Wavelet Analysis of Long Range Dependent Traffic

Patrice Abry [1], Darryl Veitch [2]

(1) *CNRS URA 1325 - Laboratoire de Physique - Ecole Normale Supérieure de Lyon - 46, allée d'Italie 69 364 LYON Cedex 07 - France*

tel: (+33) 4 72 72 84 93 - Fax: (+33) 4 72 72 80 80 - E-mail: pabry@physique.ens-lyon.fr

(2) *INRIA, 2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis, Cedex France*

tel: (+33) 4 93 65 79 18 - E-mail: darryl.veitch@sophia.inria.fr

Abstract

A Wavelet based tool for the analysis of long range dependence is introduced and a related semi-parametric estimator of the Hurst parameter. The estimator is shown to be unbiased under very general conditions, and efficient under Gaussian assumptions. It can be implemented very efficiently allowing the direct analysis of very large data sets, and is highly robust against the presence of deterministic trends, as well as allowing their detection and identification. Statistical, computational and numerical comparisons are made against traditional estimators including that of Whittle. The estimator is used to perform a thorough analysis of the long range dependence in Ethernet traffic traces. New features are found with important implications for the choice of valid models for performance evaluation. A study of mono *vs* multi-fractality is also performed, and a preliminary study of the stationarity with respect to the Hurst parameter and deterministic trends.

Keywords: Packet Traffic, Wavelet Decomposition, Time-Scale Analysis, Long Range Dependence, Hurst Parameter, Parameter Estimation, Telecommunications Networks, Stationarity.

1 Introduction

There is now ample evidence that long term correlations are present in a wide range of generalised data types, including many of those likely to form major components of telecommunications traffic in high speed networks. The best known example is given by the high quality Local Area Network Ethernet traces of Leyland *et.al.* [24] recorded at Bellcore over a number of years under a variety of conditions. These large data sets show very clearly the scale dependent properties of Ethernet traffic, and in particular the presence of long range dependence (LRD), for instance in the point process describing frame arrival instants. Video traffic [12] is another case of note. Beran *et.al.* demonstrate the presence of long range dependence over a wide range of time scales in variable bit rate traffic (VBR). Long range dependence has also been found in other traffic contexts, notably in Wide Area Networks [29], and in Common Channel Signalling (CCSN/SS7) traffic [17].

The question of the impact of such characteristics on network performance is the subject of much current research, as well as considerable confusion and debate. What seems irrefutable however is that LRD in real data does indeed impact significantly on queueing delays [18], and that certain simplified analytical models of single server queues incorporating LRD corroborate this by exhibiting virtual work distributions with tails which decay slower than the exponential decay familiar from Markovian models [27] [11] [32] [13] [28]. In the absence however of a complete programme of data analysis, informed model selection, parameter estimation and finally model verification against relevant performance criteria, the connection between LRD and performance metrics cannot be fully understood.

Among the outstanding issues in the list above, there are two of particular importance which we contribute to in the present paper. The first is the primordial problem of data-analysis/parameter-estimation in the presence of LRD. It is well known [10] that even the most elementary of classical statistics require significant revision in the face of long range dependence. For example, let x be a wide sense stationary process in discrete time with LRD, by which we mean that the covariance function $\gamma_x(k)$ takes the form

$$\gamma_x(k) \sim c_\gamma k^{-(2-2H)}, \quad k \to +\infty \tag{1.1}$$

c_γ a positive constant, $H \in (0.5, 1)$. The Hurst parameter H measures long range dependence, $H = 0.5$ corresponding to the classical case of short range dependence. In this context, Beran ([10], p. 160) shows that the distribution of $\bar{x}_n = (\sum_{i=k+1}^{k+n} x_i)/n$ tends to a normal variable with increasing sample size n, but at a slower rate than the classical $\sqrt{n}$. In fact the statistic

$$z(\mu, c_\gamma, H) = \frac{\bar{x}_n - \mu}{\sigma_\mu} n^{1-H} \tag{1.2}$$

where $\sigma_\mu^2 = c_\gamma(H(2H-1))^{-1}$, is asymptotically standard normal. Thus even the estimation of the mean of a stationary process, because of the appearance of H as an *exponent* in (1.2), depends strongly on the LRD phenomenon. The fact that this is true in general for statistics of LRD processes shows that H is of central importance. It is vital that it be estimated well, and if joint parameter estimation is impossible or impractical, that it be estimated first.

The main aim of the paper is to introduce an estimation tool from wavelet analysis [2, 3] which provides a natural, statistically and computationally efficient, estimator of the Hurst parameter H. It is known [10] that simple traditional estimators can be seriously biased. Asymptotically unbiased estimators derived from Gaussian Maximum Likelihood Estimation are available [10], [33], but these are parametric methods which require a parameterized family of model processes to be chosen a priori, and which cannot be implemented exactly in practice for large data sets (such as those here) due to high computational complexity and memory requirements. Furthermore, they are not naturally matched to the essentially *simple* scale behaviour of LRD. In contrast, wavelet analysis is first of all a tool which studies the scale dependent properties of data directly via the coefficients of a joint scale-time wavelet decomposition. As such, very little needs to be assumed about the underlying process. Should evidence of LRD be found, it then offers an unbiased semi-parametric estimator which can be very efficiently implemented using techniques from non-redundant multi-resolution analysis [3].

The wavelet based estimator has the additional virtue of robustness against an important class of non-stationarity, namely the addition of deterministic trends. This is a particularly important advantage in a LRD context where it is very difficult in theory and in practice to distinguish between real trends and long term sample path variations due to LRD. We show how the nature of trends in data can be determined, and their effects on H estimation greatly reduced or even rigorously eliminated, by varying a characteristic of the analysing wavelet known as the number of vanishing moments. A full discussion of stationarity is beyond the scope of this paper, however we believe that a real attempt to deal with this issue is indispensable to any useful application of stationary LRD models to data. We therefore discuss deterministic trends in some detail, and also perform a preliminary investigative analysis of the variation of H with time in Ethernet data.

The second objective of the paper is the analysis of Ethernet data using the wavelet based estimator. Although substantial work has already been done, and the presence and impact of LRD clearly established, we show that there are important things to learn about the structure of the data before quantitatively accurate models for performance evaluation purposes can be proposed. The issues we consider are important for traffic modelling in general, and not only for the Ethernet traffic we consider here. Thus we go further than the published studies [24] [18] which concentrate on the time series of interarrival instants. We measure in addition other aspects of the data to gain insight into the process of the arriving *work* itself. This is motivated by insights from analytical studies [27] [11], which show that the precise structure of arriving work is crucial, LRD in itself cannot determine performance. For example [11] [23] [36], the GI—M—1 queue with interarrivals of infinite variance has a LRD arrival process but classical exponentially distributed queue tails. On the other hand Weibullian (stretched exponential) tails are found for the Fractional Brownian Storage Model of Norros [27] and for the superposition of a large number of small peak-rate ON/Off fluid sources [11]. Even more extreme behaviour is found for superpositions of large peak-rate ON/Off fluid sources, that is power law tails with infinite expectation [32] (See also [13] and [28] for related results). Our extended analysis of the Bellcore data reveals some striking features with important implications for the choice of a model capable of capturing the work arrival process. We also discuss the issue of *mono vs multi-fractal* modelling, and conclude that the data traces are well described by mono-fractals.

In a second paper we use these insights to select a specific compact model, and test it extensively against real data in simulation experiments.

The remainder of the paper is set out as follows. In section 2 we introduce wavelets, wavelet analysis, and the wavelet based estimator of H. Its theoretical properties are compared to that of the Whittle estimator and its Discrete version, as well as classical alternatives. Numerical comparisons against the Discrete Whittle estimator are given.

In section 3 stationarity is briefly discussed, and a full account of the wavelet estimator's properties with respect to deterministic trends. Numerical examples compare its performance against the Discrete Whittle estimator.

In section 4 we introduce briefly the nature of the Ethernet data sets and their first order statistics. We then show how each trace can be analysed from different points of view, and give the results of a wavelet based LRD analysis for each. The implications for model choice are discussed. The multi-fractal nature of the data is investigated and finally a preliminary result on the stationarity with respect to H.

Finally in section 5 we summarize the main results.

2 Wavelet-based Analysis of Long-Range Dependence

2.1 The long-range dependence phenomenon

Although new to telecommunications, the long-range dependence phenomenon has long been recognised in many fields including hydrology, turbulence, biology and semiconductor physics. It is commonly accepted that the definition of LRD is the slow power-law decrease of the autocorrelation function of a wide sense stationary process expressed in equation (1.1) with $H > 0.5$, corresponding to the divergence of the autocorrelation sum. An equivalent statement for the spectrum $\Gamma_x(\nu)$ of x is

$$\Gamma_x(\nu) \sim c_f |\nu|^{1-2H}, \;\; \nu \to 0 \tag{2.1}$$

where $c_f = \pi^{-1} c_\gamma \Lambda(2H-1)\sin(\pi - \pi H)$, and Λ is the Gamma function. Thus the LRD processes belong to the class of random processes which take the form $1/|\nu|^\alpha$ for a range of frequencies ν close to 0. The LRD phenomenon is also closely related to the properties of scale invariance, self-similarity and hence fractals, and is therefore often associated with statistically self-similar processes such as the Fractional Brownian Motion [26].

Note that we denote by x both a wide sense stationary process, and a realisation of it referred to as a *signal*, depending on the context.

