

Impact of Data Quantization on Empirical Multifractal Analysis.

Herwig Wendt, Stéphane G. Roux, Patrice Abry

▶ **To cite this version:**

Herwig Wendt, Stéphane G. Roux, Patrice Abry. Impact of Data Quantization on Empirical Multifractal Analysis.. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), IEEE Signal Processing Society, Apr 2007, Honolulu, Hawaî, United States. ensl-00144569

HAL Id: ensl-00144569

https://ens-lyon.hal.science/ensl-00144569

Submitted on 3 May 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IMPACT OF DATA QUANTIZATION ON EMPIRICAL MULTIFRACTAL ANALYSIS

Herwig Wendt, Stephane G. Roux and Patrice Abry

Physics Lab., CNRS UMR 5672, Ecole Normale Supérieure de Lyon, France
Herwig.Wendt@ens-lyon.fr, Stephane.Roux@ens-lyon.fr, Patrice.Abry@ens-lyon.fr

ABSTRACT

Multifractal Analysis is nowadays commonly used in real-life data analyses and involved in standard signal processing tasks such as detection, identification or classification. In a number of situations, mostly in Image Processing, the data are available for the analyses only in (possibly severely) quantized versions. The present contribution aims at analyzing the robustness of standard multifractal estimation procedures against quantization. To this end, we analyze the behaviors and statistical performance of these procedures when applied to a large number of realizations of known synthetic multifractal processes subject to various quantization levels. Our study shows that immunity against quantization can be obtained by restricting the range of scales involved in multifractal parameter estimation to the largest ones. Comparing multifractal analyses based on different multiresolution quantities, increments, wavelet coefficients and leaders, we show that wavelets, thanks to their good frequency localization, bring robustness against quantization when increments do not. This study provides the practitioner with a clear guide line to perform multifractal analysis over quantized data.

***Index Terms*—** Fractals, Multifractal Analysis, Quantization, Wavelet Leader, Increments

1. MOTIVATION

Self-similar and multifractal stochastic processes are nowadays commonly used to model or describe real-life data coming from a variety of applications of different natures and possessing some form of scale invariance or scaling property. Empirical scaling analysis usually amounts to measuring scaling exponents that fully characterize the process used for the modeling. In turns, these scaling attributes are involved in standard signal processing tasks, such as detection, identification or classification. Most stochastic models used to describe scaling in real-life data are continuous time and continuous valued processes. However, for most applications, the analyzed data are sampled in time, and the impact of this sampling of the estimation of the multifractal parameters has been analyzed in various articles (cf. e.g., [1, 2, 3, 4]). In a number of situations, the data available for the analysis also present quantization in amplitude. This is very often the case in Image Processing where the necessarily limited sizes of images yield quantized boundary lines (separating various regions in the image). An informative example is provided by the analysis of crack propagations, where the data consist of boundary lines that split images into two binary regions. It is conjectured that the characteristics of the crack propagation can be inferred from the analysis of the scaling properties of these boundary lines. Because it is often needed that a large number of such images are captured along time, this may impose that sensors are used at poor resolution levels, hence resulting into the fact that the boundary lines are available for the analysis only through (possibly severely) quantized versions (cf. [4, 5] for a thorough description of this application).

To the best of our knowledge, the impact of this quantization effect on the statistical performance of the procedures aiming at estimating multifractal attributes received little attention and precisely constitutes the goal of the present contribution.

For this study, we make use of a synthetic multifractal process, commonly referred to as Multifractal Random Walk [6], chosen both for its simplicity and its ability to relevantly model a large class of scaling properties observed in real-life data. The empirical multifractal analysis we use here consists of the estimation of specific multifractal attributes, recently shown to be interesting and referred to as the *log-cumulant* [7]. Also, we compare analysis procedures designed from different multiresolution quantities, increments, wavelet coefficients and wavelet leaders. These latest were very recently shown to be the most relevant quantities multifractal analysis should be based on [8]. Multiresolution quantities, multifractal analysis, log-cumulants and Multifractal Random Walks are introduced and detailed in Section 2. One of the goals of the present contribution consists of studying the robustness against quantization that the choice of a particular multiresolution quantity brings (or not) to empirical multifractal analysis. Quantization and numerical simulations are described in Section 3, while results are reported in Section 4.

