Measuring Long-Range Dependence under Changing Traffic Conditions

Matthew Roughan Darryl Veitch

Software Engineering Research Centre, RMIT University
110 Victoria St, Carlton, Vic 3053, Australia
email: {matt,darryl}@serc.rmit.edu.au

Abstract—**Recent measurements of various types of network traffic have shown evidence consistent with long-range dependence and self-similarity. However, an alternative explanation for these measurements is non-stationarity. Standard estimators of LRD parameters such as the Hurst parameter H assume stationarity and are susceptible to bias when this assumption does not hold. Hence LRD may be indicated by these estimators when none is present, or alternatively LRD taken to be non-stationarity. The recently developed Abry-Veitch (AV) joint estimator has much better properties when a time-series is non-stationary. In particular the effect of polynomial trends in data may be intrinsically eliminated from the estimates of LRD parameters. This paper investigates the behavior of the AV estimator when there are non-stationarities in the form of a level shift in the mean and/or the variance of a process. We examine cases where the change occurs both gradually or as a single jump discontinuity, and also examine the effect of the size of the shift. In particular we show that although a jump discontinuity may cause bias in the estimates of the H, the bias is negligible except when the jump is sharp, and large compared with the standard deviation of the process. We explain these effects and suggest how any introduced errors might be minimized. We define a broad class of non-stationary LRD processes so that LRD remains well defined under time varying mean and variance. The results are tested by applying the estimator to a real data set which contains a clear non-stationary event falling within this class.**

I. Introduction

In the last few years the discovery of the *self-similar* nature of many kinds of packet traffic [12], [15] has inspired a small revolution in the way that high-speed traffic is viewed. Although no single model is accepted as definitive, the *Hurst* parameter H, which describes the degree of self-similarity, holds a central place in the description of such traffic. Its accurate measurement is therefore of considerable importance for the provision of quality of service as well as for capacity planning.

The existing literature on traffic modeling, and indeed on teletraffic performance analysis, is dominated by *stationary* models. However, there are good reasons to suppose that traffic conditions *do* change; for example the concept of a *busy hour* is important both in voice and data networks. Although the discovery of self-similarity in packet traffic has led to a much wider range of traffic models, they are nonetheless stationary.

In many traffic situations the changing conditions or non-stationarity can be safely ignored because, over the time scales of interest, they have little effect – the process may be well approximated by a stationary model. However scaling properties such as self-similarity and Long-Range Dependence (LRD) are inherently defined over a *range* of scales, which may well encompass periods where stationarity is a poor approximation. In this event, the question arises as to whether LRD is well defined, and if so, how to estimate its parameters accurately. Standard approaches – for example the R/S plot and Whittle estimators – may be inaccurate to the point of indicating LRD exists when in fact it does not [13], [18]. This failure of standard estimators has led some to question the extensive body of data demonstrating LRD and self-similarity in data traffic. Furthermore the very nature of LRD processes can cause confusion – the long term correlations cause apparent trends, encouraging the erroneous conclusion that the data is non-stationary.

The difficulties of distinguishing LRD and non-stationarity are not avoided by measuring other features of the data. Even the perennial sample mean is far more variable for LRD processes, so a test for non-stationarity based on the sample mean under Short Range Dependent (SRD) assumptions would lead to incorrect conclusions if the process is, in fact, LRD. However a test for stationarity of the mean under LRD assumptions requires a reliable estimate of the parameters of LRD! Hence it is important to be able to measure LRD meaningfully and accurately without *a priori* knowledge of whether or not a data set is non-stationary, or the exact form a non-stationary may take.

We present a set of tests of the Abry-Veitch (AV) estimator for the parameters of LRD [21], [20], and demonstrate that it is robust to a broad class of non-stationary behavior; that is, the AV estimator remains accurate even when the assumption of stationarity upon which it was based is invalid. We explain the reasons for its robustness, determine its limitations, and suggest methods for mitigating any residual bias.

We take care to define, in Section II, a class of non-stationary processes where LRD is well defined. Within this class the mean and/or variance are allowed to change whilst the parameters of LRD, including the Hurst parameter, remain well defined and constant. In previous papers the AV estimator has been shown to be robust to polynomial trends in the mean [4]. In this paper we focus on the case where the mean and/or variance undergoes a level shift, one of the simplest ways to produce a large bias in estimators of the Hurst parameter. In this context, the measurement of the Hurst parameter corresponds to measuring the 'stationary part' of the traffic under non-stationary conditions. The AV estimator allows this to be achieved robustly.

We describe the AV estimator in Section III, and in Sections IV its robustness is verified through simulation and well substantiated arguments. The major result is that the AV estimator for H remains unbiased except in the case of jumps in the mean which are both large and sharp, where small biases (~ 0.05) can be introduced. The sensitivity of the results to parameters of the analysis such as the wavelet basis is also discussed.

In addition to simulations we examine a real Ethernet dataset in Section V which appears to contain a non-stationary level shift. We show that the robustness property holds for this data set, simultaneously verifying the robustness of the AV estimator for real data, and the conclusion that Ethernet traffic is consistent with a non-stationary LRD model.

The non-stationary LRD traffic models and robustness results presented here lend credence to recent studies such as [8], [9], [10], [4], [21] and [16] which use the AV estimator to demonstrate LRD in data traffic, and to the study of LRD in traffic in general. Furthermore, this study adds to the list of benefits of using the AV estimator, which already includes a run time complexity of only $O(n)$, negligible bias, statistical efficiency [21], the ability to be performed in real time [16], joint estimation of LRD parameters other than just the Hurst parameter [21], [3],

[20], known confidence intervals for estimates [21], [4], [2], and the possibility of performing a test of the constancy of H and other scaling exponents [19].

II. The Traffic Models

A. Preliminaries

In this paper we deal with second order traffic modeling, that is Gaussian models, where the autocovariance function and the mean specify the model completely. In general terms the results of the paper are also valid for non-Gaussian processes, however in that case the second order statistics cannot specify the processes fully. The models will be defined in discrete time, corresponding to the discrete time series obtained from real data.

We define the *mean* of a process $X(t)$ to be $m_X(t) = \mathbb{E}[X(t)]$, and its *variance* as $\sigma_X^2(t) = \mathbb{E}[(X(t) - m_X(t))^2]$. The *autocovariance* is given by $R_X(t,s) = \mathbb{E}[(X(t) - m_X(t))(X(s) - m_X(s))]$, and the *autocorrelation* is defined to be $\Gamma_X(t,s) = R_X(t,s)/\sigma_X(t)\sigma_X(s)$. If X is stationary then the mean and variance are the constants m_X and σ_X^2 respectively, and the autocovariance and autocorrelation are functions of the *lag* $k = |t - s|$ only, and we denote them by $R_X(k)$ and $\Gamma_X(k) = R_X(k)/\sigma_X^2$ respectively. In the stationary case the Fourier Transform of R_X is known as the *spectral density* and we denote it by f_X.

