Comparison of Various Estimators of Hurst Parameter in Simulated FGN

Hae-Duck J. Jeong [a,*], Don McNickle [b], and Krzysztof Pawlikowski [c]

[a] *Department of Internet and Information Science*
Korean Bible University, Seoul, South Korea

[b] *Department of Management, University of Canterbury*
Christchurch, New Zealand

[c] *Department of Computer Science and Software Engineering*
University of Canterbury, Christchurch, New Zealand

Abstract

The Hurst parameter is the simplest numerical characteristic of self-similar long-range dependent stochastic processes. Such processes have been identified in many natural and man-made systems. In particular, since they were discovered in the Internet and other multimedia telecommunication networks a decade ago, they have been the subject of numerous investigations. Typical quantitative assessment of self-similarity and long-range dependency, begins with the estimation of the Hurst parameter H. There have been a number of techniques proposed for this. This paper reports results of a comparative analysis of six the most frequently used estimators of H. To set up a credible framework for this, the minimal acceptable sample size is first determined. The Hurst parameter estimators are then compared for bias and variance. Our experimental results have confirmed that the Abry-Veitch Daubechies Wavelet-Based (DWB) and the Whittle ML (Maximum Likelihood) estimators of H are the least biased. However, the latter has significantly smaller variance and can be applied to shorter data samples than the Abry-Veitch DWB estimator. On the other hand, the Abry-Veitch DWB estimator is computationally simpler and faster than the Whittle ML estimator.

Key words: Hurst parameter estimation, self-similar processes; long-range dependence; teletraffic analysis.

* This work was partially supported by a grant F03-2006-000-10058-0 from the Korea Science and Engineering Foundation (KOSEF). Corresponding author.
Email address: joshua@bible.ac.kr (Hae-Duck J. Jeong).

1 Introduction

Self-similar (or fractal) *long-range dependent* stochastic processes occur in many natural and man-made systems. In particular, since such processes were discovered in the Internet and other multimedia telecommunication networks a decade ago ([14], [15]) they have become the subject of numerous research investigations of their nature and consequences.

Theoretically, one can distinguish two types of stochastic self-similarity. A continuous-time stochastic process $\mathbf{Y}_t$=$\{Y_{t_1}, Y_{t_2}, \ldots\}$ is *strictly self-similar* with *Hurst parameter*, $H, 0.5 < H < 1$, if $\mathbf{Y}_{ct}$ and $c^H\mathbf{Y}_t$ (the rescaled process with time scale ct) have identical finite-dimensional distribution for any stretching factor c, $c > 0$, [5], [17], [27]. This means that, for any sequence of time points $t_1, t_2 \ldots t_n$,

$$\{Y_{ct_1}, Y_{ct_2}, \ldots, Y_{ct_n}\} \stackrel{d}{=} \{c^H Y_{t_1}, c^H Y_{t_2}, \ldots, c^H Y_{t_n}\}, \tag{1}$$

where $\stackrel{d}{=}$ denotes equivalence in distribution. Thus, a strictly self-similar process (or "self-similar in a narrow sense") is self-similar in the sense of probability distribution.

Weak self-similarity of stochastic processes (or "self-similarity in a broad sense") becomes relevant when one restricts analysis of stochastic processes to their first two moments only , i.e. to their means, variances and co-variances. Because of that, it is also known as second-order self-similarity. For example, let a sequence $\{X_1, X_2, \ldots\}$ be a time-stationary stochastic process, defined at discrete time instances $i = 1, 2, 3 \ldots$. Let $E[X_i] = EX$, $Var[X_i] = VarX$, and $\rho_k = E[(X_i - EX)(X_{i+k} - EX)]/VarX$ denote its mean, variance and autocorrelation coefficient of lag k, respectively. Having grouped this sequence of random variables into batches of size m, $m \geq 1$, we can define the aggregated process of $\mathbf{X}^{(m)} = \{X_1^{(m)}, X_2^{(m)}, \cdots\}$, at a given aggregation level m, where $X_i^{(m)} = \frac{1}{m}(X_{im-m+1} + \cdots + X_{im}), i \geq 1$. If $\{X_1, X_2, \ldots\}$ is a weakly self-similar time-stationary process, with $0.5 < H < 1$, then

$$Var[X_i^{(m)}] = m^{2H-2} VarX, \tag{2}$$

and

$$\rho_k^{(m)} = \rho_k, k \geq 0, \tag{3}$$

where $\rho_k^{(m)}$ is the autocorrelation coefficient of lag k of the aggregated process $\mathbf{X}^{(m)}$. Equation (3) means that the original process and its aggregated version have identical correlation structure. Further, it can be proved that in any weakly self-similar process with $0.5 < H < 1$,

$$\rho_k \sim H(2H-1)k^{(2H-2)}, \tag{4}$$

as $k \to \infty$, and

$$\sum_{k=-\infty}^{k=+\infty} \rho_k = \infty, \tag{5}$$

see for example [17]. Thus the autocorrelation function decays very slowly (hyperbolically) with k. Because of these properties, these processes are also known as *long-range dependent*, or strongly correlated.

The Hurst parameter H, known also as the *self-similarity parameter*, is the simplest numerical characteristic of stochastic self-similarity and long-range dependence. Because of that, typical quantitative studies of self-similarity and long-range dependencies, either in data traces recorded in existing networks or in data samples collected during computer simulations of networks, begin with estimation of H. At least a dozen have been different estimation techniques proposed. Only a few partial comparisons of these H estimators have been reported; cf. [6], [13], [28], [29], [30], [31]. The aim of this paper is to fill this gap by reporting results of a comparative analysis of six the most frequently used estimators of H.

To assess the quality of various estimators of H experimentally, we needed to apply them to data samples collected from self-similar stochastic processes of known value of H. This requirement was met by drawing samples from the exactly self-similar stochastic processes known as Fractional Gaussian Noise (FGN).

The six estimators of H considered in this paper belong to three main classes of estimation techniques of H: (i) techniques based on data analysis in the time domain, (ii) techniques based on data analysis in the frequency domain, and (iii) techniques based on wavelet transforms of data. Later we will narrow our discussion to the two best H estimators (the Abry-Veitch DWB estimator [2] and the Whittle ML estimator [31]), that are theoretically asymptotically unbiased and efficient, at least in the FGN case.

This paper is organised as follows. In Section 2 we describe the way in which exact pseudo-random self-similar sequences, used for studying the quality of H estimators, were generated. An appropriate sample size for the study must first be determined. This is discussed in Section 3. Section 4 describes the most frequently used estimation techniques of the Hurst parameter. Quality of the considered estimators of H is discussed in Section 5.

2 Generation of Exact Self-Similar Sequences

As mentioned, we will use an FGN process as the source of exactly self-similar sequences, with known H values. Following the recommendation given in [1],

[7], [30], these exact self-similar FGN sequences were generated by the Durbin-Levinson algorithm. For our purpose, this algorithm, with complexity $O(n^2)$, was found to be sufficiently efficient [21]. A faster generator of FGN sequences, based on the circulant matrix embedding method, has been discussed in [3]. It has complexity $O(M\log M)$, where $M \geq 2(n-1)$. Generation of a sequence of one million numbers representing an FGN process took approximately 8 hours on an Intel Pentium 4 with CPU clock at 2.4 GHz. A number of such sequences were pre-computed, to be used later in the analyse reported in this paper. This rate of generation of pseudo-random self-similar sequences would, of course, be too slow for practical simulation. Practical generators of (approximately) self-similar sequences are discussed in [12].

FGN processes are appropriate for this study as they are frequently used in simulation practise as the primary pseudo-random self-similar processes generated during simulation, to be later transformed into other self-similar processes, with a non-normal marginal probability distribution and a required autocorrelation function; see e.g. [12].

3 Determining Sample Size

Many reported studies of H parameter estimators have used sequences that are obviously too short. Long-range dependence can be observed only if a satisfactorily large sample of data is available, and such a sample needs to be substantially larger than a sample considered to be satisfactory for processes with weaker dependencies. When undertaking a comparative study of different estimators, we should at least ensure that adequate sample sizes are used.

While Mandelbrot and Wallis [16] used a sample of 9 000 numbers for the R/S-statistic, more recent studies have involved larger samples. For example, sequences of 360 000 observations, with each observation representing the number of bytes sent over the Ethernet within 10 milliseconds were used by Leland et al. [15], [32]. A two-hour long empirical sample of VBR video data recorded within 171 000 frames was used in a statistical analysis reported in [9] by Garrett and Willinger; ten samples of 32 768 numbers for estimating the Hurst parameter were used by Paxson [18]; sequences of 40 000 frames of MPEG traffic data were used by Rose [23] in studying a model of VBR MPEG video and its impacts on ATM networks.

