

Wavelets, spectrum analysis and $1/f$ processes

*Patrice Abry*¹, *Paulo Gonçalves*^{1,2} and *Patrick Flandrin*¹

¹ Ecole Normale Supérieure de Lyon,
Laboratoire de Physique (URA 1325 CNRS),
46 allée d'Italie, 69364 Lyon Cedex 07, France

² Rice University,
Department of ECE,
P.O. Box 1892, Houston, TX 77005-1892, USA

Abstract

The purpose of this paper is to evidence why wavelet-based estimators are naturally matched to the spectrum analysis of $1/f$ processes. It is shown how the revisiting of classical spectral estimators from a time-frequency perspective allows to define different wavelet-based generalizations which are proved to be statistically and computationally efficient. Discretization issues (in time and scale) are discussed in some detail, theoretical claims are supported by numerical experiments and the importance of the proposed approach in turbulence studies is underlined.

Published in "Lecture Notes in Statistics", 103.

Wavelets and Statistics.

Anestis Antoniadis, Georges Oppenheim (Editors)

Springer-Verlag, 1995,

pp. 15–30.

1 On $1/f$ processes

1.1 Power-law spectra and self-similarity

Signals with power-law spectra — or “ $1/f$ processes,” i.e., stochastic signals $X(t)$ such that their power spectrum density $\Gamma_X(\nu)$ is proportional to $|\nu|^{-\alpha}$ over some decades — are ubiquitous in fields such as physics, biology, engineering or economics, to name but a few [K2]. In turbulence for instance [B1], the spectrum of the velocity field is known to obey a power-law decay over a wide range of frequencies (the so-called inertial range) with an exponent $\alpha \sim 5/3$. In this case, a very accurate measurement of the spectral exponent is highly desirable, since its value is of a key importance for discriminating between competing theories. Generally speaking, $1/f$ behaviors at low frequencies are associated with slowly-decaying correlations, and the interest for $1/f$ processes has also been recently renewed by the development of chaos phenomenology, according to which an apparently stochastic behavior can in fact result from a deterministic mechanism with long-range dependent characteristics [MC].

From another perspective, power-law spectra indicate that a signal exists at all scales and, hence, has no characteristic scale. This results in *self-similar features*, according to the following definition:

Definition 1 *A process $\{X(t), t \in \mathbb{R}\}$ is said to be (statistically) self-similar of parameter H if and only if, for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$, where $\stackrel{d}{=}$ denotes equality of finite-dimensional distributions.*

Self-similarity (more precisely, self-affinity) indicates that the graph $(t, X(t))$ remains statistically unchanged when the time axis and the amplitude are simultaneously scaled by a factor a and a^{-H} , respectively [F1]. Ordinary Brownian motion is such an example, with sample paths evidencing self-similar features of a random fractal, with $H = \frac{1}{2}$.

1.2 The fractional Brownian motion (fBm) model

A useful and common model of self-similar process is given by the so-called fractional Brownian motion (in short, fBm), defined as follows [MV]:

Definition 2 *Fractional Brownian motion of (Hurst) exponent $0 < H < 1$ is the zero-mean Gaussian process $\{B_H(t), t \in \mathbb{R}\}$ such that³:*

$$\begin{aligned} (i) \quad & B_H(0) = 0 \\ (ii) \quad & B_H(t + \delta) - B_H(t) \stackrel{d}{=} \mathcal{N}(0, \sigma |\delta|^H) \end{aligned}$$

From this definition, we get $B_H(t) \stackrel{d}{=} \mathcal{N}(0, \sigma |t|^H)$ and, more specifically,

$$\mathbb{E} B_H(t) B_H(s) = \frac{\sigma^2}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}). \quad (1)$$

This evidences both the nonstationarity of fBm and the fact that it generalizes ordinary Brownian motion, obtained as a special case for $H = \frac{1}{2}$. Moreover, it can be easily checked that

$$\mathbb{E} B_H(at) B_H(as) = \mathbb{E} (a^H B_H(t)) (a^H B_H(s)),$$

property which, together with Gaussianness, guarantees the self-similarity of fBm. A by-product of this self-similarity is that sample paths of fBm are fractal curves, with Hausdorff dimension $\dim_H = 2 - H$ [F1].

Although nonstationary, fBm does have stationary increments. The increment process of fBm is referred to as fractional Gaussian noise (in short, fGn) and is defined according to the following definition [MV]:

Definition 3 *Fractional Gaussian noise of (Hurst) exponent $0 < H < 1$ is the zero-mean Gaussian process $\{G_{H,\delta}(t), (t, \delta) \in \mathbb{R} \times \mathbb{R}_+\}$ defined by:*

$$G_{H,\delta}(t) = \frac{1}{\delta} (B_H(t + \delta) - B_H(t)). \quad (2)$$

By definition, this is a stationary process, since $G_{H,\delta}(t) \stackrel{d}{=} \mathcal{N}(0, \sigma \delta^{H-1})$. Moreover, for large enough lags, i.e., when $|\tau| \gg \delta$, we get

$$\mathbb{E} G_{H,\delta}(t + \tau) G_{H,\delta}(t) \sim \sigma^2 H(2H - 1) |\tau|^{2H-2},$$

thus corresponding to a 1/f power spectrum density at low frequencies, namely:

$$\Gamma_{B_H,\delta}(\nu) \sim |\nu|^{1-2H}, \quad 0 < |\nu| \ll \delta^{-1}.$$

1.3 Wavelets and fBm

Definition 4 *The continuous wavelet transform of a signal $X(t) \in L^2(\mathbb{R})$ is the two-dimensional function of time $t \in \mathbb{R}$ and scale $a \in \mathbb{R}_+^*$*

$$T_X(t, a) = a^{-1/2} \int_{-\infty}^{\infty} X(s) \overline{\psi\left(\frac{s-t}{a}\right)} ds, \quad (3)$$

where the (possibly complex-valued) analyzing wavelet $\psi(\cdot)$ has a Fourier transform $\Psi(\cdot)$ such that $\Psi(0) = 0$.

