

# A BAYESIAN FRAMEWORK FOR THE MULTIFRACTAL ANALYSIS OF IMAGES USING DATA AUGMENTATION AND A WHITTLE APPROXIMATION

S. Combrexelle<sup>1</sup>, H. Wendt<sup>1</sup>, Y. Altmann<sup>2</sup>, J.-Y. Tournet<sup>1</sup>, S. McLaughlin<sup>2</sup>, P. Abry<sup>3</sup>

<sup>1</sup> IRIT - ENSEEIHT, CNRS, Univ. of Toulouse, F-31062 Toulouse, France, `firstname.lastname@enseeiht.fr`

<sup>2</sup> School of Engineering and Physical Sciences, Heriot-Watt Univ., Edinburgh, UK, `initial.lastname@hw.ac.uk`

<sup>3</sup> CNRS, Physics Dept., Ecole Normale Supérieure de Lyon, F-69364 Lyon, France, `patrice.abry@ens-lyon.fr`

## ABSTRACT

Texture analysis is an image processing task that can be conducted using the mathematical framework of multifractal analysis to study the regularity fluctuations of image intensity and the practical tools for their assessment, such as (wavelet) leaders. A recently introduced statistical model for leaders enables the Bayesian estimation of multifractal parameters. It significantly improves performance over standard (linear regression based) estimation. However, the computational cost induced by the associated nonstandard posterior distributions limits its application. The present work proposes an alternative Bayesian model for multifractal analysis that leads to more efficient algorithms. It relies on three original contributions: A novel generative model for the Fourier coefficients of log-leaders; an appropriate reparametrization for handling its inherent constraints; a data-augmented Bayesian model yielding standard conditional posterior distributions that can be sampled exactly. Numerical simulations using synthetic multifractal images demonstrate the excellent performance of the proposed algorithm, both in terms of estimation quality and computational cost.

**Index Terms**— Multifractal Analysis, Wavelet Leaders, Bayesian Estimation, Whittle Likelihood, Data Augmentation

## 1. INTRODUCTION

**Context.** Texture analysis is an important technique in image processing and many different paradigms have been proposed to quantify it. The mathematical framework of multifractal analysis has proven particularly relevant, providing standard processing tools also used in a large range of other applications, cf., e.g., [1, 2] and references therein. Multifractal analysis is an instance of scale invariance analysis that considers an image through the prism of the fluctuations of the pointwise smoothness of its amplitude. More precisely, the texture of an image  $X$  is encoded by the *multifractal spectrum*  $D(h)$  defined as the Hausdorff dimension of the sets of points that have the same pointwise regularity  $h$ , classically measured with the Hölder exponent, cf., e.g., [3–6].

From a practical point of view, multifractal models translate to the power law behaviors of the sample moments of adequate multiresolution quantities  $T_X(j, \mathbf{k})$  of  $X$  (i.e., quantities depending jointly on scale  $2^j$  and spatial position  $\mathbf{k}$ ) over a range of scales  $2^j$

$$S(q, j) \triangleq \frac{1}{n_j} \sum_{\mathbf{k}} |T_X(j, \mathbf{k})|^q \simeq (2^j)^{\zeta(q)}, \quad j_m \leq j \leq j_M \quad (1)$$

where  $n_j \approx \lfloor N^2/2^{2j} \rfloor$  is the number of  $T_X(j, \mathbf{k})$  at scale  $j$  and  $\lfloor \cdot \rfloor$  stands for the truncation to integer values. Here, *wavelet leaders*  $l(j, \mathbf{k})$  are used as multiresolution quantities, which are specifically tailored for the purpose of multifractal analysis [1, 4] and defined in Section 2. The exponents  $\zeta(q)$  in (1), termed *scaling exponents*, are closely linked to the multifractal properties of the image  $X(\mathbf{k})$  via a Legendre transform  $\mathcal{L}$  such that,  $D(h) \leq \mathcal{L}(h) \triangleq \inf_{q \in \mathbb{R}} [2 + qh - \zeta(q)]$ . The scaling exponents  $\zeta(q)$  enable the distinction of the two most prominent classes of scale invariance models: while  $\zeta(q)$  is linear in the vicinity of  $q = 0$  for self-similar processes [7], it is strictly concave for multifractal multiplicative cascade (MMC) based processes [8]. The linearity of  $\zeta(q)$  can be efficiently tested by considering the development of  $\zeta(q)$  as a polynomial at  $q = 0$ ,  $\zeta(q) = \sum_{m \geq 1} c_m q^m / m!$  [1, 9, 10]. Notably, it can be shown that  $c_2 < 0$  for multiplicative cascades whereas  $c_2 = 0$  for self-similar processes (cf., e.g. [10]) and  $c_2 = 0$  implies that  $c_m = 0, \forall m \geq 3$  [4]. Therefore, the estimation of  $c_2$ , termed the *intermittency* or *multifractality* parameter, enables the identification of the model that best fits the data and is thus of paramount importance in multifractal analysis. For details on multifractal analysis, the reader is referred to, e.g., [3–6].

