A Statistical Test for the Time Constancy of Scaling Exponents

Darryl Veitch(1) and Patrice Abry(2).

(1) EMULab, Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia (d.veitch@ee.mu.oz.au) [1].
(2) CNRS UMR 5672, ENS Lyon, 46 allée d'Italie 69 364 LYON, France (Patrice.Abry@ens-lyon.fr) [2].

Abstract— **A statistical test is described for determining if scaling exponents vary over time. It is applicable to diverse scaling phenomena including long range dependence and exactly self-similar processes in a uniform framework, without the need for prior knowledge of the type in question. It is based on the special properties of wavelet-based estimates of the scaling exponent, strongly motivating an idealised inference problem: the equality or otherwise of means of independent Gaussian variables with known variances. A uniformly most powerful invariant test exists for this problem and is described. A separate UMPI test is also given for when the scaling exponent undergoes a level change. The power functions of both tests are given explicitly and compared. Using simulation the effect in practice of deviations from the idealisations made of the statistical properties of the wavelet detail coefficients are analysed and found to be small. The tests inherit the significant robustness and computational advantages of the underlying wavelet-based estimator. A detailed methodology is given describing its use in practical situations. The use and benefits of the test are illustrated on the Bellcore Ethernet data sets.**

Keywords— **Self-similarity, long-range dependence, scaling exponent, wavelets, hypothesis testing, stationarity, telecommunications networks.**

I. Introduction

Stochastic processes exhibiting *scaling behaviour*, such as exact self similarity or long-range dependence, have been recognised in many fields as relevant models of time series data with scale invariance features. A prominent new example of the latter is telecommunications network traffic, the extraordinary scaling properties of which have stimulated much new work in the area [27], [17].

An interest in modelling data necessarily leads to issues of measurement and statistical estimation. One of the difficulties in the analysis of data with scaling features is the poor and non-standard performance of many statistical tools, which rely typically on stationarity of the model or on a short range correlation structure, or both. It has been recently shown however how estimation approaches based on wavelet analysis [2], [3], [24], [4] can overcome many of these disadvantages. Notably, efficient semi-parametric estimates of the *scaling exponent*, the key parameter describing scaling, are possible with negligible bias, and $O(n)$ computational complexity.

There is another key difficulty however which, although raised from time to time [7], has as yet received little attention and no satisfactory solution. It is the fact that the very high variability inherent in scaling processes is very easily confused with non-stationarity. This difficulty affects both the fundamental issue of the choice of model class, and the reliable estimation of model parameters in the presence of polluting non-stationarities. It is entirely possible for example that, by using an inappropriate statistical tool to detect self-similarity and measure its *Hurst* parameter H, results will be obtained which seem to indicate scaling behaviour, when in fact the data is not scaling but is non-stationary (in a non-scaling sense). Conversely, the large excursions in sample paths of stationary scaling processes can be erroneously taken as evidence of non-stationarity. This could lead for example to the attempted removal of deterministic trends which do not in fact exist.

The main contribution of this paper is the description of a simple yet optimal (in an idealised sense to be detailed) statistical test for the central problem of determining the constancy or otherwise in time of the scaling exponent. It is based on a wavelet domain estimator described in [2], [24], [4] which enables many of the statistical difficulties due to scaling and non-stationarity to be avoided in a natural way. Essentially the test consists of splitting the data into m non-overlapping blocks, and separately estimating the exponent over each. Under well motivated idealisations the wavelet based estimates can be taken as uncorrelated Gaussian variables with unknown means but known variances. It turns out that there exists a test which is *Uniformly Most Powerful Invariant* for this problem. A second test, also UMPI, is provided for the special case when the scaling exponent undergoes a level shift. A second aim is to provide a methodology to allow the test to be meaningfully applied in practical situations. To this end a detailed discussion is given on the vital question of the choice of m, and a global analysis procedure is provided[3].

II. The Wavelet Approach to Scaling Analysis

A. Scaling phenomena

Under the term *scaling* we gather several different phenomena. We first consider *Long-Range Dependence* (LRD), a long-memory property of second-order stationary stochastic processes. Its simplest definition is given by the power-law divergence at the origin of the spectrum:

[1]This work was sponsored by Ericsson Australia.
[2]Partially supported by *CNRS programme télécommunications TL99035* and *French MENRT ACI Jeune Chercheur, 2329*

[3]Matlab code for the estimation and test available upon request.

$$f_x(\nu) \sim c_f |\nu|^{-\alpha}, \quad |\nu| \to 0. \tag{1}$$

This asymptotic definition involves two parameters, the dimensionless *scaling exponent* α, and the 'power parameter' c_f which has the dimensions of variance (see [24] for a detailed discussion of the role of c_f).

An important class of scaling processes which are non-stationary are the exactly self similar processes, characterised by the famous scaling exponent H, the *Hurst* parameter. A process $x = \{x(t),\ t \in \mathcal{R}\}$ is *self-similar with parameter* $H > 0$ (H-ss) if $x(0) = 0$ and $\{x(at),\ t \in \mathcal{R}\}$ and $\{a^H x(t),\ t \in \mathcal{R}\}$ have the same finite-dimensional distributions. For $H < 1$, such processes may have *stationary increments*, and if so, for $1/2 < H < 1$ the increment processes are LRD processes, with $\alpha = 2H - 1$. We also consider self-similar processes whose increments of order $p \geq 1$ (i.e., increments of increments, p-times) are stationary [4].

In what follows we let α denote generically the second order scaling exponent of the process $x(t)$ under study, regardless of the exact kind of scaling it describes, and denote by *scaling processes* either a LRD process or a self-similar one with stationary increments of order p. There are other kinds of scaling processes, such as *fractal*, $1/f$, and *multifractal* (see [21], [10], [17]), for which the test procedure developed here could be straightforwardly applied, at least for moments of second order, but for simplicity we do not do so here.

B. Wavelets: essential properties

The coefficients $d_x(j,k)$ of the discrete wavelet transform (DWT) [8], [15] result from the comparison, by means of inner products, of the process to be analysed x and a family of functions $\{\psi_{j,k}\}$, called the wavelet basis: $d_x(j,k) = \langle x, \psi_{j,k} \rangle$. The wavelet basis $\{\psi_{j,k}\}$ consists of shifted and dilated templates of a single reference pattern ψ_0, usually called the *mother-wavelet*. The mother-wavelet has a time support and frequency support which are strongly concentrated: it therefore acts as an elementary atom of information. From it the time-shift operator and the dilation (or change of scale) operator together generate the full, two parameter set of basis functions:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi_0 \left(\frac{t - 2^j k}{2^j} \right),$$

centred on a sparse set of points in the time-scale plane known as the *dyadic grid*, that is the points $\{(\text{scale} = 2^j, t = 2^j k),\ j,k \in \mathcal{Z}\}$.

The mother-wavelet is moreover characterised by an integer N, called the number of vanishing moments, defined as:

$$k = 0, 1, 2 \ldots, N-1, \quad \int t^k \psi_0(t) dt \equiv 0 .$$

Consider first self-similar processes, which are scaling at **all** scales. The wavelet coefficients satisfy the following two key properties.

