Wavelets for scaling processes

Patrick Flandrin and Patrice Abry

Ecole Normale Supérieure de Lyon
Laboratoire de Physique (UMR 5672 CNRS)
46 allée d'Italie, 69364 Lyon Cedex 07 - FRANCE
{flandrin,pabry}@physique.ens-lyon.fr

Abstract. Depending on the considered range of scales, different *scaling processes* may be defined, which correspond to different situations connected with self-similarity, fractality or long-range dependence. Wavelet analysis is shown to offer a unified framework for dealing with such processes and estimating the corresponding scaling parameters. Estimators are proposed and discussed on the basis of representations in the wavelet domain. Statistical (and computational) efficiency can be obtained not only for second-order processes, but also for linear fractional stable motions with infinite variance.

Keywords. Wavelets, scaling, self-similarity, long-range dependence, stable processes.

1 Introduction

Signals presenting some form of scaling behaviour can be observed in a wide variety of fields, ranging from physics (turbulence, hydrology, solid-state physics, ...) or biology (DNA sequences, heart rate variability, auditory nerves spike trains, ...) to human-operated systems (telecommunications network traffic, finance, ...). While sharing a common property of *scale invariance*, processes used to model such observations may however differ according to the range of scales over which the invariance is effective. Indeed, relevant concepts can be either *self-similarity* (the part is, in some sense, identical to the whole), *long-range dependence* (algebraically decaying and non-integrable correlations at "large scales" result in power-law diverging spectra at low frequencies) or *irregularity of sample paths* ("small scale" scaling results in non-integer fractal dimensions). In each case, no characteristic scale exists (in a given range), the important feature being rather the existence of some invariant relation *between* scales.

Because they may correspond to non-standard situations in signal processing or time series analysis (non-stationarity, long-range dependence, ...), *scaling processes* raise challenging problems in term of analysis, synthesis, and processing (filtering, prediction, ...). A number of specific tools have however been developed over the years and, in a recent past, it has been realized that a natural approach was to consider scaling processes from the perspective of the multiresolution tools which had been introduced since the mid-eighties around the fruitful concept of *wavelet.* In fact, wavelets have, by construction, a built-in ability to look at a signal or a process at different scales, and to reveal potential invariant relations between scales within a proper and well-understood mathematical setting. The purpose of this paper is therefore to show which advantages can be gained from using a wavelet-based perspective when dealing with scaling processes.

More precisely, the paper is organized as follows. First, basics of wavelet analysis are briefly recalled in Section 2. Section 3 addresses the crucial issue of discussing scaling processes in the wavelet domain, showing that the proposed framework allows to consider in a unified way different types of situations (specifically, self-similar process with or without variance, as well as long-range dependent processes). Section 4 is devoted to the wavelet-based estimation of scaling parameters. Finally, a number of applications are briefly mentioned in the Conclusion, in order to support the effectiveness of the methods previously discussed.

2 Wavelets

Wavelet analysis [22] formalizes the idea of looking at a signal (or a process) at different *scales* or *levels of resolution.* This is achieved by decomposing any signal onto a set of elementary "building blocks" which are all deduced from a unique waveform—supposed to be reasonably well localized in both time and frequency—by means of shifts and dilations. More precisely, the *Continuous Wavelet Transform* (CWT) of a signal $X(t) \in L^2(\mathbf{R})$ is given by

$$T_X(a,t) := \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} X(s)\, \psi_0 \left(\frac{s-t}{a} \right) ds, \tag{2.1}$$

with $a \in \mathbf{R}_+$, $t \in \mathbf{R}$ and where $\psi_0(.)$ stands for the *mother wavelet* of the analysis. Provided that this function is zero-mean, the wavelet transform can be inverted as

$$X(t) = C_{\psi_0} \int\!\!\int_{-\infty}^{+\infty} T_X(a,s)\, \frac{1}{\sqrt{a}}\, \psi_0 \left(\frac{t-s}{a} \right) \frac{da\, ds}{a^2}. \tag{2.2}$$

By varying the t variable, the wavelet transform allows for a local analysis in time, whereas by varying the a variable, it offers the possibility on "zooming in" on details, thus playing the role of a "mathematical microscope" and making of it a tool which is *a priori* naturally adapted to scaling processes.

Whereas the wavelet transform is highly redundant in its continuous form (2.1), it can be efficiently discretized on the dyadic grid $a = 2^j, t = 2^j k$ (j and $k \in \mathbf{Z}$) of the time-scale plane. The framework developed for defining the corresponding *Discrete Wavelet Transform* (DWT) is referred to as *MultiResolution Analysis* (MRA). Roughly speaking, the MRA of a signal space is defined by a sequence of nested subspaces $\ldots \subset V_j \subset V_{j-1} \subset \ldots$, each associated with a given level of resolution indexed by j and such that passing from one subspace to the neighbouring one is obtained by a dilation of factor 2. Provided that a basis (given by integer translates of a low-pass waveform $\phi_0(.)$) exists for some subspace V_0 chosen as reference, the existence of a MRA guarantees that any signal can be decomposed under the form "signal = approximation + detail", with the possibility of iterating the process at coarser and coarser scales, by further decomposing successive approximations. For a decomposition of depth J, a signal $X(t) \in V_0$ can thus be written as

$$X(t) = \sum_{k=-\infty}^{\infty} a_X(J,k)\, \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_{k=-\infty}^{\infty} d_X(j,k)\, \psi_{j,k}(t), \tag{2.3}$$

where the a_X and d_X stand for the approximation and detail coefficients, respectively, while $\phi_{j,k}(t) := 2^{-j/2}\phi_0(2^{-j}t-k)$ and $\psi_{j,k}(t) := 2^{-j/2}\psi_0(2^{-j}t-k)$, the wavelet $\psi_0(.)$ being such that its integer translates are a basis of W_0, the complement of V_0 in V_{-1}. Up to the coarser approximation coefficients $a_X(J,k)$, the discrete wavelet transform is therefore given by the set of all detail coefficients $d_X(j,k)$, which measure indeed a *difference in information* between two successive approximations. Thanks to the dyadic structure of the sampling, they can be written as

$$d_X(j,k) := 2^{-j/2} \int_{-\infty}^{+\infty} X(t)\, \psi_0\left(2^{-j}t-k\right)\, dt, \tag{2.4}$$

and they are therefore obtained as a projection of the analyzed signal onto the corresponding wavelet subspace W_j.

