A Comparison of Bayesian Estimators for the Parameters of the Bivariate Multifractal Spectrum

Lorena Leon, Herwig Wendt, Jean-Yves Tourneret
Institut de Recherche en Informatique de Toulouse
(IRIT), Université de Toulouse, CNRS,
Toulouse INP, UT3, UT2, Toulouse, France
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 provides the theoretical and practical tools for describing the fluctuations of pointwise regularity in data and has led to many successful applications in signal and image processing. Originally limited to the analysis of single time series or images, a definition of multivariate multifractal analysis, i.e., the joint multifractal analysis of several data components, was recently proposed and was shown to effectively quantify local or transient dependencies in data regularity, beyond linear correlation. However, the accurate estimation of the associated matrix-valued joint multifractality parameters is notoriously difficult, thus limiting its practical usefulness. Leveraging a recent statistical model for bivariate multifractality, the goal of this work is to define and study Bayesian estimators designed to bypass this difficulty. Specifically, we study the original use of two different priors, combined with two different averages (arithmetic and Karcher means), for bivariate multifractal analysis. Monte Carlo simulations with synthetic data allow us to appreciate their relative performance and to conclude that our novel and original estimator based on a scaled inverse Wishart prior and the Karcher mean yields particularly favorable results with up to 5 times smaller root-mean-squared error than previous formulations.**

***Index Terms*—Multifractal analysis, multivariate data, Bayesian estimation, wavelet leaders, Karcher mean**

I. Introduction

Context. Multifractal analysis (MFA) is a rich and versatile mathematical analysis and modeling framework that makes it possible to characterize a function $X(t) : \mathbb{R}^d \to \mathbb{R}$ based on the geometry of the fluctuations of its *pointwise regularity* [1]. Over the last decades, MFA has been successfully used in a broad range of contexts of different natures (cf., e.g., [2] and references therein). However, this success has been mainly limited to the analysis of univariate data (scalar time series or single channel images). Yet, they are often the constituting parts of multivariate data (e.g., physical quantities jointly registered by several sensors, or several bands in a color space). Conducting a joint analysis instead of individually analyzing the data components provides a richer characterization of multifractality, and in particular unravels the dynamics, coupling mechanisms and dependencies among the different components. To this end, a theoretical foundation for a multivariate multifractal analysis has recently been proposed [3], [4]. First tests on synthetic data confirmed that the multivariate multifractality parameters can indeed effectively capture and quantify transient or local data dependencies that cannot be accounted for by second order statistics such as the linear correlation function [5], [6]. However, these experiments also revealed that the accurate estimation of the associated cross-multifractality parameters is challenging, in particular for small sample sizes.

Related works. The estimation of multifractality parameters is standardly performed by means of linear regressions for multiscale statistics of specific multiresolution coefficients in log-diagrams, see Section II-B and, e.g., [7]. To reduce the estimation variance for small sample size, a statistical modeling and estimation framework for the multifractality parameters of bivariate data has recently been proposed [8] (see, e.g., [9], [10] for earlier works for univariate data). The strategy consists in conducting estimation with a Bayesian model for the matrix-valued multivariate multifractality parameters of interest. The preliminary numerical results obtained in [8] using independent inverse Wishart prior distributions are encouraging and indicate that the Bayesian estimation can significantly outperform standard linear regression based estimation. However, [8] also showed that estimations can be significantly biased, in particular for small sample sizes.

Goals, contributions and outline. The goal of this work is to propose and study several alternative Bayesian models and estimators for bivariate multifractal analysis and to compare their estimation performance, with specific focus on the critical issue of accurate estimation for small sample size. To this end, we first briefly summarize in Section II theoretical bivariate multifractal analysis and practical formalisms and the statistical model proposed in [8]. Our main contributions are as follows. First, we devise a novel Bayesian model for bivariate multifractal analysis, making use of more flexible scale inverse Wishart priors than the standard inverse Wishart priors of [8], and specify the conditional distributions of a Gibbs sampler for approximating the parameters of interest via simulation (cf. Section III). Second, we advocate the use of a mean associated with a Riemanian metric (aka, the Karcher mean [11]) instead of the usual arithmetic mean for the numerical evaluation of the minimum mean square error (MMSE) estimator for the matrix-valued multifractality parameters. Finally, in Section IV, we compare in detail four different Bayesian estimators (combining inverse Wishart/scaled inverse Wishart priors and

arithmetic/Karcher means) by means of various Monte Carlo simulations performed on synthetic data. Our results show that the proposed novel estimator combining a scaled inverse Wishart prior distribution and the Karcher mean significantly improves estimation performance for small sample sizes.

