



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Signal, Image, Acoustique et Optimisation

Présentée et soutenue par :

M. SEBASTIEN COMBREXELLE

le mercredi 12 octobre 2016

Titre :

MULTIFRACTAL ANALYSIS FOR MULTIVARIATE DATA WITH
APPLICATION TO REMOTE SENSING

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. JEAN YVES TOURNERET

M. STEVE MCLAUGHLIN

Rapporteurs :

M. GABRIEL PEYRE, UNIVERSITE PARIS 9

M. PHILIPPE CIUCIU, CEA SACLAY

Membre(s) du jury :

Mme LAURE BLANC-FERAUD, CNRS COTE D'AZUR, Président

M. HERWIG WENDT, INP TOULOUSE, Membre

M. JEAN YVES TOURNERET, INP TOULOUSE, Membre

Mme VÉRONIQUE SERFATY, DGA, Membre

M. PATRICE ABRY, ECOLE NORMALE SUPERIEURE DE LYON, Membre

M. STEVE MCLAUGHLIN, HERIOT WATT UNIVERSITY, Membre

Remerciements

J'aimerais en premier lieu remercier mes directeurs de thèse M. Jean-Yves Tournet et M. Stephen McLaughlin de m'avoir donné la chance d'effectuer cette thèse à leurs côtés et de m'avoir aidé à la compléter en se rendant toujours disponible. Je tiens aussi à remercier tout particulièrement mon encadrant de thèse M. Herwig Wendt aussi bien pour son encadrement et ses conseils scientifiques que pour son soutien et son aide tout au long de ces trois ans de thèse. Je lui souhaite une très bonne continuation dans sa carrière, qui, j'en suis sûre, sera pleine de succès.

J'aimerais ensuite remercier le jury de m'avoir fait l'honneur d'accepter d'évaluer mon travail et d'assister à ma soutenance de thèse. Mes plus sincères remerciements vont à mes rapporteurs M. Philippe Ciuciu et M. Gabriel Peyré qui ont pris le temps de lire en détails et de juger mon manuscrit. Je suis également reconnaissant envers M. Patrice Abry, Mme Laure Blanc-Féraud et Mme Véronique Serfaty d'avoir accepté d'être examinateurs de mon travail de thèse. Enfin, je remercie Mme Victoria Cox d'avoir accepté mon invitation à ma soutenance de thèse.

Ce travail de thèse n'aurait pas été aussi fructueux sans la collaboration scientifique de nombreuses personnes. Je pense en particulier à M. Patrice Abry, M. Yoann Altmann et M. Nicolas Dobigeon, que je remercie ici d'avoir participé à mes recherches.

Je suis profondément reconnaissant à la Direction Générale de l'Armement (DGA) d'avoir financé et encadré cette thèse dans le cadre du programme Franco-Britannique de coopération entre la DGA et le Defence Science and Technology Laboratory. J'espère que ce programme continuera d'exister encore longtemps et donnera à d'autres étudiants comme moi la chance d'effectuer leur thèse entre la France et le Royaume-Uni.

Au sujet de mes séjours en Écosse, j'aimerais de nouveau remercier M. Stephen McLaughlin pour son hospitalité et son effort quotidien pour rendre l'accent écossais compréhensible à mes oreilles peu habituées. Je remercie également Mme Christie Gwen pour son efficacité et son aide qui ont toujours facilité mes séjours en Écosse. Bien sûr je ne peux que remercier à nouveau M. Yoann Altmann pour ses commentaires et ses idées mais également pour sa disponibilité et ses précieux conseils sur la vie en Écosse.

Ma vie à Toulouse aura aussi été grandement simplifiée grâce à l'aide des nombreuses secrétaires de l'IRIT, Mme Sylvie Armengaud, Mme Sylvie Eichen, Mme Annabelle Sansus et dorénavant Mme Isabelle Vasseur, qui effectuent pour nous un travail remarquable. Je n'oublie pas de remercier tous ceux que j'ai pu croiser de près ou de loin au cours de ma thèse. Merci donc à mes anciens et actuels collègues de bureau Abderrahim, Qi, Ning-Ning, Pierre-Antoine, Nora, Adrien et aux nombreux membres, anciens et actuels, de l'IRIT, Julien, Bilel, Yanna, Vinicius, Etienne, Olivier, Bouchra, Matthieu, Thomas, Charly, Marie, Benoît, Corinne et tous ceux que j'oublie.

Quelques petits mots de remerciement vont aussi à ma famille pour leur soutien et leur réaction admirative face à la lecture du titre de ma thèse.

Enfin, ces remerciements ne sauraient être complets sans un grand merci à ma moitié - aka mon petit Confucius *Bubu* - envers qui je suis reconnaissant de son indéfectible soutien et de toutes ses petites attentions de tous les jours qui font de la vie à ses côtés un vrai bonheur. Je pense aussi à mes beaux-parents, à qui j'aimerais adresser ces quelques mots de remerciement :

非常感谢你们对我的支持以及你们无时无刻的关心。最后，我感谢你们生了这么好的女儿，她是你们给我在这世界上最珍贵的礼物。

Résumé

La caractérisation de texture est centrale dans de nombreuses applications liées au traitement d'images. L'analyse de textures peut être envisagée dans le cadre mathématique de l'analyse multifractale qui permet d'étudier les fluctuations de la régularité ponctuelle de l'amplitude d'une image et fournit les outils pratiques pour leur évaluation grâce aux coefficients d'ondelettes ou aux coefficients dominants. Bien que mise à profit dans de nombreuses applications, l'analyse multifractale souffre à présent de deux limitations majeures. Premièrement, l'estimation des paramètres multifractaux reste délicate, notamment pour les images de petites tailles. Deuxièmement, l'analyse multifractale a été jusqu'à présent uniquement considérée pour l'analyse univariée d'images, alors que les données à étudier sont de plus en plus multivariées. L'objectif principal de cette thèse est la mise au point de contributions pratiques permettant de pallier ces limitations. La première limitation est abordée en introduisant un modèle statistique générique pour le logarithme des coefficients dominants, paramétrisé par les paramètres multifractaux d'intérêt. Ce modèle statistique permet de contrebalancer la variabilité résultant de l'analyse d'images de petite taille et de formuler l'estimation dans un cadre bayésien. Cette approche aboutit à des procédures d'estimation robustes et efficaces, que ce soit pour des images de petites ou grandes tailles. Ensuite, l'analyse multifractale d'images multivariées est traitée en généralisant ce cadre bayésien à des modèles hiérarchiques capables de prendre en compte l'hypothèse d'une évolution lente des propriétés multifractales d'images multi-temporelles ou multi-bandes. Ceci est réalisé en définissant des lois *a priori* reliant les propriétés dynamiques des paramètres multifractaux des différents éléments composant le jeu de données. Différents types de lois *a priori* sont étudiés dans cette thèse au travers de simulations numériques conduites sur des images multifractales multivariées synthétiques. Ce travail est complété par une étude du potentiel apport de l'analyse

multifractale et de la méthodologie bayésienne proposée pour la télédétection à travers l'exemple de l'imagerie hyperspectrale.

Abstract

Texture characterization is a central element in many image processing applications. Texture analysis can be embedded in the mathematical framework of multifractal analysis, enabling the study of the fluctuations in regularity of image intensity and providing practical tools for their assessment, the wavelet coefficients or wavelet leaders. Although successfully applied in various contexts, multifractal analysis suffers at present from two major limitations. First, the accurate estimation of multifractal parameters for image texture remains a challenge, notably for small image sizes. Second, multifractal analysis has so far been limited to the analysis of a single image, while the data available in applications are increasingly multivariate. The main goal of this thesis is to develop practical contributions to overcome these limitations. The first limitation is tackled by introducing a generic statistical model for the logarithm of wavelet leaders, parametrized by multifractal parameters of interest. This statistical model enables us to counterbalance the variability induced by small sample sizes and to embed the estimation in a Bayesian framework. This yields robust and accurate estimation procedures, effective both for small and large images. The multifractal analysis of multivariate images is then addressed by generalizing this Bayesian framework to hierarchical models able to account for the assumption that multifractal properties evolve smoothly in the dataset. This is achieved via the design of suitable priors relating the dynamical properties of the multifractal parameters of the different components composing the dataset. Different priors are investigated and compared in this thesis by means of numerical simulations conducted on synthetic multivariate multifractal images. This work is further completed by the investigation of the potential benefits of multifractal analysis and the proposed Bayesian methodology for remote sensing via the example of hyperspectral imaging.

Abbreviations and notation

Abbreviations

1D	One dimensional
2D	Two dimensional
CMC	Canonical Mandelbrot cascade
CPC	Compound Poisson cascade
DFT	Discrete Fourier transform
DGA	Direction générale de l'armement
DSTL	Defence science and technology laboratory
DWT	Discrete wavelet transform
fBm	Fractional Brownian motion
fMRI	Functional magnetic resonance imaging
GMM	Generalized method of moments
GMRF	Gamma Markov random field
GS	Gibbs sampler
IG	Inverse-gamma
LF	Linear fit
LF _w	Weighted linear fit
LF _{nw}	Non weighted linear fit
LN	Log-normal

LP	Log-Poisson
MAP	Maximum a posteriori
MCMC	Markov chain Monte Carlo
MFA	Multifractal analysis
MwG	Metropolis-within-Gibbs
ML	Maximum likelihood
MLE	Maximum likelihood estimator
MMC	Multifractal multiplicative cascade
MMSE	Minimum mean-square error
MRW	Multifractal random walk
QQ	Quantile-quantile
RMSE	Root mean-square error
SAR	Simultaneous autoregression
STD	Standard deviation
TSG	Two-stage Gibbs

Main notation

Matrix notation

\cdot^*	Complex conjugation
\cdot^T	Transpose operator
\cdot^H	Hermitian transpose operator
$\text{rank}(\mathbf{A})$	Rank of matrix \mathbf{A}
$\det(\mathbf{A})$	Determinant of matrix \mathbf{A}
$\text{Tr}(\mathbf{A})$	Trace of matrix \mathbf{A}
$\ \mathbf{a}\ $	Standard L^2 -norm $\ \mathbf{a}\ = \sqrt{\mathbf{a}^T \mathbf{a}}$
$\lfloor \cdot \rfloor$	Integer truncating operator
$[\mathbf{a}]_n$	n -th element of the vector \mathbf{a}

$[\mathbf{A}]_{n_1, n_2}$	Element of the matrix \mathbf{A} in the n_1 -th row and n_2 -th column
$\ \mathbf{a}\ _{\mathbf{M}}$	$\mathbf{a}^H \mathbf{M} \mathbf{a}$
$\llbracket n_1, n_2 \rrbracket$	Set of integers ranging from n_1 to n_2
$\mathbb{1}_{\mathcal{A}}(\cdot)$	Indicator function of the set \mathcal{A}
$\mathbf{1}_N$	$N \times 1$ one vector
$\text{diag}(a_1, \dots, a_N)$	$N \times N$ diagonal matrix with (diagonal) entries specified by a_1, \dots, a_N

Random variable notation

$x \sim$	"The variable x is distributed according to"
$p(x) \propto$	"The distribution of x is proportional to"
$\mathcal{N}(m, \sigma^2)$	Gaussian distribution with mean m and variance σ^2
$\mathcal{CN}(m, \sigma^2)$	Circularly-symmetric complex Gaussian distribution with mean m and variance σ^2
$\mathcal{IG}(\alpha, \beta)$	Inverse-gamma distribution with shape parameter α and scale parameter β
$\mathcal{G}(\alpha, \beta)$	Gamma distribution with shape parameter α and rate parameter β
$\text{Cum}_p[\cdot], \widehat{\text{Cum}}_p[\cdot]$	p -th order cumulant and sample cumulant
$\mathbb{E}[\cdot], \widehat{\mathbb{E}}[\cdot]$	Mean and sample mean
$\text{Var}[\cdot], \widehat{\text{Var}}[\cdot]$	Variance and sample variance
$\text{Cov}[\cdot], \widehat{\text{Cov}}[\cdot]$	Covariance and sample covariance

Main notation of Chapter 1

$X(\mathbf{u})$	Locally bounded 2D function
$\mathbf{u} = (x_1, x_2)$	Spatial location
$h(\mathbf{u})$	Hölder exponent at position \mathbf{u}
\dim_H	Hausdorff dimension
$D(h)$	Multifractal spectrum
h_m and h_M	Minimum and maximum Hölder exponents
$X(\mathbf{k})$	Square discretized version of the function X

$\mathbf{k} = (k_1, k_2)$	Discret spatial location, $k_i = 1, \dots, N$
H_0 and G_0	Low-pass and high-pass filters of a 1D DWT
ψ	Mother wavelet
N_ψ	Vanishing moments
$G^{(0)}$	Low-pass filter of a 2D DWT
$G^{(m)}, m = 1, \dots, 3$	High-pass filters of a 2D DWT
$D_X^{(0)}(j, \mathbf{k})$	Approximation coefficient at position \mathbf{k} and scale 2^j
$D_X^{(m)}(j, \mathbf{k}), m = 1, \dots, 3$	Wavelet coefficient at position \mathbf{k} and scale 2^j
$d_X^{(m)}(j, \mathbf{k}), m = 1, \dots, 3$	L^1 -normalized wavelet coefficient
$\lambda_{j, \mathbf{k}}$	Dyadic cube of side length 2^j centered at $\mathbf{k}2^j$
$L(j, \mathbf{k})$	Wavelet leader at position \mathbf{k} and scale j
$S(j, q)$	q -th order structure function of the wavelet leaders at scale 2^j
$\zeta(q)$	Scaling exponents
$\mathcal{L}(h)$	Legendre spectrum
c_p	p -th order log-cumulant
H	Hurst index of fBm
$\{w_j\}_{j=j_1}^{j_2}$	Linear regression weights
n_j	Number of wavelet leaders at scale j

Main notation of Chapter 2

$l(j, \mathbf{k})$	Logarithm of wavelet leaders (log-leaders) at scale j
ℓ_j	Vectorized log-leaders at scale j
ℓ	Vectorized log-leaders for $j = j_1, \dots, j_2$
m_j	Square root of the number of wavelet leaders at scale j , $m_j = \sqrt{n_j}$
\mathbf{w}	Mean parameter $\mathbf{w} = [c_1, c_1^0]^T$
$\boldsymbol{\theta}$	Covariance parameter $\boldsymbol{\theta} = [c_2, c_2^0]^T$ (before reparametrization)
\mathbf{v}	Covariance parameter $\mathbf{v} = \psi(\boldsymbol{\theta})$ (after reparametrization)
$\boldsymbol{\lambda}$	Unknown parameters $\{\boldsymbol{\theta}, \mathbf{w}\}$ or $\{\mathbf{v}, \mathbf{w}\}$

$\boldsymbol{\mu}_{j,\mathbf{w}}$	Parametric mean model
$\varrho_{j,\boldsymbol{\theta}}$	Parametric covariance model
$\boldsymbol{\Sigma}_{j,\boldsymbol{\theta}}$	Covariance matrix induced by parametric covariance model $\varrho_{j,\boldsymbol{\theta}}$
$\phi_{j,\boldsymbol{\theta}}, \phi_{j,\mathbf{v}}$	Parametric spectral densities
y_{ℓ_j}	Discret Fourier coefficient of log-leaders at scale j
$\boldsymbol{\omega}_{\mathbf{m}}$	Frequency $\boldsymbol{\omega}_{\mathbf{m}} = 2\pi\mathbf{m}/m_j$
J_j	Half of the total frequency grid
\mathbf{y}_{ℓ_j}	Vectorized (non-zero frequencies) discret Fourier coefficients at scale j
\mathbf{y}_{\emptyset}	Vectorized (non-zero frequencies) discret Fourier coefficients for $j = j_1, \dots, j_2$
\mathbf{y}_0	Vectorized discret (zero frequency) Fourier coefficients for $j = j_1, \dots, j_2$
\mathbf{y}	Vectorized discret (all frequencies) Fourier coefficients for $j = j_1, \dots, j_2$
N_{\emptyset}	Length of vector \mathbf{y}_{\emptyset}
N_0	Length of vector \mathbf{y}_0
$\boldsymbol{\mu}$	Latent variable vector associated with the augmented likelihood
η	Bandwidth parameter in the Whittle approximation
N_{mc}	Length of Markov chains in an MCMC algorithm
N_{bi}	Length of the burn-in period in an MCMC algorithm

Main notation of Chapter 3

\mathbf{X}	Dataset under analysis
$\{\mathbf{X}_{\underline{\mathbf{k}}}\}_{\underline{\mathbf{k}}}$	Decomposition of the dataset \mathbf{X} indexed by $\underline{\mathbf{k}}$
k_t, k_{λ}	Temporal and spectral indexation $k_t, k_{\lambda} \in \Omega_1 = \llbracket 1, M_t \rrbracket$
\mathbf{k}	Spatial indexation $\mathbf{k} \in \Omega_2 = \llbracket 1, M_x \rrbracket \times \llbracket 1, M_y \rrbracket$
$\underline{\mathbf{k}} = (k_t, \mathbf{k})$ or $(k_{\lambda}, \mathbf{k})$	Spatio-temporal/spectral indexation
$\ell_{\underline{\mathbf{k}}}$	Log-leaders associated with the element $\mathbf{X}_{\underline{\mathbf{k}}}$
$\mathbf{y}_{\underline{\mathbf{k}}} = [\mathbf{y}_{\emptyset, \underline{\mathbf{k}}}^T, \mathbf{y}_{0, \underline{\mathbf{k}}}^T]$	Fourier coefficients associated with the element $\mathbf{X}_{\underline{\mathbf{k}}}$
$\boldsymbol{\mu}_{\underline{\mathbf{k}}} = [\boldsymbol{\mu}_{\emptyset, \underline{\mathbf{k}}}^T, \boldsymbol{\mu}_{0, \underline{\mathbf{k}}}^T]$	Latent variables associated with the element $\mathbf{X}_{\underline{\mathbf{k}}}$

$\mathbf{w}_{\underline{k}} = [w_{1,\underline{k}}, w_{2,\underline{k}}]^T$	Mean parameters associated with the element $\mathbf{X}_{\underline{k}}$
$\mathbf{v}_{\underline{k}} = [v_{1,\underline{k}}, v_{2,\underline{k}}]^T$	Covariance parameters associated with the element $\mathbf{X}_{\underline{k}}$
$\lambda_{\underline{k}} = \{\mathbf{v}_{\underline{k}}, \mathbf{w}_{\underline{k}}\}$	Unknown parameters associated with the element $\mathbf{X}_{\underline{k}}$
\mathbf{w}_i	Vector containing all $w_{i,\underline{k}}$ organized in the lexicographic ordering
v_i	Vector containing all $v_{i,\underline{k}}$ organized in the lexicographic ordering
$\mathbf{L} = \{\ell_{\underline{k}}\}_{\underline{k}}$	Collection of the log-leaders
$\mathbf{Y} = \{\mathbf{y}_{\underline{k}}\}_{\underline{k}}$	Collection of the Fourier coefficients
$\mathbf{M} = \{\mu_{\underline{k}}\}_{\underline{k}}$	Collection of the latent variables
$\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$	Collection of the mean parameters
$\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2\}$	Collection of the covariance parameters
$\mathbf{\Lambda} = \{\mathbf{V}, \mathbf{W}\}$	Collection of the unknown parameters
$\mathbf{z}_i = \{\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)}\}$	GMRF latent variables
$\mathbf{a}_i = [a_i^{(1)}, a_i^{(2)}]$	GMRF hyperparameters
$\Phi_{\mathcal{I}}, \Phi_{\mathcal{G}}, \Phi$	GMRF potentials
$C(\mathbf{a}_i)$	GMRF normalizing constant
$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2)$	Collection of the GMRF latent variables
$\mathbf{D}^{(1)}, \mathbf{D}^{(2)}$	Laplacian operators associated with a SAR prior
$\epsilon = \{\epsilon_1, \epsilon_2\}, \tilde{\epsilon} = \{\tilde{\epsilon}_1, \tilde{\epsilon}_2\}$	SAR hyperparameters
$\tau(\cdot)$	Lexicographic ordering operator

Contents

Remerciements	iii
Résumé	v
Abstract	vii
Abbreviations and notation	ix
Introduction	1
1 Multifractal analysis	11
1.1 Introduction	12
1.2 Introduction (in French)	12
1.3 Multifractal analysis of images	13
1.3.1 Motivation	13
1.3.2 Hölder exponent	13
1.3.3 Multifractal spectrum	14
1.4 Wavelet leader multifractal formalism	15
1.4.1 Wavelet leaders	16
1.4.2 Wavelet leader multifractal formalism	18
1.4.3 Negative regularity	20
1.5 Two-dimensional scaling processes	20

1.5.1	Fractional Brownian motion and self-similar processes	21
1.5.2	Multifractal multiplicative cascade based processes	21
1.6	Estimation procedures	26
1.6.1	Linear regression	26
1.6.2	Alternative estimation procedures	28
1.7	Conclusion	30
1.8	Conclusion (in French)	30
2	Statistical model and univariate Bayesian estimation	31
2.1	Introduction	32
2.2	Introduction (in French)	33
2.3	Statistical model for log-leaders	34
2.3.1	Second-order statistical model	35
2.3.2	Approximation of the likelihood	40
2.4	Bayesian models	47
2.4.1	Whittle approximation based Bayesian model	48
2.4.2	Augmented likelihood based Bayesian model	49
2.5	Estimation algorithms	53
2.5.1	Metropolis-within-Gibbs algorithm	54
2.5.2	Two-stage Gibbs sampler	56
2.6	Numerical experiments	58
2.6.1	General setup	58
2.6.2	Adjusting the frequency range in the Whittle approximation	59
2.6.3	Numerical assessment of equivalence between Bayesian models	60
2.6.4	Estimation performance for c_1	63
2.6.5	Estimation performance for c_2	63
2.6.6	Convergence and computational cost	69
2.7	Conclusions and perspectives	71

2.8	Conclusions et perspectives (in French)	72
3	Bayesian multifractal analysis of multivariate images	75
3.1	Introduction	76
3.2	Introduction (in French)	78
3.3	Problem formulation: Multivariate image scenarios	80
3.4	Regularization of \mathbf{v}_i via gamma Markov random field priors	83
3.4.1	Bayesian model	83
3.4.2	Estimation algorithm	90
3.4.3	Numerical experiments	91
3.4.4	Preliminary conclusions	100
3.5	Regularization of \mathbf{v}_i via simultaneous autoregressive priors	101
3.5.1	Bayesian model	101
3.5.2	Estimation algorithm	105
3.5.3	Numerical experiments	108
3.5.4	Preliminary conclusions	112
3.6	Regularization of \mathbf{w}_i via simultaneous autoregressive priors	112
3.6.1	Prior and hyperprior specification	113
3.6.2	Sampling of \mathbf{W} and $\tilde{\epsilon}$	114
3.6.3	Numerical experiments	116
3.6.4	Preliminary conclusions	118
3.7	Conclusions and perspectives	118
3.8	Conclusions et perspectives (in French)	119
4	Application to real-word data: Illustration for hyperspectral images	123
4.1	Introduction	124
4.2	Introduction (in French)	124
4.3	Multifractal analysis and hyperspectral imaging	125
4.3.1	Context	125

4.3.2	Goals and scope	126
4.4	Numerical experiments	126
4.4.1	Analyzed hyperspectral image	126
4.4.2	Model fitting	127
4.4.3	Spectral evolution of multifractal features for hyperspectral images	129
4.4.4	Spatio-spectral evolution of multifractal features for hyperspectral images . . .	131
4.5	Conclusions and perspectives	137
4.6	Conclusions et perspectives (in French)	138
5	Conclusions and future work	141
	Appendices	153
A	Model fit for synthetic data	153
B	Marginalization of the latent variables	157
C	Appendix of Chapter 3	159
C.1	Derivates of the potential energy	159
C.2	Sampling of \mathbf{W}	160
C.2.1	Conditional prior distribution	160
C.2.2	Conditional posterior distribution	161
D	Whittle approximation for time series	163
E	Expectation-maximization	167
	Bibliography	189

List of Figures

1.1	Definition of wavelet leaders of images	17
1.2	Realizations of 2D fBm with different values of H	22
1.3	Realizations of 2D CMC-LN with different values of c_2	23
1.4	Realizations of 2D CMC-LP with different values of c_2	24
1.5	Realizations of 2D MRW with different values of c_2	25
1.6	Realizations of 2D CPC-LN with different values of c_2	26
1.7	Realizations of 2D CPC-LP with different values of c_2	27
2.1	Quantile-quantile plots of the empirical distributions of multifractal processes and associated wavelet coefficients and log-leaders	36
2.2	Fitting between the sample covariance and the proposed covariance model (2D CMC-LN)	39
2.3	Frequency grid and set J_j	42
2.4	Fitting between the periodogram and the proposed parametric spectral density (2D CMC-LN)	45
2.5	Functions $f_{1,j}$ and $f_{2,j}$ for different image sizes N	47
2.6	Influence of the bandwidth parameter η	61
2.7	Marginal posterior distribution of c_2	62
2.8	Estimation performance for c_1 for 2D MRW	64
2.9	Estimation performance for c_2 for 2D MRW	65
2.10	Estimation performance for c_2 for different 2D MMC processes	66
2.11	Estimation performance for c_2 for 2D MRW (small image sizes)	67

2.12	Computational time versus image size	70
2.13	Convergence diagnosis of MCMC algorithms	70
3.1	Multivariate images scenario	82
3.2	Proposed bipartite conditional independence graphs	88
3.3	Heterogeneous multifractal synthetic dataset	94
3.4	Illustration of the influence of the GMRF hyperparameters on the estimation of c_2 . . .	95
3.5	Estimation results for a temporal sequence of heterogeneous 2D MRWs	97
3.6	Estimation performance for c_2 with GMRF prior on temporal sequences of heteroge- neous 2D MRWs	99
3.7	Estimation performance for c_2 with GMRF prior on temporal sequences heterogeneous 2D MRWs (lines)	100
3.8	Estimation performance for c_2 with SAR prior on sequences of multi-temporal images	111
3.9	Estimation performance for c_2 and c_1 on sequences of multi-temporal images	121
4.1	Real hyperspectral data acquired by the Hypspx hyperspectral scanner.	127
4.2	Fitting between data and proposed statistical model (forested area, large patch) . . .	128
4.3	Fitting between data and proposed statistical model (forested area, small patch) . . .	129
4.4	Multifractal features of hyperspectral images	131
4.5	Estimate \hat{c}_1 of the hyperspectral image (bands 80 – 160) from the Madonna project. .	134
4.6	Discriminative power of estimation procedures for c_1	135
4.7	Estimate \hat{c}_2 of the hyperspectral image (bands 80 – 160) from the Madonna project. .	139
4.8	Discriminative power of estimation procedures for c_2	140
A.1	Quantile-quantile plots of the empirical distributions of multifractal processes and associated wavelet coefficients and log-leaders	154
A.2	Fitting between the sample covariance and the proposed covariance model (MRW) . .	155
A.3	Fitting between the sample covariance and the proposed covariance model (CPC-LN)	156
E.1	Computational time versus sample size (EM included)	169

List of Tables

2.1	Estimation performance of Bayesian estimators	62
2.2	2D FBm estimation performance for different image sizes	68
3.1	Influence of the GMRF hyperparameters on the estimation performance for c_2	95
3.2	Overall estimation performance for c_2	98

Introduction

Context and objectives of the thesis

Over the last decades, multifractal analysis (MFA) has matured into a powerful signal and image processing tool and enables the study of the scale invariance and pointwise regularity properties of data. To be more specific, MFA describes data through the collections of its pointwise regularity exponents, which quantify the degree of smoothness at each point. Their geometric properties are then summarized in the so-called multifractal spectrum, which provides a global description of the fluctuations of the pointwise regularity of data magnitude along time or space [Jaf04, JAW15]. The estimation of the multifractal spectrum constitutes the primary goal of MFA. This is in practice achieved via a so-called multifractal formalism, which essentially establishes a link between the multifractal spectrum and the scale invariance properties of the statistics of multiresolution coefficients derived from the data under analysis.

MFA has been successfully used in a large range of signal processing applications from a wide array of applicative domains, e.g., biomedical applications (body rhythms [HPP⁺96], infra slow brain activity [CVA⁺12]), geophysics [FGK14], finance [MS98] or Internet traffic [ABF⁺02], to name but a few, and more recently its use has been also reported in an increasing number of image processing applications from various fields including, for instance, texture classification [XYLJ10, WAJ⁺09], biomedical imaging [BPL⁺01, KLSJA01, LB09], physics [PBA⁺06, RAD00], biology [SS01], climate research [LS13], art investigation [CERW08, AJW13, JMS⁺14] and cloud image analysis [RAD00].

Despite these past successes, MFA suffers at present from two major limitations. First, the current

benchmark estimation procedures fail at providing accurate estimates of parameters characterizing multifractal properties, notably for small sample sizes, which prevents the relevant use of MFA in applications where sample size is limited (e.g., for the characterization of biomedical time series or for a patch-based analysis of images). Second, these estimation procedures are intrinsically univariate and can only be used to process one single data at a time. However, in an increasing number of applications, the acquired data are multivariate (e.g., multi-temporal, multi-band, multi-modal ...). The diversity of such data constitutes a rich resource for information about the underlying data generation mechanisms. Extracting this information requires the whole dataset to be processed jointly instead of each component individually.

Therefore, the main goal of this work is to develop practical contributions to overcome these two fundamental limitations. The strategy adopted in this thesis consists of embedding the estimation of the multifractal parameters in a Bayesian framework. To this end, we introduce in Chapter 2 a novel second-order statistical model for the logarithm of wavelet leaders, which are the multiresolution coefficients yielding the current benchmark multifractal formalism. This model is parametrized by multifractal parameters of interest, allowing hence their estimation to be formulated in a Bayesian framework. Then, the multifractal analysis of multivariate data is addressed in Chapter 3 by generalizing this Bayesian framework to hierarchical models able to account for the assumption that multifractal properties evolve smoothly in the dataset under analysis. This is achieved via the design of suitable priors relating the dynamical properties of the multifractal parameters of the different components composing the dataset.

Furthermore, this thesis studies the potential benefits of MFA in general and of the proposed Bayesian methodology in particular for remote sensing via the example of hyperspectral imaging. This study is motivated by the increasing spatial resolution of hyperspectral remote sensors, which requires the development of new processing methods capable of combining both spectral and spatial information. In Chapter 4, we investigate the use of multifractal parameters to extract spatial information in terms of the fluctuations of the point-wise regularity of image amplitudes.

Finally, note that, in view of the targeted application, all developments are stated for images in this manuscript. However, the proposed methodology carries over to other data straightforwardly

(e.g., time series, as detailed in Appendix D).

The work presented in this thesis has been carried out in the Institut de Recherche en Informatique de Toulouse (France), within the Signal and Communication group, and the School of Engineering and Physical Sciences of Heriot-Watt University (Scotland). This thesis has been funded by the Direction Générale de l'Armement (DGA), French ministry, within the UK-France PhD program which is a PhD funding program started in 2012 that is jointly managed by Defence Science and Technology Laboratory (DSTL) and DGA.

Structure of the manuscript

Chapter 1 recalls the key theoretical and practical concepts of MFA. The current benchmark multifractal formalism, constructed on wavelet leaders, is notably defined. This formalism yields a polynomial expansion of the multifractal spectrum in terms of the coefficients c_p termed log-cumulants. The leading order log-cumulants provide an efficient way to summarize the multifractal properties and can hence be used in applications instead of the multifractal spectrum. The present thesis focuses on the estimation of the two first log-cumulants c_1 and c_2 , respectively quantifying the average pointwise regularity and the degree of fluctuation of the pointwise regularity across space.

Chapter 2 introduces a Bayesian approach for the estimation of c_1 and c_2 for a single image. The proposed strategy relies on the construction of 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 mean and variance-covariance structures are respectively controlled by the log-cumulants c_1 and c_2 . To bypass the practically infeasible computation of the associated Gaussian likelihood, a Whittle approximation is devised that enables its efficient evaluation in the spectral domain. We introduce Bayesian models building on this approximation and incorporating prior distributions for the log-cumulants. The associated Bayesian estimators are computed via Markov Chain Monte Carlo (MCMC) algorithms. The benefits of the Bayesian approach are illustrated via the investigation of estimation performance, assessed using a large number of synthetic

2D multifractal processes for several image sizes. Results presented in this chapter show that the proposed estimation procedures provide excellent estimation performance for c_2 both for small and large image sizes. On the opposite, they indicate no significant improvement for the estimation of c_1 when non-informative priors are designed.

Chapter 3 introduces an operational approach for the multifractal analysis of multivariate data. To do so, we formulate a hierarchical Bayesian model elaborating on the statistical model formulated in Chapter 2 for a single image and incorporating multivariate priors to encode beliefs on the dynamics of multifractal properties. We focus in this chapter on the prior assumption that multifractal properties evolve slowly between neighboring components of the multivariate image. We investigate different smoothing priors for the collections of the first and second log-cumulants of the components composing the dataset. More precisely, we introduce two probabilistic models to describe the collection of the second log-cumulants c_2 : gamma Markov random fields (GMRF) and simultaneous autoregressions (SAR). For the collection of the first log-cumulants c_1 , we consider the use of a SAR prior. MCMC algorithms allowing to build Bayesian estimators are then presented. The performance of the smoothing priors is assessed and validated by means of numerical experiments conducted on multivariate images composed of synthetic multifractal processes with prescribed multifractal properties. Reported results clearly show a significant improvement in parameter estimation performance when compared to univariate formulations.

Chapter 4 investigates the potential use and benefits of MFA (and of the multivariate Bayesian multifractal methodology introduced in Chapter 3) for the extraction of spatial information in hyperspectral imaging. We present experiments conducted on a real-world hyperspectral image. Results presented in this chapter suggest that multifractal parameters are relevant spatial/textural attributes which can in turn potentially be employed in tasks such as classification or segmentation. They moreover demonstrate that the proposed multivariate Bayesian methodology is an operational MFA tool applicable to the analysis of real-world multivariate images.

Chapter 5 concludes the present thesis and provides a guideline for future work.

Main contributions

Chapter 2. The main contribution of this chapter lies, on one hand, in the formulation of a novel parametric statistical model that is central for embedding the estimation of multifractal parameters for images in a Bayesian framework. On the other hand, this chapter contains two contributions enabling efficient estimation procedures to be devised. The first key contribution resides in the formulation and use of a Whittle approximation in which the numerically problematic Gaussian likelihood associated with the model is evaluated in the spectral domain. The second contribution is the reformulation of this approximation as the marginal likelihood of an augmented model. This data augmentation scheme yields an augmented likelihood whose convenient form enables the design of conjugate priors, allowing hence the parameter inference to be achieved efficiently by means of straightforward sampling. The devised procedure constitutes the first operational Bayesian estimator for the log-cumulants c_1 and c_2 for images that is effective both for small and large image sizes.

Chapter 3. The overarching methodological contribution of this chapter resides in the proposition of procedures for performing the joint MFA of multivariate images in a Bayesian framework via hierarchical models. The large numbers of unknowns induced by the analysis of multivariate images requires a special care in the specification of the priors and in the design of inference algorithms. To that end, we develop several contributions. First, the combination of a GMRF prior for c_2 and the augmented likelihood introduced in Chapter 2 enables the exploitation of conjugacy and yields an efficient estimation procedure tailored for the handling of large numbers of unknowns. Second, a SAR prior for c_2 , which is not conjugate with the proposed statistical model, requires the design of appropriate MCMC algorithms but enables a straightforward estimation of the regularization hyperparameters. We introduce a Hamiltonian Monte Carlo (HMC) scheme providing good convergence properties for high-dimensional variable spaces. This constitutes another contribution of this chapter. Finally, a SAR prior for c_1 is here proposed since it is conjugate with the model and hence yields an efficient inference algorithm. Overall, these contributions lead to the first operational MFA tool

applicable to the joint analysis of multivariate images.

Chapter 4. This chapter contains two original contributions. First, it illustrates that the Bayesian methodology developed in this thesis provides an operational MFA tool that can be applied on real-world multivariate images. Second, it investigates for the first time the use of multifractal parameters for the extraction of spatial/textural information in a context of hyperspectral imaging.

Contexte et objectifs de la thèse (in French)

Au cours de ces dernières décennies, l'analyse multifractale est devenue un outil puissant de traitement du signal et des images, permettant l'étude des propriétés d'invariance d'échelle et de la régularité locale des données. Plus précisément, l'analyse multifractale décrit un signal ou une image à travers la description du comportement de ses exposants de régularité ponctuelle. Leurs propriétés géométriques sont résumées dans le ainsi nommé spectre multifractal, qui décrit de manière globale les fluctuations de la régularité locale [Jaf04, JAW15]. L'estimation du spectre multifractal est au cœur de l'analyse multifractale. Elle est en pratique réalisée via un formalisme multifractal, qui, en substance, établit un lien mathématique entre le spectre multifractal et les propriétés d'invariance d'échelle des statistiques des coefficients multirésolutions calculés à partir du signal ou de l'image en cours d'analyse. L'analyse multifractale a été mise à profit dans un large spectre d'applications de traitement du signal liées à des domaines variés, comme par exemple des applications biomédicales (concernant le rythme corporel [HPP⁺96] ou l'activité cérébrale [CVA⁺12]), la géophysique [FGK14], la finance [MS98], le contrôle du trafic Internet [ABF⁺02]. Plus récemment son utilisation a été rapportée dans un nombre croissant d'applications de traitement des images, incluant la classification de textures [XYLJ10, WAJ⁺09], l'imagerie biomédicale [BPL⁺01, KLSJA01, LB09], la physique [PBA⁺06, RAD00], la biologie [SS01], la climatologie [LS13], l'étude d'oeuvres d'art [CERW08, AJW13, JMS⁺14] ou encore l'analyse d'images de nuages [RAD00].

Malgré cette profusion d'exemples de son utilisation fructueuse, l'analyse multifractale souffre à présent de deux limitations majeures. Premièrement, les procédures d'estimation de référence ne permettent pas une estimation précise des paramètres caractérisant les propriétés multifractales, notamment pour les petites tailles d'échantillon, empêchant notamment l'utilisation pertinente de l'analyse multifractale à des applications où les données sont de petite taille (comme par exemple l'étude de série temporelles biomédicales, ou une analyse d'images à partir de patches). Deuxièmement, ces procédures d'estimation sont intrinséquement univariées et ne peuvent être appliquées que pour l'analyse d'un signal ou d'une image à la fois. Cependant, dans un nombre croissant d'applications les données recueillies sont multivariées (par exemple multi-temporelles, multi-bandes ou multi-modales).

La diversité de telles données constitue une source importante d'information sur le mécanisme à l'origine des données. Avoir accès à cette information nécessite de conduire une analyse conjointe du jeu de données dans son ensemble plutôt qu'une analyse image par image.

Ainsi, l'objectif principal de cette thèse est la mise en place de contributions permettant de pallier ces deux limitations fondamentales. La stratégie adoptée dans cette thèse consiste à formuler l'estimation des paramètres multifractaux dans un cadre bayésien. A cet effet, nous introduisons dans le Chapitre 2 un nouveau modèle statistique de second-ordre pour le logarithme des coefficients dominants, qui sont les coefficients multirésolutions soutenant le formalisme multifractal de référence. Ce modèle est paramétrisé par les paramètres multifractaux d'intérêt, permettant ainsi d'effectuer leur estimation dans un cadre bayésien. Ensuite, l'analyse multifractale d'images multivariées est abordée dans le Chapitre 3 via la généralisation de ce cadre bayésien à des modèles hiérarchiques capables de prendre en compte l'hypothèse d'une évolution lente des propriétés multifractales dans le jeu de données. Ceci est réalisé en définissant des lois *a priori* adaptées reliant les propriétés dynamiques des paramètres multifractaux des différents éléments composant le jeu de données.

De plus, cette thèse étudie le potentiel apport de l'analyse multifractale en général et de la méthode bayésienne proposée en particulier pour la télédétection à travers l'exemple de l'imagerie hyperspectrale. Cette étude est motivée par la constante amélioration de la résolution spatiale des actuels capteurs hyperspectraux, nécessitant dès lors le développement de nouvelles méthodes de traitement capables de combiner à la fois l'information spatiale et spectrale. Dans le Chapitre 4, nous étudions l'utilisation des paramètres multifractaux pour l'extraction de l'information spatiale en terme de fluctuations de la régularité locale de l'amplitude des images.

Le travail présenté dans cette thèse a été conduit à l'Institut de Recherche en Informatique de Toulouse (France), au sein de l'équipe Signal et Communication, et le département ingénierie et physique de l'université de Heriot-Watt (Ecosse). Cette thèse a été financée par la Direction Générale de l'Armement (DGA), ministère de l'armement, dans le cadre du programme franco-britannique de financement de thèses, lancé en 2012 et conjointement dirigé par le Defence Science and Technology Laboratory (DSTL) et la DGA.

List of publications

Journals.

1. S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, P. Abry, *Multifractal analysis of multivariate images using gamma Markov random field priors. To be submitted.*
2. S. Combrexelle, H. Wendt, N. Dobigeon, J.-Y. Tourneret, S. McLaughlin, P. Abry, *Bayesian estimation of the multifractality parameter for image texture using a Whittle approximation*, IEEE Trans. Image Proces., vol. 24, no. 8, pp. 2540-2551, Aug. 2015.

International conferences.

1. S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, P. Abry, *Bayesian joint estimation of the multifractality parameter of image patches using gamma Markov random field priors*, IEEE International Conference on Image Processing (ICIP), Phoenix, Arizona, USA, September 2016. *To appear.*
2. S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, P. Abry, *Bayesian estimation for the local assessment of the multifractality parameter of multivariate time series*, 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, August 2016. *To appear.*
3. S. Combrexelle, H. Wendt, J.-Y. Tourneret, P. Abry, S. McLaughlin, *Bayesian multifractal analysis of multi-temporal images using smooth priors*, IEEE Workshop on Statistical Signal Processing (SSP), Palma de Mallorca, Spain, June 2016.
4. S. Combrexelle, H. Wendt, J.-Y. Tourneret, Y. Altmann, S. McLaughlin, P. Abry, *Bayesian approach for the multifractal analysis of spatio-temporal data*, International Conference on Systems, Signal and Image Processing (IWSSIP), Bratislava, Slovakia, May 2016.
5. S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, P. Abry, *A Bayesian*

- framework for the multifractal analysis of images using data augmentation and a Whittle approximation*, IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP), Shanghai, China, March 2016.
6. S. Combrexelle, H. Wendt, J.-Y. Tournet, P. Abry, S. McLaughlin, *Bayesian estimation of the multifractality parameter for images via a closed-form Whittle likelihood*, 23rd European Signal Processing Conference (EUSIPCO), Nice, France, August 2015.
 7. S. Combrexelle, H. Wendt, J.-Y. Tournet, S. McLaughlin, P. Abry, *Hyperspectral image analysis using multifractal attributes*, IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, June 2015.
 8. S. Combrexelle, H. Wendt, P. Abry, N. Dobigeon, S. McLaughlin, J.-Y. Tournet, *A Bayesian approach for the joint estimation of the multifractality parameter and integral scale based on the Whittle approximation*, IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP), Brisbane, Australia, April 2015.

National conferences.

1. S. Combrexelle, H. Wendt, J.-Y. Tournet, N. Dobigeon, S. McLaughlin, P. Abry, *Estimation bayésienne locale du paramètre de multifractalité à l'aide d'un algorithme de Monte Carlo Hamiltonien*, GRETSI Symposium Signal and Image Processing, Lyon, France, September 2015.

Chapter 1

Multifractal analysis

Contents

1.1	Introduction	12
1.2	Introduction (in French)	12
1.3	Multifractal analysis of images	13
1.3.1	Motivation	13
1.3.2	Hölder exponent	13
1.3.3	Multifractal spectrum	14
1.4	Wavelet leader multifractal formalism	15
1.4.1	Wavelet leaders	16
1.4.2	Wavelet leader multifractal formalism	18
1.4.3	Negative regularity	20
1.5	Two-dimensional scaling processes	20
1.5.1	Fractional Brownian motion and self-similar processes	21
1.5.2	Multifractal multiplicative cascade based processes	21
1.6	Estimation procedures	26
1.6.1	Linear regression	26
1.6.2	Alternative estimation procedures	28
1.7	Conclusion	30
1.8	Conclusion (in French)	30

1.1 Introduction

Multifractal analysis is a widely used signal and image processing tool that aims to characterize data via the fluctuations of their local regularity along time or space. Nowadays, MFA is a theoretically well-grounded tool that has been successfully used in a large number of applications, including texture classification [XYLJ10, WAJ⁺09], biomedical applications [BPL⁺01, KLSJA01], physics [PBA⁺06, RAD00] and art investigation [JSM06, CERW08, AJW13, JMS⁺14], to name but a few.

This chapter aims at providing a brief summary of MFA (for an exhaustive presentation, the reader is referred to [Jaf97a, Jaf97b, DOT03]). In Section 1.3 we define the main theoretical notions underlying MFA. Practical aspects are developed in Section 1.4. In Section 1.5, we introduce a representative selection of scaling processes that will be used throughout this thesis to validate our estimation procedures (cf. Chapters 2 and 3). Finally, we present in Section 1.6 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, JLA06], the choice has been taken to focus the presentation on 2D data, i.e., on images.

1.2 Introduction (in French)

L'analyse multifractale est un outil largement utilisé dans le traitement des signaux et des images. Cet outil permet de caractériser les données à travers les fluctuations (spatiales ou temporelles) de leur régularité locale. De nos jours, l'analyse multifractale jouit d'une solide fondation théorique et a été mise à profit dans de nombreuses applications, incluant par exemple la classification de textures [XYLJ10, WAJ⁺09], l'imagerie médicale [BPL⁺01, KLSJA01], la physique [PBA⁺06, RAD00] ou encore l'analyse d'oeuvres art [JSM06, CERW08, AJW13, JMS⁺14].

Ce chapitre vise à donner une brève introduction à l'analyse multifractale (pour une présentation plus exhaustive, le lecteur est invité à consulter [Jaf97a, Jaf97b, DOT03]). Dans la Section 1.3 nous définissons les principales notions sous-tendant l'analyse multifractale. Les aspects pratiques sont

développés dans la Section 1.4. Dans la Section 1.5 nous introduisons une sélection représentative de processus bidimensionnels à invariance d'échelle qui seront utilisés au cours de la thèse pour valider nos procédures d'estimation (cf. Chapitre 2 et 3). Enfin, nous présentons dans la Section 1.6 les procédures existantes pour l'estimation des attributs caractérisant les propriétés multifractales des images et nous en soulignons les limites.

A noter que, bien que les concepts théoriques et les outils pratiques de l'analyse multifractale puissent être en principe donnés pour toutes les dimensions, cf., [Jaf04, JLA06], le choix a été fait d'axer la présentation sur les données bidimensionnelles, ou autrement dit, sur les images.

1.3 Multifractal analysis of images

1.3.1 Motivation

Since the early origins of digital image processing, texture has been recognized as one of the central characteristic features in images. Texture is a perceptual attribute and different paradigms have been introduced in the literature to characterize it [Har79]. Several authors have proposed to model texture using random fractals, scale invariance or self-similarity, see, e.g., [KCC89, PPLV02]. Indeed, it has been reported in the literature that scale invariant processes are relevant and effective models for textures associated with a large class of natural images, see, e.g., [Cha07, WAJ⁺09, XYLJ10]. The notion of scale invariance is deeply tied to the spatial fluctuations of the pointwise singular behavior, or in other terms, of the local regularity, of the image amplitudes [LVMB92, ADKR03]. MFA provides both the theoretical and the practical framework for its study.

1.3.2 Hölder exponent

In what follows, we denote as $X(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^2 \rightarrow X \in \mathbb{R}$ a 2D function under analysis, where $\mathbf{u} = (u_1, u_2)$ stands for the spatial variable. Throughout, X is assumed to be locally bounded (cf. Section 1.4.3 for further comments on this assumption).