- **P1:** Provided $N \geq (\alpha - 1)/2$, the sequences $\{d_x(j,k), k \in \mathcal{Z}\}$ are stationary, and their variances reproduce precisely the power law underlying the scale invariance of the process:

$$\mathbb{E} d_x(j,k)^2 = C 2^{j\alpha},$$

where C is a constant which can be calculated [4].

- **P2:** Any two wavelet coefficients exhibit a correlation that is asymptotically controlled by N such that the larger N, the weaker the correlation:

$$\mathbb{E} d_x(j,k) d_x(j',k') \sim C_1 |2^j k - 2^{j'} k'|^{\alpha - 1 - 2N}, \tag{2}$$

$|2^j k - 2^{j'} k'| \to \infty$, where C_1 is a constant. Thus, the non-stationarity of x is reduced to short-range dependent stationarity of each $d_x(j,\cdot)$ in the wavelet domain provided $N \geq \alpha/2$. These correlations are not only short-range, but are weak. It is therefore useful to consider the following idealised property:

– **ID1:** The $d_x(j,k)$ are strictly uncorrelated.

Remarks. Properties **P1** and **P2** exactly hold for self-similar processes with stationary increments of some order $p \geq 1$ [12], [9], [22], [2], [4]. In the LRD case, **P1** does not hold identically, but holds approximately over a range of large enough scales: $j \in [j_1, \infty]$. The j dependence thereby introduced in C by the scales outside of this range is weak and can be ignored, as shown numerically in [2], [24], [5]. For LRD processes (2) does not hold strictly, however the key result that the $d_x(j,\cdot)$ are short range dependent provided $N \geq \alpha/2$ remains valid, and is heuristically summarized as **P2**, see [5] for a complete discussion. Finally, it is worth noting that (2) does not hold, in any case, when $2^j k = 2^{j'} k'$, or more generally when the two wavelet coefficients are within their so-called cone of influence, where the time support of the wavelets $\psi_{j,k}$ and $\psi_{j',k'}$ overlap significantly. It is known however that for self similar processes one then has an exponential decrease in correlation along lines of constant time [12]. In this respect, **ID1**, which overlooks such correlations, may appear unrealistic. In the sequel, it will be shown from numerical simulations that the constancy tests inspired by the idealisation have a statistical performance close to the theoretical predictions assuming **ID1**.

C. Estimation of the scaling exponent

Definition of the estimator From **P1**, one can think of estimating the scaling exponent from a linear fit in a $\log_2 \mathbb{E} d_x(j,k)^2$ *vs* $\log_2(2^j) = j$ plot or *Logscale Diagram*. The wavelet based estimator reads [2], [3], [24]:

$$\begin{array}{rcl} y_j & = & \log_2 \left(1/n_j \sum_k d_x(j,k)^2 \right) - g_j \\ \hat{\alpha} & = & \sum_j w_j y_j \end{array} \tag{3}$$

where the sum is over $j \in [j_1, j_2]$, the range of octaves over which the scaling phenomenon is observed and the linear regression performed. The g_j are deterministic quantities that account for the fact $\log_2 \mathbb{E} d_x(j,k)^2 \neq \mathbb{E} \log_2 d_x(j,k)^2$, see [24], [4] for details and complete expressions. The weights w_j follow the standard formulae for weighted linear regression using the variances of the y_j.

The choice of the cutoff scales j_1, j_2, is an important issue, discussed in detail in [4], which is beyond the scope

of the paper. In the case of LRD, where only j_1 need be chosen, [5] provides a rigorous way to define it for a given model. We use this method below when estimating α from synthetic data. In section 5, a robust method based on a goodness of fit statistic was used to select j_1 from data [25].

Statistical performance of $\hat{\alpha}$ The following properties have been established in [2], [24], [4]. We assume that **P1** and **ID1** hold. The effect of deviations from these are small as documented in the references above. We also assume that $\{j_1, j_2\}$ have been well chosen.

1. **Procedure:** the estimation and its statistical performance are the same regardless of the precise nature of the scaling existing in x.
2. **Bias:** the estimator is unbiased.
3. **Variance:** the Cramer-Raò lower-bound of the corresponding idealised estimation problem is attained [24], [26], with a *known* variance.
4. **Robustness:** i) It is semi-parametric, so does not require a priori knowledge of the model, ii) Via N, it is insensitive to superimposed deterministic trends [3], [4], or even to smooth time evolution of the variance of x itself [20].
5. **Computational load:** complexity of order $O(n)$, and can be implemented in real-time [19].

We now require that two additional properties be addressed: the Gaussianity of $\hat{\alpha}$, and the independence of estimates obtained from adjacent non overlapping blocks.

D. Gaussianity of the estimator

If we assume (**ID1**), we have exact decorrelation of the y_j that enter in the definition of $\hat{\alpha}$ in equation (3). For all processes x such that the y_j have finite variance, and this includes many with infinite variance [4], [6], one can apply a generalised central limit argument (see e.g., [11], theorem 3, section VII) to suggest that $\hat{\alpha}$ is asymptotically normally distributed. To test the effect in practice of residual correlation between wavelet coefficients on the asymptotic gaussianity of $\hat{\alpha}$, numerical simulations were performed.

We studied fractional Gaussian noise (fGn), whose LRD is deeply related to self-similarity, and a particular fractional AutoRegressive Integrated Moving Average (fARIMA) process, whose LRD has no relation to self-similarity. More precisely, a so-called fARIMA$(1, d, 1)$ model with $c_f = 1$, $\psi = 0.3$, $\theta = 0.7$ was chosen for its strong short range dependence [5], requiring $j_1 = 6$, in contrast to the $j_1 = 3$ for fGn. For simplicity these values are used in all cases, although they are in fact functions of n and α, and may vary slightly (see [5] for a complete discussion). In what follows fARIMA denotes the specific process with the parameters given above. We synthesized, by the so-called *spectral synthesis* method (see [16] for one implementation) K realisations for both types of processes, for different values of α and of various lengths n, and recorded an estimate of α for each (for uniformity we use α rather than the customary choice of H: $\alpha = 2H - 1$). For each $\{\alpha, n\}$ pair the empirical probability distribution function of $\hat{\alpha}$ given by the K independent estimates was compared with that of a Gaussian random variable in *quantile-quantile* plots. For the results shown in figure 1 we have $K = 10000$, $\{\alpha, n\} = \{0.5, 2^{17}\}, \{0.5, 2^9\}, \{-0.5, 2^{17}\}, \{-0.5, 2^9\}$ for the fGn, and $\{\alpha, n\} = \{0.5, 2^{17}\}, \{0.5, 2^{12}\}, \{-0.5, 2^{17}\}, \{-0.5, 2^{12}\}$ for the fARIMA, and (here and later) *Daubechies3* ($N = 3$) wavelets were used. The agreement is not only excellent for long data sets, being close in the range corresponding to $\pm 3\sigma$ for fGn with $n = 2^{17}$, but also for much shorter data sets: between $\pm 2\sigma$ for $n = 2^9$ for fGn. Similar conclusions hold in the fARIMA cases where, since j_1 is 3 octaves larger, the smallest n used is also: 2^{12} rather than 2^9, to maintain the same minimum range of scales for estimation. This numerical study reveals that asymptotic Gaussianity for $\hat{\alpha}$ remains valid under realistic departures from **ID1**, and also that it can hold even when the scaling range is very narrow.