II. Bivariate multifractal analysis and statistical modeling

A. Bivariate multifractal spectrum

Bivariate multifractal analysis amounts to characterizing the *joint* and the single-component fluctuations along time of the pointwise regularity of a signal or function $\boldsymbol{X}(t) = (X_1(t), X_2(t)) \in \mathbb{R}^2$, $t \in \mathbb{R}$. The pointwise regularity of the rth component of $\boldsymbol{X}$ is usually quantified by the so-called Hölder exponent $h_{X_r}(t) \geq 0$, see, e.g., [1] for details; the closer $h_{X_r}(t)$ to 0, the more irregular X_r around position t. Therefore, a global geometrical description of the pointwise regularity of $\boldsymbol{X}$ is provided through the so-called *bivariate multifractal spectrum* $\mathcal{D}(h_1, h_2)$ of $\boldsymbol{X}$ [3], [4], [12]. It is defined as the collection of Hausdorff dimensions $\dim_H$ of the sets of points t with the same pair of Hölder exponents $\boldsymbol{h} = (h_1, h_2)$, i.e.,

$$\mathcal{D}(\boldsymbol{h}) \triangleq \dim_H \big\{ t : (h_{X_1}(t), h_{X_2}(t)) = \boldsymbol{h} \big\}.$$

The state-of-the-art procedure for the estimation of the multifractal spectrum is constructed from multiscale statistics of wavelet leaders and is summarized next [1], [3], [4], [7].

B. Bivariate multifractal formalism using wavelet leaders

Wavelet leaders. Let ψ denote a mother wavelet, which is an oscillating reference pattern that is characterized by its number of vanishing moments N_ψ, a positive integer defined as $\psi \in \mathcal{C}^{N_\psi - 1}$ and $\forall n = 0, \ldots, N_\psi - 1$, $\int_{\mathbb{R}} t^n \psi(t) dt \equiv 0$ and $\int_{\mathbb{R}} t^{N_\psi} \psi(t) dt \neq 0$. It is designed 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 dilated and translated templates forms an orthonormal basis of $L^2(\mathbb{R})$ [13]. The L^1 normalized discrete wavelet transform coefficients $d_X(j,k)$ of $X \in \mathbb{R}$ are defined as $d_X(j,k) = 2^{-j/2} \langle \psi_{j,k} | X \rangle$, and the *wavelet leaders* of X are defined as

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

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

Bivariate multifractal formalism. The bivariate cumulants $C_{p_1 p_2}(j)$ of the vector of log-leaders at scale j, $\boldsymbol{\ell}_{\boldsymbol{X}}(j,k) \triangleq (\ln L_{X_1}(j,k), \ln L_{X_2}(j,k)) \in \mathbb{R}^2$, $k \in \{1, \ldots, n_j\}$, can be shown to take the form $C_{p_1 p_2}(j) = c^0_{p_1 p_2} + j\, c_{p_1 p_2} \ln 2$, with $p_1 + p_2 \geq 1$ [5], [14]. The coefficients $c_{p_1 p_2}$ yield the following approximation of $\mathcal{D}(\boldsymbol{h})$

$$\mathcal{D}(h_1, h_2) \approx 1 + \frac{c_{02} b}{2} \left(\frac{h_1 - c_{10}}{b} \right)^2 + \frac{c_{20} b}{2} \left(\frac{h_2 - c_{01}}{b} \right)^2 - c_{11} b \left(\frac{h_1 - c_{10}}{b} \right) \left(\frac{h_2 - c_{01}}{b} \right), \quad (1)$$

where $c_{20}, c_{02} < 0$, $b \triangleq c_{20} c_{02} - c_{11}^2 \geq 0$ [6] and (c_{10}, c_{01}) indicates the position of the maximum of $\mathcal{D}(\boldsymbol{h})$, which corresponds to the average degrees of data regularity. However, this position does not convey information on the joint multifractality (aka joint regularity fluctuations). The coefficients c_{20} and c_{02} quantify the amount of pointwise regularity fluctuations (multifractality) for each component independently and c_{11} characterizes the coupling between the regularity fluctuations of the components. The standard estimation strategy for the coefficients $c_{01}, c_{10}, c_{02}, c_{20}, c_{11}$ relies on linear regressions of sample cumulants across scale j.

C. Statistical model

The centered vector of the logarithm of wavelet leaders $\boldsymbol{l}_{\boldsymbol{X}}(j,k) = [l_{X_1}(j,k), l_{X_2}(j,k)]^T$ can be well modeled as a Gaussian random vector [8], with a point covariance matrix parametrized by two 2×2 positive definite (p.d.) real-valued matrices $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma}_2$, i.e., $\boldsymbol{l}_{\boldsymbol{X}}(j,k) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_1} \ln 2^{-j} + \boldsymbol{\Sigma}_2)$. Specifically, $\boldsymbol{\Sigma_1} \triangleq -\begin{pmatrix} c_{20} & c_{11} \\ c_{11} & c_{02} \end{pmatrix}$ contains the multifractal parameters of interest and $\boldsymbol{\Sigma}_2$ is used for model adjustment. The joint distribution for the vectors $\boldsymbol{l}_{\boldsymbol{X}}(j,k)$ and $\boldsymbol{l}_{\boldsymbol{X}}(j',k')$ at same scale, $j = j'$, has a fixed non-trivial temporal covariance, while vectors at different scales, $j \neq j'$, are modeled as independent. We resorted to a Whittle approximation [15], [16] to diagonalize the temporal covariance and the joint distribution in the Fourier domain, leading to a model for the Fourier coefficients $\boldsymbol{z}_j = \mathcal{F}_{I_j}(\boldsymbol{l}_{\boldsymbol{X}}(j,\cdot)) \in \mathbb{C}^{(n_j - 1) \times 2}$. The operator $\mathcal{F}_{I_j}(\cdot)$ computes the discrete Fourier coefficients contained in $I_j = [\![\lfloor (-n_j - 1)/2 \rfloor, n_j - \lfloor n_j/2 \rfloor]\!] \setminus \{0\}$ ($[\![a_1, a_2]\!]$ denotes the set of integers ranging from a_1 to a_2) with frequencies $\{\omega_m = 2\pi m / n_j\}_{m \in I_j}$ and $\lfloor \cdot \rfloor$ truncates to integer values. The likelihood for $\boldsymbol{z} = \big[\boldsymbol{z}_{j_1}^T, \ldots, \boldsymbol{z}_{j_2}^T\big]^T \in \mathbb{C}^{M \times 2}$ reads