**Related work.** The coefficients  $c_m$  have been shown to be directly tied to the cumulants of the logarithm of the multiresolution quantities, see [9]. Specifically, for the multifractality parameter  $c_2$ ,

$$C_2(j) \triangleq \text{Var} [\ln l(j, \cdot)] = c_2^0 + c_2 \ln 2^j. \quad (2)$$

Relation (2) motivates the classical estimation procedure of  $c_2$  which is based on a linear regression of the sample variance  $\widehat{\text{Var}}[\cdot]$  of the log-leaders with respect to scale  $j$

$$\hat{c}_2 = \frac{1}{\ln 2} \sum_{j=j_1}^{j_2} w_j \widehat{\text{Var}} [\ln l(j, \cdot)] \quad (3)$$

where  $[j_1, j_2]$  is the range of considered scales and  $w_j$  are appropriate regression weights [1, 11, 12]. This linear regression based estimation is appealing due to its simplicity and low computational cost. However it suffers from poor performance for images (or patches) of small size due to the limited number of available scales.

In [13], a generalized method of moments has been proposed to address this difficulty. However, it relies on fully parametric models that are often too restrictive in real-world applications. Another alternative has been recently proposed in [2, 14] and embeds the estimation of  $c_2$  in a Bayesian framework. It relies on a generic semi-parametric model for the multivariate statistics of the log-leaders in which the parameters of interest, notably  $c_2$ , are encoded in the variance-covariance structure of a Gaussian likelihood. A Whittle approximation was used in [2] to numerically assess the Gaussian likelihood and a closed-form expression was proposed in [14] for

This work was supported by ANR BLANC 2011 AMATIS BS0101102. S. Combrexelle was supported by the Direction Générale de l'Armement (DGA). SML acknowledges the support of EPSRC via grant EP/J015180/1.

its efficient evaluation. In both [2, 14], the Bayesian inference was accomplished by a Markov chain Monte Carlo (MCMC) algorithm. More precisely, a Metropolis-Hasting within Gibbs (MHG) scheme was considered to bypass the difficulties resulting from non-standard conditional distributions associated with the posterior of interest. This method yielded excellent estimation performance at the price of significantly increased computational cost compared to (3).

**Goals and contributions.** The goal of this paper is to propose a novel Bayesian estimation procedure for  $c_2$  for images that inherits the excellent estimation performance of the Bayesian framework introduced in [2, 14] while significantly reducing the associated computational cost. Starting from the statistical model introduced in [2, 14] (recalled in Section 2), this is achieved through the following original key developments. First, instead of using a Whittle approximation exclusively for numerically evaluating the Gaussian likelihood more efficiently in the Fourier domain as in [2, 14], we propose here to rely on a statistical interpretation of the Whittle approximation and formulate the statistical model directly in the Fourier domain (see Section 3.1). Second, we propose a reparametrization of the problem for handling the parameter constraints more efficiently and, third, we express the Fourier domain likelihood as the marginal of a data augmented likelihood [15, 16] (cf. Section 3.2). The key property of this data augmentation scheme lies in the fact that the use of standard priors (e.g., Jeffreys or inverse Gamma) leads to conditional posterior distributions that can be sampled exactly. Therefore, unlike the framework in [2, 14], it does not require the use of MHG steps for the numerical approximation of the posterior distribution. The computation of an estimator associated with the proposed Bayesian model and the marginalization of the latent variables are then achieved efficiently by an MCMC algorithm (cf. Section 4). In Section 5, the performance of the resulting estimation procedure is assessed by means of Monte Carlo simulations on synthetic multifractal images. The proposed algorithm significantly outperforms the linear regression (3), reducing root mean squared error (RMSE) values up to one quarter, while its total computational cost is only 2 (large images) to 5 (small images) times larger.

## 2. MULTIVARIATE STATISTICAL MODEL FOR LEADERS

### 2.1. Statistical model of log-leaders

**Wavelet leaders.** Given a scaling function  $\phi(x)$  and a mother wavelet  $\psi(x)$  for a 1D multiresolution analysis, 2D wavelets can be defined as tensorial products  $\psi^{(0)}(\mathbf{x}) = \phi(x_1)\phi(x_2)$ ,  $\psi^{(1)}(\mathbf{x}) = \psi(x_1)\phi(x_2)$ ,  $\psi^{(2)}(\mathbf{x}) = \phi(x_1)\psi(x_2)$ ,  $\psi^{(3)}(\mathbf{x}) = \psi(x_1)\psi(x_2)$  [17, 18]. For a suitable  $\psi$ , the dilated and translated templates of  $\psi^{(m)}$ , denoted  $\psi_{j,k}^{(m)}(\mathbf{x}) = 2^{-j} \psi^{(m)}(2^{-j}\mathbf{x} - k)$ , form a basis of  $L^2(\mathbb{R}^2)$  with  $a = 2^j$  and  $\mathbf{x} = 2^j \mathbf{k}$ . The ( $L^1$ -normalized) discrete wavelet transform coefficients of the image  $X$  are defined as  $d_X^{(m)}(j, k) = \langle X, \psi_{j,k}^{(m)} \rangle$ ,  $m = 0, \dots, 3$  [17]. Denote as  $\lambda_{j,k}$  the dyadic cube of side length  $2^j$  centered at  $\mathbf{k}2^j$  and  $3\lambda_{j,k} = \bigcup_{n_1, n_2 = \{-1, 0, 1\}} \lambda_{j, k_1+n_1, k_2+n_2}$  the union of this cube with its eight neighbors. The wavelet leaders are defined as the supremum of the wavelet coefficients within this spatial neighborhood over all finer scales [1, 4], i.e.,

$$l(j, \mathbf{k}) \triangleq \sup_{m \in \{1, 2, 3\}, \lambda' \subset 3\lambda_{j, \mathbf{k}}} |d_X^{(m)}(\lambda')|. \quad (4)$$

