WAVELET BASED ESTIMATOR FOR FRACTIONAL BROWNIAN MOTION: AN EXPERIMENTAL POINT OF VIEW

Gérard Jacquet, Rachid Harba***

* Laboratoire Traitement du Signal et Instrumentation
UMR 5516 and CNRS GDR ISIS
Université de Saint Etienne, 23, rue Paul Michelon
42023 Saint-Etienne Cedex 2, France
Phone : 33 477 48 50 24
jacquet@univ-st-etienne.fr
** Laboratoire d'Electronique, Signaux, Images
UPRES 1715 and CNRS GDR ISIS
Polytech'Orléans - 12, rue de Blois,
BP 6745, 45067 Orléans, France
Phone : 33 238 41 72 29
Rachid.Harba@univ-orleans.fr

Abstract: Wavelet based estimators of the H parameter for fractional Brownian motion (fBm) is known to have interesting asymptotical properties. In this communication, we are studying the practical point of view by testing this estimator on finite length fBm signals of N samples. These signals are generated using the circulant embedding method (CEM). CEM is fast and exact such that the synthesis of true fBm signals of long size is easy. Results show that above N=4096 wavelet based estimator is unbiased and very close to the Cramer-Rao lower bound. At the light of this study, and by combining it to recent results, a practical user guide to estimate the H parameter of fBm can be provided: if N is lower than 512 classical maximum likelihood (ML) should be chosen. For N in the interval]512,4096] Whittle ML is to be preferred. For N above 4096 wavelet based should be selected. Finally, a precise confidence interval of the true H parameter can be given.

I. Introduction

Fractional Brownian motion (fBm) of H parameter in]0,1[is a stochastic model for nonstationary fractal data [1]. FBm is very helpful for modelling numerous real-world phenomena [2][3][4]. The main problem that occurs when using fBm as a model is to properly estimate the H parameter. Many H estimators are available and the choice of a method is a difficult issue. Among them, the wavelet based is one of the most interesting since it naturally matches the structure of the fBm process for two reasons. First, although fBm is nonstationary, its wavelet transform is stationary. Second, even if fBm is long range dependant, its wavelet coefficients are almost uncorrelated. From a practical point of view, two reasons also can motivate the use of this method: its complexity is only O(N) and it is known that the wavelet based estimator has interesting asymptotical properties. Namely for 1/f processes this estimator is efficient [5]. However, it is difficult to predict the experimental performances of the wavelet estimator for fBm, *e. g.* when time-limited data of N samples are of interest. To assess the quality of an analysis method, experimental studies have to be carried out [6]. In this last reference, analysis methods including the wavelet one were compared on fBm synthetic data for which the true H value is known. Unfortunately, some of the syntheses were not exact in principle, and the performance evaluation of analysis methods was difficult. The quality assessment of an estimation technique requires exact fBm signals than can be generated at low computational cost. Recently, a new method called the Circulant Embedding Method (CEM) was proposed to synthesise long and exact fBm data very easily [7].

In this work, we will experimentally evaluate the practical efficiency of wavelet based estimators on true fBm data of N samples generated by the CEM algorithm. Steaming from this study, and by combining these results to recent ones [8], a general framework to precisely measure the H parameter from a single realization will be proposed. Finally, a precise confidence interval of the true H parameter will be given.

This communication is organized as follows. First, the general background of the study will be recalled. It includes presentations of fBm main properties, of the CEM method, of the wavelet based estimators of the H parameter (focusing on practical remarks), and on the statistical tests to assess the efficiency of the estimator. Results will be presented and a general framework for a precise estimation of the H parameter will finally be given as well as the confident interval for the H parameter.