$$\begin{aligned} p(\boldsymbol{z} \mid \boldsymbol{\Sigma_1}, \boldsymbol{\Sigma}_2) &= \prod_{j=j_1}^{j_2} p(\boldsymbol{z}_j \mid \boldsymbol{\Sigma_1}, \boldsymbol{\Sigma}_2) \\ &\propto (\det \boldsymbol{\Omega})^{-1} \exp\big(-\tilde{\boldsymbol{z}}^H \boldsymbol{\Omega}^{-1} \tilde{\boldsymbol{z}}\big), \end{aligned} \quad (2)$$

where $\tilde{\boldsymbol{z}}$ is the $2M \times 1$ vector resulting from the concatenation of the two components of $\boldsymbol{z}$ and $\cdot^H$ denotes the conjugate transpose operator. The matrix $\boldsymbol{\Omega} \triangleq \boldsymbol{\Sigma_1} \otimes \boldsymbol{G}_1 + \boldsymbol{\Sigma}_2 \otimes \boldsymbol{G}_2$ is a real-valued block diagonal p.d. covariance matrix, where the operator $\otimes$ denotes the Kronecker product. In particular, the matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are deterministic diagonal matrices that subsume the covariance model in time.

In order to facilitate inference, we used in [8] a data augmentation strategy for model (2), introducing a complex-valued vector of latent variables $\boldsymbol{u} \in \mathbb{C}^{M \times 2}$, with the following conditional distributions $\boldsymbol{z} \mid \boldsymbol{u}, \boldsymbol{\Sigma_1} \sim \mathcal{CN}(\boldsymbol{u}, \boldsymbol{\Sigma_1} \otimes \boldsymbol{G}_1)$ and $\boldsymbol{u} \mid \boldsymbol{\Sigma}_2 \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_2 \otimes \boldsymbol{G}_2)$, where $\mathcal{CN}$ denotes the complex Gaussian distribution, where the pseudo-covariance matrix is omitted since it is the zero matrix. This results in the augmented likelihood

$$p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Sigma_1}, \boldsymbol{\Sigma}_2) \propto p(\boldsymbol{z} \mid \boldsymbol{u}, \boldsymbol{\Sigma_1}) p(\boldsymbol{u} \mid \boldsymbol{\Sigma}_2). \quad (3)$$

By construction, the likelihood (2) is obtained by marginalizing (3) with respect to $\boldsymbol{u}$. The advantage of using (3) with

respect to (2), is that, in (3), the matrix-valued parameters $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are no longer additively tied but conditionally independent.

III. Bayesian models and estimation

A. Bayesian models

Likelihood. The Bayesian models investigated in this paper are based on the augmented likelihood (3), which is the product of complex Gaussian distributions with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

Inverse Wishart. The natural *conjugate prior* for $\boldsymbol{\Sigma}_i$ is the inverse Wishart (IW) prior [17], i.e., for $i \in \{1,2\}$, $\boldsymbol{\Sigma}_i \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$, with $\nu_i > 3$ ($\nu_i \in \mathbb{R}$) degrees of freedom and a $2-$dimensional p.d. real-valued scale matrix $\boldsymbol{\Lambda}_i$.

Scaled inverse Wishart. An alternative to the IW prior is the scaled inverse Wishart (SIW) prior studied in [18], which decouples the matrix $\boldsymbol{\Sigma}_i$ as $\boldsymbol{\Sigma}_i \triangleq \boldsymbol{\Delta}_i \boldsymbol{Q}_i \boldsymbol{\Delta}_i$, where the random matrices $\boldsymbol{Q}_i$ and $\boldsymbol{\Delta}_i$ are independent, $\boldsymbol{Q}_i \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$, and $\boldsymbol{\Delta}_i$ is a diagonal matrix with independent diagonal elements $\delta_{ir} = [\boldsymbol{\Delta}_i]_{rr}$ and $\delta_{ir} \sim \mathcal{LN}(\beta_{ir}, \alpha_{ir}^2)$, where $\mathcal{LN}$ is the log-normal distribution, i.e., $p(\boldsymbol{\Sigma}_i) = p(\boldsymbol{Q}_i)p(\delta_{i1})p(\delta_{i2})$, $i \in \{1,2\}$. It was shown that using this decomposition and semi-separate priors on standard deviations and correlation coefficients provides greater flexibility than the IW prior [18].

