

A Bayesian Framework for Multivariate Multifractal Analysis

Lorena Leon, Herwig Wendt, Jean-Yves Tournet, Patrice Abry

Abstract—Multifractal analysis has become a reference tool for signal and image processing. Grounded in the quantification of local regularity fluctuations, it has proven useful in an increasing range of applications, yet so far involving only univariate data (scalar valued time series or single channel images). Recently the theoretical ground for multivariate multifractal analysis has been devised, showing potential for quantifying transient higher-order dependence beyond linear correlation among collections of data. However, the accurate estimation of the parameters associated with a multivariate multifractal model remains challenging, especially for small sample size data. This work studies an original Bayesian framework for multivariate multifractal estimation, combining a novel and generic multivariate statistical model, a Whittle-based likelihood approximation and a data augmentation strategy allowing parameter separability. This careful design enables efficient estimation procedures to be constructed for two relevant choices of priors using a Gibbs sampling strategy. Monte Carlo simulations, conducted on synthetic multivariate signals and images with various sample sizes and multifractal parameter settings, demonstrate significant performance improvements over the state of the art, at only moderately larger computational cost. Moreover, we show the relevance of the proposed framework for real-world data modeling in the important application of drowsiness detection from multichannel physiological signals.

Index Terms—multifractal analysis, multivariate, Bayesian estimation, multiscale, wavelet leaders

I. INTRODUCTION

Context: Multivariate multifractal analysis. Over the last decades, multifractal analysis has grown into a standard tool that has been successfully involved in various signal/image processing tasks (classification, detection, etc, see e.g., [1] for a review). In essence, multifractal analysis probes the temporal dynamics in time series, or the spatial dynamics (textures) in images, by quantifying the strengths and topological-geometrical structures of the fluctuations of the pointwise regularity of the data. This is achieved with the multifractal spectrum, the object of central interest to multifractal analysis, see [2], [3] and below for details. Numerical implementations for the estimation of the multifractal spectrum or the associated multifractal parameters commonly involve multiscale (e.g., wavelet-based) representations of the data [4]–[6]. Multifractal analysis led to significant successes in many real-world applications in very different contexts (cf., e.g., [7]–[13] or [14]

and references therein). However, its theoretical grounding remained, up to a recent past, fundamentally univariate in principle, hence tied to the independent analysis of single time series or images. This is a severe limitation for its practical use because many recent applications entail the joint analysis of signals or images recorded for the same system using different sensors, and crucial information is potentially conveyed in the coupling and dependencies between components. While this limitation had been recognized early on and partially addressed in specific applicative contexts [15], [16], the theoretical foundation for multivariate multifractal analysis was laid only recently [17], [18]. Its first practical use showed that the multivariate multifractal spectrum can effectively capture and quantify transient, local dependencies in data that cannot be considered by second order statistics [14], [19]–[21]. The first experiments also revealed that the estimation of the associated cross-multifractality parameters is a challenging issue, which we propose to tackle in the present work.

Related works: Multifractal estimation. Practical multifractal analysis, or multifractal estimation, fundamentally relies on multiscale representations of the data. While sometimes based on multifractal detrended fluctuation analysis [22], multifractal estimation is also often based on wavelets. In particular, elaborating on the early proposition of using wavelet transform modulus maxima representations [4], it is now well documented that multifractal analysis benefits from being constructed on wavelet leaders, defined as local nonlinear transformations of the wavelet coefficients [2], [3], [5], [6]. Whatever the multiscale representations used, multifractal estimation relies in essence on log-log regressions, intrinsically requiring the data to have a long enough sample size in order to allow their dynamics to develop along a set of scales ranging across several orders of magnitude. These regressions lead to large estimation variances, notably for limited sample size data, and become a critical challenge, for example, in several bivariate settings as explored in [19]–[21]. This limitation was first addressed for 2D images by adopting a Bayesian approach for the estimation of the scalar-valued multifractality parameters [23], [24]. The preliminary work presented in [25] considered, for the first time, the joint analysis of bivariate signals (i.e., pairs of time series) with the recently theoretically defined bivariate multifractal spectrum and proposed a Bayesian estimation framework for the estimation of the associated multifractality parameter triplet. In the present work, we propose and study a general framework for the estimation of the matrix-valued parameters associated with the joint multifractal spectrum of multivariate signals and images.

Manuscript received 11/08/2021, revised 4/3/2022, accepted 20/06/2022.

Lorena Leon, Herwig Wendt and Jean-Yves Tournet are with the Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse, CNRS, Toulouse INP, UT3, UT2, TêSA, Toulouse, France (firstname.lastname@irit.fr).

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

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

Goals, contributions and outline. The goal of this work is to propose a Bayesian framework for multivariate multifractal signal and image analysis and to assess its estimation performance. To that end, a theoretical multivariate multifractal analysis and log-log regression-based multifractal estimation are briefly recalled in Section II. Then, the first key contribution presented in Section III consists in devising an original Bayesian model for the multivariate log-leader representations of the data. This model combines the validation of a multivariate Gaussian distribution, the multiscale modeling of its covariances, the derivation of a Whittle-type approximation [26] of the corresponding likelihood and a data augmentation strategy leading to a likelihood that is separable with respect to the multifractal parameters to be estimated. Separability further permits, as a second main contribution, to propose efficient estimation algorithms, based on relevant choices of prior distributions, the derivation of the conditional distributions of the joint distribution of interest simplifying the use of Gibbs sampling strategies (Section IV). As a third contribution, the estimation performance of the proposed Bayesian frameworks for multivariate multifractal estimation are assessed and compared against classical linear regressions, and for two different priors, using extensive Monte Carlo simulations relying on synthetic multivariate multifractal processes for signal and images as defined in [19], [20]. Performance results, reported in Section V for different multifractal parameter settings and sample sizes, and for both signals and images, demonstrate a significant improvement in estimation performance achieved at moderate extra computational cost. This opens the way for the practical use of multivariate multifractal analysis on real-world data. Finally, as a fourth contribution, Section VI studies the application of the proposed framework to drowsiness detection from polysomnographic data. Our results are comparable to the state of the art, demonstrating the relevance of our approach for real-world data modeling. Bayesian multivariate multifractal estimation procedures (together with synthesis procedures for multivariate multifractal processes) will be made publicly available via a documented toolbox at the time of publication.

Notation. Bold lowercase symbols represent vectors, whereas bold uppercase symbols denote matrices. The determinant and the trace of a matrix are denoted as $\det(\cdot)$ and $\text{tr}(\cdot)$. Superscripts \cdot^H, \cdot^* and \cdot^T denote the Hermitian transpose, the complex conjugation and the transpose operators. The matrix denoted $\text{diag}(a_1, a_2, \dots, a_M)$ is the $M \times M$ diagonal matrix with a_1, a_2, \dots, a_M along its diagonal. The element at row u and column v of a matrix \mathbf{A} is denoted as $[\mathbf{A}]_{u,v}$. Further, \otimes denotes the Kronecker product, $\text{Re}(\cdot)$ stands for the real part operator, $\llbracket a_1, a_2 \rrbracket$ is the set of integers ranging from a_1 to a_2 and the operator $\lfloor \cdot \rfloor$ truncates to integer values. The symbol \sim is used for “is distributed according to” and \propto means “up to a multiplicative constant”. The normal, log-normal, complex normal (CN), uniform, inverse Wishart (IW) and scaled inverse Wishart (SIW) distributions are denoted as $\mathcal{N}, \mathcal{LN}, \mathcal{CN}, \mathcal{U}, \mathcal{IW}$ and \mathcal{SIW} .

II. MULTIVARIATE MULTIFRACTAL ANALYSIS

A. Multifractal spectrum

For a univariate function $X(\mathbf{t}) \in \mathbb{R}$, $\mathbf{t} \in \mathbb{R}^d$ (e.g., $\mathbf{t} \in \mathbb{R}$ for

time series and $\mathbf{t} \in \mathbb{R}^2$ for images), multifractal analysis provides a quantification of the fluctuations along time or space of its regularity. The Hölder exponent $h(\mathbf{t}) \geq 0$ is used to measure the pointwise regularity (see, e.g., [2] for details). The closer $h(\mathbf{t})$ to 0, the more irregular X around position \mathbf{t} . For an R -variate function $\mathbf{X}(\mathbf{t}) = [X_1(\mathbf{t}), \dots, X_R(\mathbf{t})]^T \in \mathbb{R}^R$, the Hölder exponent is denoted as $\mathbf{h}_{\mathbf{X}}(\mathbf{t}) \triangleq [h_1(\mathbf{t}), \dots, h_R(\mathbf{t})]^T$ where $h_r(\mathbf{t})$ is the exponent associated with the r th component $X_r(\mathbf{t})$. The *multivariate multifractal spectrum* $\mathcal{D}_R(\mathbf{h})$ of \mathbf{X} is defined as the collection of Hausdorff dimensions \dim_H of the sets of points \mathbf{t} at which $\mathbf{h}_{\mathbf{X}}(\mathbf{t})$ takes the value $\mathbf{h} = [h_1, \dots, h_R]^T$, i.e.,

$$\mathcal{D}_R(\mathbf{h}) \triangleq \dim_H \{ \mathbf{t} : \mathbf{h}_{\mathbf{X}}(\mathbf{t}) = \mathbf{h} \}, \quad (1)$$

see [15], [17], [18] for details. It provides a global, geometrical description of the pointwise regularity of \mathbf{X} . Specifically, its precise shape, width, and orientation with respect to the \mathbf{h} -axes quantify the degree of local fluctuation of the regularity of the components of \mathbf{X} , and to what extent these fluctuations are coupled between components. The state of the art procedure for the estimation of the multifractal spectrum is constructed from the multiscale statistics of wavelet leaders and is summarized in the next section [2], [3], [17], [18].

B. Multifractal formalism using wavelet leaders

Wavelet leaders. For $d = 1$ and $R = 1$, 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 $\forall n = 0, \dots, N_\psi - 1$, $\int_{\mathbb{R}} t^n \psi(t) dt \equiv 0$ and $\int_{\mathbb{R}} t^{N_\psi} \psi(t) dt \neq 0$. It is designed such that the collection $\{ \psi_{j,\mathbf{k}}(\mathbf{t}) = 2^{-j/2} \psi(2^{-j} \mathbf{t} - \mathbf{k}) \}_{(j,\mathbf{k}) \in (\mathbb{Z}, \mathbb{Z}^d)}$ of its dilated and translated templates forms an orthonormal basis of $L^2(\mathbb{R})$ [27]. The L^1 normalized discrete wavelet coefficients $d_X(j, \mathbf{k})$ of $X \in \mathbb{R}$ are defined as $d_X(j, \mathbf{k}) = 2^{-j/2} \langle \psi_{j,\mathbf{k}}, X \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product. Then, for $d = 1$, the *wavelet leaders* of X are defined as

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

where $\lambda_{j,\mathbf{k}} = [\mathbf{k}2^j, (\mathbf{k} + 1)2^j]$ denotes the dyadic interval of size 2^j and $3\lambda_{j,\mathbf{k}}$ stands for the union of $\lambda_{j,\mathbf{k}}$ with its 2 neighbors. In higher dimensions $d \geq 2$, Q coefficients $d_X^{(q)}(j, \mathbf{k})$, $q = 1, \dots, Q$, are computed using the tensor product of 1D transforms (see, e.g., [28] for details), and the supremum in the definition of the wavelet leaders is also taken for q and for all $3^d - 1$ direct neighbors.

