Scaling and Wavelets: An Introductory Walk*

Patrice Abry

CNRS, UMR 5672, Physics Lab., Ecole Normale Supérieure de Lyon, 46, allée d'Italie, F-69364, Lyon cedex 7, France

Abstract. This chapter offers, first, an introductory walk through the notions related to scaling phenomena and intuitions behind are gathered to formulate a tentative definition. Second, it introduces the mathematical model of self-similar processes with stationary increments, understood as the canonical reference to describe scaling. Then, it shows how and why the wavelet transform constitutes a powerful and relevant tool for the analysis (detection, identification, estimation) of self-similarity. It is finally explained why self-similarity is too restrictive a model to account for the large variety of scaling encountered in empirical data and a review of the various models related to scaling – long range dependence, local Hölder regularity, fractal and multifractal processes, multiplicative or cascade processes – is proposed. Their interrelations and differences, as well as estimation issues, are discussed. A set of *Matlab* routines has been developed to implement the wavelet-based analysis for scaling described here. It is available at `www.ens-lyon.fr/~pabry`.

1 Introduction and Motivation

1.1 Scaling Phenomena

Power laws, scaling laws, scaling phenomena or, simply, scaling, recently became a very fashionable topic. Indeed, scaling behaviors were observed or studied or used as description paradigms in a large collection of research works covering a wide variety of different domains or applications. It is worth noting that those applications may be related either to natural phenomena or to data resulting from mankind's activities. For the first category, one can, for instance, mention hydrology [10] with the study of variabilities of water levels in rivers, hydrodynamic with the study of developed turbulence [24,20], statistical physics with the study of systems having long range interactions [44], microelectronics with $1/f$ noises in semi-conductors [9,23], geophysics and fault repartitions or geological layers [43,42], biology and physiology [52] with human rhythms variabilities, heart beat [45] or gait [26] for instance. For the second category, one can find human geography and population repartition in cities or continents [19], information flows on network and mainly computer network teletraffic [35], stock

* This article reviews works done in past years and still in progress, mainly in collaboration with D. Veitch, EMUlab, Univ. of Melbourne, Australia, and with Pierre Chainais, Patrick Flandrin, ENS Lyon, France. These works have been partially supported by the French CNRS grant *TL99035, Programme Télécommunications* and French MENRT grant *ACI jeune Chercheur 2329.*

G. Rangarajan, M. Ding (Eds.): LNP 621, pp. 34–60, 2003.

market volatility or currency change rates fluctuations [29,51]. Very often, scaling in data is a crucial observation since it can be tied to key properties of the systems (e.g., high volatility in markets or large waiting times and congestions in computer network traffic or pathologies in body rythms...).

The notion of scaling, however, remains defined poorly or in a loose way and may be related to a variety of different properties of a system or a process. A possible common tentative definition for scaling can be formulated through a negative statement: there is no characteristic scale (of time or space...) in the studied system or process. In other words, this is no longer possible to identify any scale that plays a privileged role compared to others, or equivalently, all scales play identical roles and are of equal importance in the dynamics of the analyzed system or process. Scaling, therefore, correspond to situations where the *whole* can not be (statistically) distinguished from any of its *subpart*. This is commonly associated to the picture of geometric fractal object, obtained through the iteration of an identical construction procedure.

1.2 Random Walks and Self-similarity

From a data analysis (or signal processing) point of view, scaling in time series implies that the usual intuitive search techniques for characteristic scales are to abandoned and replaced instead by new ones aiming at evidencing relations, mechanisms between scales or involving a wide range of scales. This also means to abandon the use of models relying on the existence and definition of a characteristic scale (e.g, Markow chains, Poisson models, models with exponential autocorrelation functions,...). The canonical reference mathematical models for scaling are that of Random Walks and Self-Similar processes, and more particularly, the popular fractional Brownian motion. This will be introduced with details in Sect. 2.1.

1.3 Wavelets

The practical use and analysis of self-similar processes present however two major difficulties: they are non stationary and are characterized by long range dependence or long term correlations or long memory. Such statistical features turn the analysis of self-similar processes into an uneasy and non standard task. To overcome such difficulties, it has been shown recently in a collection of papers [32,33,46,11,31,2,4,47] that wavelet transforms constitute ideal tools for the analysis of scaling. Wavelet analysis will be introduced in Sect. 2.2. More precisely, the wavelet analysis can be considered as "matched" to self-similar processes in the sense that wavelet coefficients exactly reproduce, from scale to scale, the self-replicating statistical structure of such processes. This will be made explicit in Sect. 2.3. Section 2.4 explains how those statistical properties of the wavelet coefficients are to be used to design tools for the analysis (detection, identification and estimation) of scaling phenomena.

• Beyond self-similarity

Self-similarity is a mathematically well-behaved model. Its definition however implies numerous constraints seen as limitations in the practical 0 of empirical data. Obviously, it cannot account for the large variety of scaling existing amongst actual empirical data. Section 3 will therefore allow a larger part to variations around self-similarity commonly used to describe scaling, such as long range dependence, fractal sample path, multifractal processes, multiplicative processes, infinitely divisible processes and cascades... and will underline their mutual interrelations, common denominators and differences.

• Note

This chapter mainly intends to be an introductory walk in the land of scaling phenomena and scaling laws. Its aim is to provide the reader with a synthetic and comprehensive overview of their related mathematical models and with a quick start to their wavelet-based analysis. Technical details as well as mathematical proofs can be found in references given along the text.

• Matlab Routines

All the analysis procedures (detection, identification, estimation) described here as well as synthesis ones presented elsewhere are implemented in Matlab routines available at `www.ens-lyon.fr/~pabry` or `www.emulab.ee.mu.oz.au/~darryl`.

2 Self-similarity and Wavelets

2.1 Self-similarity

• Random Walks

The simplest model that can be thought of to model scaling phenomena is that of the standard random walk commonly involved is the pedagogical(!) description of a drunkard walk or more generally in that of diffusion phenomena. Let $X(t)$ denote some physical quantity of interest, a random walk consists in going from position X at time t to position $X + \delta X$ at time $t + \tau$ by making an elementary step (or increment) $\delta X(\tau, t)$:

$$X(t + \tau) = X(t) + \delta X(\tau, t)\,, \forall \tau \geq 0\,. \tag{1}$$

Without loss of generality, we assume in this whole text that $X(0) \equiv 0$ and $\mathbb{E}\delta X(\tau, t) \equiv 0, \forall\, \tau, \forall t$.

For normal (drunkard and) diffusion, one usually assumes for the increments the following statistical properties:

A1 : The $\{\delta X(\tau, t), t \in \mathbb{R}\}$ form random processes that are stationary with respect to the t variable. Their distributions are identical, do not depend on t but functionally depend on τ,

A2 : The steps $\{\delta X(\tau, t), t \in \mathbb{R}\}$ are mutually independent, i.e., for $t_1 \leq t_2 \leq t_3 \leq t_4$,

$$p_2[(X(t_4) - X(t_3)).(X(t_2) - X(t_1))] = p_1(X(t_4) - X(t_3))\, p_1(X(t_2) - X(t_1)).$$

where the $p_i(\cdot)$ denote the (joint) probability density functions. This means that the random variables $\delta X(\tau, t)$ and $\delta X(\tau', t')$ are statistically independent as soon as $t' > t + \tau$. In other words, once the step $\delta X(\tau, t)$ has been performed, one gains no extra information on the next following step.
A3 : The $\{\delta X(\tau, t), t \in \mathbb{R}\}$ form jointly Gaussian processes.

Though apparently simple and intuitive, those three properties together impose severe constraints on the walk X, and even define it in a unique manner as the ordinary Brownian motion. For instance, they imply a linear behavior of the variance of X (or of its increments) with respect to time:

$$\mathbb{E}|X(t) - \underbrace{X(0)}_{\equiv 0}|^2 = 2D_X|t|, \text{ or, equivalently } \mathbb{E}|\delta X(\tau, t)|^2 = 2D_X|\tau|. \quad (2)$$

Those behaviors, known as the celebrated Einstein's relations, constitute the signature of scale invariance or scaling phenomenon in the random walk: no characteristic scale exists or can be identified that would limit, or bound, or indicate a cut-off in the development of the walk nor plays any specific role.

However, empirical data very often exhibit significant departures from those linear behaviors. The so-called anomalous diffusion phenomena, for instance, are characterized by:

$$\mathbb{E}|X(t) - X(0)|^2 = 2D_X|t|^\gamma, \ 0 < \gamma < 2, \quad (3)$$

which can be seen as generalizations to (2) above, read as a power-law with exponent that takes the specific value 1. To account for the departure from a linear behavior, to bypass limitations resulting from **A1**, **A2** and **A3** and more generally to enlarge the framework of ordinary random walk and Brownian motion, one is naturally lead to that of self-similarity.

• Self-similar processes

A process X is said to be statistically self-similar, with self-similarity parameter $H > 0$, if [41]:

$$\forall c > 0, \{c^H X(t/c), t \in \mathbb{R}\} \stackrel{fdd}{=} \{X(t), t \in \mathbb{R}\} \quad (4)$$

where $\stackrel{fdd}{=}$ means equality of all the finite dimensional distributions. This means that the sample paths (t, X) of the process $X(t)$ and those $(t/c, c^H X)$ of the process $c^H X(t/c)$ are statistically indistinguishable. In other words, the process X is statistically similar to any of its dilated templates. Therefore, no characteristic scale of time can be identified on these processes, self-similarity is hence a model for scaling behavior. This is illustrated on Fig. 1.