**Statistical model.** Let the log-leaders at scale  $j$ ,  $\ell(j, \cdot) \triangleq \ln l(j, \cdot)$ , be stacked in the vector  $\ell_j$  of which the mean has been removed (since we focus on the estimation of  $c_2$ , cf. (2)). It has been reported in [2, 14] that the statistics of  $\ell_j$  for MMC based processes can be

well approximated by a multivariate Gaussian distribution with covariance  $C_j(\mathbf{k}, \Delta \mathbf{k}) \triangleq \text{Cov}[\ell(j, \mathbf{k}), \ell(j, \mathbf{k} + \Delta \mathbf{k})]$  modeled by a radial symmetric function parametrized only by  $\theta = (c_2, c_2^0)$ ,

$$C_j(\mathbf{k}, \Delta \mathbf{k}) \approx \varrho_j(\|\Delta \mathbf{k}\|; \theta) \triangleq \begin{cases} \varrho_j^0(\|\Delta \mathbf{k}\|; \theta) & \|\Delta \mathbf{k}\| \leq 3 \\ \varrho_j^1(\|\Delta \mathbf{k}\|; \theta) & 3 < \|\Delta \mathbf{k}\| \end{cases} \quad (5)$$

where  $\|\cdot\|$  is the Euclidian norm. The function  $\varrho_j^1$  is given by a logarithmic decay  $\varrho_j^1(r; \theta) \triangleq c_2 \ln(r/r_j) \mathbb{I}_{[0, r_j]}(r)$  where  $r_j = \lfloor \sqrt{n_j}/4 \rfloor$  and  $\mathbb{I}_A$  is the indicator function of the set  $A$ . The short-term correlation is modeled by  $\varrho_j^0(r; \theta) \triangleq a_j \ln(1+r) + c_2^0 + c_2 \ln 2^j$  where  $a_j \triangleq (\varrho_j^1(3; \theta) - c_2^0 - c_2 \ln 2^j) / \ln 4$  (see [2] for details). With the above assumptions, the likelihood of  $\ell_j$  reads

$$p(\ell_j | \theta) \propto |\Sigma_{j, \theta}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \ell_j^T \Sigma_{j, \theta}^{-1} \ell_j\right) \quad (6)$$

where the matrix  $\Sigma_{j, \theta}$  is defined by  $\varrho_j(\|\Delta \mathbf{k}\|; \theta)$  [2] and where  $|\cdot|$  denotes the determinant and  $T$  the transpose operator.

### 2.2. Closed-form Whittle likelihood

The evaluation of the Gaussian likelihood (6) requires computing the matrix inverse  $\Sigma_{j, \theta}^{-1}$ , which is problematic even for small images. Therefore, (6) is evaluated using the Whittle approximation [19–22]

$$p_W(\ell_j | \theta) = \exp\left(-\frac{1}{2} \sum_{\mathbf{m} \in J_j} \ln \phi_j(\omega_{\mathbf{m}}; \theta) + \frac{y_j^*(\omega_{\mathbf{m}}) y_j(\omega_{\mathbf{m}})}{\phi_j(\omega_{\mathbf{m}}; \theta)}\right) \quad (7)$$

as proposed in [2, 14], where  $y_j(\cdot)$  is the discrete Fourier transform of  $\ell(j, \cdot)$ ,  $\omega_{\mathbf{m}} = 2\pi \mathbf{m} / \sqrt{n_j}$ ,  $J_j$  is the grid of integers  $J_j \triangleq \llbracket [(-\sqrt{n_j}-1)/2], \dots, \sqrt{n_j}-\lfloor \sqrt{n_j}/2 \rfloor \rrbracket^2 \setminus \{0\}$  and  $(\cdot)^*$  stands for complex conjugation. In (7),  $y_j^*(\omega_{\mathbf{m}}) y_j(\omega_{\mathbf{m}})$  corresponds to the periodogram of  $\{\ell(j, \mathbf{k})\}$  and  $\phi_j(\omega_{\mathbf{m}}; \theta)$  to the discretized parametric spectral density associated with the model (5). It has been shown in [14] that the function  $\phi_j(\omega_{\mathbf{m}}; \theta)$  has a closed-form parametric expression given by

$$\phi_j(\omega_{\mathbf{m}}; \theta) = c_2 f_j(\omega_{\mathbf{m}}) + c_2^0 g_j(\omega_{\mathbf{m}}) \quad (8)$$

where the two functions  $f_j(\cdot)$  and  $g_j(\cdot)$  do not depend on  $\theta$ , which enables the efficient evaluation of (8) and hence (7) (cf. [14]).

