1 Wavelets for the analysis, estimation and synthesis of scaling data

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

(1) - *CNRS UMR 5672 - Laboratoire de Physique - Ecole Normale Supérieure de Lyon - 46, allée d'Italie 69 364 LYON Cedex 07 - France*
tel: (+33) 4 72 72 84 93 - Fax: (+33) 4 72 72 80 80
pabry,flandrin@physique.ens-lyon.fr - http://www.physique.ens-lyon.fr/ts
Partially supported by the CNRS grant TL97035, Programme Télécommunications

(2) - *Department of Mathematics - 111 Cummington Street, Boston University, Boston, MA 02215-2411, USA*
tel: (+1) (617) 353-3022 - Fax: (+1) (617) 353-8100
Email : murad@math.bu.edu - *Web :* http://math.bu.edu/people/murad
Partially supported by the NSF grant DMS-94044093 and ANI-9805623 at Boston University

(3) - *Software Engineering Research Centre - Level 2, 723 Swanston Street, Carlton, Victoria 3053, - Australia*
tel: (+613) 9282 2440 - Fax: (+1) (+613) 9282 2444
darryl@serc.rmit.edu.au - http://www.serc.rmit.edu.au/~darryl

To be published as chapter 2, in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, eds. , Wiley Interscience, to appear, 2000.

Keywords: Network teletraffic - Ethernet/Internet traffic - Self-similarity - Long-range dependence - Scaling phenomena - (Multi)fractal - Wavelet analysis - Scaling analysis - Scaling parameters estimation - Robustness - Fractional Brownian motion synthesis - Fano factor - Aggregation procedure - Allan variance.

1.1 THE SCALING PHENOMENA

1.1.1 Scaling issues in traffic

The presence of scaling behaviour in telecommunications traffic is striking not only in its ubiquity, appearing in almost every kind of packet data, but also in the wide range of scales over which the scaling holds (see e.g. [43], [18], [78]). It is rare indeed that a physical phenomenon obeys a consistent law over so many orders of magnitude. This may well extend further, as increases in network bandwidth over time progressively 'reveal' higher scales.

While the presence of scaling is now well established, its impact on teletraffic issues and network performance is still the subject of some confusion and uncertainty. Why is scaling in traffic important for networking? It is clear, as far as modelling of the traffic itself is concerned, that a feature as prominent as scaling should be built into models at a fundamental level, if these are to be both accurate and parsimonious. Scaling, therefore, has immediate implications for the choice of classes of traffic models, and consequently on the choice, and subsequent estimation, of model parameters. Such estimation is required for initial model verification, for fitting purposes, as well as for traffic monitoring.

Traffic modelling however, does not occur in isolation but in the context of performance issues. Depending on the performance metric of interest, and the model of the network element in question, the impact and therefore the relevance of scaling behaviour will vary. As a simple example, it is known that in certain infinite buffer fluid queues fed by long range dependent (LRD) On/Off sources, that the stationary queueing distribution has infinite mean, a radically non-classical result. Such infinite moments disappear however if the buffer is finite, intuitively because a finite reservoir cannot 'hold' long memory. The LRD of the input stream will strongly affect the overflow loss process, but cannot seriously exacerbate the conditional delay experienced by packets which are *not* lost, as this is bounded by the size of the buffer. The importance of scaling in the performance sense, apart from being as yet unknown in a great many cases, is therefore context dependent.

We focus here on the fundamental issues of detection, identification, and measurement of scaling behaviour. These cannot be ignored even if one is interested in *performance questions which are not directly related to scaling.* This is because scaling induces non-classical statistical properties which affect the estimation of *all* parameters, not merely those that describe scaling. This in turn, affects the predictive abilities of performance models, and therefore their usefulness in practice.

The reliable detection of scaling should thus be our first concern. By detecting the absence or presence of scaling, one will know whether the data need be analysed using traditional statistics, or by special statistical techniques that take the presence of scaling into account. Here is it vital to be able to distinguish artifacts due to non-stationarities with the appearance of

scaling, from true scaling behaviour. Identification is necessary since more than one kind of scaling exists, with differing interpretations and implications for model choice. Finally, should scaling of a given kind be present, an accurate determination of the parameters which describe it must be made. These parameters will control the statistical properties of estimates made of all other quantities, such as the parameters needed in traffic modelling, or quality of service metrics.

As a simple yet powerful example of the above, consider a second order process $X(t)$ which we know to be stationary, and whose mean μ_X we wish to estimate from a given data set of length n. For this purpose the simple sample mean estimator is a reasonable choice. The classical result is that asymptotically for large n the sample mean follows a normal distribution, with expectation equal to μ_X, and variance σ_X^2/n, where σ_X^2 is the variance of X. In the case where X is LRD the sample mean is also asymptotically normally distributed with mean μ_X, however the variance is given by $\frac{2c_r n^\alpha}{(1+\alpha)\alpha} \cdot \frac{1}{n}$, where $\alpha \in [0,1)$ and $c_r \in (0,\infty)$ are the parameters describing the LRD ([17], p.160). This expression reveals that the variance of the sample mean decreases with the sample size n at a rate which is slower than in the classical case. Noting that the ratio of the size of the LRD based variance to the classical one grows to infinity with n, it becomes apparent that confidence intervals based on traditional assumptions, even for a quantity as simple as the sample mean, can lead to serious errors when in fact the data is LRD.

We focus here on how a wavelet based approach allows the threefold objective of the detection, identification, and measurement of scaling to be efficiently achieved. Fundamentally, this is due to the non-trivial fact that the analysing wavelet family itself possesses a scale invariant feature, a property not shared by other analysis methods. A key advantage is that quite different kinds of scaling can be analysed by the same technique, indeed by the same set of computations. The semi-parametric estimators of the scaling parameters which follow from the approach have excellent properties: negligible bias and low variance, and in many cases compare well even against parametric alternatives. The computational advantages, based on the use of the Discrete Wavelet Transform (DWT), are very substantial and allow the analysis of data of arbitrary length. Finally, there are very valuable robustness advantages inherent in the method, particularly with respect to the elimination of superposed smooth trends (deterministic functions).

Another important issue connected with modelling and performance studies concerns the generation of time series for use in simulations. Such simulations can be particularly time consuming for long memory processes where the past exerts a strong influence on the future, disallowing simple approximations based on truncation. Wavelets offer in principle a parsimonious and natural way to generate good approximations to sample paths of scaling processes, which benefit from the same DWT-based computational advantages enjoyed by the analysis method. This area is less well developed than is the case for analysis however.

1.1.2 Mapping the land of scaling and wavelets

The remainder of the paper is organized as follows.

Section 2, "WAVELETS AND SCALING: THEORY", discusses in detail the key properties of the wavelet coefficients of scaling processes. It starts with a brief, yet precise, introduction to the continuous and discrete wavelet transforms, to the multiresolution analysis theory underlying the latter, and the low complexity decomposition algorithm made possible by it. It recalls concisely the definitions of two of the main paradigms of scaling, self-similarity and long range dependence. The properties of the wavelet coefficients of self similar, long range dependent, and fractal processes are then given, and it is shown how the analysis of these various kinds of scaling can be gathered into a single framework within the wavelet representation. Extensions to more general classes of scaling processes requiring a collection of scaling exponents, such as multifractals, are also discussed.

The aim of section 3, "WAVELETS AND SCALING: ESTIMATION", is to indicate how and why this wavelet framework enables the efficient analysis of scaling processes. This is achieved through the introduction of the *Logscale Diagram*, where the key analysis tasks of the detection of scaling, interpretation of the nature of scaling, and estimation of scaling parameters, can be performed. Practical issues in the use of the Logscale Diagram are addressed, with reference to examples from real traffic data and artificially generated traces. Definitions, statistical performance, and pertinent features of the estimators for scaling parameters are then studied in detail. The Logscale Diagram, first defined with respect to second order statistical quantities, is then extended to statistics of other orders. It is also indicated how the tool allows for and deals with situations/processes departing from pure scaling, such as superimposed deterministic non-stationarities. Finally, clear connections between the wavelet tool and a number of more classical statistical tools dedicated to the analysis of scaling are drawn, showing how the latter can be profitably generalized in their wavelet incarnations.

Section 4, "WAVELETS AND SCALING: SYNTHESIS", proposes a wavelet-based synthesis of the fractional Brownian motion. It shows how this process can be naturally and efficiently expanded in a wavelet basis, allowing, provided that the wavelets are suitably designed, its accurate and computationally efficient implementation.

Finally in Section 5, "WAVELETS AND SCALING: PERSPECTIVES", a brief indication is given of what may lay ahead in the broad land of scaling and wavelets.

1.2 WAVELET AND SCALING: THEORY

1.2.1 Wavelet analysis: a brief introduction

1.2.1.1 The (continuous) wavelet transform The continuous wavelet decomposition (CWT) consists of the collection of coefficients

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

that compares (by means of inner products) the signal X to be analyzed with a set of analysing functions

$$\{\psi_{a,t}(u) \equiv \frac{1}{\sqrt{a}} \psi_0(\frac{u-t}{a}),\ a \in \mathbb{R}^+,\ t \in \mathbb{R}\}.$$

This set of analysing functions is constructed from a reference pattern ψ_0, called the mother-wavelet, by the action of a time-shift operator $(\mathcal{T}_\tau \psi_0)(t) \equiv \psi_0(t-\tau)$ and a dilation (change of scale) operator

$$(\mathcal{D}_a \psi_0)(t) \equiv 1/\sqrt{a}\psi_0(t/a).$$

ψ_0 is chosen such that both its spread in time and frequency are relatively limited. It consists of a small wave defined on a support which is almost limited in time and having most of its energy within a limited frequency band. While the time support and frequency band cannot be both finite, there is an interval on which they are *effectively* limited. The time-shift operator enables the selection of the time instant around which one wishes to analyze the signal, whilst the dilation operator defines the scale of time (or equivalently, the range of frequencies) over which it will be observed. The quantity $|T_X(a,t)|^2$, referred to as a "scalogram", can therefore be interpreted as the energy content of X around time t within a given range of frequencies controlled by a. In addition to being well-localized both in time and frequency, the mother wavelet is required to satisfy the *admissibility condition*, whose weak form is

$$\int \psi_0(u) du = 0\,, \tag{1.1}$$

which shows it is a band-pass or oscillating function, whence the name "wavelet".

Wavelets that are often used in practice include the Haar wavelet, the Daubechies wavelets, indexed by a parameter $N = 1, 2, \ldots$, and the Meyer wavelets. The Haar wavelet $\psi_0(u)$ is discontinuous; it equals 1 at $0 \leq u < 1/2$, -1 at $1/2 \leq u \leq 1$ and 0 otherwise. The Daubechies wavelet with $N = 1$ is in fact the Haar wavelet, but the other Daubechies wavelets with $N > 1$ are continuous with bounded support, and have N vanishing moments (that is, they satisfy Eq. (1.5) below). The Meyer wavelets do not have bounded support, neither in the time nor frequency domain, but all their moments vanish and they belong to the Schwartz space, that is, they are infinitely

differentiable and decrease very rapidly to 0 as u tends to $\pm\infty$.

On the condition that the wavelet be admissible, the transform can be inverted:

$$X(t) = C_\psi \int\int T_X(a,\tau)\psi_{a,\tau}(t)\frac{dad\tau}{a^2}$$

where C_ψ is a constant depending on ψ. This reconstruction formula expresses X in terms of a weighted integral of wavelets (acting as elementary atoms) located around given times and frequencies, thereby constituting quanta of information in the time-frequency plane. For a more general presentation of the wavelet analysis see, for example, [24].

Because the wavelet transform represents in a plane (i.e. a 2D space) the information contained in a signal (i.e., 1D space), it is a redundant transform, which means that neighbouring coefficients in the time-scale plane share a certain amount of information. A mathematical theory, the *Multiresolution Analysis* (MRA), proves that it is possible to critically sample the time-scale plane, i.e., to keep, among the $\{T_X(a,t),\ a \in \mathbb{R}^+,\ t \in \mathbb{R}\}$, only a discrete set of coefficients whilst still retaining the total information in X. That procedure defines the so-called Discrete (or non redundant) Wavelet Transform.

1.2.1.2 Multiresolution Analysis and Discrete Wavelet Transform

A multiresolution analysis (MRA) consists in a collection of nested subspaces $\{V_j\}_{j\in\mathbb{Z}}$, satisfying the following set of properties [24]:

1. $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$
2. $V_j \subset V_{j-1}$
3. $X(t) \in V_j \iff X(2^j t) \in V_0$
4. There exists a function $\phi_0(t)$ in V_0, called the *scaling function*, such that the collection $\{\phi_0(t-k), k \in \mathbb{Z}\}$ is an unconditional Riesz basis for V_0

To understand the significance of these properties, observe that, from Property 1, the V_j's are approximation subspaces of the space of square integrable functions $L^2(\mathbb{R})$. Property 4 expresses the fact that the set of shifted scaling functions $\{\phi_0(t-k),\ k \in \mathbb{Z}\}$ form a "Riesz basis" for V_0, that is, they are linearly independent and span the space V_0, but they are not necessarily orthogonal nor do they have to be of unit length. Finding such a function $\phi_0(t)$ is hard, but many candidates for $\phi_0(t)$ are known in the literature.

Similarly, Properties 3 and 4 together imply that the scaled and shifted functions

$$\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k),\ k \in \mathbb{Z}\}$$

constitute a Riesz basis for the space V_j. The multiresolution analysis involves successively projecting the signal X to be studied into each of the approxi-

mation subspaces V_j:

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

Since, from Property 2, $V_j \subset V_{j-1}$, approx_j is a coarser approximation of X than is approx_{j-1} (Note that some authors use the opposite convention and set $V_j \subset V_{j+1}$.) Property 1 moreover indicates that in the limit of $j \to +\infty$, all information is removed from the signal. The key idea of the MRA, therefore, consists in studying a signal by examining its coarser and coarser approximations, by cancelling more and more high frequencies or details from the data.

The information that is removed when going from one approximation to the next, coarser one, is called the detail:

$$\text{detail}_j(t) = \text{approx}_{j-1}(t) - \text{approx}_j(t).$$

The MRA shows that the detail signals detail_j can be directly obtained from projections of X onto a collection of subspaces, the $W_j = V_j \ominus V_{j-1}$, called the wavelet subspaces. Moreover, the MRA theory shows that there exists a function ψ_0, called the mother wavelet, to be derived from ϕ_0, such that its templates

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k),\ k \in \mathbb{Z}\}$$

constitute a Riesz basis for W_j:

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

For example, if the scaling function $\phi_0(t)$ is the function which equals 1 if $0 \le t \le 1$ and 0 otherwise, then the corresponding mother wavelet $\psi_0(u)$ is the Haar wavelet.

Theoretically, this projection procedure can be performed from $j \to -\infty$ up to $j \to +\infty$. *In practice*, one limits the range of indices j to $j = 0, \ldots, J$ and thus only considers

$$V_J \subset V_{J-1} \subset \ldots \subset V_0.$$

This means that we restrict the analysis of X to that of its (orthogonal) projection $\text{approx}_0(t)$ onto the reference space V_0, labeled as zero by convention, and rewrite this fine scale approximation as a collection of details at different resolutions together with a a final low-resolution approximation which belongs to V_J:

$$\begin{aligned} \text{approx}_0(t) &= \text{approx}_J(t) + \textstyle\sum_{j=1}^{J} \text{detail}_j(t) \\ &= \textstyle\sum_k a_X(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_X(j,k)\psi_{j,k}(t). \end{aligned} \tag{1.2}$$

If X is in V_0, one can obviously replace approx_0 by X in the above relation.

Except in the case where X actually belongs to V_0, selecting V_0 implies some unavoidable information loss [11]. This is entirely analogous to the loss induced by the necessary prefiltering operation involved in Shannon-Whittaker sampling theory to band-limit a process prior to sampling. Note however that there is no additional information loss after the initial projection. Varying J simply means deciding if more or less information is written in details as opposed to the final approximation approx_J.

Since the approx_j are essentially coarser and coarser approximations of X, ϕ_0 needs to be a low-pass function. The detail_j, being an information 'differential', indicates rather that ψ_0 is a band-pass function, and therefore a small wave, a *wavelet*. More precisely, the MRA shows that the mother wavelet must satisfy $\int \psi_0(t)dt = 0$ [24].

