MULTIPLE-WINDOW WAVELET TRANSFORM AND LOCAL SCALING EXPONENT ESTIMATION

Paulo Gonçalvès

INRIA Rocquencourt - Projet Fractales
Domaine de Voluceau
B.P. 105 Le Chesnay 78153 Cedex, France
tel: (33) 1 39 63 52 79, Fax: (33) 1 39 63 57 71
Email: paulo.goncalves@inria.fr

Patrice Abry

Laboratoire de Physique - CNRS URA 1325
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon cedex 07, France
tel: (33) 4 72 72 84 93, Fax: (33) 4 72 72 80 80
Email: pabry@physique.ens-lyon.fr

ABSTRACT

We propose here a *multiple-window wavelet transform* for the purpose of identifying *non-stationary self-similar structures* in random processes and estimating the time-varying scaling exponent $H(t)$ that controls the local regularity and correlation of the process. More specifically, our final aim is to be able to track even rapidly varying trajectories $(t, H(t))$. The solution described here combines analysis obtained from scalograms computed with a set of multi-windows designed so as to satisfy to a decorrelation condition. We derive here the statistics for the estimate of $H(t)$, compare it against numerical simulations and show that we obtain a substantial reduction of variance in estimation, without introducing bias.

1. MOTIVATION

The problem of tracking the local regularity $H(t)$ of a function arises in many real world applications (two examples are local scaling properties in high speed telecommunication traffic and time-varying self similarity of physiologic signals). Very often also, sharp variations of the Hölder function $H(t)$ precludes the use of smoothed-wavelet based estimators as proposed in [1], as the bias/variance trade-off is penalized by the smearing effect [2]. In this paper, we propose a *multiple-window wavelet transform* inspired by Thomson's method for classical spectral analysis [3]. Following the idea of projecting a unique observation x onto several orthogonal subspaces, we consider each projection as a different realization of the same random process. After having derived a shift cross-correlation condition on the wavelet sets to span (almost) orthogonal subspaces, we turn, in a second step, to bases designed from the multiresolution analysis theory [4]. This framework not only simplifies considerably the derivation of the wavelet sets, but also supports fast and efficient algorithms.

At last, we identify the probability density function underlying the coefficients of the *Log*-time-scale distribution based on the multiple-window wavelet transform. The stabilized variance of the *Log*-scalogram across scales carries over to that of the estimate $\widehat{H(t)}$. Simulation experiments evidence the dramatic gain induced on the bias/variance trade-off.

This work was supported by the National Science Foundation, grant no. MIP–9457438, and NATO Collaborative Research Program, grant GRG–950202.

2. LOCAL SCALING EXPONENTS AND WAVELET TRANSFORM

• **A model for local regularity.** A locally self-similar processes is defined by the equality (in distribution) $x(at) \overset{\mathrm{D}}{=} a^{H(t)}x(t)$, where $H(t)$ is the time-varying scaling (or Hölder) exponent to be estimated ($H(t)$ will also be referred to as the local regularity of the process). In the course of our development we will use the *Multifractional Brownian Motion* (MBM) [1, 5], a generalization of the constant$-H$ fractional Brownian motion, as a paradigm for locally self-similar processes. Assuming continuity of the Hölder function $H(t)$, a limit expression for the covariance function of a MBM is

$$\begin{aligned}\gamma_x(t;\tau) \;&=\; \mathbf{E}\{x(t)x^*(t+\tau)\} \\ =\; \frac{\sigma_t^2}{2}\,&\left[|t|^{2H(t)} + |t+\tau|^{2H(t)} - |\tau|^{2H(t)}\right],\; \tau \to 0, \quad (1)\end{aligned}$$

whereas the variance of its increment process is

$$\mathbf{E}\{|x(t+\tau) - x(t)|^2\} \;=\; \left(\sigma_t^2/2\right)\, |\tau|^{2H(t)},\; \tau \to 0.$$

• **Scaling behavior of the scalogram.** Let us denote the wavelet transform by[1]

$$T_x(t,a) = \int x(\tau)\, \psi_{t,a}(\tau)\, d\tau \quad (2)$$

where $\psi_{t,a}(\tau) \;=\; |a|^{-\frac{1}{2}}\psi(|a|^{-1}(\tau - t))$. Let us moreover define the scalogram as the squared-magnitude of the wavelet transform: $S_x(t,a) = |T_x(t,a)|^2$.