Multifractal formalism. Wavelet leaders reproduce Hölder exponents in the limit of fine scales, $L_X(j, \mathbf{k}) \simeq C 2^{-j h(\mathbf{t})}$ for $C > 0$, $2^j \rightarrow 0$, $\mathbf{k}2^{-j} \rightarrow \mathbf{t}$ (\mathbf{t} fixed) [2], [3]. The empirical moments of $L_X(j, \mathbf{k})$ for $2^j \rightarrow 0$ behave as [18]

$$\frac{1}{n_j} \sum_{\mathbf{k} \in \llbracket 1, n_j^{1/d} \rrbracket^d} \prod_{r=1}^R L_{X_r}(j, \mathbf{k})^{p_r} \approx K_{\mathbf{p}} 2^{j \zeta_R(\mathbf{p})}, \quad (2)$$

where $\mathbf{p} = [p_1, \dots, p_R]^T$, $K_{\mathbf{p}} > 0$ and $n_j \approx \lfloor N^d / 2^{dj} \rfloor$ is the number of wavelet leaders at scale j of a single data component. The so-called scaling exponents $\zeta_R(\mathbf{p})$ in (2) are

tightly related to $\mathcal{D}_R(\mathbf{h})$ via their Legendre transform referred to as the multivariate Legendre spectrum

$$\mathcal{L}_R(\mathbf{h}) \triangleq \inf_{\mathbf{p}} (d + \langle \mathbf{p}, \mathbf{h} \rangle - \zeta_R(\mathbf{p})). \quad (3)$$

Finally, $\mathcal{L}_R(\mathbf{h})$ provides an upper-bound for $\mathcal{D}_R(\mathbf{h})$ for large classes of processes and is in practice used as an estimator of $\mathcal{D}_R(\mathbf{h})$, see, e.g., [18]. Let $\ell_{X_r}(j, \mathbf{k}) \triangleq \ln L_{X_r}(j, \mathbf{k})$ denote the logarithm of the wavelet leaders (aka log-leaders) associated with X_r at a fixed scale j and position \mathbf{k} . The first order (mean) and second order (auto- and cross-covariances) cumulants of $\ell_{X_r}(j, \mathbf{k})$ and $\ell_{X_{r'}}(j, \mathbf{k})$, with $r, r' \in \{1, \dots, R\}$, take the form [19], [29]

$$\mathbb{E}[\ell_{X_r}(j, \mathbf{k})] = c_1^0(r) + c_1(r) \ln 2^j, \quad (4)$$

$$\text{Cov}(\ell_{X_r}(j, \mathbf{k}), \ell_{X_{r'}}(j, \mathbf{k})) = c_2^0(r, r') + c_2(r, r') \ln 2^j, \quad (5)$$

where c^0 are model adjustment parameters not related to the multifractal properties. Similar expressions are obtained for the higher-order R -variate cumulants [14], [30]. The use of (4) and (5) implicitly amounts to a parabolic approximation of the pairwise scaling exponents $\zeta_2(p_r, p_{r'})$ which yields a pairwise parabolic approximation of the multifractal spectrum $\mathcal{D}_2(h_r, h_{r'})$ around its maximum [14], [18]:

$$\begin{aligned} \mathcal{D}_2(h_r, h_{r'}) \approx & d + \frac{c_2(r', r')b}{2} \left(\frac{h_r - c_1(r)}{b} \right)^2 \\ & + \frac{c_2(r, r')b}{2} \left(\frac{h_{r'} - c_1(r')}{b} \right)^2 \\ & - c_2(r, r')b \left(\frac{h_r - c_1(r)}{b} \right) \left(\frac{h_{r'} - c_1(r')}{b} \right), \quad (6) \end{aligned}$$

where $c_2(r, r) < 0$, $b \triangleq c_2(r, r)c_2(r', r') - c_2(r, r')^2 \geq 0$ [20] and

- $(c_1(r), c_1(r'))$ indicates the position of the maximum of $\mathcal{D}_2(h_r, h_{r'})$, which corresponds to the average degrees of data regularity. Note that it does not convey information on the joint multifractality (aka joint regularity fluctuations) and will not be considered in the proposed model.
- $c_2(r, r)$ quantifies the amount of pointwise regularity fluctuations (multifractality) for the r th component.
- $c_2(r, r')$ characterizes the coupling between the regularity fluctuations of the r th and r' th components.

The standard estimation procedure for the coefficients c_1 and c_2 relies on linear regressions of sample cumulants across scale j , as suggested by (4) and (5).

It is natural to define the normalized coupling parameter

$$\rho_{\text{mf}}(r, r') \triangleq - \frac{c_2(r, r')}{\sqrt{c_2(r, r)c_2(r', r')}} \in [-1, 1]. \quad (7)$$

In view of the model defined in Section II-C, it can be interpreted as a *multifractal correlation* and can be shown to quantify *higher-order dependence beyond linear correlation* among the data components $\{X_r\}_{r=1}^R$ [19]–[21]. An estimator of ρ_{mf} can be defined by replacing the coefficients in (7) by estimates [19]. The type of information that can be captured by ρ_{mf} is illustrated in Fig. 1, which shows a 3-variate synthetic multifractal image (defined in the following paragraph), and the magnitude of its isotropic image gradients, with positive and negative values for ρ_{mf} among its components. The linear correlation equals zero for all components. Nevertheless, the

$$X_1, \rho_{\text{mf}}(1, 3) = -0.99 \quad X_2, \rho_{\text{mf}}(1, 2) = 0.99 \quad X_3, \rho_{\text{mf}}(2, 3) = -0.99$$

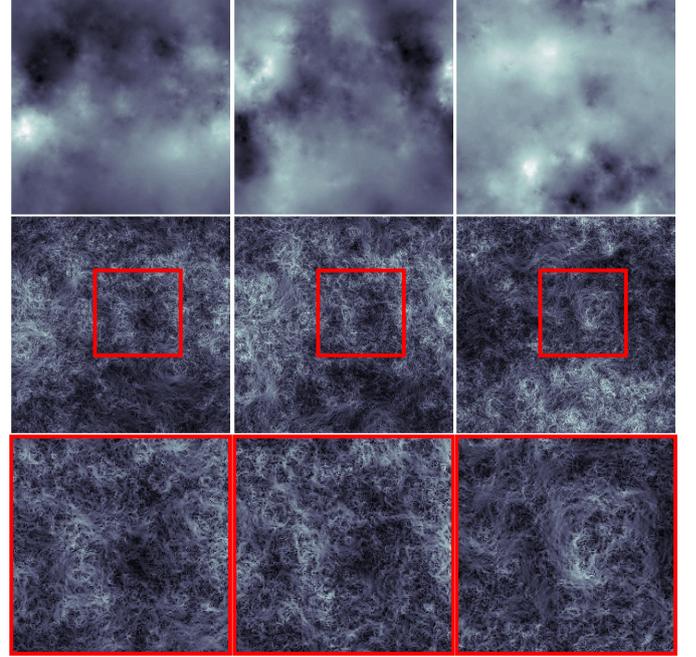


Fig. 1. Illustrations for multifractal correlation: synthetic 3-variate multifractal image (top row, defined in [19], [31] and Section II-C), magnitude of gradients (second row) and zooms of the patch marked by red square (bottom row). X_1 and X_2 have multifractal correlation $\rho_{\text{mf}}(1, 2) = 0.99$, and $\rho_{\text{mf}}(1, 3) = \rho_{\text{mf}}(2, 3) = -0.99$ with component X_3 . The linear correlation equals zero for all components.

image gradient magnitudes reveal the strong co-organization and dependence between the components. Indeed, for components with positive ρ_{mf} , large gradients tend to co-occur at the same locations, while they tend to be coupled with small gradients when ρ_{mf} is negative.

C. Multivariate multifractal random walk

The multivariate multifractal random walk (MV-MRW) [19], [31], [32] is the canonical multifractal model process for multivariate data and is used here to illustrate the proposed approach and assess its performance.

Definition. The construction of an MV-MRW for R components requires two collections of stochastic processes: (i) a collection of increments of fractional Brownian motions $(G_1(\mathbf{t}), \dots, G_R(\mathbf{t}))$, which is determined by the self-similarity parameters H_1, \dots, H_R and an $R \times R$ point covariance Σ_{ss} , with corresponding correlation coefficients $\rho_{ss}(r, r')$, and (ii) a collection of Gaussian processes $(\omega_1(\mathbf{t}), \dots, \omega_R(\mathbf{t}))$ with prescribed covariance function Σ_{mf} , with entries given by $[\Sigma_{\text{mf}}]_{rr'}(\mathbf{k}, \mathbf{l}) = [\mathbf{\Pi}_{\text{mf}}]_{rr'} \lambda_r \lambda_{r'} \ln \left(\frac{T}{\|\mathbf{k} - \mathbf{l}\|_2 + 1} \right)$, $r, r' \in \{1, \dots, R\}$, for $\|\mathbf{k} - \mathbf{l}\|_2 \leq T - 1$ and 0 otherwise, where T is an arbitrary integral scale, equal to the data sample size in the rest of the paper. To simplify notations, we consider $[\mathbf{\Pi}_{\text{mf}}]_{rr} = 1$ and $[\mathbf{\Pi}_{\text{mf}}]_{rr'} = \rho_{\text{mf}}(r, r')$. These processes are numerically synthesized as described in [33]. Each component X_r , $r \in \{1, \dots, R\}$, of an MV-MRW is then defined as the primitive of the product $G_r e^{\omega_r}$.

Multifractal properties. The multifractality parameters of an MV-MRW are given by $c_1(r) = H_r + \lambda_r^2/2$, $c_2(r, r) = -\lambda_r^2$, and $c_2(r, r') = -\rho_{\text{mf}}(r, r') \lambda_r \lambda_{r'}$ [19], [31], whereas

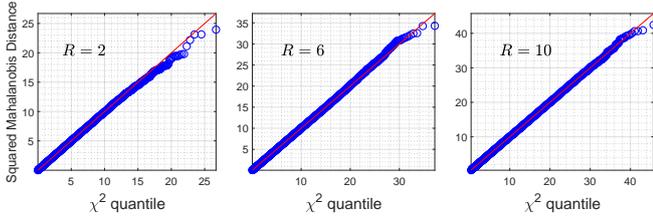


Fig. 2. Gamma plots for the joint distribution of the empirical log-leaders at scale $j = 4$, associated with 100 independent $2^{10} \times 2^{10} \times R$ synthetic images generated using a 2D MV-MRW, with $R = 2, 6, 10$ (from left to right; $\lambda_r = \sqrt{r}/100$, $\rho_{mf}(r, r')$ uniform in $[0, 0.5]$). The closer to the red line, the better the approximation of the distribution by an R -variate normal distribution.

its higher-order cumulants are equal to zero. Examples of realizations of an MV-MRW (2D) for $R = 3$ are plotted in Fig. 1. Typical values of c_2 for real-world data range from zero (no multifractality) down to $c_2 \approx -0.25$, which corresponds to an extremely intermittent signal that is rarely observed for a non-pathological physical signal.

III. STATISTICAL MODEL FOR MULTIVARIATE LOG-LEADERS

Estimations based on log-log regressions lead to large variance, in particular for the second order cumulants c_2 and for ρ_{mf} [19]–[21]. Instead, we propose a second-order statistical model for the vector of log-leaders $\ell(j, \mathbf{k}) \triangleq [\ell_{X_1}(j, \mathbf{k}), \dots, \ell_{X_R}(j, \mathbf{k})]^T \in \mathbb{R}^R$ for multivariate multifractal data. This model will be shown to be useful to estimate the multifractality parameters in multivariate scenarios using a Bayesian framework.