## 3. EXTENDED FOURIER DOMAIN STATISTICAL MODEL

One limitation of the model above is that the parameters of interest are implicitly encoded in  $\Sigma_{j, \theta}^{-1}$ . Thus, the conditional distributions of these parameters are not standard, and sampling the posterior distribution of interest with an MCMC method [2, 14] requires MHG moves. Here we propose a solution that yields a more efficient algorithm. First, we use a generative model based on a statistical interpretation of (7). Second, we reparametrize its constraints in order to enable the use of data augmentation [15, 16]. Third, we propose an extended likelihood for which the design of classical priors leads to standard conditional distributions.

### 3.1. Statistical model in the Fourier domain

The statistical model of [2, 14] summarized in Section 2 exploits the Whittle approximation only to *numerically evaluate* (6). Here, we use a statistical interpretation of the Whittle approximation and replace the likelihood (6) by a statistical model for the Fourier coefficients  $y_j(\omega_{\mathbf{m}})$ . The central symmetry properties of  $y_j(\omega_{\mathbf{m}})$  and

of the parametric spectral density  $\phi_j(\boldsymbol{\omega}_m; \boldsymbol{\theta})$  (due to properties of Fourier transform of real functions) imply that only half of the frequency plane, denoted  $\underline{J}_j$ , needs be considered in the sum in (7). The expression (7) can hence be rewritten as

$$p_W(\ell_j|\boldsymbol{\theta}) = |\Gamma_{j,\boldsymbol{\theta}}|^{-1} \exp\left(-\mathbf{y}_j^H \Gamma_{j,\boldsymbol{\theta}}^{-1} \mathbf{y}_j\right), \quad \mathbf{y}_j \triangleq \mathcal{F}(\ell_j) \quad (9)$$

where the operator  $\mathcal{F}(\cdot)$  computes and vectorizes the Fourier coefficients contained in the half-plane  $\underline{J}_j$ ,  $^H$  is the conjugate transpose operator and  $\Gamma_{j,\boldsymbol{\theta}}$  is the diagonal matrix defined by

$$\Gamma_{j,\boldsymbol{\theta}} \triangleq c_2 \mathbf{F}_j + c_2^0 \mathbf{G}_j, \quad \mathbf{F}_j \triangleq \text{diag}(\mathbf{f}_j), \quad \mathbf{G}_j \triangleq \text{diag}(\mathbf{g}_j) \quad (10)$$

with  $\mathbf{f}_j \triangleq (f_j(\boldsymbol{\omega}_m))_{m \in \underline{J}_j}$  and  $\mathbf{g}_j \triangleq (g_j(\boldsymbol{\omega}_m))_{m \in \underline{J}_j}$ . In view of (9), using  $p_W(\ell_j|\boldsymbol{\theta})$  thus amounts to modeling  $\mathbf{y}_j$  by a random vector with a non-degenerate centered circular-symmetric complex Gaussian distribution  $\mathcal{CN}(\mathbf{0}, \Gamma_{j,\boldsymbol{\theta}})$ , provided the covariance matrix  $\Gamma_{j,\boldsymbol{\theta}}$  is positive definite (PD).

In the rest of this paper, we thus consider the Fourier coefficients  $\mathbf{y}_j$ ,  $j = j_1, \dots, j_2$  with likelihood

$$p(\mathbf{y}_j|\boldsymbol{\theta}) = |\Gamma_{j,\boldsymbol{\theta}}|^{-1} \exp\left(-\mathbf{y}_j^H \Gamma_{j,\boldsymbol{\theta}}^{-1} \mathbf{y}_j\right) \quad (11)$$

as the observed data (rather than the log-leaders  $\ell_j$ , modeled by (6)). Assuming independence between scales  $j$  leads to the likelihood

$$p(\mathbf{y}|\boldsymbol{\theta}) \triangleq \prod_{j=j_1}^{j_2} p(\mathbf{y}_j|\boldsymbol{\theta}) \propto |\Gamma_{\boldsymbol{\theta}}|^{-1} \exp\left(-\mathbf{y}^H \Gamma_{\boldsymbol{\theta}}^{-1} \mathbf{y}\right) \quad (12)$$

for the vector  $\mathbf{y} \triangleq [\mathbf{y}_{j_1}^T, \dots, \mathbf{y}_{j_2}^T]^T$ , with diagonal covariance matrix

$$\Gamma_{\boldsymbol{\theta}} \triangleq c_2 \mathbf{F} + c_2^0 \mathbf{G} \quad \mathbf{F} \triangleq \text{diag}(\mathbf{f}) \quad \mathbf{G} \triangleq \text{diag}(\mathbf{g}) \quad (13)$$

$$\mathbf{f} \triangleq [\mathbf{f}_{j_1}^T, \dots, \mathbf{f}_{j_2}^T]^T \quad \mathbf{g} \triangleq [\mathbf{g}_{j_1}^T, \dots, \mathbf{g}_{j_2}^T]^T.$$

To ensure that the  $N_Y \times N_Y$  matrix  $\Gamma_{\boldsymbol{\theta}}$  is PD, where  $N_Y \triangleq \text{card}(\mathbf{y})$ , the parameters  $\boldsymbol{\theta} = (c_2, c_2^0)$  must belong to the admissible set

$$\mathcal{A} = \{\boldsymbol{\theta} \in \mathbb{R}_*^- \times \mathbb{R}_*^+ | c_2 \mathbf{f}(k) + c_2^0 \mathbf{g}(k) > 0, k = 1, \dots, N_Y\}. \quad (14)$$

### 3.2. Reparametrization and data augmentation

Finally we propose to replace the likelihood  $p(\mathbf{y}|\boldsymbol{\theta})$  with an extended likelihood  $p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\theta})$  using an appropriate set of latent variables  $\boldsymbol{\mu}$ . To do so, we first need to reparametrize the model.

