Self-similarity and long-range dependence through the wavelet lens *†‡

P. Abry[1], P. Flandrin[1], M. S. Taqqu[2] and D. Veitch[3]

January 25, 2000

(1) CNRS UMR 5672, 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 or (+33) 4 72 72 81 60 - Fax: (+33) 4 72 72 80 80
Patrice.Abry@ens-lyon.fr Patrick.Flandrin@ens-lyon.fr
http://www.ens-lyon.fr/PHYSIQUE/Signal

(2) Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215-2411, USA
tel: (+1) (617) 353-3022 - Fax: (+1) (617) 353-8100
murad@math.bu.edu - http://math.bu.edu/people/murad

(3) Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia
tel: (+613) 9344 6791
d.veitch@ee.mu.oz.au - http://www.serc.rmit.edu.au/~darryl

Abstract

Self-similar and long-range dependent processes are the most important kinds of random processes possessing scale invariance. We describe how to analyze them using the discrete wavelet transform. We have chosen a didactic approach, useful to practitioners. Focusing on the Discrete Wavelet Transform, we describe the nature of the wavelet coefficients and their statistical properties. Pitfalls in understanding and key features are highlighted and we sketch some proofs to provide additional insight. The Logscale Diagram is introduced as a natural means to study scaling data and we show how it can be used to obtain unbiased semi-parametric estimates of the scaling exponent. We then focus on the case of long-range dependence and address the problem of defining a lower cutoff scale corresponding to where scaling starts. We also discuss some related problems arising from the application of wavelet analysis to discrete time series. Numerical examples using many discrete time models are then presented to show the quality of the wavelet-based estimator and how it compares with alternative ones. The examples include strong short range dependence, and non-Gaussian series with both finite and infinite variance. We conclude with a historical and synthetic section discussing the close and natural connection between wavelets and scaling.

*We would like to thank our student Gershon Gacs who worked on the data analysis. This work has been partially supported by the CNRS grant TL97035, Programme Télécommunications, the Bede Morris French Embassy Fellowship 1999, and by the NSF grant ANI-9805623 at Boston University

†*AMS Subject classification.* Primary 60G18; secondary 62G07.

‡*Keywords and phrases:* Self-similarity, long-range dependence, wavelet, scaling parameter estimation, discrete time time series, fractional Brownian motion, fractional Gaussian noise, ARFIMA, non Gaussian innovations.

1 Motivation

In the past twenty years, a very wide collection of mathematical processes and systems, as well as data from physical systems in diverse fields covering both natural phenomena (biology: DNA sequences, heart rate variability, auditory nerves spike trains; physics: turbulence, hydrology, solid-state physics) and human activity (telecommunications network traffic, finance) have been seen to exhibit properties of *scale invariance*. Roughly speaking, scale invariance means that, within a wide range of scales between some upper and lower cut-off limits, no characteristic scale can be identified nor plays a privileged role. On the contrary, the behavior at different scales is in a sense equivalent, and they can be related to each other by a scale invariance property or renormalization operation. In other words, the paradigm of scale invariance is "relation between scales" rather than "dominance of a characteristic scale". Scale invariance has been shown to manifest in multiple forms of which the most important are "self-similarity" (two subsets of a whole observed at different scales are identical, either in themselves or statistically) and "long-range dependence" (very slowly decaying correlations allow the far past to influence the future, rather than being independent of it beyond some relaxation time). Other models exhibiting scale invariance, such as *fractal*, *multifractal* and $1/f$ processes, and *multiplicative cascades*, will not be addressed here, see Abry, Flandrin, Taqqu and Veitch (1998) for a review.

To study scale invariance in data, that is to evidence its presence, and to estimate its relevant parameters, wavelet analysis and wavelet transforms have been shown to be tools of particular interest. The aim of this chapter is to provide a brief guided tour of the wavelet analysis of the self-similarity and long-range dependence phenomena. Since it also intends to give practical operational tools to practitioners who are not familiar with wavelets, it starts with a didactic introduction to wavelet techniques in section 2. The following section, section 3, gathers results regarding the statistical properties of the wavelet coefficients of self-similar and long-range dependent (LRD) processes. The results are presented in a synthetic and constructive manner, with only qualitative outlines of some of the proofs, with accessible rather than original references. Section 4 introduces the idea of a wavelet based 'spectrum', as both a natural way of viewing scaling data through the wavelet lens, and as a natural basis for estimation. It then describes in detail the theoretical and practical issues in the estimation of the scaling exponent, the key parameter of scaling phenomena. A more detailed treatment is then given for LRD processes and data, including the special issues associated with the analysis of discrete time data. The section concludes with a theoretical and empirical analysis of the quality of estimation using the wavelet method in the case of LRD, with an emphasis on the bias-variance trade off arising from the necessity to select, indeed to define, a lower cutoff scale where the long-range dependence 'begins'. Finally, section 5 offers historical and thematic perspectives on the development of the topic "wavelet and scaling".

Note that throughout this paper the following convention is used: $f(x) \sim g(x)$ as $x \to a$ means that $\lim_{x\to a} f(x)/g(x) = 1$, and $f(x) \approx g(x)$ as $x \to a$ means that $\lim_{x\to a} f(x)/g(x) = C$ where C is some finite constant. Moreover, unless specified, $\sum_n$ will stand for $\sum_{n=-\infty}^{+\infty}$.

2 The discrete wavelet transform : a didactic introduction

A thorough presentation of the (various) wavelet transforms is not intended here, instead, we choose to propose a brief, constructive, didactic and operational introduction to the discrete wavelet transform (DWT) and to the corresponding underlying mathematical theory, the multiresolution analysis (MRA). For more complete introductions, the reader is referred to, e.g., Daubechies (1992), Mallat (1998).

2.1 Wavelet

A *wavelet* is a function $\psi(t),\ t \in \mathbb{R}$ such that

$$\int_{\mathbb{R}} \psi(t)dt = 0 \ , \tag{1}$$

which also satisfies some integrability conditions, for instance $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Typically, one requires in addition that the wavelet be bounded, centered around the origin, and have time support (resp., frequency support) that is either finite or decreases very fast as $|t| \to \infty$ (resp., $|\nu| \to \infty$). Time and frequency concentrations are restricted by the Gabor- Heisenberg uncertainty principle (see e.g. Mallat (1998, page 33)), and ideally, one would like the joint time AND frequency concentration of ψ to be as close as possible to the theoretical lower bound. Intuitively, the term "wavelet" is reminiscent of these requirements: an oscillating nature (as imposed by (1)), with short duration of the time support.

A wavelet ψ is said to have N *zero moments* (also called *vanishing moments*) if

$$\int_{\mathbb{R}} t^k \psi(t)dt = 0, \quad k = 0, 1, \ldots, N-1. \tag{2}$$

In view of the fundamental condition (1), the integer N is always at least equal to 1. Commonly, an increase in the number of vanishing moments comes with an enlargement of the time support, but brings more regularity (smoothness, continuity, derivability) and an improved concentration of the spectral content around a given frequency ν_0.

Typical examples of wavelets include the derivatives $\psi(t) = (d^n/dt^n)((2\pi)^{-1/2}e^{-t^2/2})$ of the standard normal density, the so-called *Haar wavelet*

$$\psi(t) = \begin{cases} 1 & 0 \le t < 1/2, \\ -1 & 1/2 \le t < 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

and the *Daubechies* wavelets or the *spline* wavelets, both constructed from a multiresolution analysis (see below). The *mexican hat wavelet* (2nd derivative of the normal function), for instance, has two vanishing moments and time and frequency supports that are, strictly speaking, infinite, but are actually very well concentrated. The Daubechies wavelets constitute a family of wavelets which is indexed by their number of vanishing moments, and which gives rise to an orthonormal wavelet basis. The Haar wavelet is the special element of both the Daubechies and the spline families that has one vanishing moment, and both families have strictly finite time supports. The so-called *poor man's wavelet* $\psi(t) = \delta(t) - \delta(t-\tau)$ is also sometimes used though it has poor regularity and frequency resolution. In this case the wavelet coefficients (as defined below) are equivalent to the increments of the process.

2.2 Wavelet coefficients

The functions

$$\psi_{j,k}(t) = \frac{1}{2^{j/2}}\psi(2^{-j}t - k) = 2^{-j/2}\psi(2^{-j}(t - 2^j k)), \ \ j \in \mathbb{Z}, \ k \in \mathbb{Z}, \tag{4}$$

are "dilations" and "translations" of ψ. The function $\psi(2^{-1}t)$, for example, is a dilation of ψ by a factor 2, the function $\psi(t-k)$ is the translation of ψ to the right by k units, and thus if ψ has support on $[0,1]$ then, for example, $\psi(2^{-1}t-3)=\psi(2^{-1}(t-6))$ has support on the interval $[6,8]$. The factors 2^j and j are called respectively the *scale* and the *octave*. Positive values of j correspond to dilations and negative values correspond to contractions. The normalization factor $2^{j/2}$ ensures that for all $j\in\mathbb{Z}$ and $k\in\mathbb{Z}$,

$$\int_{\mathbb{R}} \psi_{j,k}^2(t)dt = \int_{\mathbb{R}} \psi^2(t)dt,$$

that is, it allows the $L^2(\mathbb{R})$ norm to be preserved.

Using the functions $\{\psi_{j,k},\ j,k\in\mathbb{Z}\}$ in (4) as a set of filters, we can now define the *discrete wavelet transform* DWT of a function (or of the sample path of a stochastic process) $\{X(t),\ t\in\mathbb{R}\}$ as

$$d_{j,k} = \int_{\mathbb{R}} X(t)\psi_{j,k}(t)dt, \quad j,k\in\mathbb{Z}\,. \tag{5}$$

The coefficients $d_{j,k}$ are called *details*, as they encode an information differential between adjacent scales centered about scale 2^j, and the time instant 2^jk. This 'detail' nature of the details is described further below, and traces back to the band-pass nature of ψ (cf. equation (1)). To the scale 2^j can be associated a frequency $2^{-j}\nu_0$, where ν_0 is a central frequency characteristic of the wavelet ψ. The details are sometimes written $d_X(j,k)=d_{j,k}$ to emphasize that they correspond to X.

The adjective "discrete" in DWT notwithstanding, X is a function of the continuous time parameter t and the relation (5) involves an integration. The adjective "discrete" refers to the fact that the indices j,k take discrete values, in contrast to the *continuous wavelet transform*, where they take real values.

Remarks on conventions: When reading a paper or book on wavelets, we suggest that the reader immediately checks the conventions that the author makes about N, the wavelet normalization factor and the sign attached to j. This will avoid many pitfalls. Here are the typical alternative conventions:

- $\int_{\mathbb{R}} t\psi(t)dt$ is called the first moment, which means that our N becomes $N+1$.
- Instead of (4), one sets $\psi_{j,k}(t)=2^{-j}\psi(2^{-j}t-k)$. In this case $\int_{\mathbb{R}}|\psi_{j,k}(t)|dt=\int_{\mathbb{R}}|\psi(t)|dt$, and hence it is the $L^1(\mathbb{R})$ norm which is preserved.
- Instead of (4), one sets $\psi_{j,k}(t)=2^{j/2}\psi(2^jt-k)$, so that positive values of j now correspond to contractions.