A. Direct model

Marginal distributions. In the univariate case, theoretical arguments suggest that the marginal distributions of multiresolution coefficients of multifractal processes are approximately log-normal [34]. This has been studied and confirmed numerically for univariate wavelet leaders in [23], [24]. It is hence natural to extend this modeling to the multivariate case $R > 1$. Numerical simulations for synthetic multivariate multifractal processes as defined in Section II-C, for large ranges of sample sizes and multifractal parameter values, suggest that the empirical distribution of the log-leaders $\ell(j, \mathbf{k})$ can indeed be well approximated by an R -variate Gaussian distribution. Illustrative examples are given in Fig. 2, which shows gamma plots (cf. e.g. [35], [36]) for scale $j = 4$, $d = 2$ and $R = 2, 6, 10$.

Covariance. The theoretical results derived for univariate random wavelet cascades in [37] suggest a linear asymptotic decay for the auto-covariance of log-leaders, see also [23], [24] for an empirical study for a larger class of single-variable multifractal processes. Theoretical arguments available in, e.g., [32], [38] suggested similar linear asymptotic behavior also for the cross-covariance terms for multivariate multifractal processes. Inspired by that, we propose a generic covariance model for log-leaders for multivariate multifractal processes, with the key novel ingredient of a cross-term that describes the covariance between log-leaders of different components, parametrized by $c_2(r, r')$. Assuming that the vector of log-leaders associated with the signal/image under analysis is

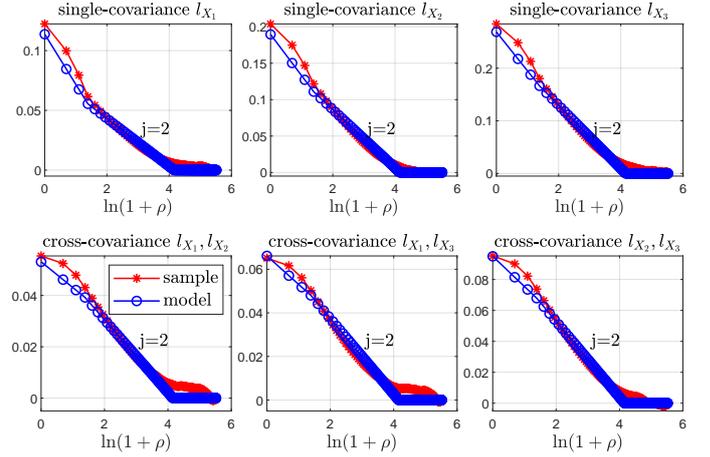


Fig. 3. Comparison between the proposed model (blue line) and the sample covariance (red line) at scale $j = 2$, averaged over 100 independent copies of an $2^{10} \times 2^{10} \times 3$ synthetic MV-MRW image with $(\lambda_1, \lambda_2, \lambda_3) = (\sqrt{0.02}, \sqrt{0.04}, \sqrt{0.06})$ and $\rho_{mf}(1, 2) = \rho_{mf}(1, 3) = \rho_{mf}(2, 3) = 0.5$. Plots correspond to each covariance matrix element for slice $\rho = \|[0, \Delta \mathbf{k}]\|_2$.

stationary and isotropic, the $R \times R$ auto-covariance matrix of the vectors $\ell(j, \mathbf{k})$ at a fixed scale j can be approximated by a radially symmetric function $\mathbf{S}_j(\rho)$ with $\rho \triangleq \|\Delta \mathbf{k}\|_2$ as

$$\text{Cov}(\ell(j, \mathbf{k}), \ell(j, \mathbf{k} + \Delta \mathbf{k})) \approx \mathbf{S}_j(\rho), \quad (8)$$

with

$$\mathbf{S}_j(\rho) = \mathbf{\Sigma}_1 f_1(j, \rho) + \mathbf{\Sigma}_2 f_2(j, \rho), \quad (9)$$

where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are two $R \times R$ symmetric real-valued matrices containing the unknown multifractal parameters to be estimated. Specifically, upon a change of sign, the elements of $\mathbf{\Sigma}_1$ equal the auto- and cross-multifractal parameters c_2 , i.e., $-\mathbf{[\Sigma}_1]_{rr'}$ = $c_2(r, r')$, with $r, r' \in \{1, \dots, R\}$. The matrix-valued parameter $\mathbf{\Sigma}_2$ is used for model adjustment at small lags $\rho \leq 3$, whose precise shape was found not to depend on the multifractality parameters and was modeled by a simple single parameter affine function. Moreover,

$$f_1(j, \rho) = \max(0, -\ln((\rho + 1)/(\rho_j + 1))), \quad (10)$$

$$f_2(j, \rho) = \max(0, 1 - \ln(\rho + 1)/\ln 4), \quad (11)$$

with $\rho_j = \lfloor n_j/\kappa \rfloor$ (with $\kappa = 5$ (1D) and $\kappa = 4$ (2D) fixed using cross-validation). For $i \in \{1, 2\}$ and all scales j , $f_i(j, \cdot)$ is a non-negative function. Therefore, assuming that $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive definite (p.d.) ensures that \mathbf{S}_j is positive semi-definite (p.s.d.) for all scales j , as it is the sum of two p.s.d. matrices [39]. Illustrative examples of this covariance model are provided in Fig. 3, which indicate good fit for the zero lag (5) and the covariance decay (10) that convey information on the multifractal parameters, and show slightly larger discrepancies for lags 1 and 2 covered by the model adjustment (11).

Likelihood. Let $\mathbf{l}(j, \mathbf{k})$ denote the centered vector of the log-leaders at a fixed scale j and position \mathbf{k} . The vector $\mathbf{l}_j \triangleq [\mathbf{l}(j, \mathbf{k}_1)^T, \dots, \mathbf{l}(j, \mathbf{k}_{n_j})^T]^T \in \mathbb{R}^{Rn_j}$ stacks all the vectors $\{\mathbf{l}(j, \mathbf{k})\}$ at scale j organized in lexicographic order. We assume here that \mathbf{l}_j and $\mathbf{l}_{j'}$ at different scales $j' \neq j$ are

independent.¹ Together with the above model, the likelihood of the vector $\mathbf{l} = [\mathbf{l}_{j_1}^T, \dots, \mathbf{l}_{j_2}^T]^T$ of the log-leaders at scales $j = \llbracket j_1, j_2 \rrbracket$ given Σ_1 and Σ_2 can be written as

$$p(\mathbf{l} \mid \Sigma_1, \Sigma_2) = \prod_{j=j_1}^{j_2} p(\mathbf{l}_j \mid \Sigma_1, \Sigma_2),$$

$$\propto \prod_{j=j_1}^{j_2} (\det \Xi_j)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{l}_j^T \Xi_j^{-1} \mathbf{l}_j\right), \quad (12)$$

where $\Xi_j = \Sigma_1 \otimes \mathbf{F}_{1j} + \Sigma_2 \otimes \mathbf{F}_{2j}$ is an $Rn_j \times Rn_j$ covariance matrix with $[\mathbf{F}_{ij}]_{u,v} = f_i(j, \|\mathbf{k}_u - \mathbf{k}_v\|_2)$, $i = 1, 2$. For any scale j , the positive definiteness of \mathbf{F}_{ij} can be assessed numerically to check that $\Sigma_i \otimes \mathbf{F}_{ij}$ and Ξ_j are p.d.

B. Whittle approximation

The numerical evaluation of the likelihood (12) requires the inversion of the matrix Ξ_j . For large sample size, this inversion can become computationally and numerically challenging. Even when the inversion is not problematic, one may want to use a faster solution based on an approximation. Therefore, we make use of a Whittle approximation [26] and approximate the time-domain likelihood (12) in the frequency domain by

$$p(\mathbf{l} \mid \Sigma_1, \Sigma_2) \approx \prod_{j=j_1}^{j_2} p_W(\mathbf{l}_j \mid \Sigma_1, \Sigma_2) \quad (13)$$

with

$$p_W(\mathbf{l}_j \mid \Sigma_1, \Sigma_2) \propto \prod_{\mathbf{m} \in I_j} (\det \tilde{\mathbf{S}}_{j,\mathbf{m}})^{-1} \exp\left(-\mathbf{z}_{j,\mathbf{m}}^H \tilde{\mathbf{S}}_{j,\mathbf{m}}^{-1} \mathbf{z}_{j,\mathbf{m}}\right), \quad (14)$$

where $\mathbf{z}_{j,\mathbf{m}} \in \mathbb{C}^R$ denotes the R -variate normalized discrete Fourier coefficient of $\mathbf{l}(j, \mathbf{k})$ at frequency $\omega_{j,\mathbf{m}} = 2\pi\mathbf{m}/n_j^{1/d}$,

$$\mathbf{z}_{j,\mathbf{m}} = n_j^{-1/d} \sum_{\mathbf{k} \in \llbracket 1, n_j^{1/d} \rrbracket^d} \mathbf{l}(j, \mathbf{k}) \exp(-i\mathbf{k}^T \omega_{j,\mathbf{m}}), \quad (15)$$

where $\mathbf{m} \in I_j \triangleq \llbracket \lceil (-n_j^{1/d} - 1)/2 \rceil, n_j^{1/d} - \lfloor n_j^{1/d}/2 \rfloor \rrbracket^d \setminus \{\mathbf{0}\}$. Moreover, the power spectral matrix $\tilde{\mathbf{S}}_j$ forms a Fourier pair with the covariance matrix \mathbf{S}_j [40], which can be approximated using a discrete Fourier transform of $f_i(j, \cdot)$,

$$g_i(\omega_{j,\mathbf{m}}) = \sum_{\mathbf{k} \in \llbracket -n_j^{1/d}, n_j^{1/d} \rrbracket^d} f_i(j, \|\mathbf{k}\|_2) \exp(-i\mathbf{k}^T \omega_{j,\mathbf{m}}), \quad (16)$$

with $i = 1, 2$. Thus,

$$\tilde{\mathbf{S}}_{j,\mathbf{m}} = \Sigma_1 g_1(\omega_{j,\mathbf{m}}) + \Sigma_2 g_2(\omega_{j,\mathbf{m}}). \quad (17)$$

For $i \in \{1, 2\}$ and any scale j , $f_i(j, \cdot)$ is a non-negative even function. Thus, $g_i(\omega_{j,\mathbf{m}})$ is real-valued and strictly positive. Since Σ_1 and Σ_2 are assumed to be p.d., $\tilde{\mathbf{S}}_j$ is a real-valued p.d. matrix for any scale j .

¹Note that the theoretical results in [37] and our numerical results suggest that the inter scale dependence between log-leaders could also be described by a model reminiscent of (8), with proper adjustments for decimation. This paper assumes independence to simplify the proposed model, which leads to reduced computational cost. The results reported in Section V demonstrate that this assumption is reasonable and leads to excellent performance.

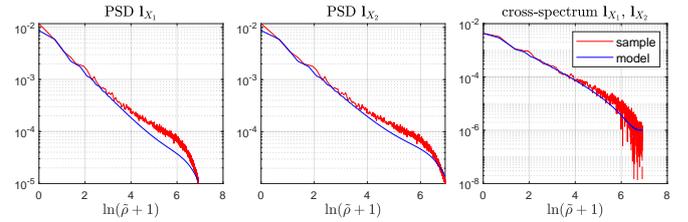


Fig. 4. Comparison between the sample power spectral matrix (red line), averaged over 100 independent realizations of a 2D MV-MRW for $R = 2$, $N = 2^{15}$, $(\lambda_1, \lambda_2) = (\sqrt{0.04}, \sqrt{0.08})$ and $\rho_{mf}(1, 2) = 0.6$, and the proposed model (blue line) obtained using (17). Plots correspond to slice $\tilde{\rho} \triangleq \|\mathbf{m}\|_2$ when $\mathbf{m} = [0, m_2]^T$.