Given a scaling function ϕ_0 and a mother-wavelet ψ_0, the discrete (or non redundant) wavelet transform (DWT) consists of the collection of coefficients

$$X(t) \to \{\{a_X(J,k),\, k \in \mathbb{Z}\},\, \{d_X(j,k),\, j = 1, \ldots, J,\, k \in \mathbb{Z}\}\} \tag{1.3}$$

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

$$\left.\begin{array}{rcl} a_X(j,k) & = & \langle X, \mathring{\phi}_{j,k} \rangle \\ d_X(j,k) & = & \langle X, \mathring{\psi}_{j,k} \rangle \end{array}\right\} \tag{1.4}$$

where $\mathring{\psi}_{j,k}$ (resp., $\mathring{\phi}_{j,k}$) are shifted and dilated templates of $\mathring{\psi}_0$ (resp., $\mathring{\phi}_0$), called the dual mother wavelet (resp., the dual scaling function), and whose definition depends on whether one chooses to use an orthogonal, semi-orthogonal or bi-orthogonal DWT (see e.g., [24]). In Eqs. (1.2) and (1.4) above, the role of the wavelet and its dual can be arbitrarily exchanged, and similarly for the scaling function and its dual. In what follows this exchange is performed for simplicity of notation. The $d_X(j,k)$ constitute a subsample of the $\{T_X(a,t),\, a \in \mathcal{R}^+,\, t \in \mathbb{R}\}$, located on the so-called dyadic grid

$$d_X(j,k) = T_X(2^j, 2^j k).$$

The logarithm (base 2) of the scale $\log_2(a = 2^j) = j$ is called the *octave* j, and a scale will often be referred to by its corresponding octave. For the sake of clarity, we henceforth restrict our presentation to the DWT (characterized by the $d_X(j,k)$), which brings with it considerable computational advantages. However, the fundamental results based on the wavelet approach hold for the CWT, see [3, 4].

1.2.1.3 Key features of the wavelet transform In the study of the scaling processes analysed below, the following two features of the Wavelet Transform play key roles:

- **F1:** The wavelet basis is constructed from the dilation (change of scale) operator, so that the analyzing family itself exhibits a scale invariance feature.

- **F2:** ψ_0 has a number $N \geq 1$ of *vanishing moments*:

$$\int t^k \psi_0(t) dt \equiv 0, \quad k = 0, 1, 2, \ldots, N-1. \tag{1.5}$$

The value of N can be freely chosen by selecting the mother wavelet ψ_0 accordingly. The Fourier transform $\Psi_0(\nu)$ of ψ_0 satisfies $|\Psi_0(\nu)| \approx |\nu|^N$, $|\nu| \to 0$ [24].

1.2.1.4 Fast pyramidal algorithm In all of what follows, we always assume that we are dealing with continuous time stochastic processes, and therefore that the wavelet (and approximation) coefficients are defined through continuous time inner products (Eq. (1.4)). One major consequence of the nested structure of the MRA consists in the fact that the $d_X(j,k)$ and the $a_X(j,k)$ can actually be computed through a discrete time convolution involving the sequence $a_X(j-1,k)$ and two discrete time filters h_1 and g_1. The DWT can therefore be implemented using a recursive filter-bank-based pyramidal algorithm, as sketched on Fig. 1.1, which has a lower computational cost than that of a FFT [24]. The coefficients of the filters h_1 and g_1 are to be derived from ϕ_0 and ψ_0 [24]. The use of the discrete time algorithm to compute the continuous time inner products $d_X(j,k) = \langle X, \psi_{j,k} \rangle$ requires an initialization procedure. It amounts to computing an initial discrete time sequence to feed the algorithm (see Fig. 1.1): $a_X(0,k) = \langle X, \mathring{\phi}_{0,k} \rangle$ which corresponds to the coefficients of the expansion of the projection of X on V_0. From a practical point of view, one deals with sampled versions of X, which implies that the initialization stage has to be approximated. More details can be found in [75, 27]. The fast pyramidal algorithm is not only scalable because of its linear complexity, $O(n)$ for data of length n, but is simple enough to implement online and in real-time in high speed packet networks. An online wavelet-based estimation method for the scaling parameter with small memory requirements is given in [62].

1.2.2 Scaling processes: self-similarity and long range dependence

We can define *scaling behaviour* broadly as a property of scale invariance, that is, when there is no controlling characteristic scale, or equivalently, when all scales have equal importance. There is no one simple definition which can capture all systems or processes with this property, rather there are a set of known classes open to expansion. In this section we briefly introduce the most well known of these, namely *self-similar*, self-similar with stationary

increments, and *long-range dependent* processes. Please note that throughout this paper we will use the following convention: $f(x) \sim g(x)$ as $x \to a$ means that $\lim_{x\to a} f(x)/g(x) = 1$, and $f(x) \approx g(x)$ as $x \to a$ means that $\lim_{x\to a} f(x)/g(x) = C$ where C is some finite constant.

Recall that a process $X = \{X(t),\ t \in \mathbb{R}\}$ is *self-similar with parameter* $H > 0$ (H-ss) if $X(0) = 0$ and $\{X(ct),\ t \in \mathbb{R}\}$ and $\{c^H X(t),\ t \in \mathbb{R}\}$ have the same finite-dimensional distributions. Such a process, obviously, cannot be stationary. The process X is H-sssi if it is H-ss and if, in addition, it has stationary increments, that is, if the finite-dimensional distributions of its increments $\{X(t+h) - X(h),\ t \in \mathbb{R}\}$ do not depend on h. A H-sssi process with $H < 1$ has zero mean and a variance which behaves as $\mathbb{E}X^2(t) = \sigma^2|t|^{2H}$. The *Fractional Brownian Motion* (FBM), for example, is the (unique) Gaussian H-sssi process, which is simply Brownian Motion for $H = 1/2$.

Long-range dependence (LRD)[1], on the other hand, is associated with *stationary* processes. A stationary finite variance process X displays *long-range dependence* if its spectral density $\Gamma_X(\nu)$ satisfies

$$\Gamma_X(\nu) \sim c_f|\nu|^{-\alpha} \text{ as } \nu \to 0 \tag{1.6}$$

where $0 < \alpha < 1$ and where c_f is a non-zero constant [2]. Eq. (1.6) implies that the autocovariance $r(k)=\mathbb{E}Z(j)Z(j+k)$ satisfies

$$r(k) \sim c_r k^{\alpha-1} \text{ as } k \to \infty, \tag{1.7}$$

where $c_r = c_f 2\Gamma(1-\alpha)\sin(\pi\alpha/2)$, Γ being (here) the Gamma function ([17], p. 43). Eqs. (1.7) and (1.6) imply that the covariances $r(k)$ decay so slowly, that $\sum_{k=-\infty}^{\infty} r(k) = \infty$, or equivalently, $\Gamma_Z(0) = \infty$.

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

$$\alpha = 2H - 1. \tag{1.8}$$

In particular, fractional Gaussian noise (FGN), which is the increment process of fractional Brownian motion[3] (FBM) [50] with $1/2 < H < 1$, has long-range dependence. FGN is close to an "ideal" model because its spectral density is close to $\nu^{1-2H} \equiv \nu^{-\alpha}$ for a large range of frequencies ν in the interval $[0, 1/2]$,

[1] Long-range dependence is sometimes referred to as "long memory" or "second-order asymptotic self-similarity".

[2] The index f indicates that this constant is in force in the frequency domain. The corresponding constant appearing in the autocovariance is denoted c_r. One can also replace these constants by slowly-varying functions but for the sake of simplicity, we will not do this here.

[3] Discrete standard FGN is the time series $X(j) = B_H(j+1) - B_H(j),\ j = 0, 1, \ldots$, where B_H is FBM. Its spectral density satisfies $\Gamma_X(-\nu) = \Gamma_X(\nu)$, and because it is a discrete-time sequence, $\Gamma_X(\nu)$ is concentrated on the interval $[-1/2, 1/2]$.

and because its correlation function:

$$r(k) = 1/2\{(k+1)^{2H} - 2k^{2H} + |k-1|^{2H}\} \tag{1.9}$$

is invariant under *aggregation* (see section 3.5.1).

We now recall the properties of the wavelet coefficients of H-sssi processes (such as FBM), and LRD processes (such as FGN), and show that they can be gathered into a unified framework. We subsequently show that other stochastic processes exhibiting scaling behaviour also fit into this framework, opening up the prospect of a single approach covering diverse forms of scaling.

1.2.3 Wavelet transform of scaling processes

1.2.3.1 Discrete wavelet transform of stochastic processes Whereas the wavelet theory was first established for deterministic finite energy processes, it has been clearly demonstrated in the literature that the wavelet transform can be applied to stochastic processes, see e.g. [20, 49]. More specifically, for the second-order random processes of interest here, it is well-known that the wavelet transform is a second-order random field, on the condition that the scaling function ϕ_0 (and hence the wavelet ψ_0) satisfy certain mild conditions [20, 49] related to the covariance structure of the analysed process. We will hereafter assume that the scaling functions and wavelets decay at least exponentially fast in the time domain, so that the second-order statistics of the wavelet transform exist for all of the random processes we discuss here.

1.2.3.2 WT of H-ss and H-sssi processes Let X be a H-ss process. Its wavelet coefficients $d_X(j,k)$ exactly reproduce the self-similarity through the following central scaling property, see [25, 26] or also [57]:

- **P0 SS:** For the DWT, $d_X(j,k) = \langle X, \psi_{j,k} \rangle$, so that

$$\begin{aligned}(d_X(j,0), d_X(j,1), \ldots, d_X(j,N_j-1)) &\stackrel{d}{=} \\ 2^{j(H+\frac{1}{2})}(d_X(0,0), d_X(0,1), \ldots, d_X(0,N_j-1)).\end{aligned} \tag{1.10}$$

  For the CWT, $T_X(a,t) = \langle X, \psi_{a,t} \rangle$, and hence

$$\begin{aligned}(T_X(ca,ct_1), \ldots, T_X(ca,ct_n)) &\stackrel{d}{=} \\ c^{H+1/2}(T_X(a,t_1), \ldots, T_X(a,t_n)), \forall c > 0.\end{aligned}$$

  These equations mimic the self-similarity of the process. Let us emphasize that this, non trivially, results from the fact that the analysing wavelet basis is designed from the dilation operator and is therefore, by nature, scale invariant (**F1**). For second-order processes, a direct

consequence of Eq. (1.10) is

$$\mathbb{E}d_X(j,k)^2 = 2^{j(2H+1)}\mathbb{E}d_X(0,k)^2\,. \tag{1.11}$$

If we, moreover, add the requirement that X has stationary increments (i.e., X is H-sssi), ingredients (**F1**) and (**F2**) combine resulting in:

- **P1 SS:** The wavelet coefficients with fixed scale index $\{d_X(j,k), k \in \mathbb{Z}\}$ form a stationary process and moreover are identically distributed. This follows from the stationary increments property of the analysed processes [20, 49, 25]. This property is not trivial given that self-similar processes are non stationary processes and is a consequence of $N \geq 1$ (**F2**). In this case, Eq. (1.11) reduces to the fundamental result:

$$\mathbb{E}d_X(j,k)^2 = 2^{j(2H+1)}C(H,\psi_0)\sigma^2, \forall k\,, \tag{1.12}$$

  with $C(H,\psi_0) = \int |t|^{2H}\left(\int \psi_0(u)\psi_0(u-t)du\right)dt$ and $\sigma^2 = \mathbb{E}X(1)^2$.

- **P2 SS:** Using the specific covariance structure of an H-sssi process $X(t)$, namely,

$$\mathbb{E}X(t)X(s) = \frac{\sigma^2}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}, \tag{1.13}$$

  it can be shown [32, 73] that the correlations between wavelet coefficients located at different positions is extremely small as soon as $N \geq H + 1/2$ and their decay can be controlled by increasing N:

$$\mathbb{E}d_X(j,k)d_X(j',k') \approx |2^{-j}k - 2^{-j'}k'|^{2H-2N},\ \ |2^{-j}k - 2^{-j'}k'| \to +\infty. \tag{1.14}$$

These two results have been obtained and illustrated originally in the case of the FBM in [31, 32, 33, 34] (see also [73]) and have been stated in more general contexts in [20, 49, 25, 26].

1.2.3.3 WT of LRD processes Let X be a second-order *stationary* process, its wavelet coefficients $d_X(j,k)$ satisfy:

- **P0 LRD:**

$$\mathbb{E}d_X(j,k)^2 = \int \Gamma_X(\nu)2^j|\Psi_0(2^j\nu)|^2 d\nu \tag{1.15}$$

where $\Gamma_X(\nu)$ and $\Psi_0(\nu)$ stands for the power spectrum of X and the Fourier transform of ψ_0. This can be understood as the classical *interference* formula of the linear filter theory and receives a spectral estimation interpretation: $\mathbb{E}d_X(j,k)^2$ is a measure of $\Gamma_X(\cdot)$ at frequency $\nu_j = 2^{-j}\nu_0$ (ν_0 depends on ψ_0) through the constant relative bandwidth wavelet filter [2, 3, 1, 34].

In the specific context of LRD processes, (**F1**) and (**F2**) together yield the two following key properties:

- **P1 LRD:** Using $\Gamma_X(\nu) \sim c_f|\nu|^{-\alpha}, \nu \to 0$ in (1.15) above, we obtain:

$$\mathbb{E}d_X(j,k)^2 \sim 2^{j\alpha} c_f C(\alpha, \psi_0),\ j \to +\infty, \tag{1.16}$$

  where $C(\alpha, \psi_0) = \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu$, $\alpha \in (0,1)$. The case of $\alpha = 0$ is well defined, corresponding to trivial scaling at large scales, leaving only short range dependence at small scales. Again, this asymptotic recovering of the underlying power-law is not a trivial result. It would not, for instance, be obtained with periodogram based spectral estimates [3] and is due to (**F1**).

- **P2 LRD:** It can also be shown [3] that the covariance function of any two wavelet coefficients is controlled by N and therefore can decay much faster than that of the LRD process itself and is no longer LRD as soon as $N \geq \alpha/2$. Since $\alpha \in [0,1)$, this is in fact always satisfied.

$$\mathbb{E}d_X(j,k)d_X(j',k') \approx |2^{-j}k - 2^{-j'}k'|^{\alpha-1-2N},\ |2^{-j}k - 2^{-j'}k'| \to +\infty. \tag{1.17}$$

Observe that the exponents in **P1 LRD** and **P2 LRD** are different from those in **P1 SS** and **P2 SS** respectively.

1.2.3.4 WT of generalized scaling processes The results above can be generalized in a straightforward manner to processes which are neither strictly H-sssi nor LRD but whose wavelet coefficients share *equivalent scaling* properties. Some important cases are detailed here.

- Start with a H-sssi process X, and define Y as

$$Y(t) = \underbrace{\int_0^t dt_{p-1} \int_0^{t_{p-1}} \dots dt_1 \int_0^{t_1} du}_{p-integrals} X(u).$$

  Then Y is a H_Y-ss process with self-similarity parameter $H_Y = H + p$ and with stationary increments of order $p+1$. We say that Z is the p^{th}-order ($p > 0$) increment process of Y if $Z(t) = Y^{(p-1)}(t+1) - Y^{(p-1)}(t)$ and $Y^{(p-1)}(t) = d^{(p-1)}Y/dt^{(p-1)}$ (note that we use such a "mixed" definition because an H-sssi process (i.e., with $0 < H < 1$) is not differentiable, whereas its integrals are). Then, properties (**P1 SS**) and (**P2 SS**) still hold replacing H by H_Y. The condition for (**P1 SS**) becomes $N \geq p+1$ [10] and can be rewritten as $N \geq H_Y$ [10]. We hereafter say that X is a H-sssi(p) process if it is H-ss and has stationary increments of order $p+1$. Note that with this definition H-sssi($p=0$) and H-sssi are equivalent.

- Let X be a second-order stationary $1/f$-*type process*, i.e., $\Gamma_X(\nu) = c_f|\nu|^{-\alpha}$, $\nu_1 \leq |\nu| \leq \nu_2$, $\alpha \geq 0$. Note that the term $1/f$ implicitly implies

the physicist point of view where the power-law behaviour is supposed to hold for a wide range of frequencies, i.e., $\nu_1 \ll \nu_2$. Recall that the mother wavelet is a band-pass function whose frequency content is essentially concentrated between ν_A and ν_B, and negligible elsewhere, if non zero. In the case of $1/f$ processes, it is therefore assumed that $|\nu_2 - \nu_1| \gg |\nu_B - \nu_A|$. We henceforth have:

$$\mathbb{E}d_X(j,k)^2 \simeq \int_{2^{-j}\nu_A<|\nu|<2^{-j}\nu_B} \Gamma_X(\nu) 2^j |\Psi(2^j\nu)|^2 d\nu.$$

This means that for all j's such that $\nu_1 \leq 2^{-j}\nu_A \leq 2^{-j}\nu_B \leq \nu_2$, the wavelet coefficients of X will reproduce the power-law: $\mathbb{E}d_X(j,k)^2 \simeq 2^{j\alpha} c_f C(\alpha, \psi_0)$. Strictly speaking, this last relation holds for wavelets whose frequency support is finite, but it is generally valid to an excellent approximation. $1/f$-type processes with $\alpha < 1$ and $\nu_1 \equiv 0$ can be seen as a special case of LRD processes. Note that the definition of $1/f$ processes naturally extends to include $\alpha < 0$.