LRD is commonly defined by the slow, power-law decrease in the autocovariance function of a second order stationary process: $R_X(k) \sim c_r|k|^{-(1-\alpha)}$, $k \to \infty$, $\alpha \in (0,1)$, or equivalently as the power-law divergence at the origin of its spectrum: $f_X(\nu) \sim c_f|\nu|^{-\alpha}$, $|\nu| \to 0$, ([6], p.160). The power-law decay is such that the sum of all correlations is always appreciable, even if individually the correlations are small. The past therefore exerts a long term influence on the future, exaggerating the impact of traffic variability and rendering statistical estimation problematic. The main parameter of LRD is the dimensionless scaling exponent α. It describes the qualitative nature of scaling – how behavior on different scales is related. The second parameter, c_r or c_f, is a quantitative parameter which gives a measure of the magnitude of LRD induced effects. The two are related by $c_f = 2(2\pi)^{-\alpha} c_\gamma \Gamma(\alpha) \sin((1-\alpha)\pi/2)$, where Γ is the Gamma function.

As an example of the importance of each parameter, consider the statistical behavior of the sample mean estimator of the mean of a stationary process $X(t)$ with data length n. The classical result is that for large n the sample mean follows a normal distribution, with expectation equal to m_X, and variance σ_X^2/n. In the case where X is LRD the sample mean is also asymptotically normally distributed with mean m_X, however the variance is given by $\frac{2c_r n^\alpha}{(1+\alpha)\alpha} \cdot \frac{1}{n}$ [6]. Note that both c_r and α appear in this expression, but the variance does not. Note also that the variance in the LRD case shrinks at a slower rate with n than in the classical case, so that for large n the confidence intervals will be far larger than classical theory would predict.

Although LRD is typically defined in relation to the autocovariance function, an entirely equivalent definition could be made in terms of the autocorrelation function:

$$\Gamma_W(k) \sim \underline{c}_r|k|^{-(1-\alpha)},\ k \to \infty \quad \alpha \in (0,1), \tag{1}$$

where the dimensionless constant $\underline{c}_r \equiv c_r/\sigma_X^2$ has replaced c_r, which has the dimensions of variance. We adopt this normalized way of defining LRD, as it is central to our generalization to non-stationary LRD models. Note that both $\underline{c}_r$, and the frequency domain equivalent $\underline{c}_f$, take values in $(0,1)$. Second order stationary processes which are not LRD are called Short Range Dependent (SRD), corresponding to $\alpha = 0$.

It is common practice to describe LRD through the *Hurst* parameter $H = (1+\alpha)/2$, though in fact H is the parameter of *self-similarity* and is properly used to describe only self-similar processes, which are non-stationary. The connection to LRD is that if a process Y (with finite second moments) is self-similar with parameter $H \in (0,1)$, then its increment process $X(t) = Y(s+t) - Y(s)$ is LRD with $\alpha = 2H - 1$. We follow this convention of writing H instead of α.

For simulation purposes sample paths of *Fractional Gaussian Noise* (FGN) are generated using a standard spectral technique. The FGN is a well known canonical Gaussian LRD process which is the increment process of the *Fractional Brownian Motion*. The FGN $Z(t)$ has autocorrelation function

$$\Gamma_Z(k) = \frac{1}{2}\left(|k+1|^{2H} - 2k^{2H} + |k-1|^{2H}\right), \tag{2}$$

for $k \geq 0$. Note that if $H = \frac{1}{2}$ then $\Gamma_Z(k) = 0$ for all $k \geq 1$, corresponding to white noise, but when $H \neq \frac{1}{2}$

$$\Gamma_Z(k) \sim H(2H-1)k^{2H-2}\ , \quad k \to \infty, \tag{3}$$

identifying $\underline{c}_r$ as $\underline{c}_r = H(2H-1)$. The identity given above yields the following useful relation between the variance of a FGN and its value of c_f:

$$c_f = \sigma_Z^2 \cdot 2(2\pi)^{1-2H} H(2H-1)\Gamma(2H-1)\sin(\pi(1-H)), \tag{4}$$

In this paper, by FGN we refer to FGN with $\sigma_Z = 1$, the so called *standard* FGN. Further details on the FGN process can be found in [6], [17].

B. A Stationary Class of LRD Models

Stationary models dominate traffic modeling, and performance analysis in general. Formally, a stochastic process $X(t)$ is stationary if, for each m, the m dimensional joint distribution of $\{X(t_1+\tau), X(t_2+\tau), \ldots, X(t_m+\tau)\}$ is independent of τ for any set of m times $\{t_1, t_2, \ldots, t_m\}$.

A stationary Gaussian model for $X(t)$ can be expressed as two simple transformations of a stationary Gaussian process $W(t)$ with zero mean and unit variance. Namely the variance of the normalized $W(t)$ may be changed by multiplication, and the mean changed by addition, yielding $X(t; m, \sigma, \Gamma_W) = m + \sigma\, W(t)$, where m and σ are positive constants. The above parametrisation separates out the *location* parameter $m_X = m$, and the *scale* parameter $\sigma_X = \sigma$ of the process, from the *shape* parameter, which is the role played by the entire, as yet unspecified, autocorrelation function $\Gamma_X(k) = \Gamma_W(k) = R_W(k) = \mathbb{E}[W(t+k)W(t)]$. The autocovariance of X is just $R_X(k) = \sigma^2\Gamma_W(k)$.

We now partially specify Γ_X by requiring that X be LRD, that is, we assume that it obeys (1). A semi-parametric class of LRD traffic models can therefore be defined as

$$X(t; m, \sigma, H, \underline{c}_r) = m + \sigma\, W(t; H, \underline{c}_r). \tag{5}$$

C. A Non-Stationary Class of LRD models

Despite the dominance of stationary modeling, it has long been known that traffic conditions *do* change. Stationary models, even fractal ones, are not always adequate. It is difficult however to move to non-stationary paradigms, as there are so many kinds of non-stationarity, and which ones are most appropriate involves many unanswered empirical issues. In this section we present a class of *Non-Stationary LRD* models (NS LRD), where long range dependent processes are generalized to allow non-stationarities of certain well defined kinds. More specifically, we begin with a stationary LRD model, and define a class of non-stationary variations by transforming it to induce a change in the mean and/or variance, whilst the parameters measuring the LRD, including H, remain well defined and constant. In this way some time-varying properties are allowed, and are well defined, but important features of the original stationary model remain, and remain well defined also.