**Reparametrization.** To construct the extended likelihood  $p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\theta})$ ,  $\Gamma_{\boldsymbol{\theta}}$  must be expressed as the sum of two PD diagonal matrices, which is not the case in (13) because  $c_2 \mathbf{F}$  is not PD since  $\exists k : \mathbf{f}(k) > 0$  (while  $\forall k : \mathbf{g}(k) > 0$ ) [14]. We thus propose a reparametrization defined by the mapping

$$\psi: \boldsymbol{\theta} \mapsto \tilde{\boldsymbol{\theta}} \triangleq (-c_2, c_2^0/\gamma + c_2), \quad \gamma = \sup_k \mathbf{f}(k)/\mathbf{g}(k). \quad (15)$$

It is easy to show that  $\psi$  is a one-to-one transformation from  $\mathcal{A}$  to  $\mathbb{R}_*^{+2}$  and that the likelihood (12) expressed with  $\tilde{\boldsymbol{\theta}}$  is given by

$$p(\mathbf{y}|\tilde{\boldsymbol{\theta}}) \propto |\Gamma_{\tilde{\boldsymbol{\theta}}}|^{-1} \exp\left(-\mathbf{y}^H \Gamma_{\tilde{\boldsymbol{\theta}}}^{-1} \mathbf{y}\right) \quad (16)$$

$$\Gamma_{\tilde{\boldsymbol{\theta}}} = \tilde{\theta}_1 \tilde{\mathbf{F}} + \tilde{\theta}_2 \tilde{\mathbf{G}} \quad \tilde{\mathbf{F}} = -\mathbf{F} + \mathbf{G}\gamma \quad \tilde{\mathbf{G}} = \mathbf{G}\gamma \quad (17)$$

where, by construction, the two diagonal matrices  $\tilde{\theta}_1 \tilde{\mathbf{F}}$  and  $\tilde{\theta}_2 \tilde{\mathbf{G}}$  are PD for  $\tilde{\boldsymbol{\theta}} \in \mathbb{R}_*^{+2}$ . This last property is essential for the data augmentation scheme proposed in the next paragraph. Furthermore, since  $\tilde{\boldsymbol{\theta}} \in \mathbb{R}_*^{+2}$ , (15) yields independent positivity constraints.

**Data augmentation.** We can now define the extended model

$$\mathbf{y}|\boldsymbol{\mu}, \tilde{\boldsymbol{\theta}}_2 \sim \mathcal{CN}(\boldsymbol{\mu}, \tilde{\theta}_2 \tilde{\mathbf{G}}), \quad \boldsymbol{\mu}|\tilde{\boldsymbol{\theta}}_1 \sim \mathcal{CN}(0, \tilde{\theta}_1 \tilde{\mathbf{F}}) \quad (18)$$

where the vector  $\boldsymbol{\mu}$  is an additional latent variable. The model (18) is associated with the extended likelihood

$$p(\mathbf{y}, \boldsymbol{\mu}|\tilde{\boldsymbol{\theta}}) \propto \tilde{\theta}_2^{-N_Y} \exp\left(-\frac{1}{\tilde{\theta}_2} (\mathbf{y} - \boldsymbol{\mu})^H \tilde{\mathbf{G}}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right) \\ \times \tilde{\theta}_1^{-N_Y} \exp\left(-\frac{1}{\tilde{\theta}_1} \boldsymbol{\mu}^H \tilde{\mathbf{F}}^{-1} \boldsymbol{\mu}\right). \quad (19)$$

Here,  $\boldsymbol{\mu}$  has been chosen such that the likelihood (16) is found by marginalizing (19) with respect to  $\boldsymbol{\mu}$ . Moreover, it is easy to see that (19) leads to standard conditional distributions when used with classical priors for  $\tilde{\theta}_i \in \mathbb{R}_*^+$  (e.g., inverse Gamma or Jeffreys priors).

## 4. BAYESIAN ESTIMATION

### 4.1. Prior and posterior distribution

Assuming *a priori* independence between  $(\boldsymbol{\mu}, \tilde{\theta}_1)$  and  $\tilde{\theta}_2$  and given priors  $p(\tilde{\theta}_i)$  for the parameters  $\tilde{\theta}_i$ , the joint posterior distribution for  $(\tilde{\boldsymbol{\theta}}, \boldsymbol{\mu})$  is given by Bayes' theorem

$$p(\tilde{\boldsymbol{\theta}}, \boldsymbol{\mu}|\mathbf{y}) \propto p(\mathbf{y}|\tilde{\boldsymbol{\theta}}, \boldsymbol{\mu}) p(\boldsymbol{\mu}|\tilde{\theta}_1) p(\tilde{\theta}_1) p(\tilde{\theta}_2).$$

