

On the Detection of LRD Phenomena

Houssain Kettani

ECECS Department

Polytechnic University of Puerto Rico

San Juan, Puerto Rico 00919

hkettani@upr.edu

John A. Gubner

ECE Department

University of Wisconsin

Madison, Wisconsin 53706

gubner@engr.wisc.edu

Abstract—A new model-testing paradigm is introduced. The proposed method uses the structure of the autocorrelation function of the model that we would like to fit the data to. We start by estimating the autocorrelation function of the data, and then apply a curve-fitting criterion, which we call the optimization method. If the resultant error is high, then the given process fails the test and may not be considered to follow that particular model. Otherwise, the process is assumed to be long-range dependent following the assumed model with the parameter as estimated. The criteria that we followed to decide whether the error is large or not is the probability of false alarm. For a wide range of the parameter, we develop a relation between the the probability of false alarm and the cutoff to decide how large the error can be. This paradigm is illustrated through the Second-Order Self-Similar (SOSS) model which is an example of a Long-Range Dependent (LRD) model. We perform an empirical study using artificial SOSS data to make a proper decision on the cutoff and to obtain empirical confidence intervals and bias. To better evaluate this proposed method, we tested the method on two sets of real data that are known to be LRD, and on artificial and real data that is known *not* to be LRD. Although we focus on processes that are LRD, the new method is readily generalizable to other processes. In other words, the new method can be used as a tool to check the validity of a model in characterizing certain process.

Index Terms—Estimation, Second-Order Self-Similarity (SOSS), Long-Range Dependence (LRD), Network Traffic.

I. INTRODUCTION

Long-range dependence processes were observed as early as 1895 in astronomical data sets studied by the astronomer Newcomb. More concrete work on long-range dependence and self-similar processes was presented by Hurst and Mandelbrot in early 1950s. Since then, this phenomenon was empirically shown to exist in a number of fields, such as agronomy, astronomy, chemistry, economics, engineering, environmental sciences, geosciences, hydrology, mathematics, physics, statistics and computer network.

Internet users are well-acquainted with congestion, sometimes known as the “World-Wide Wait” (WWW). Due to variable amounts of data sent over the Internet, buffers at routers can become quite full. Thus, you may experience delays in downloading files from the Internet because your data is waiting at a congested router until it can be forwarded to you. Sometimes, buffers at routers are full, and the router must drop packets. In this case, your wait may be even longer because the dropped packets must be retransmitted.

Internet routers were originally designed using statistical models of telephone traffic. However, in early 1990s it was shown that telephone traffic models, e.g., Poisson models, do not hold for Local Area Network (LAN) traffic [16], Wide Area Network (WAN) traffic [23], and World Wide Web (WWW) traffic [4]. Instead, models exhibiting long-range dependence (LRD) and self-similarity are more appropriate. Hence, there has been much recent work on modifying TCP algorithms, buffer sizes, algorithms for smart routers, etc., to deal better with the observed congestion and packet loss.

The most well-known models of long-range dependent processes are fractional Gaussian noise [19] (thus second-order self-similarity) and fractional ARIMA [5], [10]. Each of these models has a corresponding long-range dependence parameter β ($\beta \in (0, 2)$). The smaller the value of β , the more long-range dependent the process is. Since the value of the parameter β indicates the intensity of this dependence structure, it is important to have a better tool to estimate it, so that the estimate is not biased and the confidence intervals are as small as possible. Moreover, if we would like to estimate β on-line, then the estimation tool should be as fast as possible.

LRD model degrades the quality of service of the corresponding network traffic due to longer queueing delays and higher packet loss than the Poisson model [16], [23]. This necessitates network bandwidth provisioning and traffic controls that take the newer model into consideration. For instance, the effective bandwidth [20], [21], the server service rate [14], the network utilization factor [14] and asymptotic bounds for buffer overflow probability [17] have been expressed in terms of the parameter β . Thus, the detection of LRD phenomenon in a record of data gives researchers an access to a wealth of theoretical and empirical results. Moreover, a more accurate estimate of the LRD parameter results in a more accurate estimation of other parameters that enhance the analysis and modelling of the studied process.

In network traffic, long-range dependence corresponds to slowly decaying autocorrelation functions and heavy-tailedness. The former shows the existence of nontrivial correlation structure at large scales. This in turn leads to the “ $1/f$ noise,” which implies larger contributions of low frequency components. Heavy-tailedness on the other hand, indicates that large sample values have a nonnegligible probability. Thus, samples drawn from a heavy-tailed distribution result in a bulk of small values and another bulk of relatively very large values.

Not surprisingly, this corresponds to extreme variability and slows down the convergence rate of sample statistics. This in turn explains the burstiness observed in network traffic. Such burstiness forces packets to experience long delays and some packets are even dropped due to buffer overflow. This unpleasant behavior of network traffic introduces difficulty and complexity into traffic and resource management. Nevertheless, the long-range dependence structure helps predicting future sample values.

