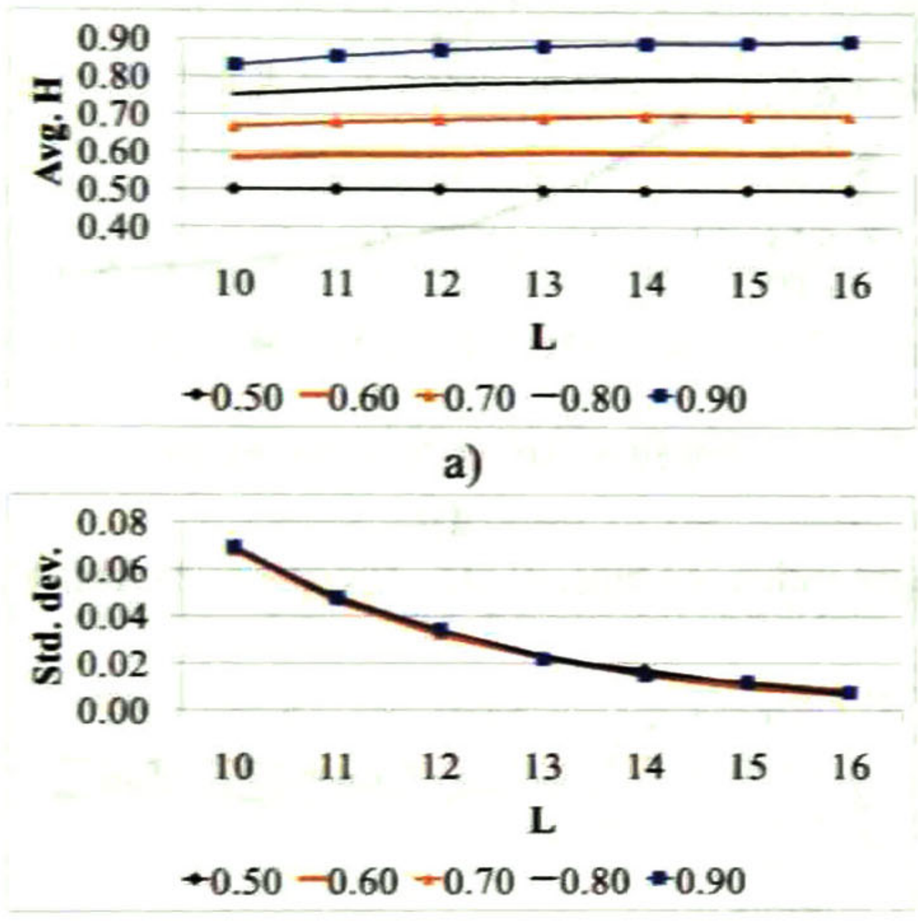
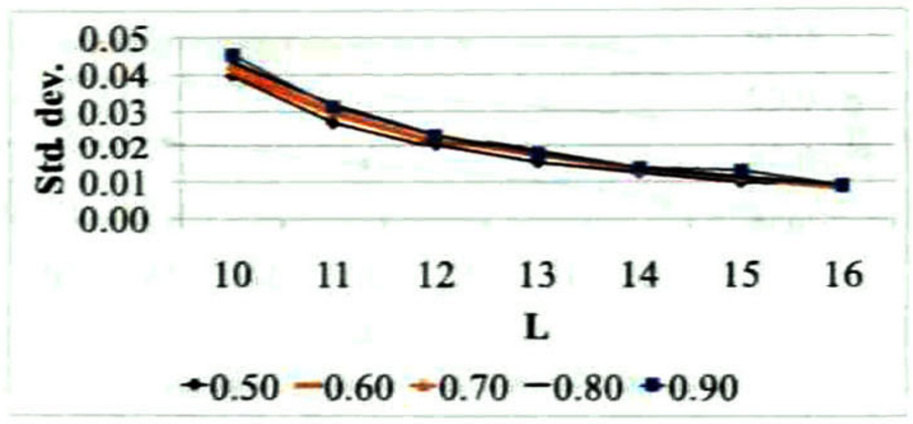
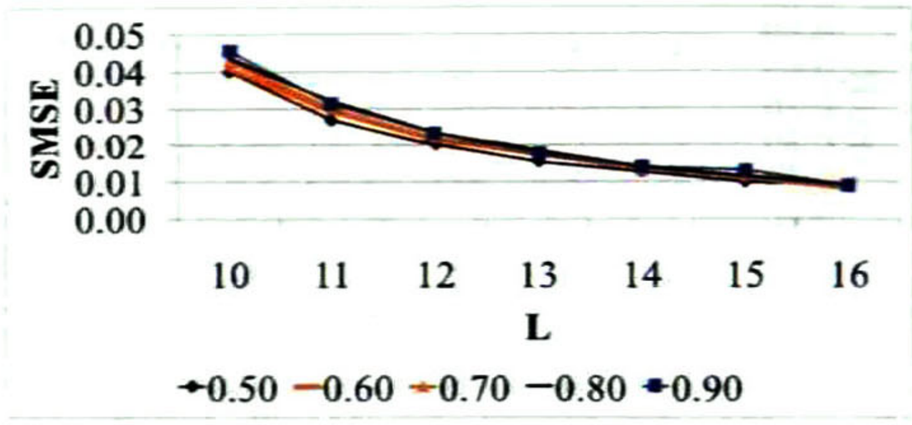


Figure 30. Estimations with PER method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.

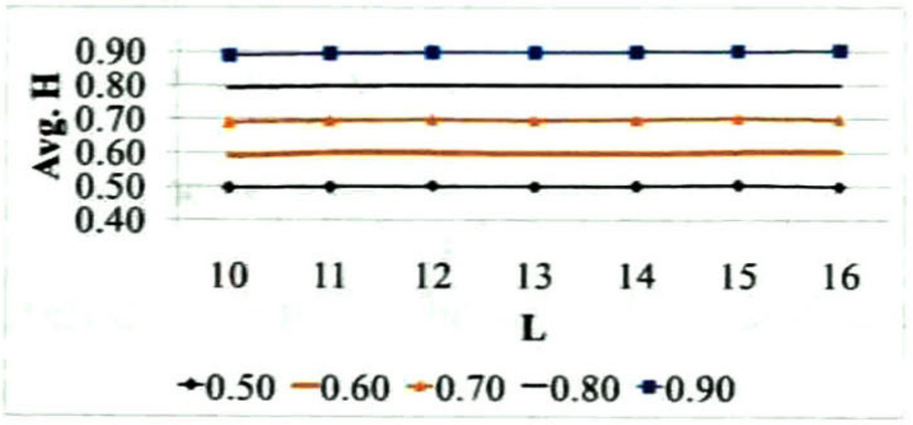


c)

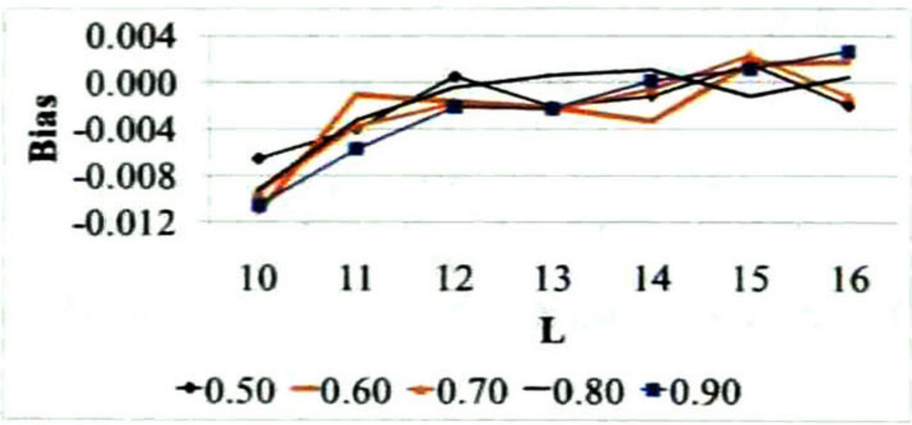


d)

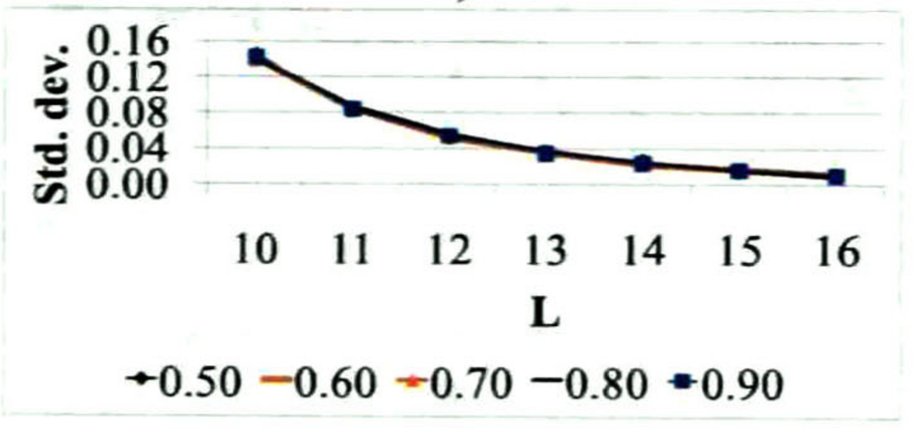
Figure 31. Estimations with MAVAR method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.



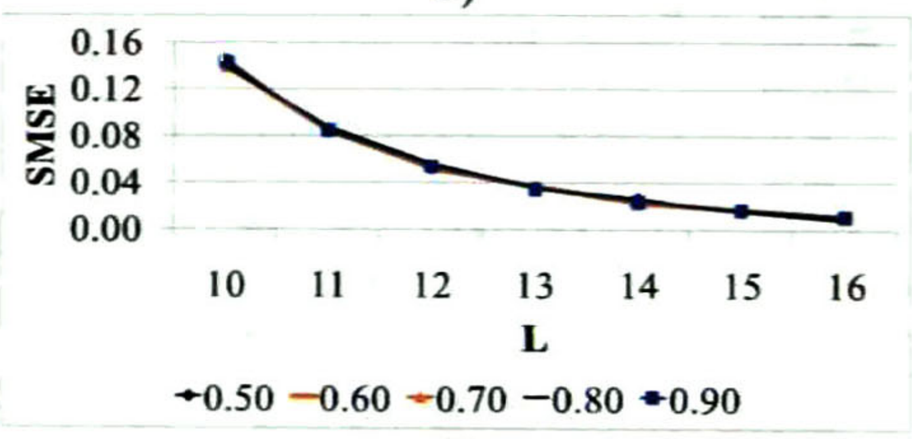
a)



b)

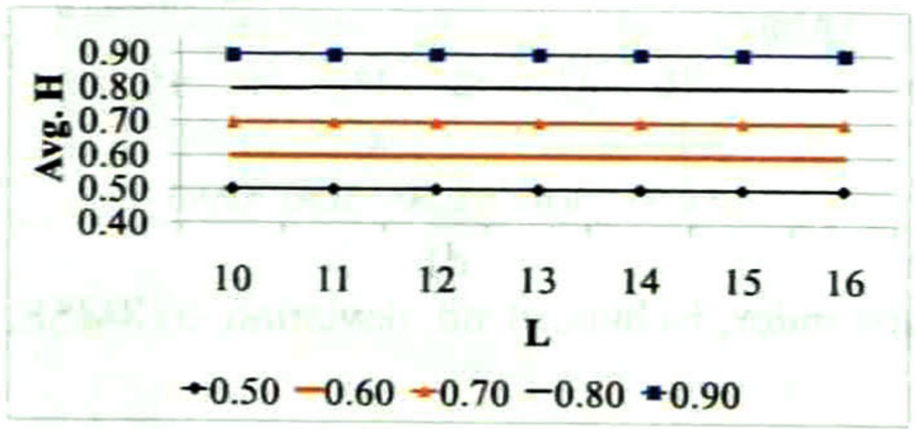


c)

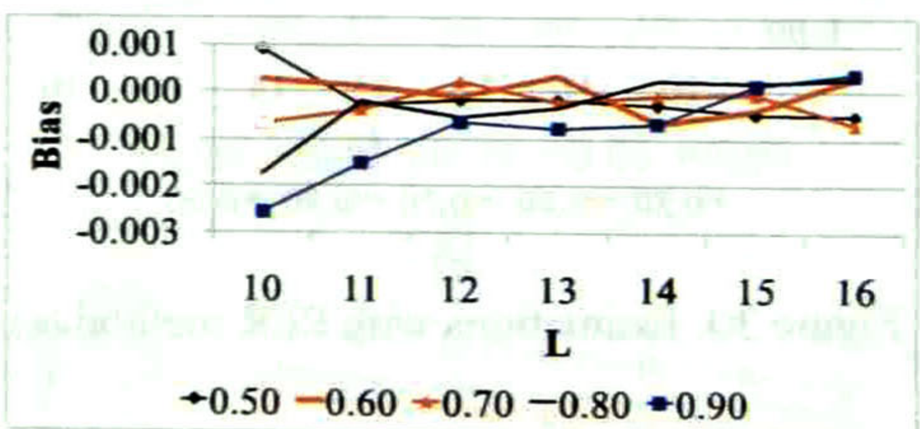


d)

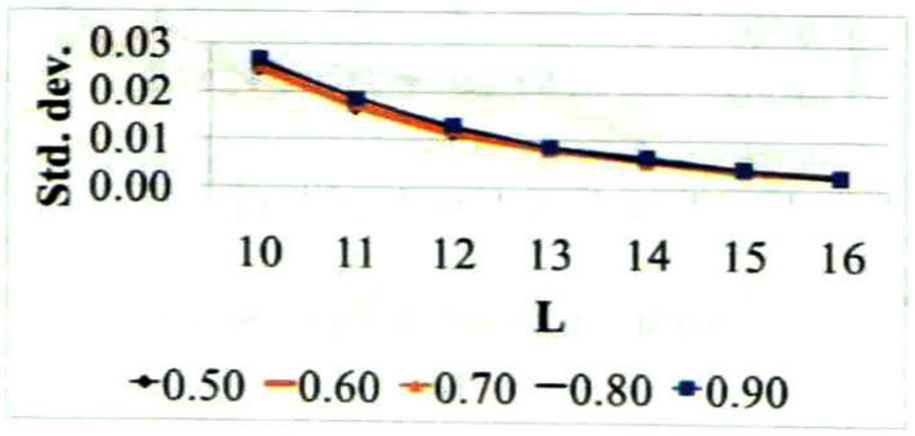
Figure 32. Estimations with LWHI method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.



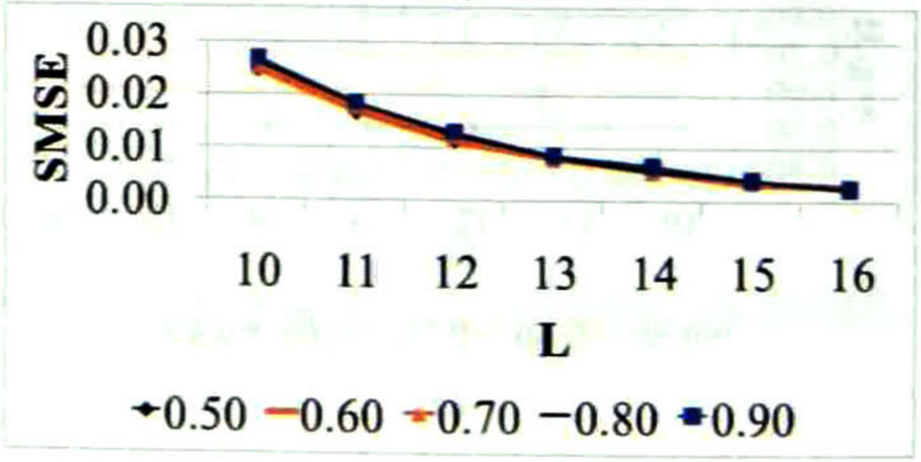
a)



b)



c)



d)

Figure 33. Estimations with WAV method. a) Hurst index, b) bias, c) std. deviation, d) SMSE.

The main findings are:

- For R/S, AM, VAR and PER, the shorter the trace and the larger theoretical H, the worse their bias (see Figure 27b, Figure 28b, Figure 29b and Figure 30b), while for MAVAR, LWHI and WAV $|\hat{\delta}| < 0.01$ in almost all cases (the exceptions are a few cases for LWHI when $L = 10$) and their bias is almost constant for all values of H in the range from 0.50 to 0.90 (see Figure 31b, Figure 32b and Figure 33b).

  The bias of all estimators except R/S was, in most cases, negative valued. R/S statistic presented positive bias for those traces with $H \leq 0.70$ and negative bias for $H > 0.70$, as shown in Figure 27b.

  Although the estimations obtained by LWHI presented the largest standard deviation (as shown in Figure 32c), the behavior of this estimator is very similar for all values of H at a certain length, as well as PER, MAVAR and WAV (see Figure 30c, Figure 31c and Figure 33c).

  The standard deviation of R/S, AM and VAR vary for different values of H, but, except for R/S, it is smaller than 0.035 for all lengths from 1024 to 65536, as shown in Figure 27c, Figure 28c, and Figure 29c.

  The smallest SMSE was obtained by WAV estimator, followed by VAR, for $H \leq 0.70$ and MAVAR, for $H > 0.70$.The average bias, average standard deviation and average SMSE of the estimations vs. theoretical H of all seven estimators are shown in Figure 34, Figure 35 and Figure 36, respectively.

The averages bias, standard deviation and SMSE are shown, respectively in Figure 34, Figure 35 and Figure 36. Figure 37 shows the average standard deviation vs. average bias plot. The behavior of the estimators for different values on H in the range between 0.50 and 0.90 can be observed.

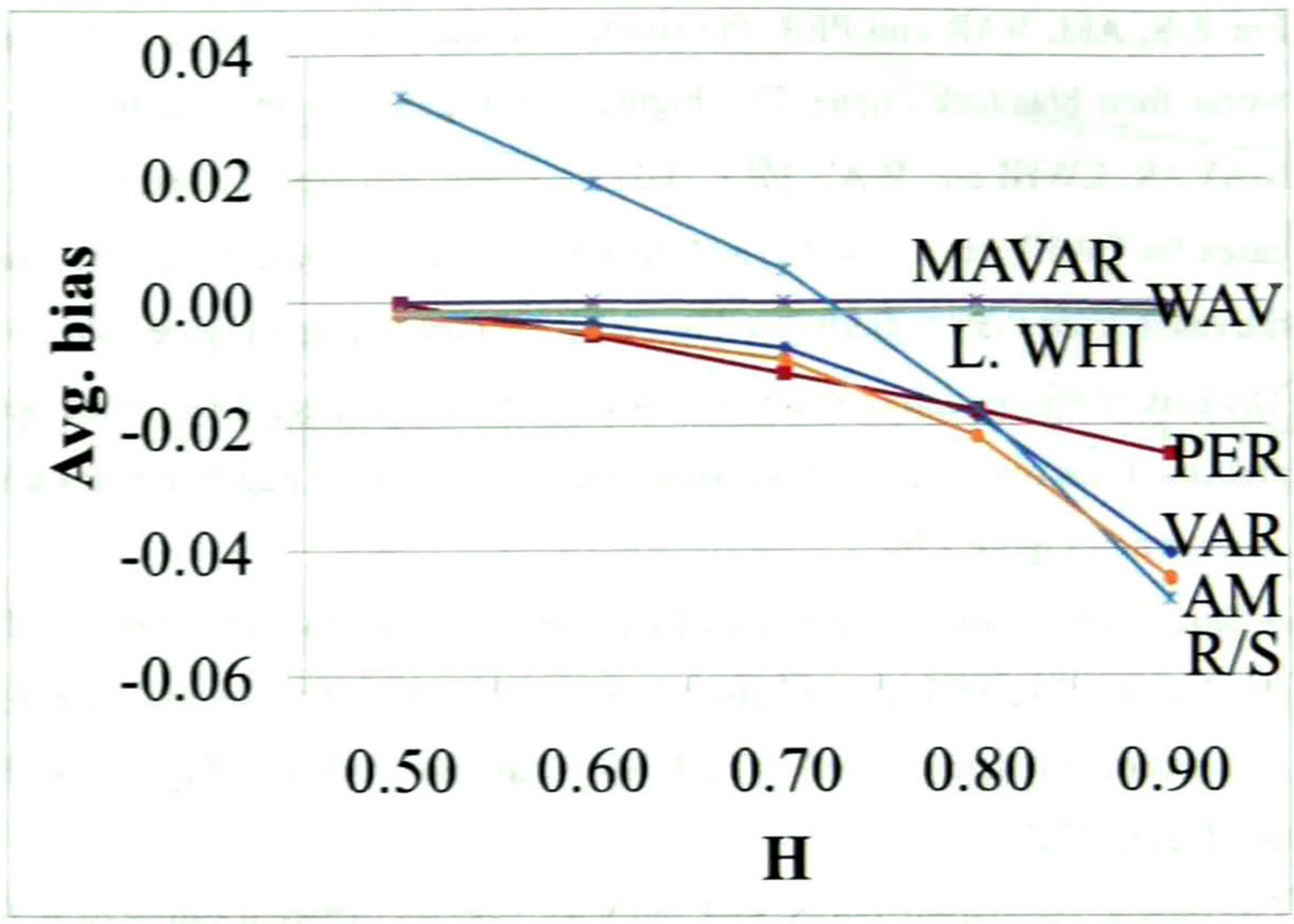


Figure 34. Average bias of the Hurst index estimations.