Definition 5 *The discrete wavelet transform of a signal $X(t) \in L^2(\mathbb{R})$ is the two-dimensional function of time index $n \in \mathbb{Z}$ and scale index $j \in \mathbb{Z}$*

$$d_X[n, j] = 2^{-j/2} \int_{-\infty}^{\infty} X(t) \psi(2^{-j}t - n) dt, \quad (4)$$

where the analyzing wavelet $\psi(\cdot) \in \mathbb{R}$ has a Fourier transform $\Psi(\cdot)$ such that $\Psi(0) = 0$.

³ $\mathcal{N}(m, \sigma)$ denotes a normal law with mean m and variance σ^2 .

Wavelets, both continuous and discrete (for definitions and properties, see e.g., [D]) have been extensively used for the analysis of fBm and fGn [AG, F2, F3, F5, M1, RZ, TK, W2, W3, WO], since they match the structure of such processes for at least two reasons:

1. First, whereas fBm is nonstationary, its wavelet transform (which is a random field over $\mathbb{R} \times \mathbb{R}_+^*$ in the continuous case and over \mathbb{Z}^2 in the dyadic case) is a stationary function of time at each scale. We get for instance [F2] from eqs. (1) and (3)

$$\mathbb{E} T_{B_H}(t, a) \overline{T_{B_H}(s, a)} = -\frac{1}{2} \sigma^2 a^{2H+1} \int_{-\infty}^{\infty} |u|^{2H} \gamma_\psi \left(\frac{t-s}{a} - u \right) du,$$

where $\gamma_\psi(\cdot)$ stands for the deterministic correlation function of the analyzing wavelet (i.e., the inverse Fourier transform of $|\Psi(\cdot)|^2$). Taking the Fourier transform on the $t-s$ variable of the above equation, it follows [F4] that each (band-pass) filtered process $T_{B_H}(\cdot, a)$ has for power spectrum

$$\Gamma_{T_{B_H}}(\nu) \propto \sigma^2 |\nu|^{-(2H+1)} a |\Psi(a\nu)|^2.$$

Interpreting this equation as the input-output relationship of the “wavelet filter,” this allows to attach to (nonstationary) fBm a pseudo-spectrum according to

$$\Gamma_{T_{B_H}}(\nu) \propto \sigma^2 |\nu|^{-(2H+1)}. \quad (5)$$

2. A second reason is that although fBm is long-range dependent, its wavelet coefficients are almost uncorrelated, and hence almost independent for a large enough spacing in the time-scale plane. In the dyadic case, it can be shown more precisely that the correlation $\mathbb{E} d_X[n, j] d_X[m, k]$ decays as $O(|2^j n - 2^k m|^{2(H-R)})$, where R is the number of vanishing moments of the analyzing wavelet [TK]. Therefore, the higher R , the lower the correlation between wavelet coefficients.

2 On spectrum analysis

2.1 The classical Welch estimator

One of the most classical nonparametric spectral estimators for the power spectrum density $\Gamma_X(\nu)$ of a wide-sense stationary process $\{X(t), 0 \leq t \leq T\}$ was introduced by P.D. Welch [W1]. It consists in averaging short-time periodograms and can be defined as

$$\hat{\Gamma}_{1,X}(\nu) \equiv \frac{1}{N} \sum_{n=1}^N \left| \int_0^T X(t) \overline{w_\theta(t - \tau_n)} e^{-i2\pi\nu t} dt \right|^2, \quad (6)$$

where $w_\theta(t)$ is some arbitrary (unit energy) weighting function of equivalent duration θ and where the time instants $\{\tau_n, n = 1, \dots, N\}$ are chosen such that the different sub-series $\{X_n(t) \equiv X(\tau_n - \theta/2 \leq t \leq \tau_n + \theta/2)\}$ are almost uncorrelated. This is a biased estimator, since

$$\mathbb{E} \hat{\Gamma}_{1,X}(\nu) = \int_{-\infty}^{\infty} \Gamma_X(f) |W_\theta(\nu - f)|^2 df,$$

where $W_\theta(\nu)$ is the Fourier transform of $w_\theta(t)$. Bias being convolutive, the frequency resolution is fixed by the spectral width of $W_\theta(\nu)$, which is of the order of $1/\theta$. The variance of the estimator is approximatively (for Gaussian processes)

$$\text{var} \hat{\Gamma}_{1,X}(\nu) \simeq \frac{1}{N} |\Gamma_X(\nu)|^2,$$

thus imposing to take large N 's for a sake of variance reduction. However, in the case where the time support of the observation is finite, increasing N necessarily decreases the length of each of the sub-series, with a corresponding increase in bias as a by-product: this is the well-known bias-variance trade-off [K1].