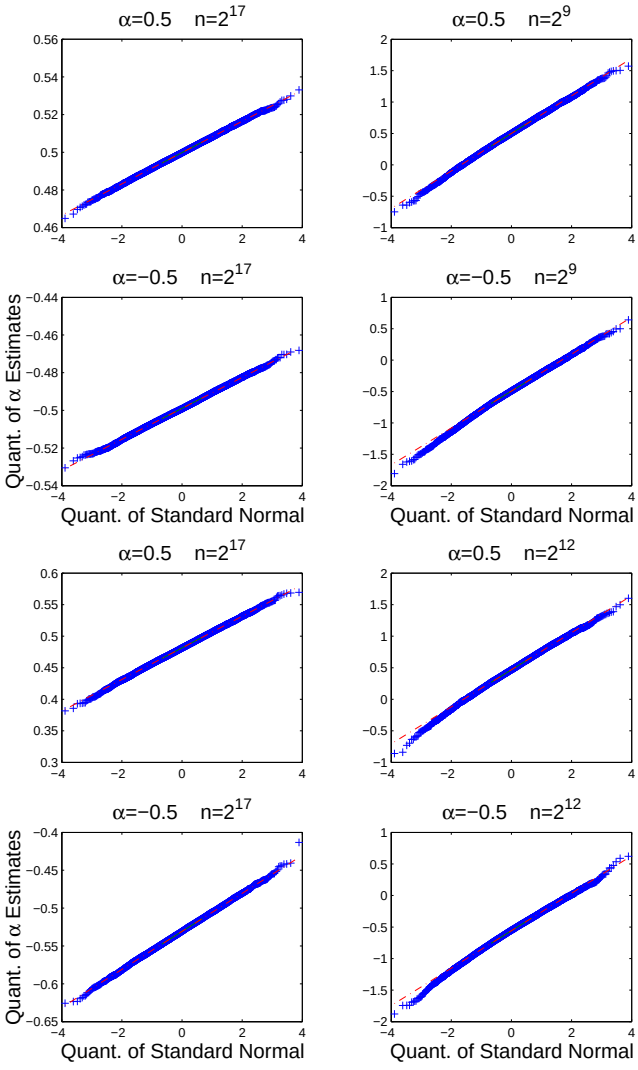


Fig. 1. **Distributional comparison of $\hat{\alpha}$ against Gaussian.** Quantile-quantile plots of $\hat{\alpha}$ against a standard Gaussian variable, obtained from 10000 realisations of fGn (4 top plots), and fARIMA (4 bottom plots) for various values of $\{\alpha, n\}$.

E. Correlation between adjacent blocks

If estimates on adjacent blocks are computed using a time domain estimator, they will be strongly dependent because of the non-stationarity or long-range dependence of the original process. In contrast to this, exact decorrelation (**ID1**) would imply that wavelet-based estimates were mutually uncorrelated.

Using the spectral synthesis technique, K realisations of fGn and fARIMA (as previously defined) were synthesized for each of various lengths n, with a common value of α corresponding to strong LRD: $\alpha = 0.6$ ($H = 0.8$). Each series was split in half and an estimation of α performed independently on each. Let $\{\hat{\alpha}_1^n(k), \hat{\alpha}_2^n(k)\}$, $k = 1, \ldots, K$ denote these series of estimates. For each n the Fischer z-statistic [18] is computed, and the corresponding test applied to examine the null hypothesis of complete decorrelation. In the numerical simulations performed: $K = 2000$, $N = 3$, for both processes, and $n = \{2^9, 2^{10}, \cdots, 2^{18}\}$, $j_1 = 3$ for the fGn, $n = \{2^{12}, 2^{13}, \cdots, 2^{18}\}$, $j_1 = 6$ for the fARIMA. Again, in the fARIMA case smaller n cannot be explored because of the large ARMA components. Figure 2 shows, as a function of the length of the original series n, that the z values all fall well within the 95% confidence interval (dashed lines) corresponding to zero correlation: we accept the hypothesis of decorrelation. Experiments for other values of α, not reproduced here, yielded identical conclusions. Although the details of residual correlations are model dependent, they tend to be stronger for larger α. The example given here shows them to be small even then.

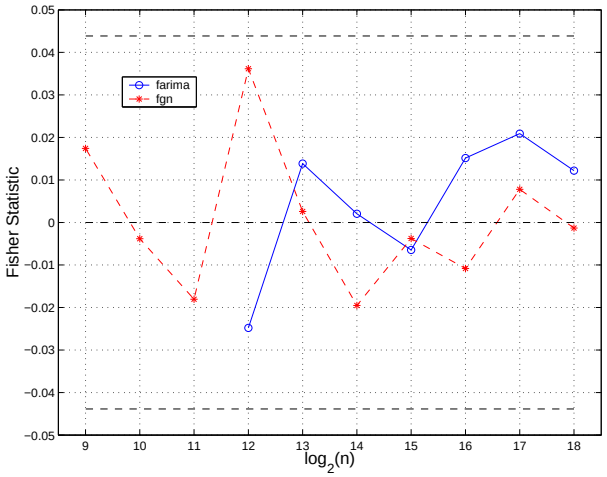


Fig. 2. **Testing for correlation between adjacent α estimates.** Fisher's z-statistics (fGn: $*$, fARIMA: $\circ$) computed from α estimates from pairs of adjacent blocks, as a function of n (block length equals $n/2$). For each n the z-values fall within the 95% confidence interval (dashed lines) indicating that there is no reason to reject the exact decorrelation idealisation.

III. Two Inference Problems, Tests and Properties

In this section we discuss optimal tests pertaining to two idealised parametric inference problems inspired by the attractive properties, discussed above, of wavelet-based α estimates made over m adjacent blocks. In each problem there are m independent Gaussian variables $\mathbf{X} = \{X_i\}$ with *unknown* means $\boldsymbol{\mu} = \{\mu_i\}$ taking real values, and *known* common variance $\{\sigma_i^2 = \sigma^2\}$, $i = 1, 2, \ldots, m$. By μ we denote the mean of the components of $\boldsymbol{\mu}$, and similarly for the vectors $\mathbf{X}$ and $\mathbf{Y}$ below. Our main focus is on the 'general' problem where no structure is assumed between the unknown means. Thus the null hypothesis H_0 is that the means share a common (unknown) value μ, against the alternative hypothesis H_1 that they are not all the same. The second problem has the same H_0 but considers that the unknown means must obey the structure of a *level shift*, so that H_1 is more restrictive. Note that in each case both the null and alternative hypotheses are *composite*, that is that they are sets in parameter space with more than a single element, so that they are a priori difficult problems.