MFA envisages the characterization of the function X through the fluctuations along space of the

local regularity of its amplitude. In MFA, the local regularity of the (locally bounded 2D) function X at position \mathbf{u}_0 is most commonly measured using the so-called Hölder exponent $h(\mathbf{u}_0)$ [Jaf04, Rie03], which is defined as follows. The function X is said to belong to $C^\alpha(\mathbf{u}_0)$ at position \mathbf{u}_0 if there exist $K, \alpha > 0$ and a polynomial $P_{\mathbf{u}_0}$ of degree smaller than α such that

$$\|X(\mathbf{u}) - P_{\mathbf{u}_0}(\mathbf{u})\| \leq K\|\mathbf{u} - \mathbf{u}_0\|^\alpha \quad (1.1)$$

where $\|\cdot\|$ is the Euclidian norm. The Hölder exponent at position \mathbf{u}_0 is the largest value of α such that this inequality holds, i.e.,

$$h(\mathbf{u}_0) = \sup\{\alpha : X \in C^\alpha(\mathbf{u}_0)\}. \quad (1.2)$$

Qualitatively, the smaller $h(\mathbf{u}_0)$, the rougher X at spatial location \mathbf{u}_0 and the larger $h(\mathbf{u}_0)$, the smoother X at \mathbf{u}_0 . Note that the definition (1.2) is meaningful only for non-negative exponents, which is guaranteed by the assumption that X is locally bounded. However, it has been reported that real-world images often violate this assumption, see, e.g., [WAJ⁺09, WRJA09]. We discuss in Section 1.4.3 a practical solution to relax this assumption.

1.3.3 Multifractal spectrum

MFA characterizes a function through a global description of the spatial fluctuations of the local regularity $h(\mathbf{u})$ [Jaf04, JAW15]. This is achieved via the so-called multifractal spectrum $D(h)$. The multifractal spectrum is defined as the Hausdorff dimension (denoted \dim_H) of the sets of points at which the Hölder exponent takes the same value, i.e.,

$$D(h) = \dim_H(\mathbf{u} : h(\mathbf{u}) = h) \quad (1.3)$$

where the Hausdorff dimension implies $0 \leq D(h) \leq 2$ and, by convention, $D(h) = -\infty$ for $h \notin [h_m, h_M]$, where h_m and h_M correspond to the minimum and maximum Hölder exponents observed in X , respectively. Broadly speaking, $D(h)$ is a measure of the relative importance of the regularity h in X . For more technical details on the multifractal spectrum and a precise definition of the Hausdorff dimension, see, e.g., [Jaf04, JAW15].

The estimation of the multifractal spectrum $D(h)$ is the central goal of MFA. However, the mathematical definition of $D(h)$ cannot be used for its estimation as, in practice, only a digital image is available, i.e., in other words, a discretized version of X with finite resolution. This is instead achieved via a so-called multifractal formalism that allows us to estimate $D(h)$ from practically computable quantities [Jaf04].

1.4 Wavelet leader multifractal formalism

In essence, multifractal formalisms establish a link between the multifractal spectrum and the moments of multiresolution coefficients, which are quantities that capture the content of the image X around a given spatial location for a given frequency scale. Historical examples are the increments [FP85] or the wavelet coefficients [AAD⁺02]. These choices of multiresolution quantities lead to multifractal formalisms with both theoretical and practical limitations studied in, e.g., [LJA05, WAJ07, WRJA09, Jaf04]. The multifractal formalism developed in [MBA93, ADKR03] is based on the wavelet coefficient modulus maxima. It overcomes the most important numerical limitations but at the price of an increased computational cost. In a different spirit, let us mention the multifractal detrended fluctuation analysis which is a technique derived from the detrended fluctuation analysis, originally introduced in [KZKB⁺02] (see [LWA⁺16] for recent theoretical and practical investigations).

The multiresolution coefficients considered in the present thesis are the recently introduced wavelet leaders [LJA05, Jaf04], which are derived from the wavelet coefficients. While wavelet coefficients tend to take values close to 0, preventing in practice negative moments to be considered in the estimation of the multifractal spectrum, wavelet leaders are specifically tailored for the robust estimation of the entire multifractal spectrum (see [LJA05, Jaf04, WAJ07] for details). They yield the current benchmark multifractal formalism, which is briefly recalled in what follows.

1.4.1 Wavelet leaders

We denote as $\{X(\mathbf{k}), \mathbf{k} = (k_1, k_2)\}$ the discretized version of the locally bounded function X . For simplicity, and without loss of generality, the (discretized) image is assumed to be square, i.e., $k_i = 1, \dots, N$ for $i = 1, 2$.

Wavelet coefficients. A two-dimensional (2D) orthonormal discrete wavelet transform (DWT) can be obtained as the tensor product of one-dimensional (1D) DWT as follows. Let $H_0(k)$ and $G_0(k)$ denote the low-pass and high-pass filters defining a 1D DWT. These filters are associated with a mother wavelet ψ , characterized by its number of vanishing moments $N_\psi > 0$ ¹. Four 2D filters $G^{(m)}(\mathbf{k})$, $m = 0, \dots, 3$, are defined by tensor products of $H_0(k)$ and $G_0(k)$. The 2D low-pass filter

$$G^{(0)}(\mathbf{k}) = H_0(k_1)H_0(k_2) \quad (1.4)$$

yields the approximation coefficients $D_X^{(0)}(j, \mathbf{k})$, whereas the high-pass filters defined by

$$G^{(1)}(\mathbf{k}) = H_0(k_1)G_0(k_2), \quad G^{(2)}(\mathbf{k}) = G_0(k_1)H_0(k_2) \quad \text{and} \quad G^{(3)}(\mathbf{k}) = G_0(k_1)G_0(k_2) \quad (1.5)$$

yield the wavelet (detail) coefficients $D_X^{(m)}(j, \mathbf{k})$, $m = 1, 2, 3$ as follows: at the finest scale $j = 1$, the $D_X^{(m)}(j, \mathbf{k})$, $m = 0, \dots, 3$ are obtained by convolving the image X with $G^{(m)}$, $m = 0, \dots, 3$, and decimation; for the coarser scales $j \geq 2$ they are obtained iteratively by convolving $G^{(m)}$, $m = 0, \dots, 3$, with $D_X^{(0)}(j-1, \cdot)$ and decimation. For scaling and MFA purposes, the approximation coefficients $D_X^{(0)}$ are discarded and it is common to normalize the wavelet coefficients according to the L^1 -norm

$$d_X^{(m)}(j, \mathbf{k}) = 2^{-j} D_X^{(m)}(j, \mathbf{k}), \quad m = 1, 2, 3 \quad (1.6)$$

so that they reproduce the self-similarity exponent for self-similar processes [ADKR03]. For a formal definition and details on (2D) wavelet transforms, the reader is referred to [Mal08, AMVA04].

Wavelet leaders. Denote as

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

¹In the rest of this thesis, all numerical results are obtained with a Daubechies mother wavelet ψ with $N_\psi = 2$ vanishing moments.

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

$$3\lambda_{j,\mathbf{k}} = \bigcup_{n_1, n_2 \in \{-1, 0, 1\}} \lambda_{j, \mathbf{k}_1 + n_1, \mathbf{k}_2 + n_2}$$

the union of this cube with its eight neighbors. The wavelet leaders are defined as the largest wavelet coefficient modulus within this neighborhood over all finer scales [Jaf04]

$$L(j, \mathbf{k}) = L(\lambda_{j,\mathbf{k}}) = \sup_{m \in (1, 2, 3), \lambda' \subset 3\lambda_{j,\mathbf{k}}} |d_X^{(m)}(\lambda')|. \quad (1.7)$$

The definition above is illustrated in Fig. 1.1. A key property of wavelet leaders is that they reproduce

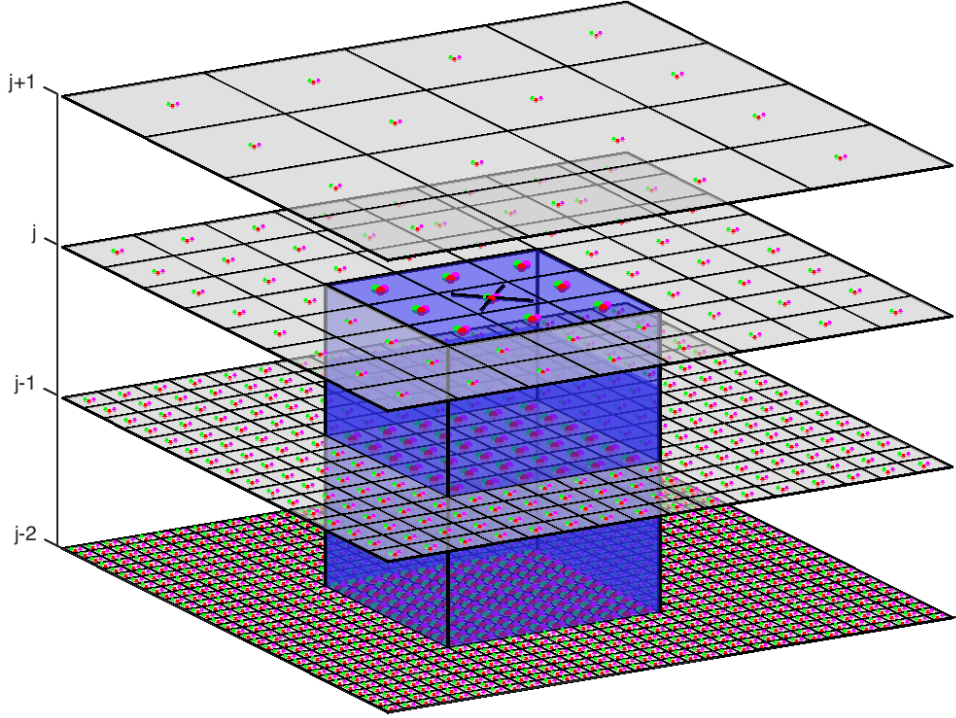


Figure 1.1: Definition of wavelet leaders of images: at scale 2^j and spatial position \mathbf{k} , the wavelet leaders $L(j, \mathbf{k})$ (\times) is taken as the largest modulus of the wavelet coefficients $d_X^{(m)}(j, \mathbf{k})$, $m = 1, 2, 3$ (\bullet , \bullet , \bullet) within the spatial neighborhood $3\lambda_{j,\mathbf{k}}$ (\blacksquare).

the Hölder exponent in the limit of fine scales as follows

$$h(\mathbf{u}_0) = \liminf_{j \rightarrow -\infty} \frac{\log L(\lambda_{j,\mathbf{k}}(\mathbf{u}_0))}{\log 2^j} \quad (1.8)$$

where $\lambda_{j,\mathbf{k}}(\mathbf{u}_0)$ denotes the cube at scale j including the spatial location \mathbf{u}_0 [Jaf04]. It has been shown that (1.8) is the theoretical property required for constructing a multifractal formalism, see [Jaf04] for details. The wavelet leader multifractal formalism is described in the next section and establishes the link between the statistics of the leaders and the multifractal spectrum $D(h)$.

1.4.2 Wavelet leader multifractal formalism

Scaling function and Legendre spectrum. Let us define the structure functions as the spatial averages of the q -th order of the leaders at scale 2^j , i.e.,

$$S(j, q) = \frac{1}{n_j} \sum_{\mathbf{k}} L(j, \mathbf{k})^q \quad (1.9)$$

where $n_j \approx \lfloor N^2/2^{2j} \rfloor$ denotes the number of wavelet leaders at scale 2^j . The scaling function is then defined as

$$\zeta(q) = \liminf_{j \rightarrow -\infty} \frac{\log S(j, q)}{\log 2^j}. \quad (1.10)$$

Eq. (1.10) suggests that, in the limit of fine scales, $S(j, q)$ exhibits a power law behavior with respect to the scale 2^j

$$S(j, q) \approx K_q 2^{j\zeta(q)} \quad j \rightarrow -\infty \quad (1.11)$$

where exponents $\zeta(q)$, termed the scaling exponents, characterize the mechanisms relating scales. Empirical MFA can be sketched as follows: estimating the function $q \rightarrow \zeta(q)$ from the image and then computing its Legendre transform

$$\mathcal{L}(h) = \min_{q \neq 0} (2 + qh - \zeta(q)). \quad (1.12)$$

The interest for $\mathcal{L}(h)$, termed the Legendre spectrum, lies in the fact that, under the local boundness condition for X , it provides an upper bound of the multifractal spectrum [Jaf04, JLA06]

$$D(h) \leq \mathcal{L}(h). \quad (1.13)$$

In practice the Legendre spectrum $\mathcal{L}(h)$ is the (only) quantity that can be computed (in a robust way) and practical MFA focuses on its estimation. For this reason, and as is common practice,

the Legendre spectrum $\mathcal{L}(h)$ is not distinguished from and will be used instead of the theoretical spectrum $D(h)$ from now on.

Log-cumulants. In the seminal work [CGM93], it has been proposed to consider a Taylor expansion of $\zeta(q)$ at $q = 0$, yielding the following polynomial expansion (see also [DMA01, AAD⁺02])

$$\zeta(q) = \sum_{p=1}^{+\infty} c_p \frac{q^p}{p!} \quad (1.14)$$

where the so-called log-cumulants c_p can be shown to be tied to the logarithm of leaders $\log L(\cdot, j)$ through the key relation

$$\text{Cum}_p[\log L(\cdot, j)] = c_{0,p} + c_p \log 2^j \quad (1.15)$$

with $\text{Cum}_p[\cdot]$ denoting the p -th order cumulant. When injecting (1.14) in (1.12), it can be shown that (calculations are reported in [Wen08] for the wavelet leader multifractal formalism) this translates into an expansion of the multifractal spectrum of the image X in terms of c_p , $p \geq 1$, [WRJA09, Wen08, JAW15]

$$\mathcal{L}(h) = 2 + \frac{c_2}{2!} \left(\frac{h - c_1}{c_2} \right)^2 + \frac{-c_3}{3!} \left(\frac{h - c_1}{c_2} \right)^3 + \frac{-c_4 + 3c_3^2/c_2}{4!} \left(\frac{h - c_1}{c_2} \right)^4 + \dots \quad (1.16)$$

where $c_2 \leq 0$ due to the concavity of $\zeta(q)$. As a result, measuring the multifractal spectrum $D(h)$, which would theoretically require to estimate $\zeta(q)$ for a whole range of values of q in a finite interval including $q = 0$ and then compute its Legendre transform, can be rephrased in estimating c_p . Notably, the leading order coefficients c_p provide a relevant summary of the multifractal properties of X in applications where it would often not be convenient to handle an entire function $D(h)$ [CGM93, WAJ07, WRJA09, JAW15]. The first log-cumulant c_1 , for instance, is the mode of $D(h)$ and can be read as a measure for the “average” smoothness of X . The coefficient c_2 , referred to as the multifractality or intermittency parameter, is directly related to the width of $D(h)$ and captures the degree of (multifractal) fluctuations of the image regularity. The multifractality parameter is central in MFA since it enables the identification of the two major classes of scale invariant model processes used in applications: self-similar processes for which $c_2 = 0$ and processes based on multifractal multiplicative cascades (MMC) for which c_2 is strictly negative (cf., e.g., [WJA12]). While the former

class is deeply tied to additive constructions, the latter is based on multiplicative constructions and is hence linked to fundamentally different physical principles [Fri95, Man74, Rie03]. Representative examples of processes within each class are briefly described in Section 1.5.

1.4.3 Negative regularity

The wavelet leader multifractal formalism presented above is well defined for locally bounded images only, see [WAJ07, WRJA09, AJW15] for precise definitions and for procedures for assessing this condition in practice. However, as stated in Section 1.3, it has been reported that a large number of real-world images do not satisfy this prerequisite [WAJ⁺09, WRJA09]. In these cases, a practical solution consists of constructing the multifractal formalism using the modified wavelet coefficients

$$d_X^{(m),\alpha}(j, \mathbf{k}) = 2^{\alpha j} d_X^{(m)}(j, \mathbf{k}), \quad \alpha > 0 \quad (1.17)$$

instead of $d_X^{(m)}$ in (1.7). The parameter α can be chosen sufficiently large to ensure that the multifractal formalism is properly defined (see [WRJA09] for details about the theoretical and practical consequences implied by this modification).

It was observed that $\alpha = 0.7$ was a sufficiently large value for the synthetic processes considered in Chapters 2 and 3 and for real-world data analysed in Chapter 4. Therefore, α is not further discussed in the rest of this thesis.

1.5 Two-dimensional scaling processes

There exists in the literature a large variety of scale invariant models to account for the richness of scaling properties observed in real-world data (see, e.g., [Rie03, CRA05, LT05] and references therein, for an overview). The two most prominent classes are the self-similar processes and the multifractal multiplicative cascade processes. In this section, we briefly summarize a representative selection of such processes that will be used in Chapter 2 and 3 to validate the proposed estimation procedures. The synthesis of the realizations of these processes was conducted using numerical synthetis procedures implemented by ourselves.

1.5.1 Fractional Brownian motion and self-similar processes

Self-similar processes are deeply tied to random walks and additive constructions. As opposed to MMC processes, their multifractal properties are controlled by one single parameter. For a precise definition of self-similar processes and details on their properties the reader is referred to [Lam62, ST94] and references therein. Here, we will make (only) use of fractional Brownian motion (fBm), which is the emblematic example of self-similar processes. It was originally introduced by Kolmogorov [Kol41] in a 1D formulation and studied by Mandelbrot [Man74] in a context of hydrodynamic turbulence. FBm is the only Gaussian self-similar process with stationary increments (see, e.g., [ST94]). In the present work, we use the 2D extension of fBms as defined in [Ste02]. Its multifractal spectrum reduces to a dirac as follows

$$D(h) = \begin{cases} 2 & \text{for } h = H \\ -\infty & \text{otherwise} \end{cases} \quad (1.18a)$$

$$c_1 = H \quad (1.18b)$$

$$\forall p \geq 2 \quad c_p = 0. \quad (1.18c)$$

where H is called the Hurst index. Realizations of 2D fBms for different values of H are displayed in Fig. 1.2. In this thesis, we set $H = 0.7$ for numerical experiments reported in Chapter 2.

1.5.2 Multifractal multiplicative cascade based processes

As opposed to self-similar processes, multiplicative cascade processes are based on multiplicative constructions [KP76, Man74, DOT03]. MMC are processes commonly used in applications to model richer multifractal properties than self-similar processes can account for.

a) Canonical Mandelbrot Cascade (CMC)

CMCs are the historical archetypes of multifractal measures, originally introduced in [Yag66] and studied in [Man74, KP76]. In essence, their construction is based on an iterative split-and-multiply

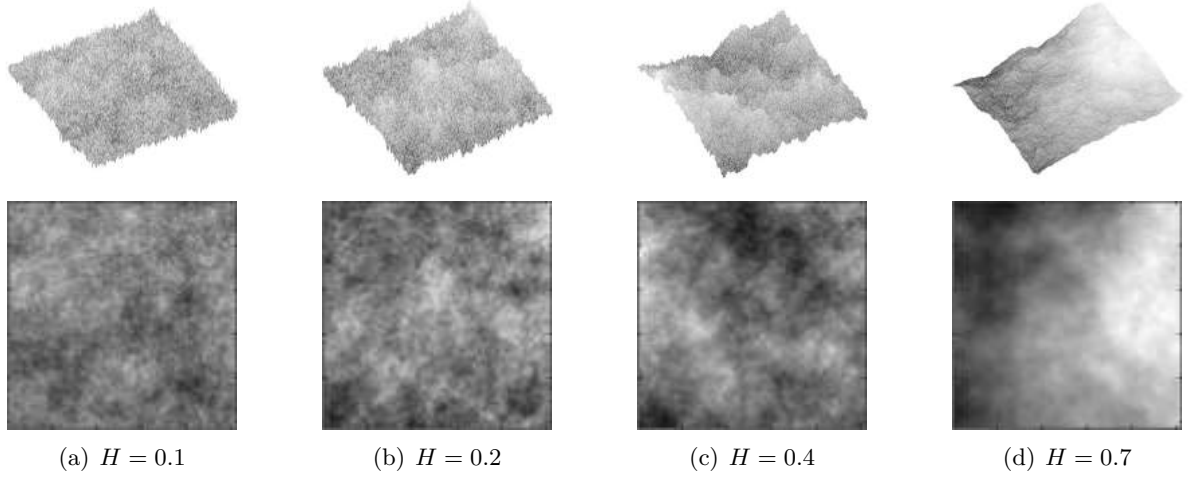


Figure 1.2: Realizations of 2D fBm of size $2^8 \times 2^8$ with different prescribed values of H .

random procedure on an interval involving random multipliers whose statistics determine the multifractal properties of the resulting measure. In this thesis, we use CMCs with two different type of multipliers.

Log-normal multipliers. First, canonical Mandelbrot cascades with log-normal (CMC-LN) multipliers $W = 2^{-U}$, where $U \sim \mathcal{N}(m, 2m/\log 2)$ is a Gaussian random variable parametrized by m . For CMC-LN, the multifractal properties are given by

$$D(h) = \begin{cases} 2 - \frac{(h-m)^2}{4m} & \text{for } h \in [h_m, h_M] \\ -\infty & \text{otherwise} \end{cases} \quad (1.19a)$$

$$c_1 = m > 0; \quad c_2 = -2m \quad (1.19b)$$

$$\forall p \geq 3 \quad c_p = 0. \quad (1.19c)$$

Thus, the value of m can be used to adjust c_1 and c_2 . Realizations of 2D CMC-LN with different values of c_2 are displayed in Fig. 1.3.

Log-Poisson multipliers. Second, canonical Mandelbrot cascades with log-Poisson (CMC-LP) multipliers $W = 2^\gamma \exp(\log(\beta)\pi_\lambda)$, where π_λ is a Poisson random variable with parameter $\lambda =$

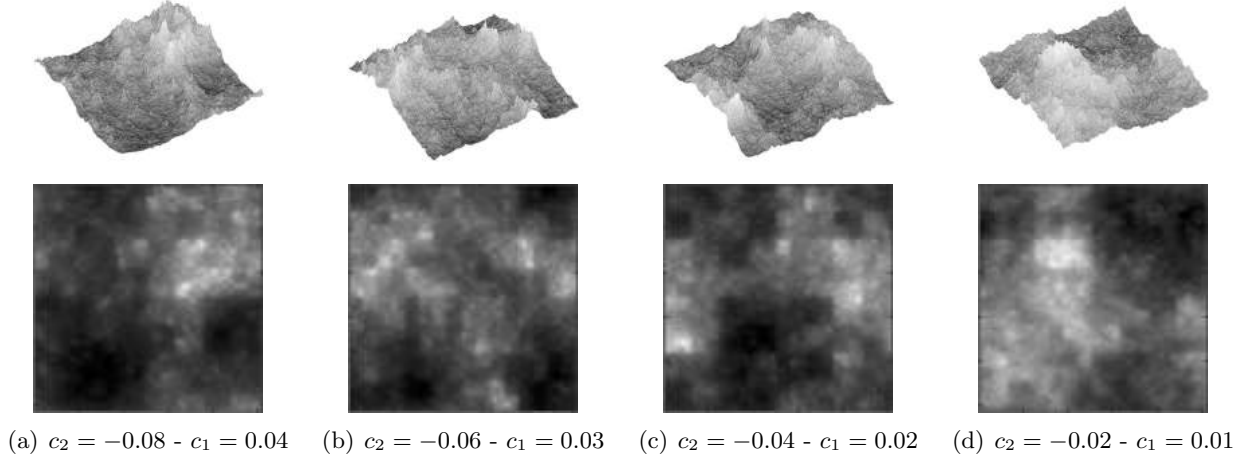


Figure 1.3: Realizations of 2D CMC-LN of size $2^8 \times 2^8$ with different prescribed values of c_p .

$-\frac{\gamma \log 2}{(\beta-1)}$. For CMC-LN, higher-order log-cumulants are non-zero and their multifractal properties are given by

$$D(h) = \begin{cases} 2 + \frac{\gamma}{\beta-1} + \frac{\gamma+h}{\log \beta} \left(\log \left(\frac{(\gamma+h)(\beta-1)}{\gamma \log \beta} \right) - 1 \right) & \text{for } h \in [h_m, h_M] \\ -\infty & \text{otherwise} \end{cases} \quad (1.20a)$$

$$c_1 = \gamma \left(\frac{\log \beta}{\beta-1} - 1 \right) \quad c_2 = -\frac{\gamma}{\beta-1} (\log \beta)^2 \quad (1.20b)$$

$$\forall p \geq 3 \quad c_p = -\frac{\gamma}{\beta-1} (-\log \beta)^p. \quad (1.20c)$$

In Chapter 2, $\gamma = 1.05$ and β is varied according to the value of c_2 (with $\beta > 1$ to ensure well-defined log-cumulants). Realizations of 2D CMC-LP with different values of c_2 are illustrated in Fig. 1.4.

b) Multifractal random Walk (MRW)

MRW is another prominent member of the class of multifractal multiplicative cascade based processes. MRW is a non Gaussian process with stationary increments, originally introduced in a 1D formulation in [BDM01]. Details on its 2D construction can be found in [RV10], where it was introduced as $X(\mathbf{k}) = \sum_{\mathbf{k}} G_H(\mathbf{k}) \exp(\omega(\mathbf{k}))$ where $G_H(\mathbf{k})$ are the increments of a fractional Brownian motion (cf. Section 1.5.1) with parameter H and ω is a Gaussian process that is independent of G_H and has

non trivial covariance $\text{Cov}[\omega(\mathbf{k}_1), \omega(\mathbf{k}_2)] = \beta^2 \log \left(\frac{L}{\|\mathbf{k}_1 - \mathbf{k}_2\| + 1} \right)$, for $L > 0$. Its multifractal properties mimic those of the Mandelbrot's log-normal cascades CMC-LN with

$$D(h) = \begin{cases} 2 - \frac{1}{2\beta^2} (h - H - \frac{\beta^2}{2})^2 & \text{for } h \in [h_m, h_M] \\ -\infty & \text{otherwise} \end{cases} \quad (1.21a)$$

$$c_1 = H + \frac{\beta^2}{2}; \quad c_2 = -\beta^2 \quad (1.21b)$$

$$\forall p \geq 3 \quad c_p = 0. \quad (1.21c)$$

These relations show the meaning of parameters H and β , which can be used to adjust c_1 and c_2 . Typical realizations of 2D MRW are plotted in Fig. 1.5 for a wide range of values of c_2 (and $H = 0.7$). In this thesis (in particular in Chapter 3), MRW will be used for its ease of numerical synthesis.

c) Compound Poisson cascade (CPC)

CPCs were introduced to overcome certain limitations of the CMCs that are caused by their discrete split-and-multiply construction (see, e.g., [BM02, Cha07]). In the construction of CPCs, the localization of the multipliers in the space-scale volume follows a Poisson random process with specific prescribed density. We use CPCs with two different types of multipliers.

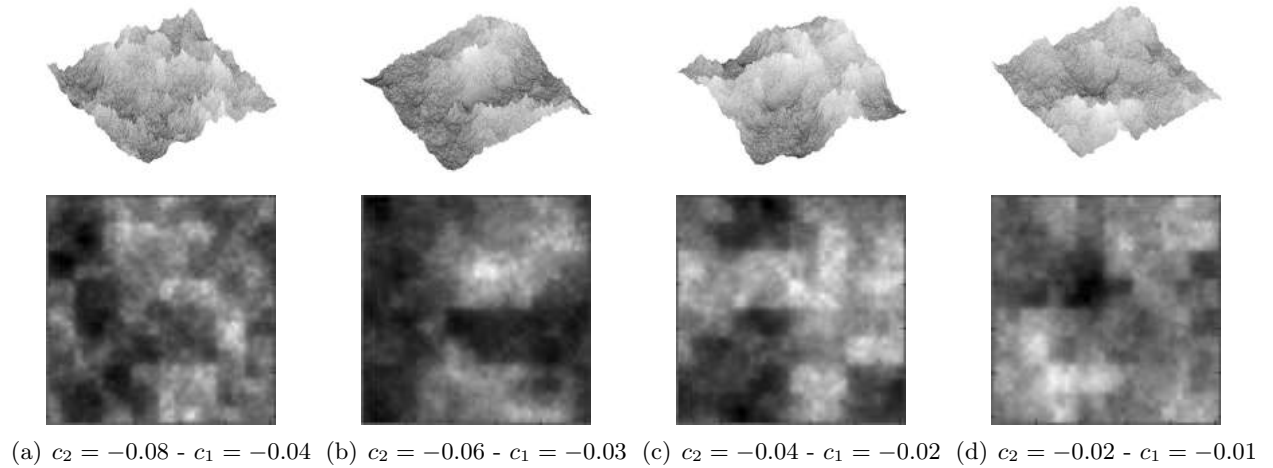


Figure 1.4: Realizations of 2D CMC-LP of size $2^8 \times 2^8$ with different prescribed values of c_2 ($\gamma = 1.05$).

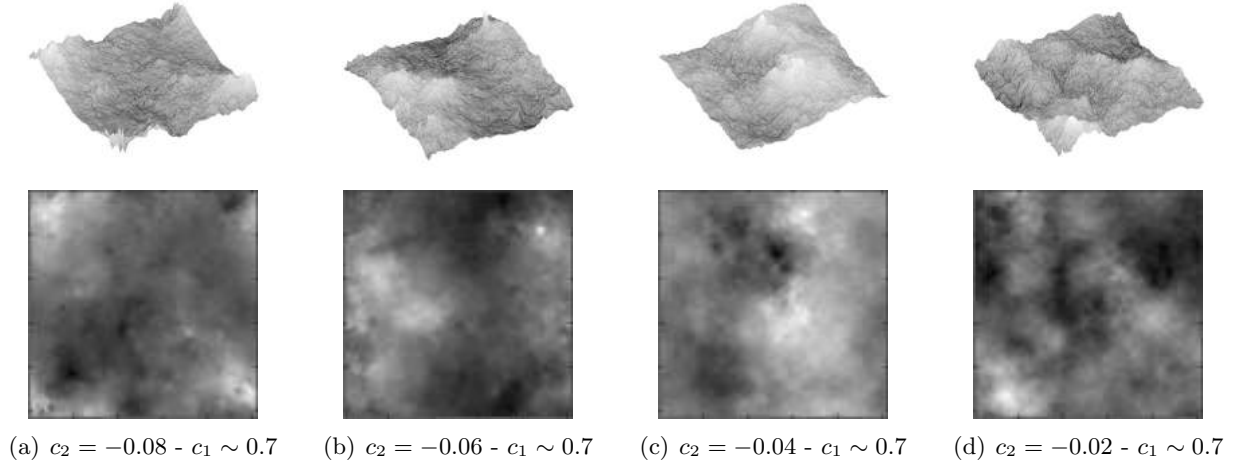


Figure 1.5: Realizations of 2D MRW of size $2^8 \times 2^8$ with different prescribed values of c_2 and $H = 0.7$.

Log-normal multipliers. First the compound Poisson cascades with log-normal (CPC-LN) multipliers $W = \exp(Y)$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$ is a Gaussian random variable. The multifractal properties of CPC-LN are given by

$$D(h) = \begin{cases} \min_{q \neq 0} (2 + qh - \left[\left(1 - \exp\left(\mu q + q^2 \frac{\sigma^2}{2}\right) \right) - q \left(1 - \exp\left(\mu + \frac{\sigma^2}{2}\right) \right) \right]) & \text{for } h \in [h_m, h_M] \\ -\infty & \text{otherwise} \end{cases} \quad (1.22a)$$

$$c_1 = \mu + 1 - \exp\left(\mu + \frac{\sigma^2}{2}\right); \quad c_2 = -(\mu^2 + \sigma^2) \quad (1.22b)$$

$$\forall p \geq 3 \quad c_p \neq 0. \quad (1.22c)$$

In this thesis, $\mu = -0.1$ and σ is varied according to c_2 . Realizations of 2D CPC-LN are illustrated in Fig. 1.6.

Log-Poisson multipliers. Second, the compound Poisson cascades with log-Poisson (CPC-LP) multipliers for which multipliers W are reduced to a constant w . The multifractal properties of the

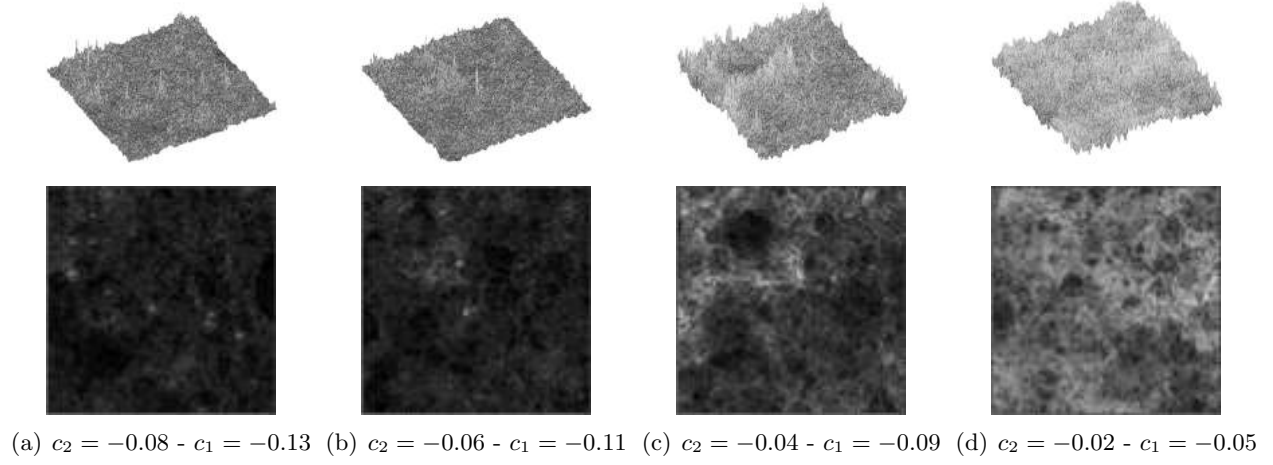


Figure 1.6: Realizations of 2D CPC-LN of size $2^8 \times 2^8$ with different prescribed values of c_2 .

resulting cascades are

$$D(h) = \begin{cases} \min_{q \neq 0} (2 + qh - [1 - w^q + q(1 - w)]) & \text{for } h \in [h_m, h_M] \\ -\infty & \text{otherwise} \end{cases} \quad (1.23a)$$

$$c_1 \neq 0; \quad c_2 = -(\log w)^2 \quad (1.23b)$$

$$\forall p \geq 3 \quad c_p \neq 0. \quad (1.23c)$$

In this thesis, w is varied according to the value of c_2 ². Illustrations of CPC-LP cascades are displayed in 1.7.

1.6 Estimation procedures

1.6.1 Linear regression

As suggested by (1.15) the estimation of log-cumulants c_p can be performed by means of linear regressions in $\log 2^j$ against the sample cumulants denoted $\widehat{\text{Cum}}_p[\log L(j, \cdot)]$, i.e.,

$$\hat{c}_p = \frac{1}{\log 2} \sum_{j=j_1}^{j_2} w_j \widehat{\text{Cum}}_p[\log L(j, \cdot)] \quad (1.24)$$

²No closed-form of c_1 is available for CPC-LP.

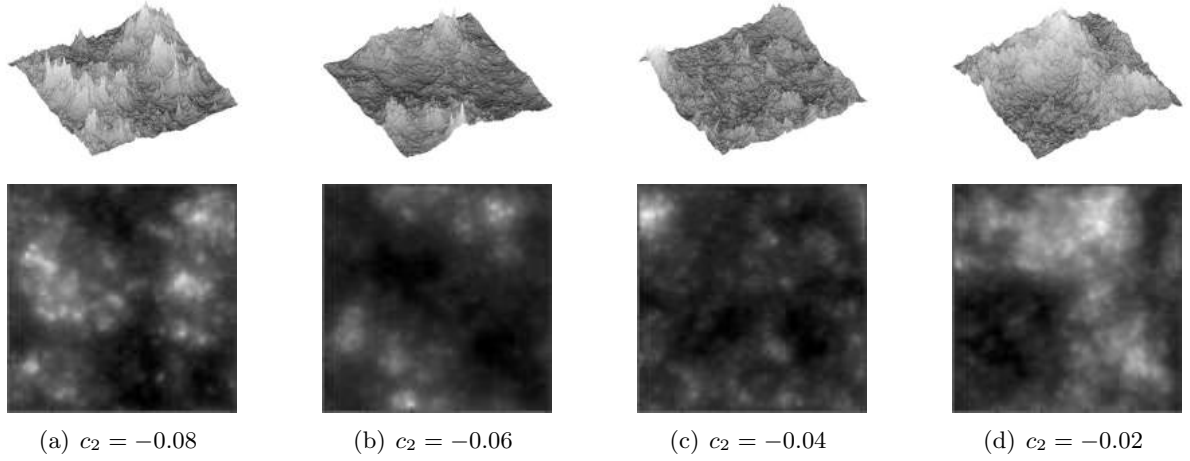


Figure 1.7: Realizations of 2D CPC-LP of size $2^8 \times 2^8$ with different prescribed values of c_2 .

where $\llbracket j_1, j_2 \rrbracket$ is the range of scales for which scale invariance (1.15) holds. In (1.24) the linear regression weights w_j have to satisfy the usual constraints $\sum_{j=j_1}^{j_2} j w_j = 1$ and $\sum_{j=j_1}^{j_2} w_j = 0$. More specifically, weights can be expressed as [VAT03, AFTV00]

$$w_j = b_j \frac{V_0 j - V_1}{V_0 V_2 - V_1^2} \quad (1.25)$$

where

$$V_i = \sum_{j=j_1}^{j_2} j^i b_j, \quad i = 0, 1, 2. \quad (1.26)$$

In (1.25), b_j are positive numbers that can be chosen to reflect the confidence granted to each sample cumulant $\widehat{\text{Cum}}_p[\log L(j, \cdot)]$. In this thesis, we consider the weighted and non-weighted linear regressions defined by $b_j = 1/n_j$ and $b_j = 1$ in (1.25) respectively (see, e.g., [VAT03, AFTV00]). Finally the definition (1.24) directly leads to the following estimators for the first two log-cumulants c_1 and c_2

$$\hat{c}_1 = \frac{1}{\log 2} \sum_{j=j_1}^{j_2} w_j \widehat{\mathbb{E}}[\log L(j, \cdot)] \quad (1.27)$$

and

$$\hat{c}_2 = \frac{1}{\log 2} \sum_{j=j_1}^{j_2} w_j \widehat{\text{Var}}[\log L(j, \cdot)] \quad (1.28)$$

where $\widehat{\mathbb{E}}[\cdot]$ and $\widehat{\text{Var}}[\cdot]$ denote the sample mean and variance.

Limitations. This method is appealing for its simplicity and low computational cost. Yet (1.24) has limited practical estimation performance. In particular, (1.24) yields modest performance for images when compared with 1D signals of equivalent sample size [WRJA09], making it for instance difficult to discriminate between $c_2 \equiv 0$ and values $c_2 < 0$ that are encountered in applications (typically, c_2 lies between -0.01 and -0.08). Moreover, because a sufficient number of scales j must be available to perform the linear regression (1.24), its use remains in practice restricted to images of relatively large size (of order 256^2 pixels). While a similar issue is encountered for the analysis of 1D signals, it is significantly more severe for images: 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 (1.24).

1.6.2 Alternative estimation procedures

There are a limited number of reports in the literature that attempt to overcome the limitations of MFA for images described above. Some of them are mentioned below.

Generalized method of moments (GMM). The generalized method of moments has been proposed and studied in, e.g., [Lux07, Lux08] and formulates parameter inference as the solution (in the least squares sense) of an over-determined system of equations that are derived from the moments of data (see [BKM13, BADML12, BKM08] for applications to financial time series). Yet, the method depends strongly on fully parametric models and yields, to the best of our knowledge, only limited benefits in practical applications.

Maximum likelihood (ML) estimation. Although classical in parameter inference, maximum likelihood estimation has mostly been formulated for a few specific self-similar and multifractal processes (see, e.g., [Ber94]). The main reason for this lies in the complex statistical properties of most of these processes, which exhibit distributions that are strongly non-Gaussian and have intricate algebraically decaying dependence structures that remain poorly studied to date. The same remark is true for their wavelet coefficients and wavelet leaders, see, e.g., [OW00, VWAJ10].

One exception is given by the fractional Brownian motion (in 1D) and fractional Brownian fields (in 2D), that are jointly Gaussian self-similar (i.e., $c_2 \equiv 0$) processes with fully parametric covariance structure appropriate for ML and Bayesian estimation. Examples of ML for 1D fBm formulated in the spectral or wavelet domains can be found in [Ber94, CP06, MRS08]. For images, an ML estimator has been proposed in [LOKS86] (note, however, that the estimation problem is reduced to a univariate formulation for the rows/columns of the image there). As far as MMC processes are concerned, [LR12] proposed an ML approach in the time domain for one specific process. However, the method relies strongly on the particular construction of this process and cannot easily accommodate more general model classes. Moreover, the method is formulated for 1D signals only.

Bayesian inference. Similarly, the complex statistical properties of multifractal processes precludes the development of Bayesian estimation procedures for MFA. A Bayesian estimation method has been introduced in [WO92] but is relevant only for the analysis of 1D self-similar signals.

Recently, a Bayesian estimation procedure for the parameter c_2 of multifractal time series has been proposed in [WDTA13]. Unlike the ML and Bayesian methods above, it does not rely on specific assumptions on the process or its wavelet coefficients but instead employs a generic parametric model for the statistics of the logarithm of wavelet leaders, empirically validated for MMC processes. This method has been reported to yield significant improvement in the estimation of c_2 . However, it is designed for and can only be applied to univariate time series of small sample size.

1.7 Conclusion

In this chapter, we have introduced the main theoretical and practical concepts of MFA. The primary goal of MFA is the estimation of the multifractal spectrum $D(h)$ (1.3), which is achieved in practice via a 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 encoded in $D(h)$ can be well summarized by a limited collection of parameters, namely log-cumulants c_p . The present thesis will focus on the estimation of the first two log-cumulants c_1 and c_2 , respectively quantifying the mode (the average regularity) and the width (the degree of fluctuations of the local regularity in space) of the multifractal spectrum.

1.8 Conclusion (in French)

Dans ce chapitre nous avons introduit les principaux concepts théoriques et pratiques de l'analyse multifractale. L'objectif central de l'analyse multifractale est l'estimation du spectre multifractal $D(h)$ défini dans (1.3), qui est effectuée en pratique via un formalisme multifractal. Les coefficients dominants sont les coefficients multirésolutions à la base du formalisme multifractal de référence. Les propriétés multifractales d'une image peuvent être très bien décrites via une collection limitée de paramètres, les log-cumulants c_p . Cette thèse se concentrera sur l'estimation des deux premiers log-cumulants c_1 et c_2 , respectivement liés au mode (la régularité moyenne) et la largeur (le degré des fluctuations spatiales de la régularité locale) du spectre multifractal.

Chapter 2

Statistical model and univariate Bayesian estimation

Contents

2.1	Introduction	32
2.2	Introduction (in French)	33
2.3	Statistical model for log-leaders	34
2.3.1	Second-order statistical model	35
2.3.2	Approximation of the likelihood	40
2.4	Bayesian models	47
2.4.1	Whittle approximation based Bayesian model	48
2.4.2	Augmented likelihood based Bayesian model	49
2.5	Estimation algorithms	53
2.5.1	Metropolis-within-Gibbs algorithm	54
2.5.2	Two-stage Gibbs sampler	56
2.6	Numerical experiments	58
2.6.1	General setup	58
2.6.2	Adjusting the frequency range in the Whittle approximation	59
2.6.3	Numerical assessment of equivalence between Bayesian models	60
2.6.4	Estimation performance for c_1	63
2.6.5	Estimation performance for c_2	63
2.6.6	Convergence and computational cost	69
2.7	Conclusions and perspectives	71
2.8	Conclusions et perspectives (in French)	72

2.1 Introduction

We introduce in this chapter a Bayesian approach for the estimation of the two leading log-cumulants c_1 and c_2 of univariate images. The strategy adopted in this chapter follows intuitions developed in [WDTA13], which proposed a Bayesian procedure for the estimation of the multifractal parameters of univariate time series. This approach, relying on a parametric model for the statistics of the log-leaders, was however fundamentally designed only for time series. This chapter alleviates this limitation and contains several contributions that enable the Bayesian estimation of the two first log-cumulants for images.

First, by means of numerical simulations for a large selection of 2D multifractal reference processes, we propose in Section 2.3.1 an empirical second-order parametric statistical model for the logarithm of wavelet leaders (referred to as log-leaders for short) of 2D multifractal multiplicative cascades based (MMC) processes. This generic model consists of assuming that, at each scale, the distributions of the log-leaders can be well approximated by a multivariate Gaussian distribution whose mean and variance-covariance structure are respectively controlled by c_1 and c_2 . The construction of such a model is the starting point for embedding estimation in a Bayesian framework. The likelihood induced by this model is however problematic due to the computation of the inverse of large covariance matrices, which is numerically unstable and computationally too demanding, even for small images. To overcome this issue, we resort in Section 2.3.2 to a Whittle approximation to efficiently evaluate the scale-wise Gaussian likelihood in the spectral domain. With the specification of prior distributions for the unknown parameters, we then build in Section 2.4 a Bayesian model addressing the estimation of c_1 and c_2 , wherein the exact likelihood is replaced by a Whittle approximation. However, the particular form of this approximation prevents the design of conjugate priors. In Section 2.4.2, we resort to a data augmentation scheme to express the Whittle approximation as the marginal of an augmented likelihood whose convenient form enables the design of conjugate priors and hence yields simplified parameter inference algorithms. Finally, the Bayesian estimators associated with posterior distributions are approximated by means of samples generated via suitable Markov chain Monte Carlo algorithms described in Section 2.5. In Section 2.6, the proposed

estimation procedure is validated and assessed on a large collection of synthetic multifractal images and shown to provide excellent estimation performance both for small and large image sizes. In particular, the standard deviation (STD) values for the estimator of c_2 is reduced by a factor close to 4 when compared to the classical linear regression-based estimators.

Developments and results presented in this chapter have been reported in [CWD⁺15, CWT⁺15a, CWT⁺15b, CWA⁺16b].

2.2 Introduction (in French)

Nous introduisons dans ce chapitre une approche bayésienne pour l'estimation des deux premiers log-cumulants c_1 et c_2 d'images univariées. La stratégie adoptée dans ce chapitre reprend des intuitions développées dans [WDTA13], où une procédure bayésienne pour l'estimation des paramètres multifractaux de séries temporelles univariées a été proposée. Cette approche, reposant sur un modèle paramétrique des statistiques du logarithme des coefficients dominants, était néanmoins fondamentalement conçue uniquement pour des séries temporelles. Ce chapitre pallie cette limitation et contient plusieurs contributions permettant une estimation bayésienne des deux premiers log-cumulants des images.

Premièrement, grâce à des simulations numériques conduites sur une large sélection de processus multifractaux de référence, nous proposons dans la Section 2.3.1 un modèle empirique et paramétrique de deuxième ordre pour le logarithme des coefficients dominants associés à des processus multifractaux bidimensionnels construits sur des cascades multiplicatives. Ce modèle générique consiste à supposer que, à chaque échelle, les distributions du logarithme des coefficients dominants peuvent être bien approximées par des lois gaussiennes multivariées, dont la moyenne et la structure de variance-covariance sont respectivement contrôlées par c_1 et c_2 . La construction d'un tel modèle est la pierre angulaire pour formuler l'estimation dans un cadre bayésien. La vraisemblance associée à ce modèle est cependant problématique à cause du calcul de l'inverse de matrices de covariance de grandes dimensions, ce qui est numériquement instable et requiert une trop grande puissance de calcul (et ce, même pour des petites images). Afin de remédier à ce problème, nous faisons appel

dans la Section 2.3.2 à une approximation de Whittle pour évaluer efficacement la vraisemblance gaussienne à chaque échelle dans le domaine spectral. Ensuite, via la spécification de lois *a priori* sur les paramètres inconnus, nous construisons dans la Section 2.4 un modèle bayésien pour l'estimation de c_1 et c_2 , où la vraisemblance exacte est remplacée par une approximation de Whittle. Cependant, la forme particulière de cette approximation empêche l'utilisation de lois *a priori* conjuguées. Dans la Section 2.4.2, nous utilisons une technique d'augmentation des données pour exprimer l'approximation de Whittle comme la marginale d'une vraisemblance augmentée, dont la forme autorise cette fois l'utilisation de lois *a priori* conjuguées. Cette technique permet d'aboutir à des algorithmes d'estimation efficaces. Enfin, les estimateurs bayésiens associés aux lois *a posteriori* sont approximatés grâce à des méthodes de Monte Carlo par chaînes de Markov, qui sont détaillées dans la Section 2.5. Dans la Section 2.6, cette procédure d'estimation est évaluée et validée pour une large collection de processus multifractaux de référence. Il est notamment montré que notre procédure offre d'excellentes performances d'estimation pour les petites et grandes images, avec des écarts types pour l'estimateur de c_2 divisés par un facteur proche de 4 par rapport à l'estimateur classique construit sur une régression linéaire.