We searched for a sample size that would produce estimates of H with negligible sample-size bias, by studying convergence of the estimates to their limit values, as the length of data sequences increased. This process is illustrated in Figs. 1 and 2. The two figures show the convergence of H estimates, calculated for the exactly self-similar FGN process with $H = 0.6$, 0.7, 0.8 and 0.9,

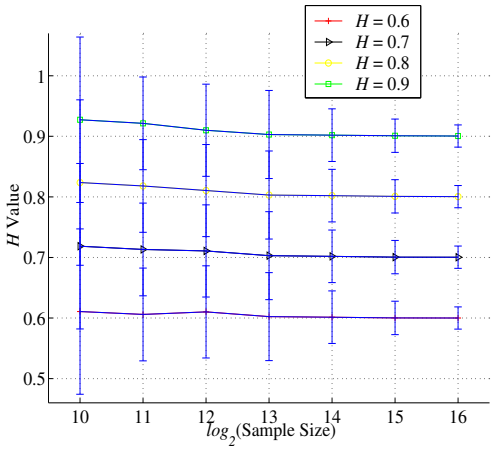

Fig. 1. Estimates of H obtained from the FGN process using the Abry-Veitch DWB estimator as the sample size increases from 2^{10} to 2^{16}.

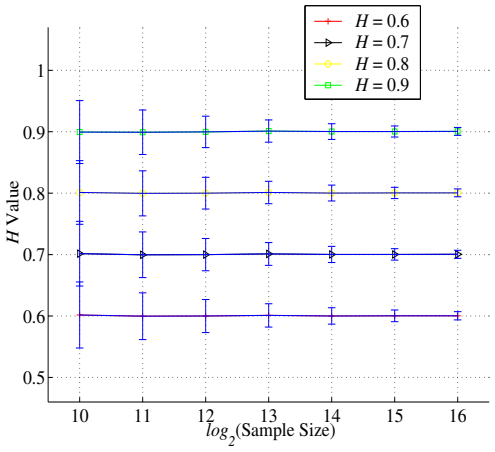

Fig. 2. Estimates of H obtained from the FGN process using the Whittle ML estimator as the sample size increases from 2^{10} to 2^{16}. The vertical bars give 95% confidence intervals for the H values.

Table 1
Relative inaccuracy ΔH of mean values of Abry-Veitch DWB estimates of H in the FGN processes as the sample size increases.

Sample Size	ΔH (%)			
n	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
2^{10}	+1.7870	+2.6530	+2.9670	+3.0340
2^{11}	+0.9794	+1.8880	+2.2650	+2.3760
2^{12}	+1.6810	+1.5360	+1.3170	+1.1110
2^{13}	+0.4068	+0.4232	+0.3838	+0.3292
2^{14}	+0.2352	+0.2762	+0.2615	+0.2278
2^{15}	+0.0258	+0.0920	+0.1096	+0.1075
2^{16}	+0.0095	+0.0593	+0.0672	+0.0603

when using the Abry-Veitch DWB estimator and the Whittle ML estimator, respectively. (These estimators are defined in the next section.) The sample size ranges from 2^{10} to 2^{16}. Fig. 1 shows that for all H values, the Abry-Veitch DWB estimates converge to the required values in all cases. They are positively biased for $n \leq 2^{12}$, asymptotically reaching their appropriate exact values as n increases.

As a measure of bias of the estimators we used their relative inaccuracy ΔH,

Table 2
Relative inaccuracy ΔH of mean values of Whittle ML estimates of H in the FGN processes as the sample size increases.

Sample Size	ΔH (%)			
n	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
2^{10}	+0.2873	+0.2193	+0.1459	- 0.0601
2^{11}	- 0.0512	- 0.0381	- 0.0314	- 0.0891
2^{12}	+0.0032	+0.0050	+0.0046	- 0.0235
2^{13}	+0.1594	+0.1501	+0.1419	+0.1252
2^{14}	+0.0155	+0.0238	+0.0310	+0.0332
2^{15}	+0.0435	+0.0443	+0.0441	+0.0435
2^{16}	+0.0656	+0.0647	+0.0663	+0.0661

defined as:

$$\Delta H = \frac{\hat{H} - H}{H} * 100\%, \tag{6}$$

where H is the exact value and $\hat{H}$ is the mean of the estimates. Further, since we know that these two estimators are asymptotically unbiased, we assume that a sample size is acceptable when the relative inaccuracy drops below one percent. Table 1 shows that the relative inaccuracies of all estimates of H are less than one percent from $n = 2^{13}$. Fig. 2 shows that all Whittle ML estimates converge to the exact values faster than Abry-Veitch DWB estimates. In this case the sign of the bias can change as n increases. However, as Table 2 shows the relative inaccuracies are smaller than one percent for $n \geq 2^{10}$.

We have done similar studies for other estimators considered in this paper, and, with appropriate "margin of safety", have decided to base our comparative analysis of the selected estimators of H on samples containing $2^{15} = 32\ 768$ data points.

4 Hurst Parameter Estimators

Various Hurst parameter estimators have been proposed. They can be divided into three basic classes, depending on whether data analysis is done in the time-domain, the frequency-domain or by using wavelets.

4.1 Time-domain based estimators of H

This class of estimators contains the first estimator of H parameter ever proposed by Hurst, known as the R/S-statistic [5]. It also contains estimators which are derived from formulae for moments. They include the variance-time estimator [8], estimators based on absolute moments [30], and their modification known as fractal dimension estimators [11]. Other estimators on this class include the estimators based on variance of residuals [19] and the index of dispersion for counts [8]. The R/S-statistic and the variance-time estimator, and the estimator based on the index of dispersion for counts are some of the most frequently used estimators in practical studies of teletraffic in telecommunication networks. The last one captures the variability of teletraffic over different time scales in a given count process. They have been selected for comparative analysis in this paper and are discussed in detail in the following sections.

4.1.1 R/S-Statistic

An R/S-statistic is used to estimate the Hurst parameter H based on the *rescaled adjusted range* estimator (or *R/S-statistic*), as proposed by Hurst in 1951 [5]. Given a tested time series $x_1, x_2, \ldots, x_n$, calculation of the R/S-statistic follows the following steps:

Step 1. Divide the sequence $x_1, x_2, \ldots, x_n$ into $k = n, \lfloor n/2 \rfloor, \lfloor n/3 \rfloor, \ldots, 1$ non-overlapping batches of size $m = \lfloor n/k \rfloor$.[1] For each m and $t = (i-1)m$, $1 \leq i \leq k$, calculate its mean

$$\bar{x}(t,m) = \frac{1}{m} \sum_{i=t+1}^{t+m} x_i \tag{7}$$

and its standard deviation $S(t,m)$

$$S(t,m) = \sqrt{\frac{1}{m} \sum_{i=t+1}^{t+m} (x_i - \bar{x}(t,m))^2}, \text{ and} \tag{8}$$

[1] Some data items at the end of the time series are not included if n is not divisible by k. While in the original definition of R/S-statistic overlapping batches are used too, non-overlapping batches are assumed here for data reduction.

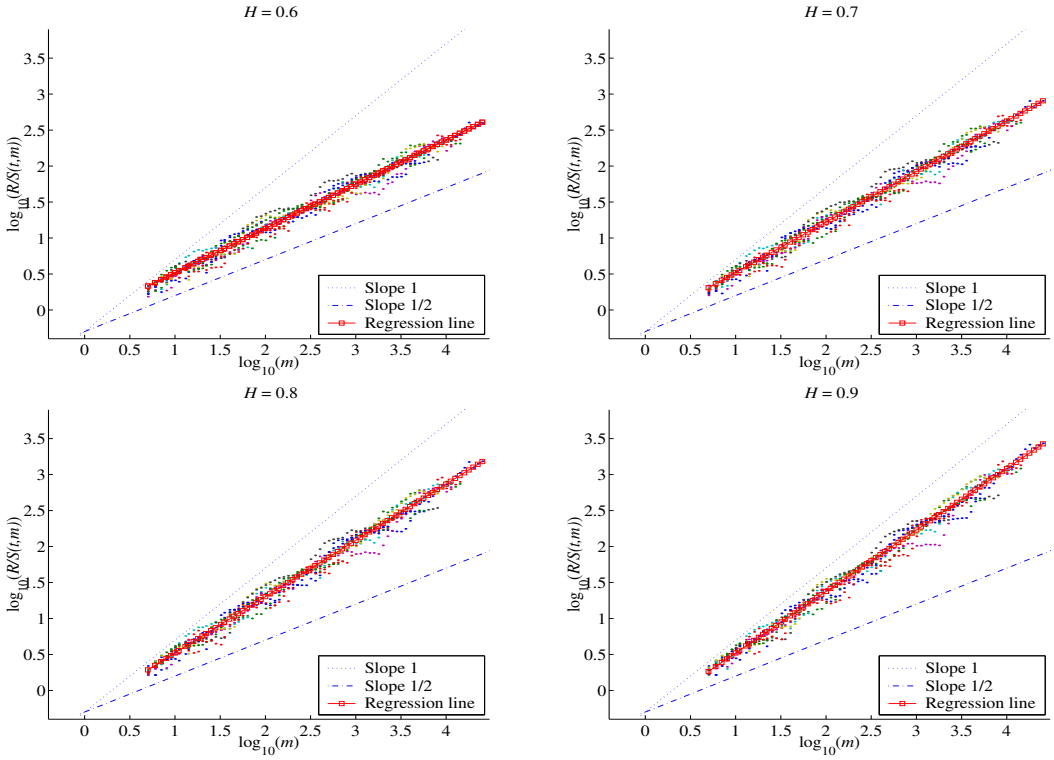

Fig. 3. R/S-statistic plots for time series generated by FGN processes, for $H = 0.6$, 0.7, 0.8 and 0.9.

$$R(t,m) = \max_i [Y_{t+i} - Y_t - \frac{i}{m}(Y_{t+m} - Y_t), \quad 0 < i \leq m] - \min_i [Y_{t+i} - Y_t - \frac{i}{m}(Y_{t+m} - Y_t), \quad 0 < i \leq m], \tag{9}$$

where $Y_t = \sum_{i=1}^{t} x_i$ and m is the current batch size. Compute the R/S-statistic as the ratios $\frac{R(t,m)}{S(t,m)}$, $m = 1, 2, \ldots, n$.