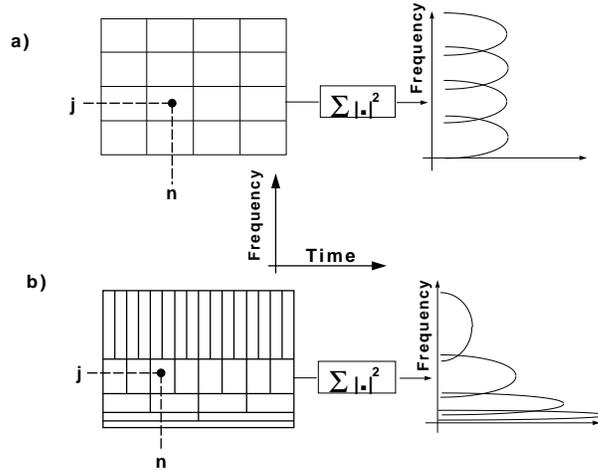


Figure 1: **Tiling of the time-frequency plane** a) the Welch estimator is the time marginal of a particular bilinear time-frequency distribution called the spectrogram (squared modulus of a Gabor transform); b) we propose another spectral estimator defined as the time marginal of a particular time-scale distribution called the scalogram (squared modulus of the wavelet transform). Because the way energy is spread over the time-frequency plane is changed, the properties of this new spectral estimator are different from those of the classical one.

2.2 A time-frequency perspective

Let us have a closer look at the way this estimator is constructed. Given a finite observation interval $0 \leq t \leq T$, the signal $X(t)$ is first chopped in time over N intervals of width θ , and each of the resulting sub-series is then Fourier transformed, with an imposed frequency resolution of the order of $1/\theta$. From a time-frequency perspective, this corresponds to a tiling of the plane by means of almost uncorrelated cells — each of unit area — centered on equispaced nodes $\{(\tau_n \sim n\theta, \nu_j \sim j/\theta), (n, j) \in \mathbb{Z}^2\}$ of a rectangular lattice [AG], see Fig. 1. The energetic contribution to each node can be measured by $|\langle X, g_{\tau_n \nu_j} \rangle|^2$, with $g_{\tau_n \nu_j}(t) \equiv w_\theta(t - \tau_n) \exp(i2\pi\nu_j t)$ and those different contributions are finally averaged over time:

$$\hat{I}_{1,X}(\nu_j) \equiv \frac{1}{N} \sum_{n=1}^N |\langle X, g_{\tau_n \nu_j} \rangle|^2. \quad (7)$$

Such a rewriting of the Welch estimator enables us to interpret it as the time marginal of a transform that spreads the energy of the signal over a two dimensional — time and frequency — space. This particular bilinear distribution is the so-called spectrogram, which is the squared modulus of the short-time Fourier transform (or equivalently, a Gabor-like transform).

2.3 A wavelet generalization

Using the time-frequency interpretation of Welch estimator, it becomes very easy to modify its definition — e.g., by changing the way energy is spread over the time-frequency plane — while preserving its overall structure in terms of time averaging of spectral features.

Considering for instance the wavelet transform as a bank of constant- Q band-pass filters, all derived by dilations from a prototype filter of central frequency ν_ψ and bandwidth $\Delta\nu_\psi$, changing the scale parameter a amounts to explore the frequency axis with the approximate correspondence $\nu(a) = \nu_\psi/a$ and the local frequency resolution $\Delta\nu(a) = \Delta\nu_\psi/a = \nu_\psi/aQ$, where Q is the quality factor of the wavelet filter. For dyadic scales $a = 2^j, j \in \mathbb{Z}$, the corresponding tiling of the plane inherits of a dyadic structure according to which contributions are located on nodes $\{(\tau_n \sim 2^{-j}n\theta, \nu_j \sim 2^j/\theta), (n, j) \in \mathbb{Z}^2\}$ (see Fig. 1), with a number of independent cells N_j which is scale — and, hence, frequency — dependent.

This naturally leads to the definition [AG] of a time-scale based spectral estimator which can be written as the time marginal of the scalogram (squared modulus of the wavelet transform)

$$\hat{T}_{2,X}(\nu_j) \equiv \frac{1}{N_j} \sum_{n=1}^{N_j} d_X^2[n, j] \equiv \frac{1}{N} \sum_{n=1}^N |\langle X, \psi_{nj} \rangle|^2, \quad (8)$$

with $\psi_{nj} = 2^{-j/2} \psi(2^{-j}t - n)$.

This is again a biased spectral estimator, since

$$\mathbb{E} \hat{T}_{2,X}(\nu_j) = \int_{-\infty}^{\infty} \Gamma_X(f) (\nu_\psi/\nu_j) |\Psi(f \nu_\psi/\nu_j)|^2 df,$$

the main difference with the previous — time-frequency based — estimator being that bias is frequency-dependent in a constant- Q fashion. Note that the above spectral estimator could have been defined using other various versions of the wavelet transform, to which we will come in the last section. Yet, the main properties of this time-scale based spectral estimator — adequation to $1/f$ processes — presented in the next section hold for any variations and are therefore given in details with the discrete transform.

3 Estimating the spectral exponent of $1/f$ processes

3.1 Bias: why using a wavelet-based estimator?

Let us assume that the observation $X(t)$ corresponds to a $1/f$ process whose power spectrum is of the form $\Gamma_X(\nu) = \sigma^2 |\nu|^{-\alpha}$, and let us address the problem of estimating the spectral exponent α .