The coefficient $\mathbf{z}_{j,\mathbf{m}}$ has the central symmetry property $\mathbf{z}_{j,\mathbf{m}} = \mathbf{z}_{j,-\mathbf{m}}^*$ since the log-leaders $\mathbf{l}(j, \mathbf{k})$ are real-valued. Also, the power spectral matrix $\tilde{\mathbf{S}}_{j,\mathbf{m}}$ has the same property. Thus, the product in (14) can be taken over the positive half of the total frequency grid.

Fig. 4 illustrates the fit between the models used for the power spectral densities (PSDs) and the cross power spectral density (CPSD) (computed using (17)) and their estimates. Simulation results obtained for a wide range of multifractal parameters evidence that the proposed model yields an excellent fit at low frequencies but larger deviation from the estimated (C)PSDs at high frequencies because of the coarser modeling of short time lags discussed above and potential aliasing due to the slow decay of the correlation function. Therefore, we propose a high-frequency cutoff introducing a bandwidth parameter η to control the fraction of the spectral grid that is actually used. Thus, the product in (14) is conducted using $\mathbf{m} \in I_j^\dagger = \{\mathbf{m} \in I_j : 0 < \tilde{\rho} \leq \sqrt{\eta} \lfloor n_j^{1/d}/2 \rfloor\}$ with $\tilde{\rho} \triangleq \|\mathbf{m}\|_2$. The value for η is obtained using cross validation in order to meet a bias-variance trade-off of the estimates. In particular, we set $\eta = 1$ (1D) and $\eta = 0.25$ (2D) in this paper.

C. Model in the Fourier domain

Expression (14) can be interpreted as a spectral likelihood, see, e.g., [41], [42], leading to model the Fourier coefficients $\mathbf{z}_{j,\mathbf{m}}$ by a frequency-independent random vector with a non-degenerate centered circular-symmetric complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \tilde{\mathbf{S}}_{j,\mathbf{m}})$. To simplify notation, we replace the sub-index $\cdot_{j,\mathbf{m}}$ of Section III-B by a single sub-index \cdot_s defined as a bijective function of (j, \mathbf{m}) on the set $\{1, \dots, M\}$, where M is the number of elements of the set $\{(j, \mathbf{m}) : j = \llbracket j_1, j_2 \rrbracket \text{ and } \mathbf{m} \in I_j^\dagger\}$. Therefore, the density of the vector $\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_M^T]^T \in \mathbb{C}^{MR}$, can be written as

$$p(\mathbf{z} \mid \Sigma_1, \Sigma_2) \propto \prod_{s=1}^M (\det \Omega_s)^{-1} \exp(-\mathbf{z}_s^H \Omega_s^{-1} \mathbf{z}_s), \quad (18)$$

where $\Omega_s \triangleq \Sigma_1 \mathbf{g}_{1,s} + \Sigma_2 \mathbf{g}_{2,s}$, with $g_{i,s} \triangleq g_i(\omega_s)$.

D. Data augmentation

Model (18) is simple and cheap to evaluate numerically compared to (12). However, its main inconvenience regarding the estimation of Σ_1 and Σ_2 is that these matrices are additively tied together in Ω_s , so that it is not possible to design

conjugate priors leading to simple conditional distributions (that will be used in the estimation algorithm). To bypass this difficulty, we use data augmentation (see, e.g., [43], [44] for more details) and introduce a complex-valued vector of latent variables $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_M^T] \in \mathbb{C}^{MR}$ as the hidden mean of the observed data \mathbf{z} leading to the augmented likelihood

$$p(\mathbf{z}, \mathbf{u} \mid \Sigma_1, \Sigma_2) = p(\mathbf{z} \mid \mathbf{u}, \Sigma_1)p(\mathbf{u} \mid \Sigma_2), \\ \propto (\det \Sigma_1)^{-M} \exp(-\text{tr}(\Sigma_1^{-1} \tilde{\Phi}_1)/2) \\ (\det \Sigma_2)^{-M} \exp(-\text{tr}(\Sigma_2^{-1} \tilde{\Phi}_2)/2), \quad (19)$$

with

$$\tilde{\Phi}_1 = 2 \text{Re} \left(\sum_{s=1}^M (\mathbf{z}_s - \mathbf{u}_s)(\mathbf{z}_s - \mathbf{u}_s)^H g_{1,s}^{-1} \right), \quad (20)$$

$$\tilde{\Phi}_2 = 2 \text{Re} \left(\sum_{s=1}^M \mathbf{u}_s \mathbf{u}_s^H g_{2,s}^{-1} \right). \quad (21)$$

By construction, the likelihood (18) is obtained by marginalizing (19) with respect to \mathbf{u} . Note that the augmented likelihood (19) is separable in Σ_1 and Σ_2 , which will simplify the estimation of these matrix-valued parameters significantly.

IV. BAYESIAN ESTIMATION

The matrices Σ_1 and Σ_2 of the model introduced in Section III can be estimated using Bayesian estimators. Bayesian inference consists in assigning prior distributions to the unknown model parameters and estimating them using the resulting posterior distribution. The estimation of covariance matrices using Bayesian estimators has been considered in several studies, motivated by the regularizing effect of the prior distribution (see, e.g., [45]). This section presents the Bayesian model investigated in this work and a procedure for the estimation of Σ_1 and Σ_2 for an arbitrary number R of components.

A. Likelihood

The proposed Bayesian model is based on the augmented likelihood (19), which is the product of complex Gaussian distributions having Σ_1 and Σ_2 as covariance matrices.

B. Priors

Inverse Wishart. The natural *conjugate prior* for Σ_i is the inverse Wishart (IW) prior [46], i.e., $\Sigma_i \sim \mathcal{IW}(\nu_i, \Lambda_i)$, with probability density function (pdf)

$$p(\Sigma_i \mid \nu_i, \Lambda_i) \propto (\det \Sigma_i)^{-\frac{1}{2}(\nu_i + R + 1)} \exp(-\text{tr}(\Lambda_i \Sigma_i^{-1})/2), \quad (22)$$

where $\nu_i \in \mathbb{R}$ is the degree of freedom and Λ_i is an $R \times R$ p.d. scale matrix. The IW prior is proper for $\nu_i > R - 1$ and its mean $\mathbb{E}[\Sigma_i] = \Lambda_i / (\nu_i - R - 1)$ only exists if $\nu_i > R + 1$. One of the main disadvantages of this prior is that the uncertainty for the variance parameters in Σ_i is only controlled by ν_i and thus does not permit to incorporate different prior information for the different variance components. Moreover, if ν_i is larger than one, the variance estimates are biased because the implied scaled distribution on each individual variance has extremely low density in a region near zero [47]. Finally, the IW prior imposes a dependency between the correlations and the

variances (since larger variances are associated with absolute values of the correlations near one while small variances are associated with correlations near zero [48]), which is not a desired property.

Scaled inverse Wishart. An alternative to the IW prior is the scaled inverse Wishart (SIW) prior proposed in [49]. The idea is to decompose the matrix Σ_i , using two independent random matrices \mathbf{Q}_i and Δ_i defined as

$$\Sigma_i \triangleq \Delta_i \mathbf{Q}_i \Delta_i, \quad (23)$$

where $\mathbf{Q}_i \sim \mathcal{IW}(\nu_i, \Lambda_i)$ and Δ_i is a diagonal matrix such that its diagonal elements $\delta_{ir} = [\Delta_i]_{rr}$ are independent and log-normal distributed, i.e., $\delta_{ir} \sim \mathcal{LN}(\beta_{ir}, \alpha_{ir}^2)$ with pdf

$$p(\delta_{ir} \mid \beta_{ir}, \alpha_{ir}^2) \propto (\delta_{ir} \alpha_{ir})^{-1} \exp(-(\ln \delta_{ir} - \beta_{ir})^2 (2\alpha_{ir}^2)^{-1}), \quad (24)$$

for $i \in \{1, 2\}$ and $r \in \{1, \dots, R\}$. The decomposition of Σ_i in (23) allows the priors of the standard deviations and correlation coefficients to be defined semi-separately, providing more flexibility than the IW prior.

Section V will study and compare the use of the IW and SIW priors for multivariate multifractal analysis.

C. Posterior distribution

The posterior distribution associated with the proposed Bayesian model for Σ_1 , Σ_2 and the latent vector \mathbf{u} can be computed from Bayes' theorem

$$p(\Sigma_1, \Sigma_2, \mathbf{u} \mid \mathbf{z}) \propto p(\mathbf{z}, \mathbf{u} \mid \Sigma_1, \Sigma_2) p(\Sigma_1) p(\Sigma_2). \quad (25)$$

This distribution can be used to define the marginal minimum mean square error (MMSE) estimator $\Sigma_i^{\text{MMSE}} \triangleq \mathbb{E}[\Sigma_i \mid \mathbf{z}, \mathbf{u}]$. This estimator is difficult to be expressed using a simple closed form expression. Thus, we propose to compute an approximation resulting from a Markov chain Monte Carlo (MCMC) algorithm [50]. In particular, we consider Gibbs samplers for the inverse Wishart prior and its scaled version, which are presented in the next section.

D. Gibbs sampler

The Gibbs sampler consists in generating samples $\{\mathbf{u}^{(\lambda)}, \Sigma_1^{(\lambda)}, \Sigma_2^{(\lambda)}\}_{\lambda=1}^{N_{\text{mc}}}$ according to the conditional distributions of (25) when using an IW prior for Σ_i or its scaled version.

Inverse Wishart prior. Assuming $\Sigma_i \sim \mathcal{IW}(\nu_i, \Lambda_i)$ for all $i \in \{1, 2\}$, the conditional distribution of $\Sigma_i \mid \mathbf{z}, \mathbf{u}$ is

$$p(\Sigma_i \mid \mathbf{z}, \mathbf{u}) \propto (\det \Sigma_i)^{-(2M + \nu_i + R + 1)/2} \\ \exp(-\text{tr}(\Sigma_i^{-1}(\Lambda_i + \tilde{\Phi}_i))/2), \quad (26)$$

which is the following IW distribution:

$$\Sigma_i \mid \mathbf{z}, \mathbf{u} \sim \mathcal{IW}(\nu_i + 2M, \Lambda_i + \tilde{\Phi}_i). \quad (27)$$

The conditional distribution of $\mathbf{u} \mid \Sigma_1, \Sigma_2, \mathbf{z}$ can be shown to be the following complex normal distribution:

$$\mathbf{u} \mid \mathbf{z}, \Sigma_1, \Sigma_2 \sim \mathcal{CN}(\mathbf{y}, \tilde{\Sigma}), \quad (28)$$

where $\tilde{\Sigma}$ is a block diagonal matrix whose s th block is defined as $\tilde{\Sigma}_s = [(\mathbf{g}_{1,s} \Sigma_1)^{-1} + (\mathbf{g}_{2,s} \Sigma_2)^{-1}]^{-1}$ and $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_M^T]^T$ with $\mathbf{y}_s = \tilde{\Sigma}_s (\mathbf{g}_{1,s} \Sigma_1)^{-1} \mathbf{z}_s$. Using the conditional distributions (27) and (28), Algorithm 1 summarizes the

ALGORITHM 1: Gibbs sampler using the IW prior

Initialization:
Set $\mathbf{u} = \mathbf{0}$ and draw $\Sigma_i^{(0)} \sim \mathcal{IW}(\nu_i, \mathbf{\Omega}_i)$ for $i = 1, 2$
MCMC iterations:
for $\lambda = 1 : N_{mc}$ **do**
 for $i = 1 : 2$ **do**
 Draw $\Sigma_i^{(\lambda)}$ from the IW distribution (27)
 end
 Draw $\mathbf{u}^{(\lambda)}$ according to the CN distribution (28)
end
return $\{\Sigma_1^{(\lambda)}, \Sigma_2^{(\lambda)}\}_{\lambda=1}^{N_{mc}}$

different steps of the proposed Gibbs sampler used to generate samples according to the posterior of interest (25).