Les développements et les résultats présentés dans ce chapitre ont fait l'objet de plusieurs publications [CWD⁺15, CWT⁺15a, CWT⁺15b, CWA⁺16b].

2.3 Statistical model for log-leaders

In this section, we introduce a novel empirical second-order statistical model for the log-leaders of 2D multifractal multiplicative cascade (MMC) based processes (cf. Section 1.5). More precisely, at each scale, the collection of log-leaders is assumed to be a stationary Gaussian random field for which we design a mean and variance-covariance structure, respectively parametrized by c_1 and c_2 . The validity and relevance of this model is assessed numerically for a selection of MMC processes. We moreover propose to resort to a Whittle approximation to enable the otherwise infeasible evaluation of the likelihood associated with the model.

2.3.1 Second-order statistical model

a) Marginal distribution

We first numerically investigate the marginal distribution of log-leaders

$$l(j, \mathbf{k}) = \log L(j, \mathbf{k}) \quad (2.1)$$

associated with 2D MMC processes. To this end, the selection of MMC processes introduced in Chapter 1, namely CMC-LN, CMC-LP, CPC-LN, CPC-LP and MRW, have been analyzed (see Section 1.5 for details on the parameters used), as well as fBm, a member of self-similar processes by means of quantile-quantile (QQ)-plots. In short, QQ-plots enable a visual assessment of the fit between two distributions by plotting their quantiles against each other. Fig. 2.1 displays QQ-plots of the quantiles of the empirical distributions of log-leaders associated with the different processes ($c_2 = -0.04$ for MMC processes and $H = 0.7$ for fBm) of size $2^9 \times 2^9$ against the quantiles of the standard normal distribution. It also reports the QQ-plots of the processes themselves and of the logarithm of the absolute value of wavelet coefficients.

These QQ-plots indicate that the normal distribution provides, within ± 3 standard deviations, a reasonable approximation for the marginal distribution of log-leaders of images for all members of the MMC class. It is also the case for the fBm, a Gaussian self-similar process that is not a member of MMC. Note that the fact that the marginal distributions of the log-leaders are approximately Gaussian for scale invariant processes confirms the intuitions formulated by Mandelbrot in [Man90]. However, it is not a trivial finding: There is no a priori reason for this property to hold even if the analyzed stochastic process has log-normal multipliers (as is the case for CMC-LN, for instance). Indeed, it is not the case for the logarithm of the absolute value of wavelet coefficients whose marginal distributions are found to be significantly more complicated and strongly depart from Gaussian, cf. Fig. 2.1 (center column). Finally, complementary results for a wide range of process parameters (c_2 and H) and sample sizes N were found to be consistent with those reported here and yield similar conclusions. Such results can be found in Fig. A.1 (see Appendix A).

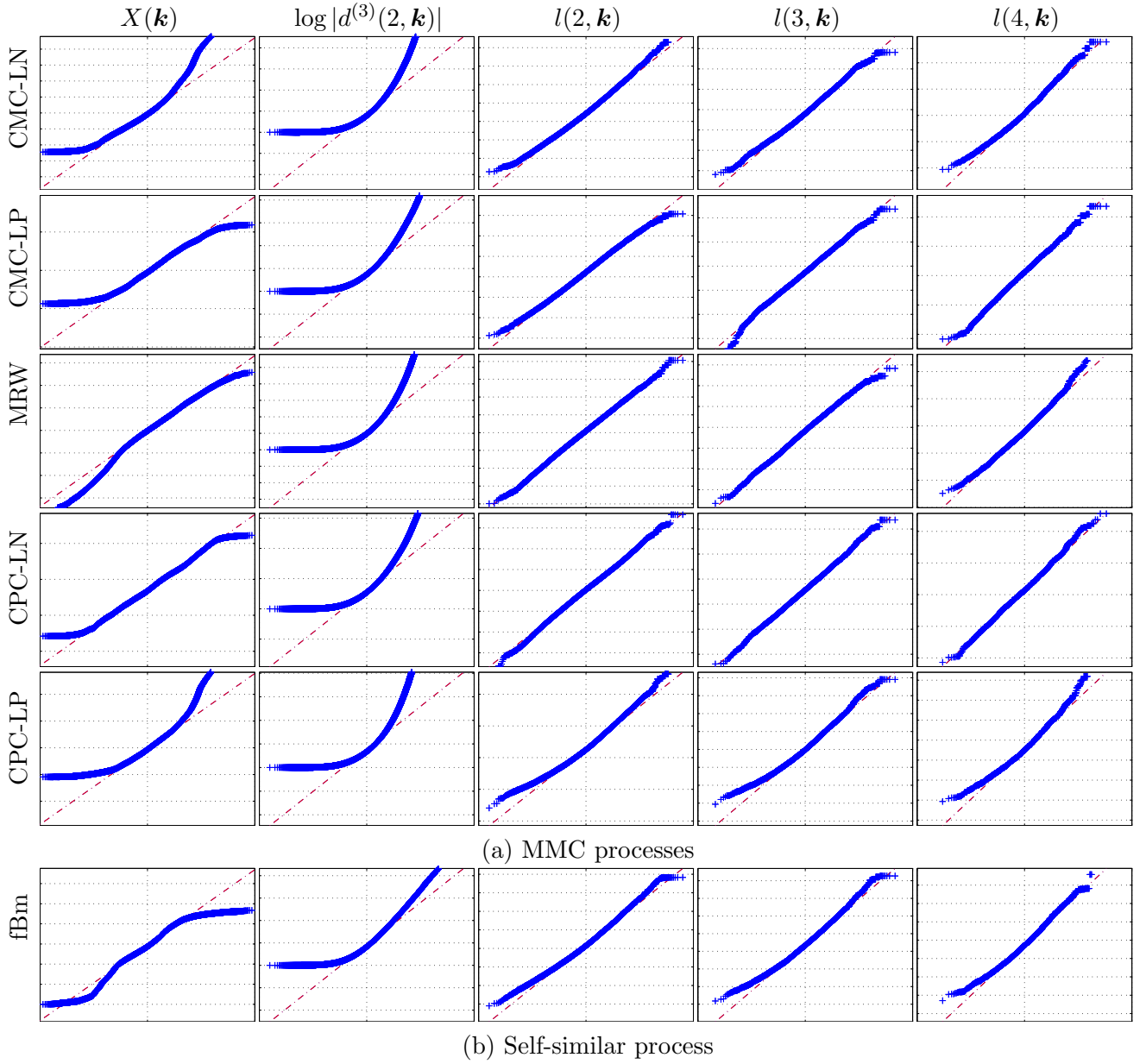


Figure 2.1: Quantile-quantile plots of the empirical distributions (+) of the process $X(\mathbf{k})$, the log-wavelet coefficients $\log |d^{(3)}(2, \mathbf{k})|$ and the log-leaders $l(j, \mathbf{k})$, $j = 2, 3, 4$ against standard normal distribution (---) for a selection of MMC processes ($c_2 = -0.04$) (a) and self-similar processes ($H = 0.7$) (b) of size $2^9 \times 2^9$. The closer to the red dashed line, the better the fit to the normal distribution.

b) Parametric statistical model

We furthermore complete the statistical model by assuming that, at each scale j , the collection of log-leaders $\{l(j, \cdot)\}$ is a stationary Gaussian random field for which we specify the mean and the covariance structure.

Mean. For the first-order statistics, we only use (1.15), which yields for the mean a linear evolution across scales with a slope controlled by c_1

$$\mathbb{E}[l(j, \mathbf{k})] = \mu_{j, \mathbf{w}} = c_1^0 + j c_1 \log 2 \quad (2.2a)$$

$$\mathbf{w} = [w_1, w_2]^T = [c_1, c_1^0]^T. \quad (2.2b)$$

Variance-covariance structure. For the second-order statistics, we build our model using the power law scaling across scales (1.15) for the variance but include in addition a parametric model for the covariance structure. It is motivated by the asymptotic covariance of the logarithm of multiscale quantities generically associated with multiplicative construction (cf. [Man74]), studied in detail for wavelet coefficients of 1D random wavelet cascades in [ABM98], and also by recent numerical results obtained for the covariance of the logarithm of 1D wavelet leaders for MMC processes [WDTA13]. These results suggest a linear decay of the covariance of log-leaders for MMC processes at fixed scale j , denoted as $\text{Cov}[l(j, k), l(j, k + \Delta k)]$, in log coordinates $\log \Delta k$, with slope given by the parameter c_2 . Numerical simulations with 2D MMC processes for a wide range of process parameters (detailed in Section 1.5) indicate that the intra-scale covariance, denoted as $\text{Cov}[l(j, \mathbf{k}), l(j, \mathbf{k} + \Delta \mathbf{k})] = \mathcal{C}_j(\Delta \mathbf{k})$, can be well approximated by a radially symmetric function parametrized by the parameter vector

$$\boldsymbol{\theta} = [\theta_1, \theta_2]^T = [c_2, c_2^0]^T \quad (2.3)$$

and defined as follows

$$\mathcal{C}_j(\Delta \mathbf{k}) = \varrho_{j, \boldsymbol{\theta}}(r) = \begin{cases} \varrho_{j, \boldsymbol{\theta}}^S(r) & 0 \leq r < 3 \\ \varrho_{j, \boldsymbol{\theta}}^L(r) & 3 \leq r \end{cases} \quad (2.4)$$

where $\varrho_{j,\boldsymbol{\theta}}^S$ and $\varrho_{j,\boldsymbol{\theta}}^L$ respectively model the short and long-term covariance and $r = \|\Delta\mathbf{k}\|$. First, the long-term covariance $\varrho_{j,\boldsymbol{\theta}}^L$ consists of a decaying function as $c_2 \log r$ (since $c_2 < 0$) given by

$$\varrho_{j,\boldsymbol{\theta}}^L(r) = \max(0, \Delta\varrho + c_2(\log r + \log 2^j)), \quad \text{for } r \geq 3 \quad (2.5)$$

where the restriction to positive values is here considered for numerical reasons (conditioning of the covariance matrix). The parameter $\Delta\varrho$ is a constant found to be well approximated by using the heuristic condition

$$\varrho_{j,\boldsymbol{\theta}}^L(r_j) = 0 \quad \text{with} \quad r_j = \lfloor m_j/4 \rfloor \quad (2.6a)$$

$$m_j = \sqrt{n_j} \quad (2.6b)$$

where the operator $\lfloor \cdot \rfloor$ truncates to integer values and n_j is defined in Chapter 1 as the total number of wavelet leaders at scale j . The short-term covariance model is then defined as a line connecting the variance $\text{Var}[l(j, \mathbf{k})] = c_2^0 + c_2 \log 2^j$ prescribed by (1.15) at $r = 0$ and $\varrho_{j,\boldsymbol{\theta}}^L(r)$ at $r = 3$ as follows

$$\varrho_{j,\boldsymbol{\theta}}^S(r) = \frac{\log(r+1)}{\log 4} (\varrho_{j,\boldsymbol{\theta}}^L(3) - (c_2^0 + c_2 \log 2^j)) + c_2^0 + c_2 \log 2^j. \quad (2.7)$$

The sample covariance $\hat{\mathcal{C}}_j(\Delta\mathbf{k})$, computed over 100 realizations of CMC-LN processes with $N = 2^9$ and $[c_1, c_2] = [0.02, -0.04]$, is illustrated and compared to the proposed covariance model $\varrho_{j,\boldsymbol{\theta}}$ in Fig. 2.2. The radial symmetry obviously appears as a good approximation for the covariance structure. Moreover, Fig. 2.2 (c), (d) and (e) indicate that the covariance model $\varrho_{j,\boldsymbol{\theta}}$ provides overall good fits. Similar results have been obtained for a wide range of c_2 values and for other MMC processes. Additional results for MRW and CPC-LN are reported in Appendix A.

c) Likelihood

Let $\boldsymbol{\ell}_j$ denote the vector gathering the collection of log-leaders $\{l(j, \cdot)\}$ at scale j , organized in lexicographic order. With the above notation and statistical model, the likelihood of $\boldsymbol{\ell}_j$ with respect to

$$\boldsymbol{\lambda} = \{\boldsymbol{\theta}, \mathbf{w}\} \quad (2.8)$$

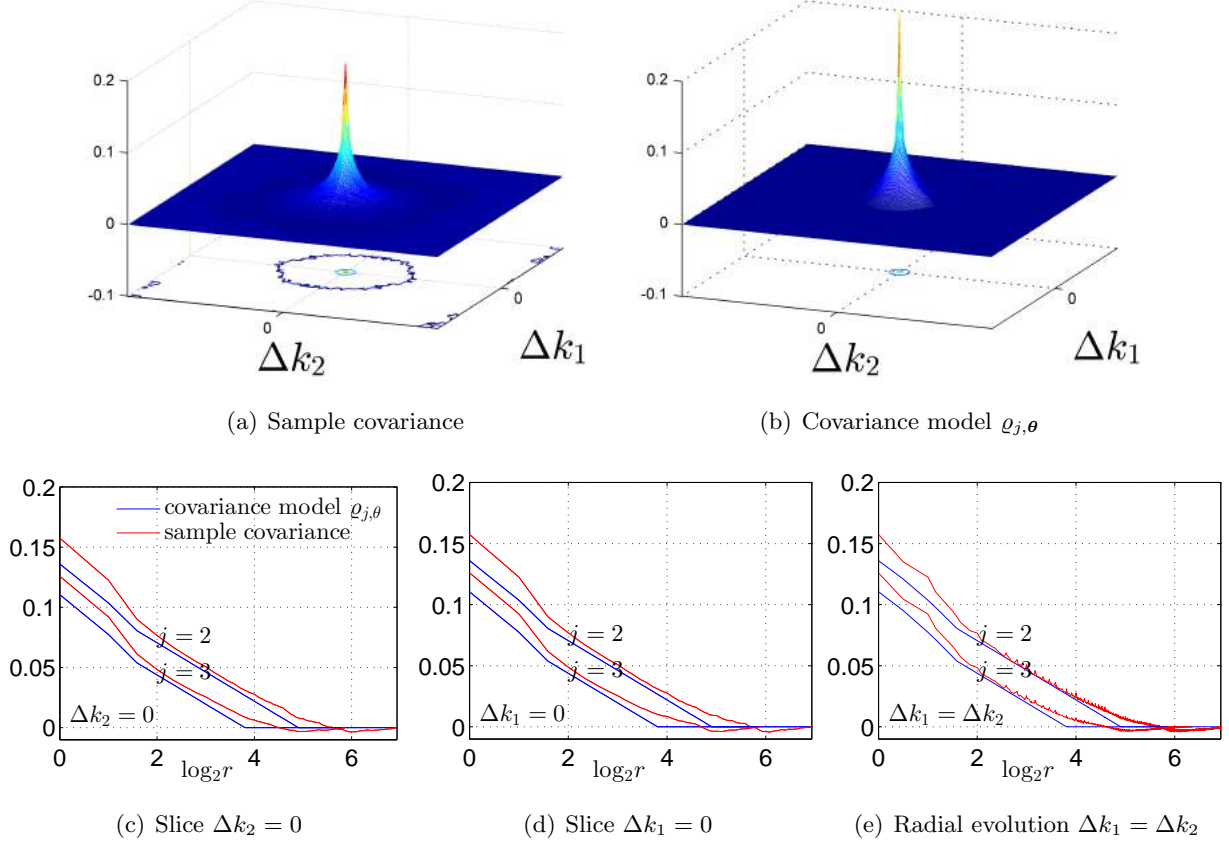


Figure 2.2: Fitting between the sample covariance (a), averaged over 100 realizations of 2D CMC-LN ($N = 2^9$ and $[c_1, c_2] = [0.02, -0.04]$), and the parametric covariance (b) at scale $j = 2$; (c), (d) and (e) compare the model (blue) and the sample covariance (red) for different directions at two scales $j = 2, 3$.

is given by

$$p(\ell_j | \boldsymbol{\lambda}) \propto (\det \boldsymbol{\Sigma}_{j,\theta})^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\ell_j - \boldsymbol{\mu}_{j,\mathbf{w}})^T \boldsymbol{\Sigma}_{j,\theta}^{-1} (\ell_j - \boldsymbol{\mu}_{j,\mathbf{w}}) \right) \quad (2.9)$$

where $\boldsymbol{\Sigma}_{j,\theta}$ corresponds to the $n_j \times n_j$ covariance matrix whose entries are induced by the 2D parametric covariance function model (2.4), i.e., $[\boldsymbol{\Sigma}_{j,\theta}]_{u,v} = \varrho_{j,\theta}(\|\mathbf{k}_u - \mathbf{k}_v\|)$ and $\boldsymbol{\mu}_{j,\mathbf{w}} = \mu_{j,\mathbf{w}} \mathbf{1}_{n_j}$. In this thesis, independence between log-leaders at different scales j is assumed¹, which leads to the

¹Taking into account the inter-scale dependence could however yield a more realistic model. This point is further discussed in the perspectives (cf. Chapter 5).

following likelihood for the vector $\ell = [\ell_{j_1}^T, \dots, \ell_{j_2}^T]^T$ of log-leaders at scale $j = j_1, \dots, j_2$

$$\begin{aligned} p(\ell|\lambda) &= \prod_{j=j_1}^{j_2} p(\ell_j|\lambda) \\ &\propto \prod_{j=j_1}^{j_2} (\det \Sigma_{j,\theta})^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\ell_j - \mu_{j,\mathbf{w}})^T \Sigma_{j,\theta}^{-1} (\ell_j - \mu_{j,\mathbf{w}}) \right). \end{aligned} \quad (2.10)$$

To ensure that (2.10) is a valid likelihood, the covariance matrices $\Sigma_{j,\theta}$ must be positive definite for $j = j_1, \dots, j_2$. This condition implicitly enforces constraints on the parameter vector θ that can be assessed only numerically. We denote by

$$\mathcal{A} = \{\theta \in \mathbb{R}_*^- \times \mathbb{R}_*^+ | \forall j \in \llbracket j_1, j_2 \rrbracket, \Sigma_{j,\theta} \text{ is positive definite}\} \quad (2.11)$$

the admissible set for θ associated with the likelihood (2.10), where $\llbracket n_1, n_2 \rrbracket$ denotes the set of integers ranging from n_1 to n_2 .

2.3.2 Approximation of the likelihood

The numerical evaluation of the scale-wise Gaussian likelihood (2.9) (and hence of the joint likelihood (2.10)) is problematic since it requires the inversion of a large covariance matrix $\Sigma_{j,\theta}$. This inversion is computationally prohibitive for images, even for very modest sizes (for instance, a 64×64 image would require the inversion of a dense matrix of size $\sim 1000 \times 1000$ for the scale $j = 1$). In addition, it is numerically unstable for larger images (due to growing condition number). In order to bypass this difficulty, we propose in this section to evaluate the scale-wise Gaussian likelihood (2.9) in the spectral domain by using an asymptotic approximation due to Whittle, see [Whi53, Whi54].

a) Whittle approximation

Whittle approximation. With the assumptions of Section 2.3.1, the collection of log-leaders $\{l(j, \cdot)\}$ is a realization of a non-zero mean stationary Gaussian random field on a regular lattice. Up to a multiplicative constant, the Whittle approximation for the scale-wise Gaussian likelihood (2.9)

was studied in [Whi54, Fue07, Ber94, Whi53, AL95, Cha97], and leads to the following result

$$p_W(\ell_j|\boldsymbol{\lambda}) = \prod_{\mathbf{m} \in J_j} (\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_{\mathbf{m}}))^{-1} \exp\left(-\frac{y_{\ell_j}^*(\boldsymbol{\omega}_{\mathbf{m}})y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}})}{\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_{\mathbf{m}})}\right) \\ \times (\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_0))^{-\frac{1}{2}} \exp\left(\frac{(y_{\ell_j}(\boldsymbol{\omega}_0) - m_j\mu_{j,\mathbf{w}})^2}{2\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_0)}\right) \quad (2.12)$$

where $y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}})$ is the 2D discrete Fourier transform (DFT) coefficient of $\{l(j, \cdot)\}$ defined with the following normalization

$$y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}}) = \frac{1}{m_j} \sum_{\mathbf{k} \in \llbracket 0, m_j \rrbracket^2} l(j, \mathbf{k}) \exp(-i\mathbf{k}^T \boldsymbol{\omega}_{\mathbf{m}}) \quad (2.13)$$

with $\boldsymbol{\omega}_{\mathbf{m}} = 2\pi\mathbf{m}/m_j$ such that

$$I_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}}) = y_{\ell_j}^*(\boldsymbol{\omega}_{\mathbf{m}})y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}}) \quad (2.14)$$

corresponds to the periodogram of $\{\ell(j, \mathbf{k})\}$. The operator $(\cdot)^*$ stands for complex conjugation and $\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the spectral density associated with the covariance function $\varrho_{j,\boldsymbol{\theta}}(r)$ (its computation is discussed in Section 2.3.2 b) below). The product is taken over half of the total frequency grid

$$J_j = \llbracket \lfloor \frac{-m_j - 1}{2} \rfloor, -1 \rrbracket \times \llbracket 1, \lfloor \frac{m_j}{2} \rfloor \rrbracket \cup \llbracket 0, \lfloor \frac{m_j}{2} \rfloor \rrbracket^2 \setminus \{(0, 0)\} \quad (2.15)$$

where the zero frequency is removed since it is included separately as the second term in (2.12). An illustration of the set J_j is given in Fig. 2.3 for $m_j = 8$.

Note that in (2.12), the contribution of the mean of the Gaussian field reduces to the zero frequency term [Cha97, Jes12]. If only $\boldsymbol{\theta}$ is of interest, it is therefore possible to make the Whittle approximation mean invariant by discarding the zero frequency term as it is commonly reported in [Whi54, Fue07, Ber94, Whi53, AL95].

Finally, in the rest of this thesis, we denote by $p_W(\ell|\boldsymbol{\lambda})$ the approximation of the joint likelihood $p(\ell|\boldsymbol{\theta})$ obtained by replacing $p(\ell_j|\boldsymbol{\lambda})$ with $p_W(\ell_j|\boldsymbol{\lambda})$ in (2.10)

$$p_W(\ell|\boldsymbol{\lambda}) \propto \prod_{j=j_1}^{j_2} p_W(\ell_j|\boldsymbol{\lambda}) \quad (2.16)$$

which is associated with the following admissible set for θ

$$\mathcal{A}_W = \{\theta \in \mathbb{R}_\star^- \times \mathbb{R}_\star^+ | \forall j \in \llbracket j_1, j_2 \rrbracket, \forall \mathbf{m} \in J_j, \phi_{j,\theta}(\omega_{\mathbf{m}}) > 0\} \quad (2.17)$$

to ensure valid parametric spectral densities.

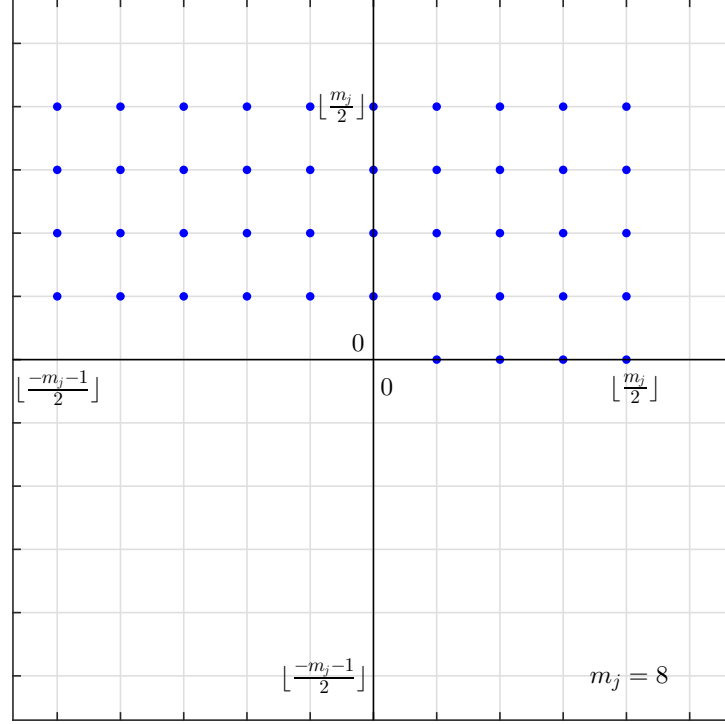


Figure 2.3: Frequency grid and set J_j (•) for $m_j = 8$.

b) Computation of the parametric spectral density

The virtue of the Whittle approximation is to replace the inversion of the covariance matrix $\Sigma_{j,\theta}$ with the computation of the parametric spectral density induced by the covariance model. Here we propose two different ways to compute $\phi_{j,\theta}(\omega)$, both of which make use of the rewriting of $\varrho_{j,\theta}$ as

$$\varrho_{j,\theta}(r) = c_2 \varrho_{1,j}(r) + c_2^0 \varrho_{2,j}(r) \quad (2.18)$$

where

$$\varrho_{1,j}(r) = \left(\frac{\log(\frac{3}{r_j 2^j})}{\log 4} \log(1+r) + j \log 2 \right) \times \mathbb{1}_{[0,3]}(r) + \log\left(\frac{r}{r_j}\right) \times \mathbb{1}_{[3,r_j]}(r) \quad (2.19)$$

$$\varrho_{2,j}(r) = \left(1 - \frac{\log(1+r)}{\log 4}\right) \times \mathbb{1}_{[0,3]}(r) \quad (2.20)$$

with r_j defined in (2.6). This yields a decomposition of $\phi_{j,\theta}(\omega_{\mathbf{m}})$ into

$$\phi_{j,\theta}(\omega_{\mathbf{m}}) = c_2 f_{1,j}(\omega_{\mathbf{m}}) + c_2^0 f_{2,j}(\omega_{\mathbf{m}}) \quad (2.21)$$

where the functions $f_{i,j}$, $i = 1, 2$ do not depend on the parameters c_2 and c_2^0 and can be pre-calculated.

Analytical expression. To begin with, the continuous spectral density associated with the covariance model $\varrho_{j,\theta}(r)$ is given by Bochner's theorem [GDGF10]

$$\bar{\phi}_{j,\theta}(\omega) = \int_{\mathbb{R}^2} \varrho_{j,\theta}(\|\mathbf{u}\|) \exp(-i\mathbf{u}^T \omega) d\mathbf{u}. \quad (2.22)$$

Second, since $\varrho_{j,\theta}(\|\mathbf{u}\|)$ is a radial symmetric function, its Fourier transform $\bar{\phi}_{j,\theta}(\omega)$ is also radial symmetric. It can therefore be expressed as a Hankel transform [Pou99], given by

$$\bar{\phi}_{j,\theta}(\omega) = \bar{\phi}_{j,\theta}(\|\omega\|) = 2\pi \int_0^\infty r \varrho_{j,\theta}(r) J_0(r\|\omega\|) dr \quad (2.23)$$

where $J_n(\cdot)$ is the n -th order Bessel function. By injecting (2.18) in (2.23), we obtain

$$\bar{\phi}_{j,\theta}(\|\omega\|) = c_2 \bar{f}_{1,j}(\|\omega\|) + c_2^0 \bar{f}_{2,j}(\|\omega\|) \quad (2.24)$$

where

$$\bar{f}_{i,j}(\|\omega\|) = 2\pi \int_0^\infty r \varrho_{i,j}(r) J_0(r\|\omega\|) dr. \quad (2.25)$$

To evaluate the integral in (2.25), we make use of the identities

$$(i) \int_0^R r J_0(r\rho) dr = \frac{RJ_1(R\rho)}{\rho}$$

$$(ii) \int_0^R r \log\left(\frac{r}{R}\right) J_0(r\rho) dr = -\frac{(1-J_0(R\rho))}{\rho^2}$$

that are valid for $R > 0$, $\rho > 0$, cf. [Pou99, Tab. 17.1]. Straightforward calculations lead to

$$\bar{f}_{1,j}(\|\omega\|) = 2\pi \left(\frac{J_0(r_j\|\omega\|) - J_0(3\|\omega\|)}{\|\omega\|^2} + \frac{\log(\frac{3}{r_j 2^j})}{\log 4} \left(\frac{3J_1(3\|\omega\|)}{\|\omega\|} + \mathcal{I}(\|\omega\|) \right) \right) \quad (2.26)$$

$$\bar{f}_{2,j}(\|\omega\|) = 2\pi \left(\frac{3J_1(3\|\omega\|)}{\|\omega\|} - \frac{\mathcal{I}(\|\omega\|)}{\log 4} \right) \quad (2.27)$$

with $\mathcal{I}(\|\boldsymbol{\omega}\|) = \int_0^3 r \log(1+r) J_0(r\|\boldsymbol{\omega}\|) dr$. With these expressions, we obtain an analytic expression for (2.21) by discretizing (2.24) using spectral aliasing as in [GF14]

$$\phi_{j,\boldsymbol{\theta}}^H(\boldsymbol{\omega}_{\mathbf{m}}) = \sum_{\mathbf{p} \in \mathbb{Z}^2} c_2 \bar{f}_{1,j}(\boldsymbol{\omega}_{\mathbf{m},\mathbf{p}}) + c_2^0 \bar{f}_{2,j}(\boldsymbol{\omega}_{\mathbf{m},\mathbf{p}}) \quad (2.28)$$

where $\boldsymbol{\omega}_{\mathbf{m},\mathbf{p}} = \boldsymbol{\omega}_{\mathbf{m}} + 2\pi\mathbf{p}$. At last, the infinite summation in (2.28) is truncated to $\llbracket -K, K \rrbracket^2$, yielding the approximation

$$\phi_{j,\boldsymbol{\theta}}^H(\boldsymbol{\omega}_{\mathbf{m}}) \approx c_2 f_{1,j}^H(\|\boldsymbol{\omega}_{\mathbf{m}}\|) + c_2^0 f_{2,j}^H(\|\boldsymbol{\omega}_{\mathbf{m}}\|) \quad (2.29)$$

with

$$f_{i,j}^H(\|\boldsymbol{\omega}_{\mathbf{m}}\|) = \sum_{\mathbf{p} \in \llbracket -K, K \rrbracket^2} \bar{f}_{i,j}(\|\boldsymbol{\omega}_{\mathbf{m},\mathbf{p}}\|), \quad i = 1, 2. \quad (2.30)$$

The two partial sums can be pre-calculated and stored for the discrete set of frequencies $\boldsymbol{\omega}_{\mathbf{m}}$, using a quadrature rule, such as the trapezoid or Simpson's rules, for the computation of the integral \mathcal{I} .

Discrete Fourier transform. It is also possible to obtain a numerical approximation of the spectral density by using a discrete Fourier transform

$$\phi_{j,\boldsymbol{\theta}}^D(\boldsymbol{\omega}_{\mathbf{m}}) = \sum_{\mathbf{k} \in \llbracket -m_j, m_j \rrbracket^2} \varrho_{j,\boldsymbol{\theta}}(\|\mathbf{k}\|) \exp(-i\mathbf{k}^T \boldsymbol{\omega}_{\mathbf{m}}) \quad (2.31)$$

which straightforwardly leads to

$$\phi_{j,\boldsymbol{\theta}}^D(\boldsymbol{\omega}_{\mathbf{m}}) = c_2 f_{1,j}^{DFT}(\boldsymbol{\omega}_{\mathbf{m}}) + c_2^0 f_{2,j}^{DFT}(\boldsymbol{\omega}_{\mathbf{m}}) \quad (2.32)$$

where

$$f_{i,j}^{DFT}(\boldsymbol{\omega}_{\mathbf{m}}) = \sum_{\mathbf{k} \in \llbracket -m_j, m_j \rrbracket^2} \varrho_{i,j}(\mathbf{k}) \exp(-i\mathbf{k}^T \boldsymbol{\omega}_{\mathbf{m}}), \quad i = 1, 2. \quad (2.33)$$

In Fig. 2.4, we illustrate the parametric spectral density $\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_{\mathbf{m}})$ computed by using (2.28) (with $K = 3$) and (2.31). Both methods are found to yield results numerically very close, suggesting that none is preferable to the other. As a result, in the rest of this thesis, we generically denote the spectral density as $\phi_{j,\boldsymbol{\theta}}(\boldsymbol{\omega}_{\mathbf{m}})$ without specifying which method is used for its computation.

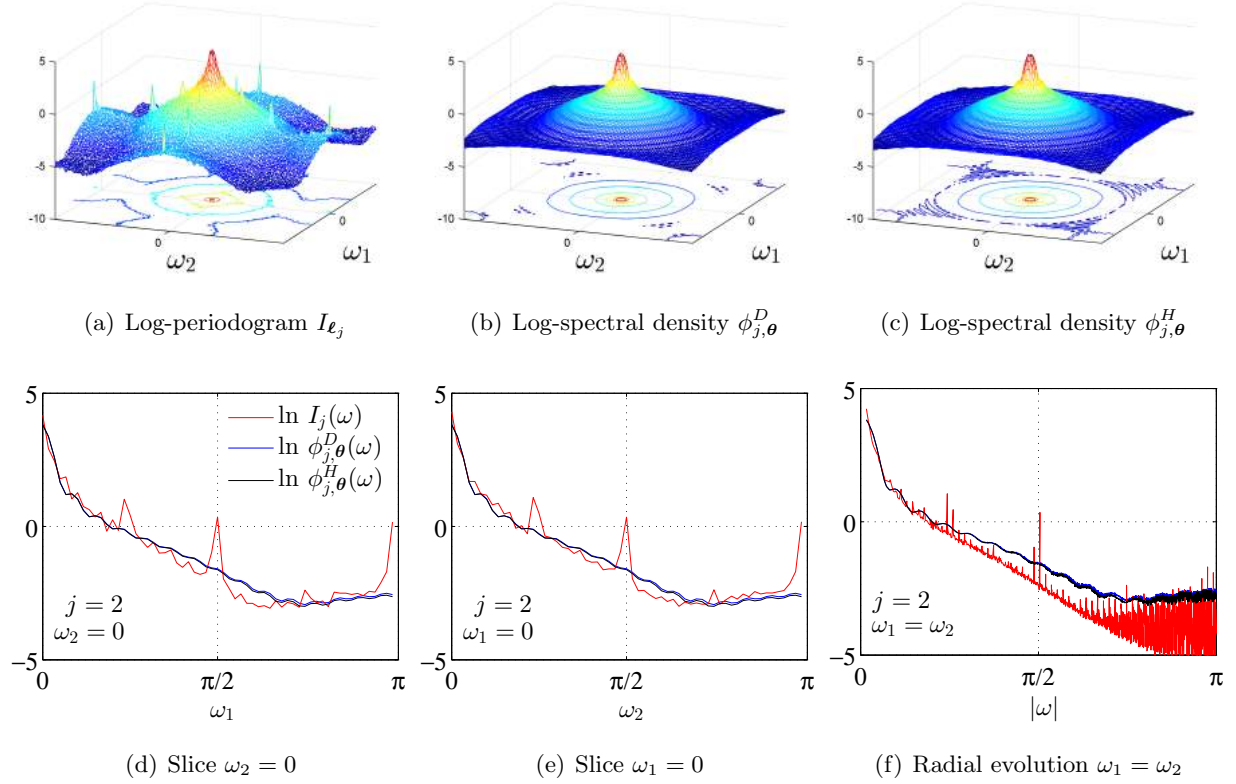


Figure 2.4: Fitting between the periodogram (a), averaged over 100 realizations of 2D CMC-LN ($N = 2^9$ and $[c_1, c_2] = [0.02, -0.04]$), and the spectral density models $\phi_{j,\theta}^D$ (b), obtained from a DFT of $\varrho_{j,\theta}(r)$, and $\phi_{j,\theta}^H$, obtained analytically, at scale $j = 2$; (d), (e) and (f) compare the models (blue and black) and the periodogram (red) for different directions.

c) Reparametrization

In the light of the decomposition (2.21) of the parametric spectral density $\phi_{j,\theta}(\omega_{\mathbf{m}})$, the admissible set (2.17) associated with the Whittle approximation $p_W(\ell|\lambda)$ can be reexpressed as

$$\mathcal{A}_W = \{\theta \in \mathbb{R}_*^- \times \mathbb{R}_*^+ | \forall j \in \llbracket j_1, j_2 \rrbracket, \forall \mathbf{m} \in J_j, c_2 f_{1,j}(\omega_{\mathbf{m}}) + c_2^0 f_{2,j}(\omega_{\mathbf{m}}) > 0\}. \quad (2.34)$$

We note that while $f_{1,j}$ can be either positive or negative, $f_{2,j}$ is always positive for any reasonable image size, as illustrated in Fig. 2.5, which displays functions $f_{i,j}$, $i = 1, 2$, for different image sizes. This property can be exploited to define a reparametrization that enables the efficient handling of the

constraints on covariance parameters $\boldsymbol{\theta}$ by transforming them into independent positivity constraints. More precisely, we consider the mapping defined by

$$\psi : \boldsymbol{\theta} \mapsto \mathbf{v} = [v_1, v_2]^T = (-c_2, c_2^0/\gamma + c_2), \quad \gamma = \sup_{j, \mathbf{m}} \frac{f_{1,j}(\boldsymbol{\omega}_{\mathbf{m}})}{f_{2,j}(\boldsymbol{\omega}_{\mathbf{m}})}. \quad (2.35)$$

It can be seen that, by construction, ψ is a one-to-one transformation from the admissible set \mathcal{A}_W to \mathbb{R}_*^{+2} and hence maps the joint constraints into independent positivity constraints. Finally, the Whittle approximation (2.16), expressed with the new set of parameters ²

$$\boldsymbol{\lambda} = \{\mathbf{v}, \mathbf{w}\} \quad (2.36)$$

directly reads

$$p_W(\boldsymbol{\ell}|\boldsymbol{\lambda}) = \prod_{j=j_1}^{j_2} p_W(\ell_j|\boldsymbol{\lambda}) \quad (2.37)$$

where

$$\begin{aligned} p_W(\ell_j|\boldsymbol{\lambda}) &= \prod_{\mathbf{m} \in J_j} (\phi_{j,v}(\boldsymbol{\omega}_{\mathbf{m}}))^{-1} \exp \left(-\frac{y_{\ell_j}^*(\boldsymbol{\omega}_{\mathbf{m}}) y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{m}})}{\phi_{j,v}(\boldsymbol{\omega}_{\mathbf{m}})} \right) \\ &\times (\phi_{j,v}(\boldsymbol{\omega}_{\mathbf{0}}))^{-\frac{1}{2}} \exp \left(-\frac{(y_{\ell_j}(\boldsymbol{\omega}_{\mathbf{0}}) - m_j \mu_{j,\mathbf{w}})^2}{2\phi_{j,v}(\boldsymbol{\omega}_{\mathbf{0}})} \right). \end{aligned} \quad (2.38)$$

In (2.38), the reparametrized spectral density is defined by

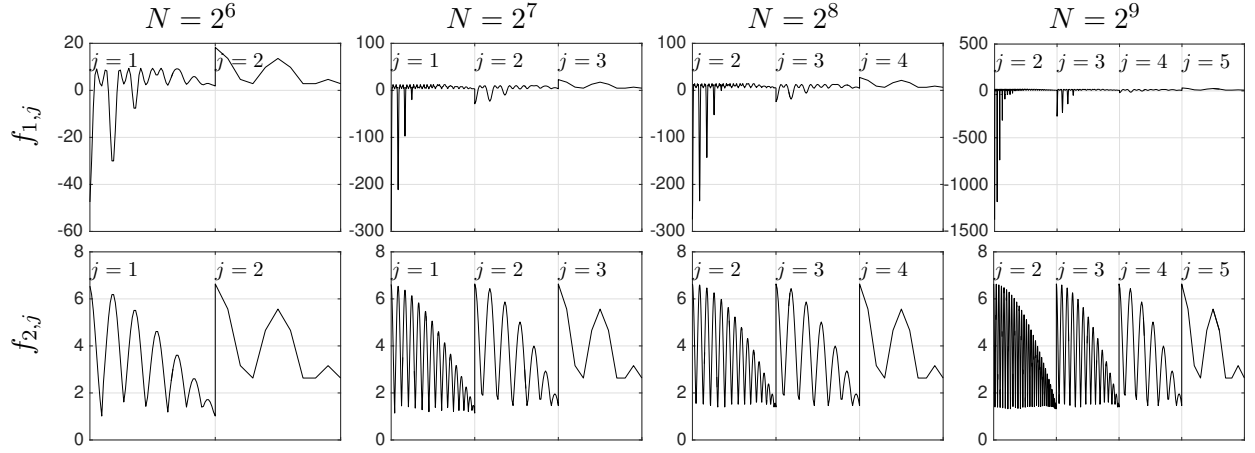
$$\phi_{j,v}(\boldsymbol{\omega}_{\mathbf{m}}) = v_1 \tilde{f}_{1,j}(\boldsymbol{\omega}_{\mathbf{m}}) + v_2 \tilde{f}_{2,j}(\boldsymbol{\omega}_{\mathbf{m}}) \quad (2.39)$$

with

$$\begin{cases} \tilde{f}_{1,j}(\boldsymbol{\omega}_{\mathbf{m}}) &= -f_{1,j}(\boldsymbol{\omega}_{\mathbf{m}}) + \gamma f_{2,j}(\boldsymbol{\omega}_{\mathbf{m}}) \\ \tilde{f}_{2,j}(\boldsymbol{\omega}_{\mathbf{m}}) &= \gamma f_{2,j}(\boldsymbol{\omega}_{\mathbf{m}}) \end{cases} \quad (2.40)$$

where, by construction, it can be shown that $\tilde{f}_{i,j}(\boldsymbol{\omega}_{\mathbf{m}}) > 0$ for $\forall \boldsymbol{\omega}_{\mathbf{m}}$.

²Here, by an abuse of notation, we keep the same symbol for the original and new set of parameters.

Figure 2.5: Functions $f_{1,j}$ and $f_{2,j}$ for different image sizes N .

2.4 Bayesian models

Using the novel statistical model developed in Section 2.3, we are now ready to specify a Bayesian model addressing the estimation of log-cumulants c_1 and c_2 via the specification of priors on \mathbf{w} and \mathbf{v} as in [WDTA13]. In contrast to the model for time series in [WDTA13], the exact likelihood (2.10) is here yet replaced with the Whittle approximation (2.37) in order to obtain stable and efficient Bayesian estimation algorithms that can actually be applied to images.

In this section, we propose two different exploitations of the Whittle approximation that translate into two Bayesian models. The first model directly uses the Whittle approximation (2.37) as such for inference. However, due to the particular form of (2.37), conjugate priors can be designed only for \mathbf{w} within this model. The second model interpretes (2.37) as a likelihood for the Fourier coefficients of the log-leaders, which is then, after augmentation of the model, expressed as the marginal of an augmented likelihood. Its convenient form enables to jointly design conjugate priors for \mathbf{w} and \mathbf{v} . For both models, inference is achieved by resorting to suitable Markov chain Monte Carlo (MCMC) algorithms, which are detailed in the next section.

2.4.1 Whittle approximation based Bayesian model

a) Likelihood

As mentioned in the introduction, since the exact likelihood (2.10) cannot be numerically evaluated, it is here replaced with the Whittle approximation (2.37).

b) Prior distribution

In this thesis, we assume a priori independence between \mathbf{v} and \mathbf{w} , which implies

$$\pi(\boldsymbol{\lambda}) = \pi(\mathbf{v})\pi(\mathbf{w}) \quad (2.41)$$

where $\pi(\mathbf{v})$ and $\pi(\mathbf{w})$ are independent priors specified below.

Mean parameter \mathbf{w} . Due to the quadratic form in (2.38), and hence (2.37), Gaussian distributions naturally arise as conjugate priors for \mathbf{w} . We therefore assign to \mathbf{w} the following prior

$$\pi(\mathbf{w}) = (2\pi \det \boldsymbol{\Sigma}_0)^{-\frac{1}{2}} \exp \left(-\frac{1}{2}(\mathbf{w}_0 - \mathbf{w})^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{w}_0 - \mathbf{w}) \right) \quad (2.42)$$

where \mathbf{w}_0 and $\boldsymbol{\Sigma}_0$ are hyperparameters respectively controlling the mean and the covariance matrix of the Gaussian prior. These hyperparameters can be tuned by the practitioner to either inject prior information, or, on the opposite, to specify a non-informative prior to reflect the absence of knowledge.

Covariance parameter \mathbf{v} . The scale-wise Whittle approximation (2.38), and hence joint Whittle approximation (2.37), is not separable in v_1 and v_2 , so that, unfortunately, the identification of conjugate priors for \mathbf{v} is very challenging. Due to the reparametrization, the constraints associated with the problem are satisfied when priors with support \mathbb{R}_*^{+2} are assigned to \mathbf{v} . We moreover assume a priori independence between v_1 and v_2 , which yields

$$p(\mathbf{v}) = p(v_1)p(v_2) \quad (2.43)$$

with $p(v_i)$ generically referring to a prior with support on \mathbb{R}_*^+ (such as, for instance, improper uniform priors on \mathbb{R}_*^+ , chi-square distributions or inverse-gamma distributions).

c) Posterior distribution and Bayesian estimators

The joint posterior distribution of the unknown parameters given the data ℓ is obtained from the Bayes rule where the exact likelihood is replaced by the Whittle approximation (2.37), i.e.,

$$\begin{aligned} p(\boldsymbol{\lambda}|\ell) &\propto p_W(\ell|\boldsymbol{\lambda})p(\boldsymbol{\lambda}) \\ &\propto p_W(\ell|\boldsymbol{\lambda})p(v_1)p(v_2) \exp\left(-\frac{1}{2}(\mathbf{w}_0 - \mathbf{w})^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{w}_0 - \mathbf{w})\right). \end{aligned} \quad (2.44)$$

The posterior (2.44) can then be used to define Bayesian pointwise estimators. We consider here two Bayesian estimators: the maximum a posteriori (MAP) and minimum mean squared error (MMSE) estimators, respectively defined by

$$\boldsymbol{\lambda}^{\text{MMSE}} = \mathbb{E}[\boldsymbol{\lambda}|\ell] \quad (2.45)$$

and

$$\boldsymbol{\lambda}^{\text{MAP}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} p(\boldsymbol{\lambda}|\ell). \quad (2.46)$$

Unfortunately, it is difficult to obtain closed-form expressions for (2.45) and (2.46). To compute (2.45) and (2.46), we resort in Section 2.5 to MCMC methods to generate a large collection of samples that are distributed according to the posterior of interest. These samples are used in turn to approximate the Bayesian estimators.

