Bayesian estimation of the parameters of the joint multifractal spectrum of signals and images

Lorena Leon

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Informatique et Télécommunication

Présentée et soutenue par :

Mme LORENA LEON ARENCIBIA

le mardi 13 décembre 2022

Titre :

Estimation bayésienne des paramètres du spectre multifractal multivarié pour le traitement du signal et des images

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur(s) de Thèse :

M. JEAN-YVES TOURNERET

M. HERWIG WENDT

Rapporteurs :

M. ANDRE FERRARI, UNIVERSITE COTE D'AZUR

MME CAROLINE CHAUX-MOULIN, CNRS

Membre(s) du jury :

MME JULIE DELON, UNIVERSITE DE PARIS, Président

M. HERWIG WENDT, TOULOUSE INP, Membre

M. JEAN-CHRISTOPHE PESQUET, UNIVERSITE PARIS-SACLAY, Membre

M. JEAN-YVES TOURNERET, TOULOUSE INP, Membre

M. MARCELO PEREYRA, HERIOT WATT UNIVERSITY, Membre

M. MATTHIEU JONCKHEERE, LAAS TOULOUSE, Invité(e)

MME CLOTHILDE MELOT, AIX-MARSEILLE UNIVERSITE, Membre

M. PATRICE ABRY, ECOLE NORMALE SUPERIEURE DE LYON, Invité(e)

Acknowledgements

First, I would like to thank my advisors Herwig WENT and Jean-Yves TOURNERET for their generosity since the beginning. I will always be in their debt for their support in the most difficult times, their advice, trust and friendship. Working with both of them has been very pleasant, they have been able to get the best out of me, and I appreciate that so much. I cannot feel more grateful for having the opportunity to work with them. I have learned a lot from them and I admire them infinitely.

Also, I would like to thank the jury for doing me the honor of accepting to review my work and attend my thesis defense. My most sincere thanks go to my rapporteurs Caroline CHAUX and Andre FERRARI, who took the time to read in detail and judge my manuscript. I am also grateful to Julie DELON, Clothilde MELOT, Jean-Christophe PESQUET and Marcelo PEREYRA for agreeing to be reviewers of my thesis. In addition, I thank Matthieu JONCKHEERE and Patrice ABRY for accepting my invitation to my thesis defense.

I cannot thank enough to CIMPA (Centre International de Mathématiques Pures et Appliquées), which indirectly is the reason why I am here today. Thanks to its full fellowship I was able to participate in the research program "The mathematics of imaging" and I visited France for the first time. Back then I got the opportunity to know the academic world outside of what I knew and met professors and students who inspired me to pursue a PhD. In particular, I would like to thank Marcelo again, this time for his advice, inspiration and help to start this thesis 3 years ago.

I am also very grateful to Patrice for his scientific guidance and the opportunity to collaborate with him on several occasions during this thesis. Moreover, I would like to acknowledge Prof. Gerard

LETAC for enthusiastically sharing his expertise on Wishart distribution.

I am very grateful to have been part of the MUTATION project and therefore I thank the ANR for its funding. I would also like to thank the CNES in Toulouse for providing the satellite image used in this manuscript.

I would also like to thank all the teachers who have taken part in my education over the years, there are many names I could mention. In particular, I am very grateful to Angela for her support and advice during these last 6 years of my career.

Thanks to my former and current IRIT colleagues, Mohamed, Oumaima, Farah, Sayantan, Kenule, Kareth, and all those I might have forgotten to mention here. I also want to thank the secretaries Vanessa and Sylvie for always helping me, my TéSa colleagues, and especially Corinne, for welcoming me so many times and always being so kind.

Thanks to my lifelong friends for their support, Betty, Glenda, Oscar Luis, Pedro Luis, Sergio, Marchena, Jeny and Betty. They know how much I appreciate and miss them.

I cannot be more grateful to my Cuban friends in Toulouse: Yeny, Willy, Anays, Suse, Mia, Ale, Anaysis, Armando and especially to Josue, who opened the doors of his home for me, included me in this Cuban-Toulousian family and gave me incredible support always. All of them have made me feel at home even though we are far away.

I am extremely grateful because during these three years I have had the opportunity to meet wonderful and interesting people in research schools, workshops and conferences. I especially want to thank Dalia and Julia for being around and helping me to distract myself.

These 3 years would not have been the same without Charlotte, to whom I am grateful for her friendship, and Thomas, to whom I infinitely thank for his support and wish him all the best.

Last but not least, infinite thanks to my family for their support and patience, my parents to whom I owe everything, my sister and my niece, and my grandfather, wherever he is. I dedicate this thesis to them.

Résumé

L'analyse multifractale est devenue un outil de référence pour le traitement des signaux et des images. Fondée sur la quantification des fluctuations de la régularité locale, elle s'est avérée utile dans un nombre croissant d'applications, qui ne concernaient pourtant jusqu'à présent que des données univariées (séries temporelles à valeurs scalaires ou images acquises dans une seule bande spectrale). Récemment, les bases théoriques de l'analyse multifractale multivariée ont été élaborées, montrant un potentiel pour quantifier les dépendances entre plusieurs collections de données allant au-delà de la corrélation linéaire. Cependant, l'estimation précise des paramètres associés à un modèle multifractal multivarié reste un défi, limitant sévèrement leur utilisation réelle dans les applications. L'objectif principal de cette thèse est de proposer et d'étudier des méthodes d'analyse multifractale multivariée pour le traitement du signal et des images. Plus précisément, l'approche proposée s'appuie sur un nouveau modèle Gaussien multivarié adapté au logarithme des coefficients dominants d'ondelettes. Ce modèle utilise une approximation de la vraisemblance basée sur les résultats de Whittle et une augmentation de données pour les paramètres d'intérêt à valeurs matricielles. Ce modèle permet de construire des procédures d'estimation efficaces pour deux choix pertinents de lois a priori dans un contexte d'inférence Bayésienne. Des algorithmes basés sur des stratégies de Monte Carlo par chaines de Markov et d'Expectation-Maximization sont conçus et utilisés pour calculer les estimateurs bayésiens. Des simulations de Monte Carlo, réalisées sur des images et des signaux synthétiques multivariés avec différentes tailles d'échantillon, différents nombres de composantes et différents jeux de paramètres, montrent des améliorations significatives des performances par rapport à l'état de l'art. En outre, des limites inférieures théoriques sur la variance des estimateurs sont

déterminées pour étudier le comportement asymptotique de ces estimateurs. Enfin, la pertinence du cadre d'estimation multifractale multivariée proposé est démontrée par l'application à deux exemples de données réelles : la détection de la somnolence à partir de signaux physiologiques multicanaux et l'imagerie satellitaire multispectrale.

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, severely limiting their actual use in applications. The main goal of this thesis is to propose and study practical contributions on multivariate multifractal analysis of signals and images. Specifically, the proposed approach relies on a novel and original joint Gaussian model for the logarithm of wavelet leaders and leverages on a Whittle-based likelihood approximation and data augmentation for the matrix-valued parameters of interest. This careful design enables efficient estimation procedures to be constructed for two relevant choices of priors using Bayesian inference. Algorithms based on Monte Carlo Markov Chain and Expectation Maximization strategies are designed and used to approximate the Bayesian estimators. Monte Carlo simulations, conducted on synthetic multivariate signals and images with various sample sizes, numbers of components and multifractal parameter settings, demonstrate significant performance improvements over the state of the art. In addition, theoretical lower bounds on the variance of the estimators are designed to study their asymptotic behavior. Finally, the relevance of the proposed multivariate multifractal estimation framework is shown for two real-world data examples: drowsiness detection from multichannel physiological signals and potential remote sensing applications in multispectral satellite imagery.

Abbreviations and notation

Abbreviations

1D	One dimensional
2D	Two dimensional
BB	Bayesian Cramér–Rao bound
CN	Complex normal
CRB	Cramér–Rao bound
DFT	Discrete Fourier transform
DWT	Discrete wavelet transform
EM	Expectation-Maximization
FIM	Fisher information matrix
fBm	Fractional Brownian motion
GMM	Generalized method of moments
IW	Inverse Wishart
WLR	Weighted linear fit
LN	Log-normal
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MFA	Multifractal analysis
ML	Maximum likelihood

MLE	Maximum likelihood estimator
MMSE	Miminum mean-square error
MRW	Multifractal random walk
MSE	Mean squared error
MV-MRW	Multivariate multifractal random walk
QQ	Quantile-quantile
RMSE	Root mean-square error
SIW	Scaled inverse Wishart
STD	Standard deviation

Main notation

Matrix notation

a	Scalar value
$\mathbf{a}$	Column vector
$\mathbf{A}$	Matrix
$.^T$	Transpose operator
$.^H$	Hermitian transpose operator
$\det \mathbf{A}$	Determinant of matrix $\mathbf{A}$
$\mathrm{Tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\|\|\mathbf{a}\|\|$	Standard L^2-norm $\|\|\mathbf{a}\|\| = \sqrt{\mathbf{a}^H\mathbf{a}}$
$\lfloor\cdot\rfloor$	Integer truncating operator
$[\mathbf{A}]_{n_1,n_2}$	Element of the matrix $\mathbf{A}$ in the n_1th row and n_2th column
$[\![n_1, n_2]\!]$	Set of integers ranging from n_1 to n_2
$\mathbb{I}_R$	$R \times R$ identity matrix
$\mathrm{diag}(a_1, \ldots, a_N)$	$N \times N$ diagonal matrix with (diagonal) entries specified by $a_1, \ldots, a_N$

Random variable notation

$x \sim$	"The variable x is distributed according to"
$p(x) \propto$	"The distribution of x is proportional to"
$\mathcal{LN}(\beta, \alpha^2)$	Log-Gaussian distribution with mean β and variance α^2
$\mathcal{CN}(\boldsymbol{m}, \boldsymbol{\Sigma})$	Circularly-symmetric complex Gaussian distribution with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$
$\mathcal{IW}(\nu, \boldsymbol{\Lambda})$	Inverse Wishart distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Lambda}$
$\mathbb{E}[\cdot]$, $\widehat{\mathbb{E}}[\cdot]$	Mean and sample mean
$\mathrm{Var}[\cdot]$, $\widehat{\mathrm{Var}}[\cdot]$	Variance and sample variance
$\mathrm{Cov}[\cdot]$, $\widehat{\mathrm{Cov}}[\cdot]$	Covariance and sample covariance

Contents

List of Figures

List of Tables

Introduction

Context and objectives of the thesis

Multifractal analysis (MFA) 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 MFA, see, e.g., [Jaf04, WAJ07], which is estimated using numerical implementations that commonly involve multiscale (e.g., wavelet-based) representations of the data [MBA93, JML+16, LWA+16].

MFA has been successfully used in a large range of different signal processing applications, e.g., biomedical applications (infra slow brain activity [CVA+12b]), geophysics [FGK94], finance [MS98] or internet traffic [ABF+02], to name but a few. More recently, it has also been proved to be useful paradigm in an increasing number of different image processing applications, e.g., texture classification [XYLJ10, WAJ+09], biomedical imagery [BPL+01, KLSJA01, LB09], physics [PBA+06, RAD00], biology [SS01], climate research [LS13] and art investigation [CERW08, AJW13, JJHB+08]. For further examples, the reader is referred to [JAW15, AWJD19] and references therein.

However, such past successes have been limited to the analysis of univariate data (scalar valued time series or single channel images). Yet, these are often the parts of naturally multivariate data, e.g., physical quantities jointly registered by several sensors, multispectral images, multitemporal images, etc. Conducting a joint analysis instead of individually analyzing the data components could provide a richer characterization, and in particular unravel the dynamics, coupling mechanisms and dependencies among the different registered components. While this limitation had been recognized

early on and partially addressed in specific applicative contexts [MSKF90, Lux07], the theoretical foundation for multivariate MFA was laid only recently [JSW+19a, JSW+19b]. Its first practical use showed that the multivariate (or joint) multifractal spectrum can effectively capture and quantify transient local dependencies in data that cannot be considered by second-order statistics [WLA+18, LAR+18, ALW+19, AWJD19].

Unfortunately, the accurate estimation of the associated multivariate multifractal parameters is extremely challenging, severely limiting the use of multivariate MFA in applications. In essence, whatever the multiscale representations that are used, multifractal estimation relies 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 [WLA+18, LAR+18, ALW+19].

The main goal of this thesis is to empower multivariate multifractal signal and image analysis by complementing it with an adequate statistical modeling and estimation framework. The first key difficulty stems from the intricate statistical nature of the multifractal models, characterized by highly non Gaussian marginals, scale-free joint distributions and strong dependence, which lead to difficult estimation problems. A second major difficulty arises from the need for models that lead to computationally efficient estimators for the multivariate multifractal parameters in order to deal with large data with potentially many data components.

The strategy adopted in this thesis consists of embedding the estimation of the multivariate multifractal parameters in a Bayesian framework. To this end, we introduce in Chapter 2, a novel and original model for the joint statistics of nonlinear multiscale representations, the wavelet leaders, that arise in the multifractal analysis of discrete data. The clever design of this model, parametrized by multifractal parameters of interest, allows the estimation problem to be formulated in a Bayesian framework. In Chapters 2 and 3, several numerical algorithms are proposed to approximate the associated Bayesian estimators and their performance is studied on synthetic data using Monte Carlo simulations. In Chapter 4, we study the asymptotic behavior of the proposed estimators by establishing fundamental lower bounds for the mean squared error (MSE) for the multifractal parameters

of interest. This can explain and predict their behavior in specific scenarios and facilitate the design of experiments (number of subjects in study, recording length, etc) to yield certain accuracy. Furthermore, we investigate in Chapter 5 the potential benefits of multivariate MFA in general and of the proposed Bayesian methodology in particular for drowsiness detection for the example of a four-channel physiological signal, and for remote sensing applications via the example of a four-band satellite image.

The proposed Bayesian multifractal estimation algorithms, synthesis procedures for multivariate multifractal processes, and procedures for computing lower bounds on the variance of the proposed estimators is available in https://www.irit.fr/~Herwig.Wendt/software.html via a documented toolbox.

This thesis has been carried out in the Institut de Recherche en Informatique de Toulouse (France), within the Computational Imaging and Vision research group and funded by the French ANR JCJC research grant MUTATION (ANR-18-CE45-0007).

Structure of the manuscript and main contributions

Chapter 1 recalls the key theoretical and practical concepts of MFA, from univariate to multivariate modeling and analysis of multifractal time series and images. The main contribution of this chapter lies in the synthetic overview exposition itself. As a basis for the following chapters the current benchmark multivariate multifractal formalism based on wavelet leaders is defined. This formalism leads to a polynomial expansion of the pairwise joint multifractal spectrum whose coefficients are given by the log-cumulants of wavelet leaders. The leading order log-cumulants summarize the multifractal properties and can hence be used in applications instead of the multifractal spectrum. The focus of the present thesis is on the estimation of the second-order log-cumulant c_2, quantifying the degree of fluctuation of the pointwise regularity individually and the coupling of these fluctuations across components, and the multifractal correlation parameter ρ_{mf}.

Chapter 2 introduces a parametric model for the multivariate statistics of the logarithm of wavelet leaders (termed log-leaders for short). This model consists of multivariate Gaussian distributions

whose variance-covariance structures are controlled by two symmetric positive definite (p.d.) matrices containing the second-order log-cumulants. This model and its validation constitute a key contribution of this chapter and the core of this thesis. We then complement the model with two original and interwoven threads: first, a spectral approximation for the efficient evaluation of the corresponding likelihood, and second, a data augmentation strategy in the spectral domain that enables its factorization in parameter-wise components. Separability further allows us to introduce Bayesian models building on this approximation and two relevant choices of prior distributions for the multifractal matrix-valued parameters of interest. In particular, we propose the use of an inverse Wishart prior distribution due to its conjugacy for covariance parameters of Gaussian distributions and a scaled inverse Wishart prior due to its greater flexibility to incorporate prior information. The derivation of the conditional distributions simplifies the use of Markov Chain Monte Carlo (MCMC) algorithms to approximate the minimum mean square error (MMSE) estimator associated with the resulting posterior distributions. We also investigate the use of a geometric mean to obtain the approximation of the MMSE estimator and compare it against the classical arithmetic mean. The combination of the two priors and the two ways of calculating the average of symmetric p.d. matrices leads to a total of four different estimators. Their estimation performance is studied and compared against classical linear regressions, using extensive Monte Carlo simulations relying on synthetic multivariate multifractal signals and images for several sample sizes and numbers of data components. The methods and algorithms developed in this chapter constitute additional valuable contributions and lead to the first operational tool for practical multivariate MFA of multivariate signals and images.

Chapter 3 proposes and studies new methods and algorithms based on Expectation-Maximization (EM) strategies for multivariate multifractal parameter estimation. The EM methodology is appealing for having a reduced computational time compared to MCMC-based methods. The proposed procedures are built on the previously devised statistical model of the log-leaders. Specifically, we propose EM-based algorithms to approximate the maximum likelihood and the maximum a posteriori estimator. Various Monte Carlo simulations are used to assess and study their estimation performance. Furthermore, they are compared against the former Bayesian estimators and classical linear

regression-based algorithms, in terms of accuracy and computational time. The design of these estimators and their evaluation are the main methodological and practical contributions of this chapter.

Chapter 4 provides lower bounds on the MSE performance of estimators of the matrix-valued parameters of the statistical model introduced in Chapter 2. To the best of our knowledge, this has not be achieved so far for multivariate multifractal parameter estimation and, specifically, for the model proposed in this thesis. First, considering the parameters to be unknown and deterministic matrices, the Cramér–Rao bound is determined. Second, assuming that the parameters are unknown random matrices that are modeled as inverse Wishart distributed matrices, the Bayesian Cramér–Rao bound is derived. This constitutes a key contribution of this chapter. In addition, as a second contribution, a novel closed-form expression for computing non-trivial expectations involving Wishart random matrices is derived and proved. As another main contribution, the properties of these bounds are analytically studied and also numerically investigated via Monte Carlo simulations. Finally, we illustrate the use of the proposed bounds for the parameters associated with the bivariate multifractal spectrum.

Chapter 5 investigates for the first time the use of multivariate multifractal parameters in two applications of very different natures: i) drowsiness detection in multichannel physiological signals and ii) quantification of spatial/textural information in multispectral satellite images. The obtained results demonstrate that the Bayesian methods and algorithms developed in this thesis provide an operational multivariate MFA tool that can be applied on real-world multivariate datasets.

Chapter 6 concludes the present thesis and provides a guideline for future work.

List of publications

Journal papers.

1. L. Leon, H. Wendt, J.-Y. Tourneret and P. Abry, *A Bayesian Framework for Multivariate Multifractal Analysis*, in IEEE Transactions on Signal Processing, vol. 70, pp. 3663-3675, Jun. 2022

2. L. Leon, H. Wendt and J.-Y. Tourneret, *Bounds for the Estimation of Matrix-Valued Parameters of a Gaussian Random Process. Submitted.*

International conferences.

1. H. Wendt, L. Leon, J.-Y. Tourneret and P. Abry, *Multifractal Anomaly Detection in Images via Space-Scale Surrogates*, IEEE International Conference on Image Processing (ICIP), pp. 556-560, Bordeaux, France, Oct. 2022.

2. L. Leon, H. Wendt, J.-Y. Tourneret and P. Abry, *A Comparison of Bayesian Estimators for the Parameters of the Bivariate Multifractal Spectrum*, 30th European Signal Processing Conference (EUSIPCO), pp. 2116-2120, Belgrade, Serbia, Aug. 2022.

3. L. Leon, H. Wendt, J.-Y. Tourneret and P. Abry, *Bayesian Estimation for the Parameters of the Bivariate Multifractal Spectrum*, 29th European Signal Processing Conference (EUSIPCO), pp. 1930-1934, Dublin, Ireland, Aug. 2021.

National conferences.

1. L. Leon, H. Wendt, J.-Y. Tourneret, P. Abry, *Estimation du paramètre de multifractalité : régression linéaire, maximum de vraisemblance ou inférence Bayésienne*, GRETSI Symposium Signal and Image Processing, Nancy, France, Sept. 2022.

Chapter 1

Multifractal analysis

Contents

1.1 Introduction

This chapter aims to provide a synthetic presentation of the main theoretical and practical concepts for the multivariate modeling and analysis of multifractal times series and images.

Over the last decades, multifractal analysis (MFA) has grown into a standard signal and image processing tool, characterizing data in terms of the fluctuations of their pointwise regularity in time or space. In the past, MFA has led to significant successes in many real-world applications in very different contexts, see, e.g., [Man74, JS01, KSAY06, DHA+11, CVA+12a, FAF+17]. However, these successes have been mainly limited to the analysis of univariate data (scalar valued 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). While this limitation had been recognized early on and partially addressed in specific applicative contexts [MSKF90, Lux07], 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. The scope of this thesis is to propose and study statistical models and estimation algorithms based on the theoretical foundations of multivariate MFA [JSW+19a, JSW+19b], to overcome this deadlock.

These theoretical foundations are briefly recalled in this chapter which is organized as follows. In Section 1.2, we introduce the main mathematical notions of univariate and multivariate MFA. For an exhaustive presentation, the reader is referred to [Jaf97a, Jaf97b, DMA01] for a univariate formulation and to [MSKF90, JSW+19a, JSW+19b] for a multivariate formulation. In Section 1.3, the main practical aspects are presented. In Section 1.4, we introduce the multifractal processes that will be used throughout this thesis for the numerical validation of the proposed contributions and tools. Finally, in Section 1.5, we comment on the existing procedures for the estimation of parameters characterizing the multifractal properties of data and we highlight their limitations.

Note that, while the theoretical concepts and practical tools for MFA can, in principle, be given

for arbitrary dimension, cf., [Jaf04], the choice has been taken to focus the presentation on signals (1D) and images (2D).

1.2 Multifractal analysis

Let $X(\boldsymbol{t}) \in \mathbb{R}$, $\boldsymbol{t} \in \mathbb{R}^d$, denote a univariate locally bounded function under analysis with $\boldsymbol{t} \in \mathbb{R}$ ($d = 1$) for time series and $\boldsymbol{t} = (t_1, t_2) \in \mathbb{R}^2$ ($d = 2$) for images.

MFA is a mathematical tool which provides a quantification of the fluctuations along time or space of the local regularity of X. The *Hölder exponent* $h_X(\boldsymbol{t}_0) > 0$ is most commonly used to measure the local regularity (see, e.g., [Rie03, Jaf04] for details) and is defined as follows. The function X is said to belong to $\mathcal{C}^\alpha(\boldsymbol{t}_0)$ at position $\boldsymbol{t}_0$ if there exist constants $K, \alpha > 0$ and a polynomial P of order strictly smaller than α such that

$$||X(\boldsymbol{t}) - P(\boldsymbol{t} - \boldsymbol{t}_0)|| \leq K||\boldsymbol{t} - \boldsymbol{t}_0||^\alpha, \tag{1.1}$$

for $\boldsymbol{t}$ sufficiently close to $\boldsymbol{t}_0$ ($||\cdot||$ stands for the Euclidean norm). The Hölder exponent is then defined as the largest value of α such that the inequality (1.1) is satisfied, i.e.,

$$h_X(\boldsymbol{t}_0) \triangleq \sup\{\alpha : X \in \mathcal{C}^\alpha(\boldsymbol{t}_0)\}. \tag{1.2}$$

Fig. 1.1 illustrates the definition (1.2) for a univariate function $X(\boldsymbol{t})$ with $\boldsymbol{t} \in \mathbb{R}$.

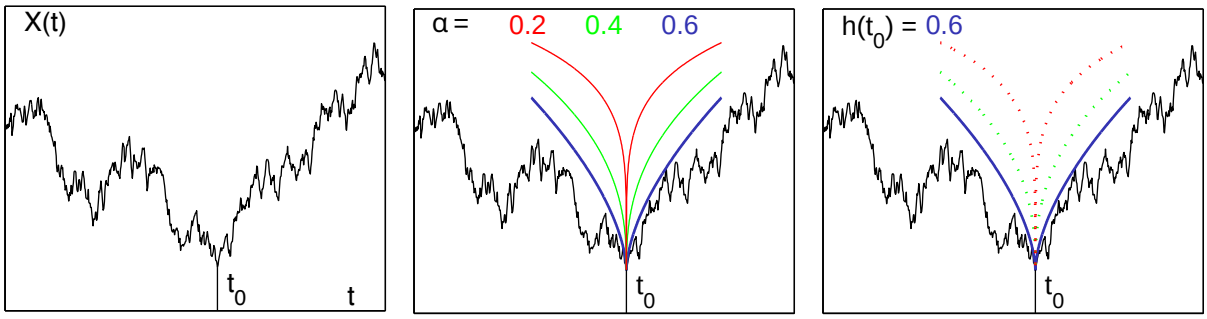

Figure 1.1: Definition of Hölder exponent of a 1D function $X(\boldsymbol{t})$ at $\boldsymbol{t}_0$ (left plot). Different values of α (center plot) such that the inequality (1.1) is satisfied. Hölder exponent equals the largest value of α (right plot).

The Hölder exponent is a special case of p-exponents $h_X^p(\boldsymbol{t}_0)$, where $p > 0$ is a parameter, which is a recently studied family of regularity exponents, see, e.g., [JML+16, LWA+16]. Qualitatively, the closer $h_X(\boldsymbol{t}_0)$ to 0, the rougher X at position $\boldsymbol{t}_0$ and the larger $h_X(\boldsymbol{t}_0)$, the smoother X at $\boldsymbol{t}_0$. The definition (1.2) comprises only non-negative exponents, which is guaranteed by the assumption that X is locally bounded. However, it has been reported that real-world data often violate this assumption, see, e.g., [WAJ+09, WRJA09]. Further comments on this assumption and a practical solution to relax it are discussed in Section 1.3.3.

The focus of MFA is to provide a global description of the fluctuations of the local regularity of the function X, rather than providing the function $h_X(\boldsymbol{t})$ for each time or space instance $\boldsymbol{t}$. This is achieved via the so-called multifractal spectrum $\mathcal{D}(h)$, which is formally defined as the collection of fractal (Hausdorff) dimension ($\dim_H$) of the sets of points that have identical exponent h, i.e.,

$$\mathcal{D}(h) \triangleq \dim_H \big\{ \boldsymbol{t} : h_X(\boldsymbol{t}) = h \big\}, \tag{1.3}$$

where theoretically $0 \leq \mathcal{D}(h) \leq d$ and, by convention, $\mathcal{D}(h_0) = -\infty$ if h_0 is not a Hölder exponent observed in X. Broadly speaking, $\mathcal{D}(h)$ is a measure of the geometrical importance of different Hölder exponents, disregarding any information on their precise geometric distribution. For more technical details, the reader is referred to [Jaf04, JAW15].

The above framework permits the MFA for one single function and regularity exponent. Multivariate MFA deals with the simultaneous MFA of several pointwise exponents derived from one or several functions and is briefly recalled in what follows.

1.2.1 Multivariate multifractal analysis

Let $\boldsymbol{X}(\boldsymbol{t}) = (X_1(\boldsymbol{t}), \ldots, X_R(\boldsymbol{t})) \in \mathbb{R}^R$ denote an R-variate function under analysis with R locally bounded components $X_r(\boldsymbol{t}) : \mathbb{R}^d \to \mathbb{R}$, with $r \in \{1, \ldots, R\}$. The Hölder exponent of $\boldsymbol{X}$ at $\boldsymbol{t}_0$ is defined as

$$\boldsymbol{h_X}(\boldsymbol{t}_0) \triangleq (h_{X_1}(\boldsymbol{t}_0), \ldots, h_{X_R}(\boldsymbol{t}_0)), \tag{1.4}$$

where $h_{X_r}(\boldsymbol{t}_0) > 0$ is the Hölder exponent associated with the rth component X_r, defined as in (1.2).

The R-variate (or joint) multifractal spectrum, denoted as $\mathcal{D}_R(\boldsymbol{h})$, of $\boldsymbol{X}$ is defined as the collection of Hausdorff dimensions of the sets of points $\boldsymbol{t}$ at which $\boldsymbol{h}_{\boldsymbol{X}}(\boldsymbol{t})$ takes the same value $\boldsymbol{h} = (h_1, \ldots, h_R)$, i.e.,

$$\mathcal{D}_R(\boldsymbol{h}) \triangleq \dim_H \big\{\boldsymbol{t} : \boldsymbol{h}_{\boldsymbol{X}}(\boldsymbol{t}) = \boldsymbol{h}\big\}, \tag{1.5}$$

see [MSKF90, JSW+19a, JSW+19b] for details. Note that if $R = 1$ we are in the univariate setting. The shape, width, and orientation of the function $\mathcal{D}_R(\boldsymbol{h})$ with respect to the $\boldsymbol{h}$-axes quantify the degree of local fluctuation of the regularity of the components of $\boldsymbol{X}$, and to what extent these fluctuations are coupled between components.

Therefore, the multifractal spectrum $\mathcal{D}_R(\boldsymbol{h})$ is the object of central interest to MFA. However, its practical estimation from real-world data cannot be conducted based on its formal definition (1.5) as, in practice, only a discretized version of $\boldsymbol{X}$ with finite resolution is available. Indeed, its estimation requires to use robust numerical methods well-known as multifractal formalisms [FP85, Jaf04].

1.3 Multifractal formalism using wavelet leaders

Multifractal formalisms provide a link between the multiscale statistics of specifically tailored multiresolution quantities and the multifractal spectrum. Several different univariate multifractal formalisms have been proposed in the literature, relying on different multiresolution coefficients. Some historical examples are the increments [FP85], the wavelet coefficients [AAD+02], the wavelet transform modulus maxima representations [MBA93, KLSJA01], the wavelet leaders [Jaf04, LJA05, WAJ07] or most recently the p-leaders [JML+16, LWA+16].

In this thesis, we make use of wavelet leaders, derived from the wavelet coefficients, which have been proven to possess the key theoretical and practical properties for MFA purposes based on the Hölder exponent and have resulted in the current state of the art multifractal formalism, see, e.g., [Jaf04, LJA05, WAJ07] for details. The multifractal formalism constructed from the wavelet leaders has been first developed for multivariate data in [JSW+19a, JSW+19b] and is briefly recalled in this section.

1.3.1 Wavelet leaders

In this manuscript, we only consider the analysis of 1D ($d = 1$) and 2D ($d = 2$) data. Therefore, we only provide the theoretical definition of 1D and 2D wavelet leaders. The extensions to the case $d > 2$ are straightforward, see, e.g., [Jaf04, JLA06, JAR+10].

Let $X(\boldsymbol{k})$ denote the discretized version of the locally bounded univariate function $X(\boldsymbol{t}) \in \mathbb{R}$. For simplicity and without loss of generality, $X(\boldsymbol{k})$ is assumed to be square, i.e., $\boldsymbol{k} \in \{1, 2, \ldots, N\}^d$.

Wavelet coefficients in 1D. For $d = 1$, let $\psi \in \mathcal{C}^{N_\psi - 1}$ denote a *mother wavelet*, which is an oscillating reference pattern that is characterized by its number of vanishing moments, a positive integer $N_\psi \geq 1$ such that

$$\forall n = 0, \ldots, N_\psi - 1 : \int_{\mathbb{R}} \boldsymbol{t}^n \psi(\boldsymbol{t}) d\boldsymbol{t} \equiv 0 \text{ and } \int_{\mathbb{R}} \boldsymbol{t}^{N_\psi} \psi(\boldsymbol{t}) d\boldsymbol{t} \neq 0. \tag{1.6}$$

It is designed such that the collection

$$\{\psi_{j,\boldsymbol{k}}(\boldsymbol{t}) = 2^{-j/2} \psi(2^{-j}\boldsymbol{t} - \boldsymbol{k})\}_{(j,\boldsymbol{k}) \in (\mathbb{Z},\mathbb{Z})} \tag{1.7}$$

of its dilated and translated templates forms an orthonormal basis of $L^2(\mathbb{R})$ [Mal98], i.e.,

$$X(\boldsymbol{t}) = \sum_{j,\boldsymbol{k}} D_X(j, \boldsymbol{k}) \psi_{j,\boldsymbol{k}}(\boldsymbol{t}), \tag{1.8}$$

where $D_X(j, \boldsymbol{k}) = \langle \psi_{j,\boldsymbol{k}}, X \rangle$ are the discrete wavelet coefficients and $\langle \, . \, , \, . \, \rangle$ is the inner product.

Wavelet coefficients in 2D. For $d = 2$, the construction of an orthonormal discrete wavelet transform (DWT) can be practically defined via the use of four 2D filters $G^{(m)}(\boldsymbol{k}), m = 0, 1, 2, 3$, with $\boldsymbol{k} = (k_1, k_2)$, obtained as tensor products of the quadrature mirror filters H_0 and G_0 (low-pass and high-pass, respectively) defining a 1D orthonormal DWT relying on a mother wavelet ψ. By convention, $G^{(0)}(\boldsymbol{k}) = H_0(k_1)H_0(k_2)$ corresponds to the 2D low pass filter yielding the approximation coefficients $D_X^{(0)}(j, \boldsymbol{k})$, whereas $G^{(1)}(\boldsymbol{k}) = G_0(k_1)H_0(k_2)$, $G^{(2)}(\boldsymbol{k}) = H_0(k_1)G_0(k_2)$ and $G^{(3)}(\boldsymbol{k}) = G_0(k_1)G_0(k_2)$ correspond to the high pass filters yielding the wavelet (detail) coefficients $D_X^{(m)}(j, \boldsymbol{k}), m = 1, 2, 3$. Specifically, $D_X^{(m)}(j, \boldsymbol{k}), m = 0, 1, 2, 3$, are obtained by, first at the finest scale $j = 1$, convolving the discrete image X with $G^{(m)}, m = 0, 1, 2, 3$, and decimation; for the coarser scales

$j \geq 2$ they are iteratively obtained by convolving $G^{(m)}, m = 0, 1, 2, 3$ with $D_X^{(0)}(j-1, \cdot)$ and decimation. For formal definition details on wavelet transforms, the reader is referred to [Mal98, AMVA04].

MFA is robust to the choice of wavelet as long as it is smooth enough [VA99, TA03]. In this thesis, we use a Daubechies' least asymmetric wavelet [Dau88] with a sufficient number of vanishing moment N_ψ. This is a common choice since its mother wavelet ψ gathers a number of attractive theoretical properties for scaling analysis.

Normalization of wavelet coefficients. For scaling and MFA purposes, it is common to normalize the wavelet coefficients according to a L^1-norm [BMA93, ABF+02] as follows

$$d_X(j, \boldsymbol{k}) = 2^{-j/2} D_X(j, \boldsymbol{k}),\ \boldsymbol{k} \in \mathbb{Z}, \tag{1.9}$$

$$d_X^{(m)}(j, \boldsymbol{k}) = 2^{-j} D_X^{(m)}(j, \boldsymbol{k}),\ \boldsymbol{k} \in \mathbb{Z}^2,\ m = 1, 2, 3. \tag{1.10}$$

Therefore, we will refer to d_X and $d_X^{(m)}$, $m = 1, 2, 3$, as the (discrete) wavelet coefficients.

Wavelet leaders in 1D. For $d = 1$, the *wavelet leader* $L_X(j, \boldsymbol{k})$, is defined as the largest coefficient modulus, taken over finer scales and within a short temporal neighborhood $3\lambda_{j,\boldsymbol{k}}$, with

$$\lambda_{j,\boldsymbol{k}} = [\boldsymbol{k}2^j, (\boldsymbol{k}+1)2^j)$$

denoting the dyadic interval of size 2^j and $3\lambda_{j,\boldsymbol{k}}$ standing for the union of $\lambda_{j,\boldsymbol{k}}$ with its 2 neighbors, i.e.,

$$L_X(j, \boldsymbol{k}) \triangleq \sup_{\lambda' \subset 3\lambda_{j,k}} |d_X(\lambda')|. \tag{1.11}$$

The definition (1.11) is illustrated in Fig. 1.2.

Wavelet leaders in 2D. For $d = 2$, the supremum in the definition of the wavelet leaders (1.11) is taken for the detail coefficients and for all eight direct neighbors, i.e.,

$$L_X(j, \boldsymbol{k}) \triangleq \sup_{m \in \{1,2,3\},\ \lambda' \subset 3\lambda_{j,\boldsymbol{k}}} |d_X^{(m)}(\lambda')|, \tag{1.12}$$

where

$$\lambda_{j,\boldsymbol{k}} = [k_1 2^j, (k_1+1)2^j) \times [k_2 2^j, (k_2+1)2^j)$$

denotes the dyadic cube of side length 2^j centered at position $\boldsymbol{k}$ and

$$3\lambda_{j,\boldsymbol{k}} = \bigcup_{n_1, n_2 \in \{-1,0,1\}} \lambda_{j,(k_1+n_1, k_2+n_2)}$$

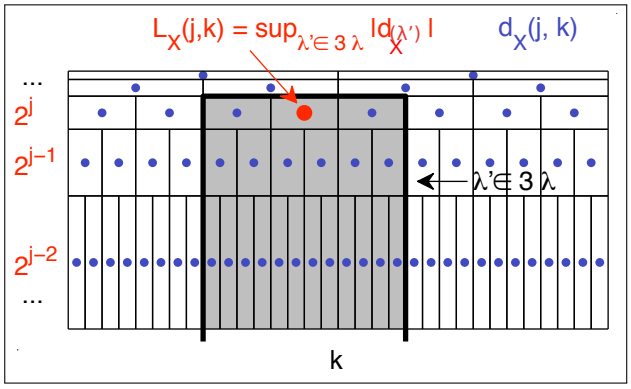

Figure 1.2: Definition of wavelet leaders (1D) [Wen08]: The wavelet leader $L_X(j,\boldsymbol{k})$ (1.11), at scale 2^j and position $\boldsymbol{k}$ (red fat dot), is defined as the largest $|d_X(j,\boldsymbol{k})|$ (blue dots) within the time neighborhood $3\lambda_{j,\boldsymbol{k}}$ over all finer scales $2^{j'} < 2^j$ (area in gray, truncated at fine scales).

the union of this cube with its eight neighbors. The definition (1.12) is illustrated in Fig. 1.3.

Multivariate wavelet leaders. For a discrete R-variate function $\boldsymbol{X}(\boldsymbol{k}) = (X_1(\boldsymbol{k}), \ldots, X_R(\boldsymbol{k}))$, let $L_{\boldsymbol{X}}(j,\boldsymbol{k})$ denote the vector of wavelet leaders such that its rth component $L_{X_r}(j,\boldsymbol{k})$ is the wavelet leader associated with X_r at position $\boldsymbol{k}$ and scale 2^j, i.e.,

$$L_{\boldsymbol{X}}(j,\boldsymbol{k}) = L_{\boldsymbol{X}}(\lambda_{j,\boldsymbol{k}}) \triangleq (L_{X_1}(j,\boldsymbol{k}), \ldots, L_{X_R}(j,\boldsymbol{k})). \tag{1.13}$$

Wavelet leaders can be shown to reproduce Hölder exponents in the limit of fine scales as follows

$$h_{\boldsymbol{X}}(\boldsymbol{t}_0) = \liminf_{j\to+\infty} \ln L_{\boldsymbol{X}}(\lambda_{j,\boldsymbol{k}}(\boldsymbol{t}_0))/\ln 2^{-j}, \tag{1.14}$$

where $\lambda_{j,\boldsymbol{k}}(\boldsymbol{t}_0)$ denotes the dyadic cube of width 2^{-j} which contains $\boldsymbol{t}_0$. The construction of the multifractal formalism described in the following section fundamentally relies on the key property (1.14). For an exhaustive presentation, the reader is referred to [JSW+19a].

