Long-Range Dependence: revisiting Aggregation with Wavelets.

Patrice ABRY (1) Darryl VEITCH (2) Patrick FLANDRIN (1)

(1) - *CNRS URA 1325 - Laboratoire de Physique - Ecole Normale Supérieure de Lyon - 46, allée d'Italie 69 364 LYON Cedex 07 - France*
tel: (+33) 4 72 72 84 93 - Fax: (+33) 4 72 72 80 80 - E-mail : pabry@physique.ens-lyon.fr

(2) - *BELLCORE, MCC 1G-309B, 445 South Street, Morristown, NJ 07962-1910, USA*
tel: (+1) 201 829 4033 - Fax: (+1) 201 829 2645 - E-mail : darryl@bellcore.com

Abstract. The aggregation procedure is a natural way to analyse signals which exhibit long-range dependent features and has been used as a basis for estimation of the Hurst parameter, H. In this paper it is shown how aggregation can be naturally rephrased within the wavelet transform framework, being directly related to approximations of the signal in the sense of a Haar-multiresolution analysis. A natural wavelet based generalisation to traditional aggregation is then proposed: "a-aggregation". It is shown that a-aggregation cannot lead to good estimators of H, and so a new kind of aggregation, "d-aggregation", is defined, which is related to the details rather than the approximations of a multiresolution analysis. An estimator of H based on d-aggregation has excellent statistical and computational properties, whilst preserving the spirit of aggregation. The estimator is applied to telecommunications network data.

Key-words. Long-range dependence, self-similarity, aggregation, multiresolution analysis, wavelet transform, parameter estimation.

1 Motivation

It has now been amply demonstrated [14, 12, 16, 8] that long-range dependence (LRD) is present in telecommunications traffic data recorded from high speed networks, and that this phenomenon is likely to have major consequences for the performance of the said networks. A full analysis of LRD is therefore of considerable importance. By LRD we understand not only that correlations persist over very long time scales, but also that they possess a certain *asymptotically self similar* structure. It is therefore very natural to analyze, or even define, LRD by performing a rescaling operation at different scales, and then to observe how measured properties vary as a function of scale. This is the essence of the *aggregation procedure* [14], an idea which has served as a basis for much of the study of the LRD phenomenon. In particular, a major concern is to accurately estimate the parameters characterising LRD, which basically amounts to the determination of the second-order properties of the data. The aggregation procedure has been used as a natural starting point for the design of such estimators, however the inherent difficulties associated with LRD has prevented them from having desirable statistical properties.

The aim of this paper is to show that the aggregation procedure has close connections to the multiresolution analysis underlying the wavelet transform, and therefore can be naturally rephrased in this framework. We then present the extensions made possible by the use of wavelets and show how we arrive at an estimator for the LRD parameter which is particularly simple to use and which enjoys excellent statistical properties, whilst preserving the spirit of aggregation analysis [4, 22].

2 Long-Range Dependence and the Aggregation point of view.

• **Definitions.** We consider second-order stationary random processes x, which for convenience will be defined in discrete time $k = 1, 2, 3, \cdots$. The covariance function $\gamma_x(\tau)$ of x is defined to be

$$\gamma_x(\tau) = \mathbb{E}\, x(k)x(k+\tau), \quad \tau = 0, 1, 2, \cdots$$

and the autocorrelation function $c_x(\tau)$ is just its normalised form:

$$c_x(\tau) = \gamma_x(\tau)/\gamma_x(0)\,.$$

In this paper we will define $f(x) \sim g(x)$ as $x \to \infty$ to mean that $\lim_{x\to\infty} f(x)/g(x) \to C$ where C is some finite constant.

• **Long-Range Dependence.** The process x is long-range dependent (LRD) if its autocorrelation function for large lag τ decreases as a power-law (We do not consider here the more general case where a slowly varying function modulates the power-law),

$$c_x(\tau) \sim \tau^{-\beta},\ \tau \to +\infty,\ \beta \in (0,1). \tag{1}$$

Essentially, the LRD phenomenon states that the ratio of the correlations at any two sufficiently large lags τ_1 and τ_2 remains appreciable no matter how large the interval $\Delta\tau = |\tau_2 - \tau_1|$ separating them. Crucially, it implies that there is no possibility of defining a characteristic time-scale τ_0, beyond which correlations would have essentially disappeared, as would be the case for a process whose autocorrelation function reads $c_x(\tau) \sim \exp(-\tau/\tau_0)$. Thus, one can not find a reference unit of time over which, for instance, some property of the data could be reliably measured. Instead of a single prominent timescale, LRD is characterised by scale invariance properties governed by the parameter β which describes the relationship *between* scales.