To study network traffic, we start by considering it as a random process $Y_i, i \in \mathbf{Z}$. It is of interest to develop a model of this process so that we can predict the future values with certain probability. This will allow us to develop better congestion control mechanisms, like Transmission Control Protocol (TCP), buffer sizes, packet spacing, routers, better algorithms for smart routers, etc. This in turn will reduce the loss of data being transmitted and increase the speed of transmission. In dealing with network traffic, we are most interested in the corresponding increment process. This process was shown to be long-range dependent in the work of Leland *et al.* [16]. Thus, to study network traffic, we need to invoke the theory of probability and random processes.

Several methods for measuring the long-range dependence parameter β have been proposed. The most well-known methods are the R/S method [19], [15], variance-time analysis [15], [16], the periodogram method [15], [16], the Whittle estimator [29], and the wavelet method [1], [22]. All these methods except Whittle's are graphical. Only the last two give confidence intervals. The least computationally intensive method of these methods is the wavelet method.

The rest of this paper is organized as follows. Section II covers background definitions and theorems to be used later. Section III proposes a method for deciding whether or not a sample of a process has a given parametric long-range dependence correlation structure. Section IV presents an empirical study to validate the new method. A summary of the proposed method and concluding remarks are presented in Section V.

II. PRELIMINARIES

A. Second-Order Stationarity

Consider a discrete-time stochastic process $X_i, i \in \mathbf{Z}$ [6], [22], where X_i is viewed as the increment process of network traffic, measured in packets, bytes, or bits. We say that X_i is *strongly stationary* if the families $(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ and $(X_{i_1+k}, X_{i_2+k}, \dots, X_{i_n+k})$ have the same joint distribution for all $i_1, i_2, \dots, i_n, k \in \mathbf{Z}$ and positive integers n . This form of stationarity turns out to be highly restrictive, and a weaker condition suffices for our purposes.

We say that X_i is *second-order stationary* if its mean and autocovariance function, respectively, satisfy

$$E(X_i) = E(X_j), \quad (1)$$

and

$$\gamma(X_i, X_j) = \gamma(X_{i+k}, X_{j+k}), \quad (2)$$

for all i, j and $k \in \mathbf{Z}$. The notation γ denotes the autocovariance function defined as

$$\gamma(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]. \quad (3)$$

Thus, if X_i is a second-order stationary process, it has a constant mean, and its autocovariance function is a function of $k = i - j$ only, which allows us to write $\gamma(X_i, X_j) = \gamma(k)$. Note that $\gamma(k)$ is an even function for real processes, i.e., $\gamma(k) = \gamma(-k)$. We put $\sigma^2 = \gamma(0)$ and $\rho(k) = \frac{\gamma(k)}{\sigma^2}$ to denote the variance and autocorrelation function of the process X_i .

For traffic-modeling purposes, we would like X_i to be at least second-order stationary so that its behavior or structure is invariant with respect to shifts in time. Without this property, a model loses much of its usefulness as a compact description of the assumed tractable phenomena [22].

B. Second-Order Self-Similarity

The notion of self-similarity or scale-invariance arises in many fields. To get a deeper understanding of this notion we start by explaining the self-similarity phenomenon on geometric images. A geometric image is said to be self-similar if there exists a piece of this image (a self-similar piece) that if magnified properly will give exactly the same image. An example of such is the Sierpinski triangle.

Statisticians and probabilists on the other hand used this scale-invariant phenomenon to define self-similar processes, which are defined as follows [22]. Y_t is a self-similar process with self-similarity parameter H if for all $a > 0$ and $t \geq 0$,

$$Y_t =_d a^{-H} Y_{at},$$

where the notation $=_d$ denotes equality in distribution. This means that Y_t and its normalized (by a^{-H}) time scaled version Y_{at} have the same distribution.

In the network traffic modeling context, Y_t can be thought of as the cumulative process or the total traffic up to time t . Analogous to fractals, for $a > 1$ where time is dilated, a contraction factor a^{-H} is applied so that the magnitudes of Y_{at} and Y_t are comparable. Likewise, for $a < 1$ the opposite holds true.

The self-similarity parameter H is also known as the Hurst parameter (which explains the notation) named after the British hydrologist H. E. Hurst (1880–1978) who was studying the Nile river minima and published his observations in [12]. Negative values of H are prohibited since the corresponding Y_t is not a measurable process [27], [28]. The value $H = 0$ is not interesting since it implies that for all $t > 0$, $Y_t = Y_1$ with probability one.