A class of non-stationary LRD models for the traffic rate $X(t)$ is again given by transformation of the mean zero, unit variance LRD $W(t; H, \underline{c}_r)$, resulting in

$$X(t; m, \sigma, H, \underline{c}_r) = m(t) + \sigma(t)\, W(t; H, \underline{c}_r), \qquad (6)$$

where $m(t)$ and $\sigma(t)$ are positive functions of time. Comparing with (5), we see that the location and scale parameters have become time varying, but the shape function Γ_W, and its associated parameters $(H, \underline{c}_r)$, do not change. In fact $m_X(t) = m(t)$, and $\sigma_X^2(t) = \sigma^2(t)$ and

$$\begin{aligned} R_X(t,s) &= \sigma(t)\sigma(s)\Gamma_W(t-s; H, \underline{c}_r), \\ \Gamma_X(t,s) &= \Gamma_W(t-s; H, \underline{c}_r) \equiv \Gamma_W(k; H, \underline{c}_r). \end{aligned} \qquad (7)$$

Thus, although the autocovariance function is no longer a function of the lag only, the autocorrelation function retains this property despite the non-stationarities in location and scale. Since we have used a definition of LRD based on such an autocorrelation function, it remains well defined, and gives a precise meaning to the notion of *non-stationary LRD models*, where the LRD parameters $(H, \underline{c}_r)$ retain their physical meanings, and remain constant. Thus, in this framework the estimation of $(H, \underline{c}_r)$ has the meaning of measuring the 'stationary part' of the non-stationary traffic model. In this paper we concentrate on the robust estimation of H. Although the estimation of c_r in the normal stationary context is well understood ([21], [20]), the estimation of $\underline{c}_r$ in the non-stationary context is more difficult and will be studied elsewhere.

For the remainder of the paper we will consider a particular form of $m(t)$ and $\sigma(t)$, namely that of a *level shift* – a monotone transition from one constant level to another, with a smoothness parameter specifying the abruptness of the transition. The motivation for this is two-fold. First, by studying such changes of level, we consider a conceptually simple yet quite extreme form of non-stationarity which is also physically meaningful. Second, as discussed in more detail below, it has already been shown [4] how polynomial trends and trends well approximated by polynomials are eliminated using a wavelet based estimator. Discontinuous changes on the other hand are poorly approximated by polynomials and require a separate treatment.

The level shifts are defined using the following family of transition functions:

$$T(t; J, S, L) = 1 + \frac{J}{2} + \frac{J}{\pi} \arctan\left(\frac{t-L}{S}\right), \qquad (8)$$

where $J \in I\!R$ is the size of the level shift, $S \geq 0$ is a smoothness parameter, and $L \in I\!R$ is a location parameter. The transition is made from a level of 1 to $1 + J$. In *Figure 1* four

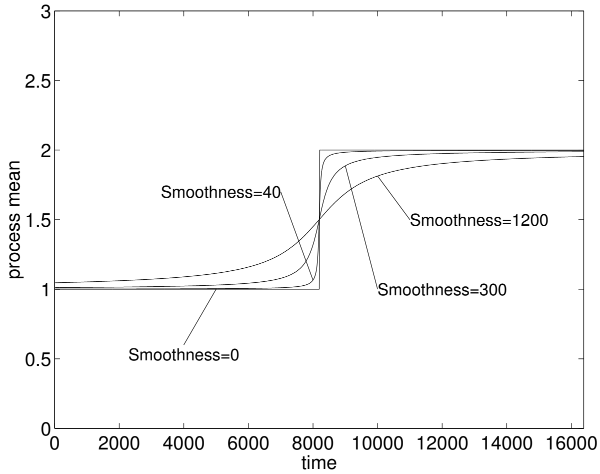


Fig. 1. The transition functions, with jump size $J = 1.0$.

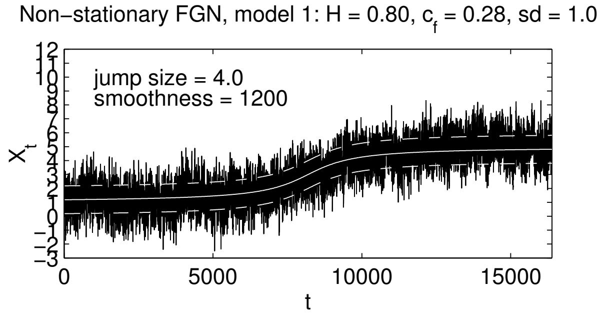


Fig. 2. Non-stationary FGN (parameters shown on each subplot). The white lines show the mean, while the dashed lines show one standard deviation about the mean. The left (right) figure shows NS FGN's constructed according to Model I (resp. II).

members of the family are illustrated with smoothness values $S = \{0, 40, 300, 1200\}$, each with $J = 1$ and $L = 8192$. The same smoothness values are used in simulations although, due to space limitations, typically only results for $S = \{0, 300\}$ will be shown. The case $S = 0$ corresponds to the limit of the above function as $S \to 0$ from above, namely a step function. The smoothness parameter has the dimensions of time and gives a measure of the duration of the 'transition region'. A dimensionless measure of the rapidity of change across the region, a shape-like parameter, is given by J/S. Two jump sizes are considered here, $J = 2\sigma$ and $J = 4\sigma$. For simplicity we set $\sigma = 1$ in what follows. Data sequences are typically of length $n = 2^{14} = 16384$, with a level shift occurring in the middle, so that $L = n/2$ as in *Figure 1*. *Figure 2* shows an example data sequence.

III. The Abry-Veitch Joint Estimator

In [21], [20] a semi-parametric joint estimator of LRD in the frequency domain, i.e. of (H, c_f), is described based on the *Discrete Wavelet Transform* (DWT). We now summarize this approach and the properties of the estimator.

A. Wavelets and the Dyadic Grid

The Wavelet transform can be understood as a more flexible form of a Fourier transform, where $X(t)$ is transformed, not into a frequency domain, but into a time-scale wavelet domain

(a,t), $a \in I\!R^+$, $t \in I\!R$. The sinusoids of Fourier theory are replaced by wavelet basis functions $\psi_{a,t}(u) \equiv \psi_0(\frac{u-t}{a})/\sqrt{a}$ generated by simple translations and dilations of the *mother wavelet* ψ_0, a band pass function with limited spread in both time and frequency. The wavelet transform can thus be thought of as a method of simultaneously observing a time series at a full range of different scales a, whilst retaining the time dimension of the original data. Multiresolution analysis theory [7], [1] shows that no information is lost if we sample the continuous wavelet coefficients at a sparse set of points in the time-scale plane known as the *dyadic grid*, defined by $(a,t) = (2^j, 2^j k)$, $j,k \in I\!N$, leading to the *Discrete Wavelet Transform* with discrete coefficients $d_X(j,k)$ known as *details*. By using the DWT very significant computational advantages are gained, as the details can be computed by a fast pyramidal algorithm with complexity of only $O(n)$. In fact the computational load and memory requirements are so low that real-time implementations are possible with inexpensive hardware [16]. Henceforth we deal exclusively with the details of the DWT. The *octave* j is simply the base 2 logarithm of scale $a = 2^j$, and k plays the role of time (although a time whose rate varies with j). For finite data of length n, j will vary from $j = 1$, up to some $j_2 \approx \log_2(n)$. The number of coefficients available at octave j is denoted by $n_j \simeq n/2^j$.