2.3 Multiresolution analysis

There exists a specific class of wavelets, called the multiresolution wavelets, which is of particular interest. The analytic form of such wavelets is not usually known, however as explained below this is not an impediment, and in return one has that i) their key properties, such as the number of zero moments, regularity, time or frequency support, can be easily and flexibly tuned, and ii) fast pyramidal algorithms to compute the wavelet coefficients are available. The construction of such wavelets, of which the Daubechies and spline families are famous examples, is integral to the so-called *multiresolution analysis theory* (see Daubechies (1992), (1998)). We do not intend to detail the whole theory here but will state the key facts:

- the wavelet ψ (the *mother wavelet*) is defined through a *scaling function*, ϕ;

- both ϕ and ψ satisfy so-called two-scale equations:

$$\begin{array}{rcl} \phi(t/2) & = & \sqrt{2}\sum_n u_n \phi(t-n) \\ \psi(t/2) & = & \sqrt{2}\sum_n v_n \phi(t-n) \ . \end{array} \tag{6}$$

  For instance, for the Haar wavelet, the scale function $\phi(t)$ is the indicator function of the interval $[0,1]$, $u_0 = u_1 = 1/\sqrt{2}$ and $u_n = 0$ otherwise, $-v_1 = v_0 = 1/\sqrt{2}$ and $v_n = 0$ otherwise;

- the sequences u and v are said to generate the multiresolution analysis. In particular, they generate ϕ and ψ and control all their properties, notably the number of zero moments, regularity, and joint time frequency resolution;

- in addition to the wavelet coefficients $d_{j,k}$ in (5), one defines the *approximation coefficients* $a_{j,k}$:

$$a_{j,k} = \int_{\mathbb{R}} X(t)\phi_{j,k}(t)dt, \ j \in \mathbb{Z}, \, k \in \mathbb{Z}\,, \tag{7}$$

  where $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k)$. The relation (7) is similar to (5) but where ψ is replaced by ϕ.

A major consequence of the two-scale equations is the fact that both the approximation and wavelet coefficients $a_{j,k}$ and $d_{j,k}$ at scale j can be computed from the approximation sequence $\{a_{j-1,k}, k \in \mathbb{Z}\}$ at the finer resolution $j-1$:

$$\begin{array}{rcl} a_{j,k} & = & \int_{\mathbb{R}} X(t) 2^{-j/2}\phi(2^{-j}t - k)dt \\ & = & \int_{\mathbb{R}} X(t) 2^{-j/2}\sqrt{2}\sum_n u_n \phi(2(2^{-j}t - k) - n)dt \\ & = & \sum_n u_n \int_{\mathbb{R}} X(t) 2^{(-j+1)/2}\phi(2^{-j+1}t - 2k - n)dt \\ & = & \sum_n u_n a_{j-1,2k+n} \\ & = & \sum_n u_n^{\vee} a_{j-1,2k-n} \\ & = & (u_{\cdot}^{\vee} * a_{j-1,\cdot})(2k) \end{array} \tag{8}$$

and, in a similar way,

$$d_{j,k} = (v_{\cdot}^{\vee} * a_{j-1,\cdot})(2k) \tag{9}$$

where $*$ stands for the discrete time convolution operation, that is $(x_{\cdot} * y_{\cdot})(k) = \sum_n x(n)y(k-n)$, $u_n^{\vee} = u_{-n}$, and similarly $v_n^{\vee} = v_{-n}$.

The two relations (8) and (9) can be rewritten using the decimation operator $\downarrow_2$ ($y = \downarrow_2 x$ means that $y_k = x_{2k}$, i.e., that one sample in x out of two is dropped):

$$\begin{array}{rcl} a_{j,k} & = & \left[\downarrow_2 (u_{\cdot}^{\vee} * a_{j-1,\cdot})\right](k) \\ d_{j,k} & = & \left[\downarrow_2 (v_{\cdot}^{\vee} * a_{j-1,\cdot})\right](k) \ . \end{array} \tag{10}$$

Though they may appear to be the result of trivial calculations, these equations yield the following fundamental interpretation, central to the intuition of the multiresolution analysis: an approximation $a_{j,k}$ of the data at octave j is split into a coarser approximation $a_{j+1,k}$ and a detail $d_{j+1,k}$. This procedure is recursively repeated on each approximation $a_{j,k}$ so that the initial sequence $\{a_{0,k}, \, k \in \mathbb{Z}\}$ is rewritten as a collection of (increasingly coarse) detail sequences $\{\,\{d_{j,k}, k \in \mathbb{Z}\}, \, j = 1, \ldots, J\}$ and a final most coarse approximation $\{a_{J,k}, \, k \in \mathbb{Z}\}$.

The equations (10) call for the following comments.

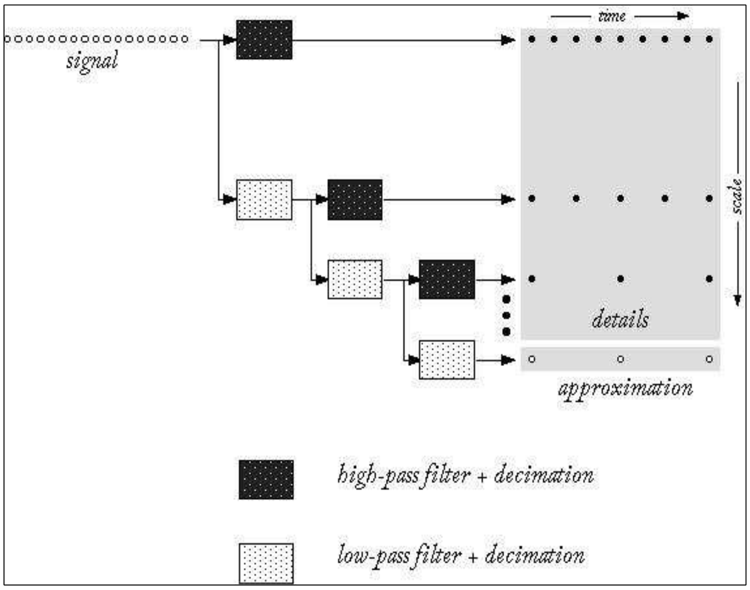

Figure 1: **Fast pyramidal filter-bank based wavelet decomposition algorithm.** The continuous signal X is first reduced to an approximation $a_{0,k}$ (at scale 0). Continuous time calculus is only involved in this initialization step, where $a_{0,k}$ is obtained from X. The "signal" illustrated in the figure is in fact $a_{0,k}$. In a multiresolution analysis, this approximation $a_{0,k}$ is decomposed into a detail $d_{1,k}$ and a further approximation $a_{1,k}$ which now has scale $j = 1$. The approximation $a_{1,k}$ is then decomposed into a detail and an approximation, and this procedure is repeated. Each step increases the scale of the approximation by 1, and in so doing, reduces the high frequencies. The details $d_{j,k}$ and the approximations $a_{j,k}$ can be computed iteratively from $a_{0,k}$ using discrete time convolutions and decimations. The impulse responses of the filters depend on the generating sequences u_{-n} and v_{-n}. The filter u_{-n} which creates the rougher approximations is a low-pass filter because it tends to preserve the low frequencies, whereas the filter v_{-n} which generates the details is a high-pass filter which tends to preserve the high frequencies.

- The computation of the $d_{j,k}$ can be performed using a recursive pyramidal discrete-time filter bank based algorithm, sketched in figure 1. Its elementary block consists of two discrete-time filters with impulse responses $u^\vee$ and $v^\vee$, and decimation operators. Because the $d_{j,k}$ are not computed from the original continuous time process X, but from the discrete-time previous approximation $a_{j-1,k}$, the resulting algorithm has a particularly low computational cost, of the order of $O(n)$ to compute n wavelet coefficients (see, for instance, Vetterli and Kovacevic (1995)).
- The only stage where continuous time calculus is involved is the computation of the initial sequence $a_{0,k}$ which requires a continuous time inner product, the integral

$$a_{0,k} = \int_{\mathbb{R}} X(t)\phi(t-k)dt\,. \tag{11}$$

The remainder of the calculus involves only discrete time convolutions. The derivation of this initial sequence is known as the initialization step and might be difficult because of its continuous time nature. A classic example which arises in practice is when only a discrete sampling of the continuous time sample path is available, rather than the entire path. Specific comments on this step in the case of long-range dependence in discrete time series will be made later in the text.

- The sequences $u_{\cdot}^{\vee}$ and $v_{\cdot}^{\vee}$ play the role of low-pass and high-pass filters respectively. The sequence $u_{\cdot}^{\vee}$ gives rise to a coarser approximation (letting the low frequencies go through), whereas the sequence $v_{\cdot}^{\vee}$ tends to preserve the details (high frequencies).

  From a more abstract perspective, these sequences allow a decomposition of the space $L^2(\mathbb{R})$ into nested subspaces,
$$L^2(\mathbb{R}) \supset \cdots \supset V_{-1} \supset V_0 \supset V_1 \supset \cdots$$
  For a given j, the $\phi_{j,k},\ k \in \mathbb{Z}$, project a signal $X(t)$ on the j^{th} "approximation space" V_j and $\psi_{j,k},\ k \in \mathbb{Z}$, on the "detail space" $V_{j-1} \ominus V_j$.

3 Wavelet coefficients of self-similar and LRD processes

The aim of this section is to describe the statistical properties of the wavelet coefficients of self-similar and long-range dependent (LRD) processes, to underline to what extent they themselves exhibit self-similarity and LRD, and thus how the scale invariance may be studied in the wavelet domain. Let $\{X(t), t \in \mathbb{R}\}$ denote a stochastic process and $d_{j,k}$ its wavelet coefficients. If the wavelet ψ decays in time sufficiently fast and satisfies other mild conditions, whose exact formulation depends on details of the process in question, the wavelet coefficients are well-defined random variables. Let us now review some more precise results.

• The wavelet coefficients $d_{j,k}$ are the same for $X(t)$ and for $X(t) + P(t)$ where P is a polynomial of degree $N-1$, when ψ has N vanishing moments. Indeed,

$$\int_{\mathbb{R}} P(t)\psi_{j,k}(t)dt = 2^{j/2} \int_{\mathbb{R}} P(2^j(u+k))\psi(u)du = 0\,,$$

since $P(2^j(u+k))$ is a polynomial in u of degree $N-1$. Hence, the DWT is blind to polynomial trends. In particular, $X(t) + \mu$ and $X(t)$ share the same $d_{j,k}$. If P is a smooth function which can be well approximated by a polynomial, then its effect will also be small, and often negligible compared to that of X (Abry and Veitch (1998)). This can be useful in cases where P is a nuisance function, for example, in economics, where it may correspond to a change of threshold.