From a model point of view, it is clear that we have $\log \Gamma_X(\nu) = \log \sigma^2 - \alpha \log |\nu|$, thus suggesting to get the value of α from a slope measurement in a log-log plot of the power spectrum. From an estimation point of view however, such a procedure can only be applied to some estimate of the power spectrum, with no guarantee that a similar linear relation still holds.

Considering first the time-frequency based spectral estimator $\hat{T}_{1,X}(\nu)$, we get, when applied to a $1/f$ process,

$$\mathbb{E} \hat{T}_{1,X}(\nu) = \sigma^2 |\nu|^{-\alpha} \left(\int_{-\infty}^{\infty} |1 + f/\nu|^{-\alpha} |W_\theta(f)|^2 df \right). \quad (9)$$

This evidences that such an estimation of the power spectrum is affected by a bias which is multiplicative and furthermore frequency-dependent. The consequence is that, when rewritten in log-log coordinates, eq. (9) does not reduce to a linear function, inducing therefore a bias in the estimation of α when using a linear regression in a log-log plot (cf. Fig. 2).

On the contrary, if we consider the time-scale based spectral estimator $\hat{T}_{2,X}(\nu)$, we get, when applied to the same $1/f$ process,

$$\mathbb{E} \hat{T}_{2,X}(\nu) = \sigma^2 |\nu|^{-\alpha} \left(\int_{-\infty}^{\infty} |f/\nu_\psi|^{-\alpha} |\Psi(f)|^2 df \right). \quad (10)$$

This shows that, while the estimation of the power spectrum is still affected by a multiplicative bias, this bias is no more frequency-dependent, thus making of the linear regression in a log-log plot an effective way to get an unbiased estimate of α . Let us remark that, for the integral (10) to exist, the number R of vanishing moments of the analyzing wavelet $\psi(\cdot)$ must be such that $R \geq (\alpha - 1)/2$.

In order to illustrate the behavior of both estimators with respect to bias, we can make use of the oversimplification of a rectangular shape for the spectrum of either the short-time window (time-frequency case) and the wavelet (time-scale case). The result is that

$$\frac{\mathbb{E} \hat{T}_X(\nu)}{\Gamma_X(\nu)} = \frac{Q(\nu) \Delta\nu_0}{1 - \alpha} \cdot \left[\left(1 + \frac{1}{2Q(\nu)} \right)^{1-\alpha} - \left(1 - \frac{1}{2Q(\nu)} \right)^{1-\alpha} \right],$$

where $\Delta\nu_0$ and $Q(\nu)$ are respectively the frequency resolution and the quality factor of the mother window or wavelet. In the time-frequency case, $Q(\nu)$ increases with ν , thus inducing a decrease in bias. As a result, the linear regression in a log-log plot leads systematically to an overestimation of the slope α (cf. Fig. 2), in clear contrast with the time-scale case for which $Q(\nu)$ is constant.

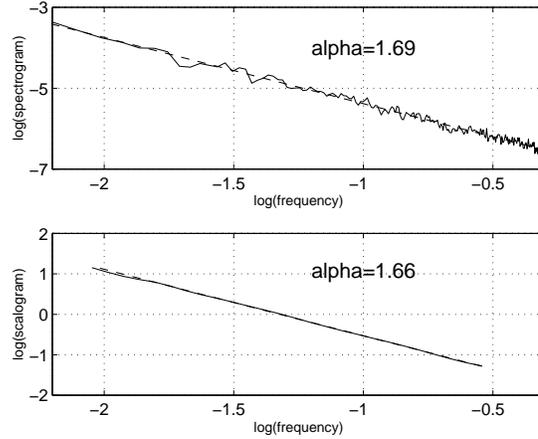


Figure 2: **Bias in spectrum estimation of $1/f^\alpha$ -processes — Time-frequency versus time-scale.** Given a fBm with $H = 1/3$ (and, hence, a spectral exponent $\alpha = 5/3$), the figure illustrates the theoretical result according to which a time-frequency based estimator of the parameter α (top) is biased, whereas a time-scale based estimator (bottom) is not. (For the comparison to be fair, the time-frequency based estimation has been performed with the most favorable window, i.e. the one corresponding to the largest scale factor in the time-scale based analysis.)

3.2 Efficiency

According to the fact that $\hat{I}_{2,X}(\nu_j)$ provides us with an unbiased estimation of the long-dependence parameter α , we can introduce the quantity

$$\delta_j \equiv \sum_{n=1}^{N_j} \left(\frac{d_X[n, j]}{\sqrt{N_j} \sigma_j} \right)^2,$$

with $\sigma_j^2 \equiv \mathbb{E} d_X^2[n, j]$. This quantity consists in a sum of squared Gaussian variables which are all zero-mean. For wavelets with a high enough number of vanishing moments, they are furthermore uncorrelated — as recalled previously in Section 1.3 —, with a variance $1/N_j$. The result is that δ_j is approximatively distributed according to a chi-squared law with N_j degrees of freedom. In analogy with the classical spectrum estimation problem, we can therefore consider the (base 2) logarithm of δ_j — a quantity which could be referred to as a normalized log-scalogram —: $\eta_j \equiv \log_2 \delta_j$, whose probability density function is given by

$$p(\eta_j) = \frac{(N_j/2)^{(N_j/2) \ln 2}}{\Gamma(N_j/2)} \exp \left\{ (N_j/2) (\eta_j \ln 2 - e^{\eta_j \ln 2}) \right\}.$$

Given an observation of N_0 data points, we have $N_j = 2^{-j} N_0$ and, asymptotically — i.e., when $N_0 \rightarrow \infty$ —, we get from Stirling's formula that

$$\eta_j \xrightarrow{d} \mathcal{N}(0, S_j),$$

with $S_j^2 \equiv 2^{j+1}/(N_0 \ln^2 2)$. Estimation of α can therefore be achieved via least-squares fit techniques applied to the pair $(j, \log_2(\eta_j))$.