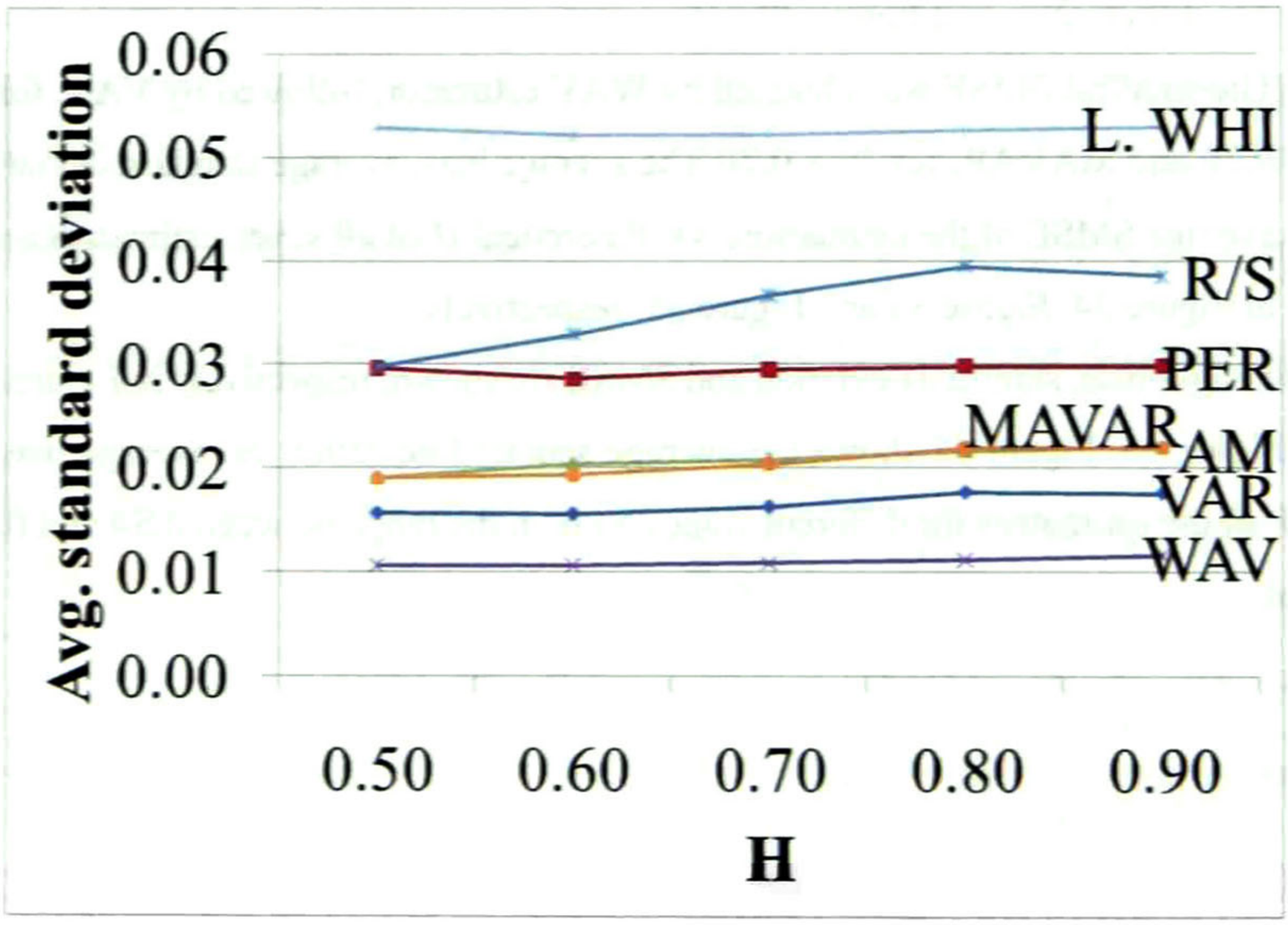


Figure 35. Average standard deviation of the Hurst index estimations.

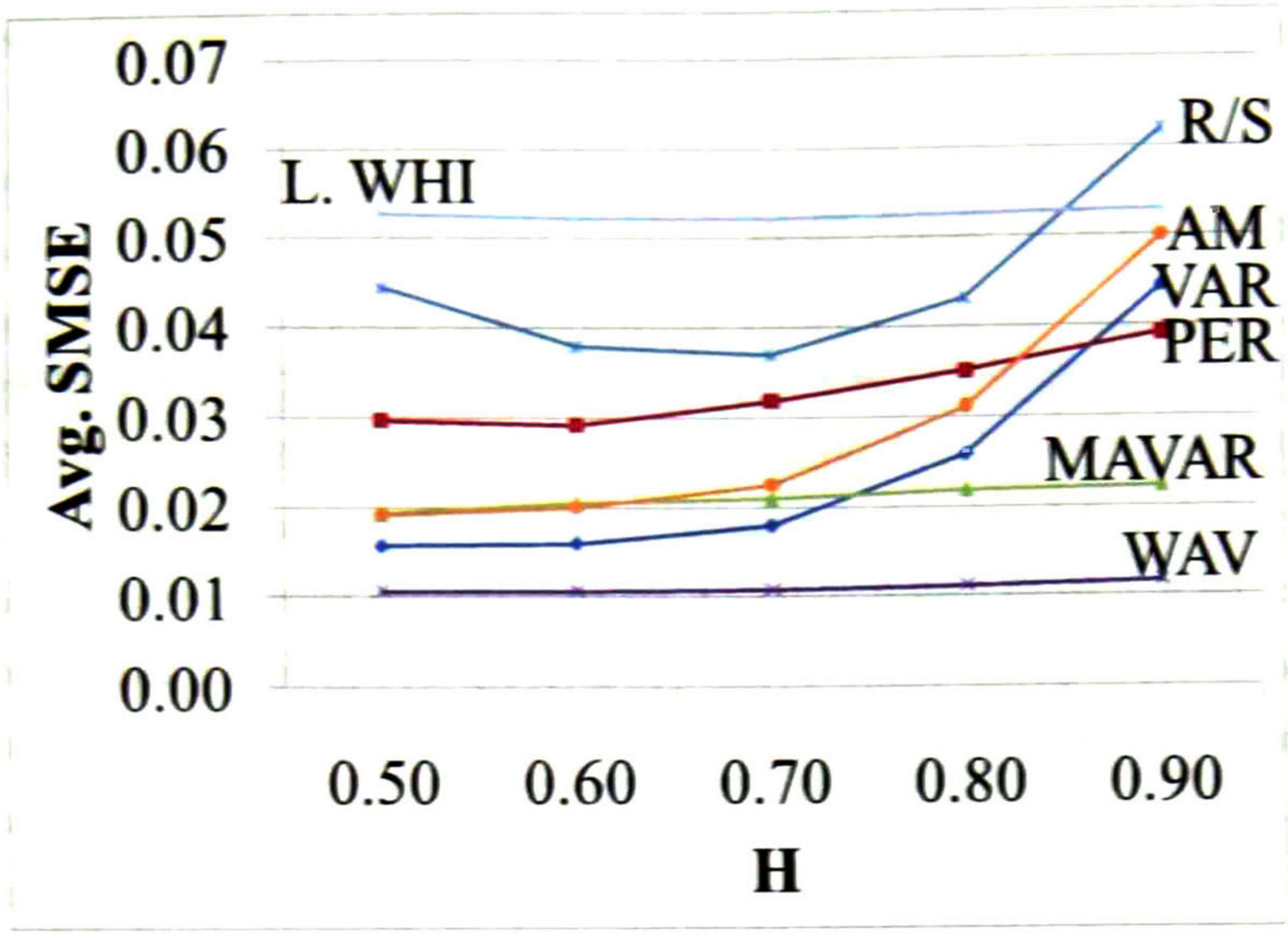


Figure 36. Average SMSE of the Hurst index estimations.

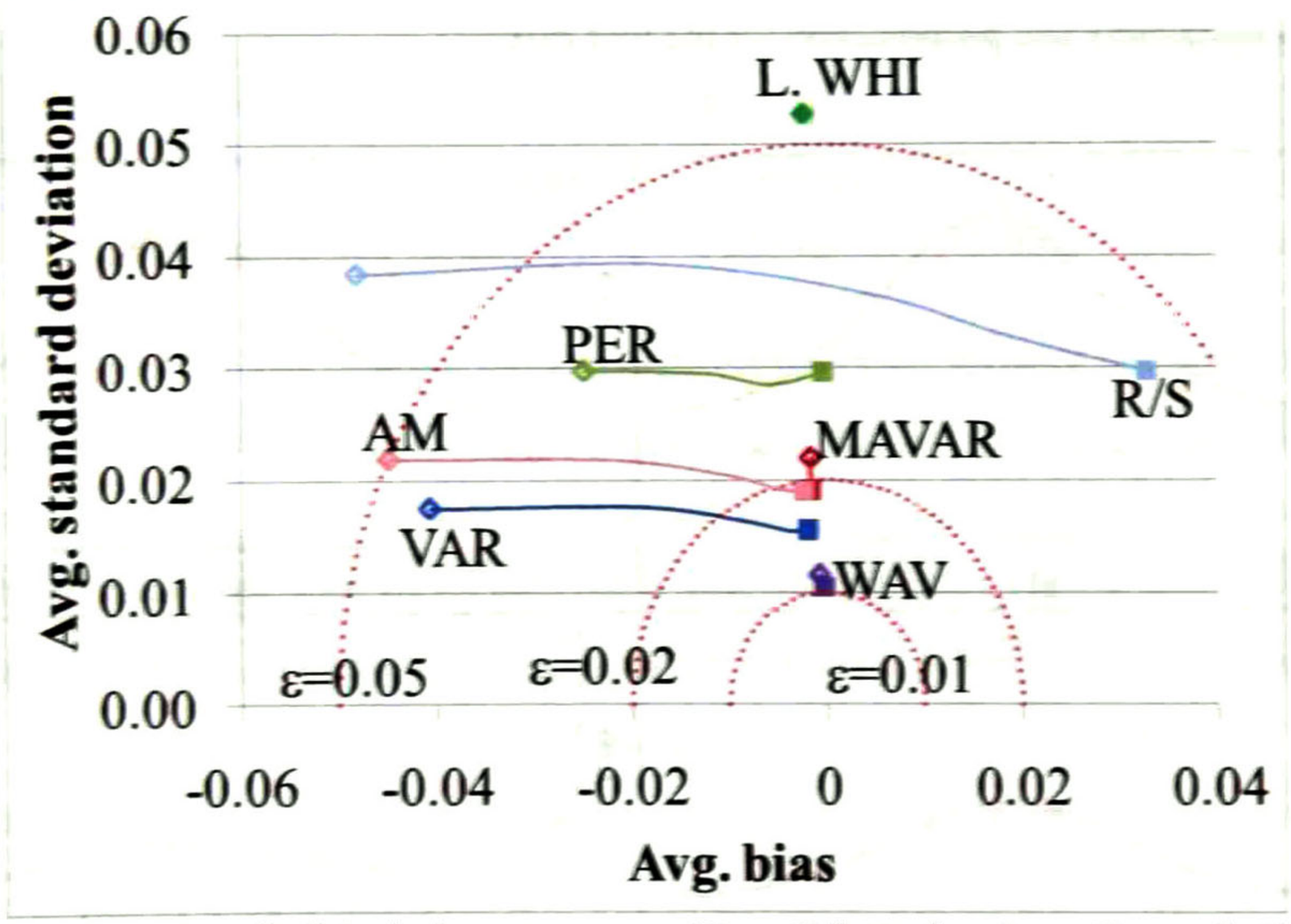


Figure 37. Average standard deviation vs. average bias of the estimations. Square indicates $H = 0.50$; diamond, $H = 0.90$

As Figure 37 shows, although the estimations of WAV did not presented the lowest average bias (the lowest are MAVAR's), they are smaller than 0.012 for H between 0.50 and 0.90. LWHI presents almost the same behavior for different values of H but the SMSE of its estimations was greater than 0.05, due to its large standard deviation ($\hat{\sigma}_H > 0.05$)

5.2 Sample Mean and Variance Estimation

5.2.1 Estimation of the Sample Mean

In order to verify the equations that describe the mean and variance of the sample mean, a set of zero-mean, unitary variance, FGN time series of size $N_p = 10^6$ observations are generated using an implementation of the generator proposed by Davies and Harte in [67], each for a different Hurst index for $H = \{0.30, 0.50, 0.70, 0.90\}$. Note that although several estimators are available in the literature, even for the multivariate case, e.g., [68], this algorithm is chosen because it is easy to implement and because it has a complexity of $O(n\ log\ n)$ .Then, the mean is estimated from blocks of size $N = 100$ and the empirical PDF is obtained from the estimations and compared to the classical (132) and proposed (131) estimators. Figure 38a shows that the variance of the estimated mean does not fits the classical model when SRD or LRD is present (Figure 38a, c, and d), but only for the uncorrelated case (Figure 38b). Only proposed estimator (131) represents adequately this phenomenon for the four cases.

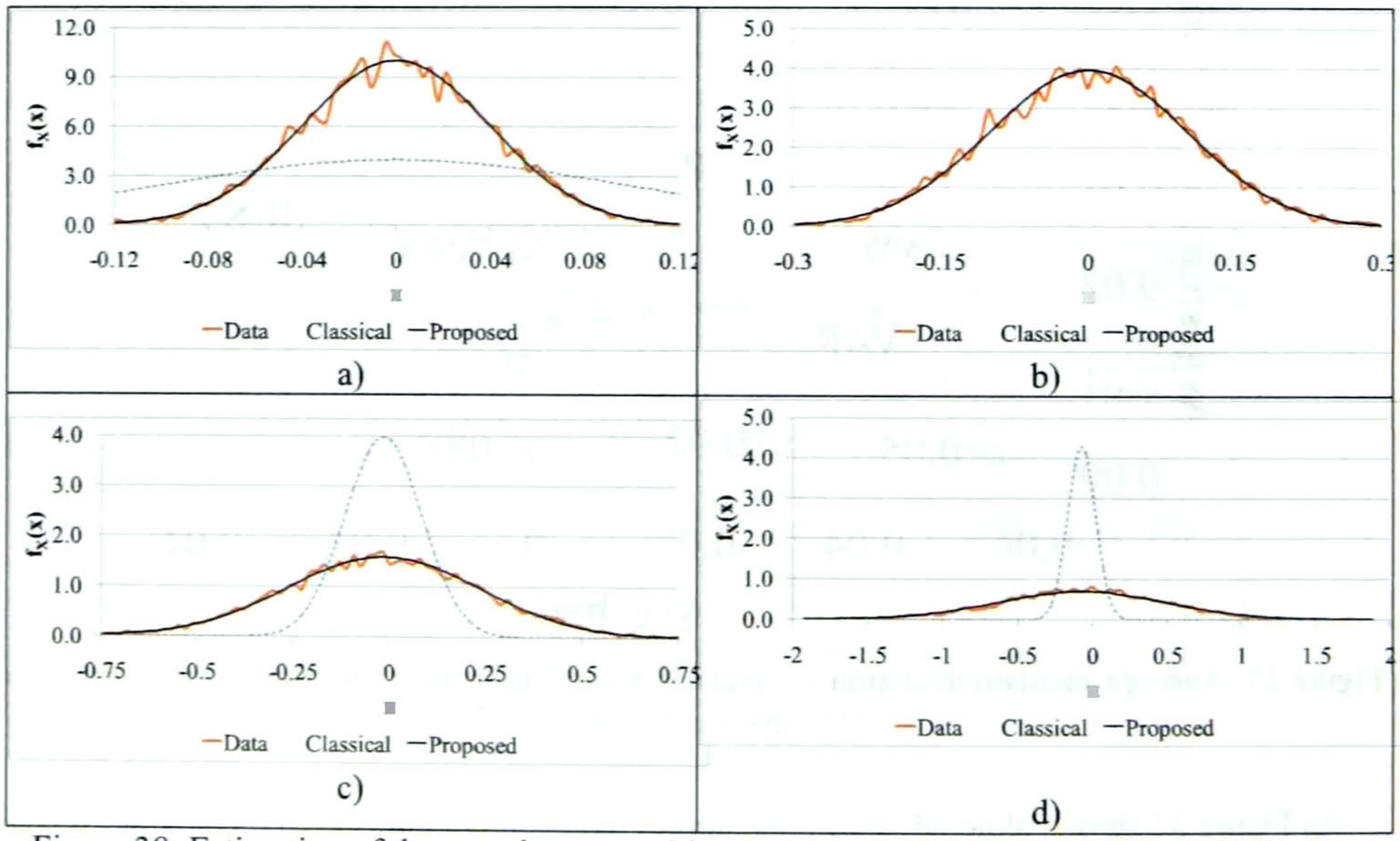


Figure 38: Estimation of the sample mean with $n = 100$ for four cases: a) $H = 0.30$, b) $H = 0.50$, c) $H = 0.70$ and d) $H = 0.90$.

5.2.2 Estimation of the Sample Variance

The followed procedure to verify the proposed estimator of the sample variance ($\hat{\sigma}_X^2$) consists of the generation of a set of 100 FGN samples of size 1024 for each value of $H = \{0.05, 0.10, 0.15, \dots, 0.95\}$ and the estimation of the variance using the classical formula (134) and the proposed estimator (139). The respective mean of both estimations for each set was obtained. For the estimation of the Hurst index the wavelet estimator of Veitch and Abry [11] is used.