Step 2. Plot all $log_{10}(\frac{R(t,m)}{S(t,m)})$ against $log_{10}(m)$, known as the *pox diagram of R/S*; see also Fig. 3.

Step 3. Fit a regression line in the pox diagram of R/S to find the slope β_1 of the R/S-statistic plot. The estimate of H equals β_1.

Having generated four exact self-similar time series of 32 768 numbers, with $H = 0.6$, 0.7, 0.8 and 0.9, representing the FGN process, as discussed in Section 2, we produced the R/S-statistic plots shown in Fig. 3. The obtained estimates of H equal 0.615, 0.702, 0.783 and 0.855, respectively. When H is well defined, a typical R/S statistic plot starts with a transient zone representing short-range dependencies in the sample, but eventually settles down and fluctuates around a straight line of a slope β_1. The graphical representation of the R/S-statistic does help to determine whether asymptotic long-range dependence

behaviour appears to be supported by the data.

4.1.2 Variance-Time Estimator

The variance-time estimator, proposed by Cox and Smith in 1953 [8], is based on the property of self-similar processes that variances of the aggregated processes (defined in the Introduction) decrease at the rate m^{2H-2} as the batch size m increases; see Equation (2).

Thus, in log-log plots, if the process $\mathbf{X}$ is self-similar then the logarithm of the variance of the aggregated processes $X^{(m)}, m \geq 1$, decrease linearly with $log_{10}(m)$, with slope equal $-\beta_2$, $0 < \beta_2 < 1$. An estimate of the Hurst parameter is given by $\hat{H} = 1 - \beta_2/2$. Thus, given a tested time series $x_1, x_2, \ldots, x_n$, this procedure follows three steps [5]:

Step 1. Divide the sequence $x_1, x_2, \ldots, x_n$ into l non-overlapping batches of equal size $m = \lfloor \frac{n}{l} \rfloor$ [2], for $l_{min} < l < \lfloor n/2 \rfloor$. For each batch size m, calculate the variance

$$Var(X^{(m)}) = \frac{1}{l-1} \sum_{j=1}^{l} (\bar{x}_j^{(m)} - \bar{x}^{(m)})^2, \tag{10}$$

where $\bar{x}_j^{(m)}, j = 1, \ldots, l$, are means over batches of size m, and

$$\bar{x}^{(m)} = \frac{1}{l} \sum_{j=1}^{l} \bar{x}_j^{(m)}. \tag{11}$$

Step 2. Plot $log_{10}(Var(X^{(m)}))$ against $log_{10}(m)$.

Step 3. Fit a simple regression line through the resulting points. Its slope equals β_2, and $\hat{H} = 1 - \beta_2/2$.

Having used the same time series of 32 768 numbers as in Section 4.1.2, the numerical results obtained by this procedure are depicted in Fig. 4. The obtained estimates of H were 0.588, 0.681, 0.768 and 0.844, respectively.

If the sequence does not have long-range dependence and has finite variance, the estimate of $\hat{H}$ would be 0.5 and the slope of the fitted line would be -1. In practice, points at the very low and high ends of the plots should not be used to fit the least-squares line. This is because (1) short-range effects present at the very low end of the plot can distort the estimates of H, and (2) at the very high end of the plot there are too few blocks of data to obtain reliable estimates of the variance. Since this is probably one of the most popular estimators of H, its properties will be further investigated.

[2] For showing existence of self-similar effects in batched processes a sufficient number of batches should be used; see [5], [15], [30]. In this paper we have assumed that $l_{min} \geq 30$.

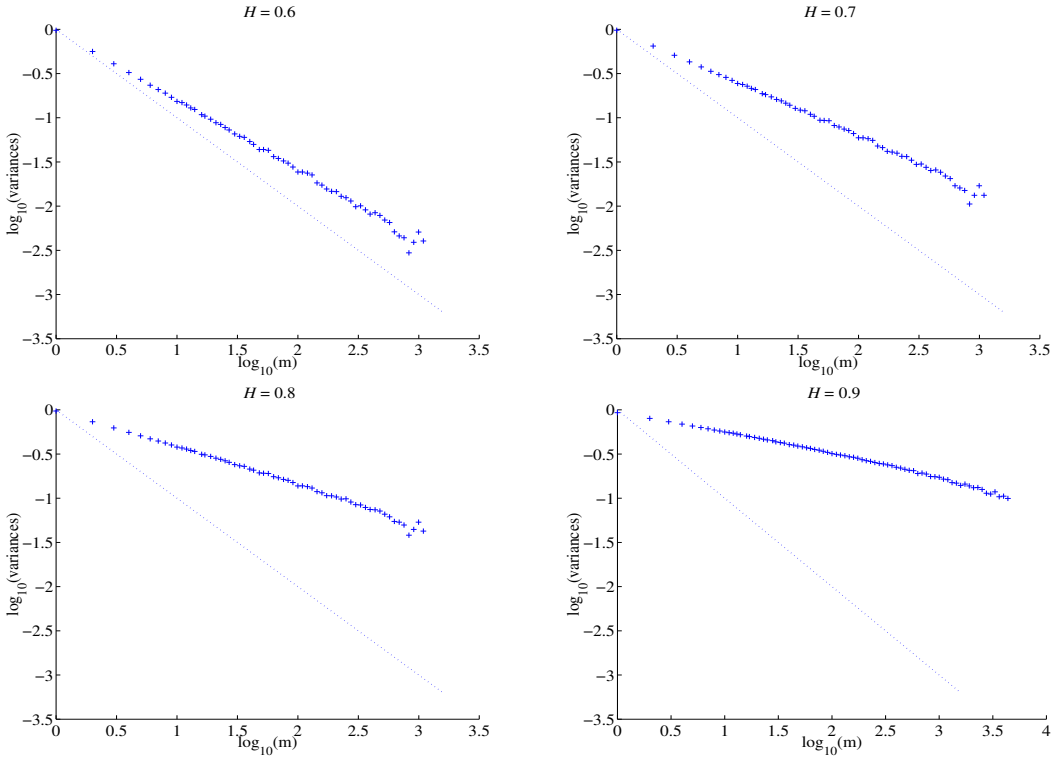

Fig. 4. Variance-time plots for time series generated by FGN processes, for $H = 0.6$, 0.7, 0.8 and 0.9.

4.1.3 Index of Dispersion for Counts

The index of dispersion for counts (IDC), proposed by Rao and Chakravati in 1956 [8], is often used to describe burstiness in teletraffic data. It can be relatively straightforward measured and it conveys more information about the structural dependencies in data than for example the coefficient of variation or lexis ratio.

IDC of a sequence $x_1, x_2, \ldots, x_n$ is defined as

$$\mathrm{IDC}(t) = \frac{S^2(t)}{\bar{x}(t)}, \tag{12}$$

where $2 \leq t \leq n$, and

$$\bar{x}(t) = \frac{1}{t}\sum_{i=1}^{t} x_i, \tag{13}$$

$$S^2(t) = \frac{1}{t-1}\sum_{i=1}^{t}(x_i - \bar{x}(t))^2. \tag{14}$$

A self-similar process produces a monotonically increasing $\mathrm{IDC}(t)$ of the form $ct^{2\hat{H}-1}$, where c is a finite positive constant that does not depend on t. $\mathrm{IDC}(t)$, as a function of t, is either constant or converges to a fixed value quite rapidly.