- Let X be such that $\Gamma_X(\nu) \sim c_f|\nu|^{-\alpha}, \nu \to 0, \alpha \geq 0$. For $\alpha \geq 1$, the variance does not exist (the integral of the spectrum diverges). X can however be seen as a generalized second-order stationary $1/f$-type process, in the sense that the variance of the wavelet coefficients remains finite $\mathbb{E}d_X(j,k)^2 = \int \Gamma_X(\nu) 2^j |\Psi_0(2^j\nu)|^2 d\nu = 2^{j\alpha} c_f \int |\nu|^{-\alpha} |\Psi_0(\nu)|^2 d\nu < \infty$, on condition that $N > (\alpha - 1)/2$. This is possible as the power law decrease of the spectrum of the wavelet at the origin $|\Psi_0(\nu)| \approx \nu^N, |\nu| \to 0$ balances the divergence of $\Gamma_X(\nu)$ (see [3, 4] for details). Then, just as before, we have: $\mathbb{E}d_X(j,k)^2 \sim 2^{j\alpha} c_f C(\alpha, \psi_0)$, $j \to +\infty$.

- Let X be such that $\Gamma_X(\nu) \sim c_f|\nu|^{-\alpha}, \nu \to \infty, \alpha \geq 1$, (i.e., $\nu_2 = \infty$). Its autocovariance function reads: $\mathbb{E}X(t)X(t+\tau) \sim \sigma^2(1 - C|\tau|^{2h}), \tau \to 0$, with $h = (\alpha - 1)/2$. Equivalently, it implies that $\mathbb{E}(X(t+\tau) - X(t))^2 \approx |\tau|^{2h}, \tau \to 0$. If X is moreover Gaussian, this implies that the sample path of each realization of the process is fractal, with fractal dimension (strictly speaking Hausdorff dimension) $D = (5 - \alpha)/2$ [28]. This means that the local regularity of the sample path of the process, or equivalently, its local correlation structure, exhibits scaling behaviour. Such processes are called *fractal*. Fractality is reproduced in the wavelet domain (generalization of **P1**) through $\mathbb{E}d_X(j,k)^2 \approx 2^{j(2h+1)}$, $j \to -\infty$, or equivalently for the CWT: $\mathbb{E}|T_X(a,t)|^2 \approx a^{2h+1}$, $a \to 0$ [36, 35], which allows an estimation of the fractal dimension through that of the scaling exponent $\alpha = 2h + 1 = 5 - 2D$.

1.2.3.5 Summary for scaling processes Let X be either a H-sssi(p) process, or a LRD process, or a (possibly generalized) second-order stationary $1/f$-type process or a fractal process, then the wavelet coefficients will, due to

the combined effects of (**F1**) and (**F2**), exhibit the two following properties, which will play a key role in the estimation of the scaling exponent presented below:

- **P1:** The $\{d_X(j,k), k \in \mathbb{Z}\}$ is a stationary process if $N \geq (\alpha-1)/2$ and the variance of the $d_X(j,k)$ accurately reproduces, *within a given range of octaves* $j_1 \leq j \leq j_2$, the underlying scaling behaviour of the data:

$$\mathbb{E}d_X(j,k)^2 = 2^{j\alpha} c_f C(\alpha, \psi_0), \tag{1.18}$$

  where

  - in the case of an H-sssi(p) process, $\alpha = 2H+1$, $C(\alpha,\psi_0)$ is to be identified from Eq. (1.12), and $j_1 = -\infty$ and $j_2 = +\infty$;
  - in the case of a LRD process, α is defined as in (1.6), $C(\alpha,\psi_0)$ is to be identified from Eq. (1.16), and $j_2 = +\infty$, j_1 is to be identified from the data;
  - in the case of a (generalized) second-order stationary $1/f$-type process, α is defined from $\Gamma_X(\nu) = c_f|\nu|^{-\alpha}$, $\nu_1 \leq |\nu| \leq \nu_2$, $C(\alpha,\psi_0) = \int |\nu|^{-\alpha}|\Psi_0(\nu)|^2 d\nu$ and (j_1, j_2) are to be derived from (ν_1, ν_2);
  - in the case of a fractal process, $\alpha = 2h+1$, expressions for $C(\alpha,\psi_0)$ can be found in [36, 35] and $j_1 = 1$, j_2 is to be identified from the data.

- **P2:** $\{d_X(j,k), k \in \mathbb{Z}\}$ is stationary and no longer exhibits long-range statistical dependences but only short term residual correlations, i.e., it is short-range dependent (SRD) and not LRD, on condition that $N \geq \alpha/2$. Moreover the higher N the shorter the correlation:

$$\mathbb{E}d_X(j,k)d_X(j,k') \approx |k-k'|^{\alpha-1-2N}, \ \ |k-k'| \to +\infty.$$

Note that these two properties of the wavelet coefficients do not rely on an assumption of Gaussianity. In **P2** above, we used only weak reformulations (setting $j = j'$) of **P2 SS** and **P2 LRD.** Their general versions (j not necessarily equal to j') can be used to formulate a stronger idealisation of strict decorrelation:

ID1: $\mathbb{E}d_X(j,k)d_X(j',k') = 0$ if $(j',k') \neq (j,k)$.

The relevance of this idealisation has already been illustrated in, for instance, [32, 33, 3, 5], and will play a key role in the next section.

1.2.3.6 Multiple exponents, Multifractal processes. Property **P1** (wavelet reproduction of the power law) extends further to classes of generalized scaling processes whose behaviour cannot be described by a single scaling exponent, but which requires a collection, even an infinite collection, of exponents. We briefly describe three classes of examples.

The first example is in the spirit of the simple fractal processes described in section 2.3.4. Consider a generalisation where the exponent h, which describes the statistics of local scaling properties, is no longer constant in time: $\mathbb{E}(X(t+\tau)-X(t))^2 \approx |\tau|^{2h(t)}$, $\tau \to 0$. One consequence is that the local regularity of sample paths is no longer uniform, but depends on t. A class of processes called *Multifractional Brownian Motion* has been proposed in [56] which satisfies such a property, with h being a continuous function of t. As detailed in [35, 36], the time evolution of h can be traced through an analysis of the continuous wavelet transform coefficients at small scales: $\mathbb{E}|T_X(a,t)|^2 \approx a^{2h(t)+1}$, $a \to 0$. This relation is to be understood as a time dependent generalization of **P1.**

The second class, *multifractal processes*, is one which allows an extremely rich scaling structure at small scales, far richer than simply fractal in general. There is not the space here to give precise definitions of such processes, nor of the related *multifractal formalism*. We aim rather to give some intuition of their relation to wavelets, and refer the reader to [60, 59] and to chapter **??** of the present volume, and references therein, for a thorough presentation. For multifractal processes, the local regularity of almost every (that is with probability one) sample path, which we write as $|X(\omega,t+\tau)-X(\omega,t)| \approx |\tau|^{h(\omega,t)}$, $\tau \to 0$ (where ω denotes an element of the probability space underlying the process) exhibits an extraordinary variability over time, indeed it is itself fractal-like. One therefore abandons the idea of *following* the time variations of h, since this is realisation dependent and in any case is too complex, but instead studies it statistically. Classically this has been done through the so-called *Hausdorff multifractal spectrum* $D(h)$, which consists of the *Hausdorff dimension* of the set of points where $h(\omega,t)=h$. The same multifractal spectrum is obtained for almost all realisations and is therefore a useful invariant describing the scaling properties of the process. A classical tool to obtain the multifractal spectrum is to calculate, from any typical sample path, the *structure functions* or *partition functions*: $S_q(\tau) = \int |X(\omega,t+\tau)-X(\omega,t)|^q dt$. It is known that for given classes of multifractal processes [42], such $S_q(\tau)$ exhibit power-law behaviour $S_q(\tau) \approx |\tau|^{\zeta(q)}, \tau \to 0$, $q \in \mathbb{R}$ which is deeply related to their multifractal nature. Another multifractal spectrum, namely the *Legendre multifractal spectrum*, can then be obtained by taking the Legendre transform of $\zeta(q)$. Although it is possible that the Legendre spectrum be different and in fact less rich than the Hausdorff spectrum, it is used as it is far more numerically accessible. The connection between multifractals and wavelets arises from the fact that the increments involved in the study of the local regularity of a sample path can be seen as simple examples of wavelet coefficients [52]. It has therefore been proposed heuristically [52] to replace increments by wavelet coefficients in the partition functions and shown theoretically that, in some cases, the multifractal formalism can be based directly on wavelet coefficients [42, 60, 16]. For the Legendre multifractal spectrum, this amounts to using wavelet-based partition functions that exhibit, for small scales, power-law behaviour: $\int |T_{X(\omega)}(a,t)|^q dt \approx a^{\zeta(\omega,q)+q/2}$, $a \to 0$. This last

relation can be thought of as a generalization of **P1** to statistics of order both above and below 2. In addition, it is important to understand that even though the relation describes a property of a single (typical) realisation, it deals directly with the object $\zeta(\omega, q)$ central to the description of the scaling, and **not** to an estimator of it. This is in contrast to self-similar processes for example, and the fractal class of the previous paragraph, where the fundamental scaling relations and exponents are defined at the level of the ensemble. Such a change of perspective is meaningful for multifractals as almost all realisations yield a *common* function $\zeta(q)$. Finally let us note that more refined wavelet-based partition functions have been proposed to overcome various difficulties arising in signal processing, the reader is referred to e.g., [52, 16].

The third example is that of *multiplicative cascades*, a paradigm introduced by Mandelbrot [51] in 1974. It involves a recursive procedure whereby an initial mass is progressively subdivided accorded to a geometric rule and assigned to subsets of an initial set, typically an interval. It provides a powerful tool to define multifractal processes and was originally considered as a natural synthesis procedure for them. Indeed cascade based methods of generating multifractals have been the preferred option thus far in tele-traffic applications (see Chapter **??**). However the infinitely divisible model proposed by Castaing [21] shows that multiplicative cascade processes can also very effectively model scaling phenomena in other cases, even where the scaling is barely observable in the time domain. Again, the wavelet tool has proved useful for the analysis of such situations, as comprehensively detailed in [15, 14]. This tool has been applied for instance in the study of turbulence, [63, 22].

1.2.3.7 Processes with infinite second-order statistics - $\alpha-$stable processes. The existence of the wavelet coefficients, the extensions of **P0 SS**, **P1 SS** and **P2 SS**, to H-sssi processes without second-order statistics, such as $\alpha-$stable processes, for instance, have recently been obtained [25, 26, 58] (see also [57]) but will not be detailed here.

1.3 WAVELETS AND SCALING: ESTIMATION

In this section it is shown in detail how the statistical properties of the wavelet detail coefficients, summarized in the previous section in the form of properties **P1** and **P2**, can be applied to the related tasks of the detection, identification, and measurement of scaling. The estimation of scaling exponents, 'magnitude of scaling' parameters, and the multifractal spectrum is discussed. Practical issues in the use of the estimators are addressed and comparisons are made with other estimation methods. Robustness of different kinds is also discussed. It is shown how wavelet methods allow statistics other than second order to be analysed, with applications in the identification of self-similar and *multifractal*

processes. It is explained how the wavelet framework allows a re-interpretation and a fruitful extension of the natural idea of *aggregation* in the study of scaling. It is shown how the *Allan variance*, an effective time domain estimator of scaling, belongs in fact to this framework. Finally it is shown how the same analysis methods can be applied to the measurement of generalised forms of the *Fano factor*, a well known descriptor of the burstiness of point processes.

1.3.1 An Analysis Tool: The Logscale Diagram

1.3.1.1 The legacy of P1 and P2. The property **P2** is the key to the statistical advantages of analysis in the wavelet domain. In sharp contrast to the problematic statistical environment in the time domain due to the long-range dependence, non-stationarity, or fractality of the original process $X(t)$, in the wavelet domain we need only deal with the stationary, short-range dependent (SRD) processes $d_X(j,\cdot)$ for each j. (Due to the admissibility condition of the mother wavelet these processes each have zero mean.) The stationarity allows us to meaningfully average across 'time' within each process to reduce variability. The short-range dependence results in these average statistics having small variance. An example of central importance here is given by

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j,k)|^2, \tag{1.19}$$

where n_j is the number of coefficients at octave j available to be analysed. The random variable μ_j is a non-parametric, unbiased estimator of the variance of the process $d_X(j,\cdot)$. Despite its simplicity, because of the SRD the variance of μ_j decreases as $1/n_j$ and it is in fact asymptotically efficient (of minimal variance). The variable μ_j can therefore be thought of as a near-optimal way of concentrating the gross second order behaviour of X at octave j. Furthermore, again from **P2**, the μ_j are themselves only weakly dependent, so the analysis of each scale is largely decoupled from that at other scales. To analyse the second order dependence of $X(t)$ on scale therefore, we are naturally lead to study μ_j as a function of j.

Property **P1** now enters by showing explicitly, in the case of scaling, the underlying power-law dependence in j of the variance (second moment) of the processes at each scale, of which the μ_j are estimates. The importance of **P1** is that its pure power-law form suggests that the scaling exponent α could be extracted simply by considering the slope in a plot of $\log_2(\mu_j)$ against j. Here it is essential to understand that, although log-log plots are a natural and familiar tool whenever exponents of power-laws are at issue, using them as a basis for semi-parametric estimation of the exponent is only effective statistically if properties equivalent to **P1-P2** hold. This is typically not the case. For example for the *correlogram*, a time domain semi-parametric estimator [17] based on direct estimation of the covariance function, covariance estimates at fixed lag are biased, resulting in bias in the exponent estimate.

Furthermore, across lags the covariance estimates are strongly correlated, resulting in misleadingly impressive 'straight lines' in the log-log plot, which in reality are symptomatic of high variance in the resulting estimate. In addition to these issues, the complication that in general $\mathbb{E}[\log(\cdot)] \neq \log(\mathbb{E}[\cdot])$ is overlooked in the correlogram and in many other estimators based on log-log plots. For simplicity of presentation we set $y_j = \log(\mu_j)$ for the moment, but address this refinement in the estimation section below. We now introduce a wavelet-based analysis tool, the *Logscale Diagram*, which exploits the key properties **P1** and **P2**, and serves as an effective and intuitive central starting point for the analysis of scaling.

Definition. The (second order) *Logscale Diagram* (LD) consists in the graph of y_j against j, together with confidence intervals about the y_j.

Examples of Logscale Diagrams analysing synthesized scaling data are given in Fig. 1.2, where the plot on the left is of a LRD series, and that on the right of a self-similar series. It follows from the nature of the dilation operator generating the wavelet basis that the number n_j of detail coefficients at octave j halves with each increase in j (in practice the presence of border effects results in slightly lower values). Confidence intervals about the y_j therefore increase monotonically with j as one moves to larger and larger scales, as seen in each of the diagrams in Fig. 1.2. The exact size of these intervals depend on details of the process and in practice are calculated using additional distributional and quasi-decorrelation assumptions. If necessary they could also be estimated from data.

Generalisations to q^{th} order Logscale Diagrams can be defined, $q > 0$, where the second moment of the details in Eq. (1.19) is replaced by the q^{th}. Here we mainly concentrate on the second order Logscale Diagram or simply 'Logscale Diagram', both as an illustrative example, and because it is the most important special case, being central for LRD and $1/f$ processes by definition, definitive for Gaussian processes, and sufficient for exactly self-similar processes. Like any second order approach, it is of course insufficient for processes whose second moments do not determine all the properties of interest. We discuss this further in section 4.3 in the particular context of multifractals.

The Logscale Diagram is first of all a means to visualize the scale dependence of data with a minimum of preconceptions. Scaling behaviour is not assumed but *detected*, through the region(s) of *alignment*, if any, observed in the log-log plot. By an alignment region we mean a range of scales where, up to statistical variation, the y_j fall on a straight line. *Estimation* of scaling parameters, if relevant, can then be effectively performed through weighted linear regression over the region(s). Finally, the *identification* of the kind of scaling is made by interpreting the estimated value in the context of the observed range. These different aspects of the aims and use of the Logscale Diagram are expanded upon below.

1.3.1.2 The detection of scaling. A priori it is not known over which scales, if any, a scale invariant property may exist. By the *detection* of scaling in the Logscale Diagram we mean the identification of region(s) of *alignment*, and the determination of their lower and upper cutoff octaves, j_1 and j_2 respectively, which are taken to correspond to scaling regimes. In a sense this is an insoluble problem, as scaling often occurs asymptotically, or has an asymptotic definition, with no clear way to define how a scaling range begins or ends. Nonetheless experience shows that good estimates are possible. Note the semantic difference between the term *scaling region* or range, a theoretical concept which refers to where scaling is truly present (an unknown in real data), and *alignment region* or range, an estimation concept corresponding to what is actually observed in the Logscale Diagram for a given set of data.

The first essential point here is that the concept of alignment is relative to the confidence intervals for the y_j, and **not** to a close alignment of the y_j themselves. Indeed, an undue alignment of the actual estimates y_j indicates strong correlations between them, a highly undesirable feature typical of time domain log-log based methods such as variance-time plots. As mentioned earlier the μ_j, and hence the y_j, are weakly dependent, resulting in a natural and desirable variation around the calculated regression line as seen, for example, in Fig. 1.2. Using weighted regression incorporates the varying confidence intervals into the estimation phase, however the selection of the *range of scales* defining the alignment region is prior to this, and great care is required to avoid poor decisions.