2.4.2 Augmented likelihood based Bayesian model

With the Bayesian model above, MCMC methods achieving inference on the posterior distribution (2.44) involve acceptance/reject steps, such as Metropolis-Hasting moves, to handle sampling according to non-standard conditional distributions (see Section 2.5). In order to bypass these acceptance/reject procedures, we resort here to a data augmentation technique (see [TW87, DM01] for more details) and express the Whittle approximation (2.37) as the marginal likelihood of an augmented likelihood that lends itself well for the exploitation of conjugacy.

a) Augmented likelihood

Statistical interpretation of the Whittle approximation. The first step of the development consists of a statistical interpretation of the Whittle approximation (2.38) as a generative model for

the Fourier coefficients of the log-leaders. To that end, we rewrite (2.38) as

$$p_W(\ell_j|\boldsymbol{\lambda}) = (\det \boldsymbol{\Gamma}_{j,v})^{-1} \exp \left(-\mathbf{y}_{\ell_j}^H \boldsymbol{\Gamma}_{j,v}^{-1} \mathbf{y}_{\ell_j} \right) \times (\phi_{j,v}(\boldsymbol{\omega}_0))^{-\frac{1}{2}} \exp \left(-\frac{(y_{\ell_j}(\boldsymbol{\omega}_0) - m_j \mu_{j,\mathbf{w}})^2}{2\phi_{j,v}(\boldsymbol{\omega}_0)} \right) \quad (2.47)$$

with $\mathbf{y}_{\ell_j} = \mathcal{F}_{J_j}(\ell_j)$, where the operator $\mathcal{F}_{J_j}(\cdot)$ computes and vectorizes the Fourier coefficients contained in the plane J_j defined in (2.15), $(\cdot)^H$ is the conjugate transpose operator and $\boldsymbol{\Gamma}_{j,v}$ is the diagonal matrix defined by

$$\boldsymbol{\Gamma}_{j,v} = v_1 \tilde{\mathbf{F}}_{1,j} + v_2 \tilde{\mathbf{F}}_{2,j} \quad \text{with} \quad \tilde{\mathbf{F}}_{i,j} = \text{diag} \left(\tilde{\mathbf{f}}_{i,j} \right) \quad \text{and} \quad \tilde{\mathbf{f}}_{i,j} = \left(\tilde{f}_{i,j}(\boldsymbol{\omega}_{\mathbf{m}}) \right)_{\mathbf{m} \in J_j}. \quad (2.48)$$

By inspection of (2.47), the use of the Whittle approximation $p_W(\ell_j|\boldsymbol{\lambda})$ to numerically evaluate (2.9) amounts to jointly modeling $\mathbf{y}_{\ell_j} = \mathcal{F}_{J_j}(\ell_j)$ by a random vector with a non-degenerate centered circularly-symmetric complex-valued Gaussian distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}_{j,v})$ (see, e.g., [Goo63, RC96]) and $y_{\ell_j}(\boldsymbol{\omega}_0)$ by a random variable with a normal distribution $\mathcal{N}(m_j \mu_{j,\mathbf{w}}, \phi_{j,v}(\boldsymbol{\omega}_0))$, respectively³. Details on the interpretation of the Whittle approximation as a spectral likelihood can be found in [BR12, Cha97], where the asymptotic normality of Fourier coefficients is discussed. It follows that, when introducing

$$\mathbf{y}_{\emptyset} = [\mathbf{y}_{\ell_{j_1}}^T, \dots, \mathbf{y}_{\ell_{j_2}}^T]^T \quad (2.49a)$$

$$\mathbf{y}_0 = [y_{\ell_{j_1}}(\boldsymbol{\omega}_0), \dots, y_{\ell_{j_2}}(\boldsymbol{\omega}_0)]^T \quad (2.49b)$$

we can develop (2.16) into

$$p_W(\ell|\boldsymbol{\lambda}) \propto (\det \boldsymbol{\Gamma}_v)^{-1} \exp \left(-\mathbf{y}_{\emptyset}^H \boldsymbol{\Gamma}_v^{-1} \mathbf{y}_{\emptyset} \right) \times (\det \boldsymbol{\Gamma}_{v,0})^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{y}_0 - \mathbf{X}\mathbf{w})^T \boldsymbol{\Gamma}_{v,0}^{-1} (\mathbf{y}_0 - \mathbf{X}\mathbf{w}) \right) \quad (2.50)$$

which indicates that the approximation of the joint likelihood (2.37) is equivalent to the following generative statistical model for the Fourier coefficients

$$\begin{cases} \mathbf{y}_{\emptyset}|\boldsymbol{\lambda} & \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}_v) \\ \mathbf{y}_0|\boldsymbol{\lambda} & \sim \mathcal{N}(\mathbf{X}\mathbf{w}, \boldsymbol{\Gamma}_{v,0}) \end{cases} \quad (2.51)$$

³Note that the frequency coefficient $y_{\ell_j}(\boldsymbol{\omega}_0)$ is by definition real-valued, since $y_{\ell_j}(\boldsymbol{\omega}_0) = \sum_{\mathbf{k}} l(j, \mathbf{k})/m_j$, while the other coefficients are complex-valued.

where $\mathbf{\Gamma}_v$ is the $N_\theta \times N_\theta$ diagonal covariance matrix, $N_\theta = \text{card}(\mathbf{y}_\theta)$, defined as

$$\mathbf{\Gamma}_v = v_1 \tilde{\mathbf{F}}_1 + v_2 \tilde{\mathbf{F}}_2 \quad \text{with} \quad \tilde{\mathbf{F}}_i = \text{diag}(\tilde{\mathbf{f}}_i) \quad \text{and} \quad \tilde{\mathbf{f}}_i = [\tilde{\mathbf{f}}_{i,j_1}^T, \dots, \tilde{\mathbf{f}}_{i,j_2}^T]^T, \quad (2.52)$$

$\mathbf{\Gamma}_{v,0}$ is the $N_0 \times N_0$ diagonal covariance matrix, $N_0 = \text{card}(\mathbf{y}_0)$, defined as

$$\mathbf{\Gamma}_{v,0} = v_1 \tilde{\mathbf{F}}_{0,1} + v_2 \tilde{\mathbf{F}}_{0,2} \quad \text{with} \quad \tilde{\mathbf{F}}_{0,i} = \text{diag}(\tilde{\mathbf{f}}_{0,i}) \quad \text{and} \quad \tilde{\mathbf{f}}_{0,i} = [\tilde{f}_{i,j_1}(\boldsymbol{\omega}_0), \dots, \tilde{f}_{i,j_2}(\boldsymbol{\omega}_0)]^T \quad (2.53)$$

and \mathbf{X} is a $N_0 \times 2$ matrix defined as

$$\mathbf{X} = \begin{bmatrix} j_1 m_{j_1} \log 2 & m_{j_1} \\ \vdots & \vdots \\ j_2 m_{j_2} \log 2 & m_{j_2} \end{bmatrix}. \quad (2.54)$$

To sum up, the statistical model (2.51) for the spectral observations $\mathbf{y} = [\mathbf{y}_\theta^T, \mathbf{y}_0^T]$ associated with the likelihood

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\lambda}) &= (\det \mathbf{\Gamma}_v)^{-1} \exp\left(-\mathbf{y}_\theta^H \mathbf{\Gamma}_v^{-1} \mathbf{y}_\theta\right) \\ &\times (\det \mathbf{\Gamma}_{v,0})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_0 - \mathbf{X}\mathbf{w})^T \mathbf{\Gamma}_{v,0}^{-1}(\mathbf{y}_0 - \mathbf{X}\mathbf{w})\right) \end{aligned} \quad (2.55)$$

is equivalent to the model (2.37) for the log-leaders ℓ . We finally note that, due to definitions (2.52) and (2.53), diagonal matrices $v_i \tilde{\mathbf{F}}_i$ and $v_i \tilde{\mathbf{F}}_{0,i}$ contain only strictly positive terms for $\mathbf{v} \in \mathbb{R}_*^{+2}$ (since $\tilde{f}_{i,j}(\boldsymbol{\omega}_m) > 0$ for $\forall \boldsymbol{\omega}_m$) and are hence positive definite. This key property is leveraged in the next paragraph for constructing an augmented model where these matrices individually act as covariance matrices of Gaussian distributions.

Data augmentation. By using the positive definiteness of $v_i \tilde{\mathbf{F}}_i$ for $i = 1, 2$, we can define the following augmented model $\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}$ (see [TW87, DM01] for more details)

$$\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda} \left\{ \begin{array}{ll} \mathbf{y}_\theta|\boldsymbol{\lambda}, \boldsymbol{\mu}_\theta \sim \mathcal{CN}(\boldsymbol{\mu}_\theta, v_1 \tilde{\mathbf{F}}_1), & \boldsymbol{\mu}_\theta|\boldsymbol{\lambda} \sim \mathcal{CN}(\mathbf{0}, v_2 \tilde{\mathbf{F}}_2) \\ \mathbf{y}_0|\boldsymbol{\lambda}, \boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, v_1 \tilde{\mathbf{F}}_{0,1}), & \boldsymbol{\mu}_0|\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{X}\mathbf{w}, v_2 \tilde{\mathbf{F}}_{0,2}) \end{array} \right. \quad (2.56)$$

where $\boldsymbol{\mu} = [\boldsymbol{\mu}_\theta^T, \boldsymbol{\mu}_0^T]$ is a latent variable vector introduced such that the likelihood (2.55) is by construction the marginal likelihood of the augmented model (see Appendix B for details), i.e.,

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \int p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}) d\boldsymbol{\mu} \quad (2.57)$$

where $p(\mathbf{y}, \boldsymbol{\mu} | \boldsymbol{\lambda})$ is the augmented likelihood given by

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\mu} | \boldsymbol{\lambda}) &\propto v_1^{-N_\emptyset} \exp \left(-\frac{1}{v_1} (\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset)^H \tilde{\mathbf{F}}_1^{-1} (\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset) \right) v_2^{-N_\emptyset} \exp \left(-\frac{1}{v_2} \boldsymbol{\mu}_\emptyset^H \tilde{\mathbf{F}}_2^{-1} \boldsymbol{\mu}_\emptyset \right) \\ &\times v_1^{-\frac{N_0}{2}} \exp \left(-\frac{1}{2v_1} (\mathbf{y}_0 - \boldsymbol{\mu}_0)^T \tilde{\mathbf{F}}_{0,1}^{-1} (\mathbf{y}_0 - \boldsymbol{\mu}_0) \right) \\ &\times v_2^{-\frac{N_0}{2}} \exp \left(-\frac{1}{2v_2} (\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})^T \tilde{\mathbf{F}}_{0,2}^{-1} (\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w}) \right). \end{aligned} \quad (2.58)$$

Moreover, and most importantly, it can be seen that the particular form of the augmented likelihood (2.58) allows inverse-gamma and Gaussian distributions to be jointly used as conjugate priors for v_i and \mathbf{w} , respectively. As a result, inference on the resulting Bayesian model can be efficiently achieved without acceptance-reject moves, which is the main motivation behind this data augmentation scheme.

b) Prior distributions

Mean parameter \mathbf{w} . The conjugate prior for \mathbf{w} is still the Gaussian distribution and we therefore assign to \mathbf{w} the prior (2.42).

Covariance parameter \mathbf{v} . As stated previously, due to the form of the augmented likelihood (2.55), it makes sense to assign inverse-gamma priors $\mathcal{IG}(\alpha_i, \beta_i)$ to v_i , i.e.,

$$p(v_i) \propto v_i^{-(\alpha_i+1)} \exp \left(-\frac{\beta_i}{v_i} \right) \mathbf{1}_{\mathbb{R}_+^+}(v_i) \quad (2.59)$$

since they are conjugate, where α_i and β_i are hyperparameters that are respectively linked to the scale and the shape of the inverse gamma distribution (2.59).

c) Posterior distribution and Bayesian estimators

The posterior distribution building on the augmented likelihood (2.58) is directly given by

$$\begin{aligned}
p(\boldsymbol{\lambda}, \boldsymbol{\mu} | \mathbf{y}) &\propto v_1^{-N_\phi} \exp\left(-\frac{1}{v_1}(\mathbf{y}_\phi - \boldsymbol{\mu}_\phi)^H \tilde{\mathbf{F}}_1^{-1}(\mathbf{y}_\phi - \boldsymbol{\mu}_\phi)\right) v_2^{-N_\phi} \exp\left(-\frac{1}{v_2} \boldsymbol{\mu}_\phi^H \tilde{\mathbf{F}}_2^{-1} \boldsymbol{\mu}_\phi\right) \\
&\times v_1^{-\frac{N_0}{2}} \exp\left(-\frac{1}{2v_1}(\mathbf{y}_0 - \boldsymbol{\mu}_0)^T \tilde{\mathbf{F}}_{0,1}^{-1}(\mathbf{y}_0 - \boldsymbol{\mu}_0)\right) \\
&\times v_2^{-\frac{N_0}{2}} \exp\left(-\frac{1}{2v_2}(\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})^T \tilde{\mathbf{F}}_{0,2}^{-1}(\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})\right) \\
&\times v_1^{-(\alpha+1)} \exp\left(-\frac{\beta}{v_1}\right) \mathbb{1}_{\mathbb{R}_+^+}(v_1) \times v_2^{-(\alpha+1)} \exp\left(-\frac{\beta}{v_2}\right) \mathbb{1}_{\mathbb{R}_+^+}(v_2) \\
&\times \exp\left(-\frac{1}{2}(\mathbf{w}_0 - \mathbf{w})^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{w}_0 - \mathbf{w})\right).
\end{aligned} \tag{2.60}$$

Finally, since the vector of latent variables $\boldsymbol{\mu}$ is not interesting for multifractal analysis purposes and is here artificially introduced for computational convenience, we consider the marginal MMSE and MAP estimators respectively defined by

$$\boldsymbol{\lambda}^{\text{MMSE}} = \mathbb{E}[\boldsymbol{\lambda} | \mathbf{y}] \tag{2.61}$$

and

$$\boldsymbol{\lambda}^{\text{MAP}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} p(\boldsymbol{\lambda} | \mathbf{y}) \tag{2.62}$$

where the expectation and maximization are taken with respect to the marginal posterior distribution

$$p(\boldsymbol{\lambda} | \mathbf{y}) = \int p(\boldsymbol{\lambda}, \boldsymbol{\mu} | \mathbf{y}) d\boldsymbol{\mu} = p(\mathbf{y} | \boldsymbol{\lambda}) p(\boldsymbol{\lambda}). \tag{2.63}$$

Their computation is achieved in the next section with an MCMC algorithm specifically tailored for inference on augmented models.

2.5 Estimation algorithms

In this section, we introduce two MCMC algorithms addressing the generation of a large collection of samples distributed according the posterior distributions (2.44) and (2.60). These samples will be then used to approximate the associated Bayesian estimators.

2.5.1 Metropolis-within-Gibbs algorithm

The Metropolis-within-Gibbs algorithm is an MCMC algorithm using the principle of the Gibbs sampler and Metropolis-Hasting moves. More precisely, the Gibbs sampler consists of sequential sampling according to the conditional distributions associated with the posterior of interest (2.44) as explained in [GG84, RC05]. When some or all conditional distributions are not standard distributions and cannot be directly sampled, we can resort to Metropolis-Hasting moves to sample according to them [RC05].

a) Sampling of \mathbf{w}

From (2.44), it can be shown that the conditional distribution of \mathbf{w} is a Gaussian distribution defined by

$$\mathbf{w}|\boldsymbol{\theta}, \ell \sim \mathcal{N}((\mathbf{X}^T \boldsymbol{\Gamma}_{\boldsymbol{\theta},0}^{-1} \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}(\boldsymbol{\Sigma}_0^{-1} \mathbf{w}_0 + \mathbf{X}^T \boldsymbol{\Gamma}_{\boldsymbol{\theta},0}^{-1} \mathbf{y}_0), (\mathbf{X}^T \boldsymbol{\Gamma}_{\boldsymbol{\theta},0}^{-1} \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}) \quad (2.64)$$

which is easy to sample. As a result, this step does not require Metropolis-Hasting moves.

b) Sampling of v

Starting from (2.44) and using definitions (2.52) and (2.53), straightforward computations lead to the following conditional distribution for v_i given $v_{i' \neq i}, \mathbf{w}$ and ℓ

$$\begin{aligned} p(v_i|v_{i' \neq i}, \mathbf{w}, \ell) &\propto (\det v_i \tilde{\mathbf{F}}_i + v_{i' \neq i} \tilde{\mathbf{F}}_{i' \neq i})^{-1} \exp \left(-\mathbf{y}_\emptyset^H (v_i \tilde{\mathbf{F}}_i + v_{i' \neq i} \tilde{\mathbf{F}}_{i' \neq i})^{-1} \mathbf{y}_\emptyset \right) \\ &\quad \times (\det v_i \tilde{\mathbf{F}}_{0,i' \neq i} + v_{i' \neq i} \tilde{\mathbf{F}}_{0,i' \neq i})^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{y}_0 - \mathbf{X} \mathbf{w})^T (v_i \tilde{\mathbf{F}}_{0,i} + v_{i' \neq i} \tilde{\mathbf{F}}_{0,i' \neq i})^{-1} (\mathbf{y}_0 - \mathbf{X} \mathbf{w}) \right) \\ &\quad \times p(v_i) \end{aligned} \quad (2.65)$$

where we recall that $p(v_i)$ generically refers to a prior with support on \mathbb{R}_*^+ . Regardless the definition of $p(v_i)$, sampling from (2.65) is not straightforward but can be handled using a Metropolis-Hasting move. More specifically, we use here a random walk to generate a candidate v_i° which is then accepted or reject with a probability r_{v_i} derived from (2.65) and detailed in Algo. 1. The instrumental distributions for the random walks are Gaussian distributions whose variances $\sigma_{v_i}^2$ are adjusted to

ensure acceptance rates between 0.4 and 0.6 (to ensure good mixing properties) as advised in [RC05, RC98] for this kind of low dimensional problems.

Algorithm 1 Metropolis-within-Gibbs algorithm MwG to sample according to (2.44)

```

1: Initialization
2: Draw  $v_i^{(0)} \sim \mathcal{IG}(\alpha_i, \beta_i)$  for  $i = 1, 2$  and draw  $\mathbf{w}^{(0)} \sim \mathcal{N}(\mathbf{w}_0, \Sigma_0)$ 
3: MCMC iterations
4: for  $p = 1 : N_{mc}$ 
5:   Sample  $\mathbf{w}$ 
6:     Draw  $\mathbf{w}^{(p)}$  according to the Gaussian law (2.64)
7:   Sample  $\mathbf{v}$ 
8:     Set  $\mathbf{v}^\circ = [v_1^\circ, v_2^\circ]^T = \mathbf{v}^{(k-1)}$ 
9:     for  $i = 1 : 2$ 
10:       Draw  $v_i^* \sim \mathcal{N}(v_i^\circ, \sigma_{v_i}^2)$  and  $u \sim \mathcal{U}_{[0,1]}$ 
11:       Compute  $r_{v_i} = \min(1, \frac{p(v_i^* | v_{i' \neq i}^\circ, \mathbf{w}, \ell)}{p(v_i^\circ | v_{i' \neq i}^\circ, \mathbf{w}, \ell)})$  using (2.65)
12:       Set  $v_i^\circ = v_i^{(*)}$  if  $u < r_{v_i}$ , otherwise  $v_i^\circ = v_i^\circ$ 
13:     end for
14:     Set  $\mathbf{v}^{(p)} = [v_1^\circ, v_2^\circ]^T$ 
15: end for
16: Return  $\{\mathbf{v}^{(p)}, \mathbf{w}^{(p)}\}_{p=1}^{N_{mc}}$ 

```

c) Approximation of the Bayesian estimators

After a burn-in period defined by $p = 1, \dots, N_{bi}$, the proposed Gibbs sampler generates samples $\{\boldsymbol{\lambda}^{(p)}\}_{p=N_{bi}+1}^{N_{mc}}$ that are distributed according to the posterior distribution (2.44). These samples are then used to approximate the MAP and MMSE estimators as follows

$$\hat{\boldsymbol{\lambda}}^{\text{MMSE}} = \frac{1}{N_{mc} - N_{bi}} \sum_{p=N_{bi}+1}^{N_{mc}} \boldsymbol{\lambda}^{(p)} \quad (2.66)$$

$$\hat{\boldsymbol{\lambda}}^{\text{MAP}} = \underset{\boldsymbol{\lambda}^{(p)}, p > N_{bi}}{\operatorname{argmax}} p(\boldsymbol{\lambda}^{(p)} | \ell). \quad (2.67)$$

2.5.2 Two-stage Gibbs sampler

The reformulation of the spectral likelihood (2.55) as the marginal likelihood of an augmented model in Section 2.4.2 leads to a two-stage Gibbs sampler as in [TW87, DM01] comprising two steps: First sampling the latent variables $\boldsymbol{\mu}$ given the parameters $\boldsymbol{\lambda}$ and the observations \mathbf{y} ; Second, sampling the parameters $\boldsymbol{\lambda}$ given the complete data, namely the observations \mathbf{y} and the latent variables $\boldsymbol{\mu}$.

a) Step 1: Sampling the latent variables $\boldsymbol{\mu}$

Sampling $\boldsymbol{\mu}_\emptyset$. Simple manipulations on (2.60) yield the following conditional distribution

$$p(\boldsymbol{\mu}_\emptyset | \boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\mu}_0) \propto \exp(-(\boldsymbol{\mu}_\emptyset - \tilde{\mathbf{x}}_v)^H \tilde{\boldsymbol{\Sigma}}_v^{-1} (\boldsymbol{\mu}_\emptyset - \tilde{\mathbf{x}}_v)) \quad (2.68)$$

where

$$\tilde{\mathbf{x}}_v = \tilde{\boldsymbol{\Sigma}}_v (\tilde{\mathbf{F}}_1 v_1)^{-1} \mathbf{y}_\emptyset \quad (2.69)$$

$$\tilde{\boldsymbol{\Sigma}}_v = ((\tilde{\mathbf{F}}_1 v_1)^{-1} + (\tilde{\mathbf{F}}_2 v_2)^{-1})^{-1} \quad (2.70)$$

i.e., the conditional distribution $\boldsymbol{\mu}_\emptyset | \boldsymbol{\lambda}, \ell$ is the following complex-valued Gaussian distribution

$$\boldsymbol{\mu}_\emptyset | \boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\mu}_0 \sim \mathcal{CN}(\tilde{\mathbf{x}}_v, \tilde{\boldsymbol{\Sigma}}_v). \quad (2.71)$$

Sampling of $\boldsymbol{\mu}_0$. Similarly, it can be shown that the conditional distribution $\boldsymbol{\mu}_0 | \boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\mu}_\emptyset$ is a (real-valued) Gaussian distribution given by

$$\boldsymbol{\mu}_0 | \boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\mu}_\emptyset \sim \mathcal{N}(\tilde{\mathbf{x}}_{0,v,\mathbf{w}}, \tilde{\boldsymbol{\Sigma}}_{0,v}) \quad (2.72)$$

with

$$\tilde{\mathbf{x}}_{0,v,\mathbf{w}} = \tilde{\boldsymbol{\Sigma}}_{0,v} ((\tilde{\mathbf{F}}_{0,1} v_1)^{-1} \mathbf{y}_0 + (\tilde{\mathbf{F}}_{0,2} v_2)^{-1} \mathbf{X} \mathbf{w}) \quad (2.73)$$

$$\tilde{\boldsymbol{\Sigma}}_{0,v} = ((\tilde{\mathbf{F}}_{0,1} v_1)^{-1} + (\tilde{\mathbf{F}}_{0,2} v_2)^{-1})^{-1}. \quad (2.74)$$

b) Step 2: Sampling of the parameters λ

Sampling of \mathbf{w} . From (2.60), it can easily be shown that the conditional distribution of $\mathbf{w}|\mathbf{v}, \mathbf{y}, \boldsymbol{\mu}$ is the following Gaussian distribution

$$\mathbf{w}|\mathbf{v}, \mathbf{y}, \boldsymbol{\mu} \sim \mathcal{N} \left(\left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \left(\boldsymbol{\Sigma}_0^{-1} \mathbf{w}_0 + \frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \boldsymbol{\mu}_0}{v_2} \right), \left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \right). \quad (2.75)$$

Sampling of \mathbf{v} . The conditional distributions of covariance parameters v_i are easy to sample. Indeed, by inspection of the augmented posterior distribution (2.60), it is straightforward to see that $v_i|v_{i' \neq i}, \mathbf{w}, \mathbf{y}, \boldsymbol{\mu}$ have inverse-gamma distributions defined by

$$v_1|v_2, \mathbf{w}, \mathbf{y}, \boldsymbol{\mu} \sim \mathcal{IG} \left(N_\emptyset + \frac{N_0}{2} + \alpha_1, \|\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset\|_{\tilde{\mathbf{F}}_1^{-1}} + \frac{1}{2} \|\mathbf{y}_0 - \boldsymbol{\mu}_0\|_{\tilde{\mathbf{F}}_{0,1}^{-1}} + \beta_1 \right) \quad (2.76)$$

and

$$v_2|v_1, \mathbf{w}, \mathbf{y}, \boldsymbol{\mu} \sim \mathcal{IG} \left(N_\emptyset + \frac{N_0}{2} + \alpha_2, \|\boldsymbol{\mu}_\emptyset\|_{\tilde{\mathbf{F}}_2^{-1}} + \frac{1}{2} \|\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w}\|_{\tilde{\mathbf{F}}_{0,2}^{-1}} + \beta_2 \right) \quad (2.77)$$

where we introduce the notation $\|\mathbf{x}\|_{\mathbf{M}} = \mathbf{x}^H \mathbf{M} \mathbf{x}$.

c) Approximation of Bayesian estimators

The sampling scheme above provides a Markov chain $\{\boldsymbol{\lambda}^{(p)}, \boldsymbol{\mu}^{(p)}\}_{p=1}^{N_{mc}}$ whose limiting distribution is the augmented posterior distribution $p(\boldsymbol{\lambda}, \boldsymbol{\mu}|\mathbf{y})$ [TW87, DM01]. The marginalization of the latent variable $\boldsymbol{\mu}$ is then numerically achieved by averaging the samples as follows

$$\hat{\boldsymbol{\lambda}}^{\text{MMSE}} = \frac{1}{N_{mc} - N_{bi}} \sum_{p=N_{bi}+1}^{N_{mc}} \boldsymbol{\lambda}^{(p)} \quad (2.78)$$

for the computation of the MMSE estimator (2.61). Finally, the MAP estimator (2.62) is simply computed as follows

$$\hat{\boldsymbol{\lambda}}^{\text{MAP}} = \underset{\boldsymbol{\lambda}^{(p)}, p > N_{bi}}{\operatorname{argmax}} p(\boldsymbol{\lambda}^{(p)}|\mathbf{y}). \quad (2.79)$$

Algorithm 2 Two-stage Gibbs sampler TSG to sample according (2.60)

- 1: Initialization
 - 2: Draw $v_i^{(0)} \sim \mathcal{IG}(\alpha_i, \beta_i)$ for $i = 1, 2$ and draw $\mathbf{w}^{(0)} \sim \mathcal{N}(\mathbf{w}_0, \Sigma_0)$
 - 3: MCMC iterations
 - 4: **for** $p = 1 : N_{mc}$
 - 5: Step 1: Sample latent variables $\boldsymbol{\mu}$
 - 6: Draw $\boldsymbol{\mu}_{\emptyset}^{(p)}$ according to the complex-valued Gaussian law (2.71)
 - 7: Draw $\boldsymbol{\mu}_0^{(p)}$ according to the real-valued Gaussian law (2.72)
 - 8: Step 2: Sample parameters $\boldsymbol{\lambda}$
 - 9: Sample \mathbf{w}
 - 10: Draw $\mathbf{w}^{(p)}$ according to the Gaussian law (2.75)
 - 11: Sample \mathbf{v}
 - 12: Draw $v_1^{(p)}$ according to the inverse-gamma law (2.76)
 - 13: Draw $v_2^{(p)}$ according to the inverse-gamma law (2.77)
 - 14: **end for**
 - 15: Return $\{\mathbf{v}^{(p)}, \mathbf{w}^{(p)}\}_{p=1}^{N_{mc}}$
-

2.6 Numerical experiments

We have introduced in the previous sections a Bayesian approach for the estimation of the two leading log-cumulants c_1 and c_2 relying on a parametric statistical model for the log-leaders of images. In this section, we numerically investigate the performance of this approach. To do so, we apply MCMC algorithms to several types of scale invariant and multifractal 2D stochastic processes for different image sizes and a large range of values for c_1 and c_2 , and compare them to the current standard and benchmark estimators (defined in Chapter 1, cf., (1.27) and (1.28)) relying on linear regressions. Those are generically referred to as LF estimators in what follows.

2.6.1 General setup

Wavelet transform. A Daubechies' mother wavelet with $N_\psi = 2$ vanishing moments is used, and $\alpha = 0.7$ in (1.17).

Analysis scales. The coarsest scale j_2 used for estimation is set such that $n_{j_2} \geq 100$ (i.e., the coarsest potentially available scale, for which $n_{j_2} < 100$, is discarded), yielding $j_2 = \{2, 3, 4, 5, 6, 7\}$, respectively, for the considered image sizes $N \in \{2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}\}$. Unless specified otherwise, the finest scale j_1 is set to $j_1 = 2$ in order to avoid pollution from improper initialization of the wavelet transform (see, e.g., [VTA00]).

Prior specification. For these simulations, in order to account for the absence of prior information about the mean parameter \mathbf{w} , we specify a non-informative prior (2.42) by setting $\mathbf{w}_0 = [0, 0]^T$ and $\Sigma_0 = \text{diag}(\sigma^2, \sigma^2)$, with $\sigma^2 \gg 1$. Similarly, the hyperparameters of the inverse-gamma priors $\mathcal{IG}(\alpha_i, \beta_i)$ (2.59) are set to $(\alpha_i, \beta_i) = (10^{-3}, 10^{-3})$, which ensures that they are close to a non-informative Jeffreys' prior.

Performance assessment. We apply the LF estimators and the proposed Bayesian estimators to a large number ($R = 100$) of independent realizations of 2D multifractal processes. Performance is then evaluated using the sample mean or bias, the sample standard deviation (STD) and the root mean squared error (RMSE) of the estimates averaged across realizations

$$m_{c_1} = \widehat{\mathbb{E}}[\hat{c}_1], \quad b_{c_1} = m_{c_1} - c_1, \quad s_{c_1} = \sqrt{\widehat{\text{Var}}[\hat{c}_1]}, \quad r_{c_1} = \sqrt{b_{c_1}^2 + s_{c_1}^2} \quad (2.80a)$$

$$m_{c_2} = \widehat{\mathbb{E}}[\hat{c}_2], \quad b_{c_2} = m_{c_2} - c_2, \quad s_{c_2} = \sqrt{\widehat{\text{Var}}[\hat{c}_2]}, \quad r_{c_2} = \sqrt{b_{c_2}^2 + s_{c_2}^2}. \quad (2.80b)$$

2.6.2 Adjusting the frequency range in the Whittle approximation

Before studying the performance of the proposed Bayesian approach, we investigate here a degree of freedom of the Whittle approximation, which lies in the selection of the range of frequency used in (2.12). The use of only a part of the available frequencies is commonly reported in the literature, especially for the periodogram-based estimation of the memory coefficient of long-range dependent time series for which only the low frequencies are used, see, e.g., [Ber94, VR00, Rob95]. Fig. 2.4

compares the parametric spectral density model $\phi_{j,\theta}(\omega)$ to the periodogram of $l(j, \mathbf{k})$, denoted as $I_{\ell_j}(\omega_{\mathbf{m}})$ and defined in Eq. (2.14), averaged over 100 realizations of CMC-LN ($N = 2^9$ and $[c_1, c_2] = [0.02, -0.04]$). This figure indicates that the proposed model yields an excellent fit at low frequencies that degrades at higher frequencies. As a result, we propose here to only use the low frequencies in (2.81), i.e.,

$$p_W(\ell_j | \lambda) = \prod_{\mathbf{m} \in J_j^\dagger(\eta)} (\phi_{j,\theta}(\omega_{\mathbf{m}}))^{-1} \exp \left(-\frac{y_{\ell_j}^*(\omega_{\mathbf{m}}) y_{\ell_j}(\omega_{\mathbf{m}})}{\phi_{j,\theta}(\omega_{\mathbf{m}})} \right) \\ \times (\phi_{j,\theta}(\omega_0))^{-\frac{1}{2}} \exp \left(\frac{(y_{\ell_j}(\omega_0) - m_j \mu_{j,\mathbf{w}})^2}{2\phi_{j,\theta}(\omega_0)} \right) \quad (2.81)$$

with

$$J_j^\dagger(\eta) = \left\{ \mathbf{m} \in J_j \mid \mathbf{m} \leq \sqrt{\eta} \lfloor \frac{m_j}{2} \rfloor \right\} \quad (2.82)$$

and we discuss below the influence of the bandwidth parameter η controlling the fraction of the spectral grid that is actually used ⁴. Fig. 2.6 reports estimation performance of the MMSE estimator (computed with the TSG Algo. 2) for c_2 as a function of η for 2D CMC-LN processes ($N = 2^8$ top, $N = 2^9$ bottom) with two different values of c_2 (0.02 in red, 0.08 in blue). As expected, η is found to tune a classical bias-variance tradeoff: a large value of η leads to a large bias and small standard deviation and vice versa. The choice $\eta = 0.25$ yields a robust practical compromise and is therefore used in the rest of this thesis.

2.6.3 Numerical assessment of equivalence between Bayesian models

In Section 2.4, we have formulated two Bayesian models building either on the Whittle approximation (2.37) itself or the augmented spectral likelihood (2.58). In this section, we show by means of numerical experiments that these two Bayesian models are equivalent when they incorporate the same prior information on \mathbf{w} and \mathbf{v} . To do so, we use for both Bayesian models the non-informative Gaussian and inverse-gamma distributions as defined in the general setup (see Subsection 2.6.1) as

⁴It is important to note that the use of low frequencies only in (2.81) is not a restriction to a limited frequency content of the image (indeed, scales $j \in [j_1, j_2]$ are used) but only concerns the numerical evaluation of the likelihood (2.38) in the spectral domain.

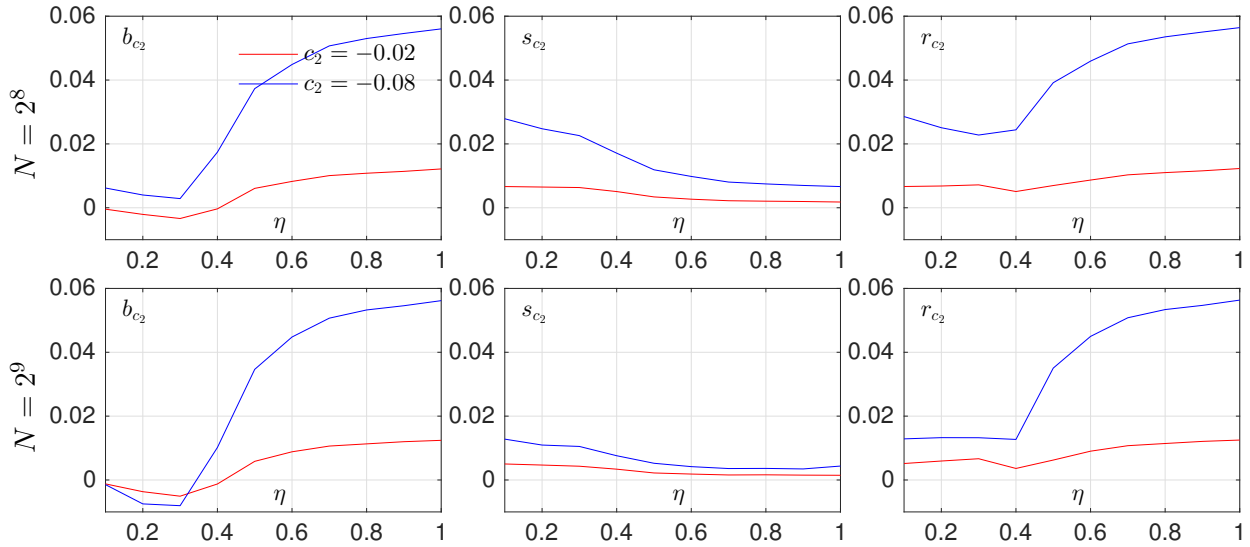


Figure 2.6: Influence of the bandwidth parameter η on estimation performance for two different sizes of 2D CMC-LN and two different values of c_2 values (-0.02 in red and -0.08 in blue).

priors and run the MCMC algorithms MwG and TSG, respectively designed for the Whittle approximation based Bayesian model (associated with the posterior (2.44)) and the augmented likelihood based one (associated with the posterior (2.60)).

We first display in Fig. 2.7, the estimated marginal posterior distributions of c_1 and $c_2 = -v_1$ obtained by running the MCMC algorithms on a single realization of 2D MRW ($N = 2^8$ and $[c_1, c_2] = [0.7, -0.02]$) and computing the normalized histograms. A visual inspection of Fig. 2.7 clearly indicates that the marginal posterior distributions induced by the two posterior distributions (2.44) and (2.60) are very similar.

We moreover report in Tab. 2.1 estimation performance of the Bayesian estimators assessed on 100 realizations of 2D MRW. It is observed that MCMC algorithms yield estimators (MAP and MMSE) with numerically very close performance. This confirms that, when the same prior distributions are used, both approaches (either directly working with the Whittle approximation (2.37) or the augmented spectral likelihood (2.58)) lead to posterior distributions and associated Bayesian estimators that are numerically equivalent.

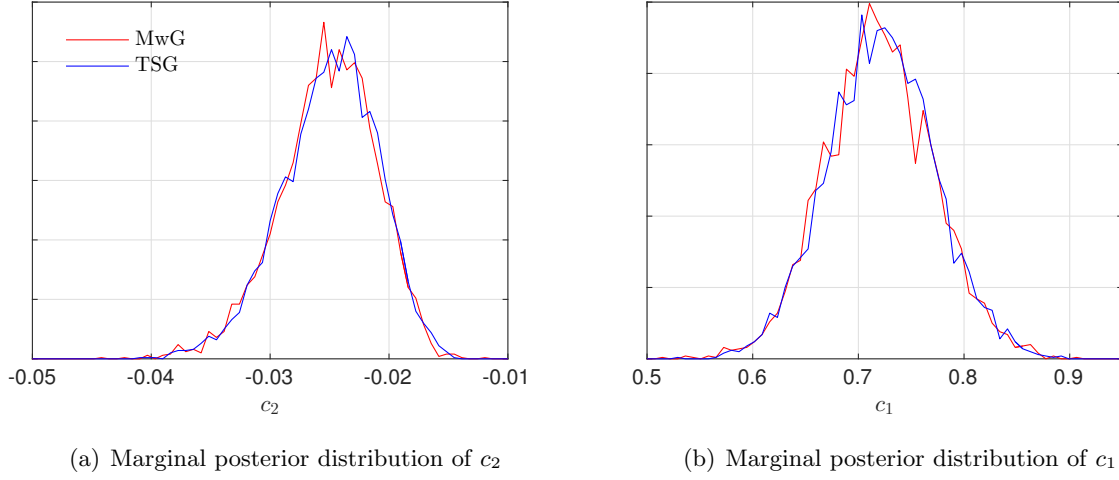


Figure 2.7: Marginal posterior distribution of (a) c_2 and (b) c_1 associated with the posterior distribution (2.44) (red) and (2.60) (blue) obtained by running MCMC algorithms on a single realization of 2D MRW ($N = 2^8$ and $[c_1, c_2] = [0.7, -0.02]$).

Table 2.1: Estimation performance of Bayesian estimators associated with posterior distributions (2.44) (computed with MwG) and (2.60) (computed with TSG) assessed on 100 realizations of 2D MRW ($N = 2^8$ and $[c_1, c_2] = [0.7, -0.02]$)

		(2.44) - MwG		(2.60) - TSG	
		c_1	c_2	c_1	c_2
MEAN	MMSE	0.715	-0.024	0.720	-0.024
	MAP	0.713	-0.023	0.713	-0.023
STD	MMSE	0.046	0.005	0.046	0.005
	MAP	0.045	0.005	0.046	0.005

In the rest of this thesis, the use of either the augmented model (2.58) or the Whittle approximation (2.38) will depend on the nature of the prior on \mathbf{v} . When inverse-gamma type priors are assigned to \mathbf{v} (cf. Section 3.4 in the next chapter), the augmented model (2.58) yields an efficient sampling scheme whose advantages in terms of computational cost and convergence properties are illustrated in Subsection 2.6.6. In other cases (cf. Section 3.5 in Chapter 3), the augmented spectral

likelihood loses much of its attraction and we hence prefer to work directly with the Whittle approximation (2.37) to reduce the number of variables to be sampled.

2.6.4 Estimation performance for c_1

In this first set of experiments, we consider realizations of 2D MRW processes ($N = 2^8$) for which we set $c_1 = 0.7$ and make c_2 vary from strong (-0.1) to weak (-0.01) multifractality. We report in Fig. 2.8 estimation performance for Bayesian estimators of c_1 (computed with TGS) and non-weighted and weighted linear regression based estimators, respectively denoted as LF_{nw} and LF_{w} . Bayesian estimators are found to offer overall estimation performance very alike that of the linear regression approach (in particular LF_{nw}), both in terms of the bias and the standard deviation. Similar results have been observed for different members of MMC processes, different image sizes and different values of c_1 . This finding is not surprising since the proposed statistical model assumes essentially as much as the linear regression for the mean of the log-leaders, namely a linear evolution with a slope controlled by c_1 (cf. (1.15)).

This experiment leads us to the conclusion that the Bayesian approach does not yield improvements for the estimation of c_1 when it incorporates a non-informative prior on \mathbf{w} (and hence on c_1). As a result, in the rest of this thesis, when a non-informative prior is assigned to \mathbf{w} , estimation performance for c_1 is not discussed. Estimation of c_1 will be discussed again in Section 3.6 of Chapter 3, where we consider an informative smoothing prior for the collection of \mathbf{w} of multivariate images..

2.6.5 Estimation performance for c_2

In this last set of experiments, we investigate in more detail the performance of the proposed Bayesian approach for the estimation of c_2 and compare it to the linear regression based approach. Note that in this section, only performance of the weighted linear regression based estimator LF_{w} are reported for simplicity, since it has been found to yield better performance as observed in [WAJ07].

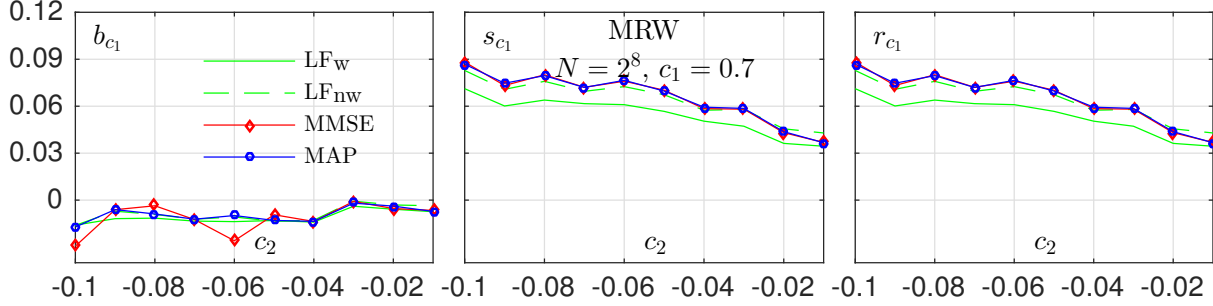


Figure 2.8: Estimation performance for c_1 assessed on 100 realizations of 2D MRW ($N = 2^8$, $c_1 = 0.7$ and varying c_2): bias, STD and RMSE from left to right.

a) Estimation performance

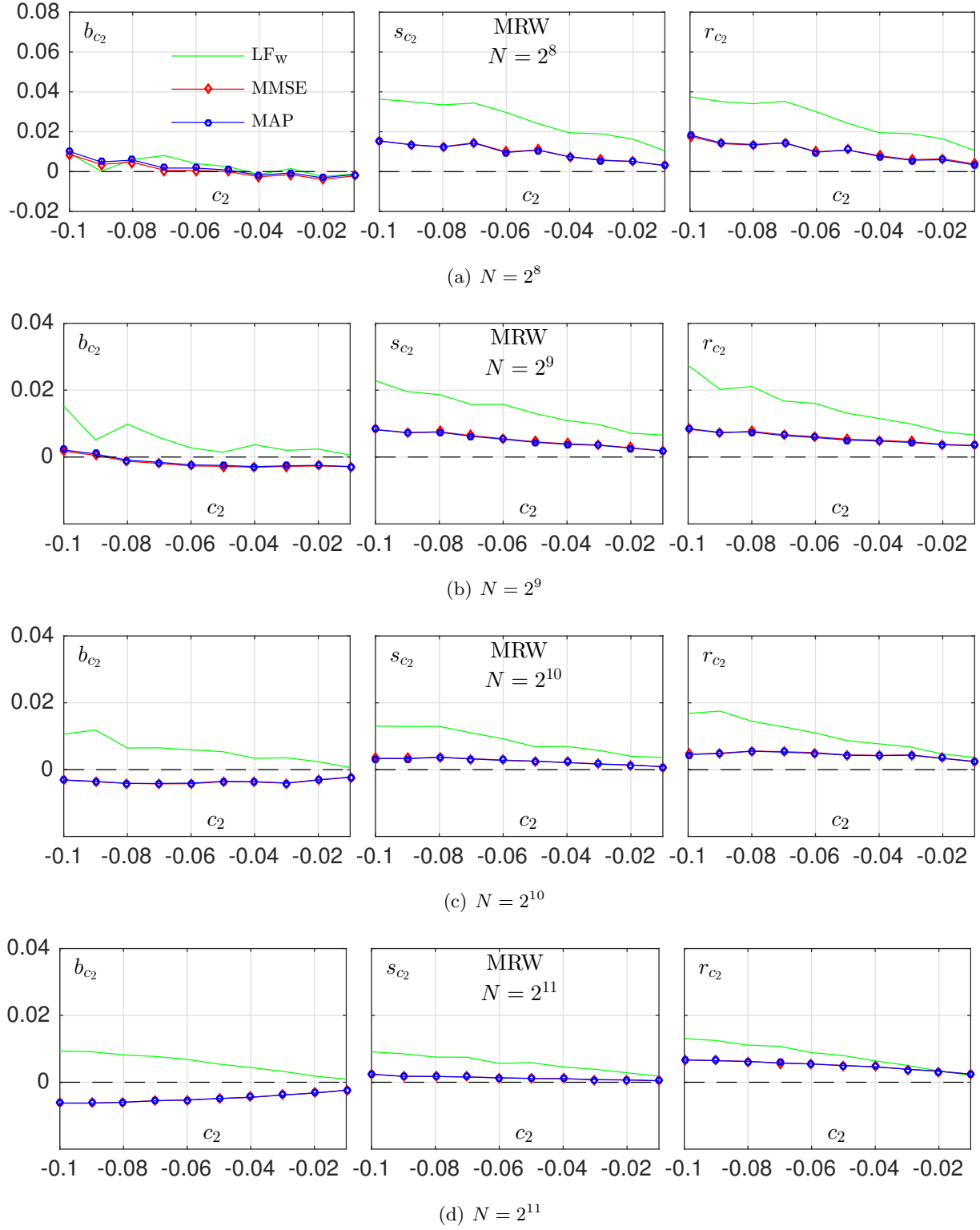
Fig. 2.9 summarizes the estimation performance of LF and Bayesian estimators (computed with TSG) for 2D MRW for medium to large image sizes $N = \{2^8, 2^9, 2^{10}, 2^{11}\}$ whereas Fig. 2.10 displays estimation performance for various 2D MMC processes (CMC-LN, CMC-LP, CPC-LN and CPC-LP) for $N = 2^8$.

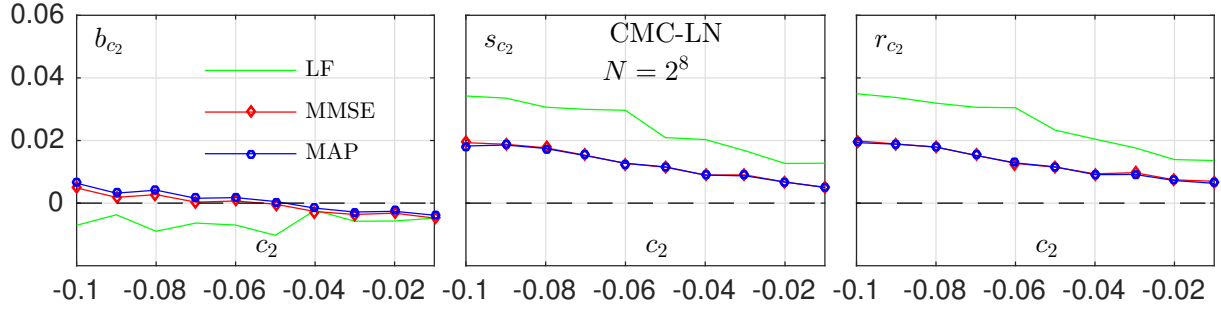
First, it is observed that the Bayesian approach tends to slightly outperform LF in terms of bias for almost all considered multifractal processes and image sizes. Second, and most strikingly, the proposed Bayesian estimators systematically yield significantly reduced standard deviations, with a reduction of up to a factor of 3 as compared to linear regressions. The standard deviation reduction is more important for small values of $|c_2|$ yet remains above a factor of 1.5 for large values of $|c_2|$.