[1]Throughout this paper, integration bounds run from $-\infty$ to $+\infty$. Without loss of generality, we deal with real wavelets only.

Then, it is known from [1, 2, 6] that the local scaling structure echoes the self-similarity of those processes as

$$\begin{aligned}\mathbf{E}\{S_x(t,a)\} &= a^{2H(t)+1}\,\sigma_t^2 \int |u|^{2H(t)}\gamma_\psi(u)\,du, \;\; a\to 0 \\ &= a^{2H(t)+1}\,C_\psi(t), \qquad a\to 0. \end{aligned} \tag{3}$$

where $\gamma_\psi(u) = \int \psi(v)\psi(v-u)\,dv$ stands for the autocorrelation of the analyzing wavelet. For the range $0 \le H(t) \le 1$ of interest, this behavior holds provided the wavelet ψ verifies the usual admissibility condition, i.e. $\int \psi(t)\,dt = 0$. When $H(t)$ exceeds this range, this condition needs to be extended up to a sufficient number of vanishing moments of the wavelets [7]. In this expression, the major problem is to access a good estimate of the ensemble average on the scalogram given a single observation of the random process x. Whereas in [1] a local time smoothing was proposed, we adopt here a different solution that relates to the technique of multiple-window spectral analysis to reduce the variance on the estimate of $\mathbf{E}\{S_x(t,a)\}$.

3. MULTIPLE-WINDOW WAVELET TRANSFORM

• **Principle and definition.** In [3], Thomson deals with the problem of spectral analysis of a single short-length random observation x by proposing a set of orthogonal analyzing functions ω_i as windows for periodograms. Then, each window ω_i spans a signal subspace $\mathcal{S}_i$ orthogonal to any other subspace $\mathcal{S}_j$ derived from a different window $\omega_{j\neq i}$. Then, all periodograms $P_i(f) = \left|\int x(t)\,\omega_i(t)\,e^{-i2\pi f t}\,dt\right|^2$ can be viewed as uncorrelated spectra of the same random variable. So, the weighted sum[2] $\widehat{\Gamma}_x(f) = \sum_i \beta_i\,P_i(f)$ corresponds to a sample mean estimate and has a variance lowered by a factor equal to the number of used windows.

Based on the same idea, in [8, 9] the *multiple-window scalogram* is defined as follows[3]

$$\Omega_x(t,a) = \frac{1}{L}\sum_{i=1}^{L}\left|\int x(\tau)\,\psi_{t,a}^{(i)}(\tau)\,d\tau\right|^2 \tag{4}$$

where $\{\psi^{(i)}\}_{i=1\ldots L}$ is a set of chosen mother wavelets.

• **Decorrelation condition.** In order to reduce the variance on the average scalogram Ω_x applied to MBM of the form (1), we must choose a set $\{\psi^{(i)}\}_{i=1\ldots L}$ leading to uncorrelated wavelet transforms, in other terms we want

$$\mathbf{E}\{T_x^{(i)}(t,a)\,T_x^{(j)}(t,a)\} \equiv 0 \qquad \forall i\neq j,\ \forall t,\ a\to 0.$$

Solution to this constraint imposes the set $\{\psi^{(i)}\}_{i=1\ldots L}$ to verify the following equality

$$\int |u|^{2H(t)}\gamma_{i,j}(u)\,du \equiv 0, \qquad \forall i\neq j \tag{5}$$

[2] The weights β_i are such that $\sum_i \beta_i = 1$.