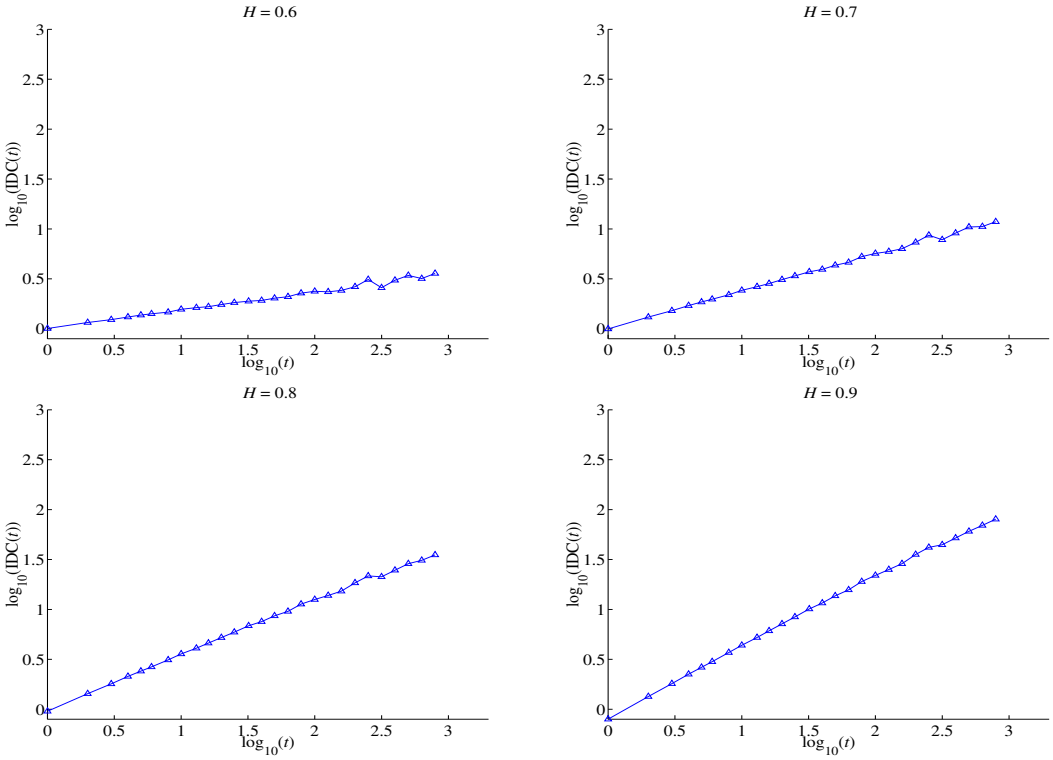

Fig. 5. Plots of IDC(t) for time series generated by FGN processes, for $H = 0.6$, 0.7, 0.8 and 0.9.

This should result in an asymptotic straight line with slope $\beta_3 = 2\hat{H}$ - 1; [10], [15]. Thus, given a tested time series $x_1, x_2, \ldots, x_n$, this estimator of H follows three steps:

Step 1. Use Equation (12) to calculate the IDC(t), for $t = 2, \ldots, n$.
Step 2. Plot $log_{10}(\text{IDC}(t))$ against $log_{10}(t)$.
Step 3. Find a regression line for $log_{10}(\text{IDC}(t))$. Its slope equals β_3, and $\hat{H} = \frac{1}{2}(1 + \beta_3)$.

Applying this procedure to the same time series of 32 768 numbers as in previous sections, we obtained the numerical results depicted in Fig. 5. The obtained estimates of H were 0.594, 0.684, 0.768 and 0.842, respectively. We have selected this estimator for our further studies since analysis of burstiness of teletraffic is widely used in performance evaluation of modern telecommunication networks.

4.2 **Frequency-domain based estimators of** H

This class of estimators is based on the fact that periodograms of long-range dependent self-similar processes increase hyperbolically, proportional to $|\lambda|^{-2H+1}$, as frequency $\lambda \longrightarrow 0$. An estimator of H employing graphical assess-

ment of the slope of the periodogram has been originally proposed by Daniell in 1948 [8]. Since then a few modifications of this estimator, such as the one based on the modified periodogram or on the cumulative periodogram, have been proposed [30]. However, as shown in [30], the original Daniell periodogram-based estimator proposed in [8], further called the Daniell PB estimator, is consistently more accurate.

Another subclass of frequency-domain based estimators applies maximum likelihood techniques. This approach was originally applied by Whittle [5], who proposed an estimator that is now known as the Whittle Maximum Likelihood (or Whittle ML) estimator. This estimator is discussed in more details in Section 4.2.2. Combining this Whittle ML estimator with data aggregation leads to a so-called aggregated Whittle ML estimator, proposed in [32]. Another semi-parametric, Whittle-type estimator, known as the local Whittle ML estimator was proposed in [22]. However, comparative studies of these estimators as reported in [28] and [29] show that the original Whittle ML estimator remains by far the most accurate, and because of that only the original estimator will be further analysed in this paper.

4.2.1 Daniell Periodogram-Based Estimator

Let $P(\lambda), -\pi \leq \lambda \leq \pi$, denote the periodogram of a given sequence $x_1, x_2, \ldots, x_n$ defined by

$$P(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^{n} x_j e^{ij\lambda} \right|^2 , \tag{15}$$

where $i^2 = -1$. As mentioned in the Introduction, if Equation (5) holds, i.e. if the autocorrelations are non-summable, then the time series comes from a long-range dependent process, and the periodogram near the origin should be randomly scattered around a straight line with a negative slope. If the autocorrelations in the tested time series $x_1, x_2, \ldots, x_n$ were summable, we would have a short-range dependent process, and the periodogram near the origin would be scattered randomly around a constant level.

An estimate of H is given by $\hat{H} = (1 - \beta_4)/2$, where β_4 is the slope of a regression line which is fitted to a number of values at low frequencies [5]. For determining an H estimate using this estimator, given a tested time series $x_1, x_2, \ldots, x_n$, one needs to follow the following steps:

Step 1. Calculate the periodogram $P(\lambda)$ as defined in Equation(15).
Step 2. Plot $log_{10}(P(\lambda))$ against $log_{10}(\lambda)$, for $\lambda > 0$.
Step 3. Find a regression line fitted to $log_{10}(P(\lambda))$. Its slope equals β_4, and $\hat{H} = \frac{1}{2}(1 - \beta_4)$.

In some cases, this plot will lead to a poor estimate of H since this periodogram-

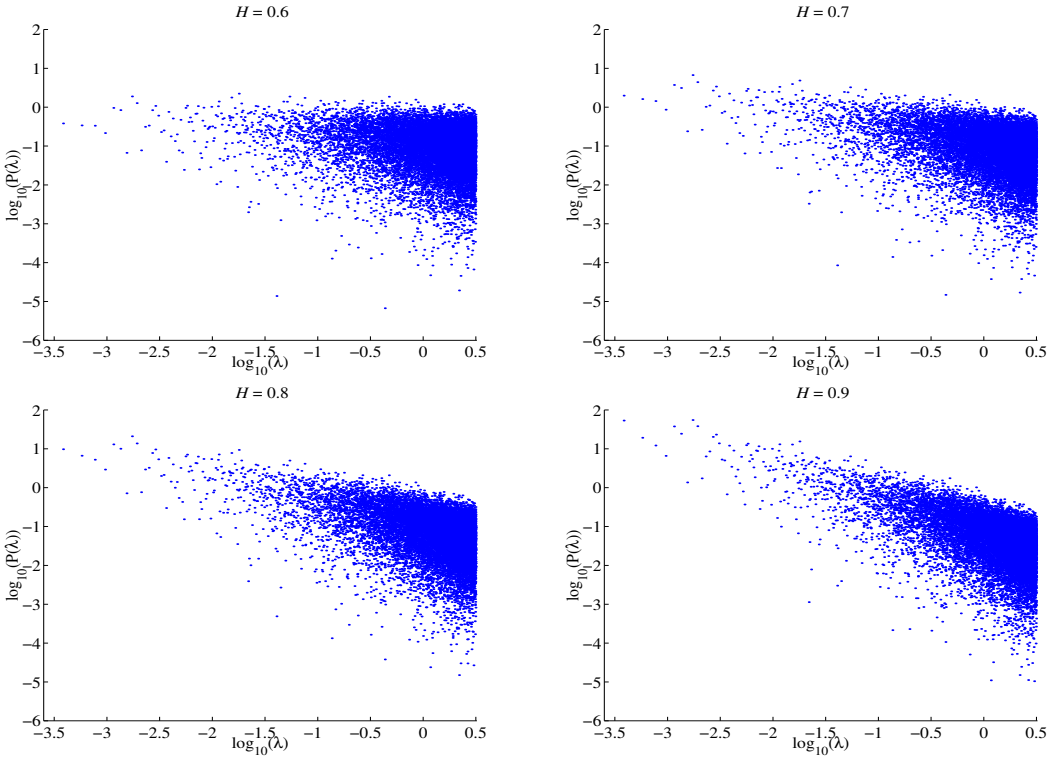

Fig. 6. Plots for determining Daniell PB estimators for FGN processes with $H = 0.6$, 0.7, 0.8 and 0.9.

based estimator has been shown to be biased and inconsistent [5]. In practice one should restrict analysis of periodogram to its values around $log_{10}(\lambda) = 0$. This can lead to elimination of as many as 90% of data points, since this phenomenon holds only for frequencies close to $\lambda = 1$ [15].

Applying this procedure to the same time series of 32 768 numbers as in previous sections, we obtained the numerical results depicted in Fig. 6. The obtained estimates of H were 0.590, 0.692, 0.793 and 0.893, respectively.