These performance gains are directly reflected in the overall RMSE values, which remain up to a factor of 2.5 below those of linear fits. Finally, note that the estimation performance for CMCs and CPCs with log-Poisson multipliers reported in Fig. 2.10 are found to be slightly inferior to those with log-normal multipliers. This may be due to an arguably slightly stronger departure from Gaussian for the former, cf. Fig. 2.1.

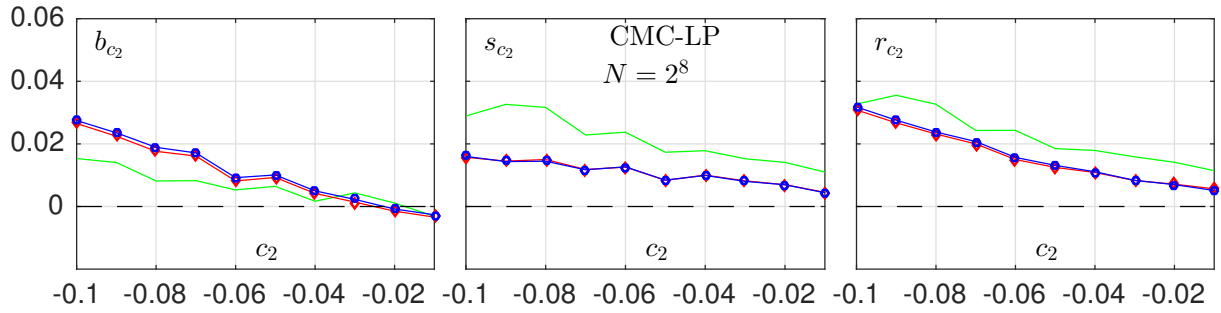
b) Performance for small image size

For the analysis of small images, $N \leq 2^7$, the limited number of available scales forces the choice $j_1 = 1$. Results for image sizes $N = \{2^6, 2^7\}$ (for which $j_2 = 2, 3$ respectively) are reported in Fig. 2.11

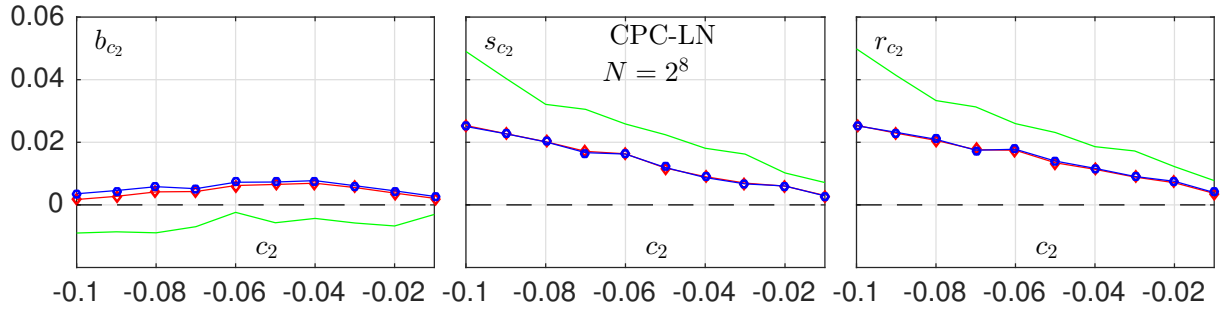
Figure 2.9: Estimation performance for c_2 for 2D MRW for different image sizes $N \in \{2^8, 2^9, 2^{10}, 2^{11}\}$.



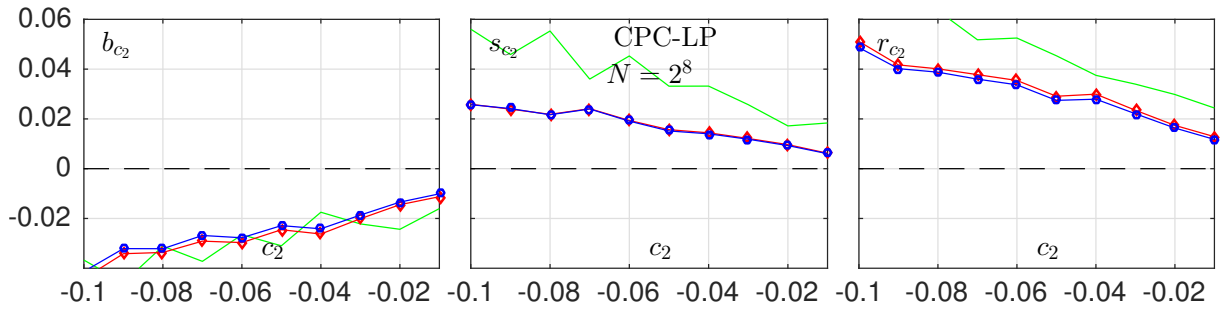
(a) CMC-LN



(b) CMC-LP



(c) CPC-LN



(d) CPC-LP

Figure 2.10: Estimation performance for c_2 for image size $N = 2^8$ for different 2D MMC processes.

for 2D MRW. They indicate that the performance gains of the proposed Bayesian estimators with respect to LF estimators are even more pronounced for small images, both in terms of bias and standard deviations, yielding a reduction of RMSE values of up to a factor of 4. In particular, LF is found to yield standard deviations that are prohibitively large to be useful in real-world applications, cf. [WRJA09]. Notably, values $c_2 = 0$ cannot be reliably detected with LF. In contrast, the proposed Bayesian procedure yields sufficiently small bias and standard deviations to enable, for the first time, the relevant estimation of the multifractality parameter c_2 even for very small images (or image patches) of size 64×64 .

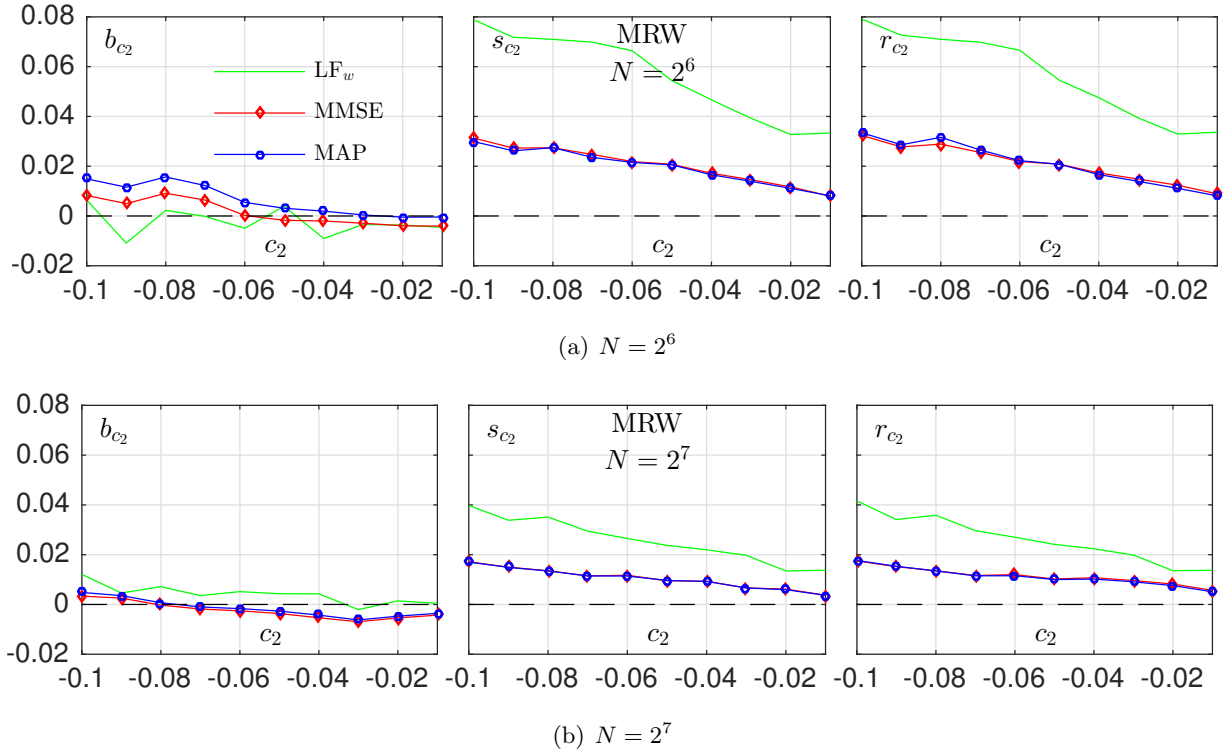


Figure 2.11: Estimation performance for c_2 for 2D MRW for small image sizes $N \in \{2^6, 2^7\}$ with $j_1 = 1$.

c) Performance for fractional Brownian motion

Self-similar fBms do not belong to the class of MMC processes for which the proposed estimation procedure was designed. The correlation structure of the wavelet coefficients of fBms has been studied in, e.g., [Fla92]. This correlation is weak, i.e., it goes to zero fast with the distance between wavelet coefficients in the time-scale plane. However, we apply here the proposed approach to test its robustness and summarize the corresponding results in Tab. 2.2. They indicate that the performance of the LF estimator is comparable to MMC for $c_2 = -0.01$ for the corresponding image size. In contrast, the proposed Bayesian estimators are strikingly practically unbiased and have standard deviation (resp. RMSE) values that significantly outperform those of LF by up to a factor 20 (resp. 10). Therefore, the Bayesian approach is found to be actually much more likely to be able to identify a model for which $c_2 = 0$ than the classical linear regression approach.

Table 2.2: 2D Fbm estimation performance for image sizes $N = \{2^6, 2^7, 2^8, 2^9, 2^{10}\}$ and $j_1 = \{1, 1, 2, 2, 2\}$, $j_2 = \{2, 3, 4, 5, 6\}$. The lower the better and best results are marked in bold.

N		2^6	2^7	2^8	2^9	2^{10}
MEAN	LF	-0.0254	-0.0127	0.0043	0.0052	0.0044
	MMSE	-0.0040	-0.0013	-0.0011	-0.0006	-0.0005
	MAP	-0.0015	-0.0008	-0.0008	-0.0004	-0.0003
STD	LF	0.0184	0.0061	0.0070	0.0028	0.0016
	MMSE	0.0022	0.0006	0.0005	0.0002	0.0001
	MAP	0.0017	0.0003	0.0003	0.0001	0.0001
RMSE	LF	0.0313	0.0140	0.0082	0.0059	0.0047
	MMSE	0.0046	0.0014	0.0012	0.0006	0.0005
	MAP	0.0022	0.0009	0.0009	0.0004	0.0003

The set of the three experiments above provides strong evidence for the benefits of the Bayesian approach over the linear regression for the estimation of the second log-cumulant c_2 . These benefits mainly lie in a significant reduction of the standard deviation values, which has been observed regardless the considered scale invariant model, the image size or the value of c_2 .

2.6.6 Convergence and computational cost

In this last set of experiments, we finally discuss the convergence of MCMC methods and the computational cost of the proposed methods.

a) Convergence of MCMC algorithms

We first investigate the convergence properties of the proposed MCMC algorithms. To do so, we display in Fig. 2.13 (a) the evolution of the draws $c_2^{(p)}$ provided by the MwG and TSG algorithms (with identical random initializations) averaged over 100 realizations of 2D MRW ($N = 2^8$ and $[c_1, c_2] = [0.7, -0.01]$). Clearly, the augmented model used in TSG leads to Markov chains (MCs) that converge almost immediately to the stationary distribution (whose mean is displayed in black), while MwG requires a much longer burn-in period. Indeed, the accept-reject procedure of MwG requires tuning of the (adaptive) step size of the random walk propositions, which is precisely bypassed by the data augmentation approach.

An analysis of the mixing properties of the MCMC algorithms is given in Fig. 2.13 (b), which plots, as a function of the lag Δp , the correlation between two draws that are distant of Δp iterations. These figures show that the TSG algorithm succeeds in generating samples with low correlation, indicating hence good mixing properties [CC96, RC98]. In contrast, the samples $c_2^{(p)}$ provided by MwG remain correlated for a significant range of values of lags Δp .

The results above suggest that the number of samples can be significantly reduced for TSG as compared to MwG. Complementary simulations for a wide range of multifractal parameters have shown that MCs of length $N_{mc} = 3000$ were required to ensure convergence for all image sizes under study with MwG, but only $N_{mc} = 1000$ with TGS ($N_{bi} = N_{mc} - 700$ for both). This reduction is directly reflected in the computational cost of TSG, which is reported in Fig. 2.12 and discussed below.

b) Computational cost

Fig. 2.12 investigates the computational time T for estimation procedures LF, MwG and TSG for different image sizes N . The computational time includes the 2D DWT for all methods. The

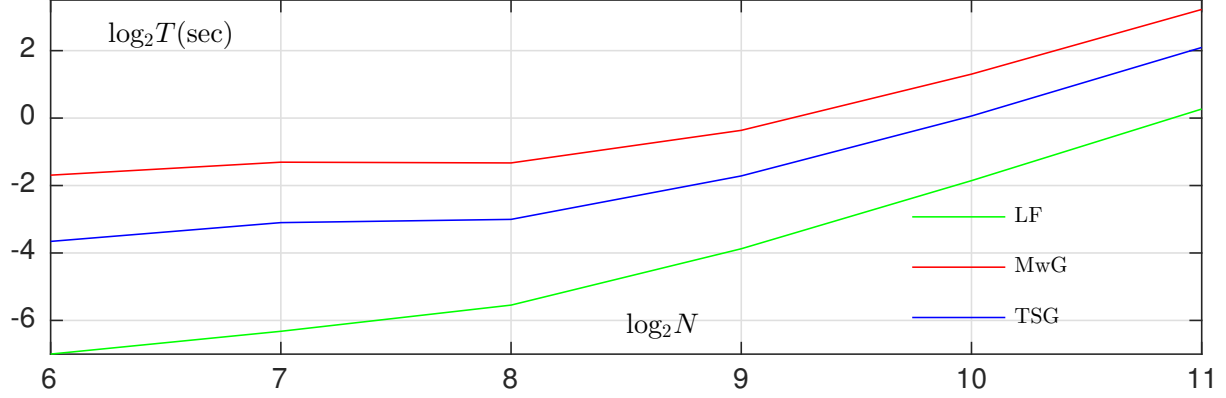


Figure 2.12: Computational time T (in seconds) versus image size N with $j_1 = \{1, 1, 2, 2, 2\}$ and $j_2 = \log_2 N - 4$ for all methods

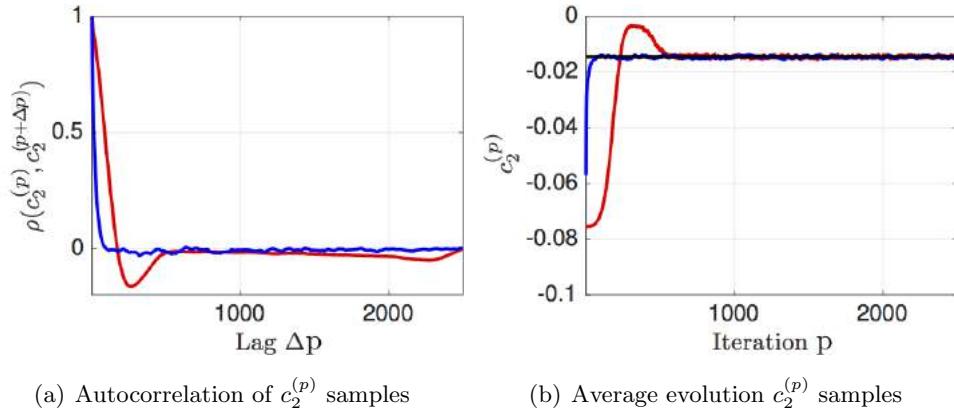


Figure 2.13: Convergence diagnosis of MCMC algorithms assessed on 100 realizations of 2D MRW ($N = 2^8$ and $[c_1, c_2] = [0.7, -0.01]$): (a) autocorrelation of $c_2^{(p)}$ samples; (b) average evolution of $c_2^{(p)}$ samples with in black the mean of the stationary distribution.

estimator LF unsurprisingly exhibits the lowest computational cost. Among the MCMC methods, TSG yields a significant reduction of the computational time as compared to MwG, by a factor ranging from 4 (small images) to 2 (large images). As a result, while the computational time for MwG is up to 25 times greater than that of LF, TSG is only between 5 (small N) and 2 (large N) times slower than the linear regression based estimation LF, which clearly demonstrates the efficiency of the proposed extended Fourier domain model and associated algorithm.

Overall, these results illustrate that the Bayesian approach using the TSG algorithm is an operational alternative to linear regression for the analysis of univariate images, significantly improving estimation performance at only ~ 2 -5 times the computational cost.

2.7 Conclusions and perspectives

Conclusions. We presented in this chapter a Bayesian estimation procedure for the two leading log-cumulants c_1 and c_2 of univariate images. The procedure relied on the formulation of a simple yet generic parametric statistical model for the log-leaders of images. This model was empirically shown to be generically valid for MMC processes. Building on the proposed statistical model and its numerical approximation via the Whittle approximation, this chapter introduced two alternative Bayesian models. The associated Bayesian estimators were approximated by means of samples generated by appropriate Markov chain Monte algorithms. The benefits of the Bayesian approach were numerically assessed via the investigation of estimation performance, assessed using a large number of multifractal processes for several image sizes. While no improvement was observed for the estimation of c_1 , this procedure was found to yield for the estimation of c_2 significant improvements in RMSE of up to a factor of 4 for MMC processes, and up to a factor of 10 for fBms when compared to the current benchmark estimator.

The proposed procedure constitutes, to the best of our knowledge, the first operational Bayesian estimator enabling the reliable estimation of the multifractality parameter for images, even for images (or image patches) of size equal to 64×64 pixels.

Finally, we mention that the evaluation of a Gaussian likelihood via a Whittle approximation is not limited to images and can be exploited for time series as well. Details on the Whittle approximation for time series can be found in Appendix D. Consequently, all developments formulated in this chapter (reparametrization, data augmentation and estimation procedures) can be transposed for time series along with multivariate extensions proposed in Chapter 3. This has been partly reported in [CWA⁺15] for univariate time series (yielding a faster estimation procedure than the one initially proposed in [WDTA13]) and in [CWT⁺16b, CWA⁺16a] for multivariate time series.

Perspectives. The Bayesian methodology enables the estimation of confidence intervals, which can be used to summarize the posterior distribution of a parameter. A direct application of the work presented in this chapter would be the development of routines for the construction of credibility intervals and hypothesis tests for c_1 and c_2 (see, e.g., [EC03, CS99]). More challenging, the incorporation of higher order log-cumulants c_p , $p \geq 3$, would constitute an important continuation of this work. Leads to explore for this purpose are further discussed in the perspectives concluding this thesis in Chapter 5. Finally, it would be of great practical interest to develop suitable algorithms enabling parameter inference to be achieved at even lower computational cost. This will also be discussed in more details in the concluding Chapter 5.

2.8 Conclusions et perspectives (in French)

Conclusions. Nous avons présenté dans ce chapitre une procédure bayésienne pour l'estimation des deux premiers log-cumulants c_1 et c_2 d'images univariées. Cette procédure reposait sur la formulation d'un modèle paramétrique simple et générique pour le logarithme des coefficients dominants des images. Il a été montré empiriquement que ce modèle était valide pour les processus multifractaux construits sur des cascades multiplicatives. À partir de ce modèle, et de son approximation numérique découlant de l'approximation de Whittle, nous avons introduit dans ce chapitre deux modèles bayésiens alternatifs. Les estimateurs bayésiens associés ont été approximés grâce à des méthodes de Monte Carlo par chaînes de Markov appropriées. Les performances d'estimation de notre approche bayésienne ont été évaluées numériquement grâce à des simulations conduites sur plusieurs processus multifractaux de référence. Bien qu'aucune amélioration n'ait été notée pour l'estimation du premier log-cumulant c_1 , il a été montré que notre procédure améliorait significativement l'estimation du deuxième log-cumulant c_2 . Il a notamment été observé que la racine de l'erreur quadratique moyenne de notre estimateur était diminuée d'un facteur 4 pour les processus multifractaux construits sur des cascades multiplicatives et d'un facteur 10 pour le mouvement brownien fractionnaire par rapport à l'estimateur de référence.

La procédure d'estimation proposée dans ce chapitre constitue, à notre connaissance, la première

procédure d'estimation bayésienne opérationnelle permettant l'estimation pertinente du second log-cumulant c_2 pour les images, applicables à des petites tailles images allant jusqu'à 64×64 pixels.

Enfin, nous soulignons ici que l'évaluation de la vraisemblance gaussienne via une approximation de Whittle n'est pas spécifique aux images et peut être également exploitée pour les séries temporelles. Des détails concernant l'approximation de Whittle pour les séries temporelles sont données en Appendice D. Par conséquent, les développements formulés dans ce chapitre (la reparamétrisation, l'augmentation des données et les procédures d'estimation), ainsi que ceux formulés dans le Chapitre 3, peuvent directement être transposés pour l'étude des séries temporelles. Ce point a fait l'objet de plusieurs publications dans [CWA⁺15] pour les séries temporelles univariées et dans [CWT⁺16b, CWA⁺16a] pour les séries temporelles multivariées.

Perspectives. La méthodologie bayésienne permet l'estimation d'intervalles de confiance, qui peuvent être utilisés pour quantifier la loi *a posteriori* d'un paramètre. Une application directe du travail présenté dans ce chapitre serait le développement de procédures de construction d'intervalles de crédibilité et de tests d'hypothèse pour c_1 et c_2 (voir, e.g., [EC03, CS99]). Plus difficile, l'incorporation des log-cumulants d'ordre supérieur c_p , $p \geq 3$, constituerait une suite majeure de ce travail. Des pistes à explorer à cet effet sont discutées plus en détail dans les perspectives concluant cette thèse dans le Chapitre 5. Enfin, il serait d'un grand intérêt pratique de développer des algorithmes d'estimation ayant un coût calculatoire encore plus faible. Ce point est également discuté dans les conclusions du Chapitre 5.

Chapter 3

Bayesian multifractal analysis of multivariate images

Contents

3.1	Introduction	76
3.2	Introduction (in French)	78
3.3	Problem formulation: Multivariate image scenarios	80
3.4	Regularization of v_i via gamma Markov random field priors	83
3.4.1	Bayesian model	83
3.4.2	Estimation algorithm	90
3.4.3	Numerical experiments	91
3.4.4	Preliminary conclusions	100
3.5	Regularization of v_i via simultaneous autoregressive priors	101
3.5.1	Bayesian model	101
3.5.2	Estimation algorithm	105
3.5.3	Numerical experiments	108
3.5.4	Preliminary conclusions	112
3.6	Regularization of w_i via simultaneous autoregressive priors	112
3.6.1	Prior and hyperprior specification	113
3.6.2	Sampling of \mathbf{W} and $\tilde{\epsilon}$	114
3.6.3	Numerical experiments	116
3.6.4	Preliminary conclusions	118
3.7	Conclusions and perspectives	118
3.8	Conclusions et perspectives (in French)	119

3.1 Introduction

As mentioned in the general introduction of this thesis, MFA has been successfully used in various fields of image processing. Yet, its application remains so far conceptually limited to the independent analysis of one single image with homogeneous scale invariance properties at a time. The main reason for this resides in the definition of the multifractal spectrum (1.3) introduced in Chapter 1, which is intrinsically univariate. Although definitions of a multivariate multifractal spectrum have been studied, cf., e.g., [HJK⁺86, SSLB96, MSKF90], these attempts remain essentially limited to pairs of time series and are of little relevance for M -uplets of data for $M \gg 2$. This limitation has become increasingly urgent in view of the number of recent applications in which the acquired images are multivariate, i.e., they consist of a set or sequences of images (multitemporal, multispectral, multimodal, ...) or spatially organized collections of image patches. Such data can provide a rich resource for information, on condition that they are analyzed jointly rather than individually [PMdJF11]. This chapter aims at introducing an operational approach for the multifractal analysis of multivariate images that alleviates this practical limitation. Instead of characterizing the data through a joint multifractal spectrum, yet to be defined and studied theoretically, we propose here to perform a joint estimation of the collection of multifractal parameters associated with the multifractal spectra of each individual data component. More specifically, starting from the univariate statistical model proposed in Chapter 2, we build multivariate Bayesian models jointly describing the collection of multifractal parameters via the design of suitable multivariate prior distributions. The formulation of the joint estimation of multifractal parameters in a Bayesian framework has never been considered before and constitutes an important contribution of this thesis. It provides flexibility for the incorporation of prior information while enabling us to resort to efficient estimation procedures. To illustrate our approach, we consider in this chapter priors encoding the assumption that the multifractal properties evolve slowly across the collection of images (cf. Chapter 5 for a discussion on other assumptions).

The remainder of this chapter is organized as follows. In Section 3.3, we define the multivariate image scenarios considered in this manuscript and the notation used throughout this chapter.

In Section 3.4, we introduce gamma Markov random fields (GMRF), see, e.g., [DC10]) as probabilistic models to describe the collections of parameters \mathbf{v} (related to c_2 via (2.35)). This type of prior induces positive correlation between parameters of different data or image components and hence promotes smoothness. It relies on the use of inverse-gamma distributions for \mathbf{v} which are precisely conjugate when the augmented likelihood (2.58) is used. Therefore, the combination of GMRF priors and (2.58) results in an efficient Bayesian inference algorithm, without acceptance-reject procedure, that is perfectly tailored for the handling of large numbers of unknowns. Within this model, the amount of induced smoothness depends, however, on hyperparameters whose estimation cannot be embedded in a fully Bayesian approach. Those are manually tuned in this chapter, see Chapter 5 for a discussion on methods for their estimation.

As an alternative, we propose in Section 3.5, to model the collections of parameters \mathbf{v} by simultaneous autoregressions (SAR), see, e.g., [Cre15]. In essence, this type of prior promotes a smooth evolution of the multifractal properties by penalizing the magnitude of the second order difference of the evolution of \mathbf{v} . Contrary to GMRF models, the estimation of the associated hyperparameters can be embedded in a fully Bayesian approach, which enables their automatic adjustment. However, since SAR priors do not build on inverse-gamma distributions, the resulting posterior distribution involves non-standard conditional distributions. Thus, parameter inference requires to design appropriate acceptance-reject based MCMC algorithms. We consider in this chapter a Hamiltonian Monte Carlo (HMC) scheme which provides good convergence properties for the high-dimensional variable spaces induced by the analysis of multivariate data. The resulting algorithm has nevertheless a significantly larger computational cost than the one obtained when using GMRF.

Finally, in Section 3.6, we propose to encode the assumption of smoothness for parameters \mathbf{w} (related to c_1 via (2.2)) via the design of SAR priors. The motivations are twofold: First, SAR priors allow the estimation of hyperparameters via a hierarchical Bayesian model; Second, and more importantly, they rely on Gaussian distributions, which are precisely the natural conjugate priors for \mathbf{w} , as stated in Chapter 2. Consequently, the use of SAR priors for the collections of parameters \mathbf{w} leads to an efficient algorithm with no additional acceptance-reject procedure.

The benefits of each of the above multivariate priors are assessed and validated by means of

numerical experiments conducted on multivariate datasets composed of 2D synthetic multifractal processes with prescribed multifractal properties. To the best of our knowledge, the method introduced in this chapter constitutes the first operational MFA tool applicable to the joint analysis of multivariate collections of images, able to jointly or separately perform regularization on c_1 and c_2 .

The work presented in this chapter has been reported in [CWT⁺16a, CWA⁺16c] and a journal paper is in preparation [CWA⁺16d].

3.2 Introduction (in French)

Comme précisé dans l'introduction générale de cette thèse, l'analyse multifractale a été mise à profit dans de nombreux domaines du traitement des images. Cependant, son utilisation reste jusqu'à présent limitée conceptuellement à l'analyse individuelle d'images avec des propriétés d'invariance d'échelle homogènes. La principale raison de cette limitation vient de la définition même du spectre multifractal (1.3) introduit dans le Chapitre 1, qui est intrinsèquement univariée. Bien que des définitions de spectre multifractal multivarié aient été étudiées, cf., e.g., [HJK⁺86, SSLB96, MSKF90], ces tentatives restent essentiellement limitées à des paires de séries temporelles et présentent peu d'intérêt pour l'analyse de M -uplet de données, avec $M \gg 2$. Il devient de plus en plus nécessaire de pallier cette limitation au vu du nombre croissant d'applications pour lesquelles les données récoltées sont multivariées, i.e., elles consistent en un set ou une séquence d'images (multitemporelles, multispectrales, multimodales, ...) ou à une collection de patches spatialement organisés. De telles données sont une riche source d'information, à condition que l'ensemble des données soit traité conjointement plutôt que séparément [PMdJF11]. Ce chapitre vise à introduire une approche opérationnelle pour l'analyse multifractale d'images multivariées. Au lieu de caractériser les données via un spectre multifractal multivarié, qui reste à définir et à étudier théoriquement, nous proposons ici d'effectuer une estimation conjointe de l'ensemble des paramètres multifractaux décrivant les spectres multifractaux de chaque composant du jeu de données. Plus précisément, élaborant sur le modèle statistique univarié introduit dans le Chapitre 2, nous construisons des modèles bayésiens multivariées décrivant l'ensemble des paramètres multifractaux grâce à la définition de lois *a priori*

adéquates. La formulation d’une estimation conjointe des paramètres multifractaux dans un cadre bayésien n’a jamais été considérée auparavant et constitue une importante contribution de cette thèse. Elle autorise une certaine flexibilité dans le choix de l’information *a priori* via la définition des lois *a priori* tout en permettant d’avoir recourt à des procédures d’estimation efficaces. Pour illustrer notre approche, nous considérons dans ce chapitre l’hypothèse que les propriétés des paramètres multifractaux évoluent lentement dans le jeu de données (cf. les perspectives dans le Chapitre 5 pour une discussion sur d’autres hypothèses).

Le reste de chapitre est organisé de la façon suivante. Dans la Section 3.3, nous définissons les scénarios des images multivariées considérés dans ce manuscript et nous introduisons les utilisées tout au long de ce chapitre. Dans la Section 3.4, nous introduisons des champs gamma markoviens (GMRF), voir, e.g., [DC10], comme modèles statistiques pour décrire la collection des paramètres \mathbf{v} (qui, nous le rappelons, sont liés à c_2 via (2.35)). Ce type de loi *a priori* introduit une corrélation positive entre les paramètres de différents données ou images et favorisent une évolution lisse des paramètres. Par ailleurs cette loi *a priori* repose sur des distributions inverse-gamma pour \mathbf{v} qui sont précisément conjuguées lorsque la vraisemblance augmentée (2.58) est utilisée. Par conséquent, la combinaison des modèles GMRF et de (2.35) aboutit à un algorithme d’estimation efficace, ne nécessitant aucune procédure d’acceptation-rejet et ainsi parfaitement adapté à l’inférence sur un grand nombre d’inconnues. Néanmoins, le degré de lissage introduit par ce modèle dépend d’hyperparamètres dont l’estimation ne peut pas être formulée via une approche totalement bayésienne. Ces derniers sont ajustés manuellement dans ce chapitre et nous discutons dans le Chapitre 5 de plusieurs méthodes possibles pour leur estimation.

Une alternative est étudiée dans la Section 3.5, où nous proposons de modéliser les collections des paramètres par des modèles d’auto-regressions simultanées (SAR). En substance, ce type de loi *a priori* favorise une évolution lente des propriétés multifractales en pénalisant l’amplitude des différences du deuxième ordre de l’évolution de \mathbf{v} . Contrairement aux modèles GMRF, l’estimation des hyperparamètres associés aux modèles SAR peut être formulée via une approche purement bayésienne, ce qui permet leur réglage automatique. Cependant, les modèles SAR n’étant pas construits sur des lois inverse-gamma, la loi *a posteriori* associée implique des lois conditionnelles non standard.

Par conséquent, l'inférence requiert des méthodes d'échantillonnage reposant sur des procédures d'acceptation-rejet. Dans ce chapitre nous considérons un schéma de Monte Carlo hamiltonien (HMC), qui jouit d'excellentes propriétés de convergence pour l'échantillonnage sur des espaces de grandes dimensions. L'algorithme résultant est néanmoins associé à un coût calculatoire beaucoup plus élevé que celui résultant de l'utilisation de modèles GMRF.

Enfin, dans la Section 3.6, nous proposons d'incorporer l'hypothèse d'une évolution lisse des paramètres \mathbf{w} (reliés à c_1 via (2.2)) à travers l'utilisation de modèles SAR. Les motivations d'un tel choix sont doubles : Premièrement, les modèles SAR permettent le réglage automatique des hyperparamètres associés; Deuxièmement, et avant tout, ils sont construits sur des lois gaussiennes, qui sont précisément les lois *a priori* conjuguées pour \mathbf{w} , comme précisé dans le Chapitre 2. Par conséquent, l'utilisation de lois *a priori* modélisant les collections des paramètres \mathbf{w} par des modèles SAR aboutit à des algorithmes efficaces, ne nécessitant pas de procédures d'acceptation-rejet supplémentaires.

Les apports de chacun des modèles ci-dessus ont été évalués et validés à partir de simulations numériques conduites sur des jeux de données d'images multivariées composées de processus multifractaux synthétiques dont nous contrôlons les propriétés multifractales. À notre connaissance, la méthodologie introduite dans ce chapitre constitue le premier outil d'analyse multifractale applicable à des images multivariées.

Le travail présenté dans ce chapitre a fait l'objet de plusieurs publications en conférences [CWT⁺16a, CWA⁺16c] et d'un article journal en préparation [CWA⁺16d].

3.3 Problem formulation: Multivariate image scenarios

In this chapter, we consider the multifractal analysis of a dataset \mathbf{X} consisting of either a single image or a sequence of images, and we generically denote by $\{\mathbf{X}_{\mathbf{k}}\}_{\mathbf{k}}$ a partition of \mathbf{X} . Depending on the nature of \mathbf{X} and the structure of the decomposition $\{\mathbf{X}_{\mathbf{k}}\}_{\mathbf{k}}$, different aspects of the evolution of the multifractal properties can be studied, and assumptions on the evolution can be used to perform joint estimation. In this thesis, we investigate in particular the situation where the evolution of the multifractal properties between the elements $\mathbf{X}_{\mathbf{k}}$ is assumed to be smooth. To illustrate our

approach, we consider the three scenarios detailed below.

Temporal/spectral decomposition. For a dataset corresponding to a (temporal/spectral) sequence of images as depicted in Fig. 3.1(a), the evolution of the multifractal properties across images can be investigated. In this case, $\underline{\mathbf{k}}$ reduces to a single index, denoted here $\underline{\mathbf{k}} = k_t \in \Omega_1 = \llbracket 1, M_t \rrbracket$, where M_t is the length of the sequence.

Spatial decomposition. When a single image is under analysis, a spatial decomposition into non-overlapping patches, as illustrated in Fig. 3.1(b), can be considered for the assessment of the local (patch-wise) multifractal properties. In this case, we use the indexation $\underline{\mathbf{k}} = \mathbf{k} = (k_x, k_y) \in \Omega_2$, with $\Omega_2 = \llbracket 1, M_x \rrbracket \times \llbracket 1, M_y \rrbracket$ where M_x and M_y are the sizes of the grid used for the spatial decomposition. This scenario particularly makes senses for the multifractal analysis of images composed of several zones of different textures, for which the analysis needs to be performed locally by means of small patches. Moreover, this situation also comprises collections of images with spatial rather than temporal/spectral organization.

Spatio-temporal/spectral decomposition. Third, given a (temporal or spectral) sequence of images, both approaches above can be combined to yield a spatio-temporal/spectral decomposition (cf. Fig. 3.1(c)), in which case $\underline{\mathbf{k}} = (k_t, \mathbf{k}) \in \Omega_1 \times \Omega_2$. This distinction between the spatial and the temporal/spectral component in the decomposition is meaningful because the dimensions indexed by \mathbf{k} and k_t have different physical roles (since we deal with a sequence of images and not an isotropic 3D data cube).

Notations. We give here the notation used throughout this chapter. We denote by $\ell_{\underline{\mathbf{k}}}$ and $\mathbf{y}_{\underline{\mathbf{k}}} = [\mathbf{y}_{\emptyset, \underline{\mathbf{k}}}^T, \mathbf{y}_{0, \underline{\mathbf{k}}}^T]^T$ the log-leaders and Fourier coefficients associated with the element $\mathbf{X}_{\underline{\mathbf{k}}}$. We moreover denote by $\mathbf{v}_{\underline{\mathbf{k}}} = [v_{1, \underline{\mathbf{k}}}, v_{2, \underline{\mathbf{k}}}]^T$, $\mathbf{w}_{\underline{\mathbf{k}}} = [w_{1, \underline{\mathbf{k}}}, w_{2, \underline{\mathbf{k}}}]^T$, respectively tied to c_2 and c_1 via (2.35) and (2.2), and $\boldsymbol{\lambda}_{\underline{\mathbf{k}}} = \{\mathbf{v}_{\underline{\mathbf{k}}}, \mathbf{w}_{\underline{\mathbf{k}}}\}$ the unknown parameters characterizing the multifractal properties of the element $\mathbf{X}_{\underline{\mathbf{k}}}$. Depending on which Whittle approximation based formulation of the model (either the Whittle approximation (2.37) or the augmented likelihood (2.58)) is being used, we denote the observations

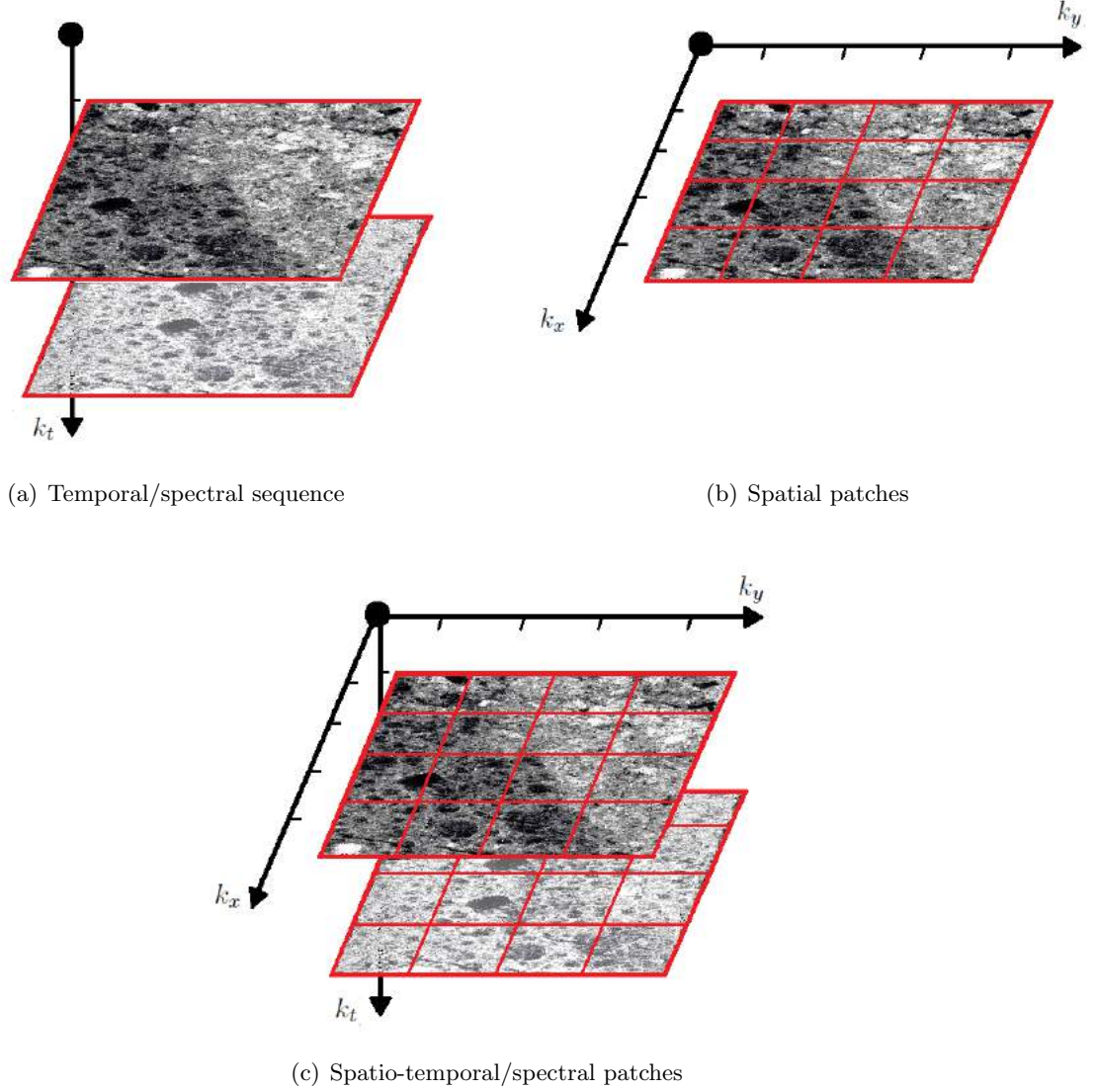


Figure 3.1: Different decompositions $\{\mathbf{X}_{\mathbf{k}}\}_{\mathbf{k}}$ of the dataset \mathbf{X} into a temporal/spectral sequence of images (a), spatial patches (b) and spatio-temporal/spectral patches (c).

by $\mathbf{L} = \{\ell_{\mathbf{k}}\}_{\mathbf{k}}$ or $\mathbf{Y} = \{\mathbf{y}_{\mathbf{k}}\}_{\mathbf{k}}$. When the augmented model is considered, the latent variables $\boldsymbol{\mu}_{\mathbf{k}} = [\boldsymbol{\mu}_{\emptyset, \mathbf{k}}^T, \boldsymbol{\mu}_{0, \mathbf{k}}^T]$ associated with the element $\mathbf{X}_{\mathbf{k}}$ are gathered in $\mathbf{M} = \{\boldsymbol{\mu}_{\mathbf{k}}\}_{\mathbf{k}}$. The parameters to be estimated are stacked in $\boldsymbol{\Lambda} = \{\mathbf{V}, \mathbf{W}\}$ with $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$ and $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2\}$, where the vectors \mathbf{v}_i and \mathbf{w}_i contain $v_{i, \mathbf{k}}$ and $w_{i, \mathbf{k}}$ organized in the lexicographic ordering.

3.4 Regularization of \mathbf{v}_i via gamma Markov random field priors

In this section, we introduce a joint Bayesian model for multivariate images that encodes the assumption of smooth evolution of the multifractal properties via the specification of gamma Markov random field (GMRF) priors for the parameters \mathbf{v}_i . The contribution and the originality of this section lie in the combination of GMRF priors and the augmented likelihood introduced in Chapter 2. These priors involve inverse-gamma distributions and are hence conjugate with respect to the augmented likelihood. This gives rise to an efficient estimation procedure able to handle a large number of data components. Numerical simulations conducted using sequences of synthetic multifractal images demonstrate that the proposed approach significantly outperforms previous univariate benchmark formulations, at competitive computational cost.

3.4.1 Bayesian model

a) Likelihood

As stated in Chapter 2, the augmented likelihood (2.58) has the advantage of allowing conjugate inverse-gamma priors for $v_{i,\mathbf{k}}$ to be used for the construction of Bayesian models. In this section, we propose to consider a multivariate Bayesian model building on (2.58). More precisely, we assume a priori independence between the elements $\mathbf{X}_{\mathbf{k}}$ and we consider the following joint likelihood for the observations \mathbf{Y}

$$p(\mathbf{Y}, \mathbf{M} | \Lambda) \propto \prod_{\mathbf{k}} p(\mathbf{y}_{\mathbf{k}}, \boldsymbol{\mu}_{\mathbf{k}} | \boldsymbol{\lambda}_{\mathbf{k}}) \quad (3.1)$$

with (cf. (2.58))

$$\begin{aligned} p(\mathbf{y}_{\mathbf{k}}, \boldsymbol{\mu}_{\mathbf{k}} | \boldsymbol{\lambda}_{\mathbf{k}}) &\propto v_{1,\mathbf{k}}^{-N_{\phi}} \exp \left(-\frac{1}{v_{1,\mathbf{k}}} (\mathbf{y}_{\phi,\mathbf{k}} - \boldsymbol{\mu}_{\phi,\mathbf{k}})^H \tilde{\mathbf{F}}_1^{-1} (\mathbf{y}_{\phi,\mathbf{k}} - \boldsymbol{\mu}_{\phi,\mathbf{k}}) \right) v_{2,\mathbf{k}}^{-N_{\phi}} \exp \left(-\frac{1}{v_{2,\mathbf{k}}} \boldsymbol{\mu}_{\phi,\mathbf{k}}^H \tilde{\mathbf{F}}_2^{-1} \boldsymbol{\mu}_{\phi,\mathbf{k}} \right) \\ &\times v_{1,\mathbf{k}}^{-\frac{N_0}{2}} \exp \left(-\frac{1}{2v_{1,\mathbf{k}}} (\mathbf{y}_{0,\mathbf{k}} - \boldsymbol{\mu}_{0,\mathbf{k}})^T \tilde{\mathbf{F}}_{0,1}^{-1} (\mathbf{y}_{0,\mathbf{k}} - \boldsymbol{\mu}_{0,\mathbf{k}}) \right) \\ &\times v_{2,\mathbf{k}}^{-\frac{N_0}{2}} \exp \left(-\frac{1}{2v_{2,\mathbf{k}}} (\boldsymbol{\mu}_{0,\mathbf{k}} - \mathbf{X} \mathbf{w}_{\mathbf{k}})^T \tilde{\mathbf{F}}_{0,2}^{-1} (\boldsymbol{\mu}_{0,\mathbf{k}} - \mathbf{X} \mathbf{w}_{\mathbf{k}}) \right). \end{aligned} \quad (3.2)$$

b) Prior distributions

As in Chapter 2, we assume prior independence between parameters characterizing the first and second log-cumulants c_1 and c_2 .

Mean parameter \mathbf{W} . In this section, we focus on \mathbf{V} only and we design a prior on \mathbf{W} assuming no dependence between $\mathbf{w}_{\underline{k}}$. More precisely, we consider the following prior distribution

$$p(\mathbf{W}) = \prod_{\underline{k}} p(\mathbf{w}_{\underline{k}}) \quad (3.3)$$

where $p(\mathbf{w}_{\underline{k}})$ is a non-informative Gaussian distribution $\mathcal{N}(\mathbf{w}_0, \mathbf{\Sigma}_0)$ as in (2.42). The regularization of \mathbf{W} is studied separatably in Section 3.6.

Multifractality parameter \mathbf{V} . The augmented likelihood (3.2) enables, for each element $v_{i,\underline{k}}$, the design of a conjugate inverse-gamma prior $\mathcal{IG}(\alpha_{i,\underline{k}}, \beta_{i,\underline{k}})$. A careful design of the hyperparameters $(\alpha_{i,\underline{k}}, \beta_{i,\underline{k}})$, rather than setting them a priori to constant values (as in Chapter 2), can enforce the parameters of interest \mathbf{v}_i to vary slowly in some privileged directions. To that end, we propose here to specify $(\alpha_{i,\underline{k}}, \beta_{i,\underline{k}})$ such that the resulting prior for \mathbf{v}_i is a hidden gamma Markov random field (GMRF), see [DC10] for details. The GMRF prior is a prior used in Bayesian inference to enforce a smooth evolution of the variances of Gaussian variables with applications, e.g., in source separation and non-negative matrix factorisation [DC10, VCG08] or hyperspectral imaging [APM15].