2.2 Multiresolution analysis and discrete wavelet transform: a short review

A multiresolution analysis (MRA) consists in a collection of nested subspaces $\{V_j\}_{j\in\mathcal{Z}}$, satisfying the following set of properties [14]:

- i) $\bigcap_{j\in\mathcal{Z}} V_j = \{0\}$, $\bigcup_{j\in\mathcal{Z}} V_j$ is dense in $L^2(\mathcal{R})$
- ii) $V_j \subset V_{j-1}$
- iii) $x(t) \in V_j \longleftrightarrow x(2^j t) \in V_0$
- iv) There exists a function $\phi_0(t)$ in V_0, called the *scaling function*, such that the collection $\{\phi_0(t-k), k \in \mathcal{Z}\}$ is an unconditional Riesz basis for V_0

Similarly, the scaled and shifted functions $\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k), k \in \mathcal{Z}\}$ constitute a Riesz basis for the space V_j. Performing a multiresolution analysis of the signal x means successively projecting it into each of the approximation subspaces V_j:

$$\mathrm{approx}_j(t) = (\,\mathrm{Proj}_{V_j} x)(t) = \sum_k a_x(j,k)\phi_{j,k}(t)\,.$$

Since $V_j \subset V_{j-1}$, approx_j is a coarser approximation of x than is approx_{j-1} and therefore, the key idea of the MRA consists in examining the loss of information, that is the detail, when going from one approximation to the next, coarser one: $\mathrm{detail}_j(t) = \mathrm{approx}_{j-1}(t) - \mathrm{approx}_j(t)$. The MRA analysis shows that the detail signals detail_j can be directly obtained from projections of x onto a collection of subspaces, the W_j, called the wavelet subspaces. Moreover, the MRA theory shows that there exists a function ψ_0, called the mother wavelet, to be derived from ϕ_0, such that its templates $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k),\ k \in \mathcal{Z}\}$ constitute a Riesz basis for W_j:

$$\text{detail}_j(t) = (\ \text{Proj}_{W_j} x)(t) = \sum_k d_x(j,k)\psi_{j,k}(t)\,.$$

Basically, the MRA consists in rewriting the information in x as a collection of details at different resolutions and a low-resolution approximation:

$$\begin{aligned} x(t) &= \text{approx}_J(t) + \textstyle\sum_{j=1}^{j=J} \text{detail}_j(t) \\ &= \textstyle\sum_k a_x(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_x(j,k)\psi_{j,k}(t) \end{aligned} \tag{2.2}$$

The approx_j essentially being coarser and coarser approximations of x means that ϕ_0 needs to be a low-pass function. The detail_j, being an information 'differential', indicates rather that ψ_0 is a band-pass function, and therefore a small wave, a *wavelet.* More precisely, the MRA shows that the mother wavelet must satisfy $\int \psi_0(t)dt = 0$ and that its Fourier transform obeys $|\Psi_0(\nu)| \sim \nu^N$, $\nu \to 0$ where N is a positive integer called the number of vanishing moments of the wavelet [14].

Given a scaling function ϕ_0 and mother-wavelet ψ_0, the discrete (or non redundant) wavelet transform (DWT) is a mapping from $L^2(\mathcal{R}) \to l^2(\mathcal{Z})$ given by

$$x(t) \to \{\{a_x(J,k),\ k \in \mathcal{Z}\},\ \{d_x(j,k),\ j = 1,\ldots,J,\ k \in \mathcal{Z}\}\} \tag{2.3}$$

These coefficients are defined through inner products of x with two sets of functions:

$$\left.\begin{aligned} a_x(j,k) &= \langle x, \mathring{\phi}_{j,k}\rangle \\ d_x(j,k) &= \langle x, \mathring{\psi}_{j,k}\rangle \end{aligned}\right\} \tag{2.4}$$

where $\mathring{\psi}_{j,k}$ (resp., $\mathring{\phi}_{j,k}$) are shifted and dilated templates of $\mathring{\psi}_0$ (resp., $\mathring{\phi}_0$), called the dual mother wavelet (resp., the dual scaling function), and whose definition depends on whether one chooses to use an orthogonal, semi-orthogonal or bi-orthogonal DWT (see e.g., [14, 6]). They can practically be computed by a fast recursive filter-bank-based pyramidal algorithm whose computational cost is extremely low (see e.g., [14]).

2.3 The Wavelet-based H Estimator

Definition of the estimator. The coefficient $|d_x(j,k)|^2$ measures the amount of energy in the analyzed signal about the time instant 2^jk and frequency $2^{-j}\nu_0$, where ν_0 is an arbitrary reference frequency selected by the choice of ψ_0. It has been suggested [2, 3] that a useful spectral estimator can be designed by performing a time average of the $|d_x(j,k)|^2$ at a given scale, that is

$$\hat{\Gamma}_x(2^{-j}\nu_0) = \frac{1}{n_j}\sum_k |d_x(j,k)|^2\,, \tag{2.5}$$

where n_j is the available number of wavelet coefficients at octave j. Essentially $n_j = 2^{-j}n$ where n is the length of the data. $\hat{\Gamma}_x(\nu)$ is therefore a measure of the amount of energy that lies within a given bandwidth around the frequency ν and can therefore be regarded as a statistical estimator for the spectrum $\Gamma_x(\nu)$ of x. In fact one can show [3] that, when x is a wide sense stationary process, the expectation of $\hat{\Gamma}_x$ is

$$\mathbb{E}\hat{\Gamma}_x(2^{-j}\nu_0) = \int \Gamma_x(\nu) 2^j |\Psi_0(2^j\nu)|^2 d\nu \tag{2.6}$$

where Ψ_0 denotes the Fourier transform of the analysing wavelet ψ_0. From this relation, one sees that $\hat{\Gamma}_x$ suffers from the standard convolutive bias, that is the spectrum to be estimated is mixed within a frequency range corresponding to the frequency width of the analysing window at scale j. The crucial point here is that for LRD signals this bias reduces naturally to a simple form, enabling an unbiased estimation of H. To see this, recall the spectral behaviour (2.1) and assume for the moment that this form holds for all frequencies. The bias equation (2.3) can then be rewritten as

$$\left.\begin{array}{rcl} \mathbb{E}\hat{\Gamma}_x(2^{-j}\nu_0) & = & c_f|2^{-j}|^{(1-2H)} \int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu \\ & = & \Gamma_x(2^{-j}\nu_0)|\nu_0|^{(2H-1)} \int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu \end{array}\right\} \tag{2.7}$$

From equation (2.7), one sees that in the case of $1/|\nu|^{\alpha}$ processes the standard convolutive bias turns into a multiplicative one. Moreover, this multiplicative constant is independent of the analyzing scale j. It is therefore possible to design an estimator $\hat{H}$ for the parameter H from a simple linear regression of $\log_2(\hat{\Gamma}_x(2^{-j}\nu_0))$ on j, that is

$$\log_2(\hat{\Gamma}_x(2^{-j}\nu_0) = \log_2(\frac{1}{n_j}\sum_k |d_x(j,k)|^2) = (2\hat{H}-1)j + \hat{c} \tag{2.8}$$

where $\hat{c}$ estimates $\log_2(c_f \int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu)$, provided that the integral

$$\int |\nu|^{(1-2H)}|\Psi_0(\nu)|^2 d\nu \tag{2.9}$$

converges. Performing a weighted-least squares fit between the scales (octaves) j_1 and j_2 yields the following explicit formula for the estimator of H

$$\hat{H}(j_1, j_2) \equiv \frac{1}{2}\left[\frac{\sum_{j=j_1}^{j_2} S_j j \eta_j - \sum_{j=j_1}^{j_2} S_j j \sum_{j=j_1}^{j_2} S_j \eta_j}{\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} S_j j^2 - (\sum_{j=j_1}^{j_2} S_j j)^2} + 1\right] \tag{2.10}$$

where $\eta_j = \log_2\left(\frac{1}{n_j}\sum_k |d_x(j,k)|^2\right)$ and the weight $S_j = (n\ln^2 2)/2^{j+1}$ is the inverse of the theoretical asymptotic variance of η_j[1, 3].

Bias of $\hat{H}$. The above definition for $\hat{H}$ holds provided that (2.1) holds for all frequencies and that (2.6) converges. We can relax the first condition since in (2.7) we are free to choose only the range of scales over which (2.1) *does* hold.

Now consider the convergence of (2.9). In fact estimation problems in the presence of LRD often arise from the singular behaviour of $1/|\nu|^\alpha$ spectra at $\nu = 0$ which causes such integrals to diverge. When designing [6] the mother-wavelet ψ_0, one is free to select one of its important characteristics, namely the number N of vanishing moments. N is an integer such that

$$\forall k = 0, 1, \ldots, N-1, \quad \int t^k \psi_0(t) dt \equiv 0. \tag{2.11}$$

Clearly, this parameter also controls the behaviour of the Fourier transform of the wavelet about $\nu = 0$:

$$|\Psi_0(\nu)| = O(\nu^N), \ \nu \to 0\,.$$

It is easy to check that provided

$$N > H - 1, \tag{2.12}$$

the behaviour of $|\Psi_0(\nu)|^2$ at the origin will be *flat* enough to balance the singularity of the long-range dependent spectrum, thus ensuring the convergence of (2.6). When this inequality is satisfied, we have shown that the log-log regression-based H estimator is asymptotically unbiased, and in practice has very low bias even for short data sets [3].

Efficiency of $\hat{H}$. It is known [10, 34] that in the presence of LRD the standard sample estimator $(1/n \sum_k x_k^2))$ for second order statistics such as the variance of the process x has very poor statistical properties, because a time-average is performed over strongly correlated data. In the wavelet coefficient representation space, it has been shown [20] that for LRD processes with parameter H:

$$\mathbb{E} d_x(j,k) d_x(j,k') = O\Big(|k-k'|^{2H-2-2N} \Big), \ \ |k-k'| \to \infty\,. \tag{2.13}$$

This clearly shows that the correlation structure of the transformed data, that is the data represented through the wavelet coefficients, is not LRD provided the no-bias condition $N > H - 1$ is satisfied, in contrast to the LRD of the original data. This reduction is a non-trivial effect due to the combination of the analysing-wavelet being band-pass with N vanishing moments (i.e, $\Psi_0(\nu) \sim \nu^N, \nu \to 0$), and the wavelet basis being built from the dilation operator (i.e., $2^{j/2}\Psi_0(2^j\nu) \sim \Psi_0(\nu), \nu \to 0$). Such a reduction in the correlation range allows the use of the standard sample variance estimator $1/n_j \sum_k |d_x(j,k)|^2$ to estimate $\Gamma_x(2^j\nu_0)$. More precisely, under Gaussian and quasi-decorrelation of the wavelet coefficient hypotheses and in the asymptotic limit, a closed-form for the variance of the estimate of H can be obtained and is given by

$$\sigma^2_{\hat{H}} = \mathrm{var}\, \hat{H}(j_1, j_2) = \frac{2}{n_{j_1} \ln^2 2} \frac{1 - 2^J}{1 - 2^{-(J+1)}(J^2+4) + 2^{-2J}} \tag{2.14}$$

where $J = j_2 - j_1$ is the number of octaves involved in the linear fit and $n_{j_1} = 2^{-j_1} n$ is the number of available coefficients at scale j_1. It can be shown that this variance is the smallest possible, that is equal to the Cramer-Rao bound, for a given J. For further details on this estimator, see [2, 3].

Confidence intervals. From the above closed-from for the variance estimation (therefore under Gaussian and asymptotic assumptions), one can derive a confidence interval:

$$\hat{H} - \sigma_{\hat{H}} z_\beta \leq H \leq \hat{H} + \sigma_{\hat{H}} z_\beta$$

where z_β is the $1-\beta$ quantile of the standard Gaussian distribution, i.e. $P(z \geq z_\beta) = \beta$. All the results presented below, both in numerical simulations and actual data analysis, were computed with $\beta = 0.025$ (i.e., 95% confidence intervals), based on the above hypotheses.