The simplest (unweighted) procedure consists in computing:

$$\hat{\alpha}_1(J) \equiv \frac{\sum_{j=1}^J j \eta_j - \sum_{j=1}^J j \sum_{j=1}^J \eta_j}{\sum_{j=1}^J j^2 - (\sum_{j=1}^J j)^2},$$

J being the number of scales involved in the linear regression, with the constraint that N_j be kept large enough for the previous asymptotic assumptions to hold. The variance of this — unbiased and asymptotically normally distributed — estimator can be evaluated in closed form, leading to

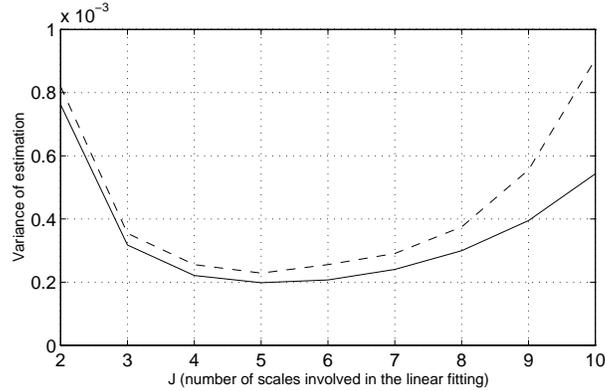


Figure 3: **Wavelet-based spectrum estimation of $1/f^\alpha$ processes – Variance of the unweighted estimator.** The time-scale based estimation of parameter α shows a minimum of variance for a small number of scales ($J = 5$). The reduction of variance one could expect from an increase of the number of scales is balanced by the exponential growth of the own variance of the data on larger scales. We obtain an excellent agreement between the asymptotic closed form (solid line) and numerical simulations (dashed line, the simulations are performed on about 3000 repeated trials of $1/f^\alpha$ processes, with $\alpha = 5/3$ and $N_0 = 2^{15}$). The quantitative difference that occurs at larger scales is caused by a too small number of points N_j involved in numerical simulations.

$$\mathbb{E} \hat{\alpha}_1^2(J) = \frac{144}{N_0 \ln^2 2} \frac{2^J(J^2 - 6J + 17) - (J^2 + 6J + 17)}{J^2(J^2 - 1)^2}.$$

A simulation experiment supporting this theoretical expression is given in Fig. 3. It also illustrates a typical behavior of the estimator according to which an optimum number of scales to be used exists, beyond which performance of the estimation is degraded. This is simply due to the fact that, whereas increasing the number of scales J allows to base the linear regression on more variables η_j , the variability in the estimation of these variables is increased — because less and less samples are available at coarser scales —, so that their use turns out to be a penalization.

Taking into account the increase of the variance S_j^2 as a function of scale, an improved (weighted) procedure consists in computing:

$$\hat{\alpha}_2(J) = \frac{\sum_{j=1}^J S_j^{-2} \sum_{j=1}^J j \eta_j S_j^{-2} - \sum_{j=1}^J j S_j^{-2} \sum_{j=1}^J \eta_j S_j^{-2}}{\sum_{j=1}^J S_j^{-2} \sum_{j=1}^J j^2 S_j^{-2} - (\sum_{j=1}^J S_j^{-2} j)^2}.$$

This corresponds in fact to a maximum likelihood estimate, whose variance is now given by

$$\mathbb{E} \hat{\alpha}_2^2(J) = \frac{1}{N_0 \ln^2 2} \frac{1 - 2^{-J}}{1 - 2^{-(J+1)}(J^2 + 4) + 2^{-2J}}.$$

This result — which is supported by the simulation experiment reported in Fig. 4 — indicates that, in contrast with the previous case, variance is a strictly decreasing function of J , with an asymptotic limit given by

$$\lim_{J \rightarrow \infty} \mathbb{E} \hat{\alpha}_2^2(J) = \frac{1}{N_0 \ln^2 2}.$$

It is worthwhile to remark that only few scales ($J \sim 6$) are sufficient for this limit to be almost attained. In any case (i.e., even for finite J 's), and under the assumptions which have been made, the variance of $\hat{\alpha}_2(J)$ is the best achievable, since it can be shown that it is equal to the Cramér-Rao lower bound [WO, LO].