Scaled inverse Wishart prior. The SIW prior used for Σ_i is defined by independent priors for the scale parameters $\{\delta_{ir}\}_{r=1}^R$ and the matrix parameter \mathbf{Q}_i [49]:

$$p(\Sigma_i) = p(\mathbf{Q}_i) \prod_{r=1}^R p(\delta_{ir}). \quad (29)$$

Thus, the posterior distribution (25) can be rewritten after substituting both Σ_1 and Σ_2 by their decompositions (23). For $i = \{1, 2\}$, assuming $\mathbf{Q}_i \sim \mathcal{IW}(\nu_i, \mathbf{\Lambda}_i)$, the following result is obtained

$$p(\mathbf{Q}_i | \mathbf{\Delta}_i, \mathbf{z}, \mathbf{u}) \propto (\det \mathbf{Q}_i)^{-(2M + \nu_i + R + 1)/2} \exp(-\text{tr}(\mathbf{Q}_i^{-1}(\mathbf{\Lambda}_i + \mathbf{\Delta}_i^{-1} \mathbf{\Phi}_i \mathbf{\Delta}_i^{-1}))/2), \quad (30)$$

which means that the conditional distributions of $\mathbf{Q}_i | \mathbf{\Delta}_i, \mathbf{z}, \mathbf{u}$ are IW distributions:

$$\mathbf{Q}_i | \mathbf{\Delta}_i, \mathbf{z}, \mathbf{u} \sim \mathcal{IW}(\nu_i + 2M, \mathbf{\Lambda}_i + \mathbf{\Delta}_i^{-1} \mathbf{\Phi}_i \mathbf{\Delta}_i^{-1}). \quad (31)$$

Similarly, after some manipulations of (25), the log-conditional posterior distribution of δ_{ir} (r th diagonal element of $\mathbf{\Delta}_i$) can be determined:

$$\begin{aligned} \ln p(\delta_{ir} | \mathbf{Q}_i, \{\delta_{ir'}\}_{r'=1, r' \neq r}^R, \mathbf{z}, \mathbf{u}) &= -(2M + 1) \ln \delta_{ir} \\ &- (\ln \delta_{ir} - \beta_{ir})^2 / (2\alpha_{ir}^2) - [\mathbf{Q}_i^{-1}]_{rr} [\mathbf{\Phi}_i]_{rr} (2\delta_{ir}^2)^{-1} \\ &- \delta_{ir}^{-1} \sum_{r' \neq r} \delta_{ir'}^{-1} [\mathbf{Q}_i^{-1}]_{rr'} [\mathbf{\Phi}_i]_{r'r} + \text{constant}, \end{aligned} \quad (32)$$

which is not a standard distribution. To sample according to (32) a Metropolis-Hastings random walk procedure is used for updating each component 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 [50]. The draws of \mathbf{Q}_i and $\{\delta_{ir}\}_{r=1}^R$ are finally used to generate samples of Σ_i using (23).

Finally, the conditional distribution of $\mathbf{u} | \Sigma_1, \Sigma_2, \mathbf{z}$ is the same as in (28). Algorithm 2 summarizes the different steps of the proposed sampling method.

Approximation of the Bayesian estimator. Observe that since both Σ_1 and Σ_2 are generated from IW distributions or their scale versions, these matrices are guaranteed to be p.d. matrices along the iterations. Therefore, for $i \in \{1, 2\}$ after a burn-in period (where the first N_{bi} samples are discarded),

ALGORITHM 2: Gibbs sampler using the SIW prior

Initialization:
Set $\mathbf{u} = \mathbf{0}$ and draw $\mathbf{Q}_i^{(0)} \sim \mathcal{IW}(\nu_i, \mathbf{\Omega}_i)$ and $\delta_{ir}^{(0)} \triangleq 1$ for all $i = 1, 2$ and $r = 1, \dots, R$
MCMC iterations:
for $\lambda = 1 : N_{mc}$ **do**
 for $i = 1 : 2$ **do**
 Draw $\mathbf{Q}_i^{(\lambda)}$ from the IW distribution (31)
 Set $\delta_{ir}^{\circ} = \delta_{ir}^{(\lambda-1)}$ for all $r = 1, \dots, R$
 for $r = 1 : R$ **do**
 Draw $\delta_{ir}^* \sim \mathcal{N}(\delta_{ir}^{\circ}, \sigma_{\delta_{ir}}^2)$ and $\mu \sim \mathcal{U}_{[0,1]}$
 Compute the acceptance ratio a using (32):

$$a = \frac{p(\delta_{ir}^* | \{\delta_{ir'}^{\circ}\}_{r'=1, r' \neq r}, \beta_{ir}, \sigma_{\delta_{ir}}^2)}{p(\delta_{ir}^{\circ} | \{\delta_{ir'}^{\circ}\}_{r'=1, r' \neq r}, \beta_{ir}, \sigma_{\delta_{ir}}^2)}$$

 Set $\delta_{ir}^{(\lambda)} = \begin{cases} \delta_{ir}^* & \text{if } \mu < a \\ \delta_{ir}^{\circ} & \text{otherwise} \end{cases}$
 end
 Set $\mathbf{\Delta}_i^{(\lambda)} = \text{diag}(\delta_{i1}^{(\lambda)}, \dots, \delta_{iR}^{(\lambda)})$
 Compute $\Sigma_i^{(\lambda)} = \mathbf{\Delta}_i^{(\lambda)} \mathbf{Q}_i^{(\lambda)} \mathbf{\Delta}_i^{(\lambda)}$
 end
 Draw $\mathbf{u}^{(\lambda)}$ according to the CN distribution (28)
end
return $\{\Sigma_1^{(\lambda)}, \Sigma_2^{(\lambda)}\}_{\lambda=1}^{N_{mc}}$

the MMSE estimator of Σ_i is approximated by computing the averages of the matrices $\{\Sigma_i^{(\lambda)}\}_{\lambda=N_{bi}+1}^{N_{mc}}$, i.e.,

$$\hat{\Sigma}_i = \frac{1}{N_{mc} - N_{bi}} \sum_{\lambda=N_{bi}+1}^{N_{mc}} \Sigma_i^{(\lambda)}, \quad i = 1, 2, \quad (33)$$

which are also p.d. matrices.

The proposed estimators (27-28), (33) and (28-31-32), (33), resulting from the IW and SIW priors of Section IV-B, will be denoted as IW-Bay and SIW-Bay in the following.

V. PERFORMANCE ASSESSMENT

The IW-Bay and SIW-Bay estimators are numerically verified for 100 independent realizations of the synthetic multivariate multifractal processes described in Section II-C for different sample sizes, numbers of data components and a large range of values of multifractal parameters. We compare the estimates $\hat{\theta}$ of $\theta \in \{-c_2(r, r'), \rho_{mf}(r, r')\}_{r, r'=1}^R$ obtained by Bayesian estimation to those obtained using a weighted linear regression (see, e.g., [3]), denoted as WLR. The performance is quantified using the bias (BIAS), standard deviation (STD) and the root-mean-square error (RMSE) across realizations, defined as $b(\hat{\theta}) = \widehat{\mathbb{E}}[\hat{\theta}] - \theta$, $s(\hat{\theta}) = \sqrt{\widehat{\text{Var}}[\hat{\theta}]}$ and $r(\hat{\theta}) = \sqrt{b(\hat{\theta})^2 + s(\hat{\theta})^2}$, where $\widehat{\mathbb{E}}[\cdot]$ and $\widehat{\text{Var}}[\cdot]$ are the sample mean and the sample variance.

A. Monte Carlo simulations and parameter settings

A detailed performance analysis was conducted for the bivariate case ($R = 2$) and for $R \in \{1, \dots, 10\}$.

TABLE I
1D MV-MRW ESTIMATION PERFORMANCE FOR $R = 2$, $N = 2^{12}$, $\rho_{mf}(1, 2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ AND $-c_2(2, 2) \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$.
BEST RESULTS ARE MARKED IN BOLD.

$\rho_{mf}(1, 2)$		0.1	0.3	0.5	0.7	0.9	$-c_2(2, 2)$		0.02	0.04	0.06	0.08	0.1	
$-c_2(1, 1) = 0.02$	BIAS	WLR	0.0033	0.0006	0.0044	0.0026	0.0005	$-c_2(1, 1) = 0.02$	WLR	0.0035	0.0029	0.0040	0.0039	0.0040
		SIW-Bay	0.0063	0.0055	0.0050	0.0043	0.0028		SIW-Bay	0.0057	0.0054	0.0048	0.0044	0.0035
		IW-Bay	0.0124	0.0117	0.0116	0.0118	0.0113		IW-Bay	0.0121	0.0120	0.0117	0.0113	0.0105
	STD	WLR	0.0135	0.0119	0.0125	0.0124	0.0131		WLR	0.0114	0.0129	0.0124	0.0145	0.0133
		SIW-Bay	0.0063	0.0046	0.0055	0.0045	0.0035		SIW-Bay	0.0049	0.0058	0.0050	0.0051	0.0047
		IW-Bay	0.0046	0.0035	0.0040	0.0035	0.0029		IW-Bay	0.0037	0.0043	0.0040	0.0038	0.0035
	RSME	WLR	0.0139	0.0120	0.0132	0.0127	0.0131		WLR	0.0119	0.0132	0.0130	0.0150	0.0139
		SIW-Bay	0.0089	0.0072	0.0074	0.0062	0.0044		SIW-Bay	0.0076	0.0080	0.0070	0.0067	0.0059
		IW-Bay	0.0132	0.0122	0.0122	0.0123	0.0117		IW-Bay	0.0127	0.0128	0.0124	0.0120	0.0111
$-c_2(2, 2) = 0.08$	BIAS	WLR	0.0094	0.0087	0.0070	0.0107	0.0080	$-c_2(2, 2) = 0.08$	WLR	0.0026	0.0022	0.0066	0.0114	0.0062
		SIW-Bay	0.0066	0.0010	0.0033	0.0055	0.0045		SIW-Bay	0.0055	0.0038	0.0007	0.0013	0.0075
		IW-Bay	0.0004	0.0004	0.0016	0.0029	0.0014		IW-Bay	0.0122	0.0077	0.0035	0.0004	0.0066
	STD	WLR	0.0231	0.0291	0.0283	0.0270	0.0259		WLR	0.0115	0.0182	0.0248	0.0298	0.0309
		SIW-Bay	0.0110	0.0096	0.0109	0.0084	0.0096		SIW-Bay	0.0055	0.0071	0.0082	0.0092	0.0106
		IW-Bay	0.0100	0.0090	0.0100	0.0082	0.0090		IW-Bay	0.0041	0.0062	0.0073	0.0086	0.0100
	RSME	WLR	0.0249	0.0304	0.0292	0.0290	0.0271		WLR	0.0118	0.0183	0.0257	0.0319	0.0315
		SIW-Bay	0.0110	0.0097	0.0114	0.0100	0.0106		SIW-Bay	0.0078	0.0080	0.0083	0.0093	0.0130
		IW-Bay	0.0100	0.0090	0.0102	0.0087	0.0091		IW-Bay	0.0128	0.0099	0.0081	0.0087	0.0120
$\rho_{mf}(1, 2)$	BIAS	WLR	0.1077	0.0572	0.0130	0.2770	0.0242	$\rho_{mf}(1, 2) = 0.5$	WLR	0.1348	0.1377	0.0997	0.1933	0.2670
		SIW-Bay	0.0309	0.0917	0.1551	0.1588	0.1801		SIW-Bay	0.1433	0.1559	0.1475	0.1329	0.1435
		IW-Bay	0.0433	0.1356	0.2282	0.2957	0.3730		IW-Bay	0.2692	0.2513	0.2320	0.2158	0.2232
	STD	WLR	0.7091	0.7971	0.6143	1.1903	0.3964		WLR	0.8834	0.7594	0.8646	1.0178	0.8363
		SIW-Bay	0.1323	0.1069	0.1073	0.0839	0.0604		SIW-Bay	0.1260	0.1224	0.1023	0.1087	0.0985
		IW-Bay	0.1001	0.0794	0.0799	0.0596	0.0476		IW-Bay	0.0800	0.0844	0.0743	0.0747	0.0714
	RSME	WLR	0.7172	0.7992	0.6144	1.2221	0.3971		WLR	0.8936	0.7718	0.8704	1.0360	0.8779
		SIW-Bay	0.1359	0.1408	0.1885	0.1796	0.1899		SIW-Bay	0.1909	0.1982	0.1795	0.1717	0.1740
		IW-Bay	0.1090	0.1571	0.2418	0.3016	0.3760		IW-Bay	0.2808	0.2651	0.2436	0.2283	0.2344