The tests given below can be shown to be *Uniformly Most Powerful Invariant* (UMPI) [13]. Essentially this means they have optimal power, that is their probability of accepting H_1, if true, is greater than that of any (reasonable) alternative test, regardless both of *which* specific H_1 is true, and of the significance level δ. The reader is referred to [23] and specifically [13] for more details on invariant tests in general and their application to the present inference problems.

A. An UMPI test for the constancy of means

The UMPI test can be defined as rejection of H_0 in the critical region $V > C$, where

$$\begin{aligned} H_0: \quad \theta &= \sum (\mu_i - \mu)^2/\sigma^2 = 0 \ \text{(means equal)} \\ H_1: \quad \theta &> 0 \qquad\qquad\qquad\quad \text{(means unequal)} \\ \text{Critical Region: } V &= \sum (X_i - X)^2/\sigma^2 > C \ \text{(one-sided)}, \end{aligned} \tag{4}$$

where θ measures the distance of $\boldsymbol{\mu}$ from H_0. Due to the cylindrical symmetry [13], a m dimensional problem has been reduced to one dimension, both the statistic V and the (unknown) parameter θ are scalars, and H_0 is now *simple*, being just the point $\theta = 0$.

The distribution of V under H_0 is just that of a Chi-squared variable with $m - 1$ degrees of freedom [23], and is thus independent of the common mean μ. The constant C is therefore determined from the significance level δ via $\int_C^\infty f_{m-1}(v)\, dv = \delta$, where $f_{m-1}(v)$ is the density function of the Chi-squared variable.

In the case of different variances σ_i^2, the test readily generalises ([13], p.377) to $V > C$ where $V = \sum \frac{1}{\sigma_i^2}\Big(X_i - \frac{\sum X_j/\sigma_j^2}{\sum 1/\sigma_j^2}\Big)^2$, and θ becomes $\theta = \sum \frac{1}{\sigma_i^2}\Big(\mu_i - \frac{\sum \mu_j/\sigma_j^2}{\sum 1/\sigma_j^2}\Big)^2$. Under H_0 we still have $\theta = 0$ with C determined exactly as before. For clarity we will continue to concentrate on the simpler case.

The power of the test Under H_1 the distribution of V becomes that of a *non-central* Chi-squared variable with $m - 1$ degrees of freedom and a non-centrality parameter equal to θ. The power is therefore given by $\int_C^\infty f_{m-1,\theta}(v)\, dv$, where $f_{m-1,\theta}(v)$ is the non-central Chi-squared density. This integral can be readily evaluated in

practice as a sum of central Chi-squared distributions ([1], equation 26.4.25, p.942.). Exactly the same facts hold in the case of general σ_i^2, provided the generalised definition of θ above is used. Although the above test is optimal with respect to the power of competing tests, actual values of power depend on the closeness of the unknown parameters to H_0, and can be very poor, in fact arbitrarily close to δ.

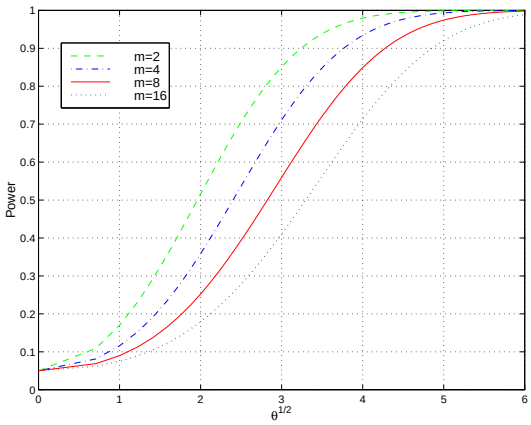


Fig. 3. **Power functions.** Power functions for different m values as a function of $\sqrt{\theta}$, corresponding to the distance of $\boldsymbol{\mu}$ from H_0.

Figure 3 gives power functions for several values of m as functions of $\sqrt{\theta}$. Note however that for each m the θ parameter is really a separate quantity θ_m, identified in the plot for ease of display, corresponding to separate inference problems. Thus, in figure 3, it does not necessarily follow that 'increasing m results in a lower power'. Comparisons across m can only be made in a context where further details/assumptions of the experimental situation allow the specification of the dependence of θ_m on m: $\theta_m = \theta(m, \sigma_i(m), \mu_i(m))$, including how the different θ_m can be meaningfully compared. As power is not defined under H_0 there is no question of how it varies with m, but only of chosing $C(\delta) = C_m(\delta)$ to fix the significance level. Since $\theta = \theta_m$ is zero under H_0, independently of m, it follows that $C_m(\delta)$ is a function of m **only**, and can always be found.

B. An UMPI test for the two-level problem

An important special case of non-constant scaling is that of a 'level change'. If there are physical reasons to expect behaviour of this type in data, or if there is compelling empirical evidence, then it may be desirable to exploit the additional structure and test directly for it. Let the first m_1 variables, denoted by $\{X_i\}$, $i = 1, \ldots, m_1$, each have mean μ_1, and the remaining $m_2 = m - m_1$ variables $\{Y_i\}$, $i = 1, \ldots, m_2$ have common mean μ_2. The UMPI two-level test can be defined as

$$\begin{array}{rrcll} H_0: & \theta_2 & = & (\mu_1 - \mu_2)/\sigma = 0 & \text{(means equal)} \\ H_1: & \theta_2 & > & 0 & \text{(means unequal)} \\ \text{Critical Region:} & V_2 & = & |X - Y|/\sigma > C & \text{(one-sided)} \end{array}$$

where C is chosen from the significance level δ using the fact that V_2 is normally distributed with mean $\theta_2 = (\mu_1 - \mu_2)/\sigma$ and variance $\mathrm{Var}[V_2] = m/(m_1 m_2) = (\beta(1-\beta))^{-1}$, where $\beta = m_1/m$. Again the null hypothesis, originally composite, has become simple through the symmetry.

Note that it is β, the relative **number** of points in the first group, but not their order, that determines the distribution of V_2, so the same analysis applies to situations other than level shift. When applying the test however *which* are in which group must be specified.

Because of the form of $\mathrm{Var}[V_2](\beta)$, at fixed m the power functions are uniformly (in θ_2) monotonically increasing in $m_1 = 1, 2 \cdots \lfloor m/2 \rfloor$. Power between the tests can be compared when $\boldsymbol{\mu}$ obeys a two-level scenario, where one can show that the two-level test has the higher power, the more so for higher m. In such a comparison, note that $\theta_2 = \sqrt{\mathrm{Var}[V_2]\,\theta/m}$. The exception is the trivial case of $m = 2$ where the tests are identical. Finally, note that this test can also be generalised to allow for different σ_i, which for $m = 2$ would again be identical to the general test (with different σ_i). This $m = 2$ different-σ_i test is used in section 4.3 to give an upper bound on the power of the more practical constant σ tests.