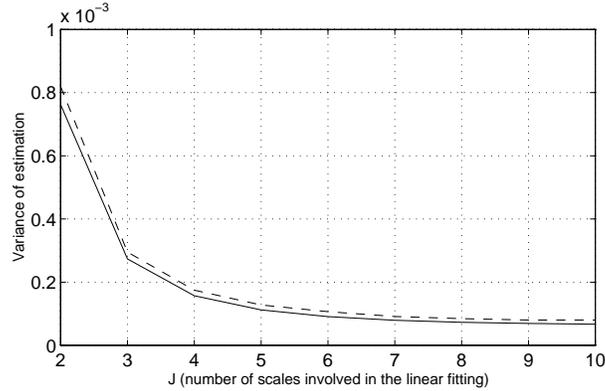


Figure 4: **Wavelet-based spectrum estimation of $1/f^\alpha$ processes – Variance of the weighted estimator.** Both the asymptotic closed form (solid line) and numerical simulations (dashed line) show that there is no use to involve a large number of scales to obtain the minimum variance estimation.

3.3 Number of vanishing moments of the wavelet: an illustration of the bias-variance trade-off.

The number of vanishing moments R of the analysing wavelet needs to be high enough to balance the divergence of $f^{-\alpha}$ at the null frequency and therefore to insure the convergence of the multiplicative integral in eq. (10). From a practical point of view, if the inequality $R \geq (\alpha - 1)/2$ is not satisfied, this results in a systematic underestimation of α . Moreover, practically, a relevant estimation is achieved only if R is chosen slightly above $(\alpha - 1)/2$.

From the variance point of view, one theoretically expects less and less correlation between wavelet coefficients and therefore less variance from the increase of R . For finite length data, yet, the number of points polluted by border effects also increases with R , therefore limiting the effective number of points involved in the summation and the reduction of variance. This may even lead to an increase of variance with increasing R 's.

We, therefore, have a practical trade-off in the choice of R : once the inequality $R \geq (\alpha - 1)/2$ is comfortably satisfied, there is no advantage (and it can even turn to be a drawback) to increase R .

4 Further remarks on the use of various wavelet transforms

More on discrete wavelets: semi- and bi-orthogonal transforms. Most of the time, one restricts the label of discrete wavelet transform to the use of orthonormal wavelets decomposition, yet, in the present framework, there is no need to use orthogonal wavelets: it works as well with semi- or bi-orthogonal transforms.

Let us recall that [JS], whereas orthogonal transforms make use of only one family of waveforms $\{\psi_{nj}(t) = 2^{-j/2}\psi(2^{-j}t - n), (n, j) \in \mathbb{Z}^2\}$, semi- and bi-orthogonal transforms use simultaneously one extra family $\{\check{\psi}_{nj}(t) = 2^{-j/2}\check{\psi}(2^{-j}t - n), (n, j) \in \mathbb{Z}^2\}$, both families being said to be dual, according to the orthogonality property

$$\langle \psi_{nj}, \check{\psi}_{mk} \rangle = \delta_{nm} \delta_{jk}, \forall n, m, j, k.$$

In the semi-orthogonal case, orthogonality between the wavelet subspaces W_j (or equivalently, between wavelets with different j 's) is preserved and therefore $\langle X, \psi_{nj} \rangle$ still stands for the orthogonal projection of X onto W_j . This contrasts with the biorthogonal case, for which the projection onto some multiresolution space is performed along the direction of one of its duals and is therefore no longer orthogonal.

Moreover, we can propose a modified version of the spectral estimator (8), given by:

$$\hat{T}_{3,X}(\nu_j) = \frac{1}{N_j} \sum_{n=1}^{N_j} d_X[n, j] \check{d}_X[n, j], \quad (11)$$

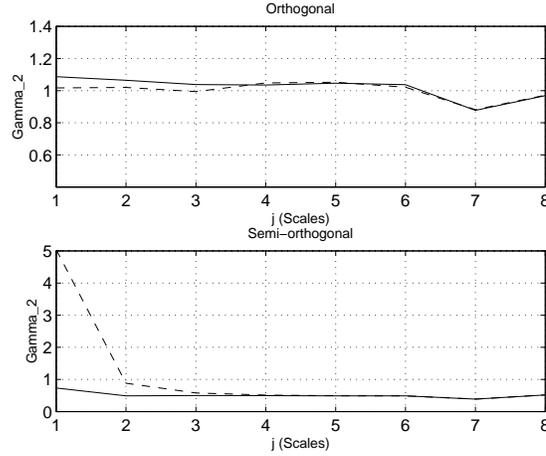


Figure 5: **Initialization of the fast pyramidal algorithm and spectral estimation** Example of a white noise wavelet-based spectral estimation, with an orthogonal (*Daubechies6*) decomposition (top), with a semi-orthogonal (*cubic spline*) decomposition (bottom), when the initialization of the fast algorithm is performed (solid line) and when it is not (dashed line). With the orthogonal transform, the whiteness of the data is preserved and therefore the initialization is not necessary. On the contrary, in the semi-orthogonal case, if no initialization is made, the data are given a *color*, that of the analysing family of atoms.

where $\hat{d}_X[n, j] = \langle X, \hat{\psi}_{nj} \rangle$. Because we get

$$\mathbb{E} \hat{T}_{3,X}(\nu_j) = \int_{-\infty}^{\infty} \Gamma_X(f) 2^j \Psi(2^j f) \overline{\hat{\Psi}(2^j f)} df,$$

this estimator still provides an unbiased estimate of α .