II. General background

II-1 fBm main properties

Continuous fBm of parameter H in]0,1[, denoted $B_H(t)$, is defined as an extension of Brownian motion B(t):

$$B_H(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{(i\omega)^{H+1/2}}(e^{it\omega}-1)dB(\omega). \quad (1)$$

When H=1/2, fBm is reduce to Brownian motion. From now on, we will focus on properties of discrete processes denoted $B_H[i]$. With a starting value $B_H[0]=0$, fBm is zero mean, Gaussian and second order nonstationary as attested by its autocorrelation function:

$$r_{B_H}[i,j]=\frac{\sigma^2}{2}\left(|i|^{2H}+|j|^{2H}-|i-j|^{2H}\right). \quad (2)$$

fBm has no derivative, and thus its increments for a time lag 1 are of interest. They are named fractional Gaussian noises (fGn), denoted G, and defined as:

$$G[i]=\Delta B_H[i]=B_H[i]-B_H[i-1]. \quad (3)$$

They are zero mean, Gaussian and stationary processes since their autocorrelation can be written:

$$r_G[k]=\frac{\sigma^2}{2}\left(|k+1|^{2H}-2|k|^{2H}+|k-1|^{2H}\right). \quad (4)$$

II-2. The CEM method

The object of this section is to briefly explains the Circulant Embeding Method (CEM) to easily generate exact fBm data [7]. As fBm is non stationary, it is more convenient to first generate stationary fGn sequences, and to recover from these signals fBm time series. CEM factorizes an extension of the covariance matrix **R** of fGn to produce random vectors with exactly the required correlation structure via FFT. The elements of this Toeplitz matrix **R** are given by the autocorrelation function of fGn and are such that:

$$R_{pq}=r_G[|p-q|]=r_G[k] \text{ for } k=0,\dots,N-1. \quad (5)$$

CEM consists in embedding **R**, the N×N correlation matrix, in a larger 2M×2M nonnegative definite matrix **S** such that M=N-1. The first row of **S,** denoted **s**, consists in the entries:

$$\begin{aligned} s[k] &= r[k], \quad k=0,\dots,N-1, \\ s[2M-k] &= r[k], \quad k=1,\dots,N-2. \end{aligned} \quad (6)$$

S is then circulant, and any N×N matrix extracted along the diagonal is a copy of **R**. Being circulant, **S** can be decomposed as $\mathbf{S}=\mathbf{FDF}^T$ where **F** is the standard 2M×2M FFT matrix. **D** is diagonal, and its diagonal $\tilde{\mathbf{s}}=\mathbf{Fs}$ is obtained by 1D FFT of the covariance function of the desired process [9]. In other words, $\tilde{\mathbf{s}}$ is the discrete PSD of the stationary model. The final step is the following: form $\mathbf{y}=\mathbf{FD}^{1/2}\mathbf{x}$ with $\mathbf{x}\sim N(0,\mathbf{I})$, a zero mean Gaussian random vector with the identity matrix **I** as covariance matrix. Then, **y** has the desired covariance since

$$\begin{aligned} E(\mathbf{yy}^T)&=E(\mathbf{FD}^{1/2}\mathbf{xx}^T\mathbf{D}^{1/2}\mathbf{F}^T)= \\ &\mathbf{FD}^{1/2}E(\mathbf{xx}^T)\mathbf{D}^{1/2}\mathbf{F}^T=\mathbf{FD}^{1/2}\,\mathbf{I}\,\mathbf{D}^{1/2}\mathbf{F}^T=\mathbf{S}. \end{aligned} \quad (7)$$

Any vector of N samples extracted out of either the imaginary or real part of **y** is fGn since its covariance matrix is equal to **R**. The only distributional errors of the method are due to inaccuracies in computer arithmetic and to the use of pseudo-random numbers instead of genuine ones. Exact realizations of fBm can be recovered from fGn ones with the initial condition $B_H[0]=0$ following:

$$B_H[i]=\sum_{j=1}^{i}G_1[j]=G_1[i]+B_H[i-1]. \quad (8)$$

II.3. Wavelet estimator

The wavelet based estimator of the H parameter is one of the most interesting since it naturally matches the structure of the fBm process. Although fBm is nonstationary, its wavelet transform is stationary [10]. Indeed, the variance of the details signals D_j at a dyadic scale j follows:

$$\mathrm{Var}(D_j[n]) \propto (2_j)^{2H+1}. \quad (9)$$

∝ means proportional to. The H parameter can be directly estimated by linear regression from last equation:

$$\mathrm{Log}_2\left(\mathrm{Var}\left(D_j[i]\right)\right)=-(2H+1)j+\text{constant}. \quad (10)$$

If the number of vanishing moments is greater than 2, detail signals are almost uncorrelated [10], which is better if H is to be estimated using last equation. Any Daubechies wavelet with more than 2 vanishing moment is suitable [11].