Posterior. The posterior distribution associated with the proposed Bayesian model for $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and the latent vector $\boldsymbol{u}$ can be computed from the Bayes theorem

$$p(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{u} \mid \boldsymbol{z}) \propto p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) p(\boldsymbol{\Sigma}_1) p(\boldsymbol{\Sigma}_2), \quad (4)$$

which can be used to define the marginal MMSE estimator $\widehat{\boldsymbol{\Sigma}}_i^{\text{MMSE}} \triangleq \mathbb{E}[\boldsymbol{\Sigma}_i \mid \boldsymbol{z}, \boldsymbol{u}]$. This estimator is difficult to be expressed using a simple closed form expression. Thus, we proposed to compute an approximation resulting from a Markov chain Monte Carlo (MCMC) algorithm [19].

B. Gibbs samplers

In order to estimate the unknown model parameters from the posterior, we consider a Gibbs sampler, successively generating samples $\{\boldsymbol{\Sigma}_1^{(\lambda)}, \boldsymbol{\Sigma}_2^{(\lambda)}, \boldsymbol{u}^{(\lambda)}\}_{\lambda=1}^{N_{\text{mc}}}$ according to the conditional distributions of (4), either using the IW prior or its scaled version.

Inverse Wishart. When using an IW prior for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, the resulting conditional distribution are

$$\boldsymbol{\Sigma}_i \mid \boldsymbol{z}, \boldsymbol{u} \sim \mathcal{IW}(\nu_i + 2M, \boldsymbol{\Lambda}_i + \tilde{\boldsymbol{\Phi}}_i), \quad (5)$$

$$\boldsymbol{u} \mid \boldsymbol{z}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \sim \mathcal{CN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}), \quad (6)$$

where $\tilde{\boldsymbol{\Phi}}_1 = 2\,\text{Re}(\sum_{s=1}^{M} (\boldsymbol{z}_s - \boldsymbol{u}_s)(\boldsymbol{z}_s - \boldsymbol{u}_s)^H g_{1s}^{-1})$ and $\tilde{\boldsymbol{\Phi}}_2 = 2\,\text{Re}(\sum_{s=1}^{M} \boldsymbol{u}_s \boldsymbol{u}_s^H g_{2s}^{-1})$, with $\boldsymbol{z}_s = \boldsymbol{z}^T(s), \boldsymbol{u}_s = \boldsymbol{u}^T(s) \in \mathbb{C}^{2\times 1}$ and $g_{is} = [\boldsymbol{G}_i]_{ss}$ for $i = \{1,2\}$. Moreover, $\tilde{\boldsymbol{\Sigma}}$ is a block diagonal matrix whose sth block is $\tilde{\boldsymbol{\Sigma}}_s = \left[(g_{1s}\boldsymbol{\Sigma}_1)^{-1} + (g_{2s}\boldsymbol{\Sigma}_2)^{-1}\right]^{-1}$, and $\tilde{\boldsymbol{\mu}}_s = \tilde{\boldsymbol{\Sigma}}_s (g_{1s}\boldsymbol{\Sigma}_1)^{-1} \boldsymbol{z}_s$.

Scaled inverse Wishart. Assuming an SIW prior for $\boldsymbol{\Sigma}_i$, the conditional distributions of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in (5) are replaced by the following sampling steps for all $i, r = \{1,2\}$

$$\boldsymbol{Q}_i \mid \boldsymbol{\Delta}_i, \boldsymbol{z}, \boldsymbol{u} \sim \mathcal{IW}(\nu_i + 2M, \boldsymbol{\Lambda}_i + \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Phi}}_i \boldsymbol{\Delta}_i^{-1}), \quad (7)$$

$$\delta_{ir} \sim f(\delta_{ir}) = p(\delta_{ir} \mid \boldsymbol{Q}_i, \{\delta_{ir'}\}_{r' \neq r}, \boldsymbol{z}, \boldsymbol{u}). \quad (8)$$

Indeed, $f(\delta_{ir})$ is not a standard distribution and is sampling using the log-conditional distribution

$$\begin{aligned} \ln f(\delta_{ir}) = &-(2M+1)\ln \delta_{ir} - [\boldsymbol{Q}_i^{-1}]_{rr}[\tilde{\boldsymbol{\Phi}}_i]_{rr}(2\delta_{ir}^2)^{-1} \\ &- \delta_{ir}^{-1}\delta_{ir'}^{-1}[\boldsymbol{Q}_i^{-1}]_{rr'}[\tilde{\boldsymbol{\Phi}}_i]_{rr'} \\ &- (\ln \delta_{ir} - \beta_{ir})^2/(2\alpha_{ir}^2) + \text{constant}. \end{aligned}$$

We thus use a Metropolis-Hastings random walk procedure for sampling each δ_{ir} in turn. The proposal distribution is chosen here as a real-valued Gaussian distribution whose location parameter is the current value δ_{ir}° and the scale parameter $\sigma_{\delta_{ir}}^2$ is adaptively chosen to ensure an acceptance rate between 0.4 and 0.6 [19]. We insert the draws of $\boldsymbol{Q}_i$ and $\{\delta_{ir}\}_{r=1}^2$ into $\boldsymbol{\Sigma}_i \triangleq \boldsymbol{\Delta}_i \boldsymbol{Q}_i \boldsymbol{\Delta}_i$ to generate samples of $\boldsymbol{\Sigma}_i$. The conditional distribution of $\boldsymbol{u} \mid \boldsymbol{z}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ is similar to (6).