Importance of N, the number of vanishing moments. The possibility of choosing N provides a very powerful means of detecting and identifying trends in the data and of cancelling their effects, as explained in detail in section 3.

N also plays a central role in the variance of $\hat{H}$. The above theoretical relations regarding bias and variance shows that the larger N is, the better the estimation. However, this theoretical improvement with increasing N is balanced by the increase of the number of wavelet coefficients polluted by border effects (due to the finite length of the data), resulting in a diminishing number of available wavelet coefficients and therefore an increase of variance. A good practical compromise seems to be $N \simeq H+1$. An elegant solution to this trade-off would be the use of *wavelets that live on an interval* [31] which take care of border effects in a clever way so as to avoid the decrease of available coefficients when increasing N[1].

Choosing the wavelet or the wavelet transform. Up to now we have restricted, for the sake of clarity, our definition of the $H-$parameter estimation to the DWT framework. It can however be extended to any redundant wavelet transform (the continuous wavelet transform, for instance) without difficulty. We have shown in detail in [3, 4] that despite its redundancy and a much higher computational cost, the performance of the estimator using a redundant wavelet transform is not superior to that given by the DWT, neither theoretically nor practically, except in some specific situations.

Within the DWT, another interesting question is the choice of the mother-wavelet. In [1, 3] it was shown that for the estimation of H, the only property of the wavelet that matters is its number of vanishing moments. Whether the wavelets are symmetrical or not, form an orthonormal, semi- or bi-orthonormal basis or not, makes no theoretical nor significant practical difference. The reason why we chose Daubechies wavelets is not orthonormality but the fact that they have a finite time support (that eases the handling of border effects), and that they form a basis where the number of vanishing moments can be naturally increased. Moreover increasing N does not result in an excessive extension of their time support. They are however certainly not the only choice.

2.4 Comparison with Standard Estimators

Time-domain estimators. By definition, the LRD phenomenon is related to the power-law behaviour of certain second-order statistics (variance, covariance,...) of the process with respect to the duration T of observation. Many estimators of H are therefore based on the idea of measuring the

[1]We thank a reviewer for suggesting this improvement which will be implemented in the next version of the tool.

slope of a linear fit in a log-log plot. The so-called variogram or R/S estimators are famous examples of this approach but are known to have poor statistical performance, notably high bias and sub-optimal variance. For a detailed review of these and other estimators, see [10], or [35]. There is however a time domain estimator with better properties known as the *Allan variance* [7], which consists in measuring the expectation of squared difference of averages of the data within windows of length T,

$$V_A(T) = \frac{1}{K}\sum_{k=1}^{K}\left(\int_{t_k-T}^{t_k} x(u)du - \int_{t_k}^{t_k+T} x(u)du\right)^2 ,$$

where K is the number of segments of size T in the data. This quantity also behaves as a power-law of T when LRD is present and allows a *non-biased* estimation of H. In fact [21], the Allan variance can be rephrased in term of the wavelet-based estimator, provided that the *Haar* wavelet is used, thus

$$V_A(T = 2^{j-1}T_0) = \mathbb{E}|d_x(j,k)|^2 ,$$

where T_0 is an arbitrary (small) sampling period. The improved properties of the Allan variance-based H estimator are explained by its belonging to the wavelet-based framework, which can therefore be seen as a generalization whereby the simple $N = 1$ Haar analysing function, which takes the value $2^{-1/2}$ over $[0, T_0/2]$, $-2^{-1/2}$ over $[T_0/2, 1]$, and zero elsewhere, is replaced by functions (or wavelets) of higher N.

Frequency-domain estimation. LRD causes the spectrum of a process to behave as a power-law for frequencies close to 0. It is therefore natural to think of using spectral estimation to measure H. A standard spectral estimator consists in averaging smoothed periodograms computed on different pieces of the data,

$$\hat{\Gamma}_2(\nu) = \sum_{k=1}^{P} |\int x(t-kL)w_L(t)\exp(\imath 2\pi\nu t)dt|^2$$

where P is the number of data pieces, L their length, and w_L a weighting window. It has been shown in detail [2, 3] that when applied to $1/|\nu|^\alpha$ signals, such a spectral estimator results in an estimator of H, based on a linear fit in a $\log(\nu)$ *vs* $\log(\hat{\Gamma}_2(\nu))$ plot, which is strongly biased. Basically, this is because a constant-bandwidth ($\Delta\nu$ = constant) spectral estimation is performed which in no way matches the $1/|\nu|^\alpha$ structure of the spectrum to be analyzed. In contrast, the wavelet-based quantity $1/n_j \sum_k |d_x(j,k)|^2$ can be read as a spectral estimate with a constant relative bandwidth ($\nu/\Delta\nu$ = constant), which perfectly matches a power-law shaped spectrum [2, 3].

2.5 Wavelet *vs* Discrete-Whittle Estimation

The Discrete-Whittle Estimator. Maximum likelihood Estimation (MLE), the best known fully parametric method, offers a coherent approach to estimator design which is capable of producing an unbiased, asymptotically efficient estimator for H (as well as for other parameters). The Whittle estimator consists of two analytic approximations to the exact Gaussian MLE, suggested by Whittle in

1953 in order to avoid the huge computational complexity of the exact algorithm. In the 1980's [10] it was shown that nothing is lost in this approximation, in the sense that asymptotically the estimator is unbiased and efficient, just as in the exact case.

The approximation essentially replaces the covariance matrix by an integral of a function of the spectrum. As described in detail elsewhere [10], computational difficulties remain, motivating a further approximation: the discretization of the frequency domain integration rewritten in terms of the periodogram. It is this discrete version which we compare against in this section. Being based on a parametric estimator, a specific family of processes must be chosen. We compare against the fractional Brownian noise (fGn) and fractional ARIMA(0,d,0) processes ($d = H - 1/2$). For the numerical comparisons here and in the next section, we use the `Splus` implementations described in Beran [10].

Statistical Comparison. Non-rigorous results of Graf [22, 10] suggest that the Discrete Whittle (D-Whittle) estimator is also asymptotically unbiased and efficient, at least in the fGn case. Assuming that this is in fact the case for Gaussian process in general, the second order statistics of the D-Whittle and wavelet estimators would then be asymptotically equivalent under Gaussian assumptions, both attaining the Cramer-Rao minimum variance. They will not be identical in practice however, since in the wavelet case not all scales are used in the estimation, but only the J scales where the asymptotic scaling behaviour is actually observed, resulting in larger confidence intervals. A numerical comparison of the two estimators under non Gaussian conditions is given by Testsignal8 in table 1, for which the marginal distributions are bimodal. We see that a failure of the Gaussian assumption biases the D-Whittle estimator slightly, but does not effect the estimation by wavelet.
Regarding the first order statistics, it is important to note that the D-Whittle estimator is only asymptotically unbiased (due to the periodogram) whilst the wavelet estimator is unbiased. This is illustrated in figure 1, where a comparison is made for different values of H. We see that except for low H values, that the D-Whittle point estimates fall further from the true value than the width of their confidence intervals. This is not the case for the wavelet estimates.

Computational issues. The D-Whittle estimator relies on the periodogram, which has a low computational cost. A minimization procedure is involved however which requires many repetitive evaluations, leading to a significantly higher overall cost. Moreover, problems of convergence to local minima rather than to the absolute minimum may be encountered. On the other hand, the wavelet-based estimator requires only the simple calculation of a DWT, which can be done in $O(n)$ operations (even less than that of a FFT) using the fast pyramidal filter-bank based algorithm [25], followed by time-averaging. For very large data sets the filter-bank algorithm has another significant advantage with respect to memory usage. The signal can be split into blocks of a treatable size, and the separate calculations combined to obtain the exact result for the signal as a whole. This does *not* imply any reduction in the range of scales analysed. Because of these advantages, we say that the wavelet estimator is capable of analysing continuous time signals. By this we mean that we are free to use a very high sampling period T_0 in the numerical discretization of the continuous time data, to capture accurately the fine detail of the signal. We cannot do this for the D-Whittle case as the size of the resulting discretized signal would be prohibitive. In summary, The wavelet-based H estimator therefore leads to a simple, low cost, scalable algorithm.

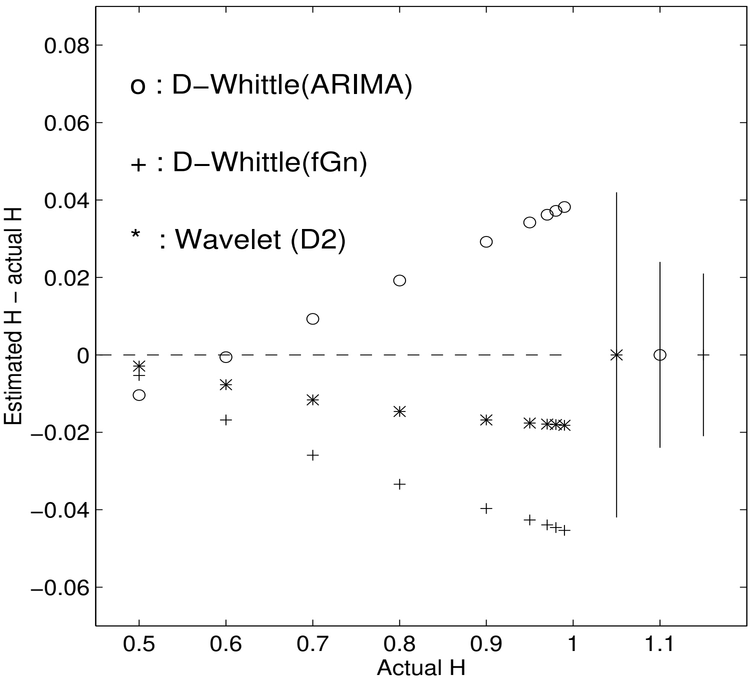

Figure 1: **Bias performance for finite sample size.** The wavelet estimator gives point estimates, here based on $(j_1, j_2) = (2, 6)$, where the true value always lies within the confidence interval. This is not the case for the two D-Whittle estimators, which are biased for finite sample sizes, here 4096 points of simulated fGn. .