As in the continuous case, wavelets have to be zero-mean for being admissible, but their design can be controlled by a number of additional degrees of freedom. One important wavelet property—which will prove essential in the following—is its *number of vanishing moments*, i.e., the number $N \geq 1$ such that

$$\int_{-\infty}^{+\infty} t^k\, \psi_0(t)\, dt \equiv 0, \qquad k = 0, 1, \ldots N-1. \tag{2.5}$$

Other properties may be appealing from the point of view of computational efficiency. This is especially the case of the *compact support property*, which is intimately related to the existence of FIR filterbank implementations. In fact, a key feature of the discrete wavelet transform is that the computation of the coefficients (2.4) can be actually achieved in a fully discrete and recursive way, leading to extremely efficient algorithms, which even outperform FFT-type algorithms.

The interested reader is referred to, e.g., [22] for a thorough presentation of wavelet transforms.

3 Scaling processes in the wavelet domain

In the previous Section, the wavelet transforms—either continuous or discrete—were defined for deterministic signals or functions belonging to $L^2(\mathbf{R})$. They have, however, been naturally and widely applied to the analysis of stochastic processes. In such cases, the wavelet coefficients are (continuous or discrete) random fields, raising issues concerning their existence and statistical properties. Such questions will not be addressed in detail here. Let us simply note that the wavelet coefficients will basically inheritate of the properties of both the analyzed process and the mother wavelet. The statistical properties of the wavelet coefficients (existence, finiteness of moments, dependence structure) will hence depend on joint conditions on the mother wavelet and on the statistics of the analyzed process. Self-similarity or long-range dependence, which are under study here, generally involve (whenever they exist) covariance functions that are not bounded. We will assume that the mother wavelet decays at least exponentially fast in the time domain, so as to guarantee the existence of the wavelet coefficients. The interested reader is referred to, e.g., [9, 12, 16, 17, 23, 26, 27] for further details.

3.1 Self-similar processes with stationary increments

Self-similarity. A process $X = \{X(t), t \in \mathbf{R}\}$ is said to be *self-similar* with self-similarity parameter $H > 0$ (hereafter, "H-ss") if and only if the processes $X_1 := X$ and $X_c := \left\{c^{-H} X(ct), t \in \mathbf{R}\right\}$ have the same finite dimensional distributions for any $c > 0$. Self-similarity means that the process is statistically scale-invariant: it does not possess any characteristic scale of time or, equivalently, it is not possible to distinguish between a suitably scaled version (in time and amplitude) of the process and the process itself. Self-similarity also implies that X is a nonstationary process, since it is obvious from the definition that the variance of X, when it exists, reads: $\mathbf{E}X^2(t) = |t|^{2H}\,\mathbf{E}X^2(1)$.

The (DWT) wavelet coefficients of an H-ss process X exactly reproduce its self-similarity through the key scaling property:

$$\textbf{P1:}\ (d_X(j,0),\ldots,d_X(j,N_j-1)) \stackrel{d}{=} 2^{j(H+1/2)}(d_X(0,0),\ldots,d_X(0,N_j-1)). \tag{3.6}$$

In the CWT framework, the same property reads (for any $c > 0$):

$$\textbf{P1c:}\ (T_X(ca,ct_1),\ldots,T_X(ca,ct_n)) \stackrel{d}{=} c^{H+1/2}(T_X(a,t_1),\ldots,T_X(a,t_n)). \tag{3.7}$$

Let us emphasize that this results (non trivially) from the fact that the analyzing wavelet basis is designed from the dilation operator and is therefore, by nature, scale invariant. Such a property has been originally established in the case of the fractional Brownian motion (FBM) in [19, 23, 33], more recently in the case of the linear fractional stable motion in [16, 17, 26, 27] and more generally, for any self-similar process, in [9, 16].

The fundamental result **P1** can be given two special forms that will play key roles in the section dedicated below to the estimation of the self-similarity parameter. For processes whose wavelet coefficients have finite second-order statistics, one has:

$$\textbf{P1var:}\ \mathbf{E}d_X^2(j,k) = 2^{j(2H+1)}\mathbf{E}d_X^2(0,k), \tag{3.8}$$

whereas, for processes for which the quantity $\mathbf{E}\log_2|d_X(j,k)|$ exists, one obtains:

$$\mathbf{P1log:}\ \mathbf{E}\log_2|d_X(j,k)| = j(H+1/2) + \mathbf{E}\log_2|d_X(0,k)|. \qquad (3.9)$$

Stationary increments. A process $X = \{X(t), t \in \mathbf{R}\}$ is said to have *stationary increments* (hereafter, "si") if and only if, for any $h \in \mathbf{R}$, the finite-dimensional distributions of the processes $X^{(h)} = \{X^{(h)}(t) := X(t+h) - X(t), t \in \mathbf{R}\}$ do not depend on t. In the case of both the DWT and the CWT, this results in a *stationarization property*, according to which

P2: the $\{d_X(j,k), k \in \mathbf{Z}\}$ form, at each octave j, a stationary sequence.

P2c: the $\{T_X(a,t), t \in \mathbf{R}\}$ form, at each scale a, a stationary process.