1.3.2 Multivariate wavelet leader multifractal formalism

Let $S(j,\boldsymbol{q})$ denote the empirical moments of $L_{\boldsymbol{X}}(j,\boldsymbol{k})$ at scale 2^j, referred to as the *multivariate structure functions* and defined as

$$S(j,\boldsymbol{q}) = \frac{1}{n_j} \sum_{\boldsymbol{k}} \prod_{r=1}^{R} L_{X_r}(j,\boldsymbol{k})^{q_r}, \tag{1.15}$$

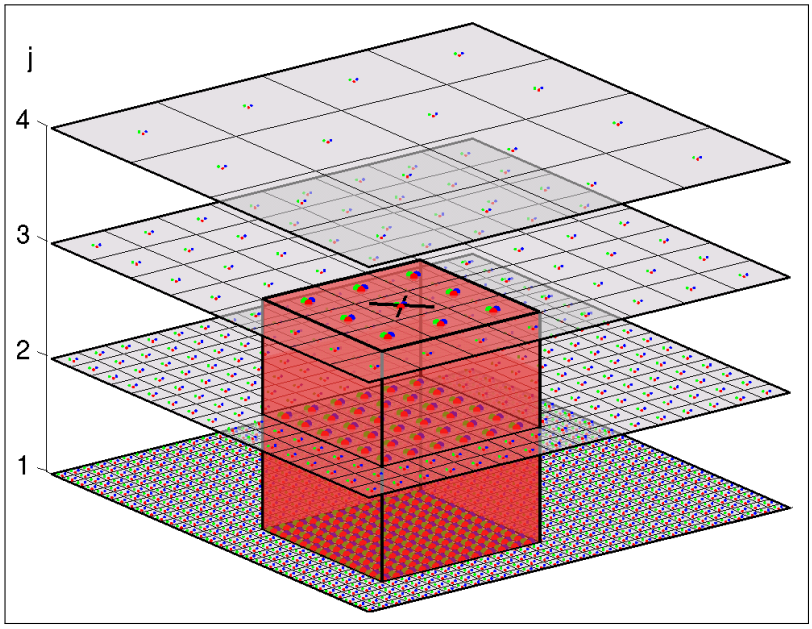

Figure 1.3: Definition of wavelet leaders (2D) [Wen08]: The wavelet leader $L_X(j,\boldsymbol{k})$ (1.12), at scale 2^j and spatial position $\boldsymbol{k}$ (black cross), is defined as the largest $|d_X^{(m)}(j,\boldsymbol{k})|$, $m \in \{1,2,3\}$, (red, green, blue dots) within the spatial neighborhood $3\lambda_{j,\boldsymbol{k}}$ over all finer scales $2^{j'} < 2^j$ (red volume, truncated at fine scales). The wavelet coefficients over which the supremum is taken are marked by fat dots.

where $\boldsymbol{q}=(q_1,\ldots,q_R)$ and $n_j \approx \lfloor N^d/2^{dj} \rfloor$ is the number of wavelet leaders available at scale 2^j for a single data component.

The corresponding *scaling function* is defined as

$$\zeta_R(\boldsymbol{q}) = \liminf_{j\to+\infty} \ln S(j,\boldsymbol{q}) / \ln 2^{-j}, \tag{1.16}$$

and is founded to be tightly related to $\mathcal{D}_R(\boldsymbol{h})$ via their Legendre transform, referred to as the *multivariate Legendre spectrum*

$$\mathcal{L}_R(\boldsymbol{h}) \triangleq \inf_{\boldsymbol{q}} (d + \langle \boldsymbol{q}, \boldsymbol{h} \rangle - \zeta_R(\boldsymbol{q})). \tag{1.17}$$

Finally, $\mathcal{L}_R(\boldsymbol{h})$ provides an upper-bound for $\mathcal{D}_R(\boldsymbol{h})$

$$\mathcal{D}_R(\boldsymbol{h}) \leq \mathcal{L}_R(\boldsymbol{h}), \tag{1.18}$$

for large classes of processes and is in practice used as an estimator of $\mathcal{D}_R(\boldsymbol{h})$. This statement is more restricted in the multivariate case than the univariate one, but is nevertheless valid for many

process constructions that necessarily require a *synchronicity* condition. For technical details, the reader is referred to [JSW+19a].

Log-cumulants. Let

$$\ell_{X_r}(j,\boldsymbol{k}) \triangleq \ln L_{X_r}(j,\boldsymbol{k}) \tag{1.19}$$

denote the logarithm of the wavelet leaders (aka log-leaders) associated with X_r at a fixed scale 2^j and position $\boldsymbol{k}$. The first-order (mean) and second-order (auto- and cross-covariances) cumulants of $\ell_{X_r}(j,\boldsymbol{k})$ and $\ell_{X_{r'}}(j,\boldsymbol{k})$, with $r,r' \in \{1,\dots,R\}$, take the form [CGM93, WLA+18]

$$\mathbb{E}[\ell_{X_r}(j,\boldsymbol{k})] = c_1^0(r) + c_1(r)\ln 2^j, \tag{1.20}$$

$$\mathrm{Cov}(\ell_{X_r}(j,\boldsymbol{k}),\ell_{X_{r'}}(j,\boldsymbol{k})) = c_2^0(r,r') + c_2(r,r')\ln 2^j, \tag{1.21}$$

where $c_\cdot^0$ are model adjustment parameters not related to the multifractal spectrum. Similar expressions are obtained for the higher-order R-variate cumulants [SL87, AWJD19]. The use of (1.20) and (1.21) 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 [JSW+19a, AWJD19]:

$$\begin{aligned}\mathcal{D}_2(h_r,h_{r'}) \approx d + \frac{c_2(r',r')b}{2}\left(\frac{h_r-c_1(r)}{b}\right)^2 + \frac{c_2(r,r)b}{2}\left(\frac{h_{r'}-c_1(r')}{b}\right)^2 \\ - c_2(r,r')b\left(\frac{h_r-c_1(r)}{b}\right)\left(\frac{h_{r'}-c_1(r')}{b}\right),\end{aligned} \tag{1.22}$$

where $c_2(r,r) < 0$ and $b \triangleq c_2(r,r)c_2(r',r') - c_2(r,r')^2 \geq 0$ [LAR+18]. Notably, the leading order c-coefficients provide a relevant summary of the multifractal properties of $\boldsymbol{X}$ in applications where it would often not be convenient to handle an entire function $\mathcal{D}(\boldsymbol{h})$. Specifically,

- $(c_1(r),c_1(r'))$ indicates the position of the maximum of $\mathcal{D}_2(h_r,h_{r'})$, which corresponds to the average degrees of $X_r, X_{r'}$ regularity,
- $c_2(r,r)$ quantifies the amount of pointwise regularity fluctuations (multifractality) for the rth component,
- $c_2(r,r')$ characterizes the coupling between the regularity fluctuations of the rth and r'th components.

Note that the projections of $\mathcal{D}_2(h_r, h_{r'})$ on the $h_r = 0$ and $h_{r'} = 0$ planes correspond to the marginal univariate spectra and are entirely controlled by the univariate parameters $c_1(r), c_2(r,r)$ and $c_1(r'), c_2(r',r')$.

In the multivariate setting, it is natural to seek to define the normalized coupling parameter

$$\rho_{\mathrm{mf}}(r,r') \triangleq -\frac{c_2(r,r')}{\sqrt{c_2(r,r)c_2(r',r')}} \in [-1,1]. \tag{1.23}$$

In view of the model defined in Section 1.4.1, it can be interpreted as a *multifractal correlation* and can be shown to quantify *higher-order dependence beyond linear correlation* among the data components $\{X_r\}_{r=1}^R$ [WLA+18, LAR+18, ALW+19]. An estimator of ρ_{mf} can be defined by replacing the coefficients in (1.23) by estimates as in (1.28) [WLA+18]. The type of information that can be captured by ρ_{mf} is illustrated in Fig. 1.4, which shows a 3-variate synthetic multifractal image (defined in Section 1.4.1), 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 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.

Therefore, conducting a multivariate MFA rather than analyzing the data components individually can potentially provide a richer characterization, and in particular unravel the dynamics, coupling mechanisms and dependencies between the different registered components. Unfortunately, the accurate estimation of the associated multivariate multifractal parameters is extremely challenging, which severely limits their actual use in applications (see Section 1.5).

This thesis addresses these limitations, proposes and studies practical tools for the estimation of parameters associated with the joint multifractal spectrum (cf. Chapters 2 and 3) and investigates the potential benefits of multivariate MFA for real-world multivariate signal and image processing (cf. Chapter 5).

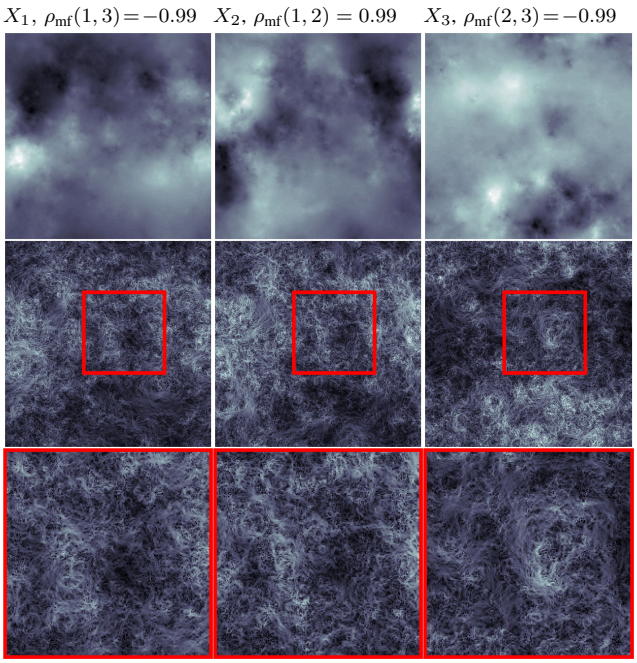

Figure 1.4: Illustrations for multifractal correlation: synthetic 3-variate multifractal image (top row, defined in [BDM01, WLA+18] and Section 1.4.1), magnitude of gradients (second row) and zooms of the patch marked by red square (bottom row). X_1 and X_2 have multifractal correlation $\rho_{\mathrm{mf}}(1,2) = 0.99$, and $\rho_{\mathrm{mf}}(1,3) = \rho_{\mathrm{mf}}(2,3) = -0.99$ with component X_3. The linear correlation equals zero for all components.

1.3.3 Negative regularity

The wavelet leader multifractal formalism presented above is only well defined for locally bounded functions. However, it has been reported that a large number of real-world signals and images do not satisfy this prerequisite, see, e.g., [WAJ+09, WRJA09]. In these cases, a practical solution consists of constructing the multifractal formalism using the modified wavelet coefficients

$$d_X^{\gamma}(j,\boldsymbol{k}) = 2^{j\gamma/2} d_X(j,\boldsymbol{k}),\ \boldsymbol{k} \in \mathbb{Z}, \tag{1.24}$$

$$d_X^{(m),\gamma}(j,\boldsymbol{k}) = 2^{j\gamma} d_X^{(m)}(j,\boldsymbol{k}),\ \boldsymbol{k} \in \mathbb{Z}^2,\ m = 1,2,3, \tag{1.25}$$

with $\gamma > 0$ instead of $d_X(j,\boldsymbol{k})$ and $d_X^{(m)}(j,\boldsymbol{k})$ in (1.9) and (1.10). The parameter γ can be chosen sufficiently large to ensure that the multifractal formalism is properly defined. For precise definitions

and details about the theoretical and practical consequences implied by this modification, the reader is referred to [WAJ07, WRJA09, AJW15].

Another alternative solution is the use of a multifractal formalism based on p-leaders [JML+16, LWA+16], which allow us to measure the regularity of non-locally bounded functions. This approach is however beyond the scope of this thesis and shall be considered for future work.

1.4 Multifractal processes

The Mandelbrot cascades [Man99], consisting of split/multiply iterative constructions that induce multifractal properties, are historical and benchmark models for multifractal processes. Multivariate extensions of cascades were however barely considered and used (see a contrario [MSKF90]). Alternatively, multifractal random walk (MRW) processes were constructed as more realistic models for real world data [BDM01], notably with signed increments. Their construction is based on the increments of fractional Brownian motion (fBm), the reference Gaussian self-similar process [ST94], whose variance is modulated using an independent process whose properties mimic those of Mandelbrot cascades, and hence impart their multifractality to the MRW [BDM01]. Its multivariate extension was first considered in the unpublished work [BDM00] and used in [WLA+18, LAR+18] to illustrate the nature of the information captured in the bivariate multifractal spectrum. Nowadays, the multivariate multifractal random walk (MV-MRW) [BDM00, MDB00, WLA+18] has become a canonical multifractal model process for multivariate data. In this thesis, we will use the MV-MRW to illustrate the proposed approach and assess its performance.

1.4.1 Multivariate multifractal random walk

Definition. The construction of an MV-MRW for R components requires two collections of those stochastic processes: (i) a collection of increments of fractional Brownian motions $(G_1(\boldsymbol{t}), \ldots, G_R(\boldsymbol{t}))$, which is determined by the self-similarity parameters $H_1, \ldots, H_R$ and an $R \times R$ point covariance $\boldsymbol{\Sigma}_{ss}$, with corresponding correlation coefficients $\rho_{ss}(r, r')$, and (ii) a collection of Gaussian processes

$(\omega_1(\boldsymbol{t}), \ldots, \omega_R(\boldsymbol{t}))$ with prescribed covariance function $\boldsymbol{\Sigma}_{\mathrm{mf}}$, with entries given by

$$[\boldsymbol{\Sigma}_{\mathrm{mf}}]_{r,r'}(\mathbf{k},\mathbf{l}) = [\boldsymbol{\Pi}_{\mathrm{mf}}]_{r,r'} \lambda_r \lambda_{r'} \ln\left(\frac{T}{||\boldsymbol{k}-\boldsymbol{l}||+1}\right),\ r,r' \in \{1,\ldots,R\},$$

for $||\boldsymbol{k}-\boldsymbol{l}|| \leq T-1$ and 0 otherwise, where T is an arbitrary integral scale, equal to the data sample size in the rest of the manuscript. To simplify notations, we consider $[\boldsymbol{\Pi}_{\mathrm{mf}}]_{r,r} = 1$ and $[\boldsymbol{\Pi}_{\mathrm{mf}}]_{r,r'} = \rho_{\mathrm{mf}}(r,r')$. These processes are numerically synthesized as described in [HPA11]. Each component X_r, $r \in \{1,\ldots,R\}$, 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 $c_2(r,r') = -\rho_{\mathrm{mf}}(r,r')\lambda_r\lambda_{r'}$ [BDM00, WLA+18], whereas its higher-order cumulants are equal to zero. Examples of realizations of a 2D MV-MRW for $R=3$ are plotted in Fig. 1.4.

Typical values of $c_2(r,r)$ for real-world data range from zero (no multifractality) down to $c_2(r,r) \approx -0.25$, which corresponds to an extremely intermittent signal that is rarely observed for a non-pathological physical signal. This range of values $-0.25 \lesssim c_2(r,r) \leq 0$ covers, however, a huge range of observed intermittencies. Let us frame this with a statistical example: for $c_2(r,r) \to 0$, MRW converges to Gaussian fBm, with no multifractality, whose moments of all orders exist; for a multifractality parameter such that $|c_2(r,r)| = 0.08$, the moments of (the increments of) MRW exist only to order 5 (included); for larger values even less moments exist, which illustrates the strong degree and variability of intermittency that is covered by the orders of magnitude considered for the parameter $c_2(r,r)$.

1.5 Estimation procedures

In this section we provide a brief review of the state of the art of the most important estimators used in MFA. We point out their main limitations and highlight the absence of efficient estimators in the multivariate formulation.

1.5.1 Estimators tied to specific models

Several parametric model-based methods have been proposed for univariate MFA and are briefly recalled below.

These approaches include the generalized method of moments (GMM) proposed and studied, e.g., in [Lux07, Lux08]. In essence, the GMM formulates parameter inference as the solution (in the least squares sense) of an over-determined system of equations derived from the moments of the data. However, this method relies heavily on fully parametric models for the data and achieves, to the best of our knowledge, only limited benefits in practical applications as shown in [BKM08, BKM13, BADML12] for financial time series.

Maximum likelihood (ML) methods are classical in parameter inference, yet their use in this context has been scarce and their definitions have been tied to specific instances of self-similar or multifractal processes only, see, e.g., [WO92, Ber94]. The key difficulty for their use stems from the intricate statistical nature of most of the multifractal models, characterized by highly non Gaussian marginals, scale-free distributions and strong dependence that remain poorly studied to date. The same remark is true for their wavelet coefficients and wavelet leaders [OW00, VWAJ10]. On the other hand, the fBm (in 1D) and fractional Brownian fields (in 2D) are jointly Gaussian self-similar processes with fully parametric covariance function suitable for ML and Bayesian estimation. Formulated in the spectral or wavelet domains, ML-based methods associated with 1D fBm process have been investigated in [Ber94, CP06, MRS08]. For images, an ML estimator has been proposed in [LOKS86], where the estimation problem is however reduced to a univariate formulation for the rows/columns of the image there. In [LR12], an ML approach formulated in the time domain for a specific 1D multifractal multiplicative cascade process was also proposed. Yet, this method strongly relies on the particular construction of this process and cannot be easily accommodated to more general classes of models or dimensions. In this context, a Bayesian method also has been introduced in [WO92] but is only relevant for the analysis of 1D self-similar signals.

1.5.2 Generic estimators

Linear regression. Regardless of the multiscale representations used, classical multifractal estimation essentially relies on log-log regressions. As suggested by (1.20) and (1.21), the estimation of the coefficients c_1 and c_2 can be conducted by linear regressions of the sample cumulants across scale j

$$\widehat{c}_1(r) = \frac{1}{\ln 2} \sum_{j=j_1}^{j_2} w_j \widehat{\mathbb{E}}[\boldsymbol{\ell}_{X_r}(j,\cdot)], \tag{1.26}$$

$$\widehat{c}_2(r,r') = \frac{1}{\ln 2} \sum_{j=j_1}^{j_2} w_j \widehat{\mathrm{Cov}}(\boldsymbol{\ell}_{X_r}(j,\cdot), \boldsymbol{\ell}_{X_{r'}}(j,\cdot)), \tag{1.27}$$

$$\widehat{\rho}_{\mathrm{mf}}(r,r') = -\frac{\widehat{c}_2(r,r')}{\sqrt{\widehat{c}_2(r,r)\widehat{c}_2(r',r')}}, \tag{1.28}$$

where $\widehat{\mathbb{E}}$ and $\widehat{\mathrm{Cov}}$ denote the sample mean and covariance, and $[\![j_1, j_2]\!]$ is the range of scales such that (1.20) and (1.21) hold, see, e.g., [WAJ07]. The linear regression weights w_j have to satisfy the usual constraints $\sum_{j=j_1}^{j_2} j w_j = 1$ and $\sum_{j=j_1}^{j_2} w_j = 0$. In this thesis, we consider the weighted linear regression (WLR) [AFTV00, VAT03] defined by

$$w_j = \frac{1}{n_j} \frac{V_0 j - V1}{V_0 V_2 - V_1^2} \tag{1.29}$$

where

$$V_i = \sum_{j=j_1}^{j_2} j^i b_j, i = 0, 1, 2. \tag{1.30}$$

Estimations based on log-log regressions are appealing for their simplicity and low computational cost. However, they have a limited practical estimation performance. In particular, these regressions intrinsically require the data to have a large enough sample size (of order 256^2 pixels for a single channel image) in order to allow their dynamics to develop along a set of scales ranging over several orders of magnitude. This challenge is significantly more severe for images than for signals: indeed, modulo border effects of the wavelet transform, the number of available scales is proportional to the logarithm of the number of samples for 1D signals and to the logarithm of the square root of the number of pixels for an image. For instance, for a 1D signal with $256 \times 256 = 65536$ samples, $j_2 = 13$

or 14 scales can be computed, while $j_2 = 4$ or 5 for an image of $N \times N = 256 \times 256$ pixels. As a consequence, images of sizes smaller than 256×256 and thus image patches cannot be relevantly analyzed in practice using (1.26) and (1.27).

In general, the regression-based methods lead to large estimation variances notably for limited sample size data. This issue becomes a critical challenge in particular for the second (and higher) order cumulants c_2 and for ρ_{mf}, for example, in several bivariate settings as explored in [WLA+18, ALW+19]. Moreover, these regressions can wrongly lead to positive values of $c_2(r, r)$ and, thus, to complex values of ρ_{mf} when using (1.28) for their estimation. They can also give values outside the interval of definition $[-1, 1]$ and even infinite values.

Statistical model for log-leaders and Bayesian inference. Most recently, a proposed heuristic semi-parametric model for the statistics of the log-leaders associated with univariate multifractal signals and images permitted the formulation of Bayesian frameworks to estimate the parameter associated with the univariate multifractal spectrum, see, e.g., [WDTA13, CWD+15]. Unlike the methods mentioned in the beginning of this section, these approaches do not rely on the assumption of specific model processes for the data. The developed Bayesian methodology showed excellent estimation performance, outperformed the classical linear regression and permitted to process small sample sized real-world data. In the same line of work, [WCA+18] proposed to conduct the analysis within a hierarchical Bayesian model that jointly describes the collection of multifractality parameters associated with the univariate multifractal spectra of different individual data components. This method combining the univariate statistical model of log-leaders with suitable multivariate priors that encode prior information on the univariate multifractal parameters. Moreover, it was shown to be useful for real-world hyperspectral image processing. These practical contributions to the univariate MFA were gathered in the PhD thesis [Com16].

All these works constitute an inspiration for this thesis. However, we here face a more general modeling problem (a multivariate analysis that requires the modeling and validation of the cross-component covariance behavior) and a different estimation problem (matrix-valued parameters instead of single-variables). The only technically related points to our current work are the use of a Gaussian model for the marginal distributions of log-leaders, the modeling of the auto-covariance

(note, however, that we use a different and better suited parametrization in the present work) and the strategies for evaluating the likelihood efficiently (using a Whittle approximation and data augmentation).

To sum up, the practical multivariate MFA remained so far based on log-log linear regression algorithms, which can be extremely challenging for small sample sizes that severely limits their actual use in real-world applications. The present thesis proposes to empower the multivariate MFA by complementing it with an adequate statistical modeling and Bayesian estimation framework.

1.6 Conclusions

In this chapter, we have introduced the main theoretical and practical concepts of MFA from the univariate to the multivariate modeling and analysis of multifractal times series and images. The main object of interest of MFA, the estimation of the multifractal spectrum, is achieved in practice via a multivariate multifractal formalism. Wavelet leaders are the multiresolution quantities that yield the current benchmark multifractal formalism, which will be used throughout this manuscript. The multifractal properties of multivariate multifractal data can be well summarized by the log-cumulants c_1 and c_2. The present thesis will focus on the joint estimation of second-order log-cumulants c_2 (auto- and cross-multifractality) and the multifractal correlation ρ_{mf} associated with the joint multifractal spectrum of multivariate multifractal signals and images.

Chapter 2

Multivariate statistical model and Bayesian estimation

Contents

2.1 Introduction

Multifractal analysis (MFA) has been successfully used in many signal and image processing applications (classification, detection, etc, see, e.g., [JAW15] for a review). However, these successes have been limited to the analysis of univariate data. This limitation is crucial nowadays, given the increasing number of applications where the acquired data are multivariate, e.g., multitemporal, multispectral or multimodal images. MFA of multivariate data has been addressed in the past, but still limited to the individual analysis of the data components, see, e.g., [WCA+18]. The theoretical foundations of multivariate MFA have only recently been settled [JSW+19a, JSW+19b]. It aims to enrich the data analysis by providing information potentially hidden in the coupling of the regularity data components. Yet, the estimation of the parameters associated with the multivariate multifractal spectrum, the core of multivariate MFA, still relies on classical linear regression-based algorithms which suffer from limited practical performance (see Section 1.5.2).

The main goal of this chapter is to provide practical tools for accurately performing multivariate MFA. Due to the statistical nature of multifractal processes (strongly non-Gaussian and with intricate dependence), the design of accurate estimators of multivariate multifractal parameters becomes very challenging, in particular, when the sample size is small (e.g., notably including a range of biomedical applications). Therefore, we require generic models that need as little assumptions as possible from the data and also to design computationally efficient estimators to deal with large data sizes and potentially many data components.

To this end, we propose to elaborate a semi-parametric model for the multivariate multiscale statistics of the logarithm of wavelet leaders (referred to as log-leaders for short) and make use of Bayesian inference for the corresponding estimation problem. The methodology adopted here follows the strategies developed in [WDTA13, CWD+15, WCA+18] where Bayesian approaches were proposed for the estimation of the multifractal parameters associated with individual components of time series or images. However, here we face a different problem, the estimation of the matrix parameters associated with the joint multifractal spectrum of multivariate signals and images, and new methods and algorithms are developed to address it.

The remainder of this chapter is organized as follows.

In Section 2.2.1, we propose an empirical second-order statistical model for the log-leaders associated with multiplicative cascades, exemplified by the multivariate multifractal random walk (MV-MRW) process. This generic model combines the validation of a multivariate Gaussian distribution with a multiscale covariance structure controlled by two symmetric positive definite (p.d.) matrix-valued parameters containing the multifractality parameters of interest $\{c_2(r,r')\}_{r,r'=1}^{R}$. This model is the cornerstone of the inference methods developed in this thesis.

The evaluation of the likelihood arising from the proposed model requires the inversion of large covariance matrices, which can become computationally and numerically challenging, even for small sample sizes. To bypass this issue, we resort in Section 2.2.2 to a Whittle-type approximation [Whi53] to efficiently evaluate the likelihood in the spectral domain. Consequently, a suitable model in the Fourier domain is presented in Section 2.2.3 and a data augmentation strategy is used in Section 2.2.4 to express the Whittle approximation as the marginal distribution of an augmented likelihood leading to a function that is separable with respect to (w.r.t.) the multifractal matrix-valued parameters to be estimated.

Separability further permits to propose efficient estimation algorithms (Section 2.3) based on relevant choices of prior distributions. In particular, we propose to study the use of the inverse Wishart prior which is appealing for its conjugacy property that leads to closed-form conditional distributions that simplifies the parameter inference. Its use however intrinsically induces dependence between the parameter estimates and can lead to biased results. As an alternative, we also propose to study the scaled inverse Wishart prior, which consist in a specific decomposition of the matrix-valued parameters that enables us to incorporate semi-independently prior information to each matrix element, yet at cost of not having closed-form expressions anymore.

In Section 2.3, we specify the resulting conditional distributions and use two different Markov chain Monte Carlo (MCMC) algorithms for approximating the minimum mean square error (MMSE) estimator associated with the resulting posterior distributions. The approximation of the MMSE estimator is conducted by averaging of the generated symmetric p.d. matrix-valued samples. This average is usually computed using the classical arithmetic mean associated with the Euclidean metric.

As an alternative, we propose and study the use of a geometric mean associated with an appropriate Riemannian metric.

Finally, the combination of the inverse Wishart/scaled inverse Wishart priors and arithmetic/ geometric means results in four different Bayesian estimators. Their performance for multivariate multifractal estimation are assessed and compared against classical linear regressions, using extensive Monte Carlo simulations relying on synthetic multivariate multifractal processes for signals and images as defined in Section 1.4.1. Performance results, reported in Section 2.4 for different multifractal parameter settings and sample sizes, demonstrate a significant improvement in estimation performance achieved at moderate extra computational cost. This opens the way for the practical use of multivariate MFA on real-world data (see Chapter 5).

Developments and results presented in this chapter have been reported in [LWTA21, LWTA22b, LWTA22a].

2.2 Statistical model for log-leaders

In this section, we introduce a novel empirical second-order statistical model for the vector of log-leaders of multivariate multifractal processes (cf. Section 1.4.1)

$$\boldsymbol{\ell}(j,\boldsymbol{k}) \triangleq (\boldsymbol{\ell}_{X_1}(j,\boldsymbol{k}),\ldots,\boldsymbol{\ell}_{X_R}(j,\boldsymbol{k})) \in \mathbb{R}^R \tag{2.1}$$

where $\boldsymbol{\ell}_{X_r}(j,\boldsymbol{k})$ is defined as in expression (1.19). This model will be shown to be useful to estimate the multifractality parameters in multivariate scenarios using Bayesian frameworks.

2.2.1 Direct model

Marginal distributions. We first numerically investigate the joint distribution of the vector of log-leaders (2.1) associated with synthetic R-variate MRW processes introduced in Chapter 1. In the univariate case, theoretical arguments suggest that the marginal distributions of multiresolution coefficients of multifractal processes are approximately log-normal [Man90]. This has been studied and confirmed numerically for univariate wavelet leaders in [CWT+15a, WCA+18]. It is hence natural to seek to extend this modeling to the multivariate case $R > 1$. Illustrative examples are given in

Fig. 2.1, which shows gamma plots for scale $j=4$, $d=2$ and $R=2,6,8,10,20,50$ associated with 100 independent $2^{10}\times 2^{10}\times R$ synthetic images generated using a 2D MV-MRW process with $\lambda_r = \sqrt{r/100}$ and $\rho_{\mathrm{mf}}(r,r')$ uniform in $[0,0.5]$.

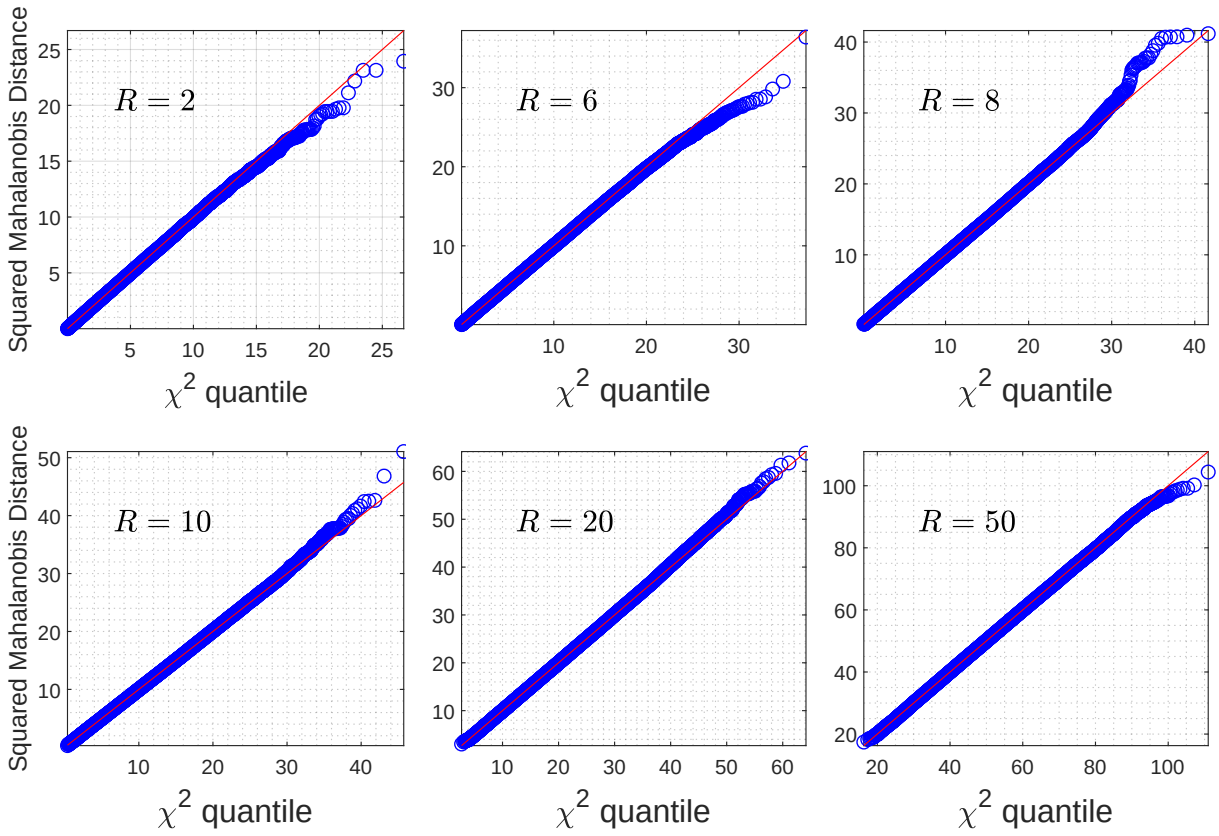

Figure 2.1: Gamma plots for the joint distribution ('$\circ$') of the empirical log-leaders at scale $j=4$, associated with 100 independent $2^{10}\times 2^{10}\times R$ synthetic images generated using a 2D MV-MRW, with $R=2,6,8,10,20,50$ (from left to right; $\lambda_r = \sqrt{r/100}$, $\rho_{\mathrm{mf}}(r,r')$ uniform in $[0,0.5]$). The closer to the red line, the better the approximation of the distribution by an R-variate normal distribution.

In short, a gamma plot is a generalization of the well-known QQ-plot to multivariate data and enables a visual assessment of the fit between the ordered squared generalized distances, from the smallest to largest, against the corresponding percentile of the chi-squared distribution. The closer to the red line, the better the approximation of the distribution by an R-variate normal distribution.

For further details on the assessment of the Gaussianity of multivariate data, the reader is referred to [JW02, OA16].

Similar results, not reported here for space reasons, were consistently obtained for different sample sizes, numbers of data components and multifractal parameter values, for both signals and images. As an additional illustration for one real-world data scenario, Fig. 2.2 (center plots) shows the corresponding gamma plots for a specific 256×256 patch of two bands (3 and 21) of a hyperspectral image acquired by the Hyspex hyperspectral scanner during the Madonna project [SFL+11]. All these results suggest that, overall, the empirical distribution of the log-leaders $\boldsymbol{\ell}(j, \boldsymbol{k})$ can indeed be well approximated by an R-variate Gaussian distribution.

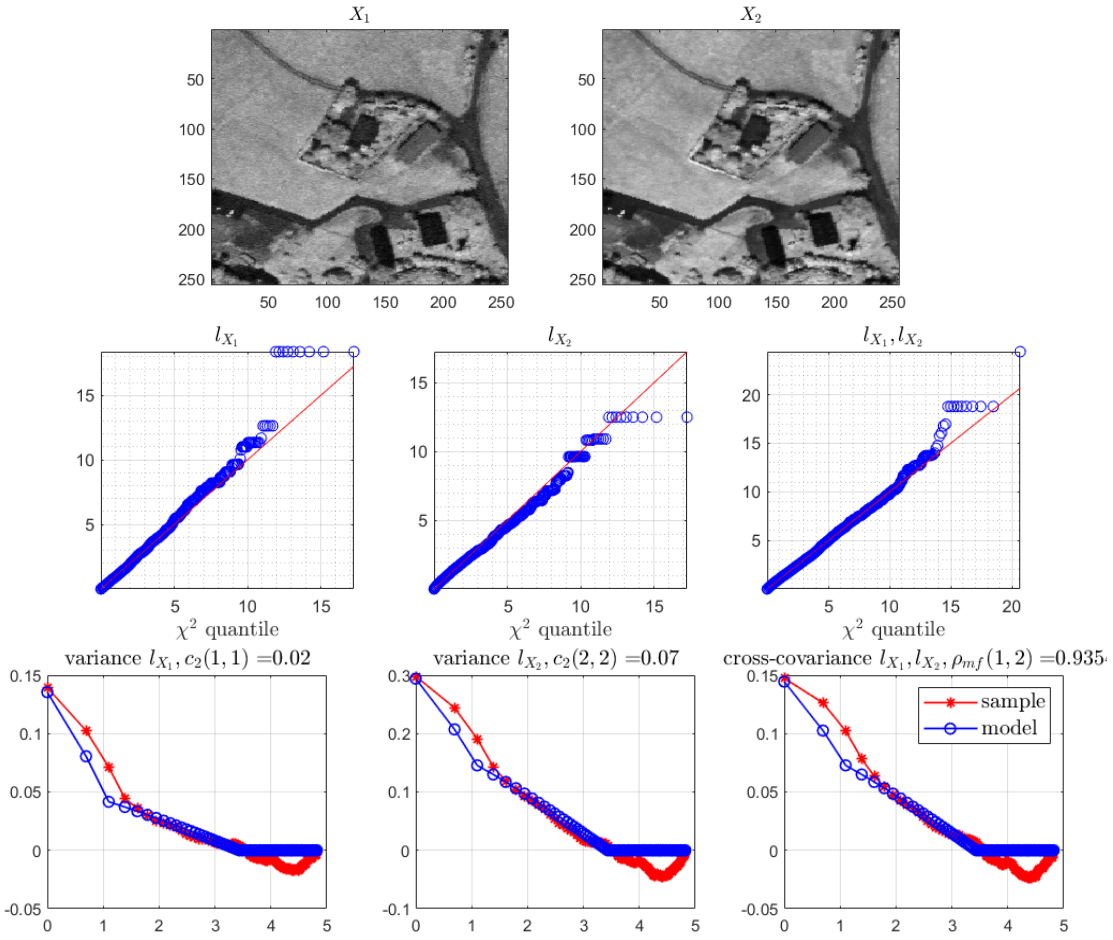

Figure 2.2: On top, patches of size 256×256 for two bands (3 and 21) of a hyperspectral image acquired by the Hyspex hyperspectral scanner during the Madonna project [SFL+11]. On center, gamma plots for the empirical log-leaders marginal and joint distributions at scale $j = 1$, associated with $\boldsymbol{X} = (X_1, X_2)$. On bottom, comparison between the proposed model (blue line) and the sample covariance (red line) at scale $j = 1$.

We furthermore propose the following model for the second-order statistics of the vector of the log-leaders.

Covariance. The theoretical results derived for univariate random wavelet cascades in [ADKR03] suggest a linear asymptotic decay for the auto-covariance of log-leaders, see also [CWT+15a, WCA+18] for an empirical study for a larger class of single-variable multifractal processes. Theoretical arguments available in, e.g., [MDB00, MSDA01] suggested similar linear asymptotic behavior also for the cross-covariance terms for multivariate multifractal processes. Inspired by that, we propose a covariance model for log-leaders of multivariate multifractal processes, with the key novel ingredient of a cross-term that describes the covariance between log-leaders of different components, parametrized by $c_2(r, r')$.