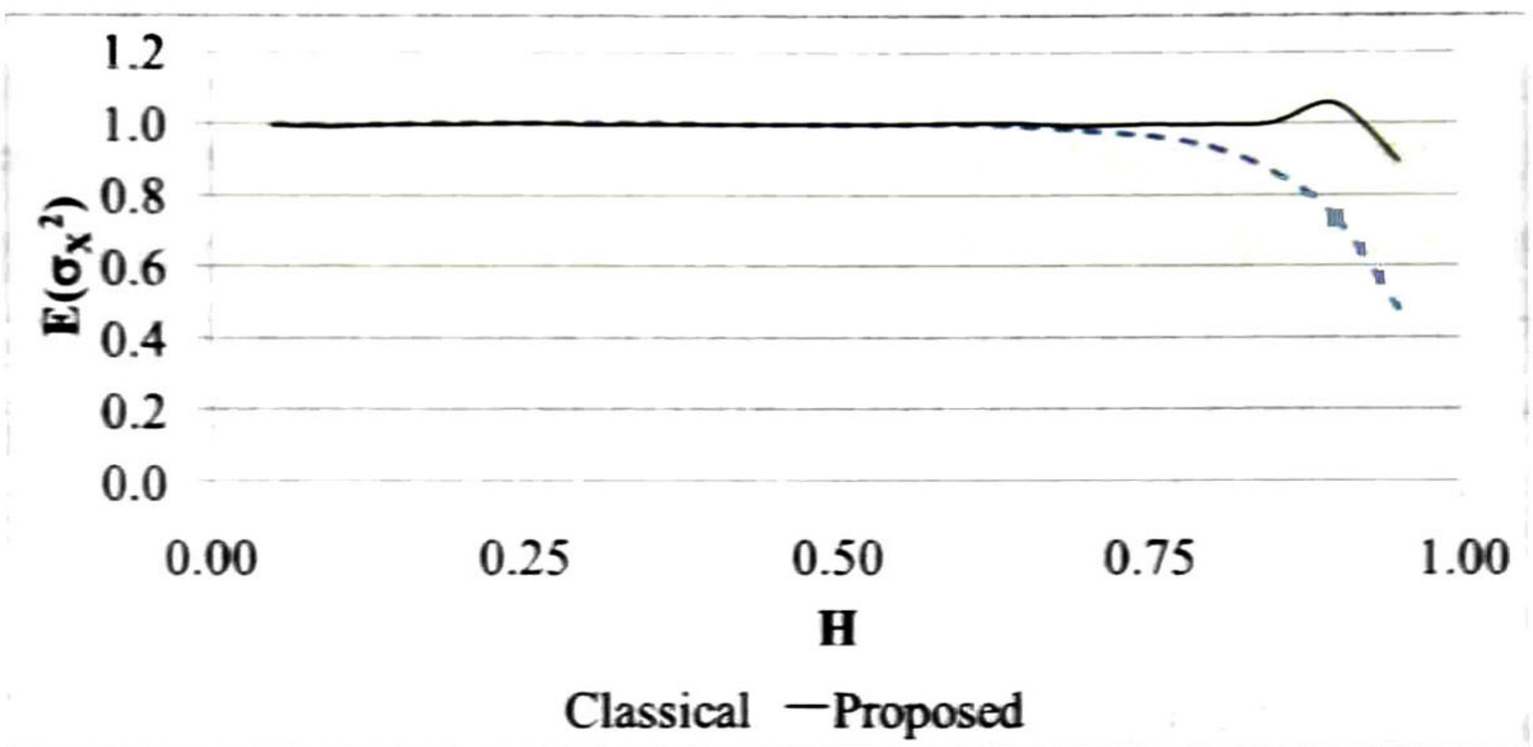


Figure 39: Mean of the estimated variance for $n = 1024$ and $H = \{0.05, 0.10, 0.15, \dots, 0.95\}$.

Figure 39 shows that the classical formula underestimates the variance noticeably for higher values of H. For $H > 0.95$ the estimated variance is less than the half of the process variance. The proposed formula (139) does not underestimate the variance, but for high values of H the mean of the estimated variance varies noticeable from one realization to another. This variation results from the estimation of the Hurst index, as the statistic $1 - N^{2\hat{H}-2}$ is very sensitive to the variations of $\hat{H}$, which depends, in turn, of the efficiency of the sample generator. This is an indicator then that the generator proposed by Davies and Harte may be less accurate as H is closer to 1.

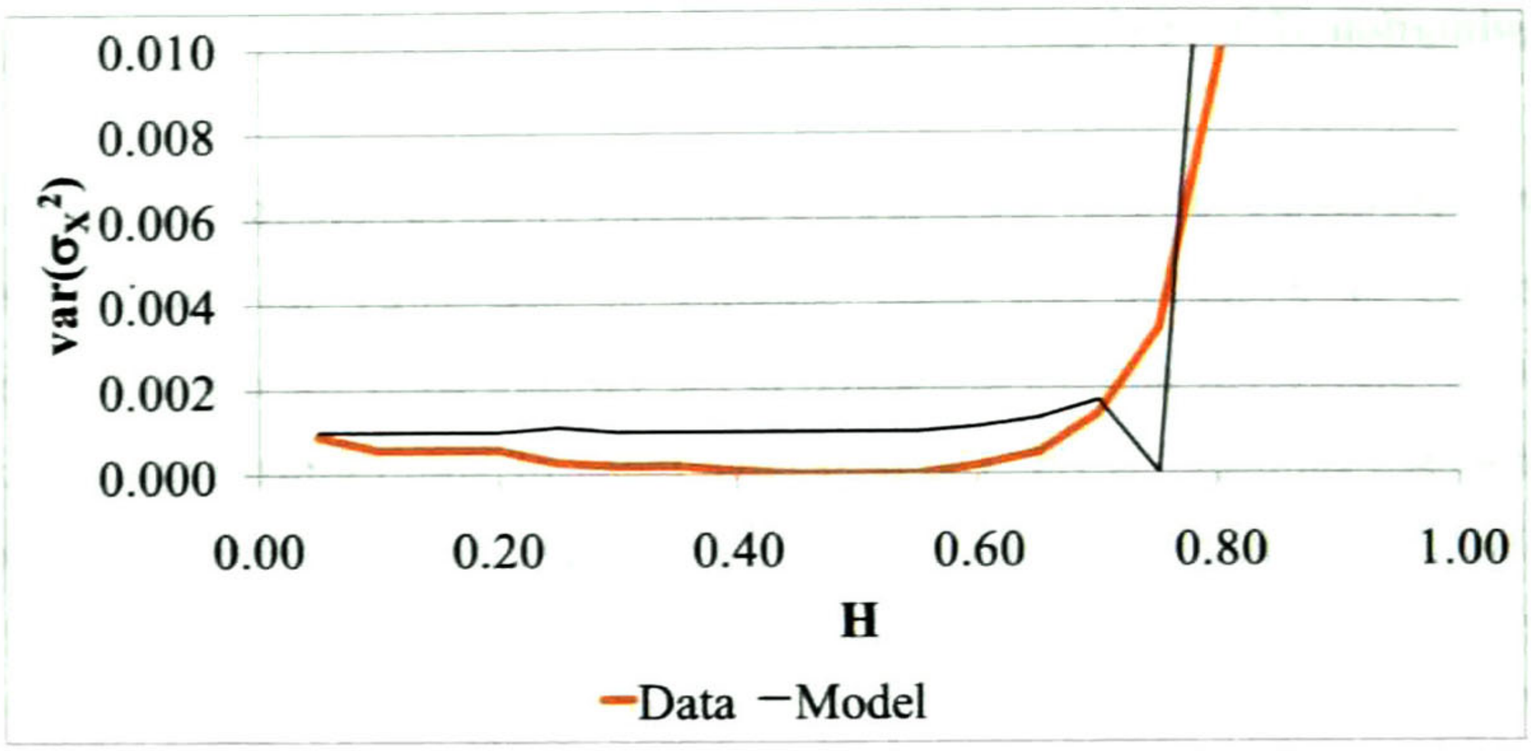
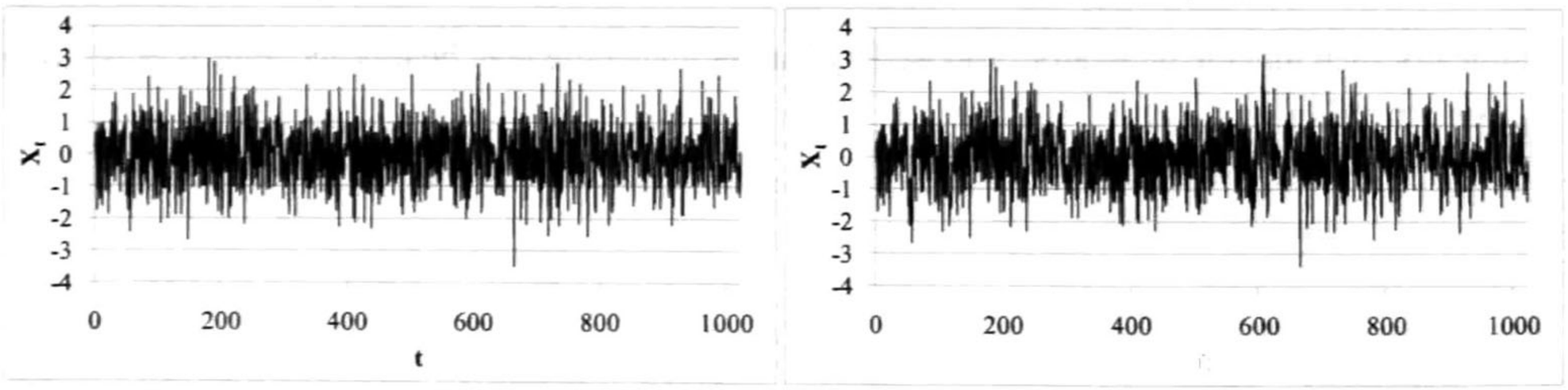


Figure 40: Variance of the estimated variance for $n = 1024$ and $H = \{0.05, 0.10, 0.15, \ldots, 0.95\}$.

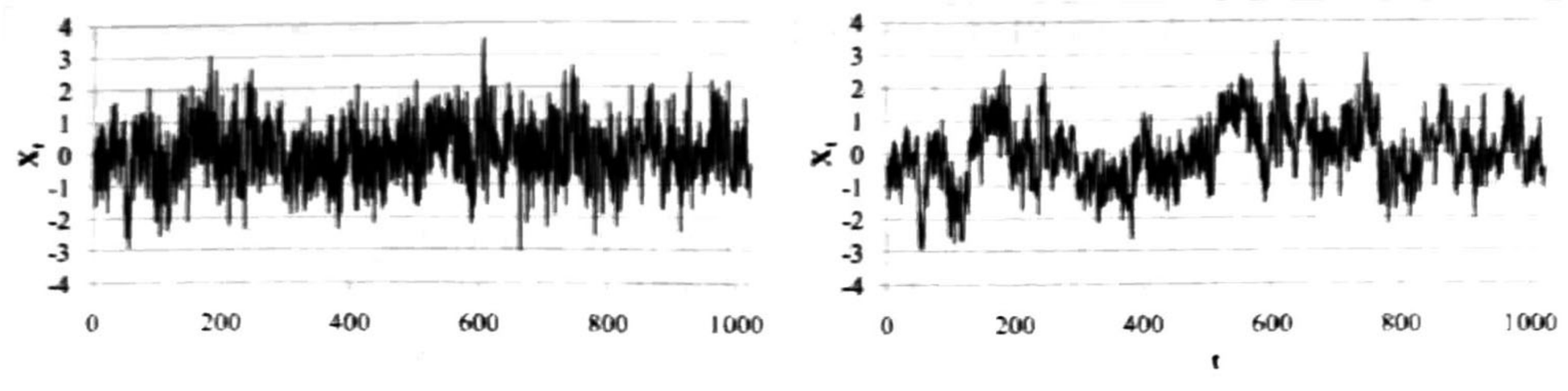
Figure 40 shows the variance of the estimated variance obtained with the proposed formula (139) compared to the approximation proposed by Yunhua in [50]. As it is expected, the variance of the estimation is close to zero (lower than 0.001) for $H < 0.5$; but, as the Hurst index increases, it becomes noticeable when $H > 0.75$. This verify the observation of [50], which says that beyond $H = 0.75$ the precision of the autocorrelation is about one order lower than when $H < 0.75$.

5.3 Synthesis of *H*-SOSS Time Series

To exemplify the proposed wavelet-based synthesis (described in Section 2.3), four time series with respective Hurst index 0.3, 0.5, 0.7 and 0.9 were synthesized from a FGN sample of size 1024. The *LD-Diagram* of the four new time series were obtained and compared to that of the original sample. The plot X_t vs. t for each one the four synthesized series is shown in Figure 41. One can visually check the presence of positive correlation in Figure 41c and Figure 41d.



a) b)



c) d)

Figure 41: Plot vs. time of the four synthesized time series. a) $H = 0.30$, b) $H = 0.50$, c) $H = 0.70$ and d) $H = 0.90$.

The *LD-Diagram* of these artificial series is shown in Figure 42. Note that the original *LD-Diagram* of the source sample is not a straight line, but it so is for the synthesized series. Also, the estimated Hurst index of this series is $\widehat{H} = 0.56$ and its estimated *LD-Diagram* is not a straight line but the estimated Hurst index of the four generated series is exactly the desired, e.g., $\widehat{H} = 0.30$ for the series shown in Figure 41a and the same for the others, and their respective *LD-Diagram* is a straight line.

This "ideal" behavior is just because the synthesizer and the estimator match each other. That does not mean that the several Hurst index estimators would return the same result; in fact, they would produce similar, but yet different, values.

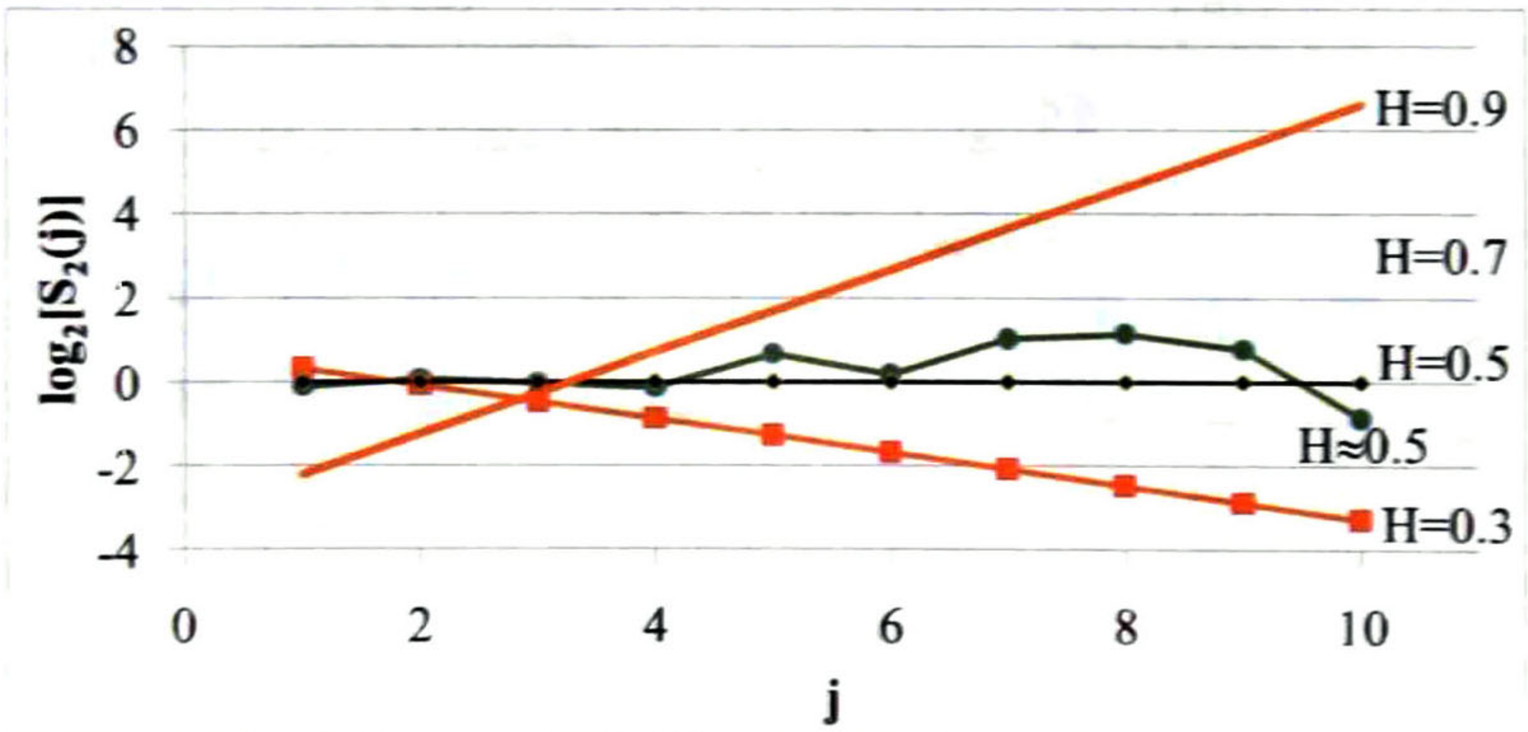


Figure 42: *LD-Diagram* of synthesized time series.

5.4 Monitored VoIP Test Calls

A set of VoIP test calls were established with the Alliance FXS PCI Voice Cards developed at CTS CINVESTAV with the following characteristics:

- H.323 architecture
- Four ports

- Codec G.711-A law [69] / G.729[70], [30]

The voice data length used for the VoIP calls is shown in Table 4.

Table 4: Used codec types and voice data length.

Voice data length		
ms	***Bytes***	
	G.711	**G.729**
10	80	10
20	160	20
40	320	40
60	480	60

5.4.1 Test Calls Scenario

The measurement scenario consists of two LANs [71]:

LAN *A*: CINVESTAV IPN

LAN *B*: Local Cable-ISP Network

Both LANs are in Guadalajara, Mexico, they have different *Internet service providers* (ISP) and are interconnected by the Internet backbone.

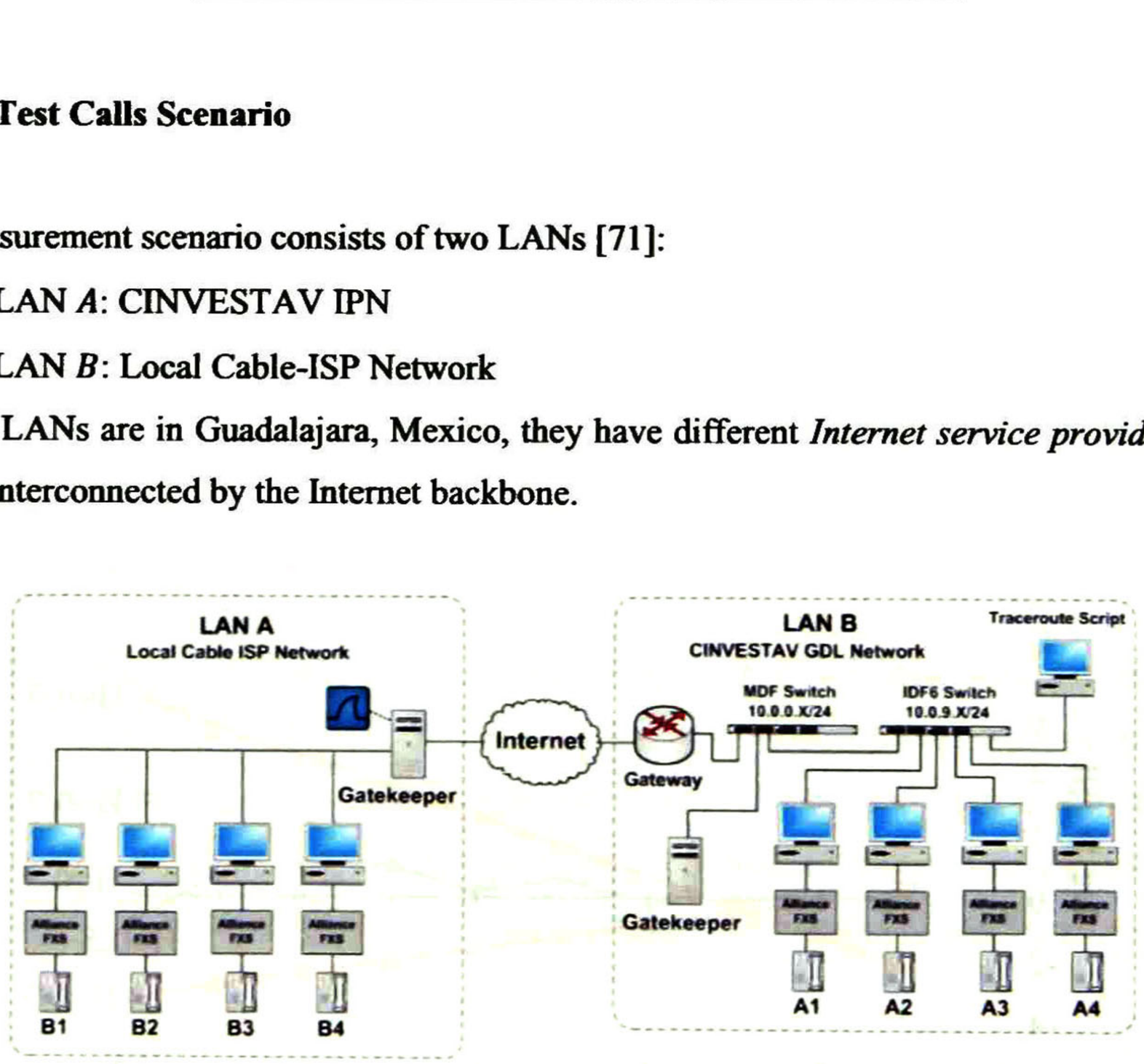


Figure 43. Test calls scenario.