We consider the case where Y_t has finite variance and stationary increments. We also take $Y_0 = 0$ with probability one. Thus, define the increment process $X_i = Y_i - Y_{i-1}$ ($i = 1, 2, \dots$). With this setup, it can be shown that X_i has zero mean and a correlation given by

$$\rho(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), \quad k \geq 1. \quad (4)$$

A process with the same second-order statistics as X_n is referred to as a *Second-Order Self-Similar* (SOSS) process.

Define the *aggregated process* $X_n^{(m)}$ of X_i at *aggregation level* m as

$$X_n^{(m)} = \frac{1}{m} \sum_{i=m(n-1)+1}^{mn} X_i. \quad (5)$$

That is, to create $X_n^{(m)}$, partition X_i into non-overlapping blocks of size m , average their values, and then use n to index these blocks. In network traffic, this can be viewed as “zooming out” in time, i.e., since many increments of X_i are juxtaposed in time, zooming out makes the juxtaposed increments appear as one increment, as shown in [15].

Let $\rho^{(m)}(k)$ denote the autocorrelation function of $X_n^{(m)}$. Then, X_i is said to be *asymptotically second-order self-similar* if for $k \geq 1$

$$\lim_{m \rightarrow \infty} \rho^{(m)}(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}). \quad (6)$$

Let $\rho(k)$ denote the *autocorrelation function* of X_i , i.e., $\rho(k) = \frac{\gamma(k)}{\sigma^2}$. A value $H > 1$ is prohibited since it contradicts the fact that $|\rho(k)| \leq 1$ for all k . The case $H = 1$ implies that $\rho(k) = 1$ for every k , which is of no practical importance. Note also that with $H = \frac{1}{2}$, the X_i 's are uncorrelated. Hence, throughout this paper, we consider the range $0 < H < 1$ only. The presence of self-similarity in network traffic was observed by Leland *et al.* [15], [16]. Since then, second-order self-similarity has become a dominant framework for modeling network traffic.

An example of the self-similar process Y_t is fractional Brownian motion, first introduced by Mandelbrot [19]. The corresponding increment process X_i is fractional Gaussian noise. When $H = \frac{1}{2}$, fractional Brownian motion coincides with the ordinary Brownian motion.

C. Implications and Long-Range Dependence

Let X_i be a second-order self-similar or a fractional ARIMA(0, d , 0) process. If the aggregated process $X^{(m)}$ is viewed as the sample mean of X , i.e.,

$$X^{(m)} = \frac{1}{m} \sum_{i=1}^m X_i,$$

then it is not hard to show that

$$(\sigma^{(m)})^2 \approx \sigma^2 m^{-\beta}, \quad (7)$$

as $m \rightarrow \infty$, where $(\sigma^{(m)})^2$ is the variance of $X^{(m)}$, $\beta = 2 - 2H$ in the SOSS case. In fact, for exact SOSS processes, (7) holds with equality for all values of m .

It also follows from (4) that the autocorrelation functions decay hyperbolically rather than exponentially fast. More precisely,

$$\rho(k) \sim c_\rho k^{-\beta} \quad \text{as } k \rightarrow \infty, \quad (8)$$

where β is as before, and the constant $c_\rho = H(2H - 1)$ in the SOSS case.

Let $f(\lambda)$ denote the spectral density function corresponding to the increment process X_i . By definition, the spectral density

function $f(\lambda)$ is the discrete-time Fourier-transform of the autocorrelation function $\rho(k)$, i.e.,

$$f(\lambda) = \sum_{k=-\infty}^{\infty} e^{-j2\pi k\lambda} \rho(k),$$

where $j^2 = -1$.

As noted in [3, p. 43], (8) is equivalent to

$$f(\lambda) \sim c_f |\lambda|^{-\alpha} \quad \text{as } \lambda \rightarrow 0, \quad (9)$$

where $\alpha = 1 - \beta$ and

$$c_f = \frac{\sigma^2 c_\rho}{\Gamma(1 - \alpha) \sin(\frac{\alpha\pi}{2})}.$$

We note here that the aforementioned equivalence is in the sense that (8) and (9) each imply (10). Strictly speaking, from a purely mathematical point of view, this equivalence does not hold; see [7] for more details.

We say the process or time series X_i is *long-range dependent* if

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty. \quad (10)$$

The process is *short-range dependent* otherwise.

Thus, in view of (8), for $\beta \in (0, 1]$ the process X_i is long-range dependent and it is short-range dependent for $\beta \geq 1$. From (9), when X_i is long-range dependent, the spectral density $f(\lambda)$ diverges around the origin implying larger contributions of low frequency components. In this case we say that the spectral density obeys a power law near the origin and we deal with “1/ f noise” [19].