Model. Assuming that the vector of log-leaders associated with the signal/image under analysis is stationary and isotropic, the $R \times R$ auto-covariance matrix of the vectors $\boldsymbol{\ell}(j, \boldsymbol{k})$ at a fixed scale j is approximated by a radially symmetric function $\boldsymbol{S}_j(\rho)$ with $\rho \triangleq ||\boldsymbol{\Delta k}||$ as

$$\mathrm{Cov}(\boldsymbol{\ell}(j, \boldsymbol{k}), \boldsymbol{\ell}(j, \boldsymbol{k} + \boldsymbol{\Delta k})) \approx \boldsymbol{S}_j(\rho), \tag{2.2}$$

with

$$\boldsymbol{S}_j(\rho) = \boldsymbol{\Sigma}_1 f_1(j, \rho) + \boldsymbol{\Sigma}_2 f_2(j, \rho), \tag{2.3}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ 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 $\boldsymbol{\Sigma}_1$ equal the auto- and cross-multifractality parameters c_2, i.e., $-[\boldsymbol{\Sigma}_1]_{r,r'} = c_2(r, r')$, with $r, r' \in \{1, \ldots, R\}$. The matrix-valued parameter $\boldsymbol{\Sigma}_2$ 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 $f_2(j, \rho)$. Specifically,

$$f_1(j, \rho) = \max\left(0, -\ln \frac{\rho + 1}{\rho_j + 1}\right), \tag{2.4}$$

$$f_2(j, \rho) = \max\left(0, 1 - \frac{\ln(\rho + 1)}{\ln 4}\right), \tag{2.5}$$

where $\rho_j = \lfloor n_j^{1/d}/\kappa \rfloor$, with $\kappa = 5$ (1D) and $\kappa = 4$ (2D) fixed using cross-validation. The operator $\lfloor \cdot \rfloor$ truncates to integer values and n_j is defined in Chapter 1 as the total number of wavelet leaders of

a single data component at scale j. For $i \in \{1,2\}$ and all scales j, $f_i(j,\cdot)$ is a non-negative function and its shape resembles that used in the auto-covariance model described in e.g. [CWD+15], except that the new parameterization proposed here ensures that no change of coordinates is necessary. Therefore, assuming that $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are p.d. ensures that $\boldsymbol{S}_j$ is positive semi-definite (p.s.d.) for all scales j, as it is the sum of two p.s.d. matrices [HJ12].

As an illustrative example, the sample covariance 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 $\rho_{\mathrm{mf}}(1,2)=\rho_{\mathrm{mf}}(1,3)=\rho_{\mathrm{mf}}(2,3)=0.5$, is compared to the proposed covariance model $\boldsymbol{S}_j(\rho)$ in Fig. 2.3. Each plot corresponds to each covariance matrix element for slice $\rho=||(0,\Delta k_2)||$. One can observe a good fit for the zero lag (1.21) and the covariance decay (2.4) that convey information on the multifractal parameters, and slightly larger discrepancies for lags 1 and 2 covered by the model adjustment (2.5). Similar results have been obtained for a wide range of multifractal parameters values, sample sizes and numbers of data components.

Likelihood. Let $\boldsymbol{l}(j,\boldsymbol{k})$ denote the centered vector of the log-leaders at a fixed scale j and position $\boldsymbol{k}$. The vector

$$\boldsymbol{l}_j \triangleq (\boldsymbol{l}(j,\boldsymbol{k}_1)^T,\ldots,\boldsymbol{l}(j,\boldsymbol{k}_{n_j})^T)\in\mathbb{R}^{Rn_j} \tag{2.6}$$

stacks all the vectors $\{\boldsymbol{l}(j,\boldsymbol{k})\}$ at scale j organized in lexicographic order. The likelihood of $\boldsymbol{l}_j$ w.r.t.

$$\mathbf{\Gamma}=(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2) \tag{2.7}$$

is given by

$$p(\boldsymbol{l}_j\mid\mathbf{\Gamma})\propto(\det\mathbf{\Xi}_j)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{l}_j^T\mathbf{\Xi}_j^{-1}\boldsymbol{l}_j\right), \tag{2.8}$$

where $\mathbf{\Xi}_j=\mathbf{\Sigma}_1\otimes\boldsymbol{F}_{1j}+\mathbf{\Sigma}_2\otimes\boldsymbol{F}_{2j}$ is an $Rn_j\times Rn_j$ covariance matrix with $[\boldsymbol{F}_{ij}]_{u,v}=f_i(j,||\boldsymbol{k}_u-\boldsymbol{k}_v||)$, $i=1,2$. For any scale j, the positive definiteness of $\boldsymbol{F}_{ij}$ can be assessed numerically to check that $\mathbf{\Sigma}_i\otimes\boldsymbol{F}_{ij}$ and $\mathbf{\Xi}_j$ are also p.d.

We assume here that $\boldsymbol{l}_j$ and $\boldsymbol{l}_{j'}$ at different scales $j'\neq j$ are independent. Note that the theoretical results in [ADKR03] and our numerical results suggest that the interscale dependence between log-leaders could also be described by a model reminiscent of (2.2), with proper adjustments for

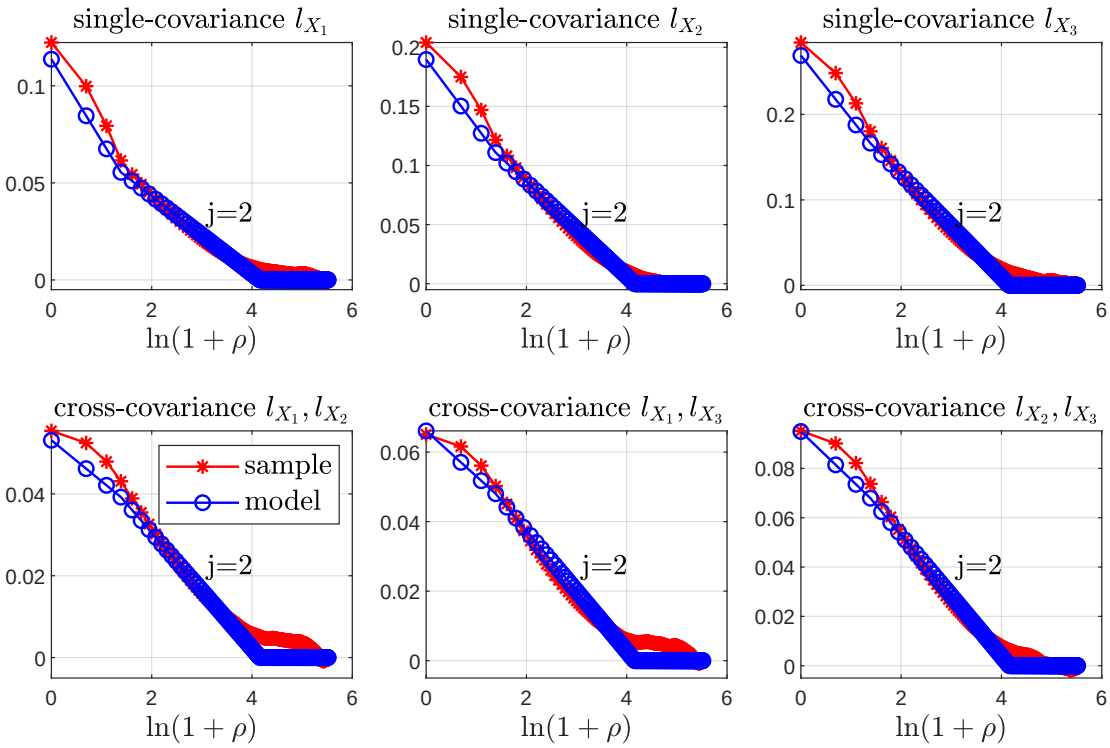

Figure 2.3: Comparison between the proposed model (blue line) and the sample covariance (red line) at scale $j = 2$, averaged over 100 independent copies of an $2^{10} \times 2^{10} \times 3$ synthetic MV-MRW image with $(\lambda_1, \lambda_2, \lambda_3) = (\sqrt{0.02}, \sqrt{0.04}, \sqrt{0.06})$ and $\rho_{\mathrm{mf}}(1,2) = \rho_{\mathrm{mf}}(1,3) = \rho_{\mathrm{mf}}(2,3) = 0.5$. Plots correspond to each covariance matrix element for slice $\rho = ||(0, \Delta k_2)||$.

decimation. This work assumes independence to simplify the proposed model and reduce the computational cost. The results reported in Section 2.4 demonstrate that this assumption is reasonable and leads to state of the art performance.

Together with the above model, the likelihood of the vector $\boldsymbol{l} = (\boldsymbol{l}_{j_1}^T, \ldots, \boldsymbol{l}_{j_2}^T)$ of the log-leaders at scales $j \in [\![j_1, j_2]\!]$ ($[\![n_1, n_2]\!]$ denotes the set of integers ranging from n_1 to n_2) can be written as

$$\begin{aligned} p(\boldsymbol{l} \mid \boldsymbol{\Gamma}) &= \prod_{j=j_1}^{j_2} p(\boldsymbol{l}_j \mid \boldsymbol{\Gamma}), \\ &\propto \prod_{j=j_1}^{j_2} (\det \boldsymbol{\Xi}_j)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{l}_j^T \boldsymbol{\Xi}_j^{-1} \boldsymbol{l}_j\right). \end{aligned} \tag{2.9}$$

2.2.2 Whittle approximation

The likelihood (2.9) is a mixture of Gaussians with large $(Rn_j \times Rn_j)$ full covariance matrices $\boldsymbol{\Xi}_j$ requiring the inversion of these matrices for its numerical evaluation. For large sample sizes, this inversion can become computationally and numerically challenging due to growing condition number. 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 [Whi53] to approximate the time-domain likelihood (2.9) in the frequency domain by a likelihood with a simpler structured covariance matrix as follows

$$p(\boldsymbol{l} \mid \boldsymbol{\Gamma}) \approx \prod_{j=j_1}^{j_2} p_W(\boldsymbol{l}_j \mid \boldsymbol{\Gamma}), \tag{2.10}$$

with

$$p_W(\boldsymbol{l}_j \mid \boldsymbol{\Gamma}) \propto \prod_{\boldsymbol{m} \in I_j} (\det \boldsymbol{R}_{j,\boldsymbol{m}})^{-1} \exp\left(-\boldsymbol{z}_{j,\boldsymbol{m}}^H \boldsymbol{R}_{j,\boldsymbol{m}}^{-1} \boldsymbol{z}_{j,\boldsymbol{m}}\right). \tag{2.11}$$

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

$$\boldsymbol{z}_{j,\boldsymbol{m}} = \frac{1}{n_j^{1/d}} \sum_{\boldsymbol{k} \in [\![1, n_j^{1/d}]\!]^d} \boldsymbol{l}(j,\boldsymbol{k}) \exp\left(-\mathrm{i}\boldsymbol{k}^T \boldsymbol{\omega}_{j,\boldsymbol{m}}\right), \tag{2.12}$$

where $\boldsymbol{m} \in I_j \triangleq [\![\lfloor(-n_j^{1/d} - 1)/2\rfloor, n_j^{1/d} - \lfloor n_j^{1/d}/2 \rfloor]\!]^d \setminus \{\boldsymbol{0}\}$. Moreover, the power spectral matrix $\boldsymbol{R}_{j,\cdot}$ forms a Fourier pair with the covariance matrix $\boldsymbol{S}_{j,\cdot}$ [SOLE17], which can be approximated using a discrete Fourier transform of $f_i(j,\cdot)$, as follows

$$\boldsymbol{R}_{j,\boldsymbol{m}} = \boldsymbol{\Sigma}_1 g_1(\boldsymbol{\omega}_{j,\boldsymbol{m}}) + \boldsymbol{\Sigma}_2 g_2(\boldsymbol{\omega}_{j,\boldsymbol{m}}), \tag{2.13}$$

where

$$g_i(\boldsymbol{\omega}_{j,\boldsymbol{m}}) = \sum_{\boldsymbol{k} \in [\![-n_j^{1/d}, n_j^{1/d}]\!]^d} f_i(j, ||\boldsymbol{k}||) \exp\left(-\mathrm{i}\boldsymbol{k}^T \boldsymbol{\omega}_{j,\boldsymbol{m}}\right),\ i = 1, 2. \tag{2.14}$$

For $i \in \{1, 2\}$ and any scale j, $f_i(j,\cdot)$ is a non-negative even function. Thus, $g_i(\boldsymbol{\omega}_{j,\boldsymbol{m}})$ is real-valued and strictly positive. Since $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are assumed to be p.d., $\boldsymbol{R}_{j,\cdot}$ is a real-valued p.d. matrix for any scale j.

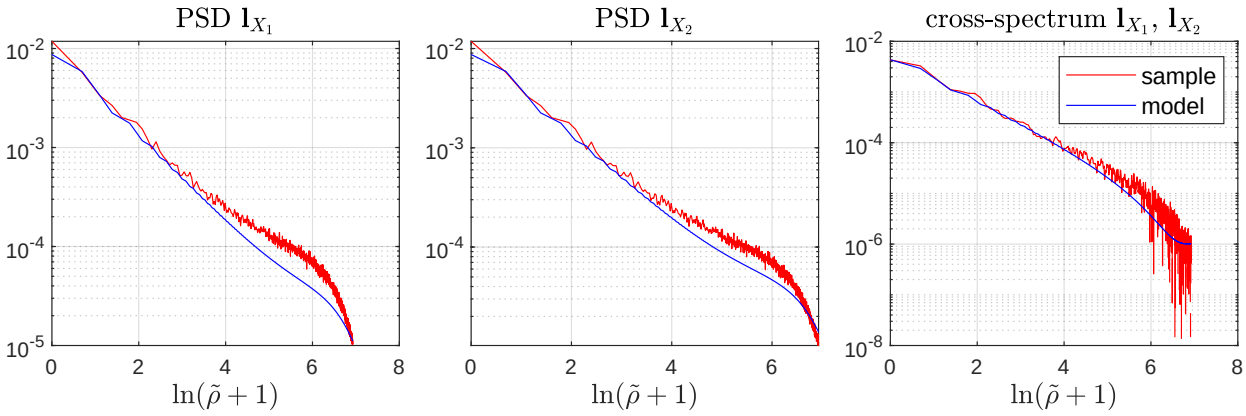

Figure 2.4: Comparison between the sample power spectral matrix (red line), averaged over 100 independent realizations of a 2D MV-MRW for $R = 2$, $N = 2^{15}$, $(\lambda_1, \lambda_2) = (\sqrt{0.04}, \sqrt{0.08})$ and $\rho_{\mathrm{mf}}(1,2) = 0.6$, and the proposed model (blue line) obtained using (2.13). Plots correspond to slice $\tilde{\rho} = ||(0, m_2)||$.

The coefficient $\boldsymbol{z}_{j,\boldsymbol{m}}$ and the power spectral matrix $\boldsymbol{R}_{j,\boldsymbol{m}}$ have a central symmetry property since the log-leaders $\boldsymbol{l}(j,\boldsymbol{k})$ and the functions $f_i(j,\cdot)$ are real-valued. Thus, the product in (2.11) can be taken over the positive half of the total frequency grid.

Fig. 2.4 illustrates the fit between the models used for the power spectral densities (PSDs) and the cross power spectral density (CPSD) (computed using (2.13)) 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 (2.11) is conducted using $\boldsymbol{m} \in I_j^{\dagger} = \{\boldsymbol{m} \in I_j : 0 < \tilde{\rho} \leq \sqrt{\eta}\lfloor n_j^{1/d}/2 \rfloor\}$ with $\tilde{\rho} \triangleq ||\boldsymbol{m}||$. 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 thesis, see Section 2.4.1.

2.2.3 Model in the Fourier domain

Expression (2.11) can be interpreted as a spectral likelihood, see, e.g., [GF14, BR12, Cha97], leading to model the Fourier coefficients $\boldsymbol{z}_{j,\boldsymbol{m}}$ by independent random vectors with a non-degenerated centered circular-symmetric complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{R}_{j,\boldsymbol{m}})$. To simplify notation, we replace the sub-index $\cdot_{j,\boldsymbol{m}}$ of Section 2.2.2 by a single sub-index $\cdot_s$ defined as a one-to-one function of $(j, \boldsymbol{m})$ on the set $\{1, \dots, M\}$, where M is the number of elements of the set $\{(j, \boldsymbol{m}) : j = [\![j_1, j_2]\!] \text{ and } \boldsymbol{m} \in I_j^\dagger\}$. Therefore, the density of the spectral observation vector $\boldsymbol{z} = (\boldsymbol{z}_1^T, \dots, \boldsymbol{z}_M^T) \in \mathbb{C}^{MR}$, can be written as

$$\begin{aligned} p(\boldsymbol{z} \mid \boldsymbol{\Gamma}) &= \prod_{s=1}^{M} p(\boldsymbol{z}_s \mid \boldsymbol{\Gamma}) \\ &\propto \prod_{s=1}^{M} (\det \boldsymbol{R}_s)^{-1} \exp(-\boldsymbol{z}_s^H \boldsymbol{R}_s^{-1} \boldsymbol{z}_s), \end{aligned} \tag{2.15}$$

where $\boldsymbol{R}_s = \boldsymbol{\Sigma}_1 g_{1,s} + \boldsymbol{\Sigma}_2 g_{2,s}$, with $g_{i,s} = g_i(\boldsymbol{\omega}_s)$ and $i = 1, 2$.

2.2.4 Data augmentation

Model (2.15) is simple and cheap to evaluate numerically compared to (2.9). However, its main inconvenience regarding the estimation of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ is that these matrices are additively tied together in $\boldsymbol{R}_s$, 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., [TW87, DM01] for more details) and introduce a complex-valued vector of latent variables $\boldsymbol{u} = (\boldsymbol{u}_1^T, \dots, \boldsymbol{u}_M^T) \in \mathbb{C}^{MR}$ as the hidden mean of the observed data $\boldsymbol{z}$

$$\boldsymbol{z} \mid \boldsymbol{u}, \boldsymbol{\Sigma}_1 \sim \mathcal{CN}(\boldsymbol{u}, \boldsymbol{\Sigma}_1 \otimes \boldsymbol{G}_1), \qquad \boldsymbol{u} \mid \boldsymbol{\Sigma}_2 \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_2 \otimes \boldsymbol{G}_2), \tag{2.16}$$

where $\boldsymbol{G}_i \triangleq \operatorname{diag}(g_{i,1}, \dots, g_{i,M})$, for $i = 1, 2$. This leads to the augmented likelihood

$$\begin{aligned} p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) &= p(\boldsymbol{z} \mid \boldsymbol{u}, \boldsymbol{\Sigma}_1) p(\boldsymbol{u} \mid \boldsymbol{\Sigma}_2), \\ &\propto (\det \boldsymbol{\Sigma}_1)^{-M} \exp\left(-\frac{1}{2} \operatorname{Tr}(\boldsymbol{\Sigma}_1^{-1} \tilde{\boldsymbol{\Phi}}_1)\right) (\det \boldsymbol{\Sigma}_2)^{-M} \exp\left(-\frac{1}{2} \operatorname{Tr}(\boldsymbol{\Sigma}_2^{-1} \tilde{\boldsymbol{\Phi}}_2)\right), \end{aligned} \tag{2.17}$$

with

$$\tilde{\boldsymbol{\Phi}}_1 = 2\sum_{s=1}^{M}(\boldsymbol{z}_s - \boldsymbol{u}_s)(\boldsymbol{z}_s - \boldsymbol{u}_s)^H g_{1,s}^{-1}, \tag{2.18}$$

$$\tilde{\boldsymbol{\Phi}}_2 = 2\sum_{s=1}^{M}\boldsymbol{u}_s\boldsymbol{u}_s^H g_{2,s}^{-1}. \tag{2.19}$$

By construction, the likelihood (2.15) is obtained by marginalizing (2.17) w.r.t. $\boldsymbol{u}$, i.e.,

$$p(\boldsymbol{z} \mid \boldsymbol{\Gamma}) = \int p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) d\boldsymbol{u}. \tag{2.20}$$

The advantage of using (2.17) w.r.t. (2.15), is that, in (2.17), the matrix-valued parameters $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are no longer additively tied but conditionally independent. Several estimators for $\boldsymbol{\Gamma}$ can be defined based on the likelihood (2.17) (cf. Section 2.3 and Chapter 3). In a Bayesian formulation, the shape of (2.17) allows (scaled) inverse Wishart priors to be used as (semi-) conjugate priors for $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, which will simplify the estimation of these matrix-valued parameters using Bayesian inference.

2.3 Bayesian estimation

The matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ of the model introduced above can be estimated using Bayesian estimators. Bayesian inference consists in assigning prior distributions to the unknown model parameters and estimating them using Bayes' theorem. 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., [SL05]). This section presents Bayesian approaches for the estimation of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ for an arbitrary number R of data components.

2.3.1 Bayesian models

a) Likelihood

The proposed Bayesian models are based on the augmented likelihood (2.17), which is the product of two complex Gaussian distributions having $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as $R \times R$ covariance matrices.

b) Priors

Inverse Wishart. The natural *conjugate prior* for $\mathbf{\Sigma}_i$ is the inverse Wishart (IW) prior [BMM00], i.e., $\mathbf{\Sigma}_i \sim \mathcal{IW}(\nu_i, \mathbf{\Lambda}_i)$. It has the probability density function (pdf)

$$p(\mathbf{\Sigma}_i|\nu_i, \mathbf{\Lambda}_i) \propto (\det \mathbf{\Sigma}_i)^{-\frac{1}{2}(\nu_i+R+1)} \exp\left(-\frac{1}{2}\operatorname{Tr}(\mathbf{\Lambda}_i\mathbf{\Sigma}_i^{-1})\right), \tag{2.21}$$

where $\nu_i \in \mathbb{R}$ is the degree of freedom and $\mathbf{\Lambda}_i$ is an $R \times R$ real-valued p.d. scale matrix. The IW prior is proper for $\nu_i > R-1$ and its mean $\mathbb{E}[\mathbf{\Sigma}_i] = \mathbf{\Lambda}_i/(\nu_i - R - 1)$ only exists if $\nu_i > R+1$. The conjugacy property of this prior facilitates the parameter inference, since the corresponding conditional distributions have closed-form, which simplifies the use of MCMC algorithms to numerically evaluate the Bayesian estimators. However, the variance parameters in $\mathbf{\Sigma}_i$ are only controlled by ν_i and thus do not allows us to incorporate accurate 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 [Gel06]. Finally, the IW prior imposes a dependency between the correlations and the variances which is not a desired property (larger variances are associated with absolute values of the correlations near one while small variances are associated with correlations near zero [TGM+11]).

Scaled inverse Wishart. An alternative to the IW prior is the scaled inverse Wishart (SIW) prior proposed in [OZ08]. The idea is to use a decomposition of the matrix $\mathbf{\Sigma}_i$ to allow the priors of the standard deviations and correlation coefficients to be defined semi-separately, providing more flexibility than the IW prior. In essence, $\mathbf{\Sigma}_i$ is modeled using two independent random matrices $\boldsymbol{Q}_i$ and $\mathbf{\Delta}_i$ defined below as

$$\mathbf{\Sigma}_i \triangleq \mathbf{\Delta}_i\boldsymbol{Q}_i\mathbf{\Delta}_i, \tag{2.22}$$

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

$$p(\delta_{ir}|\beta_{ir}, \alpha_{ir}^2) \propto \frac{1}{\delta_{ir}\alpha_{ir}} \exp\left(-\frac{(\ln \delta_{ir} - \beta_{ir})^2}{2\alpha_{ir}^2}\right), \tag{2.23}$$

for $i \in \{1, 2\}$ and $r \in \{1, \ldots, R\}$. Yet, the induced flexibility comes at the price of having to deal with non-closed-form expressions for the conditional distributions, which makes parameter estimation

more complicated.

c) Posterior distribution

The posterior distribution associated with the proposed Bayesian models for $\boldsymbol{\Gamma}$ and the latent vector $\boldsymbol{u}$ can be computed from Bayes' theorem

$$\begin{aligned} p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z}) &\propto p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) p(\boldsymbol{\Gamma}) \\ &\propto p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) p(\boldsymbol{\Sigma}_1) p(\boldsymbol{\Sigma}_2). \end{aligned} \tag{2.24}$$

Finally, since the vector of latent variables $\boldsymbol{u}$ is not interesting for MFA purposes and is here introduced for computational convenience only, we consider the marginal minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators respectively defined by

$$\hat{\boldsymbol{\Gamma}}^{\text{MMSE}} = \mathbb{E}[\boldsymbol{\Gamma} \mid \boldsymbol{z}], \tag{2.25}$$

and

$$\hat{\boldsymbol{\Gamma}}^{\text{MAP}} = \underset{\boldsymbol{\Gamma}}{\operatorname{argmax}}\, p(\boldsymbol{\Gamma} \mid \boldsymbol{z}), \tag{2.26}$$

where the expectation and maximization are taken w.r.t. the marginal posterior distribution

$$p(\boldsymbol{\Gamma} \mid \boldsymbol{z}) = \int p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z}) d\boldsymbol{u} \propto p(\boldsymbol{z} \mid \boldsymbol{\Gamma}) p(\boldsymbol{\Gamma}). \tag{2.27}$$

Those estimators are difficult to be expressed using simple closed-form expressions. As an alternative, the computation of the MAP estimator using an Expectation-Maximization (EM) algorithm will be studied in Chapter 3. In what follows, we explain how the MMSE estimator can be computed via MCMC algorithms [RC05].

2.3.2 Estimation algorithms

We consider Gibbs samplers to approximate the MMSE estimator (2.25), consisting of generating a large number N_{mc} of samples $\{\boldsymbol{\Sigma}_1^{(\lambda)}, \boldsymbol{\Sigma}_2^{(\lambda)}\}_{\lambda=1}^{N_{\text{mc}}}$ according to the conditional distributions of (2.24) when using an IW prior for $\boldsymbol{\Sigma}_i$ or its scaled version. These samples will then be used to approximate the MMSE estimator by computing the average of p.d. matrices. We will study two different ways to compute the matrix average.

a) Two-stage Gibbs sampler using an IW prior

Assuming $\boldsymbol{\Sigma}_i \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$ for all $i \in \{1, 2\}$, the posterior (2.24) can be expressed as

$$p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z}) \propto (\det \boldsymbol{\Sigma}_1)^{-\frac{1}{2}(2M+\nu_1+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}\left(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Lambda}_1 + \tilde{\boldsymbol{\Phi}}_1)\right)\right) (\det \boldsymbol{\Sigma}_2)^{-\frac{1}{2}(2M+\nu_2+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}\left(\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Lambda}_2 + \tilde{\boldsymbol{\Phi}}_2)\right)\right). \tag{2.28}$$

Thus, the conditional distribution of $\boldsymbol{\Sigma}_i \mid \boldsymbol{z}, \boldsymbol{u}$ is

$$p(\boldsymbol{\Sigma}_i \mid \boldsymbol{z}, \boldsymbol{u}) \propto (\det \boldsymbol{\Sigma}_i)^{-\frac{1}{2}(2M+\nu_i+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Lambda}_i + \tilde{\boldsymbol{\Phi}}_i))\right), \tag{2.29}$$

which is the following IW distribution:

$$\boldsymbol{\Sigma}_i \mid \boldsymbol{z}, \boldsymbol{u} \sim \mathcal{IW}(\nu_i + 2M, \boldsymbol{\Lambda}_i + \tilde{\boldsymbol{\Phi}}_i). \tag{2.30}$$

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

$$\boldsymbol{u} \mid \boldsymbol{z}, \boldsymbol{\Gamma} \sim \mathcal{CN}(\boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}}), \tag{2.31}$$

where $\tilde{\boldsymbol{\Sigma}}$ is a block diagonal matrix whose sth block is defined as

$$\tilde{\boldsymbol{\Sigma}}_s = \left[(\boldsymbol{g}_{1,s}\boldsymbol{\Sigma}_1)^{-1} + (g_{2,s}\boldsymbol{\Sigma}_2)^{-1}\right]^{-1}, \tag{2.32}$$

and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_M^T) \in \mathbb{C}^{MR}$ with

$$\boldsymbol{\mu}_s = \tilde{\boldsymbol{\Sigma}}_s (g_{1,s}\boldsymbol{\Sigma}_1)^{-1} \boldsymbol{z}_s. \tag{2.33}$$

Using the conditional distributions (2.30) and (2.31), Algorithm 1 summarizes the two steps of the proposed Gibbs sampler used to generate samples according to the posterior of interest (2.24) when using IW priors.

b) Metropolis-within-Gibbs sampler using SIW prior

The SIW prior used for $\boldsymbol{\Sigma}_i$ is defined by independent priors for the scale parameters $\{\delta_{ir}\}_{r=1}^R$ and the matrix parameter $\boldsymbol{Q}_i$ [OZ08]:

$$p(\boldsymbol{\Sigma}_i) = p(\boldsymbol{Q}_i) \prod_{r=1}^{R} p(\delta_{ir}). \tag{2.34}$$

Algorithm 1 Two-stage Gibbs sampler using IW prior

Initialization:
Set $\boldsymbol{u} = \mathbf{0}$ and draw $\boldsymbol{\Sigma}_i^{(0)} \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$ for $i = 1, 2$
MCMC iterations:
for $\lambda = 1 : N_{\text{mc}}$ **do**
    Step 1: Sample parameters $\boldsymbol{\Gamma}$
    for $i = 1 : 2$ **do**
        Draw $\boldsymbol{\Sigma}_i^{(\lambda)}$ from the IW distribution (2.30)
    end for
    Step 2: Sample latent vector $\boldsymbol{u}$
    Draw $\boldsymbol{u}^{(\lambda)}$ according to the CN distribution (2.31)
end for
return $\{\boldsymbol{\Sigma}_1^{(\lambda)}, \boldsymbol{\Sigma}_2^{(\lambda)}\}_{\lambda=1}^{N_{\text{mc}}}$

Thus, the posterior distribution (2.24) can be rewritten after substituting both $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ by their decompositions (2.22) as

$$\begin{aligned} p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z}) &\propto p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) \prod_{i=1}^{2} p(\boldsymbol{Q}_i) \prod_{r=1}^{R} p(\delta_{ir}) \\ &\propto (\det \boldsymbol{\Delta}_1)^{-2M} (\det \boldsymbol{Q}_1)^{-\frac{1}{2}(2M+\nu_1+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}\left(\boldsymbol{Q}_1^{-1}(\boldsymbol{\Lambda}_1 + \boldsymbol{\Delta}_1^{-1}\tilde{\boldsymbol{\Phi}}_1\boldsymbol{\Delta}_1^{-1})\right)\right) \\ &\quad (\det \boldsymbol{\Delta}_2)^{-2M} (\det \boldsymbol{Q}_2)^{-\frac{1}{2}(2M+\nu_2+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}\left(\boldsymbol{Q}_2^{-1}(\boldsymbol{\Lambda}_2 + \boldsymbol{\Delta}_2^{-1}\tilde{\boldsymbol{\Phi}}_2\boldsymbol{\Delta}_2^{-1})\right)\right) \\ &\quad \prod_{i=1}^{2}\prod_{r=1}^{R} \frac{1}{\delta_{ir}\alpha_{ir}} \exp\left(\frac{-(\ln \delta_{ir} - \beta_{ir})^2}{2\alpha_{ir}^2}\right). \end{aligned} \tag{2.35}$$

For $i = \{1, 2\}$, assuming $\boldsymbol{Q}_i \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$, the following result is obtained

$$p(\boldsymbol{Q}_i \mid \boldsymbol{\Delta}_i, \boldsymbol{z}, \boldsymbol{u}) \propto (\det \boldsymbol{Q}_i)^{-\frac{1}{2}(2M+\nu_i+R+1)} \exp\left(-\frac{1}{2} \operatorname{Tr}(\boldsymbol{Q}_i^{-1}(\boldsymbol{\Lambda}_i + \boldsymbol{\Delta}_i^{-1}\tilde{\boldsymbol{\Phi}}_i\boldsymbol{\Delta}_i^{-1}))\right), \tag{2.36}$$

which means that the conditional distribution of $\boldsymbol{Q}_i \mid \boldsymbol{\Delta}_i$, $\boldsymbol{z}$, $\boldsymbol{u}$ is the IW distribution:

$$\boldsymbol{Q}_i \mid \boldsymbol{\Delta}_i, \boldsymbol{z}, \boldsymbol{u} \sim \mathcal{IW}(\nu_i + 2M, \boldsymbol{\Lambda}_i + \boldsymbol{\Delta}_i^{-1}\tilde{\boldsymbol{\Phi}}_i\boldsymbol{\Delta}_i^{-1}). \tag{2.37}$$

Similarly, after some manipulations of (2.24), the log-conditional posterior distribution of δ_{ir} (rth diagonal element of $\boldsymbol{\Delta}_i$) can be determined:

$$\ln p(\delta_{ir} |\, \boldsymbol{Q}_i, \{\delta_{ir'}\}_{r'=1,r'\neq r}^{R}, \boldsymbol{z}, \boldsymbol{u}) = -(2M+1)\ln\delta_{ir} - (\ln\delta_{ir} - \beta_{ir})^2(2\alpha_{ir}^2)^{-1} - [\boldsymbol{Q}_i^{-1}]_{r,r}[\tilde{\boldsymbol{\Phi}}_i]_{r,r}(2\delta_{ir}^2)^{-1} - \delta_{ir}^{-1}\sum_{r'\neq r}\delta_{ir'}^{-1}[\boldsymbol{Q}_i^{-1}]_{r,r'}[\tilde{\boldsymbol{\Phi}}_i]_{r,r'} + \text{constant}, \tag{2.38}$$

which is not a standard distribution. To sample according to (2.38) 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 [RC05]. It is interesting to mention here that the Metropolis steps do not significantly reduce the convergence speed of the sampler since they only apply to R parameters. The draws of $\boldsymbol{Q}_i$ and $\{\delta_{ir}\}_{r=1}^{R}$ are finally used to generate samples of $\boldsymbol{\Sigma}_i$ using (2.22).

Finally, the conditional distribution of $\boldsymbol{u} \mid \boldsymbol{\Gamma}, \boldsymbol{z}$ is the same as in (2.31). Algorithm 2 summarizes the different steps of the proposed sampling method.

c) Approximation of the MMSE estimator

The chosen prior distributions for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ guarantee that the sampled matrices are p.d. 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 real-valued p.d. 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. The *average* over the space of the real-valued p.d. matrices can be computed using the arithmetic mean associated with the Euclidean metric,

$$\widehat{\boldsymbol{\Sigma}}_i^{\text{A}} = \frac{1}{N_{\text{mc}} - N_{\text{bi}}} \sum_{\lambda=N_{\text{bi}}+1}^{N_{\text{mc}}} \boldsymbol{\Sigma}_i^{(\lambda)}, \tag{2.39}$$

which is by construction also p.d.

Alternatively, we have proposed to use the geometric mean (also called Karcher or Riemannian mean) associated with a Riemannian metric, see, e.g., [Moa05] for details. The Karcher mean $\widehat{\boldsymbol{\Sigma}}_i^{\text{K}}$ is

Algorithm 2 Metropolis-within-Gibbs sampler using SIW prior

Initialization:

Set $\boldsymbol{u} = \mathbf{0}$ and draw $\boldsymbol{Q}_i^{(0)} \sim \mathcal{IW}(\nu_i, \boldsymbol{\Lambda}_i)$ and $\delta_{ir}^{(0)} \sim \mathcal{LN}(\beta_{ir}, \alpha_{ir}^2)$ for all $i = 1, 2$ and $r = 1, \ldots, R$

MCMC iterations:

for $\lambda = 1 : N_{\text{mc}}$ **do**

  Step 1: Sample parameters $\boldsymbol{\Gamma}$

  for $i = 1 : 2$ **do**

    Draw $\boldsymbol{Q}_i^{(\lambda)}$ from the IW distribution (2.37)

    Set $\delta_{ir}^{\circ} = \delta_{ir}^{(\lambda-1)}$ for all $r = 1, \ldots, R$

    for $r = 1 : R$ **do**

      Draw $\delta_{ir}^{\star} \sim \mathcal{N}(\delta_{ir}^{\circ}, \sigma_{\delta_{ir}}^2)$ and $\mu \sim \mathcal{U}_{[0,1]}$

      Compute $a = \min \left\{ 1, \dfrac{p(\delta_{ir}^{\star} \mid \{\delta_{ir}^{\circ}\}_{r=1, r \neq r'}^{R}, \beta_{ir}, \sigma_{ir}^2)}{p(\delta_{ir}^{\circ} \mid \{\delta_{ir}^{\circ}\}_{r=1, r \neq r'}^{R}, \beta_{ir}, \sigma_{ir}^2)} \right\}$ using (2.38)

      Set $\delta_{ir}^{(\lambda)} = \begin{cases} \delta_{ir}^{\star} & \text{if } \mu < a \\ \delta_{ir}^{\circ} & \text{otherwise} \end{cases}$

    end for

    Set $\boldsymbol{\Delta}_i^{(\lambda)} = \text{diag}(\delta_{i1}^{(\lambda)}, \ldots, \delta_{iR}^{(\lambda)})$

    Compute $\boldsymbol{\Sigma}_i^{(\lambda)} = \boldsymbol{\Delta}_i^{(\lambda)} \boldsymbol{Q}_i^{(\lambda)} \boldsymbol{\Delta}_i^{(\lambda)}$

  end for

  Step 2: Sample latent vector $\boldsymbol{u}$

  Draw $\boldsymbol{u}^{(\lambda)}$ according to the CN distribution (2.31)

end for

return $\{\boldsymbol{\Sigma}_1^{(\lambda)}, \boldsymbol{\Sigma}_2^{(\lambda)}\}_{\lambda=1}^{N_{\text{mc}}}$

defined as the unique p.d. symmetric matrix-valued solution to the nonlinear matrix equation

$$\sum_{\lambda=N_{\mathrm{bi}}+1}^{N_{\mathrm{mc}}} \ln\left((\boldsymbol{\Sigma}_i^{(\lambda)})^{-1}\boldsymbol{X}\right) = \mathbf{0}. \tag{2.40}$$

This way of building the estimator takes advantage of the geometry of the problem, i.e., that the estimators of $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ should be p.d. matrices. Unfortunately, due to the noncommutative nature of the matrix multiplication, in general, it is not possible to obtain the geometric mean in closed-form for more than two matrices [Moa05]. As an alternative, (2.40) can be solved using the iterative algorithm proposed in [BI13]. Unlike the arithmetic mean, the Karcher mean has the interesting property that it commutes with the matrix inverse [Moa05].

Note that the estimation of the multifractal correlation ρ_{mf} conducted by replacing the c_2 coefficients in (1.23) by their estimates as in (1.28), is equivalent to compute the correlation coefficient of the covariance matrix $\boldsymbol{\Sigma}_1$, which ensures that $-1 \leq \rho_{\mathrm{mf}} \leq 1$.

Below, we denote by $\mathrm{SIW}_{(\cdot)}$ and $\mathrm{IW}_{(\cdot)}$ the MMSE estimators resulting from using the IW (2.30-2.31) or SIW priors (2.31-2.37-2.38), and use the sub-indexes $(\cdot)_A$ and $(\cdot)_K$ for the arithmetic and Karcher means.