4.2.2 Whittle ML Estimator

The maximum likelihood estimator, proposed by Whittle in 1953 [5], has been studied extensively and shown to have desirable statistical properties for Gaussian processes, as presented in [4] and [32]. It is defined as follows.

Let $f(\lambda, \theta)$, for $-\pi \leq \lambda \leq \pi$, be the spectral density of a self-similar process represented by a given time series $x_1, x_2, \ldots, x_n$, with $\theta = (Var(\varepsilon_i), H, \theta_3, \ldots, \theta_k)$, where $Var(\varepsilon_i)$ is the variance of the white noise ε_i in the infinite autoregressive representation of the process; see [32]. The parameters $(\theta_3, \ldots, \theta_k)$ describe the SRD behaviour of the process.

Given a tested time series $x_1, x_2, \ldots, x_n$, the Whittle ML estimator is obtained

in the following three steps:

Step 1. Calculate the periodogram $P(\lambda)$ defined by Equation (15) and the power spectrum

$$f(\lambda, H) = 2c_f(1 - cos(\lambda))\mathcal{B}(\lambda, H), \tag{16}$$

for $0.5 < H < 1$ and $-\pi \leq \lambda \leq \pi$, where

$$c_f = \sigma^2(2\pi)^{-1}sin(\pi H)\Gamma(2H + 1), \tag{17}$$

$$\mathcal{B}(\lambda, H) = \sum_{k=-\infty}^{\infty} |2\pi k + \lambda|^{-2H-1}, \tag{18}$$

where $\Gamma(\cdot)$ is the gamma function; see [5].

Step 2. The Whittle Approximate ML estimator is defined as the value of H which minimises:

$$Q(H) = \int_{-\pi}^{\pi} \frac{P(\lambda)}{f(\lambda, H)} d\lambda. \tag{19}$$

Step 3. Calculate the estimate of variance σ_H^2 given by

$$\hat{\sigma}_H^2 = 4\pi \left[\int_{-\pi}^{\pi} \left(\frac{\partial \log f(\omega_i)}{\partial H} \right)^2 d\omega \right]^{-1}, \tag{20}$$

where $\omega_i = 2\pi i/n (i = 1, 2, \ldots, n^*), n^* = (n-1)/2$ if $n - 1$ is even and $n^* = (n-1)/2 - 1/2$ if $n - 1$ is odd.

As we see this technique determines both a point and an interval estimate, since $\hat{\sigma}_H^2$ can be also found.

4.3 Wavelet-based estimators of H

The wavelet transform appears to be a natural tool for studying scaling properties of self-similar processes [31]. Because of that, a number of H estimators based on wavelets have been proposed. While these estimators could be based on different families of wavelets, the estimators based on Daubechies wavelets have been more extensively studied. The first such estimator was proposed by Abry and Veitch in 1998 [2]. It uses a fast pyramidal algorithm for the wavelet transform, of order $O(n)$. However, as argued in [31], this estimator suffers from a bias associated with its log-log regression component.

Veitch and Abry [31] proposed the so-called wavelet-based joint estimator, which estimates both H and the power parameter, an independent quantitative parameter with dimension of variance; see [24] and [31] for details. The resulting wavelet-based estimator, further called the Abry-Veitch Daubechies

Wavelet-Based (or Abry-Veitch DWB) estimator, is asymptotically unbiased and (almost) the most efficient [31]. This estimator is further discussed in Section 4.3.1.

4.3.1 Abry-Veitch Daubechies Wavelet-Based Estimator

Let a wavelet transform be used to calculate the wavelet coefficients $d_x(i,j)$ for a given sequence $x_1, x_2, \ldots, x_n$. For a long-range dependent process, the variance of the wavelet coefficients at each level i is defined by

$$E[d_x(i,\cdot)^2] = \mathcal{C}\mathcal{A}_f 2^{i(2H-1)}, \tag{21}$$

where

$$\mathcal{A}_f = 2(2\pi)^{1-2H} c_\gamma \mathcal{E}(2H-1)\sin((1-H)\pi) \tag{22}$$

is the power parameter, $\mathcal{E}(\cdot)$ is the Euler function, and $\mathcal{C}$ and c_γ are positive constants. $\mathcal{A}_f$, which plays a major role in fixing the absolute size of long-range dependent effects in applications, is an independent quantitative parameter with the dimension of variance.

Abry and Veitch ([2], [31]) suggested an H estimator based on Daubechies wavelets [20]. To estimate H, we have to first calculate a time average μ_i of the $d_x(i,\cdot)$ at a given scale, defined as

$$\mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} d_x^2(i,j), \tag{23}$$

where n_i is the number of wavelet coefficients at scale i, i.e., $n_i = 2^{-i}n$, and n is the number of the data points in a given time series $x_1, x_2, \ldots, x_n$. The estimated value of H can be then obtained from the slope β_5 of the weighted linear regression curve defined in [31], for

$$\log_2(\mu_i) = \log_2\left(\frac{1}{n_i}\sum_{j=1}^{n_i} d_x^2(i,j)\right), \tag{24}$$

$i = 1, 2, \ldots, \lfloor \log_2 n \rfloor$; where c is a constant. As shown in [31], $\beta_5 = 2\hat{H} - 1$.

In practice, the regression curve should be fitted over a narrower range of scales. Namely, we need to select the range of scales (i_1, i_2) over which the power-law behaviour in Equation (21) holds, since long-range dependence is an asymptotic property of the power spectrum $f(\lambda, H)$, i.e., $f(\lambda, H) \to c\lambda^{1-2H}$ as $\lambda \to 0$. As Abry et al. suggested in [1], a useful heuristic for selection of the scale range (i_1, i_2) is to limit scales to the values for which the regression line fitted to a given subset of y_i goes through confidence intervals of y_i; see [1] for more detail. We refer to this rule as AFTV heuristic.

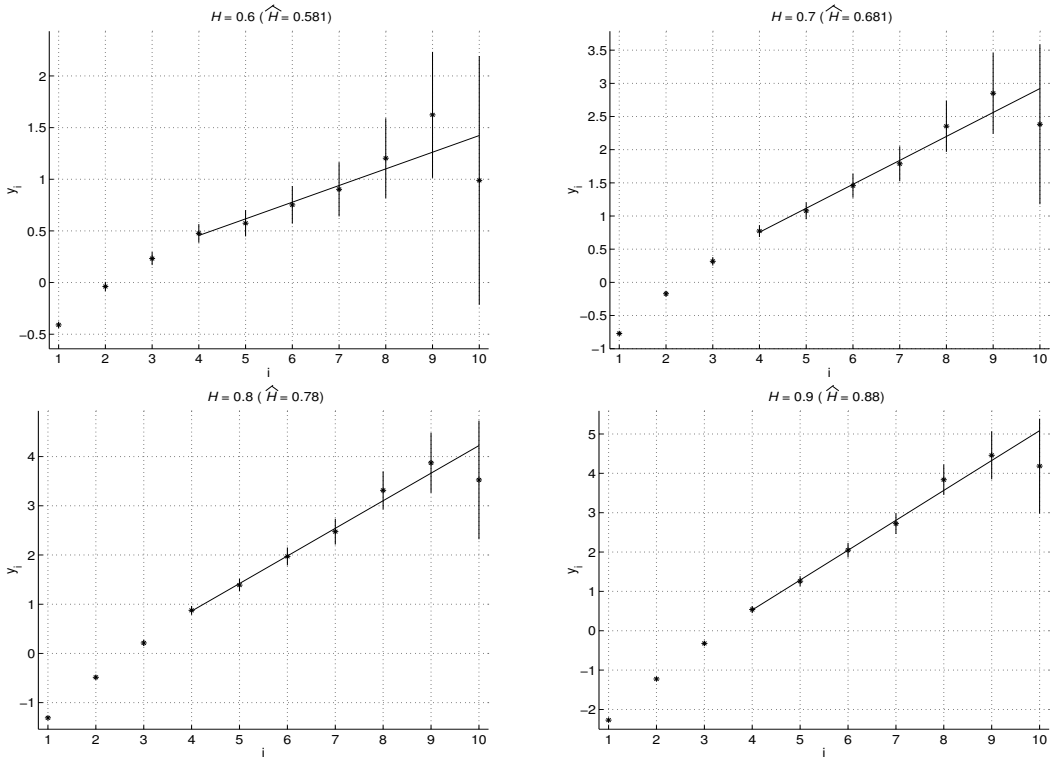

Fig. 7. Mean values of $d_x(i,\cdot)$, and their confidence intervals, for $1 \leq i \leq 10$, and regression curves for Abry-Veitch DWB estimates in the case of FGN processes, for scale $(i_1, i_2) = (4{,}10)$.