• Suppose that the process $\{X(t),\ t \in \mathbb{R}\}$ has stationary increments, i.e., the finite-dimensional probabilities of $\{X(t+h) - X(t), t \in \mathbb{R}\}$ do not depend on t. Then, for fixed $j \in \mathbb{Z}$, $\{d_{j,k},\ k \in \mathbb{Z}\}$ is a stationary sequence. It is easy to see that the one dimensional marginal of the sequence does not depend on k. Take $j = 0$ for ease of notation, then

$$\begin{aligned}
d_{0,k+k_0} &= \int_{\mathbb{R}} X(t)\psi(t-k-k_0)dt \\
&= \int_{\mathbb{R}} X(u+k_0)\psi(u-k)du \\
&= \int_{\mathbb{R}} [X(u+k_0) - X(k_0)]\psi(u-k)du \qquad (12)\\
&\overset{d}{=} \int_{\mathbb{R}} [X(u) - X(0)]\psi(u-k)du \qquad (13)
\end{aligned}$$

$$= \int_{\mathbb{R}} X(u)\psi(u-k)du = d_{0,k}, \tag{14}$$

The step (12) makes fundamental use of the wavelet property $\int_{\mathbb{R}} \psi(u)du = 0$ (cf. (2)). The step (13) is justified by approximating the integral by sums and using the stationarity of the increments (see e.g. Houdré (1993)). Similarly, $\sum_{i=1}^{n} \theta_i d_{0,k_i+k_0} \stackrel{d}{=} \sum_{i=1}^{n} \theta_i d_{0,k_i}$ for arbitrary integer $n \geq 1$ and real constants θ_i, which implies that the characteristic functions are equal and hence that n-dimensional distributions do not depend on k_0, that is, $\{d_{j,k},\ k \in \mathbb{Z}\}$ is stationary in k.

In a similar way, it can be shown that if the wavelet ψ has N zero moments then $\{d_X(j,k),\ k \in \mathbb{Z}\}$ is stationary if the increments of order N of X are stationary. One also has that $\{d_X(j,k),\ k \in \mathbb{Z}\}$ is stationary if $\{X(t),\ t \in \mathbb{R}\}$ is itself stationary.

• Let us suppose that $\{X(t),\ t \in \mathbb{R}\}$ is a self similar process, with self-similarity parameter H (henceforth, H-ss), that is $\forall c > 0, \{c^{-H}X(ct), t \in \mathbb{R}\} \stackrel{d}{=} \{X(t), t \in \mathbb{R}\}$. Then for fixed $j \in \mathbb{Z}$,

$$d_{j,k} \stackrel{d}{=} 2^{j(H+1/2)} d_{0,k} \tag{15}$$

as a process in $k \in \mathbb{Z}$. This can be roughly argued as follows: After a change of variables,

$$d_{j,k} = \int X(2^j u)2^{-j/2}\psi(u-k)2^j\, du \stackrel{d}{=} 2^{j(H+1/2)} d_{0,k} \tag{16}$$

since $X(2^j u) \stackrel{d}{=} 2^{jH}X(u)$. In (15), the factor of j in the exponent is $H + 1/2$. The 1/2 would not have been there, had we used the $L^1(\mathbb{R})$ normalization for the wavelet, namely $\psi_{j,k}(t) = 2^{-j}\psi(2^{-j}t - k)$.

• Let us assume now that $X(t)$ is a self-similar process with stationary increments (henceforth, H-sssi), $0 < H < 1$, with mean zero and finite variance, for example, the fractional Brownian motion. Then $\mathbb{E}d_{j,k} = 0$, and it follows immediately from (14) and (16) that

$$\mathbb{E}d_{j,k}^2 = C2^{j(2H+1)} \tag{17}$$

where $C = \mathbb{E}d_{0,0}^2$. Taking the logarithm in both sides of (17) yields a linear function of j whose slope is $2H + 1$. This observation is the basis of the wavelet estimation method for H, which is described in the sequel.

Beyond the variance, the covariance structure of the wavelet coefficients of a H-sssi process can be studied. For a given j, the $\{d_{j,k},\ k \in \mathbb{Z}\}$ are correlated but this correlation tends rapidly to 0 with large lag if N is sufficiently large:

$$\mathbb{E}d_{j,k_1}d_{j,k_2} \leq C|k_1 - k_2|^{2(H-N)} \tag{18}$$

where the constant C depends on j. Relation (18) is best regarded as asymptotic, as the bound can be quite large for small $|k_1 - k_2|$, especially if N is large. In view of (18), to avoid long-range dependence for the $d_{j,k}$, i.e., to ensure that $\sum_{k=0}^{\infty} \mathbb{E}|d_{j,k}d_{j,0}| < \infty$, one needs to choose $N > H + 1/2$, that is, have at least $N = 2$. The Haar wavelet does not yield enough decorrelation.

Let us now turn to the Fourier domain. In this paper, we use normalized frequencies and define the Fourier transform $\widetilde{\psi}$ of a function ψ as

$$\widetilde{\psi}(\nu) = \int_{-\infty}^{\infty} e^{-i2\pi\nu t}\psi(t)dt. \tag{19}$$

We shall focus on the spectral density $f_j(\nu)$ of the stationary sequence $\{d_{j,k},\ k \in \mathbb{Z}\}$ for each j. When $X(t)$ is H-sssi with $\mathbb{E}X^2(1) = \sigma^2 < \infty$, one can obtain it in two steps. First consider the CWT (Continuous Wavelet Transform) $\{D(a,\tau),\ a,\ \tau \in \mathbb{R}\}$ defined by

$$D(a,\tau) = a^{-1/2} \int_{-\infty}^{\infty} X(t)\psi\left(\frac{t-\tau}{a}\right) dt.$$

Assume the following properties of the wavelet: fairly fast decay at infinity, $(1+t^2)\psi(t) \in L^1(\mathbb{R})$, controlled amplitude at low frequencies, $\sup_{|\nu|\leq\epsilon} |\widetilde{\psi}(\nu)/\nu| \leq M < \infty$ for some $\epsilon > 0$, and finally the fundamental relation (1). Then, as shown by Kato and Masry (1999), the spectral density of $\{D(a,\tau),\ \tau \in \mathbb{R}\}$ is

$$f_a^{(1)}(\nu) = (\sigma^2\Gamma(2H+1)\sin\pi H)\, a \frac{|\widetilde{\psi}(a\nu)|^2}{|2\pi\nu|^{2H+1}}$$

where Γ is the gamma function. Since $\{d_{j,k},\ k \in \mathbb{Z}\,\}$ is the sampled version of $D(j,\tau)$ at integer values of τ, its spectral density is

$$f_j^{(2)}(\nu) = (\sigma^2\Gamma(2H+1)\sin\pi H)\, 2^j \left\{ \frac{|\widetilde{\psi}(2^j\nu)|^2}{|2\pi\nu|^{2H+1}} + \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{|\widetilde{\psi}(2^j(\nu+m))|^2}{|2\pi(\nu+m)|^{2H+1}} \right\},\ -1/2 < \nu < 1/2. \quad (20)$$

Although $f_a^{(1)}(\nu)$ with $a = 2^j$ and $f_j^{(2)}(\nu)$ are different even for $-1/2 < \nu < 1/2$, $f_{2^j}^{(1)}(\nu) \sim f_j^{(2)}(\nu)$ as $\nu \to 0$, and they have essentially the same behavior for small ν if the Fourier transform $\widetilde{\psi}(\nu)$ is well-localized around 0. Since $\widetilde{\psi}(\nu)$ is $O(|\nu|^N)$ as $\nu \to 0$, we see that $f_j^{(2)}(\nu)$ is bounded as $\nu \to 0$ and hence each detail process has short range dependence if $N > H + 1/2$, in accordance with the comment following (18) above.

• Let $\{Y(t), t \in \mathbb{R}\}$ denote a second-order stationary stochastic process and r_Y and g_Y its covariance function and spectrum. Since the relation $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ implies $\widetilde{\psi}_{j,k}(\nu) = 2^{j/2}\widetilde{\psi}(2^j\nu)e^{-i2\pi k 2^j \nu}$, the covariance function of the wavelet coefficients of Y becomes

$$\begin{aligned} \mathbb{E}d_{j,k}d_{j',k'} &= \int_{\mathbb{R}}\int_{\mathbb{R}} r_Y(u)\psi_{j,k}(v)\psi_{j',k'}(u+v)dvdu \\ &= \int_{\mathbb{R}} g_Y(\nu)2^{j/2}2^{j'/2}\widetilde{\psi}(2^j\nu)\widetilde{\psi}^*(2^{j'}\nu)e^{-i2\pi(k2^j - k'2^{j'})\nu}d\nu, \end{aligned} \quad (21)$$

where $\widetilde{\psi}^*$ denotes the complex conjugate of $\widetilde{\psi}$. The variance of the wavelet coefficients is

$$\mathbb{E}d_{j,k}^2 = \int_{\mathbb{R}} g_Y(\nu)2^j|\widetilde{\psi}(2^j\nu)|^2 d\nu\,. \quad (22)$$

The second-order stationary stochastic process ss $\{Y(t), t \in \mathbb{R}\}$ is said to be long-range dependent if either

$$r_Y(\tau) \sim c_r|\tau|^{\gamma-1}, \tau \to +\infty,\ \gamma \in (0,1) \quad (23)$$

or if,

$$g_Y(\nu) \sim c_g|\nu|^{-\gamma}, \nu \to 0,\ \gamma \in (0,1)\,. \quad (24)$$

In most practical situations, r_Y is asymptotically monotone in which case these relations are in fact equivalent (Zygmund (1979), Chapter V 2).

There is a close relationship between LRD and self-similar processes. Indeed, the increments of any finite variance H-sssi process have LRD, as long as $1/2 < H < 1$, with H and γ related through

$$\gamma = 2H - 1. \tag{25}$$

For instance, the increments of the fractional Brownian motion define the so-called fractional Gaussian noise. Long range dependent processes may however very well exist by themselves, without reference to self-similarity, for example the FARIMA processes used below. Note moreover that most of the LRD processes commonly used are discrete time processes, while the arguments developed here are dedicated to continuous time processes. We will return to this question later.

From equation (22) above, we obtain that the variance of the wavelet coefficients reproduce the asymptotic power-law underlying the definition of the LRD:

$$\mathbb{E}d_{j,k}^2 \sim c_g 2^{j\gamma} \int_{\mathbb{R}} |\nu|^{-\gamma} |\widetilde{\psi}(\nu)|^2 d\nu, \ j \to +\infty. \tag{26}$$

In the LRD case, this (asymptotic) relation is the basis of the estimation of γ (or H, if $\gamma = 2H - 1$), as described below. One can also study the covariance structure of the wavelet coefficients. From equation (21), let us start with wavelet coefficients located at the same scale ($j' = j$),

$$\mathbb{E}d_{j,k}d_{j,k'} = \int_{\mathbb{R}} g_Y(\nu) 2^j |\widetilde{\psi}(2^j\nu)|^2 e^{-i2\pi(k-k')2^j\nu} d\nu \ . \tag{27}$$

It is clear that $\mathbb{E}d_{j,k}d_{j,k'}$ is a function of $k' - k$ only, whose asymptotic behavior ($k' - k \to +\infty$) is controlled by that of its Fourier transform $g_Y(\nu)2^j|\widetilde{\psi}(2^j\nu)|^2$ at the origin. Because the mother-wavelet has N zero moments, its Fourier transform behaves at the origin as $|\nu|^N$, provided that the Fourier transform is N times differentiable at the origin. Indeed, N zero moments translate into $\widetilde{\psi}^{(j)}(0) = 0$, $0 \le j \le N-1$, and by Taylor, $|\widetilde{\psi}(\nu)| \approx |\nu|^N |\widetilde{\psi}^{(N)}(0)|$. Therefore, the singular behavior $|\nu|^{-\gamma}$ of the long-range dependent spectrum at the origin is balanced by the regularity $|\nu|^{2N}$ contributed by the wavelet, and hence the stationary sequence $\{d_{j,k}, \ k \in \mathbb{Z}\}$ is not LRD as soon as $2N > \gamma$, i.e., provided that the number of vanishing moments is high enough. One then expects that the covariance of the detail process at octave j decreases asymptotically as

$$\mathbb{E}d_{j,k}d_{j,k'} \approx |k' - k|^{\gamma - 2N - 1}, \ |k' - k| \to +\infty\,, \tag{28}$$

which shows that the higher N the shorter the range of dependence. Relation (28) is consistent with (18) because, going from an H-sssi process $X(t)$ to its increments $Y(t) = X(t) - X(t-1)$, involves differencing the process, an operation which introduces in (20) a multiplicative factor of $|1 - e^{-i2\pi\nu}|^2 \sim (2\pi\nu)^2$ as $\nu \to 0$. The term $|\widetilde{\psi}(2^j\nu)|^2/|2\pi\nu|^{2H+1}$ in (20) becomes $|1 - e^{-i2\pi\nu}|^2 |\widetilde{\psi}(2^j\nu)|^2/|2\pi\nu|^{2H+1} \sim |\widetilde{\psi}(2^j\nu)|^2/|2\pi\nu|^{2(H-1)+1}$ as $\nu \to 0$, effectively changing the original H to $H - 1$. Then Y is LRD, $\gamma = 2H - 1$ and the relation $2N > \gamma$ holds even for $N = 1$, that is, for Haar wavelets. In practice, one chooses $N = 2$ or 3.