2.4 Numerical experiments

We have introduced in the previous sections a Bayesian approach for the estimation of the second-order log-cumulants c_2 (auto- and cross-multifractality parameters) and the multifractal correlation ρ_{mf} relying on a multivariate parametric statistical model for the log-leaders of multivariate signals and images. In this section, we numerically investigate the performance of the $\mathrm{IW_A}$, $\mathrm{IW_K}$, $\mathrm{SIW_A}$, $\mathrm{SIW_K}$ estimators using Monte Carlo simulations for the canonical multifractal process as defined in Section 1.4.1, for different sample sizes and a large range of values of the multifractal parameters. The different Bayesian estimators are compared to the current standard and benchmark estimators (defined in Section 1.5.2, cf., (1.27) and (1.28)) relying on linear regressions. Those are generically referred to as WLR estimators in what follows.

2.4.1 General setting

Wavelet transform. Wavelet analysis is conducted with a Daubechies' least asymmetric wavelet [Dau88], a common choice for scaling analysis, with $N_\psi = 3$ vanishing moments and $\gamma = 0$ in (1.24) and (1.25).

Analysis scales. One commonly discards the finest scale $j = 1$ to remove corruption by improper initialization of the discrete wavelet transform [VTA00], and the last scales because they contain a statistically insignificant number of wavelet coefficients. Unless specified otherwise, the range of scales have been adjusted as $j_1 = 2$ to $j_2 = \log_2 N - 4$, see, e.g., [WAJ07, WRJA09].

Prior specification. The hyperparameters of the prior distributions are set to $\nu_i = R + 2$, $\mathbf{\Lambda}_i = \mathbb{I}_R$ ($R \times R$ identity matrix), $(\beta_{ir}, \alpha_{ir}^2) = (0.1, 1)$, for $i \in \{1, 2\}$ and $r \in \{1, \ldots, R\}$, see, e.g., [OZ08, ANS14]. As an illustrative example for $R = 2$, Fig. 2.5 (left plot) shows the empirical cumulative distribution function (cdf) associated with 2000 samples of the multifractal parameter $-c_2(1, 1)$ (the first diagonal element of $\mathbf{\Sigma}_1$) when using the above setting. We can observe that these choices lead to a reasonable amount of prior probability mass ($\approx 1/4 - 1/3$) assigned to a conservatively large range for multifractality $-0.25 \lesssim c_2(r, r) \leq 0$ (see Fig. 2.5 (left plot)), yet the priors are sufficiently non-informative so as not to bias our performance analysis. Concerning the multifractal correlation parameter ρ_{mf}, our choice of priors and parameter values induce a flat, unimodal distribution on the interval $[-1, 1]$, as can be observed in Fig. 2.5 (right plot).

Gibbs sampler parameters. The parameters for the Gibbs samplers are set to $N_{\mathrm{mc}} = 2000$ for the length of the Markov chains and $N_{\mathrm{bi}} = 1000$ for the number of samples that are discarded in the burn-in period. These values were found to be suitable to ensure the convergence of the proposed estimation algorithms and small variance for the approximation of the estimators, see, e.g., [RC05].

Performance assessment. The four Bayesian estimators (combining IW/scaled IW priors and arithmetic/Karcher means respectively) are numerically verified for 100 independent realizations of the synthetic multivariate multifractal processes described in Section 1.4.1 for different sample sizes, numbers of data components and a large range of values of multifractal parameters. We compare the estimates $\hat{\theta}$ of $\theta \in \{\{-c_2(r, r')\}_{r,r'=1}^R, \{\rho_{\mathrm{mf}}(r, r')\}_{r,r'=1, r\neq r'}^R\}$ obtained by Bayesian estimation to those obtained using the WLR estimator. The performance is quantified using the sample mean, the

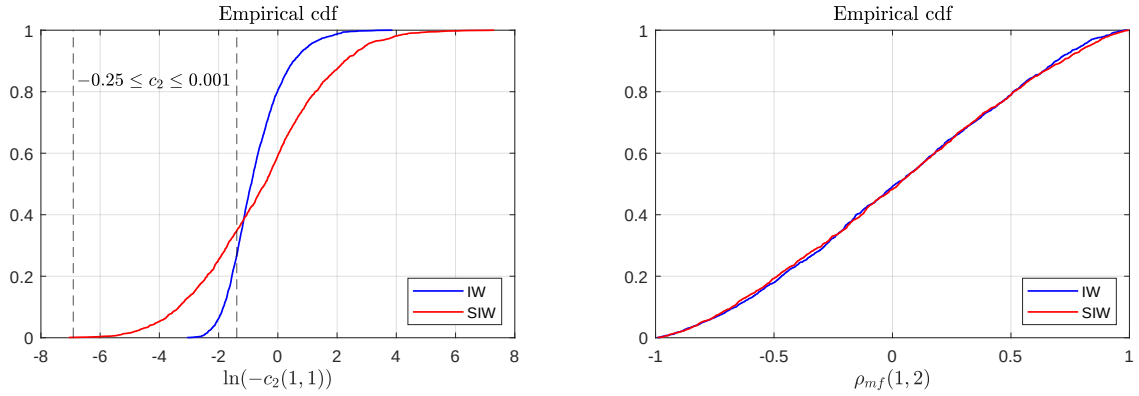

Figure 2.5: Empirical cumulative distribution function associated with 2000 samples of the multifractal parameter $-c_2(1,1)$ (left) and the multifractal correlation parameter $\rho_{\mathrm{mf}}(1,2)$ (right) as the first diagonal element and the correlation coefficient of $\mathbf{\Sigma}_1$, which is distributed according an IW or an SIW distribution.

bias, the sample standard deviation (STD) and the root mean square error (RMSE) of the estimates averaged across realizations

$$m(\hat{\theta}) = \widehat{\mathbb{E}}[\hat{\theta}],\ b(\hat{\theta}) = m(\hat{\theta}) - \theta,\ s(\hat{\theta}) = \sqrt{\widehat{\mathrm{Var}}[\hat{\theta}]} \text{ and } r(\hat{\theta}) = \sqrt{b(\hat{\theta})^2 + s(\hat{\theta})^2}, \tag{2.41}$$

where $\widehat{\mathbb{E}}[]$ and $\widehat{\mathrm{Var}}[]$ are the sample mean and the sample variance.

Bandwidth parameter. The values of η for the high-frequency cutoff, mentioned in Section 2.2.2, are obtained using cross validation. Specifically, we studied the global RMSE behavior of the estimates using numerical simulations for a large range of multifractal parameter values. Fig. 2.6 shows illustrative examples using 100 independent copies of bivariate synthetic multifractal signals of size $2^{10} \times 2$ (left plot) and bivariate synthetic multifractal images of size $2^9 \times 2^9 \times 2$ (right plot), with $\rho_{\mathrm{mf}}(1,2) = 0.5$, $c_2(1,1) = c_2(1,2) = -0.02$ and $c_2(2,2) = -0.08$. One can observe that the range of values of η for which the RMSE is the smallest are different for times series and for images ($0.75 \leq \eta \leq 1$ (1D) and $\eta \leq 0.25$ (2D), respectively). However, the estimation is robust w.r.t. the precise choice of η within these ranges.

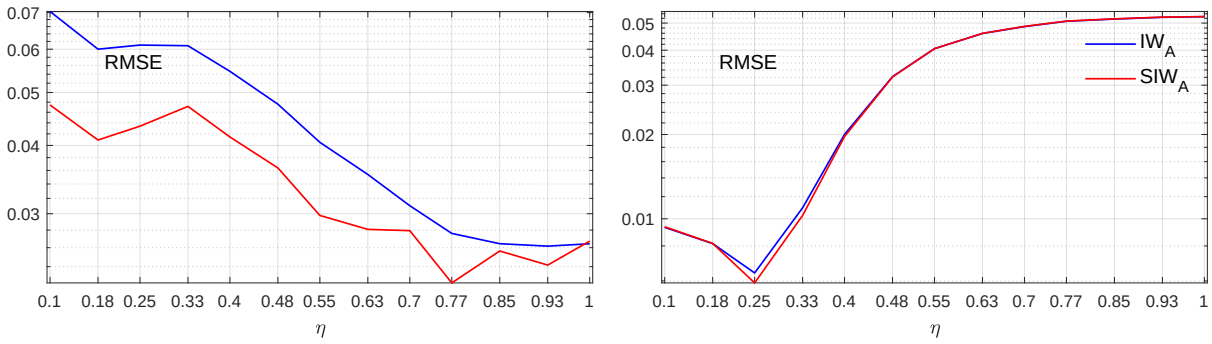

Figure 2.6: RMSE performance of the estimates of the multifractal parameters $c_2(1,1), c_2(1,2), c_2(2,2)$ obtained using $\mathrm{IW_A}$ and $\mathrm{SIW_A}$ algorithms for $\eta \in \{0.1, 0.18, 0.25, 0.33, \ldots, 0.7, 0.78, 0.85, 0.93, 1\}$. These estimates were obtained from 100 independent $2^{10} \times 2$ synthetic multifractal signals (left plot) and $2^{10} \times 2^{10} \times 2$ synthetic multifractal images (right plot), with $\rho_{\mathrm{mf}}(1,2) = 0.5$, $c_2(1,1) = c_2(1,2) = -0.02$ and $c_2(2,2) = -0.08$.

2.4.2 Estimation performance

Our experiments aim at evaluating the estimation performance of the proposed algorithms depending on the values of the multifractal parameters, sample sizes and numbers of data components. We conducted a detailed performance analysis for the bivariate case ($R = 2$), and a study for $R \in \{1, ..., 10\}$. The parameters of the MV-MRW process for $R = 2$ 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 $\rho_{\mathrm{mf}}(1,2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. 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.

Figures 2.7, 2.8, 2.9 and 2.10 summarize the estimation performance of WLR, $\mathrm{SIW_A}$ and $\mathrm{IW_A}$ estimators for 1D (signals) and 2D (images). We further compare the arithmetic mean and Karcher mean for a small sample size where their performance was found to be interesting.

a) Linear regression vs. Bayesian estimation

We observe that while bias is in general smaller for WLR than for Bayesian estimators, the latter produce smaller STD for all multifractality and multifractal correlation parameters. Specifically, WLR has $2-7$ times larger STD than the Bayesian estimators for c_2 parameters and $5-27$ times

larger STD for ρ_{mf}. Consequently, the RMSE values of the Bayesian estimators are smaller than those of WLR for c_2 and ρ_{mf}, respectively. 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).

b) Comparison of priors

Figures 2.7, 2.8, 2.9 and 2.10 show that the use of the scaled inverse Wishart prior globally leads to smaller bias, but larger STD than the inverse Wishart prior. This was 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 $\mathrm{SIW_A}$. In particular, $\mathrm{SIW_A}$ always yields better estimation performance than $\mathrm{IW_A}$ for the multifractal correlation parameter $\rho_{\mathrm{mf}}(1,2)$, and for the univariate parameters $c_2(1,1)$ and $c_2(2,2)$ when $-c_2(2,2) < 0.05$, for which $\mathrm{IW_A}$ yields RMSE values up to twice as large as those for $\mathrm{SIW_A}$.

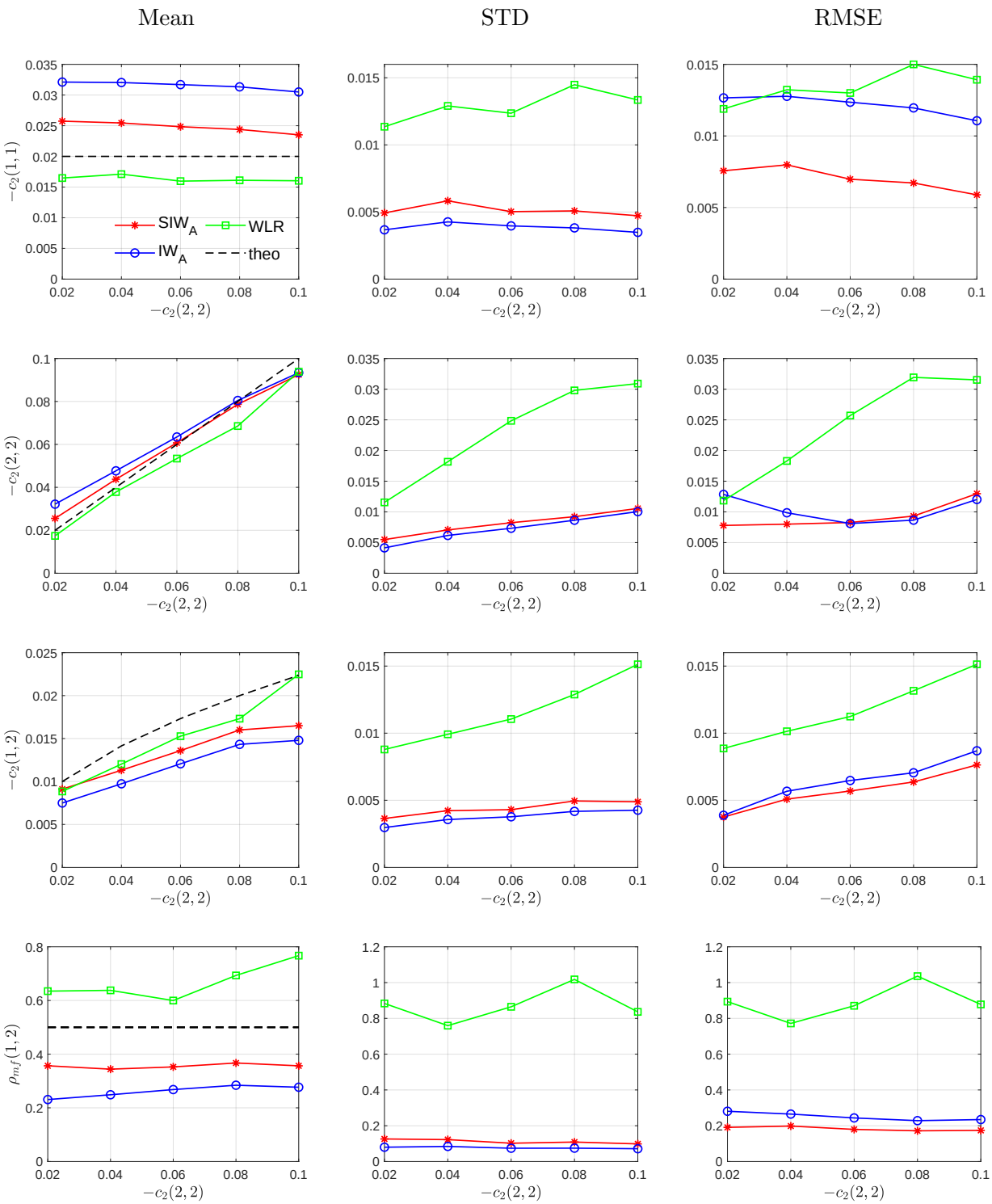

Figure 2.7: 1D MV-MRW estimation performance for $R = 2$, $N = 2^{12}$, $\rho_{\text{mf}}(1,2) = 0.5$ and $-c_2(2,2) \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$

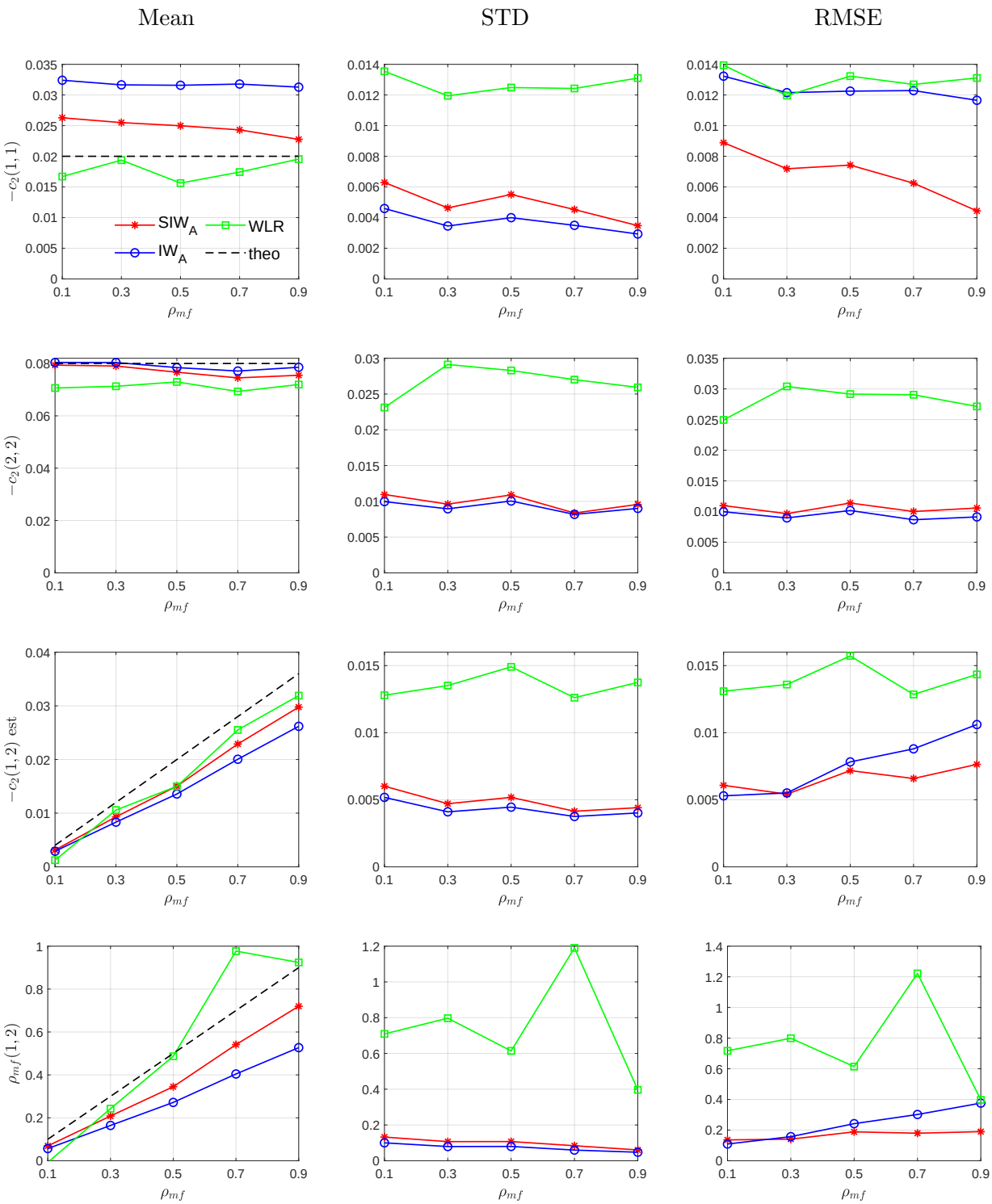

Figure 2.8: 1D MV-MRW estimation performance for $R = 2$, $N = 2^{12}$, $\rho_{\text{mf}}(1,2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$

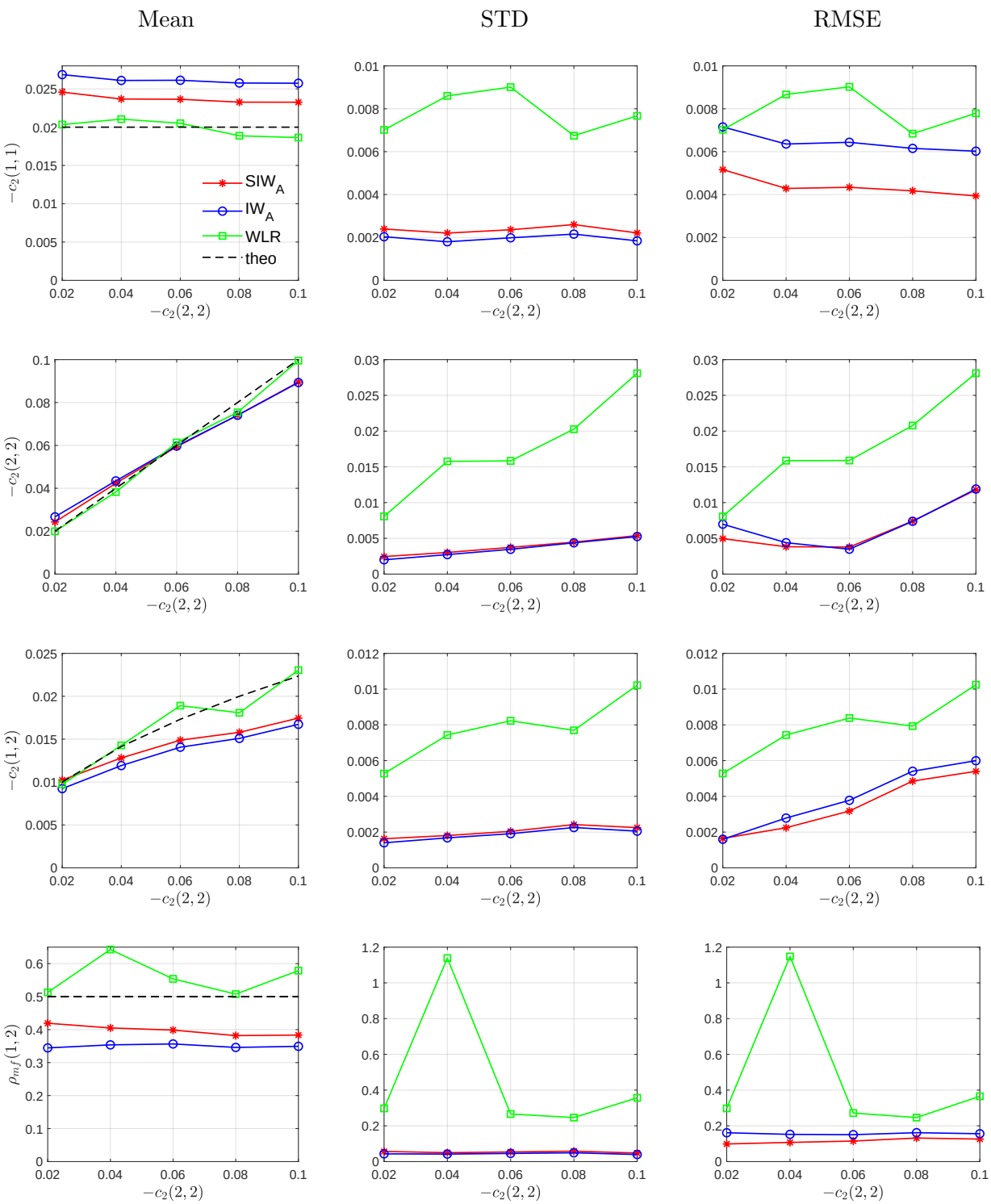

Figure 2.9: 2D MV-MRW estimation performance for $R = 2$, $N = 2^9$, $\rho_{\text{mf}}(1,2) = 0.5$ and $-c_2(2,2) \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$

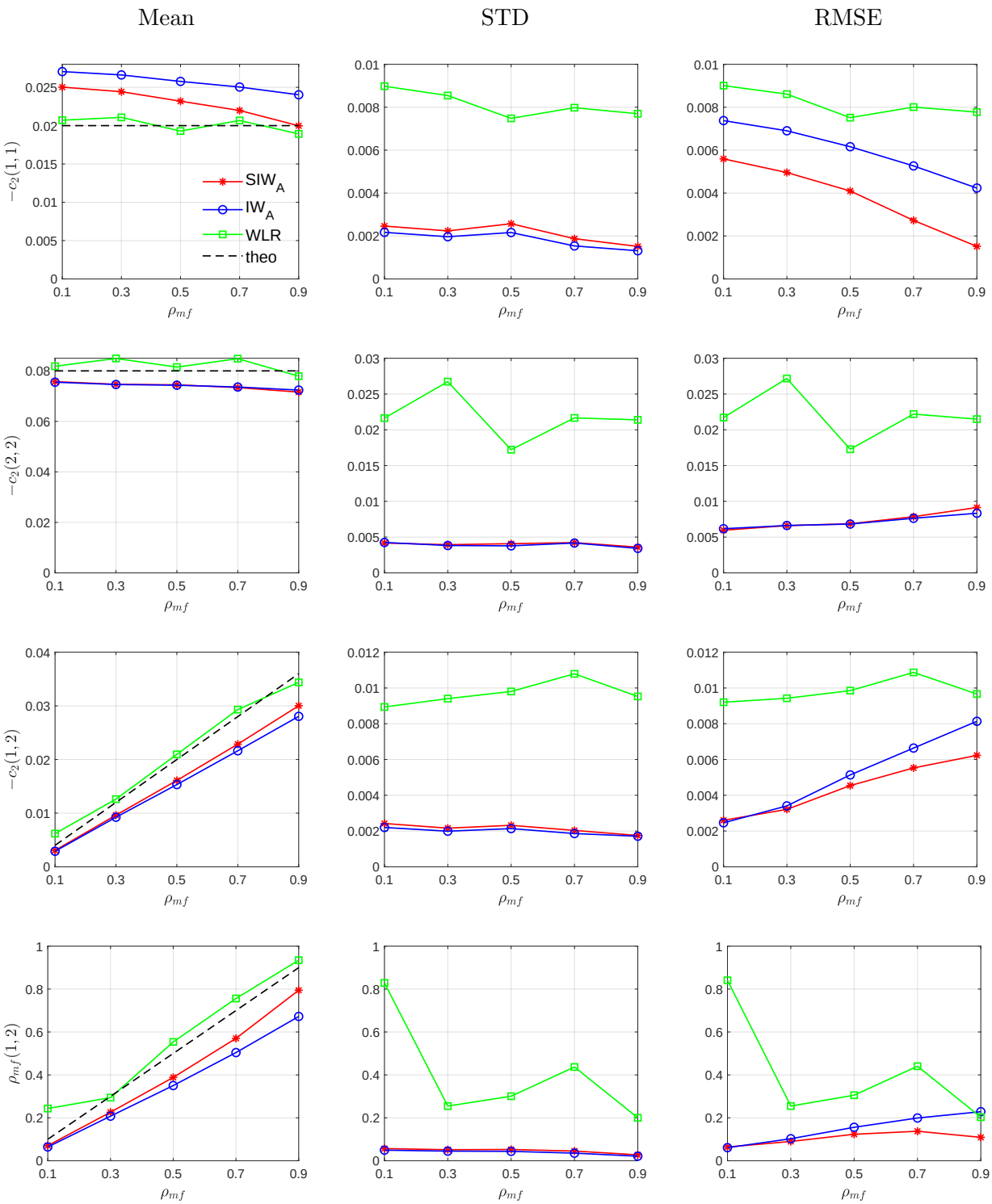

Figure 2.10: 2D MV-MRW estimation performance for $R = 2$, $N = 2^9$, $\rho_{\text{mf}}(1,2) \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$

c) Estimation performance vs. sample sizes

Fig. 2.11 summarizes the RMSE performance of WLR, $\mathrm{SIW}_{\mathrm{A}}$, $\mathrm{SIW}_{\mathrm{K}}$, IW_{A} and IW_{K} estimators for synthetic multifractal $N \times N \times 2$ images as a function of sample sizes $N \in \{2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}\}$. The coarser scale is set as $j_2 \in \{3, 4, 5, 6, 7, 8\}$ and the other parameters are set as in the above sections. For all estimators, the RMSE decreases when the sample size N increases, as expected. However, the RMSE values are significantly smaller for Bayesian estimators when compared to WLR. The latter performs worse when the sample size is small ($N < 2^8$), explained by its large estimate variances. As above, $\mathrm{SIW}_{(\cdot)}$ leads to smaller RMSE values than $\mathrm{IW}_{(\cdot)}$, and significantly so for multifractality parameters $c_2(2,2)$ and $c_2(1,1)$. For large sample sizes $N \geq 2^{10}$, all estimators tend to have a similar behavior in terms of RMSE. Similar results and conclusions are obtained for 1D time series.

d) Comparison of averages

Fig. 2.11 shows that the estimators using Kacher mean and arithmetic mean lead to similar RMSE performance, except for very small sample image size ($2^6 \times 2^6$). In particular, the estimator using the IW prior and Kacher mean shows a slightly improved estimation performance compared to its counterpart using the classical arithmetic mean. This is not the case for the SIW prior.

We further compare $\mathrm{SIW}_{\mathrm{A}}$, $\mathrm{SIW}_{\mathrm{K}}$, IW_{A} and IW_{K} estimation performance for 1D bivariate MRW processes for small sample size $N = 2^6$ and for several multifractality settings, i.e., $-c_2(2,2) \in \{0.02, 0.04, \ldots, 0.1\}$. Results are summarized in Table 2.1 and lead to the following conclusions.

The estimators $\mathrm{SIW}_{(\cdot)}$ have significantly reduced BIAS for $c_2(1,1)$ and $c_2(2,2)$, and similar BIAS for $c_2(1,2)$ when compared to $\mathrm{IW}_{(\cdot)}$. This conclusion is valid for all levels of multifractality for $c_2(2,2)$. Moreover, the STD values are smaller for $\mathrm{SIW}_{(\cdot)}$ than for $\mathrm{IW}_{(\cdot)}$, leading overall to significantly reduced RMSE for $\mathrm{SIW}_{(\cdot)}$ in all cases. As far as the matrix averages are concerned, smaller RMSE values are consistently obtained when the Karcher mean is used to approximate the MMSE estimator.

Conversely to the 2D scenario, for small sample sizes, 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.

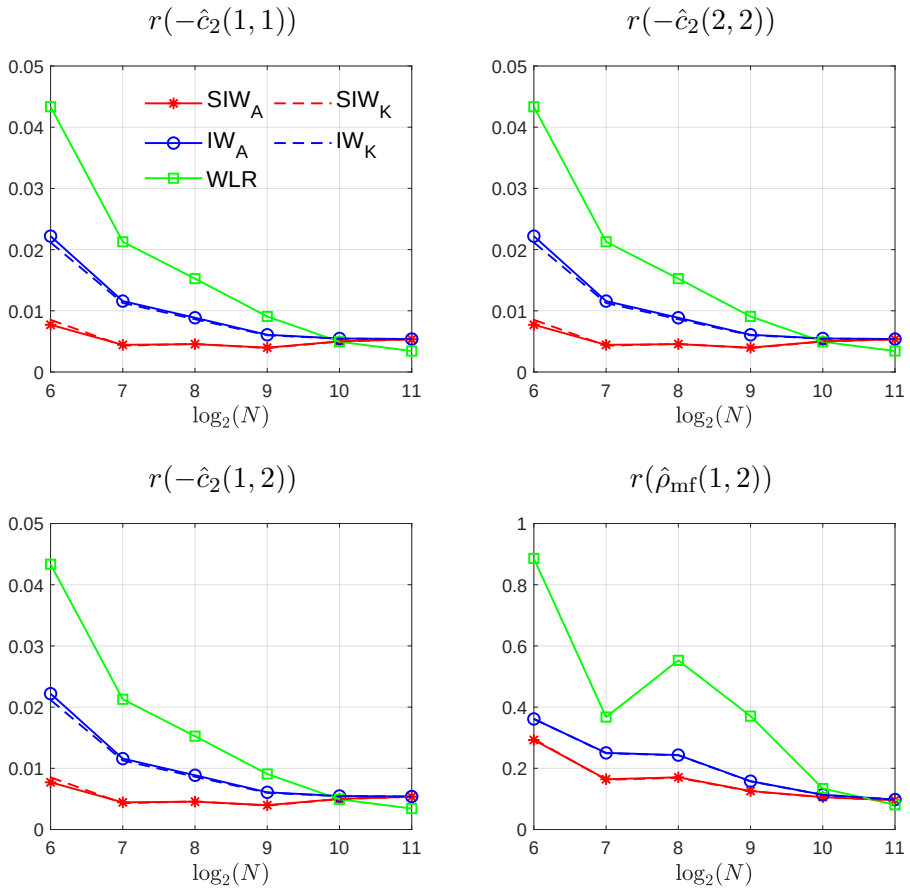

Figure 2.11: RMSE performance for 2D MV-MRW with $R = 2$, $\rho_{\mathrm{mf}}(1,2) = 0.5$, $c_2(1,1) = c_2(1,2) = -0.02$, $c_2(2,2) = -0.08$, $N = \{2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}\}$, $j_1 = 2$ and $j_2 \in \{3,4,5,6,7,8\}$.

Table 2.1: 1D MV-MRW estimation performance for $N = 2^6$, $R = 2$ and $-c_2(2,2) \in \{0.02, 0.04, 0.06, 0.08, 0.1\}$ (best results in **bold**).

$-c_2(2,2)$			0.02	0.04	0.06	0.08	0.1
$-c_2(1,1) = 0.02$	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
$-c_2(2,2)$	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
$-c_2(1,2)$	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

e) Estimation performance vs. numbers of data components

Fig. 2.12 displays the RMSE values of the estimated multifractal parameters, computed as the root square of the average over realizations of the trace of the matrix $(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^T$, where the vector $\boldsymbol{x}$ contains the diagonal and upper triangle of $\boldsymbol{\Sigma}_1$. 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 case $R=2$: The Bayesian estimators perform significantly better than WLR, and $\mathrm{SIW_A}$ has slightly lower RMSE than $\mathrm{IW_A}$.

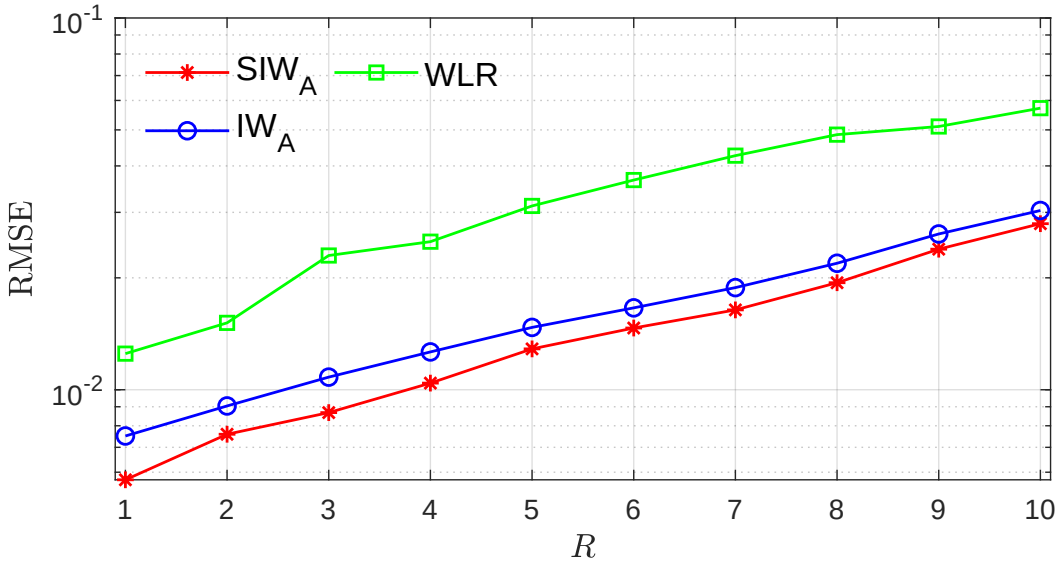

Figure 2.12: 2D MV-MRW estimation performance for $R \in \{1, 2, \ldots, 10\}$, $N = 2^9$, $j_1 = 2$ and $j_2 = 5$.

f) 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 $O(R^3 n \ln n)$, i.e., nearly linear in sample size n and cubic in R. For comparison, the method based on linear regression has complexity $O(nR^3)$ (without the $\ln n$ factor due to our additional use of a fast Fourier transform).

The iterative algorithm introduced in [BI13] to compute the Karcher mean to approximate the

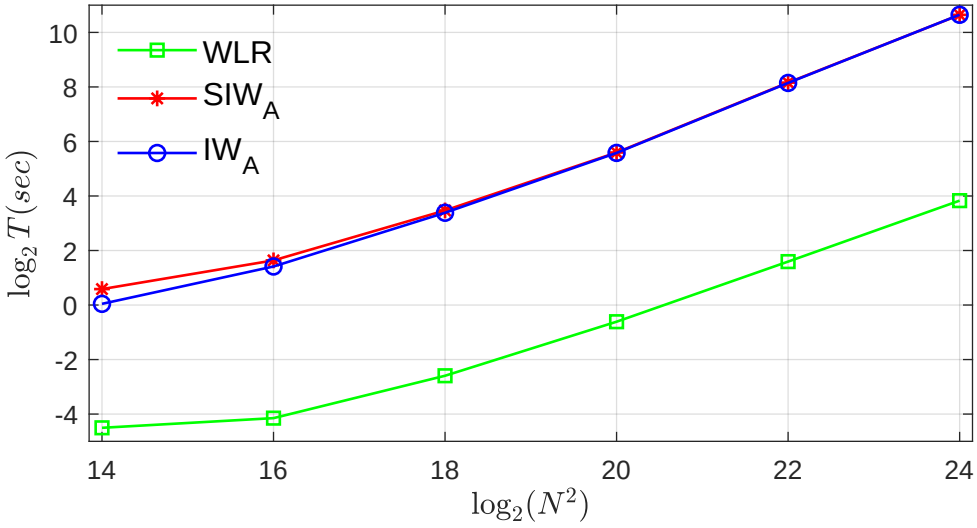

Figure 2.13: Computational time T (in seconds) versus image sizes $N \times N \times 2$ with $N \in \{8, 9, 10, 11, 12\}$, $j_1 = 2$ and $j_2 = \log_2 N - 4$ for all methods.

MMSE estimator does not depend on n but depends linearly on the number of matrices being averaged. Moreover, it has the same complexity in R ($O(R^3)$). The computational time comparison in this section is made as a function of n, thus the algorithms using the Karcher mean are not included.

Fig. 2.13 compares the execution times for the estimation of the multivariate multifractal parameters from a bivariate image ($R = 2$) of size $N \times N$ using WLR, IW_A and SIW_A estimators, as a function of N. The results confirm the leading order complexity estimates mentioned above. Specifically, 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_A is found to be only marginally larger than that of IW_A.

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 MFA, at reasonably larger computational cost than linear regression.

2.5 Conclusions and perspectives

Conclusions. In this chapter, we proposed and studied Bayesian estimation procedures for the multifractality parameters associated with the recent theoretical definition of the joint multifractal spectrum of multivariate data. Specifically, we introduced and validated an original and versatile model for the joint statistics of the log-leaders of several data components that combines a Whittle-type spectral approximation and a data augmentation strategy. Based on the proposed statistical model, two alternative Bayesian models are derived assuming IW and SIW prior distributions for the multifractal matrix-valued parameters of interest. The MMSE estimator is then approximated using standard Gibbs sampling and two ways to compute the average p.d. matrix estimates: the classical arithmetic mean and the Karcher mean. This leads to four different Bayesian estimators that were numerically studied and compared against the WLR through numerous experiments performed on synthetic multifractal data. The results clearly demonstrated the significant advantages of the proposed framework in terms of estimation accuracy notably for small sample sizes and the multivariate multifractal parameters, at moderate computational cost.