Estimation *vs* analysis- the semi-parametric approach. The D-Whittle estimator solves many of the computational problems of the Gaussian MLE approach, but the essential disadvantage of parametric methods remain, that a specific parametric family of processes must be chosen. In most practical data analysis situations however, one has no idea of the exact correlation structure of the data, and an inappropriate choice of parametric family can result in biased estimation, as seen in the D-Whittle fARIMA results for fGn data of figure 1. Compare also columns 4 and 5 of table 1. The wavelet estimator on the other hand enables an *analysis* of the scale behaviour in a nearly hypothesis-free manner. Thus the $\log_2(2^j)$ *vs* $\log_2(1/n_j \sum_k |d_x(j,k)|^2)$ plot is a tool for the detection of LRD, the determination of the range of scales (or equivalently, of frequencies) over which the power-law behaviour holds, and a test for the presence of spurious trends as discussed in the next section. If scaling behaviour across a range of scales j_1 to j_2 is observed, the wavelet estimator is then an effective semi-parametric technique for the estimation of the corresponding H.

In common with all semi-parametric methods however, there is the problem of the arbitrary nature of the cutoff scales or frequencies at which the LRD behaviour is taken to hold, i.e. the exact choice of j_1 and j_2. Here the wavelet based estimator has a significant advantage in practical terms because of the reduction in variance property of the wavelet coefficients *across scales*. In fact a relation similar to that of equation (2.10) for the reduction across *time* holds true also for scales. It implies that the estimation within a $1/|\nu|^\alpha$ regime is relatively independent across scales, allowing the beginning of such a regime to be clearly identified, if present. In other words, the short range correlation structure does not pollute the estimation of the long range dependent structure, simplifying greatly the task of the choice of cutoff scale.

3 Stationarity

3.1 General Considerations

As discussed in the introduction, the testing of the stationarity hypothesis is particularly difficult in the presence of LRD where many classical statistical approaches cease to hold. Even without LRD however, there is the fundamental problem that there are an infinity of ways in which a process can be non-stationary. Normally we must choose a particular model framework and test for stationarity only against the types of non-stationarity encompassed by it. To assist in the process it is important to include a priori information concerning the known "physics" of the problem. For example, in the context of Ethernet traffic the first thing to note is that it is clear that the data is *not* stationary, because of the diurnal cycle, lunch breaks etc. On the other hand it is reasonable to expect that for smaller timescales where network conditions are relatively stable, that stationarity will be a natural and useful assumption. Linear trends for instance are not to be expected on a priori physical grounds. The task in such a context is therefore to determine the time-scale at which we can reasonably assume stationarity. Clearly we would like this to be as large as possible for both practical and statistical reasons. Unfortunately we cannot explore this issue properly here, however we present in section 5 some simple results on the stationarity of the Ethernet traces with respect to the variation of H with time. We now turn to address in detail a special but important case of non-stationarity, that of generalised deterministic trends.

3.2 Deterministic Trends

Assume that the signal $x(t)$ consists of stationary 'data' $s(t)$, plus some contaminating deterministic function of time $p(t)$ such that $x(t) = s(t) + p(t)$. We would like to be able to measure H correctly for the data, and detect and identify the trend.

One example is the fact that a slowly (power-law shaped) decaying trend added to a short range dependent process s can generate autocovariance estimations with slowly decaying tails, which could be incorrectly mistaken for evidence that x is stationary with LRD. The opposite problem can also occur, that the data is LRD but this is masked by the overall non-stationary behaviour of x due to the trend. Alternatively the presence of the trend may not be recognised, being confused with the local statistical trends characteristic of LRD. In each case the trend may drastically bias the estimation of H for the data s. We will show both from numerical simulations and theoretical arguments that the wavelet-based estimator enables us to detect and even identify such trends, and to avoid their adverse effects on the estimation of H.

In the paragraphs below we compare the wavelet and D-Whittle estimators over a number of test signals where deterministic trends have been added to a stationary signal of known H. Testsignal1 is a realisation of Fractional Gaussian Noise (fGn) with $H = 0.82$, synthezised by a spectral method. Testsignal2a to Testsignal2j consists of Testsignal1 plus linear deterministic trends of increasing amplitudes. In Testsignal3 to Testsignal7, quadratic (t^2), quartic (t^4), increasing power-law ($t^{3/2}$), sinusoidal ($\sin 2\pi f$), and decreasing power law ($t^{-1/4}$) trends have been added. In Testsignal8 the marginal of the process was chosen to be bimodal and hence strongly non-Gaussian, whilst retaining $H = 0.82$ (see section 2.5 for a discussion of this case). In Testsignal9 and Testsignal10 decreasing and increasing power-law trends respectively were added to ordinary Gaussian noise ($H = 0.5$). The magnitude of the trend for (most of) these test signals is given in the third column of Table 1 as a multiple of their respective sample variances.

For each of these signal types we studied the effect of increasing the trend amplitude (the magnitude of the trend across the entire trace in units of the trace sample variance is shown in the third column of Table 1). We found that the D-Whittle estimates were at first unaffected, and then became increasingly biased, finally saturating at absurd values. This is to be expected of a parametric estimator, as it is based in an essential way on the assumption that no trend is present. The main purpose then of the following discussion, and of Table 1 in particular, is to illustrate and contrast the robustness of the wavelet estimator across a wide range of conditions, rather than to demonstrate the failure of the D-Whittle estimator. In particular we are not interested in trying to quantify at what trend amplitudes the D-Whittle estimation become seriously affected as a function of trend type. Note that since we are interested in evaluating the polluting effect of a deterministic addition to a single realisation of a process, we must compare the *estimates* of H obtained with and without the trend against each other, and not against the real value of H!

Linear trends. Let us focus first of all on linear trends, since this is the simplest case which is most often of interest. Figure 2 clearly shows that the D-Whittle estimates in the presence of a linear trend depart progressively from the correct value as the amplitude of the trend increases. This is because there is no allowance for a trend in the underlying parametric model. When the trend grows to the same magnitude as the data, Whittle's method does its best to interpolate between the two. On the other hand, the wavelet-based estimates remain constant whatever the amplitude of the trend, provided that one uses wavelets with $N \geq 2$.

Polynomial trends. Now consider polynomial trends of higher degree (see Testsignal3 and Testsignal4 in table 1). Once again we observe that whereas Whittle estimates fail whenever the amplitude of the trend becomes too large, the wavelet-based estimation remains accurate provided that the N of the wavelet is tuned to the degree P of the polynomial trend $p(t)$, that is provided

$$N \geq P + 1\,.$$

To explain this, recall that a wavelet ψ_0 with N vanishing moments is, by definition (equation 2.11), orthogonal to the space of polynomials of degree less than or equal to $N-1$. Hence the details $d_p(j,k)$ corresponding to $p(t)$ vanish provided that $N \geq P+1$. It follows that the estimation of H will not be affected by the presence of the trend, as it is entirely absent from the details of the signal, i.e. $d_x(j,k) = d_s(j,k)$. This can also be given a useful spectral interpretation. The Fourier transform of a polynomial of order P consists, within the distribution theory framework, in the $P-$th derivative of the Dirac impulse function $\delta^{(P)}(\nu)$. The frequency content of a polynomial is therefore concentrated at

Testsignal	Signal type	Trend size	Whittle-fGn	Wh-fARIMA	Wavelet	D♯
1	fGn (H=0.82)	-	0.80	0.86	0.82	D1
2f	$+at$	4.61	0.88	0.94	0.81	D2
3	$+at^2$	0.676	0.80	0.86	0.82	D3
4	$+at^4$	3.8×10^{11}	1.00	1.50	1.97	D1
					2.95	D2
					3.71	D3
					0.80	D4
					0.82	D5
					0.82	D6
5	$+at^{3/2}$	432	1.00	1.50	0.86	D1
					0.81	D4
6	$+a \sin 2\pi ft$	2.70	0.86	0.92	0.78	D1
	$nf \sim 3$				**0.82**	D3
7	$+at^{-1/4}$	15.8	0.88	0.95	0.99	D1
					0.81	D3
8	non-Gaussian	-	0.79	0.85	0.81	D1
9	Gn $+at^{-1/4}$	7	0.63	0.66	0.501	D2
10	Gn $+at^{1/4}$	7	0.76	0.80	0.507	D2

Table 1: **Bias comparison with deterministic trends.** Comparison of D-Whittle and wavelet-based H estimates for test signals consisting of deterministic trends $p(t)$ of different kinds superimposed on simulated fractional Gaussian noise with $H = 0.82$ (Testsignal1). Each test signal has $n = 4096$ points. The confidence intervals are not reproduced as they are of the same order of magnitude (~ 0.02) for all three estimators. D♯ stands for the standard *Daubechies* wavelets with ♯ vanishing moments. Note that $D1$ is also the Haar wavelet. The trend size is calculated as $(\max_t p(t) - \min_t p(t))/\sigma$, where σ is the sample variance of the data $s(t)$. For Testsignal6 there are ~3 periods across the trace.

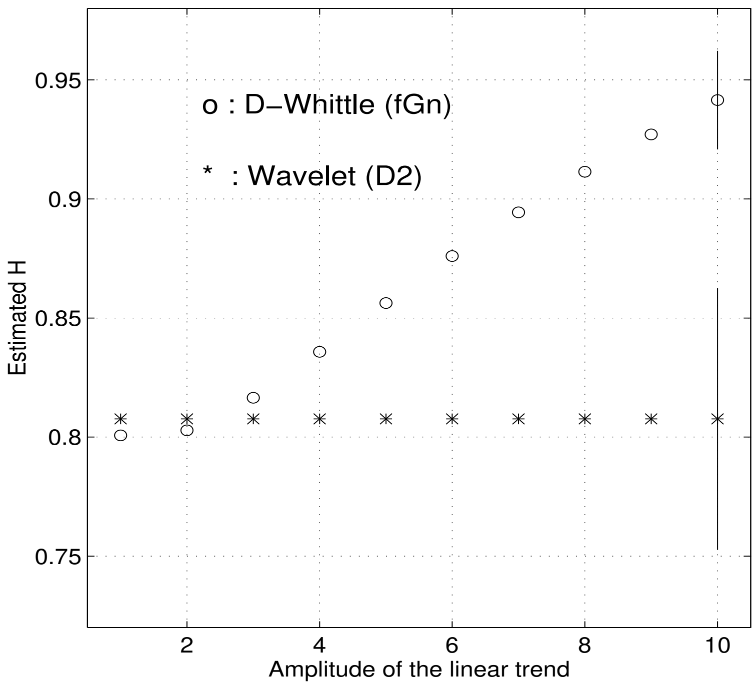

Figure 2: **Bias comparison with linear trends.** D-Whittle fGn and wavelet-based H estimates for a fractional Gaussian noise ($H = 0.82$) with superimposed linear trends of increasing amplitude. The bias in the Whittle estimate drastically increases with increasing trend, whilst the wavelet estimation remains essentially constant, provided that the analysing wavelet possesses at least two vanishing moments.

the null frequency and since wavelets are band-pass functions, in fact satisfying $|\Psi(\nu)| = O(\nu^N), \nu \to 0$, they will be blind to a given polynomial for N sufficiently large.