We now discuss the selection, in practice, of the cutoff scales j_1 and j_2. A preliminary comment is that for the regression to be well defined at least two scales are required, for a Chi-squared goodness of fit test three, and in practice four are needed before any estimate can be taken seriously: it is simply too easy for three points to align fortuitously if the confidence intervals are not very small. A useful heuristic in the selection of a range is that the regression line should cut, or nearly so, each of the confidence intervals within it. This can help avoid the following two errors: (1) the non-detection of an alignment region due to the apparently wild variation of the y_j, when in fact to within the confidence intervals the alignment is good (this typically occurs when the slope is small, such as in the right hand plot in Fig. 1.4, as the vertical scale on plots is reduced, increasing the apparent size of variations), and (2) the erroneous inclusion of extra scales to the left of an alignment region, as to the eye they appear to accurately continue a linear trend, whereas in fact the small confidence intervals about the y_j for small j reveal that they depart significantly from it.

The above heuristic can be formalised somewhat by a Chi-squared goodness of fit test [9], where the critical level of the goodness of fit statistic is monitored as a function of the endpoints of the alignment range. At least in the case of the lower scale, this can make a very clear and relatively objective choice of cutoff possible, eliminating the error of type (2) above. An example of this is afforded by the left hand plot of Fig. 1.2, where the octave $j = 3$,

if included, results in a drop in the Chi-squared goodness of fit of several orders of magnitude! An even subtler example is that of the right hand plot in Fig. 1.3 where $j = 2$ was excluded from the alignment region for the same reason, whereas in the left hand plot in the same figure it is clear even to the eye that, given the small size of the confidence interval about octave $j = 8$, that it should not be included. Further work is required to develop reliable automated methods of cutoff scale determination. This is especially true for upper cutoff scales, where the difficulties are compounded by a lack of data. On the other hand at smaller scales the technical assumptions used in the calculation of the confidence intervals (see below) may be less reliable, whereas at large scales the data is highly aggregated and therefore Gaussian approximations are reasonable.

1.3.1.3 The interpretation of scaling. By the *interpretation* of scaling we mean the identification of the kind of underlying scaling phenomenon, LRD, H-ss, etc, generating the observed alignment in the Logscale Diagram. The task is one of the meaningful interpretation of the estimated value of α in the context of the range of scales defining the alignment region, informed where possible by other known or assumed properties of the time series such as stationarity. It is in fact partly a question of model choice, and there may be no unique solution. We now consider, non-exhaustively, a number of important cases.

If an estimate of the scaling exponent α is found to lie in $(0, 1)$, and the range of scales is from some initial value j_1 up to the largest one present in the data, then the scaling could be said to correspond to LRD with a scaling exponent which is simply the measured α. Examples are afforded by the left hand plots in each of Fig. 1.2 and Fig. 1.3. If there were a priori physical reasons to believe that the data were stationary, then LRD would be an especially relevant conclusion. This applies to the left hand plot in Fig. 1.3, as the series corresponds to successive inter-arrival times of Ethernet packets, which under steady traffic conditions one would expect to be stationary (the Ethernet data in Fig. 1.3 is from from the 'pAug' Bellcore trace [43]).

Another key example, illustrated in the right hand plots in Fig. 1.2 and Fig.1.3, is a value of α *greater* than 1 but also measured over a range including the largest scales. Such a value precludes LRD, and may indicate that a self-similar or asymptotically self-similar model is required, implying that the data is non-stationary. The exponent would then be re-expressed as $H = (\alpha - 1)/2$, the Hurst parameter. Again conclusions should be compared with a priori physical reasoning. The right hand plot in Fig. 1.3 is the analysis of a cumulative work process for Ethernet traffic, that is the total number of bytes having arrived by time t. Such a series is intrinsically non-stationary, though under steady traffic conditions one would expect it to have stationary increments. Thus a conclusion of non-stationarity is a natural one, and the estimated value of $\hat{H} = 0.80$, being in $(0, 1)$, indeed corresponds to a H-sssi process. It would have been problematic however if underlying physical

reasons had indicated that in fact stationarity was to be expected. Such an apparent paradox could be resolved in one of two ways. It may be that the underlying process is indeed stationary and exhibits $1/f$ noise over a wide range of scales, but that the data set is simply not long enough to include the upper cutoff scale. The alternative is to accept that empirical evidence has shown the physical reasoning concluding stationarity to be invalid.

If on the other hand the scaling were concentrated at the lowest scales (high frequencies), that is $j_1 = 1$ with some upper cutoff j_2, then the scaling may best be understood as indicating the fractal nature of the sample path. The observed α should then be re-expressed as $h = (\alpha - 1)/2$, the local regularity parameter. Values of h in the range $(0, 1]$ for example would then be interpreted as indicating continuous but non-differentiable sample paths (under Gaussian assumptions [28]), as observed in the leftmost alignment region in the Internet delay data in the left plot in Fig. 1.4. The stationarity or otherwise of the data in such a case may then not be relevant. Note that $j = 1$ has been excluded from the leftmost alignment region in each plot in Fig. 1.4. This is not in contradiction with interpretations of fractality, as it is known that the details at $j = 1$ can be considerably polluted due to errors in the initialisation of the multiresolution algorithm (see section 2.1.4).

If scaling with $\alpha > 1$ is found over all or almost all of the scales in the data, such as in the right hand plot in Fig. 1.2, then exact self-similarity could be chosen as a model, again with $H = (\alpha - 1)/2$ being the relevant exponent. However in this case one could equally well use the local regularity parameter $h = (\alpha - 1)/2$, with the interpretation that the fractal behaviour at small scales is constant over time, and happens to extend right up to the largest scales in the data.

Finally, more than one alignment region is certainly possible within a single Logscale Diagram, a phenomenon which we refer to as *biscaling*. One could imagine for example fractal characteristics leading to an alignment at small scales with one exponent, and LRD resulting in alignment at large scales with a separate scaling exponent. Examples of this phenomenon are shown in Fig. 1.4 in the context of delays (left) and losses (right) experienced by consecutive User Datagram Protocol (UDP) packets sent over the Internet in a regular stream (see [12] for details of such data sets). In both figures the alignment at large scales corresponds to LRD, whereas at small scales it is associated with highly irregular sample paths. Note that when second order properties are insufficient to fully describe the scaling nature of the process (an extreme example is afforded by multifractals), then the correct interpretation of each branch of the biscaling will require the examination of Logscale Diagrams across a range of orders, as discussed in section 3.3.

1.3.2 Estimation within the Logscale Diagram

In this sub-section it is assumed that a scaling range $j \in [j_1, j_2]$ has been correctly identified. Sums over j, and regressions, are always taken over this

range. The estimators to be defined are semi-parametric, as they depend on the range of scales $j \in [j_1, j_2]$ where scaling is deemed to be present, and the scaling property **P1** valid there, but not on any tightly specified parametric model.

1.3.2.1 Estimating the scaling exponent α. Because of property **P1**, the measurement of α is reduced to the determination of the slope over the alignment region in the Logscale Diagram. A natural way to achieve this in a statistical estimation context is through linear regression, the defining hypothesis of which is $\mathbb{E}y_j = \alpha j + a$ where a is a real constant. Because in general $\mathbb{E}[\log(\cdot)] \neq \log(\mathbb{E}[\cdot])$, this condition is not exactly satisfied however. We therefore introduce small corrective deterministic factors $g(j)$, discussed below, and redefine the y_j as $y_j = \log(\mu_j) - g(j)$, so that $\mathbb{E}y_j = \alpha j + a$ by definition.

Any kind of linear regression of y_j on j constitutes an unbiased estimator of α, as the lack of bias does not require decorrelation between the y_j, nor the knowledge of their variances or distributions. A weighted regression where the weights are related to the variances σ_j^2 of the y_j is preferable however, as this is the minimum variance unbiased estimator (MVUE) for the regression problem [48]. Intuitively this refinement is significant as we know that the σ_j^2 are far from equal. To exploit the optimality however, the correction factors $g(j)$ and the variances σ_j^2 must be calculated, a difficult task. They can nonetheless be well approximated, provided simplying idealised properties are adopted. The particular idealisations chosen here are:

ID1a: For each fixed j the $d_X(j,\cdot)$ are stationary sequences of uncorrelated variables.

ID1b: The processes $d_X(j,\cdot)$ and $d_X(j',\cdot)$, $j \neq j'$, are uncorrelated.

ID2: The process X, and hence the processes $d_X(j,\cdot)$, are Gaussian.

The above conditions may appear unduly restrictive at first, however the underlying effectiveness of the method is based on **P1** and **P2**, **ID1** – **ID2** being added mainly to extend the quantitative analysis. Robustness with respect to departures from these idealisations is discussed in detail below. Note that **ID1a** and **ID1b** together make **ID1**, the idealisation of complete decorrelation. It is split here to highlight the fact that **ID1b**, independence between scales, is not needed for the key results.

Under **ID1a** and **ID2** it can be shown [75] that $g(j)$ is a negative, increasing function of n_j only, given by

$$g(j) = \psi(n_j/2)/\ln 2 - \log_2(n_j/2), \tag{1.20}$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the so called Psi function and $\Gamma(z)$ the Gamma function. This function be easily calculated for all values of n_j.

Under **ID1a** and **ID2**, with $g(j)$ as above, the variables $y_j = \log(\mu_j) - g(j)$

are scaled and shifted logarithms of Chi-squared variables, satisfying

$$\begin{aligned} \mathbb{E}y_j &= j\alpha + \log_2 c_f C, & (1.21)\\ \mathrm{Var}\,(y_j) \equiv \sigma_j^2 &= \zeta(2, n_j/2)/\ln^2 2, & (1.22) \end{aligned}$$

where $\zeta(z,\nu)$ is a generalised Riemann Zeta function (see [38], page 1072).

There may be other sets of idealisations under which the $g(j)$ and σ_j^2, if they cannot be expressed in closed form, may nonetheless be accurately calculated. In such cases the estimator would have the same properties including a variance which is known, although no longer expressible in terms of standard functions.

The estimator $\hat{\alpha}$ of α is the slope of a weighted linear regression of y_j on j given by

$$\hat{\alpha} = \frac{\sum y_j(Sj - S_j)/\sigma_j^2}{SS_{jj} - S_j^2} \equiv \sum w_j y_j, \qquad (1.23)$$

where $S = \sum 1/\sigma_j^2$, $S_j = \sum j/\sigma_j^2$ and $S_{jj} = \sum j^2/\sigma_j^2$.

1.3.2.2 Properties of the estimator. By construction, $\hat{\alpha}$ is **unbiased** under **ID1a** and **ID2**, and if in addition we assume **ID1b** its variance is simply

$$\mathrm{Var}(\hat{\alpha}) = \sum \sigma_j^2 w_j^2. \qquad (1.24)$$

Note that the variance is a function of the *amount* of data, the n_j, but is independent both of the data itself (1.22), and of the actual (unknown) value of α. It is also independent of the precise choice of mother wavelet, except indirectly through the choice of N, the number of vanishing moments. A quantitative study of this dependence is given in [27].

It has been shown [75], in the limit of n_j large for each j in $[j_1, j_2]$, that the Cramér-Rao bound for the full problem is attained, showing that $\hat{\alpha}$ is asymptotically the minimum variance unbiased estimator under **ID1–ID2**. The decrease in variance of $\hat{\alpha}$ as a function of the size of the data is then explicitly seen to be $1/n$, a remarkable result, being the rate of decay typical of short-range dependent problems, yet appearing in a difficult scaling context. Numerical comparisons [75] show that away from the limit the variance of $\hat{\alpha}$ remains extremely close to the Cramér-Rao bound. This is not surprising as the assumption that n_j is large is a very good one, except possibly for the n_j corresponding to the largest j, since $n_{j+1} \approx n_j/2$. Examining the limit in more detail, for large n_j we have [75]:

$$\begin{aligned} g(j) &\to \frac{\log_2 e}{n_j} & (1.25)\\ \sigma_j^2 &\to \frac{2(\log_2 e)^2}{n_j}. & (1.26) \end{aligned}$$

The first of these relations indicates that for n_j large, y_j can be identified with $\log_2(\mu_j)$. It has moreover been shown [3] that, under **ID1** and **ID2**, $\log_2(\mu_j)$ is asymptotically normally distributed:

$$\log_2(\mu_j) \stackrel{\mathrm{d}}{\sim} N\Big(j\alpha + \log_2(c_f C), \frac{2(\log_2 e)^2}{n_j} \Big).$$

Since $\hat{\alpha}$ consists of a sum of the y_j, most of which are approximately Gaussian and weighted according to their (known) variances, $\hat{\alpha}$ can be considered as approximately Gaussian distributed. Confidence intervals for the y_j and $\hat{\alpha}$ have been calculated using these arguments.

1.3.2.3 Robustness with respect to ID1 and ID2. Simulation studies show [5, 75] that the above properties hold to an excellent approximation, even for small size data, upon the mild departures from **ID1** characteristic of the FGN series used.

Numerical simulations presented in [5] as well as those described below show that the above properties also hold to an excellent approximation when the Gaussian hypothesis **ID2**, as well as **ID1**, is dropped. The robustness with respect to **ID2** can be justified using the following asymptotic arguments. Let $Y = f(X)$, and σ_X^2 and σ_Y^2 be the variances of X and Y respectively. Standard approximation formulae for a change of variable [54] are $\mathbb{E}f(X) \simeq f(\mathbb{E}X) + f''(\mathbb{E}X)\sigma_X^2/2$ and $\sigma_Y^2 \simeq |f'(\mathbb{E}X)|^2\sigma_X^2$. Because Var μ_j decreases as $1/n_j$ in the limit of large n_j, we can apply those formulae to $\log_2(\mu_j)$. Using **ID1** we obtain:

$$\begin{aligned} \mathbb{E}\log_2 \mu_j &\simeq \log_2 \mathbb{E}\mu_j - \tfrac{(\log_2 e)}{2} \frac{\mathrm{Var}\ \mu_j}{(\mathbb{E}\mu_j)^2} \\ &= \log_2 \mathbb{E}\mu_j - \tfrac{(\log_2 e)}{2} \frac{\mathrm{Var}\ d_X^2(j,\cdot)}{n_j (\mathbb{E}d_X^2(j,\cdot))^2} \\ &= \log_2 \mathbb{E}\mu_j - (\log_2 e) \tfrac{1+C_4(j)/2}{n_j} \end{aligned} \qquad (1.27)$$

where $C_4(j)$ is the (normalised) fourth-order cumulant of the random process $d_X(j,\cdot)$ given by $C_4(j) = \left(\mathbb{E}d_X^4(j,k) - 3(\mathbb{E}d_X^2(j,k))^2\right)/(\mathbb{E}d_X^2(j,k))^2$. From Eq. (1.25) it can be seen that the term $(\log_2 e)\,\frac{1+C_4(j)/2}{n_j}$ plays the role of $g(j)$ from the Gaussian case and, up to the $C_4(j)$ term, has the same form. Performing the regression of $y_j = \mathbb{E}\log_2 \mu_j$ on j we obtain, using $n_j = 2^{-j}n$ and under **ID1**:

$$\mathbb{E}\hat{\alpha} = \mathbb{E}\textstyle\sum_j w_j \log_2 \mu_j \simeq \alpha - \left((\log_2 e) \sum_j (1 + C_4(j)/2) w_j 2^j \right)/n \qquad (1.28)$$

which i) shows that the estimate is asymptotically unbiased irrespective of the Gaussian hypothesis and ii) allows us to subtract a first order approximation

of that bias. Similarly

$$\text{Var } \log_2 \mu_j \quad \simeq \quad 2\,(\log_2 e)^2\, \tfrac{1+C_4(j)/2}{n_j} \tag{1.29}$$

which again shows the similarity of form with the corresponding Gaussian case, Eq. (1.26), in the asymptotic limit. It follows that

$$\text{Var } \hat{\alpha} = \text{Var } \textstyle\sum_j w_j \log_2 \mu_j, \quad \simeq \quad \left(2\,(\log_2 e)^2 \textstyle\sum_j (1 + C_4(j)/2) w_j^2 2^j\right)/n \tag{1.30}$$

which again is identical, up to the C_4 term, to what would be obtained asymptotically from Eq. (1.24) under **ID2**.

The C_4 term can be estimated for each octave using the sample moment estimators of the fourth and second moments of the details, combined as per the definition. Due to the quasi-decorrelation in the wavelet plane this simple estimator will have low bias.