The chosen prior distributions for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ guarantee the sampled matrices are p.d. matrices along the iterations. Finally, after a burn-in period, where the first N_{bi} samples are discarded, the MMSE estimator of $\boldsymbol{\Sigma}_i$ is approximated by averaging over the set of p.d. real-valued matrices $\{\boldsymbol{\Sigma}_i^{(\lambda)}\}_{\lambda=N_{\text{bi}}+1}^{N_{\text{mc}}}$, where $\boldsymbol{\Sigma}_i^{(\lambda)}$ is the λth matrix generated by the sampler.

C. Approximation of the MMSE estimator

The *average* over the space of the p.d. real-valued matrices can be computed using the arithmetic mean associated with the Euclidean metric, $\widehat{\boldsymbol{\Sigma}}_i^{\text{A}} = (N_{\text{mc}} - N_{\text{bi}})^{-1} \sum_{\lambda=N_{\text{bi}}+1}^{N_{\text{mc}}} \boldsymbol{\Sigma}_i^{(\lambda)}$, which is by construction also p.d. Alternatively, one can use the mean associated with the Riemannian metric (the geometric mean or Karcher mean), see [11] for details. The Karcher mean $\widehat{\boldsymbol{\Sigma}}_i^{\text{K}}$ is the unique p.d. symmetric matrix-valued solution to the nonlinear matrix equation $\sum_{\lambda=N_{\text{bi}}+1}^{N_{\text{mc}}} \ln((\boldsymbol{\Sigma}_i^{(\lambda)})^{-1}\boldsymbol{X}) = \boldsymbol{0}$ that can be solved, e.g., using the iterative algorithm proposed in [20]. Unlike the arithmetic mean, the Karcher mean has the interesting property that it commutes with the matrix inverse [11].

Below, we denote by $\text{SIW}_{(\cdot)}$ and $\text{IW}_{(\cdot)}$ the MMSE estimators resulting from using the IW or SIW priors, and use the sub-indexes $(\cdot)_A$ and $(\cdot)_K$ for the arithmetic and Karcher means.

IV. Performance evaluation

The estimation performance is evaluated by comparing the estimates $\hat{\theta}$ of $\theta \in \{-c_{20}, -c_{02}, -c_{11}\}$ obtained using 100 realizations of a bMRW (defined in the next paragraph) for different sample sizes and a large range of values of multifractality parameters. More precisely, we computed the bias (BIAS), $b(\hat{\theta}) = \widehat{\mathbb{E}}[\hat{\theta}] - \theta$, the standard deviation (STD), $s(\hat{\theta}) = \sqrt{\widehat{\text{Var}}[\hat{\theta}]}$, and the root mean square error (RMSE), $r(\hat{\theta}) = \sqrt{b(\hat{\theta})^2 + s(\hat{\theta})^2}$ of the estimates, where $\widehat{\mathbb{E}}[.]$ and $\widehat{\text{Var}}[.]$ denote the sample mean and the sample variance.

Bivariate multifractal random walk (bMRW). The construction of a bMRW process $\boldsymbol{X}(t) = (X_1(t), X_2(t))$ [5] is based on two pairs of stochastic processes: a pair of increments of fractional Brownian motions $(G_1(t), G_2(t))$, which