Tests in simple hypothesis cases It may be required to perform tests against specific values of $\boldsymbol{\mu}$. Optimal tests for these simpler situations are also available.

IV. A Wavelet Test for the Constancy of Scaling

A. Definition of the test

Let x denote the series to be analyzed of length n. Compute the estimates $\{\hat{\alpha}_1, \ldots, \hat{\alpha}_i, \ldots, \hat{\alpha}_m\}$, according to the definitions given in section 2, for each of m adjacent blocks with, possibly, unequal lengths n_i, using a common scaling range $[j_1, j_2]$. From section 2, we know that the $\{\hat{\alpha}_1, \ldots, \hat{\alpha}_m\}$ can be considered as uncorrelated Gaussian variables, with unknown means, but known, possibly different variances, that is

$$\begin{aligned} \hat{\alpha}_i &\sim N(\alpha_i, \sigma_i^2), \quad \text{and} \\ \sigma_i^2 &= \left(2^{j_1-1}(1-2^{-J})/(\ln^2 2(1-2^{-J}(J^2+4)+2^{-2J}))\right)/n_i \end{aligned} \tag{5}$$

to an excellent approximation, n_i being the number of samples in the ith block, and $J = j_2 - j_1 + 1$ the width of the scaling range (see [24] for the exact expression). We wish to test the null hypothesis H_0: the means are identical, against H_1: the means differ. The tests from section 2 then apply immediately upon making the identification $\alpha_i = \mu_i$. Note that this choice of H_0, which in fact puts a priority on high confidence in decisions to *reject* rather than accept the proposition that scaling is constant, allows us to accept a simple constant α model whenever this is reasonable, in keeping with the principle of parsimonious modelling.

B. Statistical properties

The idealised tests from the previous section are based on the central axiom of a well defined set of m random variables, with known variances. In this section we examine only the effect of residual correlations among the wavelet coefficients on the corresponding wavelet tests (both the general and two-level tests with constant σ_i), and so it

is essential to satisfy this axiom, which corresponds to α values being constant on each block.

To study the type I error (reject H_0 when true), K sample paths of fGn and fARIMA of length n and constant parameter α were synthesized. Each sample path was split into m blocks of equal length and the above tests applied to each, counting the number of times that V fell within the critical region $V > C(\delta)$, with a significance level of $100(1-\delta) = 95\%$. In the results presented in figure 4 (top plots), we have $K = 2000$, $m = \{2, 4, 8, 16\}$, $\alpha = \{0.6, 0.8\}$ ($H = \{0.8, 0.9\}$), and $n = 2^{13}$ for the fGn, but $n = 2^{16}$ for the fARIMA, so that the range of octaves $[j_1, j_2]$ are of identical width. The plot shows that the two wavelet tests closely reproduce the theoretical 5% rejection rate, so we conclude that residual correlations only slightly affect the type I error probability.

To study the type II error (accept H_0 when false), K sample paths of fGn and fARIMA of length n were synthesized with α abruptly changing from α to $\alpha+\Delta\alpha$ at sample $n/2+1$. The samples were then split into m blocks of equal length and the above tests applied to each as described above, again with a significance level of 95%. In the results presented in figure 4, (lower plots), we have $K = 400$, $m = 10$, $\alpha = \{0.6, 0.8\}$ ($H = \{0.8, 0.9\}$) and $n = 2^{14}$ for the fGn, $n = 2^{17}$ for the fARIMA, and $\Delta\alpha$ was set to four different values: $\Delta\alpha = \{-0.1, -0.17, -0.23, -0.3\}$, corresponding to four evenly spread values of power. Figure 4 shows that, for both tests, the resulting values of the power for the four different values of θ are close to those derived theoretically (solid curves), as well as being insensitive to the initial α values. Note that the fact that the change point occurs at sample $n/2+1$ is only a convenient way to achieve the targeted power values and is in no way a simplification of the test. A different situation will be addressed in the next subsection.

This study reveals that, despite residual correlations (i,e., departures from **ID1** leading to asymptotic gaussianity, non-zero correlations between blocks, and only approximate expressions for estimation variances), the properties of the wavelet-based hypothesis tests are very close to those of the idealised problem for which exact theoretical results are available. The tests can therefore be regarded as close to UMPI in practice.

C. Choosing the number of blocks m

When applying the test to a set of data of length n, one has to select the number m and sizes $\{n_i\}$ of the blocks into which the data is split. For simplicity we will mainly discuss the general test in the case where the blocks are of equal size, leading to the question, how is m to be chosen? An essential difference from the ideal situation is that here, in order to use the test for an m fixed, it is necessary to *assume* that scaling is constant over each of the blocks separately. Such an assumption must of course be tested, and this can only be done by increasing m to examine the data at a higher time resolution. Only if m is high enough so that scaling is effectively constant over each block – and this may never be the case – can the test be

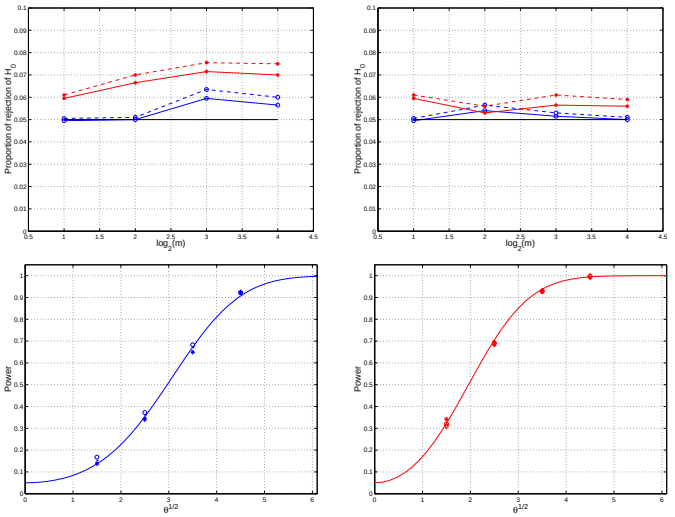


Fig. 4. **Statistical properties of the wavelet-based constancy tests. Left plots:** General test, **Right plots:** Two level test. **Top plots – Type I error**: the proportion of rejections of H_0 (when true) in simulations using fGn ($\circ$'s) and fARIMA ($*$'s) as a function of the number of blocks m, is very close to the chosen target of 5% (horizontal solid line), independently of m and the α value ($\alpha = \{0.6, 0.8\}$ given by $\{\text{solid}, \text{dashed}\}$ lines). **Lower plots – Type II error**: for each of four θ values the actual power of four wavelet tests (for $\alpha = \{0.6, 0.8\}$; with fGn: $\{\diamond, \circ\}$; with fARIMA: $\{+, *\}$) are compared against the idealised tests (solid curves). They are very close despite the residual correlations between the wavelet coefficients.

applied rigorously in essentially the same way as in the idealised context.