This is a direct consequence of the fact that the mother wavelet has at least one vanishing moment (i.e., $N \geq 1$). This has been shown in its most general form in [9, 16] and, in a more specific context, in [19, 23, 33] for the case of the FBM and in [16, 17, 26, 27] for the case of the linear fractional stable motion (LFSM).

The stationarization property **P2** can be extended to processes that do not possess stationary increments but have increments of higher order that are stationary. If we denote by p the number of times one has to take increments to obtain a stationary process, the wavelet coefficients form themselves a stationary process under the condition that $N \geq p$. By stationary increments, we hereafter mean that there exists an integer p such that the increments of order p of X are stationary and that the condition $N \geq p$ is satisfied.

Let us also note that the increments of a process X can be read as a specific example of wavelet coefficients, since we have

$$X^{(ah_0)}(t) := X(t+ah_0) - X(t) \equiv \frac{1}{\sqrt{a}} T_X(a,t),$$

with $\psi_0(t) = \delta(t+h_0) - \delta(t)$, where $\delta(t)$ denotes the Dirac distribution. Such a mother wavelet has however poor regularity, possesses only one vanishing moment (i.e., $N = 1$) and cannot be constructed from a multiresolution analysis.

Self-similarity with stationary increments. Among all self-similar processes, there exists a subclass of particular interest over which we hereafter concentrate: the class of *self-similar processes with stationary increments* (hereafter, "H-sssi"). For the increments $Y(h,t) := X^{(h)}(t)$ of a H-sssi process, one can show that the processes $\{Y(h,t), t \in \mathbf{R}\}$ and $\{c^{-H}Y(ch,ct), t \in \mathbf{R}\}$ have the same finite-dimensional distributions, for all $c > 0$, a property that has an analagous form to—and is highly reminiscent of—the property **P1c** of the wavelet coefficients.

In the remainder of the paper, and for a sake of simplicity, we will concentrate on the coefficients of the DWT (the $d_X(j,k)$) rather than on those of the CWT (the $T_X(a,t)$), despite the fact that most results could be written in either framework.

Finite variance and Gaussian processes Variance. Let X denote a zero-mean H-sssi process with finite variance (i.e., such that $\mathbf{E}X(t) \equiv 0$ and Var $X(t) < \infty$, for any $t \in \mathbf{R}$). From the fundamental property **P1** (eq.(3.6) or eq.(3.8)), from $\mathbf{E}d_X(j,k) \equiv 0$ and from the stationarity of the wavelet coefficients (property **P2**), it is straighforward to show that the wavelet coefficients of X satisfy:

$$\text{Var } d_X(j,k) = 2^{j(2H+1)}\text{Var } d_X(0,0). \tag{3.10}$$

Covariance. Moreover, adding the usual convention that $X(0) \equiv 0$, one gets that the covariance of an H-sssi process reads:

$$\mathbf{E}X(t)X(s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{3.11}$$

From this form of the covariance of the process, the asymptotic behaviour of the covariance of the wavelet coefficients can be obtained:

$$\textbf{P3: } \mathbf{E}d_X(j,k)d_X(j',k') \sim |2^{-j}k - 2^{-j'}k'|^{2(H-N)}, \; |2^{-j}k - 2^{-j'}k'| \to \infty. \tag{3.12}$$

This result has been established originally for the FBM in [19, 23, 33]. It shows that the range of correlation is controlled by the number N of vanishing moments of the mother wavelet: the higher N, the shorter the range.

Finally, from the expression (3.11) of the covariance of X, it can be readily derived [1, 3, 6, 19] that:

$$\text{Var } d_X(j,k) = 2^{j(2H+1)}\sigma^2 C(\psi_0, H), \tag{3.13}$$

where $\sigma^2 := \mathbf{E}X^2(1)$ and $C(\psi_0, H) := \int_{-\infty}^{+\infty} |u|^{2H} \left(\int_{-\infty}^{+\infty} \psi_0(v)\psi_0(v+u)dv\right) du$.

Gaussianity. If ones moreover requires that the zero-mean H-sssi process is Gaussian, one is led to the only FBM [24]. In this case, the wavelet coefficients are Gaussian too.

From Gaussian to stable processes Stable processes. Let us suppose that we are now interested in (zero-mean) H-sssi processes with (possibly) infinite variance. *Stable motions* [30] offer an interesting framework to model H-sssi processes whose second-order statistics do not exist. By definition, such processes admit the representation

$$X(t) = \int_{-\infty}^{+\infty} f(t,u)M(du),$$

with $M(du)$ some *symmetric* α*-stable* (hereafter, "SαS") measure, with scale parameter σ, and $f(t,u)$ an integration kernel that controls the time dependence of the statistics of X. For instance, if $f(t,u) \equiv f(0, u-t)$, then $X(t)$ is a stationary process; if $f(ct, cu) = c^{-H}f(t,u)$ for any $c > 0$, then $X(t)$ is a H-ss process. Two examples are of particular interest, namely the *Lévy flights* and the *linear fractional stable motion* (LFSM) [30]. The first one is defined through $f(t,u) := 1$ if $u \le t$, and 0 otherwise. It is a H-ss process with $H = 1/\alpha$, and its increments are stationary and independent. The second one is defined through a parameter $d \le 1/2$ and the kernel $f(t,u) := (t-u)_+^d - (-u)_+^d$, where $(t)_+ = t$ if $t \ge 0$, and 0 otherwise. It is a H-ss process with $H = d + 1/\alpha$. Its increments are stationary but dependent, the dependence being controlled by d. It has been shown [9, 16, 17, 27] that, under mild joint conditions on the mother wavelet ψ_0 and the kernel $f(t,u)$, the wavelet coefficients of a stable motion are SαS random variables with integral representation:

$$d_X(j,k) = \int_{-\infty}^{+\infty} h_{j,k}(u) M(du)\,; \qquad h_{j,k}(u) := \int_{-\infty}^{+\infty} f(t,u)\psi_{j,k}(t)dt.$$

H-sssi stable processes. If X is a H-sssi stable process, then, from the stationarity of its increments and from property **P1**, one can show that the scale parameters of its wavelet coefficients satisfy the following scaling relation:

$$\textbf{P1infvar:}\ \sigma_{j,k}^{\alpha} = 2^{j(H+1/2)\alpha}\sigma_{0,0}^{\alpha}\,, \tag{3.14}$$

where $\sigma_{0,0}$ is a quantity that depends on both the mother wavelet $\psi_0(t)$ and the function $f(t,u)$, and therefore on d for the LFSM [4, 16, 17]. This relation plays a role that is equivalent to eq.(3.10) in the finite variance case.