2. MULTIFRACTAL ANALYSIS AND PROCESSES

Multiresolution Quantities. Performing the empirical multifractal analysis (EMA), or scaling analysis, of some data X implies first to chose a *multiresolution quantity*, $T_X(a,t)$, i.e., a quantity that depends jointly on the time position t and the analysis scale a. Historically, EMA was based on first order increments [6]. The use of higher order increments as a generalization has also been proposed. Wavelet coefficients are nowadays widely used as standard quantities for EMA [9]. Very recently, new theoretical results showed that EMA has to be based on wavelet leaders instead of wavelet coefficients. Because of their being hierarchical quantities, leaders can be shown to enable a theoretically exact and practically accurate analysis of the multifractal properties for any type of multifractal processes [8]. Such results have never been proven for increments or for wavelet coefficients.

Let us now define the different $T_X(a,t)$. Let $X(t)$, $t \in [0,n)$ denote the process or data under analysis and n its observation duration. Let $\psi_0(t)$ denote a reference pattern with fast exponential decay, called the *mother-wavelet* and $\{\psi_{j,k}(t) = 2^{-j}\psi_0(2^{-j}t - k), j \in \mathcal{Z}, k \in \mathcal{Z}\}$ its templates, dilated to scales 2^j, and translated to time positions $2^j k$. The wavelet coefficients of X are defined as $d_X(j,k) = \langle \psi_{j,k} | X \rangle$. The mother-wavelet is further characterized by its *number of vanishing moments*, a strictly positive integer $N_\psi \geq 1$ defined as: $\forall k = 0, 1, \ldots, N_\psi - 1$, $\int_{\mathcal{R}} t^k \psi_0(t) dt \equiv 0$ and $\int_{\mathcal{R}} t^{N_\psi} \psi_0(t) dt \neq 0$. Let us also introduce the indexing $\lambda_{j,k} =$

$[k2^j,(k+1)2^j)$ and the union $3\lambda_{j,k}=\lambda_{j,k-1}\cup\lambda_{j,k}\cup\lambda_{j,k+1}$. The wavelet leaders are defined as $L_X(j,k)=\sup_{\lambda'\subset 3\lambda_{j,k}}|d_{\lambda'}|$, where the supremum is taken on the discrete wavelet coefficients $d_X(\cdot,\cdot)$ in the time neighborhood $3\lambda_{j,k}$ over all finer scales $2^{j'}<2^j$ [8]. The 4 $T_X(a,t)$ studied here (Increments of orders 1 and 2, wavelet coefficients and leaders) are hence defined as follows, for dyadic analysis scales $a=2^j$ (τ_0 stand for arbitrary units):

$$T_X^{(I_1)}(2^j,t) = X(t+2^j\tau_0)-X(t), \quad (1)$$

$$T_X^{(I_2)}(2^j,t) = X(t+2\cdot 2^j\tau_0)-2X(t+2^j\tau_0)+X(t), \quad (2)$$

$$T_X^{(W)}(2^j,t) = d_X(j,k)=\langle\psi_{j,k}|X\rangle, \quad (3)$$

$$T_X^{(L)}(2^j,t) = L_X(j,k)=\sup_{\lambda'\subset 3\lambda_{j,k}}|d_{\lambda'}|. \quad (4)$$

Empirical Multifractal Analysis: Log-Cumulants. A process X is said to possess *scale invariance* or *scaling* properties if, for some statistical orders q, the time averages of the (q-th power of the modulus of the) $T_X(a,t)$ taken at fixed scales display power law behaviors with respect to scales:

$$S_n(q,a)=\frac{1}{n_a}\sum_{k=1}^{n_a}|T_X(a,ka)|^q\simeq F_q|a|^{\zeta(q)}, \quad (5)$$

over a wide range of scales $a\in[a_m,a_M]$, $a_M/a_m\gg 1$. The $\zeta(q)$ are referred to as the scaling exponents of X and are closely related to its theoretical multifractal spectrum [8, 9]. The $\zeta(q)$ can be naturally expanded as a polynomial,