The above equation implies the divergence of the correlation sum $\int c_x(\tau)d\tau = \infty$, a defining feature of LRD. Note that equation (1) with $\beta \in (1,2)$ is also of interest in other contexts and can be studied within the framework we discuss. In this case however the correlation sum is finite and so the process is not LRD, but short-range dependent, despite the power-law form.

• **Aggregation: a natural analysis for LRD.** To exploit the time-scale invariance properties of a LRD process x, it is natural to think of forming the averaged or *aggregated* discrete time process $x^{(T)}$ with *aggregation level* T as

$$x^{(T)}(k) = (1/T)\int_{kT-T}^{kT} x(u)du.$$

The short range averaging effect of aggregation does not change the asymptotic dependence structure, so $x^{(T)}$ remains LRD with index β for any T. The short range structure on the other hand, being increasingly mixed with the long range structure as T increases, takes on its power law character. More precisely, with increasing T the correlation structure $c_x^{(T)}(\tau)$ of $x^{(T)}$ tends to a non-degenerate limiting form $c(\tau)$, itself invariant under aggregation [10], given by

$$\begin{aligned} c(1) &= 2^{1-\beta} \\ c(\tau) &= \tfrac{1}{2}\left((\tau+1)^{2-\beta} - 2\tau^{2-\beta} + (\tau-1)^{2-\beta}\right), \quad \tau = 2, 3 \cdots \end{aligned} \tag{2}$$

Note that although the *correlation* structure of $x^{(T)}$ tends to a nondegenerate limit with increasing T, the *covariances* at each lag decrease to zero, though more slowly than in the short range dependent case.

• **Long-range dependence *vs* self-similarity.** LRD processes are also called asymptotically second order self-similar, and exactly second order self-similar if in addition (2) is satisfied. The self-similarity referred to here is a description of the invariance of second order statistics under rescaling as described by (2), however there is a natural connection to processes which are self-similar in the strict sense. A process $Y(t)$ with continuous time parameter t is called self-similar (in the strict sense) with self-similarity parameter H, if for any positive scale factor c the rescaled process $c^{-H}Y(ct)$ is distributed as $Y(t)$ [7, 20]. The most important class of such processes is that where $H > 0$ and Y has stationary increments. Now assuming in addition the existence of second moments and restricting to $H > 1/2$, the increment process $Y(k+1)-Y(k)$ is just the LRD stationary process $x(k)$ above obeying equation (2). The LRD and self-similarity parameters are related as $H = 1 - \beta/2$, and we often use H instead of β with this understanding. The fractional Brownian motion (fBm), which is the only self-similar Gaussian process with stationary increments and finite second moments, is the paradigm for this class of processes.

• **Aggregation: a renormalisation tool.** The above discussion underlines one of the main features of aggregation, namely that it is a natural renormalisation tool which reveals the LRD or asymptotically self-similar aspect of the signal, via the convergence to second order exact self-similarity it induces. Informally this can be seen in plots of $x^{(T)}$, which have a similar appearance for different T, up to amplitude and time rescaling. Another important aspect of aggregation is the fact that, through averaging, the aggregated process becomes increasingly Gaussian-like with increasing T. In particular, since there is only one second order exactly self-similar Gaussian process with finite variance, namely the discrete Fractional Gaussian Noise, the well documented properties of this latter process can be used in the study of highly aggregated LRD signals. A drawback of aggregation is that it in effect reduces the length of a finite data set without making use of the high frequency behaviour lost.