Arbitrary trends. When the trend is not polynomial but some smooth function, we observe that increasing N still helps to cancel its influence and very accurate estimates for H are recovered, whereas again the D-Whittle estimator gives values whose bias increases with the amplitude of the trend. Testsignal5 to Testsignal7 in table 1 are examples for very large, moderate, and large amplitude respectively. This can be interpreted in at least two ways. First, selecting a N will effectively cancel the part of the trend which can be efficiently approximated by polynomials of degree $N-1$. Increasing N would therefore approximately cancel any smooth function. The second interpretation is again from the spectral viewpoint. Smooth trends have, in most cases, an important frequency content near $\nu = 0$. The LRD phenomenon basically consists of a power-law behaviour of the spectrum near $\nu = 0$ and in general this overlap significantly complicates analysis and estimation. To see why choosing wavelets with high N significantly improves this situation consider the power-law trend $p(t) = at^{\alpha}$, a a constant. The wavelet coefficients read $d_p(j,k) = 2^{j(\alpha+1/2)} C \int |\nu|^{-(\alpha+1)} \Psi_0(\nu) \exp(\imath 2\pi k\nu) d\nu$ where C is a constant independent of the scale j. It can checked numerically that for a given j, the magnitude of these coefficients decreases with increasing N. Increasing N therefore enlarges the range of scales where $|d_p(j,k)| \ll |d_s(j,k)|$, that is where the effect of the trend is negligible.

The upper plot in figure 3 shows that when using the Haar Wavelet ($N = 1$) to analyze Testsignal5, a large number of scales, $j = 4$ to $j = 9$, are corrupted by the presence of the $t^{3/2}$ trend, preventing an accurate estimate of H or even the detection of the LRD. However with a *Daubechies4*-wavelet ($N = 4$), the log-log plot falls close to a straight line (note that there is no data point at scale 9 for $D4$ due to insufficient data). In Testsignal6 the corrupting trend consists of a low-frequency sinewave. The bottom plot in figure 3 shows that when analysed with the Haar Wavelet the polluting effect of the sine is numerically important across scales $j = 6$ to $j = 9$. With a *Daubechies4*-wavelet ($N = 4$), the effect of the trend is concentrated in scale $j = 8$, enabling the inclusion of octaves 6 and 7 in the regression. This can be explained using the spectral interpretation above. Increasing N causes the Daubechies wavelet to tend to an ideal band-pass filter, thereby concentrating the effect of the sine waveform onto a single scale.

The results for Testsignal7 in table 1 show that the D-Whittle estimator can be biased even by a trend that tends to zero. Testsignal9 gives the same result for $H = 0.5$, a particularly interesting case as it illustrates that the D-Whittle estimator is biased by a power-law decreasing trend for short range dependent data of finite length. This is in contrast to the asymptotic lack of bias for the exact MLE and Whittle estimators (see [10], p143). Testsignal10 is similar to Testsignal5 but with $H = 0.5$ and a trend of smaller amplitude.

Experimental procedure and the identification of trends. In practice one does not usually know the form of the trend, nor even necessarily that there is one. The correct experimental procedure therefore consists of performing estimations with increasing N. Successive estimates will change rapidly until N is large enough to eliminate the polynomial components of the trend, followed by a slower convergence to the final estimate as trend components of a more general form are gradually eliminated.

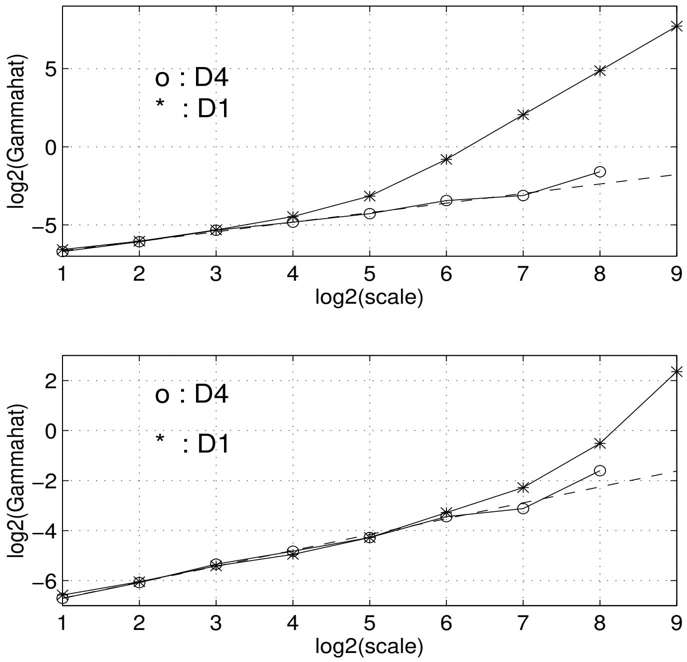

Figure 3: **Eliminating non-polynomial trends.** We compare the wavelet analysis of a fractional Gaussian noise ($H = 0.82$) using wavelets with different numbers of vanishing moments N, in the presence of corrupting deterministic trends, top: Testsignal5 ($t^{3/2}$ with large amplitude), bottom: Testsignal6 ($\sin 2\pi ft$ with moderate amplitude and $nf \sim 3$, $n = 4096$). When $N = 1$, the effect of the trend is felt over a wide range of scales, preventing a correct estimation of H. Increasing N concentrates their influence onto a very small number of scales, thereby widening the range of scales J available for the estimation.

Heuristic yet reliable identification of trends can therefore be obtained by examining the speed and nature of the convergence with N. For example a sharp transition to stable estimates can be taken as evidence of a polynomial trend of the appropriate degree, as clearly illustrated in table 1 with Testsignal4, where the entries stabilize for $N = P + 1 = 5$. This procedure will fail in the presence of discontinuous deterministic trends, for instance sharp level changes in the mean. The D-Whittle estimator will also fail in such a case.

We see that the wavelet tool allows estimation of the Hurst parameter without a prior detrending procedure. This is a very important advantage as detrending in the presence of LRD raises difficult statistical issues regarding the joint estimation of regression parameters and H.

3.3 Stationarity *vs* Stationary increments

So far we have considered wide sense stationary processes. Another fundamental feature of the wavelet-based analysis of the LRD phenomenon is the fact that it can be used to meaningfully analyse the important class of non-stationary processes $X(t)$ with *stationary increments*, that is $Y_\tau(t) = X(t+\tau) - X(t)$ is stationary. The fractional Brownian motion (fBm) is the canonical reference for such processes. Its corresponding increment-process is the fGn, which has been used as the typical example throughout this paper. The properties and implications of the wavelet tool in the study of processes with stationary increments is fully detailed in [20, 21]. Here we simply note from the fact that the spectrum of fGn reads $\Gamma_{\text{fGn}}(\nu) \simeq |\nu|^{-(2H-1)}$, that wavelet analysis can give a consistent meaning to the idea of defining the 'spectrum' $\Gamma_{\text{fBm}}(\nu) \simeq |\nu|^{-(2H+1)}$ of fBM. Thus, whereas to estimate the H parameter of a fBm signal using Whittle's technique one must first compute the increments, with the wavelet estimator one can work with the process itself. In this case, the no-bias equation (2.9) becomes $N > H$. This feature is particularly useful in the context of stationary point processes, where it may be more natural and/or more convenient to analyze the corresponding counting processes, which have stationary increments. [2, 3].

4 Wavelet-based Analysis of Ethernet data

4.1 Ethernet Data

Aims of the analysis. Ethernet data has now been extensively studied and the presence of scaling behaviour established beyond question for the great majority of traces collected. Nonetheless there are important issues which remain uncovered. We outline a more thorough approach to the analysis of LRD which reveals features of importance for model building. These ideas are valuable for the modelling of packet data in general, not only Ethernet data. More specifically, we study the LRD structure of the work itself, since it is this which is essential for performance evaluation purposes. We show that some aspects of the data analysis approach used to date are well justified, whereas others are not. We also tackle the question of the mono-fractality *vs* multi-fractality of the data. Finally we make a preliminary contribution to the stationarity question. Throughout we make use of properties of the wavelet based estimator which allow approaches which were previously very difficult or even impossible.

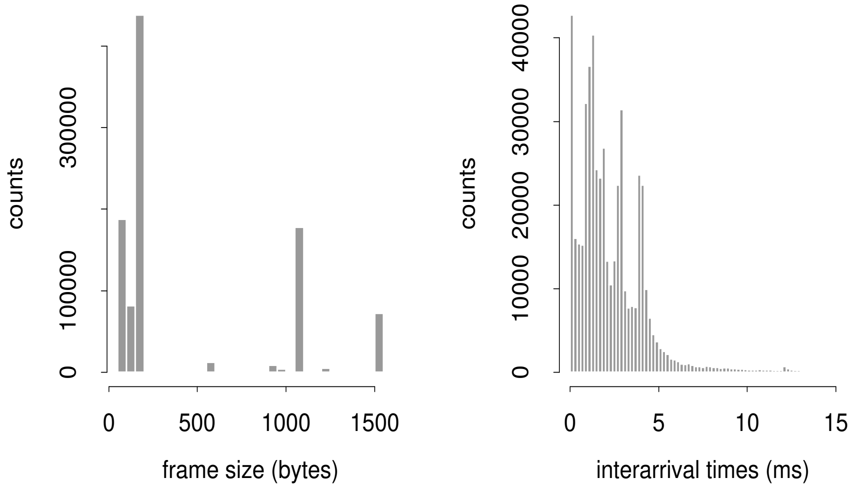

Figure 4: **Marginal frame and interarrival distributions for pAug.** The frame distribution is highly non-Gaussian, taking only a few values. The interarrival distribution is roughly exponential.

Brief description for the data. The data sets analysed are three Ethernet traces from the well known set collected at Bellcore in the late 1980's and early 1990's. They are available from anonymous ftp from ftp.bellcore.com in the files lan_traffic{pAug, pOct, OctExt}.TL.Z. These very high quality traces (no loss, jitter$<100\mu$sec.) have been described in great detail elsewhere [24]. Here we restrict ourselves to a basic description of the contents of the above files.