A major consequence of self-similarity also lies in the fact that the moments of the process, when they exist (we do nit discuss existence issues in this text), behave as power laws with respect to time,

$$\mathbb{E}|X(t)|^q = \mathbb{E}|X(1)|^q |t|^{qH}, \quad (5)$$

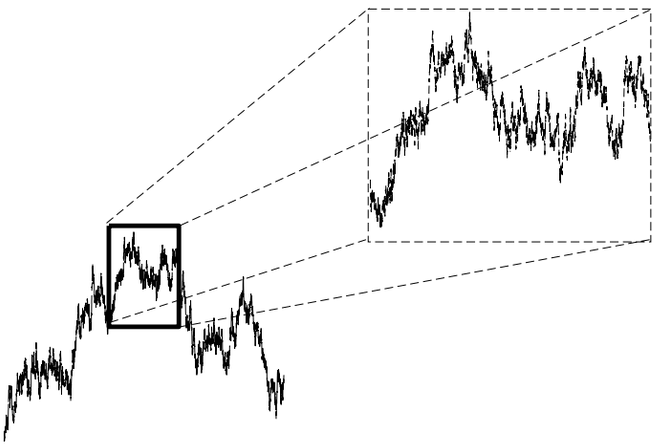

Fig. 1. Sample path of a self-similar process. Starting with the sample path of a self-similar process, if one performs a dilation of the time axis of factor $1/c$ and a dilation on the amplitude axis of factor c^H, one obtains a new sample path that is (statistically) indistinguishable from the original one.

whose exponents are all controlled by the self-similarity parameter H. Besides the connection between scaling and power law, those relations also show that self-similar processes are non stationary ones.

• Self-similar processes with stationary increments

Because with actual empirical data, the use of non stationary models is a major difficulty, one often restricts the class of self-similar processes to that of self-similar processes with stationary increments (hereafter, H-sssi processes). A process is said to posses stationary increments if:

$$\{\delta X(\tau,t) \equiv X(t+\tau) - X(t), t \in \mathbb{R}\} \stackrel{fdd}{=} \{\delta X(\tau,0) \equiv X(0+\tau) - X(0) \equiv X(\tau)\}, \forall \tau. \tag{6}$$

Self-similarity and stationary increments together yield that the fundamental equation (5) above can be rewritten on the increments (for all finite moments):

$$\mathbb{E}|X(t+\tau) - X(t)|^q = \mathbb{E}|X(1)|^q |\tau|^{qH}. \tag{7}$$

In keeping with the general framework of random walks (cf. (1)), the class of H-sssi processes enlarges that of ordinary Brownian motion and ordinary random walk through the replacements of assumptions **A1, A2** and **A3** with:
B1: The $\{\delta X(\tau,t), t \in \mathbb{R}\}$ form random processes that are stationary with respect to the t variable.
B2: The steps $\{\delta X(\tau,t), t \in \mathbb{R}\}$ satisfy, $\forall H > 0$,

$$\forall c > 0, \{c^H \delta X(\tau/c, t/c), t \in \mathbb{R}\} \stackrel{fdd}{=} \{\delta X(\tau,t), t \in \mathbb{R}, \}. \tag{8}$$

Obviously, **B1** is identical to **A1** while **B2** gathers **A2** and **A3** . This latest implies that the steps of the random walk no longer need to be independent (they can even exhibit long memory as detailed below) nor Gaussian (they can have any *stable* [41] marginal distribution, i.e., infinite variance distributions, see section 3). B. Mandelbrot, who significantly contribute to the promotion of the use of H-sssi processes (fractional Brownian motion [27], Lévy stable motions [29]) in applications (turbulence, finance, ...), named those two variations the *Joseph's effect* and the *Noah's effect*, respectively [29], after the celebrated biblical characters. One can, moreover, consider situations, problems or time series where the two difficulties occur jointly and simultaneously. Technically, H-sssi processes can be written as [41]:

$$X(t) = \int_{\mathbb{R}} f(t,u)M(du), \tag{9}$$

where $M(du)$ denotes an α stable stochastic measure (where α, $0 < \alpha \leq 2$, stands for the stability index, and $\alpha = 2$ corresponds to the Gaussian case). This definition means that the process X can be read as a weighted sum of independent α stable random variables. Technically, this implies that, for a fixed t, $X(t)$ is a α stable random variable, M therefore controls the marginals of the process, or in other words, its static properties. The weighting kernel function $f(t,u)$ controls the statistical dependences of X and hence its joint statistics, or in other words, its dynamical properties. For well chosen forms of the kernel, [18,34,41], X is a self-similar process with stationary increments. For instance, the celebrated fractional Brownian motion (fBm), which is, up to a normalization, the only Gaussian H-sssi process, corresponds to the choice of a Gaussian measure: $M(du) = dB(u)$ and of the specific *fractional integration* kernel $f(t,u) = (t-u)_+^d - (-u)_+^d$, where $(u)_+ = u$ if $u \geq 0$ and 0 else and where $d = H - 1/2$ controls the long memory of the process (hence its dependence structure) [27]. Lévy stable processes (that include the ordinary Brownian motion) are characterized by independent increments and correspond to $f(t,u) = 1$ if $0 \leq u \leq t$ and 0 elsewhere. Linear fractional stable motion is characterized by an infinite variance stable distribution $M(du)$ (i.e., α strictly smaller than 2) and by the same kernel as that of fBm with $d = H - 1/\alpha$; it therefore gathers simultaneously the two difficulties mentioned above.

• Self-similar processes with stationary increments and finite variance

Self-similarity, stationary increments and finite variance together impose that [41,27]:

$$0 < H < 1. \tag{10}$$

Hence, for H-sssi processes with finite variance, the choice $q = 2$ in the relation above (5) accounts for the *anomalous* behavior reported in (3) with $0 < \gamma = 2H < 2$.

Moreover, for H-sssi process X with finite variance (and with zero mean and $X(0) \equiv 0$), one can show that the covariance function necessarily takes the

following form:

$$\mathbb{E}X(t)X(s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{11}$$

with $\sigma^2 = \mathbb{E}|X(1)|^2$ and that the correlation function of the increment process reads:

$$\mathbb{E}\delta X(\tau,t)\delta X(\tau,t+s) = \frac{\sigma^2}{2}\left(|\tau+s|^{2H} + |\tau-s|^{2H} - 2|s|^{2H}\right). \tag{12}$$

• Long range dependence or long term correlations or long memory

From this relation, one can infer the asymptotic behavior of the covariance function on the increment process in the limit of large s (i.e., $s \to +\infty, s \gg \tau$):

$$\mathbb{E}\delta X(\tau,t)\delta X(\tau,t+s) \sim \frac{\sigma^2}{2}\, 2H(2H-1)\,\tau^2 s^{2(H-1)}. \tag{13}$$

Such a power law decrease of the covariance function above refers to a notion known as long range dependence, or long term correlations [10,41]. A stochastic stationary process Y is said to be long range dependent if its spectrum behaves asymptotically as a power law in the limit of small frequencies or, equivalently, if its covariance function behaves asymptotically in the limit of large lag as a power law:

$$\left.\begin{array}{r} \Gamma_Y(\nu) \sim C|\nu|^{-\gamma},\ |\nu| \to 0, \quad 0 < \gamma < 1, \\ \mathbb{E}Y(t)Y(t+s) \sim C'|s|^{-\beta},\ |s| \to +\infty,\ 0 < \beta = 1-\gamma < 1. \end{array}\right\} \tag{14}$$

This asymptotic power law decrease of the autocorrelation function is to be compared to the exponential one encountered in more usual processes (like Markov processes): an exponential decrease, by definition, implies a characteristic time while a power law behavior does not, meeting again the intuition of the absence of characteristic scale of time. For long range dependent processes, the autocovariance function decreases so slowly that its sum diverges, i.e., for $A > 0$,

$$\int_A^\infty \mathbb{E}Y(t)Y(t+s)ds = \infty.$$

This implies that the correlation between any two samples of the process cannot be neglected without missing something crucial in the analysis of the process, no matter how far apart one from the other they are. Long range dependence among its increments constitute a major difficulty in the analysis of a self-similar process.

Among the statistical community, where its formal definition was first proposed, long memory is often studied through the so-called aggregation procedure. Let $Y^{(T)}(t)$ be the version of the process Y aggregated in a window of size T:

$$Y^{(T)}(t) = \frac{1}{T}\int_{tT}^{(t+1)T} Y(u)du. \tag{15}$$

The long memory of the process $Y(t)$ results in a asymptotical power law behavior of the variance of the aggregated process:

$$\mathbb{E}|Y^{(T)}(t)|^2 \sim T^{-\beta}, T \to \infty, \tag{16}$$

A large family of estimators has been based on this asymptotic property.

Long memory can be defined independently from self-similarity and in itself, is a model for scaling observed in the limit of large scales (see Sect. 3 for details). However, increments of H-sssi processes, when $H > 1/2$, exhibit long memory as proven by (13) above. Therefore, the two notions are subtly related.