Table 5: Test calls protocol.

Set	A1/B1	A2/B2	A3/B3	A4/B4
1	G.711-10*ms*	G.711-20*ms*	G.711-40*ms*	G.711-60*ms*
2	G.729-10*ms*	G.729-20*ms*	G.729-40*ms*	G.729-60*ms*
3	G.711-10*ms*	G.711-20*ms*	G.729-10*ms*	G.729-20*ms*
4	G.711-40*ms*	G.711-60*ms*	G.729-40*ms*	G.729-60*ms*

As shown in Figure 43, the H.323 zone is composed by the endpoints *A*1, *A*2, *A*3 and *A*4 located in LAN *A*, the gatekeeper and the endpoints *B*1, *B*2, *B*3 and *B*4, both located in LAN *B*, each endpoint has an Alliance FXS PCI Voice Card and a conventional cord phone. The

measurements protocol is shown in Table 5. The measurements were monitored at LAN *A* using the Network Protocol Analyzer Wireshark [72].

Additionally, a Traceroute-based script [73] was implemented in LAN *B*, in a parallel fashion to the VoIP measurements, in order to sample the path followed by the VoIP packets.

5.4.2 Collected Data Sets

The measurement protocol is described in Table 6. The number of packets sent and the total payload (in bytes) for each set is approximated.

Table 6: Description of the VoIP calls.

Set	Period (hours)	Number of data samples	Total number of sent packets*	Total payload (Bytes)*
1	6	24	4140000	910620000
2	6	24	4140000	305820000
3	6	24	6480000	840240000
4	6	24	1800000	484200000
*Values are approximated				

5.4.3 Post-processing and Filtering

The captured RTP streams were processed with Wireshark and filtered with a script to obtain, from each test VoIP call, the following characteristics:

- The respective series of sequence numbers of received packets.
- The respective series of *delay jitter*.

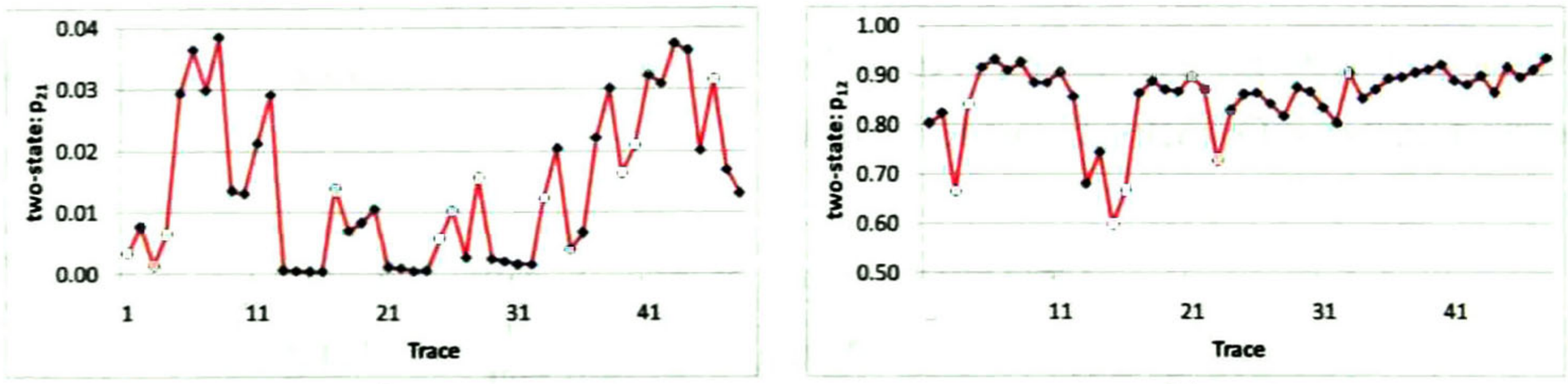
From the series of sequence numbers, the loss sequence (as defined by equation (178)) is extracted. Section 5.4.4 presents the estimation of the probabilities of transitions for the 2- and 4-state modes from these loss sequences, for which the algorithms described in 4.4.2 and 4.4.3 were applied. Also, the respective gap and burst length distributions ($f_g(k)$ and $f_b(k)$) were obtained. The respective SMSE of the gap and burst length distributions, i.e., the obtained with the estimated parameters vs. the measured, is calculated in order to evaluate the modeling.

In Section 5.4.5, the modeling of these *jitter* series, i.e., the estimation of Cauchy parameters and Hurst index, is presented.

5.4.4 Modeling of Measured Loss Sequences

The results presented in this work correspond to the 48 VoIP data traces of sets 3 and 4 which were the ones that presented higher PLR [71].

The estimated parameters of the two-state model (p_{21} and p_{12}) are shown in Figure 44a and Figure 44b. The respective minimum, maximum, average and standard deviation of these two parameters are shown in Table 7.



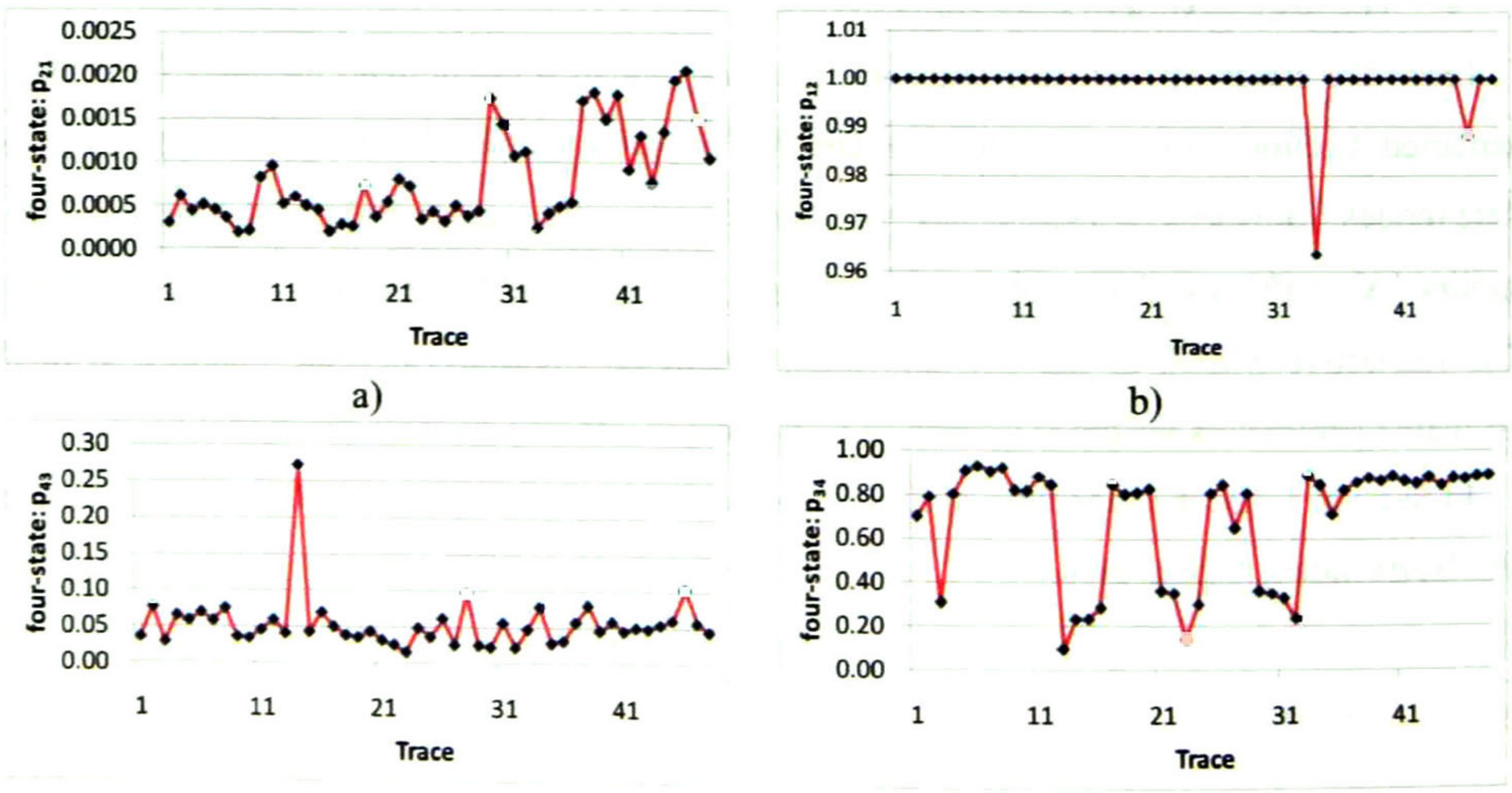
a) b)

Figure 44: Two-state model parameters a) p_{21} and b) p_{12}.

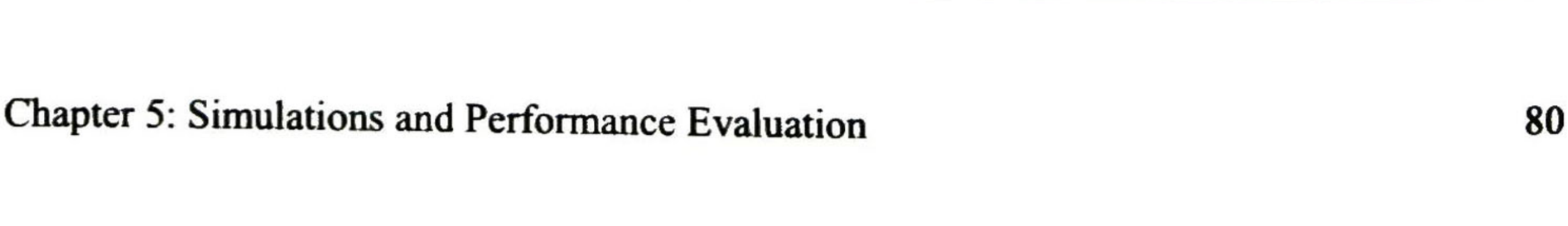
Table 7: Two-state model parameters

	p_{21}	p_{12}
Minimum	0.000322	0.595744
Maximum	0.038605	0.934703
Average	**0.013923**	**0.852838**
Std. deviation	0.012234	0.074697

The estimated parameters of the four-state model are shown in Figure 45a-f, respectively. The respective minimum, maximum, average and standard deviation of these two parameters are shown in Table 8. The loss rate threshold used for the algorithm is 1%.



a) b)



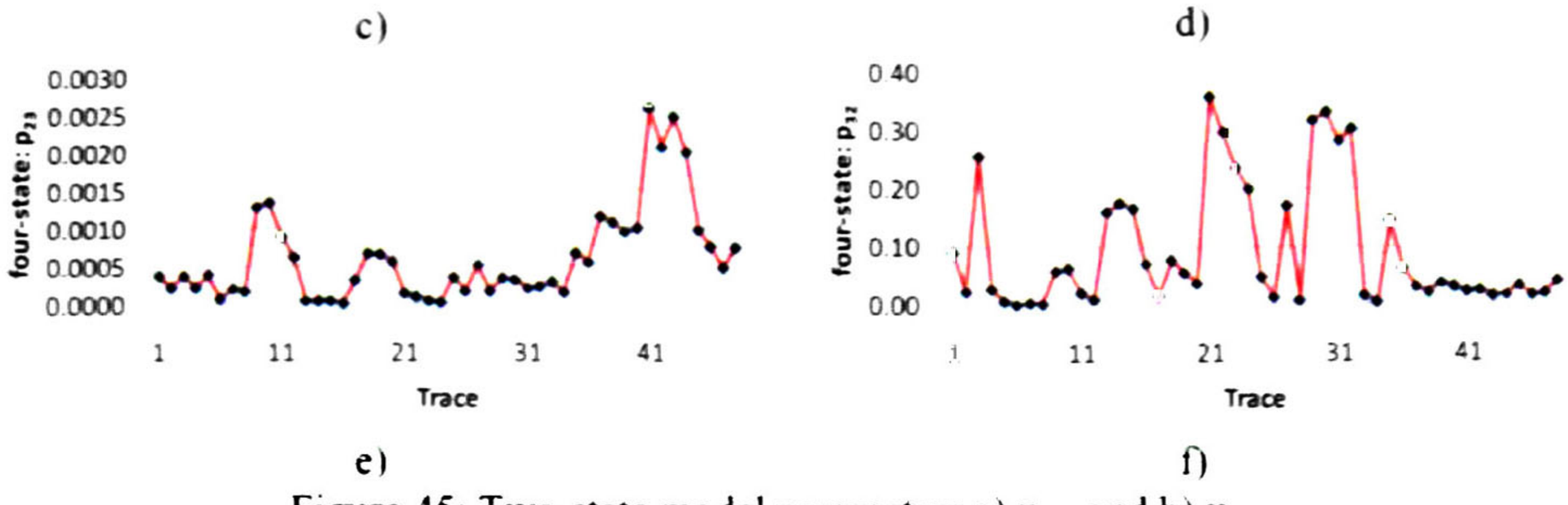


Figure 45: Two-state model parameters a) p_{21} and b) p_{12}.

Table 8: Four-state model parameters

	p_{21}	p_{12}	p_{43}	p_{34}	p_{23}	p_{32}
Minimum	0.000188	0.963636	0.015503	0.096774	0.000016	0.000890
Maximum	0.002053	1.000000	0.272727	0.931464	0.002551	0.361111
Average	**0.000793**	**0.998997**	**0.054215**	**0.688747**	**0.000602**	**0.092536**
Std. deviation	0.000527	0.005424	0.037093	0.259845	0.000609	0.106315

The burst and gap length distributions of one of the captured traces obtained from a VoIP call with codec G.711 and packet inter-departure time of $20ms$ are shown in Figure 46 and Figure 47, respectively.

In Figure 46 it is shown that the burst length decays rapidly, e.g., to zero probability for burst of length lower than 5 packets. It is also observed that both two-state and four-state models can characterize this decay.

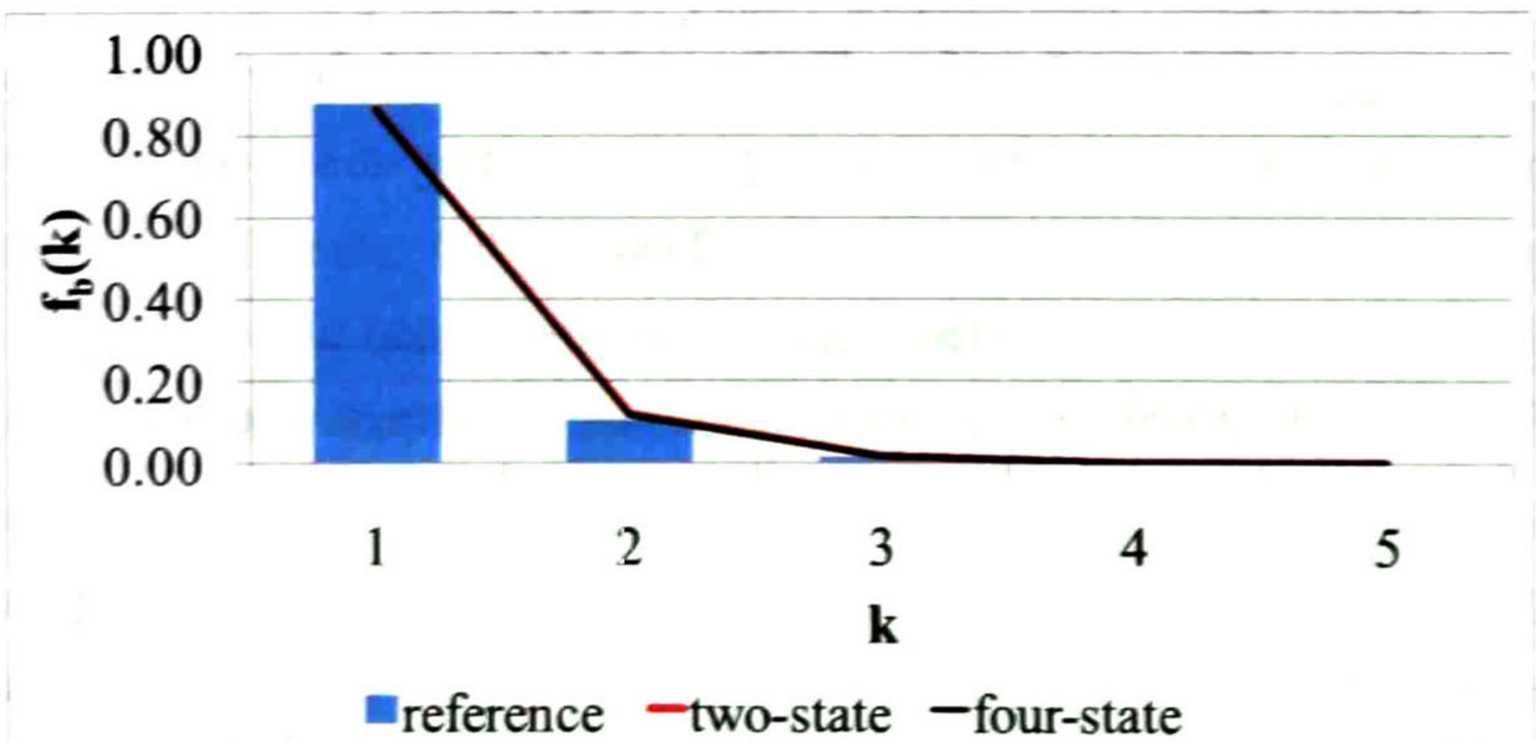


Figure 46. Burst length distribution of one of the loss sequences.

The gap length distribution decays slower than the burst length distribution. There exist gaps of tens and hundreds of packets with non-negligible probability and, in this case, the less flexible one-parameter formula of the two-state model cannot fit the measured distribution, in contrast with the four-state model, which fits it adequately.

The SMSE for burst length distribution of both two-state and four-state models are quite similar (less than 0.01) for most traces, as seen in Figure 48. But there is a remarkable difference between both models in the gap length distribution. In Figure 49 it can be observed that the SMSE four-state model fits remarkably better the gap length distribution for most traces (its maximum SMSE is 0.002).