Logarithm of H-sssi processes. The property (and mainly the dependence structure) of the wavelet coefficients of H-sssi stable processes will not be further detailed here, and the reader is referred to [4, 5, 16, 18]. Instead, we will turn to *logarithmic transformations* of H-sssi processes. It is well-known [25, 30] that, if X is a stable variable, then the variable $Y = \log_2 |X|$ has a finite variance, yielding the idea of considering the random variable $\log_2 |d_X(j,k)|$.

Let X denote a H-sssi process with arbitrary (finite or infinite) variance, then it results from the fundamental properties **P1** and **P2** that [16]:

$$\mathbf{E} \log_2 |d_X(j,k)| = j(H+1/2) + \mathbf{E} \log_2 |d_X(0,0)|\,. \tag{3.15}$$

This equation plays a role analogous to that of eq.(3.10). Let us moreover note that, while eq.(3.10) involves the variance of $d_X(j,k)$ because of their mean which is identically zero, the equation above directly reproduce the self-similarity of X through the means of $\log_2 |d_X(j,k)|$.

The covariance of $\log_2 |d_X(j,k)|$ has also been studied in the case of the LFSM. It has been shown in [16, 18] that:

$$\textbf{P3log:}\ |\text{Cov } \log_2 |d_x(j,k)|, \log_2 |d_x(j,k')|| \le \tag{3.16}$$

$$C|k-k'|^{-(\alpha/4)(N-H)},\ |k-k'| \to \infty\,, \tag{3.17}$$

evidencing again that the number of vanishing moments N controls the correlation of the log-coefficients $\log_2 |d_X(j,k)|$, which can be made as small as desired by increasing N.

3.2 Long-range dependent processes

Second-order stationarity. Let X be a second-order stationary stochastic process (supposed to be zero-mean, for a sake of simplicity), with stationary covariance function $c_X(\tau)$ and spectrum $\Gamma_X(\nu)$. The covariance of its wavelet coefficients can be expressed as [19]:

$$\mathbf{E}d_X(j,k)d_X(j',k') = \int\int_{-\infty}^{+\infty} c_X(u-v)\,\psi_{j,k}(u)\,\psi_{j',k'}(v)\,du\,dv\,,$$

whereas its variance reads:

$$\mathbf{E}d_X^2(j,k) = \int_{-\infty}^{+\infty} \Gamma_X(\nu)\,2^j|\Psi_0(2^j\nu)|^2\,d\nu\,,$$

where $\Psi_0(\nu)$ stands for the Fourier transform of $\psi_0(t)$. The above relation may receive a standard *spectral estimation* interpretation within the time-invariant linear filtering theory, with k seen as a time index and ψ_0 as a band-pass filter [1].

Long-range dependence. A second-order stationary process X is said to be *long-range dependent* (LRD) if its covariance satisfies [11]:

$$c_X(\tau) \sim c_r\,\tau^{-\beta},\ \tau \to +\infty,$$

with $0 < \beta < 1$. An equivalent definition amounts to saying that the spectrum of a LRD process satisfies:

$$\Gamma_X(\nu) \sim c_f\ |\nu|^{-\gamma},\ \nu \to 0, \tag{3.18}$$

with $0 < \gamma < 1$.

From the Fourier transform, we have $\beta = 1-\gamma$ and $c_r = c_f 2\Gamma(1-\alpha)\sin(\pi\alpha/2)$, Γ denoting (here and only here) the Gamma function [11, p. 43]. Both definitions imply that $\int_{-\infty}^{+\infty} c_X(\tau)d\tau = \infty$ or, equivalently, that $\Gamma_X(0) = \infty$, the specific marks of LRD. Note that LRD is sometimes also referred to as "long memory" or "second-order asymptotic self-similarity," for a reason made clearer in the following.

The variance of the wavelet coefficients of a LRD process reproduces the underlying power-law:

$$\mathrm{Var}\ d_X(j,k) \sim 2^{j\gamma}c_f C(\psi_0,\gamma)\,,\,j \to \infty, \tag{3.19}$$

with

$$C(\psi_0,\gamma) := \int_{-\infty}^{+\infty} |\nu|^{-\gamma}|\Psi_0(\nu)|^2 d\nu. \tag{3.20}$$

This relation shows that the wavelet coefficients catch the power-law behaviour of the LRD spectrum, and it is equivalent —in form, spirit and consequences— to the eqs.(3.10) or (3.13), which hold for H-sssi processes with finite variance. Moreover, the covariance of wavelet coefficients can be explicitly computed, yielding the following asymptotic behaviour:

$$\textbf{P3LRD: } \mathbf{E}d_X(j,k)d_X(j',k') \sim |2^{-j}k - 2^{-j'}k'|^{\gamma-1-2N}, \ |2^{-j}k - 2^{-j'}k'| \to \infty. \tag{3.21}$$

This shows again that the number of vanishing moments N controls the range of correlation and that long-range correlation within X can be turned to short-range correlation within its wavelet coefficients, under the condition that $N > \gamma/2$, an inequality which is satisfied for any wavelet since, by definition, $N \geq 1$.