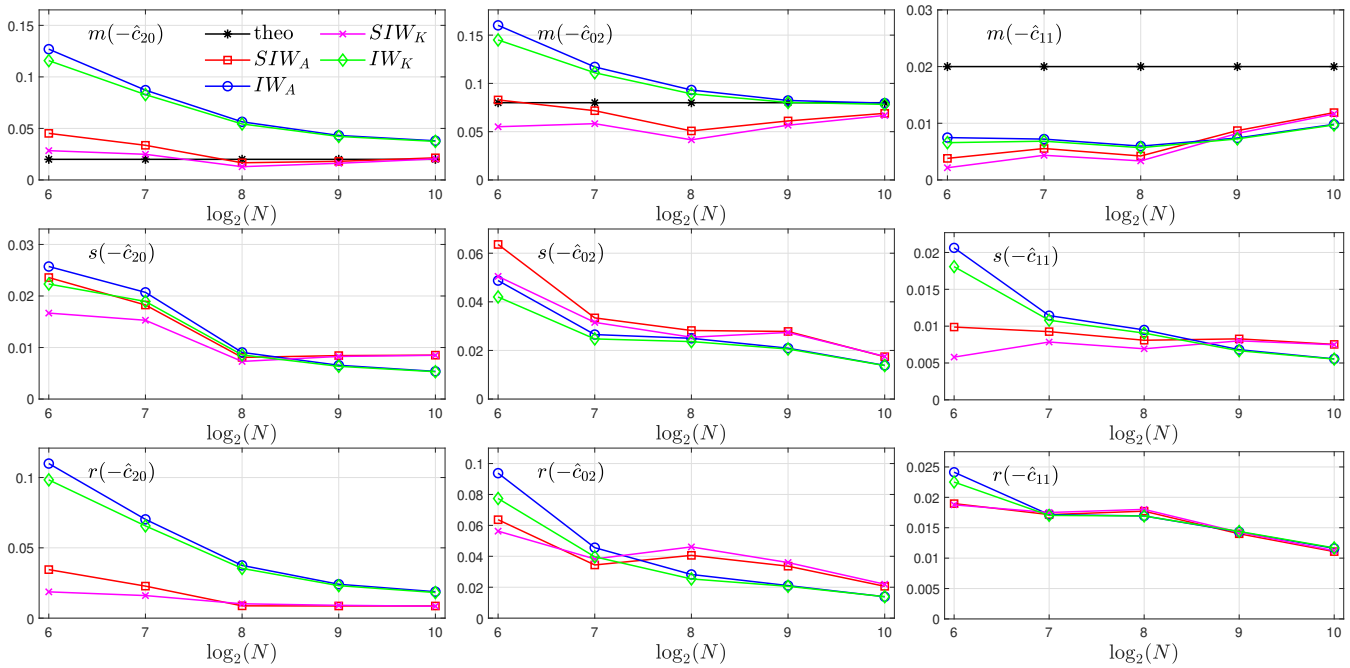

Fig. 1. bMRW estimation performance for $N \in \{2^6, 2^7, 2^8, 2^9, 2^{10}\}$, $j_1 = \{1, 1, 2, 2, 2\}$ and $j_2 = \log_2(N) - 4$.

is determined by two self-similarity parameters H_1 and H_2 and a point covariance $\mathbf{\Sigma}_{ss}$, and a pair of Gaussian processes $(\omega_1(t), \omega_2(t))$ with prescribed covariance function $\mathbf{\Sigma}_{\text{mf}}$, with entries given by $[\mathbf{\Sigma}_{\text{mf}}]_{rv}(k,l) = [\boldsymbol{\rho}_{\text{mf}}]_{rv} \lambda_r \lambda_v \ln\left(\frac{T}{|k-l|+1}\right)$, with $r, v \in \{1, 2\}$, for $|k-l| \leq T-1$ and 0 otherwise, where T is an arbitrary integral scale. To simplify notations, we consider $[\boldsymbol{\rho}_{\text{mf}}]_{11} = [\boldsymbol{\rho}_{\text{mf}}]_{22} = 1$, and $[\boldsymbol{\rho}_{\text{mf}}]_{12} = \rho_{\text{mf}}$. Both pairs of processes are numerically synthesized as described in [21]. Finally, each component $\{X_r\}_{r=1}^2$, of bMRW is defined as the primitive of the product process $G_r e^{\omega_r}$. The bivariate multifractality properties of bMRW are given by $c_{10} = H_1 + \lambda_1^2/2$, $c_{01} = H_2 + \lambda_2^2/2$, $c_{20} = -\lambda_1^2$, $c_{02} = -\lambda_2^2$, and $c_{11} = -\rho_{\text{mf}}\lambda_1\lambda_2$ [5], [22]; in addition, $\forall(p_1, p_2)$ such that $p_1 + p_2 \geq 3$, $c_{p_1 p_2} \equiv 0$.

Monte Carlo simulations and parameter settings. The parameters of the bMRW are set to $(H_1, H_2) = (0.72, 0.72)$, $\lambda_1 = \sqrt{0.02}$, $\lambda_2 \in \{\sqrt{0.02}, \sqrt{0.04}, \sqrt{0.06}, \sqrt{0.08}, \sqrt{0.1}\}$, $\mathbf{\Sigma}_{ss} = \mathbb{I}_2$, $\rho_{\text{mf}} = 0.5$, sample size $N \in \{2^6, 2^7, 2^8, 2^9, 2^{10}\}$ and the integral scale is set to $T = N$. The wavelet analysis is conducted with a Daubechies least asymmetric wavelet, with $N_\psi = 3$ and scales ranging from $j_1 = \{1, 1, 2, 2, 2\}$ to $j_2 = \log_2(N) - 4$. The Gibbs samplers use a burn in of $N_{\text{bi}} = 1000$ samples and a total number of iterations equal to $N_{\text{mc}} = 2000$. The hyperparameters for the (scaled) inverse Wishart priors are set to $\nu_1 = \nu_2 = 4$, $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \mathbb{I}_2$ as in [23]. In addition, for the scaled inverse Wishart prior, we used cross validation to set $(\beta_{ir}, \alpha_{ir}^2) = (0.1, 1)$, for all $i, r \in \{1, 2\}$.