• $1/f$-Processes

So-called $1/f$-processes have also been widely used to model scaling phenomena. A stationary process is said to posses a $1/f$-spectrum if its spectrum behaves as a power law in a wide range of scales:

$$\Gamma_X(\nu) \sim C|\nu|^{-\gamma}, \, 0 \leq \nu_m \leq |\nu| \leq \nu_M, \nu_M/\nu_m \gg 1, \gamma > 0. \tag{17}$$

In this setting, all frequencies are playing equivalent roles. Moreover, such spectrum satisfy $\forall \lambda > 0, \Gamma_X(\lambda\nu) \sim \lambda^{-\gamma}\Gamma_X(\nu)$, hence the connection with scale invariance.

Despite its being a non-stationary process, fBm is often *naturally* considered as the reference for $1/f$-Processes. However, connections are somehow intricate... The increment process of fBm can be regarded as the output of a linear time invariant filter whose input is fBm and whose impulse response reads $\psi_{\delta\tau}(t) = \delta(t+\tau) - \delta(t)$:

$$\delta X(\tau, t) = (\psi_{\delta\tau}(\cdot) * X(\cdot))(t),$$

where $*$ stands for the convolution. The standard Fourier relation of the linear time invariant filter yields:

$$\Gamma_{\delta X(\tau,\cdot)}(\nu) = |\Psi_{\delta\tau}(\nu)|^2 \Gamma_X(\nu),$$

where $\Psi_{\delta\tau}(\nu) = 1 - \exp(\imath 2\pi\tau\nu)$. In the limit $|\tau\nu| \ll 1$, $|\Psi_{\delta\tau}(\nu)| \sim |2\pi\tau\nu|$ and $\Gamma_{\delta X(\tau,\cdot)}(\nu) \sim |\nu|^{-(2H-1)}$, and hence, from the relation above one *heuristically* or *qualitatively* associate to fBm a spectrum of the form $\Gamma_X(\nu) \sim |\nu|^{-(2H+1)}$, hence a $1/f$-spectrum. This correspondence has been formulated in various frameworks including that of wavelets, see, e.g., [32].

• Local regularity of the process

Exploring the other limit, that of fine scales, i.e., $s \ll \tau, s \to 0$, one obtains that the autocovariance function of X behaves as:

$$\mathbb{E}\delta X(\tau, t)\delta X(\tau, t+s) \sim \sigma^2 \, |\tau|^{2H} \, (1 - |\tau|^{-2H}|s|^{2H}). \tag{18}$$

Such a power law behavior again traces back to the absence of characteristic scale (in the limit of small scales). Since $\delta X(\tau, t)$ is a stationary process, the equation above also straightforwardly yields:

$$\mathbb{E}|\delta X(\tau, t+s) - \delta X(\tau, t)|^2 \sim \sigma^2 \, |s|^{2H}, s \to 0, \tag{19}$$

which gives indications with respect to the local (ir)regularity of the sample paths of X.

Indeed, local regularity of sample paths of stochastic processes or of functions is usually measured in terms of Hölder exponent: this consists in comparing X at time t against a power law function. A process X is said to have Hölder regularity $h \geq 0$ at time t if their exists a local polynomial $P_t(s)$ of degree $n = \lfloor h \rfloor$ and a constant C such that:

$$|X(t+s) - P_t(s)| \leq C\,|s|^h. \tag{20}$$

For $0 \leq h < 1$, the regular part of X at time t reduces to $P_t(t) = X(t)$, yielding the following regularity characterization:

$$|X(t+s) - X(t)| \leq C\,|s|^h. \tag{21}$$

Heuristically, the Hölder h exponent describes the roughness of the sample path of X: h between 0 and 1 indicates that the sample path is everywhere continuous but nowhere differentiable, h close to 1 betrays a smooth and regular behavior and conversely, h close to 0 implies sharp roughness and large variability. For $1 < h < 2$, the same arguments apply to the first derivative of the sample path, and so on.

Self-similarity with stationary increments and finite variance, and more precisely the central relation (7) for the increments together with Kolmogorov's regularity criterion[1] shows that the local regularity of each sample path of the fractional Brownian motion (for which all moments exist for $q > -1$) is constant along time and controlled by the parameter H: $h = H$. From (19) above, one sees that the same holds for the increment process of fractional Brownian motion. Processes with sample paths characterized by a constant Hölder exponent, are often referred to as *monofractal* processes. Monofractal processes constitute therefore a model for scaling observed in the limit of small scales. For further details, see e.g., [37].

2.2 Wavelet Analysis

• Continuous wavelet transform

The wavelet coefficients of the so called continuous wavelet transform (CWT) [17,30] are defined as the results of comparisons, by means of inner products, between the process to be analyzed X and a family of functions, the wavelets $\psi_{a,t}$:

$$T_X(a,t) = \langle X, \psi_{a,t} \rangle, \quad (a,t) \in (\mathbb{R}^+, \mathbb{R}). \tag{22}$$

[1] Kolmogorov's criterion (see for example [38]): If $\{X(t) : t \in R\}$ is a stochastic process with values in a complete separable metric space (S,d), and if there exists positive constants β, C, ϵ such that for all $s,t \in R$

$$\mathbb{E}d(X_s, X_t)^\beta \leq C|s-t|^{1+\epsilon}$$

then there exists a continuous version of X. This version is Hölder continuous of order θ for each $\theta < \epsilon/\beta$.

The wavelets are dilated and translated templates of a reference pattern ψ called the mother wavelet:

$$\psi_{a,t}(u) = \frac{1}{|a|}\psi(\frac{u-t}{a}). \tag{23}$$

Figure 2 shows dilated templates of a single mother wavelet. Note that some definitions prefer a $1/\sqrt{|a|}$ instead of $1/|a|$ normalization term, mainly because it ensures energy preservation. For the analysis of scaling phenomena, however, the choice $1/|a|$ is more convenient. The function ψ is usually required to be bounded and to have time and frequency supports that are either bounded or decrease very fast, jointly in both domains, time and frequency. To ensure that

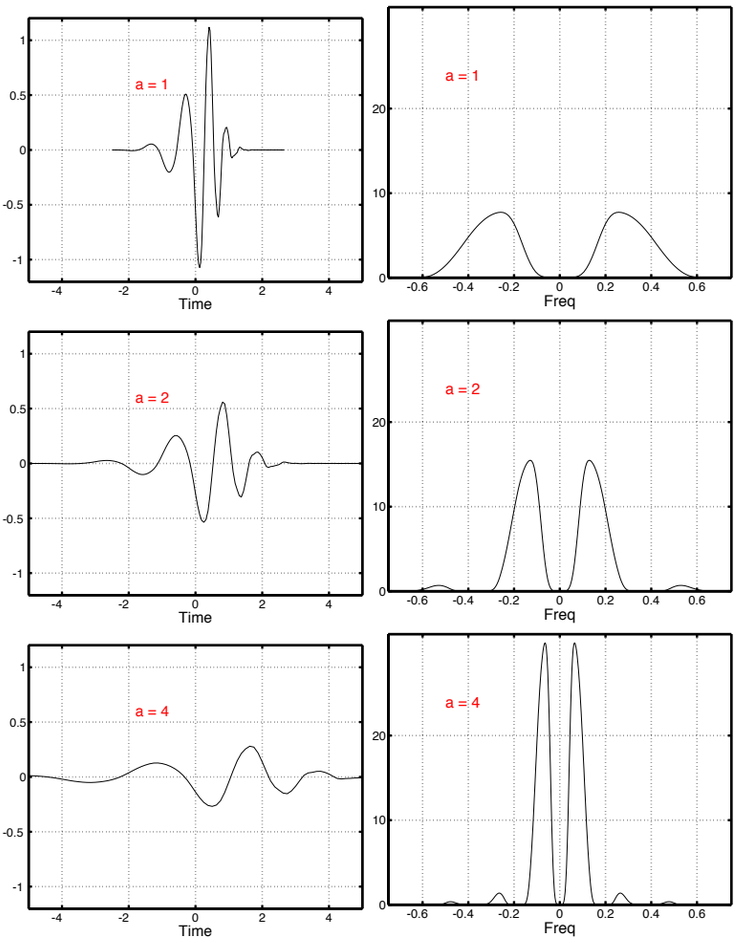

Fig. 2. Translated and dilated wavelets and their corresponding Fourier transforms. Left, dilated versions of the same mother-wavelet (Daubechies6) with dilation factors 1, 2 and 4, and, right, the corresponding Fourier transforms.

the wavelet transform is invertible, ψ moreover has to satisfy a so-called admissibility condition whose weak form reads:

$$\int_{\mathbb{R}} \psi(t)dt = 0. \tag{24}$$

Hence, ψ is a zero-mean function that has to oscillate and exist on a narrow time support. This is therefore a small wave: a *wavelet*.

This is because the mother wavelet ψ has a joint localization in time and frequency that the wavelet coefficients can be given the signification of frequency content of the data at a given time or of joint time-frequency content of the information in X.