Several elements should be taken into account for an efficient implementation [10]. First of all, border effects due to the filtering can lead to wrong values. To reduce these effects, we chose the smallest possible filter, that is Daubechies's wavelet of 4 coefficients. In addition, polluted samples are discarded. Furthermore, and it is the main cause of error, the relation (9) is only valid in the continuous case. The sampling introduces a bias in the estimation of H, especially in the early stages of decomposition. This leads to underestimate the H parameter, this difference being smaller when H is close to 1. We have implemented a correction which improves the quality of the results [10]. It consists in balancing the variance of the détail signals as a function of the chosen wavelet, of the scale and of the H value according to the following :

$$\mathrm{Var}_{corrected}(D_j)=\frac{\mathrm{Var}(D_j)}{\dfrac{C_0(j)}{2^{(2H+1)j}}} \quad (11)$$

where

$$C_0(j)=-\sum_{n=-\infty}^{\infty}\gamma_g[-n]D_{j-1}[n] \quad (12)$$

and

$$D_j[n]=\sum_{n'=-\infty}^{\infty}\gamma_h[2n-n']D_{j-1}[n'] \text{ if } j\geq 1 \quad (13)$$

with

$$D_0[n]=|n|^{2H}. \quad (14)$$

γ_h et γ_g are the autocorrelation function of the impulse response of filter H and G of the wavelet decomposition. It is clear that a precise estimation

H/N	128	256	512	1024	2048	4096	8192	16384	32768
0.2	0.0430	0.0300	0.0212	0.0153	0.0105	0.0075	0.0052	0.00375	0.0027
0.5	0.0553	0.0394	0.0270	0.0193	0.0134	0.0092	0.0065	0.0046	0.0032
0.8	0.0542	0.0395	0.0288	0.0289	0.0145	0.0189	0.0077	0.0056	0.0035

Table 1 : Square root of the Cramer-Rao lower bound for the estimation of the H parameter as a function of H and N.

H/N	128	256	512	1024	2048	4096	8192	16384	32768
0.2	0.1702	0.1859	0.1907	0.1938	**0.19744**	0.19753	**0.19926**	**0.19905**	**0.1995**
	0.1248	0.0684	0.0437	0.0291	0.0186	0.0147	0.0090	0.0067	0.0045
0.5	0.4752	0.4847	0.4923	0.4964	**0.4980**	**0.4989**	**0.4990**	**0.4990**	**0.4995**
	0.1293	0.0754	0.0468	0.0312	0.0203	0.0141	0.0103	0.0071	0.0045
0.8	0.7788	0.7894	0.7929	**0.7957**	**0.7986**	0.7971	**0.7998**	**0.7990**	**0.7996**
	0.1321	0.0747	0.0488	0.0334	0.0221	0.0170	0.0100	0.0075	0.0048

Table 2 : Mean H value and standard deviation of the wavelet based estimation of the H parameter as a function of H and N. When unbiased or when reaching the CRLB, values are bold.

by this method can only be iterative since the H value is necessary to realize the corrections. The convergence is very fast, 3 or 4 iterations are enough. Finally, to improve the variance, it is recommended to make a weighted regression [12].

III. Practical efficiency

The practical efficiency of the wavelet based estimator is going to be tested. An estimator of the H parameter denoted $\hat{H}$ is efficient if unbiased and if its variance reaches the Cramer-Rao Lower Bound (CRLB). Here, it means that two joint hypothesis tests at a level of significance α have to be carried out to check this efficiency:

- a first test related to the null hypothesis $E(\hat{H})=H$, against the two-sided alternative hypothesis $E(\hat{H})\neq H$,
- a second test related to the null hypothesis $Var(\hat{H})=CRLB(H)$, versus the one-sided alternative hypothesis $Var(\hat{H})>CRLB(H)$.