Performance vs. sample size. Fig. 1 plots averages, STD and RMSE for the estimates obtained from the SIW_A, SIW_K, IW_A and IW_K algorithms as a function of the sample size N. We observe that when the sample size increases, the average for each estimation approaches the true value of the parameter, indicating that the estimators are asymptotically unbiased. One can also observe that the STD and RMSE values decrease as N increases, as expected.

A closer look on the respective performance leads to the following observations. First, when the sample size is small, the BIAS of the $\text{SIW}_{(\cdot)}$ estimators of c_{20} and c_{02} is significantly smaller (up to 40 times) when compared to the $\text{IW}_{(\cdot)}$ estimators, whereas for c_{11} the BIAS values of these estimators are similar. As mentioned above, the estimators converge to the same values for large sample sizes. In general, the STD are slightly larger for $\text{SIW}_{(\cdot)}$. Overall, this leads to comparable RMSE values for large sample size, and to better performance for $\text{SIW}_{(\cdot)}$ for small sample sizes.

The arithmetic and Karcher mean yield similar performance for large sample sizes. For small sample size, the Karcher mean has better performance. This small sample regime will be investigated with more details in the next paragraphs.

Performance vs. multifractal parameter values. Table I shows the estimation performance for small sample size $N = 2^6$ and for several multifractality settings, i.e., $-c_{02} \in \{0.02, 0.04, \ldots, 0.1\}$ and leads to the following conclusions.

A comparison of the results obtained with the different priors shows that the estimators $\text{SIW}_{(\cdot)}$ have significantly reduced BIAS for c_{20} and c_{02}, and similar BIAS for c_{11} when compared to $\text{IW}_{(\cdot)}$. This conclusion is valid for all levels of multifractality for c_{02}. Moreover, the STD values are smaller for $\text{SIW}_{(\cdot)}$ than for $\text{IW}_{(\cdot)}$, leading overall to significantly reduced RMSE for $\text{SIW}_{(\cdot)}$ in all cases. As far as the matrix averages are concerned, smaller RMSEs are consistently obtained when the Karcher mean is used to approximate the MMSE estimator.

TABLE I

bMRW ESTIMATION PERFORMANCE FOR $N = 2^6$ AND $-c_{02} \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$ (BEST RESULTS IN **BOLD**).

		$-c_{02}$	0.02	0.04	0.06	0.08	0.1
$-c_{20} = 0.02$	BIAS	SIW_A	0.0252	0.0266	0.0299	0.0278	0.0237
		SIW_K	**0.0075**	**0.0099**	**0.0130**	**0.0107**	**0.0073**
		IW_A	0.1052	0.1045	0.1063	0.1061	0.1029
		IW_K	0.0945	0.0938	0.0953	0.0948	0.0922
	STD	SIW_A	0.0272	0.0307	0.0386	0.0394	0.0301
		SIW_K	**0.0181**	**0.0229**	**0.0295**	**0.0278**	**0.0208**
		IW_A	0.0339	0.0304	0.0376	0.0410	0.0330
		IW_K	0.0295	0.0267	0.0328	0.0349	0.0286
	RMSE	SIW_A	0.0370	0.0407	0.0488	0.0482	0.0383
		SIW_K	**0.0196**	**0.0249**	**0.0322**	**0.0298**	**0.0221**
		IW_A	0.1105	0.1088	0.1128	0.1138	0.1080
		IW_K	0.0990	0.0975	0.1008	0.1011	0.0965
$-c_{02}$	BIAS	SIW_A	0.0301	0.0209	**0.0047**	**0.0027**	**0.0015**
		SIW_K	**0.0118**	**0**	0.0179	0.0270	0.0335
		IW_A	0.1080	0.0985	0.0852	0.0728	0.0703
		IW_K	0.0968	0.0863	0.0719	0.0589	0.0540
	STD	SIW_A	0.0378	0.0376	0.0408	0.0610	0.0717
		SIW_K	**0.0280**	**0.0291**	**0.0301**	**0.0481**	**0.0518**
		IW_A	0.0386	0.0334	0.0419	0.0483	0.0620
		IW_K	0.0337	0.0292	0.0362	0.0420	0.0537
	RMSE	SIW_A	0.0483	0.0430	0.0411	0.0611	0.0717
		SIW_K	**0.0304**	**0.0291**	**0.0350**	**0.0551**	**0.0616**
		IW_A	0.1147	0.1040	0.0950	0.0874	0.0937
		IW_K	0.1025	0.0911	0.0805	0.0723	0.0762
$-c_{11}$	BIAS	SIW_A	**0.0097**	0.0109	0.0146	0.0171	0.0185
		SIW_K	0.0098	0.0123	0.0158	0.0184	0.0204
		IW_A	0.0105	**0.0079**	**0.0116**	**0.0121**	**0.0140**
		IW_K	0.0105	0.0085	0.0122	0.0130	0.0150
	STD	SIW_A	0.0060	0.0086	0.0085	0.0106	0.0103
		SIW_K	**0.0028**	**0.0058**	**0.0053**	**0.0069**	**0.0063**
		IW_A	0.0165	0.0168	0.0172	0.0189	0.0198
		IW_K	0.0145	0.0148	0.0151	0.0166	0.0171
	RMSE	SIW_A	0.0115	0.0139	0.0169	0.0201	**0.0212**
		SIW_K	**0.0102**	**0.0136**	**0.0167**	**0.0197**	0.0213
		IW_A	0.0196	0.0185	0.0208	0.0224	0.0242
		IW_K	0.0179	0.0171	0.0194	0.0211	0.0228

Overall, this leads us to conclude that for large sample size, a precise choice of the prior and the averaging operation has little impact on estimation performance. Conversely, for small sample size, the estimator with best performance is the MMSE estimator that combines the SIW prior with Karcher mean, with up to 5 times smaller RMSE values.

