

# Contributions à l'étude des performances statistiques des estimateurs multifractals

Stéphane G. ROUX<sup>1</sup>, Bruno LASHERMES<sup>1</sup>, Patrice ABRY<sup>1</sup>, Stéphane JAFFARD<sup>2</sup>

<sup>1</sup>Laboratoire de Physique, CNRS UMR 5672  
École Normale Supérieure de Lyon 46, allée d'Italie, 69364 Lyon, France

<sup>2</sup>CNRS UMR, dépt. d'Analyse et de Mathématiques Appliquées,  
Université Paris XII, créteil, France.

stephane.roux@ens-lyon.fr, bruno.lashermes@ens-lyon.fr  
patrice.abry@ens-lyon.fr, jaffard@univ-paris12.fr

**Résumé** – Nous nous intéressons à différents estimateurs des exposants des lois d'échelle, construits sur des analyses en ondelette. L'un de ces estimateurs est notamment défini à partir d'une technique récente d'analyse multifractale, dite des *coefficients leaders*. Nous comparons leurs performances statistiques sur un processus multifractal particulier, la marche aléatoire multifractale, choisi pour son caractère simple et représentatif d'une large classe de multifractals. Nous mettons en évidence que l'erreur quadratique moyenne des estimateurs prend une forme générale qui ne dépend que de l'écart à la monofractalité du processus étudié. Nous illustrons l'intérêt de l'estimateur leaders.

**Abstract** – We consider various wavelet based estimators for scaling exponents of multifractal processes. One of these estimators is based on a new multifractal formalism the wavelet leaders. Their statistical performance are studied using a specific multifractal process, the Multifractal Random Walk, chosen for its being representative though simple of a large class of multifractals. We show that the mean square error presents a general shape which depends only on the departure of the process from monofractality. We show also the relevance of the wavelet leaders.

## 1 Motivations

Au cours de cette dernière décennie, des comportements en loi d'échelle, des phénomènes d'invariance d'échelle ont été observés dans de nombreuses applications de natures très différentes. Pour appréhender l'invariance d'échelle, on remplace couramment le signal  $X(t)$  à étudier par des coefficients multirésolution  $T_X(a, t)$  qui dépendent à la fois du temps  $t$  et de l'échelle d'analyse  $a$ . On définit alors l'invariance d'échelle à travers le comportement en loi de puissance en fonction de  $a$  des puissances  $q$ -èmes d'une fonctionnelle des  $T_X(a, t)$  :

$$S(a, q) = \frac{1}{n_a} \sum_{k=1}^{n_a} F(|T_X(a, t(k))|)^q = c_q |a|^{\zeta(q)}, \quad (1)$$

où  $n_a$  désigne le nombre de coefficients à l'échelle  $a$ . Les  $T_X(a, t)$  peuvent en particulier être les coefficients d'ondelettes ou les accroissements de  $X$ .

L'analyse pratique des phénomènes d'invariance d'échelle sur des données expérimentales, consiste à mettre en évidence ces lois de puissance et à mesurer les exposants  $\zeta(q)$  correspondants. Souvent, cette estimation est réalisée à partir de la mesure de la pente d'une droite observée dans le diagramme  $\ln S(a, q)$  versus  $\ln a$ . Alors que les performances statistiques de tels estimateurs ont été largement étudiées pour les processus autosimilaires [1, 2], celles-ci sont bien moins connues pour les processus multifractals, notamment pour ceux dont les exposants  $\zeta(q)$  suivent une

loi non linéaire en  $q$ .

Dans cet article, nous nous concentrons sur quatre déclinaisons d'estimateurs *ondelette* d'exposants  $\zeta(q)$ . Nous étudions leurs performances statistiques à partir d'une mise en œuvre sur un processus multifractal particulier, la *Marche Aléatoire Multifractale* [3]. Ce dernier est choisi à la fois pour son caractère représentatif d'une large classe de multifractals et pour son caractère simple (ses propriétés multifractales sont définies par un nombre réduit de paramètres) et facile à simuler numériquement.