• **Aggregation: a mathematical implication.** As mentioned above, the covariances at each lag decrease to zero more slowly than the classical $O(1/T)$. More precisely, for the variance we have

$$\mathrm{var}\left(x^{(T)}\right) \sim T^{-\beta}, \qquad T \to +\infty. \tag{3}$$

• **Estimating the LRD parameter.** In many situations, it is of major importance for analysis or modelling to accurately measure the index β, which both reveals the existence of the LRD phenomenon, quantifies its importance, and determines the normalisation factors for statistics of all kinds. A natural idea arising from the aggregation point of view is to use equation (3) to estimate β from a linear fit in a $\log(T)$ *vs* $\log\left(\mathrm{var}(x^{(T)})\right)$ plot. However, it is well-known [7] that estimation from a single realization of a LRD process is problematic because of the persistent serial correlations in the data. The sample variance estimator for instance has very poor statistical properties: strong bias and high variance. A variety of more sophisticated estimators have been designed to overcome this difficulty, of which the most famous and efficient is known as the Whittle estimator [7, 17, 4], which is based on a Gaussian maximum-likelihood principle. Much of statistical analysis is dependent upon Gaussian assumptions, including most versions of the Whittle estimator, and so aggregation is advantageous in this regard as it increases the Gaussian character of the signal. For a review of several estimators and a numerically based comparison of their performance, see [18].

• **Revisiting aggregation.** We are now going to show how the aggregation property can be understood in the general framework of wavelet (or time-scale) analysis. We show how this framework leads to natural generalizations, which lead to an elegant and efficient estimator of the LRD parameter which remains very close to the initial, intuitive idea of aggregation [4, 22].

3 Multiresolution Analysis and Discrete Wavelet Transform

• **Multiresolution analysis.** A multiresolution analysis (MRA) consists in a collection of nested subspaces $\{V_j\}_{j\in\mathcal{Z}}$, satisfying the following set of properties [11]:

- i) $\bigcap_{j\in\mathcal{Z}} V_j = \{0\}$, $\bigcup_{j\in\mathcal{Z}} V_j$ is dense in $L^2(\mathcal{R})$
- ii) $V_j \subset V_{j-1}$
- iii) $x(t) \in V_j \longleftrightarrow x(2^j t) \in V_0$

- iv) There exists a function $\phi_0(t)$ in V_0, called the *scaling function*, such that the collection $\{\phi_0(t-k), k \in \mathcal{Z}\}$ is an unconditional Riesz basis for V_0

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

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

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

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

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

$$\begin{aligned} x(t) &= \text{approx}_J(t) + \textstyle\sum_{j=1}^{j=J} \text{detail}_j(t) \\ &= \textstyle\sum_k a_x(J,k)\phi_{J,k}(t) \ + \ \sum_{j=1}^{J}\sum_k d_x(j,k)\psi_{j,k}(t) \end{aligned} \qquad (4)$$

Since the approx_j are essentially coarser and coarser approximations of x, ϕ_0 needs to be a low-pass function. The detail_j, being an information 'differential', indicates rather that ψ_0 is a band-pass function, and therefore a small wave, a *wavelet*. More precisely, the MRA shows that the mother wavelet must satisfy $\int \psi_0(t)dt = 0$ and that its Fourier transform obeys $|\Psi_0(\nu)| \sim \nu^N$, $\nu \to 0$ where N is a positive integer called the number of vanishing moments of the wavelet [11].

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

$$x(t) \to \{\{a_x(J,k),\ k \in \mathcal{Z}\},\ \{d_x(j,k),\ j = 1,\ldots,J,\ k \in \mathcal{Z}\}\} \qquad (5)$$

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

$$\left.\begin{aligned} a_x(j,k) &= \langle x, \mathring{\phi}_{j,k}\rangle \\ d_x(j,k) &= \langle x, \mathring{\psi}_{j,k}\rangle \end{aligned}\right\} \qquad (6)$$

where $\mathring{\psi}_{j,k}$ (resp., $\mathring{\phi}_{j,k}$) are shifted and dilated templates of $\mathring{\psi}_0$ (resp., $\mathring{\phi}_0$), called the dual mother wavelet (resp., the dual scaling function), and whose definition depends on whether one chooses to use an orthogonal, semi-orthogonal or bi-orthogonal DWT (see e.g., [11, 5]). They are computed in practice from a fast recursive filter-bank-based pyramidal algorithm which has a lower computational cost than that of a FFT [11].