This estimator of H uses the weighted linear regression as the variances of $\log_2(\mu_i)$ vary with $\log_2(2^i)$. The plots of

$$y_i = \log_2(\mu_i) + 1/(n_i \ln 2) \tag{25}$$

against $\log_2(2^i)$ can be used for detection of long-range dependencies. Thus, given a tested time series $x_1, x_2, \ldots, x_n$, the Abry-Veitch DWB estimator is obtained in the following five steps:

Step 1. For a given time series $x_1, x_2, \ldots, x_n$, calculate coefficients of Daubechies wavelets $d_x(i,j)$ for $i = 1, 2, \ldots, \lfloor \log_2 n \rfloor$, and $j = 1, 2, \ldots, 2^{-i}n$.

Step 2. Use Equations (23) and (25) to determine y_i for $i = 1, 2, \ldots, \lfloor \log_2 n \rfloor$. The variance of y_i can be estimated as $\hat{\sigma}^2_{y_i} = 2/(n_i \ln^2 2)$.

Step 3. Plot y_i, and their confidence intervals, against $\log_2(2^i)$, for $i = 1, 2, \ldots, \lfloor \log_2 n \rfloor$.

Step 4. Find experimentally the range (i_1, i_2) of scales over which the H parameter should be estimated. Do it by systematic elimination of small and large scales that do not satisfy the AFTV heuristic, for the weighted regression line defined in [31].

Step 5. Draw the weighted regression line for the subset of y_i over (i_1, i_2). If the slope of the regression line is β_5, then $\hat{H} = \frac{1}{2}(1 + \beta_5)$.

Applying this procedure to the same time series of 32 768 numbers as in

Table 3
ΔH of mean values of the Abry-Veitch DWB estimates for the FGN process for different scales.

Scale	ΔH (%)			
(i_1, i_2)	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
(1,10)	+5.8590	+6.1380	+6.0870	+5.8800
(2,10)	+1.3000	+1.7920	+1.9410	+1.9290
(3,10)	+0.2200	+0.4090	+0.4639	+0.4628
(4,10)	+0.0258	+0.0920	+0.1096	+0.1075
(5,10)	-0.0783	-0.0788	-0.0904	-0.1033
(6,10)	-0.4554	-0.3627	-0.3088	-0.2806

Table 4
$Var[\hat{H}]$ of the Abry-Veitch DWB estimator for the FGN process for different scales.

Scale	Estimated Variances			
(i_1, i_2)	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
(1,10)	1.713e-05	1.700e-05	1.697e-05	1.703e-05
(2,10)	4.356e-05	4.362e-05	4.383e-05	4.420e-05
(3,10)	7.630e-05	7.716e-05	7.851e-05	8.058e-05
(4,10)	1.892e-04	1.922e-04	1.951e-04	1.985e-04
(5,10)	4.863e-04	4.912e-04	4.952e-04	4.955e-04
(6,10)	1.228e-03	1.277e-03	1.337e-03	1.393e-03

previous sections, we obtained the numerical results depicted in Fig. 7. The obtained estimates of H were 0.581, 0.681, 0.780 and 0.880, respectively. Fig. 7 illustrates the previous discussed phenomenon that data used to estimate H at higher scales are highly variable, and so can cause significant bias of $\hat{H}$ [25], [26].

Table 3 shows that the mean values of $\hat{H}$ obtained from 100 self-similar sequences are more biased when a smaller scale i_1 is assumed. Table 4 shows $\mathrm{Var}[\hat{H}]$ for different scales and H values. The variances gradually increase as the value of i_1 increases. For our data the Abry-Veitch DWB estimator appeared to be the least biased at scale $i_1 = 4$ and 5. Therefore, scale $i_1 = 4$ has been chosen for all our experiments.

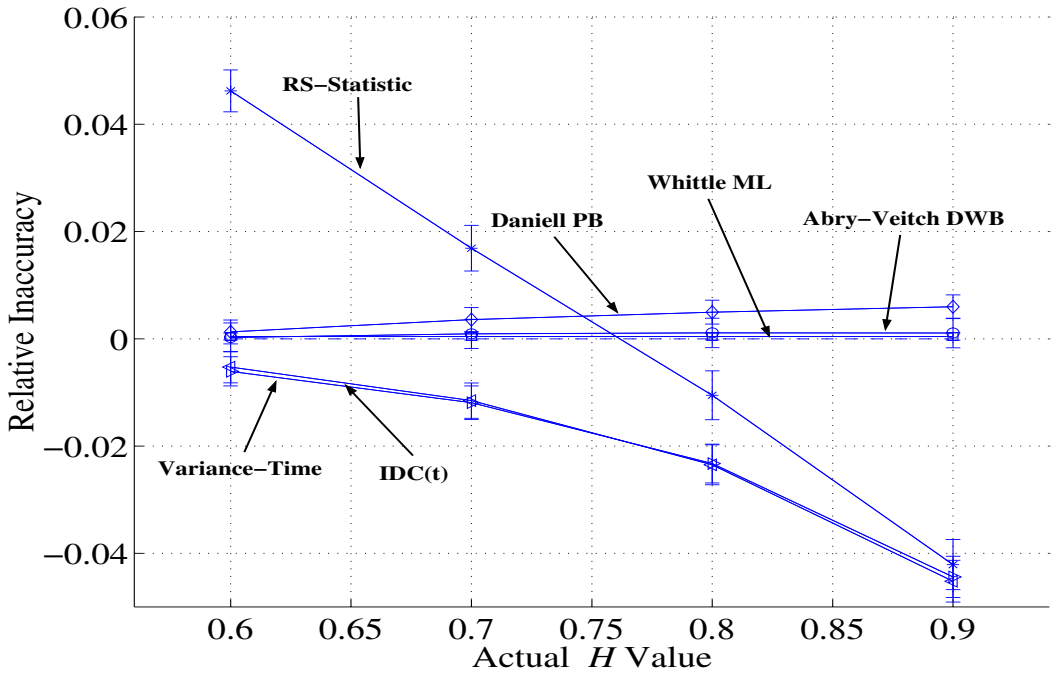

Fig. 8. Bias performance of the considered H estimators. Bars indicate 95% confidence intervals for the relative inaccuracy

5 Comparison of H Estimators

All reported here results are averages over 100 independent data samples.

Tables 5 and 6, and Figure 8, show the relative inaccuracies ΔH of mean values of $\hat{H}$, and mean bias, for the six estimators considered in this paper. 95% confidence intervals for the means of the Hurst parameter estimators are given in parentheses. The main findings can be summarised as follows:

- For $H \leq 0.7$, the R/S-statistic estimates are positively biased (with $|\Delta H| \leq 4.63\%$). They become negatively biased for $H \geq 0.8$, with $1\% \leq |\Delta H| \leq 4.2\%$.
- The variance-time estimates become increasingly negatively biased as H increases, with $|\Delta H| < 4.44\%$.
- Estimates of H based on IDC(t) become increasingly negatively biased as H increases, with $|\Delta H| < 4.53\%$.
- The Daniell PB estimates become increasingly positively biased as H increases, with $|\Delta H| < 0.6\%$.
- The bias of the Whittle's ML estimator does not depend strongly on H and is very small ($|\Delta H| < 0.05\%$). Thus, this is the most accurate estimator.
- The positive bias of the Abry-Veitch DWB estimator increases with H, with $|\Delta H| < 0.12\%$.

Table 5
Relative inaccuracy ΔH of mean values of $\hat{H}$ for $H = 0.6$ and 0.7. 95% confidence intervals are given in parentheses.

Methods of Estimation	Mean Values of Estimated H and ΔH			
	$H = 0.6$		$H = 0.7$	
	$\hat{H}$	$\Delta H(\%)$	$\hat{H}$	$\Delta H(\%)$
R/S-statistic	.6277 (.624,.632)	+4.623	.7118 (.708,.716)	+1.689
Variance-time	.5964 (.594,.599)	−0.608	.6917 (.689,.695)	−1.192
IDC(t)	.5968 (.594,.600)	−0.528	.6919 (.689,.695)	−1.153
Daniell PB	.6008 (.599,.603)	+0.128	.7025 (.700,.705)	+0.358
Whittle ML	.6003 (.591,.610)	+0.043	.7003 (.691,.710)	+0.044
Abry-Veitch DWB	.6002 (.573,.628)	+0.026	.7006 (.673,.728)	+0.092

Tables 5 and 6 also show that all confidence intervals produced by the Abry-Veitch DWB and Whittle ML estimators contain the exact values of H , while in the case of the confidence intervals of other estimators the exact values of H are usually outside these intervals. Thus, Abry-Veitch DWB and Whittle ML estimators of H are also superior in the sense of coverage of their confidence intervals, at least on the considered cases.

So far we have concentrated on bias of the H estimators considered. A "good" estimator is not only the one which produces estimates with expected value close to the true value (low bias), but also the one which is more efficient, i.e., has smaller variance. Table 7 shows that $Var[\hat{H}]$ for Whittle ML estimates is by far the smallest of all estimators considered.

As the Abry-Veitch DWB and Whittle ML estimators are substantially more accurate than others, they have been chosen for further analysis. Figure 9 shows that the estimates using the Whittle ML estimator are less biased in all cases considered, except $H = 0.6$. Figs. 10 – 13 show that, as indicated in Table 7, Whittle ML estimates have a much lower variance than those from Abry-Veitch DWB estimator.