B. The Logscale Diagram

The wavelet approach is so effective for the analysis of scaling phenomenon because the wavelet basis themselves possess a scaling property, and therefore generate a matched 'co-ordinate system' naturally suited to the study of scaling. The practical outcome is that the LRD in the time domain is reduced to residual **short** range correlation in the wavelet coefficient plane $\{j,k\}$, thus removing entirely the special estimation difficulties. In fact for each fixed j, we can regarded the detail series $d_X(j,\cdot)$ as a stationary process with weak short range dependence, and the series as uncorrelated across scales. To exploit these properties, averages are taken across time, to form

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j,k)|^2. \tag{9}$$

The random variable μ_j is a non-parametric, unbiased estimator of the variance of $d_X(j,\cdot)$ (the means of the details are zero), and can be thought of as a way of concentrating the gross second order behavior of X at octave j. Furthermore, the μ_j are themselves only weakly dependent, so the analysis of each scale is largely decoupled from that at other scales. To analyze the second order dependence of $X(t)$ on scale therefore, we are naturally lead to study μ_j as a function of j. Since we consider LRD to be essentially a power-law behavior of second order moments, this is naturally done in a log-log plot, called the *Logscale Diagram*, examples of which appear throughout the paper, for example in *Figure 3*. Note that the vertical confidence intervals about each of the $y_j = \log_2(\mu_j)$ increase with j since $n_{j+1} = n_j/2$. (Note that in forming y_j small corrective terms are in fact subtracted from $\log_2(\mu_j)$). Prior to any estimation LRD must first be detected by detecting regions of *alignment* in the Logscale Diagram from some lower scale j_1 up to the largest scale in the data. Further discussion on the use of the Logscale Diagram and detailed issues which cannot be entered into here can be found in [4], [21], [5], [3].

C. The Estimator

Assuming that a valid alignment has been detected between octaves j_1 and j_2, the Abry-Veitch joint estimator of the LRD parameters (α, c_f) can then be used by performing a weighted linear regression over the scales $j \in [j_1, j_2]$. Exact expressions for the weights are available in terms of special functions [21], however for moderate to large n_j they are very well approximated by $2(\log_2 e)^2/n_j$ at octave j. The slope of the regression is simply the exponent α, and H is estimated as $\hat{H} = (1+\hat{\alpha})/2$. The estimator $\hat{c}_f$ of c_f is related to the intercept of the regression: $\hat{c}_f = p2^{\hat{a}}$ where $\hat{a}$ is the intercept and p a known bias correction factor (see [21] for details). The normalized form can be estimated using $\hat{\underline{c}}_f = \hat{c}_f/S^2$ where S^2 is the unbiased sample variance estimator of the variance of X. It can be shown that this joint estimator, under some additional technical hypotheses [21], [3], is unbiased and has very close to minimal variance. It performs well under deviations from the said hypotheses, and is close to unbiased in practice even for short sequences. Note that the multiplication of $X(t)$ by a factor σ induces corresponding factors of σ^2 in c_f and $\hat{c}_f$, but does not affect H or $\underline{c}_f$, nor their estimates. Further details of the wavelet based estimation of H, can be found in [4], [21], [5], [2], [3].

An important flexibility inherent in the wavelet based analysis is the ability to freely choose a property of the mother wavelet, the number N of *vanishing moments* [7]. This property has important implications with respect to robustness to smooth additive trends [4]. More precisely, if $p(t)$ is a polynomial of order s with $s < N$, then the details of $X(t) = p(t) + W(t)$ will be the same as those of $W(t)$, as wavelets with $N > s$ are 'blind' to such polynomials. The polluting polynomial does not have to be small in magnitude, it can in fact be far larger than the random signal itself. In practice, estimation bias due to the presence of deterministic 'trends' which are smooth, though not polynomial, can also be largely eliminated [4], [3]. Such trends include sinusoidal, power-law decreasing, and even power-law increasing functions (provided their exponent does not exceed that of the stochastic component). Discontinuous 'trends' however cannot be eliminated in this way, motivating us to study them here.

IV. Robustness Tests

In this section we investigate level changes in mean and variance separately. Models which are particular combinations of level changes in mean and variance can be naturally understood by combining the individual effects of the mean and variance transformations in the appropriate way.

A. Robustness to Mean Level Changes

We begin by investigating the robust estimation of (H, c_f) of a standard FGN to which a transition function has been **added**, corresponding to an increase in mean with constant variance:

$$X(t) = T(t; J, S, n/2) + W(t; H, \underline{c}_f). \tag{10}$$

Note that for this model c_f remains well defined, as the variance is constant. We therefore include estimates for c_f using the joint AV estimator [21], though the focus will remain on H.

Apart from the work mentioned above where use is made of the vanishing moments of the wavelets, there is some previous work on the efficient detection of LRD and measurement of H in the presence of deterministic non-stationarities in the

			Stationary FGN Estimates		NS FGN: Mean Shift		Variance Shift
H	S	J	$\hat{H}$	$\hat{c}_f$	$\hat{H}$	$\hat{c}_f$	$\hat{H}$
0.50	0	4.0	0.498 ± 0.0032	0.9792	0.555 ± 0.0029	0.7743	0.497 ± 0.0034
		2.0			0.521 ± 0.0029	0.8907	0.498 ± 0.0031
	40	4.0			0.515 ± 0.0029	0.9096	0.497 ± 0.0033
		2.0			0.505 ± 0.0028	0.9515	0.498 ± 0.0031
	300	4.0			0.498 ± 0.0029	0.9782	0.497 ± 0.0032
		2.0			0.498 ± 0.0029	0.9790	0.498 ± 0.0030
	1200	4.0			0.498 ± 0.0029	0.9793	0.498 ± 0.0030
		2.0			0.498 ± 0.0029	0.9793	0.498 ± 0.0029
0.80	0	4.0	0.799 ± 0.0032	0.2716	0.811 ± 0.0035	0.2581	0.799 ± 0.0051
		2.0			0.802 ± 0.0036	0.2679	0.799 ± 0.0047
	40	4.0			0.801 ± 0.0036	0.2689	0.799 ± 0.0051
		2.0			0.799 ± 0.0036	0.2710	0.799 ± 0.0046
	300	4.0			0.799 ± 0.0036	0.2715	0.799 ± 0.0049
		2.0			0.799 ± 0.0036	0.2716	0.799 ± 0.0045
	1200	4.0			0.799 ± 0.0036	0.2716	0.799 ± 0.0046
		2.0			0.799 ± 0.0036	0.2716	0.799 ± 0.0043

TABLE I

HURST PARAMETER ESTIMATES WITH $j_1 = 2$ BEFORE AND AFTER A MEAN CHANGE OR A VARIANCE CHANGE. EACH OF THE RESULTS $\hat{H}$ AND $\hat{c}_f$ ARE THE AVERAGE OF 30 TESTS. IN ADDITION FOR H THE 95% CONFIDENCE INTERVALS ARE SHOWN, BASED ON THE KNOWN PERFORMANCE OF THE ESTIMATOR FOR THE STATIONARY FGN AND THE VARIANCE OF THE 30 TESTS IN THE NS FGN CASES.

mean. In [18] modifications to estimators of Whittle type allow LRD and two kinds of non-stationarities in the mean: levels changes and decreasing power-law trends, to be distinguished. We show below that the wavelet-based approach outlined here is more powerful as it allows for robust estimation without the need for a preliminary analysis to check for and to determine the type of non-stationarity, and a wider range of non-stationarities are allowed. It also leads to estimates where the bias due to the non-stationarities is lower.