The presence of long-range dependence has both positive and negative effects. For such processes, the forecasting becomes easier in the sense that good short- and long-term predictions can be obtained when a long record of past values is available (see [3] for details). On the other hand, the classical time series approach to estimating the second-order statistics assumes short-range dependence. The presence of long-range dependence has been shown to significantly slow the speed of convergence of the estimates. The rate of convergence becomes slower as the long-range dependence increases (β decreases). The following section discusses this issue in more detail.

D. Statistical Sampling

Since we are dealing with measurements and no a priori probability density function, we use statistical sampling to estimate the second-order statistics of the process.

Let X_i be a second-order stationary process with mean, variance, and covariance μ , σ^2 , and $\gamma(k)$, respectively. The *sample mean* and the *sample covariance* are given by the following formulas:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad (11)$$

and

$$\hat{\gamma}_n(k) = \frac{1}{n} \sum_{i=1}^{n-k} (X_i - \hat{\mu}_n)(X_{i+k} - \hat{\mu}_n), \quad (12)$$

where n is the number of samples to be used. The *sample variance* is given by

$$\hat{\sigma}_n^2 = \hat{\gamma}_n(0).$$

Likewise, the *sample autocorrelation* is given by

$$\hat{\rho}_n(k) = \frac{\hat{\gamma}_n(k)}{\hat{\sigma}_n^2}. \quad (13)$$

The relationship between these estimates and the estimated parameters (the asymptotic distribution of the difference, bias, etc.) was studied extensively for both the short-range and long-range dependence cases (see [2] for details on the former and [11] on the latter). Hosking's results [11] focused on processes with hyperbolically decaying autocorrelation functions as in (8). In this paper, we will be interested in the relationship between $\hat{\rho}$ and ρ .

In [11], Hosking assumes that the process X_i has the representation

$$X_i = \mu + \sum_{j=0}^{\infty} \psi_j a_{i-j}, \quad (14)$$

where

$$\psi_j \sim \delta j^{-\frac{1}{2}(1+\beta)}, \quad \delta > 0, \quad \text{as } j \rightarrow \infty, \quad (15)$$

and a_i is a white-noise process consisting of independent, identically distributed $N(0, \sigma^2)$ random variables. As noted by Hosking [11], both fractional Gaussian noise and fractional ARIMA processes have the representation (14) and (15). For such process, the following two theorems (see [11, Theorems 6 and 7]) provide the asymptotic bias, covariance, and limiting distribution of the sample autocovariances $\hat{\rho}_n(k)$, $k \geq 1$.

Theorem 2.1: Let X_i be a time series satisfying (14) and (15). Then as $n \rightarrow \infty$, the *asymptotic bias* and *covariance* of $\hat{\rho}(k)$, $k \geq 1$, is given by

$$E(\hat{\rho}_n(k)) - \rho(k) \sim \frac{-2c_\rho(1-\rho(k))n^{-\beta}}{(1-\beta)(2-\beta)}, \quad (16)$$

while the *asymptotic covariance* of $\hat{\rho}(k)$, $k \geq 1$, is given by

$$\text{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim 2c_\rho^2(1-\rho(k))(1-\rho(l))K_2 n^{-2\beta}, \quad (17)$$

if $0 < \beta < \frac{1}{2}$,

$$\text{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim 4c_\rho^2(1-\rho(k))(1-\rho(l))n^{-1} \log(n), \quad (18)$$

if $\beta = \frac{1}{2}$, and

$$\text{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim n^{-1} \sum_{s=-\infty}^{\infty} \left[\rho(s)\rho(s+k-l) + \rho(s)\rho(s+k+l) + 2\rho(k)\rho(l)\rho(s)^2 - 2\rho(k)\rho(s)\rho(s+l) - 2\rho(l)\rho(s)\rho(s+k) \right], \quad (19)$$

if $\frac{1}{2} < \beta < 2$, where K_2 is defined shortly.

Theorem 2.2: Let X_i be a time series satisfying (14) and (15). Then as $n \rightarrow \infty$,

- 1) If $0 < \beta < \frac{1}{2}$, and $R_k := n^\beta(\hat{\rho}_n(k) - \rho(k))/(1 - \rho(k))$, then as $n \rightarrow \infty$, the common limiting distribution of the R_k has r th cumulant

$$\kappa_r = c_\rho^r 2^{r-1} (r-1)! K_r, \quad (20)$$

where

$$K_1 = \frac{-2}{(1-\beta)(2-\beta)},$$

$$K_r = \int_0^1 \cdots \int_0^1 g(x_1, x_2) g(x_2, x_3) \cdots g(x_{r-1}, x_r) g(x_r, x_1) dx_1 dx_2 \cdots dx_r, \quad r \geq 2,$$

with

$$g(x, y) = |x - y|^{-\beta} + \frac{2}{(1-\beta)(2-\beta)} \frac{x^{1-\beta} + (1-x)^{1-\beta} + y^{1-\beta} + (1-y)^{1-\beta}}{1-\beta}.$$

- 2) If $\beta = \frac{1}{2}$, and $R_k := (n/\log n)^{\frac{1}{2}}(\hat{\rho}_n(k) - \rho(k))/(1 - \rho(k))$, then as $n \rightarrow \infty$, the common limiting distribution of the R_k is $N(0, 4c_\rho^2)$.