To begin, first note from 6 that for a given m the common variance σ_m^2 of the estimates $\hat{\alpha}_i$ are roughly inversely proportional to the number of analyzed samples. Hence $\sigma_m^2 \simeq \sigma^2 m$, where $\sigma^2 = \sigma_1^2$ is the variance of the estimate over the full series.

Under H_0 the assumption of scaling over each block is clearly satisfied, so the discussion at the end of section III-A applies: there is no preferred value of m. To understand the role of m under H_1, consider the following simple toy problem: the data to be analysed consists of 4 concatenated subseries, each exhibiting scaling with scaling parameters $\{\alpha_A, \alpha_B, \alpha_A, \alpha_B\}$ over a shared scaling range.

$*$ If one choses $m = 2$, scaling is not constant over each block, violating the assumptions of the test and yielding estimates $\{\hat{\alpha}_1, \hat{\alpha}_2\}$ which do not in fact correspond to anything meaningful. Nonetheless, the two estimates are statistically identical, and via the test procedure one is therefore very likely to accept H_0, as it is not possible from looking at $m = 2$ alone to determine if scaling is constant over each block or not. Essentially the number of blocks is not large enough to *see* or *follow* precisely enough the variation in time of α.

$*$ If one chooses $m = 4$ one obtains $\{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4\}$, estimates which **are** meaningful, and the assumptions of the test are satisfied, as the scaling is constant over each block with exponents $\{\alpha_A, \alpha_B, \alpha_A, \alpha_B\}$. The variances are $\sigma_4^2 = 4\sigma^2$ yielding $\theta_4 = (\alpha_A - \alpha_B)^2/(4\sigma^2)$. In this case, the power of the test (probability of accepting H_1 when true) is $P_4(\theta_4) = \int_{C_4}^{+\infty} f_{3,\theta_4}(x)dx$ and can be read off the $m = 4$

curve of figure 3.

$*$ If one now chooses $m = 8$, we obtain the estimates $\{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \hat{\alpha}_5, \hat{\alpha}_6, \hat{\alpha}_7, \hat{\alpha}_8\}$, which again are meaningful, the exponents being $\{\alpha_A, \alpha_A, \alpha_B, \alpha_B, \alpha_A, \alpha_A, \alpha_B, \alpha_B\}$. The variances are $\sigma_8^2 = 8\sigma^2$, yielding a θ_8 which from equation (4) can be shown to be $\theta_8 = 2\theta_4/2 = \theta_4$. In this case the power of the test is $P_8(\theta_8) = \int_{C_8}^{+\infty} f_{7,\theta_8}(x)dx$ and can be read off the $m = 8$ curve of figure 3. Since $\theta_8 = \theta_4$, it is valid to compare the two power functions using a common θ coordinate, and so from figure 3 one sees that for any fixed value of $\theta > 0$, the power of the test decreases when m is increased from 4 to 8. This is valid in general: *there is no benefit in increasing m beyond a given m_0, provided that the assumptions of the test are valid at $m = m_0$.*

Trade-off We see that the choice of m is subject to trading-off the need for m to be large enough to resolve time variations of the scaling parameter; but small enough to retain useful power. The *optimal* choice is therefore that the data be split into the largest possible (unequal) blocks within which the scaling parameter is *not* varying. This of course can not be achieved in practice as it depends on the specific unknown H_1, imposing an experimental methodology where m is varied. Thus, if the decision is *accept* H_0 for all m, then the final decision is *accept* H_0. Under H_1, the recommendation of the test may be *accept* H_0 for small m (because the blocks are too wide to see the variations of α), then *reject* H_0 for a given set of median values of m as the time variation of α is resolved, then again *accept* H_0 for large values of m because the statistical fluctuations of the estimates have become so high that they mask that variation. In other words, the power has become so low that one can no longer reliably detect that H_1 is true. In this case, the final decision should be *reject* H_0. We resist the urge to offer a definitive algorithm for the determination of the optimal m. We believe that so many factors are involved in practice that any simple criterion could lead to erroneous conclusions. Instead we continue to explore the basic factors that must be taken into account.

The above argument overlooks the fact that varying m implies varying the range of octaves of analysis $[j_1(m), j_2(m)]$, which implies a maximum m in practice to ensure sufficient scales per block for a reasonable estimate, another reason why m should be chosen as small as possible. A limitation intrinsic to the semi-parametric nature of the underlying estimator is that this maximum m may not be sufficient to resolve the time variation.

To illustrate the above issues, figure 5 gives a measure of how the tests behave in a more realistic context. As in section 4.2, the full wavelet tests are applied to fGn series generated according to a level change scenario for α. However, the change point is no longer in the middle, but at n^*, chosen to be 27% into the series, a value which ensures that it does not fall on a block boundary for any m used. Again, for simplicity a uniform value of $j_1 = 3$ is taken for each m. The range of m chosen is the largest possible satisfying the constraint that there be at least 4 scales, the minimum practical number. Thus, since $j_1 = 3$, we require $j_2 \geq 6$ which, with $n = 2^{14}$, results in $m = 2, 3 \cdots, 43$. Two sets of α values are used each with $\Delta\alpha = 0.2$: $\{\alpha_1, \alpha_2\} = \{0.6, 0.8\}$ and $\{0.4, 0.6\}$. Very similar results are obtained for both scenarios, as required, since they share the same θ, even though the second crosses the white noise frontier of $\alpha = 0.5$.

Figure 5 plots the percentage of rejections in $K = 500$ realisations. This (estimated) empirical measure of power is lower bounded by the significance level of 0.05, and upper bounded by the power of the optimal test (topmost line), obtained by splitting the data about n^* into $m = 2$ unequal blocks, with constant α in each, and calculating analytically the power of the test with **unequal** variances (see section 3.2). By inspecting Logscale Diagrams and estimates for several m values, one is led to suspect a change point around 1/4 into the trace, motivating the use of the two-level test. Thus we give results (using equal sized blocks) both for the general test, and the two-level test with the change point as near as possible to $m/4$ (in fact $m_1 = \max(1, \lfloor m/4 \rfloor)$). Between the bounds, the power varies with m according to the discussion above: low power when insufficient resolution mixes and hides changes in α, a peak at the minimum m value when resolution is adequate, and then a steady decline as variance increases (though with some fluctuations at very large m due to the small amount of effective data available). The oscillatory nature of the curves is an integer arithmetic effect, corresponding to when a block boundary passes particularly close to n^*. In particular, power is relatively higher at the 'harmonics' of $m = 4$, where power peaks. The general test has lower power than the two-level test, consistent with the fact that the 'guessed' change point of $n/4$ is close to n^*. In conclusion, despite the 'difficult' choice of n^*, the wavelet tests can successfully detect the lack of constancy in α over a reasonable range of m values (together with a reasonable choice of change point in the case of the two-level test), with power a good fraction of the theoretical maximum.