• **Discrete Wavelet Transform of Stochastic Processes.** It has been clearly demonstrated in the literature that the wavelet transform can be applied to stochastic processes, see e.g. [9, 15]. More specifically, for the second-order random processes of interest in the LRD context, it is well-known that the wavelet transform is a second-order random field, on the condition that the scaling function ϕ_0 (and hence the wavelet ψ_0) satisfy certain mild conditions [9, 15] related to the statistical properties of the analysed process. For instance, if the covariance sum of the random process is bounded, ϕ_0 and ψ_0 have to be in L^2; if the covariance sum diverges (while the covariance remains bounded), ϕ_0 and ψ_0 must be in L^1 [9]. Let us denote by $\mathbb{E}$ and $R_x(u,v)$ respectively the expectation and the covariance of the random process. We then have the following set of relations that govern the statistics of the wavelet transform:

$$\left.\begin{array}{lcl} \mathbb{E}a_x(j,k) & = & \mathbb{E}x\int\phi_0(t)dt \\ \mathbb{E}d_x(j,k) & = & 0 \\ \mathbb{E}a_x(j,k)a_x(j',k') & = & \int\int R_x(u,v)\phi_{j,k}(u)\phi_{j',k'}(v)dudv \\ \mathbb{E}d_x(j,k)d_x(j',k') & = & \int\int R_x(u,v)\psi_{j,k}(u)\psi_{j',k'}(v)dudv \end{array}\right\} \tag{7}$$

In the case of second-order stationary processes $R_x(u,v) = \gamma_x(u-v)$, and it is convenient to work in the frequency domain. Let $\Gamma_x(\nu)$ (the power spectrum), Φ_0 and Ψ_0 be the Fourier Transforms of γ_x, ϕ_0 and ψ_0 respectively. The variances can be written as

$$\left.\begin{array}{lcl} \mathbb{E}|a_x(j,k)|^2 & = & \int\Gamma_x(\nu)2^j|\Phi_0(2^j\nu)|^2d\nu \\ \mathbb{E}|d_x(j,k)|^2 & = & \int\Gamma_x(\nu)2^j|\Psi_0(2^j\nu)|^2d\nu \end{array}\right\} \tag{8}$$

These relations, which can be given a spectral estimation interpretation [2], constitute the starting points for deriving equations (10) and (12).

As mentioned above, LRD (stationary) processes are closely connected to self-similar (non-stationary) processes such as fBm. The autocovariance functions of these latter processes are not bounded, however the second order statistics of their wavelet transforms exist provided that the scaling function and mother wavelet satisfy extra conditions, see e.g., [9, 15]. Henceforth we will assume that the scaling functions and wavelets decay at least exponentially fast in the time domain, so that the second-order statistics of the wavelet transform exist for all of the random processes we discuss.

4 Aggregation and Multiresolution

• **Equivalence of the Haar multiresolution and Aggregation.** The Haar multiresolution is designed from the scaling function $\phi_0(t) = \chi_{[0,1]}(t)$, where $\chi_{[a,b]}$ is the characteristic function for the interval $a \le t < b$. A corresponding mother wavelet reads: $\psi_0(t) = \chi_{[0,1/2]}(t) - \chi_{[1/2,1]}(t)$, a function such that the $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t-k),\ k \in \mathcal{Z}, j \in \mathcal{Z}\}$ form an orthonormal basis for L^2.

From this definition, it is straightforward to check that the aggregated process $x^{(T)}$, with $T = 2^j$, can be rewritten as an approximation of x:

$$x^{(2^j)}(k) = a_x(j,k) = \langle x, \phi_{j,k}\rangle\,. \tag{9}$$

This relation shows that there is an exact, obvious and natural identity between the aggregation procedure and the Haar multiresolution analysis: studying x over longer and longer observation periods T simply translates in the MRA vocabulary to increasing the scale of analysis 2^j, or equivalently to lowering the resolution. Amongst others, a consequence is that whenever x is a LRD process, one has [2]:

$$\mathrm{var}\,(a_x(j,k)) \sim 2^{j(1-\beta)}, \quad j \to \infty,\ \forall k\,. \tag{10}$$

• **Beyond Haar Aggregation: Changing the wavelet.** There exist infinitely many MRAs (see e.g., [5, 1]). We selected the Haar only to underline the possibility of exactly reformulating the aggregation

procedure as a multiresolution analysis. However, dropping the insistance on an exact identity, we can arbitrarily select a MRA and still retain the key idea underlying the aggregation procedure: tracking scale invariance through analysis over larger and larger scales. Any MRA can therefore be understood as an aggregation procedure.
Moreover, the Haar MRA plays the role of a starting point from which many new MRAs can be designed. It can for instance generate the higher order spline MRA [21], or the so-called Daubechies MRA [11]. In both cases, wavelets result whose number of vanishing moments N can be freely chosen.