- 3) If $\frac{1}{2} < \beta < 2$, and $R_k := n^{\frac{1}{2}}(\hat{\rho}_n(k) - \rho(k))$, then as $n \rightarrow \infty$, R_k has a limiting distribution that is multivariate normal zero mean and covariances given by n times (19).

We note here that this latter result for $\beta \in (\frac{1}{2}, 2)$ was found by Anderson [2]. From (16) it is seen that the smaller β is (or equivalently, the higher the long-range dependence) the slower the decay of the bias of the sample autocorrelations as the sample size n increases.

III. THE PROPOSED METHOD

Suppose we have a record of network traffic that records the total packets or bytes received in a period of time. Let X_i denote its increment process. There is always a vast interest in the process X_i by the networking community, since understanding the behavior of this process allows us to develop better congestion control mechanisms, like Transmission Control Protocol (TCP), buffer sizes, packet spacing, routers, better algorithms for smart routers, etc.

Define the *error function* $E_K(\beta)$ as

$$E_K(\beta) = \frac{1}{4K} \sum_{k=1}^K \{\rho(k) - \hat{\rho}_n(k)\}^2, \quad (21)$$

where $\rho(k)$ denotes the autocorrelation function of the model with parameter β that we would like to fit the data to, $\hat{\rho}_n(k)$ is the sample autocorrelation function of the data and K is the largest value of k for which $\hat{\rho}_n(k)$ is to be computed to reduce edge effects. Since we estimate the model parameter β based on optimizing (21), we call this estimation method the *optimization method*.

How small should $E_K(\beta)$ be? We expect $E_K(\beta)$ to be close to zero if X_i is close to the model. The estimated parameter $\hat{\beta}$ is chosen so that $E_K(\hat{\beta})$ is the minimum of the error function over the appropriate range of the parameter. It is easily seen that the highest $E_K(\hat{\beta})$ can be is one. Thus, we consider that the prescribed model fits the process X_i if $E_K(\hat{\beta}) = e$, where e is “much smaller than one”.

Let U_K be the row vector with entries

$$u_k = \rho(k) - \hat{\rho}_n(k), \quad k = 1, 2, \dots, K.$$

Thus, the error function can be rewritten as

$$E_K(\beta) = \frac{U_K U_K^T}{4K}. \quad (22)$$

Define the probability of false alarm as

$$P_{FA} = P(E_K(\hat{\beta}) \geq e | R = R_\beta), \quad (23)$$

where R is the covariance matrix of U_K and the condition $R = R_\beta$ denotes that the process follows the model with the corresponding covariance $R = R_\beta$. Thus, we would like to pick e so that $P_{FA} \leq 0.05$.

Next, note that

$$\begin{aligned} P_{FA} &= P(E_K(\hat{\beta}) \geq e | R = R_\beta) \\ &\leq P(E_K(\beta) \geq e | R = R_\beta) \\ &= P\left(\frac{U_K U_K^T}{4K} \geq e | R = R_\beta\right) \\ &= P(U_K U_K^T \geq 4Ke | R = R_\beta) \\ &= 1 - S_K(4Ke), \end{aligned} \quad (24)$$

where S_K is the cdf of $U_K U_K^T$.

Theorem 3.1: For $\beta \in [\frac{1}{2}, 2)$, S_K is the cumulative distribution function of *Stacy's distribution* with the corresponding probability density function given by

$$s_{K,\lambda}(t) = t^{\frac{K}{2}-1} \sum_{k=0}^{\infty} \frac{(-t)^k}{\Gamma(\frac{K}{2} + k)} \sum_{l=1}^K \prod_{m_l=k}^{n=1} \frac{\Gamma(m_n + \frac{1}{2})(2\lambda_n)^{m_n + \frac{1}{2}}}{m_n! \sqrt{\pi}},$$

where $\Gamma(\cdot)$ denotes the gamma function, and $\lambda = [\lambda_1, \dots, \lambda_K]$, where λ_n , for $n = 1, 2, \dots, K$, are the eigenvalues of the covariance matrix R_β of U_K .