$$\zeta(q)=\sum_{p\geq 1}c_pq^p/p!, \quad (6)$$

where, for truly multifractal processes, at least c_1 and c_2 differ from 0. It has been proven [7] that the multifractal parameters $c_p, p\geq 1$ can be defined from the p-th order cumulants C_p^j of $\ln|T_X(2^j,t)|$:

$$\forall p\geq 1: \quad C_p^j=c_p^0+c_p\ln 2^j. \quad (7)$$

Thus, the measurements of the scaling exponents $\zeta(q)$ is now often fruitfully replaced by those of the *log-cumulants* c_p.

Estimation procedures. Given n_j coefficients $T_X(2^j,k2^j)$, the asymptotically unbiased and consistent standard estimators are employed to obtain estimates $\hat{C}_p^j$ for the cumulants of $\ln|T_X(2^j,k2^j)|$, where all $|T_X|$ below the threshold 10^{-10} are discarded. The c_p can then be estimated by linear regression (cf. Eq. (7)),

$$\hat{c}_p=\log_2 e\sum_{j=j_1}^{j_2}w_j\hat{C}_p^j. \quad (8)$$

The weights w_j have to satisfy the constraints $\sum_{j_1}^{j_2}jw_j\equiv 1$ and $\sum_{j_1}^{j_2}w_j\equiv 0$.

Multifractal Random Walk. For empirical studies and numerical simulations, we use a multifractal process, Multifractal Random Walk (MRW), chosen for its being simple both from theoretical and numerical synthesis points of view and yet representative for a large class of multifractal processes. MRW has been introduced in [6] as a multifractal (hence non Gaussian) process with stationary increments: $X(k)=\sum_{k=1}^n G_H(k)e^{\omega(k)}$, where $G_H(k)$ consists of the increments of a normalized fractional Brownian motion (FBM) with self-similarity parameter H. The process ω is independent of G_H, Gaussian, with non trivial covariance: $\mathrm{cov}(\omega(k_1),\omega(k_2))=\lambda\ln\left(\frac{L}{|k_1-k_2|+1}\right)$ when $|k_1-k_2|<L$ and 0 otherwise. It has been shown that MRW has interesting scaling properties as in Eq. (5), with $\zeta(q)=(H+\lambda^2)q-\lambda^2q^2/2$. The simplicity of MRW hence lies in its multifractal properties being entirely controlled by the single $c_2=-\lambda^2$.

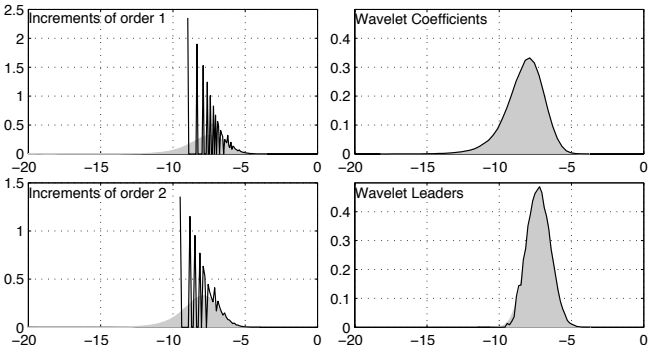


Fig. 1. **Quantization and empirical distributions.** Empirical distribution of $\ln|T_X(2^j,\cdot)|$ ($j=3$) for non quantized (gray area) and quantized, $b=12$, (solid line) signals.

3. QUANTIZATION AND NUMERICAL STUDY

Quantization. The quantized signal is defined as:

$$X^\Delta(n)=[X(n)/\Delta]\cdot\Delta, \quad b=-\log_2\Delta,$$

where $[\cdot]$ denotes the rounding operation, Δ the quantization interval width and b the quantization level (in bit).