## 2 Ondelettes versus leaders

Les coefficients de la transformée en ondelettes continue (TOC) résultent de la comparaison par produits scalaires  $T_X(a, t)$  de  $X$  contre les dilatées et translatées d'une fonction  $\psi_0$ , appelée ondelette-mère :

$$T_X(a, t) = \int X(u) |a|^{-1} \psi_0(a^{-1}(u-t)) du. \quad (2)$$

L'ondelette mère est principalement caractérisée par un entier  $N \geq 1$  appelé le nombre de moments nuls défini par  $\int t^k \psi_0(t) dt \equiv 0, \forall k = 0, 1, \dots, N-1$ .

Les coefficients de la transformée en ondelettes discrète (TOD) consistent essentiellement en une restriction de ceux de la TOC à une grille dyadique :

$$d_X(j, k) = T_X(2^j, k2^j). \quad (3)$$

Cette transformée bénéficie notamment d'un coup de calcul significativement plus faible que la précédente. Pour une présentation plus détaillée des transformées en ondelettes, le lecteur est renvoyé à [4].

On définit ensuite une version squelettisée de la TOC, les maxima du module de la transformée en ondelette, ou TOMM, développée par A. Arnéodo et ses collaborateurs, (voir par exemple [5]). Pour cela, on extrait, à chaque échelle les maxima du module de la transformée en ondelettes. De manière récursive et en partant de la plus petite échelle, on relie chaque maximum obtenu à cette échelle au maximum le plus proche de l'échelle supérieure. On obtient ainsi un ensemble de lignes de maxima, un *squelette*, noté  $\mathcal{L}$ . Sur ce squelette, on remplace ensuite les  $T_X(a, t) \in \mathcal{L}$  par leurs maxima, observés à une échelle plus fine sur la même ligne du squelette :

$$M_X(a, t) = \sup_{(t, a') \in I, a' < a} |T_X(a', t)|. \quad (4)$$

À partir de la TOD a été très récemment proposée une analyse dite des coefficients *leaders*, TOL, [6]. Ces coefficients *leaders* se définissent par :

$$L_X(j, k) = \sup \left\{ \bigcup_{0 \leq j' \leq j} \left\{ \bigcup_{m=m_1}^{m_2} \{|d(j, m)|\} \right\} \right\}, \quad (5)$$

avec  $m_1 = 2^{j-j'}(k-2)+1$ ,  $m_2 = 2^{j-j'}(k+1)$  et  $L_X(0, \cdot) = 0$ . Cela signifie simplement que le coefficient leader  $L_X(j, k)$  est obtenu comme le maximum dans un voisinage local de tous les  $d_X(j, k)$  situés à des échelles inférieures. Le voisinage temporel local est défini par les deux coefficients adjacents à celui considéré (cf. figure 1).

Dans la suite, nous utiliserons ces quatre ensembles de coefficients dans l'équation 1.

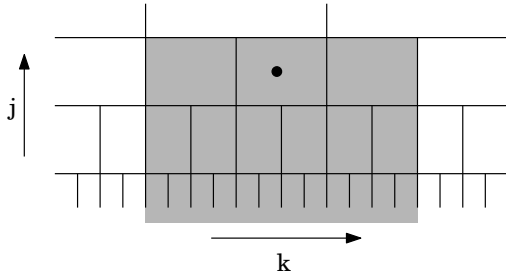


Figure 1: **Coefficient leader** : Sur la grille dyadique correspondant à la TOD, le coefficient leader matérialisé par le point est obtenu en prenant le maximum du module des  $d_X(j, k)$  localisés dans la zone grisée et ce jusqu'à la plus petite échelle.

### 3 Marche Aléatoire Multifractale

**Définition.** La Marche Aléatoire Multifractale (MAM) a été initialement définie dans [3] selon :

$$X(k) = \sum_{k=1}^n G_H(k) e^{\omega(k)}. \quad (6)$$

Les processus  $G_H(k)$  et  $\omega(k)$  sont indépendants,  $G_H(k)$  désigne le bruit gaussien fractionnaire, de paramètre  $H$  et  $\omega(k)$  un bruit gaussien corrélé selon :

$$\text{cov}(\omega(k_1), \omega(k_2)) = c_2 \ln(L/(|k_1 - k_2| + 1)), \quad (7)$$

si  $|k_1 - k_2| < L$  et 0 sinon. Ce processus est donc entièrement défini par trois paramètres  $H$ ,  $c_2$  et  $L$ , ce dernier définit l'*échelle intégrale*. Par ailleurs, la synthèse pratique des processus implique nécessairement une résolution finie, au moins due au caractère échantillonné, éventuellement à certains détails de la construction (cf. [7]). Celle-ci sera notée ici  $a_R$  et prise arbitrairement à 1 sans caractère restrictif. En réalité, seul le rapport  $L/a_R$  importe. Enfin, ce processus est, par construction, non stationnaire mais à accroissements stationnaires.