In the bivariate case ($R = 2$), the parameters of the MV-MRW process are set to $(H_1, H_2) = (0.72, 0.72)$, $\lambda_1 = \sqrt{0.02}$, $\lambda_2 \in \{\sqrt{0.02}, \sqrt{0.04}, \sqrt{0.06}, \sqrt{0.08}, \sqrt{0.1}\}$, $\rho_{ss}(1, 2) = 0$ (uncorrelated data components) and $\rho_{mf}(1, 2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Sample sizes are fixed to $N = 2^{12}$ (1D signals) and $N \times N = 2^9 \times 2^9$ (2D square images) when not mentioned otherwise. For $R \in \{1, \dots, 10\}$, $\lambda_r = \sqrt{1/r}$, ρ_{mf} is uniform in $[0, 0.5]$ and all other parameters are as above. The wavelet analysis is conducted with a Daubechies least asymmetric wavelet (a common choice for scaling analysis) with $N_\psi = 3$, and excluding the finest scale ($j_1 = 2$ to $j_2 = 7$ (1D) and $j_2 = 5$ (2D), respectively), cf, e.g., [3], [51]. The hyperparameters for our model are set to $\nu_i = R + 2$, $\Lambda_i = \mathbb{I}_R$ ($R \times R$ identity matrix), $(\beta_{ir}, \alpha_{ir}^2) = (0.1, 1)$, for $i \in \{1, 2\}$ and $r \in \{1, \dots, R\}$, see, e.g., [49], [52]. With these choices, a reasonable amount of prior probability mass ($\approx 1/4 - 1/3$) is assigned to a conservatively large range of multifractality parameters $-0.25 \lesssim c_2 \leq 0$. Note that the priors are sufficiently non-informative so as not to bias our performance analysis. The parameters of the Gibbs samplers are set to $N_{bi} = 1000$ and $N_{mc} = 2000$, which were found to be suitable values to ensure the convergence of the proposed estimation algorithms (see, e.g., [50]).

B. Estimation performance

Tables I and II summarize the estimation performance of WLR, SIW-Bay and IW-Bay estimators for 1D (signals) and 2D (images), respectively.

Linear regression vs. Bayesian estimation. We observe that while BIAS is in general smaller for WLR than for the Bayesian estimations, the latter produce smaller STD for all

parameters (univariate $c_2(r, r)$ and cross-component multifractal correlation ρ_{mf}). Specifically, WLR has 2 – 7 times larger STD than the Bayesian estimators for the univariate parameters $c_2(r, r)$ and 5 – 27 times larger STD for ρ_{mf} . Consequently, the RMSE values of the Bayesian estimators are (in median over the considered values for c_2 and ρ_{mf}) 1.8, 2.6 and 3.5 times smaller than those of WLR for c_2 and ρ_{mf} . These RMSE reductions are larger for large values of the multifractality parameter $c_2(2, 2)$, and they are especially important for small values of ρ_{mf} (up to a factor 14). Such significant performance gains for the proposed Bayesian estimators are observed for both time series (1D) and images (2D).

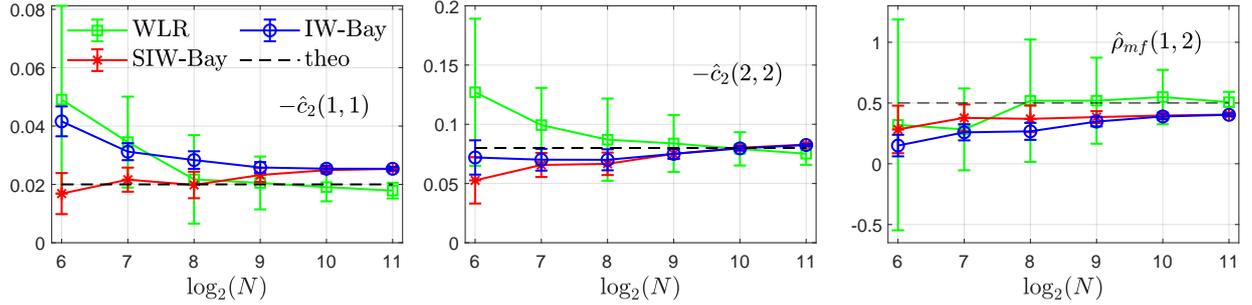
Comparison of priors. Tables I and II show that the use of the SIW prior globally leads to smaller BIAS, but larger STD than the IW prior. This is to be expected because of the extra modeling flexibility of the scaled prior with a larger number of parameters. This bias-variance trade-off leads in many cases to better estimation performance (i.e., smaller RMSE) for SIW-Bay. In particular, SIW-Bay always yields better estimation performance than IW-Bay for the multifractal correlation parameter $\rho_{mf}(1, 2)$, and for the univariate parameters $c_2(1, 1)$ and $c_2(2, 2)$ when $-c_2(2, 2) < 0.05$, for which IW-Bay yields RMSE values up to twice as large as those for SIW-Bay.

Estimation performance vs. sample size. Fig. 5 plots averages and STD-based error bars for estimates obtained for WLR, IW-Bay and SIW-Bay as a function of the sample size (for $N \times N$ size images; similar results are obtained for 1D time series and are not reported here for space reasons). The parameters for a 2D MV-MRW and $R = 2$ were set to $(H_1, H_2) = (0.72, 0.72)$, $\lambda_1 = \sqrt{0.02}$, $\lambda_2 = \sqrt{0.08}$, $\rho_{ss}(1, 2) = 0$, $\rho_{mf}(1, 2) = 0.5$, $N \in \{2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}\}$

TABLE II

 2D MV-MRW ESTIMATION PERFORMANCE FOR $R = 2$, $N = 2^9$, $\rho_{mf}(1, 2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ AND $-c_2(2, 2) \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$. BEST RESULTS ARE MARKED IN BOLD.

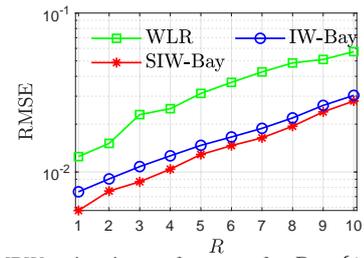
$\rho_{mf}(1, 2)$		0.1	0.3	0.5	0.7	0.9	$-c_2(2, 2)$		0.02	0.04	0.06	0.08	0.1	
$-c_2(1, 1) = 0.02$	BIAS	WLR	0.0007	0.0011	0.0007	0.0007	0.0011	$-c_2(1, 1) = 0.02$	WLR	0.0003	0.0010	0.0005	0.0011	0.0014
		SIW-Bay	0.0050	0.0044	0.0032	0.0020	0		SIW-Bay	0.0046	0.0037	0.0036	0.0033	0.0033
		IW-Bay	0.0071	0.0066	0.0058	0.0050	0.0040		IW-Bay	0.0069	0.0061	0.0061	0.0058	0.0057
	STD	WLR	0.0090	0.0085	0.0075	0.0080	0.0077		WLR	0.0070	0.0086	0.0090	0.0067	0.0077
		SIW-Bay	0.0025	0.0022	0.0026	0.0019	0.0015		SIW-Bay	0.0024	0.0022	0.0024	0.0026	0.0022
		IW-Bay	0.0022	0.0020	0.0022	0.0015	0.0013		IW-Bay	0.0020	0.0018	0.0020	0.0022	0.0018
RSME	WLR	0.0090	0.0086	0.0075	0.0080	0.0078	WLR	0.0070	0.0087	0.0090	0.0068	0.0078		
	SIW-Bay	0.0056	0.0050	0.0041	0.0027	0.0015	SIW-Bay	0.0052	0.0043	0.0043	0.0042	0.0039		
	IW-Bay	0.0074	0.0069	0.0062	0.0053	0.0042	IW-Bay	0.0072	0.0064	0.0064	0.0062	0.0060		
$-c_2(2, 2) = 0.08$	BIAS	WLR	0.0018	0.0049	0.0015	0.0048	0.0021	$-c_2(2, 2) = 0.08$	WLR	0.0001	0.0017	0.0014	0.0046	0.0004
		SIW-Bay	0.0043	0.0053	0.0055	0.0066	0.0084		SIW-Bay	0.0043	0.0023	0.0006	0.0059	0.0105
		IW-Bay	0.0045	0.0054	0.0057	0.0064	0.0076		IW-Bay	0.0067	0.0034	0.0003	0.0060	0.0107
	STD	WLR	0.0216	0.0267	0.0172	0.0217	0.0214		WLR	0.0081	0.0158	0.0158	0.0203	0.0281
		SIW-Bay	0.0042	0.0039	0.0041	0.0042	0.0036		SIW-Bay	0.0025	0.0030	0.0037	0.0045	0.0054
		IW-Bay	0.0042	0.0038	0.0038	0.0042	0.0034		IW-Bay	0.0020	0.0027	0.0035	0.0044	0.0052
RSME	WLR	0.0217	0.0272	0.0173	0.0222	0.0215	WLR	0.0081	0.0159	0.0159	0.0208	0.0281		
	SIW-Bay	0.0060	0.0066	0.0069	0.0079	0.0091	SIW-Bay	0.0050	0.0038	0.0038	0.0074	0.0118		
	IW-Bay	0.0062	0.0066	0.0068	0.0076	0.0083	IW-Bay	0.0070	0.0044	0.0035	0.0074	0.0119		
$\rho_{mf}(1, 2)$	BIAS	WLR	0.1434	0.0059	0.0537	0.0561	0.0345	$\rho_{mf}(1, 2) = 0.5$	WLR	0.0128	0.1428	0.0538	0.0076	0.0788
		SIW-Bay	0.0297	0.0738	0.1117	0.1298	0.1055		SIW-Bay	0.0807	0.0949	0.1014	0.1182	0.1165
		IW-Bay	0.0358	0.0922	0.1496	0.1960	0.2276		IW-Bay	0.1554	0.1463	0.1434	0.1540	0.1507
	STD	WLR	0.8287	0.2548	0.3007	0.4368	0.2001		WLR	0.2975	1.1387	0.2661	0.2463	0.3568
		SIW-Bay	0.0566	0.0513	0.0524	0.0452	0.0274		SIW-Bay	0.0566	0.0498	0.0537	0.0582	0.0478
		IW-Bay	0.0493	0.0448	0.0438	0.0353	0.0211		IW-Bay	0.0432	0.0420	0.0456	0.0491	0.0396
RSME	WLR	0.8410	0.2548	0.3055	0.4404	0.2030	WLR	0.2977	1.1477	0.2715	0.2465	0.3654		
	SIW-Bay	0.0640	0.0899	0.1234	0.1374	0.1090	SIW-Bay	0.0985	0.1072	0.1148	0.1317	0.1259		
	IW-Bay	0.0609	0.1025	0.1559	0.1991	0.2286	IW-Bay	0.1613	0.1522	0.1505	0.1616	0.1558		