NR independent realizations of size N power of 2 and of H value in]0,1[are first generated with the CEM method. On each of them, $\hat{H}$ is estimated. The sample mean $\overline{X}_H$ and the unbiased sample variances S_H^2 are calculated as:

$$\overline{X}_H=\frac{1}{NR}\sum_{i=0}^{NR-1}\hat{H}i \text{ and } S_H^2=\frac{1}{NR-1}\sum_{i=0}^{NR-1}(\hat{H}i-\overline{X}_H)^2. \quad (15)$$

Thus $\hat{H}$ is unbiased if [13]:

$$\left|\overline{X}_H-H\right| \le z_{\alpha/2}\sqrt{\frac{CRLB(H)}{NR}}. \quad (16)$$

z_a is the 100×a percentage point of the normal distribution. The variance of $\hat{H}$ reaches the CRLB if [13]:

$$S_H^2 \le \chi^2_{NR,\alpha}\frac{CRLB(H)}{NR}. \quad (17)$$

$\chi^2_{m,a}$ is the 100×a percentage point of the chi-square distribution with m degrees of freedom. If the two above inequalities are verified, $\hat{H}$ is efficient. For these tests, CRLB(H) values are needed. For unbiased estimators and under some regularity conditions, the CRLB(H) is such that:

$$CRLB(H)=\frac{1}{E\left[\left(\frac{\partial Ln(L(\mathbf{x};H))}{\partial H}\right)^2\right]}. \quad (18)$$

L(**x**;H) is the likelihood function of the observed fGn vector **x** depending on H. Using results in [14], the final expression is the following [15]:

$$CRLB(H)=\frac{2N}{N\times tr\left[(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial H})^2\right]-tr\left[(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial H})\right]^2} \quad (19)$$

tr is the trace operator. Table 1 shows as a function of H and N the square root of the CRLB(H), a quantity which has the same dimension as the H parameter. H values are 0.2, 0.5 and 0.8. Moreover, N values range from 128 to 32768 by steps of power of 2. For large N values, asymptotical results can be used [14]. Finally, we chose α=0.01 meaning that the tests are performed at a level of significance of 1%. We also chose the number of realizations NR=100.

V. Results

For the test on the mean H value, if $\left|\overline{X}_H - H\right|$ is lower than $z_{\alpha/2}\sqrt{\frac{\mathrm{CRLB(H)}}{\mathrm{NR}}} = 0.258\times\sqrt{\mathrm{CRLB(H)}}$, the null hypothesis is accepted meaning that the estimator is unbiased. For the test of the variance, if S_H^2 is lower than $\chi^2_{\mathrm{NR},\alpha}\frac{\mathrm{CRLB(H)}}{\mathrm{NR}} = 1.36\times\mathrm{CRLB(H)}$, the estimator reaches the CRLB. If both hypothesis are accepted, the estimator is efficient with a level of significance of α=1%. Results are presented in table 2. They indicate that for N above 4096, this estimator is unbiased and that its variance is very close to the CRLB.

VI. User guide for the practical estimation of H

Before measuring the H parameter, on should care about the fBm character of the given data. Among the properties to define fBm, only 3 are necessary and sufficient conditions: if a random process is Gaussian, and has stationary and self-similar increments, then it must be fBm [1]. 3 statistical tests are required to estimate the validity of the fBm model:

- a Gaussian test of the process,
- a stationary test of the process increments,
- a self-similarity test of the process increments.

The two first tests can be implemented following [13]. For the self-similarity test, Bardet has proposed a method in [16] that can be applied to a single realization.

If the given data is check to be fBm, the H parameter could be measured. The obtained results of this study show that the wavelet estimator is efficient above N = 4096. It can be combined to recent ones [8]. In this last study, it was shown that below N=512, only the classical maximum likelihood estimator [17] should be used because it is the only efficient one although its complexity is $O(N^2)$. Above N=512, it was recommended to use the Whittle ML version [18] since efficient and in $O(N\log_2 N)$. At the light of this new study, for N above 4096 wavelet based is to be used since unbiased, close to the CRLB and only in O(N).