We perform the AV estimation procedure both on realizations of standard FGN, and on those **same** realizations after transformation, and compare the two. As noted in [21], although FGN has an almost 'pure' power-law spectrum, not all scales can be used due mainly to the presence of initialization errors in the wavelet decomposition. The lower cutoff scale used in the estimation is therefore set to $j_1 = 2$.

The estimates of H and c_f presented in *Table I* are averages of AV estimates over 30 realizations, and can be thought of as estimates of the expectation of the estimators. The 95% confidence intervals noted in the table – measured in the NS case and known in the stationary case – indicate the variance of the average estimates, and can therefore be used to compare the stationary and non-stationary results. It is important to understand that, since the transformations are deterministic, in *Table I* we are not so much interested in the statistical performance of the estimators, but rather the systematic change in the estimates induced by the transformation. Indeed the performance before the transformation is already known for both H and c_f.

The first two result columns show the estimates of (H, c_f) obtained for the original stationary FGN, to be used as a control. The second column shows the NS FGN estimates after the mean level shift transformation. It is seen that, except in 5 cases, the NS estimates of H fall within the control confidence intervals based on the stationary FGN. The exceptions: $(H, S, J) = (0.5, 0, 2)$, $(0.5, 0, 4)$, $(0.5, 40, 2)$, $(0.5, 40, 4)$, $(0.8, 0, 4)$, occur when the transition is sharp and the level shift large, and are more severe for low H. The changes in the estimates for c_f are also only notable for sharp, large shifts. Moreover, even in the most extreme case the bias is $\sim 10\%$ – hence it is very unlikely that apparent strong evidence for LRD, such as a measurement $\hat{H} > 0.6$, is in fact due to non-stationarity in the mean. Furthermore, any non-stationarity large and sharp enough to cause a 10% bias is easy to detect as the jump size is of the order of four times the standard deviation of the process.

For a deeper understanding of these results we must examine the Logscale Diagrams of the data. The deterministic changes caused by the addition of the transition function can be observed by superposing in the same Logscale Diagram the results before and after the transformation, as shown in *Figure 3* for two values of H. Again, averages of 30 realizations are given to show the systematic changes due to the transformation. These average y_j values can be taken to be valid estimates of the expectation of the respective y_j's. The vertical 95% confidence intervals shown correspond to a single observation, and were calculated based on the 30 measurements of the NS y_j's.

When $H = 0.8$ we can see in *Figure 3*, that in most cases there is very little change between the mean y_j's for the NS FGN and the stationary FGN. Hence the accuracy of the estimates of (H, c_f). When there is a large, sharp jump (for instance $J = 4, S = 0$) there is a noticeable deviation at higher scales from the values of y_j for the stationary control. This leads to bias in the estimates, though it is limited because the regression used to estimate (H, c_f) from the y_j is weighted, giving less weight to higher scales. This corruption of the higher scale y_j's is more evident when the Hurst parameter is smaller. It is most evident in *Figure 3* when $H = 0.5$.

To explain the results we examine *Figure 4* which shows the Logscale Diagrams of the transitions functions $T(t; J, S, L)$ de-

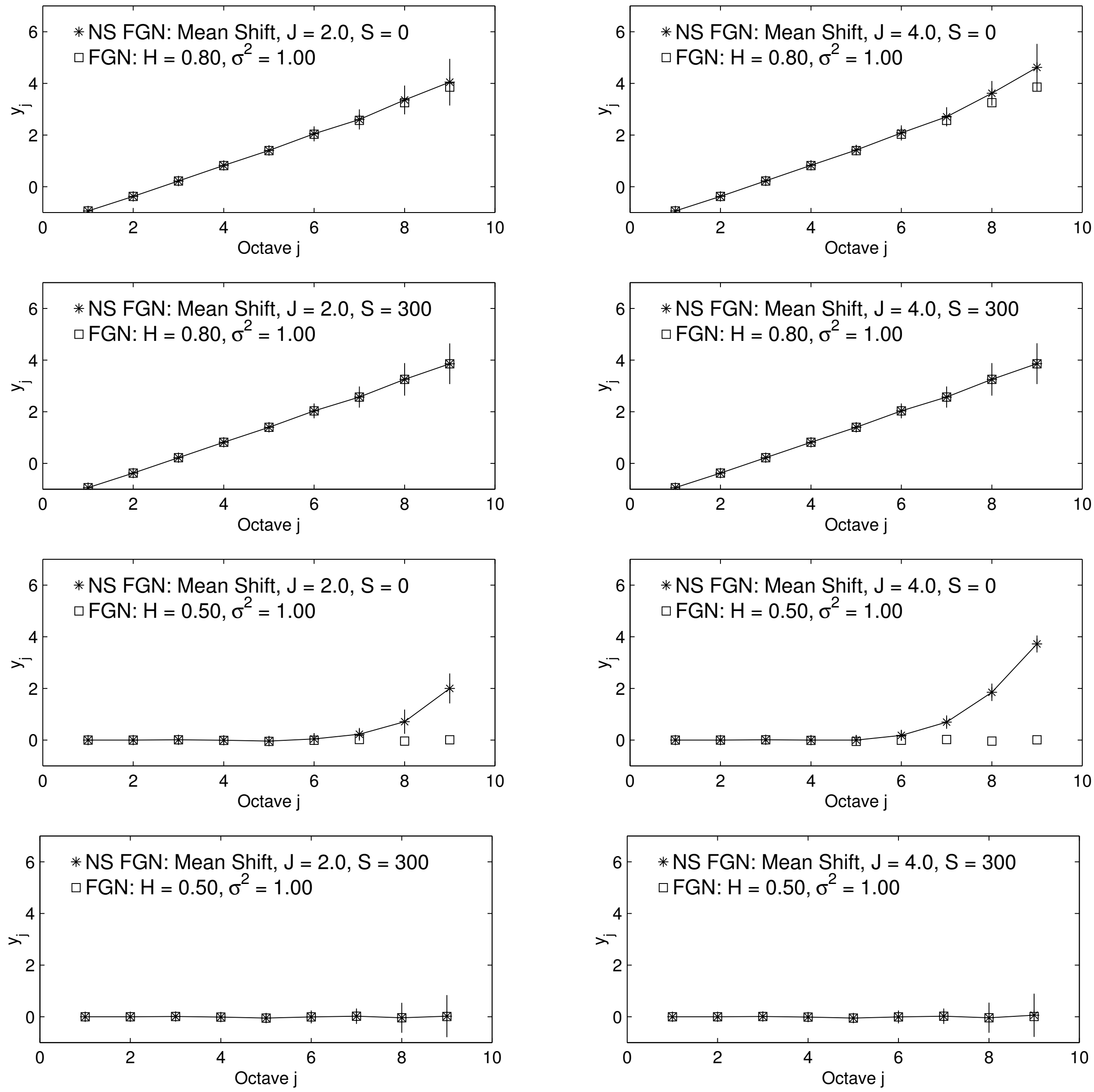


Fig. 3. Logscale Diagrams for a shift in the mean. The (*) show the sample mean of the y_j for the NS FGN, and the squares show the sample mean for the original FGN realizations. The vertical lines show the one standard deviation of the NS FGN results. The smoothness and jump size parameters are shown in the figures.

fined in (8). These are obtained exactly as if the process were stochastic, that is, the wavelet transform is performed, the average squared coefficient at each octave is computed, and the y_j's are calculated by taking the log (after removing the small logarithm generated bias term).