The mother-wavelet is characterized by an integer N, called its number of vanishing (or zero) moments, defined as:

$$\forall m \in \{0, \ldots, N-1\}, \quad \int_{\mathbb{R}} t^m \, \psi(t)dt = 0, \; \int_{\mathbb{R}} t^N \, \psi(t)dt \neq 0. \tag{25}$$

The admissibility condition above (cf. (24)) imposes $N \geq 1$. This means that, for a mother-wavelet with N vanishing moments, the wavelet coefficients of a polynomial of degree $P < N$ are strictly zero. More generally, it means that the wavelet coefficients $T_X(a,t)$ of a process X at time t are only sensitive to the part of the local behavior of X which is more irregular than that of a polynomial of degree N. In other words, the higher N, the less the wavelet coefficients are sensitive to regular parts of the time series. The number of vanishing moments of the mother wavelets also controls the behavior of its Fourier transform at the origin:

$$|\Psi(\nu)| \sim |\nu|^N, \; |\nu| \to 0. \tag{26}$$

Figure 3 shows examples of wavelets with different vanishing moments.

• Discrete wavelet transform

One also defines the coefficients of the discrete wavelet transform (DWT) as a discrete subset of the $T_x(a,t)$:

$$d_X(j,k) = T_X(a = 2^j, t = 2^j k) = \langle X, \psi_{j,k} \rangle, \;\; (j,k) \in (\mathbb{Z}^+, \mathbb{Z}), \tag{27}$$

where $\psi_{j,k}(u) = 2^{-j}\psi(2^{-j}u - k)$. This discrete subset of points is usually called the dyadic grid of the discrete wavelet transform and this definition also usually implies that the mother wavelet is constructed through a multiresolution analysis [17,30]. The major interest in the use of the discrete wavelet transform lies in the facts that the $\{\psi_{j,k}, (j,k) \in (\mathbb{Z}^+, \mathbb{Z})\}$ form (possibly orthonormal) basis of $L^2(\mathbb{R})$ (so that the DWT is a non redundant representation of X) and that the $d_X(j,k)$ can be computed with fast pyramidal recursive algorithm whose computational costs is of the order of that of a fast Fourier transform. In all the methods and algorithms proposed here, the DWT is always used. For further details on wavelet transforms, the reader is referred to [17,30].

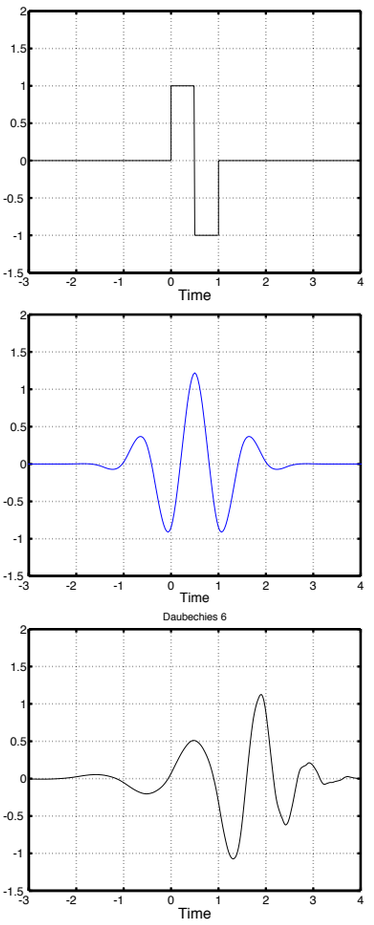

Fig. 3. Vanishing moments. Examples of mother wavelet with respectively, 1 (Haar, or poor man's, wavelet), 3 (B-spline wavelet) and 6 (Daubechies6) vanishing moments.

• Multiresolution analysis and scaling

Self-similar processes are usually studied through the analysis of their increments, and mainly (8) and so is local regularity (cf. (19)). Long memory in turn is analyzed through the aggregation procedure (cf. relation (16)). In spirit both techniques already are multiresolution analysis since they consist in analyzing the evolution of a resolution or scale dependent quantity (increments or aggregation) with that of the scale.

The wavelet based analysis of scaling proposes to replace increments or aggregation with wavelet coefficients and intuition behind this substitution can be formulated as follows. First, consider the wavelet coefficients of a process X obtained with a particular choice for the wavelet-like function $\psi(t) = \delta(t+\tau_0)-\delta(t)$

and note that they are identical to its increments:

$$X(t+\tau) - X(t) \equiv T_X(a,t), \text{ with } \tau = a\tau_0.$$

Increments can therefore be thought of as a specific example of wavelet coefficients obtained from a particular mother wavelet, which has a poor spectral localization and only one vanishing moment (i.e., $N = 1$). Moreover, let us note that selecting a mother-wavelet with N vanishing moment amounts to compute increments of order N (i.e., increments of increments of ...).

Consider now the wavelet coefficients of a long range dependent process Y, obtained with a particular choice for the mother wavelet $\psi(t) = 1$ for $-T_0/2 < t < T_0/2$ and 0 elsewhere[2] and note that they are identical to the aggregated process:

$$Y^T(t) \equiv T_Y(a,t), \text{ with } \tau = aT_0.$$

Aggregation can therefore be thought of as a specific case of wavelet decomposition which has a poor spectral localization and zero vanishing moment (i.e., $N = 0$).

While the increments and the aggregation procedure compute differences and averages respectively, a wavelet, being a band-pass function, naturally performs a difference of averages and therefore gathers the two techniques in a single one. And not only, the wavelet transform unifies these two techniques in a single framework but it also brings generalization, versatility and robustness through the choice of the mother-wavelet. One can, indeed, naturally think of using mother-wavelets with better joint time and frequency localizations and higher numbers of vanishing moment, resulting in better statistical properties, this will be further discussed in Sect. 2.4.

SUMMARY. All what the reader unfamiliar with wavelets needs to have in mind to follow the remainder of this text is that the relevance of the wavelet transform for the analysis of self-similar processes relies on two ingredients:

I1) the wavelet basis is designed from a dilation operator, $\psi_{a,0}(u) = \frac{1}{|a|}\psi(\frac{u}{a})$;

I2) the mother wavelet is characterized by a strictly positive integer N, its number of vanishing moments, cf. (25).

2.3 Self-similarity and Wavelets: Theory

Let X be a H-sssi process. Its DWT coefficients $d_X(j,k)$ and its CWT coefficients $T_X(a,t)$ have the following statistical properties (for proofs, see, e.g., [8,32,11,46,31]).

P1 Self-Similarity: The $d_X(j,k)$ (and the $T_X(a,t)$) reproduce, in an exact manner, the self-similarity of the process:

$$\{2^{-jH} d_X(j,k), k \in \mathbb{Z}\} \stackrel{d}{=} \{d_X(0,k), k \in \mathbb{Z}\}. \tag{28}$$

[2] Strictly speaking, this *box* or *indicator* function is not a wavelet, since this is not a band pass function.

$$\forall c > 0, \{c^H T_X(a/c, t/c), t \in \mathbb{R}, a \in \mathbb{R}^+\} \stackrel{fdd}{=} \{T_X(a,t), t \in \mathbb{R}, a \in \mathbb{R}^+\}. \tag{29}$$

These two relations result, fundamentally, from the fact that wavelets are designed using a dilation operator (ingredient I1 above). It is, moreover, interesting to note that this last relation has strong and obvious analogies to that satisfied by increments (cf. (8)).

P2 Non Stationarity: Though self similar processes are non stationary, their $\{d_X(j,k), k \in \mathbb{Z}\}$ form stationary sequences at each octave j. Identically, their $\{T_X(a,t), t \in \mathbb{R}\}$ form stationary processes at each scale a. This is again analogous to the stationarity of the increments and this is deeply connected to the fact that $N \geq 1$ (cf. ingredient I2 above).

P3 Long Range Dependence: It can be shown that the covariance function of any two wavelet coefficients on the dyadic grid can be asymptotically bounded as, $|2^j k - 2^{j'} k'| \to +\infty$,

$$|\text{Cov } d_X(j,k), d_X(j',k')| \leq C|2^j k - 2^{j'} k'|^{-2(N-H)}. \tag{30}$$

This shows the key role played by the number of vanishing moments N. Increasing N allows to increase the rate of decrease of the covariance function and therefore to reduce as much as desired the range of correlation amongst the wavelet coefficients. More precisely, it can be shown that, in the Gaussian case, if $N > H + 1/2$, long range dependence that exists amongst increments of X when $H > 1/2$ is turned into short range dependence. Note that obtaining this last property requires the simultaneous use of both ingredient I1 and I2 above.

P3ID Idealization: The "decorrelation effect", i.e., the reduction of the range of dependence of the wavelet coefficients under the increase of N, is idealized as follows:

any two wavelet coefficients of X, on the dyadic grid $\{d_X(j,k), k \in \mathbb{Z}, j \in \mathbb{Z}^+\}$, can be regarded as independent one from the other.

This idealization is used to provide approximated but analytical studies of the performance of the estimators proposed below.

Summary: Together, properties **P1** and **P2** imply that, for all finite moments:

$$\mathbb{E}|d_X(j,k)|^q \equiv C_q 2^{qjH}, \forall j, \tag{31}$$

where $C_q = \mathbb{E}|d_X(0,0)|^q$. Those relations are reminiscent of the fundamental equations (5) and (7) and yield the same constraints on the process X: power laws are to hold for all finite moments (e.g., for all $q > -1$ in the Gaussian case or for all $\alpha > q > -1$ in the α-stable case) and for all scales 2^j, moreover, all the exponents of the power laws are controlled by the single parameter H.

2.4 Self-similarity and Wavelets: Application

• Intuition

Self similar processes with stationary increments and finite variance are traditionally studied through their increments, mainly through (7), with $q = 2$.