Monte Carlo Simulations. The impact of quantization on the estimation procedures is assessed by applying them to a large number N_{MC} of realizations of MRW. The influence of quantization on the performance of the estimation of c_p is assessed through the mean squared error: $\mathrm{MSE}=\sqrt{\left(\widehat{\mathbb{E}}\hat{c}_p-c_p\right)^2+\widehat{\mathrm{Var}}\,\hat{c}_p}$, where $\widehat{\mathbb{E}}$ and $\widehat{\mathrm{Var}}$ denote the sample mean and sample variance over Monte Carlo realizations, respectively. To better understand the impact of quantization, we also study the empirical distributions of $\ln|T_X(j,\cdot)|$, as well as the mean over realizations of $\hat{C}_p^j$.

Simulation Setup. The results are obtained using Daubechies2 wavelets (i.e., with $N_\psi=2$). Simulation parameters are set to $N_{MC}=1000$ and $n=2^{14}$, and process parameters to $(H,\lambda)=(0.72,\sqrt{0.08})$, i.e., $(c_1,c_2)=(0.8,-0.08)$.

4. QUANTIZATION IMPACTS

4.1. Distributions of $\ln|T_X(2^j,\cdot)|$

Fig. 1 compares the empirical distributions of $\ln|T_X(2^j,\cdot)|$ (for a given j) for a non quantized and quantized at level b signals. We observe that while quantization does not have any visible impact on the distributions of $\ln|T_X^{(W)}(2^j,\cdot)|$ and $\ln|T_X^{(L)}(2^j,\cdot)|$, the distributions of $\ln|T_X^{(I_1)}(2^j,\cdot)|$ and $\ln|T_X^{(I_2)}(2^j,\cdot)|$, obtained from quantized data, are lattice and significantly different from the distribution obtained for the non quantized signal. This will, in turns, affect the estimation of the $\hat{C}_p^j$.

4.2. $\hat{C}_p^j$ as linear functions of j

Fig. 2 compares the means over Monte Carlo realizations of $\hat{C}_1^j$ and $\hat{C}_2^j$ as functions of j, for non quantized data and for data quantized at different levels. It yields a central observation: Quantization affects $\hat{C}_p^j$ at fine scales first, and then at coarser and coarser scales as Δ increases (equivalently b decreases). Also, we observe that this impact

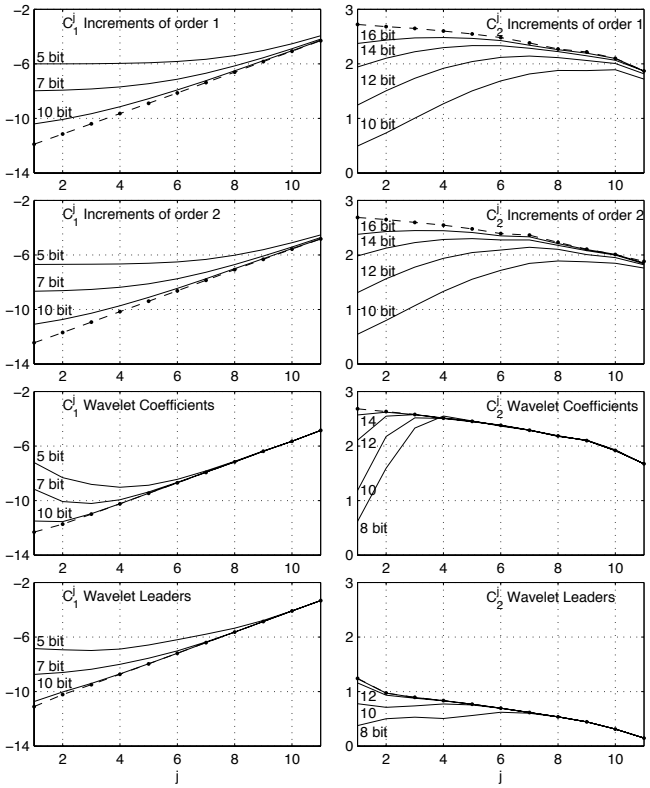


Fig. 2. $\hat{C}_p^j$ **as linear functions of** j**.** $\hat{C}_1^j$ (left column) and $\hat{C}_2^j$ (right column) vs. j for non quantized data (dashed dotted line) and data quantized at different levels b.