• The discussions above show that the wavelet coefficients of self-similar and long-range dependent processes share the same fundamental properties: i) stationarity at fixed scale, ii) short range statistical dependence (cf. equations (18) and (28)), and iii) reproduction in the wavelet domain of the power-laws defining the scale invariance phenomena (cf. equations (17) and (26)). Let us emphasize that these properties result from a deep correspondence between the analyzed phenomena — processes with scale invariance – and the analyzing tool — wavelet or multiresolution or multiscale analysis.

More precisely, they rely on the fact that $i)$ the mother wavelet has at least one zero moment, and $ii)$ that the wavelets are constructed from the mother wavelet using a dilation operator. One can add that these results extend to other types of processes with scale invariance or, in other words, without characteristic scales, such as $1/f^\gamma$ noises, fractal processes, multifractal processes and more recently even to stochastic multiplicative cascades (Abry, Flandrin, Taqqu and Veitch (1998)).

These key properties of the wavelet coefficients constitute the foundation for the detection, analysis, and estimation tools dedicated to scaling processes, as detailed in the next section.

4 Estimation

The core problem in the practical application of scaling models to data is the detection of a scaling phenomenon, followed by the estimation of its relevant exponent. In this section we discuss these issues, with a particular emphasis on the study of long range dependent processes in discrete time. We include a practical section where the wavelet estimator is used on a variety of discrete time models, allowing comparison with previous work using other methods. We begin by describing a natural way of viewing the underlying quantities which are to be estimated, followed by a description of their estimation, and leading finally to the estimation of the exponent itself.

4.1 Logscale Diagrams

Collectively, the variances of the detail processes $d_{j,\cdot}$ over all scales $\{2^j\}$, when such processes are stationary, are a second order description of the process X, which constitute a kind of 'wavelet spectrum'. It was seen in the previous section, notably in equation (26), that a recurring feature among scaling models is the power-law progression (within the scaling range) of these variances with scale. This motivates a logarithmic view of this spectrum, where a straight line over some range, if present, indicates scaling, with the slope giving the scaling exponent. Accordingly, we consider a plot of

$$s_j \equiv \log_2(\mathbb{E}d_{j,\cdot}^2) \tag{29}$$

against j, which we refer to as the *Exact Logscale Diagram* (exact LD). This 'wavelet spectrum' is a signature of the process in question (in logarithmic coordinates), which can be defined for example for stationary processes or those with stationary increments, independently of whether they are scaling or not. It is obtained by performing the integration (in practice numerically) defining the variances in question, for example equation (22) in the case of stationary processes with a known spectral density $g_Y(\nu)$ against j.

Three examples appear in the left plot in figure 2 for the LRD sequence standard fractional Gaussian noise (fGn) with $H = 0.8$, corresponding to the Daubechies wavelets with $N = 1$, $N = 3$ and $N = 6$ vanishing moments. It is not surprising that the spectra are not identical since each involves a different wavelet (the spectra are specific to the underlying family of analyzing functions, just as the traditional Fourier spectrum is tightly bound to the use of sinusoids). Note however that they are very similar in form (in fact in the case of Daubechies wavelets the spectra converge pointwise with increasing N), are all close to straight lines except at small scales, and share the same asymptotic slope $\gamma = 2H-1$ at large j which characterizes the LRD nature of fGn. This is an important property which is characteristic of the Logscale Diagrams of scaling processes in general: in the scaling regime the specific choice of wavelet ceases to matter.

Just as the Fourier spectrum can be estimated from data, so can the exact Logscale Diagram. In the right hand plot of figure 2 we again trace the $N = 3$ exact Logscale Diagram and superpose

5 sets of unbiased estimates $y_j = \hat{s_j}$ (discussed below), obtained from 5 synthetic samples paths each $n = 10000$ long. Gaussian 95% confidence intervals corresponding to the variability of the y_j (see discussion below and equation (33)) are shown as the vertical segments centered on the known values s_j. These estimated exact Logscale Diagrams, with confidence intervals, we call simply *Logscale Diagrams* (LDs). Of course in practice the true $\mathbb{E}y_j = s_j$ are not known, and the confidence intervals are drawn centered on the estimates y_j.

Note that it follows from the nature of the dilation operator generating the wavelet basis that the number n_j of detail coefficients available to be analyses at octave j halves with each increase in j (in practice the presence of border effects results in slightly lower values). Confidence intervals therefore increase monotonically with j as one moves to larger scales, as seen in the right plot in figure 2.

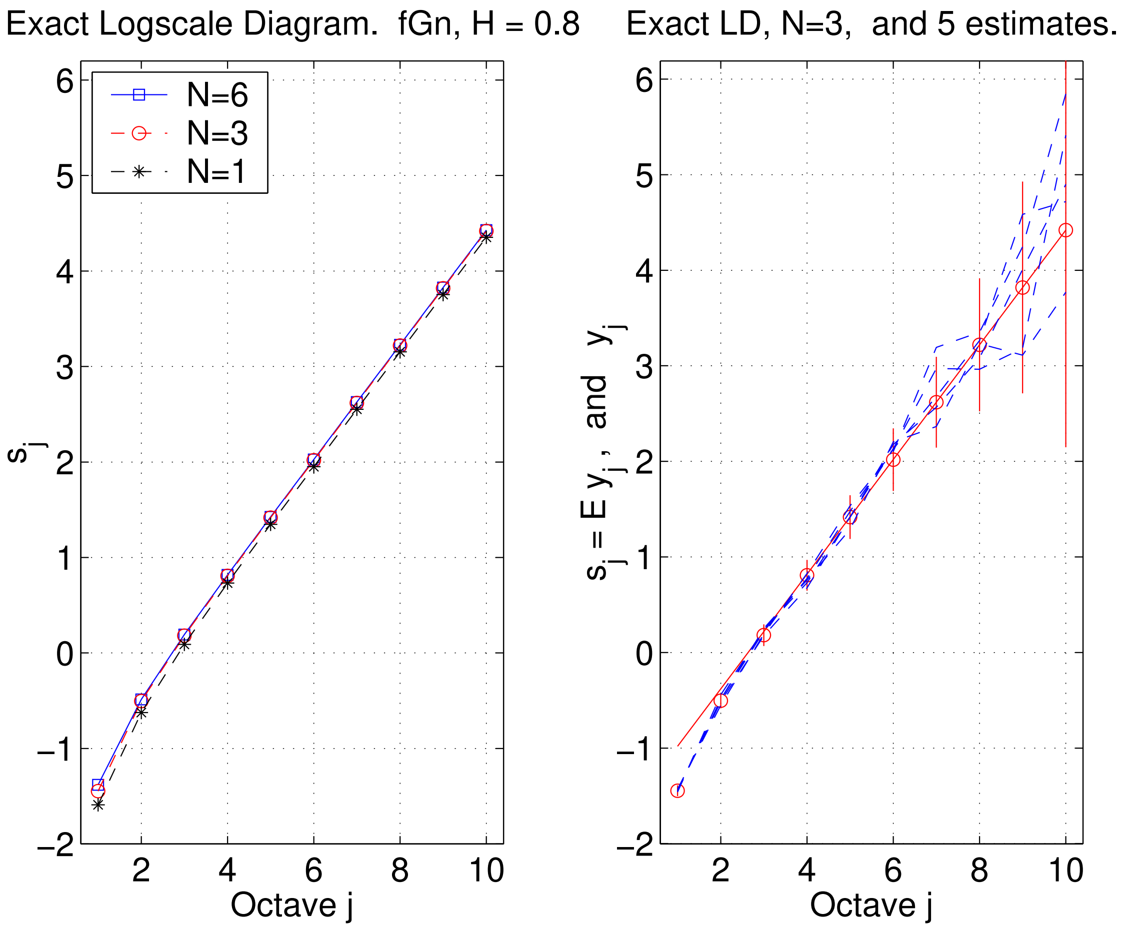

Figure 2: **Exact Logscale Diagrams for fGn,** $H = 0.8$**.** Left: the exact LD's are given for $N = 1$, 3, and 6. The 'wavelet spectrum' depends on the wavelet, however the asymptotic slope $\gamma = 0.6$ ($H = 0.8$) is invariant. Right: the exact LD is given for $N = 3$, together with confidence intervals corresponding to estimates at each scale based on $n = 10000$ points, and Logscale Diagrams superimposed from 5 sample paths. The solid line is the asymptotic slope, $\gamma = 0.6$, of the exact LD.

We now discuss the details of the estimators comprising the Logscale Diagram, and how from them an estimator can be designed for scaling exponents.

4.2 Estimation within the Logscale Diagram

We have already introduced the estimates y_j as random variables with the property $\mathbb{E}y_j = s_j$ by definition. To see how they can be constructed, first consider the following non-parametric, unbiased estimator of the variance of the detail process at octave j:

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_{j,k}|^2. \tag{30}$$

The logarithm of this variable would be an estimator of s_j, however it would be biased as the non-linearity of the logarithm means that $\mathbb{E}[\log(\cdot)] \neq \log(\mathbb{E}[\cdot])$. To cancel this we introduce a small bias correction term $g(j)$, and define the y_j as

$$y_j = \log_2(\mu_j) - g(j). \tag{31}$$

Denote the variance of y_j by σ_j^2, from which the confidence intervals in the LD can be calculated using a Gaussian approximation for y_j. Now let us assume that the *scaling range* $[j_1, j_2]$, where a scaling behavior is present, that is the exact LD is straight, has been correctly identified. The index j_1 is called the *lower-scale* or *small scale* or *high frequency* cutoff and the index j_2 is called the *higher-scale* or *large scale* or *low frequency* cutoff. The range $[j_1, j_2]$ could be detected in practice by looking for a straight line behavior (consistent with the confidence intervals), in the Logscale Diagram. An unbiased estimator of the scaling exponent with low variance can then be obtained through standard weighted linear regression over the y_j in this range. Specifically, defining first the quantities $S = \sum_{j=j_1}^{j_2} 1/\sigma_j^2$, $S_1 = \sum_{j=j_1}^{j_2} j/\sigma_j^2$ and $S_2 = \sum_{j=j_1}^{j_2} j^2/\sigma_j^2$, this estimator $\hat{\gamma}$ of γ is given by

$$\hat{\gamma} = \frac{\sum_{j=j_1}^{j_2} y_j (Sj - S_1)/\sigma_j^2}{SS_2 - S_1^2} \equiv \sum_{j=j_1}^{j_2} w_j y_j, \tag{32}$$

which is unbiased provided that the exact LD truly aligns over $[j_1, j_2]$.