Each file consists of 1 million rows in two columns. Each row relates to a single Ethernet frame. The first column gives the timestamp (measured from the beginning of the trace) for the *end* of the frame in seconds. The second column gives the integer size in bytes of the frame. The actual traffic consists therefore of a sequence of disjoint alternating frames and silent periods. The Ethernet protocol imposes a minimum silence between frames of 52.6μ seconds, or 65.75 bytes, and a minimum (maximum) frame size of 64 (1518) bytes. We will measure time in either seconds or bytes as appropriate. One byte corresponds to 8×10^{-7} seconds.

First Order Statistics. Simple summary statistics for the first half of trace pAug are shown in figure 4. The results for the other two traces are similar. The frame-size histogram in figure 4 reveals that frames take typically one of just a few values. The interarrival times however can be roughly described by an exponential distribution. The statistics are calculated assuming stationarity and that the large size of the traces allows good estimates despite the LRD.

Modelling Approach. To analyse the data we need to choose a stochastic model framework. Since silences are not restricted to be multiples of bytes, a process general enough to fully capture the Ethernet arrival process must be defined in continuous time. Denote this process by $\{X_t, t \in \mathbb{R}^+\}$ where the state space is the set $\{0,1\}$, corresponding to the presence or absence of a frame. Alternatively we may ignore features at very small time scales and discretize time. The discrete version of X we denote by

pAug	X	P	W_δ	C_δ	F	S	A
H: D1	0.798	0.822	0.793	0.804	0.805	0.804	0.797
H: D3	0.809	0.822	0.794	0.822	0.818	0.809	0.784
CI: D3	0.025	0.016	0.016	0.017	0.021	0.020	0.017

Table 2: **H values for the different aspects of the data** There is no evidence of linear or quadratic trends, the estimate being stable at D2. The results are consistent with the premise that the value of H is the same for both work and arrivals, in full detail or after aggregation, and for the component sequences.

$\{W_{\delta,n}, n = 1, 2, \cdots\}$ where $W_{\delta,n}$ is the integral of X_t over $[n\delta, (n+1)\delta)$ corresponding to the total Work which arrived during this time interval, and δ is a positive constant. To simplify the representation, we could consider the arrival instants of frames only. Let $\{P_t, t \in I\!R^+\}$ be the continuous time Point process where each frame arrival is represented by a Dirac impulse, and otherwise P_t is zero. The discrete version of this we we denote by $\{C_{\delta,n}, n = 1, 2, \cdots\}$, a discrete time, discrete state space Counts process corresponding to the number of frame arrivals in the interval $[n\delta, (n+1)\delta)$. Clearly neither P nor C_δ capture the structure of arriving *work*, and as discussed in the introduction, it is this which is essential for performance evaluation purposes. Nonetheless, it is the process C_δ which has been the preferred framework to date [24, 18]. The assumption is that it captures all of the LRD behaviour, so that the work can simply be 'added' as essentially instantaneous arrivals according to the marginal distribution of the frames. We test the validity of this assumption by examining the dependence properties of the work separately. To this end, let $\{F_n, n = 1, 2, \cdots\}$ be the sequence of Frame sizes, $\{S_n, n = 1, 2, \cdots\}$ the sequence of Silence durations, and $\{A_n, n = 1, 2, \cdots\}$ the sequence of inter-Arrival times. We will examine the LRD properties of each of X, P, W_δ, C_δ, F, S and A. We exploit the fact that the wavelet estimator allows us to study each of these in a uniform framework, independent of whether the index or the state space is continuous or discrete, the nature of the marginals, or of other finer details of the respective processes.

4.2 Results: Evidence for LRD and H estimations

The complete Work process X. We will describe results for pAug. Results for the other two traces are similar unless stated otherwise. The plot in figure 5 evidences the existence of the LRD phenomenon over a very large range of scales: $(j_1, j_2) = (14, 25)$. Performing an estimation over $(j_1, j_2) = (14, 24)$ to avoid the fluctuation at scale 25 due to border effects (lack of data), we obtain $\hat{H} = 0.798 \pm 0.025$. This estimate is insensitive to an increase in N as shown in figure 5 and table 2, indicating the absence of trends in the data. See table 3 for results for the two other signals. Recall that traditional estimators, including the D-Whittle estimator, cannot be used to analyse continuous time signals such as X or P due to computational limitations.

The continuous-time point process P. When applied to P we observe (plot(b) in figure 5) LRD behaviour over an even wider range of scales: $(j_1, j_2) = (10, 23)$, allowing a very accurate estimation of H and a clear determination of the minimum time scale T_m over which the data needs to being observed before scale-invariant behaviour begins. We find, for pAug, $T_m \simeq 10$ ms and $\hat{H} = 0.822 \pm 0.012$, consistent with that obtained for X.

Signal	D1	pAug	pOct	OctExt
X	$\hat{H}$	0.798	0.825	0.947
	CI	0.024	0.026	0.017
	(j_1, j_2)	(14,24)	(15,22)	(14,22)
P	$\hat{H}$	0.822	0.824	0.943
	CI	0.012	0.016	0.017
	(j_1, j_2)	(10,23)	(13,22)	(14,22)

Table 3: **H values for the different traces**. Estimations performed on the complete work process X or the simplified continuous-time point process P are completely consistent.

It is interesting to note that plot(b) in figure 5 is strikingly reminiscent of what is observed for certain variations of the Poisson point process for which the arrival rate λ is itself a stationary LRD continuous-time process. A canonical example of such a doubly stochastic point process which may constitute a good candidate model for data is given by taking $\lambda(t) = \mathrm{fGn}_H(t)$ [4, 1, 30]. This possibility and variants of it are currently under investigation.

The aggregated processes W_δ and C_δ. From plots(a) and (b) in figure 6, the presence of LRD in the discrete or aggregated processes W_δ and C_δ is clear, and estimates for H are obtained consistent with those found for X and P (see table 2). Although intuitively this is not a surprising result, as aggregation over a given fixed scale does not affect the properties over ranges beyond this scale, we can offer a more complete explanation from the wavelet framework. As previously indicated, the wavelet decomposition basically consists in splitting data into an approximation and details. For X this reads

$$X(t) = \sum_k a_X(j_0, k)\phi_{j_0,k}(t) + \sum_{j=1}^{j_0}\sum_k d_X(j,k)\psi_{j,k}(t) \ .$$

Suppose that j_0 is chosen so as to be smaller than the smallest scale j_1 at which the LRD phenomenon occurs, then the details for $j = 1$ to $j = j_0$ can be discarded without changing the estimation of H, and we could therefore replace the analysis of X by that of the approximation $\sum_k a_X(j_0,k)\phi_{j_0,k}(t)$. In the case of the Haar multiresolution (cf. section 2) for which the scale function ϕ_0 is the indicator function over $[0, T_0]$, where T_0 is the arbitrary (very small) sampling period, one obtains exactly $W_{\delta=2^{j_0}T_0} = \sum_k a_X(j_0,k)\phi_{j_0,k}(t)$. For wavelets of higher N, the approximation is not strictly equal to W_δ but has the same low-pass approximation interpretation. This connection allows a general reformulation of the aggregation technique used to process LRD data within the multiresolution framework [5] and justifies the usual [24] practice of using C_δ to replace P in the analysis of LRD. Moreover, it offers a framework to select δ. The same arguments and conclusion hold for W_δ and X.

Frame F, silence S, and inter-arrival A sequences. Because the same H was found for X and P, or equivalently W_δ and C_δ, we may be tempted to conclude that the process P somehow 'carries' the LRD of the signal. Indeed, this is the assumption that has been made in previous work [24, 18] when C_δ has been used to represent the data. To investigate further the 'sources' of LRD in X, we now examine F, S and A. From table 2 and figure 6 we see that, remarkably, in each case the same H is recovered. These estimates are robust with respect to an increase of N. The most significant result is that for the sequence of frame sizes F. A priori these could have been mutually independent, and indeed this

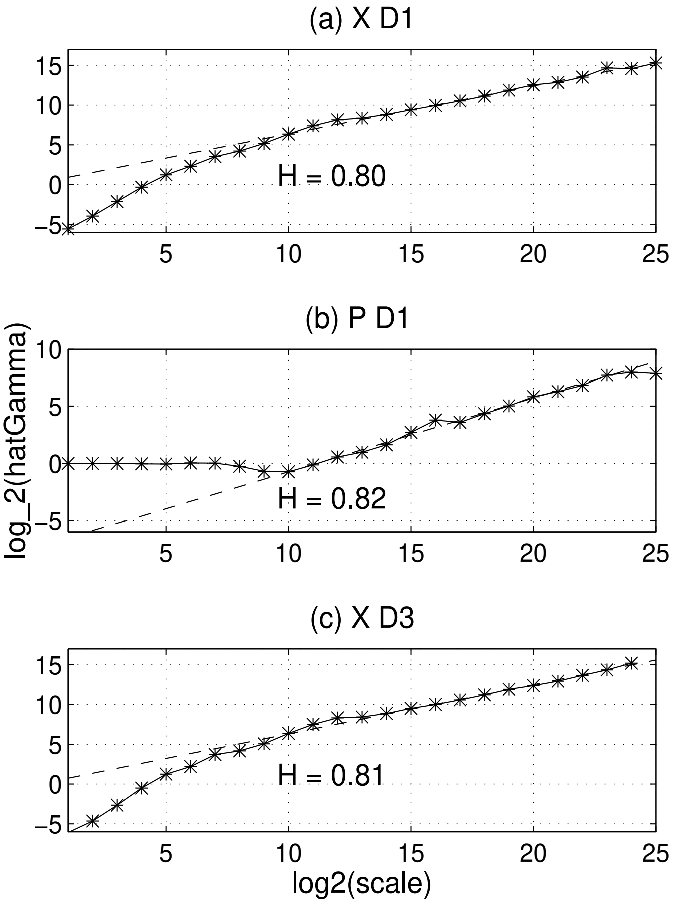

Figure 5: **Analysis of pAug: continuous time representations X and P.** Plots (a) and (c) show $\log(2^j)$ *vs* $\log_2(\hat{\Gamma}_x(2^{-j}\nu_0)) = \log_2(1/n_j \sum_k |d_x(j,k)|^2)$, with wavelets D1 and D3, for the complete work process X. It evidences a power-law behaviour and therefore LRD over a wide range of scales and enables an accurate estimation of H. (b) consists in the same plot for the P process, exhibits LRD over an even wider range of scales and yields the same estimate for H.