However, the practical use of such an equation requires that the mathematical expectation be estimated, usually from a single observation of finite length. The existence of long term correlations amongst increments, however, substantially increase difficulties in the practical issue of estimation. For instance, the use of the standard sample variance estimator (that replaces statistical averages with time averages) presents remarkably poor statistical estimation performances [10].

The wavelet rewriting of (7), see (31) above with $q = 2$ can be used as a new starting point:

$$\mathbb{E}|d_X(j,k)|^2 = C2^{2jH}. \tag{32}$$

From **P1**, wavelet coefficients exactly reproduce self-similarity. From properties **P2** and **P3**, they form, at each scale 2^j, stationary sequences with short range and weak statistical dependence. On condition that N is high enough, they do not suffer any more from long range dependence. They are therefore statistically better behaved than increments and offer a versatile and convenient tool for the analysis of self-similarity. For instance, the standard sample variance estimator of the wavelet coefficients is a very satisfactory estimator for the ensemble average.

• Log-scale Diagram

To study scaling and more specifically self-similarity with wavelets, we define the following quantities:

$$\begin{array}{l} Y_j = \log_2 \left(\frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j,k)|^2 \right) , \\ \sigma_j^2 = \mathrm{Var}\ Y_j \, , \end{array} \tag{33}$$

where the n_js denote the numbers of wavelet coefficients available at octaves js. Then, we form the plots of the $Y_j = \log_2 \left(1/n_j \sum_{k=1}^{n_j} |d_X(j,k)|^2\right)$, together with their error bars (σ_j^2), versus $\log_2 2^j = j$. In those plots, that we proposed to call *Logscale Diagrams*, straight lines evidence the existence of self-similarity and the measurement of their slopes allows for an estimation of the parameter H. Figure 4 proposes examples of logscale diagrams for the sample paths of fractional Brownian motion and of a Long Range Dependent process.

• Estimation Issues

Precisely, the estimator $\hat{H}$ for H is defined through a weighted linear fit:

$$\hat{H} = \sum_j w_j Y_j / 2, \tag{34}$$

where $\sum_j$ runs over the range $\{j_1, \ldots, j_2\}$ of octaves where the linear fit is to be performed, this range is to be chosen a priori. The w_j satisfy the usual relations,

$$\left.\begin{array}{ll} \sum_j j w_j & = 1 \\ \sum_j w_j & = 0 \\ w_j & = (1/\lambda_j)(S_0 j - S_1)/(S_0 S_2 - S_1^2) \\ S_m & = \sum_{j=j_1}^{j_2} \lambda_j^{-1} j^m (m = 0, 1, 2), \end{array}\right\} \tag{35}$$

the λ_js being arbitrary numbers.

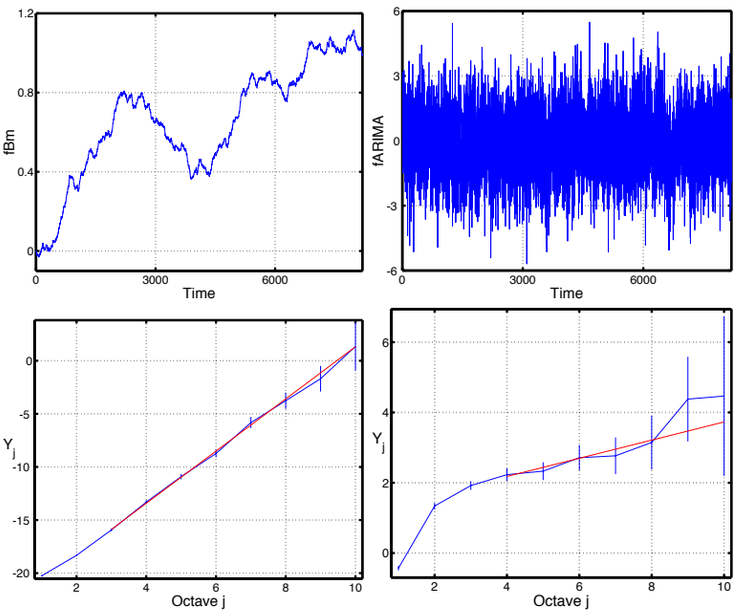

Fig. 4. Examples of Logscale Diagrams. Top left, sample path of a fractional Brownian motion; top right, sample path of a Long Range Dependent process (a fARIMA process). The fractional Brownian motion is an exact self-similar process, this can be seen on its Logscale diagram through the fact that the linear behavior of the log of the variance of the wavelet coefficients against the log of the scale holds for all scales (bottom left). The measurement of the slope enables moreover to precisely estimate the self-similarity parameter. The logscale diagram of the LRD process (bottom right) shows scaling behavior that holds only in the limit of large scales.

The statistical performance of this estimator have been studied in detail in [47,4]. For Gaussian processes, analytical calculations relying on the idealization of exact independence of the wavelet coefficients **P3ID**, show that a residual bias can be determined and therefore subtracted to produce an unbiased estimator. Let n denote the number of samples of the analyzed process X, neglecting the practical border effects resulting from the computation of the wavelet coefficients, the n_js behave as $n_j = 2^{-j}n$. For Gaussian processes and under **P3ID**, we showed that [47]:

$$\left.\begin{array}{l} \sigma_j^2 \equiv \mathrm{Var}\ Y_j \simeq = 2(\log_2 e)^2/n_j \\ \mathrm{Var}\ \hat{H} \simeq 2(\log_2 e)^2 \sum_j w_j^2 2^j/n. \end{array}\right\} \tag{36}$$

The variance of the estimator asymptotically decreases as the inverse of the analyzed number of samples. Numerical simulations showed [47] that the actual statistical performance are very close to the idealized ones, even for non Gaussian processes. This approximate, however realistic, estimation of the variance of $\hat{H}$ enables us to have confidence intervals on the estimation of the parameter H.

The choice of the weights w_j remains to be specified through the choice of the λ_js. It is well-known that the variance of the linear fit is minimal on condition that the λ_js match the covariance structure of the Y_js. Assuming the idealization **P3ID**, one obtains that the Y_js are independent. The choice $\lambda_j = \text{cste}\ n_j^{-1}$, therefore, ensures that the quantities $(\sum_j w_j^2/n_j)$ and Var $\hat{H}$ takes close to minimal values. The estimator $\hat{H}$ is systematically implemented with that particular choice.

• Additional properties

Thanks to its number of vanishing moments, the wavelet-based analysis of self-similarity, moreover, benefits from robustness against non-stationarities. For instance, if deterministic smooth trends (like a linear trend or an oscillating trend) are superimposed to self-similar processes, this may significantly complicate the detection and analysis of self-similarity. Because wavelet coefficients are blind to polynomials (of degree smaller than N) and *only feels* the most irregular parts of a process, increasing N will *cancel out* the possibly superimposed trends and therefore enable a relevant analysis of scaling in data without requiring any a priori processing. The possibility of performing various discrete wavelet transforms with different N and of comparing the resulting analysis and estimations is hence a key feature of the wavelet-based analysis of scaling against non stationarities. Comparison of wavelet-based analysis performed using mother-wavelet with different number of vanishing moments will allow to detect those trends and perform relevant analysis of self-similarity (see [2,47,4] for details).

From another perspective, non-stationarities and scaling may have practical effects and consequences that are practically very close and similar so that it may be difficult to distinguish one phenomena from the other. We have also shown that the wavelet framework offers a convenient way to design a statistical test allowing to discriminate actual scaling from some non stationary effects [49].

Finally, other interesting features of this wavelet based analysis lies in the facts that it is simple both conceptually and practically (DWT plus linear fits) and that it has a low computational cost thanks to the recursive pyramidal algorithm underlying the DWT. This is of importance when dealing with large sample of data, as is often the case when dealing with scaling and allowed us to propose real time on line algorithm for the analysis of scaling [39].

3 Beyond Self-similarity

3.1 Practical Limitations

Self-similar processes with stationary increments and finite variance, and more specifically their Gaussian version, the fractional Brownian motion, are very attractive models to describe scaling in empirical data and they are used and quoted in numerous and various applications. This is mainly because they are mathematically well-defined and well-documented and they moreover fulfill the

intuition of scaling phenomena in a very satisfactory manner. Their major practical quality is their *simplicity*: each and every of their (scaling) properties, (self-similarity, long-range dependence, fractal sample paths) is controlled by the single H. This parameter is therefore used in applications to describe, sometimes confusingly, either global or local scaling properties, i.e., either long memory or fractality... The major practical drawback of H-sssi processes lies in ... their *simplicity*: it is very unlikely that the numerous and various types of scaling encountered in the many different applications where they occur can all be described by a unique model depending on a single parameter. More precisely, exact self-similarity implies a number of specific properties (as summarized in (5) or (31)) and significant departures from those properties can be observed in the analysis of actual empirical data: *i*) moments of different orders may have scaling exponents that are not controlled by a single parameter, or more simply some moments may not present scaling or, even more simply, may not exist at all; *ii*) when scaling are observed, they may not exist over the whole range of scales as in the self-similar case, but only in a given range of scales, or only asymptotically in the limit of large scales or in the limit of small scales; *iii*) power-law behavior of the moments may not exist despite scaling behavior. In this section, we will explore those variations and describe some related mathematical models.

Scaling phenomena may also occur or exist in point processes and might be fruitfully studied through wavelets as well. This has been discussed in [1] and will not be addressed here.