On the implementation of the spectral estimator: initialization of the fast pyramidal algorithm. In any case, orthogonal, semi-orthogonal or bi-orthogonal, the discrete transform and therefore the associated spectral estimator can be implemented using the celebrated fast pyramidal algorithm [M2]. A correct use of this algorithm requires to perform an initialization step [AF] that computes the starting sequence $a_X[n, 0] = \langle X, \varphi_{n0} \rangle$ from the samples $X[k]$, $k \in \mathbb{Z}$ of the continuous-time signal $X(t)$, $t \in \mathbb{R}$ (by definition, one has $\varphi_{nj}(t) = 2^{-j/2} \varphi(2^{-j}t - n)$, $(n, j) \in \mathbb{Z}^2$, where $\varphi(t)$ is the scaling function from which the mother wavelet $\psi(t)$ is derived). If this initialization step can be omitted (i.e., if we can assume that $a_X[n, 0] = X[n]$) without important consequences when using an orthogonal decomposition to perform spectral analysis, it turns out to be crucially necessary with the two other types of transforms. Fig. 5 presents spectral estimations of a white noise with (solid line) and without (dashed line) initialization from an orthogonal (top) and a semi-orthogonal (bottom) transform. In this latter case, when no initialization is performed, the data are given a *color* which, in fact, is that of the correlation of the analysing family of atoms.

Moreover, as a side remark, one can mention the following result concerning the modified version of the spectral estimator (11). Because $\psi(t)$ and $\hat{\psi}(t)$ are dual functions, their spectral behaviors are balancing each other so that a correct spectral estimation is achieved, even in the absence of initialization.

On discrete times. In signal analysis, a common objection to the use of the discrete wavelet transform concerns its lack of time-shift invariance, due to the dyadic sampling. One can, of course, enlarge the previous definitions of the time-scale based spectral estimators to versions which, while keeping the dyadic sampling in scale, would be continuous in time (or, at least, sampled as fast as the analyzed signal at any scale):

$$T_X(t, a = 2^j) = 2^{-j/2} \int_{-\infty}^{\infty} X(s) \psi(2^{-j}(s-t)) ds.$$

All properties and results on bias for $1/f$ processes obviously still hold in this case. For a sampled signal consisting in N_0 data points, one could expect that the constant number of points at each scale (N_0 instead of $N_j = 2^{-j}N_0$, with the discrete transform) would provide some extra reduction of variance, but this actually does not occur because the inserted samples $T_X[n, a = 2^j]$ for $n \neq 2^j k$ remain strongly correlated. This qualitative statement is conformed by numerical simulations: the use of a continuous time variable provides us, on the spectrum itself, with a reduction of variance that hardly reaches a factor of 2 for $j = 8$ (instead of the expected factor $2^j!$) but does not supply any reduction of variance on the estimation of α . Hence, the use of a quasi-continuous time variable — which furthermore prevents from using the fast pyramidal algorithm — does not enable any substantial improvement in the statistical properties of the estimator.

On discrete scales. The discrete wavelet-based spectral estimator is certainly attractive because of its satisfying statistical properties as well as of its extremely low computational cost, but there are some instances for which its dyadic sampling in scale constitutes a severe limitation. This is, for instance, especially true in turbulence, a domain where much is still to be learned from systematic spectral estimations over large amounts of data. It remains, indeed, an important matter to decide whether the spectral exponent α of Kolmogorov is exactly equal to $5/3$ or slightly above, but first experiments made on real data tend to cool this enthusiasm. In laboratory experiments for instance, the inertial range (i.e., the limited range of frequencies in which the power-law model holds) exists over 1 or 2 decades at most — that is 4 or 5 octaves — with upper and lower bounds which are not precisely predicted from theory. Therefore, the fluctuations on the estimation of α which are due to the choice of the octaves involved in the linear fitting are larger than what is required to discriminate between the competing theories.

To avoid this severe drawback of only one single estimation by octave, one has to use a transform that is quasi-continuous in the scale variable. One could, of course, imagine to make use of the continuous wavelet transform (3), but this would imply the use of a (quasi-)continuous time variable which is computationally expensive and statistically useless. We can however propose [AC] to use another version of the wavelet transform based on a finer sampling of the scale axis while preserving a dyadic time sampling:

$$T_X[n = 2^j k, a = 2^{j+m/M}] = \langle x, \psi_{kj}^m \rangle,$$

with

$$\psi_{kj}^m(t) \equiv 2^{-(j/2+m/2M)} \psi(2^{-(j+m/M)}(t - 2^j k)), \quad m = 0, 1, \dots, M-1.$$

The associated spectral estimator then reads:

$$\hat{I}_{4,X} \left(\nu_j^m = 2^{-(j+m/M)} \nu_\psi \right) \equiv \frac{1}{N_j} \sum_{k=1}^{N_j} T_X^2[2^j k, 2^{j+m/M}].$$

We have shown elsewhere [AC, FC] that, using the multiresolution analysis framework, it is possible to design an algorithm that enables a fast and efficient implementation of this variation on the wavelet transform and therefore on the spectral estimator, in so far as it preserves the pyramidal structure of the Mallat algorithm. The key point of this algorithm lies in the possibility, from any arbitrarily chosen wavelet, to design a multiresolution wavelet that provides us with an excellent approximation of the starting pattern [AA].

When applied to velocity data in turbulence, this estimator highlights the existence of a small bump in the higher part of the inertial range (see Fig. 6). This phenomenon, known in the literature as the *bottle-neck* effect [LM], drastically limits the range of existence of the power-law spectrum and prevents from an accurate estimation of the parameter α . This is the most interesting result given by the time-scale based spectral estimator in the context of turbulence.