In this paper, we consider non-informative Jeffreys priors,  $p(\tilde{\theta}_i) \propto 1/\tilde{\theta}_i$  (even if other priors could be investigated), in which case the resulting posterior distribution is given by

$$p(\tilde{\boldsymbol{\theta}}, \boldsymbol{\mu}|\mathbf{y}) \propto \tilde{\theta}_2^{-N_Y} \exp\left(-1/\tilde{\theta}_2 (\mathbf{y} - \boldsymbol{\mu})^H \tilde{\mathbf{G}}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right) \\ \times \tilde{\theta}_1^{-N_Y} \exp\left(-\tilde{\theta}_1^{-1} \boldsymbol{\mu}^H \tilde{\mathbf{F}}^{-1} \boldsymbol{\mu}\right) \times \tilde{\theta}_1^{-1} \times \tilde{\theta}_2^{-1}. \quad (20)$$

### 4.2. Inference procedure

**Bayesian estimators.** Since the latent variable  $\boldsymbol{\mu}$  is of no interest for multifractal analysis purposes, we consider the marginal posterior mean estimator for  $\tilde{\boldsymbol{\theta}}$ , denoted MMSE (minimum mean square error) estimator and defined as

$$\tilde{\boldsymbol{\theta}}^{\text{MMSE}} \triangleq \mathbb{E}[\tilde{\boldsymbol{\theta}}|\mathbf{y}] \quad (21)$$

where the expectation is taken with respect to the marginal posterior density  $p(\tilde{\boldsymbol{\theta}}|\mathbf{y})$ . Unfortunately, (21) cannot be computed directly because this would require integrating over the full posterior (20). Thus, we propose here to resort to an MCMC algorithm [23] to approximate (21). More precisely, we investigate a Gibbs sampler drawing samples  $(\tilde{\boldsymbol{\theta}}^{(k)}, \boldsymbol{\mu}^{(k)})_{k=1}^{N_{mc}}$  that are asymptotically distributed according to the joint posterior  $p(\tilde{\boldsymbol{\theta}}, \boldsymbol{\mu}|\mathbf{y})$ . The marginal posterior mean can then be approximated by [23]

$$\tilde{\boldsymbol{\theta}}^{\text{MMSE}} \approx \frac{1}{N_{mc} - N_{bi}} \sum_{t=N_{bi}+1}^{N_{mc}} \tilde{\boldsymbol{\theta}}^{(t)} \quad (22)$$

where  $N_{bi}$  is the length of the burn-in period.

**Gibbs sampler.** The Gibbs sampler consists in successively generating samples from the conditional distributions  $p(\boldsymbol{\mu}|\mathbf{y}, \tilde{\boldsymbol{\theta}})$ ,  $p(\tilde{\theta}_1|\mathbf{y}, \boldsymbol{\mu}, \tilde{\theta}_2)$  and  $p(\tilde{\theta}_2|\mathbf{y}, \boldsymbol{\mu}, \tilde{\theta}_1)$ . It is easy to show that for the posterior (20), these conditional distributions are given by

$$\boldsymbol{\mu}|\mathbf{y}, \tilde{\boldsymbol{\theta}} \sim \mathcal{CN}\left(\left(\tilde{\theta}_1 \tilde{\mathbf{F}} \Gamma_{\tilde{\boldsymbol{\theta}}}^{-1}\right) \mathbf{y}, \left(\left(\tilde{\theta}_1 \tilde{\mathbf{F}}\right)^{-1} + \left(\tilde{\theta}_2 \tilde{\mathbf{G}}\right)^{-1}\right)^{-1}\right) \quad (23)$$

$$\tilde{\theta}_1|\mathbf{y}, \boldsymbol{\mu}, \tilde{\theta}_2 \sim \mathcal{IG}(N_Y, \boldsymbol{\mu}^H \tilde{\mathbf{F}}^{-1} \boldsymbol{\mu}) \quad (24)$$

$$\tilde{\theta}_2|\mathbf{y}, \boldsymbol{\mu}, \tilde{\theta}_1 \sim \mathcal{IG}(N_Y, (\mathbf{y} - \boldsymbol{\mu})^H \tilde{\mathbf{G}}^{-1} (\mathbf{y} - \boldsymbol{\mu})) \quad (25)$$

where  $\mathcal{IG}$  stands for the inverse Gamma distribution. Note that none of the conditional sampling steps requires the use of an acceptance-rejection procedure, contrary to the classical MHG algorithm proposed in [2, 14], which constitutes the main advantage of the proposed extended Bayesian model defined in (20).

## 5. NUMERICAL EXPERIMENTS

We compare the proposed Bayesian estimator (denoted as G) to the Bayesian estimator presented in [14] (denoted as MHG) and to the standard linear regression based estimator (using (3) and denoted as LF) by applying them to a large number of independent realizations of a 2D multifractal random walk (MRW). MRW is a widely used member of the class of multifractal multiplicative cascade based processes whose multifractal properties mimic those of the Mandelbrot's log-normal cascades, with scaling exponents  $\zeta(q) = (H - c_2)q + c_2q^2$  (cf., [24] for details and Fig. 1 (top) for realizations of MRW obtained with identical multipliers for different values of  $c_2$ ).