**Invariance d'échelle.** Les propriétés d'invariance d'échelle de ce processus peuvent être rassemblées dans les relations suivantes. Pour la TOC et la TOD, pour  $q > -1$  et pour  $1 \leq a = 2^j \leq L$ , nous avons :

$$\mathbb{E}|T_X(a, t)|^q = B_q a^{\zeta(q)}, \quad (8)$$

$$\mathbb{E}|d_X(j, k)|^q = B_q 2^{j\zeta(q)}, \quad (9)$$

avec

$$\zeta(q) = (H + c_2)q - c_2 q^2/2. \quad (10)$$

Cela indique que l'écart au comportement linéaire en  $q$  de la fonction  $\zeta(q)$  est entièrement caractérisé par le seul paramètre  $c_2$ . De plus, il est possible de montrer que pour la TOMM et la TOL, pour  $1 \leq a = 2^j \leq L$  et pour  $q \in [q_*^-, q_*^+]$ , nous avons :

$$\frac{1}{n_a} \sum_{k=1}^{n_a} |T_X(a, k)|^q = B_q a^{\zeta(q)}, \quad q \in [0, q_*^+],$$

$$\frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(2^j, k)|^q = B_q 2^{j\zeta(q)}, \quad q \in [0, q_*^+],$$

$$\frac{1}{n_a} \sum_{k=1}^{n_a} M_X(a, k)^q = B_q a^{\zeta(q)}, \quad q \in [-q_*^-, q_*^+],$$

$$\frac{1}{n_j} \sum_{k=1}^{n_j} L_X(2^j, k)^q = B_q 2^{j\zeta(q)}, \quad q \in [-q_*^-, q_*^+].$$

Il est essentiel de noter que la limitation de la validité de ces lois de puissance à une gamme restreinte de valeurs de  $q$  constitue une propriété fondamentale et générale des processus multifractals (cf. [8]). La *Marche Aléatoire Multifractale* correspond au cas particulièrement simple où :  $q_*^+ = -q_*^- = \sqrt{2/c_2} = q_*$ .

**Extension.** La marche aléatoire peut être vue comme un cas particulier d'une construction plus générale de processus infiniment divisibles multifractals (voir, par exemple, [7, 9]).

### 4 Estimateurs

**Définitions.** Nous allons considérer quatre estimateurs construits sur les quatre coefficients multirésolution définis précédemment. Nous noterons donc  $F_X(a_j, k)$  pour

$|T_X(a_j, k)|, |d_X(j, k)|, M_X(a_j, k), L_X(j, k)$  respectivement. Les estimateurs sont définis par régressions linéaires du logarithme de  $S(a_j, q) = 1/n_j \sum_{k=1}^{n_j} F_X(a_j, k)^q$  en fonction de  $\log_2 a_j$  par :

$$\hat{\zeta}(q) = \sum_{j=j_1}^{j_2} w_j \log_2 S(a_j, q), \quad (11)$$

où les poids  $w_j$  prennent la forme

$$w_j = (S_0 x_j - S_1) / (\alpha_j (S_0 S_2 - S_1^2)), \quad (12)$$

avec  $S_p = \sum_{j=j_1}^{j_2} x_j^p / \alpha_j$ ,  $p = 0, 1, 2$ . Ici, les  $\alpha_j$  qui quantifient la confiance accordée à chaque  $\log_2 S(a_j, q)$  sont pris tous égaux à 1. Pour les TOD et TOL, les  $a_j$  prennent naturellement les valeurs  $a_j = 2^j$ . On utilise pour TOC et TOD,  $q \in [0, q_*]$ , pour TOMM, TOL  $q \in [-q_*, q_*]$ . Nous n'aborderons pas ici la question de la sélection de la gamme d'échelle. Elle est supposée fixe et connue a priori.

**Commentaires.** Le défaut des estimateurs TOC et TOD vis-à-vis d'une analyse multifractale correcte est de ne pas offrir de solutions ni pratique ni théorique à l'estimation des  $\zeta(q)$  pour la gamme de  $q$  négatifs,  $q \in [q_*, 0]$ . En effet, d'une part, cette procédure serait instable pour le calcul de puissances  $q$ -èmes négatives de coefficients d'ondelettes proches de 0 ; d'autre part, elle n'est soutenue par aucun résultat théorique. Il est cependant connu qu'une étude pertinente des propriétés multifractales de  $X$  doit passer par l'étude des  $q$  négatifs. C'est historiquement l'une des raisons qui a présidé à la définition de la TOMM (cf. [5]). Dans cette approche, il est essentiel de noter que la somme  $S(a, q)$  est prise sur les lignes de maxima appartenant au squelette seulement, ( $l \in \mathcal{L}$ ), donc sur un ensemble de points restreints. L'intérêt et la pertinence de cette méthode, développée par A. Arneodo et ses collaborateurs, ont été établis et éprouvés dans une abondante littérature [5]. L'intérêt et la pertinence de la TOL pour l'analyse multifractale ont été établis dans [6] et discuté et illustré dans [10]. Elle présente aussi l'avantage d'avoir un coût de calcul très inférieur à la TOMM et de se généraliser facilement, théoriquement et pratiquement, à des dimensions supérieures.

## 5 Performances des estimateurs

### 5.1 Méthodologie.

Pour étudier les performances statistiques de ces estimateurs, nous procédons par simulations numériques. Nous fabriquons par synthèse numérique un grand nombre de réalisations  $n_{real}$  de MAM de longueur  $n * n_L$ , où  $n$  est le nombre de points par échelle intégrale  $L$  et  $n_L$  le nombre d'échelles intégrales. On applique chacun des quatre estimateurs ci-dessus à chacune des réalisations et on obtient une estimation  $\hat{\zeta}_l^i(q)$ ,  $i = 1, 2, 3, 4$ ;  $l = 1, \dots, n_{real}$ . En moyennant sur les réalisations, on peut mesurer la valeur moyenne  $\langle \hat{\zeta}^i(q) \rangle_l$  et la variance  $\langle \langle \hat{\zeta}^i(q) \rangle_l \rangle_l$  des estimateurs  $\hat{\zeta}^i(q)$ . On peut alors calculer l'erreur quadratique moyenne :  $E_q^i = (\langle \langle \hat{\zeta}^i(q) \rangle_l - \zeta(q) \rangle_l^2 + \langle \langle \hat{\zeta}^i(q) \rangle_l \rangle_l)^{1/2}$ . Ici,

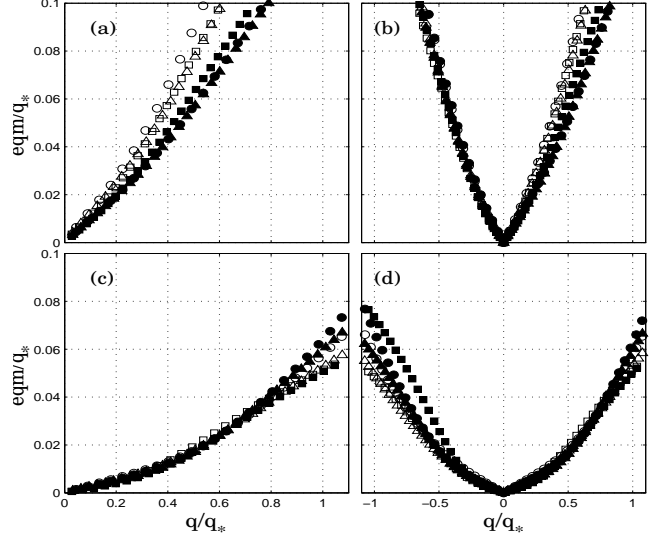


Figure 2: **Erreur quadratique moyenne  $E_q^i/q_*$  en fonction de  $q/q_*$ .** Pour  $H = 0.8$  et  $c_2 = 0.1$  ( $\circ, \bullet$ ),  $c_2 = 0.05$  ( $\square, \blacksquare$ ) et  $c_2 = 0.01$  ( $\triangle, \blacktriangle$ ); pour la TOC (symboles pleins) et la TOD (symboles vides) à gauche et la TOMM (symboles pleins) et la TOL (symboles vides) à droite. Les graphes du haut correspondent à  $(n, n_L) = (1024, 2)$ , ceux du bas à  $(1024, 64)$ .

la résolution est fixée et on fait varier la durée des signaux de deux façons : soit, à échelle intégrale fixée, on augmente le nombre d'échelles intégrales  $n_L$  disponibles à l'analyse ; soit, à nombre d'échelles intégrales fixé, on augmente l'échelle intégrale  $L$ . Nous présentons des résultats pour  $nb_{real} = 500$ ,  $(n, n_L) = (2^{12}, 2)$ ,  $(2^{12}, 2^6)$ ,  $(2^{16}, 2)$  et pour les couples  $(H, c_2)$  avec  $H = 0.5, 0.8$  et  $c_2 = 0.1, 0.05$  et  $0.01$ . Nous avons choisi  $20a_R \leq a_j \leq L/4$ . Les ondelettes utilisées sont des Daubechies2 pour TOD et TOL et des dérivées secondes de gaussienne pour TOC et TOMM. Toutes les procédures de synthèse et d'estimation ont été implantées par nos soins en MATLAB et/ou C et sont accessibles sur demande.

### 5.2 Résultats

**Forme universelle.** Pour chacun des quatre estimateurs, les figures 2 et 3 présentent les courbes  $E_q^i/q_*$  en fonction de  $q/q_*^+$  pour les différents jeux de paramètres  $H, c_2$ . Le premier résultat obtenu réside dans le fait que, à  $H$  fixé, les courbes :

$$E_q^i/q_* = f_{n, n_L}^i(H, q/q_*), \quad (13)$$

se superposent les unes aux autres quelle que soient les valeurs de  $c_2$  (ou de façon équivalente de  $q_*^+$ ). En d'autres termes, les courbes décrivant l'erreur quadratique moyenne des estimateurs sont homothétiques les unes des autres à condition d'être renormalisées par  $q_*^+$ , qui rend compte de l'écart à la monofractalité. Ceci est valable pour les quatre estimateurs étudiés. Une comparaison des figures 2 et 3 montre de plus que les fonctions  $f_{n, n_L}^i$  semblent ne dépendre que très faiblement de la valeur du paramètre  $H$ . Enfin, l'étude des  $f_{n, n_L}^i$  indique que  $f_{n, n_L}^i \sim$

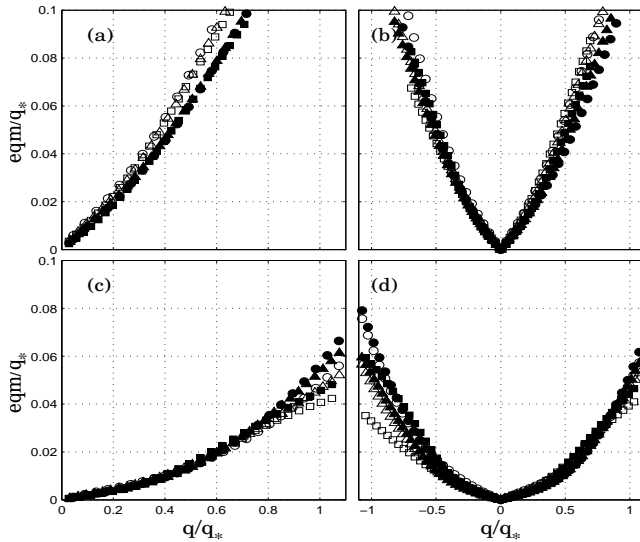


Figure 3: **Erreur quadratique moyenne  $E_q^i/q_*$  en fonction de  $q/q_*$ .** Pour  $H = 0.5$  et  $c_2 = 0.05$  ( $\circ, \bullet$ ),  $c_2 = 0.1$  ( $\square, \blacksquare$ ),  $c_2 = 0.01$  ( $\triangle, \blacktriangle$ ); pour la TOC (symboles pleins) et la TOD (symboles vides) à gauche et la TOMM (symboles pleins) et la TOL (symboles vides) à droite. Les graphes du haut correspondent à  $(n, n_L) = (1024, 2)$ , ceux du bas à  $(1024, 64)$ .

$\frac{1}{\sqrt{n_L}}(C^i(n))^2(q/q_*^+)^2$ , où les  $C^i(n)$  décroissent avec  $n$ .