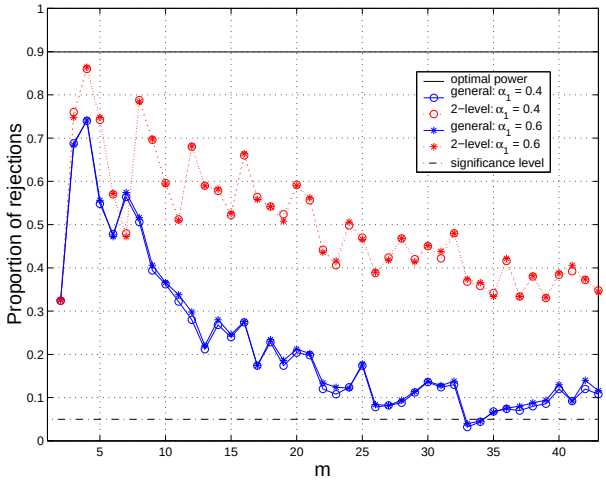


Fig. 5. **Practical example: rejection rate as a function of m.** The empirical power, estimated with 500 realisations, for the general (solid lines) and two-level tests (dotted lines). Two scenarios are given at constant $\Delta\alpha = 0.2$: $\alpha_1 = 0.4$ (o's), $\alpha_1 = 0.6$ ($*$'s) which agree closely.

D. The practical test procedure

To analyse the existence, and check for the constancy of, scaling in data, one can proceed as follows:

Initial Step Perform a global analysis of the data as proposed in [3], [4]. The choice of wavelet is not crucial but the number of vanishing moments N of the wavelets should be chosen to ensure $N \geq \alpha/2$, and as high as necessary to eliminate possible trends. ¿From the Logscale Diagram (the y_j vs j plot) of the full data set, determine an initial scaling range $[j_1, j_2]$ and estimate α according to equation (3).

If scaling is found the question is: is the observed $\hat{\alpha}$ meaningful? and if no scaling range is found, it is: can the data be split into sub-blocks over some or all of which scaling exists?

Procedure on blocks

1. Choose a significance level δ.
2. Choose a $m > 1$, but not so large as to exclude the scales of interest from each block (four is a practical minimum).
3. Examine the Logscale Diagrams for each block, and select a range of scales $[j_1(m), j_2(m)]$ common to each where scaling is observed. If no common range can be chosen, scaling is not constant and the test at this m is not defined. Go to step 5.
4. Compute and compare the threshold $C_m(\delta)$ and the test statistic V_m and record the test outcome at this m.
5. If valid m values remain go to step 2.
6. Analyse the set of m dependent test outcomes to draw the final conclusion as discussed above.

V. Application to Ethernet Data

To illustrate the use of the test on actual data, we apply it to some of the celebrated *Bellcore* Ethernet data sets. Recall briefly that these consist in lists of arrival times and Ethernet frame lengths recorded on a local area computer communications network. For a thorough description the reader is refered to [14] (see also [3]). From each of the data sets "pAug", "pOct", "OctExt" and "OctExt4" we have extracted an aggregated rate process of arriving work, that is a discrete time series corresponding to the number of bytes transmitted during contiguous constant length time intervals, here of length 12, 10, 1000, and 10 milliseconds respectively, values chosen for convenience so that the series would have approximately the same length. The time series for pAug and OctExt are plotted in the middle plots in figures 6 and 7 respectively.

For each time series the Logscale Diagram (LD) (top plot) is first computed with $N = 3$, and evidence for LRD ($0 < \hat{\alpha} < 1$) is seen in all but OctExt4. We then split into blocks. For pAug, pOct, and OctExt we are interested in checking that the evidence for LRD observed over the whole series is confirmed as being valid and constant in time, and for OctExt4 we wish to see if clear evidence of scaling appears over subsets of the series.

The $m = 12$ estimates, made in each case with $[j_1, j_2] = [7, \text{max-possible}]$, are shown in the bottom right plots of the figures together with their confidence intervals and the outcome of the test with a significance level of 95%. The observations were robust to changes of m. For each of pAug and pOct it was observed (not shown) that there is acceptable evidence that scaling is present in the same scaling range over each block, so that the assumptions of the test are satisfied and can be applied. In both cases H_0 was accepted. One can therefore return with confidence to the full series to estimate its value. For the times series OctExt and OctExt4 it was also observed, albeit less convincingly, that scaling is present in the same scaling range over each block. In both cases the test strongly indicated that H_0 be rejected.

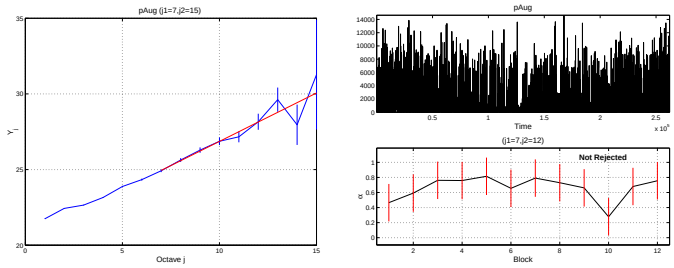


Fig. 6. **pAug.** Left: the LD of the whole time series with $N = 3$. Top right: the time series of bytes per 12ms intervals. Bottom right: the estimates from 12 adjacent blocks with $[j_1, j_2] = [7, 12]$, and the test outcome: Accept H_0. One then has $\hat{\alpha} = 0.64$.

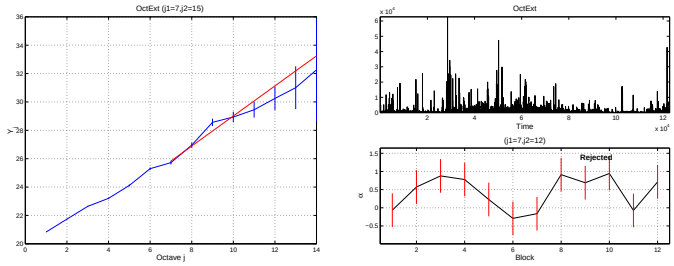


Fig. 7. **OctExt.** Left: the LD of the whole time series with $N = 3$. Top right: the time series of bytes per 1000ms intervals. Bottom right: the estimates from 12 adjacent blocks with $[j_1, j_2] = [7, 12]$, and the test outcome: Reject H_0: one cannot estimate α over the whole series.

It is instructive to examine a little further the case of OctExt. Since H_0 was rejected, we must consider in hindsight that the LD of the entire series, shown in figure 7, is meaningless from the point of view of measuring a scaling exponent. This alignment in the LD is merely an undesirable artifact resulting from the 'averaging' of the non-stationarity across the series. The fact that in reality the scaling is not constant is graphically illustrated in the extreme variability of the $\hat{\alpha}_i$ in the lower plot of figure 7, as detected in the negative test outcome.

VI. Conclusion

Two statistical tests were provided to investigate formally the question of the constancy in time of scaling exponents of scaling processes, the first in the general case and the second for changes following a level shift. The properties of the tests are very close to those of corresponding idealised inference problems, for which they are uniformly most powerful invariant, with explicitly known power functions. The reduction to such elegant tests is due the many

advantageous properties of the wavelet based estimator of scaling exponents which underlies the test design. In addition a methodology was developed for the use of the tests in practical situations, which involves non-trivial complications, and examples are given using real data. A key outcome is an ability to determine if an apparent scaling observed across a time series is in fact meaningful, and the corresponding estimate reliable.