[3] A generalization of this principle to time-varying spectral analysis can be found in [9].

where $\gamma_{i,j}(u) = \int \psi^{(i)}(v)\psi^{(j)}(v-u)dv$ is the cross-correlation function between the two wavelets $\psi^{(i)}$ and $\psi^{(j)}$. Moreover and without loss of generality, we impose that the wavelets are of unit energy, i.e. :

$$\gamma_{i,i}(0) = \int |\psi^{(i)}(v)|^2\,dv \equiv 1. \tag{6}$$

In the past, several studies (mainly in coding theory [10]) have faced, and partially solved, similar problems of finding time-shift cross-orthogonal functions, but in general this remains a difficult task. Therefore, we decide to restrict the class of solutions to the ones stemming from the procedure described below.

• **Designing the set of wavelets.** To help finding solutions to this set of constraints we choose to write each mother wavelet $\psi^{(i)}$ as a linear combination of an existing multiresolution-type wavelet $\psi^{(0)}$ (referred to as the *grand-mother wavelet*)

$$\psi^{(i)}(t) = \sum_{k=0}^{K} q_k^{(i)}\psi^{(0)}(t-k). \tag{7}$$

Here $\psi^{(0)}$ is chosen such that the collection $\{\psi_{j,k}^{(0)}(t) = 2^{-j/2}\psi^{(0)}(2^{-j}t-k),\,(j,k)\in(\mathbb{Z}^+,\mathbb{Z})\}$ defines an orthonormal basis of wavelets (strictly speaking, it only needs to form a Riesz [4] basis, but the choice of an orthonormal basis simplifies further calculations). This linear combination technique is a general procedure that enables the design of infinitely many different semi-orthogonal or bi-orthogonal multiresolution-type basis of wavelets [4, 11]. Within this framework, the decorrelation condition as expressed in equation (5) can be rewritten in terms of the coefficient series $q^{(i)}$ and reduces to

$$C_{i,j} = \sum_{l=-K+1}^{K-1} Q_{i,j}(l)\int |u|^{2H(t)}\,\gamma_0(u-l)du \equiv 0, \quad \forall i\neq j \tag{8}$$

where $Q_{i,j}(l) = \sum_k q_k^{(i)}q_{k-l}^{(j)}$ corresponds to the cross-correlation between the two sequences $q^{(i)}$ and $q^{(j)}$, and γ_0 the autocorrelation function of the grand-mother wavelet $\psi^{(0)}$. The normalization condition of equation (6) becomes :

$$Q_{i,i}(0) = 1,\ \forall i. \tag{9}$$

Hence, the identification of the set $\{\psi^{(i)}\}_{i=1\ldots L}$ simplifies to the derivation of a finite set of coefficients $\{q_k^{(i)},\ k=1,\ldots,K\}$ that verifies (8) and (9).

However, since equation (8) involves a time-varying regularity $H(t)$, the set of coefficients $\{q_k^{(i)},\ k=1,\ldots,K\}$ need to satisfy time-shift cross-orthogonality (in the sense of (8)) for all $0 < H(t) < 1$ simultaneously. To overcome this other difficulty, we propose the trivial solution satisfying :

$$|Q_{i,j}(l)| = 0, \forall i\neq j,\ \forall l \text{ and } Q_{i,i}(0) = 1, \forall i.$$

Unfortunately, numerical solutions of this new equation correspond to sequences of length $K > 2L$, yielding unacceptable long wavelets as far as time-tracking of $H(t)$ is concerned (see [2]). Nevertheless, one can easily show that the modulus of the integral

$$\int |u|^{2H(t)}\,\gamma_0(u-l)\,du\,,$$

decreases rapidly with l (see figure 1), which as a result, loosens the constraint on the solutions $q^{(i)}$. Therefore, we end up with the following constraints, solved by a *Matlab* optimization routine :

$$\begin{aligned} & |Q_{i,j}(l)| = 0, \qquad \forall i \neq j,\ |l| < K_\epsilon \\ \text{and} \quad & Q_{i,j}(l) < \epsilon, \quad \forall i \neq j,\ K_\epsilon \leq |l| \leq K. \end{aligned}$$

Wavelets designed by mean of this procedure are used in the numerical simulations reported in figure 2.