• **A new kind of Aggregation: Detail instead of Approximation.** Aggregating the data means averaging it over a time duration T, in other words, it means filtering it with a low-pass function, whose characteristic time support is of length T. However, another possibility would be to filter the data with a band-pass function, whose characteristic time support is of order T. Let us call $y^{(T)}$ such a band-pass aggregated process. There is a natural identity here also between this new type of aggregated process and the coefficients of the Discrete Wavelet Transform. This time however the coefficients are the details rather than the approximations:

$$y^{(2^j)}(k) = d_x(j,k) = \langle x, \psi_{j,k} \rangle . \tag{11}$$

Again, x being a LRD process results in a power law behaviour for the variance of the aggregated process as a function of aggregation level [2]:

$$\mathrm{var}\,(d_x(j,k)) \sim 2^{j(1-\beta)},\ j \to \infty,\ \forall k . \tag{12}$$

We now briefly sketch the proof of the above relation. From equations (8), (7), it is not difficult to see that

$$\begin{aligned} \mathrm{var}\,(d_x(j,k)) &= \mathbb{E} d_x(j,k)^2 \\ &= \int \Gamma(2^{-j}\nu)|\Psi_0(\nu)|^2 d\nu \end{aligned} \tag{13}$$

Since the spectrum of LRD processes near the origin reads $\Gamma_x(\nu) \sim |\nu|^{(\beta-1)}$, as $j \to \infty$ we have

$$\begin{aligned} \mathrm{var}\,(d_x(j,k)) &\sim \int |2^{-j}\nu|^{\beta-1}|\Psi_0(\nu)|^2 d\nu \\ &= 2^{j(1-\beta)} \int |\nu|^{\beta-1}|\Psi_0(\nu)|^2 d\nu \\ &\sim 2^{j(1-\beta)} . \end{aligned} \tag{14}$$

Equation (10) can be derived in a similar way.
We will henceforth refer to this detail related aggregation process as *d-aggregation*, and the MRA approximation based version of the more traditional aggregation process as *a-aggregation*.
We quickly note that, although defined here through the non-redundant (Discrete) Wavelet Transform, this analysis would still hold with a redundant, for instance a Continuous, Wavelet Transform. We will not enter into this technical issue here (see e.g. [2]).

• **The Allan variance from the wavelet point of view.** To more accurately estimate $\mathrm{var}(x^{(T)})$, in 1966 Allan [6] proposed replacing the estimate $V_1(T) = \frac{1}{K}\sum_{k=1}^{K}(x^{(T)}(k))^2$ by another, the *Allan variance* $V_2(T) = \frac{1}{K}\sum_{k=1}^{K-1}\left(x^{(T)}(k+1) - x^{(T)}(k)\right)^2$, and showed that it had far better statistical properties. In terms of the Haar multiresolution, the Allan variance can be written as

$$\left.\begin{aligned} V_2(2^j) &= \tfrac{1}{K}\textstyle\sum_{k=1}^{K-1}\left(a_x(j,k+1) - a_x(j,k)\right)^2 \\ &= \tfrac{1}{K}\textstyle\sum_{k=1}^{K-1}\left(\sqrt{2}d_x(j+1,k)\right)^2 \end{aligned}\right\}$$

The Allan variance is therefore clearly based on a Haar d-aggregation of the data. The improvement of its statistical performance will therefore be explained by the wavelet based arguments below. The connection between the Allan variance, the Haar wavelet and the multiresolution analysis was first reported in [13].

• **Estimating the LRD parameter in the wavelet framework.** Both the a-aggregation and d-aggregation procedures contain the key idea of rescaling with T replaced by 2^j, and yield power law relations, equations (10) and (12) respectively, which could be used as a basis for the estimation of β. We have already discussed that the use of (3) (the analogue of (10)) leads to estimators with very poor properties. We now show how an estimator from d-aggregation based on (12) can do far better. This improvement results from the following two basic properties. The first is common to both a and d-aggregation, whereas the second holds only for the latter.