References

[1] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical functions*, Dover Publications Inc., New York, 1970.

[2] P. Abry, P. Gonçalvès and P. Flandrin, "Wavelets, spectrum estimation and $1/f$ processes", in A. Antoniadis and G. Oppenheim, eds, *Wavelets and Statistics*, *Lectures Note in Statistics* **103**, pp. 15–30. Springer-Verlag, New York, 1995.

[3] P. Abry, D. Veitch, "Wavelet analysis of long-range dependent traffic", *IEEE Trans. on Info. Theory*, 44(1) pp 2–15, 1998.

[4] P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data", in "Self Similar Network Traffic Analysis and Performance Evaluation", K. Park and W. Willinger, Eds., Wiley, pp. 39–88, 2000.

[5] P. Abry, P. Flandrin, M. S. Taqqu and D. Veitch, "Self-similarity and long-range dependence through the wavelet lens", a chapter in *Long range dependence: theory and applications*, G. Oppenheim, M. Taqqu, P. Doukhan editors, Wiley 2001.

[6] P. Abry, L. Delbeke, P. Flandrin, "Wavelet-based estimator for the self-similarity parameter of α-stable processes", IEEE-ICASSP99, pp.III, 1581–1584, Phoenix, USA, May 1999.

[7] J. Beran, N. Terrin. "Estimation of the long-memory parameter, based on a multivariate central limit theorem. Journal of Time Series Analysis", Vol.15, No.3, pp.269–277, 1994.

[8] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia (PA), 1992.

[9] R. Dijkerman, R. Mazundar. "On the correlation structure of the wavelet coefficients of the fractional Brownian motion", *IEEE Trans. on Info. Theory*, 40(5), pp.1609–1612, 1994.

[10] A. Feldmann, A.C. Gilbert, W. Willinger, "Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic", Proceedings of the ACM/SIGCOMM'98, Sept. 1998, Vancouver, Canada.

[11] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley, second edition, 1971.

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

[13] E.L. Lehmann, *Testing Statistical Hypotheses*, Wiley, second edition 1986.

[14] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (Extended version)", *IEEE/ACM Trans. on Networking*, 2, pp. 1–15, 1994.

[15] S. Mallat, "A Wavelet Tour of Signal Processing", Academic Press 1998.

[16] V. Paxson, "Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic". *Computer Communications Review*, V. 27 No. 5, pp.5-18, October 1997.

[17] K. Park and W. Willinger, Eds., "Self Similar Network Traffic Analysis and Performance Evaluation", Wiley, 2000.

[18] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipies in C, The art of scientific computing*, second edition. Cambridge University Press.

[19] M. Roughan, D. Veitch, P. Abry, "On-line estimation of LRD parameters", *Proceeding Globecom '98, Sydney*, Vol.6, pp.3716-3721, Nov 1998.

[20] M. Roughan, D. Veitch, "Measuring Long-Range-Dependence under Changing Traffic Conditions", *proceedings of IEEE Infocom99*, NY, NY, pp.1513–1521, March 1999.

[21] R.H. Riedi, "An Improved Multifractal Formalism and Self-Similar Measures", *J. Math. Anal. Appl*, 189, pp.462–490, 1995.

[22] A.H. Tewfik, M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion", *IEEE Trans. Info. Theory*, 38, pp.904–909, 1992.

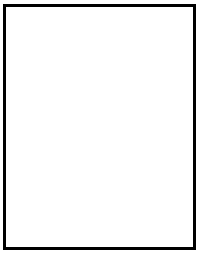
[23] S.D.Silvey, *Statistical Inference*, Chapman & Hall 1975.

[24] D. Veitch, P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence", *IEEE Trans. on Info. Theory* , Special Issue on Multiscale Statistical Signal Analysis and its Applications, April, Vol 45, no.3, pp.878–897, 1998.

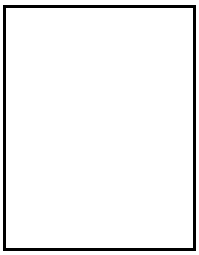
[25] D. Veitch, P. Abry and M. S. Taqqu "On the automatic selection of scaling range in the semi-parametric estimation of scaling exponents", in preparation.

[26] G.W. Wornell, A.V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets", *IEEE Trans. on Signal Proc.*, 40(3), pp.611–623, 1992.

[27] W. Willinger, M.S. Taqqu, A. Erramilli, "A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for High-Speed Networks", in Stochastic Networks: Theory and Applications, F.P.Kelly, S.Zachary, I.Ziedins, editors, Royal Statistical Lecture Notes, Vol. 4, pp.339-366, Clarendon Press, Oxford, UK, 1996.



Darryl Veitch was born in Melbourne, Australia in 1963. He completed a Bachelor of Science Honours degree at Monash University, Melbourne in 1985, and a mathematics doctorate in Dynamical Systems at the University of Cambridge, England, in 1990.
In 1991 he joined the research laboratories of Telecom Australia (Telstra) in Melbourne where he became interested in long-range dependence as a property of tele-traffic in packet networks. In 1994 he left Telstra to pursue the study of this phenomenon at the CNET in Paris (France Telecom). He then held visiting positions at the KTH in Stockholm, INRIA in the south of France, and Bellcore in New Jersey, before taking up a three year position as Senior Research Fellow at RMIT, Melbourne. He is now a Senior Research Fellow the Ericsson funded EMUlab in the Department of Electrical Engineering at the University of Melbourne. His research interests include scaling models of packet traffic, parameter estimation problems and queueing theory in a long range dependent context, and the statistical and dynamic nature of Internet traffic. His email address is: d.veitch@ee.mu.oz.au



Patrice Abry Patrice Abry was born in Bourgen-Bresse, France in 1966.
He received the degree of Professeur-Agrege de Sciences Physiques, in 1989 at Ecole Normale Supérieure de Cachan and completed a PhD in Physics and Signal Processing, at Ecole Normale Supérieure de Lyon and Université Claude-Bernard Lyon I, in 1994.
He is, since October 95, a permanent CNRS researcher, at laboratoire de Physique from Ecole Normale Supérieure de Lyon.
Patrice Abry received the AFCET-MESR-CNRS prize for best PhD in Signal Processing for the years 93-94 and is the author of a book "Ondelettes et Turbulences - Multirésolution, algorithmes de décompositions, invariance d'échelle et signaux de pression", published in october 97, by Diderot, éditeur des Sciences et des Arts, Paris, France.
His current research interest include wavelet-based analysis and modelling scaling phenomena and related topics (self-similarity, atable processes, fractal, 1/f processes, long-range dependence, local regularity of processes, inifinitely divisible cascades...). The applications of current interest are physics of turbulence and analysis and modelling of packet traffic.