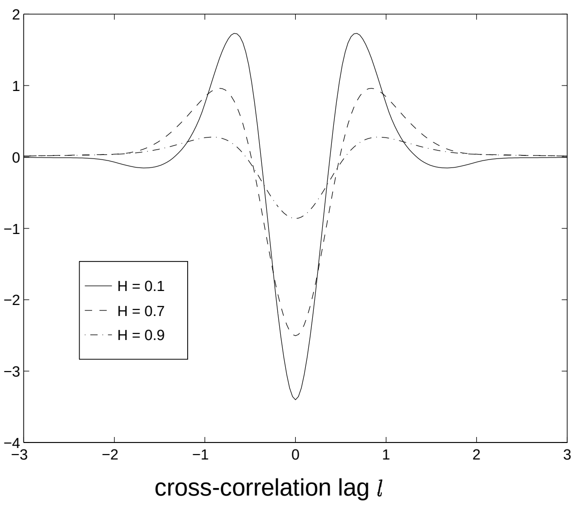

Figure 1: *Cross-correlation* $\int_{-\infty}^{\infty} |u|^{2H(t)}\,\gamma_0(u-l)\,du$, *for* $\psi_0(t)$, *a Daubechies wavelet of regularity 2 (at scale* $a = 1$*) and for three different values of the exponent* $H(t)$*.*

4. APPLICATION TO LOCAL SCALING EXPONENT ESTIMATION

We now use this new multiple-window wavelet transform to estimate the local Hölder exponent $H(t)$ of a MBM. The results regarding the statistics of the estimate $\widehat{H(t)}$ presented here are obtained assuming that a MBM is a locally Gaussian process with variance σ_t^2 [5] (details of the calculus can be found in [2]). Moreover all the results listed below are to be understood with respects to the limit of $a \to 0$.

• **Statistics of the multiple window wavelet transform.** In the following, we denote by Ω_x the multiple window wavelet transform proposed in (4), and by $S_x^{(i)}$ the scalogram based on the wavelet $\psi^{(i)}$. Because each wavelet $\psi^{(i)}$ is a linear combination of orthogonal admissible grandmother wavelets $\psi^{(0)}$, it can be shown, using notations and results of relation (3), that the statistics for $\log S_x$ reads (cf [2]) :

$$\chi_1^2\left((2H+1)\log(a) + b^{(i)}, \pi^2/2\right)$$

where $b^{(i)}$ is a constant depending on $\psi^{(i)}$ which can be made explicit.
Using theorems for non-linear transforms of asymptotically normal (AN) random variables (see e.g., [12]), it is straightforward to show that the probability density function for the variable $\log \Omega_x$ is itself AN with mean $\mu_{\log\Omega} = (2H+1)\log(a) + b_\Omega$ and variance $\sigma^2_{\log\Omega} = (2/L)(1+\sigma_C^2/\mu_C^2)$, where b_Ω is a constant, and μ_C and σ_C are respectively the mean and the standard deviation of the sequence $\{C_{\psi^{(i)}},\, i = 1,\ldots,L\}$. Basically, it amounts to say that the variance of $\log\Omega_x$ behaves as $\sigma^2_{\log\Omega} \simeq 2/L$.

• **Bias and variance on the estimate $\widehat{H(t)}$.** The statistics presented above show that both $S_x(t,a)$ and $\Omega_x(t,a)$ provide us with unbiased estimates of the local regularity, denoted $\widehat{H(t)}_S$ and $\widehat{H(t)}_\Omega$ respectively and obtained from a linear regression of the corresponding distribution *vs.* scale in a *Log-Log* plot. Moreover, the fact that the variances of $\log S_x$ and $\log\Omega_x$ are constant with respect to the scale a, indicates that non-weighted linear fits can be used as efficient estimates. More precisely, assuming that the linear fits are performed on the set $\{a_j,\, j = 1,\ldots,J\}$ defining a reasonably loose sampling of the scale axis (say octaves $a_j = 2^j$) so that $\log S_x(t,a_i)$ and $\log S_x(t,a_j)$, $(i \neq j)$ are uncorrelated (respectively, $\log\Omega_x(t,a_i)$ and $\log\Omega_x(t,a_j)$, $(i \neq j)$), we can derive the second-order statistics of the estimates, we get :