Beyond long-range dependence. In the definition (3.18) of LRD, one has assumed that $0 < \gamma < 1$. The extension to $\gamma < 0$ leads to a spectrum Γ_X which still exhibits a power-law behaviour at the origin, but the corresponding process X has no longer long-memory in this case. Nevertheless, the wavelet analysis described above still holds in a straightforward manner. The situation where $\gamma > 1$ is technically more difficult, since X has no longer a finite variance (this situation is sometimes referred to as that of "generalized processes"). Wavelet analysis is however equally valid in this case, since the wavelet coefficients still have finite second-order statistics and eq.(3.19) still holds, on condition that $C(\psi_0, \gamma)$ is finite: this is satisfied as soon as $N > (\gamma - 1)/2$, a condition which will be always assumed to hold in the following.

Long range dependence and self-similarity. Let us finally note that, despite the fact that LRD is defined independently, it has deep relations with self-similarity. It is indeed easy to check that the increments of an H-sssi process with finite variance are LRD processes, with $\gamma = 2H - 1$.

3.3 A unified framework

Let X be either a H-sssi or a LRD process, then its wavelet coefficients exhibit the following key properties for the analysis of the underlying scaling:

- The wavelet coefficients form *stationary sequences* in time, at every scale. This is true on condition that $N > p$, p being the number of times one has to take increments of a H-sssi process to obtain stationarity.
- The wavelet coefficients exactly *reproduce the scale invariance*, through either

$$\text{Var } d_X(j,k) = 2^{j\gamma} c_f C$$

or

$$\mathbf{E} \log_2 |d_X(j,k)| = j(H + 1/2) + C\,.$$

While the second of these two equations only applies to H-sssi processes with either finite or infinite variance, the first one gathers long-range dependence and self-similarity (with finite variance) into a single framework. In the case of self-similarity, γ has to be read as $2H + 1$ and c_f as σ^2. The constant

C depends on the mother wavelet and on the scaling exponent, and it has to be read according to eq.(3.10) or (3.19), accordingly. Note, moreover, that the first equality holds, gathering the various cases, on condition that $N > (\gamma - 1)/2$.

- The wavelet coefficients are *weakly correlated*—i.e., they exhibit no LRD—as soon as $N > \gamma/2$. This decorrelation is idealized into the approximation:

  ID1: the wavelet coefficients $d_X(j,k)$ are exactly uncorrelated.

  An equivalent decorrelation effect holds for $\log_2 |d_X(j,k)|$ and, again, it can be idealized into an exact decorrelation approximation:

  ID2: the log-wavelet coefficients $\log_2 |d_X(j,k)|$ are exactly uncorrelated.

These three key properties hold regardless of the precise features of the mother wavelet, except for its number of vanishing moments that plays a role for both the stationarization property, the reproduction of the power-law and the decorrelation effect. What is noteworthy is also that they do not depend on some *a priori* assumption about the nature of the analyzed process: in any of the considered situations, the existence of a scaling behaviour is indeed evidenced by the analysis, the corresponding interpretation (in terms of either self-similarity, small scale scaling or LRD) depending on the range of scales over which it is actually observed. The following section is devoted to make use of these results for *estimating* scaling parameters evidenced by a wavelet analysis.

4 A wavelet framework for the estimation of scaling parameters

4.1 Finite variance and Gaussian processes

Focusing first on finite variance processes, a wavelet-based estimation of scaling parameters can be designed, that takes full advantage of the various properties which have been established so far. In particular, and as far as estimation is concerned, the *stationarization* property (**P2**) of the wavelet transform allows for using the quantity

$$\mu_j = 3D\frac{1}{n_j} \sum_{k=3D1}^{n_j} d_X^2(j,k) \tag{4.22}$$

as an estimator of the variance $\mathrm{Var}\, d_X(j,k)$ at scale j, based on the n_j coefficients available at that scale. Moreover, the coefficients involved in (4.22) can be considered as *almost uncorrelated* for an appropriate choice of the analyzing wavelet (in terms of its number of vanishing moments, see **P3**), making of (4.22) a potentially efficient estimator.

Given these attractive properties, estimating the scaling structure of a process amounts to studying the scale dependence (in j) of the variance estimator μ_j. Since power-law behaviors are expected to occur, it proves more interesting to rather consider the quantity $y_j := 3D\log_2 \mu_j$, thus reformulating the problem in a simple linear regression framework: together with appropriate confidence intervals about the y_j, the graph of y_j against j has been referred to as the (second order) *Logscale Diagram* [6].[1]

The Logscale Diagram is based in fact on the logarithm of the *estimated* variance μ_j, but theory only guarantees that the logarithm of the *true* variance $\mathbf{E}d_X^2(j,k)$ is linear in j. Since, in general, $\mathbf{E}\log_2\{.\} \neq \log_2\{\mathbf{E}\ .\}$, the quantity y_j cannot be expected to be unbiased, unless some correction is applied. Assuming that the analyzed process is Gaussian, and that its wavelet coefficients are uncorrelated (**ID1**), a way out [3] is to redefine y_j as $y_j = 3D\log_2 \mu_j - g_j$, with

$$g_j := 3D\psi\left(n_j/2\right)/\log 2 - \log_2\left(n_j/2\right), \tag{4.23}$$

where $\psi(.)$ is the logarithmic derivative of the Gamma function. Doing so, we readily get that the theoretical scaling relation (3.10) leads to

$$\mathbf{E}y_j = 3D\gamma j + \log_2 c_f C. \tag{4.24}$$

Moreover, the Gaussianity of the process carries over to the associated wavelet coefficients, so that the y_j are scaled and shifted logarithms of chi-squared variables, with variance

$$\sigma_j^2 := 3D\mathrm{Var}\ y_j = 3D\zeta(2, n_j/2)/\log^2 2, \tag{4.25}$$

where $\zeta(z,\nu)$ is a generalized Riemann Zeta function.

