New Insights on the Estimation of Scaling Exponents

B. Lashermes[(1)], P. Abry[(1)], P. Chainais[(2)]

[(1)] CNRS UMR 5672, Laboratoire de Physique. ENS Lyon, France,
tel : 33 4 7272 8493, fax : 33 47272 8080,
{pabry, blasherm}@ens-lyon.fr, www.ens-lyon.fr/∼pabry
[(2)] ISIMA-LIMOS UMR 6158 - Université Blaise Pascal, Aubière, France
pchainai@isima.fr, www.isima.fr/∼chainais .

Abstract

Motivation and Position of the Problem. During the last ten years, scaling phenomena and scale invariance have been observed in a wide range of applications of very different nature (hydrodynamic turbulence, computer network teletrafic, body rhythms in biology,... to name but a few). Therefore, their analysis and characterisation received considerable efforts and is still an active research area.

Let $X(t)$ stand for the observed process and let $T_X(a,t)$ denote a quantity that performs some measurement of X at *scale* a (see below for complete definition). Scaling phenomena usually imply a power law behaviour of the moments of $T_X(a,t)$ with respect to scale a, for a given range of scales and a given range of exponents:

$$\mathbb{E}|T_X(a,t)|^q = C_q|a|^{\zeta_q},\ q_m < q < q_M,\ a_m \ll a \ll a_M. \tag{1}$$

The practical study of scaling phenomena mainly consists in the detection of such power laws and in the estimation of the corresponding scaling exponents. The present work is a contribution to this last issue through a study of the statistical performances of various estimators for the ζ_q based on multiresolution analysis. Despite the fact that this question already received considerable efforts, we believe that the results reported here bring insights that are likely to interestingly renew the point of view on that topic.

Processes. To test the performance of the estimators described below, we use first the well known conservative multiplicative Mandelbrot's cascades (CMC) [8] generating the measures $Q(t)$. From $Q(t)$, one also produces the cumulated process $A(t) = \int_0^t Q(u)du$ and the (fractional) Brownian motion in multifractal time $V_H(t) = B_H(A(t))$ (where $B_H(t)$ is the usual fractional Brownian motion) [12]. Such processes however are well know for their two major drawbacks resulting from their being constructed on a rigid dyadic tree: they suffer from non continuous (or discrete) scale invariance and they only have weak forms of stationarity. Therefore, we also use new processes recently introduced in the literature, known as *products of cylindrical pulses* or *compound Poisson processes* (CPC) [4] and *infinitely divisible noise and motion* (IDC) [3, 5]. Their definitions (not detailed) also imply the production of a measure $Q(t)$ and of the corresponding processes $A(t)$ and $V_H(t)$. We also used of the Multifractal Random Walk (MRW) introduced in [2] and that, to some extent, is a process of type $V_H(t)$. All those new processes have controlled scaling multifractal or multiscaling properties, they are also continuously scale invariant and Q and the increment processes of A and V_H meet the full stationarity requirements.

Multiresolution quantities. For the process X, the multiresolution or scale dependent coefficients are defined as for the process X:

$$\begin{array}{lcl} \mathbb{E}|T_X(a,t;f_0)|^q & = & \int_{\mathcal{R}} X(u) f_{a,t}(u)du, \\ f_{a,t}(u) & = & (1/a) f_0((u-t)/a). \end{array} \tag{2}$$

The mother-functions f_0 correspond to the following choices:

$$\begin{array}{llcl} \text{Local Time Average} & f_0(u) & = & 1/\tau_0 1_{\tau_0}(u), (= 1/\tau_0 \text{ if } 0 \le u < \tau_0 \text{ and } = 0 \text{ else}), \\ \text{Increments (of order } N) & f_0(u) & = & \sum_{k=1}^{N} (-1)^k C_N^k \delta(u + (k-1)\tau_0), \\ \text{Wavelet} & f_0(u) & = & \psi_0(u) \text{ Any Mother - Wavelet Function.} \end{array} \tag{3}$$

For the positive conservative measure Q, the three type of $|T_X(a,t;f_0)|$ can be used. For the motions A and V_H only the two lasts apply. Note that the order N of the increments plays a role identical to that of the number of vanishing moments in the wavelet based analysis.