We have yet to explain how the $g(j)$ or the σ_j^2 can be calculated. In general neither can be exactly, but the following expressions obtained by idealizing the weak correlations between the wavelet coefficients to complete decorrelation, are excellent approximations in the case where the $d_{j,k}$ are Gaussian:

$$\begin{aligned} g(j) &= \Gamma'(n_j/2)/(\Gamma(n_j/2)\ln 2) - \log_2(n_j/2) \sim \frac{-1}{n_j \ln 2}, \\ \sigma_j^2 &= \zeta(2, n_j/2)/\ln^2 2 \sim \frac{2}{n_j \ln^2 2} \end{aligned} \tag{33}$$

where Γ is the gamma function, Γ' its derivative, and $\zeta(2,z) = \sum_{n=0}^{\infty} 1/(z+n)^2$ a generalized Riemann Zeta function. The asymptotic large n_j formulae above hold true to a good approximation even for non-Gaussian data (see Veitch and Abry (1998), Abry, Flandrin, Taqqu, Veitch (1998) for details). At large n_j and again using the independence idealization, the variance of the estimator, even in the non-Gaussian case, can be given explicitly as

$$\mathrm{Var}(\hat{\gamma}) = \sum_{j=j_1}^{j_2} \sigma_j^2 w_j^2 \sim \frac{1}{n} \cdot \frac{(1 - 2^{-J})}{F} \tag{34}$$

where $F = F(j_1, J) = \ln^2 2 \; 2^{1-j_1} \cdot (1 - (J^2/2 + 2)2^{-J} + 2^{-2J})$, and $J = j_2 - j_1 + 1$ is the width of the scaling range.

A major practical advantage of the above methodology is that exactly the same procedure applies, the same weights and the same estimator $\hat{\gamma}$ for the slope, irrespective of the kind of scaling (H-sssi, LRD, etc.) in question. Even the detailed choice of wavelet is not important in the scaling regime, as mentioned above, provided that the number of vanishing moments of the wavelet is sufficiently high to ensure the quasi-decorrelation of the details. The first issue however in the practical application of scaling models to data is the *detection* of a scaling phenomenon, which involves the selection of the range $[j_1, j_2]$. If one has already decided for example that a H-ss model applies, then of course one would choose $j_1 = 1$ and j_2 the largest possible given the data length, and there is nothing to decide, but for other kinds of scaling behavior both upper and lower cutoffs exist which must be identified. We now explore this issue in more detail in the intermediate case of LRD processes, where j_2 is the largest possible by definition and does not pose a problem, but j_1 is not defined a priori and must be determined somehow during detection/measurement.

4.3 Estimation of LRD

A subtle but essential point is the fact that the selection of the lower cutoff scale j_1 for LRD is not merely a practical and estimation question, but involves deeper issues: the cutoff is not defined even within the exact Logscale Diagram. This is a simple consequence of the asymptotic definition of LRD, where there is no given frequency at which the scaling regime starts. To address this problem we must add a new definition. There is no 'real' cutoff which relates to the spectrum $g_X(\nu)$ or alternatively to an exact AD of the process *only*. The cutoff has meaning only with reference to some other criterion, which here we choose to be *estimation quality*. Specifically, we apply our estimation approach for each j_1, and *define* the 'true' $j_1 = j_1^{\mathrm{MSE}}$ as the one corresponding to the minimum mean square error (MSE), where the squared error is defined as

$$\mathrm{MSE}(\hat{\gamma}) \equiv \mathbb{E}(\hat{\gamma} - \gamma)^2 = (\mathbb{E}\hat{\gamma} - \gamma)^2 + \mathrm{Var}(\hat{\gamma}). \tag{35}$$

Intuitively it is clear that, at fixed n, as we move to lower frequencies with increasing j_1 the asymptotic straight line behavior is less and less polluted by departures at small scales, and so the estimation bias decreases, at the expense of an increase in variance due to the effective loss of data. An MSE criterion allows the tradeoff between these two effects to be economically expressed, as illustrated in the right plot in figure 3 for a particular model with strong SRD. If we now consider increasing n, the balance shifts toward larger j_1 values as the variances of the y_j all decrease with n, whereas the bias at each j, which depends on the exact LD only, remains constant. Thus $j_1^{\mathrm{MSE}}(n)$, far from being an absolute constant for a process indicating where the LRD 'truly' begins, is in fact a non-decreasing function of n.

We could also describe this choice of cutoff in another way, as being the crossover point between the SRD and LRD of the process. It is interesting to note that even for the canonical LRD process, the fGn, the scale invariance in this sense does not extend to all scales, and in fact for $N = 3$ and $n = 10000$ the MSE defined cutoff is $j_1^{MSE} = 3$, which is consistent with the deviation from the asymptotic slope in relation to the confidence intervals, seen in the right plot in figure 2.

4.4 Analysis of discrete data

Another issue which is often overlooked is the fact that the wavelet analysis, including the Discrete Wavelet Transform, is defined for *continuous* time processes, and not for discrete time series such as

FARIMA processes and discrete fGn. This is significant as time series models in general, including some of the most popular LRD models, and much real data, fall into this category. Recall from section 2, equation (11), that to initialize the DWT the initial approximation sequence $a_{0,k}$ must be obtained via a sequence of integrals in continuous time, a procedure that has no meaning for discrete data. Typically what is done if one wants to study a discrete process $X(k)$ is to set $a_{0,k} = X(k)$, an ad hoc procedure which makes it unclear what is in fact being studied, and introduces errors. Recent work (Veitch, Taqqu, Abry (2000)) has shown how to use the DWT in a well defined way for the study of second order properties of discrete wide sense stationary processes, such as those we consider here. It offers a simple and effective way to allow the advantages of wavelet analysis to be brought to bear on intrinsically discrete series.

The approach is conceptually very simple and can be expressed as follows. To a given discrete series $\{X(k), k \in \mathbb{Z}\}$ we associate a closely related continuous-time process $\{\widetilde{X}(t), t \in \mathbb{R}\}$. In fact the process $\widetilde{X}(t)$ is chosen such that the spectral densities g_X of X and $g_{\widetilde{X}}$ of $\widetilde{X}$ coincide on the principle interval $\nu \in (-1/2, 1/2]$. The series $\widetilde{X}(t)$, being defined in continuous time, can be studied using the DWT as usual. Since the spectra are equal in $(-1/2, 1/2]$, any conclusion on the spectral density of $\widetilde{X}$ in this range, for example estimates of γ in the case of LRD, holds automatically for the spectral density of X. The only difficulty is the initialization phase of the MRA which begins the DWT analysis of $\widetilde{X}(t)$. It can be shown that the two steps of first converting from X to $\widetilde{X}$, and then initializing the MRA for $\widetilde{X}$, can be accomplished in a single linear discrete filtering (convolution) operation: $a^{\widetilde{X}}_{0,k} = (X * I)(k)$, where I is a discrete filter given by $I(m) = \int_{-\infty}^{\infty} \sin(\pi(t+m))\phi_0(t)/\pi(t+m)\,dt$, which depends only on the wavelet used (I has infinite support but in practice can be taken to be quite small, say 100 long or less). The prefiltering is easy to perform, and corrects errors which would otherwise have been very significant on the first two octaves, but negligible at coarse scales. In this paper, for example in figure 2 which analyses discrete fGn, this filtering has been performed, and the y_j values are therefore valid at all octaves.

4.5 Estimation examples, LRD and SRD

In this subsection we present examples of the wavelet estimation of γ for several discrete time models. These include fGn, and a variety of low order Fractional Autoregressive Integrated Moving Average (FARIMA) models. We review FARIMA models rapidly below. The examples are divided into three main categories, with corresponding tables of results.

In the first category, table 1, are time series with Gaussian innovations, for which both the exact LD's, calculated through accurate numerical integration of equation (22), and the j_1^{MSE}, are known. All results in this table can therefore be taken as exact. Within the table two values of j_1 are used, the optimal j_1^{MSE} on the left which shows the best result available in the sense described above, and a fixed $j_1 = 5$ on the right, which enables a comparison with previous work in Taqqu, Teverovsky and Willinger (1995), Taqqu and Teverovsky (1996), (1998), (1997). In the second table, table 2, the same Gaussian models are studied, but this time the results are estimated through averages of 50 realizations generated using the Splus software. This allows an idea of the accuracy to be expected in estimation using 50 realizations. This is useful when considering table 3, which deals with non-Gaussian models, and even models with infinite variance, for which the theoretical performance of the estimator is not exactly known, but again has been estimated.

All time series are of length $n = 10,000$. The realizations we use are exactly the same as those in Taqqu and Teverovsky (1998) and Taqqu, Teverovsky and Willinger (1995), where other estimators were studied. The reader should examine the tables in these papers with those below in order to compare the effectiveness of the different estimators[1]. We give a brief comparison below.

The FARIMA family of models are very widely used. For our purposes they simply offer a convenient family of LRD processes in which both the LRD and SRD components can be modified, and for which the Fourier spectra can be written explicitly. We therefore restrict ourselves to a minimal description.

A Gaussian FARIMA$(0,d,0)$ process is defined by $X_i = \triangle^{-d}\epsilon_i$, $i \geq 1$, where the ϵ_i are independent, identically distributed Gaussian random variables called *innovations* with zero mean and variance σ_ϵ^2, and where $\triangle$ is the differencing operator $\triangle\epsilon_i = \epsilon_i - \epsilon_{i-1}$. For fractional d we interpret $\triangle^{-d}$ via a formal power series expansion: $\triangle^{-d} = \sum_{i=0}^{\infty} b_i(-d)B^i$, where B is the backward operator, $B\epsilon_i = \epsilon_{i-1}$, and $b_i(-d) = \Gamma(i+d)/\Gamma(d)\Gamma(i+1)$, $i = 1, 2, \ldots$, Γ being the gamma function. For $d \in (0, 1/2)$ this process is LRD, where $\gamma = 2d$. More generally, a FARIMA$(1,d,1)$ process adds a single autoregressive and moving average term to the fractional differencing:

$$X_i - \phi_1 X_{i-1} = \triangle^{-d}\epsilon_i - \theta_1 \triangle^{-d}\epsilon_{i-1} \tag{36}$$

where the ϕ_1 and the θ_1 are the autoregressive and moving average coefficient respectively. The Fourier spectrum[2] for this process is

$$g_X(\nu) = \sigma_\epsilon^2 |1 - e^{-i2\pi\nu}|^{2d} \frac{|1 - \theta_1 e^{-i2\pi\nu}|^2}{|1 - \phi_1 e^{-i2\pi\nu}|^2}, \; -1/2 < \nu < 1/2\,. \tag{37}$$

To give a greater feel for the models used in the tables, an example with strong SRD is given in figure 3, to be contrasted with the weak SRD of the fGn already seen. It is the FARIMA$(1,d,1)$ process with $H = d + 1/2 = 0.7$, $\phi_1 = 0.3$ and $\theta_1 = 0.7$, and as can be seen from the exact LD in the left plot, it deviates strongly from the asymptotic slope of $\gamma = 0.4$ over the first four octaves. With $n = 10000$ and $N = 3$, the optimal value according to the MSE criterion is $j_1^{\mathrm{MSE}} = 5$. The tradeoff between bias and variance, and the resulting minimum MSE, is shown in detail in the right hand plot.