The above arguments clearly show that in the limit of a large number of samples, the key statistical features of the estimator, namely lack of bias and low variance, are retained and are therefore not **ID2** dependent. Numerical simulations, summarized in Table 1.3.2.3, show that this asymptotic behaviour is moreover reached for a relatively small number of samples. The simulations were performed using 1000 realizations of FARIMA$(0, d, 0)$ processes (of length $n = 2^{14} = 16384$) with a variety of probability density functions for the marginals. We used a Daubechies3 [24] wavelet and j_1 was set to $j_1 = 4$ from a preliminary analysis involving the Chi-squared goodness of fit test described above. Table 1.3.2.3 shows that the performance obtained with the non Gaussian processes is quite close to that obtained from Gaussian processes. Note that the four last chosen (Pareto and α-stable) processes are infinite variance processes. The estimates remain however unbiased and the variances, though larger, remain controlled as explained in [25, 58].

1.3.2.4 Estimating the second parameter of scaling. The scaling exponent α is a dimensionless parameter which can be thought of as characterising the *qualitative* nature of the scaling phenomena in question. Although α is clearly the key, defining parameter of scaling, it is not sufficient to fully characterise a given scaling phenomenon, nor therefore the effect that scaling may have on the distributions of various statistics, nor the impact of scaling on performance issues in applications. There is a need for, at the very least, a second parameter to describe the *quantitative* aspect of the scaling, a magnitude or 'volume' of scaling parameter. This was illustrated in the introduction in the context of the variance of the sample mean of a LRD process. There c_r was introduced as a second parameter with the dimensions of variance describing the relative role that LRD plays. Similarly, for self-similar processes the variance σ^2 of the marginal at $t = 1$ is a free parameter which also requires estimation. These 'magnitude' parameters are also problematic

RV	Bias	Variance	MSE
Gaussian	0.015	0.039	0.039
Uniform	0.017	0.036	0.036
Exponential	0.016	0.039	0.039
LogNormal	0.017	0.045	0.045
Pareto $\alpha = 1.75$	0.015	0.149	0.150
Pareto $\alpha = 1.25$	0.019	0.131	0.132
α-stable $\alpha = 1.75$	0.013	0.122	0.122
α-stable $\alpha = 1.25$	-0.004	0.228	0.228

Table 1.1 Robustness with respect to Gaussianity (ID2). It can be seen that the residual bias and variance of the wavelet-based estimator of the scaling exponent α are not very sensitive to the form of the marginal of the process $X(t)$. They are moreover very close to the theoretical performance derived assuming exact decorrelation of the wavelet coefficients and Gaussianity of $X(t)$.

to estimate using traditional methods, however as with α they can be simply and effectively estimated from the Logscale Diagram. For simplicity we will continue the discussion for the LRD case only, however an essentially identical procedure, estimator definition, and properties hold in the self-similar case. For other kinds of scaling also, magnitude parameters can be defined and estimated in a similar way, however this will not be discussed here.

We briefly summarize the results of [75, 74] for a two dimensional joint estimator $(\hat{\alpha}, \hat{c}_f)$ of LRD. One of the powerful properties of $\hat{\alpha}$ is that its statistics are entirely independent of the specific form of the mother wavelet, depending only on the coefficients of the linear regression, and the amount of data n_j at each scale. It is this feature which allows an explicit expression for its variance to be obtained independently of the wavelet basis. It is clear that this property is shared by $\hat{a} = \sum v_j y_j$, the unbiased estimator of the intercept a of the same linear regression defining α. (The coefficients v_j are given by $v_j = \sum(S_{jj} - S_j j)/(\sigma_j^2(SS_{jj} - S_j^2))$, cf. equation 1.23). From Eq. (1.16) it is apparent that whereas α is simply the slope in the Logscale Diagram, the magnitude parameter c_f is related to the intercept, being essentially proportional to $2^{\hat{a}}$, a quantity which retains the wavelet independence advantages of $\hat{a}$. Unfortunately, $2^{\hat{a}}$ does not correspond exactly to c_f but rather to the dimensionless quantity $c_f C(\alpha, \psi_0)$, and an attempt to isolate the former necessarily brings in a wavelet dependence. It is advantageous however to study $c_f C$ as the largest "wavelet independent part" of c_f, and to define subsequently

the estimator of c_f as $\hat{c}_f = \widehat{c_f C}/\hat{C}$ where $\hat{C}$ is an estimator of the integral $C(\alpha, \psi_0)$ which will not be detailed here. It can be shown [75] that $\hat{C}$ has small variance, so that the properties of $\hat{c}_f$ closely resemble those of $\widehat{c_f C}$. More specifically, define $\widehat{c_f C} = p 2^{\hat{a}}$, where p is a wavelet independent bias correcting (and variance reducing) factor given by $p = \prod \frac{\Gamma(n_j/2)\exp(\psi(n_j/2)v_j)}{\Gamma(v_j+n_j/2)}$, which is typically close to 1 (here ψ is the Psi function, not the wavelet!). This estimator is unbiased, and complex but explicit expressions for the variance of $\widehat{c_f C}$ (and covariance of $(\hat{\alpha}, \widehat{c_f C})$) are obtainable, and again the Cramér-Rao bound is attained in the limit n_j large, provided the additional condition $v_j/n_j \to 0$ is also satisfied for each j in $[j_1, j_2]$. Based on these explicit expressions for $(\hat{\alpha}, \widehat{c_f C})$, very accurate approximate expressions (which to first order are also wavelet independent), can be derived for the covariance matrix of $(\hat{\alpha}, \hat{c}_f)$.

The estimator $\hat{c}_f$ is asymptotically unbiased and efficient, and approximately lognormally distributed. The correlation coefficient of $(\hat{\alpha}, \hat{c}_f)$ is negative and large in magnitude: typically around -0.9. As before, simulation studies show [75] that the above properties hold to an excellent approximation, even for small size data, upon mild departures from **ID1**. An example of the joint estimation is given in the left hand plot of Fig. 1.2.

1.3.2.5 Comparisons with other estimators. In the evaluation of an estimator both statistical and computational aspects must be considered. The Logscale Diagram based estimators are essentially optimal computationally speaking, as they have a complexity of only $O(n)$, and a direct, non-problematic implementation which can even be performed in real-time [62]. Simple estimators of scaling such as the variogram [17] also have excellent computational properties, however they suffer from significant bias and high variance [68]. At the other end of the spectrum, fully parametric maximum likelihood estimators require the inversion of a $n \times n$ autocovariance matrix, an $O(n^3)$ operation, which is unsuitable for anything but small data sets. Even approximate forms such as the Whittle estimator, or discrete versions of it [17, 5], involve numerical minimisation and are prohibitively slow for the large ($n > 2^{14}$) data sets now routinely encountered in teletraffic studies.

Statistically, the best performing estimators are fully parametric, such as those based on maximum likelihood, which offer zero bias and optimal variance *provided that the data fits the chosen parametric model.* As mentioned above, to avoid extreme computational difficulties encountered for all but small data sets, approximate forms are used in practice which retain these desirable statistical properties asymptotically [17, 5]. In [68] and [71] a comparative discussion of the statistical properties of a variety of estimators of LRD is given. It is shown that approximate maximum likelihood-based estimators such as the Whittle, Aggregated Whittle and Local Whittle Methods, still offer the best statistical performance when compared against alterna-

tives such as the Absolute Value Method, the Variance Method, the Variance of Residuals Method, the R/S Method, and the Periodogram Method (see also [69], [72], [66]). We therefore compare against such parametric alternatives. Being parametric based, such estimators can make full use of the data and will therefore outperform Logscale Diagram based estimators, which are constrained to use only those scales where the scaling is both present, and apparent. Although it is possible that all the data may be accessible to a Logscale Diagram based estimator, that is that $[j_1, j_2]$ can be correctly chosen as $[j_1, j_2] = [1, \log_2(n)]$, this is unlikely in general. It is even possible that a data set be too short to contain scales in the scaling range, in which case the Logscale Diagram based estimators, or indeed any semi-parametric estimator of scaling, will be useless. On the other hand, in practice typically one cannot know the 'true' model for the data, and parametric estimators based on the wrong model can yield meaningless results. In contrast Logscale Diagram based estimation is not sensitive to non-scaling details of the data provided the scaling range is correctly identified, and are also more robust in other ways, as discussed below. Another key advantage is the ability to measure in a uniform framework both stationary and non-stationary forms of scaling.

A detailed comparison of the performance of $\hat{\alpha}$ against that of the discrete Whittle estimator for the FGN and Gaussian FARIMA(0,d,0) processes is given in [5]. In [75] $(\hat{\alpha}, \hat{c}_f)$ is compared against a joint discrete Whittle estimator for the Gaussian FARIMA(0,d,0) process, and $(\hat{\alpha}, \widehat{c_f C})$ against a joint maximum likelihood estimator for a 'pure' scaling process defined in the wavelet domain [79]. The main conclusion is that the Logscale Diagram based estimators offer almost unbiased estimates, even for data of small length, with far greater robustness, for the price of a small to moderate increase in variance compared to that of the parametric alternatives. They can also be used to treat data of arbitrary length both from the computational complexity and memory requirement points of view [5, 75, 62]. Comparisons against parametric estimators for processes which are further from such "ideal" processes will appear elsewhere.

1.3.3 The Multiscale Diagram and multifractals

1.3.3.1 The need for statistics other than second order. It is natural and straightforward to generalize the Logscale Diagram to the study of statistics other than second order, by replacing Eq. (1.19) with $\mu_j^{(q)} = 1/n_j \sum_k |d_X(j,k)|^q$, $q \in \mathbb{R}$. The resulting q^{th} *order Logscale Diagrams* are, naturally enough, of interest in situations where information relevant to the analysis of scaling is beyond the reach of second order statistics. Let us concentrate on two important examples: self-similarity and multifractality.

Self-similarity From the definition of self-similarity, the moments of $X(t)$ satisfy $\mathbb{E}|X(t)|^q = \mathbb{E}|X(1)|^q \cdot |t|^{qH}$, $\forall t$. As for the wavelet coefficients, it

follows from Eq. (1.10) that $\mathbb{E}|d_X(j,k)|^q = \mathbb{E}|d_X(0,k)|^q \cdot 2^{j(qH+q/2)}$, implying that $\mathbb{E}\mu_j^{(q)} = C_q 2^{j(\zeta(q)+q/2)}$, $\forall j$, with $\zeta(q) = qH$. This relation suggests that self-similarity can be detected by testing the linearity of $\zeta(q)$ with q.

Multifractal processes For the class of multifractal processes, assuming that $\int |T_{X(\omega)}(a,t)|^q dt \approx a^{\zeta(q)+q/2}$, $a \to 0$, can be related to the multifractal properties of the process, $\mu_j^{(q)}$ is expected to behave according to $\mu_j^{(q)} \approx 2^{j(\zeta(q)+q/2)}$ for small j. From these relations one can measure $\zeta(q)$ in practice and therefore estimate the Legendre multifractal spectrum. A particularly interesting question in the multifractal formalism is to test whether, in the range of q where $\int |T_{X(\omega)}(a,t)|^q dt$ is finite, $\zeta(q)$ takes a simple linear form: $\zeta(q) = qh$, or not. Clearly in such cases the Legendre spectrum is somewhat degenerate as it is entirely determined by h and the range of q where $\zeta(q)$ is defined. For instance, self-similar processes, for which $\mu_j^{(q)} \approx 2^{j(qH+q/2)}$ for *all* scales, satisfy $\zeta(q) = qH$ and are therefore fractal processes with $h = H$. More specifically the self-similar Lévy processes have infinite variance and are multifractal [41], yet their spectra are parameterised by H and are therefore derivative of the strong self-similar property. The FBM is another, even simpler example, often referred to as 'monofractal', confirming the intuition that a single scaling parameter controls all of its scaling properties.

1.3.3.2 The Multiscale Diagram and its use. In both the self-similar and multifractal cases therefore, accurately measuring the deviation of $\zeta(q)$ from a simple linear form is a crucial issue. To test this, and to investigate the form of $\zeta(q)$ in general, the q^{th} order scaling exponent $\alpha_q = \zeta(q) + q/2$ can be estimated in the q^{th} order Logscale Diagram for a variety of q values, and then the q dependence examined in the following tool:

Definition. The *Multiscale Diagram* (MD) consists in the graph of $\hat{\zeta}(q) = \hat{\alpha}_q - q/2$ against q, together with confidence intervals about the $\hat{\zeta}(q)$.

A lack of alignment in the Multiscale Diagram strongly suggests multifractal scaling which is not of the simple type where $\zeta(q) = qH$. Alignment in the Multiscale Diagram may indicate the presence of the simple type of multifractal scaling, or of self-similarity, or both simultaneously as in the self-similar Lévy processes above. It is important to note that the question of alignment, just as in the Logscale Diagram, is relative to confidence intervals, and that conclusions cannot be drawn in their absence. The confidence intervals shown are derived through approximations presented in [27].

To better examine the presence of alignment, it is also of interest to use an alternative *Linear* Multiscale Diagram (LMD) where $h_q = \alpha_q/q - 1/2$ is plotted against q. In the LMD the linear form is detected as a horizontal alignment, and the value of h obtained by estimating its level. Although the LMD is more convenient for visual inspection of alignment, the two forms of Multiscale Diagram are statistically equivalent in terms of the determination of the linearity or otherwise of $\zeta(q)$.

Fig. 1.5 illustrates the use of the Multiscale Diagrams by superimposing an analysis of a synthesized FBM and that of actual Internet data (delays of UDP packets as described in section 3.1.3, see Fig. 1.4). The regressions in the underlying Logscale Diagrams were performed for small scales: $[j_1, j_2] = (2, 6)$, for each of eleven different q values (as discussed in section 3.1.3, the lowest scale is problematic due to initialisation issues). The Multiscale Diagram in the left plot shows a clear alignment for the FBM series, consistent with its self-similar nature, and a clear absence of alignment for the Internet data, indicating that it is not monofractal and that a multifractal model may be appropriate. These conclusions are confirmed in the right hand plot where the same data is analysed in the Linear Multiscale Diagram. A clear horizontal alignment of the FBM data is observed while it is confirmed that the Internet data cannot obey $\zeta(q) = qH$.

The Internet data example in the figure is given to illustrate the use of the Multiscale Diagrams, although the presence of possible multifractal scaling behaviour in such traffic is of interest in its own right [76]. More generally however, multifractals have recently attracted considerable attention in traffic modelling because of their ability to flexibly model highly variable local behaviour in a natural way. Multifractal processes can also have positive increments which are approximately lognormally distributed, in agreement with a number of empirical traffic studies. For more details on these empirical findings the reader is referred to [61, 45] where such analyses were first performed or to [60, 76, 29, 30, 37, 70] for further investigations, and where connections to multiplicative cascade models are presented. Chapter **??** of this volume and references therein can also be consulted and Chapter **??** for details of lognormal distributions in traffic.

1.3.4 Departures from scaling

There are two broad categories of departures from models possessing a pure scaling feature. One is the presence of non-stationarities which are not in themselves of a scaling nature, and in the presence of which a scaling exponent continues to be well defined. Here the issue is one of the correct measurement of the constant underlying scaling parameter(s), despite the non-stationary "noise". The other category is the possibility that the scaling is itself changing in time, that is that the parameters of constant-in-time scaling are not well defined. A central issue in that case is the reliable detection of such a variation. In the following, special properties of the wavelet approach are exploited to yield partial, but nonetheless significant and useful, replies to these two different challenges.

1.3.4.1 Superimposed trends. In many situations of practical interest, the assumption that the data is fully described by a scaling model – be it self-similar, fractal or LRD – is much too restrictive to be realistic. This is

especially the case when an observed time series is in fact the result of an additive contamination of a scaling process $X(t)$ by some extraneous contribution $T(t)$, that is $Y(t) = X(t) + T(t)$. We will not consider here the case where $T(t)$ is random (this situation of a scaling process corrupted by some "observation noise" is considered, e.g., in [79]), but will comment on the case when $T(t)$ is deterministic and can be thought of as a "trend".

A simple model for a trend amounts to choosing $T(t)$ as a polynomial, say of order p. If no care is taken when analyzing a scaling process corrupted by such a trend, important features of interest (stationarity of the increments, for instance) are likely to be lost, thus impairing the estimation of the relevant scaling parameters. Power-law trends can even mimic LRD correlations when added to a stationary short-range process, leading to entirely erroneous conclusions ([5], [17], [65]). It is therefore desirable, prior to any analysis, to eliminate possible trends or, at least, to be able to evaluate and control their effects on the final estimates: wavelets offer a versatile and easy way of doing so.