Proof: From Theorem 2.2, for $\beta \in [\frac{1}{2}, 2)$, U_K is zero mean multivariate normally distributed with covariance R with the corresponding entries

$$r_{kl} = \text{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)).$$

The distribution of $U_K U_K^T$ has the following characteristic function [25, p. 65]:

$$\begin{aligned} \Phi(\omega) &:= E[e^{j\omega U_K U_K^T}] \\ &= \prod_{n=1}^K (1 - j\omega 2\lambda_n)^{-\frac{1}{2}}. \end{aligned}$$

But this is the characteristic function of the sum of the independent random variables v_n , $n = 1, 2, \dots, K$, each distributed as $\Gamma(\frac{1}{2}, 2\lambda_n)$, where $\Gamma(a, b)$ is the Gamma distribution

with probability density function

$$\gamma_{a,b}(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{b^a \Gamma(a)}, \quad x > 0.$$

The distribution of such sums is discussed by Stacy [24], from which we get Stacy's distribution of the Theorem. This concludes the proof.

We next consider the asymptotics of $S_K(t)$. To do this, we invoke the following form of the *Central Limit Theorem* (see [18, p. 287] for proof).

Theorem 3.2: Let w_1, w_2, \dots be independent variables satisfying

$$E(w_j) = 0, \quad \text{var}(w_j) = \sigma_j^2, \quad E(|w_j^3|) < \infty$$

and such that

$$\frac{1}{\sigma(n)^3} \sum_{j=1}^n E(|w_j^3|) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (25)$$

where

$$\sigma(n)^2 = \sum_{j=1}^n \text{var}(w_j).$$

Then as $n \rightarrow \infty$,

$$\frac{1}{\sigma(n)} \sum_{j=1}^n w_j \xrightarrow{D} N(0, 1).$$

This gives us the following result.

Lemma 3.1: For $\beta \in [\frac{1}{2}, 1)$, suppose (25) holds, then for large K ,

$$S_K \sim N\left(\sum_{j=1}^K \lambda_j, 2 \sum_{j=1}^K \lambda_j^2\right).$$

Note that in our analysis, (25) is equivalent to

$$\frac{\sum_{j=1}^K \lambda_j^3}{\left(\sum_{j=1}^K \lambda_j^2\right)^{\frac{3}{2}}} \rightarrow 0 \quad \text{as } K \rightarrow \infty.$$

An empirical study shows that for $\beta \in [\frac{1}{2}, 1)$, as K increases the ratio decreases. Thus, in the remainder we apply Lemma 3.1 to find the right number e .

Figure 1 present the minimum e to get $P_{FA} < 0.05$ for different β values in $[\frac{1}{2}, 2)$ in SOSS case ($H = 1 - \frac{\beta}{2}$). To obtain the corresponding e value, the Normal distribution of Lemma 3.1 was applied with $n = 4000$ and $K = 50$. From Figure 1, we see that $e \approx 1.5 \times 10^{-4}$ for $H < 0.60$. As the LRD increases ($H \in (0.60, 0.75)$), e increases up to about 3.5×10^{-3} . As can readily be seen from (24), this e is an upper bound on the actual e . In the following section, an empirical study is performed to find the appropriate value e for the whole range of the parameter.

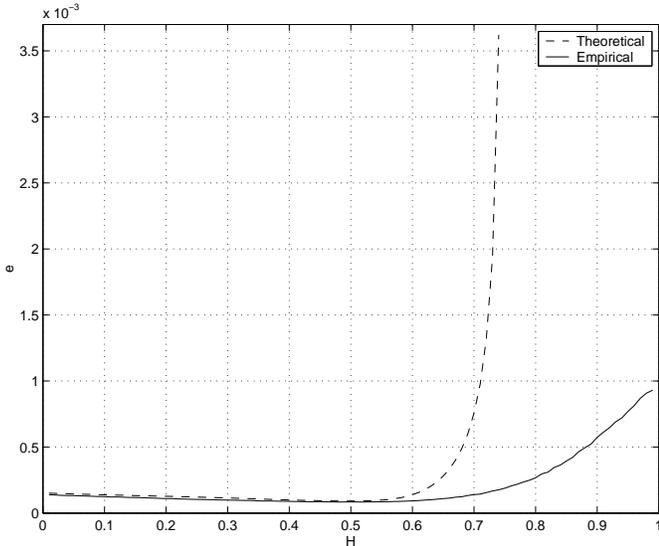


Fig. 1. A plot of $\arg\min_e 1 - S_K(4Ke)$ in (24) and the empirical upper bound e vs. the Hurst parameter $H = 1 - \frac{\beta}{2}$ in SOSS case. Here S_K is taken to be normal, based on Lemma 3.1.

IV. EMPIRICAL STUDY

A. SOSS Artificial Data

For each $H = 0.01, 0.02, \dots, 0.99$, we start by generating 10^4 realizations of a fractional Gaussian noise using MATLAB. The length of each realization is $n = 4000$ points. In our analyses we take $K = 50$ in (21) to reduce edge effects. Based on (23), for each H , e is chosen so that 95% of the obtained $E_K(\hat{H})$ are smaller than e . The plot of e versus H is given in Figure 1. For $H < 0.60$, the upper bound on e (theoretical e in (24)) and the empirical value of e almost coincide. As the long-range dependence increases, the upper-bound becomes looser and looser. On the other hand, as the number of realizations increases, the empirical plot of e gets smoother and decreases slightly on the range $H \in (0.80, 1)$ (not shown).