We now comment on the results. In table 1 we see that the j_1^{MSE} vary from 1 to 5, showing that the models chosen cover a range of SRD / LRD combinations as desired. In table 2 we first note that the results are almost identical to those from table 1, indicating that 50 realizations is a reasonable number from which to determine the j_1 corresponding to the minimum MSE, and to estimate the bias, variance, and the MSE itself. In the two cases where the j_1^{MSE} chosen empirically differed from the analytic result (the models with $H = 0.7$ in the last and third last group), the MSE for the two alternatives were found to be very close. One of these is the model illustrated in figure 3, where the closeness of the MSE at $j_1 = 5$ and 6 is apparent. The choice of $j_1 = 5$ allows a direct comparison with other estimators applied to the very same realizations in previous work Taqqu and Teverovsky (1998), whereas j_1^{MSE} allows the best possible performance of the wavelet estimator to be compared. The conclusion is that it is comparable with the best of them, namely the local Whittle estimator (compare with Taqqu and Teverovsky (1998)). The local Whittle estimator is a semiparametric estimator of γ due to Robinson (1995). It is a Whittle-type estimator, which assumes that the spectral density of the discrete-time process is of the form $g(\nu) = C|\nu|^{-\gamma}$ for $0 < \nu < 2\pi m/n$,

[1]The sample MSE in the tables is the sum of the sample bias squared and the sample variance. In this paper, the sample variance based on $I = 50$ replications, involving a division by I and not $I - 1$, is nonetheless unbiased as the known mean was used, whereas in Taqqu, Teverovsky and Willinger (1995), Taqqu and Teverovsky (1998), it involved a division by $I - 1$, and the known mean was not used. The resulting differences are negligible.

[2]Recall that we use here the definition (19) of Fourier transforms.

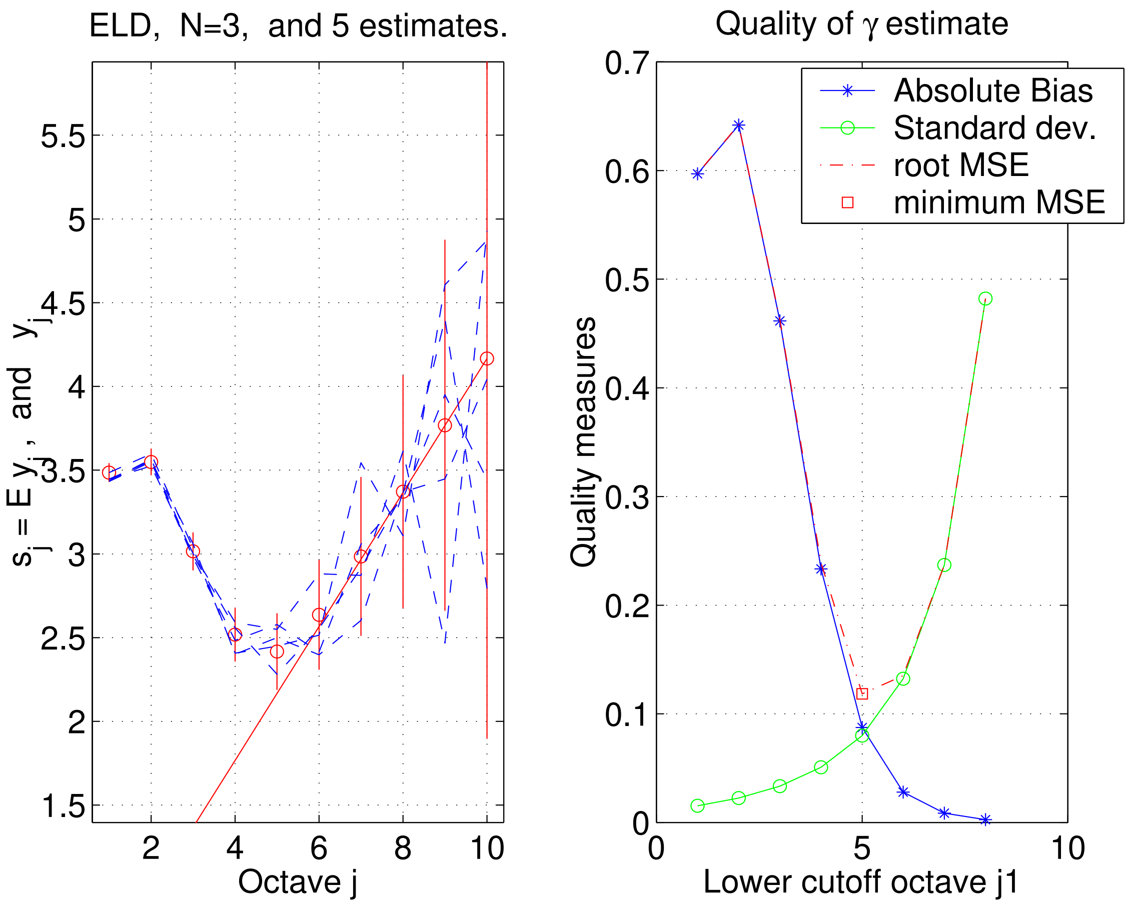

Figure 3: **Exact LD and** j_1^{MSE} **for FARIMA**$(1, d, 1)$**,** $H = 0.7$**,** $\phi_1 = 0.3$**,** $\theta_1 = 0.7$**.** Left: the exact LD for $N = 3$, and 5 sample paths based on $n = 10000$ points. Right: the quality of ideal estimates of γ based on the exact LD, as judged by the magnitude of the bias, the standard deviation, and the root MSE as a function of j_1. The value of $j_1^{\mathrm{MSE}} = 5$ is marked with the square.

where n denotes the number of observations and m is a parameter that has to be adequately chosen. Since m/n is the high-frequency cutoff, it corresponds to our scale 2^{-j_1}. In Taqqu and Teverovsky (1998), the value $m = n/32$ was chosen based on detailed studies performed in Taqqu and Teverovsky (1996, 1997). Since $1/32 = 2^{-5}$, our choice $j_1 = 5$ here is compatible with $m = n/32$.

Table 3 includes an additional column indicating the distribution of the innovations variables. The first two, exponential and lognormal, have finite variance, the others infinite variance. Note that this does **not** mean that the marginal distribution of the process itself will have the distribution indicated, but it does offer an easy way of generating a variety of non-Gaussian time series, and the finiteness or lack thereof of the variance is preserved. In the case of stable innovations the index of stability parameter, α, is given in the same column, and here $H = d + 1/\alpha$ (see Samorodnitsky and Taqqu (1994, Sect. 7.13)). Otherwise $H = d + 1/2$ as before, so there is no need to give the values of d.

An interesting observation is that, despite the variety of different distributions and even in the infinite variance case, the estimator continues to give very accurate estimates, and with comparable estimation variance in the finite variance cases. In the infinite variance cases, however, the estimation variance increases markedly, as we might expect given the far higher variability of these processes, but the bias remains very small. In particular we see that for each of the two models in table 3, namely FARIMA$(0, d, 0)$ and the FARIMA$(1, d, 1)$ which appears in figure 3, the results in each column are comparable across the different finite variance distributions, and comparable in terms of bias for all distributions. In comparison to other estimators, we again note that it is among the best and is as effective as the local Whittle estimator (again see Taqqu and Teverovsky (1998)).

5 A brief historical overview

Wedding wavelets and self-similarity has a long and rich history. In fact, as soon as they appeared in the mid-eighties, it was almost immediately recognized that, thanks to their built-in multiresolution structure, wavelet transforms were natural tools for revealing self-similar features in signals, images or processes. Whereas the potential usefulness of wavelets for studying "fractal noises" is incidentally mentioned in one of Mallat's seminal papers (Mallat (1989)), two short notes, one by Flandrin (1989), the other by Wornell (1990), pointed out key features in two distinct directions. The first study was devoted to fractional Brownian motion and established the stationarization property of the wavelet transform, while proposing a spectral interpretation. The second one proposed the construction of "almost $1/f$ processes," on the basis of uncorrelated wavelet coefficients. Both types of results were subsequently considered in more general settings: stationarization was shown to be closely linked to the existence of stationary increments (Masry (1993), Cambanis and Houdré (1995)), whereas efficient fBm synthesis procedures were further developed by Sellan (1995), Abry and Sellan (1996) and Meyer, Sellan and Taqqu (2000). Linking somewhat the analysis and synthesis viewpoints, the correlation structure of fBm wavelet coefficients was studied in greater detail by Tewfik and Kim (1992) and Flandrin (1992), evidencing that residual correlations can be reduced by increasing the number of vanishing moments of the mother wavelet, thus suggesting that simple and standard tools could be used in the transformed domain for efficiently estimating scaling parameters.

Parallel to the wavelet characterization of scaling processes, effective estimation schemes of scaling parameters were indeed developed. Most efforts relied initially on the assumption of almost uncorrelated coefficients, leading either to maximum likelihood procedures (Wornell and Oppenheim (1992)) or to regression techniques across scales (Abry, Goncalves and Flandrin (1995)). The latter approach ("Logscale Diagram") was thoroughly investigated by Abry and Veitch (1998) and Veitch and Abry (1998) (for a review, see Abry, Flandrin, Taqqu and Veitch (1998)), and is still an active area of research from the point of view of refined statistical performance evaluation, in particular when relaxing the idealized assumption of exact decorrelation (Papanicolaou and Solna (1998), Bardet, Lang, Moulines and Soulier (2000)). Bardet (1999) establishes asymptotic normality of estimators based on wavelet-like transformations $d_{j,k} = \sum_{i=1}^{p} u_i X_{j(i+k)}$, where the first N moments of $(u_1, \ldots, u_p)$ are zero: $\sum_{i=1}^{p} i^q u_i = 0,\ q = 0, \ldots, N-1$.