Table 6
Relative inaccuracy ΔH of mean values of $\hat{H}$ for $H = 0.8$ and 0.9. 95% confidence intervals are given in parentheses.

Methods of Estimation	Mean Values of Estimated H and ΔH			
	$H = 0.8$		$H = 0.9$	
	$\hat{H}$	$\Delta H(\%)$	$\hat{H}$	$\Delta H(\%)$
R/S-statistic	.7916 (.787,.796)	−1.053	.8621 (.857,.867)	−4.210
Variance-time	.7814 (.778,.785)	−2.327	.8600 (.856,.864)	−4.439
IDC(t)	.7812 (.778,.785)	−2.349	.8593 (.855,.863)	−4.520
Daniell PB	.8040 (.802,.806)	+0.496	.9054 (.903,.908)	+0.598
Whittle ML	.8004 (.791,.810)	+0.044	.9004 (.891,.909)	+0.043
Abry-Veitch DWB	.8009 (.773,.828)	+0.110	.9010 (.874,.929)	+0.108

In all reported cases, the same 100 samples taken from the exact self-similar FGN process were used. Thus we can further quantify the relative accuracy of the Whittle ML and Abry-Veitch DWB estimators by counting how often (out of 100 Bernoulli trials) one of them produced more accurate estimate than the other. The results are given in Table 8. One can see that at least 86% of the time Whittle ML estimates were more accurate for a particular value of H. The values given in brackets are the probabilities that the two estimators are in fact equally accurate, and that the superior performance of Whittle ML estimator is due to chance alone.

6 Conclusions

We have reported the results of a comparative analysis of the most frequently used estimators of H: R/S-statistic-based, variance-time based, IDC(t), Daniell PB, Whittle ML and Abry-Veitch DWB estimators. Our results have shown that while for the assumed sample size of 2^{15} all estimators produce results bi-

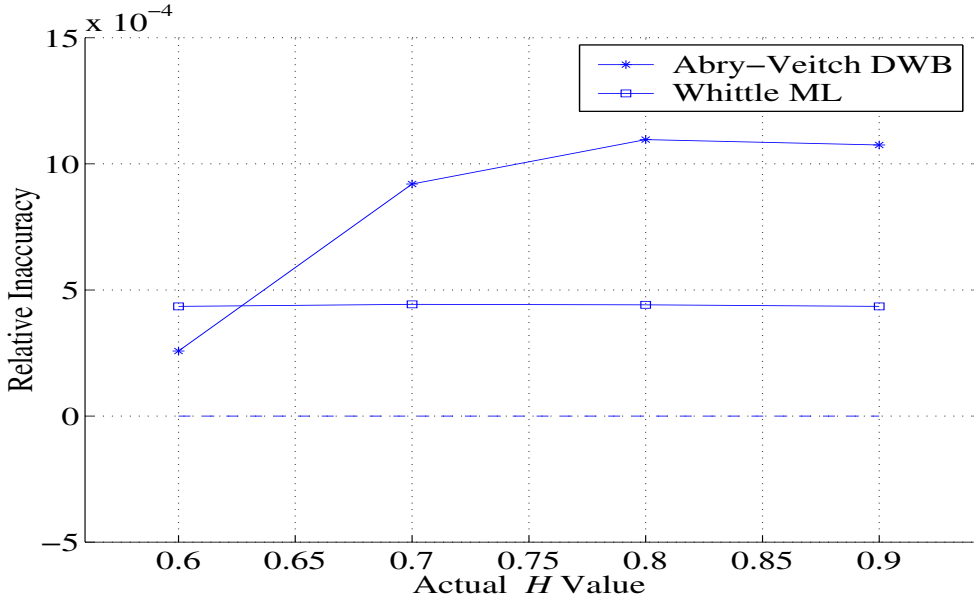

Fig. 9. Bias performance of Abry-Veitch DWB estimator and Whittle ML estimator for the FGN process.

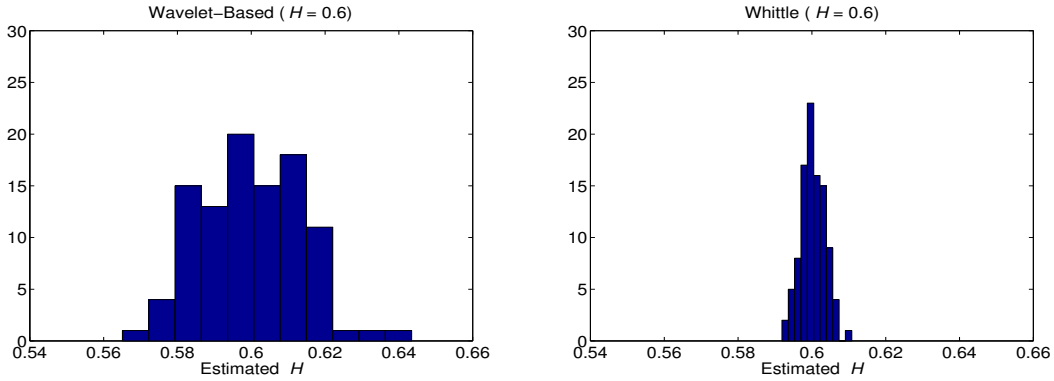

Fig. 10. Histograms of $\hat{H}$ for Abry-Veitch DWB estimator and Whittle ML estimator for $H = 0.6$; for 100 replications.

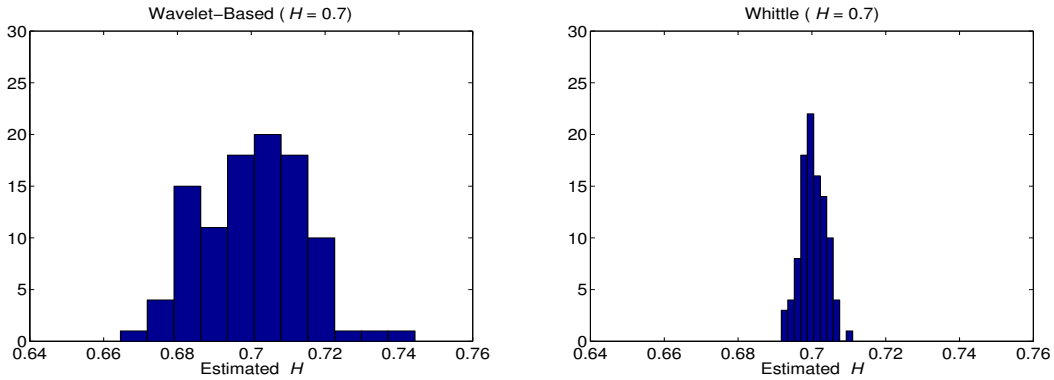

Fig. 11. Histograms of $\hat{H}$ for Abry-Veitch DWB estimator and Whittle ML estimator for $H = 0.7$; for 100 replications.

Table 7
Variance of $\hat{H}$.

Methods of Estimation	Variances of Estimated H			
	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
R/S-statistic	3.971e-04	4.724e-04	5.400e-04	5.669e-04
Variance-time	1.928e-04	2.571e-04	3.405e-04	3.826e-04
IDC(t)	2.227e-04	2.812e-04	3.613e-04	3.947e-04
Daniell PB	1.288e-04	1.306e-04	1.291e-04	1.279e-04
Whittle ML	1.093e-05	1.145e-05	1.179e-05	1.217e-05
Abry-Veitch DWB	1.892e-04	1.922e-04	1.951e-04	1.985e-04

Table 8
Comparison of Abry-Veitch DWB and Whittle ML estimates of H in 100 samples from the FGN processes with $H = 0.6$, 0.7, 0.8 and 0.9.

Estimators	How many times (out of 100) gave closer estimate of H			
	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
Abry-Veitch DWB	14 (1.0)	9 (1.0)	14 (1.0)	11 (1.0)
Whittle ML	86 (4.1422e-14)	91 (1.6610e-18)	86 (4.1422e-14)	89 (1.2704e-16)
Total	100	100	100	100

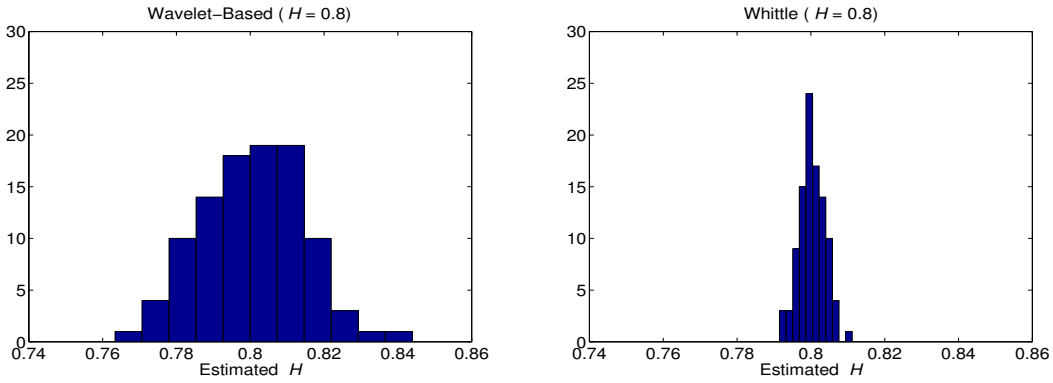

Fig. 12. Histograms of $\hat{H}$ for Abry-Veitch DWB estimator and Whittle ML estimator for $H = 0.8$; for 100 replications.