is much more dramatic for increments than for wavelet coefficients and leaders. Whereas for increments the influence of quantization propagates very fast up to the coarsest scale as Δ increases, it remains restricted to fine scales for wavelet coefficients and leaders, leaving the coarser scales unchanged and usable to perform the linear regressions yielding $\hat{c}_p$. For wavelet coefficients and leaders, a meaningful range of scales for linear regression can still be found for quantizations significantly below $b = 5$ for $\hat{c}_1$ and $b = 8$ for $\hat{c}_2$, whereas for increments, linear regression is meaningless already for $b = 7$ for $\hat{c}_1$ and $b = 12$ for $\hat{c}_2$.

This can be understood as follows. The increments $T_X^{(I_1)}(2^j, t)$ and $T_X^{(I_2)}(2^j, t)$ can be read as wavelet coefficients obtained with specific mother-wavelets: $\psi_0(t) = \delta(t - \tau_0) - \delta(t)$ and $\psi_0(t) = -\delta(t - 2\tau_0) + 2\delta(t - \tau_0) - \delta(t)$, respectively. Such ψ_0 possess respectively $N_\psi = 1$ and $N_\psi = 2$ vanishing moments and are commonly referred to as *poor man's* wavelets, because they act as band pass filters whose Fourier transforms $\Psi_0(\nu)$ are poorly localized in frequency, compared to those of standard mother wavelets, such as the Daubechies'. This is illustrated in Fig. 3 where the Fourier transforms of the increments are compared to that of a Daubechies2 wavelet. For simplicity the frequency axis is in octaves $j = -\log_2 \nu$. For small frequencies, the behavior of the Fourier transforms is controlled by N_ψ according to $|\Psi_0(\nu)| \sim C \cdot |\nu|^{N_\psi}, |\nu| \to 0$. For large frequencies, $|\Psi_0^{(W)}(\nu)|$ is characterized by a good frequency localization, while $|\Psi_0^{(I_1)}(\nu)|$ and $|\Psi_0^{(I_2)}(\nu)|$ show much poorer frequency localizations with important side lobes whose amplitudes do not decrease. This poor frequency localization turns out to have a

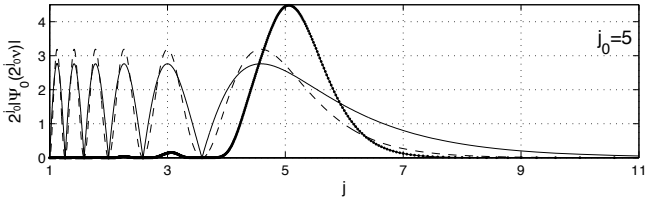


Fig. 3. **Band pass filters.** Fourier transforms of the mother wavelets $|2^{j_0}\Psi_0(2^{j_0}\nu)|$ versus $j = -\log_2(\nu)$ for Daubechies2 (solid dotted line), Increments of first (solid line) and second (dashed line) orders.

significant impact on the robustness of the multifractal parameter estimation procedures against quantization. Indeed, a major consequence of quantization consists of mimicking noise superimposition to the original non quantized data. Fig. 2 suggests that this noise mostly contributes at fine scales, or equivalently, at high frequencies. Hence, any estimates involving such scales are poor whatever the chosen multiresolution quantity. However, the well-localized in frequency nature of the wavelet band pass filters significantly limits the contamination of larger scales by the noise. Therefore, restricting the linear regressions to larger scales yields satisfactory estimates. Conversely, the poor frequency localization of the increment band pass filters results in a significant pollution of the large scales by the fine scale noise. This implies that, to perform estimation, one has to restrict the regression range to much larger scales, if there are any left that are not polluted, which hence significantly degrades the performance.