Estimators. The usual partition functions are computed:

$$S_n(q,j) = \frac{1}{n_j} \sum_{k=1}^{n_j} |T_X(2^j, k2^j; f_0)|^q \tag{4}$$

where n is the observation length and n_j the number of available samples $T_X(2^j, k2^j; f_0)$ at scale $a = 2^j$. The estimators $\hat{\zeta}_{q,n}$ of the ζ_q result from linear fits in $\log_2 2^j = j$ versus $log_2 S_n(q,j)$ plots [1]. In all cases the range of q used, $q_m < q < q_M$, is such that the moments are finite: $\mathbb{E}|T_X(a,t;f_0)|^q < \infty$, so that we do not address here the issue of existence of moments, nicely tackled in e.g., [6].

Analysis of the Statistical Performance of the Estimators. To study the statistical performance of the estimators described above $\hat{\zeta}_{q,n}$ as functions of q and n, we performed numerical simulations. For each of the category of the processes mentioned above (CMC, CPC, IDC, MRW), and for Q, A and V_H, we numerically synthetized *nbreal* copies of length n and applied to each all the relevant estimators. From this, we studied the statistical performance of the estimators. For sake of simplicity, in this abstract, we give results only for processes whose theoretical multifractal spectrum takes on negative values both on the right and left hand sides.

Results. The results presented in Figure 1 are obtained from wavelet based estimators (Daubechies3 wavelet) applied to $nbreal = 1000$ trials of cumulated processes of Log-Normal binomial cascades (CMC) of various length $2^{14} \le n \le 2^{21}$. We emphasize that plots yielding equivalent analysis, comments and conclusions, were obtained using local time averages or increment based estimators applied to CMC with different multipliers as well as to CPC, MRW or IDC, as well as to various actual empirical time series. The *generality* of the observation reported here constitute a key point of our contribution.

Critical q^* and h^*. First, we observe that there exists a critical value of q, labeled q^*, beyond which the estimators $\hat{\zeta}_{q,n}$ *do not work correctly* any more. For $q > q^*$ values, the $\hat{\zeta}_{q,n}$ do not reproduce any longer the ζ_q behaviour, as would be expected a priori but instead follow a linear behaviour. This is clearly seen in Figure 1 (first row). Precisely, this means that, asymptotically, i.e., in the limit of infinite length observation:

$$\begin{array}{llll} q < q^*, & \hat{\zeta}_{q,n} & \rightarrow & \zeta_q, \\ q > q^*, & \hat{\zeta}_{q,n} & \rightarrow & 1 + qh^*. \end{array} \tag{5}$$

Length of observation. Second, we observe that in the limit of large n,

$$\begin{array}{llll} \text{for } q < q^*/2 & \hat{\zeta}_{q,n} & \rightarrow & \zeta_q, \\ & \operatorname{Var}\hat{\zeta}_{q,n} & \sim & C_q/n, \\ \text{for } q^*/2 < q < q^*, & \hat{\zeta}_{q,n} & \rightarrow & \zeta_q, \\ & \operatorname{Var}\hat{\zeta}_{q,n} & \sim & \text{decrease with } n \text{ but much slower than } 1/n, \\ \text{for } q^* < q, & \hat{\zeta}_{q,n} & \rightarrow & 1 + qh^*, \\ & \operatorname{Var}\hat{\zeta}_{q,n} & \sim & \text{decrease with } n \text{ but much slower than } 1/n. \end{array} \tag{6}$$

Those observations are illustrated in Figure 1 (second row).
Determination of h^* and q^*. Thirdly, we observe that h^* and q^* can be obtained from the Legendre transform $D(h)$ of ζ_q seen as a function of q:

$$D(h) = \min{}_q (qh - \zeta_q) + 1, \tag{7}$$

or, equivalently, for all q where ζ_q is defined,

$$\begin{array}{lll} D(q) & = & 1 + q d\zeta_q/dq - \zeta_q, \\ h(q) & = & d\zeta_q/dq. \end{array} \tag{8}$$

Then, h^* and q^* are obtained from:

$$\begin{array}{lll} D(h^*) & = & 0, \\ h^* & = & (d\zeta_q/dq)_{q=q^*}. \end{array} \tag{9}$$

Those observations are illustrated in Figure 1 (last row). Note that for processes where $D(h) > 0$ for all h, this means that q^* is infinite and h^* undefined. Those cases will be detailled in the extended version.