 Fig. 5. 2D MV-MRW estimation performance for $R = 2$, $N = \{2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}\}$, $j_1 = 2$ and $j_2 \in \{3, 4, 5, 6, 7, 8\}$.

and $j_2 \in \{3, 4, 5, 6, 7, 8\}$, the other parameters are set as above. The results show that WLR performs also worse in terms of bias when the sample size is small: indeed, when $N < 2^8$, the bias is larger than that of SIW-Bay (and significantly so for $c_2(2, 2)$ and $c_2(1, 1)$). As above, SIW-Bay leads to smaller bias values than IW-Bay, yet it can be observed that this difference in bias disappears for large sample size. For all estimators, the standard deviation decreases when the sample size N increases, as expected. However, the standard deviations are significantly smaller for the Bayesian estimators when compared to WLR.

Estimation performance vs. number of data components.

Fig. 6 displays the root-mean-square error (RMSE) of the estimated multifractal parameters, computed as the root square of the average over realizations of the trace of the matrix $(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T$, where the vector \mathbf{x} contains the diagonal and upper triangle of Σ_1 . As expected, the RMSE of all estimators increases as the number R of data components (thus, $(R^2 + R)/2$ multifractal parameters) increases. The relative performance of the estimators remains similar to the


 Fig. 6. 2D MV-MRW estimation performance for $R = \{1, 2, \dots, 10\}$, $N = 2^9$, $j_1 = 2$ and $j_2 = 5$.

case $R = 2$: Bayesian estimators perform significantly better than WLR, and SIW-Bay has a slightly lower RMSE than IW-Bay.

Computational cost. Given an R -variate signal with sample size n ($n = N$ for time series and $n = N^2$ for images), neglecting border effects of the wavelet transform, the global complexity of our algorithms is $O(n \ln(n) R^3)$, i.e., nearly linear in sample size n and cubic in R . For comparison, the method based on linear regression has complexity $O(nR^3)$

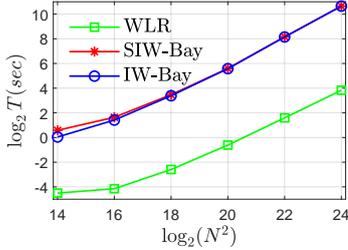


Fig. 7. Computational time T (in seconds) versus image size $N \times N \times 2$ with $N \in \{2^7, 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}\}$, $j_1 = 2$ and $j_2 = \log_2(N) - 4$ for all methods.

(without the $\ln n$ factor due to our additional use of a Fast Fourier transform). Fig. 7 compares the execution times for the estimation of c_2 and ρ_{mf} from a bivariate image ($R = 2$) of size $N \times N$ using WLR, IW-Bay and SIW-Bay, as a function of N , which confirms these leading-order complexity estimates. The cost when using the Bayesian estimators is only ~ 8 times larger when compared to linear regression and is hence no real limitation in practice. Also, the cost of SIW-Bay is found to be only marginally larger than that of IW-Bay. As an example, the processing of a 1024×1024 image takes about one minute on a standard laptop computer with a 2.11 Ghz Intel Core i7 processor and 16GB RAM, allowing even the processing of a 4096×4096 image with a reasonable execution time.

Overall, these results clearly demonstrate a significant benefit of the proposed Bayesian estimators for multivariate multifractal analysis, at reasonably larger computational cost than linear regression.

VI. REAL-WORLD POLYSOMNOGRAPHIC DATA ANALYSIS

This section finally illustrates the use of the proposed framework on real-world data. We consider the problem of detecting drowsiness, defined as an intermediate state between awake and sleep [53], from several light non-invasive modalities related to the cardiovascular, respiratory and brain states. Drowsiness is a major factor in high rates of vehicle accidents. The use of non-invasive biomedical signals for drowsiness detection is an important and open issue that has recently received a considerable interest [54]–[58].

A. Data and preprocessing

MIT-BIH dataset. The dataset used for this study is extracted from the MIT-BIH Polysomnographic database², which involves a collection of recordings of multiple physiological signals, including heart rate (HR), blood pressure (BP), electroencephalogram (EEG) and respiration (RESP), acquired during sleep at a sample rate of 250 Hz [59]. The recordings come with manual annotations for sleep stages. We consider here the states “awake” vs. “stage 1” since the transition to the latter is considered as drowsiness.

Preprocessing. All the 18 available multichannel records were used in the experiments, without a priori exclusion of subjects. Note that most studies reported in the literature use only a hand-picked subset of subjects to avoid variability

²<https://physionet.org/content/slpdb/1.0.0/>

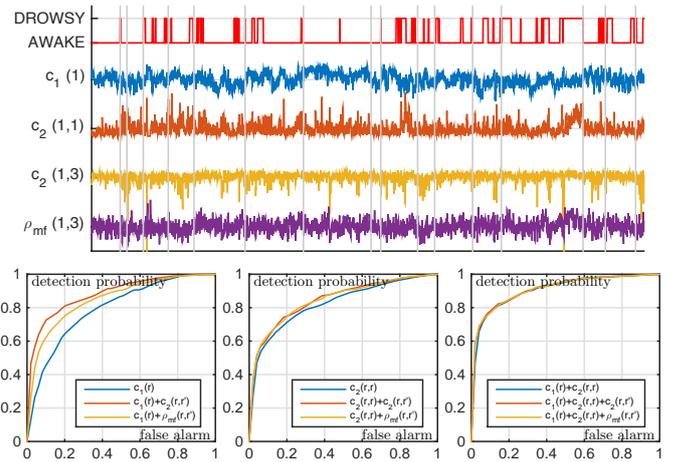


Fig. 8. Visualization of awake and drowsy stages and multifractal parameters (top). Drowsiness detection performance (ROCs, bottom): $c_1(r)$, $c_2(r, r)$ and $(c_1(r), c_2(r, r))$ (blue solid lines in left, center, right plot, respectively) and the corresponding curve when $c_2(r, r')$ (red solid line) or $\rho_{mf}(r, r')$ (green solid line) is used as an additional feature.

caused by the use of different sensors for certain subjects, and to remove subjects affected by outliers. For each multichannel recording, we consider the HR ($r = 1$), BP ($r = 2$), EEG ($r = 3$) and RESP ($r = 4$) channels yielding $R = 4$ components. HR recordings were corrected for missing QRS using the Pan Tompkins ECG QRS detector and linear interpolation. The data was resampled at 4 Hz using linear interpolation, and the analysis was performed on the 2nd primitive to avoid negative uniform regularity issues (cf., e.g., [51]).

Multifractal analysis. Fractal and multifractal models have been widely and successfully used for the analysis of single physiological time series, including sleep staging, in particular for heart rhythm [60]–[62] but also for EEG [61], [63]–[65], BP [61], [66] and RESP [65], [66] recordings. Here, we use the proposed Bayesian multivariate multifractal analysis framework to perform, for the first time, a joint multivariate analysis of these recordings. The analysis was performed using 75% overlapping windows of sample size $N = 480$, yielding a set of multifractal parameter estimates for each 30 second interval (i.e., for each annotation). We use the SIW prior and scales $j = 3 - 6$ (equivalently, 2.6s–21s), with parameters set to $N_\psi = 3$, $N_{mc} = 1000$, $N_{bi} = 500$.

B. Drowsiness detection

Detection. In total, 2381 and 561 examples are available for the awake and stage 1 states, respectively. For each example and channel $r \in \{1, 2, 3, 4\}$, the proposed algorithm estimated the values of $c_1(r)$ and $c_2(r, r)$. Likewise, for the 6 pairs of channels, $c_2(r, r')$ and $\rho_{mf}(r, r')$ (with $r \neq r'$) were estimated. The values of $c_1(r)$ were also estimated using standard linear regression as in [3], [51]. We study several sets of features: the univariate features $\{c_1(r)\}_{r=1}^4$, $\{c_2(r, r)\}_{r=1}^4$, $\{c_1(r), c_2(r, r)\}_{r=1}^4$, and their combination with either $c_2(r, r')$ or $\rho_{mf}(r, r')$, $r \neq r'$, as joint multifractality estimates. Drowsiness detection was performed using a random forest classifier with 50 trees that was trained on random subsets of 80% of the available examples (class 0 corresponds to the awake state and class 1 to stage 1). The

TABLE III
CLASSIFICATION ACCURACY (TOP, IN %), F-MEASURE (CENTER) AND
AUC (BOTTOM), THE LARGER, THE BETTER.

features	$c_1(r)$	$c_2(r, r)$	c_1 & $c_2(r, r)$
	Classification accuracy		
	82.7	86.9	90.1
with $c_2(r, r')$	88.6	87.7	90.7
with ρ_{mf}	86.1	87.7	91.0
	F-measure (detection)		
	73.7	75.6	83.2
with $c_2(r, r')$	80.7	77.2	83.7
with ρ_{mf}	77.6	77.5	83.2
	AUC (detection)		
	0.787	0.817	0.898
with $c_2(r, r')$	0.874	0.839	0.901
with ρ_{mf}	0.874	0.839	0.901

detection performance was tested on the remaining 20% of the database. The reported results are averages over 25 different random subsets. An illustration of estimates of the single-channel parameters $c_1(1)$ and $c_2(1, 1)$ (HR channel) and the cross-channel parameters $c_2(1, 3)$ and $\rho_{mf}(1, 3)$ (HR and EEG channels) is provided in Fig. 8 (top panel).

Performance. The obtained classification and detection performance are quantified using 2-class accuracies as in [58], and F-measure and area-under-curve (AUC) for the receiver operational characteristics (ROCs), as reported in Table III. To further illustrate the detection performance, the ROCs are displayed in Fig. 8 (bottom plots). The ROCs are computed by varying the relative weight for the “awake” and “stage 1” classes in the loss during training of the random forest. We observe that the use of the joint multifractal parameters $c_2(r, r')$ or $\rho_{mf}(r, r')$ consistently and significantly improves single-recording (up to 5.9%, 7.0% and 8.7% increase for classification accuracy, F-measure and AUC, respectively). The best results are obtained when the single-recording parameters $c_1(r)$ and $c_2(r, r)$ are used jointly with either the multifractal correlation parameter $\rho_{mf}(r, r')$ or the cross-multifractality parameter $c_2(r, r')$ (classification accuracy 90.7 – 91.0%, F-measure 83.2–83.7% and AUC 0.901). Thus, the performance is similar to the state of the art reported in [58] (classification accuracy of 93%). Overall, these results demonstrate robustness and relevance of the proposed joint estimation framework for the analysis of real-world data.