V. CONCLUSIONS

This paper studied and compared four different Bayesian estimators for the parameters of the bivariate multifractal spectrum. These Bayesian estimators were based on a recent statistical model for the likelihood of log-leaders. Moreover, they combined the use of either an inverse Wishart prior or its scaled version with two different averaging strategies (arithmetic mean or Karcher mean) to approximate the marginal minimum mean square error estimator of the unknown multifractality parameters. Various Monte Carlo simulations with synthetic bivariate data demonstrated that all four combinations lead to accurate parameter estimates and are essentially equivalent for large sample sizes. For small sample size, our results reveal that the novel estimator combining a scaled inverse Wishart prior and the Karcher mean yields the best results, with root mean square errors up to 5 times smaller than those of the other estimators. Future work will extend this study to multivariate images and focus on applications to multi-component physiological signals.

ACKNOWLEDGMENT

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

REFERENCES

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

[2] P. Abry, H. Wendt, S. Jaffard, and G. Didier, "Multivariate scale-free temporal dynamics: From spectral (Fourier) to fractal (wavelet) analysis," *Comptes Rendus Physique*, vol. 20, no. 5, pp. 489–501, 2019.

[3] S. Jaffard, S. Seuret, H. Wendt, R. Leonarduzzi, S. Roux, and P. Abry, "Multivariate multifractal analysis," *Applied and Computational Harmonic Analysis*, vol. 46, no. 3, pp. 653–663, 2019.

[4] S. Jaffard, S. Seuret, H. Wendt, R. Leonarduzzi, and P. Abry, "Multifractal formalisms for multivariate analysis," *Proc. Royal Society A*, vol. 475, no. 2229, 2019.

[5] H. Wendt, R. Leonarduzzi, P. Abry, S. Roux, S. Jaffard, and S. Seuret, "Assessing cross-dependencies using bivariate multifractal analysis," in *IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Calgary, Canada, April 2018.

[6] R. Leonarduzzi, P. Abry, S. G. Roux, H. Wendt, S. Jaffard, and S. Seuret, "Multifractal characterization for bivariate data," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Rome, Italy, September 2018.

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

[8] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry, "Bayesian estimation for the parameters of the bivariate multifractal spectrum," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Dublin, Ireland, August 2021.

[9] 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.

[10] 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. on Imaging Sciences (SIIMS)*, vol. 11, no. 2, pp. 1294–1316, 2018.

[11] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, Mar. 2005.

[12] C. Meneveau, K.R. Sreenivasan, P. Kailasnath, and M.S. Fan, "Joint multifractal measures - theory and applications to turbulence," *Phys. Rev. A*, vol. 41, no. 2, pp. 894–913, 1990.

[13] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, CA, 2nd edition, 2009.

[14] B. Castaing, Y. Gagne, and M. Marchand, "Log-similarity for turbulent flows," *Physica D*, vol. 68, no. 3-4, pp. 387–400, 1993.

[15] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat.*, vol. 2, no. 5, pp. 423–434, 1953.

[16] A. M. Sykulski, S. C. Olhede, J. M. Lilly, and J. J. Early, "Frequency-domain stochastic modeling of stationary bivariate or complex-valued signals," *IEEE T. Signal Process.*, vol. 65, no. 12, pp. 3136–3151, 2017.

[17] J. Barnard, R. McCulloch, and X.-L. Meng, "Modelling covariance matrices in terms of standard deviations and correlations, with application to shrinkage," *Statistica Sinica*, vol. 10, pp. 1281–1311, October 2000.

[18] A. J. O'Malley and A. M. Zaslavsky, "Domain-level covariance analysis for multilevel survey data with structured nonresponse," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1405–1418, 2008.

[19] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, USA, 2005.

[20] D. A. Bini and B. Iannazzo, "Computing the Karcher mean of symmetric positive definite matrices," *Linear Algebra and its Applications*, vol. 438, no. 4, pp. 1700–1710, 2013, 16th ILAS Conference Proceedings, Pisa 2010.

[21] H. Helgason, V. Pipiras, and Patrice Abry, "Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding," *Signal Processing*, vol. 91, no. 5, pp. 1123–1133, 2011.

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

[23] I. Alvarez, J. Niemi, and M. Simpson, "Bayesian inference for a covariance matrix," in *Conf. Applied Statistics in Agriculture*, 2014.