Interpretation and comments. In the multifractal formalism, $D(h)$ is the Hausdorff dimension of the support of the set of points where the local regularity of the analysed process takes the value h [12]. The equations above (Equation 9) can be read as the fact that one cannot expect to estimate $D(h)$ when D takes on negative values.
From the literature, we found out that such facts had already been reported in [7, 11, 9, 10] for the cases of Mandelbrot's cascades and for local time averages estimates. The definition of q^* given in [11] can be reformulated as in the equations above. Therefore, the results proposed here can be seen as an extension of the results presented in [11], first, to other class of processes (CPC, MRW, IDC), second, to other types of scaling exponents estimators (increments and wavelets). In [11], the existence of a critical q^* is deeply rooted and connected to the issue of non degeneracy of q-th powers of the measure $Q(t)$ obtained as an infinite product of positive multipliers (martingales). From the theoretical construction of the CPC and IDC processes (cf. [4, 5, 3]), we know that this issue of non degeneracy is also relevant since they as well constitute martingales. The fact that the existence of this critical q^* applies for estimators based on increments or wavelet coefficients (and not only to local averages) shows that it is

deeply rooted in the construction of the process itself and not in the estimation details.

Estimation of h^* and q^*. In the extended version, a procedure will be given which allows, from a single observation of finite length, to practically estimate h^* and q^* (since Equation (9) cannot practically be used because it requires the exact knowledge of ζ_q). Numerical simulations validate the performance of this procedure.

Mono versus Multi-Fractality. Besides that, it has often been proposed in various works (including ours) to test mono versus multi-fractality by testing whether the estimated $\hat{\zeta}_{q,n}$ have a linear behaviour in q or not. From these results, we see that it seems to be a deadend track, since the $\hat{\zeta}_{q,n}$ always follow a linear behaviour. Instead, we propose to discriminate between mono versus multi-fractality by the analysis of the intercept at the origin of the asymptotic linear fit obtained from the $\hat{\zeta}_{q,n}$ for $q \gg q^*$. This will be detailed in the extended version.

References

[1] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch. *Self-Similar Network Traffic and Performance Evaluation*, chapter Wavelets for the analysis, estimation, and synthesis of scaling data. Wiley, 2000.

[2] E. Bacry, J. Delour, and J.F. Muzy. Multifractal random walk. *Phys. Rev. E*, 64:026103, 2001.

[3] E. Bacry and J.F. Muzy. Multifractal stationary random measrues and multifractal random walks with log-infinitely divisible scaling laws. *Phys. Rev. E*, 66, 2002.

[4] J. Barral and B. Mandelbrot. Multiplicative products of cylindrical pulses. *Probab. Theory Relat. Fields*, 124:409–430, 2002.

[5] P. Chainais, R. Riedi, and P. Abry. Scale invariant infinitely divisible cascades. In *Int. Symp. on Physics in Signal and Image Processing, Grenoble, France*, 2003.

[6] P. Gonçalvès and Riedi R. Diverging moments and parameter estimation. *preprint*.

[7] V.K. Gupta and E.C. Waymire. A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, 32(2):251–267, Feb. 1993.

[8] B. B. Mandelbrot. Intermittent turbulence in self similar cascades: Divergence of high moments and dimension of the carrier. *J. Fluid. Mech.*, 62:331, 1974.

[9] G.M. Molchan. Scaling exponents and multifractal dimensions for independent random cascades. *Comm. Math. Phys.*, 179, 1996.

[10] G.M. Molchan. Turbulent cascades: limitations and a statistical test of the lognormal hypothesis. *Phys. Fluids*, 9(8):2387–2396, 1997.

[11] M. Ossiander and E.C. Waymire. Statistical estimation for multiplicative cascades. *Ann. Statistics*, 28(6):1533–1560, 2000.

[12] R. Riedi. *Lois d'échelle, fractales et ondelettes, vol. 2,* Editeurs : P. Abry, P. Gonçalvès, J. Lévy Vehel, chapter Lois d'échelles multifractales : fondements et approche par ondelettes. Lavoisier, 2002.