ased below 5%, the Abry-Veitch DWB and Whittle ML estimators are clearly superior as they are the least biased of the estimators considered. Within the range of considered values of H, the Abry-Veitch DWB estimator has a lower bias than the Whittle ML estimator only for $H = 0.6$. However, it is computationally simpler and faster than the Whittle ML estimator. The Whittle ML

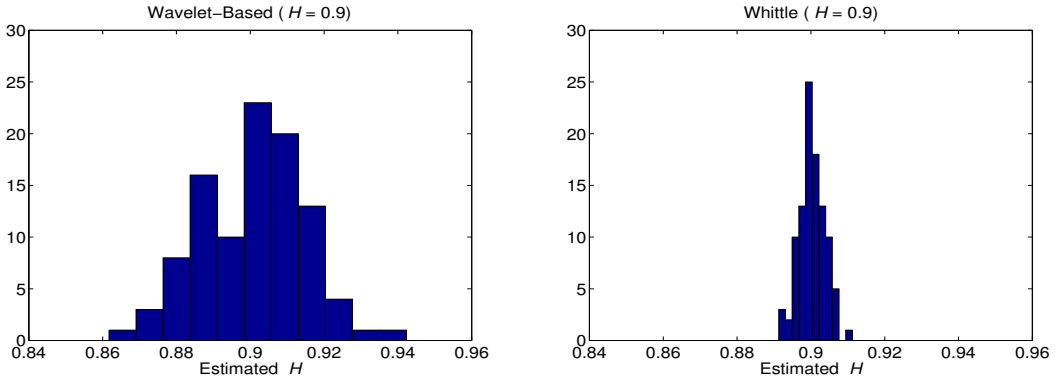

Fig. 13. Histograms of $\hat{H}$ for Abry-Veitch DWB estimator and Whittle ML estimator for $H = 0.9$; for 100 replications.

estimator is superior in the sense of accuracy and because it has the smallest variance. It also requires smaller samples for ensuring that estimates do not exceed a given level of the bias. Its relative bias was smaller than 1%, even for samples with as few as $n = 2^{10}$ data points. However, the minimisation procedure it employs requires many repetitive calculations, leading to a significantly higher overall cost than the Abry-Veitch DWB estimator, which requires simple calculations of discrete wavelet transforms, performed in $O(n)$ steps.

References

[1] Abry, P., Flandrin, P., Taqqu, M., and D.Veitch. *Theory and Applications of Long-Range Dependence*. Birkhäuser, Doukhan, Oppenheim, and Taqqu (eds), Boston, MA, 2002, ch. Self-Similarity and Long-Range Dependence Through the Wavelet Lens, pp. 527–556.

[2] Abry, P., and Veitch, D. Wavelet Analysis of Long-Range-Dependent Traffic. *IEEE Transactions on Information Theory 44*, 1 (1998), 2–15. http://www.emulab.ee.mu.oz.au/∼darryl/.

[3] Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., and Taqqu, M. *Theory and Applications of Long-Range Dependence*. Birkhäuser, Doukhan, Oppenheim, and Taqqu (eds), Boston, MA, 2002, ch. Generators of Long-Range Dependent Processes: A Survey, pp. 579–623.

[4] Beran, J. Statistical Methods for Data with Long Range Dependence. *Statistical Science 7*, 4 (1992), 404–427.

[5] Beran, J. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.

[6] Beran, J., and Terrin, N. Estimation of the Long-Memory Parameter, Based on a Multivariate Central Limit Theorem. *Time Series Analysis 15*, 3 (1994), 269–278.

[7] Brockwell, P., and Davis, R. *Time Series: Theory and Methods, Second Edition.* Springer-Verlag, New York, 1991.

[8] Cox, D., and Lewis, P. *The Statistical Analysis of Series of Events.* Spottiswoode, Ballantyne & Co. Ltd., UK, 1966.

[9] Garrett, M., and Willinger, W. Analysis, Modeling and Generation of Self-Similar VBR Video Traffic. In *Computer Communication Review, Proceedings of ACM SIGCOMM'94* (London, UK, 1994), vol. 24 (4), pp. 269–280.

[10] Gusella, R. Characterizing the Variability of Arrival Processes with Indexes of Dispersion. *IEEE Journal on Selected Areas in Communications 9* (1991), 203–211.

[11] Higuchi, T. Approach to an Irregular Time Series on the Basis of the Fractal Theory. *Physica D 31*, 2 (1988), 277–283.

[12] Jeong, H.-D. *Modelling of Self-Similar Teletraffic for Simulation.* PhD thesis, Department of Computer Science, University of Canterbury, 2003.

[13] Krunz, M., and Matta, I. Analytical Investigation of the Bias Effect in Variance-Type Estimators for Inference of Long-Range Dependence. *Computer Networks 40* (2002), 445–458.

[14] Leland, W., Taqqu, M., Willinger, W., and Wilson, D. On the Self-Similar Nature of Ethernet Traffic. In *Proc. ACM SIGCOM'93* (1993), pp. 183–193.

[15] Leland, W., Taqqu, M., Willinger, W., and Wilson, D. On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE ACM Transactions on Networking 2*, 1 (1994), 1–15.

[16] Mandelbrot, B., and Wallis, J. Computer Experiments with Fractional Gaussian Noises. *Water Resources Research 5*, 1 (1969), 228–267.

[17] Park, K., and Willinger, W. *Self-Similar Network Traffic and Performance Evaluation.* John Wiley & Sons, Inc., K. Park and W. Willinger (eds), New York, 2000, ch. Self-Similar Network Traffic: An Overview, pp. 1–38.

[18] Paxson, V. Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic. *Computer Communication Review, ACM SIGCOMM 27*, 5 (1997), 5–18.

[19] Peng, C.-K., Buldyrev, S., Havlin, S., Simons, M., Stanley, H., and Goldberger, A. Mosaic Organization of DNA Nucleotides. *Physical Review E 49*, 2 (1994), 1685–1689.

[20] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1999.

[21] Riedi, R., Crouse, M., Ribeiro, V., and Baraniuk, R. A Multifractal Wavelet Model with Application to Network Traffic. *IEEE Transactions on Information Theory 45*, 3 (1999), 992–1018.

[22] Robinson, P. Gaussian Semiparametric Estimation of Long-Range Dependence. *The Annals of Statistics 23* (1995), 1630–1661.

[23] Rose, O. *Traffic Modeling of Variable Bit Rate MPEG Video and Its Impacts on ATM Networks.* PhD thesis, Bayerische Julius-Maximilians-Universität Würzburg, 1997.

[24] Roughan, M., and Veitch, D. Measuring Long-Range Dependence Under Changing Traffic Conditions. In *Proceedings of IEEE INFOCOM'99* (New York, NY, USA, 1999), pp. 1513–1521.

[25] Roughan, M., Veitch, D., and Abry, P. On-Line Estimation of the Parameters of Long-Range Dependence. In *Proceedings of GLOBECOM'98* (Sydney, Australia, 1998), pp. 3716–3721.

[26] Roughan, M., Yates, J., and Veitch, D. The Mystery of the Missing Scales: Pitfalls in the Use of Fractal Renewal Processes to Simulate LRD Processes. In *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics* (American University, Washington, DC, 1999).

[27] Taqqu, M. Self-Similar Processes. In *Encyclopedia of Statistical Sciences*, vol. 8. John Wiley and Sons, Inc., S. Kotz and N. Johnson (eds), New York, 1988.

[28] Taqqu, M., and Teverovsky, V. Robustness of Whittle-Type Estimators for Time Series with Long-Range Dependence. *Stochastic Models 13* (1997), 723–757.

[29] Taqqu, M., and Teverovsky, V. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* R.J. Adler, R.E. Feldman and M.S. Taqqu (eds), Birkhauser, Boston, MA, 1998, ch. On Long-Range Dependence in Finite and Infinite Variance Series, pp. 177–217.

[30] Taqqu, M., Teverovsky, V., and Willinger, W. Estimators for Long-Range Dependence: an Empirical Study. *Fractals 3*, 4 (1995), 785–788. http://math.bu.edu/people/murad/methods/index.html.

[31] Veitch, D., and Abry, P. A Wavelet Based Joint Estimator of the Parameters of Long-Range Dependence. *IEEE Transactions on Information Theory 45*, 3 (1999), 878–897. Special Issue on Multiscale Statistical Signal Analysis and its Applications, http://www.emulab.ee.mu.oz.au/~darryl/.

[32] Willinger, W., Taqqu, M., Leland, W., and Wilson, D. Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements. *Statistical Science 10*, 1 (1995), 67–85.