3.2 Beyond Finite Variance

In previous sections, we assumed that the variance of the process X, as well as all higher moments, existed. However, one may encounter situations where scaling and self similarity are valid but where the variance of the process, for instance, and therefore all higher moments, are infinite. For those situations, the model of Gaussian self-similar processes as well as the analysis presented above and based on the variance of the wavelet coefficients are no longer relevant. Such situations can be modeled using α-stable self-similar processes, see Sect. 2.1. A wavelet-based analysis of α-stable self-similar processes can be conducted but the log of the variance of the wavelet coefficients, $\log_2 \mathbb{E}|d_X(j,k)|^2$, has to be replaced by the quantity $\mathbb{E}\log_2|d_X(j,k)|$, a random quantity with finite variance. This has been discussed at length in [3,5].

3.3 Beyond Scaling over All Scales: Long Range Dependence, $1/f$-Processes and Local Regularity

As said, a major consequence of self-similarity lies in the fact that the scaling behavior holds for all the scales (see (31) or (32)). Practically however, scaling may exist for the second order statistics (namely the variance) of the process, but may be observed only in a given, large but finite, range of scales, or in the asymptotic limits of small or large scales, rather than in the whole range of scales.

For instance, one may empirically observe the following asymptotic behavior:

$$\mathbb{E}|d_X(j,k)|^2 \simeq C2^{j(\gamma-1)}, 2^{j_m} \leq j \leq 2^{j_M}, 2^{j_M}/2^{j_m} \gg 1.$$

This is to be read as a weakened version of (32) and this is the signature of a scaling behavior in a wide, but finite, range of scales therefore corresponding to $1/f$-processes, with power law exponent γ (cf. (17)).

One may also empirically observe the following asymptotic behavior (see, for example, Fig. 4(right)):

$$\mathbb{E}|d_X(j,k)|^2 \simeq C2^{j2H}, 2^j \to +\infty.$$

This is to be seen as a weakened version of (32) and this is the signature of scaling that exist only for the largest scales of the process. This is reminiscent of (16) and tells us that the data are not self-similar but rather present some long term correlations properties and can therefore be modeled as a stationary long range dependent process [47,4].

Another possibility is to empirically observe an asymptotic scaling behavior in the limit of fine scales:

$$\mathbb{E}|d_X(j,k)|^2 \simeq C2^{j2H}, 2^j \to 0.$$

This is again to be seen has a weakened version of (32) and is reminiscent of Hölder regularity behavior (cf. (19)). This means that the data are not self similar but rather that their sample paths are characterized by a local regularity h controlled by H and that remains constant along time. This therefore betrays a fine scale scaling property.

For those situations, $1/f$-processes, long range dependence or local regularity, the analysis and estimation of the exponent can be performed with the logscale diagram, as in the self-similar case, except that linear fits are to be performed over a finite chosen range of scales, respectively. The question of automatically choosing or *detecting* the relevant range of scales is subtle and has been addressed e.g., in [50] for the LRD case.

3.4 Beyond Second-Order Statistics – Multiplicative and Multifractal Processes

• MultiScaling

Another practical major limitation of self-similar processes lies in the fact that the exponents of the power-laws for all the moments are controlled by the single parameter H, see (31). It is, however, quite common on empirical data to observe, in a given, finite, range of scales, a behavior of the type:

$$\mathbb{E}|d_X(j,k)|^q \simeq C2^{jH(q)}, 2^{j_m} < 2^j < 2^{j_M}, \tag{37}$$

where the exponents $H(q)$ may significantly depart from the linear qH behavior. We proposed to refer to this observation as *multiscaling*.

• Multiplicative processes

Modeling multiscaling implies a major change of paradigm: the additive structure underlying a random walk (cf. (1)) has to be abandoned and replaced by a multiplicative scheme. In other words, being at position X at time t results from a collection of elementary steps that are no longer added up together but instead multiplied one to the other. The canonical reference for multiplicative processes are the celebrated Mandelbrot's c-adic cascade processes [28]. Their construction is based on the iteration of a sequence of operations. At iteration j, one has c^j segments, to which are associated numbers $X_{j,k}$, $k = 0, \ldots, c^j - 1$. At stage $j+1$, one divides segment (j,k) into c subsegments to which are associated new numbers $X_{j+1,l} = W_{j+1,l} X_{j,k}$ where $l = (k-1) * c + p$, with $p = 0, \ldots, c-1$. The multipliers $W_{j,k}$ are i.i.d. positive random variables. Usually, $\mu(t)$ denote the measure obtained in the limit of an infinite number of iterations and $X(t) = \int_0^t \mu(du)$ the corresponding process. There exists numerous variations around this scheme that all share the spirit of multiplicative *cascade*.

• Multifractal processes

An important consequence of the Mandelbrot's multiplicative cascade procedure lies in the fact that the resulting motions $X(t) = \int_0^t \mu(du)$ are multifractal processes. In other words, they present sample paths with Hölder exponents $h(t)$ that vary very widely, irregularly and erratically from point to point and with each realizations. Those fluctuations of the local regularity are often described through the so-called multifractal spectrum $D(h)$, (which consists of the Hausdorff dimension of the set of points where the local regularity take the value h).

An important practical consequence of multifractality is that quantities called partition functions,

$$(1/n_j) \sum_{k=1}^{n_j} |d_X(j,k)|^q$$

present in the limit of small scales power law behaviors,

$$(1/n_j) \sum_{k=1}^{n_j} |d_X(j,k)|^q \sim C_q 2^{jH(q)}, 2^j \to 0.$$

Reading the partition functions $(1/n) \sum_{k=1}^{n} |d_X(j,k)|^q$ as estimators of the moments $\mathbb{E}|d_X(j,k)|^q$, the scaling relation above is very close to the equation defining multiscaling in the limit of small scales. Therefore multifractal can be seen as the very example for multiscaling.

Moreover, for the c-adic cascades, the multifractal spectrum $D(h)$ can be obtained from the function $H(q)$ through a Legendre transform. In that case, $H(q)$, and therefore $D(h)$, are controlled by the probability density function of the multipliers W. Further details on multifractal are beyond the scope of this chapter and the interested reader is referred to e.g., [36,37].

• Estimation issues

To test multiscaling or multifractal in empirical time series and estimate the corresponding $H(q)$ exponents, one forms the quantities $Y_j^{(q)}$ that can be read both as generalization of the Y_j (cf. (33)) to statistics of order different than 2 and as sample time average estimators for the ensemble averages:

$$Y_j^{(q)} = (1/n_j) \sum_{k=1}^{n_j} |d_X(j,k)|^q. \tag{38}$$

An extension of the estimation procedure described in Sect. 2 has been proposed to estimate the $H(q)$ exponents: it mainly consists in measuring slopes in $\log_2 2^j = j$ *vs* $\log_2 Y_j^{(q)}$ through **non weighted** linear regressions. For details, see [4,48]. The estimation of the $H(q)$ exponents for a multifractal process, synthesized according to the definitions proposed in [16], is illustrated in Fig. 5.

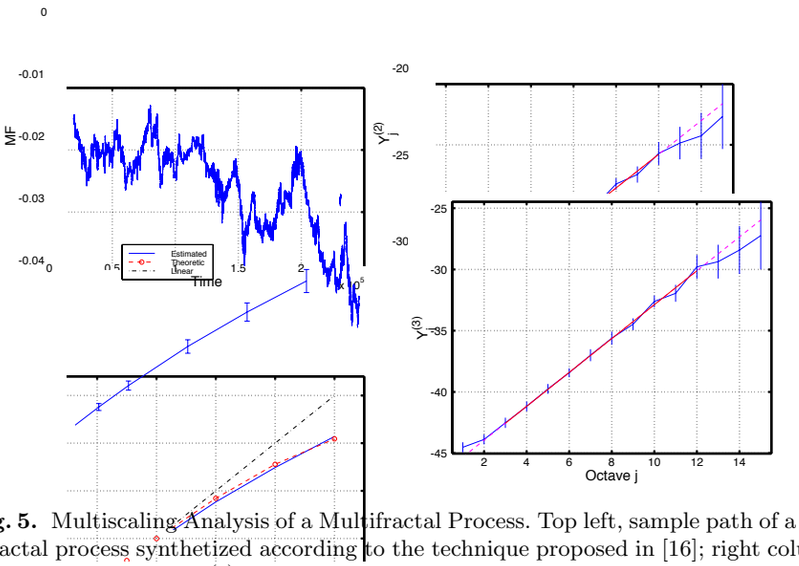

Fig. 5. Multiscaling Analysis of a Multifractal Process. Top left, sample path of a multifractal process synthetized according to the technique proposed in [16]; right column, Structure Functions ($Y_j^{(q)}$ vs j) of orders 2 and 3; bottom left, estimated exponents $H(q)$.

A key practical issue is to define a statistical test that would allow to test whether $H(q)$ is a linear function of q or not. It would enable to decide whether the analyzed data are to be modeled with an additive or multiplicative scheme. Up to our knowledge, this has hardly ever been addressed theoretically.