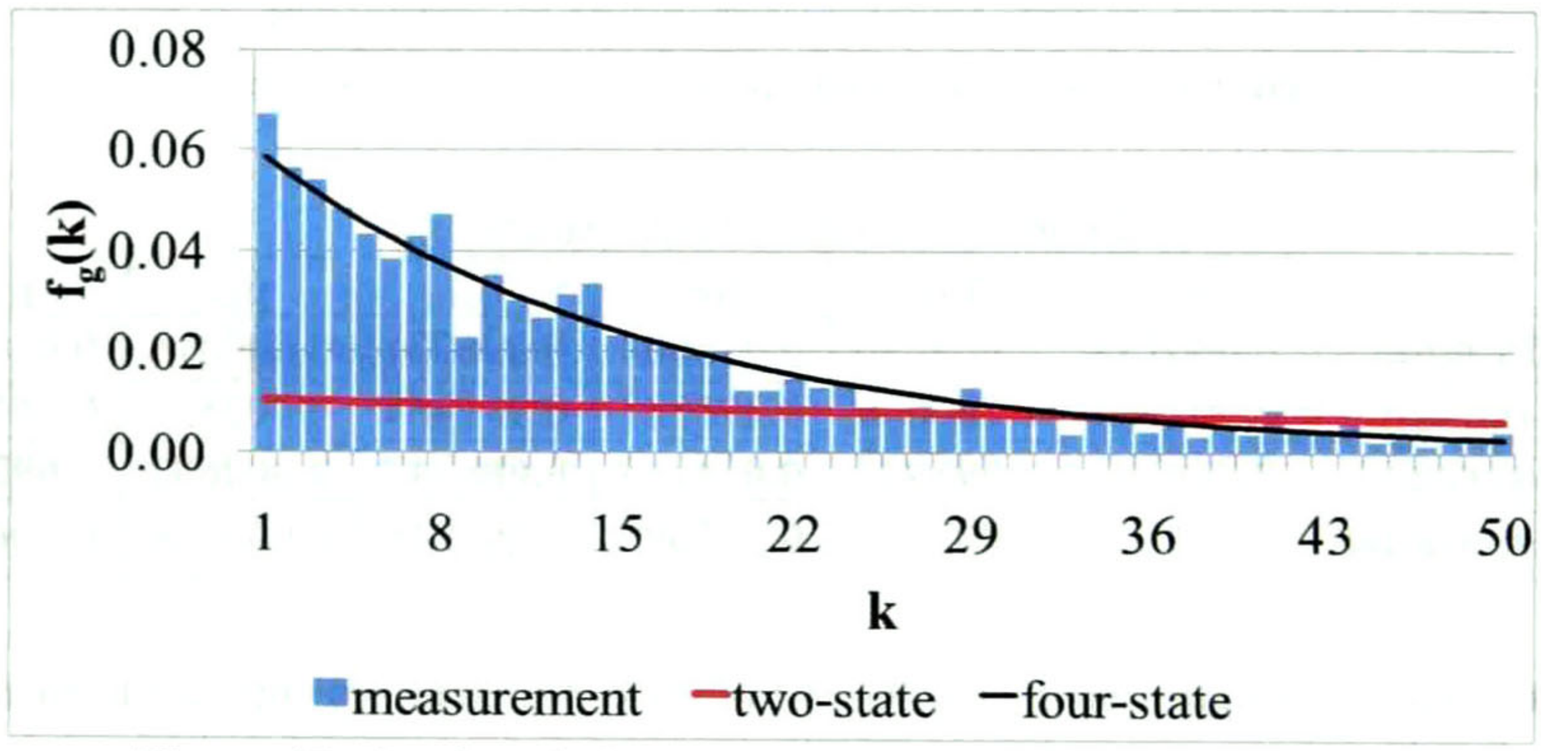


Figure 47. Gap length distribution of one of the loss sequences.

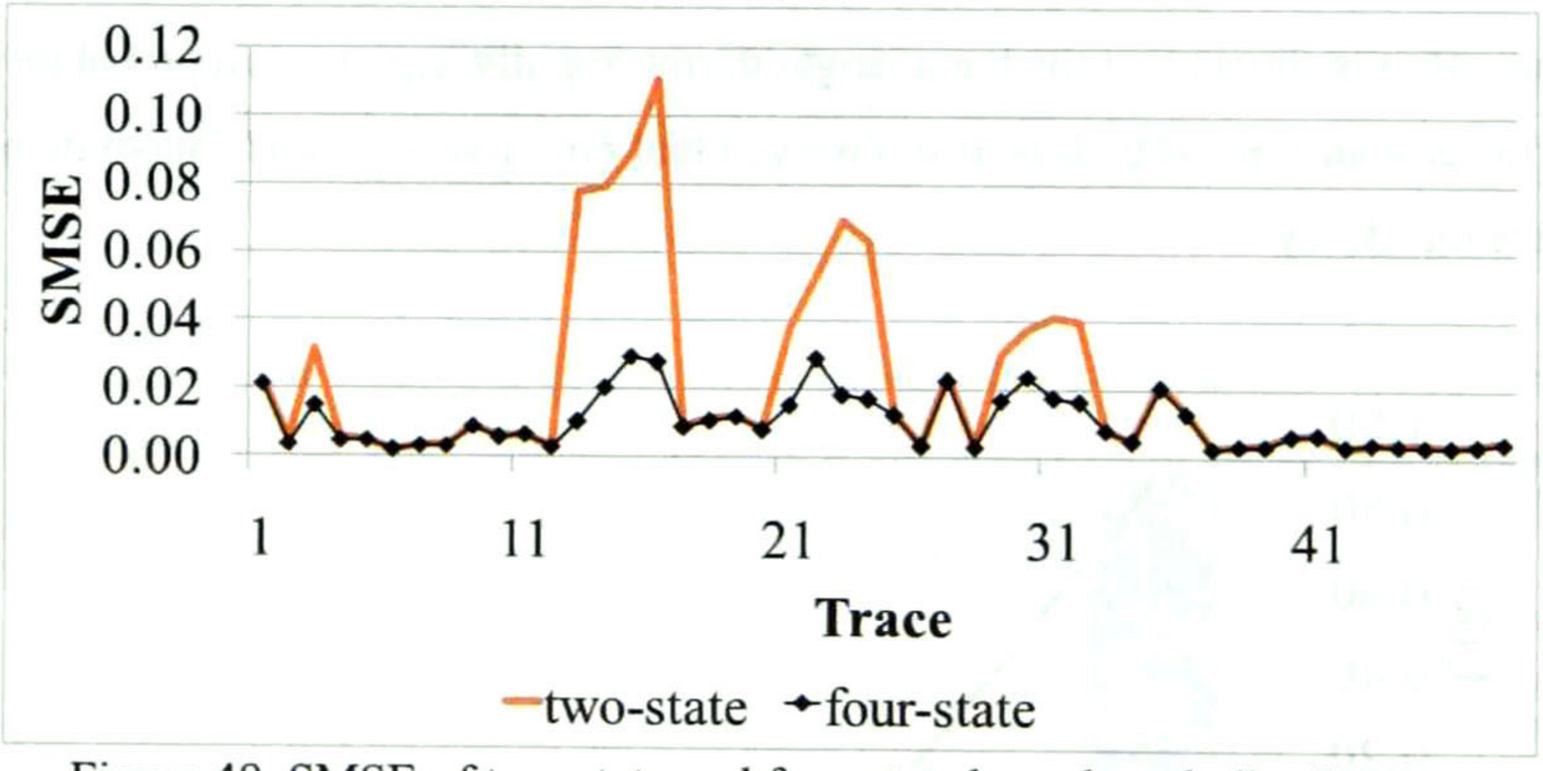


Figure 48. SMSE of two-state and four-state burst length distribution.

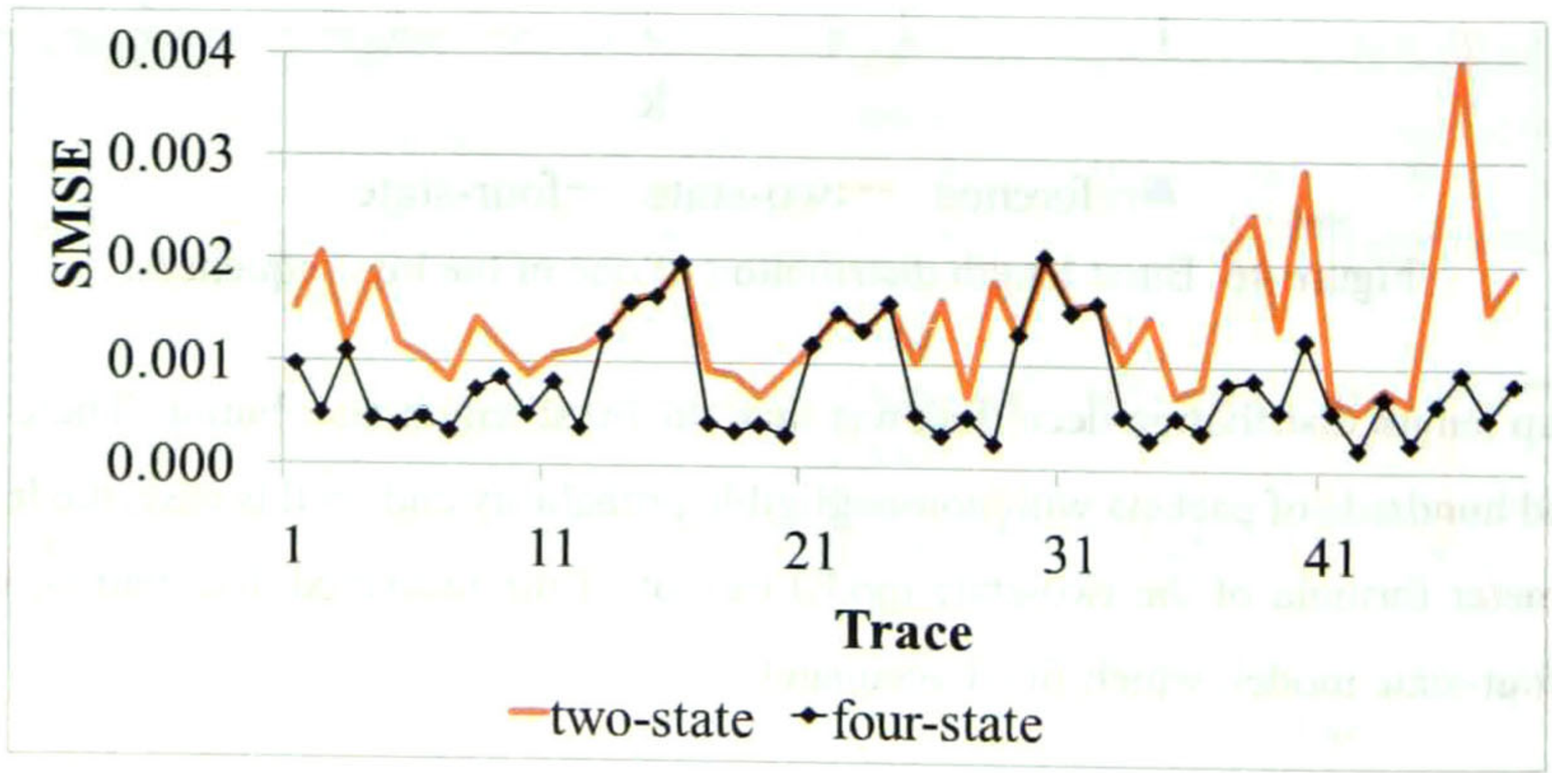


Figure 49. SMSE of two-state and four-state gap length distribution.

Figure 50 shows the PLR of the 48 studied data traces, which is calculated as the quotient of the number of lost packets and the number of sent packets. Also, by applying (161), the perceived PLR after a N-packet FEC is estimated for $N = 1, 2$ and 3.

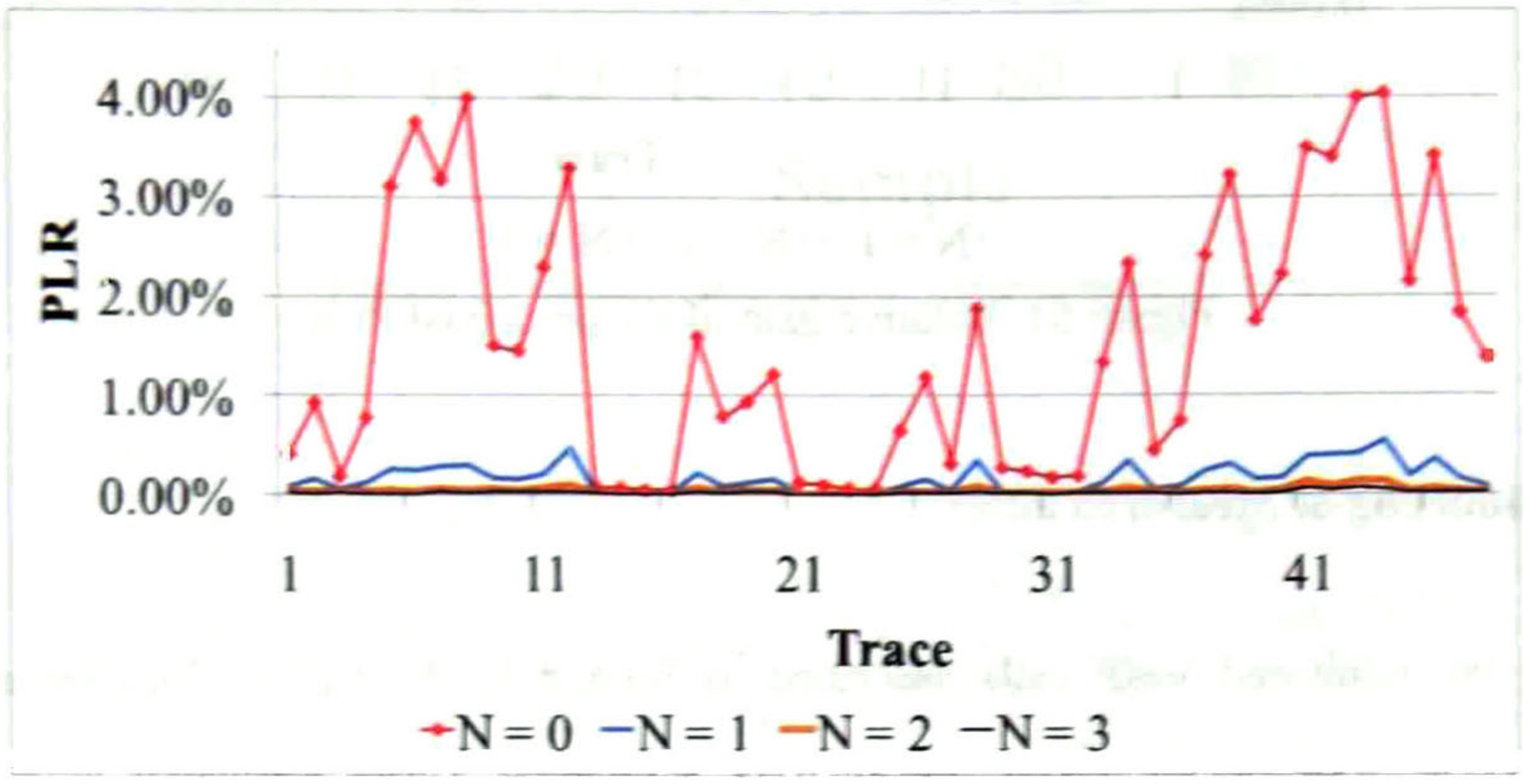


Figure 50. Perceived PLR for redundancy of $N = 0,1,2,3$ packets.

To determine how the performance is improved when increasing the level of redundancy (N), the relative gain is calculated, which defined as:

$$\Delta_{r'}(N) = \frac{-(r'_N - r'_{N-1})}{r}; N > 0 \quad (180)$$

Figure 51 shows the relative gain for the studied traces for the redundancy levels $N = 1, 2$ and 3. The major relative gain (approximately 80%) is obtained by adding redundancy of one packet, i.e., for $N = 1$. In this case the perceived PLR decreases below 0.55% for all studied traces, which is acceptable for VoIP calls. Although PLR constraints can be lower than 0.1% for Internet backbone routers or public telephony systems, a less strict limitation applies for VoIP providers and user local's ISP networks, where losses up to 1% are considered undetectable [74].

Although other communication scenario may need a different level of redundancy, e.g., $N = 0$ (no redundancy) or $N = 2$, these results are still significant. It is shown that, as the losses occur in burst of short length, e.g., one or two packets, the major gain is obtained with the first level of redundancy.

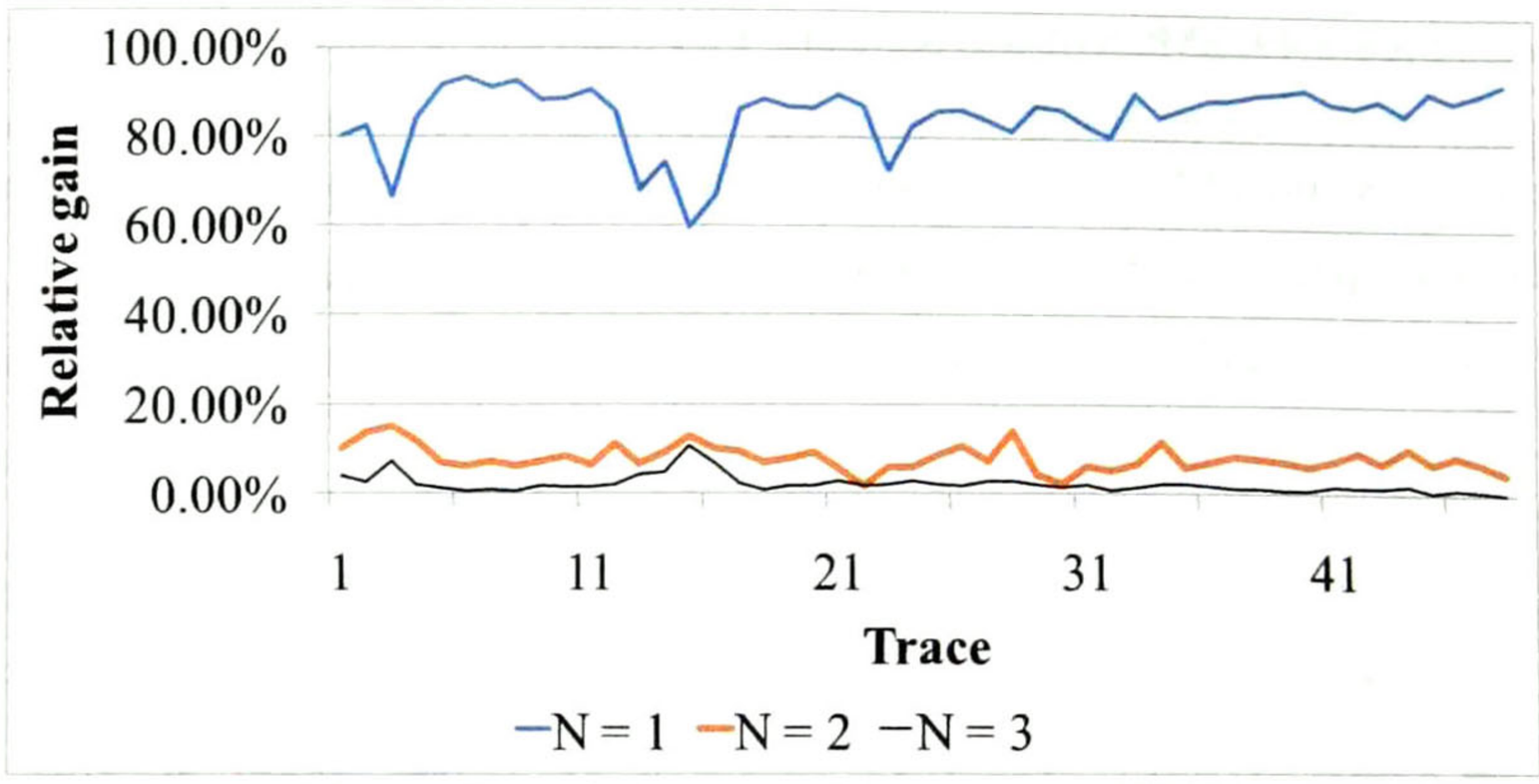


Figure 51. Relative gain of the perceived PLR.

5.4.5 Modeling of Measured Jitter

From the 96 monitored VoIP calls described in Section 0, the Cauchy-location and –scale parameters, as well as the Hurst index, were estimated. These estimated parameters are respectively shown in Figure 52, Figure 53 and Figure 54.