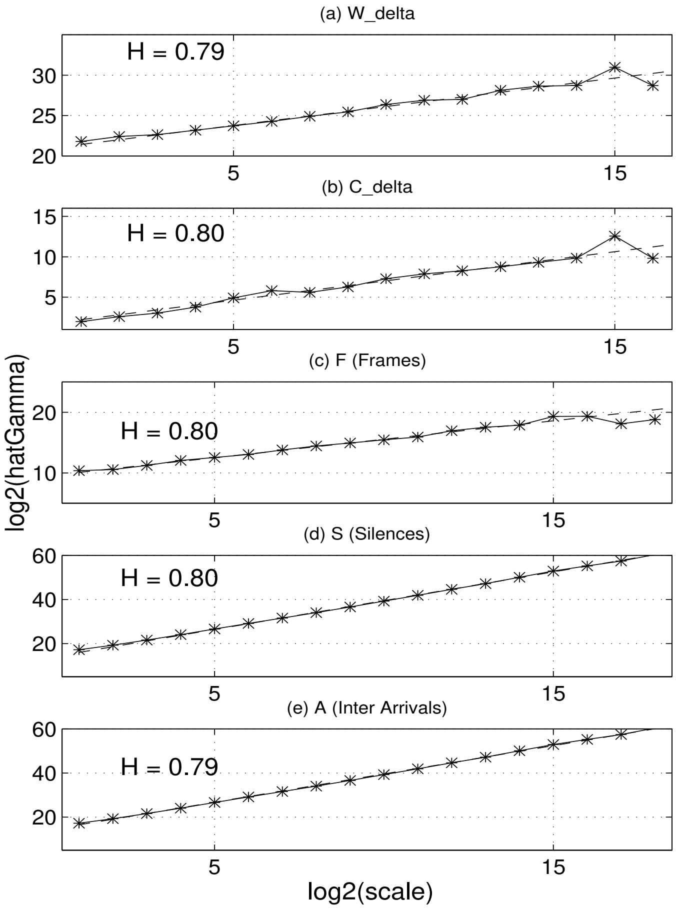

Figure 6: **Analysis of pAug: aggregated processes W_δ, C_δ, and sequences F, S, A.** Plots of $\log(\Gamma_x(2^{-j}\nu_0))$, with wavelet D1. Plots (a) and (b) show the clear presence of LRD in the aggregated processes W_δ and C_δ. The H values are very close to those of X and P respectively, illustrating the validity of the aggregation approximation. In plots (c), (d) and (e) the discrete-time processes F, S and A are analysed. Each displays LRD with almost identical H estimates, each consistent with that of X.

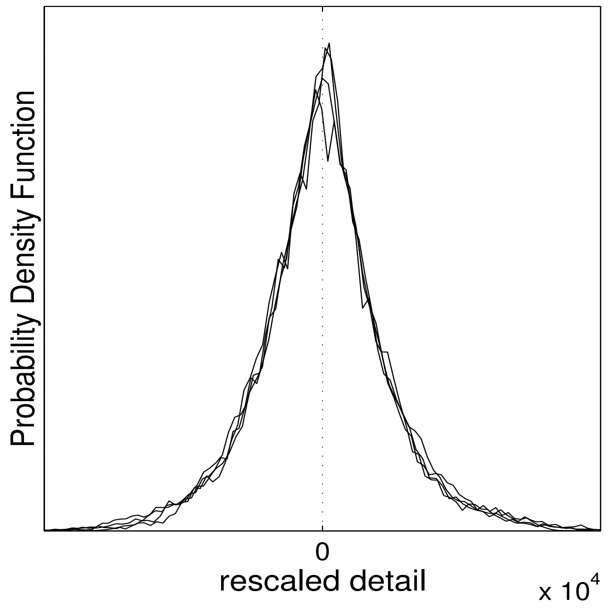

Figure 7: **Analysis of pAug: rescaled details Probability Density functions.** It shows that when rescaled by a factor $2^{-j(\hat{H}-1/2)}$, the probability density function estimates of the details of X collapse onto a single curve, close to a Gaussian function. This is evidence that a single scaling parameter fully describes the data, which is therefore mono-fractal.

has been the implicit assumption to date: that the LRD is captured by C_δ and that the frames can be added independently in a simple way. Since they are in fact LRD we see that a model based on this idea would give the incorrect estimation of arriving work, and hence misleading performance results. The lesson here, which is of general applicability, is that many aspects of a signal or data trace may be LRD, but for performance evalution purposes one must understand and include all of those which impinge on the *work* arrival process. It is not sufficient to simply model one convenient aspect which has the 'right' H. Further questions are raised here such as the nature of the cross-correlation of F and A. Work is in progress on these and on the incorporation of the results discussed here into a compact traffic model.

Monofractal *vs* **multifractality.** Now that the *self-similarity* over a wide range of scales is established, the question which immediately arises is whether this single scaling parameter H, measured from second order statistics, fully described the higher order statistics of the data. This amounts to deciding whether the data requires a *mono-* or *multi-fractal* description [8]. One way of approaching such a question is to study the probability density functions (pdf's) of the details $\{d_X(j,k), k \in \mathcal{Z}\}$ at fixed scales j. Figure 7 shows that the probability density function estimates of the rescaled coefficients $d'_X(j,k) = 2^{-j(\hat{H}-1/2)} d_X(j,k)$ collapse onto a single Gaussian function. This is known [8] to be evidence that the data are well described by the single scaling parameter H and is therefore best modelled as a mono-fractal. Similar results were found for pOct.

Stationarity with respect to H. We now make a first, non-rigorous and incomplete attempt to check the stationary of the data by examining the time variation of H. We divide W_δ into thirty-two contiguous segments and perform H estimates as if each were independent. This is justified by the

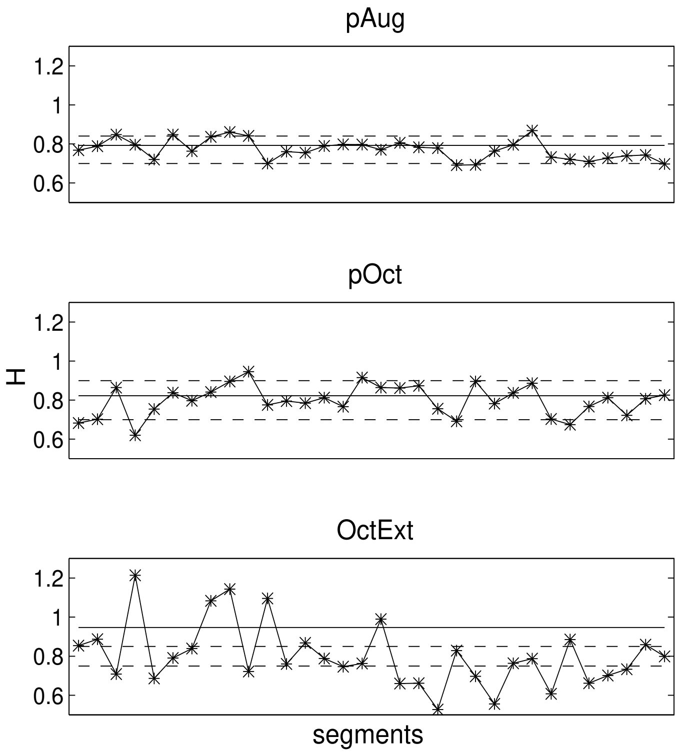

Figure 8: **Time variation of H.** Each trace is split into 32 segments and separate H estimates made for each, with a common asymptotic confidence interval shown as the interrupted lines. The solid horizontal line shows not the mean of the estimates (inferred by the confidence interval), but the H estimate taken over the entire trace. The traces pAug and pOct appear stationary, whilst OctExt does not.

quasi-decorrelation of the wavelet coefficients discussed in section 2.3 and expressed in equation (2.10). To each segment corresponds an estimation of H and a common (asymptotic) confidence interval. We refer to the mean of these 32 estimates as the 'mean', and we center the confidence interval about it. For pAug and pOct, almost all the estimates (see figure 8) lie within the confidence interval. Moreover, the mean corresponds very closely to the point estimate of H obtained from the entire trace. This is good heuristic evidence that H indeed remains constant across these traces. Hence H can remain constant over quite long intervals as pAug, the longer of the two traces, is almost an hour in duration. On the other hand the estimates for OctExt show high variability, many falling outside the confidence interval, indicating a clear non-stationarity. The mean also differs significantly from the H estimate obtained from the whole trace. This may not seem surprising as the trace is $\approx$35 hours long and the diurnal cycle is clearly visible in a time series plot, however this could have been due to mean rates varying with H remaining constant. In fact we have checked that the estimate obtained from the whole trace is not sensitive to an increase in N, excluding the possibility that the variation observed results from a smooth cyclic trend. On the contrary, the estimates being strongly varying across the segments seems to indicate that the dynamics of the process generating the traces varies with time. The results here are not conclusive because the exact nature of the correlation between estimates made in adjacent data segments is unknown.

With respect to pOct it has been noted before [16] that, as shown in the upper plot in figure 9, there is

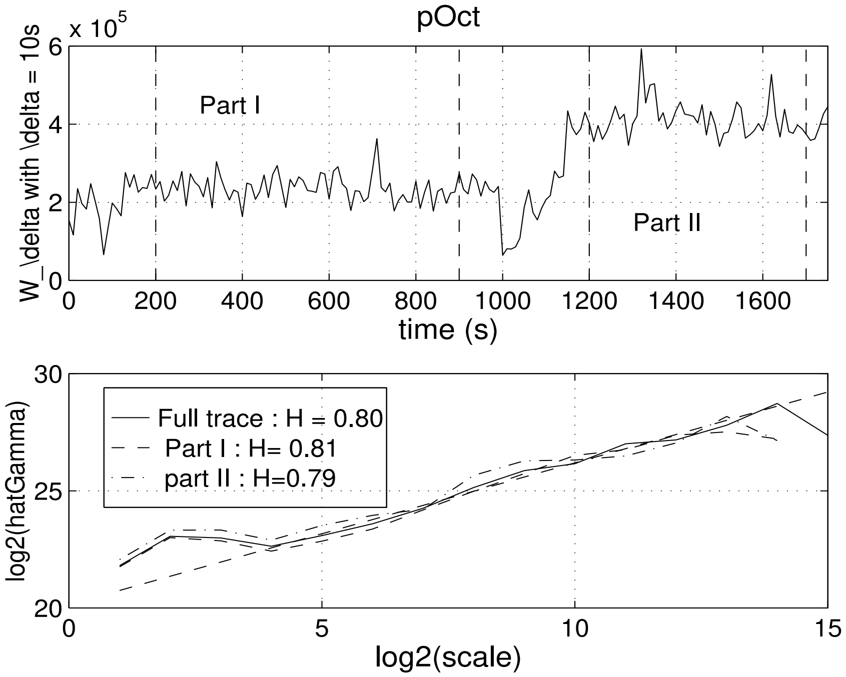

Figure 9: **Elimination of level shift in pOct**. The upper plot shows W_δ for pOct with $\delta = 10$s. A level shift seems to occur at around the 16 to 19 minute mark. The lower plot shows wavelet based estimates. Essentially identical H estimates close to 0.80 are found for the whole trace and the portions to the left and right of the shift, showing that the LRD is not an artifact of the shift.