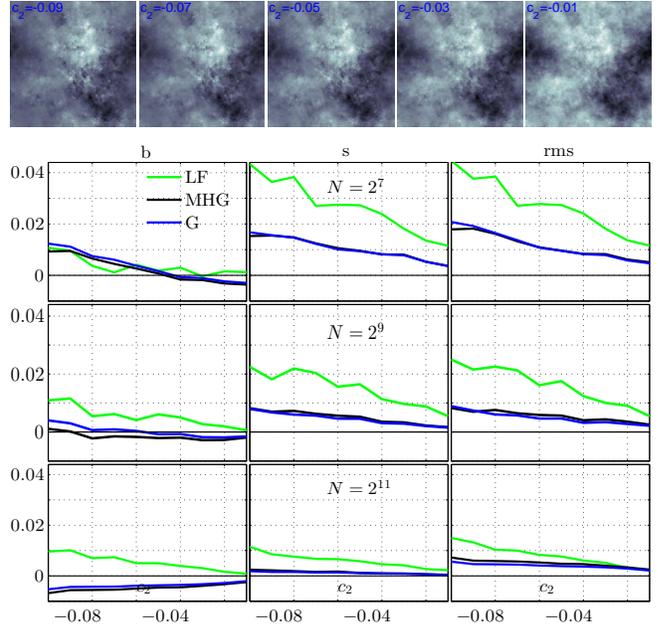
**Experimental setup.** MRW parameters are set to  $H = 0.72$  and  $c_2 \in \{-0.1, -0.09, \dots, -0.01\}$ . The 2D DWT uses a Daubechies' mother wavelet with  $N_\psi = 2$  vanishing moments and the linear regression weights  $w_j$  are chosen proportional to  $n_j$  (cf. [1, 11, 12]). For both Bayesian estimators, frequencies considered in (7) are restricted to  $|\omega_m| < \pi/4$  as in [2, 14]. The estimation performance is quantified through the bias, the standard deviation (STD) and the root mean squared error (RMSE), respectively defined by  $b = \mathbb{E}[\hat{c}_2] - c_2$ ,  $s = (\text{Var}[\hat{c}_2])^{1/2}$  and  $\text{rms} = \sqrt{b^2 + s^2}$  and computed for 200 independent realizations of MRW of size  $N \times N$ .

**Estimation of  $c_2$ .** Fig. 1 investigates estimation performance as a function of  $c_2$  for three different image sizes  $N \in \{2^7, 2^9, 2^{11}\}$ . Clearly, the Bayesian estimators compete favorably when compared to the LF estimator. Notably, the STD are up to 4 times smaller, resulting in significantly lower RMSE values. The Bayesian estimators have bias comparable to LF for the smallest image size considered, and significantly smaller bias for larger images. The estimators MHG and G build upon the same initial model (6) and thus unsurprisingly exhibit similar performance with only a slight difference for the bias which arises from the different priors (uniform for MHG and Jeffreys for G) and vanishes for large image sizes as expected.

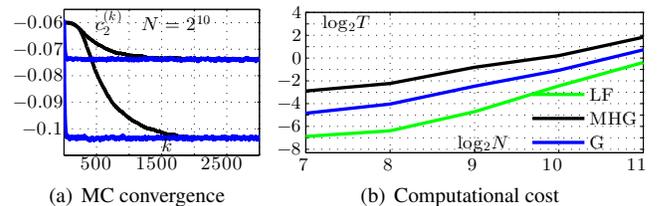
**Convergence.** Fig. 2(a) displays the evolution of Markov Chains (MCs) of G and MHG averaged over 200 realizations (with identical random initializations for both algorithms;  $N = 2^{10}$ ,  $c_2 = \{-0.07, -0.1\}$ ). Clearly, the augmented model used in G leads to MCs that converge almost immediately, while MHG requires a much longer burn-in period: Indeed, the accept-reject procedure of MHG requires tuning of the (adaptive) step size of the random walk propositions, which is not necessary for G. Consequently, G enables us to use much shorter MCs than MHG. Here, we assume that 500 samples are required in the sum (22) and hence MCs of length  $N_{mc} = \{600, 3000\}$  must be sampled for G and MHG, respectively.

**Computational cost.** Fig. 2(b) plots the computational times  $T$  (in seconds; defined as the total time to compute an estimate of  $\theta$  from an image  $X$ ) as a function of image size  $N$  for LF, MHG and G. Clearly, LF is the fastest approach. However, while the computational time for MHG is up to 25 times greater than that of LF, the Bayesian estimator G proposed here is between 5 (small  $N$ ) and 2 (large  $N$ ) times slower than the linear regression based estimation LF, which clearly demonstrates the efficiency of the proposed extended Fourier domain model and algorithm.

Overall, these results illustrate that the proposed estimator G is an operational alternative to linear regression, significantly improving estimation performance at only  $\sim 2$ – $5$  times the computational cost.



**Fig. 1.** Top: realizations of MRW for different values of  $c_2$ . Bottom: estimation performance for image sizes  $N = \{2^7, 2^9, 2^{11}\}$  with  $j_1 = \{1, 2, 2\}$  and  $j_2 = \{3, 5, 7\}$  (top to bottom rows, respectively): bias, STD and RMSE values (from left to right) for linear fit (LF), the proposed method (G) and the method in [14] (MHG).