3.5 Beyond Power Laws – Infinitely Divisible Cascade

• Intuition

Multiscaling offers an extension to self-similarity insofar as a collection of exponents rather than a single one is needed to describe data. Yet, it maintains a major feature of self-similarity: moments behaves as power laws of the scales. However, when analyzing actual data, it may very well be observed that this is not the case, see e.g., [12,20,48,52]. To account for those situations, the infinitely divisible cascade model, on which we concentrate now, proposes to gain an extra degree of freedom by giving up the requirement that moments behaves a priori as power laws of the scales. The equations below summarize the connections between self-similarity, multiscaling and infinitely divisible cascade:

$$\begin{array}{lll} \text{Self-Sim.} & \mathbb{E}|d_X(j,k)|^q = C_q(2^j)^{qH} & = C_q \exp(qH \ln(2^j)) \\ \text{MultiScaling} & \mathbb{E}|d_X(j,k)|^q = C_q(2^j)^{H(q)} & = C_q \exp(H(q) \ln(2^j)) \\ \text{Inf. Div. Casc.} & \mathbb{E}|d_X(j,k)|^q = ---- & = C_q \exp(H(q) n(2^j)), \end{array} \tag{39}$$

where the functions $n(2^j)$ and $H(q)$ need not a priori be the function $\ln 2^j$ and the linear function qH, respectively.

• Definition

The concept of infinitely divisible cascade was first introduced by B. Castaing in [12,13] and rephrased in the wavelet framework in [7,48]. We briefly recall its intuition, definition and relations to other models. Starting again from the self-similar case, one can write the probability density function (pdf) of the wavelet coefficients at scale $a = 2^j$, as a dilated version of the pdf of the wavelet coefficients at a larger scale a': $p_a(d) = (1/\alpha_0)\, p_{a'}(d/\alpha_0)$ where the dilation factor is unique: $\alpha_0 = (a/a')^H$. In the cascade model, the key change is that there is no longer a unique factor but a collection of dilation factors α; consequently p_a will result from a weighted sum of dilated incarnations of $p_{a'}$:

$$p_a(d) = \int G_{a,a'}(\ln \alpha)\, \frac{1}{\alpha}\, p_{a'}\left(\frac{d}{\alpha}\right) d \ln \alpha.$$

The weighting function $G_{a,a'}$ is called the kernel or the *propagator* of the cascade. Obviously, if $G_{a,a'}$ is a Dirac function, $G_{a,a'}(\ln \alpha) = \delta(\ln \alpha - H \ln(a/a'))$, infinitely divisible cascade reduces to self-similarity, therefore understood as a special case. The definition of the cascade above shows that the pdf's of $\underline{p}_a$ and $\underline{p}_{a'}$ of the log wavelet coefficients $\ln |d|$ are related by a convolution with the propagator:

$$\begin{aligned} \underline{p}_a(\ln \alpha) &= \int G_{a,a'}(\ln \alpha)\, \underline{p}_{a'}(\ln |d| - \ln \alpha)\, d \ln \alpha \\ &= (G_{a,a'} * \underline{p}_{a'})(\ln \alpha). \end{aligned} \tag{40}$$

If cascades exist between scales a and a'' and between scales a'' and a', then a cascade between scales a and a' exists, and the corresponding propagator results from the convolutions of the two propagators: $G_{a,a'} = G_{a,a''} * G_{a'',a'}$. Infinite

divisibility (also called continuous self similarity) means that no scale between a and a' plays any characteristic role (i.e., a'' in the above statement can be any scale between a and a'). Infinite divisibility therefore implies that the propagator consists of an elementary function G_0 convolved with itself a number of times, where that number depends on a and a':

$$G_{a,a'}(\ln \alpha) = [G_0(\ln \alpha)]^{*\left(n(a)-n(a')\right)} .$$

Using the Laplace transform $\tilde{G}_{a,a'}(q)$ of $G_{a,a'}$, this can be rewritten as $\tilde{G}_{a,a'}(q) = \exp\{H(q)(n(a) - n(a'))\}$, with $H(q) = \ln \tilde{G}_0(q)$ and $a := 2^j$; this implies that $\mathbb{E}|d_X(j,k)|^q = C_q \exp\left\{H(q)n(2^j)\right\}$, thus validating (39). The main consequences of infinitely divisible cascade read therefore:

$$\ln \mathbb{E}|d_X(j,k)|^q = H(q)n(2^j) + K_q \tag{41}$$

$$\ln \mathbb{E}|d_X(j,k)|^q = \frac{H(q)}{H(p)} \ln \mathbb{E}|d_X(j,k)|^p + \kappa_{q,p}. \tag{42}$$

This last equation implies that for any p and q, the moment of order q behave as power-law of the moment of order p. This is sometimes referred to as "extended self-similarity", in turbulence mainly. Note moreover, that, in the relation (41) above, there is some arbitrariness, indeed:

$$\begin{aligned} H(q)n(a) + K_q &= \left(\tfrac{H(q)}{\beta}\right)(\beta n(a) + \gamma) + (K_q - \tfrac{H(q)\gamma}{\beta}) \\ &= H'(q)n'(a) + K'_q \end{aligned} \tag{43}$$

where $\beta \neq 0$ and γ are arbitrary constants. It clearly indicates that the function $H(q)$ is defined up to a multiplicative constant while n is defined up to multiplicative and additive constants.

If it is moreover required that the function $n(a) \equiv \ln a$, the infinitely divisible cascade is called *scale invariant* and this implies that:

$$\tilde{G}_{a,a'}(q) = (a/a')^{\ln \tilde{G}_0(q)} \text{ and } \mathbb{E}|d_X(j,k)|^q = (2^j)^{\ln \tilde{G}_0(q)},$$

proving that scale invariant infinitely divisible cascade reduces to multiscaling with exponents being controlled by the propagator: $H(q) = \ln \tilde{G}_0(q)$. In this framework, multiscaling, or multifractal, is therefore understood as a special case of infinitely divisible cascade. In a scale invariant infinitely divisible cascade, one can also inquire on whether $H(q)$ is a linear function of q or not, in which case the cascade reduces to the even more special case of self-similarity. It is, therefore, natural to consider the function $H(q)/q$ and to test its constancy.

• A FUNDAMENTAL FEATURE

The Infinitely Divisible Cascade model is hence a natural extension to the multiscaling and self-similarity ones, it is important to note however that it maintains a fundamental feature that already existed in the two previous models (cf. the set of equations (39): the dependence of the moments in the variables q

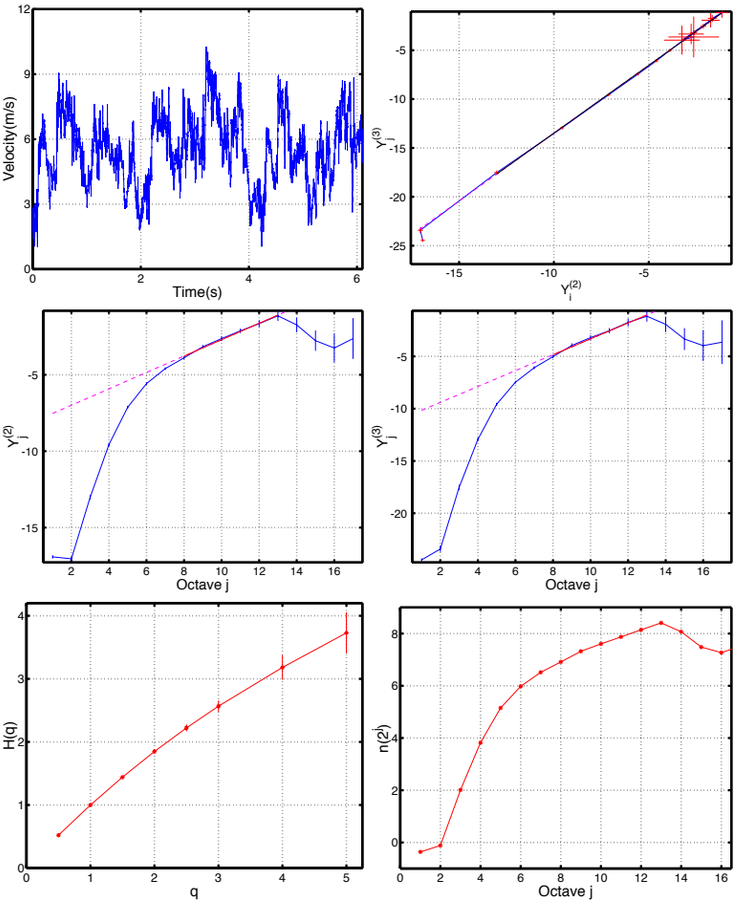

Fig. 6. Infinitely Divisible Scaling Analysis for Turbulent Velocity Signals. Top left, Velocity Time Series; Second Raw, the Structure Functions of orders 2 and 3 show clear departures from strict power laws; top right, Extended Self-Similarity ($Y_j^{(3)}$ vs $Y_j^{(2)}$), the straight line show that an Infinitely Divisible Cascade Model gives a satisfactory description of the Velocity Time Series; bottom left and right respectively, estimated functions $H(q)$ and $n(2^j)$.

(order of the moment) and 2^j (scale) is separable. This key feature can actually be considered as a *practical or operational definition* of scaling in time series.