Finally, a confidence interval for the H parameter is as follows. Since the used estimators are efficient, and that the estimates are Gaussian distributed, the true H is such that

$$\hat{H} - z_{1-\alpha/2}\sqrt{\mathrm{CRLB(H)}} \le H \le \hat{H} + z_{1-\alpha/2}\sqrt{\mathrm{CRLB(H)}}$$

with a confidence coefficient of 1-α. z_a is the 100×a percentage point of the normal distribution.

References

[1] B. B. Mandelbrot, J. W. Van Ness, "Fractional Brownian motion, fractional noises and applications," *SIAM*, vol. 10, pp. 422-438, 1968.

[2] C. W. J. Granger, "The typical spectral shape of an economic variable," *Econometrica*, vol. 34, pp. 150-161, 1966.

[3] B. B. Mandelbrot, "Noises with 1/f spectrum, a bridge between Direct current and white Noise," *IEEE Trans. on Info. Theory*, vol. IT-13, pp. 289-298, 1967.

[4] P. Abry, D. Veitch, "Wavelet Analysis of Long-Range-Dependent Traffic," *IEEE Trans. on Info. Theory*, vol. 44, pp. 2-15, 1998.

[5] G. G. Wornell, A. V. Oppenheim, Estimation of fractal signals from noisy measurments using wavelets, IEEE Trans. on Signal Proc., pp. 611-623, 1992.

[6] N. Gache, P. Flandrin, D. Gerreau, "Fractal dimension estimators for fractional Brownian motions," *Proceedings of the ICASSP*, vol. 5, pp. 3557-3560, 1991.

[7] E. Perrin, R. Harba, R. Jennane, I. Iribaren, *Fast and exact synthesis for 1D fractional Brownian motion and fractional Gaussian noises,* IEEE Signal Processing letters, pp. 382-384, 2002.

[8] R. Jennane, R. Harba, G. Jacquet, *Méthodes d'analyse du Mouvement Brownien Fractionnaire : théorie et résultats comparatifs,* Traitement du Signal, Vol. 18, N° 5-6, pp. 419-436, 2002.

[9] S. Barnet, "Matrices, Methods and Applications," *Oxford Applied Mathematics and Computing Sciences Series, Clarendon Press, Oxford*, 1990.

[10] P. Flandrin. «Wavelet Analysis and Synthesis of Fractional Brownian Motions», *IEEE Transactions on Information Theory*, Vol. 38, N° 2, pp. 910-917, March 1992.

[11] I. Daubechies. «Orthonormal Bases of Compactly Supported Wavelets», *Communications on Pure and Applied Mathematics*, Vol. XLI, pp. 909-996, 1988.

[12] P. ABRY. «Ondelettes et turbulences – Multirésolution, algorithmes de décomposition, invariance d'échelle et signaux de pression» Ed. Diderot, Paris, 1997.

[13] J. S. Bendat, A. G. Piersol, "RANDOM DATA: Analysis and measurement procedures," *Wiley-Interscience Editions*, 1971.

[14] S. M. Kay, "Fundamentals of Statistical Signal Processing," *Englewood Cliffs, New Jersey : Prentice Hall International Editions*, 1992.

[15] C. L. Benhamou, E. Lespessailles, G. Jacquet, R. Harba, R. Jennane, T. Loussot, D. Tourlière, W. J. Ohley, "A Fractal Evaluation of Trabecular Bone Microarchitecture on Radiographs," *Journal of Bone and Mineral Research*, Vol. 9, pp. 1909-1918, 1994.

[16] J. M. Bardet, "Un test d'autosimilarité pour les processus gaussiens à accroissements stationnaires," *C. R. Acad. Sci. Paris, Sér. I Math.,* vol. 6, pp. 521-526, 1999.

[17] T. Lundahl, W. J. Ohley, S. M. Kay, R. Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and its Application to Image Texture," IEEE Transactions on Medical Imaging, Vol. 5, N° 3, 1986, pp. 152-161.

[18] R. Dahlhaus, "Efficient parameter estimation for self-similar processes," *The Annals of Statistics*, vol. 17, pp. 1749-1766, 1989.