In order to understand where the effectiveness of wavelets comes from in the context of trend removal, it is worthwhile to start from the admissibility condition, Eq. (1.1), satisfied by any wavelet ψ_0: saying that a wavelet is zero-mean is in fact equivalent to saying that it is orthogonal, and therefore "blind," to non-zero mean values. A natural generalization consists of considering wavelets with more than one vanishing moment, in the sense of Eq. (1.5), since assuming that the number of vanishing moments is N allows the analysis to be blind to polynomials of order up to $N-1$. In other words, the removal of a polynomial trend of order p is guaranteed by a wavelet with $N \geq p+1$. From a practical point of view, when p is the unknown order of a polynomial approximation of a trend, trend removal amounts to analyzing the data with different wavelets such that $N = 1, 2, \ldots$. Until the effective value $N = p+1$ is reached, the analysis is governed by the trend and gives N-dependent results whereas, as soon as $N \geq p+1$, stabilized results are obtained and reveal relevant features of the detrended data. Exact detrending is expected to occur in the case of polynomial trends, but it is worthwhile to remark that the procedure still remains effective in the case of non-polynomial trends, including many power-law trends, and oscillatory functions. In Fig. 1.6 an example is given of an FGN series contaminated with linear and sinusoidal trends. Although the Logscale Diagram with $N = 2$ is free from the effects of the linear trend (top right), higher values of N are required to effectively remove the sinusoidal trend (bottom right). Note that increasing N unfortunately also decreases the number of scales available for the analysis. The usefulness of an ability to effectively remove smooth non-polynomial trends is again illustrated in Fig. 1.7, where, using a value of N as low as 2, a change in mean level in Ethernet byte data is shown to not affect the estimation of the exponent α of LRD. (see [74], Fig. 9 for further discussion. The data set is derived from the 'pOct' Bellcore trace [43]).

The versatility of wavelets with respect to the freedom of choice of N makes

them an easy and efficient tool for trend removal. This contrasts with other methods, such as the Whittle estimator and parametric estimators in general, whose performance is heavily affected by trends (see [5], Fig. 2 and Table 1). An important advantage is that one can choose N to eliminate pre-selected trends, without having to know if they are actually present a priori, nor to jointly estimate their characteristics.

1.3.4.2 Time varying scaling exponents. Because of the high variability inherent in scaling processes, instances of scaling behaviour can be incorrectly judged as 'non-stationarity', in the broad sense of unstructured time variation, or conversely, variability due to non-stationarity may be erroneously taken to be scaling in nature. There is a strong need therefore to monitor the value of scaling parameters over time to distinguish between these possibilities, and to test if constancy of scaling can be concluded or not. A basic approach is to split a data set into a number m of adjacent blocks, and to examine the corresponding Logscale Diagrams. If alignment is found in each over roughly the same range of scales, then separate estimations $\hat{\alpha}^{(m)}$ of the scaling exponent could be performed for each block. A null hypothesis might be that the (unknown) values $\alpha^{(m)}$ share a common value. The difficulty is how to combine these estimates in a well defined statistical test, given that the problematic statistical nature of scaling processes would in general imply strong correlations between the $\hat{\alpha}^{(m)}$. Because of the quasi-decorrelation in the wavelet plane however, it can be shown [77] that the $\hat{\alpha}^{(m)}$ can be treated as almost independent, and moreover that they are approximately normally distributed with known variances. The difficult problem of the constancy of the scaling exponent can thereby be reduced to a simple model inference problem, to which an optimal (uniformly most powerful invariant) test exists [77] (see also the discussion in [5]).

1.3.5 Relations to other tools

1.3.5.1 Aggregation procedure Let $X(t)$ be a centred second-order stationary process with finite variance and form the discrete-time process $X^{(T)}(n), n \in \mathbb{Z}$, defined by

$$X^{(T)}(n) = \frac{1}{T} \int_{(n-1)T}^{nT} X(t)\, dt.$$

Such a quantity, referred to as the *aggregated* process of X with aggregation level T, turns out to play a key role with respect to self-similarity and LRD. In fact, if we assume that X is LRD in the sense that its autocovariance decreases asymptotically as in eq.(1.7), it can be shown [23] that, in the limit

of arbitrarily large aggregation levels T, the normalized autocovariance

$$r^{(T)}(k) = \frac{\mathbb{E}X^{(T)}(n)X^{(T)}(n+k)}{\text{Var } X^{(T)}(n)}, \quad k = 1, 2, \ldots$$

of the aggregated process tends to a well-defined limiting form which coincides with that of a FGN with unit variance (see Eq. (1.9)), while its variance satisfies

$$\text{Var } X^{(T)}(n) \approx T^{\alpha-1}, \quad T \to +\infty. \tag{1.31}$$

Aggregation appears therefore as a natural renormalization tool which reveals, in a well-defined manner, the possible occurrence of LRD in an observed process. By construction, aggregation amounts to averaging the analyzed process over adjacent blocks of larger and larger support. This is of course reminiscent of the way the Haar MRA, whose scaling function is simply the indicator function of $[0, 1]$, is built. More precisely, we have the exact equivalence [6]

$$X^{(2^j)}(k) = 2^{-j/2}\langle X, \phi_{j,k}^{(\text{Haar})}\rangle,$$

from which we can deduce that

$$\text{Var } a_X^{(\text{Haar})}(j, k) \approx 2^{j\alpha}, \quad j \to +\infty.$$

Revisiting the aggregation procedure from the Haar perspective allows for two levels of generalization [6]. We can first relax the choice of the Haar system and move towards any MRA. Second, we can replace approximation coefficients by details, since we also have, for any admissible wavelet ψ,

$$\text{Var } d_X(j, k) \approx 2^{j\alpha}, \quad j \to +\infty.$$

1.3.5.2 Allan variance Although well rooted in wavelet theory, this idea of using details rather than approximations can itself be viewed as a generalization of a method put forward by Allan in 1966 [13]. Specifically, whereas (1.31) shows the theoretical link between variance and the scaling behavior associated with LRD, Allan showed that the estimation of α is greatly improved when the standard variance estimate $(1/N)\sum_n (X^{(T)}(n))^2$ is replaced by the "Allan variance", defined as $(1/N)\sum_n (X^{(T)}(n+1) - X^{(T)}(n))^2$. In wavelet words, this just amounts — when dyadic intervals of the form $T = 2^j$ are considered — to computing Haar details [32]. As explained previously, the reason for the improved efficiency of such an approach is to be found in the ability of wavelets to almost decorrelate LRD processes. Moreover, revisiting the Allan variance in the light of the Haar system suggests that more versatile generalizations could be considered based on wavelets with more vanishing moments. This indeed not only guarantees smaller residual correlations but also an increased robustness of the analysis to polynomial trends.

1.3.5.3 Fano factor Apart from aggregation and the Allan variance, wavelets also offer a way of generalizing standard approaches in other areas, e.g., in the case of point processes [4]. Let us consider for instance processes of the form

$$P(t) = \sum_{k=-\infty}^{+\infty} G(t - t_k),$$

where the t_k are Poisson distributed, with a stationary (but possibly time-varying) density $\lambda(t)$, and $G(t)$ is the impulse response of some filter. With such a model, it turns out that scaling features or LRD may result from the joint properties of $\lambda(t)$ and $G(t)$. This is especially the case when $G(t) = \delta(t)$ and $\lambda(t)$ is FGN with non-zero mean. A classical way of revealing the scaling structure of such a "fractal point process" amounts to looking at how it departs from an ordinary Poisson process when observed over larger and larger scales. With the ordinary Poisson process, the associated counting process $N(T)$ is such that $\mathrm{Var}\ N(T) = \mathbb{E}N(T)$, no matter how large T is, thus guaranteeing the "Fano factor," defined as

$$F(T) = \frac{\mathrm{Var}\ N(T)}{\mathbb{E}N(T)}$$

to be constant. This situation contrasts with the fractal case for which we have

$$F(T) \approx T^{2H-1}, \quad T \to +\infty. \tag{1.32}$$

We use H here rather than α as the process $N(t)$, which has stationary increments, is not LRD but rather asymptotically self-similar.

By definition, the Fano factor is a ratio between fluctuations and averages at a given observation scale. As such, it naturally admits a wavelet-based alternative definition :

$$WF(j) = 2^{j/2} \frac{\mathrm{Var}\ d_P(j,k)}{\mathbb{E}a_P(j,k)},$$

in which details (resp. approximations) play the role of fluctuations (resp. averages). Given this definition, the scaling relation (1.32) now reads

$$WF(j) \approx 2^{j(2H-1)}, \quad j \to +\infty,$$

and moreover we have the explicit equivalence $WF^{(\mathrm{Haar})}(j) = F^{(\mathrm{Allan})}(2^j)$. Again, this wavelet revisiting of the Fano factor allows for versatile generalizations with increased performance, in which the Haar wavelet is to be replaced by a wavelet with a larger number of vanishing moments.

From the above examples, it is worthwhile stressing the fact that wavelets provide the user with a unified framework which i) possesses increased versatility and performance as compared to the earlier methods it generalizes, and ii) is equally applicable to continuous processes and point processes.

1.4 WAVELETS AND SCALING: SYNTHESIS

After having allocated the greater part of this paper to the wavelet *analysis* of scaling processes, we now turn to the question of their synthesis. To do so, we focus here on a wavelet-based synthesis of the FBM. This may seem restrictive, however unfortunately generation is more difficult than analysis and the area is less well developed. Although it is not difficult to generate time series with approximate scaling properties using wavelets, there are very few examples where a specific process can be accurately reproduced. We discuss this further at the end of this section.

There exist a number of methods dedicated to the synthesis of the FBM which are not based on wavelets. Among these, the so-called Choleski or Durbin-Levinson methods [19, 39, 68] are notable as they are exact. They have severe computational drawbacks however which make them unsuitable to the generation of long traces, particularly the enormous series required in tele-traffic simulations. Various approximate synthesis methods with reasonable computational loads are known, for example the so-called spectral synthesis method (for a recent variant, see [55]), the On-Off superposition method [78], a method based on the discretization of the integral representation, and recently a refined random midpoint displacement method [53] with has a number of advantages. These methods yield tractable practical implementations (sometimes on-the-fly implementations), but often suffer from the drawback that the errors (due to approximations made) cannot be controlled and that it is not clear which property of the FBM has been lost. Wavelet-based methods have also already been proposed [79, 80] to produce stationary $1/f$-type processes. We present here a wavelet-based synthesis of the FBM, introduced in [64], which is developed and implemented in [7, 47]. The method relies on an exact wavelet expansion of the FBM. This allows the generation of approximate sample paths, but with controlled errors, in a practical implementation framework using a fast pyramidal algorithm similar to that underlying the inversion of the DWT [46, 24]. This method reproduces accurately the key features that characterize the FBM: stationarity of the increments, self-similarity, long range dependence, and allows new and clearer insights into them, as described in [7, 47]. We detail below the key steps of this synthesis.

1.4.1 Wavelet expansions for the FBM

When $0 < H < 1$, and $H \neq 1/2$, FBM has an integral representation [50, 67],

$$B_H(t) = \frac{\sigma}{C}\left\{\int_{-\infty}^{0}\left[(t-x)^{H-1/2} - (-x)^{H-1/2}\right]dB(x) - \int_{0}^{t}(t-x)^{H-1/2}dB(x)\right\} \tag{1.33}$$

From this integral representation, one can see FBM as a fractionally integrated version of a Gaussian white noise (denoted dB). The starting point of the wavelet-based representation of FBM lies in the fact that, in an orthonormal wavelet basis expansion of a Gaussian white noise process (in the sense of distributions), the weights of the expansion are simply Gaussian i.i.d. random variables. To obtain the wavelet expansion of the FBM, one can fractionally integrate the wavelet Gaussian white noise expansion, which basically amounts to fractionally integrating the wavelets. One may therefore conjecture that it can be represented as

$$B_H(t) \stackrel{(?)}{=} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} 2^{jH} \psi_H(2^{-j}t - k)\epsilon_{j,k} \tag{1.34}$$

where ψ_H is a suitable wavelet and the $\epsilon_{j,k}$ are i.i.d. $N(0, \sigma_0^2)$ random variables. One can check that the RHS of (1.34) scales correctly at dyadic points, namely $B_H(at) \stackrel{d}{=} a^H B_H(t)$ for $a = 2^{-\ell}$, $\ell = -1, 0, 1, \ldots$, as can be easily verified by making the change of variables $j \to j' = j + \ell$ and using the fact that the $\epsilon_{j,k}$'s are i.i.d. However, the RHS of (1.34), as written, does not converge because of the growth of 2^{jH} as $j \to \infty$, that is, because of its behaviour at low frequencies. The representation (1.34) is almost correct however. It is just necessary to add an "infrared correction," namely to write

$$B_H(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} 2^{jH} [\psi_H(2^{-j}(t-k)) - \psi_H(-k)]\epsilon_{j,k}. \tag{1.35}$$

It is shown in [47] that, for a suitable function ψ_H (see below), the RHS of (1.35) converges to fractional Brownian motion. Not only does the scaling by $a = 2^{-\ell}$ hold, but in fact (1.35) scales (in the sense of the finite-dimensional distributions) for any $a > 0$. Observe that the infrared correction also ensures that the resulting process has stationary increments and equals 0 at $t = 0$. We are, in fact, representing the increment $B_H(t) - B_H(0)$, which equals $B_H(t)$ because $B_H(0) = 0$, a similar idea is behind the integral representation (1.33).

Whereas the representation (1.33) involves *integration* with respect to "continuous" white noise dB, the representation (1.35) involves a summation over discrete white noise (i.i.d. $\epsilon_{j,k}$), and as such, is very close to the intuition of a so-called "Karhunen-Loève representation" of the process. It is based on translations (by k) and dilations (by j) of the function ψ_H. A value $k > 0$ corresponds to translation to the right, $k < 0$ to the left, $j > 0$ to dilation and $j < 0$ to compression. Thus the limit $j \to -\infty$ captures the high frequencies and the limit $j \to \infty$ captures the low frequencies. Note that paper [47] use $j' = -j$ instead of j. The function ψ_H is continuous, and in fact, the RHS of (1.35) converges to $B_H(t)$ uniformly on every closed bounded interval $t \in [-T, T]$.

There exist different wavelet representations which differ (a) by the way the

low frequencies are represented, and (b) by the choice of wavelets. Instead of using ψ_H for both the low and high frequencies, it is possible to generate the low frequencies by using a different function. As shown in [47], one also has

$$\begin{aligned} B_H(t,\omega) &= \sum_{k=-\infty}^{\infty} \widetilde{\phi}_H(t-k) B_H(k,\omega') \\ &\quad + \sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} 2^{jH} \psi_H(2^{-j}t-k) \epsilon_{j,k}(\omega'') - \widetilde{b}_0(\omega) \quad (1.36) \end{aligned}$$

for some suitably chosen function $\widetilde{\phi}_H$. The function ψ_H is as before and the random variable $\widetilde{b}_0(\omega)$ is a random level shift, independent of t, necessary to ensure that $B_H(0,\omega) = 0$. We write here $B_H(t,\omega)$ instead of $B_H(t)$ not only to underline the fact that B_H is random (ω is an element of the probability space), but also because $\{B_H(k,\omega'),\ k \in \mathbb{Z}\}$ and $\{\epsilon_{j,k}(\omega''),\ j \le 0,\ k \in \mathbb{Z}\}$ are assumed independent and $\omega = (\omega', \omega'')$. Thus, to generate *continuous* time $\{B_H(t),\ t \ge 0\}$, one generates first a *discrete* sequence $B_H(k,\omega')$ in addition to the i.i.d. $\epsilon_{j,k}$. It is the sum $\sum_{k=-\infty}^{\infty} \widetilde{\phi}_H(t-k) B_H(k,\omega')$ that generates the low frequency components and hence the corresponding long-range dependence. Observe that $B_H(k,\omega)$ is not equal to $B_H(k,\omega')$ because $B_H(k,\omega)$ depends also on the $\epsilon_{j,k}$. Here again, convergence to $B_H(t)$ holds uniformly for any $t \in [-T,T]$.

The sequence $\{B_H(k,\omega'),\ k \in \mathbb{Z}\}$ is a discrete time series whose increments $\{\Delta B_H(k) = B_H(k+1,\omega') - B_H(k,\omega'),\ k \in \mathbb{Z}\}$ are stationary and have long-range dependence. It is possible to replace $\{B_H(k,\omega'),\ k \in \mathbb{Z}\}$ by a sum of other stationary time series with long-range dependence, in particular, by a Gaussian FARIMA$(0,d,0)$ time series with $d = H - 1/2$ [40].

Let then $\{Z_i^{(H)},\ i \in \mathbb{Z}\}$ represent this mean zero FARIMA time series independent of the $\epsilon_{j,k}$s and let

$$S_0^{(H)} = 0,\ S_{-k}^{(H)} = -\sum_{i=-k+1}^{0} Z_i^{(H)},\ S_k^{(H)} = \sum_{i=1}^{k} Z_i^{(H)} \text{ for } k \ge 1,$$

denote the partial sums. Then fractional Brownian motion can also be represented as

$$B_H(t) = \sum_{k=-\infty}^{\infty} \phi_H(t-k) S_k^{(H)} + \sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} 2^{jH} \psi_H(2^{-j}t-k) \epsilon_{j,k} - b_0 \quad (1.37)$$

for some suitably chosen function ϕ_H. The function ψ_H is as before and the

random variable b_0 is such that $B_H(0) = 0$, that is

$$b_0 = \sum_{k=-\infty}^{\infty} \phi_H(-k) S_k^{(H)} + \sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} 2^{jH} \psi_H(-k) \epsilon_{j,k}.$$

Once again, convergence holds uniformly for $t \in [-T, T]$.