1. The power law behaviour of equation (12) (resp., (10)) holds only because the set of analysing functions, the wavelet family, (resp., the scaling functions family) is designed from the dilation or change of scale operator, $\psi_{j,0}(t) = 2^{-j/2}\psi_0(2^{-j}t)$, (resp., $\phi_{j,0}(t) = 2^{-j/2}\phi_0(2^{-j}t)$). In other words, the use of the dilation operator results in a **constant relative** bandwidth spectral analysis that exactly matches the power-law form near the origin of the spectra associated with LRD processes. In contrast, a Fourier based spectral analysis, which relies on a basis designed from the frequency shift operator, performs a **constant absolute** bandwidth analysis which does not lead to an exact power-law form in frequency of any second-order statistic. Such power laws are therefore related to the multiresolution analysis or to the fact that the variance is expressed as a function of the scale variable j [2].

2. The next question is how to estimate the above variances, since as already indicated, estimating second order statistics is difficult in the presence of LRD. The central fact is that the expectation of the details is identically zero, due to the band-pass ($N \geq 1$) nature of the wavelet, and therefore the variance of the details reduces to their second moments. Now it is generally true that the normalized sum of squares is an unbiased estimator of the second moment, so $1/N_j \sum_{k=1}^{k=N_j} d_x(j,k)^2$ is an **unbiased** estimate of the variance var $(d_x(j,\cdot))$. Here N_j is the number of available coefficients at scale j. Now whilst the correlations of $x(t)$ as expressed in (1) are long-range, in the wavelet representation of the data we have $\mathbb{E} d_x(j,k) d_x(j,k') \sim |k-k'|^{-\beta-2N}$ [13, 2, 3], which is always short range. In addition, the degree of decorrelation can be controlled by increasing N. To see this, note that the power-law behaviour of the Fourier transform of the wavelet, at frequencies near 0: $|\Psi_0(\nu)| \sim \nu^N, \nu \to 0$, cancels the power-law divergence of the spectrum at the origin, a feature of LRD processes. These crucial properties - zero bias and quasi-decorrelation of the wavelet coefficients - allow the standard sample variance estimator to asymptotically **efficiently** estimate var$(d_x(j,\cdot))$. That is, the variance of the estimator decreases as $\sim 1/N_j = 2^j/N_0$ where N_0 is the size of the data.

The two arguments above, that var $(d_x(j,\cdot))$ behaves as a power-law of the scale j, and that the mean of the square of the wavelet coefficients is an unbiased, asymptotically efficient estimate for it, leads us to propose an estimator for the LRD parameter β as the slope of a linear fit in a $\log_2(2^j)$ *vs* $\log_2\left(2^j/N_0 \sum_k d_x(j,k)^2\right)$ plot. This slope is measured using a linear weighted least squares fitting procedure. Using the quasi-decorrelation assumption, it can be shown that the resulting estimator is asymptotically **unbiased** and **efficient** [2]. Moreover, in practice, the bias turns out to be negligible even for moderate N_0. Exact analytical results and further details on the statistical performance of this wavelet-based estimator can be found in [2]. It can moreover be given the following interpretation. Whenever one expands a function over a basis, some of its properties are conveyed by the basis, and others by the coefficients. In the wavelet expansion of a LRD process (equation (4)), the LRD is captured by the wavelet basis leaving the expansion coefficients (i.e., the wavelet coefficients) almost uncorrelated. It is this and the fact that the number N of vanishing moments of the wavelet can be freely chosen in the MRA framework, which are the keys to the statistical success of the approach. It can moreover be shown that by varying N a high degree of robustness with respect to certain non stationarities (smooth deterministic trends including linear trends) which are likely to corrupt data can be achieved [4, 2, 3].

• **Estimating the LRD parameter in the a-aggregation framework.** Returning to the approximation coefficients, it can be checked that their correlation, which reads $\mathbb{E}a_x(j,k)a_x(j,k') \sim |k-k'|^{-\beta}$ [2, 3], remains long-range with the same index as the original time domain signal x. The low-pass nature of the scaling function (as opposed to the band-pass nature of the wavelet $N \geq 1$) used in the computation of the approximation coefficients does not bring any reduction in correlation, and therefore the standard variance estimator would result in a poor estimate of $\text{var}(a_x(j,\cdot))$, or equivalently of $\text{var}\ (x^{(T)})$. The multiresolution formulation of the original aggregation technique therefore gives, through the identity of equation (9), a clearer insight into the poor properties of an estimator obtained from (3).