4.3. Statistical Performance

Increments. $T_X^{(I_1)}(2^j, \cdot)$ and $T_X^{(I_2)}(2^j, \cdot)$ take on fewer and fewer different discrete values as Δ increases, until eventually they only consists of 0 or Δ values. Then, $\hat{C}_1^j = \mathbb{E} \ln |T_X^{(I)}(2^j, \cdot)| \to \ln \Delta$ and $\hat{C}_2^j = \mathrm{Var} \ln |T_X^{(I)}(2^j, \cdot)| \to 0$ (cf. Fig. 2, top row and second row). In turns, the final log-cumulant estimates become $\hat{c}_1 = 0$ and $\hat{c}_2 = 0$, no matter what values c_1 and c_2 actually take. Therefore, the estimations based on $T_X^{(I_1)}$ and $T_X^{(I_2)}$ become meaningless for severely quantized signals. Hence, for large Δ, their statistical performance are no longer discussed. Ultimately, as $\Delta \to \infty$, a similar effect occurs for all T_X however, at significantly higher Δ than for increments.

Fixed regression range. Fig. 4 (top row) compares the MSEs of the estimations of c_1 and c_2, using a fixed regression range, at coarse scales $j_1 = 5$ and $j_2 = 11$. For non quantized signal, we observe that increments and coefficients based estimators achieve comparable performance, whereas the leaders based estimation is better, and significantly so for c_2. When the signal is quantized, the performance of the increments based procedures degrade dramatically and fast when b decreases, whereas the coefficient and leaders based estimations maintain their performance over an impressive range of coarse quantization levels: For c_1, the performance of the increments based procedures start degrading at $b = 15$, while that of coefficients and leaders at $b = 9$, a factor of ≈ 60 in Δ; with a difference in MSE of up to a factor 7 ! For c_2, the situation is similar: The performance of the increments based procedures start degrading as soon as $b = 13$, while that of coefficients and leaders are maintained up to $b = 7$ and $b = 10$, respectively, a factor of ≈ 60 in Δ, with a difference in MSE of up to a factor 10 ! Further, we note that the MSE mainly reproduces standard deviation, apart from at severe

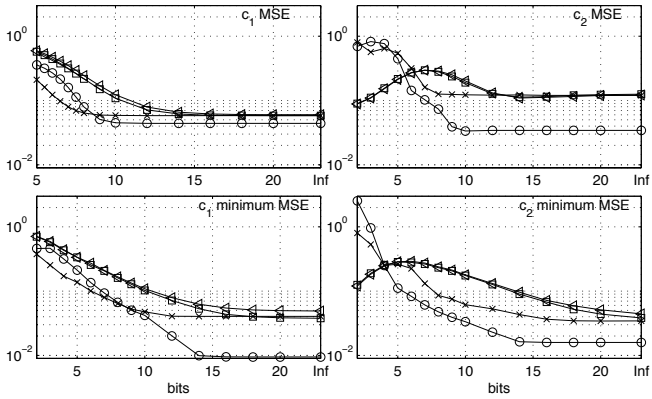


Fig. 4. **Statistical Performance.** MSE of $\hat{c}_1$ (left column) and $\hat{c}_2$ (right column), obtained for a fixed regression range $j_1 = 5$, $j_2 = 11$ (top row), and for the optimal regression range j_1, j_2 (bottom row). The symbols ($\triangleleft$, $\square$, $\times$, $\circ$) correspond to (Increments of order 1, Increments of order 2, Wavelet Coefficients, Wavelet Leaders).

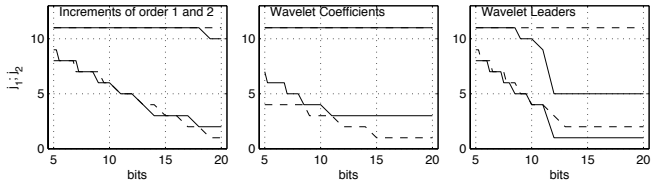


Fig. 5. **Optimal regression range.** Optimal j_1 and j_2 to obtain minimal MSE for $\hat{c}_1$ (solid lines) and $\hat{c}_2$ (dashed lines) vs. b.

quantization levels, where the bias becomes dominant.