Any kind of (weighted) linear regression of y_j on j:

$$\hat{\gamma} = 3D\sum_j w_j\, y_j, \tag{4.26}$$

with

$$w_j := 3D\frac{1}{a_j}\frac{S_0 j - S_1}{S_0 S_2 - S_1^2}, \tag{4.27}$$

where the a_j are arbitrary non-zero numbers, and $S_m := 3D\sum_j j^m/a_j$ for $m = 3D\ 0,1,2$, constitutes therefore an unbiased estimator of the scaling parameter γ. Among them, we will let the weights a_j be precisely the variances σ_j^2 of the y_j, since this choice actually leads to the minimum variance unbiased estimator for the regression problem.

By construction, (4.26) is an unbiased estimator:

$$\mathbf{E}\hat{\gamma} = 3D\gamma \tag{4.28}$$

[1] It has to be observed that such an estimator shares with many other ones the common feature of characterizing straight lines in a log-log plot. Its main originality relies on the fact that wavelets allow for a well-controlled splitting of the analyzed process in a number of sub-processes at different scales, each of those being much better behaved than the original process considered as a whole.

and, if we assume—as we did—that the sequences of wavelet coefficients at different scales are uncorrelated (**ID1**), we obtain that

$$\mathrm{Var}\ \hat{\gamma} = 3D \sum_j \sigma_j^2\, w_j^2. \tag{4.29}$$

This is an important result, since this variance only depends on the amount of data (through the n_j), but not on the data itself (the y_j), not on the chosen wavelet (provided that its number of vanishing moments is high enough), nor on the actual (unknown) value of γ. It can be shown [1, 34] that, for data of size n, Var $\hat{\gamma}$ decreases as $1/n$ in the limit of large n_j at each scale j under consideration. Moreover, the Cramér-Rao lower bound is attained in this case, and the estimate $\hat{\gamma}$ is (asymptotically) normally distributed [1, 34]. Under the assumptions made, this permits therefore to associate confidence intervals with the points on which the linear regression is performed, thus allowing for designing tests aimed at justifying the relevance of a linear fit, as well as determining the range of scales on which such a fit makes sense [6].

Estimating γ may not be the only issue when characterizing a scaling process. In particular, it may be most useful to also estimate a "magnitude" parameter (σ^2 in eq.(3.13), or c_f in eq.(3.19)), which clearly measures in practical situations the *quantitative* importance of scaling effects in observed data. According to (4.24), the magnitude parameter c_f is related to the intercept of the fitted straight line in the Logscale Diagram. Unfortunately, this intercept also involves the quantity C, which depends on both the chosen wavelet and the actual scaling parameter γ (see eq.(3.20)). A two-step procedure can however be used [34], which consists first in estimating $\widehat{c_f C}$ from the intercept of the regression and, second, in estimating $\hat{c}_f$ as $\widehat{c_f C}/\hat{C}$, where $\hat{C}$ is, given $\hat{\gamma}$, an estimator of the integral $C(\gamma, \psi_0)$. The estimator constructed this way can be shown to be asymptotically unbiased, efficient and log-normally distributed [34].

4.2 From Gaussian to stable processes

In the case of processes with arbitrary variance (not necessarily finite), the analysis conducted so far cannot be followed. However, it has been shown previously that a number of key properties of wavelet transformed processes still apply in very general situations, irrespectively of the existence of second-order moments. In particular, the reproduction identity **P1** (eq.(3.6)) for H-ss processes guarantees that (3.15) holds. This allows us again to estimate H by measuring a slope in a log-log plot, with the notable difference that $\mathbf{E} \log_2 |d_X(j,k)|$ needs now to be estimated. If we restrict ourselves to the class of H-sssi processes (such as H-sssi SαS processes), the stationarization property of the wavelet transform still holds, thus suggesting to make use of the quantity

$$Y_j := 3D \frac{1}{n_j} \sum_{k=3D1}^{n_j} \log_2 |d_X(j,k)| \tag{4.30}$$

as an estimator of $\mathbf{E}\log_2|d_X(j,k)|$, with the obvious consequence that

$$\mathbf{E}Y_j = 3D(H+1/2)j + \mathbf{E}\log_2|d_X(0,0)|. \tag{4.31}$$

We know [18] that, whereas SαS processes (and, hence, the sequences of their wavelet coefficients at any scale) have infinite variance, the log-coefficients $\log_2|d_X(j,k)|$ have finite second-order statistics. Moreover, we have seen that, in the specific LFSM case, the covariance of those log-coefficients can be made arbitrarily small when using a wavelet with a high enough number of vanishing moments (see eq.(3.16))[2]. This results in behaviors similar to what had been established for second-order processes, the exact decorrelation idealization **ID2** leading for the variance of Y_j to the closed form expression:

$$\text{Var } Y_j = 3D\left(1+\frac{2}{\alpha^2}\right)\frac{(\pi\log_2 e)^2}{12}\frac{1}{n_j}. \tag{4.32}$$

In this more general context of processes with a possibly infinite variance, stationarization and almost decorrelation are again the key ingredients for guaranteeing a relevant estimate of the scaling parameter H. More precisely, the estimate $\hat{H}$ follows from the linear relation (4.31) and can be written as:

$$\hat{H} := 3D\sum_j w_j Y_j - \frac{1}{2}, \tag{4.33}$$

with the weights w_j defined as in (4.27).

Since we have, by construction, $\sum_j w_j = 3D0$ and $\sum_j jw_j = 3D1$, it is easy to check that $\mathbf{E}\hat{H} = 3DH$, a result which is exact regardless of the data length and of $0 < \alpha \leq 2$.