Acknowledgements. This work was supported in part by the National Science Foundation, grant no MIP 94-57438.

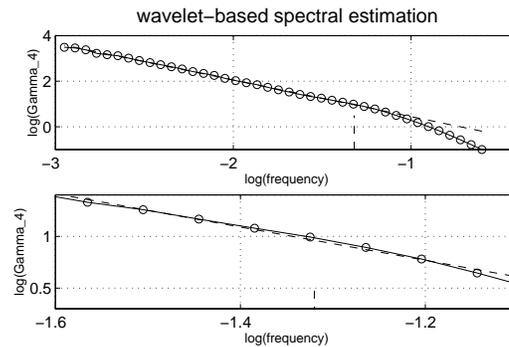


Figure 6: **Wavelet-based spectral estimation for any chosen frequency** Making use of the multiresolution framework, it is possible to define a spectral estimator for frequencies that no longer have to be dyadic. Such an estimator can be implemented with a fast pyramidal algorithm. When applied to turbulence velocity data, this estimator evidences a departure from the power-law behaviour in the higher part of the so-called inertial range that calls into question the asymptotic power-law model of turbulence.

References

- [AA] Abry, P., Aldroubi, A.: Designing multiresolution analysis-type wavelets and their fast algorithms. *J. Fourier Anal. Appl.* (to appear 1995).
- [AC] Abry, P., Chassande-Mottin, E., and P. Flandrin: Algorithmes rapides pour la décomposition en ondelettes continue. Application à l'implantation de la réallocation du scalogramme. Submitted to Colloque GRETSI, Juan les Pins, France, september 1995.
- [AF] Abry, P., Flandrin, P.: On the initialization of the discrete wavelet transform. *IEEE Signal Proc. Lett.* **SPL-1** (1994) 32–34.
- [AG] Abry, P., Gonçalves P., and P. Flandrin: Wavelet-based spectral analysis of $1/f$ processes. *Proc. IEEE-ICASSP'93* (1993) III.237–III.240.
- [B1] Batchelor, G.K.: *The Theory of Homogeneous Turbulence*. Cambridge Univ. Press (1953).
- [D] Daubechies, I.: *Ten Lectures on Wavelets*. SIAM (1992).
- [F1] Falconer, K.: *Fractal Geometry*. J. Wiley and Sons (1990).
- [F2] Flandrin, P.: On the spectrum of fractional Brownian motions. *IEEE Trans. on Info. Theory* **IT-35** (1989) 197–199.
- [F3] Flandrin, P.: Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Info. Theory* **IT-38** (1992) 910–917.
- [F4] Flandrin, P.: Temps-fréquence. Hermès, France, 1993
- [F5] Flandrin, P.: Time-scale analyses and self-similar stochastic processes. in Byrnes, J. *et al. (eds.): Wavelets and Their Applications*. Kluwer (1994) 121–142.
- [FC] Flandrin, P., Chassande-Mottin, E., and Abry, P.: Reassigned scalograms. *Proc. SPIE'95* (to appear 1995).
- [JS] Jawerth, B., Sweldens, W.: An overview of wavelet based multiresolution. *SIAM Rev.* **36** (1994) 377–412.
- [K1] Kay, S.M.: *Modern Spectral Estimation – Theory and Application*. Prentice-Hall (1988).
- [K2] Keshner, M.S.: $1/f$ noise. *Proc. IEEE* **70** (1982) 212–218.
- [LM] Lohse, D., Muller-Groeling, A.: Bottleneck effects: scaling phenomena in r -versus p -space. Preprint (1994).
- [LO] Lundhal, T., Ohley, W.J., Kay, S.M., Siffert, R.: Fractional Brownian Motion: maximum likelihood estimator and its application to image texture. *IEEE Trans. on Medical Imaging*, **MI-5** (1986) 152–161.
- [MC] McCauley, J.L.: *Chaos, Dynamics and Fractals – An Algorithmic Approach to Deterministic Chaos*. Cambridge Univ. Press (1993).
- [MV] Mandelbrot, B.B., van Ness, J.W.: Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** (1968) 422–437.
- [M1] Masry, E.: The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Trans. on Info. Theory* **IT-39** (1993) 260–264.
- [M2] Mallat, S.: A theory for multiresolution signal decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-11** (1989) 674–693.
- [RZ] Ramanathan, J., Zeitouni, O.: On the wavelet transform of fractional Brownian motion. *IEEE Trans. on Info. Theory* **IT-37** (1991) 1156–1158.

- [TK] Tewfik, A.H., Kim, M.: Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. on Info. Theory* **IT-38** (1992) 904–909.
- [W1] Welch, P.D.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short modified periodograms. *IEEE Trans. on Audio* **AU-15** (1967) 70–73.
- [W2] Wornell, G.W.: A Karhunen-Loève-like expansion for $1/f$ processes via wavelets. *IEEE Trans. on Info. Theory* **IT-36** (1990) 859–861.
- [W3] Wornell, G.W.: Wavelet-based representations for the $1/f$ family of fractal processes. *Proc. IEEE* **81** (1993) 1428–1450.
- [WO] Wornell, G.W., Oppenheim, A.V.: Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. on Signal Proc.* **SP-40** (1992) 611–623.