Observe that, just as for LRD processes, the y_j are increasing with j, and are larger for larger J, and smaller S. That is, the y_j for a transition function increase when the level shift in the mean is larger and sharper, and are approximately linear in j. *Figure 4* also shows the Logscale Diagrams for a set of realizations of FGN with $H = 0.5, 0.6, 0.7, 0.8$ and 0.9.

Now the DWT is built from linear filters, and is therefore linear. Furthermore the computation of the μ_j is approximately linear, so that, if we take M signals X_t^m with mean squared detail at scale j of μ_j^m then, as $n_j \to \infty$, the estimates of μ_j follow

$$\mu_j \to \sum_m \mu_j^m, \tag{11}$$

We can consider the NS FGN with a mean level shift to be constructed from two parts, a LRD stochastic process $W(t)$, and a deterministic mean rate $m(t)$. The sum of $X(t) = m(t) + W(t)$, is used to give the coefficients μ_j^X, which we now know can be approximated by $\mu_j^X = \mu_j^m + \mu_j^W$. Since, to obtain y_j^X, we essentially take the log of μ_j, if the ratio of μ_j^m to μ_j^W is large (resp. small), then the response of $m(t)$ (resp. $W(t)$) will dominate the result. In *Figure 4* the sizes of the y_j for both components of $X(t)$ appear in the same Logscale Diagram and can be compared. It is seen that in most cases the coefficients for the FGN are significantly larger than those for $m(t)$ (note the log scale), the exception being that those for the transition function dominate for large j when $S = 0$.

The conclusion is that, except for large j in the case of very abrupt jumps with large magnitude, the FGN dominates the Logscale Diagram, and therefore we obtain accurate estimates of y_j. The accuracy of the estimates of H and c_f clearly follows from that of the y_j. However even when the upper scale y_j's are in-

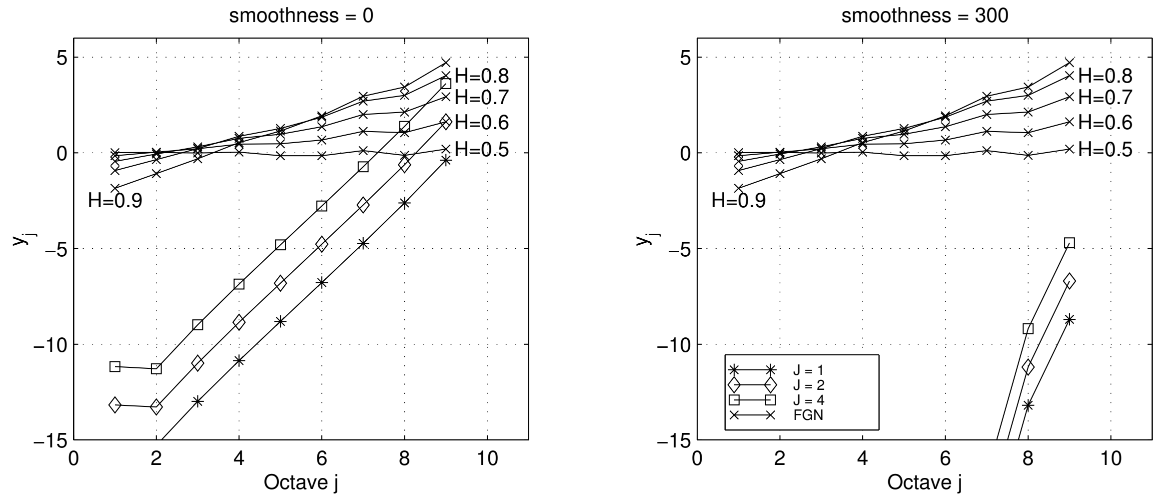


Fig. 4. Logscale Diagram of the transition functions and stationary FGN processes. The sequences are 2^{14} data points long. The $*$'s show the Logscale Diagram in the case with $J = 1$, while the $\diamond$'s show the Logscale Diagram of the mean rate with a jump $J = 1$, and the $\Box$'s show case with $J = 4$. The crosses show the Logscale Diagrams for a realization of FGN with Hurst parameter shown, and $\sigma^2 = 1$ as in the previous examples.

accurate the resultant estimate for (H, c_f) is not strongly biased because, as noted above, the weighted regression underlying the estimator places less weight on the higher scale data, as these have naturally greater variability.

B. Robustness to Variance Changes

We next consider the robust estimation of H of a standard FGN transformed by **multiplication** by a transition function, corresponding to a level increase in variance with constant mean:

$$X(t) = T(t; J, S, n/2) \cdot W(t; H, \underline{c}_f). \tag{12}$$

Now that the variance of X is time varying, c_f is no longer defined, and we do not consider it here.

Table I shows the effect of the level change in variance on the estimates of the Hurst parameter. The results are again the average of 30 realizations. Note that the change in variance introduces only very minor variation in the Hurst parameter estimates – we can conclude that no significant bias is introduced.

We can explain this result by considering the Logscale Diagram shown in *Figure 5*. The plot shows three superimposed Logscale Diagrams, each of which displays averaged results over transformations of the same 30 realizations of an underlying stationary FGN. The lower and upper rows of points correspond to the underlying FGN with standard deviations matched to that of the NS FGN at time zero and at time n (the end of the data). The NS FGN plot is the one lying between these two extremes.

The figure displays the main feature of a change in variance. Each y_j is shifted so that it lies directly between the y_j of the stationary process with the same variance as the initial (smallest) variance of the NS FGN, and the y_j of the stationary process with the final (largest) variance of the NS FGN. The effect is similar for the $H = 0.5$ case (space limitations prevent showing the plots here). This effect is to be expected. The y-intercept of the Logscale Diagram for the stationary FGN is directly related to c_f (see Section III-C) and the variance of each process is proportional to its value of c_f via (4). The NS FGN variance function lies between the constant variances of the two stationary FGNs and therefore the y_j of the NS FGN curve should lie between the y_j of the two extremes. The same conclusion fol-

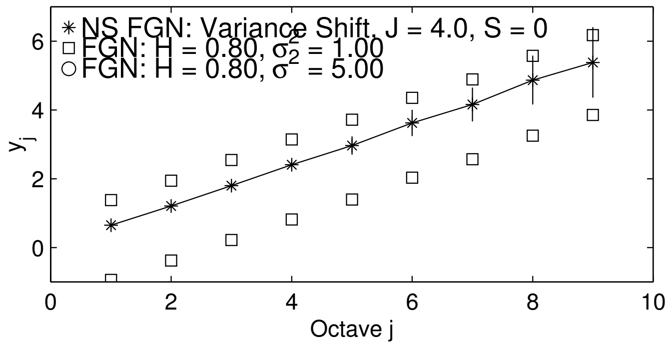


Fig. 5. An example Logscale Diagram for a change in the variance, $H = 0.8$. The (*) show the sample mean of the y_j for the NS FGN, and the ($\Box$) and ($\circ$) show the sample mean of the y_j's for FGN processes with variance matched to that of the NS FGN at time 0 and time n respectively. The vertical lines show the one standard deviation of the NS FGN results.

lows from the observation that the $\mu_j \approx 2^{y_j}$, in both the stationary or non-stationary cases, is a measure of the average energy in the data at scale j, and an increase in variance corresponds to an increase in energy.