Initial attempts for applying wavelet-based techniques to scaling processes were partly heuristic, and somewhat overlooked the fact that wavelet theory was then essentially aimed at deterministic finite energy signals. Cambanis and Houdré (1995) and later Averkamp and Houdré (1998)), Cohen, Froment and Istas (1991), and more recently Kato and Masry (1999), addressed general questions regarding both the existence, interpretation and (distributional) properties of wavelet transforms, when applied specifically to stochastic processes, and especially those with self-similar properties. This allowed the first approaches to be made more rigorous and to considerably extend their initial scope, which was mostly restricted to fBm. In this respect, departing from the fBm model can be thought of as relaxing at least one the three constructive properties of Gaussianity, stationarity of the increments and self-similarity. It turned out that key properties, essential for revealing self-similarity, still held in non-Gaussian situations, even when the variance is infinite: this led to new developments in the direction of certain non-Gaussian processes, with special attention paid by Delbeke (1998), Pesquet-Popescu (1999), Delbeke and Abry (2000) to the case of self-similar stable motions. Allowing increments to be nonstationary was also considered, although to a lesser extent (tracking time-varying scaling exponents with time-scale methods was first considered by Gonçalvès and Flandrin (1992); the concept of multifractional Brownian motion later led Cohen to also propose "quasi-wavelet" estimation techniques based on increments (Cohen (1999)), as was the possibility of softening the assumption of a global scale invariance, initially assumed to hold uniformly over all scales. The approaches proposed have been based on either a phenomenological modeling, where the variance progression from scale to scale can be controlled by a function that does not necessarily identify to a strict power-law (Kaplan and Kuo (1994)), or on deeper ideas of cascade analysis, in which probability distribution functions of detail coefficients at different scales are related to each other (Arnéodo, Muzy and Roux (1997)). Such extensions can possibly be applied for example to long-range dependent processes, for which scale invariance is only asymptotically considered in the limit of large scales. From a different perspective, classes of point processes with scale invariance properties (e.g., Poisson processes whose time-dependent density is controlled by fractional Gaussian noise) were considered too Abry and Flandrin (1996).

In all cases, wavelets progressively emerged as a unifying framework for dealing with many different types of scaling processes, and were identified as versatile generalizations of previous techniques which may have been used in specific contexts. A good example is given by the Allan variance, a refined variance estimation technique which had been specifically introduced in stability studies of atomic clocks, and whose structure turned out to be equivalently phrased in terms of the Haar transform (Flandrin (1992), Percival and Guttorp (1993). Similarly, in classical point process theory, departures from a Poisson model can be ascertained from the fact that the so-called Fano factor, i.e., the ratio between the mean and variance of the associated counting process, differs from unity (for an application to scaling point processes, see, e.g., Kumar and Johnson (1993)). Multiresolution being naturally attached to the ideas of approximation through averaging (scaling function) and detailed fluctuations (wavelet), it offers a natural framework for proposing more efficient wavelet-based analogues of the Fano factor (Abry and Flandrin (1996), Teich, Heneghan, Lowen and Turcott (1996), Thurner *et al.* (1997)). One can finally remark that the same applies to aggregation, a concept of successive approximations via windowed averages over larger and larger scales (Abry, Veitch and Flandrin (1998)).

Implicitly or explicitly, wavelet ideas therefore appeared as the naturally suitable concept for handling scaling processes. In a nutshell, the rationale of using wavelets in this context can be viewed as the result of the threefold conjunction of *i*) offering a natural language for scaling, *ii*) being versatile enough for tolerating model mismatches and/or uncertainties (such as asymptotic self-similarity, trend removal), and *iii*) being easy to implement and naturally equipped with fast and efficient algorithms.

Model	H	j_1^{MSE}	Bias	Std	$\sqrt{\mathrm{MSE}}$	$j_1 = 5$	Bias	Std	$\sqrt{\mathrm{MSE}}$
fGn	0.5	2	−0.004	0.011	0.012	5	−0.000	0.040	0.040
	0.6	2	−0.010	0.011	0.015	5	−0.000	0.040	0.040
	0.7	3	−0.005	0.017	0.017	5	−0.000	0.040	0.040
	0.8	3	−0.005	0.017	0.018	5	−0.000	0.040	0.040
	0.9	3	−0.006	0.017	0.018	5	−0.000	0.040	0.040
farima(0,d,0)	0.6	2	0.000	0.011	0.011	5	0.000	0.040	0.040
	0.7	2	0.003	0.011	0.012	5	0.000	0.040	0.040
	0.8	1	−0.008	0.008	0.011	5	0.000	0.040	0.040
	0.9	1	0.002	0.008	0.008	5	0.000	0.040	0.040
farima(1,d,0)	0.5	5	−0.014	0.040	0.042	5	−0.014	0.040	0.042
$\phi_1 = 0.5$	0.6	5	−0.013	0.040	0.042	5	−0.013	0.040	0.042
	0.7	5	−0.013	0.040	0.042	5	−0.013	0.040	0.042
	0.8	5	−0.012	0.040	0.042	5	−0.012	0.040	0.042
	0.9	5	−0.011	0.040	0.042	5	−0.011	0.040	0.042
farima(0,d,1)	0.5	5	0.016	0.040	0.043	5	0.016	0.040	0.043
$\theta_1 = 0.5$	0.6	5	0.015	0.040	0.043	5	0.015	0.040	0.043
	0.7	5	0.014	0.040	0.042	5	0.014	0.040	0.042
	0.8	5	0.013	0.040	0.042	5	0.013	0.040	0.042
	0.9	5	0.012	0.040	0.042	5	0.012	0.040	0.042
farima(1,d,1)	0.5	5	0.049	0.040	0.064	5	0.049	0.040	0.064
$\phi_1 = 0.3$	0.6	5	0.046	0.040	0.061	5	0.046	0.040	0.061
$\theta_1 = 0.7$	0.7	5	0.044	0.040	0.059	5	0.044	0.040	0.059
	0.8	5	0.041	0.040	0.058	5	0.041	0.040	0.058
	0.9	5	0.039	0.040	0.056	5	0.039	0.040	0.056
farima(1,d,1)	0.5	3	−0.008	0.017	0.019	5	−0.001	0.040	0.040
$\phi_1 = -0.3$	0.6	3	−0.007	0.017	0.018	5	−0.000	0.040	0.040
$\theta_1 = -0.7$	0.7	3	−0.005	0.017	0.018	5	−0.000	0.040	0.040
	0.8	3	−0.004	0.017	0.017	5	−0.000	0.040	0.040
	0.9	3	−0.003	0.017	0.017	5	−0.000	0.040	0.040
farima(1,d,1)	0.5	5	−0.043	0.040	0.058	5	−0.043	0.040	0.058
$\phi_1 = 0.7$	0.6	5	−0.041	0.040	0.057	5	−0.041	0.040	0.057
$\theta_1 = 0.3$	0.7	5	−0.039	0.040	0.056	5	−0.039	0.040	0.056
	0.8	5	−0.037	0.040	0.055	5	−0.037	0.040	0.055
	0.9	5	−0.036	0.040	0.054	5	−0.036	0.040	0.054

Table 1: **Estimation quality for Gaussian series: Theoretical.** On the left the optimal j_1 value (minimum $\mathrm{MSE}(j_1)$ of the H estimator) is given along with the corresponding bias, standard deviation and root MSE (these can all been calculated theoretically in the Gaussian case). The values for $j_1 = 5$ are also given for comparison.

Model	H	j_1^{MSE}	Bias	Std	$\sqrt{\mathrm{MSE}}$	$j_1 = 5$	Bias	Std	$\sqrt{\mathrm{MSE}}$
fGn	0.5	2	-0.005	0.013	0.014	5	-0.002	0.041	0.041
	0.6	2	-0.013	0.013	0.018	5	-0.005	0.034	0.035
	0.7	3	-0.006	0.017	0.018	5	-0.006	0.044	0.044
	0.8	3	-0.002	0.020	0.020	5	-0.003	0.035	0.036
	0.9	3	-0.004	0.016	0.017	5	-0.003	0.038	0.038
farima(0,d,0)	0.6	2	0.000	0.013	0.013	5	-0.003	0.042	0.042
	0.7	2	0.002	0.010	0.010	5	-0.000	0.041	0.041
	0.8	1	-0.009	0.008	0.012	5	-0.001	0.040	0.040
	0.9	1	0.006	0.008	0.010	5	0.009	0.033	0.035
farima(1,d,0)	0.5	5	-0.016	0.038	0.041	5	-0.016	0.038	0.041
$\phi_1 = 0.5$	0.6	5	-0.016	0.033	0.037	5	-0.016	0.033	0.037
	0.7	5	-0.010	0.038	0.039	5	-0.010	0.038	0.039
	0.8	5	-0.015	0.042	0.045	5	-0.015	0.042	0.045
	0.9	5	-0.022	0.043	0.048	5	-0.022	0.043	0.048
farima(0,d,1)	0.5	5	0.014	0.039	0.041	5	0.014	0.039	0.041
$\theta_1 = 0.5$	0.6	5	0.010	0.038	0.039	5	0.010	0.038	0.039
	0.7	5	0.020	0.040	0.045	5	0.020	0.040	0.045
	0.8	5	0.011	0.040	0.042	5	0.011	0.040	0.042
	0.9	5	0.019	0.036	0.041	5	0.019	0.036	0.041
farima(1,d,1)	0.5	5	0.050	0.054	0.073	5	0.050	0.054	0.073
$\phi_1 = 0.3$	0.6	5	0.038	0.043	0.057	5	0.038	0.043	0.057
$\theta_1 = 0.7$	0.7	6	0.006	0.063	0.063	5	0.052	0.041	0.066
	0.8	5	0.035	0.045	0.057	5	0.035	0.045	0.057
	0.9	5	0.037	0.046	0.060	5	0.037	0.046	0.060
farima(1,d,1)	0.5	3	-0.007	0.017	0.019	5	0.012	0.036	0.038
$\phi_1 = -0.3$	0.6	3	-0.004	0.018	0.018	5	0.006	0.043	0.044
$\theta_1 = -0.7$	0.7	3	-0.008	0.016	0.018	5	-0.010	0.039	0.041
	0.8	3	-0.005	0.017	0.017	5	0.004	0.042	0.042
	0.9	3	-0.005	0.017	0.017	5	-0.004	0.037	0.038
farima(1,d,1)	0.5	5	-0.040	0.038	0.055	5	-0.040	0.038	0.055
$\phi_1 = 0.7$	0.6	5	-0.050	0.044	0.066	5	-0.050	0.044	0.066
$\theta_1 = 0.3$	0.7	6	-0.027	0.047	0.055	5	-0.043	0.040	0.059
	0.8	5	-0.032	0.041	0.052	5	-0.032	0.041	0.052
	0.9	5	-0.040	0.044	0.059	5	-0.040	0.044	0.059

Table 2: **Estimation quality for Gaussian series: Empirical.** On the left j_1 has been chosen according to the minimum estimated (average over 50 realizations) MSE(j_1) of the H estimate, with corresponding estimated bias, standard deviation and root MSE. The values for $j_1 = 5$ are also given, and are the estimated versions of those in Table 1.