• **Generalization to self-similar processes with stationary increments: estimating the self-similarity parameter.** When applied to real data it may be that estimators for β yield negative values, calling into question the assumption of second-order stationarity. If values $\beta \in (-1,0)$ are found it is known that such data can be modelled using self similar processes with stationary increments, such as the fBm. However to measure the self-similarity parameter using tools of the *a-aggregation* type would first require the computation of the increment process, whereas the wavelet-based estimate can be directly performed on the original process. This is another major advantage of the wavelet technique. Indeed it has been shown [13, 15] that when the wavelet transform is applied to self-similar random processes with stationary increments, for each fixed scale j the details $\{d_x(j,k), k \in \mathcal{Z}\}$ are stationary processes. Moreover it has been shown [2] that equation (12) still holds so that the estimator described above can be efficiently used to measure the self-similarity parameter. The basic ingredient that produces this *stationarizing* effect is the band-pass nature of the wavelet, which shows that even disregarding the reduction of correlation effect, the number of vanishing moments of the wavelet N plays a key role.
Let B_H be a fBm with $H \in (0,1)$. As suggested in [19], the process defined by

$$Z(t) = \int_0^t B_H(s)ds\,,$$

is self-similar with parameter $H+1$ in the range 1 to 2, with *increments of second degree* (that is, in this case, the increments of the increment process) which are stationary. For such a process it can be shown that the wavelet coefficients $\{d_x(j,k), k \in \mathcal{Z}\}$ form a stationary process provided that the number of vanishing moments of the wavelet is at least 2 ($N \geq 2$), and in addition that equation (12) still holds. This can be generalized in a straightforward way to self-similar processes whose *increments of higher degree* are stationary, allowing the modelling of processes with arbitrary $H > 1$ or $\beta < 0$. Thus by increasing N, the wavelet framework enables the self-similarity parameter to be estimated efficiently without requiring the computation of increments of a degree which is unknown a priori.

• **An illustrating example on actual Ethernet data.** The plot in Figure 1 of $\log_2(2^j)$ *vs* $\log_2\left(2^j/N_0 \sum_k d_x(j,k)^2\right)$ is of one of the continuous time Ethernet data traces from the well-known set collected at Bellcore [14] in the early 1990's. The linear regime can easily be observed and persists over a large number of octaves, clearly revealing the presence of the LRD phenomenon and allowing a virtually unbiased estimate of β, here measured to be 0.4 ± 0.048 with 95% confidence. The Haar wavelet ($N=1$) was used. Estimates with higher values of N were unchanged, indicating the absence of smooth deterministic trends of any kind.
The data set describes the arrival times and sizes in continuous time of 1 million Ethernet frames. A complete analysis therefore would imply choosing a small time scale with which to discretize the

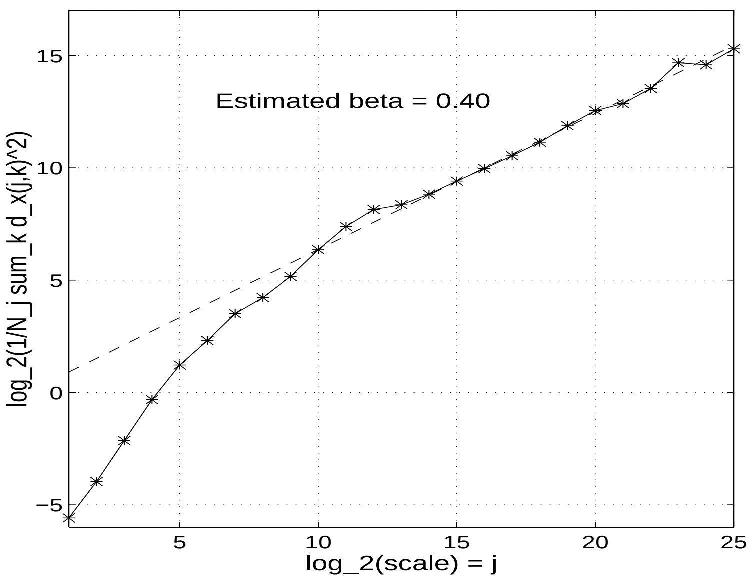

Figure 1: **Log-log plot of the sample variance of the wavelet coefficients vs scale.** LRD behaviour is seen in continuous time Ethernet data using a wavelet based formulation of the aggregation procedure. The asymptotic LRD behaviour is seen to enter at octave $j = 14$.