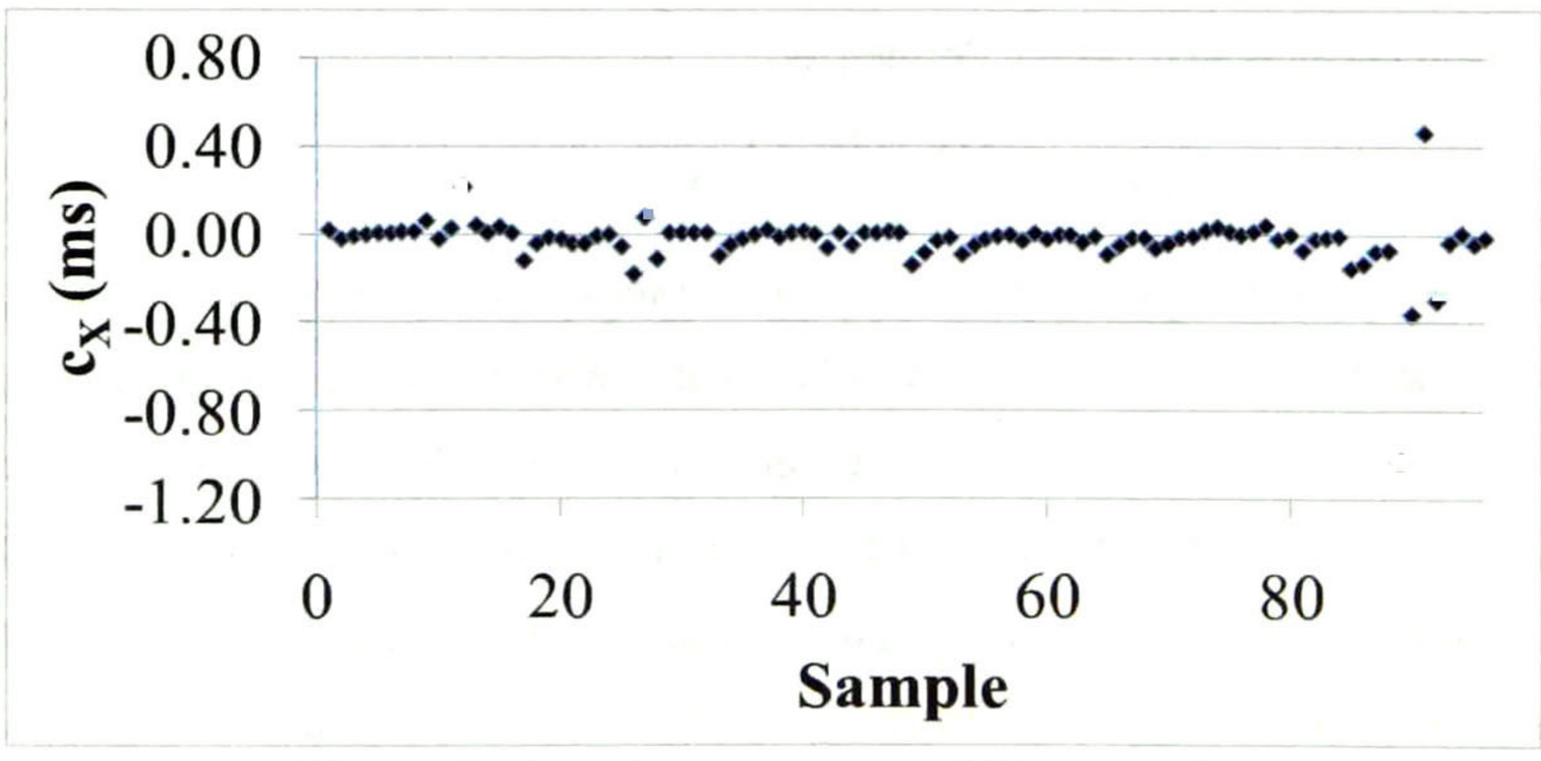


Figure 52: Location parameter of jitter samples

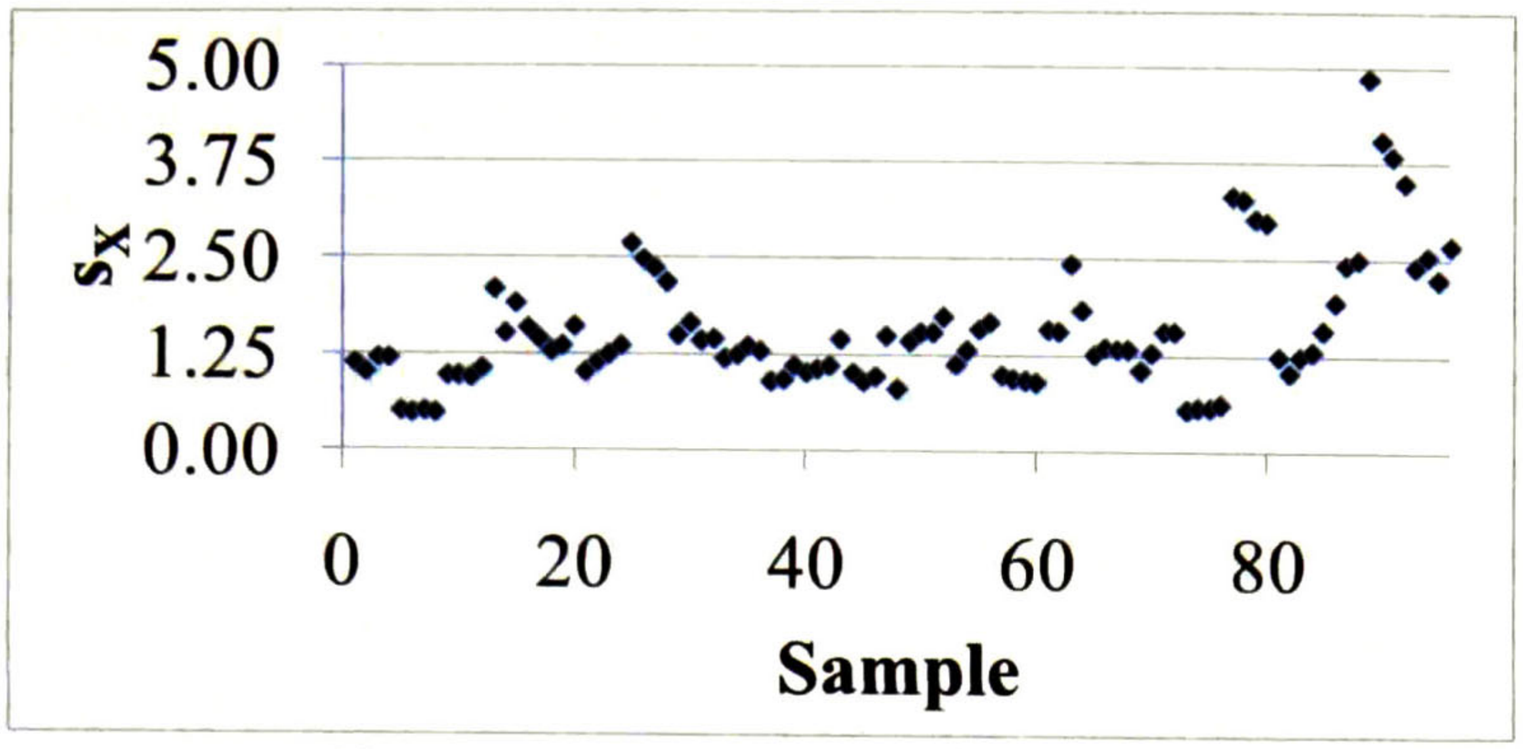


Figure 53: Scale parameter of jitter samples

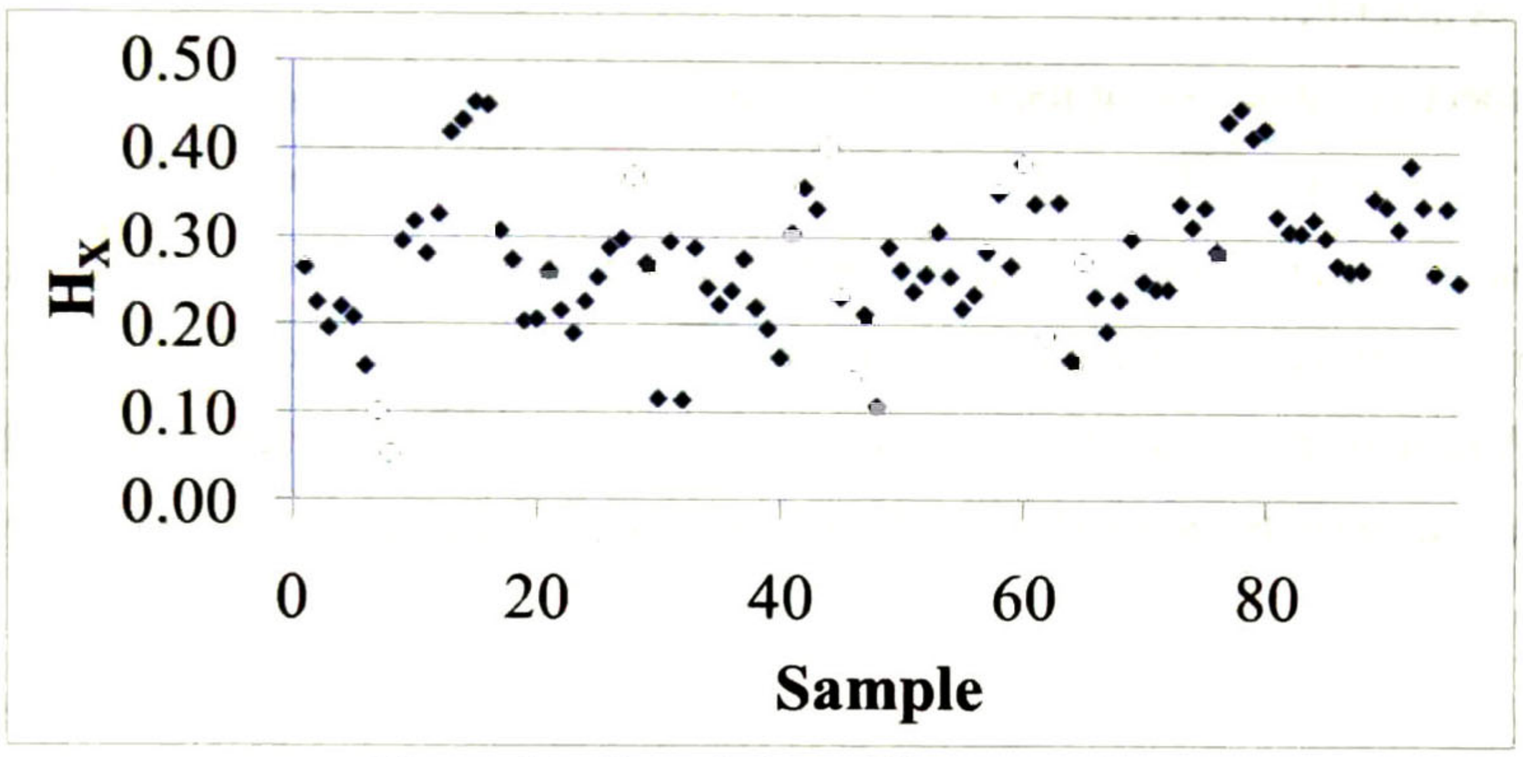


Figure 54: Hurst index of jitter samples

Table 9 shows the minimum, average, maximum and standard deviation of these three parameters. It can be observed from this table and from Figure 52 that the location parameter is near zero for most traces, which is expected due to the nature of delay jitter. Its estimated standard deviation is approximately 0.132, but a new estimation excluding outliers (those values whose magnitude is larger than 0.40) gives 0.068, which is lower than the standard deviation of Hurst estimations.

The average scale parameter is 1.55, has a standard deviation of 0.823 and presents positive *skewness*.

Table 9: Minimum, average, maximum and standard deviation of estimated parameters

	Location	**Scale**	**Hurst index**
Minimum	-1.024881207	0.473172505	0.053220736
Average	**-0.032550621**	**1.550452582**	**0.275858331**
Maximum	0.456634627	4.856410683	0.453175013
Std. deviation	0.131975764	0.823479616	0.081280047
Skewness	-4.233167324	1.597377662	-0.017680292

In order to evaluate the performance of the proposed generator, for each jitter sample described in Section 5.4.2, a new artificial series was generated with the same length, location, scale and Hurst parameters than the original, i.e., a new set of 96 samples. The efficiency of the generation was evaluated by estimating the parameters c, s and H and comparing them with the parameters of the original sample, and by calculating the square root of the MSE (SMSE) of the CDF from the new series.

The random samples consisted of FGN series generated with an implementation of the algorithm proposed by R. B. Davies and D. S. Harte [67], with Hurst index equal to that of the original jitter series. The location and scale parameters were estimated by applying a MLE estimator [75]. The Hurst index was estimated with an implementation of the Haar-wavelet based estimator, as described in [11].

The respective estimation of the c and s parameters for the series X_t, Y_t and Z_t, described in Section 4.3, are shown in Figure 55 and Figure 56. The estimated values for Y_t and Z_t are very close to that of X_t. The location parameter is in practice very close to zero, e.g., $c \in (-0.5,0.5)$ for most traces. And the estimated scale parameter is in the range between 0 and 5 for all studied samples. Also note that the ICDF transformation does not depend on whether or not the original series (X_t) is adequately modeled with a Cauchy distribution, but only in the estimation of c and s.

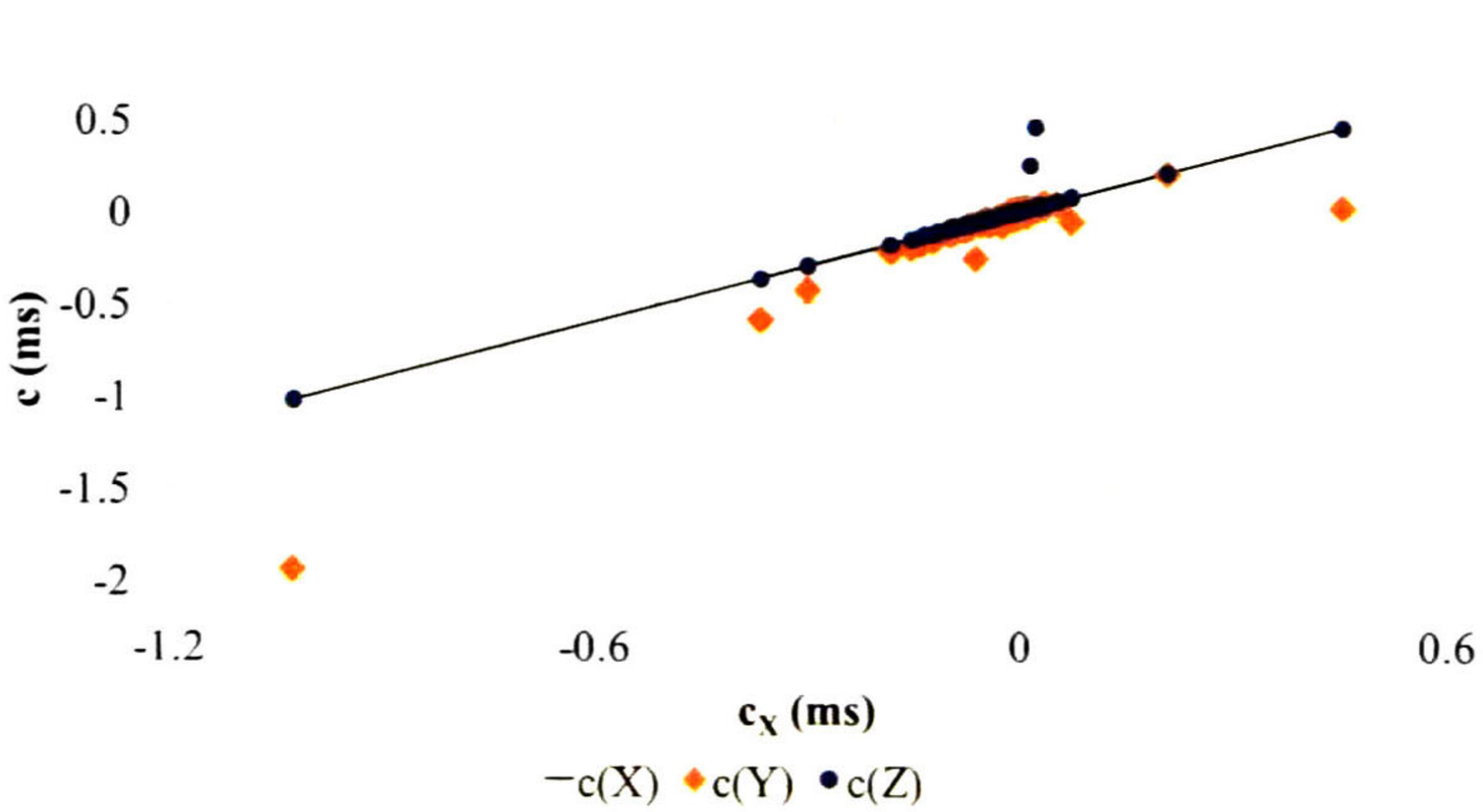


Figure 55: Estimations of the location parameter

Figure 54 shows the estimated H for the three series X_t, Y_t and Z_t. It can be observed that the correlation structure is somehow altered because of the ICDF transformation, i.e., the estimated H for Y_t is different of that of X_t and closer to 0.5. But the weighted wavelet-based synthesis, described in Section 2.3, makes the Hurst index of Z_t match that of X_t (the estimated difference $|H_Z - H_X|$ is lower than 10^{-14}).