Perspectives. The estimation of the first-order log-cumulant c_1 associated with the average regularity of each component of the data can be incorporated straightforwardly into the proposed statistical model and estimation framework, following the ideas of [Com16]. However, the incorporation of higher-order log-cumulants is more challenging and will require taking into account a more complex statistical model. This would constitute an important continuation of this work. Along another line, the acceleration and improvement of the sampling algorithms could be investigated, for example, we could use strategies to recycle the samples in the algorithm resulting from using SIW priors similar to the strategy proposed in [MECV18]. Other prior distributions with more complex parametrization and greater flexibility have been proposed for covariance matrix estimation, see, e.g., [ANS14]. Their study could be another option to further improve the multifractal parameter estimation accuracy, possibly at the cost of having more difficult estimation problems. These research directions are discussed in more detail in Chapter 6.

Chapter 3

Expectation-Maximization

Contents

3.1 Introduction

In Chapter 2, we provided practical tools for multivariate multifractal analysis using Bayesian inference. Specifically, Markov chain Monte Carlo (MCMC) algorithms were proposed and studied for approximating the minimum mean square error (MMSE) estimator of the matrix-valued multifractal parameters $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$. The devised Bayesian models use (scaled) inverse Wishart priors and are built on the augmented likelihood (2.17) which approximates the original domain-based likelihood (2.9). Numerous numerical simulations verified that the proposed Bayesian approaches outperform the standard linear regression-based algorithms. The price to be paid was extra computational time. Even though the computational time of the Bayesian algorithms was shown not to be dramatic for our concrete examples, the processing of large real-world databases (e.g., many data components, large number of windows/patches under analysis, large sample sizes, etc) and the increasing number of real-time applications involving such data call for alternative formulations to reduce these computational costs.

Expectation-Maximization (EM) based algorithms, which are attractive for their reduced computational time, are considered here as an alternative to the MCMC-based algorithms developed in the previous chapter. The use of EM-based methods for the estimation of univariate multifractal parameters was first mentioned in the PhD thesis [Com16] (Appendix E). Inspired by these ideas, in this chapter, we devise new estimators of the matrix-valued multifractal parameters associated with the joint multifractal spectrum using EM strategies. We also compare their performance and computational costs against the classical estimators constructed using linear regressions and the previously proposed Bayesian estimators. The design of these new EM-based estimators and their evaluation are the main methodological and practical contributions of this chapter.

The outline of the chapter is set as follows. Section 3.2 presents the general formulation of the EM algorithm for augmented models and discusses its theoretical convergence. Section 3.3 derives the EM algorithms to performance maximum likelihood and maximum a posteriori estimation of the matrix-valued multifractal parameters of interest. Section 3.4 compares the performance of the different EM and MCMC estimators in terms of estimation accuracy and computational cost. This

comparison is done thanks to various Monte Carlo simulations performed for several multifractality parameter values, numbers of data components and observation sizes.

Developments and results presented in this chapter have been partly reported in [LWTA22c].

3.2 EM algorithm

The EM algorithm is an iterative scheme that can be used for computing the maximum likelihood estimator (MLE) or the maximum a posteriori (MAP) estimator for models with latent variables. It consists of generating a sequence of estimates that increases the likelihood or posterior of interest at each iteration and converges to one of its local maxima [DLR77]. In this section, we recall the general formulation of MLE and MAP estimators based on EM algorithms, denoted as EM-MLE and EM-MAP respectively.

Augmented models. Consider a general scenario in which the observed data $\boldsymbol{y}$ is augmented by some hidden vector $\boldsymbol{x}$ to form the *complete* data, where $\boldsymbol{x}$ can be either missing data or cleverly introduced latent variables. The joint likelihood $p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta})$ depends on a set of parameters to be inferred, abbreviated by $\boldsymbol{\theta}$.

If $\boldsymbol{\theta}$ is unknown and deterministic, the MLE of $\boldsymbol{\theta}$ based on the observed data $\boldsymbol{y}$ is the solution of the following optimization problem

$$\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ln p(\boldsymbol{y}|\boldsymbol{\theta}), \tag{3.1}$$

where the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ is obtained by marginalizing $p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta})$ with respect to (w.r.t.) $\boldsymbol{x}$, i.e., $p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}$.

Assuming that the unknown parameters have a prior distribution $p(\boldsymbol{\theta})$, the MAP estimator of $\boldsymbol{\theta}$ is obtained by maximizing the posterior distribution as follows

$$\widehat{\boldsymbol{\theta}}^{\mathrm{MAP}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ln p(\boldsymbol{\theta}|\boldsymbol{y}), \tag{3.2}$$

where the posterior $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})$ can be computed using Bayes' theorem and $p(\boldsymbol{\theta}|\boldsymbol{y}) = \int p(\boldsymbol{\theta}, \mathbf{x}|\boldsymbol{y}) d\mathbf{x}$.

Problems (3.1) and (3.2) can be difficult to solve because the cost functions to be maximized can be complicated and have several local maxima. In practice, EM-based algorithms can be used to approximate $\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}}$ and $\widehat{\boldsymbol{\theta}}^{\mathrm{MAP}}$ as follows.

EM-MLE. Starting from some initial parameter $\boldsymbol{\theta}^{(0)}$, the EM-MLE algorithm iterates between the following two steps

1. **Expectation (E-step)**: compute the expectation of the log-augmented likelihood $\ln p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta})$ w.r.t. the conditional distribution of the latent vector $\boldsymbol{x}$ given the current estimate $\boldsymbol{\theta}^{(\lambda)}$, i.e.,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\lambda)}) = \mathbb{E}_{p(\mathbf{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(\lambda)})}[\ln p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta})]. \tag{3.3}$$

2. **Maximization (M-step)**: update $\boldsymbol{\theta}^{(\lambda)}$ with

$$\boldsymbol{\theta}^{(\lambda+1)} = \operatorname*{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\lambda)}). \tag{3.4}$$

EM-MAP. Similarly, the EM-MAP algorithm iterates between the following two steps

1. **E-step**: compute the expectation of the log-augmented posterior likelihood $p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ w.r.t. the conditional distribution of the latent vector $\boldsymbol{x}$ given the current estimate $\boldsymbol{\theta}^{(\lambda)}$, i.e.,

$$\begin{aligned} F(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\lambda)}) =& \mathbb{E}_{p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(\lambda)})}[\ln p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})] \\ =& \mathbb{E}_{p(\mathbf{x}|\boldsymbol{y}, \boldsymbol{\theta}^{(\lambda)})}[\ln p(\boldsymbol{y}, \mathbf{x}|\boldsymbol{\theta})] + \ln p(\boldsymbol{\theta}). \end{aligned} \tag{3.5}$$

2. **M-step**: update $\boldsymbol{\theta}^{(\lambda)}$ with

$$\boldsymbol{\theta}^{(\lambda+1)} = \operatorname*{argmax}_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\lambda)}). \tag{3.6}$$

The sequence of parameters $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$ obtained from EM-MLE and EM-MAP algorithms does not decrease the corresponding log-likelihood and log-posterior, which guarantees that the EM algorithms converge to local maxima of the likelihood and the posterior, respectively. Hopefully, these local maxima are good approximations of the optimal solutions of the problems (3.1) and (3.2). The proof of this statement is recalled in the next section.

3.2.1 Convergence

The intuition behind EM-MLE is that at each iteration the two steps lead to an increase of the log-likelihood, i.e., $\ln p(\boldsymbol{y}|\boldsymbol{\theta}^{(\lambda+1)}) \geq \ln p(\boldsymbol{y}|\boldsymbol{\theta}^{(\lambda)})$ which ensures that the iterations will converge to some local maximum of the likelihood function. It can be seen by using the equality $p(\boldsymbol{y}|\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}) = p(\boldsymbol{y},\mathbf{x}|\boldsymbol{\theta})$ that

$$p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\boldsymbol{y}|\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\boldsymbol{y},\mathbf{x}|\boldsymbol{\theta}) - p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}) \tag{3.7}$$

which leads, by integrating over $\mathbf{x}$, to the equality

$$\begin{aligned} \ln p(\boldsymbol{y}|\boldsymbol{\theta}) &= Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\lambda)}) - \int p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}) d\mathbf{x} \\ &= Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\lambda)}) + \mathcal{D}_{KL}\big(p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}||p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta})\big) - \int p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) d\mathbf{x} \end{aligned} \tag{3.8}$$

where $\mathcal{D}_{KL}\big(\cdot\,||\,\cdot\big)$ is the Kullback-Leibler divergence. Therefore, it can be derived that

$$\begin{aligned} \ln p(\boldsymbol{y}|\boldsymbol{\theta}^{(\lambda+1)}) &= Q(\boldsymbol{\theta}^{(\lambda+1)},\boldsymbol{\theta}^{(\lambda)}) + \mathcal{D}_{KL}\big(p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}||p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda+1)})\big) - \int p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) d\mathbf{x} \\ &\geq^{(1)} Q(\boldsymbol{\theta}^{(\lambda)},\boldsymbol{\theta}^{(\lambda)}) + \mathcal{D}_{KL}\big(p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}||p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda+1)})\big) - \int p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) \ln p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}) d\mathbf{x} \\ &=^{(2)} \ln p(\boldsymbol{y}|\boldsymbol{\theta}^{(\lambda+1)}) + \mathcal{D}_{KL}\big(p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda)}||p(\mathbf{x}|\boldsymbol{y},\boldsymbol{\theta}^{(\lambda+1)})\big) \\ &\geq^{(3)} \ln p(\boldsymbol{y}|\boldsymbol{\theta}^{(\lambda)}) \end{aligned}$$

where the inequality (1) comes from the definition of $\boldsymbol{\theta}^{(\lambda+1)}$ in (3.4), the equality (2) comes from (3.8) and the last inequality (3) is due to the positivity of the KL divergence.

Although an EM iteration does increase the likelihood, no guarantee exists that the sequence converges to the MLE. For multimodal distributions, the EM algorithm can converge to a local maximum of the observed data likelihood function, depending on starting values.

A similar analysis can be derived to prove that the log-posterior distribution is non-decreasing at each iteration when using the EM-MAP estimator.

It is well known that the rate at which the EM algorithm converges is linear. For more properties on convergence of the EM algorithm, the reader is referred to e.g. [RW84, DLR77, Wu83].

3.3 EM algorithms for multivariate multifractal analysis

3.3.1 EM-MLE estimator

In the following, we devise an EM algorithm for the approximation of the MLE estimator of $\boldsymbol{\Gamma} = (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ defined as

$$\widehat{\boldsymbol{\Gamma}}^{\mathrm{MLE}} = \underset{\boldsymbol{\Gamma}}{\operatorname{argmax}} \ln p(\boldsymbol{z}|\boldsymbol{\Gamma}), \tag{3.9}$$

where $\boldsymbol{z} \in \mathbb{C}^{RM}$ and $\ln p(\boldsymbol{z}|\boldsymbol{\Gamma})$ is the logarithm of the likelihood (2.15) which is obtained by marginalizing $p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})$ (2.17) w.r.t. $\boldsymbol{u} \in \mathbb{C}^{RM}$, i.e., $p(\boldsymbol{z} \mid \boldsymbol{\Gamma}) = \int p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) d\boldsymbol{u}$. Given an initial guess $\boldsymbol{\Gamma}^{(0)}$, the proposed EM-MLE method iterates between the following two steps.

E-step. We compute the expectation of the log-augmented likelihood

$$\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma}) = -M \ln(\det \boldsymbol{\Sigma}_1) - M \ln(\det \boldsymbol{\Sigma}_2) - \frac{1}{2} \operatorname{Tr}(\boldsymbol{\Sigma}_1^{-1} \tilde{\boldsymbol{\Phi}}_1 + \boldsymbol{\Sigma}_2^{-1} \tilde{\boldsymbol{\Phi}}_2) + \text{constant}, \tag{3.10}$$

w.r.t. the conditional distribution of the latent vector $\boldsymbol{u}$ (2.31) given the current estimate $\boldsymbol{\Gamma}^{(\lambda)} = (\boldsymbol{\Sigma}_1^{(\lambda)}, \boldsymbol{\Sigma}_2^{(\lambda)})$

$$Q(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{(\lambda)}) = \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})]. \tag{3.11}$$

Matrices $\tilde{\boldsymbol{\Phi}}_1$ and $\tilde{\boldsymbol{\Phi}}_2$, defined in (2.18) and (2.19), are recalled bellow

$$\tilde{\boldsymbol{\Phi}}_1 = 2 \sum_{s=1}^{M} (\boldsymbol{z}_s - \boldsymbol{u}_s)(\boldsymbol{z}_s - \boldsymbol{u}_s)^H g_{1,s}^{-1}, \;\; \tilde{\boldsymbol{\Phi}}_2 = 2 \sum_{s=1}^{M} \boldsymbol{u}_s \boldsymbol{u}_s^H g_{2,s}^{-1}.$$

By linearity of the trace operator, it can be shown that

$$\begin{aligned} \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})] = & - M \ln(\det \boldsymbol{\Sigma}_1) - M \ln \det(\boldsymbol{\Sigma}_2) + \text{constant} \\ & - \frac{1}{2} \operatorname{Tr}(\boldsymbol{\Sigma}_1^{-1} \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})}[\tilde{\boldsymbol{\Phi}}_1] + \boldsymbol{\Sigma}_2^{-1} \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})}[\tilde{\boldsymbol{\Phi}}_2]), \end{aligned} \tag{3.12}$$

where

$$\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})} \left[\tilde{\boldsymbol{\Phi}}_1 \right] = 2 \sum_{s=1}^{M} \left(\tilde{\boldsymbol{\Sigma}}_s^{(\lambda)} + (\boldsymbol{z}_s - \boldsymbol{\mu}_s^{(\lambda)})(\boldsymbol{z}_s - \boldsymbol{\mu}_s^{(\lambda)})^H \right) g_{1,s}^{-1}, \tag{3.13}$$

$$\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\Gamma}^{(\lambda)})} \left[\tilde{\boldsymbol{\Phi}}_2 \right] = 2 \sum_{s=1}^{M} \left(\tilde{\boldsymbol{\Sigma}}_s^{(\lambda)} + \boldsymbol{\mu}_s^{(\lambda)} (\boldsymbol{\mu}_s^{(\lambda)})^H \right) g_{2,s}^{-1}. \tag{3.14}$$

M-step. We maximize $Q(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{(\lambda)})$ by finding the zeroes of its gradient $\nabla \boldsymbol{Q} = (\nabla \boldsymbol{Q}_1, \nabla \boldsymbol{Q}_2)$, with

$$\nabla \boldsymbol{Q}_1 = \frac{\partial \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})]}{\partial \boldsymbol{\Sigma}_1} = -M\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_1^{-1}\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_1\right]\boldsymbol{\Sigma}_1^{-1}, \tag{3.15}$$

$$\nabla \boldsymbol{Q}_2 = \frac{\partial \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})]}{\partial \boldsymbol{\Sigma}_2} = -M\boldsymbol{\Sigma}_2^{-1} + \boldsymbol{\Sigma}_2^{-1}\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_2\right]\boldsymbol{\Sigma}_2^{-1}, \tag{3.16}$$

yielding the following closed-form expressions for the parameters to be updated

$$\boldsymbol{\Sigma}_1^{(\lambda+1)} = M^{-1}\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_1\right], \tag{3.17}$$

$$\boldsymbol{\Sigma}_2^{(\lambda+1)} = M^{-1}\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_2\right]. \tag{3.18}$$

3.3.2 EM-MAP estimator

The MMSE estimators, proposed in Chapter 2, rely on two Bayesian models using inverse Wishart (IW) or scaled inverse Wishart (SIW) prior distributions for $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$. Similar to the proposed MCMC-based algorithms, the use of an IW prior leads to a simpler formulation of the estimators when compared to an SIW prior (the M-step requires solving a system of nonlinear equations that is not always easy and fast to solve). Since our goal here is to investigate for the first time the usefulness of the EM strategies for a faster and accurate estimation of the matrix-valued multifractal parameters of interest, we consider the IW prior-based Bayesian model for simplicity. The study of the EM-based estimator when using SIW priors is the subject of future work.

In the following, we devise an EM algorithm for the approximation of the MAP estimator (2.26) based on the posterior distribution (2.28) using an IW prior. The proposed EM-MAP method iterates between the following two steps.

E-step. We compute the expectation of the log-augmented posterior $\ln p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z})$

$$\begin{aligned}\ln p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z}) = &-\frac{1}{2}(2M + \nu_1 + R + 1)\ln(\det \boldsymbol{\Sigma}_1) - \frac{1}{2}(2M + \nu_2 + R + 1)\ln(\det \boldsymbol{\Sigma}_2) \\ &- \frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Lambda}_1 + \tilde{\boldsymbol{\Phi}}_1) + \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Lambda}_2 + \tilde{\boldsymbol{\Phi}}_2)) + \text{constant},\end{aligned} \tag{3.19}$$

w.r.t. the conditional distribution of the latent vector $\boldsymbol{u}$ (2.31) given the current estimates $\boldsymbol{\Gamma}^{(\lambda)}$

$$\begin{aligned} F(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{(\lambda)}) =& \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z})] \\ =& \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{\Gamma})] + \ln p(\boldsymbol{\Sigma}_1) + \ln p(\boldsymbol{\Sigma}_2) \\ =& Q(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{(\lambda)}) + \ln p(\boldsymbol{\Sigma}_1) + \ln p(\boldsymbol{\Sigma}_2), \end{aligned} \tag{3.20}$$

where $p(\boldsymbol{\Sigma}_i)$ has been defined in (2.21).

M-step. We update the parameter estimates by maximizing the function $F(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{(\lambda)})$. Indeed, finding the zeroes of the gradient $\nabla F = (\nabla F_1, \nabla F_2)$ of the function F, with

$$\nabla F_i = \frac{\partial \mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}[\ln p(\boldsymbol{\Gamma}, \boldsymbol{u} \mid \boldsymbol{z})]}{\partial \boldsymbol{\Sigma}_i} = \nabla Q_i - \frac{1}{2}\left((\nu_i + R + 1)\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_i^{-1}\right), \; i = 1, 2, \tag{3.21}$$

yields the following set of equations

$$\boldsymbol{\Sigma}_1^{(\lambda+1)} = \frac{2}{2M + \nu_1 + R + 1}\left[\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_1\right] + \boldsymbol{\Lambda}_1/2\right], \tag{3.22}$$

$$\boldsymbol{\Sigma}_2^{(\lambda+1)} = \frac{2}{2M + \nu_2 + R + 1}\left[\mathbb{E}_{p(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{\Gamma}^{(\lambda)})}\left[\tilde{\boldsymbol{\Phi}}_2\right] + \boldsymbol{\Lambda}_2/2\right]. \tag{3.23}$$

3.4 Numerical experiments

This section studies the estimation performance of the EM-MLE and the EM-MAP estimators defined in Sections 3.3.1 and 3.3.2 for the estimation of the multifractal parameters associated with the joint multifractal spectrum. We also compare their performance against the four MMSE estimators developed in Chapter 2 (SIW_A, SIW_K, IW_A and IW_K) and the linear regression-based estimator (1.27) denoted as WLR.

a) Monte Carlo simulations

To validate and study the estimation performance, we use Monte Carlo simulations considering a large number of independent realizations of a 1D multivariate multifractal random walk (MV-MRW) process (cf. Section 1.4 and e.g., [BDM01]). The analysis of the results is presented for the univariate and bivariate scenarios, i.e., the numbers of data components are $R = 1$ and $R = 2$.

Parameters of the multifractal signal. The multifractality parameter $-c_2(r,r)$, with $r = 1,\dots,R$, has been chosen in the interval $[0.01, 0.16]$ so as to consider various levels of regularity. The multifractal correlation for the bivariate setting is set to $\rho_{\mathrm{mf}}(1,2) = 0.5$. The lengths of the generated signals are set to $N \in \{2^5, 2^7, 2^8, 2^9, \dots, 2^{15}\}$.

Wavelet analysis. We used a Daubechies' wavelet with $N_\psi = 2$ vanishing moments, and scales $j \in [\![2, \log_2 N - 4]\!]$ for $N \in \{2^7, 2^8, 2^9, \dots, 2^{15}\}$ and $j \in [\![1, 2]\!]$ for $N = 2^5$.

Prior specification. In the bivariate setting, the hyperparameters of the prior distributions are set to $\nu_i = 4$ and $\mathbf{\Lambda}_i = \mathbb{I}_2$, with $i \in \{1, 2\}$ (cf. Chapter 2) leading to non informative priors. In the univariate setting, we only use IW priors and its parameters were chosen to have a support equal to $[0.004, 0.1]$, covering the possible values of these parameters, i.e., $\nu_1 = \nu_2 = 20$ and $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = 0.2$.

Algorithm specification. Note that the EM algorithm is stopped when the variations of the marginal log-likelihood or the marginal log-posterior are less than 10^{-4} or when a maximum of 200 iterations has been reached. The initial guess of the parameters for the initialization of the EM algorithms has been chosen as $\mathbf{\Sigma}_i^{(0)} = \mathbf{\Lambda}_i/(\nu_i - R - 1)$, with $i = 1, 2$. The consequences of this choice is discussed in Section 3.4.1. For the MMSE estimators, we chose a number of burn-in iterations $N_{\mathrm{bi}} = 1000$ and a total number of iterations $N_{\mathrm{mc}} = 2000$.

Performance. The estimation performance is evaluated in terms of mean, standard deviation (STD) and root mean square error (RMSE), defined in (2.41), computed using the different realizations of the MV-MRW processes.

b) Estimation performance

Fig. 3.1 compares all proposed estimators (SIW_A, SIW_K, IW_A, IW_K, EM-MAP, EM-MLE and WLR) as a function of $N \in \{2^7, 2^8, 2^9, \dots, 2^{15}\}$, for $R = 2$, $c_2(1,1) = c_2(1,2) = -0.02$, $c_2(2,2) = -0.08$ and $\rho_{\mathrm{mf}}(1,2) = 0.5$, in terms of the RMSE of the multifractal parameter estimates.

WLR performance. All estimators, including EM algorithms, overall perform better than WLR, most notably for small sample sizes ($N < 2^8$). The large RMSE values of the WLR estimator are the result of the large variance produced by them, which becomes critical for large values of $c_2(r,r)$ and in the estimation of ρ_{mf} (see Section 2.4.2).

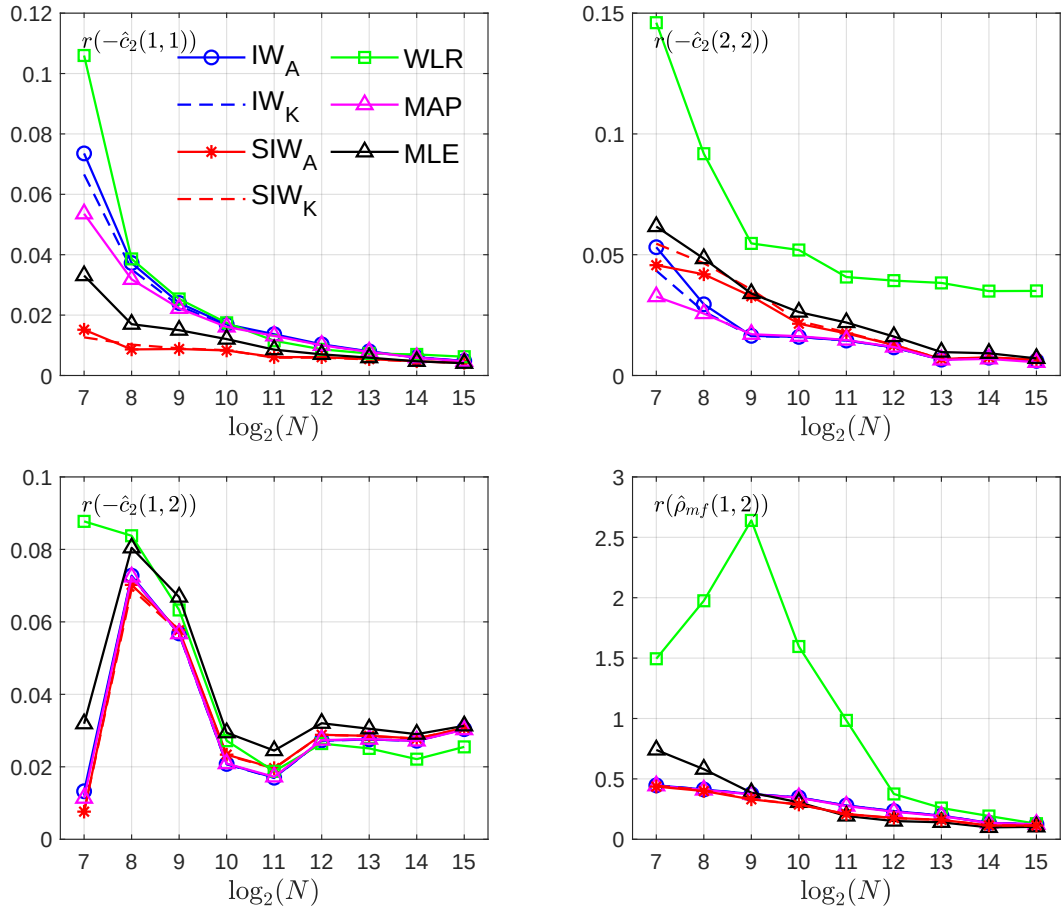

Figure 3.1: RMSE of all estimators as a function of $N \in \{2^7, 2^8, 2^9, \ldots, 2^{15}\}$ averaged over 100 copies of a 1D bivariate MRW process with $c_2(1,1) = c_2(1,2) = -0.02$, $c_2(2,2) = -0.08$ and $\rho_{\text{mf}}(1,2) = 0.5$.

EM algorithms. Overall, the MAP estimator has reduced RMSE values compared to the MLE estimator except when the univariate multifractality parameter $c_2(1,1)$ is small. This is explained by the large bias induced by the IW prior for small $c_2(r,r)$ values, as discussed in Section 2.4.2.

Bayesian estimators. The MMSE and MAP estimators have similar behavior for all values of N especially for the cross-parameter $c_2(1,2)$ and the multifractal correlation ρ_{mf}. Regarding the univariate multifractality parameters $c_2(1,1), c_2(2,2)$, the performance of the Bayesian methods depends on the chosen prior distribution. The performance gain of estimators using the SIW prior is in particular remarkable for small values of $c_2(r,r)$ and for small sample sizes (values reaching up

to 4 times smaller when $c_2(1,1) = -0.02$ and $N = 2^7$). For more details on the comparison of the two different priors, the reader is referred to Section 2.4.2.

For a better presentation of the results, only the MMSE estimator using an IW prior (the same as the one used for the MAP estimator) and arithmetic mean will be considered in what follows.

Performance vs. c_2. Fig. 3.2 compares the MMSE, WLR, EM-MLE and EM-MAP estimators as a function of $-c_2 \in [0.01, 0.16]$, for $R = 1$ and three sample sizes $N = 2^5, 2^7, 2^9$. The results are averaged over $N_{\text{mc}} = 10000$ realizations and indicate that using the models of Chapter 2 for the univariate scenario allows us to obtain significant performance gains, in particular for small values of N and for large values of c_2, with RMSE values up to 5 times smaller than with linear regression. We note that the use of Bayesian estimators improves the estimation performance compared to the MLE, especially for small sample size ($N = 2^5$). Overall, the performance of the Bayesian estimators (MMSE and EM-MAP) using the same IW prior is very similar.

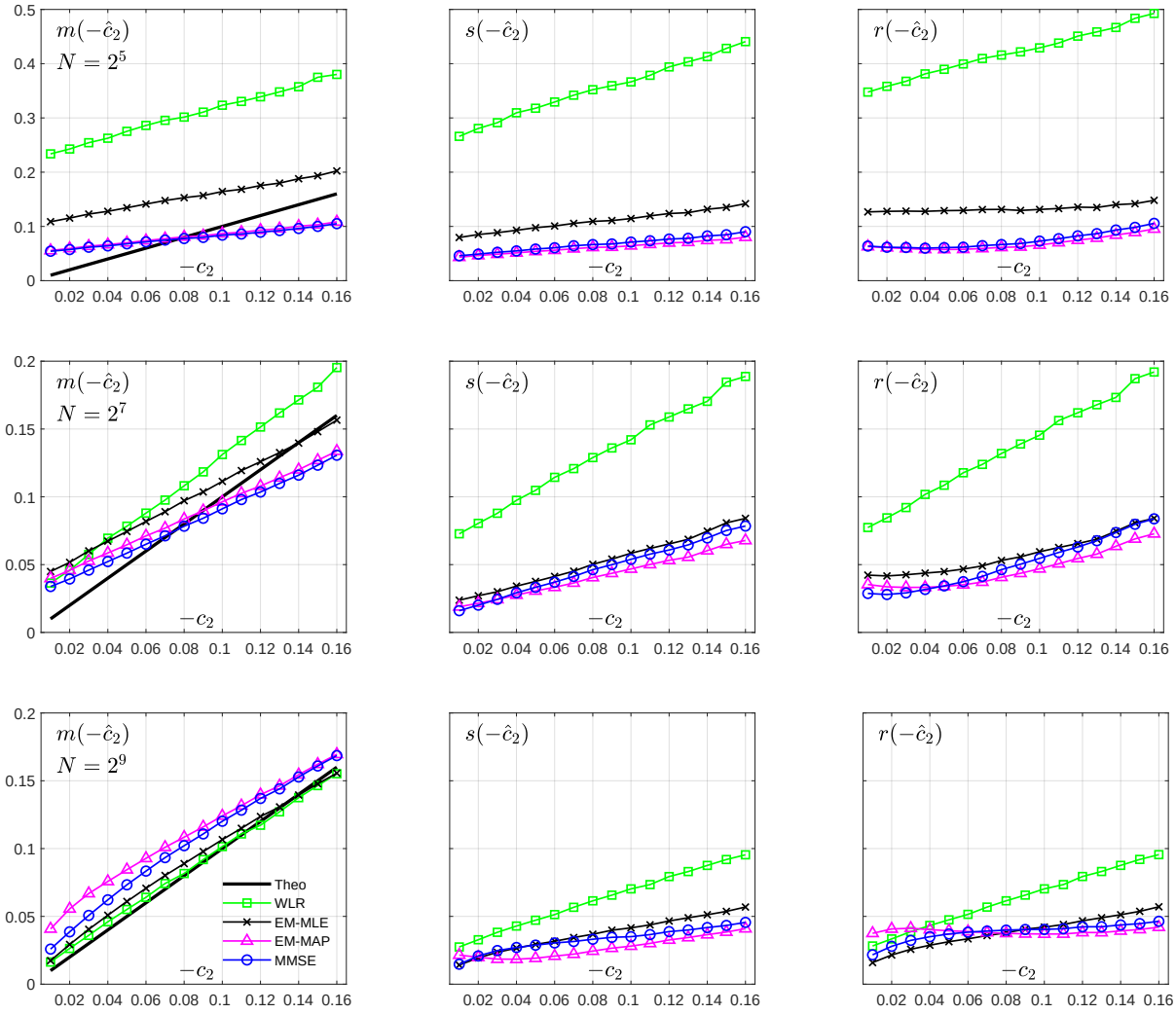

Figure 3.2: Estimation performance (from left to right: mean, STD and RMSE) as a function of c_2 for $N = 2^5$ (top), $N = 2^7$ (middle) $N = 2^9$ (bottom).

3.4.1 Computational cost and convergence

Computational cost. Fig. 3.3 shows the execution times of the different algorithms as a function of $N \in \{2^5, 2^6, \ldots, 2^{15}\}$ when $R = 1$. It is observed that the MMSE estimator is approximately 50 times slower than the EM-MAP estimator, which itself is twice as slow as a linear regression for $N = 2^{10}$. Overall, this allows us to conclude that the EM-MAP estimator allows us to obtain significant gains in estimation performance compared to a linear regression or a MLE, with a computational cost much lower than the MMSE estimator, and of the order of that obtained with a linear regression.

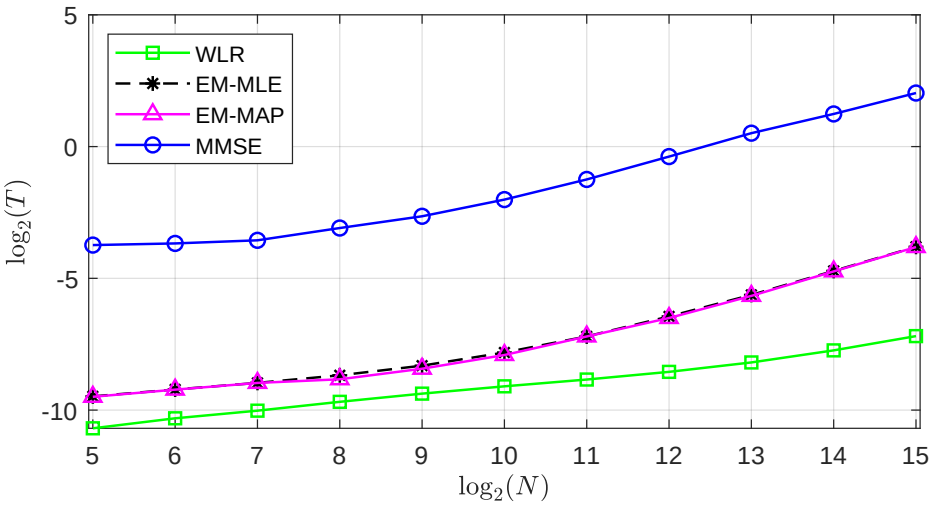

Figure 3.3: Execution time as a function of N for $R = 1$ (Intel(R) Xeon(R) Silver 4114 processor CPU 2.20GHz, 64GB RAM, single-threaded).

Convergence. As an illustration, Fig. 3.4 shows, for sample size $N \in \{2^7, 2^9, 2^{12}\}$, how the target function of the EM algorithms increases with each iteration. The initialization is the same for both methods. We observe that the MAP estimator requires less iterations to converge, in particular for small sample size ($N = 2^7$). This can be interpreted in view of the regularization introduced by the prior information in the Bayesian formulation.

Overall, the EM-MLE estimator converge slower than the EM-MAP estimator. Fig. 3.5 displays the histograms of the reached iterations for a total of 1000 independent realizations of the algorithms

for $R = 2$ and $N = 2^7$. Fig. 3.5 (left) corresponds to the initial guess $\mathbf{\Sigma}_i^{(0)} = \mathbb{I}_2$ and Fig. 3.5 (right) corresponds to a different one, $\mathbf{\Sigma}_i^{(0)} = \begin{pmatrix} 0.1181 & -0.0309 \\ -0.0309 & 0.1954 \end{pmatrix}$, with $i = 1, 2$. We can observe that the maximum and the mean of the reached iterations vary with the starting point. Therefore, the initialization strategies of the EM algorithms should be carefully chosen.

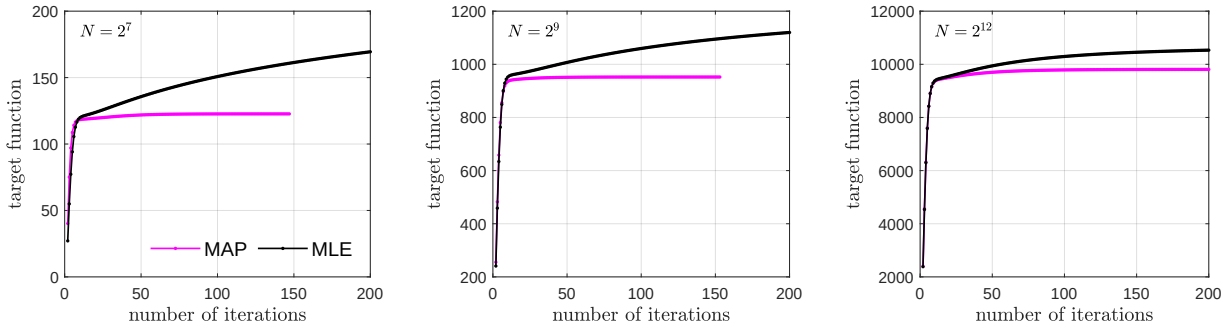

Figure 3.4: Evaluation of the target function of the EM algorithms in each iteration, for $R = 2$ and $N \in \{2^7, 2^9, 2^{12}\}$. Initialization $\mathbf{\Sigma}_i^{(0)} = [1\ 0\ ;\ 0\ 1]$, with $i = 1, 2$

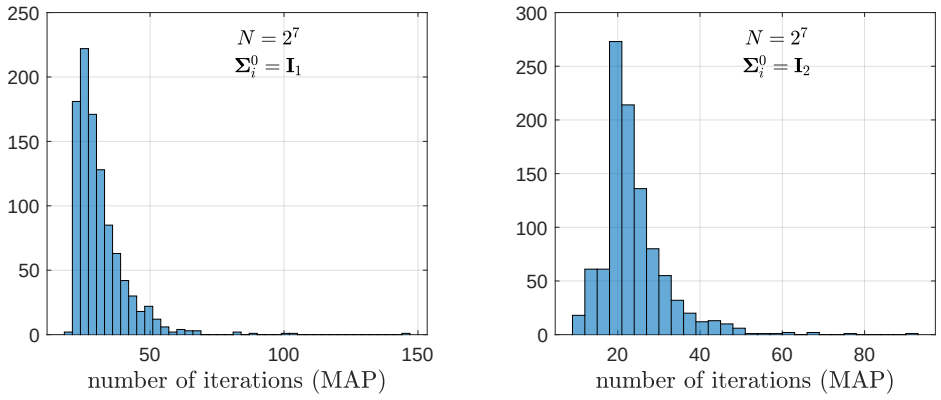

Figure 3.5: Number of iterations of EM-MAP estimator to converge over 1000 independent realizations and for two different initialization: $\boldsymbol{I}_1 = \mathbb{I}_2$ (left) and $\boldsymbol{I}_2 = \begin{pmatrix} 0.1181 & -0.0309 \\ -0.0309 & 0.1954 \end{pmatrix}$ (right).

3.5 Conclusions and perspectives

Conclusions. Building on the statistical model of log-leaders introduced in Chapter 2, new methods and algorithms based on EM strategies were proposed and studied in this chapter to approximate the MLE and MAP estimator of the multivariate multifractal parameters of interest. The relevance of EM algorithms for multifractal parameter estimation was verified on synthetic data. In particular, the proposed EM-MAP and the MMSE estimator based on the Bayesian model associated with an IW prior distribution yield overall similar estimation performance. Nevertheless, the EM-MAP estimator benefits from significantly lower computational costs.

Yet, it was again evidenced that the use of the SIW prior (only used with MMSE estimators) leads to significant estimation improvements in general, which motivates its use in the applications studied in Chapter 5.

Perspectives. One main drawback of these EM algorithms is the dependence on initialization, as observed in the experiments. Therefore, a continuation of this work could be to investigate different initialization strategies. It would be interesting to explore the use of these EM-based estimators on large real-world datasets, where the reduced computational time of such algorithms can become critical. The EM strategies developed in this chapter could also be used to design new estimation algorithms robust to noisy data, where outliers can be modeled as latent variables. Another possible line of work is to explore this methodology when an SIW prior is assigned to the parameters, a challenge being to achieve a reduced computational cost since the update of the parameters in the M-step does not have closed-form. These ideas are discussed in more detail in the concluding Chapter 6.