• Estimation issues

To analyze Infinite Divisibility in empirical time series, we propose to form the diagrams $Y_j^{(q)}$ versus $Y_j^{(p)}$, that constitute natural extensions to logscale diagrams. Again, those diagrams come with confidence intervals for the $Y_j^{(q)}$s. Testing the validity of the model empirical data amounts to test for the exis-

tence of an affine relation amongst the $Y_j^{(q)}$ versus $Y_j^{(p)}$, i.e., straight lines in the diagrams. The estimation of the corresponding $H(q)$ and $n(a)$ parameter functions is performed through a non weighted linear fit between the random variables $Y_j^{(q)}$ versus $Y_j^{(p)}$ (see (41) and (42)):

$$\begin{array}{ll} \hat{H}(q)/H(p) & = \mathrm{slope}_{q,p},\ \forall q \\ \hat{K}_q & = \mathrm{intercept}_{q,p},\ \forall q \\ \hat{n}(a=2^j) & = \frac{1}{H(p)}\left\langle \frac{1}{\mathrm{slope}_{q,p}} \left(Y_j^{(q)} - \hat{K}_q \right) \right\rangle_q + K_p, \end{array} \tag{44}$$

where $\langle . \rangle_q$ stands for simple means on q-values. Details on the difficulties (best choice of p, arbitrariness from (43), ...) of the procedure as well as on its statistical performances are given in [4,15]. Fig. 6 illustrates this estimation procedure on turbulence velocity data[3] recorded in jet turbulence at a R_λ Reynolds number of the order of 600.

4 Conclusion

In this article, we showed that scaling phenomena in empirical data may be described through a large variety of mathematical models. We gave an introductory and comprehensive overview of those models, that can be read as variations on the self-similarity reference. We pointed out however that multiscaling, multifractal and cascades imply the replacement of the additive random walk framework with multiplicative constructions.

Then, we have shown how and why the wavelet transform offers a versatile, powerful and efficient tool to perform the analysis of scaling: with only little a priori on the nature and properties of the empirical data, it allows to *detect* the existence of scaling in data, to *identify* the nature of the detected scaling, to *estimate* the corresponding parameter. Wavelets may also be used for the numerical synthesis of stochastic processes with a priori prescribed (scaling) properties. This has not been detailed here and the interested reader is referred to [4].

This article is expected to propose to the reader a *quick start* on what scaling in time series may mean and on what to do with wavelets to analyze them. A set of MATLAB routines enables a practical use of all the techniques presented here. Those techniques have been fruitfully used for the analysis of hydrodynamic turbulence [14,15] and computer network traffic [4,48].

References

1. P. Abry, P. Flandrin: 'Point processes, long-range dependence and wavelets'. In: *Wavelets in Medicine and Biology*, ed. by A. Aldroubi and M. Unser (CRC Press, Boca Raton (FL) 1996) pp.413–438

[3] Data collected by C. Baudet, in Physic Lab. at Ecole Normale Supérieure de Lyon, France. The calculation of R_λ is based on the Taylor microscale.

2. P. Abry, D. Veitch: IEEE Trans. on Info. Theory, **44**, 2 (1998)
3. P. Abry, L. Delbeke and P. Flandrin: IEEE-ICASSP-99, Phoenix (AZ) (1999)
4. P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch: *Wavelets for the analysis, estimation and synthesis of scaling data* A chapter in [35]
5. P. Abry, B. Pesquet-Popescu, M. S. Taqqu: 'Wavelet Based Estimators for Self Similar α-Stable Processes'. In: *Int. Conf. on Signal Proc., 16th World Computer Congress, Beijing, China, 2000*
6. P. Abry, P. Gonçalvès and J. Lévy Véhel: Traité Information, Commandes, Contrôle, Hermès Editions, Paris, France, April (2002)
7. A. Arnéodo, J.F. Muzy, S.G. Roux: J. Phys. II France, **7**, 363 (1997)
8. R. Averkamp and C. Houdré: IEEE Trans. on Info. Theory, Vol.**44**, 1111 (1998)
9. V. Bareikis and R. Katilius: *Noise in Physical Systems and* $1/f$ *Fluctuations* (World Scientific 1995)
10. J. Beran: *Statistics for Long-Memory Processes* (Chapman and Hall, New York 1994)
11. S. Cambanis and C. Houdré: IEEE Trans. on Info. Theory, **41**, 628 (1995)
12. B. Castaing, Y. Gagne, E. Hopfinger: Physica D, **46**, 177 (1990)
13. B. Castaing: J. Phys. II France, **6**, 105 (1996)
14. P. Chainais, P. Abry, and J. F. Pinton: Phys. Fluids, **11**, 3524 (1999)
15. P. Chainais: Cascades log-infiniment divisibles et analyse multirésolution. Application à l'étude des intermittences en turbulence. PhD Thesis, Ecole Normale Supérieure de Lyon (2001)
16. P. Chainais, R. Riedi and P. Abry: 'Infinitely Divisible Processes'. In: *Stochastic processes and Their Applications* (May 2002)
17. I. Daubechies: *Ten Lectures on Wavelets* (SIAM, Philadelphia 1992)
18. L. Delbeke and P. Abry: Stochastic Processes and their Applications, **86**, 177 (2000)
19. P. Frankhauser: Population, **4**, 1005 (1997)
20. U. Frisch: *Turbulence. The legacy of A. Kolmogorov* (Cambridge University Press, UK 1995)
21. A. C. Gilbert, W. Willinger and A. Feldmann: *IEEE Trans. on Info. Theory* **45**, 971 (1999)
22. P. Gonçalvès and R. H. Riedi: Proc. 17ème Colloque GRETSI, Vannes, France (1999)
23. M. Keshner: proc. of the IEEE, **70**, 212 (1982)
24. A. N. Kolmogorov: 'a) Dissipation of energy in the locally isotropic turbulence, b) The local structure of turbulence in incompressible viscous fluid for very large Reynolds number, c) On degeneration of isotropic turbulence in an incompressible viscous liquid'. In: *Turbulence, Classic papers on statistical theory*, ed.by S.K. Friedlander and L. Topper (Interscience publishers 1941) pp.151-161
25. W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson: IEEE/ACM Trans. on Networking, **2**, 1 (1994)
26. L. Liebovitch and A. Todorov: The American Physiological Society, 1446 (1996)
27. B. B. Mandelbrot and J. W. Van Ness: SIAM Rev, **10**, 422 (1968)
28. B. B. Mandelbrot: J. of Fluid Mech., **62**, 331 (1974)
29. B. B. Mandelbrot: *Fractales, Hasard et Finance* (Flammarion 1997)
30. S. Mallat: *A Wavelet Tour of Signal Processing* (Academic Press, Boston 1997)
31. E. Masry: IEEE Trans. on Info. Theory, **39**, 260 (1993)
32. P. Flandrin: IEEE Trans. on Info. Theory, **38**, 910 (1992)

33. P. Flandrin: 'Fractional Brownian motion and wavelets'. In: *Wavelets, Fractals, and Fourier Transforms* ed. by M. Farge, J.C.R. Hunt and J.C. Vassilicos (Clarendon Press, Oxford 1993) pp.109–122
34. B. Pesquet-Popescu: Signal Processing **75** (1999)
35. K. Park and W. Willinger: *Self-Similar Network Traffic and Performance Evaluation* (Wiley ,Interscience Division 2000)
36. R. Riedi, M.S. Crouse, V.J. Ribeiro and R.G. Baraniuk: IEEE Trans. on Info. Theory, **45**, 992 April (1999)
37. R. H. Riedi: 'Multifractal processes'. In: *Long range dependence: theory and applications* ed. by Doukhan, Oppenheim and Taqqu (2002)
38. L. C. G. Rogers and D. Williams: *Diffusions, Markov processes and martingales*, 2nd edition, vol 1 (Foundations, Wiley 1994)
39. M. Roughan, D. Veitch, and P. Abry: IEEE Trans. on Networking, **8**, 467 (2000)
40. M. Roughan, D. Veitch, J. Yates, M. Ashberg, H. Elgelid, M. Castro, M. Dwyer and P. Abry: 'Real-Time Measurement of Long-Range Dependence in ATM Networks'. In: *Passive and Active Measurement Workshop, Hamilton, New-Zealand 2000*
41. G. Samorodnitsky and M. S. Taqqu: *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance* (Chapman and Hall, New York, London 1994)
42. A. Saucier: 'Méthodes multifractales pour l'analyse d'images et de signaux'. In: [6]
43. D. Sornette: *Critical Phenomena in Natural Sciences* (Springer 2000)
44. A. Torcini and M. Antoni: Physical Review E, **59**, 2746 (1999)
45. M. Teich, S. Lowen, B. Jost and K. Vibe-Rheymer: *Heart rate Variability: Measures end Models* (preprint 2000)
46. A. H. Tewfik and M. Kim: IEEE Trans. Info. Theory, **38**, 904 (1992)
47. D. Veitch and P. Abry: IEEE Trans. on Info. Theory, **45**, 878, April (1999)
48. D. Veitch, P. Abry, P. Flandrin and P. Chainais: 'Infinitely Divisible Cascade Analysis of Network Traffic Data'. In: *Proceedings of ICASSP, Istanbul, June 2000*
49. D. Veitch and P. Abry: 'A statistical test for the constancy of scaling exponents'. In: itshape IEEE Trans. on Sig. proc. 2001
50. D. Veitch, P. Abry and M. S. Taqqu: *On the automatic selection of the onset of scaling* (Preprint 2002)
51. C. Walter: *Lois d'échelle en finance* in [6]
52. B. West and A. Goldberger: American Scientist, **75**, 354 (1987)