A further, key observation is that the size of the shift is almost the same for each j, that is the NS FGN curve appears to be simply a shifted version of the stationary curves. The slope of the NS FGN curve is therefore almost the same as before the transformation, and therefore the estimate of H will be essentially unchanged as observed above.

Again the explanation for this behavior lies in the linearity of the estimates. We consider the extreme case – a jump shift discontinuity. The process can be decomposed into two process – one which is just a stationary FGN, and a second, which is zero for the first half of the data sequence and a stationary FGN for the second half. By linearity the μ_j for the original process will be $\mu_j = \mu_j^{(1)} + \mu_j^{(2)}$, where $\mu_j^{(1)}$ and $\mu_j^{(2)}$ are those for the stationary FGN and the FGN which starts half way through the data sequence. The former is well understood (as it is just that for a stationary FGN) while the later will be those for a stationary FGN of half the length of the original data – the zero terms will not contribute anything, and the edge effects can be assumed to be minimal in this context. Hence the final μ_j are given by $\mu_j^{(1)}$ shifted by an amount such that in the Logscale plot the shift is almost constant.

C. Sensitivity to Other Parameters

The previous sections have all used the same length sequences, and the same wavelet bases for the purposes of comparison, but it is important to study the sensitivity of the results to sequence length and to wavelet basis.

First we varied the sequence length from 2^{14} to 2^{18} data points. The same transformations were performed, though the smoothness parameter was scaled by $S' = Sn'/n$ in order that the transition region takes up the same proportion of the total length of the data. We find that overall the bias remains at a similar level for this increase in sequence length leading us to the conclusion that our results are not highly sensitive to the length of the data.

We next examined the sensitivity to the wavelet basis functions. Previously in this paper we have used a single wavelet basis for the DWT, namely the Daubechies wavelet basis with $N = 5$ vanishing moments. We also examined the behavior both for the Daubechies wavelet basis with 3 vanishing moments (with filters 6 taps long), and with $N = 10$ (filters 20 taps long). The most noticeable effect is that when we have more vanishing moments in the wavelet basis the coefficients y_j for the transition function are lower. This is beneficial because it improves the performance of the estimator in the presence of non-stationarity, but it comes at the cost of a reduction in the amount of data n_j available at each scale, resulting in additional variance in the estimation, and extra computation due to the longer filters used.

This effect can be predicted, by approximating the level shift by a polynomial, from the robustness with respect to N discussed in Section III-C. Unfortunately the improvement gained by increasing the number of vanishing moments diminishes, and values beyond $N = 5$ seem to give little improvement (see [1] for more details of such effects).

V. REAL DATA EXAMPLE

The preceding results are based upon FGN, and the question arises, "How robust is the AV estimator when applied to real data?" This section demonstrates that the robustness of the AV estimator extends to real data by testing an Ethernet data set with an obvious level shift in the mean. The data set is byte counts in 10ms intervals from the 'pOct' Bellcore trace [12].

The jump can be clearly seen in *Figure 6*, which shows the data over intervals of 10 seconds for easier visualization. The data shows a distinct level shift in mean at about 1050 seconds – the mean estimates to the left and right of the transition region are 2.40 and 4.14 respectively (the exact intervals of measurements are shown in the figure). The variance also increases in the same region, going from 7.3 to 9.5 as measured over the same intervals. The mean therefore jumps by $J = 1.73$ and the variance by a multiplicative factor of $J = 1.3$, a mixture of a mean and variance shift.

It has already been shown in [4] that Hurst parameter estimates using the AV estimator to the right and the left of the level shift agree both with each other, and with the estimate made over the whole data set. The work in [4] does not, however, fully explain how to reconcile these observations, with the non-stationarity in the mean. It is now possible to recognize that H may remain constant regardless of the non-stationarity in mean and variance, and explain the AV estimator's robustness in the presence of this non-stationarity.

In addition, the model allows us to meaningfully transform the data to remove the non-stationarity, and measure the Hurst parameter under stationary conditions for further comparison. The results are shown in *Table II* where it is clear that no noticeable error has been introduced by the non-stationarity.

	$\hat{H}$	95% CI
original data	0.779	± 0.0072
corrected data	0.779	± 0.0072
data interval (150,950)	0.769	± 0.0108
data interval (1170,1750)	0.779	± 0.0128

TABLE II

THE HURST PARAMETER ESTIMATES FOR THE ORIGINAL DATA, THE DATA ONCE THE NON-STATIONARITY IN THE MEAN IS REMOVED, AND THE DATA ON THE TWO INDICATED TIME INTERVALS.

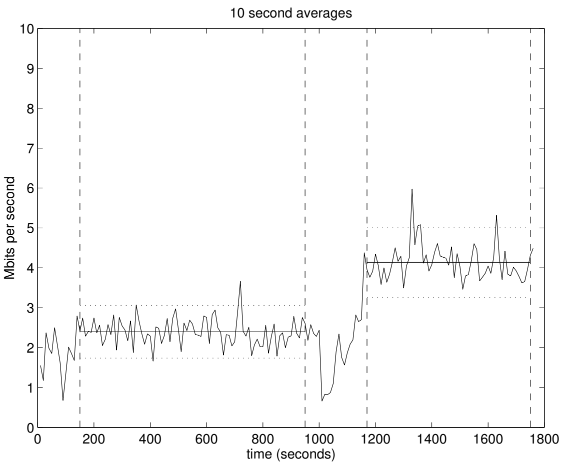
We removed the mean level shift by fitting a transition function to the data by estimating the mean over the first segment shown in *Figure 6*(a), and then the parameters of the transition function: the transition point, jump size and smoothness. The latter was done using Matlab's non-linear minimization function `fmins` which performs the Nelder-Mead simplex search described in [14], [11]. The key point here is that we wish to show robustness, not absolute accuracy, and hence a simple method for modeling the level shift is quite sufficient. Once the transition function is fitted it can be subtracted to obtain a process with an approximately stationary mean. We do not attempt to fix the variance as we have shown that this will have negligible effect on the results.