Chapter 4

Bounds for the estimation of the matrix-valued parameters

Contents

4.1 Introduction

In the earlier chapters, we proposed several estimators for the symmetric positive definite (p.d.) matrices $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ associated with the joint multifractal spectrum of multivariate data. The estimation relies on a multivariate statistical Gaussian model for the log-leaders of multivariate data. For convenience and numerical reasons, this model is diagonalized in the Fourier domain where the parameters of interest govern the covariance structure of a zero mean circularly-symmetric complex Gaussian likelihood (2.11) (or the augmented version (2.17)). For a deterministic parameter estimation, we devised an Expectation-Maximization (EM) algorithm to approximate the maximum likelihood estimator (MLE) [Kay93] in Section 3.3.1. On the other hand, we proposed Bayesian approaches assuming a prior distribution for these matrix-valued parameters and then devised Markov chain Monte Carlo (MCMC) and EM-based algorithms for approximating the minimum mean square (MMSE) and the maximum a posteriori (MAP) estimators (see Sections 2.3.2 and 3.3.2). The proposed estimators have been validated through numerous Monte Carlo simulations (see Sections 2.4 and 3.4). A theoretical analysis of the performance of these estimators can explain and predict their behavior in specific scenarios. This can facilitate the design of experiments (numbers of subjects in study, recording lengths, etc) to yield certain accuracy. This can also make it possible to study the asymptotic behavior of the estimators of the multifractal parameters of interest. This kind of theoretical analysis can be achieved by establishing fundamental lower bounds for the mean squared error (MSE) of these parameters.

Many lower bounds have been developed for deterministic settings, such as the classical Cramér–Rao [Cra46, Rao92], Hammersley–Chapman–Robbins [Ham50, CR51], Bhattacharya [Bha66] and Barankin [Bar49] bounds, as well as more recent results [Abe93, HFU96, FL02, Eld04], to name but a few. Similarly, lower bounds have been derived for parameters that are assigned an *a priori* probability distribution as well, commonly referred to as Bayesian bounds, for example, the Bayesian Cramér–Rao [Tre01], Bobrovski–Zakai [BZ76] and Weiss–Weinstein [WW85, WW88] bounds. By far the simplest and most commonly used of these approaches is the Cramér–Rao bound (CRB) and its Bayesian version. In what follows, we will refer to the Bayesian Cramér–Rao bound as BB in short.

To the best of our knowledge, none of the existing works provide a lower bound on the MSE performance for multivariate multifractal parameter estimation and specifically for the model proposed in Chapter 2. In this chapter, we close this gap by providing the CRB and the BB on the MSE performance of estimators in the multivariate multifractal analysis (MFA) framework. In the spirit of [BBT08], in this chapter, we derive the CRB and BB for the MSE of the matrix-valued parameters of the statistical model described in the Section 2.2.3.

First, in Section 4.2, we recall in a general fashion the statistical model and the estimation problem. Then, considering $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ to be unknown and deterministic matrices, Section 4.3 derives the CRB. Assuming they are unknown and random matrices with inverse Wishart (IW) prior distributions (see Section 2.3), Section 4.4 derives the BB of $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ and analytically studies their properties. In Section 4.5, we study the properties of the bounds in this framework using Monte Carlo simulations. Finally, in Section 4.5.3, we illustrate the use of the proposed bounds for the parameters associated with the bivariate multifractal spectrum. For simplicity, this chapter will be limited to an IW prior, see Section 4.6 for further comments on the use of the scaled inverse Wishart (SIW) prior.

To sum up, the main contributions of this chapter are i) the derivation of the BB for the specified statistical model, which is a new theoretical result obtained from (4.8), (4.14), (4.16) and (4.20), ii) the derivation and proof of a novel closed-form expression for computing non-trivial expectations involving Wishart random matrices, see Proposition 1, iii) the study of the analytic properties of the bounds (see Section 4.4) and iv) extensive numerical experiments and results that validate and illustrate the obtained theoretical expressions of the bounds (see Section 4.5).

The work presented in this chapter has been submitted to a journal on 23/09/2022.

4.2 Problem statement and statistical model

Consider M independent zero mean circularly-symmetric complex Gaussian random vectors $\boldsymbol{z}_s \in \mathbb{C}^R, s = 1, ..., M$, such that $\mathbb{E}[\boldsymbol{z}_s] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{z}_s \boldsymbol{z}_s^T] = \mathbf{0}$, and $\mathbb{E}[\boldsymbol{z}_s \boldsymbol{z}_s^H] = \boldsymbol{R}_s$, i.e., $\boldsymbol{z}_s \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R}_s)$. The covariance matrix $\boldsymbol{R}_s$ is assumed to be real-valued p.d. and of the form

$$\boldsymbol{R}_s = \boldsymbol{\Sigma}_1 g_{1,s} + \boldsymbol{\Sigma}_2 g_{2,s}, \; s = 1, ..., M,$$

where $g_{1,\cdot},\ g_{2,\cdot} > 0$ are known real-valued functions for all $s \in \{1, \ldots, M\}$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ are the $R \times R$ symmetric p.d. matrix-valued parameters to be estimated. This model is identical to the statistical model introduced in Chapter 2 for log-leaders. Thus, the vector of the MR samples arranged as $\boldsymbol{z} = (\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_M^T) \in \mathbb{C}^{MR}$ can be modeled as a zero mean circularly-symmetric complex Gaussian random vector with the real-valued covariance matrix

$$\boldsymbol{R} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{G}_1 + \boldsymbol{\Sigma}_2 \otimes \boldsymbol{G}_2, \tag{4.1}$$

where $\otimes$ is the Kronecker product and $\boldsymbol{G}_1, \boldsymbol{G}_2$ are known diagonal matrices whose sth diagonal entries are given by $[\boldsymbol{G}_i]_{ss} = g_{i,s}$, for $i \in \{1, 2\}$.

The expressions derived and studied in what follows can be generalized to more than 2 matrix summands. For ease of presentation, we treat here the case with 2 summands, without loss of generality.

4.3 Cramér-Rao bound

Let $\boldsymbol{\theta} \in \mathbb{R}^p$, with $p = R^2 + R$, the vector obtained by concatenating the vectors $\mathrm{vec}_{\mathrm{triu}}(\boldsymbol{\Sigma}_1)$ and $\mathrm{vec}_{\mathrm{triu}}(\boldsymbol{\Sigma}_2)$, where the matrix operator $\mathrm{vec}_{\mathrm{triu}}(\boldsymbol{A})$ returns the vector of the elements of the upper triangular part of $\boldsymbol{A}$. Note that the first and the last $\frac{p}{2}$ elements of $\boldsymbol{\theta}$, denoted as $\boldsymbol{\theta}_{1:\frac{p}{2}}$ and as $\boldsymbol{\theta}_{\frac{p}{2}+1:p}$, correspond to the main diagonal and all elements of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ above the diagonal, respectively. In the following, the matrix $\boldsymbol{R}$ will be denoted as $\boldsymbol{R}(\boldsymbol{\theta})$ to emphasize the dependence of $\boldsymbol{R}$ on $\boldsymbol{\theta}$.

We first derive the Cramér–Rao bound for the estimation of $\boldsymbol{\Gamma} = (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, assuming that the matrices are deterministic and unknown. The evaluation of the CRB requires to invert the Fisher information matrix (FIM), defined as [Tre01]

$$\boldsymbol{F}_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{z}|\boldsymbol{\Gamma}} \left[-\frac{\partial^2 L(\boldsymbol{z} \mid \boldsymbol{\Gamma})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right], \tag{4.2}$$

where $L(\boldsymbol{z} \mid \boldsymbol{\Gamma})$ is the log-likelihood of $\boldsymbol{z}$ which can be expressed as

$$L(\boldsymbol{z} \mid \boldsymbol{\Gamma}) = -M \ln \pi - \ln \det \boldsymbol{R}(\boldsymbol{\theta}) - \boldsymbol{z}^H \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \boldsymbol{z}. \tag{4.3}$$

Note that $L(\boldsymbol{z} \mid \boldsymbol{\Gamma})$ (4.3) is twice differentiable with respect to (w.r.t.) $\boldsymbol{\theta}$ and has a bounded support independent of $\boldsymbol{\theta}$. These are sufficient regularity conditions to ensure the existence of the CRB.

Since $\boldsymbol{z}$ is a zero mean circularly-symmetric complex Gaussian vector, the element of $\boldsymbol{F}_{\boldsymbol{\theta}}$ located at the kth row and lth column, for $k, l \in \{1, \ldots, p\}$, can be calculated as [Fre93, PF86]

$$[\boldsymbol{F}_{\boldsymbol{\theta}}]_{kl} = \operatorname{tr}\left\{ \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{R}(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{R}(\boldsymbol{\theta})}{\partial \theta_l} \right\}. \tag{4.4}$$

Note that in general the expression (4.4) needs $O(R^3M^3)$ operations to be computed because it requires the matrix inversion of $\boldsymbol{R}(\boldsymbol{\theta})$. That can lead to a large computational time for large values of M and R and even when it is not problematic a faster solution is preferred. We propose to overcome this limitation by exploiting the diagonal block structure of $\boldsymbol{R}(\boldsymbol{\theta})$, that is,

$$\boldsymbol{R}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{R}_1(\boldsymbol{\theta}) & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \boldsymbol{R}_M(\boldsymbol{\theta}) \end{bmatrix}, \tag{4.5}$$

whose sth block given by $\boldsymbol{R}_s(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_1 g_{1,s} + \boldsymbol{\Sigma}_2 g_{2,s}$ and $s = 1, \ldots, M$. Thus, both the matrix inverse and the derivative operators can be applied to each $R \times R$ diagonal block of $\boldsymbol{R}(\boldsymbol{\theta})$, individually. Specifically,

$$\boldsymbol{R}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{R}_1^{-1}(\boldsymbol{\theta}) & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \boldsymbol{R}_M^{-1}(\boldsymbol{\theta}) \end{bmatrix}, \tag{4.6}$$

with $\boldsymbol{R}_s^{-1}(\boldsymbol{\theta}) = (\boldsymbol{\Sigma}_1 g_{1,s} + \boldsymbol{\Sigma}_2 g_{2,s})^{-1}$ and

$$\frac{\partial \boldsymbol{R}(\boldsymbol{\theta})}{\partial \theta_l} = \begin{bmatrix} \boldsymbol{B}_{1,l} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \boldsymbol{B}_{M,l} \end{bmatrix}, \tag{4.7}$$

with

$$\boldsymbol{B}_{s,l} = \frac{\partial \boldsymbol{R}_s(\boldsymbol{\theta})}{\partial \theta_l} = \begin{cases} \boldsymbol{J}_{1,l} g_{1,s} \text{ if } l \in \{1, \ldots, \frac{p}{2}\}, \\ \boldsymbol{J}_{2,l} g_{2,s} \text{ if } l \in \{\frac{p}{2} + 1, \ldots, p\}, \end{cases}$$

where $\boldsymbol{J}_{i,l} = \dfrac{\partial \boldsymbol{\Sigma}_i}{\partial \theta_l}$ does not depend on $\boldsymbol{\theta}$, hence $\dfrac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \theta_k \partial \theta_l} = \mathbf{0}$ for $i \in \{1,2\}$. Thus, (4.4) can be rewritten as

$$[\boldsymbol{F}_{\boldsymbol{\theta}}]_{kl}(\boldsymbol{\theta}) = \sum_{s=1}^{M} \operatorname{tr}\left\{\boldsymbol{R}_s^{-1}(\boldsymbol{\theta})\boldsymbol{B}_{s,k}\boldsymbol{R}_s^{-1}(\boldsymbol{\theta})\boldsymbol{B}_{s,l}\right\}, \tag{4.8}$$

which can be computed with only $O(R^3 M)$ operations.

The MSE of any estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is defined as the trace of the error covariance matrix

$$\text{MSE} = \operatorname{Tr}(\mathbb{E}_{\boldsymbol{\theta}}[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]), \tag{4.9}$$

where $\widehat{\boldsymbol{\theta}}$ is the vector obtained by concatenating the vectors $\text{vec}_{\text{triu}}(\widehat{\boldsymbol{\Sigma}}_1)$ and $\text{vec}_{\text{triu}}(\widehat{\boldsymbol{\Sigma}}_2)$. Finally, under the assumption that the estimator is unbiased, the MSE for the entries of $\boldsymbol{\Gamma}$, when $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ are deterministic, is lower bounded by the trace of the inverse of the FIM, i.e.,

$$\text{MSE} \geq \text{CRB} = \operatorname{Tr}([\boldsymbol{F}_{\boldsymbol{\theta}}]^{-1}). \tag{4.10}$$

4.4 Bayesian Cramér-Rao bound

In this section, we derive the BB for the MSE of estimators of $\boldsymbol{\Gamma}$, when $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are assigned independent inverse Wishart priors, i.e., $\boldsymbol{\Sigma}_i \sim \mathcal{IW}(\nu_i, \boldsymbol{\Omega}_i)$, with ν_i degrees of freedom ($\nu_i \in \mathbb{R}$ and $\nu_i > R+1$), and mean matrix $(\nu_i - R - 1)^{-1}\boldsymbol{\Omega}_i$, where $\boldsymbol{\Omega}_i$ is a real-valued p.d. scale matrix. To this end, we make use of the following results.

Proposition 1 *Moments of the type* $\mathbb{E}[\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}]$.

If $\boldsymbol{\Sigma} \sim \mathcal{IW}(\nu, \boldsymbol{\Omega})$, *then* $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$ *has the Wishart distribution* $\mathcal{W}(\nu, \boldsymbol{\Delta} = \boldsymbol{\Omega}^{-1})$. *Then, for any pair of real-valued symmetric matrices* $(\boldsymbol{A}, \boldsymbol{B})$*:*

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Sigma}}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{\Sigma}^{-1}\right] &= \boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}(\nu^3 + 2\nu^2 + \nu) + \boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}(\nu^2 + 3\nu) \\ &\quad + [\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{A})]\boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}(\nu^2 + \nu) + [\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{B})]\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}(\nu^2 + \nu) \\ &\quad + \boldsymbol{\Delta}\left[(\nu^2 + \nu)\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{B}) + \nu \operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{A})\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{B})\right].\end{aligned} \tag{4.11}$$

The proof of Proposition 1 is provided in Appendix A using the approach detailed in [GLM05].

On the other hand, according to [HP20], for any real-valued symmetric matrix $\boldsymbol{A}$, we have

$$\mathbb{E}_{\boldsymbol{\Sigma}}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\right] = (\nu^2 + \nu)\Delta\boldsymbol{A}\Delta + \nu \operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{A})\boldsymbol{\Delta}. \tag{4.12}$$

The evaluation of the BB requires to invert the posterior Fisher information matrix (PFIM), defined as [Tre01]

$$\boldsymbol{F} = \mathbb{E}_{\boldsymbol{z},\boldsymbol{\Gamma}}\left[-\frac{\partial^2 L(\boldsymbol{z},\boldsymbol{\Gamma})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right], \tag{4.13}$$

where $L(\boldsymbol{z},\boldsymbol{\Gamma})$ is the joint log-likelihood of the model, which is twice differentiable w.r.t. $\boldsymbol{\theta}$ and has a bounded support independent of $\boldsymbol{\theta}$. These required regularity conditions ensure the existence of the BB. Equation (4.13) can be rewritten as

$$\begin{aligned}\boldsymbol{F} &= \mathbb{E}_{\boldsymbol{\Gamma}}\left[\mathbb{E}_{\boldsymbol{z}|\boldsymbol{\Gamma}}\left[-\frac{\partial^2 L(\boldsymbol{z}\mid\boldsymbol{\Gamma})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] - \frac{\partial^2\pi_1(\boldsymbol{\Sigma}_1)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{\partial^2\pi_2(\boldsymbol{\Sigma}_2)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] \\ &= \mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{F}_{\boldsymbol{\theta}} + \boldsymbol{F}_{\boldsymbol{\Sigma}_1} + \boldsymbol{F}_{\boldsymbol{\Sigma}_2}\right].\end{aligned} \tag{4.14}$$

Moreover, $\pi_i(\boldsymbol{\Sigma}_i)$ is the log-prior of $\boldsymbol{\Sigma}_i$ and for $i \in \{1,2\}$, one has

$$\pi_i(\boldsymbol{\Sigma}_i) = -\frac{\nu_i + R + 1}{2}\ln\det\boldsymbol{\Sigma}_i - \frac{1}{2}\operatorname{Tr}(\boldsymbol{\Omega}_i\boldsymbol{\Sigma}_i^{-1}) + \text{constant}. \tag{4.15}$$

Note that the expression in (4.15) also satisfies the regularity conditions ensuring the existence of the BB for $\boldsymbol{\Gamma}$.

As a consequence, the following result is obtained

$$\begin{aligned}[\boldsymbol{F}_{\boldsymbol{\Sigma}_i}]_{kl} = -\frac{\partial^2\pi_i(\boldsymbol{\Sigma}_i)}{\partial\theta_k\partial\theta_l} =& \frac{\nu_i + R + 1}{2}\operatorname{Tr}\left(\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,k}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,l}\right) \\ &- \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Omega}_i[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,k}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,l}\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,l}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{i,k}\boldsymbol{\Sigma}_i^{-1}]\right).\end{aligned} \tag{4.16}$$

Expectations. In order to compute (4.14), the linearity property of the expectation is used,

$$\boldsymbol{F} = \mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{F}_{\boldsymbol{\theta}}\right] + \mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{F}_{\boldsymbol{\Sigma}_1} + \boldsymbol{F}_{\boldsymbol{\Sigma}_2}\right]. \tag{4.17}$$

Moreover, since $\pi_1(\boldsymbol{\Sigma}_1)$ does not depend on θ_k if $k \in \{\frac{p}{2}+1,\ldots,p\}$ and $\pi_2(\boldsymbol{\Sigma}_2)$ does not depend on θ_k if $k \in \{1,\ldots,\frac{p}{2}\}$, then $[\boldsymbol{F}_{\boldsymbol{\Sigma}_i}]_{kl} = 0$ if $k \in \{1,\ldots,\frac{p}{2}\}$ and $l \in \{\frac{p}{2}+1,\ldots,p\}$, and viceversa. As a consequence,

$$\boldsymbol{F}_{\boldsymbol{\Sigma}_1} + \boldsymbol{F}_{\boldsymbol{\Sigma}_2} = \begin{pmatrix} -\dfrac{\partial^2 L(\boldsymbol{\Sigma_1})}{\partial\boldsymbol{\theta}_{1:\frac{p}{2}}\partial\boldsymbol{\theta}_{1:\frac{p}{2}}^T} & \boldsymbol{O} \\ \boldsymbol{O} & -\dfrac{\partial^2 L(\boldsymbol{\Sigma_2})}{\partial\boldsymbol{\theta}_{\frac{p}{2}+1:p}\partial\boldsymbol{\theta}_{\frac{p}{2}+1:p}^T} \end{pmatrix} \tag{4.18}$$

and thus the expectation $\mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{F}_{\boldsymbol{\Sigma}_1}+\boldsymbol{F}_{\boldsymbol{\Sigma}_2}\right]$ reduces to determining the expectation of $-\dfrac{\partial^2\pi_1(\boldsymbol{\Sigma}_1)}{\partial\boldsymbol{\theta}_{1:\frac{p}{2}}\partial\boldsymbol{\theta}_{1:\frac{p}{2}}^T}$ w.r.t. $\boldsymbol{\Sigma}_1$ and the expectation of $-\dfrac{\partial^2\pi_2(\boldsymbol{\Sigma}_2)}{\partial\boldsymbol{\theta}_{\frac{p}{2}+1:p}\partial\boldsymbol{\theta}_{\frac{p}{2}+1:p}^T}$ w.r.t. $\boldsymbol{\Sigma}_2$. Both expectations have a closed-form expression that can be determined using the matrix expectations (4.11) and (4.12).

The challenge here is to compute the expectation $\mathbb{E}_{\boldsymbol{\Gamma}}\left[\boldsymbol{F}_{\boldsymbol{\theta}}\right]$, which involves calculating the expectation of the expression $\boldsymbol{R}_s^{-1}\boldsymbol{B}_{s,k}\boldsymbol{R}_s^{-1}\boldsymbol{B}_{s,l}$ w.r.t. $\boldsymbol{\Gamma}$ for all $s=1,\ldots,M$. This computation is possible provided we can compute the expectation

$$\mathbb{E}_{\boldsymbol{\Gamma}}\Big[\left(a\boldsymbol{\Sigma}_1+b\boldsymbol{\Sigma}_2\right)^{-1}\boldsymbol{A}\left(a\boldsymbol{\Sigma}_1+b\boldsymbol{\Sigma}_2\right)^{-1}\boldsymbol{B}\Big], \tag{4.19}$$

for $a,b\in\mathbb{R}_+$, $\boldsymbol{\Sigma}_1\sim\mathcal{IW}(\nu_1,\boldsymbol{\Omega}_1)$, $\boldsymbol{\Sigma}_2\sim\mathcal{IW}(\nu_2,\boldsymbol{\Omega}_2)$ and any pair of symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Note that, if a or b is equal to zero, (4.19) can be calculated using (4.12). Otherwise, we propose to approximate (4.19) numerically via a Monte Carlo algorithm. Given a, b, $\boldsymbol{A}$ and $\boldsymbol{B}$, we can generate a large number L_{mc} of samples $\{\boldsymbol{\Sigma}_1^{(\lambda)},\boldsymbol{\Sigma}_2^{(\lambda)}\}_{\lambda=1}^{L_{\mathrm{mc}}}$ according to inverse Wishart distributions, compute

$$\boldsymbol{E}^{(\lambda)}=\left[\left(a\boldsymbol{\Sigma}_1^{(\lambda)}+b\boldsymbol{\Sigma}_2^{(\lambda)}\right)^{-1}\boldsymbol{A}\left(a\boldsymbol{\Sigma}_1^{(\lambda)}+b\boldsymbol{\Sigma}_2^{(\lambda)}\right)^{-1}\boldsymbol{B}\right]$$

and approximate (4.19) by the average of $\{\boldsymbol{E}^{(\lambda)}\}_{\lambda=1}^{L_{\mathrm{mc}}}$.

Finally, the inverse of the PFIM (4.13), denoted as $[\boldsymbol{F}]^{-1}$, yields the desired lower bound for the MSE of any estimator $\widehat{\boldsymbol{\Gamma}}$ of $\boldsymbol{\Gamma}$, when $\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2$ are IW distributed:

$$\mathrm{MSE}\geq\mathrm{BB}=\mathrm{Tr}([\boldsymbol{F}]^{-1}). \tag{4.20}$$

4.4.1 Properties of the bounds

Assuming that a or b are zero and $\boldsymbol{\Delta}$ is diagonal, (4.20) can be computed in closed-form. In particular, the entries of $\boldsymbol{F}$ are given by

$$[\boldsymbol{F}]_{kl}=\gamma_1\,\mathrm{Tr}(\boldsymbol{\Delta}\boldsymbol{J}_k)\,\mathrm{Tr}(\boldsymbol{\Delta}\boldsymbol{J}_l)+\gamma_2\,\mathrm{Tr}(\boldsymbol{\Delta}\boldsymbol{J}_k\boldsymbol{\Delta}\boldsymbol{J}_l), \tag{4.21}$$

where $\gamma_1=\frac{1}{2}(\nu^2+\nu(2M-R+3))$ and $\gamma_2=\frac{1}{2}(\nu^3+\nu^2(2M-R+6)+\nu(2M+9-R))$. For $\boldsymbol{\Delta}$ the identity matrix, the contribution of the first term $\mathrm{Tr}(\boldsymbol{\Delta}\boldsymbol{J}_k)\,\mathrm{Tr}(\boldsymbol{\Delta}\boldsymbol{J}_l)$ to the PFIM is a matrix with zero entries except for an $R\times R$ diagonal block with non-zero entries, that of the second term

$\mathrm{Tr}(\boldsymbol{\Delta J}_k \boldsymbol{\Delta J}_l)$ is a diagonal matrix. It can be shown that the PFIM has $R-1$ eigenvalues equal to γ_2, $R(R+1)/2 - R = (R^2-R)/2$ eigenvalues equal to $2\gamma_2$ and one eigenvalue equal to $R\gamma_1 + \gamma_2$. The trace of $[\boldsymbol{F}]^{-1}$ with the above assumptions gives an approximation of the BB (4.20),

$$\mathrm{aBB} = \mathrm{Tr}([\boldsymbol{F}]^{-1}) = \frac{R(R-1)}{4\gamma_2} + \frac{R((R-1)\gamma_1 + \gamma_2)}{\gamma_2^2 + R\gamma_1\gamma_2}. \tag{4.22}$$

This shows that the bound behaves asymptotically as:

- $O(\nu^{-3})$ as $\nu \to +\infty$
- $O(\nu^{-2})$ as $\nu \to R+1$
- $O(R)$ as $R \to +\infty$
- $O(R^2)$ as $R \to 1$
- $O(M^{-1})$ as $M \to +\infty$.

4.5 Numerical experiments

In this section, we use extensive numerical simulations to study the properties of the CRB and BB for the MSE for any estimator of $\boldsymbol{\Gamma} = (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ in the i) deterministic and ii) probabilistic frameworks, and compare the bounds against the MSE of the maximum likelihood and Bayesian estimators.

4.5.1 Monte Carlo simulations

Estimation algorithms. For the deterministic CRB studied in Section 4.3, we consider the EM-MLE estimator proposed in Section 3.3.1. For the BB determined in Section 4.4, we consider the MMSE ($\mathrm{IW_A}$) and EM-MAP estimators, proposed in Sections 2.3.2 and 3.3.2.

Simulation setup. Unless otherwise stated $R = 2$, $M = 2^8$, $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2 = \mathbb{I}_R$ ($R \times R$ identity matrix) and $\nu_1 = \nu_2 = 80$. Without loss of generality, we use the functions $g_{1,s} = 2\pi \cos^2(\boldsymbol{x}[s]) + 0.1$ and $g_{2,s} = 2\pi \sin^2(\boldsymbol{x}[s]) + 0.1$, where $\boldsymbol{x}$ is the vector of M components whose values have been generated in the interval $[0, 2]$, equi-spaced with a distance of $2/(M-1)$. In all cases, we compute

the sample MSE of the estimators as the average of the trace of the error covariance matrix over 1000 independent realizations. Gibbs samplers are run with $N_{\text{mc}} = 1000$ and $N_{\text{bi}} = 500$, and (4.19) is approximated where needed as described using $L_{\text{mc}} = 200$.

4.5.2 Performance analysis

a) Performance vs. sample sizes

Fig. 4.1 compares the CRB and the MSE of the MLE. We also display a function proportional to $1/M$ indicating the expected asymptotic behavior. The following comments are appropriate:

1) The CRB decreases almost linearly when the sample size increases.

2) The MSE is very close to the CRB for large values of M. Similar results obtained for other choices of deterministic $\mathbf{\Gamma}$ are not reproduced here because they lead to the same conclusions.

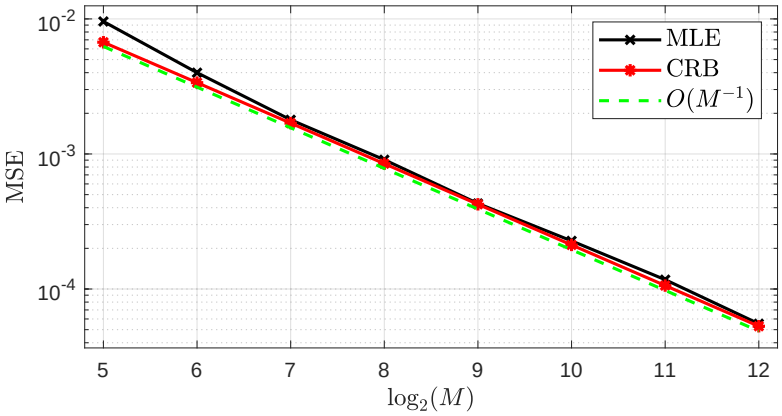

Figure 4.1: Comparison between the sample MSE of the MLE averaged over 1000 independent realizations versus the CRB for sample sizes $M \in \{2^5, 2^6, \ldots, 2^{12}\}$.

Fig. 4.2 displays the BB and its approximation aBB, and the MSE of the MMSE and MAP estimators, for various sample sizes, where $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are random matrices with inverse Wishart prior distributions. We observe that:

1) As expected, the BB decreases as M^{-1} when $M \to \infty$ and to a constant when $M \to 0$. The

approximation aBB is asymptotically similar but tends to a different constant for small sample sizes.

2) BB vs. MSE: The MSE of both estimators is approaching the BB when M increases - the more data, the tighter the bound.

3) MMSE vs. MAP: Overall, the MMSE estimator has better performance than the MAP estimator, in particular for small sample sizes. This result was expected since the MMSE estimator indeed minimizes the MSE.

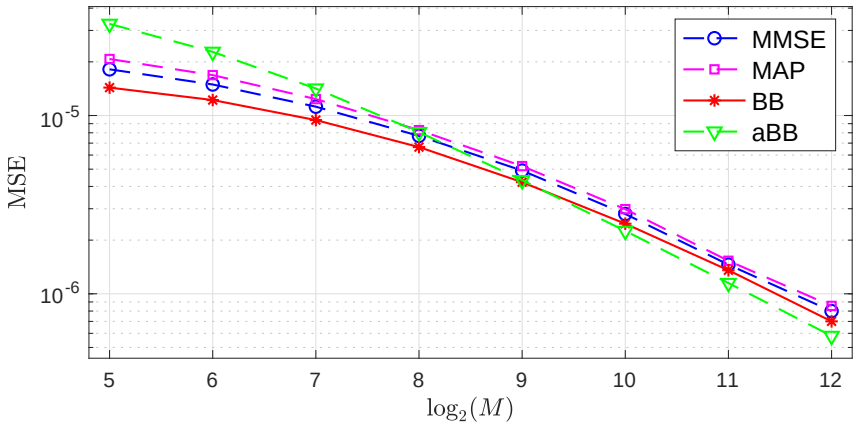

Figure 4.2: Comparison between the sample MSE of the MAP and MMSE estimators averaged over 1000 independent realizations versus the BB for sample sizes $M \in \{2^5, 2^6, \ldots, 2^{12}\}$, $\nu_1 = \nu_2 = 80$ and $R = 2$.

b) Performance vs. degrees of freedom

Fig. 4.3 compares the MSE of the MMSE and MAP estimators and the BB for various degrees of freedom, $\nu_1 = \nu_2 \in \{10, 15, 20, 25, \ldots, 120\}$. We can observe that the BB decreases when ν_1 and ν_2 increase. Indeed, in that case, the priors are more informative. The approximation aBB is very similar to BB and predicts that this decay is of order ν^{-3}. Moreover, the values of the MSE for both the MAP and MMSE estimators are observed to be significantly larger than the lower BB for small values for ν_1 and ν_2 (uninformative priors), but very close to the bound for large values of ν_1, ν_2 (informative priors).

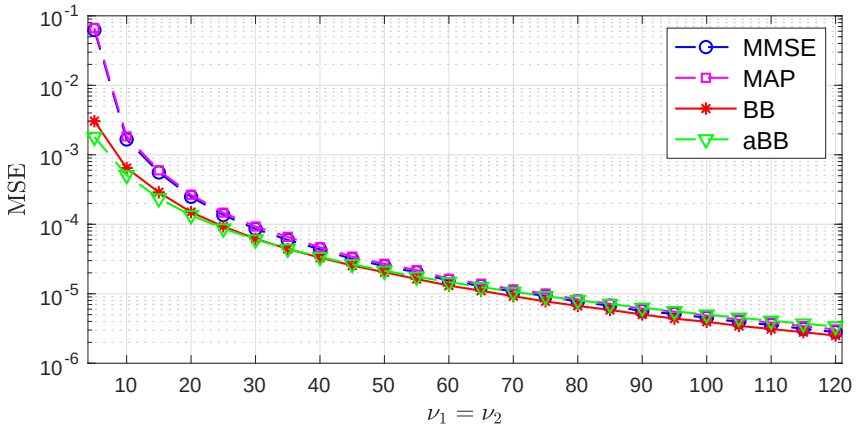

Figure 4.3: Comparison between the sample MSE of the MAP and MMSE estimators averaged over 1000 independent realizations versus the BB, varying the degrees of freedom $\nu_1 = \nu_2$, for $R = 2$ and $M = 2^8$.

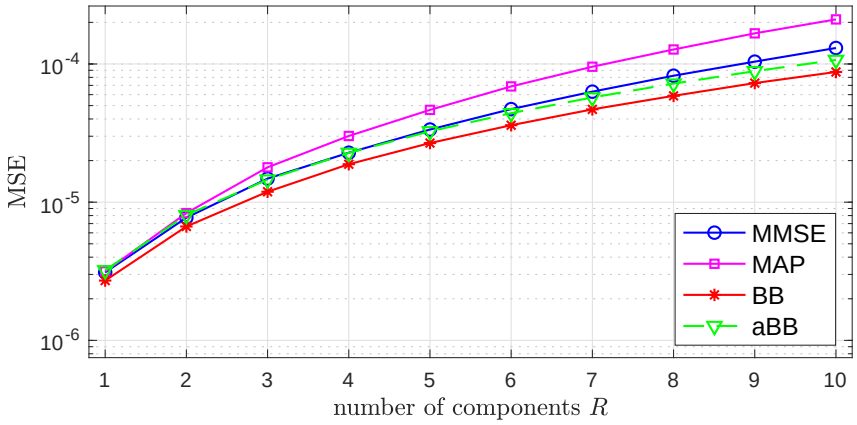

Figure 4.4: Comparison between the sample MSE of the MAP and MMSE estimators averaged over 1000 independent realizations versus the BB, varying the number of components R, for $M = 2^8$, $\mathbf{\Omega}_1 = \mathbf{\Omega}_2 = \mathbb{I}_R$ and $\nu_1 = \nu_2 = 80$.

c) Performance vs. numbers of components

Fig. 4.4 displays the MSE of the MMSE and MAP estimators and the BB when the number of components R - thus, the number of parameters p - is varied, specifically $R \in \{1, 2, 3, \ldots, 10\}$. We can observe that:

1) BB vs. R: The BB increases with increasing values for R and is very tightly approximated by

aBB, thus suggesting an asymptotically linear behavior in R.

2) BB vs. MSE: The values taken by the MSE and the BB are very similar for small numbers of components/parameters. For large values of R, the bound is slightly less tight. Since the sample size is fixed here, this behavior is coherent. Indeed, we would expect that larger sample sizes are required to converge to the BB when more parameters are estimated.

Similar results were obtained when $\mathbf{\Delta}_i$, $i = 1, 2$, is considered as a non-diagonal matrix leading to the same conclusions as above.

4.5.3 Application to a multivariate multifractal analysis

In this section, we apply the theoretical results developed in the above sections to a practical example related to the statistical model introduced in Chapter 2 for multivariate MFA. The Fourier transform of the log-leaders approximately obeys the data model considered in Section 4.2, where the elements of $\mathbf{\Sigma}_1$ are directly related to the multifractality of the data. Specifically, for a bivariate time series ($R = 2$), $\mathbf{\Sigma}_1 = -[c_2(1,1),\ c_2(1,2);\ c_2(1,2),\ c_2(2,2)]$ and thus $\boldsymbol{\theta}_{1:3} = -(\ c_2(1,1),\ c_2(1,2),\ c_2(2,2))$, where $c_2(1,1), c_2(2,2) < 0$ are related to the widths of the marginal multifractal spectra, and $c_2(1,2)$ quantifies the joint multifractality. The matrix $\mathbf{\Sigma}_2$ is an adjustment parameter that essentially subsumes the short-lag auto-correlation of log-leaders. The BB for the multifractal correlation parameter $\rho_{\mathrm{mf}}(1,2)$ defined in (1.23), has not been derived above. However, it is obtained from the BB for $\mathbf{\Sigma}_1$ based on functional invariance [Tre01].

Simulation study. We generate 2000 independent copies of $2^{10} \times 2$ time series of a canonical multifractal model process described in Section 1.4.1 to compute the sample MSE of the MMSE ($\mathrm{IW_A}$) estimator, and the BB, for different multifractal parameter settings, controlled by $\mathbf{\Sigma}_1$. The estimation is conducted for a single-scale $j = 2$ because the Bayesian model in Section 2.2.1 does not take into account the dependence between scales.

In a first experiment, $\mathbf{\Sigma}_1$ is generated using $\mathbf{\Omega}_1 = [0.5,\ 0;\ 0,\ \omega]$, with $0.37 \leq \omega \leq 1.2$ and $\nu_1 = 10$, leading to realistic expected values for the multifractal parameters, i.e., $-c_2(1,1) = 0.05$ and $-c_2(2,2) \in [0.037, 0.12]$. In a second experiment $\mathbf{\Omega}_1 = [0.4,\ \gamma;\ \gamma,\ 0.4]$ with γ tuned such that

$0 \leq \rho_{\mathrm{mf}}(1,2) \leq 0.8$ in average. The parameters of $\mathbf{\Sigma}_2$ cannot be controlled by the 2-variate MRW synthesis and are thus unknown, and we set $\mathbf{\Omega}_2 = \mathbf{\Omega}_1$ and $\nu_2 = \nu_1$.

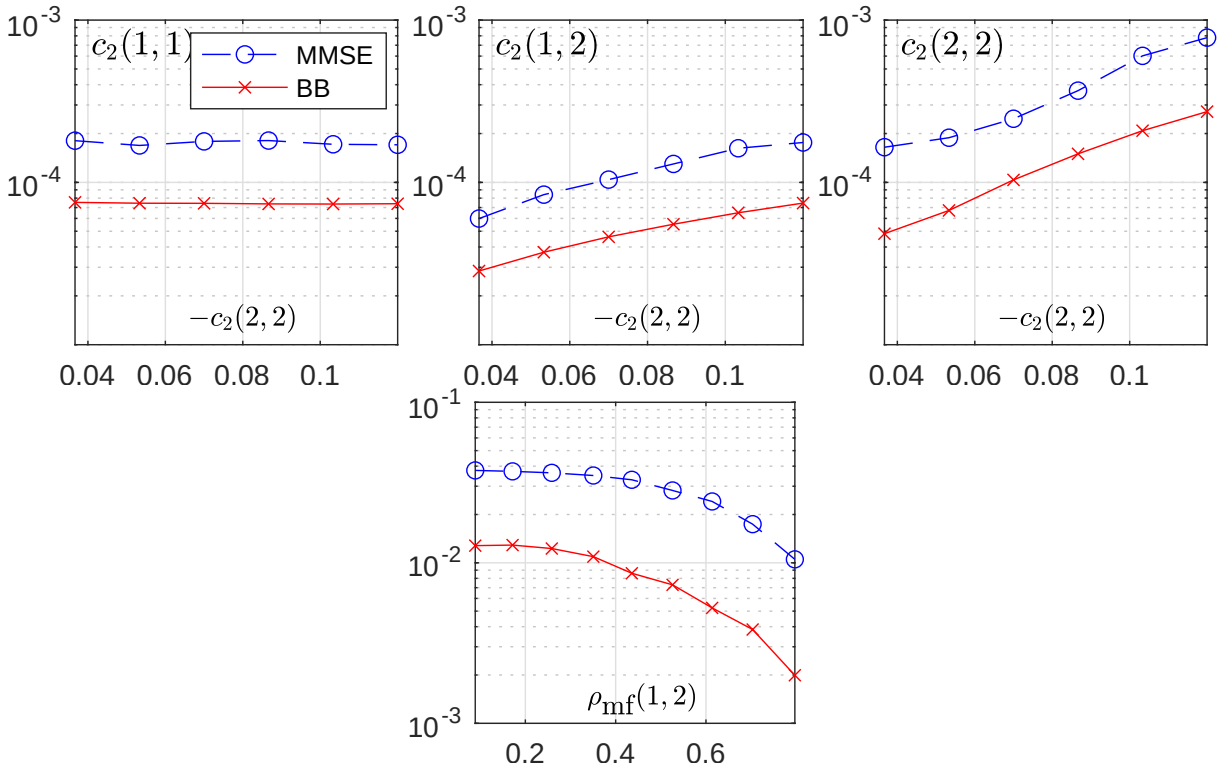

Figure 4.5: Sample MSE for multifractal parameters $c_2(1,1)$, $c_2(1,2)$, $c_2(2,2)$ as a function of $c_2(2,2)$ (top) and sample MSE for multifractal correlation $\rho_{\mathrm{mf}}(1,2)$ (bottom).