VII. CONCLUSION

This paper proposed Bayesian estimators for the multifractality parameters associated with the recent theoretical definition of the multifractal spectrum of multivariate data. The estimators leverage on an original and versatile model for the joint statistics of the logarithm of wavelet leaders of several data components. We showed that when combined with a Whittle-type spectral approximation and a data augmentation strategy, the proposed Bayesian approach yields efficient numerical estimates for the matrix-valued parameters of interest using standard Gibbs sampling. Large-scale numerical experiments were conducted with synthetic multifractal data and clearly demonstrate the significant benefits of the proposed framework in terms of estimation accuracy, at moderate

cost. The application of the proposed multivariate multifractal analysis to a benchmark real-world multivariate time series dataset also yielded promising results and highlights the actual applicability and practical relevance of our approach for real-world data modeling and analysis. Future work will focus on more theoretical investigations of our model, including the study of performance bounds, and applications to physiological signals and remote sensing data. The estimation and synthesis procedures will be made freely available via a documented toolbox.

REFERENCES

- [1] S. Jaffard, P. Abry, and H. Wendt, *Irregularities and Scaling in Signal and Image Processing: Multifractal Analysis*. Singapore: World scientific publishing, 2015, pp. 31–116.
- [2] 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., vol. 72(2). AMS, 2004, pp. 91–152.
- [3] H. Wendt, P. Abry, and S. Jaffard, “Bootstrap for empirical multifractal analysis,” *IEEE Signal Proces. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.
- [4] J. F. Muzy, E. Bacry, and A. Arneodo, “Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method,” *Phys. Rev. E*, vol. 47, no. 2, p. 875, 1993.
- [5] S. Jaffard, C. Melot, R. Leonarduzzi, H. Wendt, P. Abry, S. G. Roux, and M. E. Torres, “p-exponent and p-leaders, Part I: Negative pointwise regularity,” *Physica A: Statistical Mechanics and its Applications*, vol. 448, pp. 300–318, 2016.
- [6] R. Leonarduzzi, H. Wendt, P. Abry, S. Jaffard, C. Melot, S. G. Roux, and M. E. Torres, “p-exponent and p-leaders, Part II: Multifractal analysis. relations to detrended fluctuation analysis,” *Physica A: Statistical Mechanics and its Applications*, vol. 448, pp. 319–339, 2016.
- [7] K. Kiyono, Z. R. Struzik, N. Aoyagi, and Y. Yamamoto, “Multiscale probability density function analysis: non-Gaussian and scale-invariant fluctuations of healthy human heart rate,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 95–102, Jan. 2006.
- [8] B. B. Mandelbrot, “Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier,” *J. Fluid Mech.*, vol. 62, pp. 331–358, 1974.
- [9] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt, “Scale-free and multifractal time dynamics of fMRI signals during rest and task,” *Front. Physiol.*, vol. 3, Jun. 2012.
- [10] 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.
- [11] M. Doret, H. Helgason, P. Abry, P. Gonçalves, C. Gharib, and P. Gaucherand, “Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor,” *Am. J. Perinatol.*, vol. 28, no. 4, pp. 259–266, 2011.
- [12] R. Fontugne, P. Abry, K. Fukuda, D. Veitch, K. Cho, P. Borgnat, and H. Wendt, “Scaling in Internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections,” *IEEE/ACM Trans. Networking*, vol. 25, no. 4, pp. 1–14, 2017.
- [13] A. Johansen and D. Sornette, “Finite-time singularity in the dynamics of the world population, economic and financial indices,” *Physica A*, vol. 294, pp. 465–502, 2001.
- [14] 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.
- [15] C. Meneveau, K. Sreenivasan, P. Kailasnath, and M. Fan, “Joint multifractal measures - theory and applications to turbulence,” *Phys. Rev. A*, vol. 41, no. 2, pp. 894–913, 1990.
- [16] T. Lux, “Higher dimensional multifractal processes: A GMM approach,” *Journal of Business and Economic Statistics*, vol. 26, pp. 194–210, 2007.
- [17] 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.
- [18] 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.

- [19] H. Wendt, R. Leonarduzzi, P. Abry, S. Roux, S. Jaffard, and S. Seuret, "Assessing cross-dependencies using bivariate multifractal analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Calgary, Canada, April 2018.
- [20] 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.
- [21] P. Abry, R. Leonarduzzi, H. Wendt, S. Jaffard, and S. Seuret, "A multivariate multifractal analysis of lacunary wavelet series," in *Proc. IEEE Int. Workshop CAMSAP*, Guadeloupe, France, Dec. 2019.
- [22] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A*, vol. 316, no. 1, pp. 87–114, 2002.
- [23] S. Combexelle, 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.
- [24] H. Wendt, S. Combexelle, 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.
- [25] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry, "Bayesian estimation for the parameters of the bivariate multifractal spectrum," in *Proc. Europ. Signal Process. Conf. (EUSIPCO)*, Dublin, Ireland, August 2021.
- [26] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat.*, vol. 2, no. 5, pp. 423–434, 1953.
- [27] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 1998.
- [28] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali, *Two-Dimensional Wavelets and their Relatives*. Cambridge Univ. Press, 2004.
- [29] B. Castaing, Y. Gagne, and M. Marchand, "Log-similarity for turbulent flows," *Physica D*, vol. 68, no. 3-4, pp. 387–400, 1993.
- [30] D. Schertzer and S. Lovejoy, "Physical modelling and analysis of rain and clouds by anisotropic scaling multiplicative processes," *Journal of Geophysical Research*, vol. 92, pp. 9693–9714, 1987.
- [31] E. Bacry, J. Delour, and J.-F. Muzy, "Multifractal random walk," *Phys. Rev. E*, vol. 64: 026103, 2001.
- [32] J. Muzy, J. Delour, and E. Bacry, "Modelling fluctuations of financial time series: from cascade process to stochastic volatility model," *The European Physical Journal B*, vol. 17, 2000.
- [33] H. Helgason, V. Pipiras, and P. Abry, "Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding," *Signal Proc.*, vol. 91, no. 5, pp. 1123–1133, 2011.
- [34] B. B. Mandelbrot, "Limit lognormal multifractal measures," *Frontiers of physics*, p. 309, 1990.
- [35] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002, vol. 5, no. 8.
- [36] F. B. Opong and S. Y. Agbedra, "Assessing univariate and multivariate normality, a guide for non-statisticians," *Mathematical Theory and Modeling*, vol. 6, no. 2, pp. 26–33, 2016.
- [37] A. Arneodo, N. Decoster, P. Kestener, and S. G. Roux, "A wavelet-based method for multifractal image analysis: from theoretical concepts to experimental applications," in *Advances in Imaging and Electron Physics*, P. Hawkes, B. Kazan, and T. Mulvey, Eds., vol. 126. Academic Press, 2003, pp. 1–98.
- [38] J. Muzy, D. Sornette, J. Delour, and A. Arneodo, "Multifractal returns and hierarchical portfolio theory," *Quantitative Finance*, vol. 1, no. 1, pp. 131–148, 2001.
- [39] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.
- [40] 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 Trans. Signal Process.*, vol. 65, no. 12, pp. 3136–3151, 2017.
- [41] D. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of kth-order spectra," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, pp. 206–10, 1967.
- [42] R. E. Chandler, "A spectral method for estimating parameters in rainfall models," *Bernoulli*, vol. 3, no. 3, pp. 301 – 322, 1997.
- [43] 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.
- [44] D. A. van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.
- [45] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3554–3571, 2005.
- [46] 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, 2000.
- [47] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian Analysis*, vol. 1, no. 3, pp. 515 – 534, 2006.
- [48] T. Tokuda, B. Goodrich, I. Mechelen, A. Gelman, and F. Tuerlinckx, "Visualizing distributions of covariance matrices," Columbia University, New York; USA, Tech. Rep., 2011.
- [49] A. J. O'Malley and A. M. Zaslavsky, "Domain-level covariance analysis for multilevel survey data with structured nonresponse," *J. American Statistical Association*, vol. 103, no. 484, pp. 1405–1418, 2008.
- [50] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, USA: Springer, 2005.
- [51] H. Wendt, S. Roux, S. Jaffard, and P. Abry, "Wavelet leaders and bootstrap for multifractal analysis of images," *Signal Processing*, vol. 89, no. 6, pp. 1100–1114, 2009.
- [52] I. Alvarez, J. Niemi, and M. Simpson, "Bayesian inference for a covariance matrix," in *Conf. Applied Statistics in Agriculture*, 2014.
- [53] J. Yu, S. Park, S. Lee, and M. Jeon, "Driver drowsiness detection using condition-adaptive representation learning framework," *IEEE Trans. on Intell. Transp. Syst.*, pp. 1–13, December 2018.
- [54] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Electromyogram signal based hypovigilance detection," *Biomedical Research (India)*, vol. 25, pp. 281–288, 2014.
- [55] F. Wang, H. Wang, and R. Fu, "Real-time ECG-based detection of fatigue driving using sample entropy," *Entropy*, vol. 20, no. 3, 2018.
- [56] S. Ahn, T. Nguyen, H. Jang, J. G. Kim, and S. C. Jun, "Exploring neuro-physiological correlates of drivers mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data," *Frontiers in Human Neuroscience*, vol. 10, p. 219, May 2016.
- [57] C. Basri, K. Wan, I. Zunaidi, L. Lee, A. B. Shahrman, M. Razlan, Z. Mustafa, W. Zulkepli, and N. Noriman, "Muscle fatigue detections during arm movement using EMG signal," *IOP Conference Series: Materials Science and Engineering*, vol. 557, p. 012004, June 2019.
- [58] Y. Abichou, S. Chaabene, and L. Chaari, "A sleep monitoring method with EEG signals," in *Digital Health in Focus of Predictive, Preventive and Personalised Medicine*, L. Chaari, Ed. Springer, 2020, pp. 25–32.
- [59] A. Goldberger, L. Amaral, L. Glass, Hausdorff, I. J., R. P. C., Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiobank, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. e215e220, 2000.
- [60] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, no. 6735, pp. 461–465, 1999.
- [61] P. Castiglioni, D. Lazzeroni, P. Coruzzi, and A. Faini, "Multifractal-multiscale analysis of cardiovascular signals: a DFA-based characterization of blood pressure and heart-rate complexity by gender," *Complexity*, 2018.
- [62] H. Wendt, P. Abry, K. Kiyono, J. Hayano, E. Watanabe, and Y. Yamamoto, "Wavelet p-leader non Gaussian multiscale expansions for heart rate variability analysis in congestive heart failure patients," *IEEE Trans. Biomedical Engineering*, vol. 66, no. 1, pp. 80–87, 2019.
- [63] B. Weiss, Z. Clemens, R. Bódizs, Z. Vágó, and P. Halász, "Spatio-temporal analysis of monofractal and multifractal properties of the human sleep EEG," *Journal of neuroscience methods*, vol. 185, no. 1, pp. 116–124, 2009.
- [64] T. Zorick and M. A. Mandelkern, "Multifractal detrended fluctuation analysis of human EEG: preliminary investigation and comparison with the wavelet transform modulus maxima technique," *PLoS one*, vol. 8, no. 7, p. e68360, 2013.
- [65] Q. Ma, X. Ning, J. Wang, and J. Li, "Sleep-stage characterization by nonlinear EEG analysis using wavelet-based multifractal formalism," in *Proc. IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2006, pp. 4526–4529.
- [66] L. Angelini, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G. D. Pinna, S. Stramaglia, and S. A. Tupputi, "Multiscale analysis of short term heart beat interval, arterial blood pressure, and instantaneous lung volume time series," *Artificial Intelligence in Medicine*, vol. 41, no. 3, pp. 237–250, 2007.