underlying continuous time process, resulting in a final data set which could be far larger than the original. Estimators such as the discrete Whittle would require an initial aggregation to be performed to reduce this size. Although aggregation in general is a valid approach as discussed in this paper, in this case it amounts to the selection of an arbitrary lower scale of observation, with a corresponding loss of information. The multiresolution framework however allows the entire signal to be processed, enabling the lowest scale where LRD manifests itself to be optimally chosen. From Figure 1 this is seen to at octave 14 for this data set. This is possible in terms of memory because, using wavelets, the data does not need to be processed in a single block, and in terms of time because the recursive filter-bank based algorithm allows a computational cost of the order of the length of the data. The wavelet-based estimator therefore not only possesses excellent statistical properties, but is also of simple practical use.

5 Conclusion

We have shown here that the aggregation procedure, which is a natural way of studying Long-Range Dependence or Self-Similarity, can be rephrased within the wavelet framework. This insight allows its generalisation first to "a-aggregation" and then to the quite different "d-aggregation". We have shown how aggregation of the first type cannot lead to good estimators of the LRD parameter. However d-aggregation allows an asymptotically unbiased and efficient estimator to be defined which is essentially unbiased in practice and very effective computationally. The key ingredients behind these properties are: -1- the wavelet basis is designed from the dilation operator, which by nature is scale invariant. It is therefore in some sense matched to the scale invariance phenomenon underlying LRD. -2- the wavelets are band-pass functions balancing the divergence of the spectrum of the data at frequencies close to 0 and therefore cancelling LRD in the wavelet coefficients. The resulting short range dependence in time of the wavelet coefficients enables efficient variance estimation.

References

[1] P.Abry, A.Aldroubi, Designing Multiresolution Analysis-type Wavelets and their Fast Algorithms. *Journal of Fourier Analysis and Applications*, Vol. **2**, N. 2 (1995), pp.135–159.

[2] P.Abry, P.Gonçalvès and P.Flandrin, Wavelets, Spectrum estimation, $1/f$ processes. Wavelets and Statistics, *Lectures Note in Statistics*, Vol. **103**, (1995), pp.15–30.

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

[4] P.Abry, D.Veitch, Wavelet Analysis of Long Range dependent Traffic. submitted to *Trans. Info. Theory*, (1996).

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

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

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

[8] F. Brichet, J. Roberts, A. Simonian and D. Veitch - *Heavy Traffic Analysis of a Storage Model with Long Range Dependent On/Off Sources*, Queueing Systems **23** 197–215, 1996.

[9] S. Cambanis and C. Houdré, On the continuous wavelet transform of second-order random processes. In *IEEE Trans. on Info. Theory*, Vol. 41, No. 3, May 1995, pp. 628–642.

[10] D.R.Cox, Long-range dependence: a review. In *Statistics: An Appraisal*, H.A.David and H.T.David, Editors, Iowa State University Press, (1984), pp.55–74.

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

[12] A.Erramilli, O.Narayan, W. Willinger, Experimental Queueing Analysis with Long-Range Dependent Packet Traffic, IEEE/ACM Transactions on Networking, Vol. 2, April (1996).

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

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

[15] E. Masry, The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. In *IEEE Trans. on Info. Theory*, Vol. 39, No. 1, January 1993, pp. 260–264.

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

[17] P.M.Robinson, Semi-parametric analysis of long-memory time-series. The Annals of Statistics, Vol. 22, No. 1, (1994), pp.515–539.

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

[19] M.S.Taqqu, private communication, 1996.

[20] G. Samorodnitsky, M.S.Taqqu, Stable Non-Gaussian random Processes, Stochastic models with Infinite Variance. Chapman - Hall. 1994.

[21] M.Unser, A. Aldroubi, M. Eden, "A family of polynomial spline wavelet transforms", Signal Processing, vol. 30, (1993), pp. 141–162.

[22] D.Veitch, P.Abry, Wavelet-based joint estimation for the Long-range dependence parameters. Preprint, (1996).