Eq. (1.37) is a corrected version of a representation in [64]. It describes the FBM as a trend $\sum_{k=-\infty}^{\infty} \phi_H(t-k) S_k^{(H)}$ over which are superimposed (high frequency) details $\sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} 2^{jH} \psi_H(2^{-j}t - k) \epsilon_{j,k}$. The non stationarity property of the FBM is already apparent in the expression of the trend which involves $S^{(H)}$. The self-similarity is carried in the actual variance of the wavelet coefficients of the expansion $d_{B_H}(j,k) = 2^{j(H+1/2)} \epsilon_{j,k}$, which follows a power-law of the scale 2^j. Moreover, ψ_H is tailored in such a way that the wavelet basis it generates exactly catches the correlation of the FBM so that the wavelet coefficients are strictly uncorrelated. This specific wavelet basis therefore acts as a Karhunen-Loève basis for the high frequencies of the process.

1.4.2 Wavelet Design

It remains to indicate what are the functions ψ_H, ϕ_H and $\widetilde{\phi}_H$. They are derived from a scaling function $\phi_0(t)$ and a corresponding wavelet function $\psi_0(t)$, yielding an orthonormal wavelet basis. Using the Fourier transform notation $U(\nu) = \int_{-\infty}^{\infty} u(t) e^{-it\nu} dt$, start with the Fourier transforms Ψ_0 and Φ_0 of ψ_0 and ϕ_0 and let

$$\Psi_{0,1/2-H}(\nu) = (i\nu)^{1/2-H} \Psi_0(\nu), \quad \Phi_{0,1/2-H}(\nu) = \left(\frac{i\nu}{1 - e^{-i\nu}} \right)^{1/2-H} \Phi_0(\nu),$$

Going back to the time domain, get the functions ψ_H and ϕ_H by integration:

$$\psi_H(t) = \int_{-\infty}^{t} \psi_{0,1/2-H}(u) du = \psi_{0,-(H+1/2)}(t)$$

and

$$\phi_H(t) = \int_{t-1}^{t} \phi_{0,1/2-H}(u) du = \phi_{0,-(H+1/2)}(t).$$

The expression for $\widetilde{\Phi}_H$ is more involved and we refer the reader to [47].

There are many possible choices for the pair (ψ_0, ϕ_0). Because the spectral density [50, 67] of the increments of FBM may diverge at the frequency $\nu = 0$, it is necessary that the Fourier transform $\Psi_0(\nu)$ of the wavelet ψ_0 tend to 0 as $\nu \to 0$ relatively fast. In [47], the Meyer (or Lemarié-Meyer) wavelets are used. Their Fourier transforms vanish not just at the origin but in a whole neighborhood of the origin. The Meyer scaling and wavelet functions $\phi_0(t)$

and $\psi_0(t)$ as well as $\phi_H(t)$ and $\psi_H(t)$ are very smooth: they are infinitely differentiable and, while they do not have bounded support, they tend to 0 as $t \to \pm\infty$ faster than any polynomial.

Instead of the Meyer wavelets, one can use, to obtain ϕ_H and ψ_H, any orthonormal wavelet basis with enough vanishing moments, such as the Daubechies wavelets with $N \geq H + 1/2$, that is $N \geq 2$. One can also use the so-called Lemarié-Battle orthonormal spline wavelets [24].

1.4.3 Fast implementation

It is known from the wavelet literature [24] that the synthesis in Eq. (1.37), that we rewrite for convenience

$$\begin{aligned} B_H(t) + b_0 &= \sum_{k=-\infty}^{\infty} S_k^{(H)} \phi_H(t-k) \\ &\quad + \sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} \left(2^{j(H+1/2)} \epsilon_{j,k}\right) 2^{-j/2} \psi_H(2^{-j}t - k), \end{aligned}$$

can be implemented using a discrete-time fast pyramidal algorithm on condition that $\phi_H(t)$ and $\psi_H(t)$ are designed from a multiresolution analysis (see section 3). The implementation we present here is based on a slightly modified version of the one developed in [7].

The coefficients of the discrete time filters h_2 and g_2 involved in the synthesis depend on ϕ_H and ψ_H through

$$\begin{aligned} h_2 &= u_H \\ g_2 &= v_H \end{aligned} \tag{1.38}$$

where u_H and v_H are the generating sequences of ϕ_H and ψ_H respectively:

$$\begin{aligned} \phi_H(t/2) &= \sqrt{2} \textstyle\sum_k u_H(k) \phi_H(t-k) \\ \psi_H(t/2) &= \sqrt{2} \textstyle\sum_k v_H(k) \phi_H(t-k). \end{aligned} \tag{1.39}$$

This means that we do not necessarily need to provide explicit expressions for these functions. Let us denote by $\phi_0(t)$ and $\psi_0(t)$ a scaling function and a mother wavelet giving birth to an orthonormal wavelet basis with sufficient regularity and by u and v their generating sequences:

$$\begin{aligned} \phi_0(t/2) &= \sqrt{2} \textstyle\sum_k u(k) \phi_0(t-k) \\ \psi_0(t/2) &= \sqrt{2} \textstyle\sum_k v(k) \phi_0(t-k). \end{aligned} \tag{1.40}$$

We have shown in [7] that u_H and v_H can be obtained from u and v by:

$$\left.\begin{aligned} u_H &= f^{(s)} * u\,, \quad F^{(s)}(z) &= 2^{-s}(1+z^{-1})^s \\ v_H &= g^{(s)} * v\,, \quad G^{(s)}(z) &= 2^{-s}(1-z^{-1})^{-s} \end{aligned}\right\} \tag{1.41}$$

where the $*$ denotes the discrete time convolution, $s = H + 1/2$, $f^{(s)}$ and $g^{(s)}$ are infinite length sequences whose $z-$transform are labeled $F^{(s)}(z)$ and $G^{(s)}(z)$. From a practical point of view, the discrete time convolution above are computed using an approximation technique described in [7].

By using the pyramidal algorithm sketched in Fig. 1.8, we end up with an approximation of fractional Brownian motion which is computationally efficient and conceptually simple. The "trend" part involves the cumulated sum of a FARIMA$(0, d, 0)$ time series, which can be obtained from i.i.d. Gaussian random variables, with zero mean and variance σ_0^2. The "details" (successive high frequencies) are labeled here by the indices $j = 0, -1, \ldots$. We include them up to some level $-J$. Observe that the details at levels $j = 0, \ldots, -J$ involve independent Gaussian random variables $\eta_{j,k} = 2^{j(H+1/2)}\epsilon_{j,k}$, $j \leq 0, k \in \mathbb{Z}\}$, i.i.d. in k, with zero mean and variances that follows a power law in j: Var $2^{j(H+1/2)}\epsilon_{j,k} = \sigma_0^2 2^{j(2H+1)}$.

1.4.4 Synthesis of other scaling processes

Athough the FBM is an important scaling process, it is of course very specific, and unsuitable for many modeling purposes. It is of great interest to be able to rapidly and accurately synthesize other scaling processes that are not strictly FBM but which may have looser constraints (e.g., $1/f$ type processes) or richer scaling properties (e.g., multifractal processes), or more flexible non-scaling features together with scaling features (e.g., LRD processes with flexible short range dependent structure). Despite these broader needs, the highly focused understanding of the wavelet synthesis described above remains of interest, because it shows where to modify the synthesis scheme either to give up, or to maintain, certain properties of the FBM. For instance in [80] a wavelet based synthesis is presented which is in the same spirit as that given here, but which produces *only* $1/f$ type processes. In fact there exists in the tele-traffic literature a variety of wavelet based syntheses of scaling processes, which typically amount to producing $1/f$ type processes, though in a loose and poorly controlled manner. However, wavelets can also be used to implement multiplicative cascades, which are themselves capable of generating processes with rich scaling behaviour. This has been exploited in recent work in [60] for the synthesis of multifractal processes, and in [15] for the synthesis of a large variety of multiplicative cascades processes.

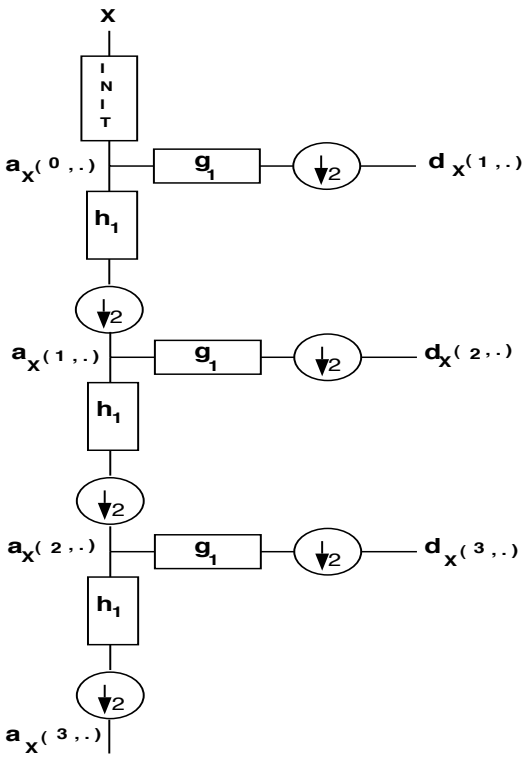

Fig. 1.1 Fast filter-bank based pyramidal algorithm. The DWT can be computed using a fast pyramidal algorithm, i.e., given that we have approximation $a_X(j-1,k)$ at level $j-1$, we obtain approximation $a_X(j,k)$ and detail $d_X(j,k)$ at level j by convolving with h_1 and g_1, respectively, and decimating. The coefficients of the filters h_1 and g_1 are derived form the chosen scaling function and wavelet ϕ_0 and ψ_0. The downarrow stands for a decimation by a factor of 2 operation: one drops the odd coefficients. An initialization step is required to go from the process X to the approximation of order 0: $a_X(0,k)$.

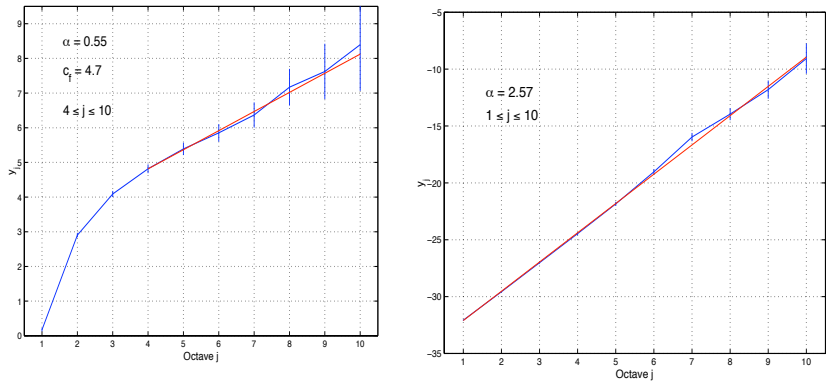

Fig. 1.2 Logscale Diagrams. Left: An example of the y_j against j plot and regression line for a LRD process with strong SRD. The vertical bars at each octave give 95% confidence intervals for the y_j. The series is simulated FARIMA(0,d,2) with $d = 0.25$ and second order moving average operator $\Psi(B) = 1 + 2B + B^2$, implying $(\alpha, c_f) = (0.50, 6.38)$. Alignment is observed over scales $[j_1, j_2] = (4, 10)$, and a weighted regression over this range allows an accurate estimation despite the strong SRD: $\hat{\alpha} = 0.53 \pm 0.07$, $\hat{c}_f = 6.0$ with $4.5 < \hat{c}_f < 7.8$. The scaling can be identified as LRD as the value is in the correct range, $\hat{\alpha} \in (0, 1)$, and the alignment region includes the largest scales in the data. **Right**: Alignment is observed over the full range of scales with $\hat{\alpha} = 2.57$, corresponding to $\hat{H} = 0.79$, consistent with the self-similarity of the simulated FBM ($H = 0.8$) series analysed.

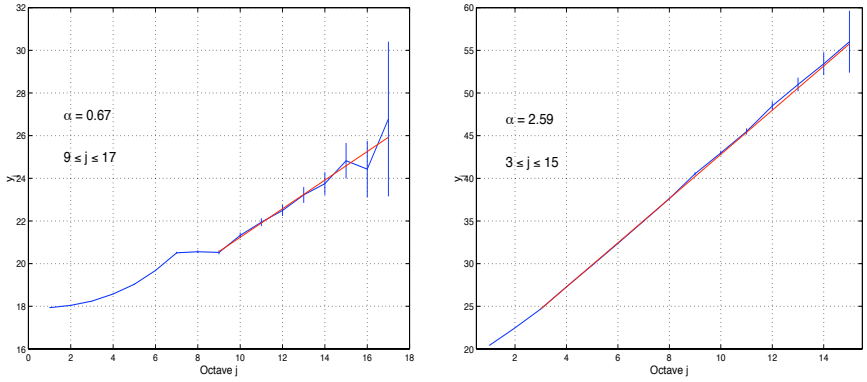

Fig. 1.3 LRD and H-sssi behaviour in Ethernet traffic data. Logscale Diagrams for **Left:** the discrete series of successive interarrival times, showing a range of alignment and an α estimate consistent with LRD, and **Right:** the cumulative work process (bytes up to time t), consistent with an asymptotically self-similar (close to exactly self-similar) process with stationary increments.

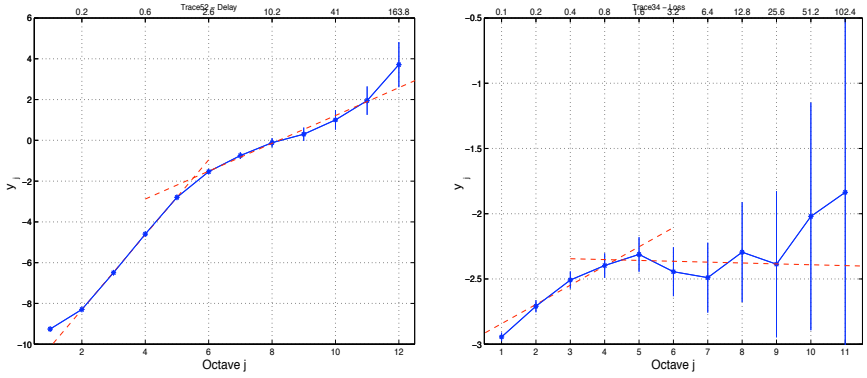

Fig. 1.4 Logscale Diagrams with biscaling. Internet UDP packet data displaying two scaling regimes, examples of *biscaling* **Left:** Delay series: Regime I on the left (small j's) is related to continuous but non-differentiable sample paths ($h \in (0,1)$) and regime II (large j's) to LRD. **Right:** Loss series (1 for lost packets, else 0). Regime I corresponds to discontinuous sample paths ($h < 0$). In regime II there is trivial white noise scaling ($\alpha = 0$) indicating stationary short-range dependent behaviour.

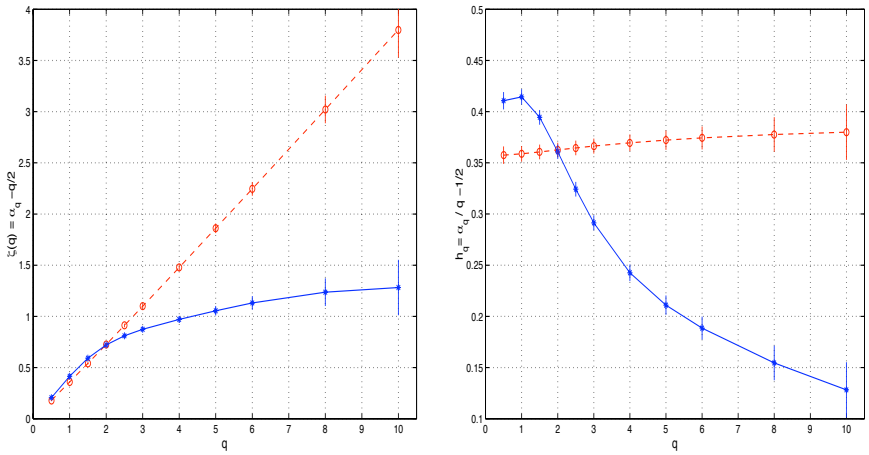

Fig. 1.5 Testing for Multifractality. Each plot superposes results for a synthesized FBM (dashed line) and actual Internet data (solid line). **Left** Multiscale Diagram: Alignment for the FBM series is consistent with its known self-similar nature, and no alignment for the Internet data suggests non-degenerate multifractal behaviour**. Right:** Linear MD: Horizontal alignment for the FBM series suggests even more clearly that $\zeta(q) = qH$, whereas it is even more clearly not the case for the Internet data.