**Fig. 2.** Evolution of MCs for  $N = 2^{10}$  and  $c_2 = \{-0.07, -0.1\}$  (a) and total computational cost as a function of image size (b).

## 6. CONCLUSIONS AND PERSPECTIVES

In this paper, a novel Bayesian framework to estimate the multifractality parameter  $c_2$  for images was proposed. It builds on a generic statistical model for the multivariate statistics of log-leaders of MMC based processes that was recently introduced in [2, 14] and relies on the following original key contributions. First, a generative model in the Fourier domain was developed based on the Whittle approximation. Then, the joint parameter constraints inherent to this model were reformulated as independent positivity constraints. This finally enabled the proposition of an extended likelihood which leads to conditional posterior distributions that are standard laws, contrary to the model in [2, 14]. Its main virtue is that the computation of the associated Bayesian estimators can be achieved efficiently with an MCMC algorithm involving only Gibbs sampling. The proposed procedure yields excellent estimation performance, both for small and large image sizes, at computational cost comparable to classical linear regression based estimation. Future work will include the analysis of multivariate images (e.g., multi-band, multi-temporal) or of multiple image patches for which more informative priors (e.g., modeling correlations between parameters of different modalities) could be efficiently handled within the framework introduced here.

## 7. REFERENCES

- [1] H. Wendt, S. G. Roux, S. Jaffard, and P. Abry, “Wavelet leaders and bootstrap for multifractal analysis of images,” *Signal Process.*, vol. 89, no. 6, pp. 1100–1114, 2009.
- [2] 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 Trans. Image Process.*, vol. 24, no. 8, pp. 2540–2551, 2015.
- [3] R. H. Riedi, “Multifractal processes,” in *Theory and applications of long range dependence*, P. Doukhan, G. Oppenheim, and M.S. Taqqu, Eds. 2003, pp. 625–717, Birkhäuser.
- [4] S. Jaffard, “Wavelet techniques in multifractal analysis,” in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, Proc. Symp. Pure Math.*, M. Lapidus and M. van Frankenhuysen, Eds. 2004, vol. 72(2), pp. 91–152, AMS.
- [5] J. F. Muzy, E. Bacry, and A. Arneodo, “The multifractal formalism revisited with wavelets,” *Int. J. of Bifurcation and Chaos*, vol. 4, pp. 245–302, 1994.
- [6] R. Lopes and N. Betrouni, “Fractal and multifractal analysis: A review,” *Medical Image Analysis*, vol. 13, pp. 634–649, 2009.
- [7] B. B. Mandelbrot and J. W. van Ness, “Fractional Brownian motion, fractional noises and applications,” *SIAM Review*, vol. 10, pp. 422–437, 1968.
- [8] B. B. Mandelbrot, “A multifractal walk down Wall Street,” *Sci. Am.*, vol. 280, no. 2, pp. 70–73, 1999.
- [9] B. Castaing, Y. Gagne, and M. Marchand, “Log-similarity for turbulent flows,” *Physica D*, vol. 68, no. 3-4, pp. 387–400, 1993.
- [10] H. Wendt, S. Jaffard, and P. Abry, “Multifractal analysis of self-similar processes,” in *Proc. IEEE Workshop Statistical Signal Process. (SSP)*, Ann Arbor, USA, 2012.
- [11] H. Wendt, P. Abry, and S. Jaffard, “Bootstrap for empirical multifractal analysis,” *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.
- [12] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, “Multiscale nature of network traffic,” *IEEE Signal Proc. Mag.*, vol. 19, no. 3, pp. 28–46, 2002.
- [13] T. Lux, “Higher dimensional multifractal processes: A GMM approach,” *J. Business & Economic Stat.*, vol. 26, pp. 194–210, 2007.
- [14] S. Combrexelle, H. Wendt, J.-Y. Tourneret, P. Abry, and S. McLaughlin, “Bayesian estimation of the multifractality parameter for images via a closed-form Whittle likelihood,” in *Proc. 23rd European Signal Process. Conf. (EUSIPCO)*, Nice, France, 2015.
- [15] M. A. Tanner and W. H. Wong, “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.
- [16] D. A. van Dyk and X.-L. Meng, “The art of data augmentation,” *J. Comput. Graph. Statistics*, vol. 10, no. 1, pp. 1–50, 2001.
- [17] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 3rd edition, 2008.
- [18] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali, *Two-Dimensional Wavelets and their Relatives*, Cambridge University Press, 2004.
- [19] P. Whittle, “On stationary processes in the plane,” *Biometrika*, vol. 41, pp. 434–449, 1954.
- [20] M. Fuentes, “Approximate likelihood for large irregularly spaced spatial data,” *J. Am. Statist. Assoc.*, vol. 102, pp. 321–331, 2007.
- [21] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York, 1994.
- [22] V. V. Anh and K. E. Lunney, “Parameter estimation of random fields with long-range dependence,” *Math. Comput. Model.*, vol. 21, no. 9, pp. 67–77, 1995.
- [23] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, USA, 2005.
- [24] R. Robert and V. Vargas, “Gaussian multiplicative chaos revisited,” *Ann. Proba.*, vol. 38, no. 2, pp. 605–631, 2010.