$$\text{Var}(\widehat{H(t)}_S) = f(J)\pi^2/2 \ \text{ and } \ \text{Var}(\widehat{H(t)}_\Omega) = f(J)2/L\,,$$

where $f(J)$ is a function depending only on the number of octaves J involved in the linear fits.

• **Comparison with numerical simulations.** To implement this new multiple-window estimation technique, we need to implement L wavelet transforms. By averaging the L estimates $\widehat{H(t)}_{S^{(i)}}$ obtained independently from the L scalograms, we may obtain a reduction of variance of a factor L, as compared to the variance of a crude estimator based on a single scalogram:

$$\text{Var}(\widehat{H(t)}_S)/\text{Var}\left(L^{-1}\sum_{i=1}^{L}\widehat{H(t)}_{S^{(i)}}\right) = L\,.$$

On the other hand, the reduction of variance on the estimate $\widehat{H(t)}_\Omega$ derived from averaging the scalograms before taking the log reads :

$$\text{Var}(\widehat{H(t)}_S)/\text{Var}(\widehat{H(t)}_\Omega) \simeq (\pi^2/4)L \,>\, L.$$

Experimental results shown in figure 2 confirm reasonably well this assertion. For a piece-wise constant MBM

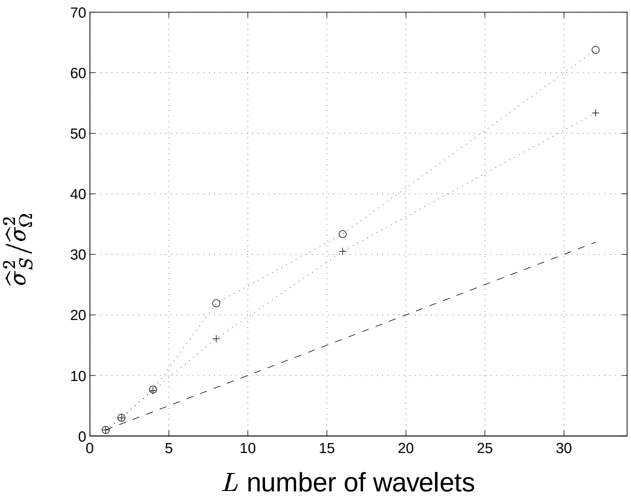

Figure 2: *For a piece-wise constant MBM, ratio $\widehat{\sigma}_S^2/\widehat{\sigma}_\Omega^2$ vs. L, where $\widehat{\sigma}_S^2$ and $\widehat{\sigma}_\Omega^2$ are the variances of $\widehat{H(t)}$ estimated respectively from a single scalogram (2) and the multiple-window method (4) based on L wavelets. The wavelets we use are of the form (7) and verify the condition (5) where a* Daubechies2 *wavelet (i.e., with 4 coefficients) as the grand-mother wavelet is chosen. Symbols $\circ$ and $+$ correspond to $H = 0.2$ and $H = 0.8$ respectively. The dashed line shows the usual improvement on the variance obtained with a sample mean estimator, assuming independence of the L realizations. The empirical variances are estimated on a 50 independent trial set of MBM.*

($H(t) = 0.2$ for $0 \leq t < t_0$ and $H(t) = 0.8$ for $t_0 \leq t < t_1$) we compute the empirical variances $\widehat{\sigma}_S^2$ of $\widehat{H(t)}_S$ and $\widehat{\sigma}_\Omega^2$ of $\widehat{H(t)}_\Omega$ based respectively on a single scalogram (2) and on the multiple-window method (4). The ratio $\widehat{\sigma}_S^2/\widehat{\sigma}_\Omega^2$ is ploted *vs.* the number L of wavelets involved in (4). Choosing a set of wavelets verifying the shift cross-correlation condition (5), the multiple-window estimator performs very well as compared to the usual decrease of variance achieved by a sample mean estimator assuming independence of the L realizations.