Finally *Figure 6* shows the Logscale Diagrams for the original data and the corrected data. We can see that the two Logscale Diagrams are almost exactly the same except for a small discrepancy at higher scales, which has little effect due to the weighted regression. The figure also shows the log-scale response to the transition function used to model the change in mean. The response is significantly lower than that of the data sequence and hence the transition has little effect on the estimates.

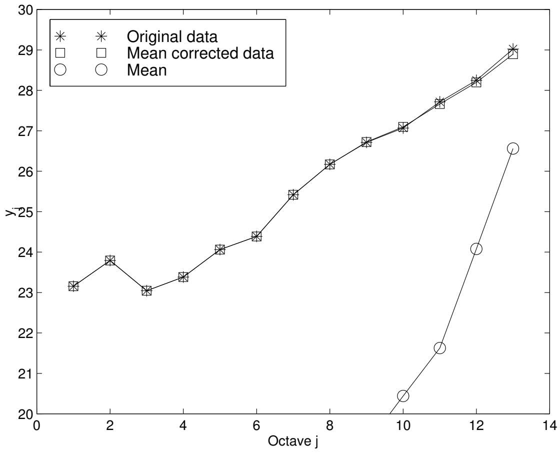
VI. MITIGATION

The effects of a mean level shift appear at the higher scales. They could therefore be almost completely eliminated by choosing an upper scale for the regression analysis (which underlies the estimation of (H, c_f) from the Logscale Diagram). By taking $j_2 = 6$ for example the bias is almost completely removed. Even in the worst case of $S = 0$ and $J = 4.0$, the average results for $\hat{H}$ after transformation are 0.509 ± 0.0035 compared to the average over the 30 stationary FGN's which yield 0.497 ± 0.0038, a bias of only 0.01.

If the properties of the jump were known, we could predict the upper scale to be used in the regression. In practice we might not know the exact nature of the jump, but the argument could be used in an approximate sense in the selection of a mitigating upper cutoff. Alternatively a procedure could be used to pick the scales to be used in the regression using heuristic arguments.



(a) Dashed vertical lines indicate estimation ranges for means and variances. Means are shown with 95% confidence intervals.



(b) The Logscale Diagrams of the original data ($*$), the corrected data ($\square$), and the fitted transition function ($\circ$).

Fig. 6. The 'pOct' Bellcore trace.

VII. Conclusion

Our main finding is that the AV estimator is very robust, allowing accurate estimates to be made of the Hurst parameter, despite non-stationarities in the mean and variance, specifically level shifts. We illustrate this robustness, explain its origin, and indicate when the residual bias due to the non-stationarities may be appreciable and how it can be minimized.

As an essential precursor to the robustness study, a broad class of *non-stationary LRD* processes were defined. They allow a well defined separation of the mean and variance, which are allowed to be time varying, from the time constant parameters, including the LRD parameters, which remain well defined despite the non-stationarity.

Non-stationarity and LRD can be confused, and therefore it is particularly important to have robust estimators that do not need the full details of the traffic before valid estimation can take place. Although the focus here was on level shifts, this represents in some sense a worst case, and we expect that the robustness found will hold for a very wide range of non-stationarities. This has already been shown in the case of the mean (additive trend) in [4]. Thus, the AV estimator allows the Hurst parameter to be measured, and in particular the question of the presence or absence of LRD decided, without any need to tackle in advance or simultaneously the difficult stationary issue. This is an enormous practical advantage. Once it is known that LRD is or is not present, then analysis of any non-stationarities can be tackled in a far more informed way.

An Ethernet data set was studied where a clear level shift non-stationarities was found. The non-stationary LRD model was used to explain the time variation of the data, and the robustness of the wavelet based estimator was used to measure its Hurst parameter, confirming previous work in [4] which had observed the robustness for this same data set without a full explanation.

Acknowledgments

The support of Ericsson Australia is gratefully acknowledged.

References

[1] P. Abry. *Ondelettes et Turbulences - Multirésolutions, Algorithmes de Décompositions, Invariance d'Echelle et Signaux de Pression.* Diderot, Editeur des Sciences et des Arts, Paris, 1997.
[2] P. Abry, P. Gonçalvès, and P. Flandrin. *Wavelets and Statistics*, volume 105 of *Lecture Notes in Statistics*, chapter Wavelets, Spectrum estimation, $1/f$ processes., pages 15–30. Springer-Verlag, New York, 1995.
[3] P. Abry, M. S. Taqqu, P. Flandrin, and D.Veitch. *Wavelets for the analysis, estimation, and synthesis of scaling data.* Wiley, 1998.
[4] P. Abry and D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Trans. on Info. Theory*, 44(1):2–15, 1998.
[5] P. Abry, D. Veitch, and P. Flandrin. Long-range dependence: revisiting aggregation with wavelets. *J. of Time Series Analysis*, 19:253–266, 1998.
[6] J. Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, 1994.
[7] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia (PA), 1992.
[8] A. Feldmann, A.C. Gilbert, and Walter Willinger. Data networks as cascades: Investigating the multifractal nature of internet wan traffic. In *ACM/Sigcomm'98, September 2–4, Vancouver, Canada*, 1998. preprint.
[9] Judith L. Jerkins and Jonathan L. Wang. A measurement analysis of ATM cell-level aggregate traffic. In *IEEE GLOBECOM'97*, 1997.
[10] J.L.Jerkins, J. Monroe, M.F.Pucci, and J.L.Wang. Carrying internet traffic over frame relay links: Frame and call level traffic analysis. preprint.
[11] J.E. Dennis Jr and D.J. Woods. *New Computing Environments: Microcomputers in Large-Scale Computing*, pages 116–122. SIAM, 1987.
[12] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, Feb 1994.
[13] S. Molnár, A. Vidács, and A.A. Nilson. Bottlenecks on the way towards fractal characterization of network traffic: Estimation and interpretation of the Hurst parameter. In *PMCCN'97*, Tsukuba, Japan, Nov 1997.
[14] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313.
[15] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modelling. In *Proceedings of SIGCOMM '94*, 1994.
[16] Matthew Roughan, Darryl Veitch, and Patrice Abry. On-line estimation of parameters of long-range dependence. In *IEEE GLOBECOM'98*, pages 3716–3721, Sydney, Australia, November 1998.
[17] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman and Hall, 1994.
[18] V. Teverovsky and M. S. Taqqu. Testing for long-range dependence in the presence of shifting means or a slowly declining trend using a variance-type estimator. *Journal of Time Series Analysis, 18:279–304*, 1997.
[19] Darryl Veitch and Patrice Abry. A statistical test for the time constancy of scaling exponents. submitted to IEEE Trans. on Info. Theory, 1998
[20] Darryl Veitch and Patrice Abry. Estimation conjointe en ondelettes des paramètres du phénomène de dépendance longue. In *Proc. 16ième Colloque GRETSI, pp.1451–1454, Grenoble, France*, 1997.
[21] Darryl Veitch and Patrice Abry. A wavelet based joint estimator of the parameters of long-range dependence. submitted IEEE Tran. on Info. Theory, 1998.