Model	innovations	H	j_1	Bias	Std	$\sqrt{\text{MSE}}$	Bias	Std	$\sqrt{\text{MSE}}$
farima(0,d,0)	exponential	0.5	2	−0.005	0.015	0.016	−0.006	0.035	0.035
		0.6	2	0.003	0.014	0.014	0.001	0.041	0.041
		0.7	2	0.005	0.012	0.013	0.003	0.041	0.041
		0.8	1	−0.009	0.007	0.012	0.001	0.035	0.035
		0.9	1	0.004	0.009	0.010	−0.005	0.039	0.040
farima(0,d,0)	lognormal	0.5	2	−0.004	0.018	0.019	−0.003	0.040	0.041
		0.6	2	0.003	0.018	0.018	0.003	0.043	0.043
		0.7	2	0.003	0.019	0.019	−0.003	0.037	0.037
		0.8	2	0.003	0.012	0.012	−0.002	0.043	0.043
		0.9	1	0.001	0.014	0.014	−0.001	0.042	0.042
farima(0,d,0)	pareto	0.5	1	−0.038	0.064	0.074	0.003	0.132	0.132
		0.6	1	−0.020	0.073	0.076	0.020	0.132	0.133
		0.7	1	−0.039	0.067	0.077	0.009	0.102	0.102
		0.8	2	0.006	0.072	0.072	−0.009	0.125	0.126
farima(0,d,0)	sym stable	0.5	1	−0.029	0.082	0.087	0.018	0.200	0.201
	$\alpha = 1.2$	0.6	1	−0.011	0.080	0.081	0.001	0.208	0.208
farima(0,d,0)	$\alpha = 1.5$	0.5	1	−0.047	0.057	0.074	0.014	0.131	0.132
		0.6	1	−0.030	0.058	0.065	0.004	0.124	0.124
		0.7	1	−0.016	0.054	0.057	0.016	0.110	0.111
		0.8	1	−0.001	0.059	0.059	0.004	0.091	0.091
farima(1,d,1)	sym stable	0.5	5	0.047	0.190	0.196	0.047	0.190	0.196
$(\phi_1, \theta_1) = (0.3, 0.7)$	$\alpha = 1.2$	0.6	5	0.045	0.183	0.188	0.045	0.183	0.188
farima(1,d,1)	$\alpha = 1.5$	0.5	6	0.025	0.155	0.157	0.062	0.172	0.183
$(\phi_1, \theta_1) = (0.3, 0.7)$		0.6	5	0.082	0.162	0.182	0.082	0.162	0.182
		0.7	5	0.004	0.118	0.118	0.004	0.118	0.118
		0.8	6	0.023	0.126	0.128	0.024	0.130	0.132
farima(0,d,0)	skewed stable	0.5	2	−0.008	0.069	0.070	0.013	0.115	0.116
	$\alpha = 1.5$	0.6	1	−0.031	0.050	0.058	0.007	0.115	0.115
		0.7	1	−0.009	0.066	0.067	0.015	0.128	0.129
		0.8	1	0.005	0.054	0.055	−0.001	0.101	0.101

Table 3: **Estimation quality for non-Gaussian series: Empirical.** The distribution of the innovations variables is given in the second column. For stable processes the index of stability α is also given. On the left j_1 has been chosen according to the minimum estimated (average over 50 realizations) MSE(j_1) of the H estimate, with corresponding estimated bias, standard deviation and root MSE. The last 3 columns, which are included for comparison, correspond to the value $j_1 = 5$. In all cases the bias remains small and comparable with the corresponding Gaussian case. The estimation variance is also comparable for finite variance innovations, but is much larger with infinite variance.

References

Abry, P. & Flandrin, P. (1996), Point processes, long-range dependence and wavelets, *in* A. Aldroubi & M. Unser, eds, 'Wavelets in Medicine and Biology', CRC Press, Boca Raton, USA, pp. 413–437.

Abry, P. & Sellan, F. (1996), 'The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation', *Applied and Computational Harmonic Analysis* **3**(4), 377–383.

Abry, P. & Veitch, D. (1998), 'Wavelet analysis of long range dependent traffic', *IEEE Transactions on Information Theory* **44(1)**, 2–15.

Abry, P., Flandrin, P., Taqqu, M. S. & Veitch, D. (1998), Wavelets for the analysis, estimation and synthesis of scaling data, *in* K. Park & W. Willinger, eds, 'Self-Similar Network Traffic and Performance Evaluation', Wiley (Interscience Division). To appear in 2000.

Abry, P., Gonçalvès, P. & Flandrin, P. (1995), Wavelets, spectrum estimation and $1/f$ processes, *in* A. Antoniadis & G. Oppenheim, eds, 'Wavelets and Statistics', Vol. 103 of *Lecture Notes in Statistics*, Springer-Verlag, pp. 15–30.

Abry, P., Veitch, D. & Flandrin, P. (1998), 'Long-range dependence: revisiting aggregation with wavelets', *Journal of Time Series Analysis* **19(3)**, 253–266.

Arnéodo, A., Muzy, J. F. & Roux, S. G. (1997), 'Experimental analysis of self-similar random cascade processes: application to fully developed turbulence', *Journal de Physique, II, France* **7**, 363–370.

Averkamp, R. & Houdré, C. (1998), 'Some distributional properties of the continuous wavelet transform', *IEEE Transactions on Information Theory* **44(3)**, 1111–1124.

Bardet, J.-M. (1999), Testing for the presence of self-similarity of Gaussian time series having stationary increments, Preprint.

Bardet, J. M., Lang, G., Moulines, E. & Soulier, P. (2000), 'Wavelet estimator of long-range dependent processes', *Statistical Inference for Stochastic Processes.* To appear.

Cambanis, S. & Houdré, C. (1995), 'On the continuous wavelet transform of second-order random processes', *IEEE Transactions on Information Theory* **41(3)**, 628–642.

Cohen, A., Froment, J. & Istas, J. (1991), 'Analyse multirésolution des signaux aléatoires', *Comptes Rendus de l'Académie des Sciences de Paris* **312, série I**, 567–570.

Cohen, S. (1999), Locally self similar processes, *in* M. Dekking, J. L. Véhel, E. Lutton & C. Tricot, eds, 'Fractals: theory and applications in engineering', Springer-Verlag.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM Philadelphia. CBMS-NSF series, Volume 61.

Delbeke, L. (1998), Wavelet based estimators for the Hurst parameter of a self-similar process, PhD thesis, KU Leuven, Belgium.

Delbeke, L. & Abry, P. (2000), 'Stochastic integral representation and properties of the wavelet coefficients of the linear fractional stable motion', *Stochastic Processes and their Applications.* To appear.

Flandrin, P. (1989), 'On the spectrum of fractional Brownian motions', *IEEE Transactions on Information Theory* **IT-35**(1), 197–199.

Flandrin, P. (1992), 'Wavelet analysis and synthesis of Fractional Brownian motion', *IEEE Transactions on Information Theory* **38**, 910–917.

Gonçalvès, P. & Flandrin, P. (1992), Scaling exponents estimation from time-scale energy distributions, *in* 'Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-92', San Francisco, pp. V.157–V.160.

Houdré, C. (1993), Wavelets, probability and statistics: some bridges, *in* J. J. Benedetto & M. W. Frazier, eds, 'Wavelets: Mathematics and Applications', CRC Press, Boca Raton, Florida, pp. 361–394.

Kaplan, L. M. & Kuo, C. C. J. (1994), 'Extending self-similarity for fractional Brownian motion', *IEEE Transactions on Signal Processing* **42(12)**, 3526–3530.

Kato, T. & Masry, E. (1999), 'On the spectral density of the wavelet transform of fractional Brownian motion', *Journal of Time Series Analysis* **20**, 559–563.

Kumar, A. H. & Johnson, D. H. (1993), 'Analyzing and modeling fractal intensity point processes', *The Journal of the Acoustical Society of America* **93**, 3365–3373.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Academic Press, Boston.

Mallat, S. G. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-11**(7), 674–693.

Masry, E. (1993), 'The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion', *IEEE Transactions on Information Theory* **39**(1), 260–264.

Meyer, Y. (1998), *Wavelets, vibrations and scalings*, Vol. 9 of *CRM Monograph Series*, American Mathematical Society, Providence.

Meyer, Y., Sellan, F. & Taqqu, M. S. (2000), 'Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion', *The Journal of Fourier Analysis and Applications*. To appear.

Papanicolaou, G. & Sølna, K. (1998), Estimation of local power law processes, Preprint.

Percival, D. & Guttorp, P. (1993), Long memory processes, the Allan variance and wavelets, *in* Foufoula-Georgiou & Kumar, eds, 'Wavelets in Geophysics', Academic Press.

Pesquet-Popescu, B. (1999), 'Statistical properties of the wavelet decomposition of some non Gaussian self-similar processes', *Signal Processing* **75**, 303–322.

Robinson, P. M. (1995), 'Gaussian semiparametric estimation of long range dependence', *The Annals of Statistics* **23**, 1630–1661.

Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, London.

Sellan, F. (1995), 'Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes', *Comptes Rendus de l'Académie des Sciences de Paris* **321**, 351–358. Série I.

Taqqu, M. S. & Teverovsky, V. (1996), Semi-parametric graphical estimation techniques for long-memory data, *in* P. M. Robinson & M. Rosenblatt, eds, 'Athens Conference on Applied Probability and Time Series Analysis. Volume II: Time Series Analysis in Memory of E. J. Hannan', Vol. 115 of *Lecture Notes in Statistics*, Springer-Verlag, New York, pp. 420–432.

Taqqu, M. S. & Teverovsky, V. (1997), 'Robustness of Whittle-type estimates for time series with long-range dependence', *Stochastic Models* **13**, 723–757.

Taqqu, M. S. & Teverovsky, V. (1998), On estimating the intensity of long-range dependence in finite and infinite variance series, *in* R. Adler, R. Feldman & M. S. Taqqu, eds, 'A Practical Guide to Heavy Tails: Statistical Techniques and Applications', Birkhäuser, Boston, pp. 177–217.

Taqqu, M. S., Teverovsky, V. & Willinger, W. (1995), 'Estimators for long-range dependence: an empirical study', *Fractals* **3**(4), 785–798. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

Teich, M. C., Heneghan, C., Lowen, S. B. & Turcott, R. G. (1996), Estimating the fractal exponent of point processes in biologic al systems using wavelet- and Fourier-transform methods, *in* A. Aldroubi & M. Unser, eds, 'Wavelets in Medicine and Biology', CRC Press, Boca Raton (Florida), pp. 383–412.

Tewfik, A. H. & Kim, M. (1992), 'Correlation structure of the discrete wavelet coefficients of fractional Brownian motions', *IEEE Transactions on Information Theory* **IT-38**(2), 904–909.

Thurner, S., Lowen, S. B., Feurstein, M. C., Heneghan, C., Feichtinger, H. G. & Teich, M. C. (1997), 'Analysis, synthesis and estimation of fractal-rate stochastic point processes', *Fractals* **5**(4), 565–595.

Veitch, D. & Abry, P. (1998), 'A wavelet-based joint estimator of the parameters of long-range dependence', *IEEE Transactions on Information Theory* **45(3)**, 878–897. Special Issue on Multiscale Statistical Signal Analysis and its Applications.

Veitch, D., Taqqu, M. S. & Abry, P. (2000), 'Meaningful MRA initialisation for discrete time series', *Signal Processing*. To appear.

Vetterli, M. & Kovacevic, J. (1995), *Wavelets and Subband Coding*, Prentice Hall.

Wornell, G. W. (1990), 'A Karhunen-Loève-like expansion for 1/f processes via wavelets', *IEEE Transactions on Information Theory* **36**, 859–861.

Wornell, G. W. & Oppenheim, A. V. (1992), 'Estimation of fractal signals from noisy measurements using wavelets', *IEEE Transactions on Information Theory* **40**(3), 611–623.

Zygmund, A. (1979), *Trigonometric Series*, Cambridge University Press, Cambridge. Volumes I, II.