5. CONCLUSIONS AND PERSPECTIVE

We propose a generalization of the multiple-window method of Thomson for constant-Q time-varying spectral analysis of *locally self-similar processes*. Applied to local scaling exponent estimation, this method performs surprisingly well, insofar as it is unbiased (with the limit $a \to 0$) and exhibits very low variance. It also prompts the use of non-linear estimators like *cross-validation* [13] to eventually replace the sample mean estimator of equation (4) by a weighted sum :

$$\Omega_x(t,a) = \frac{1}{\sum \beta_i} \sum_{i=1}^{L} \beta_i \left| \int x(\tau)\, \psi_{t,a}^{(i)}(\tau)\, d\tau \right|^2 ;$$

this particular point is under current investigation. Another interesting issue is to formalize the place of these multiple window wavelet transforms within the echelon of the affine class of time-scale distributions [14]

$$\Omega_x(t,a) \;=\; \int\!\!\int W_x(\tau,\theta)\, \Pi\left(\frac{\tau - t}{a}, a\theta\right)\, d\tau\, d\theta,$$

where $W_x(\tau,\theta)$ is the Wigner-Ville distribution of the signal x. While it is straightforward to show that the corresponding kernel $\Pi(\tau,\theta)$ writes in our case

$$\Pi(\tau,\theta) \;=\; \frac{1}{L} \sum_{i=1}^{L} W_{\psi^{(i)}}(\tau,\theta)\,,$$

it is not clear how in general, the bias-variance trade-off on the estimate $\widehat{H(t)}_\Omega$ stems from the properties of this kernel (see [2] for further discussions).

6. REFERENCES

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

[2] P. Gonçalvès, P. Abry, and P. Flandrin, "Time-scale based estimation for local regularity," 1997.

[3] D. J. Thomson, "Spectrum estimation and harmonic analysis," vol. 70, pp. 1055–1096, Sept. 1982.

[4] A. Aldroubi and M. Unser, "Families of multiresolution and wavelet spaces with optimal properties," *Num. Func. Anal. Opt.*, vol. 14, pp. 417–446, 1993.

[5] R. F. Peltier and J. Lévy-Véhel, "Multifractional Brownian motion : definition and preliminary results," *Stochastic Processes and their Applications*, 1995. submitted.

[6] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. on Info. Theory*, vol. IT-35, pp. 197–199, June 23, 1989.

[7] P. Abry, P. Gonçalvès, and P. Flandrin, "Wavelets, spectrum analysis and $1/f$ processes," in *Wavelets and Statistics (A. Antoniadis, Ed.)* (A. Antoniadis and G. Oppenheim, eds.), Lectures Notes in Statistics, 1995.

[8] J. M. Lilly and J. Park, "Multiwavelet spectral and polarization analyses of seismic records," *Geophys. J. Int.*, no. 122, pp. 1001–1021, 1995.

[9] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency and time-scale analysis," 1996.

[10] N. Suehiro and M. Hatori, "N-shift Cross-Orthogonal Sequences," vol. 34, pp. 143–146, Jan. 1988.

[11] P. Abry and A. Aldroubi, "Designing multiresolution analysis-type wavelets and their fast algorithms," *J. Fourier Anal. Appl.*, vol. 2, no. 2, pp. 135–159, 1995.

[12] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

[13] R. Nowak, "Optimal Signal Estimation Using Cross Validation," Nov. 1996. To appear.

[14] O. Rioul and P. Flandrin, "Time-scale energy distributions: A general class extending wavelet transforms," *IEEE Trans. Signal Processing*, vol. SP-40, pp. 1746–1757, July 1992.