The underlying motivation for its use lies in the fact that it succeeds in inducing positive dependence between neighboring elements of \mathbf{v}_i [DC10] while preserving an inference procedure as simple as with independent inverse-gamma priors. More specifically, the proposed GMRF defines a joint prior distribution on $(\mathbf{v}_i, \mathbf{z}_i)$ where \mathbf{z}_i is a set of positive latent variables introduced such that, conditionally on \mathbf{z}_i , the variables $v_{i,\underline{k}}$ are independent and distributed according to inverse-gamma distributions of the form

$$v_{i,\underline{k}} | \mathbf{z}_i, \mathbf{a}_i \sim \mathcal{IG}(\alpha_{i,\underline{k}}(\mathbf{a}_i), \beta_{i,\underline{k}}(\mathbf{z}_i, \mathbf{a}_i)) \quad (3.4a)$$

while, conditionally on \mathbf{v}_i , $z_{i,\mathbf{k}}$ are independent gamma variables

$$z_{i,\mathbf{k}}|\mathbf{v}_i, \mathbf{a}_i \sim \mathcal{G}(\tilde{\alpha}_{i,\mathbf{k}}(\mathbf{a}_i), \tilde{\beta}_{i,\mathbf{k}}(\mathbf{v}_i, \mathbf{a}_i)) \quad (3.4b)$$

where $\mathcal{G}(\alpha, \beta)$ stands for the gamma distribution of shape α and rate β . As a result, combining this prior to the likelihood (3.1) leads to standard conditional distributions (inverse-gamma for $v_{i,\mathbf{k}}$ and gamma for $z_{i,\mathbf{k}}$) and enables hence an efficient sampling scheme. In (3.4a) and (3.4b), the parameters $\alpha_{i,\mathbf{k}}$, $\beta_{i,\mathbf{k}}$, $\tilde{\alpha}_{i,\mathbf{k}}$ and $\tilde{\beta}_{i,\mathbf{k}}$ are determined by the definition of a bipartite conditional independence graph between \mathbf{v}_i and \mathbf{z}_i [DC10] and \mathbf{a}_i is a regularization hyperparameter vector. In this section, we handle the temporal/spectral and spatial regularization with separate sets of latent variables $\mathbf{z}_i^{(1)}$ and $\mathbf{z}_i^{(2)}$, respectively. More precisely, for the three multivariate image scenarios described above, \mathbf{z}_i is given by $\mathbf{z}_i^{(1)}$ for temporal/spectral regularization, \mathbf{z}_i is given by $\mathbf{z}_i^{(2)}$ for spatial regularization, while $\mathbf{z}_i = \{\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)}\}$ for spatio-temporal/spectral regularization. In a similar fashion, we consider $\mathbf{a}_i = [a_i^{(1)}, a_i^{(2)}]$ where $a_i^{(1)}$ and $a_i^{(2)}$ adjust the amount of temporal/spectral and spatial smoothness, respectively. The design of the vertices and edges of the bipartite graph between \mathbf{v}_i and \mathbf{z}_i for the three different cases is sketched in Fig. 3.2 and discussed in detail in the following paragraphs.

Temporal or spectral regularization. For a decomposition of \mathbf{X} into a (temporal or spectral) sequence of images $\{\mathbf{X}_{k_t}\}_{k_t \in \Omega_1}$, correlation between the parameters of consecutive images is induced by introducing auxiliary variables $z_{i,k_t}^{(1)} \in \mathbb{R}_*^+$ and a bipartite conditional independence graph between \mathbf{v}_i and $\mathbf{z}_i^{(1)} = \{z_{i,k_t}^{(1)}\}_{k_t \in \Omega_1}$ such that each v_{i,k_t} is connected to the two neighbors of $\mathbf{z}_i^{(1)}$ and vice-versa. The proposed neighborhood structure is illustrated in Fig. 3.2(a). The density of the resulting GMRF prior for $(\mathbf{v}_i, \mathbf{z}_i^{(1)})$ is given by [DC10]

$$p(\mathbf{v}_i, \mathbf{z}_i^{(1)} | a_i^{(1)}) = \frac{1}{C(a_i^{(1)})} \prod_{k_t \in \Omega_1} \Phi_{\mathcal{I}}(v_{i,k_t}; 2a_i^{(1)}) \prod_{k_t \in \Omega_1} \Phi_{\mathcal{G}}(z_{i,k_t}^{(1)}; 2a_i^{(1)}) \times \prod_{k_t \in \Omega_1} \Phi(a_i^{(1)} \sum_{k'_t \in \mathcal{V}_v^{(1)}(k_t)} z_{i,k'_t}^{(1)}, v_{i,k_t}) \quad (3.5)$$

where we have introduced, for convenience of notation, the three potentials

$$\Phi_{\mathcal{I}}(\xi; \alpha) = \exp(-(\alpha + 1) \log \xi) \quad (3.6a)$$

$$\Phi_{\mathcal{G}}(\xi; \alpha) = \exp((\alpha - 1) \log \xi) \quad (3.6b)$$

$$\Phi(\xi, \delta) = \exp(-\xi/\delta) \quad (3.6c)$$

and where $C(a_i^{(1)})$ is a normalizing constant. Eq. (3.5) yields the conditionals (3.4a-3.4b) with parameters

$$\alpha_{i,k_t} = 2a_i^{(1)}, \quad \beta_{i,k_t} = a_i^{(1)} \sum_{k'_t \in \mathcal{V}_v^{(1)}(k_t)} z_{i,k'_t}^{(1)} \quad (3.7a)$$

$$\tilde{\alpha}_{i,k_t}^{(1)} = 2a_i^{(1)}, \quad \tilde{\beta}_{i,k_t}^{(1)} = (a_i^{(1)} \sum_{k'_t \in \mathcal{V}_z^{(1)}(k_t)} v_{i,k'_t}^{-1})^{-1} \quad (3.7b)$$

where the sums are taken over

$$\mathcal{V}_v^{(1)}(k_t) = \{k_t, k_t + 1\} \quad (3.8a)$$

$$\mathcal{V}_z^{(1)}(k_t) = \{k_t - 1, k_t\}. \quad (3.8b)$$

Note that the normalizing constant $C(a_i^{(1)})$ in (3.5) cannot be expressed in closed-form except in very specific cases [DC10].

Spatial regularization. Given a decomposition of a single image into non-overlapping patches $\{\mathbf{X}_{\mathbf{k}}\}_{\mathbf{k} \in \Omega_2}$, spatial regularization is achieved by using a set of auxiliary variables $\mathbf{z}_i^{(2)} = \{z_{i,\mathbf{k}}^{(2)}\}_{\mathbf{k} \in \Omega_2}$ such that each $v_{i,\mathbf{k}}$ is connected to the four variables elements $\{z_{i,\mathbf{k}'}^{(2)}\}_{\mathbf{k}' \in \mathcal{V}_v^{(2)}(\mathbf{k})}$ with

$$\mathcal{V}_v^{(2)}(\mathbf{k}) = \{(k_x, k_y), (k_x + 1, k_y), (k_x, k_y + 1), (k_x + 1, k_y + 1)\} \quad (3.9a)$$

and vice-versa, each $v_{i,\mathbf{k}}$ is connected to $\{z_{i,\mathbf{k}'}^{(2)}\}_{\mathbf{k}' \in \mathcal{V}_z^{(2)}(\mathbf{k})}$ with

$$\mathcal{V}_z^{(2)}(\mathbf{k}) = \{(k_x - 1, k_y - 1), (k_x, k_y - 1), (k_x - 1, k_y), (k_x, k_y)\}. \quad (3.9b)$$

The bipartite conditional independence graph between \mathbf{v}_i and $\mathbf{z}_i^{(2)} = \{z_{i,\mathbf{k}}^{(2)}\}_{\mathbf{k} \in \Omega_2}$ is shown in Fig. 3.2(b) (in red). Following [DC10], it can be shown that the resulting joint GMRF prior for $(\mathbf{v}_i, \mathbf{z}_i^{(2)})$ is

associated with the density

$$p(\mathbf{v}_i, \mathbf{z}_i^{(2)} | a_i^{(2)}) = \frac{1}{C(a_i^{(2)})} \prod_{\mathbf{k} \in \Omega_2} \Phi_{\mathcal{I}}(v_{i,\mathbf{k}}; 4a_i^{(2)}) \times \prod_{\mathbf{k} \in \Omega_2} \Phi_{\mathcal{G}}(z_{i,\mathbf{k}}^{(2)}; 4a_i^{(2)}) \times \prod_{\mathbf{k} \in \Omega_2} \Phi(a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(2)}(\mathbf{k})} z_{i,\mathbf{k}'}^{(2)}, v_{i,\mathbf{k}}) \quad (3.10)$$

where $C(a_i^{(2)})$ is a (generally intractable) normalizing constant. The parameters for the associated conditional distributions (3.4a-3.4b) are given by

$$\alpha_{i,\mathbf{k}} = 4a_i^{(2)}, \quad \beta_{i,\mathbf{k}} = a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(2)}(\mathbf{k})} z_{i,\mathbf{k}'}^{(2)} \quad (3.11a)$$

$$\tilde{\alpha}_{i,\mathbf{k}}^{(2)} = 4a_i^{(2)}, \quad \tilde{\beta}_{i,\mathbf{k}}^{(2)} = (a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_z^{(2)}(\mathbf{k})} v_{i,\mathbf{k}'}^{-1})^{-1}. \quad (3.11b)$$

Spatial and temporal/spectral regularization Finally, by combining the structure of the two GMRFs (3.5) and (3.10), a spatio-temporal/spectral regularization for the data decomposition $\{\mathbf{X}_{k_t,\mathbf{k}}\}_{(k_t,\mathbf{k}) \in \Omega_1 \times \Omega_2}$ can be obtained. Specifically, we make use of the two sets of auxiliary variables $\mathbf{z}_i^{(2)} = \{z_{i,k_t,\mathbf{k}}^{(2)}\}_{(k_t,\mathbf{k}) \in \Omega_1 \times \Omega_2}$ and $\mathbf{z}_i^{(1)} = \{z_{i,k_t,\mathbf{k}}^{(1)}\}_{(k_t,\mathbf{k}) \in \Omega_1 \times \Omega_2}$ (i.e., the proposed neighborhood is a combination of the red and blue graphs shown in Fig. 3.2(c)). The resulting joint prior reads

$$\begin{aligned} p(\mathbf{v}_i, \mathbf{z}_i | \mathbf{a}_i) &= \frac{1}{C(\mathbf{a}_i)} \prod_{k_t \in \Omega_1} \prod_{\mathbf{k} \in \Omega_2} \Phi_{\mathcal{I}}(v_{i,k_t,\mathbf{k}}; 2a_i^{(1)} + 4a_i^{(2)}) \\ &\times \prod_{k_t \in \Omega_1} \prod_{\mathbf{k} \in \Omega_2} \Phi_{\mathcal{G}}(z_{i,k_t,\mathbf{k}}^{(1)}; 2a_i^{(1)}) \times \prod_{k_t \in \Omega_1} \prod_{\mathbf{k} \in \Omega_2} \Phi_{\mathcal{G}}(z_{i,k_t,\mathbf{k}}^{(2)}; 4a_i^{(2)}) \\ &\times \prod_{k_t \in \Omega_1} \prod_{\mathbf{k} \in \Omega_2} \Phi(a_i^{(1)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(1)}(k_t)} z_{i,k_t,\mathbf{k}'}^{(1)} + a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(2)}(\mathbf{k})} z_{i,k_t,\mathbf{k}'}^{(2)}, v_{i,k_t,\mathbf{k}}) \end{aligned} \quad (3.12)$$

where $\mathcal{V}_v^{(d)}(\cdot)$ and $\mathcal{V}_z^{(d)}(\cdot)$, $d = 1, 2$, are defined in (3.8) and (3.9). The parameters of the associated conditionals (3.4a) and (3.4b) can be expressed as

$$\alpha_{i,k_t,\mathbf{k}} = 2a_i^{(1)} + 4a_i^{(2)}, \quad \beta_{i,k_t,\mathbf{k}} = a_i^{(1)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(1)}(k_t)} z_{i,k_t,\mathbf{k}'}^{(1)} + a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_v^{(2)}(\mathbf{k})} z_{i,k_t,\mathbf{k}'}^{(2)} \quad (3.13a)$$

$$\tilde{\alpha}_{i,k_t,\mathbf{k}}^{(1)} = 2a_i^{(1)}, \quad \tilde{\beta}_{i,k_t,\mathbf{k}}^{(1)} = (a_i^{(1)} \sum_{\mathbf{k}' \in \mathcal{V}_z^{(1)}(k_t)} v_{i,k_t,\mathbf{k}'}^{-1})^{-1} \quad (3.13b)$$

$$\tilde{\alpha}_{i,k_t,\mathbf{k}}^{(2)} = 4a_i^{(2)}, \quad \tilde{\beta}_{i,k_t,\mathbf{k}}^{(2)} = (a_i^{(2)} \sum_{\mathbf{k}' \in \mathcal{V}_z^{(2)}(\mathbf{k})} v_{i,k_t,\mathbf{k}'}^{-1})^{-1}. \quad (3.13c)$$

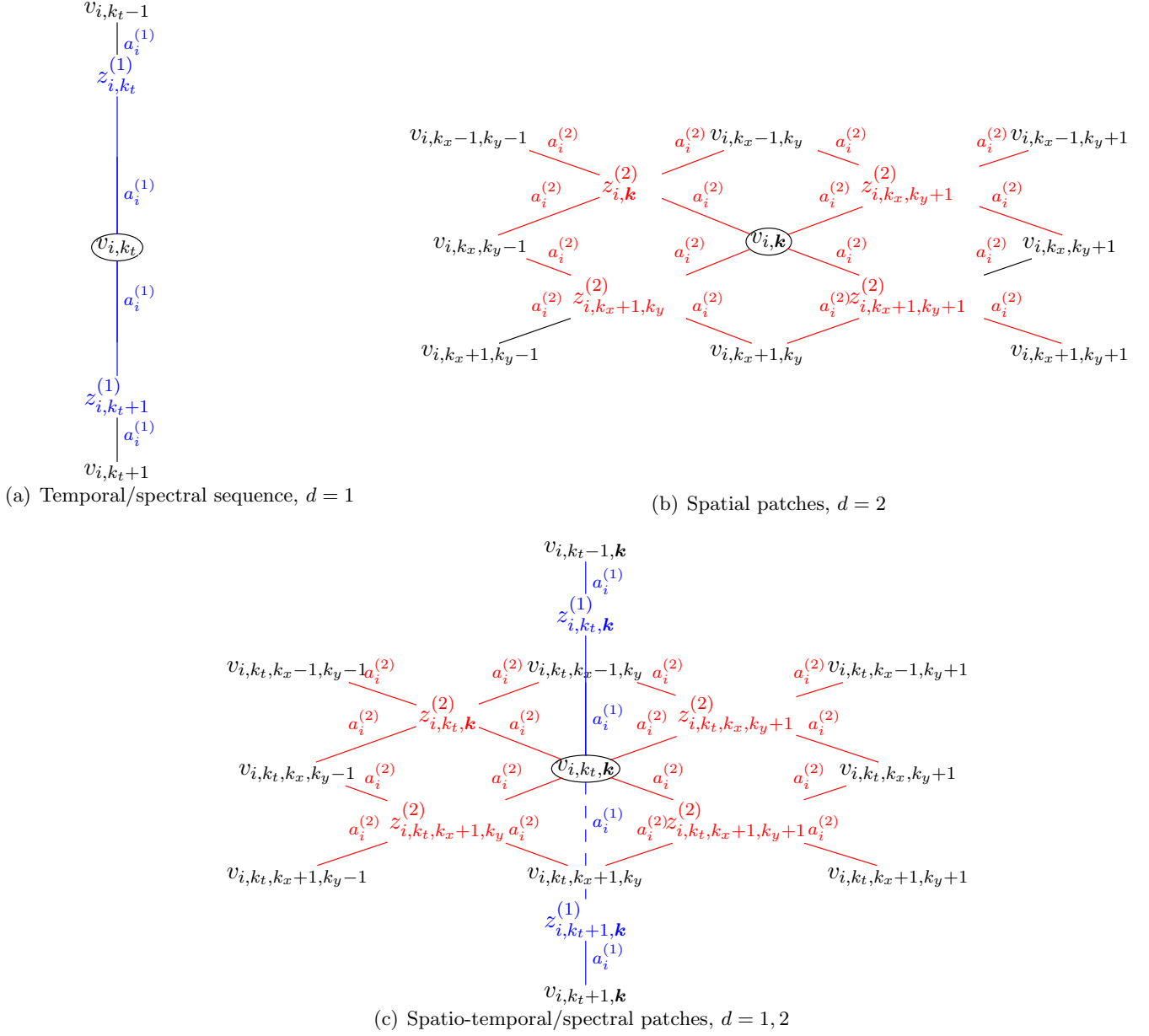


Figure 3.2: Temporal/spectral (blue) and spatial (red) components of the proposed bipartite conditional independence graphs between \mathbf{v}_i and $\mathbf{z}_i^{(d)}$ for the three different scenarios.

c) Posterior distribution and Bayesian estimators

Using Bayes' theorem and assuming prior independence between \mathbf{v}_1 , \mathbf{v}_2 , and \mathbf{W} , the joint posterior distribution associated with the proposed Bayesian model is

$$\begin{aligned} p(\mathbf{\Lambda}, \mathbf{Z}, \mathbf{M} | \mathbf{Y}, \mathbf{a}) &\propto p(\mathbf{Y}, \mathbf{M} | \mathbf{\Lambda}) p(\mathbf{W}) p(\mathbf{V}, \mathbf{Z} | \mathbf{a}) \\ &\propto p(\mathbf{Y}, \mathbf{M} | \mathbf{\Lambda}) p(\mathbf{W}) \prod_{i=1}^2 p(\mathbf{v}_i, \mathbf{z}_i | \mathbf{a}_i) \end{aligned} \quad (3.14)$$

with $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2)$ and $\mathbf{a} = \{\mathbf{a}_1, \mathbf{a}_2\}$. The posterior distribution (3.14) summarizes our knowledge about the unknowns given the observed data and the prior information assigned to the different model parameters. We consider here the marginal posterior mean estimator for $\mathbf{\Lambda}$, denoted as MMSE for minimum mean square error estimator and defined as

$$\mathbf{\Lambda}^{\text{MMSE}} = \mathbb{E}[\mathbf{\Lambda} | \mathbf{Y}, \mathbf{a}] \quad (3.15)$$

where the expectation is taken with respect to the marginal posterior density $p(\mathbf{\Lambda} | \mathbf{Y}, \mathbf{a})$. The direct computation of (3.15) is obviously intractable as it requires integrating the posterior (3.14) over all other unknown variables. However, (3.15) can be approximated by using an MCMC algorithm [RC05]. In the next section we propose a Gibbs sampler drawing samples $\{\mathbf{\Lambda}^{(p)}, \mathbf{M}^{(p)}, \mathbf{Z}^{(p)}\}_{p=1}^{N_{mc}}$ that are asymptotically distributed according to the targeted joint posterior (3.14), which are used in turn to approximate the marginal posterior mean (3.15) analogous to (2.45).

Finally, note that in the posterior distribution (3.14), the regularization hyperparameters \mathbf{a}_i are fixed a priori and their estimation is not addressed during the inference. Instead, throughout this thesis, the GMRF hyperparameters will be set using a qualitative analysis of preliminary results obtained for a range of values for \mathbf{a}_i (see Section 3.4.3 for an example). The motivation for this manual tuning stems from the intractability of the normalizing constant $C(\mathbf{a}_i)$ in the priors (3.7), (3.11) and (3.13), which prevents from addressing their estimation in a fully Bayesian model. The development of estimation procedures adapted to such situations is a research topic on its own and is not considered in this thesis. This point is discussed in more detail in the perspectives concluding this manuscript in Chapter 5.

3.4.2 Estimation algorithm

As described in Chapter 2, the strategy of the Gibbs sampler consists of successively generating samples from the conditional distributions associated with the targeted posterior distribution. In an attempt to provide a unified presentation of an estimation algorithm valid for the three cases considered in this thesis, we introduce the set ξ defined as follows

$$\xi = \begin{cases} \{1\} & \text{for a temporal/spectral decomposition} \\ \{2\} & \text{for a spatial decomposition} \\ \{1, 2\} & \text{for a spatio-temporal/spectral decomposition.} \end{cases} \quad (3.16)$$

a) Sampling of M

By using (3.14), it can be shown that, conditionally on $\Lambda, Z, Y, \mathbf{a}$, the elements $\mu_{\emptyset, \underline{k}}$ and $\mu_{0, \underline{k}}$ in M are a posteriori independent and can be sampled in a parallel manner. More precisely, as in Chapter 2, the conditional distributions of $\mu_{\emptyset, \underline{k}}$ and $\mu_{0, \underline{k}}$ are respectively a complex-valued Gaussian distribution and a real-valued Gaussian distribution

$$\mu_{\emptyset, \underline{k}} | \Lambda, Z, Y, \mathbf{a} \sim \mathcal{CN}(\tilde{x}_{v_{\underline{k}}}, \tilde{\Sigma}_{v_{\underline{k}}}) \quad (3.17a)$$

$$\mu_{0, \underline{k}} | \Lambda, Z, Y, \mathbf{a} \sim \mathcal{N}(\tilde{x}_{0, v_{\underline{k}}, \mathbf{w}_{\underline{k}}}, \tilde{\Sigma}_{0, v_{\underline{k}}}) \quad (3.17b)$$

where $\tilde{x}_{v_{\underline{k}}}$, $\tilde{\Sigma}_{v_{\underline{k}}}$, $\tilde{x}_{0, v_{\underline{k}}}$ and $\tilde{\Sigma}_{0, v_{\underline{k}}}$ follow the notation (2.69), (2.70), (2.73) and (2.74) introduced in Chapter 2.

b) Sampling of W

The sampling of W is also straightforward since all variables $\mathbf{w}_{\underline{k}}$ are conditionally independent of each other and have the following conditional Gaussian distribution

$$\mathbf{w}_{\underline{k}} | V, Z, M, Y, \mathbf{a}_i \sim \mathcal{N} \left(\left(\frac{X^T \tilde{\mathbf{F}}_{0,2}^{-1} X}{v_{2, \underline{k}}} + \Sigma_0^{-1} \right)^{-1} \left(\Sigma_0^{-1} \mathbf{w}_0 + \frac{X^T \tilde{\mathbf{F}}_{0,2}^{-1} \mu_{0, \underline{k}}}{v_{2, \underline{k}}} \right), \left(\frac{X^T \tilde{\mathbf{F}}_{0,2}^{-1} X}{v_{2, \underline{k}}} + \Sigma_0^{-1} \right)^{-1} \right). \quad (3.18)$$

c) Sampling of \mathbf{Z}

Since the auxiliary variables in \mathbf{Z} are not involved in the likelihood (3.1), the sampling of $z_{i,\underline{\mathbf{k}}}^{(d)}, d \in \xi$, from its conditional distribution amounts to sampling from (3.4b), i.e.,

$$z_{i,\underline{\mathbf{k}}}^{(d)} | \mathbf{A}, \mathbf{M}, \mathbf{Y}, \mathbf{a} \sim \mathcal{G} \left(\tilde{\alpha}_{i,\underline{\mathbf{k}}}^{(d)}, \tilde{\beta}_{i,\underline{\mathbf{k}}}^{(d)} \right), \quad d \in \xi \quad (3.19)$$

where the definitions of $\tilde{\alpha}_{i,\underline{\mathbf{k}}}^{(d)}$ and $\tilde{\beta}_{i,\underline{\mathbf{k}}}^{(d)}$ were given in Section 3.4.1 b) and ξ is described in (3.16).

d) Sampling of \mathbf{V}

Straightforward computations on (3.14) lead to the conclusion that the elements of \mathbf{V} are a posteriori independent (conditionally on $\mathbf{W}, \mathbf{Z}, \mathbf{M}, \mathbf{Y}$ and \mathbf{a}) and are distributed according to inverse-gamma distributions. More precisely, all elements of \mathbf{V} can be updated simultaneously by drawing samples from the conditional distributions

$$v_{1,\underline{\mathbf{k}}} | \mathbf{W}, \mathbf{Z}, \mathbf{M}, \mathbf{Y}, \mathbf{a} \sim \mathcal{IG} \left(N_{\emptyset} + \frac{N_0}{2} + \alpha_{1,\underline{\mathbf{k}}}, \|\mathbf{y}_{\emptyset,\underline{\mathbf{k}}} - \boldsymbol{\mu}_{\emptyset,\underline{\mathbf{k}}}\|_{\tilde{\mathbf{F}}_1^{-1}} + \frac{1}{2} \|\mathbf{y}_{0,\underline{\mathbf{k}}} - \boldsymbol{\mu}_{0,\underline{\mathbf{k}}}\|_{\tilde{\mathbf{F}}_{0,1}^{-1}} + \beta_{1,\underline{\mathbf{k}}} \right) \quad (3.20a)$$

$$v_{2,\underline{\mathbf{k}}} | \mathbf{W}, \mathbf{Z}, \mathbf{M}, \mathbf{Y}, \mathbf{a} \sim \mathcal{IG} \left(N_{\emptyset} + \frac{N_0}{2} + \alpha_{2,\underline{\mathbf{k}}}, \|\boldsymbol{\mu}_{\emptyset,\underline{\mathbf{k}}}\|_{\tilde{\mathbf{F}}_2^{-1}} + \frac{1}{2} \|\boldsymbol{\mu}_{0,\underline{\mathbf{k}}} - \mathbf{X}\mathbf{w}_{\underline{\mathbf{k}}}\|_{\tilde{\mathbf{F}}_{0,2}^{-1}} + \beta_{2,\underline{\mathbf{k}}} \right) \quad (3.20b)$$

with $\alpha_{i,\underline{\mathbf{k}}}$ and $\beta_{i,\underline{\mathbf{k}}}$ defined in Section 3.4.1 b).

As a result, all conditional distributions in the Gibbs sampler above are standard and thus can be sampled efficiently, without any acceptance-reject step. We emphasize here that this overall simplicity was the underlying motivation for the design and the combination of the augmented likelihood (2.58) and the GMRF prior. It leads to an inference procedure perfectly tailored for the estimation of unknown parameter vectors of high dimensionality. The corresponding sampling scheme is summarized in Algo. 3.

3.4.3 Numerical experiments

In this section, we compare the multivariate Bayesian approach described in Sections 3.4.1 and 3.4.2 (generically referred to as GMRF below) to an independent application of the univariate Bayesian

approach introduced in Chapter 2 for individual images (denoted below as IG) and to the weighted linear regression based method (1.28) (denoted below as LF_w). The comparison is performed by applying the different estimation methods to a large number of independent realizations of sequences of heterogeneous synthetic multifractal images. To distinguish between the different types of GMRF priors for the approach proposed in this section, we use the notation GMRF 1D, GMRF 2D and GMRF 2D+1D for the temporal, spatial and spatio-temporal/spectral priors detailed in Section 3.4.1 b), respectively.

a) Experimental setup

Scenario. The scenario considered here is summarized in Fig. 3.3. Each realization of the synthetic dataset consists of a sequence of 50 independent 2D multifractal random walks (MRWs) of size 3200×3200 , obtained using a synthesis procedure coded by ourselves. Each 2D MRW in the sequence,

Algorithm 3 Gibbs sampler to sample according to (3.14)

- 1: Initialization
 - 2: Draw $v_{i,\underline{k}}^{(0)} \sim \mathcal{IG}(\alpha_i, \beta_i) \forall \underline{k}$ for $i = 1, 2$ and draw all $\mathbf{w}_{\underline{k}}^{(0)} \sim \mathcal{N}(\mathbf{w}_0, \Sigma_0)$
 - 3: MCMC iterations
 - 4: **for** $p = 1 : N_{mc}$
 - 5: Sample \mathbf{M}
 - 6: Draw $\boldsymbol{\mu}_{\emptyset, \underline{k}}^{(p)} \forall \underline{k}$ according to the complex-valued normal law (3.17a)
 - 7: Draw $\boldsymbol{\mu}_{0, \underline{k}}^{(p)} \forall \underline{k}$ according to the real-valued normal law (3.17b)
 - 8: Sample \mathbf{W}
 - 9: Draw $\mathbf{w}_{\underline{k}}^{(p)} \forall \underline{k}$ according to the normal law (3.18)
 - 10: Sample \mathbf{Z}
 - 11: Draw $z_{i, \underline{k}}^{(d)(p)} \forall \underline{k}, d \in \xi$, according to the gamma law (3.19)
 - 12: Sample \mathbf{V}
 - 13: Draw $v_{1, \underline{k}}^{(p)} \forall \underline{k}$ according to the inverse-gamma law (3.20a)
 - 14: Draw $v_{2, \underline{k}}^{(p)} \forall \underline{k}$ according to the inverse-gamma law (3.20b)
 - 15: **end for**
 - 16: Return $\{\mathbf{\Lambda}^{(p)}, \mathbf{M}^{(p)}, \mathbf{Z}^{(p)}\}_{p=1}^{N_{mc}}$
-

indexed by k_t , has two distinct multifractal regions whose geometries have been fixed for all k_t according to Fig. 3.3(c). It comprises a background with $c_2 = -0.02$ constant throughout the sequence, in which is included an ellipse for which c_2 evolves with k_t according to the piece-wise constant profile plotted in Fig. 3.3(c). An example of a realization of such a heterogeneous 2D MRW (corresponding to frame $k_t = 30$) is displayed in Fig. 3.3(b). Note that the piece-wise constant evolution of c_2 (in space and along k_t) is intentionally chosen here as a limit test case for the robustness of the proposed approach (which is based on the assumption of smooth evolution of c_2 in the data). Multifractal analysis of this dataset is achieved by decomposing each frame into 50×50 non-overlapping patches of size 64×64 pixels, resulting in an overall decomposition into $50 \times 50 \times 50$ patches indexed by (k_t, \mathbf{k}) with $k_t \in \llbracket 1, 50 \rrbracket$ and $\mathbf{k} \in \llbracket 1, 50 \rrbracket^2$. Finally, note that although in this scenario a GMRF prior taking into account both spatial correlations and correlation along k_t (i.e., GMRF 2D+1D) appears most appropriate, we also include the Bayesian estimators GMRF 1D and GMRF 2D accounting for correlation along k_t and spatial correlation only in the comparison.

Estimation setting. For all methods, we used scales $\{j_1, j_2\} = \{1, 2\}$ for each 64×64 patch. MCMC algorithms were run with $N_{mc} = 1000$ and $N_{bi} = 300$. The hyperparameters of the independent \mathcal{IG} priors associated with the method IG were set to $(\alpha_{0,i}, \beta_{0,i}) = (10^{-3}, 10^{-3})$ to mimic a non-informative Jeffreys' prior as in Chapter 2. The GMRF hyperparameters were set to $a_i^{(2)} = 10$ and $a_i^{(1)} = 20$ based on preliminary results discussed in the next paragraph.

b) Influence of GMRF hyperparameters

Before investigating in detail the estimation performance of the different methods, we discuss preliminary results that are used to set the GMRF hyperparameters. They consist here of visual comparisons of estimates obtained with different values of $a_i^{(2)}$ (resp. $a_i^{(1)}$) on spatial patches (resp. sequences of multi-temporal images).

Fig. 3.4 compares IG estimates (b) and GMRF 2D estimates (c) for different values of $a_i^{(2)}$ obtained on a single realization of a heterogeneous 2D MRW with prescribed multifractal properties as defined for frame $k_t = 10$ (a) and decomposed into spatial patches. As expected, a small value

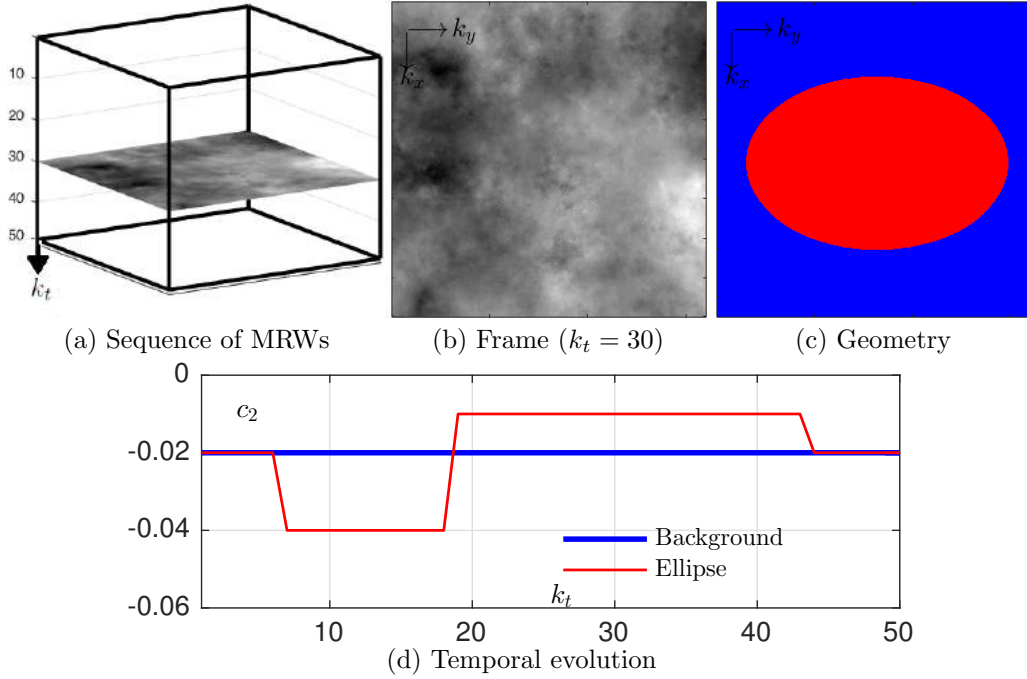


Figure 3.3: Synthetic dataset: (a) sequence of heterogeneous 2D MRWs (frame 30 out of 50 frames plotted), (b) single realization of heterogeneous 2D MRW (frame $k_t = 30$); (c) prescribed spatial geometry and (d) evolution of c_2 with k_t .

of $a_i^{(2)}$ results in estimates c_2 close to those obtained with IG, while a large value of $a_i^{(2)}$ leads to an oversmoothed estimate. The settings $a_i^{(2)} = 10$ and $a_i^{(2)} = 30$ are found to yield the visually most satisfactory estimates for c_2 . We moreover summarize in Tab. 3.1 the estimation performance assessed on 100 independent realizations of heterogeneous MRWs. Note however that these results are not an attempt to optimize $a_i^{(2)}$ in terms of a quantitative performance criterion but only aim at further illustrating the influence of the hyperparameters. Two conclusions can be drawn from this table: First, the parameter $a_i^{(2)}$ can unsurprisingly be interpreted as a tuning parameter for a bias-variance tradeoff; Second, an acceptable tradeoff is achieved for a relatively large range of values of $a_i^{(2)}$ (here between 10 and 30). This suggests that GMRF priors are robust to the choice of the hyperparameters, at least in this example, which hence can be set by means of a visual inspection of results obtained for a coarse grid of values. Finally, similar results (not reported here) obtained on a multi-temporal image sequence have yielded the setting $a_i^{(1)} = 20$.

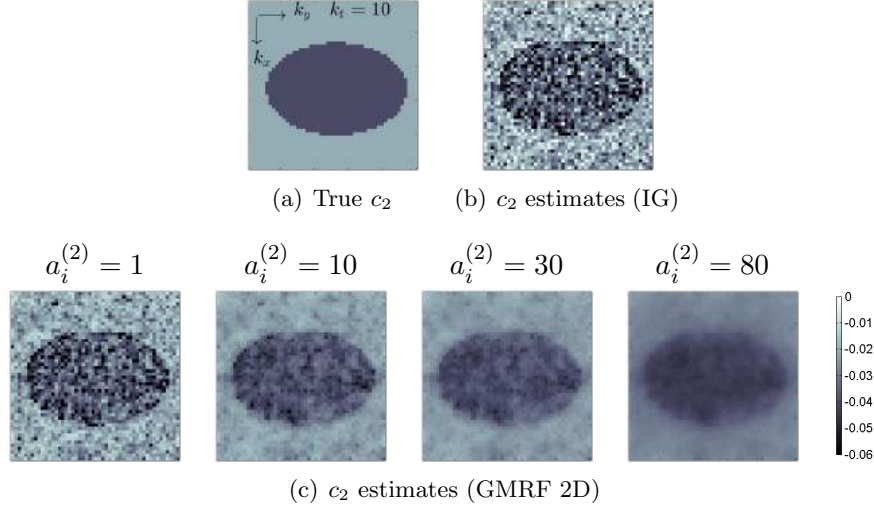


Figure 3.4: Illustration of the influence of the GMRF hyperparameter $a_i^{(2)}$ on the estimation of c_2 ; estimates IG (b) and GMRF 2D (c) obtained on a single realization of heterogeneous 2D MRW reproducing the multifractal properties of the frame $k_t = 10$ (a).

Table 3.1: Influence of the GMRF hyperparameter $a_i^{(2)}$ on the estimation performance for c_2 : absolute values of bias $|m_{c_2} - c_2|$, standard deviations s_{c_2} and RMSE values r_{c_2} ; results obtained for 100 independent realizations of heterogeneous 2D MRWs reproducing the multifractal properties of frame $k_t = 10$. The lower the better and best results are marked in bold.

		GMRF 2D			
	IG	$a_i^{(2)} = 1$	$a_i^{(2)} = 10$	$a_i^{(2)} = 30$	$a_i^{(2)} = 80$
$ m_{c_2} - c_2 $	0.0104	0.0076	0.0041	0.0046	0.0069
s_{c_2}	0.0130	0.0094	0.0037	0.0030	0.0014
r_{c_2}	0.0167	0.0121	0.0056	0.0055	0.0071

c) Illustration for a single realization

We first discuss the estimates for c_2 obtained by the different methods for one single realization of the scenario described in Section 3.4.3 a). These results allow us to appreciate the overall capability of each method to capture the spatial/temporal dynamics of the multifractality parameter c_2 . Fig. 3.5(a)

plots estimates \hat{c}_2 for frame $k_t = 10$ (first row), and for a slice along k_t for $k_y = 25$ (second row) together with the histograms of all estimates (third row) for LF_w, IG, GMRF 1D, GMRF 2D and GMRF 2D+1D (left to right column, respectively). The prescribed values of c_2 are reproduced in (b).

Clearly, LF_w exhibits strong spatial and temporal variability and fails at providing a relevant conjecture for the evolution of the multifractality in the dataset. The Bayesian estimator IG with non-informative prior improves the estimation accuracy with respect to LF_w and enables the visual identification (in time and space) of the zones of different multifractalities, confirming results reported in Chapter 2. Yet, estimates obtained with IG still display strong variability. In particular, their histogram does not permit any conclusion on the existence of three distinct zones of multifractality in the data. In contrast, the three Bayesian estimators GMRF provide more satisfactory results with increased spatial and temporal coherence of the estimates. In particular, the estimates obtained with GMRF 2D and GMRF 2D+1D display strongly reduced variability, leading to histograms in which the three different values for c_2 in the data are reflected as pronounced and well separated peaks.

A more quantitative analysis of these results is proposed in Fig. 3.5(c), which shows the results of a classification of the estimates, obtained by histogram thresholding using the k-means algorithm with 3 classes (the classes have been attributed in order to yield lowest misclassification error). The misclassification rates achieved by the different algorithms are 54% (LF), 45% (IG) and 35% (GMRF 1D) while they are only 8% and 3% for GMRF 2D and GMRF 2D+1D, respectively, which further confirms the above conclusions ¹.

d) Estimation performance

We now assess the estimation performance of the different methods by applying them on 100 independent realizations of the multivariate dataset described in Section 3.4.3 a). For each patch, we compute the average m_{c_2} , the standard deviation (STD) s_{c_2} and the root mean squared error (RMSE) r_{c_2} evaluated over the realizations, as defined in (2.80b). Fig. 3.6 illustrates estimation performance

¹A supplementary 3D animation of the results is available on <http://combrexelle.perso.enseeiht.fr/thesis/c2MRW.zip>.

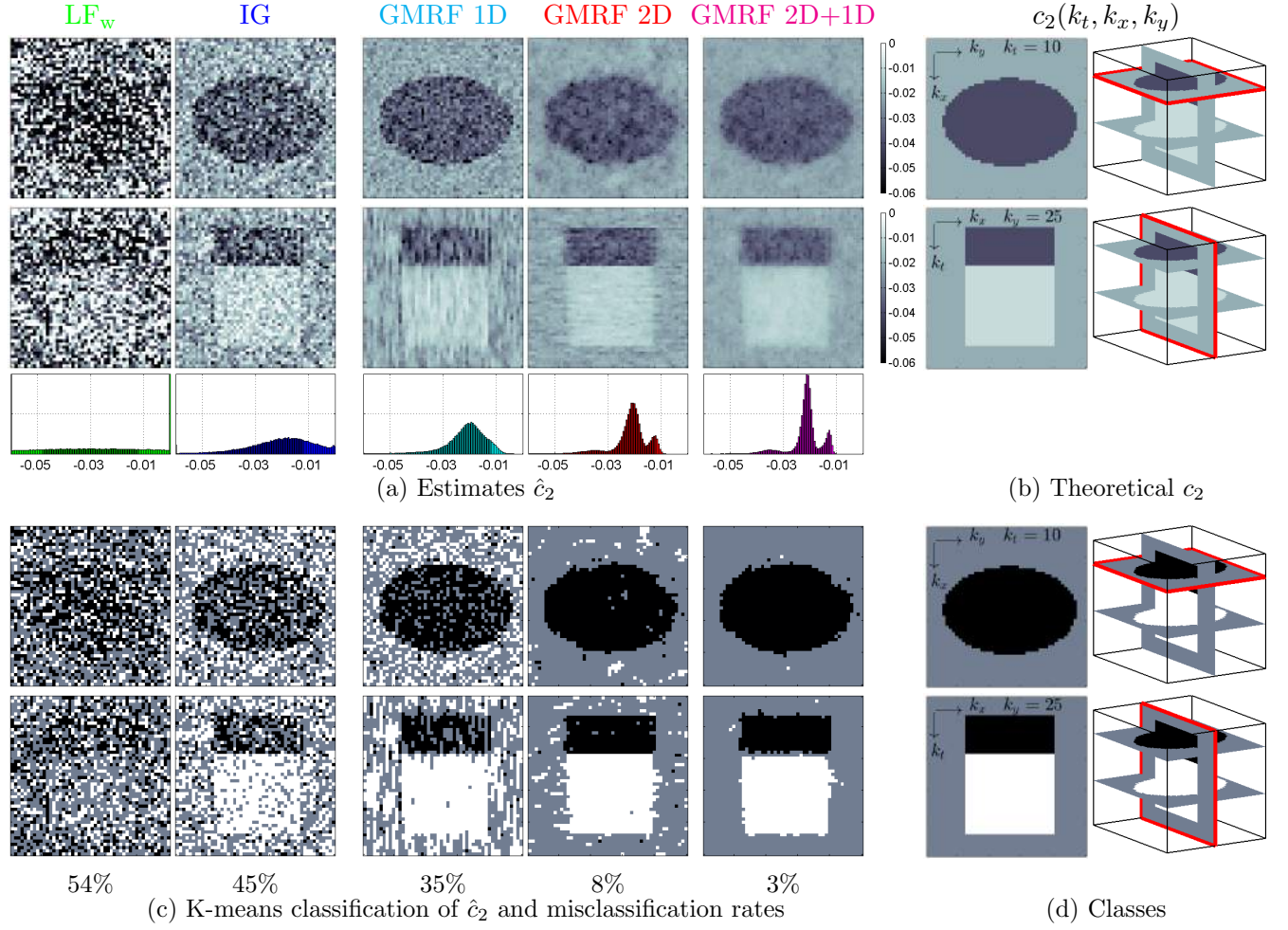


Figure 3.5: Estimation results for a temporal sequence of heterogeneous 2D MRWs decomposed into $50 \times 50 \times 50$ patches of size $2^6 \times 2^6$: estimates \hat{c}_2 for two different slices and global histograms (a) versus prescribed c_2 masks (b). Classification labels obtained by histogram thresholding (c) versus true classes (d).

for the frame $k_t = 10$ and the slice along k_t for $k_y = 25$. To simplify the visual analysis of the results, we moreover report in Fig. 3.7 estimation performance as a function of k_x for $k_t = 10, k_y = 25$ and as a function of k_t for $k_x = 25, k_y = 25$. Finally, the overall performance for the image sequence is given in Tab. 3.2. Only estimation performance for c_2 are reported here since the regularization of c_2 has been observed to have very little effect on the estimation performance for c_1 . This point is

further discussed in Section 3.6.

First, a comparison of the average of estimates for the Bayesian estimators leads to the conclusion that, despite the departure of the scenario considered here from the assumption of slow evolution for c_2 , all GMRF estimators yield average profiles close to that of IG and close to the prescribed c_2 . Only close to sharp transitions for the value of c_2 does GMRF introduce some bias due to the smoothing effect of the prior. Yet, this effect remains confined to ± 3 neighboring patches and has little impact on the overall bias reported in Tab. 3.2. Estimates obtained with LF_w are found to have the largest (by a factor of 3) overall bias.

Second, the Bayesian estimator IG with non-informative prior yields a significant reduction of variability compared to LF_w (STD values are divided by 4, as already reported in Chapter 2). GMRF further and dramatically decreases STD to values that are more than one order of magnitude below those of LF_w . This is also reflected by the overall STD and RMSE values reported in Tab. 3.2, which are more than one order of magnitude better for GMRF than for LF_w . Due to the bias introduced by GMRF close to sharp transitions of the value of c_2 , local RMSE values range from 25% (close to transitions) to only 4% (in homogeneous areas) of those of IG.

Finally, note that this significant gain in performance of GMRF are achieved at very reasonable computational cost. As an example, the analysis of a $1024 \times 1024 \times 50$ data cube using patches of size 64×64 takes about 100s for GMRF on a standard desktop computer², which is only 4 (resp. 1.1) times more than the LF_w (resp. IG) method.

Table 3.2: Absolute values of bias $|m_{c_2} - c_2|$, standard deviations s_{c_2} and RMSE values r_{c_2} for the different estimators (results obtained for 100 independent realizations). The lower the better and best results are marked in bold.

	LF_w	IG	GMRF 1D	GMRF 2D	GMRF 2D+1D
$ m_{c_2} - c_2 $	0.0057	0.0017	0.0022	0.0020	0.0023
s_{c_2}	0.038	0.011	0.0047	0.0027	0.0016
r_{c_2}	0.039	0.011	0.0054	0.0035	0.0029

²For all algorithms, the computational time includes the discrete wavelet transform (as in Chapter 2). Algorithms were run using Matlab, a 3.40 Ghz Intel Core i7 processor and 8GB RAM.