Assuming (**ID2**) an exact decorrelation between the log-coefficients, one obtains:

$$\text{Var } \hat{H} = 3D\left(1+\frac{2}{\alpha^2}\right)\frac{(\pi\log_2 e)^2}{12}\sum_j\frac{w_j^2}{n_j}. \tag{4.34}$$

This variance is minimum for the choice $a_j = 3D\text{Var } Y_j \sim 1/n_j$. Since n_j, the number of data points available at scale j, behaves basically as $n_j = 3D2^{-j}n$ for a total number n of data points, we see that Var $\hat{H}$ decreases as $1/n$, regardless of the possible LRD nature of the analyzed process. Finally, numerical investigations support the claim that the wavelet-based estimate $\hat{H}$ is (asymptotically) normally distributed, thus allowing the derivation of confidence intervals from the knowledge of the variance [4].

[2] It has however to be noted that, for a given H, the decorrelation effect requires larger N's for smaller α's, since N has basically to behave as $1/\alpha$.

4.3 Additional benefits

Besides the effectiveness of wavelet-based tools (such as the Logscale Diagram) for the analysis of "perfect" scaling processes, a number of additional benefits are offered by the proposed framework, which can prove of particular interest in more realistic situations.

Robustness to non-Gaussianity. Among the various assumptions which have been made for deriving analytic performance of the estimators, one concerns the Gaussianity of wavelet coefficients. Except for processes which are by themselves Gaussian, such an assumption has no reason to be relevant, but it has been shown in [6] that formulæ of the type (4.28) and (4.29) still apply in non-Gaussian cases, up to correction terms which can be explicitly included in the analysis, or even neglected in the limit of large data samples.

Insensitivity to polynomial trends. A common limitation of standard techniques aimed at scaling processes is that their results can be severely impaired by perturbations with respect to the ideal model of a "perfect" scaling process. This is especially the case when deterministic trends are superimposed to a process of interest, with consequences such as invalidating the stationary increments property of an actual LRD process, or mimicking LRD correlations when added to a short-range dependent process [3]. Wavelets are a nice and versatile solution to this crucial issue, since they offer the possibility of being blind to polynomial trends. As it has already been said, a wavelet needs, for being admissible, to be zero-mean (i.e., to have one, zero-th order, vanishing moment) or, in other words, to be orthogonal to constants. A natural extension of this requirement consists in imposing a higher number of vanishing moments, say N, so that the resulting wavelet be blind to polynomials up to orders $p \leq N-1$.

Computational efficiency. The analysis of scaling processes is often faced (as it is the case with network traffic) with enormous quantities of data, thus requiring methods which are efficient from a computational point of view. Because of their multiresolution structure and their pyramidal implementation, wavelet-based methods are associated with fast algorithms overperforming FFT-based algorithms (complexity $\mathrm{O}(n)$ *vs.* $\mathrm{O}(n \log n)$, for n data points).

5 Conclusion

Wavelet analysis has been shown to offer a natural and unified framework for the characterization of scaling processes. One of its main advantages is that it allows for a unique treatment of a large variety of processes, be they self-similar, fractal, long-range dependent, Gaussian or not, ... Further extensions can even be given to the results presented so far. For instance, it has been shown in [2, 32] how *point processes* with fractal characteristics can enter the same framework, on the basis of versatile generalizations of the Fano factor, which is of common use in this context. Other developments have been conducted for taking into account situations where *one* scaling exponent is not sufficient for a proper description and modelling. In such *multifractal* scenarii [21, 28], analyses based on higher-order statistics of wavelet coefficients are a key for efficiently computing quantities such as Legendre singularity spectra [10].

As far as applications are concerned, *telecommunications network traffic* is a domain in which wavelet-based analysis proved most useful in different respects. First, it permits to evidence in a well-controlled way the non-standard features of traffic (self-similarity, long-range dependence) which have been observed since a recent past, and to accurately estimate the corresponding scaling parameters [3, 6]. Second, its versatility allows for using, in a common framework, different aspects of traffic data (work process, interarrival delays, loss, ...) [3], as well as for being a basis for new (wavelet-based) models [29]. Finally, its computational efficiency allows for a fast processing (with possible on-line implementations) of the very large amounts of data which are commonly encountered in this domain.

Other areas have benefited from the same tools. Among them, one can cite *turbulence*, a field in which the identification of scaling phenomena is of primary importance [20]. The equations (Navier-Stokes, ...) that govern fluid motion are characterized by nonlinear terms that insure energy transfers from the injection to the dissipation scales, giving birth to the so-called "inertial scaling range". Scaling exponents within this range have been estimated using wavelet tools generalized to statistics different from two (see, e.g., [13]). The multifractal formalism has also been considered to model scaling within the inertial range, and refined wavelet tools have been widely used in this kind of analysis [10]. More recently, cascade models [14] have been invoked to describe a much wider variety of scaling phenomena where scaling on the time series themselves may even be barely observable. These cascades were recently rephrased in the wavelet framework [8] and used to give new insights on scaling in turbulence [7, 15].

In *biology*, scaling processes were used to model, for instance, spike trains discharges of neurons. In this context, wavelet tools such as those presented here were used to estimate the scaling parameters, directly from the observed point-processes, and to decide whether the responses are pathological or not [32].

More recently, the proposed wavelet analysis of scaling has been applied to the study of *extremal models* [31]. Such models describe in a common language a large variety of physical phenomena (wetting front motion, roughening of crack front in fracture, solid friction, fluid invasion in porous media, ...). In all these situations, key ingredients are the competition between elastic restoring force and nonlinear pining forces, and the assumption that the dynamic of the systems is controlled, at each time step, by its *extremal* part. Wavelet analysis allowed to evidence, amongst the times series produced by extremal models, the existence of temporal statistical dependence and self-similarity, to estimate the corresponding parameters and to show the relevance of the use of LFSM models.