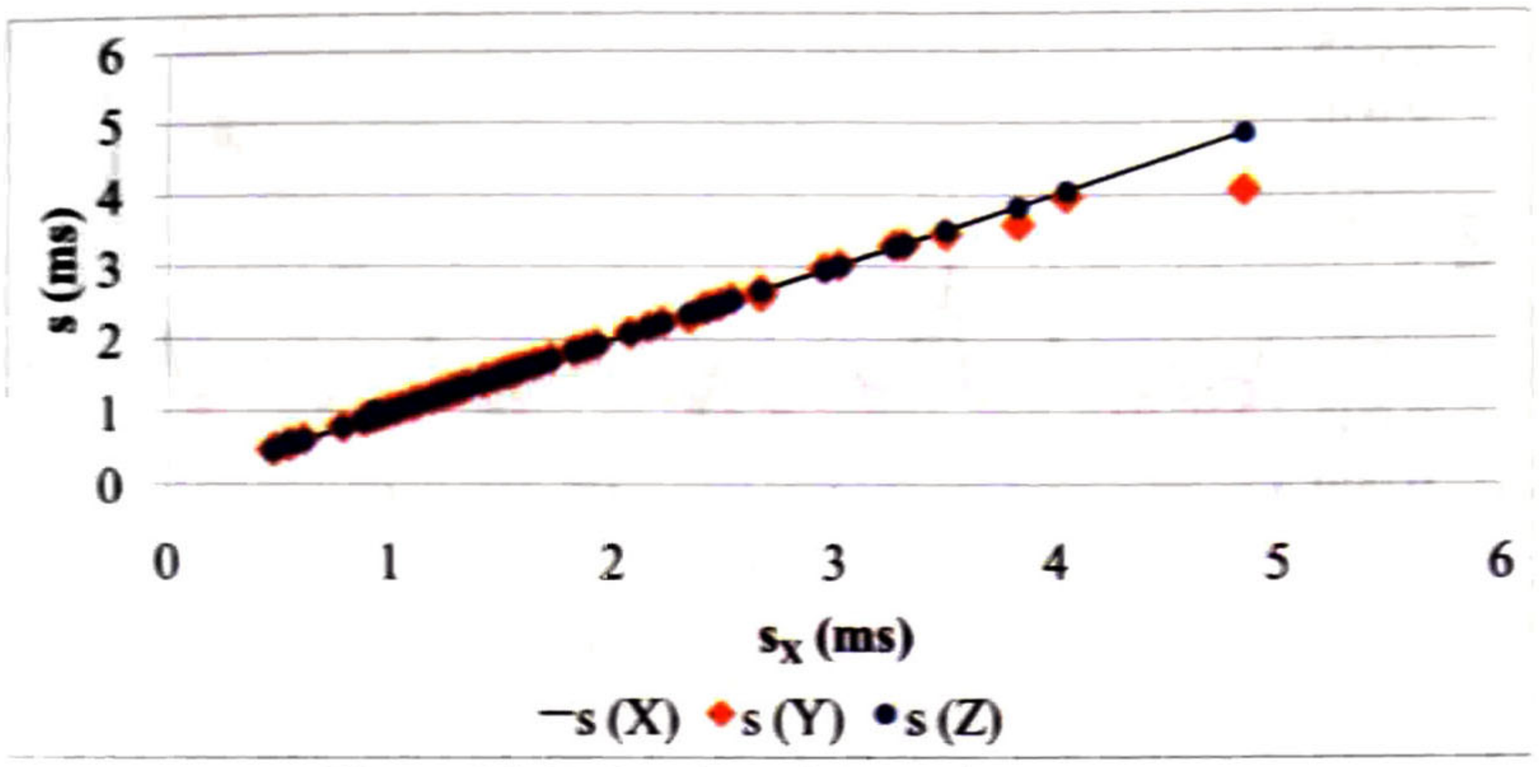


Figure 56: Estimation of the scale parameter

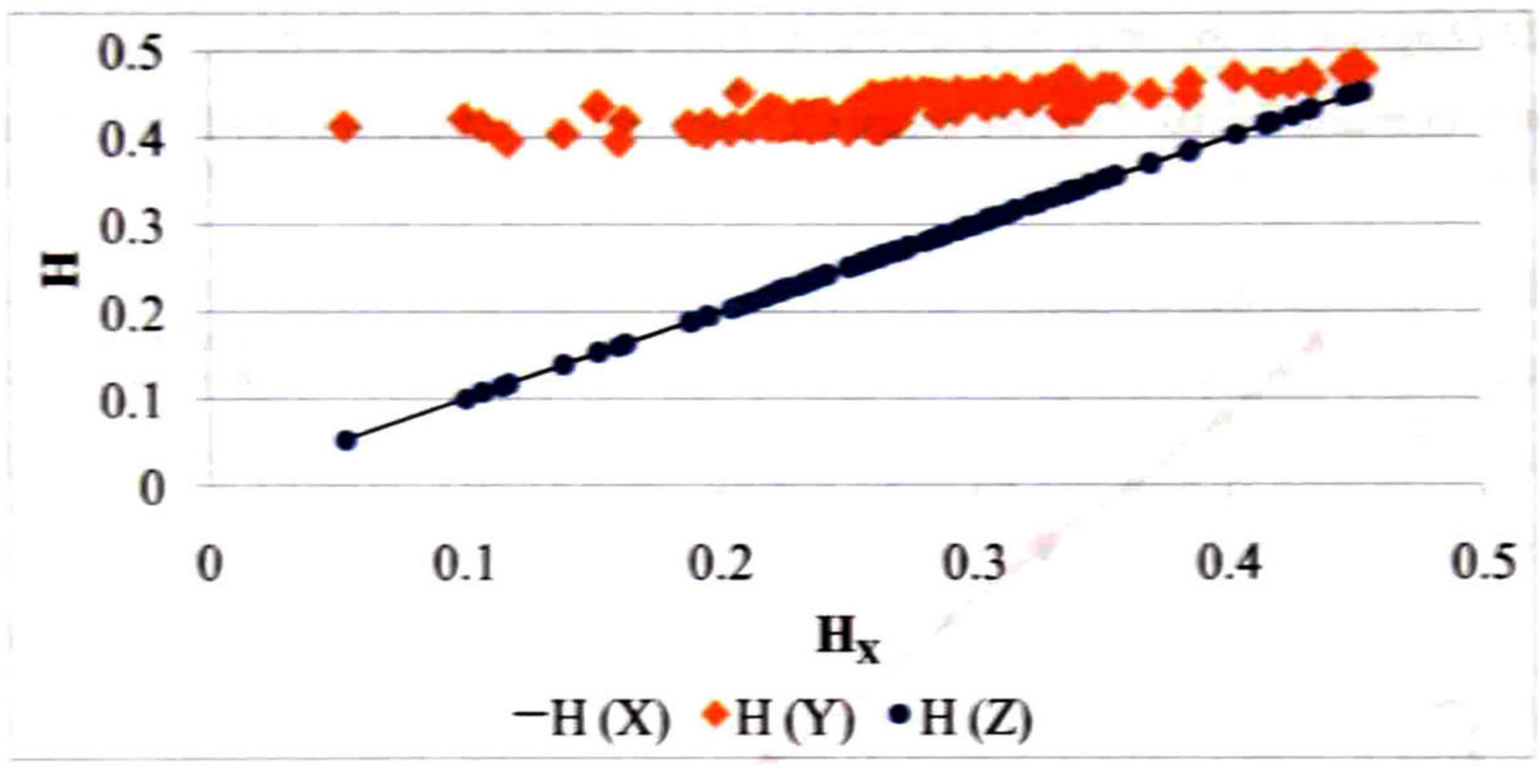


Figure 57: Estimation of the Hurst index

Figure 58 shows the respective SMSE of the CDF for Y_t and Z_t with respect to a Cauchy CDF with location and scale parameters c_X and s_X (estimated from X_t). It is observed that, although the wavelet-synthesis increases the SMSE of Z_t, it is very close to that of Y_t, i.e., Z_t is very close to Cauchy-distributed if Y_t so is. This is an indication that the efficiency of the proposed generator is very close the ICDF method, but with the advantage that the correlation structure is also adjusted.

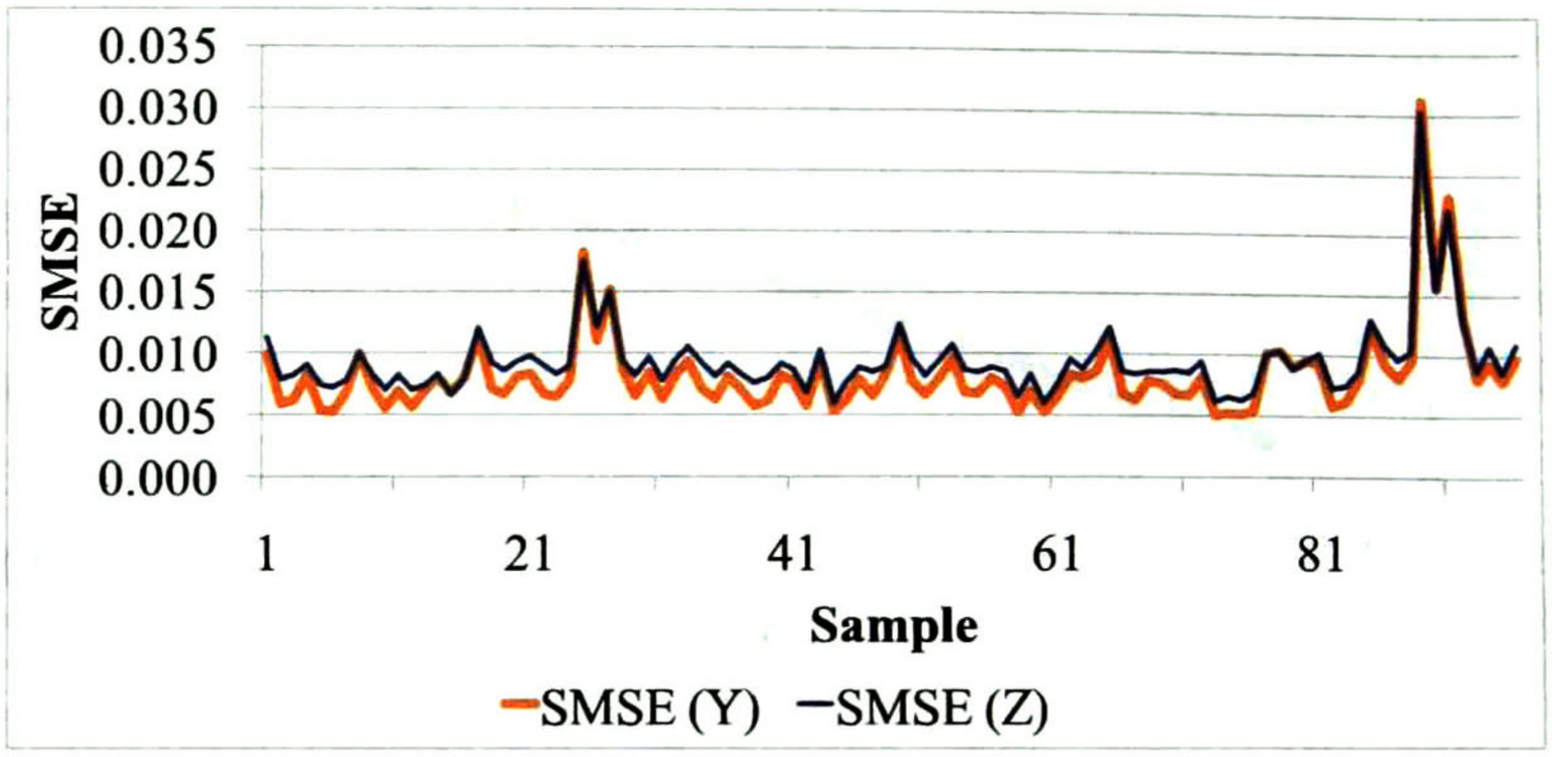


Figure 58: SMSE of the CDF of

The *LD-Diagram* of the jitter measurements helps determining the its self-similar nature, be either mono- or multi-fractal. The two cases are represented respectively by Figure 59 and Figure 60.

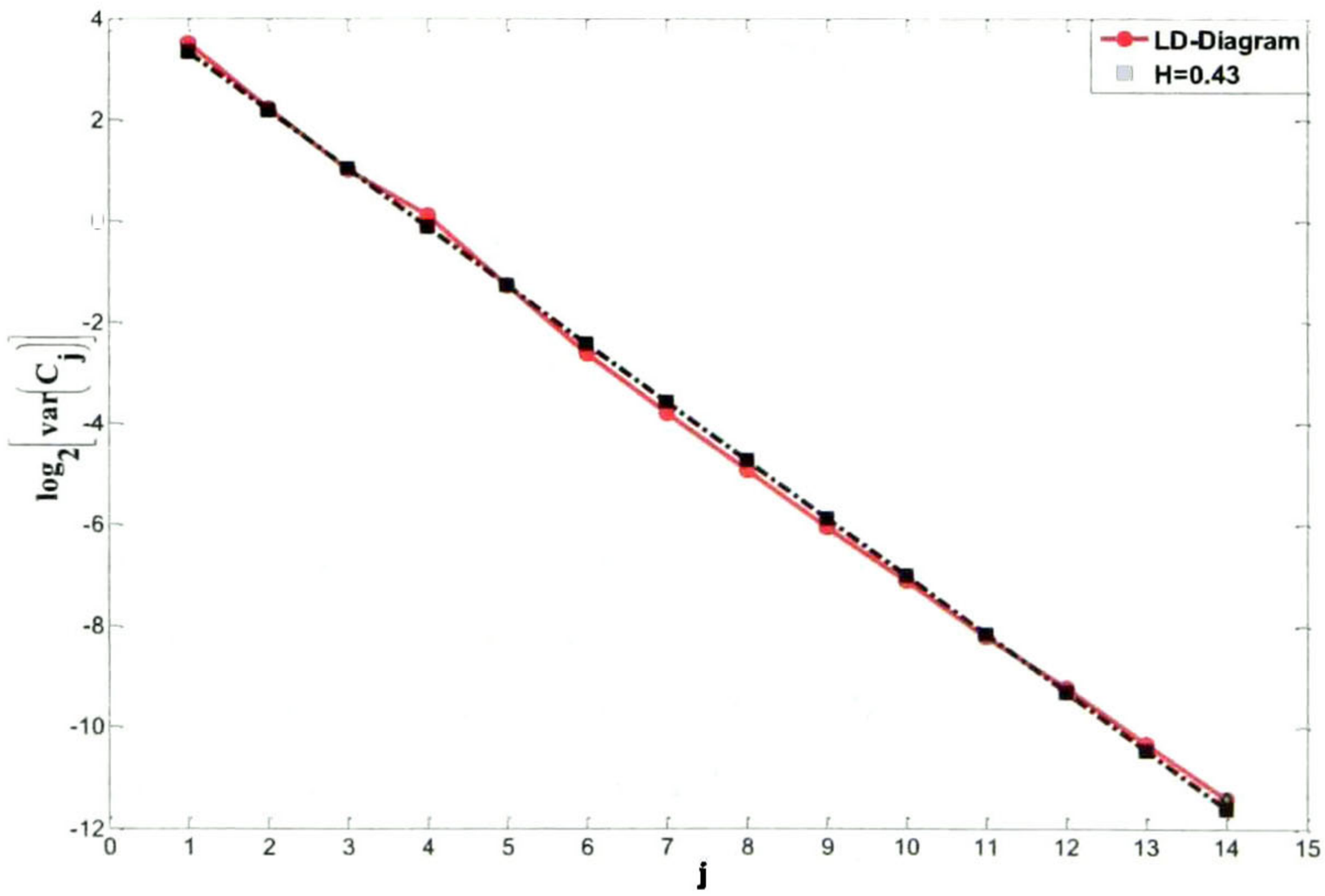


Figure 59: *LD-Diagram* behavior of a VoIP jitter data sample with mono-fractal behavior.

Figure 59 shows the components behaviors of a VoIP jitter data sample that belong to the data sets with SRD. It is observed that the variance of the components of this time series is modeled by a straight line; therefore, it is said that the time series exhibits mono-fractal behavior.

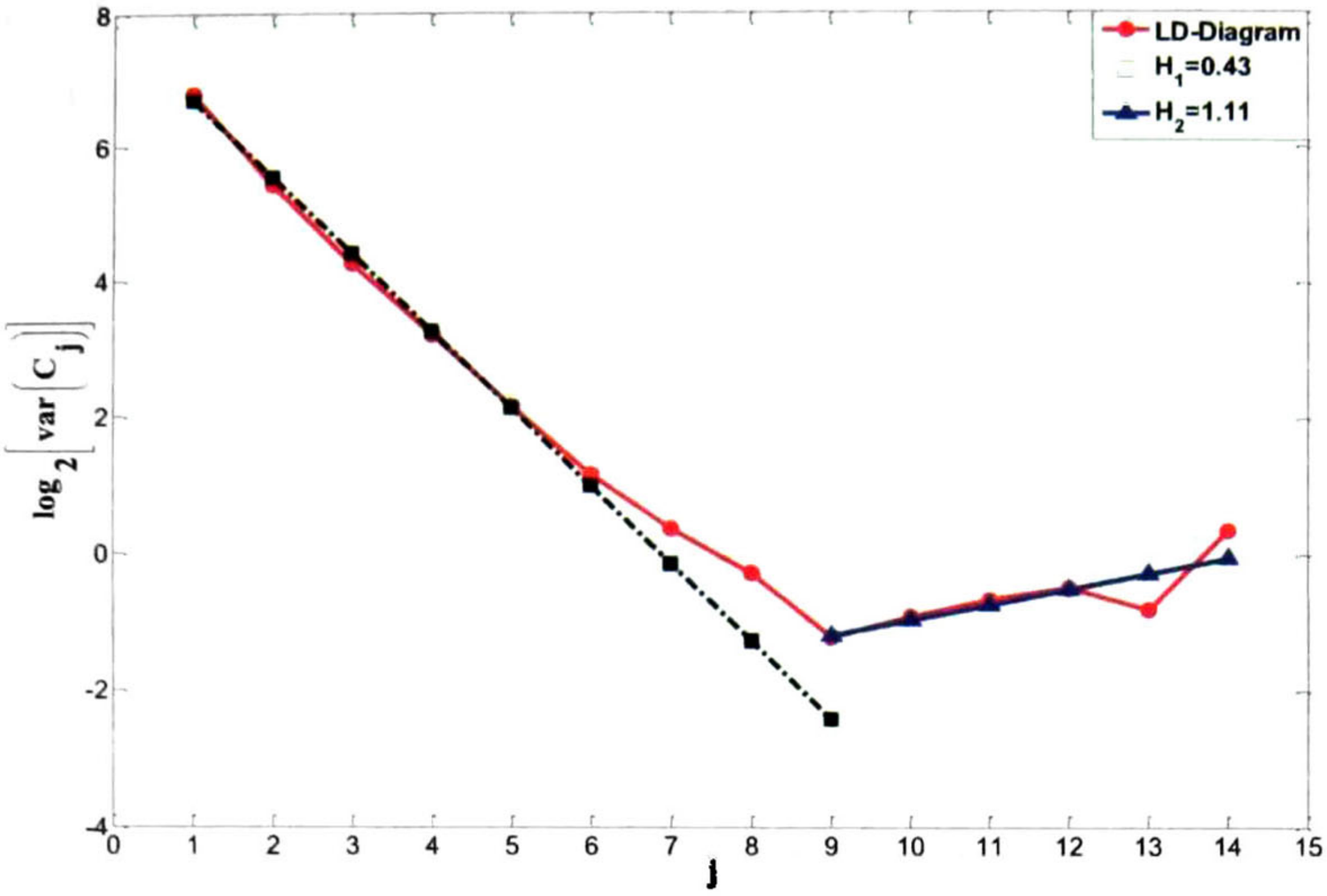


Figure 60: *LD-Diagram* behavior of a VoIP jitter data sample with multi-fractal behavior.

Figure 60 shows the components behaviors of a VoIP jitter data sample that belong to the data sets with LRD. It is observed that the variance of the components of this time series cannot be adequately modeled with a linear model, and the scaling behavior should be described with multiple scaling parameters (biscaling), therefore, it is said that this time series exhibits multi-fractal behavior. The *LD-Diagram* can be split in two segments, e.g., first for $j = 1, \ldots, 8$ and the second for $j = 9, \ldots, 14$. From each segment the slope is calculated and, from it, the Hurst index is estimated, yielding $\widehat{H}_1 = 0.43$ and $\widehat{H}_2 = 1.11$, as labeled in Figure 60. A mono-fractal process with $H > 1$ cannot exist but, in this case as the linear trend involves only certain part of the spectrum, it does not imply infinite energy.

These results show that VoIP jitter with SRD or LRD, may exhibit either mono-fractal or multi-fractal behavior. Multi-fractal behavior explains the apparently contradictory behavior of some real world time series, e.g., stationary samples that seem to present both SRD and LRD. But in any case, it is convenient to observe the *LD-Diagram* when estimating the Hurst index.

The implication of multi-fractal behavior on VoIP and other interactive multimedia services is that de-jitter buffer at the receiver may not be large enough to mask the jitter with LRD due to the persistent low frequency components.

5.5 Chapter Summary

Several experiments are described in this chapter: 1) an empirical comparison of seven estimators of the Hurst index: R/S, AM, VAR (classical), PER, MAVAR, LWHI, and WAV, (Paxson's algorithm was used here to generate the FGN samples) 2) evaluation of the formulae regarding

the sample mean and variance of *H*-SOSS samples, as described in Chapter 3, 3) evaluation of the synthesis of *H*-SOSS time series, and 4) evaluation of the modeling of packet loss and delay using, respectively, discrete Markov processes and Cauchy-distributed series.

The comparison of seven of the most extensively used Hurst index estimators showed that some estimators have a non-homogeneous behavior, e.g., the bias of R/S, AM, VAR, and PER changes across the *H* domain. MAVAR, LHWI, and WAV estimators are more consistent in this sense. And although LWHI has a very low bias, it has the greatest standard deviation of all. Generally the WAV estimator performed best, followed closely by MAVAR.

It is shown also that when working with self-similar samples, researchers should be careful when applying formulae corresponding to uncorrelated samples. The estimator of the sample mean is the same, i.e., the classical average, and it remains unbiased although the confidence interval depend on the Hurst index. The greater the *H*, the wider the confidence interval. For the variance, the classical estimator works fine for traces with SRD but, when LRD is present, it tends to underestimate the *H* as it is higher. So a more appropriate estimator, i.e., equation (139) is needed. Although when the Hurst index is close to 1, small variations on its estimation may lead to noticeable error on the variance, so possible a more detailed analysis of the sample would be needed to avoid wrong results.

The packet delay and loss are measured from real VoIP calls (96 in total) in order to estimate the necessary parameters to produce artificial samples with similar characteristics. Parameters for the 2-state and 4-state Markov models, Cauchy distribution, and the Hurst index are estimated. Some particularities can be observed from those estimations, e.g., that the parameter p_{21} from the 2-state Markov model is generally low (around 0.14), while parameter p_{12} is high (around 0.85), that the Hurst index for the delay jitter samples is always lower than 0.5, etc.

The generation of Cauchy-distributed samples with determined Hurst index performs well. Not only the Hurst index is close to the desired but also the distribution is not greatly affected, as shown in Figure 58. These results cannot be obtained using the ICDF transformation alone. These results make the proposed generator useful to test systems and prototypes with artificial samples easily, e.g., without having to take samples from real VoIP calls.

Finally, some traces are worthy plotting the *LD-diagram* because, although the estimation of *H* indicates SRD, the spectrum does not match a single linear model. Instead, it seems to be splitted in two parts, one behaving like SRD and another like LRD (see Figure 60). That constitutes an example of a multi-fractal sample. Those types of samples have also to be taken cautiously, and must be analyzed more cautiously in order to reveal its nature.

Conclusions

This work constitutes a both a theoretical and a practical reference for the analysis and synthesis of time series, especially self-similar or correlated traces. It is motivated by the importance of the self-similar processes in Internet network analysis and modeling. These type of processes are attractive because modern network are so complex that most of the times only statistical approaches are feasible, and because they can capture the both the short- and long-memory nature of traffic characteristics.