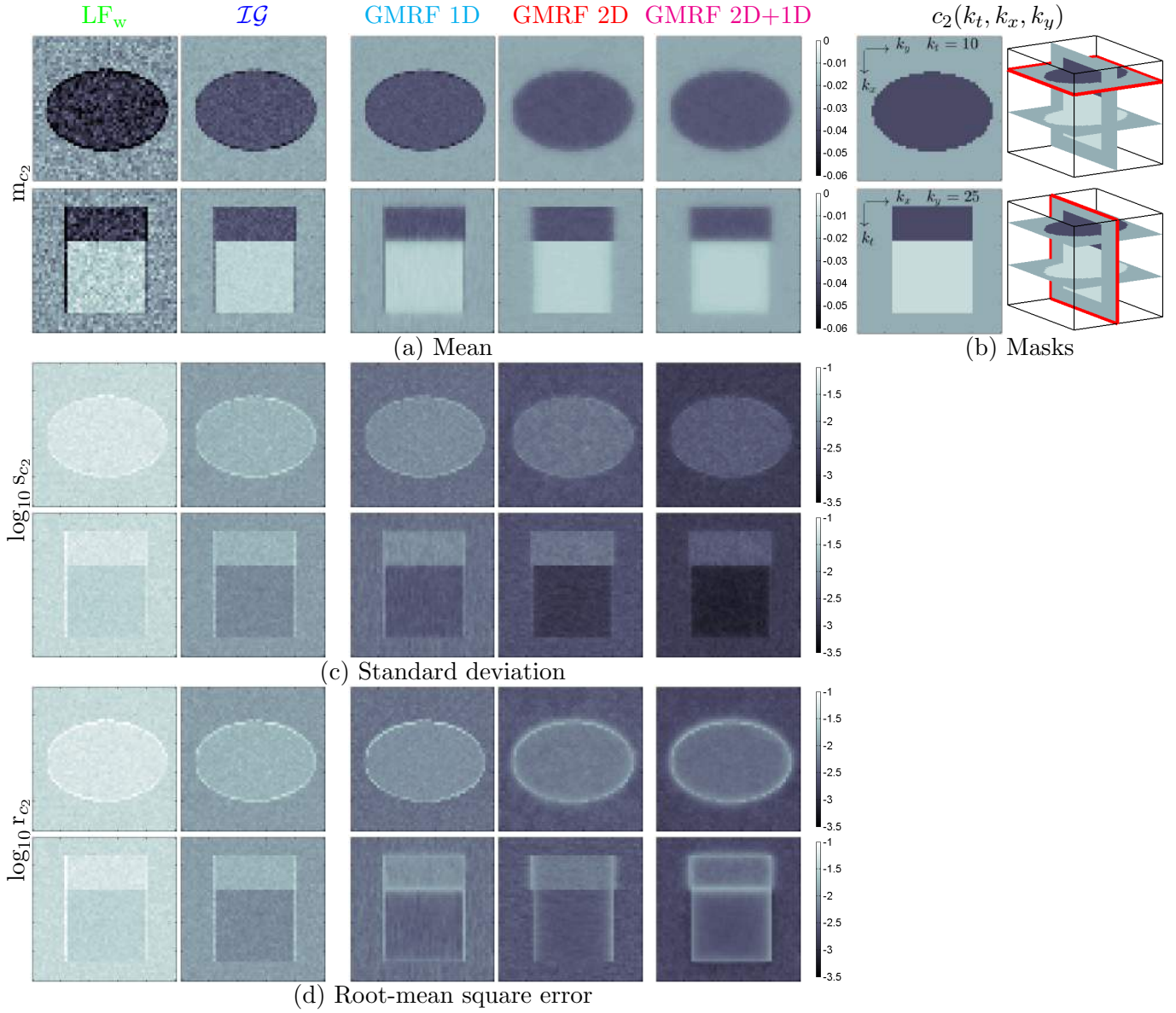


Figure 3.6: Estimation performance evaluated on 100 independent sequences of heterogeneous 2D MRWs decomposed into $50 \times 50 \times 50$ patches of size $2^6 \times 2^6$: average (a), standard deviation (c) and root-mean square error (d) for two different slices and the prescribed c_2 mask (b). The darker, the better for (c) and (d).

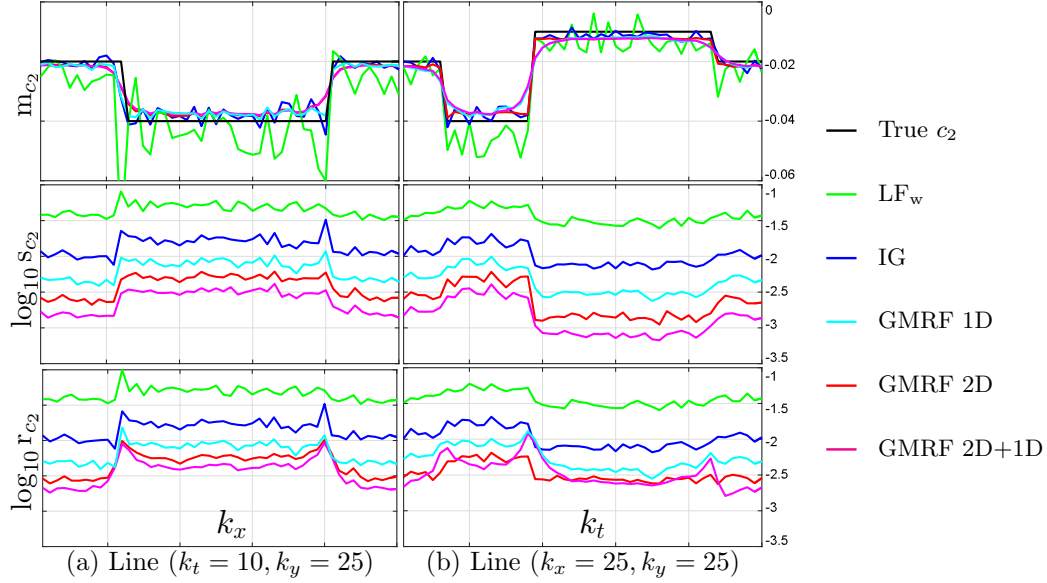


Figure 3.7: Estimation performance evaluated on 100 independent sequences of heterogeneous 2D MRWs decomposed into $50 \times 50 \times 50$ patches of size $2^6 \times 2^6$: average (first row), standard deviation (second row) and root-mean square error (third row) along (a) spatial direction k_x ($k_t = 10, k_y = 25$) and (b) temporal direction k_t ($k_x = 25, k_y = 25$).

3.4.4 Preliminary conclusions

GMRF. This section introduced a novel Bayesian procedure for the joint estimation of the multifractality parameter c_2 for (patches of) multivariate images building on the combination of the statistical model (2.58) introduced in Chapter 2 and smoothing GMRF priors involving inverse-gamma distributions. Together, these two ingredients enabled us to devise an efficient estimation algorithm able to cope with large datasets. Numerical experiments conducted on sequences of synthetic multifractal 2D processes indicate that the proposed multivariate procedure significantly outperforms both the standard benchmark linear regression estimator and the univariate Bayesian formulation of Chapter 2. It notably yields improvements in root-mean-squared error of one order of magnitude as compared to the former and of a factor 4 as compared to the latter, while having a comparable computational cost. Within this model the degree of smoothing is controlled by hyperparameters whose estimation cannot be formulated in a fully Bayesian approach, which constitutes the main

limitation of this model. Here, values were set empirically via a visual inspection of results obtained in preliminary experiments.

Further developments. Although not considered in this thesis for sake of focus, the automatic adjustment of hyperparameters \mathbf{a}_i naturally arises as the next step of this work as it would avoid leaving parameters to be selected by the user and might further improve the estimation performance of the proposed procedure (cf. Chapter 5).

3.5 Regularization of \mathbf{v}_i via simultaneous autoregressive priors

In this section, we introduce an alternative prior for \mathbf{v}_i , which, unlike the GMRF model, enables a data-driven tuning of the hyperparameters controlling the amount of smoothness. The proposed strategy consists of modeling the parameters \mathbf{v}_i by simultaneous autoregressions (SAR), cf., e.g., [Cre15]. This prior enforces smoothness by assigning a Gaussian prior to the second (spatial and/or spectral/temporal) order differences of the multifractal parameters $v_{i,\underline{\mathbf{k}}}$ associated with multivariate images. Within this model, the associated hyperparameters simply act as variances of Gaussian distributions, allowing their estimation to be embedded in a fully Bayesian approach. The performance of the resulting estimation procedure along with its computational cost are analyzed via a numerical experiment on sequences of synthetic multifractal multi-temporal images with prescribed temporal evolutions of its parameters.

3.5.1 Bayesian model

a) Likelihood

Unlike in the previous section, where the data augmented likelihood (2.58) was used, we consider here a multivariate Bayesian model building on the Whittle approximation (2.37) introduced in Chapter 2. The independence between the elements $\mathbf{X}_{\underline{\mathbf{k}}}$ is again assumed, which leads to the following likelihood

$$p(\mathbf{L}|\mathbf{\Lambda}) \propto \prod_{\underline{\mathbf{k}}} p_W(\ell_{\underline{\mathbf{k}}}|\lambda_{\underline{\mathbf{k}}}) \quad (3.21)$$

where $p_W(\ell_{\underline{k}}|\lambda_{\underline{k}})$ is the Whittle approximation (2.37), i.e.,

$$p_W(\ell_{\underline{k}}|\lambda_{\underline{k}}) = \prod_{j=j_1}^{j_2} \prod_{\mathbf{m} \in J_j} (\phi_{j,v_{\underline{k}}}(\omega_{\mathbf{m}}))^{-1} \exp \left(-\frac{y_{\ell_{j,\underline{k}}}^*(\omega_{\mathbf{m}}) y_{\ell_{j,\underline{k}}}(\omega_{\mathbf{m}})}{\phi_{j,v_{\underline{k}}}(\omega_{\mathbf{m}})} \right) \\ \times (\phi_{j,v_{\underline{k}}}(\omega_0))^{-\frac{1}{2}} \exp \left(-\frac{(y_{\ell_{j,\underline{k}}}(\omega_0) - m_j \mu_{j,\mathbf{w}_{\underline{k}}})^2}{2\phi_{j,v_{\underline{k}}}(\omega_0)} \right). \quad (3.22)$$

b) Prior distributions

As in the Bayesian model proposed in the previous section, we assume prior independence between \mathbf{W} and \mathbf{V} and design two independent priors for \mathbf{W} and \mathbf{V} .

Mean parameter \mathbf{W} . We assign to \mathbf{W} the prior (3.3) as in Section 3.5, which assumes no dependence between the different parameters $\mathbf{w}_{\underline{k}}$ (regularization of \mathbf{W} is studied in Section 3.6).

Multifractality parameter \mathbf{V} . In this model, we propose to enforce a smooth evolution of the parameters $\{v_{i,\underline{k}}\}$ by constraining the second order differences of these parameters to be small. To do so, we model the parameters \mathbf{v}_i by a simultaneous autoregression (SAR), see, e.g., [Cre15] for details. This prior model has been considered in various signal and image processing problems, e.g., in image deconvolution [CE07, MMK06], altimetry [HMTS16], hyperspectral unmixing ([SUS14]) or fMRI analysis [CVF⁺13, MIV⁺08]. In essence, this prior favors smooth variations of \mathbf{v}_i and penalizes abrupt changes by assigning a Gaussian distribution to the second order differences of \mathbf{v}_i . In the following paragraphs, we describe the three SAR models respectively enforcing temporal, spatial or spatio-temporal smoothness.

Temporal or spectral regularization. For a decomposition of \mathbf{X} into a sequence of images $\{\mathbf{X}_{k_t}\}_{k_t \in \Omega_1}$, we promote a smooth temporal/spectral evolution of v_{i,k_t} by assigning to the vector \mathbf{v}_i the prior

$$p(\mathbf{v}_i|\epsilon_i^{(1)}) \propto \left(\frac{1}{\epsilon_i^{(1)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(1)T} \mathbf{D}^{(1)})}{2}} \exp \left(-\frac{1}{2\epsilon_i^{(1)}} \|\mathbf{D}^{(1)} \mathbf{v}_i\|^2 \right) \quad (3.23)$$

where $\text{rank}(\cdot)$ denotes the rank of a matrix and $\mathbf{D}^{(1)}$ is the discrete Laplacian operator, i.e.,

$$[\mathbf{D}^{(1)}\mathbf{v}_i]_t = 2v_{i,t} - v_{i,t-1} - v_{i,t+1} \quad (3.24)$$

with $[\cdot]_n$ denoting the n -th element of a vector. Note that in Eq. (3.24), $\epsilon_i^{(1)}$ corresponds to a variance. This choice of parametrization for the Gaussian distribution enables us to avoid confusion between the square and the dimension superscript (d) , $d = 1, 2$, in the rest of this section.

Spatial regularization In a similar fashion, given a decomposition of a single image \mathbf{X} into non-overlapping patches $\{\mathbf{X}_{\mathbf{k}}\}_{\mathbf{k} \in \Omega_2}$, we favor smooth variations in \mathbf{v}_i via the prior

$$p(\mathbf{v}_i | \epsilon_i^{(2)}) \propto \left(\frac{1}{\epsilon_i^{(2)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(2)T} \mathbf{D}^{(2)})}{2}} \exp \left(-\frac{1}{2\epsilon_i^{(2)}} \|\mathbf{D}^{(2)}\mathbf{v}_i\|^2 \right) \quad (3.25)$$

where $\mathbf{D}^{(2)}$ is the discrete Laplacian operator, i.e.,

$$[\mathbf{D}^{(2)}\mathbf{v}_i]_{\tau(\mathbf{k})} = 4v_{i,\mathbf{k}} - v_{i,k_x-1,k_y} - v_{i,k_x+1,k_y} - v_{i,k_x,k_y-1} - v_{i,k_x,k_y+1} \quad (3.26)$$

where $\tau(\cdot)$ is the lexicographic ordering operator.

Spatial and spectral/temporal regularization. Finally, for a spatio-temporal/spectral decomposition $\{\mathbf{X}_{\mathbf{k},k_t}\}_{\mathbf{k},k_t \in \Omega_2 \times \Omega_1}$, we propose to assign a prior enabling a separate handling of temporal/spectral and spatial regularizations as it was the case with the GMRF prior (3.12) introduced in Section 3.4. To do so, we consider the prior defined as follows

$$p(\mathbf{v}_i | \epsilon_i) \propto \prod_{d=1}^2 \left(\frac{1}{\epsilon_i^{(d)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(d)T} \mathbf{D}^{(d)})}{2}} \exp \left(-\frac{1}{2\epsilon_i^{(d)}} \|\mathbf{D}^{(d)}\mathbf{v}_i\|^2 \right) \quad \text{with} \quad \epsilon_i = [\epsilon_i^{(1)}, \epsilon_i^{(2)}]^T \quad (3.27)$$

where $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ are the discrete Laplacian operators such that

$$[\mathbf{D}^{(1)}\mathbf{v}_i]_{\tau(\mathbf{k})} = 2v_{i,\mathbf{k}} - v_{i,k_t-1,k_x,k_y} - v_{i,k_t+1,k_x,k_y} \quad (3.28a)$$

$$[\mathbf{D}^{(2)}\mathbf{v}_i]_{\tau(\mathbf{k})} = 4v_{i,\mathbf{k}} - v_{i,k_t,k_x-1,k_y} - v_{i,k_t,k_x+1,k_y} - v_{i,k_t,k_x,k_y-1} - v_{i,k_t,k_x,k_y+1}. \quad (3.28b)$$

c) Hyperprior distributions.

Within this model, the degree of smoothing depends on the values of the hyperparameters $\epsilon_i^{(d)}$, $i = 1, 2$ and $d \in \xi$, where ξ is described in (3.16). We adopt here a fully Bayesian strategy by including the hyperparameters in the unknown parameter vector and assigning a non-informative Jeffreys' prior for them

$$p(\epsilon_i^{(d)}) \propto \frac{1}{\epsilon_i^{(d)}} \mathbb{1}_{\mathbb{R}^+}(\epsilon_i^{(d)}), \quad d \in \xi \quad (3.29)$$

where $\mathbb{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function of the set \mathbb{R}^+ .

d) Posterior distribution and Bayesian estimators

The joint posterior distribution of the unknown parameters $\mathbf{\Lambda}$ and hyperparameters $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2\}$ is then given by the hierarchical structure $p(\mathbf{\Lambda}, \boldsymbol{\epsilon} | \mathbf{L}) \propto p(\mathbf{L} | \mathbf{\Lambda}) p(\mathbf{W}) p(\mathbf{V} | \boldsymbol{\epsilon}) p(\boldsymbol{\epsilon})$. Assuming prior independence between \mathbf{v}_1 and \mathbf{v}_2 yields

$$p(\mathbf{\Lambda}, \boldsymbol{\epsilon} | \mathbf{L}) \propto p(\mathbf{L} | \mathbf{\Lambda}) p(\mathbf{W}) \prod_{i=1}^2 p(\mathbf{v}_i | \epsilon_i) p(\epsilon_i) \quad (3.30)$$

where

$$p(\epsilon_i) = \prod_{d \in \xi} p(\epsilon_i^{(d)}). \quad (3.31)$$

As in the previous section, we are only interested in the parameters $\mathbf{\Lambda}$ for MFA purposes. We therefore consider here the marginal posterior mean estimator associated with (3.30) defined by

$$\mathbf{\Lambda}^{\text{MMSE}} = \mathbb{E}[\mathbf{\Lambda} | \mathbf{L}] \quad (3.32)$$

where the expectation is taken with respect to the marginal posterior distribution $p(\mathbf{\Lambda} | \mathbf{L})$. Since (3.32) involves integrating over the posterior (3.30), its direct computation is intractable. The inference is performed by using an MCMC algorithm generating the collection of samples $\{\mathbf{\Lambda}^{(p)}, \boldsymbol{\epsilon}^{(p)}\}_{p=1}^{N_{mc}}$ that are asymptotically distributed according to (3.30). These samples are used in turn to approximate the marginal posterior mean estimator (3.32) analogous to (2.45).

3.5.2 Estimation algorithm

We consider again a Gibbs sampler [RC05] and we discuss below the conditional distributions associated with the posterior distribution $p(\mathbf{\Lambda}, \boldsymbol{\epsilon} | \mathbf{L})$. The Gibbs sampling steps are summarized in Algo. 4.

a) Sampling of \mathbf{W}

By inspection of (3.30), it is straightforward to see that, conditionally on $\mathbf{V}, \boldsymbol{\epsilon}, \mathbf{L}$, all $\mathbf{w}_{\mathbf{k}}$ are a posteriori independent and can be sampled in parallel according to the following Gaussian distributions

$$\mathbf{w}_{\mathbf{k}} | \mathbf{V}, \boldsymbol{\epsilon}, \mathbf{L} \sim \mathcal{N}((\mathbf{X}^T \boldsymbol{\Gamma}_{\mathbf{v}_{\mathbf{k}},0}^{-1} \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1} (\boldsymbol{\Sigma}_0^{-1} \mathbf{w}_0 + \mathbf{X}^T \boldsymbol{\Gamma}_{\mathbf{v}_{\mathbf{k}},0}^{-1} \mathbf{y}_{0,\mathbf{k}}), (\mathbf{X}^T \boldsymbol{\Gamma}_{\mathbf{v}_{\mathbf{k}},0}^{-1} \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}). \quad (3.33)$$

b) Constrained Hamiltonian Monte Carlo sampling of \mathbf{V}

It can be shown that the conditional distribution $p(\mathbf{v}_i | \mathbf{L}, \mathbf{v}_{i' \neq i}, \mathbf{W}, \boldsymbol{\epsilon})$ is not standard. Moreover, due to the high dimension of \mathbf{v}_i ($M_x, M_y, M_t \gg 1$), implementing a Metropolis-Hasting procedure (as in Chapter 2) would require a careful design of an appropriate multivariate proposal in order to yield a good exploration of the target distribution [RC05]. Instead, we resort to a constrained Hamiltonian Monte Carlo algorithm (HMC) [DKPR87, Nea10] whose strategy is recalled in the next paragraphs and summarized in Algo. 5.

Hamiltonian system. The HMC algorithm is a sampling scheme inspired by Hamiltonian dynamics [DKPR87, Nea10]. The target distribution, here $p(\mathbf{v}_i | \mathbf{L}, \mathbf{v}_{i' \neq i}, \mathbf{W}, \boldsymbol{\epsilon})$, is associated with a potential energy $E(\mathbf{q})$

$$E(\mathbf{q}) = -\log p(\mathbf{q} | \mathbf{L}, \mathbf{v}_{i' \neq i}, \mathbf{W}, \boldsymbol{\epsilon}) \quad (3.34)$$

with $\mathbf{q} = \mathbf{v}_i$. Moreover auxiliary momentum variables $\mathbf{p} \in \mathbb{R}^M$, with $M = M_x M_y M_t$, are introduced and associated with the following standard kinetic energy $K(\mathbf{p})$

$$K(\mathbf{p}) = \frac{\mathbf{p}^T \mathbf{p}}{2}. \quad (3.35)$$

Finally, the Hamiltonian $H(\mathbf{q}, \mathbf{p})$

$$H(\mathbf{q}, \mathbf{p}) = E(\mathbf{q}) + K(\mathbf{p}) \quad (3.36)$$

defines trajectories $(\mathbf{q}(\tau), \mathbf{p}(\tau))$, in continuous time τ , with constant total energy $H(\mathbf{q}, \mathbf{p})$ via the Hamiltonian system of equations [DKPR87, Nea10]

$$\frac{d\mathbf{q}}{d\tau} = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q}, \mathbf{p}) \quad (3.37a)$$

$$\frac{d\mathbf{p}}{d\tau} = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{q}, \mathbf{p}). \quad (3.37b)$$

Sampling. In an HMC sampling scheme, the proposal of a candidate is achieved through the discrete evaluation of the Hamiltonian equations. More precisely, at the iteration p of the overall Gibbs sampler summarized in Algo 4, given the initial state $\mathbf{q}_0 = \mathbf{v}_i^{(p)}$ and an initial momentum \mathbf{p}_0 , the system of Hamiltonian equations (3.37) is numerically integrated for a short time interval to yield a candidate $(\mathbf{q}^*, \mathbf{p}^*)$, which is then accepted or rejected with the following acceptance rate

$$\alpha = \min(1, \exp[H(\mathbf{q}_0, \mathbf{p}_0) - H(\mathbf{q}^*, \mathbf{p}^*)]). \quad (3.38)$$

Leap-frog method. In this thesis, we consider the leap-frog method [Nea10], which is the numerical integration method classically used in the MCMC literature, to discretize and numerically integrate (3.37) for a time interval $\delta \cdot L$. The standard leap-frog method consists of L iterations, indexed by $n = 1, \dots, L$, each composed of the following three steps

$$\mathbf{p}_{n-1/2} = \mathbf{p}_{n-1} - \frac{\delta}{2} \frac{\partial E}{\partial \mathbf{q}^T}(\mathbf{q}_{n-1}) \quad (3.39a)$$

$$\mathbf{q}_n = \mathbf{q}_{n-1} + \delta \mathbf{p}_{n-1/2} \quad (3.39b)$$

$$\mathbf{p}_n = \mathbf{p}_{n-1/2} - \frac{\delta}{2} \frac{\partial E}{\partial \mathbf{q}^T}(\mathbf{q}_n) \quad (3.39c)$$

where δ is an increment that is tuned during the N_{bi} first iterations of the Gibbs sampler such that $\alpha \in [0.5, 0.8]$ in (3.38) [DKPR87, Nea10]. Here, \mathbf{q} is subject to positivity constraints, since $v_{i,\underline{k}} > 0$. These constraints can be easily taken into account by considering a variation of the leap-frog method (3.39), which is given in Algo. 5 (see, e.g., [BGJM11] for more details). The candidate is finally given by the output of the leap-frog method, i.e., $(\mathbf{q}^*, \mathbf{p}^*) = (\mathbf{q}_L, \mathbf{p}_L)$.

Derivatives. The sampling scheme above requires the computation of the derivatives of the potential energy (3.34). For the model proposed in this work, the derivatives can be calculated

analytically (see Appendix C for details) and are given by the closed-form expressions

$$\begin{aligned} \frac{\partial E(\mathbf{v}_i)}{\partial v_{i,\mathbf{k}}} &= \sum_{j=1}^{j_2} \left(\sum_{\mathbf{m}} \frac{\tilde{f}_{i,j}(\boldsymbol{\omega}_{\mathbf{m}})}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{m}})} \left(1 - \frac{|y_{\ell_{j,\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{m}})|^2}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{m}})} \right) + \frac{\tilde{f}_{i,j}(\boldsymbol{\omega}_{\mathbf{0}})}{2\phi_{j,\mathbf{v}_{\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{0}})} \left(1 - \frac{(y_{\ell_{j,\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{0}}) - m_j \mu_{j,\mathbf{w}_{\mathbf{k}}})^2}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\boldsymbol{\omega}_{\mathbf{0}})} \right) \right) \\ &+ \sum_{d \in \xi} \left[\frac{\mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{v}_i}{\epsilon_i^{(d)}} \right]_{\tau(\mathbf{k})} \end{aligned} \quad (3.40)$$

where the functions $\phi_{j,\mathbf{v}_{\mathbf{k}}}$ and $\tilde{f}_{i,j}$ are given in Eqs. (2.39) and (2.40), respectively, and ξ is the set defined in (3.16).

Algorithm 4 HMC-based Gibbs sampler to sample from (3.30)

- 1: Initialization
 - 2: Draw $v_{i,\mathbf{k}}^{(0)} \sim \mathcal{IG}(\alpha_i, \beta_i) \forall \mathbf{k}$ for $i = 1, 2$ and draw all $\mathbf{w}_{\mathbf{k}}^{(0)} \sim \mathcal{N}(\mathbf{w}_0, \boldsymbol{\Sigma}_0)$
 - 3: MCMC iterations
 - 4: **for** $p = 1 : N_{mc}$
 - 5: Sample parameters $\boldsymbol{\Lambda}$
 - 6: Sample \mathbf{W}
 - 7: Draw all $\mathbf{w}_{\mathbf{k}}^{(p)}$ according to the Gaussian law (3.33)
 - 8: Sample \mathbf{V}
 - 9: Draw $\mathbf{v}_i^{(p)}$ using a constrained HMC procedure (cf. Algo 5)
 - 10: Sample hyperparameters $\boldsymbol{\epsilon}$
 - 11: Draw all $\epsilon_i^{(d)(p)}$, $d \in \xi$, according to the inverse-gamma law (3.41)
 - 12: **end for**
 - 13: Return $\{\boldsymbol{\Lambda}^{(p)}, \boldsymbol{\epsilon}^{(p)}\}_{p=1}^{N_{mc}}$
-

c) Sampling of $\boldsymbol{\epsilon}$

Due to the design of conjugate priors for the hyperparameters $\epsilon_i^{(d)}$, their sampling is straightforward. More precisely, the conditional distributions of the hyperparameters $\epsilon_i^{(d)}$ are the following inverse-gamma distributions (that are easy to sample)

$$\epsilon_i^{(d)} | \boldsymbol{\Lambda}, \mathbf{L} \sim \mathcal{IG} \left(\frac{\text{rank}(\mathbf{D}^{(d)T} \mathbf{D}^{(d)})}{2}, \frac{\|\mathbf{D}^{(d)} \mathbf{v}_i\|^2}{2} \right), \quad d \in \xi. \quad (3.41)$$

Algorithm 5 Constrained HMC procedure

```

1: Set  $\mathbf{q}_0 = \mathbf{v}_i^{(p)}$  and draw  $\mathbf{p}_0 \sim \mathcal{N}(\mathbf{0}_M, \mathbf{I}_M)$  with  $M = M_x M_y M_t$ 
2: Constrained leap-frog method
3: for  $n = 1 : L$ 
4:    $\mathbf{p}_{n-1/2} = \mathbf{p}_{n-1} - \frac{\delta}{2} \frac{\partial E}{\partial \mathbf{q}^T}(\mathbf{q}_{n-1})$ 
5:    $\mathbf{q}_n = \mathbf{q}_{n-1} + \delta \mathbf{p}_{n-1/2}$ 
6:   for  $k = 1 : M$ 
7:     if  $[\mathbf{q}_n]_k < 0$ 
8:       Set  $[\mathbf{q}_n]_k = -[\mathbf{q}_n]_k$  and  $[\mathbf{p}_{n-1/2}]_k = -[\mathbf{p}_{n-1/2}]_k$ 
9:     end if
10:  end for
11:   $\mathbf{p}_n = \mathbf{p}_{n-1/2} - \frac{\delta}{2} \frac{\partial E}{\partial \mathbf{q}^T}(\mathbf{q}_n)$ 
12: end for
13: Set  $(\mathbf{q}^*, \mathbf{p}^*) = (\mathbf{q}_L, \mathbf{p}_L)$  and draw  $u \sim U_{[0,1]}$ 
14: Acceptance-reject
15: Compute  $\alpha = \min(1, \exp[H(\mathbf{q}_0, \mathbf{p}_0) - H(\mathbf{q}^*, \mathbf{p}^*)])$ 
16: Set  $\mathbf{v}_i^{(p+1)} = \mathbf{q}^*$  if  $u < \alpha$ , otherwise  $\mathbf{v}_i^{(p+1)} = \mathbf{v}_i^{(p)}$ 

```

3.5.3 Numerical experiments

We investigate in this section the performance of the proposed multivariate Bayesian approach using SAR priors (denoted below as SAR). It is compared to the two Bayesian counterparts IG and GMRF, respectively introduced in Chapter 2 and in Section 3.4, and to the weighted linear regression estimator LF_w defined in (1.27).

a) Experimental setup

Scenario. The scenario used for this numerical experiment consists of a sequence of $M_t = 100$ multi-temporal images defined as 2D homogeneous MRWs of size $2^7 \times 2^7$ [RV10]. Four different evolutions of $c_2(k_t)$, $k_t = 1, \dots, 100$, across the image sequences are studied, with values for c_2 ranging from -0.02 to -0.08 : a slow sinusoidal profile, a fast sinusoidal profile, a chirp profile including both slow and fast evolutions as well as, as a limit benchmark case violating the slow evolution assumption, a

discontinuous evolution (cf. Fig. 3.8(a), top row, left to right columns, respectively). For all cases, the parameter c_1 is kept constant, $c_1(k_t) = 0.7$. Finally, only GMRF 1D is studied in this numerical experiment.

Estimation setting. For all methods, scales $(j_1, j_2) = (1, 3)$ are used in the analysis. For this experiment, MCMC algorithms were run with $(N_{mc} = 1000, N_{bi} = 300)$ for IG and GMRF and with $(N_{mc} = 1500, N_{bi} = 800)$ for SAR, where we used $L = 20$ steps for the leap-frog method in the HMC scheme. The regularization parameters of GMRF 1D were fixed to $a_i^{(1)} = 50$ for all profiles based on preliminary experiments as described in Section 3.4.3.

b) Estimation performance

Estimation performance is again quantified via the average, the standard deviation (STD) and the root mean squared error (RMSE) defined in (2.80b). Performance results, computed for $MC = 100$ independent realizations, are summarized in Fig. 3.8(a) for the four different temporal evolutions of c_2 and yield the following conclusions.

SAR vs. univariate methods. Estimators LF_w , IG and SAR succeed in reproducing on average the prescribed values of c_2 for the smooth evolutions (i.e., they have small bias). For the discontinuous evolution of c_2 , SAR provides a smooth estimate, as expected, and hence introduces a small bias in the vicinity of the discontinuities. More importantly, it is observed that the proposed joint Bayesian SAR estimator consistently yields a significant reduction of STD values as compared to LF_w (STD divided by up to 8) and to the univariate Bayesian estimator IG (STD divided by up to 3). This gain in STD is directly reflected in RMSE values which are, except at the locations of discontinuities, 70–400% smaller than those of IG and LF_w for SAR. This demonstrates the benefits of the proposed multivariate Bayesian procedure when compared to the univariate methods for the estimation of c_2 .

SAR vs. GMRF. Overall, SAR and GMRF estimators are found to exhibit similar estimation performance in terms of bias and (especially) standard deviation. It is however observed that, close to zones of fast variations of c_2 , SAR introduces a smaller bias than GMRF for low multifractality

($|c_2| \approx 0$). This is also reflected in the RMSE values, e.g., for the fast sinusoidal and chirp profiles. This might result from an oversmoothing since, for this experiment, the GMRF hyperparameters are set to be identical for all profiles, regardless of their dynamics. On the contrary, by enabling an efficient automatic tuning of hyperparameters in a fully Bayesian approach, SAR has the advantage of inducing an amount of smoothness corresponding the dynamics in the dataset (this point is discussed in the next subsection).

However, SAR has one major practical drawback, which is its high computational cost. For instance, the analysis of a sequence of $2^7 \times 2^7$ images of length $M_t = 100$ takes 20s with SAR, while it takes only 1s with LF_w , 4s with IG and 5s with GMRF. We emphasize that this computational cost is not induced by the estimation of the hyperparameters (sampling according to (3.41) is straightforward) but is a consequence of the HMC scheme. This leads us to the conclusion that the multivariate Bayesian model relying on GMRF priors can be preferred in practice for the analysis of large datasets (as it will be the case in Chapter 4 with hyperspectral images).

c) Estimation of hyperparameters

In order to study the effectiveness of the automatic tuning of the hyperparameters $\epsilon_i^{(1)}$, we report in Fig. 3.8(b) histograms of MMSE estimates

$$\epsilon_i^{(1)\text{MMSE}} \approx \frac{1}{N_{mc} - N_{bi}} \sum_{p=N_{bi}+1}^{N_{mc}} \epsilon_i^{(1)(p)} \quad (3.42)$$

for the four different evolutions of c_2 . Since $\epsilon_i^{(1)}$ corresponds to the variance in the prior (3.23), the smaller $\epsilon_i^{(1)}$, the smoother the evolution of estimates.

For the sinusoidal evolutions, the average value of $\epsilon_i^{(1)}$ (indicated by a vertical red line in Fig. 3.8(b)) is $10^{-6.2}$ and $10^{-4.6}$ for slow and fast evolution, respectively, thus reflecting the degree of smoothness in the evolution. For the chirp, the average value is $10^{-4.9}$ and thus slightly below that of the fast sinusoid (indeed, stronger smoothing would introduce bias and would hence be highly penalized). The values of $\epsilon_i^{(1)}$ for the discontinuous case are centered at 10^{-5} and close to the continuous cases, indicating that larger bias at the two discontinuities is traded off for small variability within the segments with constant c_2 . These results provide strong evidence that the model succeeds in adjusting

the hyperparameter to an appropriate smoothing level for the data.

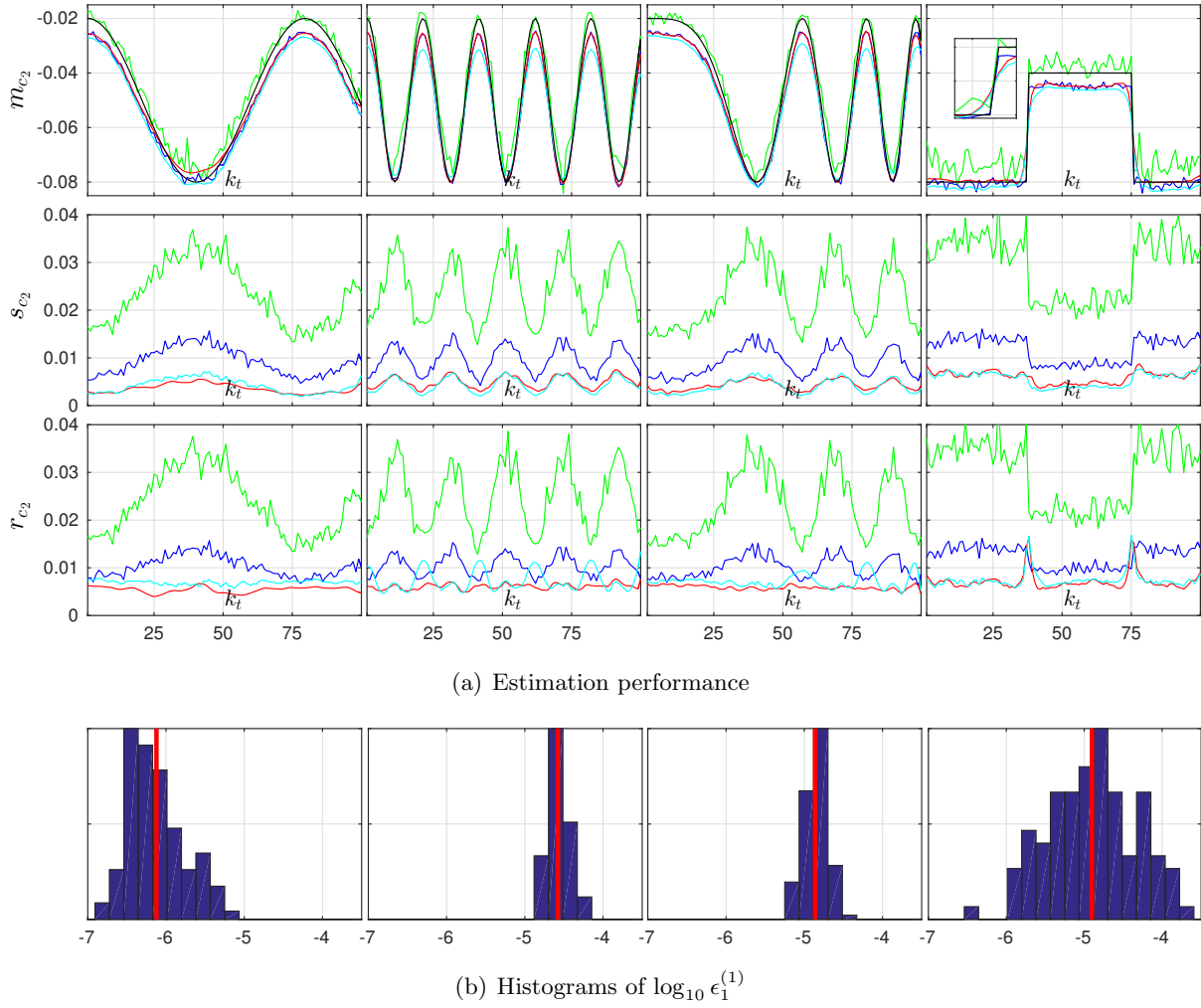


Figure 3.8: Estimation performance for c_2 (from top to bottom: mean, STD and RMSE) assessed on 100 independent realizations of a sequence of multi-temporal images defined as 2D MRWs of size $2^7 \times 2^7$ for four different profiles of c_2 ; True $c_2(k_t)$ (—), LF_w (—), IG (—), GMRF 1D (—) and SAR (—).

3.5.4 Preliminary conclusions

SAR. We introduced in this section an alternative Bayesian procedure addressing the estimation of the multifractality parameter c_2 for multivariate datasets. The smoothness assumption was encoded via a Gaussian prior on the (spatial and/or spectral/temporal) second order differences of the multifractal attributes $v_{i,\underline{k}}$. The hyperparameters controlling the amount of smoothness were embedded in a fully Bayesian model. To bypass the difficulties resulting from non-standard high-dimensional conditional distributions associated with the Bayesian model, a Hamiltonian Monte Carlo scheme was proposed. Numerical experiments conducted on sequences of synthetic multifractal multi-temporal images demonstrated that the proposed procedure yielded significantly improved estimation performance over univariate formulations.

SAR vs GMRF. We also showed that encoding the assumption of smooth evolution of the multifractality with either GMRF, introduced in Section 3.4, or SAR priors on \mathbf{v}_i yielded comparable estimation performance for c_2 . However, GMRF priors can arguably be preferred over SAR priors for computational reasons.

Further developments. The reduction of the computational cost to achieve Bayesian inference on the proposed model constitutes an important practical issue. One interesting lead could be to consider Langevin Monte Carlo (LMC) methods [Nea10, GC11, WT11], which can be seen as a special case of HMC schemes for which the number of iterations in the leap-frog method is reduced to one.

3.6 Regularization of \mathbf{w}_i via simultaneous autoregressive priors

In the two previous sections, we focused on the regularization of the parameter vectors \mathbf{v}_i (associated with the multifractality parameter c_2 via (2.35)), for which we proposed two different priors accounting for the assumption that parameters $v_{i,\underline{k}}$ of neighboring positions in time, space or frequency were correlated. In this last section, we encode such an assumption also for the parameters $w_{i,\underline{k}}$, associated with c_1 , via (2.2). To do so we propose to assign to each \mathbf{w}_i a SAR prior. The motivation for this is twofold: First, as illustrated in the previous section, it allows an efficient data-driven tuning

of the hyperparameters controlling the amount of smoothness; Second, and more importantly, since SAR models involve Gaussian distributions, they are the natural conjugate priors for \mathbf{w}_i . As a result, incorporating this prior in the Bayesian models of Sections 3.4 and 3.5 comes without noteworthy extra computational effort and only requires few changes in the associated Gibbs samplers Algo. 3 and 4 (unlike their use for \mathbf{v}_i , investigated in Section 3.5). Those are made explicit in this section. The performance of the resulting Bayesian estimation procedure for the parameter c_1 of images is finally assessed with Monte Carlo simulations for sequences of synthetic 2D multifractal processes.

3.6.1 Prior and hyperprior specification

We state here the general form of the hierarchical prior resulting from assigning SAR priors to \mathbf{w}_i .

a) Prior distributions

As in Section 3.5, for the temporal/spectral, spatial and spatio-temporal/spectral scenarios depicted in Fig. 3.1, the priors for \mathbf{w}_i are, respectively,

$$p(\mathbf{w}_i | \tilde{\epsilon}_i^{(1)}) \propto \left(\frac{1}{\tilde{\epsilon}_i^{(1)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(1)T} \mathbf{D}^{(1)})}{2}} \exp \left(-\frac{1}{2\tilde{\epsilon}_i^{(1)}} \|\mathbf{D}^{(1)} \mathbf{w}_i\|^2 \right) \quad (3.43a)$$

$$p(\mathbf{w}_i | \tilde{\epsilon}_i^{(2)}) \propto \left(\frac{1}{\tilde{\epsilon}_i^{(2)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(2)T} \mathbf{D}^{(2)})}{2}} \exp \left(-\frac{1}{2\tilde{\epsilon}_i^{(2)}} \|\mathbf{D}^{(2)} \mathbf{w}_i\|^2 \right) \quad (3.43b)$$

$$p(\mathbf{w}_i | \tilde{\epsilon}_i) \propto \prod_{d=1}^2 \left(\frac{1}{\tilde{\epsilon}_i^{(d)}} \right)^{\frac{\text{rank}(\mathbf{D}^{(d)T} \mathbf{D}^{(d)})}{2}} \exp \left(-\frac{1}{2\tilde{\epsilon}_i^{(d)}} \|\mathbf{D}^{(d)} \mathbf{w}_i\|^2 \right) \quad \text{with} \quad \tilde{\epsilon}_i = [\tilde{\epsilon}_i^{(1)}, \tilde{\epsilon}_i^{(2)}]^T \quad (3.43c)$$

where matrices $\mathbf{D}^{(d)}$, $d = 1, 2$, are the discrete Laplacian operators introduced in Section 3.5 (cf. Eqs (3.24), (3.26) and (3.28b)) and $\tilde{\epsilon}_i^{(d)}$, $d = 1, 2$, are hyperparameters controlling the amount of smoothness. By assuming moreover prior independence between \mathbf{w}_1 and \mathbf{w}_2 , the prior for \mathbf{W} directly reads

$$p(\mathbf{W} | \tilde{\epsilon}) \propto \prod_{i=1}^2 p(\mathbf{w}_i | \tilde{\epsilon}_i) \quad \text{with} \quad \tilde{\epsilon} = \{\tilde{\epsilon}_1, \tilde{\epsilon}_2\}. \quad (3.44)$$

b) Hyperprior distributions

Similarly as in Section 3.5, we adopt a fully Bayesian strategy by including the hyperparameters in the unknowns and assign to each $\tilde{\epsilon}_i^{(d)}$ a conjugate non-informative Jeffreys' prior, i.e.,

$$p(\tilde{\epsilon}_i^{(d)}) \propto \frac{1}{\tilde{\epsilon}_i^{(d)}} \mathbb{1}_{\mathbb{R}^+}(\tilde{\epsilon}_i^{(d)}), d \in \xi. \quad (3.45)$$

When assuming prior independence between hyperparameters, the hierarchical prior on $(\mathbf{W}, \tilde{\epsilon})$ is finally given by the following structure

$$p(\mathbf{W}, \tilde{\epsilon}) \propto p(\mathbf{W}|\tilde{\epsilon})p(\tilde{\epsilon}) \quad (3.46)$$

with

$$p(\tilde{\epsilon}) \propto \prod_{i=1}^2 \prod_{d \in \xi} p(\tilde{\epsilon}_i^{(d)}). \quad (3.47)$$

3.6.2 Sampling of \mathbf{W} and $\tilde{\epsilon}$

We detail here the modifications in the Gibbs sampler when the hierarchical prior (3.46) is injected in the Bayesian models (3.14) and (3.30) instead of the non-informative Gaussian prior (3.3). Those mainly concern the sampling of \mathbf{W} .

a) Sampling of \mathbf{W}

By manipulations of the prior (3.44) (details are given in Appendix C), it can be shown that, conditionally on other elements in \mathbf{W} and $\tilde{\epsilon}$, each $\mathbf{w}_{\underline{k}}$ follows a Gaussian distribution given by

$$\mathbf{w}_{\underline{k}} | \mathbf{w}_{\underline{k}' \neq \underline{k}}, \tilde{\epsilon} \sim \mathcal{N}(\mathbf{w}_{0,\underline{k}}, \Sigma_{0,\underline{k}}) \quad (3.48)$$

where we introduce the notation

$$\mathbf{w}_{0,\underline{k}} = \Sigma_{0,\underline{k}} \sum_{d \in \xi} \sum_{\underline{k}' \neq \underline{k}} -\delta_{\tau(\underline{k}), \tau(\underline{k}')}^{(d)} \Sigma^{(d)-1} \mathbf{w}_{\underline{k}'} \quad (3.49a)$$

$$\Sigma_{0,\underline{k}} = \left(\sum_{d \in \xi} \delta_{\tau(\underline{k}), \tau(\underline{k})}^{(d)} \Sigma^{(d)-1} \right)^{-1} \quad (3.49b)$$

with $\Sigma^{(d)} = \text{diag}(\tilde{\epsilon}_1^{(d)}, \tilde{\epsilon}_2^{(d)})$ and $\delta_{k_1, k_2}^{(d)}$ denotes the element in the k_1 -th row and k_2 -th column of the matrix $\mathbf{D}^{(d)T} \mathbf{D}^{(d)}$. As a result, when injecting (3.46) in (3.14) and (3.30), it can be seen that the conditional posterior distribution (3.18) is replaced by

$$\mathbf{w}_{\underline{k}} | \mathbf{w}_{\underline{k}' \neq \underline{k}}, \mathbf{V}, \mathbf{Z}, \mathbf{M}, \tilde{\epsilon}, \mathbf{Y}, \mathbf{a}_i \sim \mathcal{N}\left(\left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_{2,\underline{k}}} + \Sigma_{0,\underline{k}}^{-1}\right)^{-1} \left(\Sigma_{0,\underline{k}}^{-1} \mathbf{w}_{0,\underline{k}} + \frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \boldsymbol{\mu}_{0,\underline{k}}}{v_{2,\underline{k}}}\right), \left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_{2,\underline{k}}} + \Sigma_{0,\underline{k}}^{-1}\right)^{-1}\right) \quad (3.50)$$

and the conditional posterior distribution (3.33) by

$$\mathbf{w}_{\underline{k}} | \mathbf{w}_{\underline{k}' \neq \underline{k}}, \mathbf{V}, \epsilon, \tilde{\epsilon}, \mathbf{L} \\ \sim \mathcal{N}\left((\mathbf{X}^T \Gamma_{\mathbf{v}_{\underline{k}},0}^{-1} \mathbf{X} + \Sigma_{0,\underline{k}}^{-1})^{-1} (\Sigma_{0,\underline{k}}^{-1} \mathbf{w}_{0,\underline{k}} + \mathbf{X}^T \Gamma_{\mathbf{v}_{\underline{k}},0}^{-1} \mathbf{y}_{0,\underline{k}}), (\mathbf{X}^T \Gamma_{\mathbf{v}_{\underline{k}},0}^{-1} \mathbf{X} + \Sigma_{0,\underline{k}}^{-1})^{-1}\right). \quad (3.51)$$

Note however that, unlike in Sections 3.4 and 3.5, the elements $\mathbf{w}_{\underline{k}}$ are now no longer conditionally independent of each other due to the definition of $\mathbf{w}_{0,\underline{k}}$ in (3.49), which involves other elements $\mathbf{w}_{\underline{k}' \neq \underline{k}}$. Sampling \mathbf{W} by sequentially drawing $\mathbf{w}_{\underline{k}}$ one by one would yield a high computational complexity and potentially lead to a poor exploration of the posterior distribution. To circumvent this issue, we propose to use a colouring sampling scheme. More precisely, when taking into account the sparsity of the matrices $\mathbf{D}^{(d)T} \mathbf{D}^{(d)}$, the sum over \underline{k}' in (3.49) reduces to only few terms. This enables us to identify groups of elements $\mathbf{w}_{\underline{k}}$ that are jointly conditionally independent and that can be thus sampled in parallel. When considering (3.43a), (3.43b) and (3.43c), 3, 9 and 27 such groups can be identified, which yields an efficient sampling of \mathbf{W} in 3, 9 and 27 steps, respectively.

b) Sampling of $\tilde{\epsilon}$

As in Section 3.5, the conditional distributions for the hyperparameters $\tilde{\epsilon}_i^{(d)}$, $d \in \xi$, associated with the hierarchical model (3.46) are inverse-gamma distributions that are easy to sample

$$\tilde{\epsilon}_i^{(d)} | \bullet \sim \mathcal{IG}\left(\frac{\text{rank}(\mathbf{D}^{(d)T} \mathbf{D}^{(d)})}{2}, \frac{\|\mathbf{D}^{(d)} \mathbf{w}_i\|^2}{2}\right), \quad d \in \xi \quad (3.52)$$

where \bullet denotes either $\Lambda, \mathbf{Z}, \mathbf{M}, \mathbf{Y}, \mathbf{a}$ or $\Lambda, \epsilon, \mathbf{L}$.