For any of these applications, using wavelets is in some sense "natural", in terms of a structural adequacy between the *mathematical* framework they offer (multiresolution) and the *physical* nature of the processes under study (scaling).

Acknowledgments. This work has been supported in part by CNRS *via* its "Programme Télécommunications", under grant # TL 97035. Thanks also to L. Delbeke and D. Veitch, with whom several of the results presented here have been derived.

References

1. P. Abry, P. Gonçalvès and P. Flandrin (1995): Wavelets, spectrum estimation and $1/f$ processes. In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Lectures Notes in Statistics, **105**, pp. 15–30.
2. P. Abry and P. Flandrin (1996): Point processes, long-range dependence and wavelets. In *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, eds., pp. 413–437.
3. P. Abry and D. Veitch (1998): Wavelet analysis of long-range dependent traffic. *IEEE Trans. on Info. Theory*, **44** (1), pp. 2–15.
4. P. Abry, L. Delbeke and P. Flandrin (1999): Wavelet-based estimator for the self-similarity parameter of α-stable processes. *IEEE-ICASSP-99*, Phoenix (AZ).
5. P. Abry, L. Delbeke, P. Flandrin and B. Pesquet-Popescu (1999): Wavelet coefficients of self-similar stable motions. Preprint.
6. P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch (1999): Wavelets for the analysis, estimation and synthesis of scaling data. To appear in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, eds., Wiley Interscience.
7. A. Arnéodo, J.F. Muzy and S.G. Roux (1997): Experimental analysis of self-similar random cascade processes: application to fully developped turbulence. *J. Phys. II France*, **7**, pp. 363–370.
8. A. Arnéodo, E. Bacry and J.F. Muzy (1998): Random cascades on wavelet dyadic trees. *J. Math. Phys.*, **39** (8), pp. 4142–4164.
9. R. Averkamp and C. Houdré (1998): Some distributional properties of the continuous wavelet transform of random processes. *IEEE Trans. on Info. Theory*, **44** (3), pp. 1111–1124.
10. E. Bacry, J.F. Muzy, and A. Arnéodo (1994): Singularity spectrum of fractal signals from wavelet analysis: exact results. *J. Stat. Phys.*, **70**, pp. 635–674.
11. J. Beran (1994): *Statistics for Long-Memory Processes.* Chapman and Hall, New York.
12. S. Cambanis and C. Houdré (1995): On the continuous wavelet transform of second-order random processes. *IEEE Trans. on Info. Theory*, **41** (3), pp. 628–642.
13. R. Camussi and G. Guj (1997): Orthonormal wavelet decomposition of turbulent flows: intermittency and coherent structures. *J. Fluid Mech.*, **348**, pp. 177–199.
14. B. Castaing, Y. Gagne and M. Marchand (1993): Log-similarity for turbulent flows ? *Physica D*, **68**, pp. 387–400.
15. P. Chainais, P. Abry and J.F. Pinton (1999): Intermittency and coherent structures in a swirling flow: a wavelet analysis of joint pressure and velocity measurements. Preprint, submitted to *Phys. Fluids.*
16. L. Delbeke (1998): *Wavelet-Based Estimators for the Scaling Index of a Self-Similar Process with Stationary Increments.* PhD Thesis, KU Leuven.
17. L. Delbeke and P. Abry (1997): Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion. Preprint, submitted to *Stoch. Proc. and their Appl.*.

18. L. Delbeke and J. Segers (1998): The covariance of the logarithm of jointly symmetric stable random variables. Preprint.
19. P. Flandrin (1992): Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Info. Theory*, **38** (2), pp. 910–917.
20. U. Frisch (1995): *Turbulence: The Legacy of A.N. Kolmogorov.* Cambridge Univ. Press, Cambridge.
21. J. Lévy-Véhel (1996): Fractal approaches in signal processing. In *Fractal Geometry and Analysis*, H.G. Peitgen and R. Voss, eds., World Scientific, Singapore.
22. S. Mallat (1997): *A Wavelet Tour of Signal Processing.* Academic Press, San Diego.
23. E. Masry (1993): The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Trans. on Info. Theory*, **39** (1), pp. 260–264.
24. B.B. Mandelbrot and J.W. van Ness (1968): Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, **10**, pp. 422–437.
25. C.L. Nikias and M. Shao (1995): *Signal Processing with Alpha-Stable Distributions and Applications.* Wiley and Sons, Inc., New York.
26. B. Pesquet-Popescu (1998): *Modélisation bidimensionnelle de processus non stationnaires et application à l'étude du fond sous-marin.* Thèse de Doctorat, Ecole Normale Supérieure de Cachan.
27. B. Pesquet-Popescu (1999): Statistical properties of the wavelet decomposition of some non-Gaussian self-similar processes. *Signal Proc.*.
28. R.H. Riedi (1995): An improved multifractal formalism and self-similar measures, *J. Math. Anal. Appl.*, **189**, pp.462–490.
29. R. Riedi, M.S. Crouse, V.J. Ribeiro and R.G. Baraniuk (1999): A multifractal wavelet model with application to network traffic. To appear *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications.
30. G. Samorodnitsky and M.S. Taqqu (1994): *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York.
31. A. Tanguy, P. Abry, S. Krishnamurthy and S. Roux (1999): Extremal models and linear fractional stable motion. Preprint, submitted to *Phys. Rev. Lett.*.
32. M.C. Teich, C. Heneghan, S.B. Lowen and R.G. Turcott (1996): Estimating the fractal exponent of point processes in biological systems using wavelet- and Fourier-transform methods. In *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, eds., pp. 383–412.
33. A.H. Tewfik and M. Kim (1992): Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. on Info. Theory*, **38** (2), pp. 904–909.
34. D. Veitch and P. Abry (1999): A wavelet-based joint estimator of the parameters of long-range dependence. To appear *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications.