A Robust Model and its EM Algorithm for the Estimation of the Multifractality Parameter

Lorena Leon, Herwig Wendt, Jean-Yves Tourneret
Institut de Recherche en Informatique de Toulouse (IRIT),
Univ. de Toulouse, CNRS, Toulouse INP, UT3, UT2.
firstname.lastname@irit.fr

Patrice Abry
Univ. Lyon, ENS de Lyon, Univ. Claude Bernard,
CNRS, Laboratoire de Physique, Lyon, France.
patrice.abry@ens-lyon.fr

***Abstract*—Multifractal analysis is a powerful modeling and estimation framework for the characterization of data dynamics via the fluctuations of its pointwise regularity in time or space. Though successfully applied in a large number of applications in very different contexts, the estimation of parameters related to the data multifractality is an intricate issue for discrete real-world data in non-standard situations, e.g., for small sample size or in the presence of data corruptions. Building upon a recently introduced generic statistical model for log-leaders, specific multiresolution quantities designed for multifractal analysis, the present work proposes a novel robust model and estimator for the multifractality parameter that quantifies the degree of multifractality in data. Our model explicitly accounts for certain data corruptions as outliers in the spectral log-leader domain. Moreover, we propose an original expectation-maximization algorithm for estimating the parameters associated with the model. Several Monte Carlo simulations have been conducted to evaluate the performance of the proposed estimator, which confirm its good performance when compared to standard linear regression in the presence of different additive or multiplicative data perturbations.**

***Index Terms*—multifractal analysis, wavelet leaders, robust estimation, expectation-maximization**

I. Introduction

Context: Multifractal analysis. Multifractal analysis has become a standard signal and image processing tool dedicated to the study of scale invariance in spatial or temporal dynamics, i.e., in systems whose dynamics are controlled by a large continuum of interacting scales instead of isolated characteristic scales (or frequencies). The relationship between scales can be quantified by one (or a few) scaling exponent(s). An equivalent, dual description is given by the multifractal spectrum, which quantifies the temporal or spatial fluctuations of the data's local regularity [1], [2]. In particular, its width, quantified by the so-named multifractality parameter, can be related to spatial or temporal dynamics' intermittency and deviations from joint Gaussianity. Multifractal analysis has been widely used in many applications of very different natures (cf., [2] and the references therein). The estimation of the multifractal spectrum is a central goal of multifractal analysis. In practice, defining a robust formula for its determination from discrete data remains an important and delicate issue.

Related works: estimation of the multifractality parameter. Classical multifractality parameter estimation tools solely build on power laws of the statistics of well-chosen multiscale quantities across scales, as the wavelet leaders, and perform simple linear regressions to quantify scaling exponents. Yet, this approach requires signal or images with large sample size. This limitation spurred the proposition of alternative estimators that build on more informative statistical models [3], [4], most of which remain however tied to fully parametric models for specific multifractal processes and thus of limited practical relevance. More recently, a statistical model for log-wavelet leaders has been introduced that is generic and not tied to specific process constructions. It has paved the way for the construction of new families of estimators, by Bayesian inference [5], [6] or expectation-maximization (EM) [7], and constitutes the current state-of-the-art for estimating the multifractality parameter. However, these estimators are to varying degrees affected by data corruptions that frequently occur in applications, including noise or deterministic perturbations. Though it can critically alter parameter estimates and is thus of central practical importance, this issue has barely been considered in the context of multifractal analysis (see, a contrario, [8] for a the theoretical study on the estimation of a multifractal function in white noise).

Goals, contributions and outline. This work proposes, as a first contribution, a novel model and estimator for the multifractality parameter that is more robust to certain types of data corruptions than classical estimators. The key concepts of multifractal analysis are briefly recalled in Section II. The proposed model is described in Section III. It builds on the statistical model for log-leaders proposed and studied in [7]. However, contrary to [7], a novel key ingredient allows this "robust model" to mitigate the impact of outliers in the spectral domain of log-leaders. As a second contribution, in Section III-C, we derive an EM algorithm for the estimation of the parameters of this robust model, including the multifractality parameter. As a third contribution, we conduct large Monte Carlo simulations to study the robustness of the algorithm with respect to (w.r.t.) additive white noise and or deterministic data corruptions (cf., Section IV). Our results show that the proposed model and EM algorithm succeed in providing relevant estimates for parameter values for a large range of outlier fractions and can effectively deal with various types and degrees of data perturbations. Finally, the proposed model is investigated for the analysis of a real-world heart rate signal.

Work supported by ANR-18-CE45-0007 MUTATION.

II. MULTIFRACTAL ANALYSIS

Multifractal spectrum. Multifractal analysis characterizes the fluctuations along time (or space) of the pointwise regularity $h_X(t)$ of a function or signal $X(t) \in \mathbb{R}, t \in \mathbb{R}$. Pointwise regularity is commonly quantified using the Hölder exponent denoted as $h_X(t) \geq 0$ (see, e.g., [1] for details). The closer $h_X(t)$ to 0, the more irregular $X(t)$ around t. The information contained in the fluctuations of the pointwise regularity is then described globally via the *multifractal spectrum*, defined as the Hausdorff dimensions of the sets of points that share the same Hölder exponent, i.e., $\mathcal{D}(h) \triangleq \dim_H\{t : h_X(t) = h\}$ [1]. The estimation of $\mathcal{D}(h)$ relies on studying the statistics of multiscale quantities such as wavelet coefficients or wavelet leaders, as summarized in [1], [9] and in the next section.

Wavelet leaders. Let ψ denote a mother wavelet, an oscillating reference function with a finite number of zero moments $N_\psi \in \mathbb{N}^+$ (i.e., $\psi \in \mathcal{C}^{N_\psi - 1}$, $\forall n = 0, \ldots, N_\psi - 1$, $\int_\mathbb{R} t^n \psi(t) dt = 0$ and $\int_\mathbb{R} t^{N_\psi} \psi(t) dt \neq 0$) chosen such that the collection $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)\}_{(j,k)\in(\mathbb{Z},\mathbb{Z})}$ of its translated and dilated versions forms a basis of $L^2(\mathbb{R})$. The L_1 normalized discrete wavelet coefficients of $X(t)$ are defined as $d_X(j,k) \triangleq 2^{-j/2} \langle \psi_{j,k}|X\rangle$, and the corresponding *wavelet leaders* as

$$L_X(j,k) \triangleq \sup_{\lambda' \subset 3\lambda_{j,k}} |d_X(\lambda')|,$$

where $\lambda_{j,k} = [k2^j, (k+1)2^j)$ is a dyadic interval of size 2^j, and $3\lambda_{j,k}$ the union of this interval with its two neighbors.

Multifractal formalism. The pth order cumulants $C_p(j)$ of the logarithm of wavelet leaders $\boldsymbol{\ell}_j(k) \triangleq \ln L_X(j,k)$ (in short log-leaders) at a fixed scale 2^j are known to behave as [9]

$$C_p(j) = c_p^0 + j\, c_p \ln 2, \tag{1}$$

where the c_p are referred to as *log-cumulants*, which provides an approximation for the multifractal spectrum $\mathcal{D}(h) \approx 1 + (c_2/2)\left((h - c_1)/c_2\right)^2 + \ldots$, with $c_2 < 0$. Here, c_1 describes the average data regularity, and the *multifractality parameter* c_2 quantifies to leading order the degree of regularity fluctuations of the data and is of fundamental importance for distinguishing different multifractal models, cf., [9]. As suggested by (1), c_2 can be standardly estimated by means of linear regressions of the variances of log-leaders $C_2(j) = \mathrm{Var}[\boldsymbol{\ell}_j(k)]$ against scale j [9]. In the following, we focus on the definition of an alternative robust estimator for this parameter.

III. A ROBUST EM ALGORITHM FOR MULTIFRACTALITY PARAMETER ESTIMATION

Several works have shown that log-leaders can be well modeled as Gaussian random vectors with covariance parametrized by the multifractality parameter c_2 [5]. An equivalent model can be formulated in the spectral domain via a Whittle approximation and data augmentation, facilitating the design of efficient estimators [6]. Starting from this model, which is briefly recalled next, we propose in the following a novel model and estimator for c_2 that is robust against outliers in the spectral domain.

A. Log-leader spectral domain model

Denote as $\boldsymbol{z} \in \mathbb{C}^M$ the vector containing the Fourier coefficients obtained by a discrete Fourier transform (DFT) of the centered log-leaders from all scales under analysis, i.e.,

$$\boldsymbol{z} = \left(\mathrm{DFT}\left\{\boldsymbol{\ell}_j - \langle \boldsymbol{\ell}_j(k)\rangle_k\right\}\right)_{j=j_1}^{j_2}.$$

It can be shown that the components z_i, $i = 1, \ldots, M$, of $\boldsymbol{z}$ are independent and complex Gaussian distributed with zero mean and covariance parametrized by the multifractality $\theta_1 = -c_2$ and a model adjustment parameter θ_2, i.e.,

$$z_i \sim \mathcal{CN}(0, \theta_1 g_1(i) + \theta_2 g_2(i)). \tag{2}$$

The functions $g_1, g_2 > 0$ are known, fixed and subsume the covariance in time for log-leaders of multifractal processes [6]. Finally, introducing a complex Gaussian vector of latent variables $\boldsymbol{u}$ leads to the augmented likelihood

$$\begin{aligned} p(\boldsymbol{z}, \boldsymbol{u}|\theta_1, \theta_2) &= \prod_{i=1}^{M} p(z_i, u_i, |\theta_1, \theta_2) \\ &\propto \prod_{i=1}^{M} \theta_1^{-1} \exp\left(-(\theta_1 g_1(i))^{-1}|z_i - u_i|^2\right) \\ &\qquad \theta_2^{-1} \exp\left(-(\theta_2 g_2(i))^{-1}|u_i|^2\right), \end{aligned} \tag{3}$$

where $\propto$ means "proportional to".

B. Robust model

Outlier model. Our model builds on the assumption that due to data corruptions, the variances of some of the Fourier coefficients z_i in (3) deviate from the model g_1, g_2. This assumption is generic in the sense that it does not further specify the precise data corruptions (noise, outliers in the time domain, additive or multiplicative trends, ...). Note that models for specific data corruptions could also be designed but are left for future studies. Specifically, we assume that a proportion γ of the observations z_i are outliers that follow a zero mean complex Gaussian distribution with variance δ^2 and probability density function (pdf)

$$p(z_i|\delta^2) = \frac{1}{\pi\delta^2} \exp\left(-\frac{|z_i|^2}{\delta^2}\right). \tag{4}$$

Moreover, a distribution $p_1(u_i)$ is assigned to u_i, when z_i is an outlier, whose shape is assumed not to depend on the model parameters of interest.

Conditional distributions. Combining the model (3) for uncorrupted data and (4) for outliers leads to express the conditional distribution of z_i given u_i as the Gaussian mixture

$$\begin{aligned} p(z_i|u_i, \boldsymbol{\theta}) =& \frac{\gamma}{\pi\delta^2} \exp\left(-\frac{|z_i|^2}{\delta^2}\right) \\ &+ \frac{1-\gamma}{\pi\theta_1 g_1(i)} \exp\left(-\frac{|z_i - u_i|^2}{\theta_1 g_1(i)}\right), \end{aligned} \tag{5}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \delta^2, \gamma)$ gathers the unknown parameters of the model. Upon introduction of binary latent variables x_i,

$i = 1, \ldots, M$, such that $x_i = 1$ if z_i is an outlier and $x_i = 0$ otherwise, (5) can be rewritten as

$$p(z_i|u_i, x_i, \boldsymbol{\theta}) = \left[\frac{1}{\pi\delta^2}\exp\left(-\frac{|z_i|^2}{\delta^2}\right)\right]^{x_i} \left[\frac{1}{\pi\theta_1 g_1(i)}\exp\left(-\frac{|z_i - u_i|^2}{\theta_1 g_1(i)}\right)\right]^{1-x_i}, \quad (6)$$

where it is natural to model x_i as a Bernoulli random variable with parameter γ and distribution

$$p(x_i|\boldsymbol{\theta}) = \gamma^{x_i}(1-\gamma)^{1-x_i}. \quad (7)$$

Similarly, the conditional distribution of u_i given x_i is

$$p(u_i|x_i, \boldsymbol{\theta}) = [p_1(u_i)]^{x_i}\left[\frac{1}{\pi\theta_2 g_2(i)}\exp\left(-\frac{|u_i|^2}{\theta_2 g_2(i)}\right)\right]^{1-x_i}. \quad (8)$$

Complete likelihood. In order to obtain a closed-form expression of the maximum likelihood estimator (MLE) of the unknown vector $\boldsymbol{\theta}$, in the spirit of [10], we propose to use the EM algorithm [11] using the so-called complete likelihood

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{x}) &= \prod_{i=1}^{M} p(z_i, u_i, x_i \mid \boldsymbol{\theta}) \\ &= \prod_{i=1}^{M} \underbrace{p(z_i|u_i, x_i, \boldsymbol{\theta})}_{(6)} \underbrace{p(u_i|x_i, \boldsymbol{\theta})}_{(8)} \underbrace{p(x_i|\boldsymbol{\theta})}_{(7)} \\ &= \prod_{i=1}^{M} \left[\frac{\gamma p_1(u_i)}{\pi\delta^2}\exp\left(-\frac{|z_i|^2}{\delta^2}\right)\right]^{x_i} \\ &\quad \left[\frac{1-\gamma}{\pi^2\theta_1\theta_2 g_1(i) g_2(i)}\exp\left(-\frac{|z_i - u_i|^2}{\theta_1 g_1(i)} - \frac{|u_i|^2}{\theta_2 g_2(i)}\right)\right]^{1-x_i}. \quad (9) \end{aligned}$$

C. Proposed EM Algorithm

Starting from an initial parameter $\boldsymbol{\theta}^{(0)}$, the EM algorithm iterates between two steps called expectation (E) and maximization (M) steps.

E-step. Compute the expectation of the log-likelihood (9) w.r.t. the conditional distribution of the latent vectors $\boldsymbol{x}, \boldsymbol{u}$ given the current estimate $\boldsymbol{\theta}^{(\lambda)}$, i.e.,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\lambda)}) = \mathbb{E}_{p(\boldsymbol{u},\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}^{(\lambda)})}[\ln L(\boldsymbol{\theta}; \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{x})] \quad (10)$$

with

$$\begin{aligned} \ln L(\boldsymbol{\theta}; \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{x}) &= \sum_{i=1}^{M} x_i\big(\ln\gamma - \ln\delta^2 - \frac{|z_i|^2}{\delta^2}\big) \\ &+ \sum_{i=1}^{M}(1-x_i)(\ln(1-\gamma) - \ln\theta_1\theta_2) \\ &- \sum_{i=1}^{M}(1-x_i)\big(\frac{|z_i - u_i|^2}{\theta_1 g(i)} - \frac{|u_i|^2}{\theta_2 g_2(i)}\big) + K, \quad (11) \end{aligned}$$

where K is a constant that does not depend on $\boldsymbol{\theta}$. The conditional distribution of $\boldsymbol{u}, \boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{\theta}^{(\lambda)}$ is

$$p(\boldsymbol{u}, \boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{\theta}^{(\lambda)}) = \prod_{i=1}^{M} p(u_i, x_i|z_i, \boldsymbol{\theta}), \quad (12)$$

with

$$\begin{aligned} p(u_i, x_i|z_i, \boldsymbol{\theta}) &\propto p(z_i|u_i, x_i, \boldsymbol{\theta})p(u_i, x_i|\boldsymbol{\theta}) \\ &\propto [\tilde{\pi}_{i,1}p_1(u_i)]^{x_i}[\tilde{\pi}_{i,2}f(u_i|\alpha_i, \beta_i^2)]^{1-x_i}, \quad (13) \end{aligned}$$

where

$$\tilde{\pi}_{i,1} = \frac{\gamma}{\pi\delta^2}\exp\left(-\frac{|z_i|^2}{\delta^2}\right), \quad (14)$$

$$\tilde{\pi}_{i,2} = \frac{1-\gamma}{\pi(\theta_1 g_1(i) + \theta_2 g_2(i))}\exp\left(-\frac{|z_i|^2}{\theta_1 g_1(i) + \theta_2 g_2(i)}\right), \quad (15)$$

and

$$f(u_i|\alpha_i, \beta_i^2) = \frac{1}{\pi\beta_i^2}\exp\left(-\frac{|u_i - \alpha_i|^2}{\beta_i^2}\right) \quad (16)$$

is the density of the random variable

$$u_i|z_i, \boldsymbol{\theta} \sim \mathcal{CN}(\alpha_i, \beta_i^2) \quad (17)$$

$$\text{with} \quad \alpha_i = \beta_i^2(\theta_1 g_1(i))^{-1}z_i, \quad (18)$$

$$\beta_i^2 = \frac{\theta_1 g_1(i)\theta_2 g_2(i)}{\theta_1 g_1(i) + \theta_2 g_2(i)}. \quad (19)$$

The notations

$$\pi_{i,1} = \frac{\tilde{\pi}_{i,1}}{\tilde{\pi}_{i,1} + \tilde{\pi}_{i,2}}, \ \pi_{i,2} = 1 - \pi_{i,1} \quad (20)$$

lead to

$$\mathbb{E}_{p(\boldsymbol{u},\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}^{(\lambda)})}[x_i] = \pi_{i,1}^{(\lambda)} \quad (21)$$

$$\mathbb{E}_{p(\boldsymbol{u},\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}^{(\lambda)})}[1-x_i] = \pi_{i,2}^{(\lambda)} \quad (22)$$

$$\mathbb{E}_{p(\boldsymbol{u},\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}^{(\lambda)})}[(1-x_i)|u_i|^2] = \pi_{i,2}^{(\lambda)}(|\alpha_i^{(\lambda)}|^2 + (\beta_i^2)^{(\lambda)}) \quad (23)$$

$$\mathbb{E}_{p(\boldsymbol{u},\boldsymbol{x}|\boldsymbol{z},\boldsymbol{\theta}^{(\lambda)})}[(1-x_i)|z_i - u_i|^2] = \pi_{i,2}^{(\lambda)}(|z_i - \alpha_i^{(\lambda)}|^2 + (\beta_i^2)^{(\lambda)}) \quad (24)$$

M-step. Closed-form expressions of the different parameter updates for the *Maximization step* are then derived by canceling the gradient of the function Q:

$$\gamma^{(\lambda+1)} = \frac{\pi_1}{M} \quad (25)$$

$$\theta_1^{(\lambda+1)} = (\pi_2)^{-1}\sum_{i=1}^{M}\pi_{i,2}^{(\lambda)}(|z_i - \alpha_i^{(\lambda)}|^2 + (\beta_i^2)^{(\lambda)})/g_1(i) \quad (26)$$

$$\theta_2^{(\lambda+1)} = (\pi_2)^{-1}\sum_{i=1}^{M}\pi_{i,2}^{(\lambda)}(|\alpha_i^{(\lambda)}|^2 + (\beta_i^2)^{(\lambda)})/g_2(i) \quad (27)$$

$$(\delta^2)^{(\lambda+1)} = (\pi_1)^{-1}\sum_{i=1}^{M}\pi_{i,1}^{(\lambda)}|z_i|^2, \quad (28)$$

where $\pi_2 = \sum_{i=1}^{M}\pi_{i,2}^{(\lambda)}$ and $\pi_1 = \sum_{i=1}^{M}\pi_{i,1}^{(\lambda)}$.

Finally, $\pi_{i,2}^{(\lambda)}$ is the a posteriori probability that z_i belongs to the outlier class and can be used to define an outlier detector, e.g., stating that z_i is an outlier if $\pi_{i,2}^{(\lambda)} < 0.5$ (maximum a posteriori detector).

IV. Numerical experiments and results

A. Synthetic multifractal process: multifractal random walk

To assess the performance of the proposed algorithm, we make use of a multifractal random walk (MRW) [12]. MRW is a reference synthetic stochastic process that is constructed to mimic the multifractal properties of the celebrated multiplicative log-normal cascades of Mandelbrot and has multifractal spectrum $\mathcal{D}(h) = 1 + (c_2/2)\left((h - (H - c_2/2))/c_2\right)^2$, with $0.5 < H < 1$ and multifractality parameter $c_2 < 0$.

B. Monte Carlo simulations

To study the estimation performance of the algorithm when the data $X(t)$ are subjected to additive or multiplicative corruptions, we conduct Monte Carlo simulations for $N_{\text{mc}} = 1000$ independent copies of an MRW of size $N = 2^{10}$, with $H = 0.7$, for different values for c_2, with variance of increments normalized to the value 1. The following scenarios for data corruptions $E(t)$, $t \in [0,1]$ are studied:

1) White Gaussian noise with variance σ^2:
$Y(t) = X(t) + \sigma E(t)$, $E(t) \sim \mathcal{N}(0,1)$;
2) Sparse Gaussian outliers with variance σ^2:
$Y(t) = X(t) + \sigma E(t)$, where $E(t) \sim \mathcal{N}(0,1)$ for a fraction γ of data points, and $E(t) = 0$ otherwise;
3) Exponential harmonic perturbation:
$Y(t) = X(t) + \sigma E(t)$, $E(t) = e^{4(\sin(7.9\pi t)+\sin(32\pi t))}$.

The exponential harmonic perturbation is included as a benchmark because it has been previously considered in the context of multifractal analysis [13]. For none of these corruptions it is know which points in the spectrogram of log-leaders are affected. The performance are assessed in terms of the average (mean) and the root-mean-squared error (rmse) of estimates for the multifractality parameter c_2. We use Daubechies wavelet with $N_\psi = 2$ vanishing moment and scales $j \in [\![2, 6]\!]$ for the analysis of data $Y(t)$. Results are shown for a weighted linear regression (WLR) [9], the proposed robust EM algorithm (R-MLE), and an EM algorithm for the likelihood (9) without outlier model denoted MLE (i.e., for $\gamma = 0$, which reduces to the algorithm in [7]). The EM algorithms are stopped when the variation of the log-likelihood falls below 10^{-4}.

C. Simulation results and performance analysis

Additive white noise. Fig. 1 shows the performance of the algorithms for the estimation of c_2 as a function of the noise level σ (log scale) of additive white Gaussian noise ($H = 0.2$, $c_2 = -0.16$). Clearly, WLR has worst performance even for small noise levels and quickly degrades with increasing values of σ. Interestingly, the sole use of the likelihood (9) with $\gamma = 0$ in MLE adds robustness to noise, which has never been reported before. Moreover, MLE estimators display significantly lower RMSE values, and their performance start to deteriorate at noise levels twice as large as those for which performance of WLR degrades. The best performance in terms of bias and RMSE is obtained with the proposed R-MLE estimator.

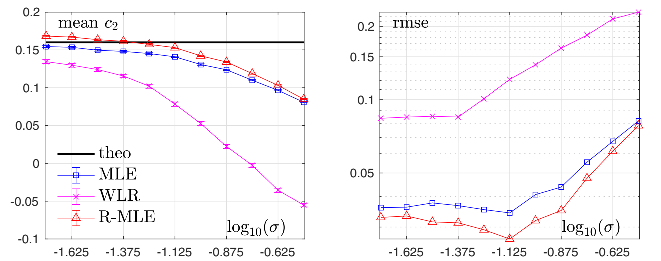

Fig. 1. Estimation performance in the presence of white Gaussian noise as a function of noise level σ ($c_2 = -0.16$).

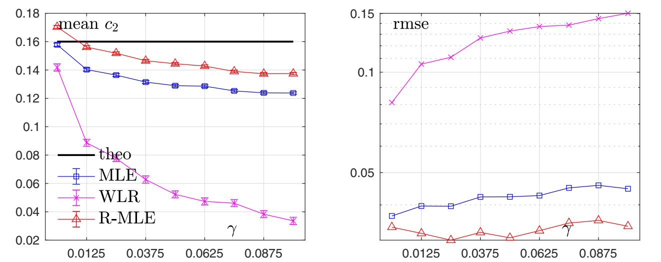

Fig. 2. Estimation performance in the presence of Gaussian outliers as a function of outlier fraction γ ($\sigma = 0.25$, $c_2 = -0.16$)

Sparse additive Gaussian outliers. As in the previous section, Gaussian noise is added to an MRW, but only to a fraction γ of the data points, thus simulating the presence of outliers corrupting the data. Fig. 2 displays the results for this case, which leads to the following conclusions. Again, WLR yields the worst results and inaccurate estimates already for small values for γ. On the contrary, the performance of MLE and R-MLE are little affected by the data corruptions. The proposed R-MLE estimator yields overall the best results in the presence of outliers.

Additive exponential harmonic corruptions. Fig. 3 (center) plots estimation results as a function of the strength σ (log scale) for an additive exponential harmonic data corruption ($H = 0.2, c_2 = -0.16$). This figure shows that WLR is extremely sensitive to such data corruptions and yields severely biased estimates for all but the smallest values of σ. The use of the likelihood (9) with $\gamma = 0$ in MLE adds some robustness, but it eventually also severely impaired by the data perturbation as σ increases. The proposed R-MLE yields significantly better results, and though its bias increases for increasing values of σ, estimates remain reasonably close to the theoretical value $c_2 = -0.16$. Fig. 3 (bottom) shows results as a function of $-c_2$ (for $\sigma = 0.125$) and further confirms these findings: for the whole range of values for c_2, the proposed R-MLE yields consistently good estimates, while WLR and MLE are severely biased.

Illustration for real-world heart rate variability data. The proposed method is finally applied to the analysis of a real-world heart rate (RR beat interval) signal extracted from

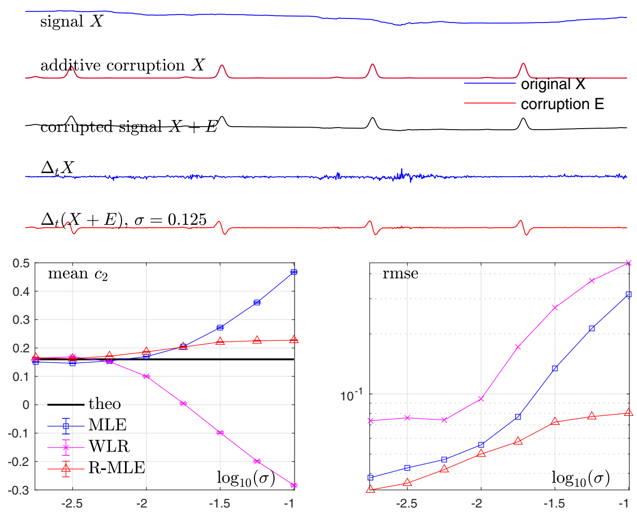

Fig. 3. Estimation performance in the presence of an exponential harmonic perturbation as a function of σ ($c_2 = -0.16$)

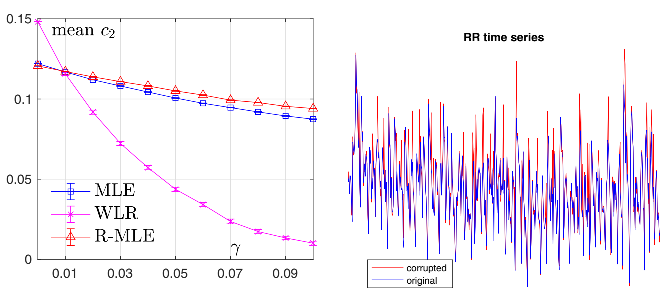

Fig. 4. Average estimates (left) and data (right) for RR intervals for one subject of the MIT-BIH Polysomnographic database with artificial outliers.

the MIT-BIH Polysomnographic database [14][1]. The signal was resampled at 4 Hz (from originally 250 Hz) and time scales corresponding to $2.6 - 41s$ were analyzed (for which multifractal properties are commonly reported in the literature for heart rate data [15]). A common issue with heart rate data is that RR beats are missed by the detection algorithm, leading to wrong and excessively long intervals. We simulate such corruptions by adding the average value of RR beat interval to a fraction γ of the data points. Results are plotted in Fig. 4 in terms of average estimates for c_2 as a function of γ (average over 100 realizations of outliers), together with an illustration for a small portion of the time series for $\gamma = 0.1$. The results are similar to those observed for the synthetic data scenario reported in Fig. 2: The estimates of c_2 obtained using WLR are practically unusable even for small portions of missed RR beats, thus must rely on extremely precise preprocessing, while those of the proposed model remain reliable for a large range of outlier fractions γ, and better for R-MLE than MLE.

Overall, these results clearly show the effectiveness of the proposed robust multifractal model and its estimator to counter the effect of various types and degrees of data corruptions on the estimation of the multifractality parameter c_2, and that it can be beneficially applied to the analysis of real-world data.

[1] Subject 2a, https://physionet.org/content/slpdb/1.0.0/

V. Conclusions

This work introduced a novel model and estimator for the multifractality parameter c_2 that is more robust to certain types of data corruptions than classical estimators. The proposed method builds on the log-leader frequency domain likelihood of [6] and models data corruptions as alterations of portions of the log-leaders' spectrum. We devised and studied an EM algorithm for the estimation of the parameters of this robust model. Our numerical results for synthetic data for several different types of data corruptions clearly indicate the benefits of the proposed method over previously existing estimators for the multifractality parameter in terms of estimation performance. As a side result, this paper also showed that the model in [6] is, to lesser degree, also more robust than classical linear regressions, which has never been reported before. Future work will include the design of models that explicitly account for specific data corruptions in the time domain, such as the outliers in our illustration for real-world heart rate data.

References

[1] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Appl.*, M. Lapidus and M. van Frankenhuijsen, Eds. 2004, vol. 72(2), pp. 91–152, AMS.

[2] S. Jaffard, P. Abry, and H. Wendt, *Irregularities and Scaling in Signal and Image Processing: Multifractal Analysis*, pp. 31–116, World scientific publishing, Singapore, 2015.

[3] T. Lux, "Higher dimensional multifractal processes: A GMM approach," *Journal of Business and Economic Statistics*, vol. 26, pp. 194–210, 2007.

[4] O. Løvsletten and M. Rypdal, "Approximated maximum likelihood estimation in multifractal random walks," *Phys. Rev. E*, vol. 85, pp. 046705, 2012.

[5] S. Combrexelle, H. Wendt, N. Dobigeon, J.-Y. Tourneret, S. McLaughlin, and P. Abry, "Bayesian estimation of the multifractality parameter for image texture using a Whittle approximation," *IEEE T. Image Proces.*, vol. 24, no. 8, pp. 2540–2551, 2015.

[6] H. Wendt, S. Combrexelle, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, and P. Abry, "Multifractal analysis of multivariate images using gamma Markov random field priors," *SIAM J. Imaging Sciences*, vol. 11, no. 2, pp. 1294–1316, 2018.

[7] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry, "Estimation du paramètre de multifractalité : régression linéaire, maximum de vraisemblance ou inférence Bayésienne ?," in *Proc. GRETSI*, Nancy, Sep 2022, number 001-0323, pp. p. 1293–1296.

[8] F. Gamboa and J.-M. Loubes, "Wavelet estimation of a multifractal function," *Bernoulli*, vol. 11, no. 2, pp. 221–246, 2005.

[9] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis," *IEEE Signal Proces. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

[10] J. Lesouple, B. Pilastre, Y. Altmann, and J.-Y. Tourneret, "Robust hypersphere fitting from noisy data using an EM algorithm," in *29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1–5.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B*, vol. 39, no. 1, pp. 1–22, 1977.

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

[13] E. Ihlen, "Multifractal analyses of response time series: A comparative study," *Behavior research methods*, vol. 45, pp. 928–945, 2013.

[14] A. Goldberger, L. Amaral, L. Glass, Hausdorff, Ivanov J., R. P. C., Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, 2000.

[15] T. Nakamura, K. Kiyono, H. Wendt, P. Abry, and Y. Yamamoto, "Multiscale analysis of intensive longitudinal biomedical signals and its clinical applications," *Proc. IEEE*, vol. 104, no. 2, pp. 242–261, 2016.