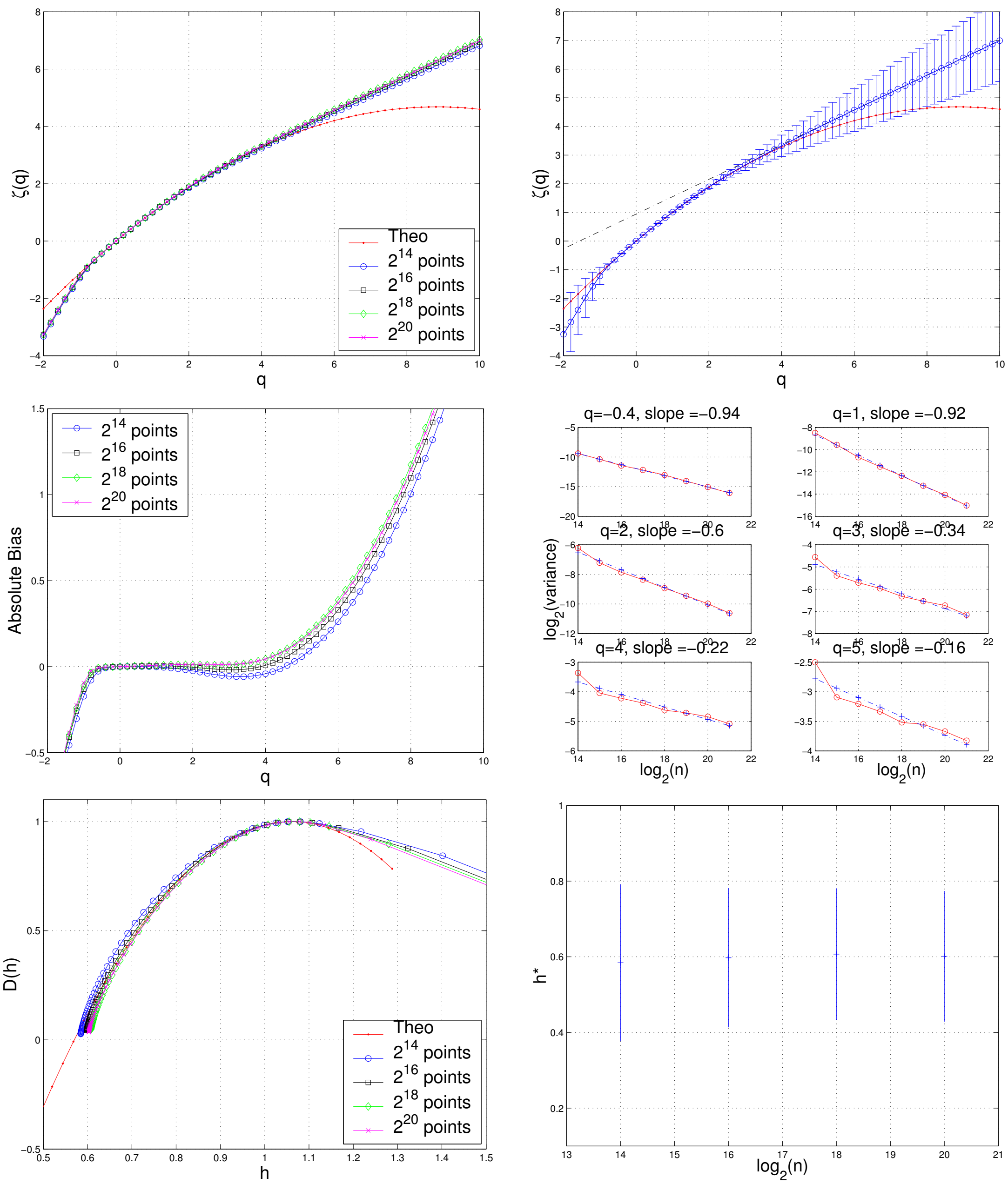

Figure 1: **Statistical behaviour of the estimators: Top Row:** left plot, $\hat{\zeta}_{q,n}$ versus ζ_q. Right plot, $n = 2^{20}$, with 95% Confidence Intervals, obtained fro $nbreal = 1000$ realizations. One clearly sees that for large q, the $\hat{\zeta}_{q,n}$ depart from the ζ_q to follow on a straight line that well matches the equation: $\hat{\zeta}_{q,n} = 1 + qh^*$. **Middle Row:** Left, Biais $\hat{\zeta}_{q,n} - \zeta_q$ as a function of q, for different n. Above a critical q^* (here $q^* = 4.08$ and $h^* = 0.57$), the bias increases with q regardless of n ; right, the (log of the) variance of $\hat{\zeta}_{q,n}$ as a function of n for different q. For $q \leq q^*/2$, the variances essentially decrease as $1/n$, while they decrease significantly more slowly for $q \geq q^*/2$. **Bottom Row:** Left, Legendre transforms $D(h)$ of the $\hat{\zeta}_q$ and of the ζ_q for different n ; Right, estimated h^* (such that $D(h^*) = 0$) as a function of n. One sees that the Legendre transforms of the estimated exponents $\hat{\zeta}_q$ accumulate around the critical values $D(h^*) = 0$ and the corresponding h^* do not depend on n.