**Performances comparées.** En termes de comparaisons des quatre estimateurs étudiés, nous observons que :  
– TOMM et TOL donnent systématiquement de meilleures estimations que TOC et TOD, et ce, quelle que soit  $n \cdot n_L$  ;  
– pour  $q > 0$ , et pour des signaux de faible longueur, les performances de TOC et TOMM sont meilleures que celles de TOD et TOL, respectivement. Lorsque la durée des signaux augmente, cet écart disparaît rapidement ;  
– pour  $q < 0$ , les performances d’estimation de TOL sont sensiblement meilleures que celles de TOMM. Cette différence semble augmenter quand la taille des signaux augmente.

De plus, il faut noter que les quatre estimateurs étudiés présentent tous un biais systématique. Pour  $H$  fixé, ce biais augmente avec  $c_2$ . Ce biais est en général moindre pour TOL et TOMM que pour TOD et TOC (courbes non présentées).

## 6 Conclusions et Perspectives

Nous avons étudié quatre estimateurs *ondelette* pour les exposants de loi d’échelle sur un processus multifractal représentatif, la Marche Aléatoire Multifractale. Nous montrons que les performances statistiques de TOL et TOMM sont systématiquement meilleures que celles de TOC et TOD. Notre étude indique aussi que TOMM est préférable pour les signaux courts et les  $q > 0$ , alors que la méthode TOL présente de meilleures performances pour les  $q < 0$  et ce quelles que soient la taille des signaux et les valeurs des paramètres les définissant. Il faut également noter que la méthode TOL a un coût de calcul signi-

ficativement plus faible que celui de la TOMM, ce qui renforce l’intérêt de leur utilisation quand la durée des signaux augmente. De plus, la méthode TOL présente l’avantage d’être aisément généralisable aux dimensions supérieures, contrairement à la TOMM. Nous montrons que pour le cas particulier de la MAM, les courbes de l’erreur quadratique moyenne prennent une forme générale à condition d’être correctement renormalisées à l’aide du paramètre  $q_*^+ = \sqrt{2/c_2}$  qui rend compte de l’écart à la monofractalité. Il convient maintenant de conforter la portée générale des résultats obtenus ici dans le cas pédagogique de la MAM à de plus larges classes de processus multifractals, notamment à des processus pour lesquels  $-q_*^- \neq q_*^+$  ou pour lesquels l’écart à la monofractalité n’est pas contrôlé par un seul paramètre. Ce travail est en cours de réalisation.

## References

- [1] D. Veitch, P. Abry, A wavelet-based joint estimator of the parameters of long-range dependence, *IEEE Trans. on Info. Theory*, 45 (3) : 878–897, 1999.
- [2] B. Audit, E. Bacry, J.F. Muzy, A. Arneodo, Wavelet-based estimators of scaling behavior, *IEEE Trans. Inform. Theory*, 48 (11), 2938–2954, 2002.
- [3] E. Bacry, J. Delour, J.F. Muzy, Multifractal Random Walk, *Phys. Res. E*, 64: 026103, 2001.
- [4] S. Mallat, A Wavelet Tour of Signal Processing *Academic Press*, San Diego, CA, 1998.
- [5] J.F. Muzy, E. Bacry, A. Arneodo, The multifractal formalism revisited with wavelets, *Int. Journal of Bifurcation and Chaos*, 4(2): 245–302, 1994.
- [6] S. Jaffard, Wavelet Techniques in multifractal analysis *Fractal Geometry and applications : a jubilee of Benoit Mandelbrot*, eds. M. Lapidus et M. van Frankenhuysen, *Proc. of Symp. in Pure Mathematics*, 2004.
- [7] P. Chainais, R. Riedi, P. Abry, On Non Scale Invariant Infinitely Divisible Cascades, *IEEE trans. on Info. Theory*, à paraître, 2005.
- [8] B. Lashermes, P. Abry, P. Chainais, New Insights into the Estimation of Scaling Exponents, “*Wavelets in Statistics*” *International Journal of Wavelets, Multiresolution and Information Processing*, 2(4): 497–523, 2004.
- [9] J.F. Muzy, E. Bacry, A. Arneodo, Multifractal stationary random measures and multifractal random walks with log-infinitely divisible scaling laws *Phys. Rev. E*, 66, 2002.
- [10] P. Abry, S. Jaffard, B. Lashermes, Revisiting Scaling, Multifractal, and Multiplicative Cascades with the Wavelet Leader Lens. *Optic East, Wavelet Applications in Industrial Applications II*, Vol. 5607:103–117, Philadelphia, USA, 2004.