Since the empirical plot of e vs. H is less than 10^{-3} for $H \in (0, 1)$, the cutoff that we take as our measuring criteria is the value $e = 10^{-3}$. Thus, if the error function is less than 10^{-3} (i.e., $E_K(\hat{H}) < 10^{-3}$), then the process X_i is accepted as fitting the prescribed second-order self-similar model, otherwise it is not (which may be also due to the lack of sufficient data).

To illustrate the power of our test, consider the following short-range dependent process. Let X_i be an $\text{ARIMA}(0.90, 0, 0) = \text{AR}(0.90)$. Then applying the optimization method results in $E_K(\hat{H}) = 7.4 \times 10^{-3}$. Thus, the optimization method declares the process X_i not second-order self-similar in agreement with the truth.

B. Real Data

In this section, we study standard measurements where second-order self-similarity was observed. Namely, the Nile river data, and the Bellcore data. Throughout this study, we

take $K = 50$ in the error function formula (21). We then consider a process that is *not* second-order self-similar, namely the Variable Bit Rate data. A summary of the results is provided in Table I.

The Nile river data consists of the yearly minimum water levels of the Nile river measured at the Roda Gauge near Cairo, Egypt, for the years 622 – 1281 [26]. This data led Hurst to the observation of what was later called the Hurst effect, or self-similarity [12]. Applying the optimization method to this data results in the error function $E = 2.15 \times 10^{-4}$. Thus, the optimization method acknowledges the fact that this process is second-order self-similar.

The Bellcore data consists of Ethernet measurements for a local area network traffic at Bellcore, Morristown, New Jersey [15], [16]. It was collected on August 29, 1989 and lasted for about fourteen minutes. Each observation represents the number of packets sent over the Ethernet per 100 ms. This data was considered by Leland *et al.* and used to show that network traffic is second-order self-similar. Passing this data through the optimization method gives the error function $E = 1.15 \times 10^{-4}$. Thus, the optimization method acknowledges again the fact that this process is second-order self-similar.

Data	Nile	BC	VBR
$E_K(\hat{H})$	2.15×10^{-4}	1.15×10^{-4}	6.74×10^{-3}

TABLE I

A SUMMARY OF THE RESULTS OF THE APPLICATION OF THE OPTIMIZATION METHOD TO DIFFERENT SETS OF REAL DATA.

The last real data that we consider, is the VBR data that is known *not* to be second-order self-similar. This data consists of the amount of codec information per frame for a certain video scene. The scene consists of a conversation between three people sitting at a table. No change in the background and no movement of the camera exist. The codec that was used is called variable-bit-rate (VBR). This data was gathered in 1991 by engineers at Siemens, Munich, Germany (see [8], [9] and [3]). Application of the optimization method to this data results in $E = 6.74 \times 10^{-3}$. Since this value is greater than 10^{-3} , the optimization method declares that this data is not second-order self-similar, in agreement with what we already know.

V. SUMMARY AND CONCLUDING REMARKS

In this paper, we have presented a new tool to decide whether or not a process has a given parametric correlation structure. The cutoff value of the error function to decide whether the given process follows the prescribed model or not was found empirically so that the probability of false alarm is less than 0.05. The new method is tested on pseudo random data over various ranges of the long-range dependence parameter and on real data. The new method was shown to successfully answer the question of whether the studied process follows the prescribed model or not.

A summary of the optimization method is as follows:

- Let X_1, X_2, \dots, X_n be a realization of a Gaussian process,
- Compute $\hat{\rho}_n(k)$ as in (13),
- Compute the error function $E_K(\beta)$ as in (21),
- Let $E_K(\hat{\beta}) = e$, where $\hat{\beta}$ is the minimizer of the error function,
- If $e < 10^{-3}$, then the process is LRD with parameter $\hat{\beta}$. Otherwise, the process is not, or the data is not sufficient (n and K are not large enough) to make the right judgment.

Many real processes possess the long-range dependent characteristics discussed in Section II. In fact, some processes were shown to follow a particular model. In particular, we mention the results of Leland *et al.* [16], where it was shown that local area network (LAN) traffic is *exactly* second-order self-similar. Thus, knowledge of the Hurst parameter determines the second-order characteristics of such traffic. With this as motivation, we formulate the ideas of this section as follows.