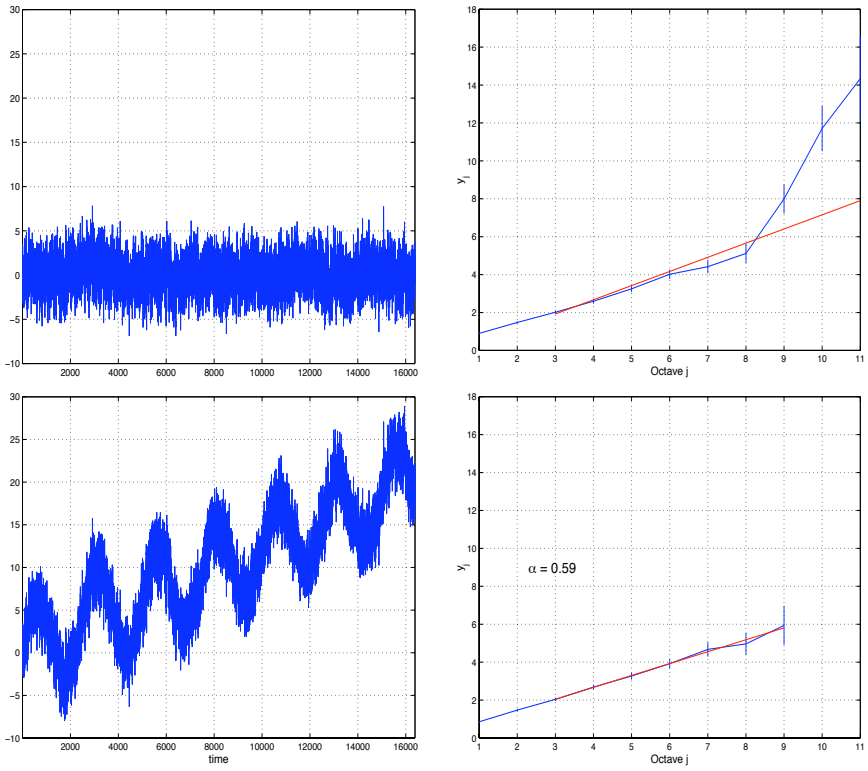

Fig. 1.6 Robustness with respect to trends. Left: Synthesized FGN with $H = 0.80$ (top), and with superimposed sinusoidal and linear trends (bottom). **Right:** Logscale Diagrams of the data *with trends* using a Daubechies2 (ie $N = 2$) wavelet (top), and Daubechies9 wavelet (bottom). It is seen that increasing N allows the effects of superimposed trends to be removed from the Logscale Diagram, enabling uncontaminated estimates of the scaling exponent. Here, with $N = 9$, $\hat{\alpha} = 0.59$ and therefore $\hat{H} = 0.795$.

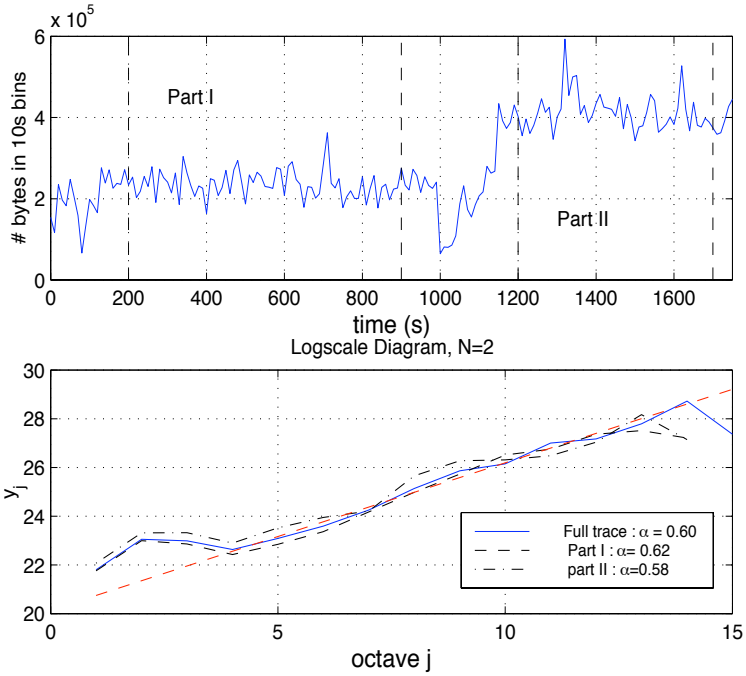

Fig. 1.7 Eliminating mean level shifts. Upper: Ethernet byte data aggregated over 10s intervals. A marked, but smooth, level shift seems to occur around 1050 seconds. **Lower:** Logscale Diagram of the byte data aggregated over 10ms intervals. With $N = 2$, estimates to the left and the right of the shift, and over the whole series, are consistent, showing that smooth level shifts can be effectively eliminated in practice.

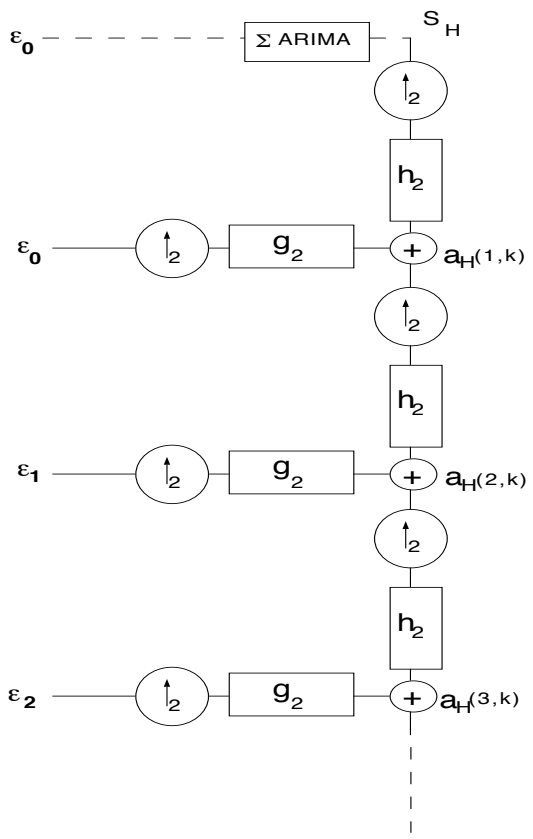

Fig. 1.8 Wavelet-based synthesis for FBM. The FBM can be numerically synthetized using a fast pyramidal algorithm based on discrete time filters. The coefficients of the filters h_2 and g_2 depend on functions ϕ_H and ψ_H obtained by the fractional integration of order $H + 1/2$ of an orthonormal wavelet basis. The inputs of the filters, denoted $\{\epsilon_j\}$ in the figure, stand for independent random vectors $\{\eta_{j,k}, k \in \mathbb{Z}\}$, labeled by j, whose components are i.i.d. Gaussian random variables with zero-mean and variance $\sigma_0^2 2^{j(2H+1)}$. The uparrow operator indicates an upsample by a factor of two operation obtained by inserting a zero between each sample. One gets, for example, $a_H(2,k)$ by adding the upsampled $a_H(1,k)$ convolved with h_2 to the upsampled ϵ_1 convolved with g_2.

References

1. P. Abry, *Ondelettes et Turbulences - Multirésolutions, Algorithmes de Décompositions, Invariance d'Echelle et Signaux de Pression*. Diderot, Editeur des Sciences et des Arts, Paris, 1997.

2. P. Abry, P. Gonçalvès and P. Flandrin, Wavelet-based spectral analysis of $1/f$ processes, Proc. IEEE–ICASSP'93, pp. III.237–III.240, Minneapolis (MN), 1993.

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

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

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

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

7. P. Abry, F. Sellan, The wavelet based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and implementation, *Applied and Computational Harmonic Analysis*, 3:377–383, 1996.

8. P. Abry, L. Delbeke, P. Flandrin Wavelet-based estimator for the self-similarity parameter of α-stable processes, to be published in *ICASSP-99*, Phoenix, USA, March 1999.

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

10. P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, unpublished note.

11. A. Aldroubi and M. Unser. Sampling procedure in function spaces and asymptotic equivalence with shannon's sampling theory. *Numer. Funct. Anal. and Optimiz.*, 15:1–21, 1994.

12. J. Andren, M. Hilding, and D. Veitch, Understanding end-to-end Internet traffic dynamics, to appear in Proc. Globecom '98, Sydney, 1998.

13. W. Allan, Statistics of atomic frequency standards, *Proc. IEEE*, 54:221–230, 1966.

14. A. Arnéodo, J.F. Muzy, S.G. Roux, Experimental analysis of self-similar random cascade processes: application to fully developped turbulence, J. Phys. II France, 7:363–370, 1997.

15. A. Arnéodo, E. Bacry, J.F. Muzy, Random cascades on wavelet dyadic trees, Journal of mathematical Physics, 39(8):4142–4164, August, 1998.

16. E. Bacry, J.F. Muzy, and A. Arnéodo, Singularity spectrum of fractal signals from wavelet analysis: Exact results, *J. Stat. Phys.*, 70:635–674, 1994.

17. J.Beran, *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.

18. J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, Long-range dependence in variable-bit-rate video traffic, *IEEE Trans. on Comm.*, 43:1566–1579, 1995.

19. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition, 1991.

20. S. Cambanis, C. Houdré, On the continuous wavelet transform of second-order random processes, *IEEE Trans. on Info. Theory*, 41(3): 628–642, 1995.

21. B. Castaing, Y. Gagne, M. Marchand, Log-similarity for turbulent flows ?, *Physica D*, 68:387–400, (1993).

22. P. Chainais, P. Abry, J.F. Pinton, Intermittency and coherent structures in a swirling flow: a wavelet analysis of joint pressure and velocity measurements. preprint, submitted to *Phys. Rev.*, 1999.

23. D.R. Cox, Long-range dependence : a review, in H.A. David and H.T. David, eds., *Statistics : an Appraisal*, pp. 55–74. Iowa State Univ. Press, Ames (IA), 1984.

24. I.Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia (PA), 1992.

25. L. Delbeke, *Wavelet based estimators for the scaling index of a self-similar with stationary increments*, PhD Thesis, KU Leuven, Belgium, 1998.

26. L. Delbeke, P. Abry, Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion, submitted to *Stochastic Processes and their Applications*, preprint, 1997.

27. L. Delbeke, P. Abry, Wavelet based estimators for the self-similarity parameter of the fractional Brownian motion, submitted to *Applied and Computational Harmonic Analysis*, preprint, 1998.

28. K. Falconer, *Fractal Geometry – Mathematical Foundations and Applications*. J. Wiley and Sons, New York, 1990.

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

30. A. Feldmann, A.C. Gilbert, W. Willinger, T.G. Kurtz, Looking behind and beyond self-similarity: On scaling phenomena in measured WAN traffic. Preprint.

31. P. Flandrin, On the spectrum of fractional Brownian motions, *IEEE Trans. on Info. Theory*, 35:197-199, 1989.

32. P. Flandrin, Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. on Info. Theory*, 38:910-917, 1992.

33. P. Flandrin, Fractional Brownian motion and wavelets, in M. Farge, J.C.R. Hunt and J.C. Vassilicos, eds., *Wavelets, Fractals, and Fourier Transforms*, pp. 109–122. Clarendon Press, Oxford, 1993.

34. P. Flandrin, *Temps-Fréquence*. Hermès, Paris, 1993.

35. P. Flandrin, P. Gonçalvès, in L.L. Schumaker and G. Webb, eds., *Recent Advances in Wavelet Analysis*, pp. 309–334. Academic Press, San Diego, 1994.

36. P. Gonçalvès, P. Flandrin, Scaling exponents estimation from time-scale energy distributions, Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-92, pp.V.157–V.160, San Francisco (CA), 1992.

37. A.C. Gilbert, W. Willinger, A. Feldmann, Scaling analysis of random cascades, with applications to network traffic, to be published in *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications, April, 1999.

38. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*. Academic Press, New York, corrected and enlarged edition, 1980.

39. M. A. Hauser, W. Hörmann, R. M. Kunst, and J. Lenneis. A note on generation estimation and prediction of stationary processes. In R. Dutter and W. Grossmann, editors, *COMPSTAT*, pages 323–329. Physica Verlag, 1994.

40. J. R. M. Hosking, Fractional differencing, *Biometrika*, 68(1):165–176, 1981.

41. S. Jaffard, Sur la nature multifractale des processus de Lévy, C. R. Acad. Sci. paris, Série 1, t. 323:1059–1064, 1996.

42. S. Jaffard, Multifractal formalism for function, Part II: self-similar functions, *SIAM J. Math. Anal.*, 28(4):971-988, 1997.

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

44. J. Lévy-Véhel, Fractal Approcahes in Signal Processing, *Fractal geometry and Analysis*, The Mandelbrot Festschrift, Curacao, 1995, Evertsz, Peitgen, Voss, eds., World Sientific, 1996.

45. J.L. Lévy Véhel, R. Riedi Fractional Brownian Motion and data traffic modelling: The other end of the spectrum, Fractals in Engeneering, J.L. Lévy Véhel, E. Lutton, C. Tricot, eds., pp., 185–203, Springer, London, 1997.

46. S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, Boston, 1997.

47. Y. Meyer, F. Sellan, and M. S. Taqqu, Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion, preprint, 1998.

48. S.M. Kay, Fundamentals of Statistical Signal Processing. Prentice-Hall, Englewoods Cliffs (NJ), 1993.

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

50. B.B. Mandelbrot, J.W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev, 10:422-437, 1968.

51. B.B. Mandelbrot, Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, Journal of fluid mechanics, 1974.

52. J.F. Muzy, E. Bacry, and A. Arnéodo, The multifractal formalism revisited with wavelets, *Int. J. of Bifurc. and Chaos*, 4(2):245–301, 1994.

53. I. Norros, P. Mannersalo, and J.L. Wang Simulation of fractional Brownian motion with conditionalized random midpoint displacement. Preprint Oct 1998.

54. A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 2nd Edition, 1984.

55. V. Paxson. Fast approximation of self-similar network traffic. Preprint, 1995.

56. R. Peltier, J. Lévy-Véhel, Multifractional Brownian motion: definition and preliminary results, submitted to *Stochastic Processes and their Applications*, preprint, 1997.

57. B. Pesquet-Popescu, Statistical properties of the wavelet decomposition of some non Gaussian self-similar processes, Invited paper, *Signal Processing*, to appear, 1999.

58. B. Pesquet-Popescu,, P. Abry, Wavelet based estimators for the self-similarity parameter of α-stable processes, in preparation, 1999.

59. R.H. Riedi, An improved multifractal formalism and self-similar measures, *J. Math. Anal. Appl., Vol.189*, pp.462–490, 1995.

60. R. Riedi, M.S. Crouse, V.J. Ribeiro, R.G. Baraniuk, A Multifractal Wavelet Model with Application to Network Traffic, to be published in *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications, April, 1999.

61. R. Riedi, J.L. Lévy Véhel, TCP traffic is multifractal: a numerical study, Preprint, submitted to *IEEE Trans. on Networking.*

62. M. Roughan, D. Veitch, and P. Abry, On-line estimation of LRD parameters, to appear in Proc. Globecom '98, Sydney, 1998.

63. Roux S., Muzy J.F., Arnéodo A. Detecting vorticity filaments using wavelet analysis : about the statistical contribution of vorticity filaments to intermittency in swirling turbulent flows. to be published in *European Phys. Journal* (1998).

64. F. Sellan, Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes, *C.R.A.S. Série I*, 321:351–358, 1995.

65. V. Teverovsky, M. S. Taqqu, Testing for long-range dependence in the presence of shifting means or a slowly declining trend using a variance-type estimator, *Journal of Time Series Analysis*, 18:279–304, 1997.

66. V. Teverovsky, M. S. Taqqu, and W. Willinger, A critical look at Lo's modified R/S statistics, to appear in the *Journal of Statistical Planning and Inference*, preprint, 1997.

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

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

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

70. M. S. Taqqu, V. Teverovsky,and W. Willinger, Is network traffic self-similar or multifractal? *Fractals*, 5:63–74, 1997.

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

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

73. A.H. Tewfik, M. Kim, Correlation structure of the discrete wavelet coefficients of fractional Brownian motion, *IEEE Trans. Info. Theory*, 38:904–909, 1992.

74. D. Veitch, P. Abry, Estimation conjointe en ondelettes des paramètres du phénomène de dépendance longue, Proc. 16ième Colloque GRETSI, pp.1451–1454, Grenoble, France, 1997.

75. D. Veitch, P. Abry, A wavelet-based joint estimator of the parameters of long-range dependence, *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications, to appear, April 1999.

76. D. Veitch, P. Abry, J. Bolot, Wavelet Tools for the Analysis of Scaling Phenomena in Traffic, preprint, 1999.

77. D. Veitch, P. Abry, A statistical test for the constancy of scaling exponents, preprint, 1998.

78. W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level, *IEEE/ACM Transactions in Networking*, 5(1):71–86, 1997. Extended Version of the paper with the same title that appeared in *Computer Communications Review*, 25:100-113, 1995.

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

80. G.W. Wornell, *Signal Processing with Fractals : a Wavelet-Based Approach. Prentice-Hall, 1995.*