a level shift (that is an apparent shift in the mean), at around the 16–19 minute mark of this half hour trace. As was shown in figure 4 from [16], variogram based H estimates [10] of the sub-series to either side of this transition are lower and markedly different to that for the whole trace. This fact lends itself to the interpretation that the value obtained from the whole trace has been corrupted by a non stationarity in the mean, perhaps to the point of giving the appearance of LRD when there is none. In the lower plot in 9 we perform a wavelet analysis of the trace and the sub-series and find H values which are in close agreement, in fact within the original confidence interval. It seems then that H does not vary across the level shift or the trace, that its value is indeed greater than 1/2, and that the wavelet estimator can measure it accurately despite the shift. This latter fact can be explained by noting that the the shift is actually quite smooth, occuring over approximately 3 minutes. We expect that we can eliminate it for sufficiently high N (assuming that a level shift can be viewed as a kind of linear trend). In the present situation, we used $N = 2$. This is a good example [2] in an experimental setting of where the wavelet estimator has a strong advantage over other estimators such as the variogram with respect to trend elimination.

5 Conclusion

We have introduced wavelet analysis and the $\log_2(1/n_j \sum_k |d_x(j,k)|^2)$ *vs* j plot as a tool for the examination of scale behaviour in data. This tool allows the detection of LRD and the semi-parametric measurement of H for stationary *or* stationary-increment data. It has been shown how the wavelet-based

[2] We thank the reviewer who pointed out this issue in the pOct time series and suggested these comparisons.

estimator derived from this plot is superior to alternatives, being unbiased, not only asymptotically unbiased, and also efficient under Gaussian assumptions. The unbiased property holds in general for processes with finite second moments. The estimator can be implemented by a direct computational rather than optimization based algorithm with low computational complexity, both with respect to time and memory, allowing the analysis of very large data sets. It can therefore be applied to continuous time as well as discrete time signals in a uniform framework. It also offers powerful advantages with respect to the problem of stationarity. By varying the number N of vanishing moments of the mother wavelet, polynomial trends can be rigorously eliminated, and in practice essentially any smooth deterministic trend. The advantages are twofold: accurate determination of H, and the detection and identification of trend types. In this way difficult problems in the joint statistical estimation of regression parameters and H can be entirely bypassed. Numerical comparisons against the D-Whittle estimator, the only alternative which is even asymptotically unbiased, show that even when the generated test signals belong to the underlying parametric family of the D-Whittle estimator (we used fractional Gaussian noise), the wavelet estimator is always at least as accurate. This is due to its being unbiased even for finite data samples. When the underlying parametric family is not appropriate, or in the presence of trends, the wavelet estimator is clearly superior.

We have used the advantages of the wavelet-based estimator to study in greater depth the LRD nature of some the Bellcore Ethernet traces. We found that the sequences of frames, silences and interarrival times each possess LRD with the same H as the full signal X, indicating that the approach of using only the discrete frame counting process C_δ to represent the data will not be adequate for performance evaluation purposes. This highlights an important general principle, that simply capturing *an* aspect of the data which possess LRD is not sufficient. We were able to justify however, and explain within the wavelet framework, the aggregation method whereby the continuous time work process X or frame arrival process P are replaced by the discretized versions W_δ and C_δ respectively. We performed a test of mono *vs* multi-fractality for the traces pAug and pOct, and found that a mono-fractal model is clearly preferable. Finally we performed a preliminary analysis of stationarity with respect the time variation of H. We found that pAug and pOct seem consistent with the hypothesis that H does not vary, but not OctExt.

Acknowledgements. The authors wish to thank P. Flandrin for continuous encouragement to undertake and achieve this collaboration, J. Bolot for financial and technical support, and W. Willinger for constructive remarks on this work during Les Houches spring school on Long-range dependence, May '96. This work has in part been supported by a grant from France Telecom (CNET).

References

[1] P.Abry, Transformées en ondelettes - Analyses multirésolution et signaux de pression en turbulence. Thèse de doctorat, Université Lyon I et Ecole Normale Supérieure de Lyon, (1994).

[2] P.Abry, P.Gonçalvès and P.Flandrin, Wavelet-based spectral analysis of $1/f$ processes. Proc. IEEE–ICASSP'93 (1993) III.237–III.240.

[3] P.Abry, P.Gonçalvès and P.Flandrin, Wavelets, Spectrum estimation, $1/f$ processes. Wavelets and Statistics, *Lectures Note in Statistics*, Vol. **105**, (1995), pp.15-30.

[4] P.Abry and P.Flandrin, "Point-processes, Long-Range Dependence and Wavelets," in *Wavelets in Medecine and Biology* (A. Aldroubi, M. Unser, eds.). CRC Press, Boca Raton, USA, (1996), pp.413-438.

[5] P.Abry, D. Veitch and P.Flandrin, "Long-Range Dependence: revisiting Aggregation with Wavelets" submited to Journal of Time Series Analysis, November 96.

[6] A.Aldroubi and M.Unser, "Families of multiresolution and wavelet spaces with optimal properties," *Num. Func. Anal. Opt.*, Vol. **14**, (1993), pp.417-446.

[7] Allan, W, "Statistics of atomic frequency standards." In *Proc. IEEE,* 54, (1996), pp.221-230.

[8] A. Arnéodo, E. Bacry, and J.F.Muzy, The Thermodynamics of Fractals revisited with Wavelets. *Physica A,* **213**, (1995) pp.232-275.

[9] B.Bensaou, J.Guibert, J.W.Roberts, A.Simonian, Performance of an ATM multiplexer queue in the fluid approximation using the Beneš approach, Annals of Operations Research, 49 (1994), pp.137-160.

[10] J.Beran, Statistics for Long-Memory Processes, Chapman & Hall (1994).

[11] F.Brichet, J.Roberts, A.Simonian, D.Veitch, Heavy Traffic Analysis of a Storage Model with Long Range Dependent On/Off Sources, submitted Queueing systems (1995).

[12] J.Beran, R.Sherman, M.S.Taqqu, W.Willinger, Long range dependence in variable bit rate traffic, IEEE Transactions on communications 43, No. 2/3/4, (1995), pp.1566-1579.

[13] G.L.Choudhury, W.Whitt, Long Tail Buffer Content Distributions in Broadband Networks, preprint September 1995.

[14] I.Daubechies, Ten Lectures on Wavelets. SIAM, (1992).

[15] N.G.Duffield, N.O'Connell, Large deviations and overflow probabilities for the general single-server queue, with applications, to appear in the Proc. Cambridge Phil. Soc. (1995).

[16] N.G.Duffield, J.T.Lewis and N.O'Connell, Predicting Quality of Service for Traffic with Long-Range Fluctuations, Proc. IEEE ICC'95, (1995) pp.473-478.

[17] D.E.Duffy, A.A.McIntosh, M.Rosenstein and W.Willinger, Statistical Analysis of CCSN/SS7 traffic data from Working subnetworks, IEEE J.Selected Areas in Communications, Vol. **1, No 3**, (1993), pp.329-343.

[18] A.Erramilli, O.Narayan, W. Willinger, Experimental Queueing Analysis with Long-Range Dependent Packet Traffic, to appear in IEEE Transactions in Networking (1995).

[19] W.Feller, *An introduction to Probability theory and its applications, Volume II,* J.Wiley (1971).

[20] P.Flandrin, On the spectrum of fractional Brownian motions. IEEE Trans. on Info. Theory **IT-35**, (1989) pp.197-199.

[21] P.Flandrin, Wavelet analysis and synthesis of fractional Brownian motion. IEEE Trans. on Info. Theory **IT-38**, (1992) pp.910-917.

[22] H.P.Graf, Long-range correlations and estimation of the self-similarity parameter. PhD Thesis, ETH Zürich (1983).

[23] L.Kleinrock, *Queueing systems, vol.1*, John Wiley (1975).

[24] W.Leland, M.Taqqu, W.Willinger, D.Wilson, On the self similar nature of Ethernet traffic (extended version), IEEE/ACM Transactions on Networking, February (1994), pp.1-15.

[25] S.Mallat, A theory for multiresolution signal decomposition. IEEE Trans. on Pattern Analysis and Machine Intelligence **PAMI-11)** (1989) pp.674-693.

[26] B.B.Mandelbrot, J.W. van Ness, Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10** (1968) pp.422-437.

[27] I Norros, A storage model with self-similar input, Queueing systems, 16 (1994), pp.387-396.

[28] A.Erramilli, R.P.Singh, P.Pruthi, An Application of Deterministic Chaotic Maps to Model Packet Traffic, Queueing Systems 20, (1995) pp.171-206.

[29] V.Paxson, S.Floyd, Wide-Area Traffic: The Failure of Poisson Modelling, Proceedings of SIGCOMM '94 (1994).

[30] B.Ryu, S.B.Lowen, Point Process Approaches for Modelling and Analysis of Self-Similar Traffic: Part II- Queueing Applications, preprint submitted to IEEE Globecom '96 (1996).

[31] A. Cohen, I. Daubechies and P. Vial, Wavelets on the Interval and Fast Wavelet Transforms. Applied and Computational Harmonic Analysis, Vol. 1, No 1, pp. 54–81, 1993.

[32] A.Simonian, D.Veitch, A Storage Model with High Rate and Long Range Dependent On/Off Sources, Preprint (1996).

[33] Statistical Sciences, *S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.2*, Seattle: StatSci, a division of MathSoft, Inc., (1993).

[34] M.S.Taqqu, " A bibliographical guide to self-similar process and long-range dependence. In *E. Eberlin and M.S. taqqu, eds., Dependence in Probability and Statistics*, Birkhauser, (1986), pp.137-162.

[35] M.S.Taqqu, V.T.Teverovsky, W.Willinger, Estimators for Long-Range Dependence: an Empirical Study. Preprint submitted to *Fractals*, (1996).

[36] D.Veitch, Novel Models of Broadband Traffic, Proc. IEEE Globecom '93, Houston, Texas, (December 1993).