Suppose that the process X_i is known to follow a particular model. Then the use of a one-lag autocorrelation function is justifiable, i.e., $K = 1$ in (21). In this case, the error function can be rewritten as

$$E_1(\beta) = [\rho(1) - \hat{\rho}_n(1)]^2. \quad (26)$$

This provides further simplifications of the optimization method. Mainly, it allows us to obtain theoretical confidence intervals of the estimated long-range dependence parameter $\hat{\beta}$ and makes the method much faster. This special case was presented in [13].

REFERENCES

- [1] P. Abry and D. Veitch (1998). Wavelet Analysis of Long-Range Dependent Traffic. *IEEE Transactions on Information Theory*, 44(1):2–15.
- [2] T. W. Anderson (1958). *The Statistical Analysis of Time Series*. John Wiley & Sons, New York, New York.
- [3] J. Beran (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York, New York.
- [4] M. Corvella and A. Bestavros (1997). Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5:835–846.
- [5] C. W. J. Granger and R. Joyeux, (1980). An Introduction to Long-Memory Models and Fractional Differencing *Journal of Time Series Analysis*, 1:15–29.
- [6] G. R. Gimmatt and D. R. Stirzaker (1994). *Probability and Random Processes*. Clarendon Press, Oxford, United Kingdom.
- [7] J. A. Gubner (2005) Fallacies in the Theory of Long-Range-Dependent Processes. *IEEE Transactions on Information Theory*, 51(3):1234–1239.
- [8] H. Heeke (1991). Statistical Multiplexing Gain for Variable Bit Rate Codecs in ATM Networks. *International Journal of Digital and Analog Communication Systems*, 3:261–268.
- [9] D. Heyman, A. Tabatabai and T. V. Lakshman (1991). Statistical Analysis and Simulation of Video Teleconferencing in ATM Networks. *IEEE Transactions on Circuits and Systems Video Technology*, 2:49–59.
- [10] J. R. M. Hosking (1981). Fractional Differencing. *Biometrika*, 68:165–176.
- [11] J. R. M. Hosking (1981). Asymptotic Distribution of the Sample Mean, Autocovariances, and Autocorrelations of Long-memory Time Series. *Journal of Econometrics*, 73:261–284.
- [12] H. E. Hurst (1951). Long-term Storage Capacity of Reservoirs *Transactions of the American Society of Civil Engineers*, 116:770–779.
- [13] H. Kettani and J. Gubner (2006). A Novel Approach to the Estimation of the Long-Range Dependence Parameter *IEEE Transactions on Circuits and Systems II*, 53(6):463–467.
- [14] Y. Koucheryavy, A. Krendzel, S. Lopatin, J. Harju. (2002) *Performance Estimation of UMTS Release 5 IM-Subsystem. Elements*, Proceedings of The 4th IEEE Conference on Mobile and Wireless Communications Networks(MWCN).
- [15] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. (1993) On the Self-Similar Nature of Ethernet Traffic. In *Proceedings of ACM SIGCOMM'93*.
- [16] W. Leland, M. Taqqu, W. Willinger, and D. Wilson (1994). On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 2(1):1–15.
- [17] N. Likhanov (2000). Bounds on the Buffer Occupancy Probability with Self-Similar Input Traffic. in K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation* pp. 193–213, Wiley-Interscience.
- [18] M. Loeve (1977) *Probability Theory*, Vol. 1 (4th edition). Springer, Berlin.
- [19] B. B. Mandelbrot (1983). *The Fractal Geometry of Nature*. W. H. Freeman and Co., New York, New York.
- [20] I. Norros (1994). A Storage Model with Self-Similar Input. In *Queueing Systems*, 16:387–396.
- [21] I. Norros (1995). On the Use of Fractional Brownina Motion in the Theory of Connectionless Networks. In *IEEE Journal on Selected Areas in Communications*, 13(6):953–962.
- [22] K. Park and W. Willinger (2000). Self-Similar Network Traffic: An Overview *Self-Similar Network Traffic and Performance Evaluation*. K. Park and W. Willinger (editors), John Wiley & Sons, New York, New York.
- [23] V. Paxson and S. Floyd (1995). Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244.
- [24] E. W. Stacy (1962) A Generalization of the Gamma Distribution. *Annals of Mathematical Statistics*, 33(3): 1187–1192.
- [25] L. L. Scharf (1991) *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* Addison-Wesley Publishing Company.
- [26] O. Tousson (1925). Mémoire Sur L'Histoire Du Nil. *Mémoires de l'institut d'Egypte*.
- [27] W. Vervaat (1985). Sample Path Properties of Self-similar Processes With Stationary Increments. *Annals of Probability*, 13:1–27.
- [28] W. Vervaat (1987). Properties of General Self-similar Processes. *Bulletin of the International Statistical Institute*, 52(4):199–216.
- [29] P. Whittle (1953). Estimation and Information in Stationary Time Series *Arkiv for Matematik*, 2:423–434.