Results for the two experiments are presented in Fig. 4.5 (top and bottom row, respectively). They indicate that the derived BB provide good indications for the variations of the observed MSE of the multifractal parameter estimates. In particular, they show that:

1) The MSE of the estimator of $c_2(1,1)$ does not depend on $c_2(2,2)$, which is to be expected because $c_2(1,1)$ corresponds to a marginal parameter of the first data component that is independent of $c_2(2,2)$.

2) The MSE of the estimator of $c_2(2,2)$ increases with $c_2(2,2)$, indeed $c_2(2,2)$ controls the variance of the marginal likelihood of the second data component.

3) The MSE of the estimator of $c_2(1,2)$ also increases with $c_2(2,2)$ because $\rho_{\mathrm{mf}}(1,2)$ is held fixed so that $c_2(1,2)$ and thus the covariance also increase.

4) The MSE of $\rho_{\mathrm{mf}}(1,2)$ decreases in a non-trivial way when $\rho_{\mathrm{mf}}(1,2)$ increases.

Overall, the results show that the shape of the BB can predict the behavior of the MSE for the parameter estimates associated with the joint multifractal spectrum. The larger gap in the plots as compared to the results obtained on synthetic data, can potentially be explained by the fact that the model for the statistics of the log-leaders is not exact, and that the $\boldsymbol{\Sigma}_2$ parameter cannot be controlled in the experiments.

4.6 Conclusions and perspectives

Conclusions. This chapter derived and studied the Crámer-Rao lower bound and its Bayesian version for the MSE of estimators of the symmetric p.d. matrix-valued parameters of a zero mean circularly-symmetric complex Gaussian model with a covariance matrix structured as the sum of two covariance matrices (extensions to more than 2 summands are straight-forward). To calculate the Bayesian bound, a novel closed-form expression for a non-trivial expectation involving Wishart random matrices was provided. The properties of the bounds were studied analytically. Various numerical simulations were used to validate the theoretical results. Their use of the derived Bayesian bound was illustrated for the estimation performance of the parameters of the bivariate multifractal spectrum.

Perspectives. The expressions derived and studied in this chapter can be extended to more than 2 matrix summands and potentially used in other important contexts where Gaussian models with a zero mean vector and a covariance matrix $\boldsymbol{R}$ structured as above can arise, e.g., [DNR11, HDS+17, CRV+19]. The derivation of this type of bounds for the case when SIW priors are used, could also be carefully studied but it involves the calculation of non-trivial expectation. The above ideas are discussed in more detail in the concluding Chapter 6.

Chapter 5

Applications to multivariate real-world data

Contents

5.1 Introduction

In Chapters 2 and 3, we devised methods and algorithms capable of integrating multivariate multifractal analysis (MFA) into a Bayesian framework. In this chapter, we illustrate and investigate the potential use and benefits of such multivariate Bayesian multifractal methodologies for real-world data processing. The principal motivation here is to illustrate the contributions of the cross-multifractality and multifractal correlation parameters to enrich the data analysis.

In the past, fractal and multifractal approaches have been proposed and successfully used for different types of single-channel physiological signal processing in numerous contexts, see, e.g., [IAG+99, AMM+07, WCB+09, ZM13, CLCF18] to name but a few. Such data are usually part of multivariate datasets. However they have been analyzed individually rather than jointly. The algorithms developed in this thesis allow us to overcome this limitation. Specifically, in Section 5.2, we investigate the use of a joint MFA of a four-channel physiological signal to address the problem of drowsiness detection. We used the multifractal properties of the data as a feature vector in a devised detection scheme, and compared the detection accuracy based on a joint MFA against that obtained for exclusive use of univariate parameters.

In the context of image processing, fractal and MFA have also been widely used, especially for the modeling of textures associated with natural images [WRJA09]. As important examples, in recent works [CWT+15b, CWD+15, WCA+18], Bayesian univariate multifractal analyses have been put to test on real-world hyperspectral images. Their results evidenced that multifractal parameters associated with univariate spectra can provide relevant spatial/textural attributes in the context of hyperspectral imagery. Moreover, these works also suggested that the combination of both spectral and spatial information can improve the performance in classical hyperspectral image processing tasks, such as classification [FTB+13, RAAF10], segmentation [GRR+09] or endmember identification [MP12]. Inspired by those works, in this chapter we also investigate for the first time the use of the proposed multivariate MFA in the context of satellite imagery. Specifically, in Section 5.3, we propose to use the multifractal parameters to extract spatial information of a multispectral image in terms of the fluctuations of the local regularity of the image amplitudes and the characterization of

these fluctuations between different bands. The computed multifractal features are then used in a spatial clustering algorithm and compared against the exclusive use of simple pixel intensity-based features.

The results shown in this chapter overall suggest the following comments. First, the Bayesian methodology presented in this thesis is an operational tool for multivariate MFA applicable to real-world multivariate data processing. Second, it can provide new relevant attributes in biomedical and remote sensing contexts, which could in turn be employed in tasks such as classification, image segmentation or data mining.

To the best of our knowledge, this is the first time that a joint MFA is performed and investigated on such real-world datasets. Therefore, this constitutes in itself the main contribution of this chapter. The presented results allow us to arrive at first conclusions and to visualize the next directions of investigation, but nevertheless remain on a preliminary analysis. In particular, the definition of a precise methodology for the incorporation of multifractal features in multispectral image processing algorithms is not discussed here.

Results presented in this chapter have been partly reported in the journal paper [LWTA22a].

5.2 Multichannel polysomnographic data

5.2.1 Context and motivation

In this section, we consider the problem of detecting drowsiness, defined as an intermediate state between awake and sleep [YPLJ18], from several light non-invasive modalities related to the cardio-vascular, 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 [SSM14, WWF18, ANJ+16, BWZ+19, ACC20].

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 rate (HR) [IAG+99, CLCF18, WAK+19] but also for electroencephalogram (EEG) [WCB+09, ZM13, MNWL06, CLCF18], blood pressure (BP) [CLCF18, AMM+07] and respiration (RESP) [MNWL06, AMM+07] recordings.

The Bayesian multivariate MFA approaches proposed in this thesis allow us to go beyond a component-wise analysis and to investigate with robust algorithms that also quantify the multifractal dependencies between components, such datasets in which the signals are recorded jointly and simultaneously. To the best of our knowledge, this is the first time that a joint multivariate analysis of such data has been performed. In what follows, we describe the data and the analysis scenario (Section 5.2.2), and propose and study a drowsiness detection scheme based on the multivariate multifractal characteristics of the data (Section 5.2.3).

5.2.2 Data and preprocessing

MIT-BIH dataset. The dataset used for this study is extracted from the MIT-BIH Polysomnographic database[1], which involves a collection of recordings of multiple physiological signals, including HR, BP, EEG and RESP, acquired during sleep at a sample rate of 250 Hz [GAG+00]. The recordings come with manual annotations for sleep stages for each 30 second segment. We consider here the states "awake" vs. "sleep stage 1" (or drowsy stage) since the transition to the latter is considered as drowsiness.

Preprocessing. All the 18 available four-channel 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 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 ($\gamma = 2$ in (1.24)) to avoid negative uniform regularity issues (cf. Section 1.3.3).

5.2.3 Multivariate multifractal analysis for drowsiness detection

Multifractal analysis. The multivariate MFA was performed using 75% overlapping windows of sample size $N = 480$, yielding a set of multifractal parameter estimates for each 30 second interval

[1] https://physionet.org/content/slpdb/1.0.0/

(i.e., for each annotation). We use the MMSE estimator associated with an SIW prior (cf. Section 2.3) and scales $j = [\![3, 6]\!]$ (equivalently, 2.6s−21s), with parameters set to $N_\psi = 3$, $N_{\mathrm{mc}} = 1000$, $N_{\mathrm{bi}} = 500$.

In total, 2381 and 561 examples are available for the awake and sleep 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 $\rho_{\mathrm{mf}}(r, r')$ (with $r \neq r'$) were estimated. The values of $c_1(r)$ were estimated using standard linear regression as defined in (1.26). 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 $\{c_2(r, r')\}_{r\neq r';r,r'=1}^4$ or $\{\rho_{\mathrm{mf}}(r, r')\}_{r\neq r';r,r'=1}^4$, as joint multifractality estimates. 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 $\rho_{\mathrm{mf}}(1, 3)$ (HR and EEG channels) are provided in Fig. 5.1.

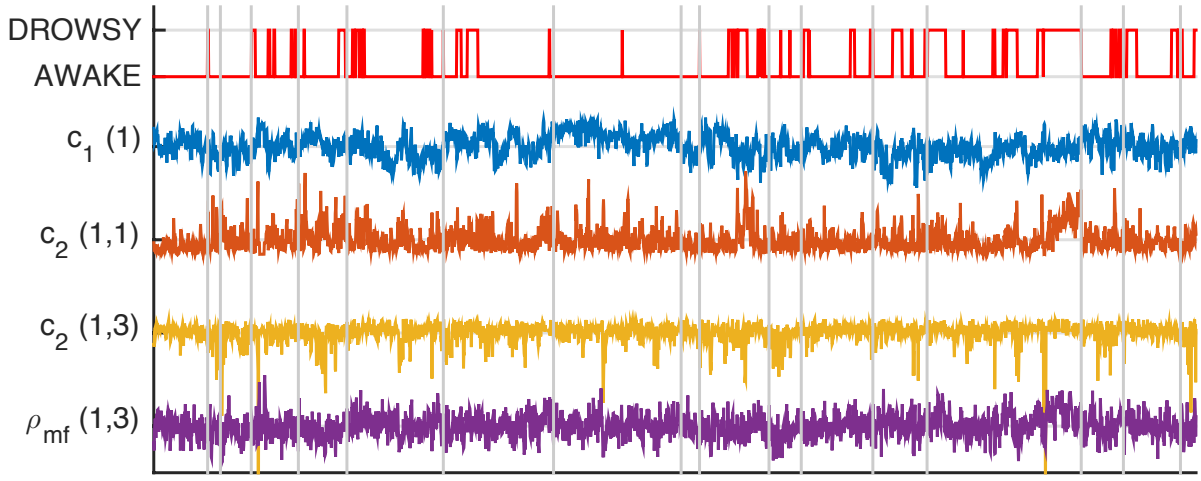

Figure 5.1: Visualization of awake and drowsy stages and multifractal estimates.

Detection performance. Drowsiness detection was performed using a random forest classifier with 50 trees that were trained on random subsets of 80% of the available examples (class 0 corresponds to the awake state and class 1 to sleep stage 1).

The detection performance was tested on the remaining 20% of the database. The reported results are averages over 25 different random subsets. The obtained classification and detection performance are quantified using 2-class accuracies as in [ACC20], and F-measure and area-under-curve (AUC) for the receiver operational characteristics (ROCs), as reported in Table 5.1. To further illustrate

the detection performance, the ROCs are also displayed in Fig. 5.1. The ROCs are computed by varying the relative weight for the "awake" and "sleep 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 $\rho_{\mathrm{mf}}(r, r')$ 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 $\rho_{\mathrm{mf}}(r, r')$ 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 [ACC20] (classification accuracy of 93%). Overall, these results demonstrate the robustness and relevance of the proposed joint estimation framework for the analysis of real-world data.

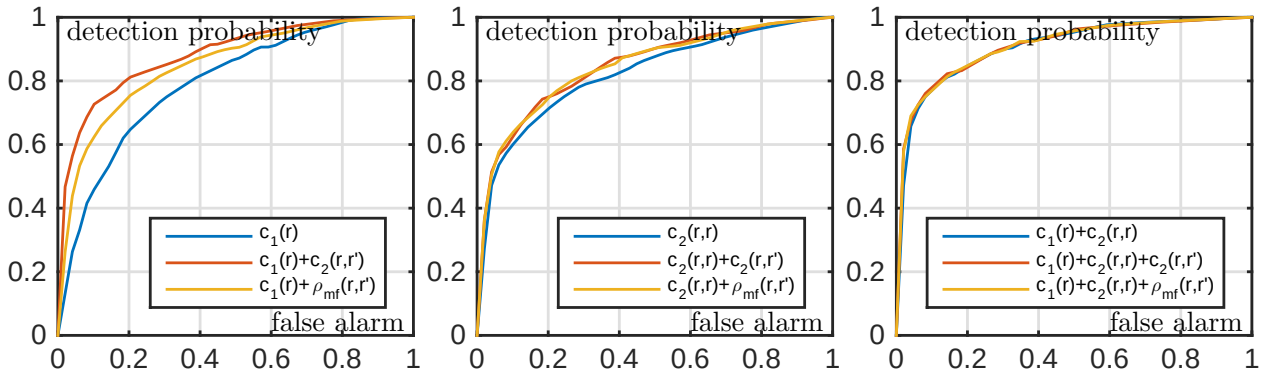

Figure 5.2: Drowsiness detection performance (ROCs, bottom): $c_1(r)$, $c_2(r, r)$ and $(c_1(r), c_2(r, r))$ (blue solid lines in left, center, right plot, respectively) and the corresponding curve when $c_2(r, r')$ (red solid line) or $\rho_{\mathrm{mf}}(r, r')$ (yellow solid line) is used as an additional feature.

Table 5.1: 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**

5.3 Multispectral satellite image

5.3.1 Context and motivation

Nowadays, remote sensing applications are of great importance for ecology, agriculture, defense, natural disaster forecasting, etc. Many of the images used in these contexts are multivariate, such as multispectral images that typically have between 3 to 15 spectral bands, hyperspectral images that typically consist of hundreds of contiguous spectral bands jointly registered, multitemporal images in which the same scene is registered at different times, or combinations thereof. Several attempts have been made for the use of fractal and multifractal concepts for these images, see, e.g., [SXGL06, KWD+20] for recent reviews.

The use of Bayesian models and estimators for univariate MFA purposes was already tested on real-world hyperspectral images, see, e.g., [CWT+15b, CWD+15, WCA+18]. Their results evidenced that multifractal parameters associated with univariate spectra can provide relevant spatial/textural attributes in the context of hyperspectral imagery. Inspired by those works, in this section we investigate for the first time the use of the proposed multivariate MFA in the context of satellite imagery. The study focuses on investigating the potential benefit in capturing complementary information using multivariate multifractal parameters as compared to traditional pixel intensity-based features.

5.3.2 Data and preprocessing

The dataset used for our study is a real-world multispectral satellite image. This image contains around 100 million pixels and includes four multispectral bands ($R = 4$): blue ($r = 1$), red ($r = 2$) and green ($r = 3$) color bands, and near infrared ($r = 4$), and is depicted in Fig. 5.3 in the RGB color space. It was provided by the CNES from Toulouse. We focus in particular on the area of size 300×300 pixels indicated by a red frame in Fig. 5.3 and depicted in Fig. 5.4 using gray scale.

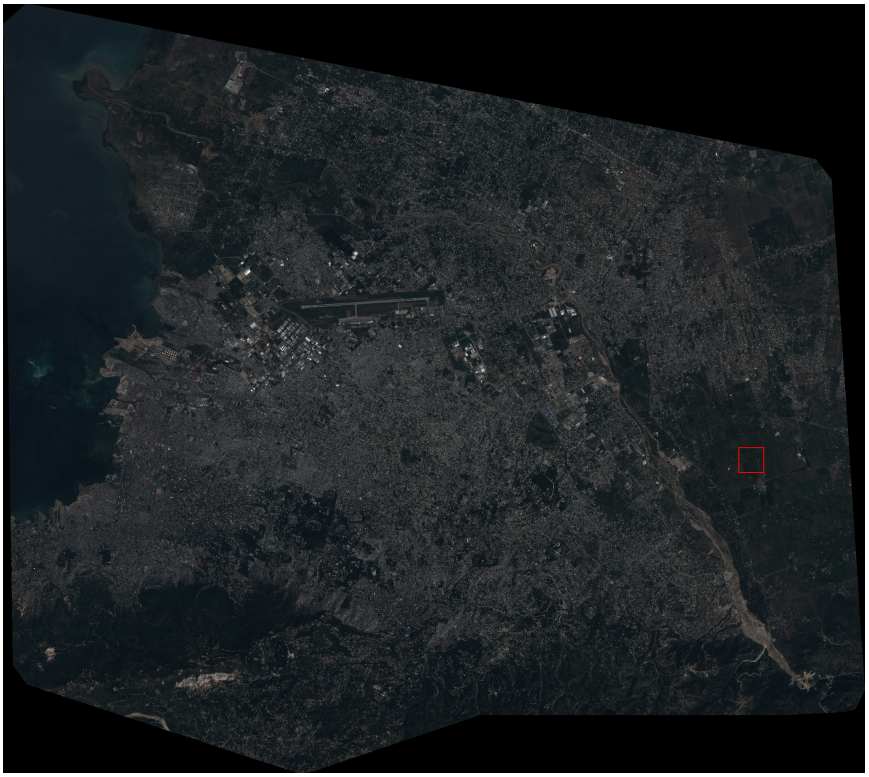

Figure 5.3: Real-world multispectral satellite image provide by the CNES from Toulouse

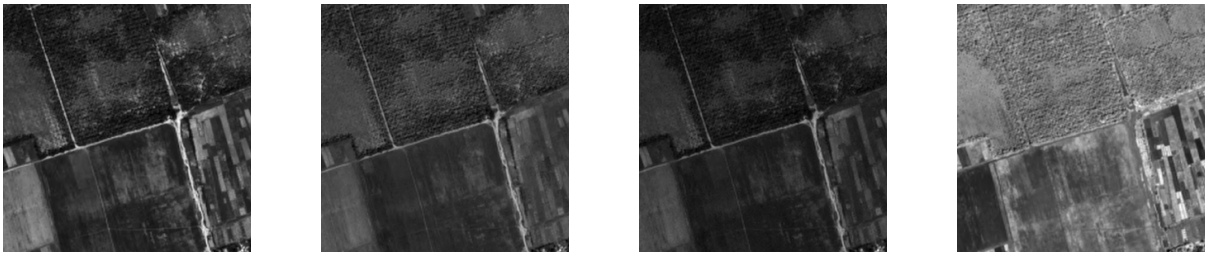

Figure 5.4: The blue, red, green and near infrared bands of the patch under analysis (gray scale).

5.3.3 Multivariate multifractal analysis for satellite imagery

Analysis scenario. In an attempt to increase the spatial resolution, we propose to conduct the analysis on small patches of the multispectral images in order to illustrate and study the characterization enabled by the multifractal properties throughout space and bands. Thus, each band is decomposed into 91^2 patches of size 30×30 pixels, with 90% overlap, resulting in a decomposition into patches of size $30 \times 30 \times 4$. This choice is motivated by the desire to obtain results as close as possible to the pixel domain to allow a better interpretation of the extracted spatial information. Note that this size is extremely small for MFA purposes where classical linear regression for example cannot be used. Dealing with such a small sample size is only possible using the proposed statistical model and algorithms.

For each patch, we use the MMSE estimator associated with an SIW prior (cf. Section 2.3) and scale $j = 1$, with parameters set to $N_\psi = 1$, $N_{\mathrm{mc}} = 4000$, $N_{\mathrm{bi}} = 2000$. The multifractal correlation $\rho_{\mathrm{mf}}(r, r')$ is estimated using (1.23) and estimates $\{\hat{c}_2(r, r')\}_{r,r'=1}^{4}$. For comparison, we compute the linear correlation $\rho(r, r')$ of each patch between pairs of bands.

Multifractal correlation vs. linear correlation. An illustration of the estimates of the multifractal correlation $\rho_{\mathrm{mf}}(r, r')$ and the linear correlation $\rho(r, r')$ are provided in Fig. 5.5 for $(r, r') = (1, 4)$ (top row) and $(r, r') = (2, 3)$ (bottom row). One can observe that ρ_{mf} estimates reproduce better the spatial structures of the image texture such as the paths between parcels of land. Similar result and conclusion were obtained for the rest of band combinations. These preliminary results suggest that the multifractal correlation is a spatial attribute that conveys complementary information that differs from, for example, the standard linear correlation. Thus, it could be incorporated, for instance, into an edge detection or segmentation scheme based on the characteristics of the multifractal spectrum to identify the roads for example.

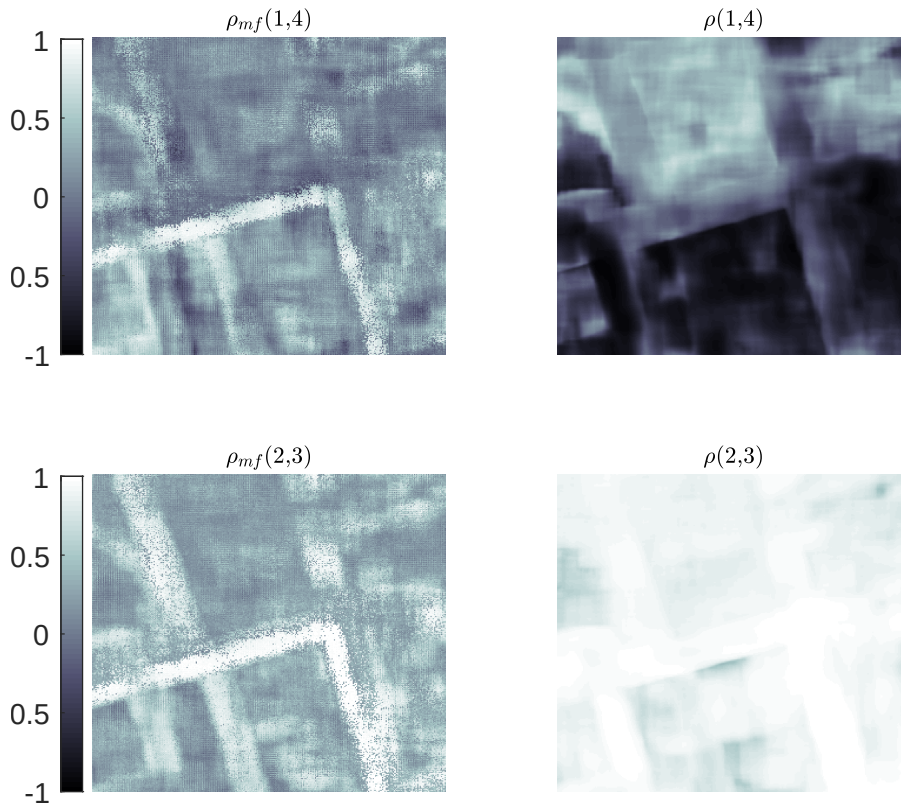

Figure 5.5: Estimates of the multifractal correlation $\rho_{\mathrm{mf}}(r, r')$ (left) and the linear correlation $\rho(r, r')$ (right) for $(r, r') = (1, 4)$ (top row) and $(r, r') = (2, 3)$ (bottom row).

Multifractal segmentation. Based on the conclusions drawn in the previous experiment, it is natural to seek to use the multivariate multifractal properties in a patch-wise classification scheme. We propose to use a simple k-means algorithm, see, e.g., [Mac67, Boc08]. The problem consists in partitioning the 91^2 feature vectors (one per patch) into k clusters in which each vector belongs to the cluster with the nearest mean. Specifically, we gather all multifractal estimates $\hat{c}_2$, $\hat{\rho}_{\mathrm{mf}}$ per patch into a vector with a total of 16 components and then use it, after normalization, as input to the k-means algorithm. As an illustration, in Fig. 5.6 (top row), results of the classification algorithm are shown for $k = 3$ (left), $k = 4$ (center) and $k = 5$ (right) clusters. Each color corresponds to a different class and the color scale changes with k. For this experiment, the ground truth is not available. Thus, we propose to compare these result against those obtained based on a normalized vector composed of classical features given by the average and standard deviation (STD) of the image amplitudes, and the linear correlation between different bands (14 features in total). Qualitatively, the segmentation obtained using the multivariate multifractal characteristics better preserves the spatial structure of the original image compared to the other approach. However the classification of the same kind of objects in the images is not consistent, for instance in the position of the roads.

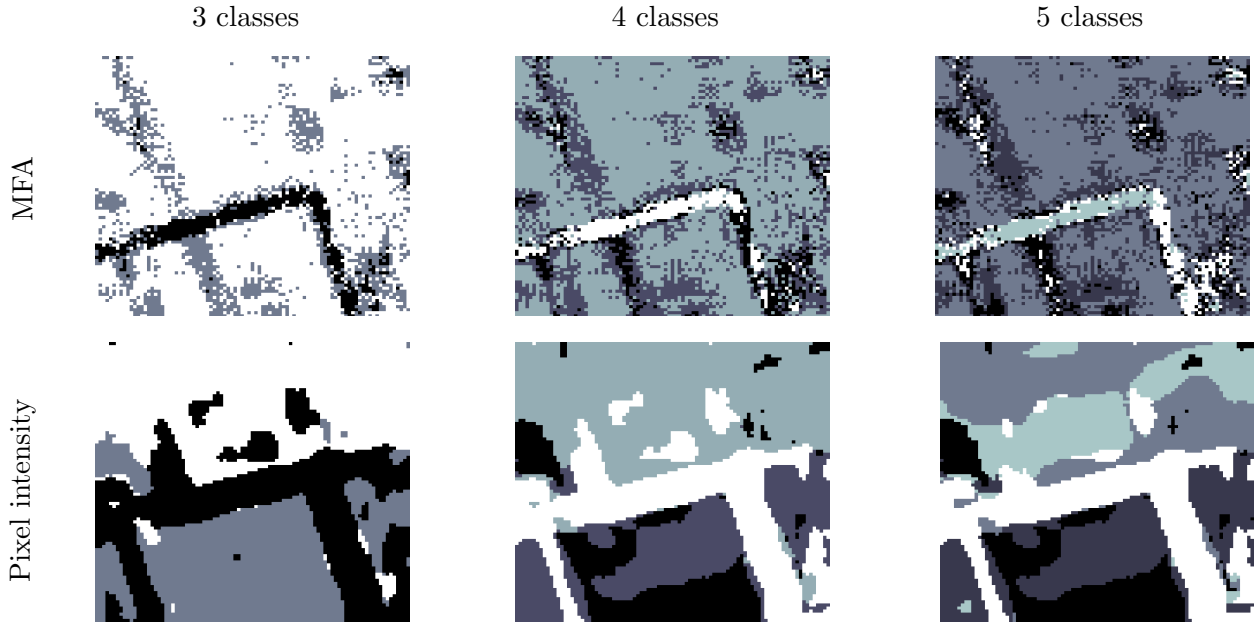

Figure 5.6: Segmentation obtained via k-means using second-order multifractality parameters and multifractal correlations estimates (top) and using the mean, STD and correlation of pixel intensity values (botton). Numbers of classes: $k = 3$ (left), $k = 4$ (center) and $k = 5$ (right).

5.4 Conclusions and perspectives

Conclusions. In this chapter, we proposed and investigated the use of a multivariate MFA for two different real-world multivariate datasets. In the analysis of the four-channel physiological signal, it was shown that the combination of the univariate multifractal properties with those characterizing the coupling between components improves the performance of the proposed drowsiness detection scheme compared to using only the univariate ones. The experiments conducted on the four-band satellite image illustrated the improvements of using the multifractal correlation parameter for the extraction of spatial/textural information with respect to the use of the standard correlation coefficient. Moreover, a qualitative analysis of the results indicated that the multivariate multifractal properties captured spatial information different from that given by the average, STD and the linear correlation of the images amplitudes, suggesting that they could be used in a complementary manner as spatial features to perform tasks such as data-mining, segmentation or classification. Overall, the results reported in this chapter enabled us to illustrate that the Bayesian methodology to estimate the parameters associated with the joint multifractal spectrum introduced in Chapter 2 is operational and relevant for the joint MFA of real-world multivariate signals and images.

Perspectives. Results reported in Chapter 5 constitute a preliminary illustration of the potential applications of the proposed Bayesian methodology for multivariate MFA of real-world data. A systematic study of practical multivariate MFA remains to be defined and tested on other datasets and applications, for example, in financial time series or hyperspectral images. Future research directions are discussed in more detail in Chapter 6.

Chapter 6

Conclusions and future work

Context

Multifractal analysis (MFA) is a powerful theoretical and practical tool for signal and image processing. It enables the characterization of data based on the dynamics of their local regularity and has found many successful applications of different natures in the past. However, these successes have been limited to the independent processing of individual data components, while in an increasing number of applications, the data to be analyzed are multivariate. The theoretical foundations of multivariate analysis and the definition of the multivariate multifractal spectrum, and its practical potential to capture higher-order transient dependencies between different components of the data, have only recently been settled. However, the accurate estimation of the multivariate multifractal parameters remains challenging, severely limiting their actual use in applications. To overcome these limitations, the primary objective of this thesis was to propose and study practical contributions to multivariate MFA of signals and images. To this end, the present thesis formulated a novel Bayesian framework for multivariate multifractal parameter estimation. In addition, the use of multivariate MFA in general and of these novel methods in particular was investigated in two different contexts: drowsiness detection from a multichannel physiological signals and potential remote sensing applications from a multispectral satellite image.

Conclusions

Chapter 1 presented the main theoretical and practical concepts of the multivariate MFA of signals and images. In practice, the estimation of the multivariate multifractal spectrum of the data is

the central object of interest of MFA and can be achieved via a multivariate multifractal formalism based on wavelet leaders. The multivariate multifractal parameters of interest considered here are the second-order log-cumulants c_2 (auto- and cross-multifractality coefficients) and the multifractal correlation ρ_{mf}. Their estimation relies on classical linear regression-based algorithms which suffer of several practical limitations. Therefore, the focus of this thesis is to devise accurate and efficient estimation algorithms for multivariate MFA of signals and images.

Chapter 2 introduced a novel and original joint Gaussian model for the log-leaders and leverages on a Whittle-based likelihood approximation and on data augmentation for the symmetric positive definite (p.d.) matrix-valued parameters of interest. This careful design enables efficient estimation procedures to be constructed in a Bayesian framework for two relevant choices of priors and two alternative ways of calculating the average of symmetric p.d. matrices. Algorithms based on Monte Carlo Markov Chain (MCMC) are designed and used to approximate the minimum mean square error (MMSE) estimator associated with the resulting posterior distributions. Monte Carlo simulations, conducted on synthetic multivariate signals and images with various sample sizes, numbers of components and multifractal parameter values, demonstrated significant performance improvements over the state of the art, at moderately increased computational cost only. The methods and algorithms developed in this chapter constitute the first operational tool for practical multivariate MFA of multivariate signals and images.

Chapter 3 proposed and studied Expectation-Maximization based algorithms to approximate the maximum likelihood and the maximum a posteriori estimators of the matrix-valued multifractal parameters of interest. These new approaches were built on the proposed statistical model of the log-leaders, with the motivation of having a reduced computational time. Using Monte Carlo simulations, their performance is assessed and compared against the MCMC-based estimators and the classical linear regression-based algorithm in terms of accuracy and computational time.

Chapter 4 derived and studied the Crámer-Rao and the Bayesian Crámer-Rao lower bounds of the

mean squared error of estimators of the matrix-valued parameters of the proposed statistical model. A novel closed-form expression for a non-trivial expectation involving Wishart random matrices was derived that is required for the calculation of the Bayesian bound. The properties of these bounds were analytically studied and also numerically investigated via Monte Carlo simulations. Finally, we illustrated the use of the proposed bounds for the estimation performance of the parameters associated with the bivariate multifractal spectrum.

Chapter 5 investigated for the first time the potential use and benefits of multivariate MFA and of the proposed Bayesian methodology in two applications of very different natures: i) drowsiness detection in multichannel physiological signals and ii) quantification of spatial/textural information in multispectral satellite images. In the analysis of the four-channel physiological signal, it was shown that the combination of the univariate multifractal properties with those characterizing the coupling between components improves the performance of the proposed drowsiness detection scheme compared to using only the univariate ones. The experiments conducted on the four-band satellite image illustrated the improvements of using the multifractal correlation parameter for the extraction of spatial/textural information with respect to the use of the standard correlation coefficient. These preliminary results indicated that the multivariate multifractal properties could potentially be beneficial for the multispectral satellite image processing. Finally, this chapter illustrated that the Bayesian methodology to estimate the parameters associated with the joint multifractal spectrum yielded robust estimation procedures that are operational and relevant for the joint MFA of real-world multivariate signals and images.

Perspectives and future work

At the end of this work, the following directions for future work can be formulated.

Model developments

To simplify the proposed model and reduce the computational cost, in Chapter 2, it is assumed that log-leaders at different scales are independent. However, this assumption is not realistic as interscale dependence of wavelet coefficients and the associated log-leaders is commonly reported in the literature, see, e.g., [BS99, LM01, SS02] and references therein. Therefore, the design of appropriate models based on the inter-scale dependence of log-leaders could be investigated in future work (see [ABM98] for results obtained for 1D random wavelet cascades process). These may requires further improvements in the estimation of the multivariate multifractal parameters.

The estimation of the first-order log-cumulant c_1 associated with the average regularity of each component of the data can be incorporated straightforwardly into the proposed statistical model and estimation framework, as in the univariate scenario, see, e.g., [Com16]. However, the incorporation of higher-order log-cumulants linked to the multivariate skewness and the kurtosis of the distribution of the log-leaders, is more challenging and potentially requires the use of non-Gaussian distributions as generative models, such as, for instance, the skew Student-t-normal distribution and its extensions, see, e.g., [CBP08, NO12].

Along a similar line, the model could be generalize to the use of p-exponent and p-leaders [JML+16, LWA+16], a versatile family of regularity exponents and multiresolution coefficients to which wavelet leaders belong as a special case ($p = +\infty$). Preliminary results indicate that log-p-leaders are in general non well modeled by multivariate Gaussians, thus require further developments beyond the model presented here.

Other prior distributions with more complex parametrization and greater flexibility have been proposed for covariance matrix estimation in the literature, such as, for instance, the hierarchical half-t prior, see, e.g., [ANS14, TGM+11] and reference therein. The investigation of such alternative priors could improve the multifractal parameter estimation accuracy, possibly at the cost of having

to deal with more difficult estimation problems.

In a different vein, it would be interesting to address multivariate MFA of collections of multivariate signals and images via the design of Bayesian hierarchical models incorporating multivariate priors. This could enforce a smooth estimation of the multifractal parameters as explored in the univariate setting, see [WCA+18]. These could be critical for future segmentation, edge detection and classification schemes in multivariate image applications. The study of such hyperpriors enabling the efficient coupling of matrix valued parameters remains as an open issue in itself.

In the spirit of [LPAT21], the EM strategies developed in Chapter 3 could be used to design new estimation algorithms robust to noise, outliers or other corruptions present in multifractal data.

Computational and algorithmic developments

The acceleration and improvement of the sampling algorithms developed in Chapter 2, could be investigated. For example, we could use strategies to recycle the samples in the algorithm resulting from using scaled inverse Wishart (SIW) priors similar to the approach proposed in [MECV18].

To expedite convergence and accuracy for the proposed EM algorithms as well as initialization strategies must be carefully study. A variety of heuristic or metaheuristic approaches exist to escape a local maximum, such as, for instance, random-restart hill climbing starting with several different random initial estimates, see, e.g., [NR02]. These EM methodologies could also be extended to the use of SIW priors, an additional challenge being to nevertheless achieve a reduced computational cost when the update of the parameters in the M-step does not have a closed-form expression.

Theoretical analysis of the estimators

The expressions derived and studied in Chapter 4 for the lower bounds can be extended for more than 2 matrix summands and potentially used beyond the multivariate MFA. Specifically, Gaussian models with a zero mean vector and a covariance matrix $\boldsymbol{R}$ structured as (4.1) can arise in other important contexts. One example is given by vector-valued additive Gaussian processes, in which the matrices $\boldsymbol{G}_1, \boldsymbol{G}_2$ subsume the kernels for the temporal/spatial isotropic covariance models, expressed

in the Fourier domain, and $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are the associated point covariance matrices for the vector-valued variates, see, e.g., [DNR11, HDS+17, CRV+19]. Thus, it would be interesting to study the performance of the derived bounds in such models. Another continuation of this work is the derivation of similar expressions when an SIW prior is used. However, this is challenging because the use of this prior induces a nonlinear parametrization of the matrices to be estimated, leading to non-closed-form expectations. This significantly complicates the derivation of analytical expressions for the lower bounds.

Applications

Results reported in Chapter 5 constitute a preliminary illustration of the potential applications of the proposed Bayesian methodology for multivariate MFA of real-world data. Its practical use on a broader scale will certainly require further research.

Methodologically, model selection/validation procedures are required to, for example, assess whether the log-leaders associated with the analyzed multivariate data are well described by the proposed statistical model and to determine the range of scales over which the model is valid. The latter is challenging since only limited results are available in the literature, even for the linear regression-based estimation (see, e.g., [VAT03, LTA14] for examples of procedures addressing the scaling range selection).

The work presented in Chapter 5 leads us to conclude that multivariate multifractal attributes could be considered for i) drowsiness detection using multivariate physiological signals, and ii) the extraction of textural information for multispectral satellite image processing applications. A systematic study of practical multivariate MFA remains to be defined and tested on other datasets and applications. In this sense, different perspectives need be explored, such as the investigation of different multifractal formalisms (e.g, using the most recent p-leaders) or of different attributes (higher-order log-cumulants), the incorporation of multifractal attributes in segmentation, detection, classification algorithms and the comparison against established methods.

Appendices

Appendix A

Proof of Proposition 1

The Proposition 1 introduced in Section 4.4 and recalled below, is proved in this section using the approach detailed in [GLM05].