The self-similar nature of discrete processes, represented by discrete time series, is defined by a single parameter: the Hurst index. Although this parameter is defined mathematically without ambiguity, in practice it is not easy to estimate. Even when there are several estimators vastly documented in the literature, they are based on different assumptions and they produce different estimations from the same sample.

One of the most accepted and most robust estimators is the wavelet-based [11]. There are also several implementations, e.g., there are different wavelet analysis functions, but it is unbiased and it has minimal variance. It is based on the *LD-Diagram*, which constitutes itself an estimation of the power spectrum by octaves, and it has the advantage that the Hurst index can be estimated also from its slope. Another estimator that is famous in the literature, although less valued, is the variance-method. It has been put aside in many works because its implementations are known to be biased. Due to this disadvantage, it is generally used for comparison only or to give a rough estimation of the Hurst index.

In Chapter 2, a wavelet-based analysis is presented. It defines its analysis functions as differences of the aggregated (and expanded) series of the sample under study. Those functions (multiplied by a corresponding factor) become the components that synthesize the original signal in sum and are pair wise orthogonal. Also, their variances form a spectrum that coincides with the *LD-Diagram*. The variances of the aggregated series and the components are studied, in order to provide formulae for their statistics such as mean and variance. One of the most important results is the demonstration that the variance-based estimator of the Hurst index is not biased by its nature, as many authors claim, but because of inadequate implementations. I.e., to estimate the Hurst index, the slope of the *Variance-Plot* has to be estimated; this implies that variances of the aggregated series have to be calculated. And this leads to an ill-conditioned problem, as the variance of a sample itself depends on the Hurst index (or in the correlation, if not self-similar). That vicious circle has been avoided by simply use the classical estimator of the variance, which does not consider the Hurst index, and that is the source of its underestimation. In Chapter 3 it is

shown a better solution to this ill-conditioned problem, i.e., by subtracting the variances of the consecutive aggregated series we obtain the variances of the wavelet components, which form the *LD-diagram* from which the Hurst index can be estimated unbiasedly. That means, the variance-based method and the wavelet-based method are very similar, being exactly the same estimator when working with the Haar-wavelet as basis.

A comparison of seven of the most common Hurst index estimations is carried out. The results show, as expected, that the estimator with lowest bias (in magnitude) and standard deviation is the wavelet-based. Some time-based estimator as R/S, AM, PER, and the classical VAR present different bias for different Hurst index, and have higher variance. MAVAR method has the lowest bias of the compared estimators, but it has twice the standard deviation than wavelet-based (WAV).

In this work also, the quality of the VoIP communication is studied. The model of quality is defined by the ITU-T Recommendation G.107 [23] through the E-model's R-factor. It defines the QoS as a function of the voice codec, packet loss, and packet delay. In Chapter 4 it is shown that the packet losses can be modeled using finite-state discrete Markov processes and the packet delay jitter using self-similar discrete Cauchy processes. Algorithms to characterize and generate artificial samples for packet losses and delay jitter are presented. Real world measurements of those metrics were obtained using the scenario is described in Section 5.4. It consists of an H.323 zone where VoIP calls were conducted and monitored. From them, the sequences of packet loss and delay jitter were extracted, and are used to evaluate the performance of the proposed models.

One of the most important advantages of using discrete Markov processes is that identifying the source of losses is not necessary, i.e., only the probabilities for a packet to be lost or received. Consecutive received packets form a gap and consecutive packet losses form a burst. In Chapter 4 also, 2-state and 4-state discrete Markov models are studied, and formulae regarding the burst length and gap length distributions are derived (including their averages). The burst length distribution is important because it helps determining the perceived PLR when communication includes FEC as a technique to recover missing packets. Simulations show that low levels of redundancy, i.e., by using 1-packet FEC, the PLR is reduced substantially (e.g., from 5% to 0.5%). A strategy to obtain the burst and gap length distributions for the four-state model presented in Section 1.3.3. It exemplifies the generalized methodology for a m-state Markov process model, which consists of finding firstly their respective CDF, i.e., $C_b(k)$ and $C_g(k)$. It is shown also, through an evaluation based on SMSE, that both Markov models, 2-state and 4-state, can capture the geometric-type decay of the distribution of the burst length (4-state is better but the difference in SMSE is negligible), but the two-state model generally fails to model the gap length distribution in many cases. This is because in the presence of non-homogeneous losses (i.e., loss rate varies by time intervals), a single geometrical series (i.e., 2-state model) is insufficient to represent the

gap distribution decay. As the 4-state model involves two geometrical series, it has better performance.

The proposed method to generate Cauchy-distributed self-similar time series produces high-quality artificial samples that represent real world measurements of delay jitter. These artificial samples can be generated more quickly than real measurements, and can be used to test algorithms or prototype systems easily. Measurements of Cauchy parameters (location and scale) as well as the Hurst index are also taken from real delay jitter samples. They serve as basis for generating the artificial samples. It is noticeable that the estimated Hurst index from all studied delay jitter samples range from 0.05 through 0.45, meaning all of them present SRD. That is expected, as jitter is defined as the difference between the respective arrival times of consecutive packets, but the proposed generation method is also useful for series with LRD, although several recommendations should be considered (see Section 2.3).

Future work

This work can be continued in the following directions:

- Proposed formulate for H-SOSS processes can be extended to the multi-process case, i.e., considering auto- and cross-correlation, and marginal distribution for the processes. Several works have been already published, see [68], but equations proposed in this work can be extended to multivariate case and combining distinct types of marginal distributions, e.g., Gaussian, Cauchy, etc.

Construct a prototype of a VoIP channel using the proposed improvements (Chapter 4). The prototype would perform the following operations:

- At the sender side: Capture and digitize voice. Then, codify the digitized sequence using a specific scheme, e.g., G-711 or G-729, or both. Finally, send packets to the destination in the network (including redundancy).
- At the receiver side: Receive packets and decode to reconstruct the voice (lost packets can be substituted by redundant information). Then, play the audio stream by adjusting the de-jitter buffer configuration. And, during all the communication, monitor the packet delay and loss to adjust configuration to improve the QoS.

Also, the statistical characteristics of the OWD can be studied in order to use the artificial Cauchy-distributed self-similar time series in order to produce OWD sequences, so that prototypes can be tested by simulations, which are generally faster than real VoIP calls.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, "On the self-similar nature of Ethernet traffic (Extended Version)", IEEE/ACM Transactions on Networking, Volume 2, issue 1, pp.1-15.

[2] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, "On the self-similar nature of Ethernet traffic", Proceedings of the ACM SIGCOMM Computer Communication Review, vol. 23, issue 4, San Francisco CA, USA (1993) 13-17.

[3] T. Karagiannis, M. Molle, F. Faloutsos, "Long-range dependence, ten years of Internet traffic modeling", IEEE Internet Computing, Vol. 8, issue 5, 2004.

[4] D. Veitch, P. Flandrin, P. Abry, R. Riedi, R. Baraniuk. "The Multiscale Nature of Network Traffic: Discovery, Analysis, and Modelling", IEEE Signal Processing Magazine, vol. 19, issue 3, (2002) 28-46.

[5] M. S. Taqqu, V. Teverovsky, W. Willinger, "Is network traffic self-similar or multifractal?" Fractals, Vol. 1, No. 1, 1997, pp. 63-73.

[6] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch. "Multiscale nature of network traffic", IEEE Signal Processing Magazine, 2002.

[7] H-D. J. Jeong, J-S. R. Lee, K. Pawlikowski, "Comparison of various estimators in simulated FGN", Simulation Modelling Practice and Theory 15, Issue 9 (2007) 1173-1191.

[8] J. Mielniczuk, P. Wojdyllo. "Estimation of Hurst exponent revisited", Computational Statistics & Data Analysis 51, Issue 9 (2007), 4510-4525.

[9] P. Shang, Y. Lu, S. Kamae. "Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis". Chaos, Solitons and Fractals 36 (2008) 82-90.

[10] G. Horn, A. Kvalbein, J. Blomskøld, E. Nilsen. "An empirical comparison of generators for self similar simulated traffic", Performance Evaluation 64 (2007) 162–190.

[11] D. Veitch, P. Abry, "A wavelet based joint estimator of the parameters of long-range dependence". IEEE Transactions on Information Theory, special issue "Multiscale statistical signal analysis and its applications", Vol. 45, No. 3, 1999, pp. 878-897.

[12] D. Radev, I. Lokshina, "Advanced Models and Algorithms for Self-Similar Network Traffic Simulation and Performance Analysis". Journal of Electrical Engineering (JEE), pp. 341-349, 2010.

[13] R. G. Clegg. "A practical guide to measuring the Hurst parameter", 21st UK Performance Engineering Workshop. School of Computing Science. Technical Report Series, pp.43-55, 2006.

[14] J. F. Kenney, "Mathematics of statistics", Van Nostrand, N. Y. 1939.

[15] J. F. Kenney, E. S. Keeping, "Mathematics of Statistics, Pt. 2", 2nd ed. Princeton, NJ: Van Nostrand, 1951

[16] T. T. Soong. "Fundamentals of Probability and Statistics for Engineers", John Wiley & Sons, Ltd., USA (2004).

[17] B. Tsybakov and N. Georganas, "Self-similar processes in communication networks", IEEE Transactions on Information Theory 44, Vol. 5 (1998) pp. 1713-1725.

[18] Sheluhin O., Smolskiy S., Osin A., "Self-similar processes in telecommunications", Wiley, April 2007.

[19] R. Dobrescu, D. Hossu, S. Mocanu, M. Nicolae, "New algorithms for QoS performance improvement in high speed networks", WSEAS Transactions on Communications, Vol. 7, No. 12, 2008, pp. 1192-1201.

[20] T. Palade, E. Puschita, "Requirements for a New Resource Reservation Model in Hybrid Access Wireless Network", WSEAS Transactions on Communications, Vol. 7, No. 3, 2008, pp. 377-382.

[21] R. Libnik, A. Svigelj, G. Kandus, R. Libnik, "Performance evaluation of SIP based handover in heterogeneous access networks", WSEAS Transactions on Communications, Vol. 7, No. 5, 2008, pp. 448-458.

[22] N. Erdöl, C. Castellucia, Ali Zilouchian, "Recovery of Missing Speech Packets Using the Short-time Energy and Zero-crossing Measurement", IEEE Transactions on Speech and Audio Processing, Vol. 1 No. 3, July 1993.

[23] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning", 2003.

[24] T. Karagiannis, M. Faloutsos, R. Riedi, "Long-range dependence: Now you see it, now you don't", Proceedings of the IEEE Global Telecommunications Conf., Global Internet Symp., 2002.

[25] L. Atzori, N. Aste, M. Isola, "Estimation of multi-fractal parameters in traffic measurement: An accuracy-based real-time approach", Computer Communications 29, issue 11, pp. 1879-1888, July 2006.

[26] S. Stoev, M. S. Taqqu, C. Park, J. S. Marron, "On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of Internet traffic", Computer Networks, special issue "Long range dependent traffic", Vol. 48, No. 3, 2005, pp. 423-445.

[27] S. Bregni, L. Primerano, "Using the modified Allan variance for accurate estimation of the Hurst parameter of long-range dependent traffic", Submitted to IEEE Transactions on Information Theory, 2005.

[28] D. Veitch, N. Hohn, P. Abry, "Multifractality in TCP/IP traffic: the case against", Computer Networks 48, pp. 293-313, 2005.

[29] I. W. C. Lee, A. O. Fapojuwo, "Stochastic processes for computer network traffic modeling", Computer Communications, Vol. 29, No. 1, 2005, pp. 1-23.

[30] ITU-T Recommendation G.1020, "Performance parameter definitions for quality of speech and other voiceband applications using IP networks", 2006.

[31] H. Schulzrinne, J. F. Kurose, D. F. Towsley, "Loss Correlation for Queues with Bursty Input Streams", Proc. IEEE ICC '92, Chicago II, 1992, pp.219-224.

[32] J. R. Yee, E. J. Weldon, "Evaluation of the Performance of Error Correcting Codes on a Gilbert Channel", IEEE transactions on Communications, Vol. 43 No. 8, 1995.

[33] W. Ching, M. K. Ng., "Markov Chains: Models, Algorithms and Applications", Springer, 2006.

[34] G. G. Yin, Q. Zhang, "Discrete-Time Markov Models, Two-Time-Scale Methods and Applications", Springer, 2005.

[35] H. Lee, H. Lee, "A Packet Loss Recovery Scheme Based on the Gap Statistics", C. Kim (Ed.): ICOIN 2005, LNCS 3391, Springer-Verlag Berlin Heidelberg, 2005, 2005, pp. 627-634.

[36] IETF Standard RFC-3550, "RTP: A Transport Protocol for Real-Time Applications", July 2003.

[37] ITU-T Recommendation G.114, "One-way transmission time", 2003.

[38] S. Madhani, S. Shah, A. Gutierrez, "Optimized Adaptive Jitter Buffer Design for Wireless Internet Telephony", Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE, 2007, pp. 5248-5253.

[39] ITU-T Recommendation G.109, "Definition of categories of speech transmission quality", 1999.

[40] G. L. Choudhurya, R G. Cole, "Design and analysis of optimal adaptive de-jitter buffers", Computer Communications 27, 2004, pp. 529-537.

[41] R.G. Cole, J.H. Rosenbluth, "Voice over IP performance monitoring", ACM J. Computer Communications Review 31, 2001.

[42] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software", IEEE Press, pp. 213-235.

[43] N. Rillo, "Introduction to wavelets theory", http://www.mat.ub.es/~soria/TAD-Wavelets.pdf, pp. 34-35.

[44] P. Abry, D. Veitch, P. Flandrin, "Long-range dependence: revisiting aggregation with wavelets", Journal of Time Series Analysis, 1998.

[45] É. Deléchelle, J-C. Nunes, J. Lemoine, "Empirical mode decomposition synthesis of fractional processes in 1D- and 2D- space", Image and Vision Computing 23 (2005) 799-806.

[46] L. Zão, R.Coelho, P. Flandrin, "Speech Enhancement with EMD and Hurst-Based Mode Selection", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 22, No. 5, May 2014.

[47] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance", Proc. Cambridge Philos. Soc. 30 (1934) 178–191.

[48] Y. Tian, G. P. H. Styan, "Cochran's statistical theorem revisited", Journal of Statistical Planning and Inference 136 (2006) 2659-2667.

[49] J. Beran. "Statistics for Long-Memory Processes". Chapman & Hall/CRC. Monographs on Statistics & Applied Probability (1994).

[50] R. Yunhua, "Evaluation and estimation of second-order self-similar network traffic", Computer Communications 27 (2004) 898-904.

[51] M. Krunz, I. Matta, "Analytical investigation of the bias effect in variance-type estimators for inference of long-range dependence", Computer Networks 40 (2002) 445-458.

[52] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion", IEEE Transactions on Information Theory, Vol. 38, No.2 (1992) 910-917.

[53] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, R. G. Baraniuk, "A multifractal wavelet model with application to network traffic", IEEE Transactions on Information Theory, Vol. 45, No. 3, April 1999.

[54] V. Teverovsky, M. Taqqu, "Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator", Journal of Time Series Analysis, Vol. 18, Issue 3 (2001) 279-304.

[55] L. Estrada, D. Torres, and H. Toral, "A Study of Wavelet Analysis and Data Extraction from Second-Order Self-Similar Time Series", Mathematical Problems in Engineering, Hindawi, Volume 2013 (2013), May 2013.

[56] R. Aldea, D. Tărniceriu. "Estimating the Hurst Exponent in Motor Imagery-based Brain Computer Interface" 7th Conference on Speech Technology and Human - Computer Dialogue (SpeD), 2013. DOI: 10.1109/SpeD.2013.6682656, 2013, pp.1 – 6.

[57] L. Estrada, D. Torres, H. Toral, "Variance Error for Finite-length Self-similar Time Series". 7th International Conference on Computing, Communications, and Control Technologies (CCCT 2009), Orlando, Florida, USA, pp.193-198, 2009.

[58] J. Bolot, S. Fosse-Parisis, D. Towsley, "Adaptive FEC-based Error Control for Interactive Audio in the Internet", Proceedings of IEEE INFOCOM, 1998, pp. 1453-1460.

[59] H. Zhang, L. Xie, J. Byun, P. Flynn, C. Shim, "Packet Loss Burstiness and Enhancement to the E-Model", Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks (SNPD/SAWN'05), 2005.

[60] H. J. Jeong, J. R. Lee, and H. Park, "Teletraffic Generation of Self-Similar Processes with Arbitrary Marginal Distributions for Simulation: Analysis of Hurst Parameters", ICCSA 2004, LNCS 3045, 2004, pp. 827–836.

[61] H. D. Wacker, J. Boercsoek, H. Hillmer, "Redundant Data Transmission and Nonlinear Codes", WSEAS Transactions on Communications, Vol. 7, No. 6, 2008.

[62] L. Estrada, D. Torres and H. Toral, "Analytical Description of a Parameter-based Optimization of the Quality of Service for VoIP Communications", WSEAS Transactions on Communications, issue 9, Vol. 8, 2009, pp. 1042-1052.

[63] L. Estrada, D. Torres, H. Toral, "Analytical Investigation of the Performance of Packet-level FEC Techniques in VoIP Communications", Recent Advances in Signals and Systems, Editors: Imre Rudas, Budapest Tech, Hungary; Metin Demiralp, Istanbul Technical University, Turkey; Nikos Mastorakis, Technical University of Sofia, Bulgaria. Ed. WSEAS Press, ISBN 978-960-474-114-4.

[64] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic", Computer Communication Review 27(5). Pp. 5-18. Oct 1997.

[65] J. Ramirez, D. Torres, "A tool for analysis of Internet metrics", Mexico, D.F., September 2005, Proceedings CIE 2005, pp. 60-63.

[66] J. Ramirez, D. Torres, "Development of a tool for basic analysis of Internet self-similar traffic", M.S. Thesis (Thesis in Spanish), CINVESTAV del IPN Unidad Guadalajara, November 2005.

[67] R. B. Davies, D. S. Harte, "Test for Hurst effect", Biometrika, Vol. 74, No. 1, 1987, pp. 95-101.

[68] H. Helgason, V. Pipiras, P. Abry, "Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding", Signal Processing, Elsevier, Volume 91, Issue 5, May 2011, pp.1123–1133.

[69] ITU-T Recommendation G.711, "Pulse Code Modulation (PCM) of Voice Frequencies", 1993.

[70] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," 1998.

[71] H. Toral, D. Torres, C. Hernandez, L. Estrada, "Self-similarity, Packet Loss, Jitter, and Packet Size: Empirical Relationships for VoIP", 18th International Conference on Electronics, Communications and Computers, CONIELECOMP, March 2008.

[72] "Wireshark: A Network Protocol Analyzer", http://www.wireshark.org/.

[73] V. Jacobson, "Traceroute", ftp://ftp.ee.lbl.gov/traceroute.tar.gz, February 1989.

[74] K. Maheswari, M. Punithavalli, "A Survey of Packet Loss in VoIP", International Journal of Computational Intelligence Research, Vol. 5, No. 1, 2009, pp. 57–66.

[75] P. Axensten, "Cauchy CDF, PDF, inverse CDF, parameter fit and random generator", http://www.mathworks.com/matlabcentral/fileexchange/1174.

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL I.P.N.
UNIDAD GUADALAJARA

El Jurado designado por la Unidad Guadalajara del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional aprobó la tesis

Análisis y modelado de procesos auto-similares con aplicación en comunicaciones de VoIP

Self-similar time series: analysis and modeling with applications to VoIP

del (la) C.

Leopoldo ESTRADA VARGAS

el día 27 de Febrero de 2015.

Dr. Deni Librado Torres Román
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara

Dr. Yuriy Shkvarko Sosnoff
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara

Dr. Ramón Parra Michel
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara

Dr. Mario Angel Siller González
Pico
Investigador CINVESTAV 3A
CINVESTAV Unidad Guadalajara

Dr. Pablo Velarde Alvarado
Profesor de Tiempo Completo
Universidad Autónoma de Nayarit