Optimal regression range. In practice, the range of scales used to perform the linear regressions yielding the final estimates $\hat{c}_p$ is not fixed a priori but by visual inspection of $\hat{C}_p^j$ vs. j in order to determine a region in which the scaling model is valid. Fig. 4 (bottom row) shows the MSEs obtained by choosing the regression range such that the MSE of the estimation is minimal. Comparing top and bottom rows in Fig. 4, we observe that, as expected, estimation can in general be improved by choosing an appropriate regression range. Whereas for increments, this improvement is small and confined to $b > 14$, relevant estimates of c_1 and c_2 are obtained still at $b = 5$ when using coefficients or leaders, a significant improvement compared to the fixed regression region. We note further that whereas leaders consistently outperform coefficients for sufficiently large b, their performance eventually degrades faster for very heavily quantized signals. Fig. 5 shows the optimal regression ranges. As expected, increasing Δ forces j_1 to increase, restricting the estimation to coarser and coarser scales. This happens much earlier for increments than for coefficients and leaders.

Conclusions. These analyses lead us to conclude that increments can not be used when the data are quantized, even for low quantization levels. We found that coefficients and leaders are significantly more robust to quantization than increments of any order. By choosing an appropriate regression region, the effects of quantization on $\hat{c}_p$ can be circumvented even for coarse Δ levels when using coefficients and leaders, whereas this is not the case for increments.

5. CONCLUSIONS AND PERSPECTIVES

We showed here that signal quantization significantly impairs empirical multifractal analysis. Mostly, it pollutes the finest scales hence implying a restriction towards the largest scales of the range of scales used in the linear regression involved in multifractal parameter estimation. However, we showed that choosing mother wavelets with a good frequency resolution contains the noise pollution to as low as possible scales, hence limiting the necessary narrowing of the regression range and the estimation performance degradation. Conversely, the absence of localization of the frequency response of the increment based band pass filters results in a stronger narrowing of the regression range for a given quantization level and hence in poorer performance. Therefore, wavelet coefficients and leaders are to be preferred to increments of any orders to analyze quantized data. Also, we showed that leaders consistently outperform coefficients for non quantized data as well as for a large range of quantization levels. It is only for very heavily quantized signals that coefficients eventually become more robust than leaders. This study, which, to the best of our knowledge had never been conducted, provides the practitioners with a careful framework for real life data analysis, in situations where quantization occurs, such as the one described in Section 1.

An automatic selection of the most relevant regression range of scales given a quantization level is currently being studied. It can be assisted by the use of promising statistitical techniques such as bootstrap [10]. Also, the impact of quantization will be further studied in image processing in situations where textures are described using multifractal models, but where the amplitudes are quantized.

6. REFERENCES

[1] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motions," *IEEE Trans. on Info. Theory*, vol. 38, pp. 910–917, 1992.

[2] D. Veitch, M. Taqqu, and P. Abry, "Meaningful MRA initialisation for discrete time series," *Signal Processing, Elsevier Science*, vol. 8, pp. 1971–1983, 2001.

[3] S. J. Mitchell, "Discontinuities in self-affine functions lead to multiaffinity," *Phys. Rev. E*, vol. 72, pp. 4065103(R), 2005.

[4] S. Santucci et al., "Gaussian statistics of fracture surfaces," *http://lanl.arxiv.org/cond-mat/pdf/0607/0607372.pdf*, 2006.

[5] N. Mallick, P. P. Cortet, S. Santucci, S. G. Roux, and L. Vanel, "Discrepancy between subcritical and fast rupture roughness: A cumulant analysis," *preprint*, 2006.

[6] E. Bacry, J. Delour, and J.F. Muzy, "Multifractal random walk," *Phys. Rev. E*, vol. 64, pp. 026103, 2001.

[7] J. Delour, J.F. Muzy, and A. Arneodo, "Intermittency of 1d velocity spatial profiles in turbulence: A magnitude cumulant analysis," *The Euro. Phys. Jour. B*, vol. 23, pp. 243–248, 2001.

[8] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus and M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics*. 2004, vol. 72(2), pp. 91–152, AMS.

[9] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data," in *Self-similar Network Traffic and Performance Evaluation,* K. Park and W. Willinger, Eds. spring 2000, Wiley.

[10] A.M. Zoubir and D.R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge University Press, 2004.