Moments of the type $\mathbb{E}[\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{B}\boldsymbol{W}]$. If $\boldsymbol{\Sigma} \sim \mathcal{IW}(\nu, \boldsymbol{\Omega})$, then $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$ has the Wishart distribution $\mathcal{W}(\nu, \boldsymbol{\Delta} = \boldsymbol{\Omega}^{-1})$. Then, for any pair of real-valued symmetric matrices $(\boldsymbol{A}, \boldsymbol{B})$:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Sigma}}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{\Sigma}^{-1}\right] = {} & \boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}(\nu^3+2\nu^2+\nu) + \boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}(\nu^2+3\nu) \\ & + [\mathrm{Tr}\,(\boldsymbol{\Delta}\boldsymbol{A})]\boldsymbol{\Delta}\boldsymbol{B}\boldsymbol{\Delta}(\nu^2+\nu) + [\mathrm{Tr}\,(\boldsymbol{\Delta}\boldsymbol{B})]\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}(\nu^2+\nu) \\ & + \boldsymbol{\Delta}\left[(\nu^2+\nu)\,\mathrm{Tr}\,(\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{B}) + \nu\,\mathrm{Tr}\,(\boldsymbol{\Delta}\boldsymbol{A})\,\mathrm{Tr}\,(\boldsymbol{\Delta}\boldsymbol{B})\right]. \end{aligned} \tag{A.1}$$

Proof. The proof is based on Eq. (4.27) of the paper by Graczyk, Letac and Massam [GLM05] recalled below

$$\begin{aligned}\mathbb{E}\left[\mathrm{Tr}(\boldsymbol{S}\boldsymbol{m}_1\ \boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}\boldsymbol{m}_3^T)\right] = {} & p^3\mathrm{tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3^T) \\ & + \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3^T) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3^T) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_3^T)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2)\right] \\ & + \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2^T\boldsymbol{\sigma}\boldsymbol{m}_3^T) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3\boldsymbol{\sigma}\boldsymbol{m}_2^T)\right] \\ & + \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_3^T) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3)\right] \\ & + \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_3^T)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2^T)\right] \\ & + \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2^T\boldsymbol{\sigma}\boldsymbol{m}_3) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3^T\boldsymbol{\sigma}\boldsymbol{m}_2) + \mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3^T\boldsymbol{\sigma}\boldsymbol{m}_2^T)\right] \\ & + \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3\boldsymbol{\sigma}\boldsymbol{m}_2)\right]. \end{aligned} \tag{A.2}$$

First, we note that for any matrix $\boldsymbol{X}$

$$\mathbb{E}[\mathrm{Tr}(\boldsymbol{X}\boldsymbol{m}_3^T)] = \mathrm{Tr}(\mathbb{E}[\boldsymbol{X}]\boldsymbol{m}_3^T). \tag{A.3}$$

Thus, if we can express the left hand side of this equation as

$$\mathbb{E}[\mathrm{Tr}(\boldsymbol{X}\boldsymbol{m}_3^T)] = \mathrm{Tr}(\boldsymbol{M}\boldsymbol{m}_3^T),$$

by identification, we obtain $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{M}$. By setting $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{m}_1\boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}$ in (A.3), we obtain

$$\mathbb{E}[\mathrm{Tr}(\boldsymbol{S}\boldsymbol{m}_1\boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}\boldsymbol{m}_3^T)] = \mathrm{Tr}(\mathbb{E}[\boldsymbol{S}\boldsymbol{m}_1\boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}]\boldsymbol{m}_3^T).$$

The left hand side of this equation can be computed by Eq. (4.27) of the paper by Graczyk, Letac and Massam. The trick is to express the right hand side of (A.2) as $\mathrm{Tr}(\boldsymbol{M}\boldsymbol{m}_3^T)$ to obtain

$$\mathbb{E}[\boldsymbol{S}\boldsymbol{m}_1\boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}] = \boldsymbol{M}.$$

There are 15 terms in the right hand side of (A.2) that need to be expressed as $\mathrm{Tr}(\boldsymbol{M_k}\boldsymbol{m}_3^T)$ with $k = 1, ..., 15$. These terms are detailed below

1. Term #1

$$p^3\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3^T) = \mathrm{Tr}(\boldsymbol{M_1}\boldsymbol{m}_3^T),$$

   with

$$\boldsymbol{M_1} = p^3\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}.$$

2. Term #2

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}\boldsymbol{m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_2}\boldsymbol{m}_3^T),$$

   with

$$\boldsymbol{M_2} = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1)\right]\boldsymbol{\sigma}\boldsymbol{m}_2\boldsymbol{\sigma}.$$

3. Term #3

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}\boldsymbol{m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_3}\boldsymbol{m}_3^T),$$

   with

$$\boldsymbol{M_3} = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma}\boldsymbol{m}_2)\right]\boldsymbol{\sigma}\boldsymbol{m}_1\boldsymbol{\sigma}.$$

4. Term #4

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_3^T)\,\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2)\right] = \mathrm{Tr}(\boldsymbol{M_4 m}_3^T),$$

with

$$\boldsymbol{M_4} = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2)\right]\boldsymbol{\sigma}.$$

5. Term #5

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_3)\right] = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{m}_3^T\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma})\right] = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_5 m}_3^T),$$

with

$$\boldsymbol{M_5} = \frac{p^2}{2}\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma}.$$

6. Term #6

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_3^T)\right] = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_6 m}_3^T),$$

with

$$\boldsymbol{M_6} = \frac{p^2}{2}\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma}.$$

7. Term #7

$$\frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3\boldsymbol{\sigma m}_2^T)\right] = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3)\right] = \frac{p^2}{2}\left[\mathrm{Tr}(\boldsymbol{m}_3^T\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma})\right] = \mathrm{Tr}(\boldsymbol{M_7 m}_3^T),$$

with

$$\boldsymbol{M_7} = \frac{p^2}{2}\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma}.$$

8. Term #8

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_8 m}_3^T),$$

with

$$\boldsymbol{M_8} = \frac{p}{4}\,\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma m}_2)\boldsymbol{\sigma}.$$

9. Term #9

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_3)\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{m}_3^T\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma})\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_3^T)\right]=\mathrm{Tr}(\boldsymbol{M_9 m}_3^T),$$

with

$$\boldsymbol{M_9}=\frac{p}{4}\,\mathrm{Tr}(\boldsymbol{\sigma m}_1)\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma}.$$

10. Term #10

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3)\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2)\,\mathrm{Tr}(\boldsymbol{m}_3^T\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma})\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2)\,\mathrm{Tr}(\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma m}_3^T)\right]=\mathrm{Tr}(\boldsymbol{M_{10} m}_3^T),$$

with

$$\boldsymbol{M_{10}}=\frac{p}{4}\,\mathrm{Tr}(\boldsymbol{\sigma m}_2)\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma}.$$

11. Term #11

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_3^T)\,\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T)\right]=\mathrm{Tr}(\boldsymbol{M_{11} m}_3^T),$$

with

$$\boldsymbol{M_{11}}=\frac{p}{4}\,\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T)\boldsymbol{\sigma}.$$

12. Term #12

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_3)\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{m}_3^T\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma})\right]\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma m}_3^T)\right]=\mathrm{Tr}(\boldsymbol{M_{12} m}_3^T),$$

with

$$\boldsymbol{M_{12}}=\frac{p}{4}\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1^T\boldsymbol{\sigma}.$$

13. Term #13

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3^T\boldsymbol{\sigma m}_2)\right]=\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3^T)\right]=\mathrm{Tr}(\boldsymbol{M_{13} m}_3^T),$$

with

$$\boldsymbol{M_{13}}=\frac{p}{4}\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma}.$$

14. Term #14

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3^T\boldsymbol{\sigma m}_2^T)\right] = \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_{14}m}_3^T),$$

with

$$\boldsymbol{M_{14}} = \frac{p}{4}\boldsymbol{\sigma m}_2^T\boldsymbol{\sigma m}_1\boldsymbol{\sigma}.$$

15. Term #15

$$\frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_3\boldsymbol{\sigma m}_2)\right] = \frac{p}{4}\left[\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_3^T)\right] = \mathrm{Tr}(\boldsymbol{M_{15}m}_3^T),$$

with

$$\boldsymbol{M_{15}} = \frac{p}{4}\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2\boldsymbol{\sigma}.$$

After summing all these terms, the following result is obtained

$$\mathbb{E}\left[\mathrm{Tr}(\boldsymbol{Sm}_1\boldsymbol{Sm}_2\boldsymbol{Sm}_3^T)\right] = \mathrm{Tr}\left(\sum_{k=1}^{15}\boldsymbol{M_k m}_3^T\right),$$

hence

$$\mathbb{E}[\boldsymbol{Sm}_1\boldsymbol{Sm}_2\boldsymbol{S}] = \sum_{i=1}^{15}\boldsymbol{M_k}.$$

Using the fact that $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are symmetric matrices, straightforward computations lead to

$$\begin{aligned}\mathbb{E}[\boldsymbol{Sm}_1\boldsymbol{Sm}_2\boldsymbol{S}] = {} & \boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2\boldsymbol{\sigma}\left(p^3+p^2+\frac{p}{4}\right) + \boldsymbol{\sigma m}_2\boldsymbol{\sigma m}_1\boldsymbol{\sigma}\left(\frac{p^2}{2}+\frac{3p}{4}\right)\\ & + \boldsymbol{\sigma m}_2\boldsymbol{\sigma}\left(\frac{p^2}{2}+\frac{p}{4}\right)\mathrm{Tr}(\boldsymbol{\sigma m}_1) + \boldsymbol{\sigma m}_1\boldsymbol{\sigma}\left(\frac{p^2}{2}+\frac{p}{4}\right)\mathrm{Tr}(\boldsymbol{\sigma m}_2)\\ & + \boldsymbol{\sigma}\left(\frac{p^2}{2}\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2) + \frac{p}{4}\mathrm{Tr}(\boldsymbol{\sigma m}_1)\,\mathrm{Tr}(\boldsymbol{\sigma m}_2) + \frac{p}{4}\mathrm{Tr}(\boldsymbol{\sigma m}_1\boldsymbol{\sigma m}_2)\right). \qquad \text{(A.4)}\end{aligned}$$

The definition of the Wishart distribution in statistics is different from the definition used in the paper by Graczyk, Letac and Massam. Indeed, in the paper by Graczyk, Letac and Massam, a random matrix $\boldsymbol{S}$ is distributed according to a Wishart distribution with parameters p and $\boldsymbol{\sigma}$ if its Laplace transform is

$$\mathbb{E}\left[e^{-\mathrm{Tr}(s\boldsymbol{S})}\right] = \frac{1}{\det\left[\boldsymbol{I}+s\boldsymbol{\sigma}\right]^p}.$$

Conversely, in statistics, a random matrix $\boldsymbol{W}$ is distributed according to a Wishart distribution with parameters ν and $\boldsymbol{\Delta}$ if its Laplace transform is

$$\mathbb{E}\left[e^{-\operatorname{Tr}(s\boldsymbol{W})}\right] = \frac{1}{\det\left[\boldsymbol{I} + 2s\boldsymbol{\Delta}\right]^{\frac{\nu}{2}}}.$$

To use an expression in the paper by Graczyk, Letac and Massam that is compatible with our definition, one has to replace $\boldsymbol{\sigma}$ by $2\boldsymbol{\Delta}$ and p by $\frac{\nu}{2}$. These changes lead to

$$\begin{aligned}\mathbb{E}[\boldsymbol{S}\boldsymbol{m}_1\boldsymbol{S}\boldsymbol{m}_2\boldsymbol{S}) =& \boldsymbol{\Delta}\boldsymbol{m}_1\boldsymbol{\Delta}\boldsymbol{m}_2\boldsymbol{\Delta}\left(\nu^3 + 2\nu^2 + \nu\right) + \boldsymbol{\Delta}\boldsymbol{m}_2\boldsymbol{\Delta}\boldsymbol{m}_1\boldsymbol{\Delta}\left(\nu^2 + 3\nu\right) \\ &+ \boldsymbol{\Delta}\boldsymbol{m}_2\boldsymbol{\Delta}\left(\nu^2 + \nu\right)\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_1) + \boldsymbol{\Delta}\boldsymbol{m}_1\boldsymbol{\Delta}\left(\nu^2 + \nu\right)\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_2) \\ &+ \boldsymbol{\Delta}\left(\nu^2\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_1\boldsymbol{\Delta}\boldsymbol{m}_2) + \nu\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_1)\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_2) + \nu\operatorname{Tr}(\boldsymbol{\Delta}\boldsymbol{m}_1\boldsymbol{\Delta}\boldsymbol{m}_2)\right), \qquad \text{(A.5)}\end{aligned}$$

which concludes the proof.

Bibliography

[AAD+02] A. Arneodo, B. Audit, N. Decoster, J. F. Muzy, and C. Vaillant. Wavelet based multifractal formalism: applications to DNA sequences, satellite images of the cloud structure, and stock market data. In *The science of Disasters*, pages 26–102. Springer, 2002.

[Abe93] J. S. Abel. A bound on mean-square-estimate error. *IEEE Transactions on Information Theory*, 39(5):1675–1680, 1993.

[ABF+02] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. Multiscale nature of network traffic. *IEEE Signal Proces. Mag.*, 19(3):28–46, May 2002.

[ABM98] A. Arneodo, E. Bacry, and J. F. Muzy. Random cascades on wavelet dyadic trees. *Journal of Mathematical Physics*, 39(8):4142–4164, 1998.

[ACC20] Y. Abichou, S. Chaabene, and L. Chaari. A sleep monitoring method with EEG signals. In L. Chaari, editor, *Digital Health in Focus of Predictive, Preventive and Personalised Medicine*, pages 25–32. Springer, 2020.

[ADKR03] 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 P.W. Hawkes, B. Kazan, and T. Mulvey, editors, *Advances in Imaging and Electron Physics*, volume 126, pages 1–98. Academic Press, 2003.

[AFTV00] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch. Wavelets for the analysis, estimation

and synthesis of scaling data. In *K. Park and W. Willinger, Eds., Self-similar Network Traffic and Performance Evaluation*. Wiley, 2000.

[AJW13] P. Abry, S. Jaffard, and H. Wendt. When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture. *Signal Proces.*, 93(3):554–572, 2013.

[AJW15] P. Abry, S. Jaffard, and H. Wendt. A bridge between geometric measure theory and signal processing: Multifractal analysis. In K. Grochenig, Y. Lyubarskii, and K. Seip, editors, *Operator-Related Function Theory and Time-Frequency Analysis, The Abel Symposium 2012*, pages 1–56. Springer, 2015.

[ALW+19] 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.

[AMM+07] 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*, 41(3):237–250, 2007.

[AMVA04] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali. *Two-Dimensional Wavelets and their Relatives*. Cambridge Univ. Press, 2004.

[ANJ+16] 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*, 10:219, May 2016.

[ANS14] I. Alvarez, J. Niemi, and M. Simpson. Bayesian inference for a covariance matrix. In *Conf. Applied Statistics in Agriculture*, 2014.

[AWJD19] P. Abry, H. Wendt, S. Jaffard, and G. Didier. Multivariate scale-free temporal dynamics: From spectral (Fourier) to fractal (wavelet) analysis. *Comptes Rendus Physique*, 20(5):489–501, 2019.

[BADML12] J. Barunik, T. Aste, T. Di Matteo, and R. Liu. Understanding the source of multifractality in financial markets. *Physica A: Statistical Mechanics and its Applications*, 391(17):4234–4251, 2012.

[Bar49] E. W. Barankin. Locally best unbiased estimates. *The Annals of Mathematical Statistics*, 20(4):477–501, 1949.

[BBT08] O. Besson, S. Bidon, and J.-Y. Tourneret. Bounds for estimation of covariance matrices from heterogeneous samples. *IEEE Transactions on Signal Processing*, 56(7):3357–3362, 2008.

[BDM00] E. Bacry, J. Delour, and J. F. Muzy. A multivariate multifractal model for return fluctuations. *arXiv preprint cond-mat/0009260*, 2000.

[BDM01] E. Bacry, J. Delour, and J. F. Muzy. Multifractal random walk. *Phys. Rev. E*, 64(2):026103, 2001.

[Ber94] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, New York, 1994.

[Bha66] P. K. Bhattacharya. Estimating the mean of a multivariate normal population with general quadratic loss function. *The Annals of Mathematical Statistics*, 37(6):1819–1824, 1966.

[BI13] D. A. Bini and B. Iannazzo. Computing the karcher mean of symmetric positive definite matrices. *Linear Algebra and Its Applications*, 438(4):1700–1710, 2013. 16th ILAS Conference Proceedings, Pisa 2010.

[BKM08] E. Bacry, A. Kozhemyak, and J. F. Muzy. Continuous cascade models for asset returns. *J. Economic Dynamics and Control*, 32(1):156–199, 2008.

[BKM13] E. Bacry, A. Kozhemyak, and J. F. Muzy. Log-normal continuous cascade model of asset returns: aggregation properties and estimation. *Quantitative Finance*, 13(5):795–818, 2013.

[BMA93] E. Bacry, J. F. Muzy, and A. Arnéodo. Singularity spectrum of fractal signals from wavelet analysis: Exact results. *Journal of Statistical Physics*, 70(3-4):635–674, 1993.

[BMM00] J. Barnard, R. McCulloch, and X.-L. Meng. Modelling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:1281–1311, 2000.

[Boc08] H.-H. Bock. Origins and extensions of the k-means algorithm in cluster analysis. *Journal Électronique d'Histoire des Probabilités et de la Statistique [electronic only]*, 4, 01 2008.

[BPL+01] C. L. Benhamou, S. Poupon, E. Lespessailles, S. Loiseau, R. Jennane, V. Siroux, W. J. Ohley, and L. Pothuaud. Fractal analysis of radiographic trabecular bone texture and bone mineral density: two complementary parameters related to osteoporotic fractures. *Journal of Bone and Mineral Research*, 16(4):697–704, 2001.

[BR12] D. R. Brillinger and M. Rosenblatt. Asymptotic theory of estimates of kth-order spectra. In *Selected Works of David Brillinger*, pages 173–177. Springer, 2012.

[BS99] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. on Image Proces.*, 8(12):1688–1701, 1999.

[BWZ+19] C. Basri, K. Wan, I. Zunaidi, L. Lee, A. B. Shahriman, 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*, 557:012004, June 2019.

[BZ76] B.-Z. Bobrovsky and M. Zakai. A lower bound on the estimation error for certain diffusion processes. *IEEE Transactions on Information Theory*, 22:45–52, 1976.

[CBP08] C. R. B. Cabral, H. Bolfarine, and J. R. G. Pereira. Bayesian density estimation using skew student-t-normal mixtures. *Computational Statistics & Data Analysis*, 52(12):5075–5090, 2008.

[CERW08] J. Coddington, J. Elton, D. Rockmore, and Y. Wang. Multifractal analysis and authentication of Jackson Pollock paintings. In *Proc. SPIE 6810*, page 68100F, 2008.

[CGM93] B. Castaing, Y. Gagne, and M. Marchand. Log-similarity for turbulent flows. *Physica D: Nonlinear Phenomena*, 68(4):387–400, 1993.

[Cha97] R. E. Chandler. A spectral method for estimating parameters in rainfall models. *Bernoulli*, pages 301–322, 1997.

[CLCF18] 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.

[Com16] S. Combrexelle. *Multifractal analysis for multivariate data with application to remote sensing*. PhD thesis, Université de Toulouse, 2016. Thèse de doctorat dirigée par Tourneret, Jean-Yves et Mclaughlin, Steve Signal, Image, Acoustique et Optimisation Toulouse, INPT 2016.

[CP06] N. H. Chan and W. Palma. Estimation of long-memory time series models: A survey of different likelihood-based methods. *Adv. Econom.*, 20:89–121, 2006.

[CR51] D. G. Chapman and H. Robbins. Minimum variance estimation without regularity assumptions. *The Annals of Mathematical Statistics*, 22(4):581–586, 1951.

[Cra46] H. Cramér. A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal*, 1946(1):85–94, 1946.

[CRV+19] L. Cheng, S. Ramchandran, T. Vatanen, N. Lietzén, R. Lahesmaa, A. Vehtari, and H. Lähdesmäki. An additive Gaussian process regression model for interpretable non-parametric analysis of longitudinal data. *Nature Communications*, 10(1):1–11, 2019.

[CVA+12a] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt. Scale-free and multifractal dynamic properties of fMRI signals during rest and task. *Frontiers in Physiology*, 3(186), 2012.

[CVA+12b] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt. Scale-free and multifractal time dynamics of fMRI signals during rest and task. *Frontiers in Physiology*, 3:186–186, 2012.

[CWD+15] S. Combrexelle, H. Wendt, N. Dobigeon, J.-Y. Tourneret, S. McLaughlin, and P. Abry. Bayesian estimation of the multifractality parameter for image texture using a Whittle approximation. *IEEE Trans. Image Proces.*, 24(8):2540–2551, 2015.

[CWT+15a] S. Combrexelle, H. Wendt, J.-Y. Tourneret, P. Abry, and S. McLaughlin. Bayesian estimation of the multifractality parameter for images via a closed-form Whittle likelihood. In *Proc. Eur. Signal Proces. Conf. (EUSIPCO)*, Nice, France, 2015.

[CWT+15b] S. Combrexelle, H. Wendt, J.-Y. Tourneret, S. McLaughlin, and P. Abry. Hyperspectral image analysis using multifractal attributes. In *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Tokyo, Japan, June 2015.

[Dau88] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and App. Math.*, 41:909–996, 1988.

[DHA+11] M. Doret, H. Helgason, P. Abry, P. Gonçalvès, C. Gharib, and P. Gaucherand. Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor. *American Journal of Perinatology*, 28(4):259–266, 2011.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[DM01] D. A. van Dyk and X.-L. Meng. The art of data augmentation. *J. of Comput. and Graphical Stat.*, 10(1):pp. 1–50, 2001.

[DMA01] J. Delour, J. F. Muzy, and A. Arneodo. Intermittency of 1D velocity spatial profiles

in turbulence: a magnitude cumulant analysis. *The European Physical Journal B*, 23(2):243–248, 2001.

[DNR11] D. Duvenaud, H. Nickisch, and C. E. Rasmussen. Additive Gaussian processes. *arXiv preprint arXiv:1112.4394*, 2011.

[Eld04] Y. C. Eldar. Minimum variance in biased estimation: bounds and asymptotically optimal estimators. *IEEE Transactions on Signal Processing*, 52(7):1915–1930, 2004.

[FAF+17] 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*, 25(4):1–14, 2017.

[FGK94] E. Foufoula-Georgiou and P. Kumar, editors. *Wavelets in Geophysics*, volume 4. Academic Press, San Diego, 1994.

[FL02] P. Forster and P. Larzabal. On lower bounds for deterministic parameter estimation. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1137–1140, 2002.

[FP85] U. Frisch and G. Parisi. Fully developed turbulence and intermittency. *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, 88:71–88, 1985.

[Fre93] R. Frehlich. Cramér-Rao bound for Gaussian random processes and applications to radar processing of atmospheric signals. *IEEE Transactions on Geoscience and Remote Sensing*, 31(6):1123–1131, 1993.

[FTB+13] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[GAG+00] A. Goldberger, L. Amaral, L. Glass, Hausdorff, Ivanov J., R. P. C., Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet:

Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[Gel06] A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515 – 534, 2006.

[GF14] J. Guinness and M. Fuentes. Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *NC State Department of Statistics Technical Reports, no. 2657*, 2014.

[GLM05] P. Graczyk, G. Letac, and H. Massam. The hyperoctahedral group, symmetric group representations and the moments of the real Wishart distribution. *Journal of Theoretical Probability*, 18(1):1–42, Jan. 2005.

[GRR+09] N. Gorretta, J. M. Roger, G. Rabatel, V. Bellon-Maurel, C. Fiorio, and C. Lelong. Hyperspectral image segmentation: The butterfly approach. In *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Grenoble, France, Aug. 2009.

[Ham50] J. M. Hammersley. On estimating restricted parameters. *Journal of the royal statistical society series b-methodological*, 12:192–229, 1950.

[HDS+17] J. Hensman, N. Durrande, A. Solin, et al. Variational Fourier features for Gaussian processes. *J. Mach. Learn. Res.*, 18(1):5537–5588, 2017.

[HFU96] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform cr bound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, 1996.

[HJ12] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2012.

[HP20] T. Holgersson and J. Pielaszkiewicz. A collection of moments of the Wishart distribution. In *Recent Developments in Multivariate and Random Matrix Analysis*, pages 147–162. Springer International Publishing, 2020.

[HPA11] H. Helgason, V. Pipiras, and P. Abry. Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding. *Signal Proces.*, 91(5):1123 – 1133, 2011.

[IAG+99] P. Ch. Ivanov, L. A. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399(6735):461–465, 1999.

[Jaf97a] S. Jaffard. Multifractal formalism for functions part I: results valid for all functions. *SIAM Journal on Mathematical Analysis*, 28(4):944–970, 1997.

[Jaf97b] S. Jaffard. Multifractal formalism for functions part II: self-similar functions. *SIAM Journal on Mathematical Analysis*, 28(4):971–998, 1997.

[Jaf04] S. Jaffard. Wavelet techniques in multifractal analysis. In M. Lapidus and M. van Frankenhuijsen, editors, *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, Proceedings of Symposia in Pure Mathematics*, volume 72(2), pages 91–152. AMS, 2004.

[JAR+10] S. Jaffard, P. Abry, S. Roux, B. Vedel, and H. Wendt. *The contribution of wavelets in multifractal analysis*, pages 51–98. Series in contemporary applied mathematics. Higher Education Press, Series in contemporary applied mathematics, China, 2010.

[JAW15] S. Jaffard, P. Abry, and H. Wendt. *Irregularities and Scaling in Signal and Image Processing: Multifractal Analysis*, pages 31–116. World scientific publishing, Singapore, 2015.

[JJHB+08] C. R. Johnson Jr., E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Processing for artist identification: Computerized

analysis of Vincent van Gogh's painting brush- strokes. *IEEE Signal Processing Magazine (Special Section - Signal Processing in Visual Cultural Heritage)*, 25:37–48, 2008.

[JLA06] S. Jaffard, B. Lashermes, and P. Abry. Wavelet leaders in multifractal analysis. In *Wavelet Analysis and Applications, T. Qian, M.I. Vai, X. Yuesheng, Eds.*, pages 219–264, Basel, Switzerland, 2006. Birkhäuser.

[JML+16] 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*, 448:300–318, 2016.

[JS01] A. Johansen and D. Sornette. Finite-time singularity in the dynamics of the world population, economic and financial indices. *Physica A: Statistical Mechanics and its Applications*, 294:465–502, 2001.

[JSW+19a] S. Jaffard, S. Seuret, H. Wendt, R. Leonarduzzi, and P. Abry. Multifractal formalisms for multivariate analysis. *Proc. Royal Society A*, 475(2229), 2019.

[JSW+19b] S. Jaffard, S. Seuret, H. Wendt, R. Leonarduzzi, S. Roux, and P. Abry. Multivariate multifractal analysis. *Applied and Computational Harmonic Analysis*, 46(3):653–663, 2019.

[JW02] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 2002.

[Kay93] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[KLSJA01] P. Kestener, J. Lina, P. Saint-Jean, and A. Arneodo. Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Analysis and Stereology*, 20(3):169–175, 2001.

[KSAY06] 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 Transactions on Biomedical Engineering*, 53(1):95–102, Jan. 2006.

[KWD+20] M. Krupiński, A. Wawrzaszek, W. Drzewiecki, M. Jenerowicz, and S. Aleksandrowicz. What can multifractal analysis tell us about hyperspectral imagery? *Remote Sensing*, 12(24), 2020.

[LAR+18] 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.

[LB09] R. Lopes and N. Betrouni. Fractal and multifractal analysis: A review. *Medical Image Analysis*, 13:634–649, 2009.

[LJA05] B. Lashermes, S. Jaffard, and P. Abry. Wavelet leader based multifractal analysis. In *IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, volume 4, pages iv–161. IEEE, 2005.

[LM01] J. Liu and P. Moulin. Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. *IEEE Trans. Image Proces.*, 10(11):1647–1658, 2001.

[LOKS86] T. Lundahl, W.J. Ohley, S.M. Kay, and R. Siffert. Fractional Brownian motion: A maximum likelihood estimator and its application to image texture. *IEEE Trans. Medical Imaging*, 5(3):152–161, Sept 1986.

[LPAT21] J. Lesouple, B. Pilastre, Y. Altmann, and J.-Y. Tourneret. Robust hypersphere fitting from noisy data using an em algorithm. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 1–5, 2021.

[LR12] O. Løvsletten and M. Rypdal. Approximated maximum likelihood estimation in multifractal random walks. *Phys. Rev. E*, 85:046705, 2012.

[LS13] S. Lovejoy and D. Schertzer. *The weather and climate: emergent laws and multifractal cascades*. Cambridge University Press, 2013.

[LTA14] R. F Leonarduzzi, M. E. Torres, and P. Abry. Scaling range automated selection for wavelet leader multifractal analysis. *Signal Proces.*, 105:243–257, 2014.

[Lux07] T. Lux. Higher dimensional multifractal processes: A GMM approach. *Journal of Business and Economic Statistics*, 26:194–210, 2007.

[Lux08] T. Lux. The Markov-switching multifractal model of asset returns. *J. Business & Economic Stat.*, 26(2):194–210, 2008.

[LWA+16] 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*, 448:319–339, 2016.

[LWTA21] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry. Bayesian estimation of the parameters of the bivariate multifractal spectrum. In *Proc. European Signal Processing Conference (EUSIPCO)*, pages 1930–1934, 2021.

[LWTA22a] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry. A Bayesian framework for multivariate multifractal analysis. *IEEE Transactions on Signal Processing*, 70:3663–3675, 2022.

[LWTA22b] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry. A comparison of Bayesian estimators for the parameters of the bivariate multifractal spectrum. In *Proc. European Signal Processing Conference (EUSIPCO)*, pages 2116–2120, Belgrade, Serbia, August 2022.

[LWTA22c] L. Leon, H. Wendt, J.-Y. Tourneret, and P. Abry. Estimation du paramètre de multifractalité : régression linéaire, maximum de vraisemblance ou inférence Bayésienne ? In *Proc. Colloque Francophone de Traitement du Signal et des Images (GRETSI 2022)*, number 001-0323, pages p. 1293–1296, Nancy, Sep 2022.

[Mac67] J. MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297, 1967.

[Mal98] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, CA, 1998.

[Man74] B. B. Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62:331–358, 1974.

[Man90] B. Mandelbrot. Limit lognormal multifractal measures. In E.A. Gotsman, Y. Ne'eman, and A. Voronel, editors, *Frontiers of Physics, Proc. Landau Memorial Conf., Tel Aviv, 1988*, pages 309–340. Pergamon Press, 1990.

[Man99] B. Mandelbrot. Iterated random multiplications and invariance under randomly weighted averaging. In *Multifractals and* $1/f$ *Noise*, pages 358–371. Springer New York, 1999.

[MBA93] 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*, 47(2):875, 1993.

[MDB00] J. F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *The European Physical Journal B*, 17, 05 2000.

[MECV18] L. Martino, V. Elvira, and G. Camps-Valls. The recycling gibbs sampler for efficient learning. *Digital Signal Processing*, 74:1–13, 2018.

[MNWL06] 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)*, pages 4526–4529, 2006.

[Moa05] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, March 2005.

[MP12] G. Martin and A. Plaza. Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, 5(2):380–395, April 2012.

[MRS08] E. Moulines, F. Roueff, and M. S.Taqqu. A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. *Ann. Stat.*, pages 1925–1956, 2008.

[MS98] B. B. Mandelbrot and I. Stewart. Fractals and scaling in finance. *Nature*, 391(6669):758–758, 1998.

[MSDA01] J. F. Muzy, D. Sornette, J. Delour, and A. Arneodo. Multifractal returns and hierarchical portfolio theory. *Quantitative Finance*, 1(1):131–148, 2001.

[MSKF90] C. Meneveau, K.R. Sreenivasan, P. Kailasnath, and M. S. Fan. Joint multifractal measures: Theory and applications to turbulence. *Physical Review A*, 41(2):894–913, 1990.

[NO12] J. Nakajima and Y. Omori. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew students t-distribution. *Computational Statistics & Data Analysis*, 56(11):3690–3704, 2012.

[NR02] P. Norvig and S.-J. Russel. Artificial intelligence: A modern approach. *Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems*, 90:33–48, 2002.

[OA16] F. B. Oppong and S. Y. Agbedra. Assessing univariate and multivariate normality, a guide for non-statisticians. *Mathematical Theory and Modeling*, 6(2):26–33, 2016.

[OW00] M. Ossiander and E. C. Waymire. Statistical estimation for multiplicative cascades. *Ann. Stat.*, 28(6):1533–1560, 2000.

[OZ08] A. J. O'Malley and A. M. Zaslavsky. Domain-level covariance analysis for multilevel survey data with structured nonresponse. *J. American Statistical Association*, 103(484):1405–1418, 2008.

[PBA+06] L. Ponson, D. Bonamy, H. Auradou, G. Mourot, S. Morel, E. Bouchaud, C. Guillot, and J. Hulin. Anisotropic self-affine properties of experimental fracture surface. *J. Fracture*, 140(1–4):27–36, 2006.

[PF86] B. Porat and B. Friedlander. Computation of the exact information matrix of Gaussian time series with stationary random components. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(1):118–130, 1986.

[RAAF10] G. Roussel, V. Achard, A. Alakian, and J.-C. Fort. Benefits of textural characterization for the classification of hyperspectral images. In *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Reykjavik, Iceland, June 2010.

[RAD00] S. G. Roux, A. Arneodo, and N. Decoster. A wavelet-based method for multifractal image analysis. III. Applications to high-resolution satellite images of cloud structure. *European Physical Journal B: Condensed Matter Physics*, 15(4):765–786, 2000.

[Rao92] C. R. Rao. *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, pages 235–247. Springer New York, New York, NY, 1992.

[RC05] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, USA, 2005.

[Rie03] R. H. Riedi. Multifractal processes. In P. Doukhan, G. Oppenheim, and M.S. Taqqu, editors, *Theory and applications of long range dependence*, pages 625–717. Birkhäuser, 2003.

[RW84] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

[SFL+11] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon. Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential

of hyperspectral imagery. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 3672–3675, Vancouver, Canada, July 2011.

[SL87] D. Schertzer and S. Lovejoy. Physical modelling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92:9693–9714, 1987.

[SL05] L. Svensson and M. Lundberg. On posterior distributions for signals in Gaussian noise with unknown covariance matrix. *IEEE Transactions on Signal Processing*, 53(9):3554–3571, 2005.

[SOLE17] 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 Transactions on Signal Processing*, 65(12):3136–3151, 2017.

[SS01] F. G. Schmitt and L. Seuront. Multifractal random walk in copepod behavior. *Physica A: Statistical Mechanics and its Applications*, 301(1):375–396, 2001.

[SS02] L. Sendur and I. W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing*, 50(11):2744–2756, 2002.

[SSM14] A. Sahayadhas, K. Sundaraj, and M. Murugappan. Electromyogram signal based hypovigilance detection. *Biomedical Research (India)*, 25:281–288, 2014.

[ST94] G. Samoradnitsky and M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. Chapman and Hall, New York, 1994.

[SXGL06] W. Sun, G. Xu, P. Gong, and S. Liang. Fractal analysis of remotely sensed images: A review of methods and applications. *International Journal of Remote Sensing*, 27(22):4963–4990, 2006.

[TA03] M. E. Torres and P. Abry. Comparison of different methods for computing scaling parameter in the presence of trends. In *Proc. XIV Congreso Argentino de Bioingenieria (SABI 2003)*, Cordoba, Argentina, 2003.

[TGM+11] T. Tokuda, B. Goodrich, I. Mechelen, A. Gelman, and F. Tuerlinckx. Visualizing distributions of covariance matrices. Technical report, Columbia University, New York; USA, 2011.

[Tre01] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley & Sons, Inc., Sept. 2001.

[TW87] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Ass.*, 82(398):pp. 528–540, 1987.

[VA99] D. Veitch and P. Abry. A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Info. Theory*, 45(3):878–897, 1999.

[VAT03] D. Veitch, P. Abry, and M. S. Taqqu. On the automatic selection of the onset of scales. *Fractals*, 11(04):377–390, June 2003.

[VTA00] D. Veitch, M. S. Taqqu, and P. Abry. Meaningful MRA initialization for discrete time series. *Signal Proces.*, 80(9):1971–1983, 2000.

[VWAJ10] B. Vedel, H. Wendt, P. Abry, and S. Jaffard. On the impact of the number of vanishing moments on the dependence structures of compound Poisson motion and fractional Brownian motion in multifractal time. In P. Doukhan, G. Lang, D. Surgailis, and G. Teyssière, editors, *Dependence in Probability and Statistics*, Lecture Notes in Statistics, pages 71–101. Springer Berlin Heidelberg, 2010.

[WAJ07] H. Wendt, P. Abry, and S. Jaffard. Bootstrap for empirical multifractal analysis. *IEEE Signal Proces. Mag.*, 24(4):38–48, 2007.

[WAJ+09] H. Wendt, P. Abry, S. Jaffard, H. Ji, and Z. Shen. Wavelet leader multifractal analysis for texture classification. In *Proc. IEEE Int. Conf. Image Proces. (ICIP)*, Cairo, Egypt, Nov. 2009.

[WAK+19] 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*, 66(1):80–87, 2019.

[WCA+18] 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)*, 11(2):1294–1316, 2018.

[WCB+09] 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*, 185(1):116–124, 2009.

[WDTA13] H. Wendt, N. Dobigeon, J.-Y. Tourneret, and P. Abry. Bayesian estimation for the multifractality parameter. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, pages 6556–6560, Vancouver, Canada, May 2013.

[Wen08] H. Wendt. *Contributions of Wavelet Leaders and Bootstrap to Multifractal Analysis: Images, Estimation Performance, Dependence Structure and Vanishing Moments. Confidence Intervals and Hypothesis Tests.* PhD thesis, Ecole Normale Supérieure de Lyon, 2008.

[Whi53] P. Whittle. Estimation and information in stationary time series. *Arkiv för matematik*, 2(5):423–434, 1953.

[WLA+18] 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 Proces. (ICASSP)*, Calgary, Canada, April 2018.

[WO92] G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Proces.*, 40(3):611–623, 1992.

[WRJA09] H. Wendt, S. G. Roux, S. Jaffard, and P. Abry. Wavelet leaders and bootstrap for multifractal analysis of images. *Signal Proces.*, 89(6):1100–1114, 2009.

[Wu83] C. F. Jeff Wu. On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11(1):95 – 103, 1983.

[WW85] A. Weiss and E. Weinstein. A lower bound on the mean-square error in random parameter estimation (corresp.). *IEEE Transactions on Information Theory*, 31(5):680–682, 1985.

[WW88] E. Weinstein and A. J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2):338–342, 1988.

[WWF18] F. Wang, H. Wang, and R. Fu. Real-time ECG-based detection of fatigue driving using sample entropy. *Entropy*, 20(3), 2018.

[XYLJ10] Y. Xu, X. Yang, H. Ling, and H. Ji. A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid. In *Proc. IEEE Conf. Comp. Vis. Pattern Recognition (CVPR)*, pages 161–168, San Francisco, CA, USA, June 2010.

[YPLJ18] J. Yu, S. Park, S. Lee, and M. Jeon. Driver drowsiness detection using condition-adaptive representation learning framework. *IEEE Trans. on Intell. Transp. Syst.*, pages 1–13, December 2018.

[ZM13] 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*, 8(7):e68360, 2013.