3.6.3 Numerical experiments

We compare here the performance of the different multivariate Bayesian approaches addressing the joint estimation of multifractal parameters c_1 and c_2 of multivariate datasets. The goals of this subsection are twofold: First, study estimation performance for c_1 when SAR priors are assigned to \mathbf{w}_i ; Second assess the potential cross-coupling between c_1 and c_2 .

For this experiment, we study the following three multivariate Bayesian approaches: GMRF/SAR (using GMRF priors on \mathbf{v}_i and SAR priors on \mathbf{w}_i), SAR/SAR (using SAR priors both on \mathbf{v}_i and \mathbf{w}_i) and IG/SAR (using non-informative inverse-gamma priors on $v_{i,\mathbf{k}}$ and SAR priors on \mathbf{w}_i). As before, we also report estimation performance for the univariate Bayesian approach using independent priors for each $v_{i,\mathbf{k}}$ and $\mathbf{w}_{i,\mathbf{k}}$ (denoted here IG/N) and of the weighted linear regression estimator LF_w .

a) Experimental setup

Scenario. As in Section 3.5, we consider here a scenario consisting of a sequence of $M_t = 100$ multi-temporal images defined as 2D homogeneous MRWs of size $2^7 \times 2^7$ [RV10] with prescribed multifractal properties. Three different combinations of evolutions $\{c_2(k_t), c_1(k_t)\}$ are studied in Fig. 3.9: (a) a discontinuous profile $c_2(k_t)$ with a smooth profile $c_1(k_t)$, (b) two synchronized discontinuous profiles $c_2(k_t)$ and $c_1(k_t)$ and (c) two delayed discontinuous profiles $c_2(k_t)$ and $c_1(k_t)$.

Estimation setting. The setting for this experiment is identical to the one described in Section 3.5, i.e., $(j_1, j_2) = (1, 3)$, $(N_{mc} = 1000, N_{bi} = 300)$ for MCMC algorithms without HMC procedure (IG/N, IG/SAR and GMRF/SAR) and $(N_{mc} = 1500, N_{bi} = 800)$ for those involving an HMC procedure (SAR/SAR, for which we use $L = 20$ steps in the leap-frog method). Finally, the regularization parameter $a_i^{(1)}$ is set to 50 for all scenarios, as in Section 3.5.3 a).

b) Estimation performance

Fig. 3.9 reports, for all the methods considered, the estimation performance for c_1 (quantified via the average, the standard deviation and the root-mean squared error) which is assessed on $MC = 100$ independent realizations of sequences of multi-temporal images. Inspection of Fig. 3.9 leads to the

following observations.

SAR vs. univariate methods. Estimators $\text{LF}_{\mathbf{w}}$, IG/N and IG/SAR are found to have similar bias for the continuous profile of $c_1(k_t)$ in (a). For the discontinuous profiles in (b) and (c), it is observed that IG/SAR is biased in the vicinity of the discontinuities due to the smoothing induced by the SAR priors on parameters \mathbf{w}_i (similar to what was observed in Section 3.5 for $c_2(k_t)$ with SAR priors on \mathbf{v}_i). The major difference between the different estimators clearly lies in the standard deviation. Indeed, IG/SAR is found to significantly and systematically outperform the univariate approaches in terms of standard deviations, in particular LF_{nw} and IG/N which exhibit here the largest standard deviation values (as reported in Chapter 2). More precisely, the average STD reduction factor is close to 2 for $\text{LF}_{\mathbf{w}}$ and reaches 2.5 for LF_{nw} and IG/N . Overall, the benefits of using SAR priors on \mathbf{w}_i for the accurate estimation of c_1 are demonstrated by the RMSE values of IG/SAR , which are in average 2 to 2.5 lower than those of univariate approaches except at the locations of the discontinuities, where the SAR priors induce a bias.

Note finally that these significant improvements are achieved at a very reasonable computational cost since, for instance, IG/SAR takes only $\sim 5\%$ longer to compute than IG/N for the considered dataset. The same remark holds when comparing GMRF/IG and SAR/IG to GMRF/SAR and SAR/SAR .

Cross-coupling. We compare estimation performance for c_1 obtained with IG/SAR (i.e., encoding no smoothness assumption on c_2) and those obtained with GMRF/SAR and SAR/SAR (i.e., encoding a smoothness assumption on c_2). These three estimators are found to provide similar performance both with respect to the mean and the standard deviation. In particular, inspection of Fig. 3.9 (a) and (c) does not reveal noticeable difference in the vicinities of the locations of discontinuities ($k_t = 38, 75$) of the c_2 profile. This suggests that the specification of the prior on \mathbf{v}_i (and hence parameters c_2) has negligible impact on the estimation performance for c_1 . Conversely, when comparing estimation performance for c_2 obtained with IG/N and GMRF/N , it is observed that the prior on \mathbf{w}_i (and hence parameters c_1) has little impact on the estimation performance for c_2 .

3.6.4 Preliminary conclusions

SAR. This section proposed the use of SAR priors for the joint estimation of c_1 for collections of images. Due to their conjugacy within the statistical models introduced in Chapter 2, SAR priors could be easily incorporated for $w_{i,\mathbf{k}}$ with no noticeable impact on the cost of the corresponding inference procedures. It moreover enabled the automatic adjustment of the amount of induced smoothness to be achieved in a fully Bayesian model. Results presented in this section provided strong evidence for the practical benefits of SAR priors for the joint estimation of c_1 . The proposed approach notably yielded improvements in RMSE of a factor between 2 and 2.5 when compared to univariate procedures.

Cross-coupling. The presented results moreover suggest that there is little to no cross-coupling between c_1 and c_2 in the sense that the prior (and hence, estimation performance) of the one does not influence the estimation performance of the other, and vice versa. Overall, the implementation of the presented estimation procedures gives rise to an operational MFA toolbox for the analysis of multivariate images, able to encode different combinations of assumptions on c_1 and c_2 .

3.7 Conclusions and perspectives

Conclusions. This chapter introduced a novel approach for the multifractal analysis of multivariate images. The proposed strategy consisted of formulating a joint estimation of the collections of multifractal parameters of data components in a Bayesian framework by combining the univariate statistical models introduced in Chapter 2 and suitable multivariate priors that encode prior information on the multifractal parameters.

This chapter illustrated the benefits of this approach with the example of priors assuming that multifractal attributes c_1 and c_2 evolve smoothly in privileged (temporal/spectral, spatial) directions. Two priors on the parameters \mathbf{v}_i (related to c_2) were investigated: GMRF and SAR priors. Assigning GMRF priors to \mathbf{v}_i yielded an efficient inference procedure, but the automatic adjustment of the associated hyperparameters via a fully Bayesian approach is difficult (cf. Chapter 5). Conversely, when SAR priors were assigned to \mathbf{v}_i , the associated hyperparameters could be easily

estimated but the model leads to non-standard conditional distributions that required the use of costly acceptance-reject based sampling in the inference procedure. Both priors were found to yield comparable estimation performance for the joint estimation of c_2 . However, inference with GMRF priors can argueably be preferred over one with SAR priors for its low computational complexity. For the joint estimation of c_1 , SAR priors on \mathbf{w}_i were shown to be appropriate since they enabled the exploitation of conjugacy with the univariate statistical models of Chapter 2.

Numerical experiments, conducted on multivariate datasets composed of synthetic multifractal processes, demonstrated the excellent performance of the multivariate Bayesian approach, which significantly outperformed previously existing (univariate) methods in terms of estimation accuracy, at reasonable computational cost.

Perspectives. The work presented in this chapter leads to the formulation of different perspectives. Two of them have been already pointed out above and consist of: Including the estimation of GMRF hyperparameters in the estimation procedure; Reducing the computational cost of the inference procedures. To that end, different research directions are discussed in Chapter 5. Finally, it would be interesting to explore how the proposed methodology, namely formulating a joint estimation via the definition of hierarchical models, could be adapted to image processing tasks such as joint estimation-segmentation or joint estimation-classification. This requires the investigation of other multivariate priors (cf. Chapter 5).

3.8 Conclusions et perspectives (in French)

Conclusions Ce chapitre a introduit une nouvelle approche pour l'analyse multifractale d'images multivariées. La stratégie adoptée dans ce chapitre a été de formuler une estimation conjointe des collections des paramètres multifractaux associées à un jeu d'images multivariées dans un cadre bayésien en combinant le modèle statistique introduit dans le Chapitre 2 et des lois *a priori* multivariées appropriées, permettant d'injecter de l'information *a priori* sur l'évolution des propriétés multifractales.

Ce chapitre a illustré les apports de notre approche sur l'exemple de lois *a priori* induisant une

évolution lisse pour les paramètres multifractaux c_1 et c_2 dans des directions de prédilection (temporelle/spectrale, spatiale). Deux modèles de loi *a priori* ont été étudiés pour les paramètres v_i (reliés à c_2), à savoir GMRF et SAR. L'utilisation de modèles GMRF permet d'aboutir à des algorithmes d'inférence efficaces, mais le réglage automatique des hyperparamètres associés au modèle ne peut pas être formulé via une approche totalement bayésienne (cf. Chapitre 5). Inversement, l'utilisation de modèles SAR permet une estimation simple de ces hyperparamètres, mais la loi *a posteriori* est associée à des lois conditionnelles non standard, ce qui requiert le recours à des méthodes d'échantillonnage coûteuses. Il a été observé que les deux modèles de loi *a priori* offraient des performances d'estimation comparables. Néanmoins, l'utilisation des modèles GMRF peut être préférée par rapport à celle des modèles SAR pour des raisons de coût calculatoire. Pour l'estimation multivariée de c_1 , il a été montré que les modèles SAR étaient appropriés car ils permettaient l'exploitation de propriétés de conjugaison avec le modèle statistique introduit dans le Chapitre 2.

Des simulations numériques, conduites sur des jeux de données multivariées composées de processus multifractaux synthétiques, ont démontré les excellentes performances d'estimation de notre approche bayésienne. Notamment, il a été montré que notre approche surpassait significativement les précédentes formulations univariées en terme de précision d'estimation, et ce, à un coût calculatoire comparable.

Perspectives. Le travail présenté dans ce chapitre permet la formulation de plusieurs perspectives. Deux d'entre elles ont été indiquées et consistent à : incorporer l'estimation des hyperparamètres des modèles GMRF dans la procédure d'estimation; Réduire le coût calculatoire des algorithmes d'estimation. A cette fin, différentes directions de recherche sont discutées dans le Chapitre 5. Enfin, il serait intéressant d'étudier dans quelle mesure notre approche, à savoir formuler une estimation jointe via des modèles hiérarchiques, pourrait être adaptée pour effectuer des tâches classiques du traitement des images, comme la segmentation et la classification. Ce point requiert notamment l'étude d'autres lois *a priori* multivariées (cf. Chapitre 5).

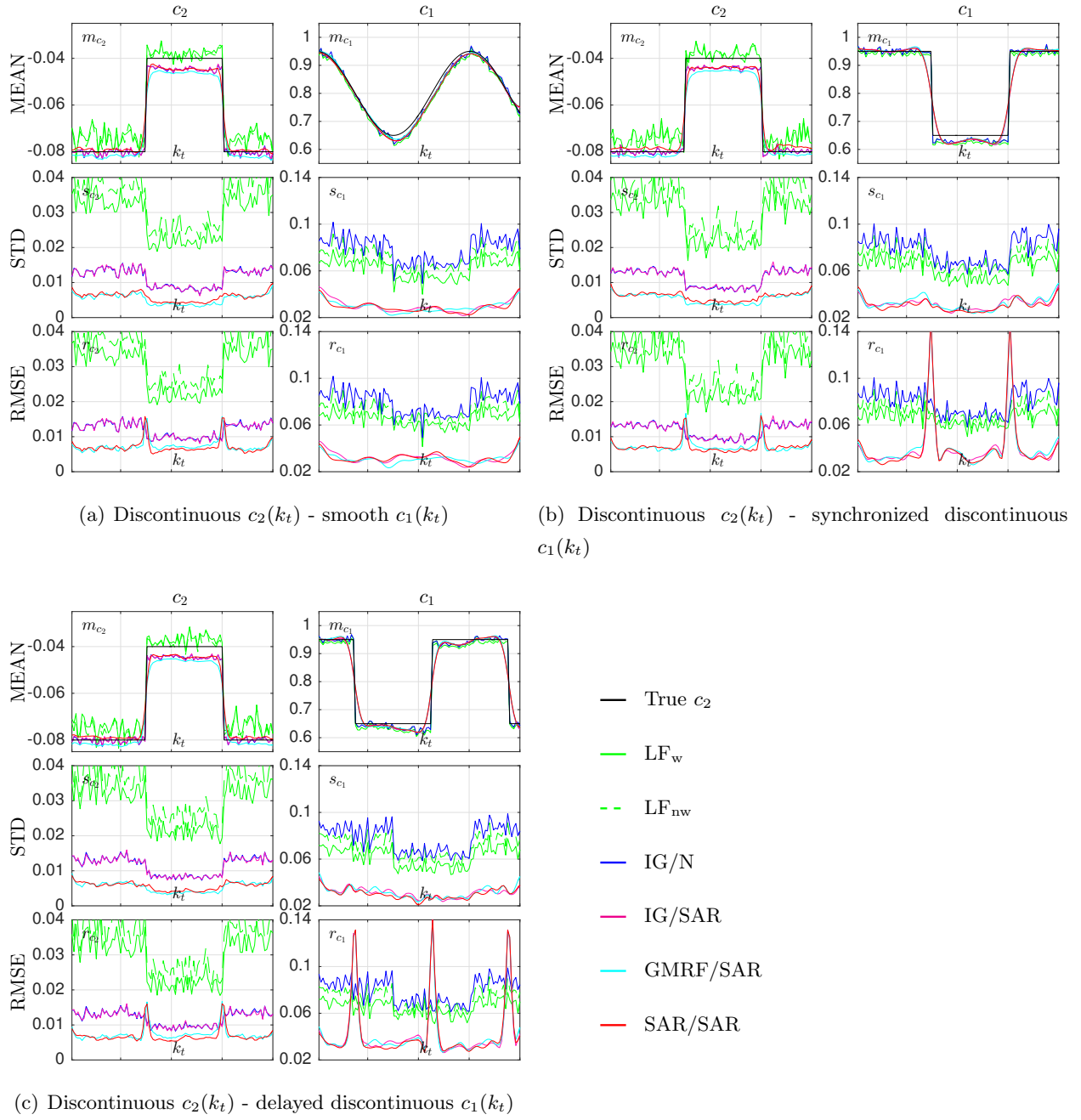


Figure 3.9: Estimation performance for c_2 (left columns) and c_1 (right columns), assessed on 100 independent realizations of a sequence of multi-temporal images defined as 2D MRWs of size $2^7 \times 2^7$ for three different evolutions $\{c_2(k_t), c_1(k_t)\}$.

Chapter 4

Application to real-world data: Illustration for hyperspectral images

Contents

4.1	Introduction	124
4.2	Introduction (in French)	124
4.3	Multifractal analysis and hyperspectral imaging	125
4.3.1	Context	125
4.3.2	Goals and scope	126
4.4	Numerical experiments	126
4.4.1	Analyzed hyperspectral image	126
4.4.2	Model fitting	127
4.4.3	Spectral evolution of multifractal features for hyperspectral images	129
4.4.4	Spatio-spectral evolution of multifractal features for hyperspectral images	131
4.5	Conclusions and perspectives	137
4.6	Conclusions et perspectives (in French)	138

4.1 Introduction

In the two previous chapters, we developed methods enabling multifractal analysis to be embedded in a Bayesian framework. In this chapter, we investigate the potential use and benefits of this Bayesian multifractal methodology for the processing of real-world data. The application considered for this investigation is hyperspectral imaging. The motivation for the use of MFA for hyperspectral images lies in the increasing spatial resolution of hyperspectral remote sensors, which requires the development of new processing methods capable of, first, extracting spatial information, and, second, combining spectral and spatial information in the processing chain. We propose here to use the multifractal parameters to extract spatial information in terms of the fluctuations of the local regularity of image intensity. To that end, we conduct in this chapter experiments on real-world hyperspectral data. These experiments suggest that: First, the Bayesian methodology introduced in this thesis is an operational multifractal analysis tool applicable to the analysis of real-world multivariate sets of images; Second, MFA can provide relevant spatial/textural attributes in the hyperspectral imaging context, which could potentially be employed in tasks such as classification, segmentation or data-mining.

Results presented in this chapter have been partly reported in publications [[CWD⁺15](#), [CWT⁺15c](#), [CWA⁺16d](#)].

4.2 Introduction (in French)

Dans les deux chapitres précédents, nous avons développé des méthodes permettant de formuler l'analyse multifractale dans un cadre bayésien. Dans ce chapitre, nous étudions les bénéfices potentiels de l'utilisation de cette analyse multifractale bayésienne pour le traitement de données réelles. Le domaine d'application considéré pour cette étude est l'imagerie hyperspectrale. Cette étude est motivée par la constante amélioration de la résolution spatiale des capteurs hyperspectraux, nécessitant le développement de nouvelles méthodes de traitement capables tout d'abord d'extraire de l'information spatiale, puis, de la combiner à l'information spectrale. Nous proposons

ici d'utiliser les paramètres multifractaux pour extraire de l'information spatiale en termes de fluctuations de la régularité ponctuelle de l'intensité de l'image. A cette fin, nous conduisons dans ce chapitre des expériences numériques sur un jeu de données hyperspectrales. Ces simulations suggèrent que : Premièrement, la méthodologie bayésienne introduite dans cette thèse fournit des procédures d'estimation opérationnelles pour l'analyse de données réelles multivariées; Deuxièmement, l'analyse multifractale permet d'obtenir des attributs spatiaux et texturaux pertinents qui pourraient potentiellement être utilisés pour des tâches telles que la classification et la segmentation.

Les résultats présentés dans cet chapitre ont fait l'objet de plusieurs publications [CWD⁺15, CWT⁺15c, CWA⁺16d].

4.3 Multifractal analysis and hyperspectral imaging

4.3.1 Context

Hyperspectral imaging is a remote sensing technique which has sparked considerable interest over the last decades. It consists of acquiring an image of a scene in many narrow contiguous spectral bands, typically of the order of one hundred bands (cf., e.g., [LdABD⁺15] for more details on hyperspectral imaging and related issues). This relative wealth of spectral information has motivated the development of processing methods that focus on the spectral information on a pixel-by-pixel-level while neglecting the spatial information contained in the image. However, with recent advances in hyperspectral sensors, achieving sub-meter spatial resolution, efficiently extracting the spatial information and including it into processing schemes has been recognized as an important topic (see, e.g., [AMB02, MLA⁺03]).

A number of studies suggests 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, ZTPP10]. Several authors have proposed to extract spatial information by using textural characterizations, see, e.g., [ZTPP10, RAAF10], for which MFA can provide both a theoretical and a practical framework for

image analysis. In the context of hyperspectral image processing, several attempts have been reported for the use of fractal and multifractal concepts, see, e.g., [Don08, SXGL06, YGJ12] for recent contributions. While these results are encouraging, most of the attempts remain limited for two reasons: conceptually because they consider essentially fractal dimensions instead of the richness of a full multifractal characterization; practically because the algorithms traditionally associated with this quantity (essentially, the box-counting algorithm) have limited accuracy.

4.3.2 Goals and scope

In this chapter, we propose to study the extraction of textural information via the log-cumulants c_p , which, as stated in Chapter 1, provide an efficient summary of the main features of the multifractal spectrum in terms of few parameters. The primary goal of this study is to illustrate the potential benefits of applying the Bayesian methodology introduced in this thesis to the MFA of real-world multivariate images. Meanwhile, it enables us to investigate the potential value of basing the extraction of spatial information in hyperspectral imaging on features derived from multifractal analysis. This investigation constitutes in itself a contribution of this PhD since, to our knowledge, such an analysis has never been reported before in the hyperspectral imaging context.

However, note that results presented in this chapter constitute only a preliminary analysis allowing us to draw first conclusions and to point out where further investigations and subsequent developments are needed. Notably, the definition of a precise methodology for the incorporation of multifractal features in the processing of hyperspectral images is not considered in this chapter.

4.4 Numerical experiments

4.4.1 Analyzed hyperspectral image

The dataset used for our study is a real-world hyperspectral image depicted in Fig. 4.1 in artificial color composition. This hyperspectral image corresponds to a forested area near a city and was acquired over Villelongue (France) by the Hypspx hyperspectral scanner during the Madonna project [SFL⁺11]. It contains 960×1952 pixels with a spatial resolution of 0.5 meters and 160 spectral bands

ranging from the visible to near infrared.



Figure 4.1: Real hyperspectral data acquired by the Hypspx hyperspectral scanner.

4.4.2 Model fitting

Before applying estimation procedures, we numerically investigate the fit between the statistical model introduced in Chapter 2 and the empirical distribution of log-leaders associated with the hyperspectral image under analysis. To that end, we focus in particular on the homogeneous forested area of size 256×256 pixels indicated by a red frame in Fig. 4.1, which will be used in Section 4.4.3. We display in Fig. 4.2 the QQ-plot of the log-leaders at subband $k_\lambda = 100$ along with the empirical covariance, averaged over 6 contiguous subbands, and the parametric covariance model (2.4)¹. The same investigation is reproduced in Fig. 4.3 for the 64×64 subregion of the forested area indicated by a blue frame in Fig. 4.1, which will be used in Section 4.4.4.

Inspection of Figs. 4.2 and 4.3 leads us to the following comments. First, the QQ-plot (a) indicates that the marginal distribution of log-leaders $l(j, \mathbf{k})$ can be indeed well approximated by a Gaussian

¹For this illustration, we injected in the model the parameters values estimated via the univariate Bayesian methodology proposed in Chapter 2, in particular $[\hat{c}_1, \hat{c}_2] = [0.35, -0.03]$.

distribution. Second, when comparing in (e) the radial evolution of the sample covariance (c) and the proposed covariance model (d), the model is found to provide an overall satisfactory fit. These first results hence indicate that the proposed model is relevant for certain areas of the hyperspectral image under analysis. In what follows, we will apply the Bayesian estimation procedures to different regions and spectral bands of the hyperspectral image without systematically assessing the fit of the statistical model. This point would require the development of appropriate model selection statistical procedures (see Chapter 5).

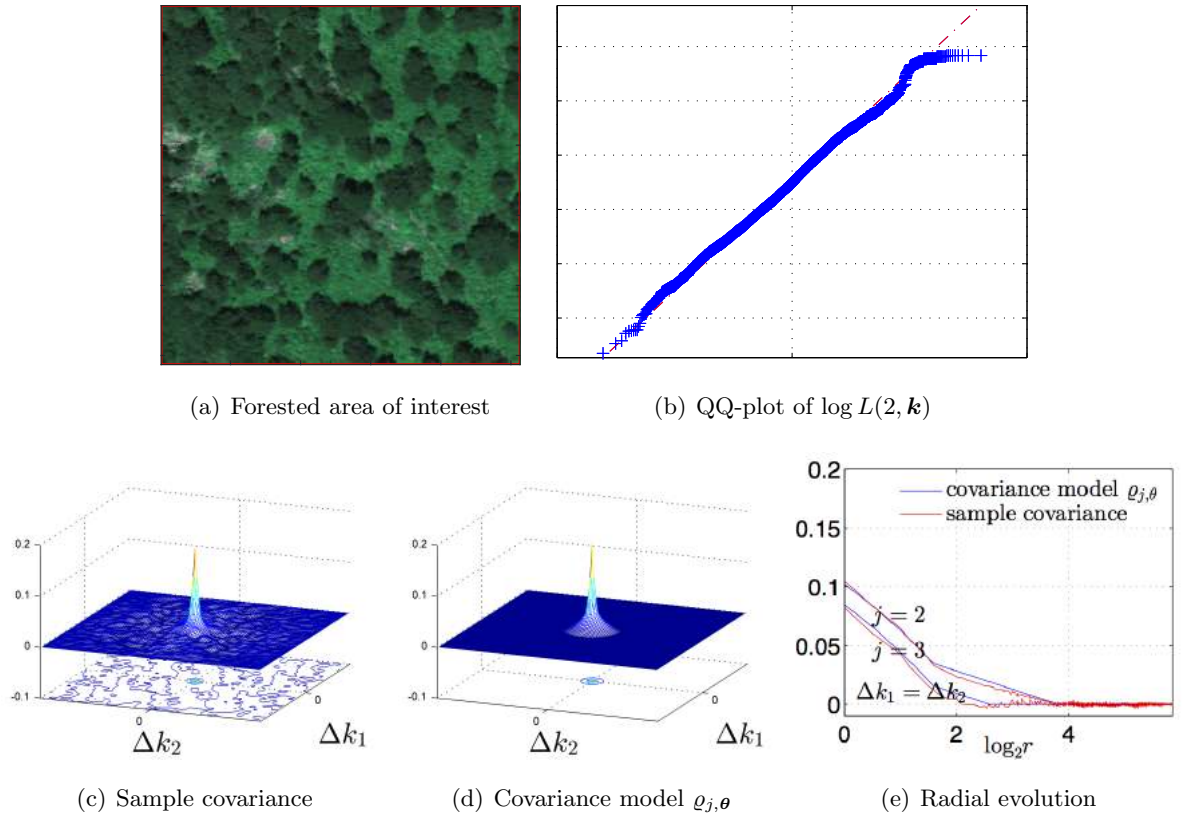


Figure 4.2: Fitting between data and proposed statistical model: Forested area of interest (256×256 pixels) (a), quantile-quantile plot (b) of the empirical distributions of associated log-leaders $\log L(j, \mathbf{k})$ at scale $j = 2$ (subband $k_\lambda = 100$); sample covariance (c), averaged over contiguous subbands $k_\lambda \in \llbracket 100, 105 \rrbracket$, parametric covariance (d); (e) compares the radial evolution of the model (blue) to the sample covariance (red) of log-leaders at scales $j = 2, 3$.

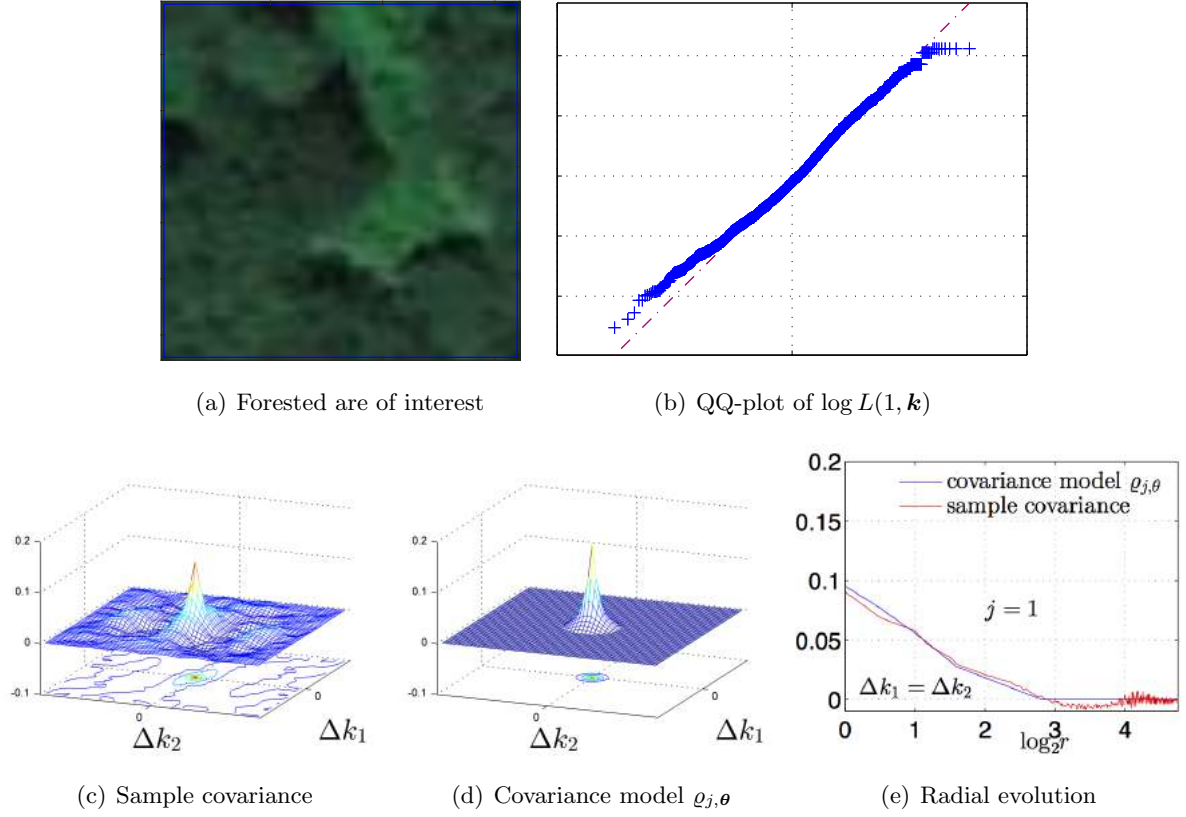


Figure 4.3: Fitting between data and proposed statistical model: Forested area of interested (a) (64×64 pixels), quantile-quantile plot (b) of the empirical distributions of associated log-leaders $\log L(j, \mathbf{k})$ at scale $j = 1$ (subband $k_\lambda = 100$); sample covariance (c), averaged over contiguous subbands $k_\lambda \in \llbracket 100, 105 \rrbracket$, parametric covariance (d); (e) compares the radial evolution of the model (blue) to the sample covariance (red) of log-leaders at scale $j = 1$.

4.4.3 Spectral evolution of multifractal features for hyperspectral images

For this experiment, we are interested in the evolution of multifractal properties across spectral bands. For this purpose, we apply independently to each subband of the homogenous forested area (red frame in Fig. 4.1) the weighted linear regression (denoted below as LF_w) and the univariate Bayesian estimation (as described in Chapter 2 and denoted below as IG/N). Results are summarized in Fig. 4.4.

Fig. 4.4 (b) plots multifractal spectra, estimated by linear regression, of the image 4.4 (a) for three

different bands $k_\lambda = \{62, 100, 120\}$. Clearly, the spectra for different bands differ in shape and position indicating that the mode, the width and the asymmetry of the spectra (respectively tied to c_1 , c_2 and c_3) have their own independent spectral evolutions. For instance, bands $k_\lambda = 100$ and $k_\lambda = 120$ are observed to have similar values of c_1 but different c_3 .

We investigate in more details the spectral evolution of the log-cumulants c_1 and c_2 , which can be estimated by either the weighted linear regression or the Bayesian approach IG/N, allowing further discussion. The spectral evolutions of c_1 and c_2 are displayed in Fig. 4.4 (c) and (d). For comparison, the average reflectance (denoted as MEAN) of the image 4.4 (a) versus spectral bands is reproduced in Fig. 4.4 (e) with its standard deviation (denoted as STD), computed over the patch. Visual inspection of Fig. 4.4 (c-e) leads to the following observations. First, as expected, IG/N and LF_w yield a very similar spectral evolution for c_1 while they produce different evolutions for c_2 . Notably, LF_w yields c_2 estimates with positive values, which are not coherent with multifractal theory since necessarily $c_2 < 0$, while IG/N estimates are consistently negative. Secondly, certain parts of the spectral evolutions of the reflectivity and of the log-cumulants are similar (e.g., the abrupt shift around $k_\lambda = 50$). However, it is worth noticing that particular fluctuations of c_1 or c_2 are observed without being identifiable on the reflectivity spectrum. For instance, the fast decay of c_1 for the last bands cannot be identified on MEAN and STD. The same remark holds for the flat evolution of c_2 displayed by IG/N for the first bands. Finally, the correlation coefficient between the reflectance and log-cumulants (both c_1 and c_2) is computed and found to be on average close to 0.7 both for the average reflectance and the standard deviation. Note however that this coefficient is lower (~ 0.6) for c_2 when using IG/N.

These preliminary results suggest that the log-cumulants c_1 and c_2 are spatial attributes whose spectral evolutions differ from that of the average reflectivity and hence convey complementary information. In that respect, they could, for instance, be incorporated in a data-mining scheme to identify bands of interest based on characteristics of the multifractal spectrum.

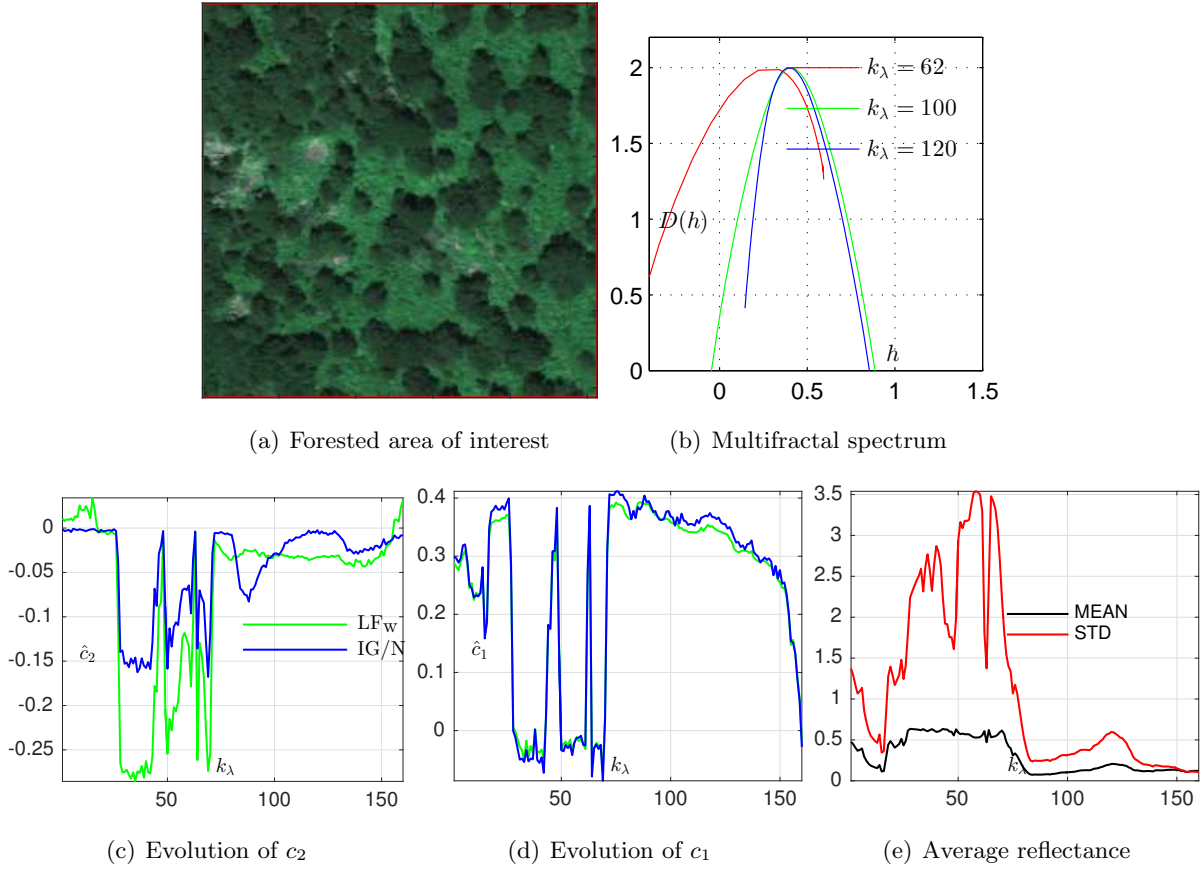


Figure 4.4: Multifractional features of hyperspectral images: Forested area from the Hypspec hyperspectral data (a), multifractional spectrum for three different bands (b), spectral evolution of c_2 (c), spectral evolution of c_1 (d) and average reflectance and standard deviation (e).

4.4.4 Spatio-spectral evolution of multifractional features for hyperspectral images

In a second set of experiments, we propose to conduct analysis on small patches of the hyperspectral images in order to illustrate and study the characterization enabled by the log-cumulants throughout space and bands. For this experiment, we make use of the multivariate Bayesian methodology introduced in Chapter 3 and compare it to univariate approaches.

Analysis scenario. Here, we restrict ourselves to the analysis of the 80 last bands for computational reasons. In an attempt to increase the spatial resolution, each band is decomposed into 29×59

patches of size 64×64 pixels, with 50% overlap, resulting in a decomposition into $29 \times 59 \times 80$ patches indexed by (k_x, k_y, k_λ) . Note that the above overlapping clearly results in a violation of the independence assumption between patches which underlies the multivariate Bayesian models introduced in Chapter 3. Yet, results (and complementary results that have been obtained for synthetic images) indicate that estimation procedures are robust with respect to these violations of model assumptions.

Estimation setting. We compare in this experiment three different estimators: the weighted linear regression based estimator LF_w , the univariate Bayesian estimator IG/N and the multivariate Bayesian estimator (relying on GMRF 2D+1D and SAR priors for \mathbf{v}_i and \mathbf{w}_i , respectively, denoted below as GMRF/SAR). Note that results for the multivariate Bayesian method SAR/SAR (using SAR priors both for \mathbf{v}_i and \mathbf{w}_i) are not reported here since, for that method, processing the whole dataset requires computational capacities beyond the one available (excessively large computational time), and SAR/SAR and GMRF/SAR have been observed to yield similar results on a subset of the hyperspectral cube. Finally, we use for the different estimators a setting similar to the one introduced in Section 3.4 of Chapter 3, where a spatio-temporal decomposition into patches of sizes 64×64 was considered, and is as follows: $(j_1, j_2) = (1, 2)$, $(a_i^{(2)}, a_i^{(1)}) = (10, 20)$.

a) Estimation results for c_1

We discuss here estimation results for c_1 , for which LF_w and IG/N are expected to be similar due to the use of non-informative prior for the latter (cf. Chapter 2). Fig. 4.5 reports estimates obtained for two spectral bands (b) (the bands 87 and 114, which are plotted in (a)) as well as for a slice along the spectral dimension (d) (the corresponding $64 \times 1962 \times 80$ portion of the hyperspectral cube is indicated by a red frame in (c))².

LF vs. Bayesian estimation. Inspection of the results for the bands 87 and 114, cf. Fig. 4.5 (b), confirms that IG/N does not yield improvements in the estimation of c_1 when compared to LF_w . Both approaches lead indeed to similar estimates for c_1 . On the opposite, the multivariate Bayesian

²A supplementary 3D animation of the results is available on <http://combrexelle.perso.enseeiht.fr/thesis/c1HS.zip>.

method GMRF/SAR is found to enable a smooth estimation of c_1 , with reduced spatial variability, yet, without altering the structures observable in LF_w and IG/N c_1 maps. Inspection of the spectral evolution of c_1 , plotted in Fig. 4.5 (d), leads to a similar conclusion. The proposed GMRF/SAR estimator permits a reduction in the variance of estimates along the spectral dimension and yields a visually less noisy evolution of c_1 across the bands of the image, while maintaining its main features.

Discriminative power. Although no ground truth is available for this illustration, a more quantitative analysis of the relative quality of the estimates of c_1 obtained with LF_w , IG/N and GMRF/SAR is proposed in Fig. 4.6. Histograms (a) of the estimates of c_1 obtained with the three methods for subband $k_\lambda = 114$ confirm the similarity between LF_w and IG/N and the slightly reduced variance of GMRF/SAR estimates. The Fisher linear discriminant criterion proposed in [DHS12, Ch. 3.8] is also calculated for the different estimators, as a function of a threshold for c_1 separating two classes of textures and plotted in Fig. 4.6 (b). It indicates that the estimates \hat{c}_1 obtained with GMRF/SAR have slightly superior discriminative power than those obtained with LF_w and IG/N. In a similar vein, we compute for the maps of c_1 in Fig. 4.5 for subband $k_\lambda = 114$ the corresponding sharpness index introduced in [BM12]. This index is a reference-free indicator which assesses image quality in terms of blur, noise or ringing by investigating the Fourier phase spectrum of the image (see [BM12] for details on its implementation and its relation to the global phase coherence based indicator of [BMR08]). These sharpness indexes are close to 3 for LF_w and IG/N and 3.2 for GMRF/SAR, which is in the line with the above remarks.

Overall, quantitative improvements yielded by GMRF/SAR for c_1 hence remain very modest for this hyperspectral image.

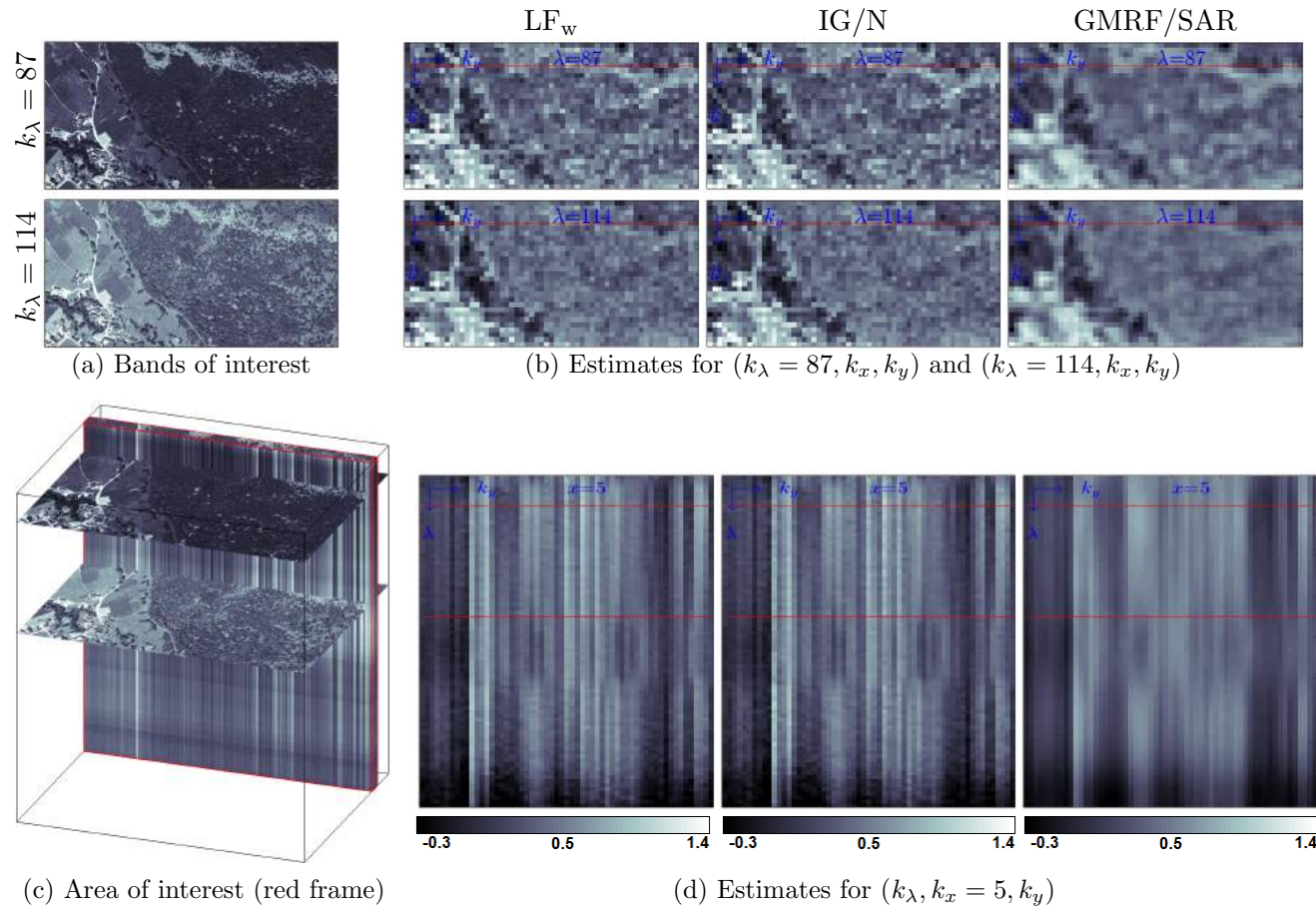


Figure 4.5: Estimate \hat{c}_1 of the hyperspectral image (bands 80 – 160) from the Madonna project.

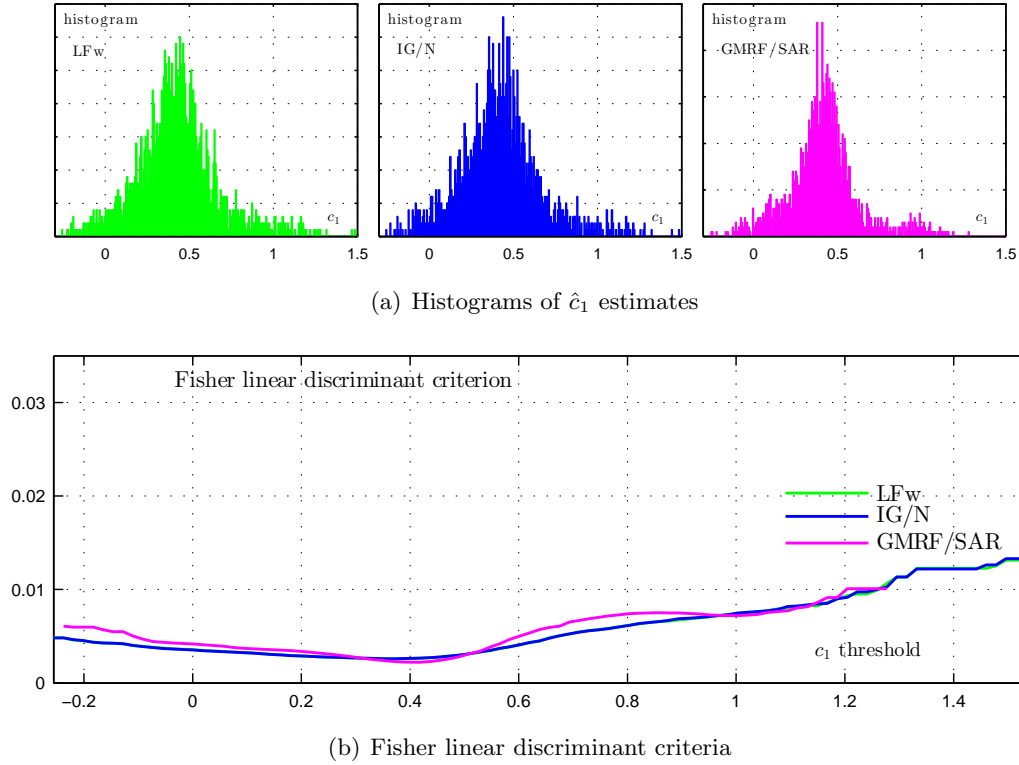


Figure 4.6: Discrimination power of estimation procedures for c_1 : histograms (a) and Fisher linear discriminant criteria (b) for estimators of c_1 obtained by LF_w (—), IG/N (—) and GMRF/SAR (—) for subband $k_\lambda = 114$.

b) Estimation results for c_2

We now investigate estimation results for c_2 , which are summarized in Fig. 4.7 with again a focus on the spectral bands 87 and 114 (b) and a slice along the spectral dimension (d)³. Since for c_2 linear regression and Bayesian approaches strongly differ, reported results allow an interesting comparison between the different approaches.

LF vs. Bayesian estimation. First, LF_w is observed to provide c_2 estimates with strong spatial and spectral variability. On the opposite, the estimator IG/N yields better spatial and spectral

³A supplementary 3D animation of the results is available on <http://combrexelle.perso.enseeiht.fr/thesis/c2HS.zip>.

coherence and clearly improves over the estimates obtained with LF_w , with notably no positive values for c_2 . Yet, the variability within visually homogeneous zones of the dataset (e.g., the forested region) is still important. In comparison with IG/N, and a fortiori with LF_w , the proposed multivariate Bayesian method GMRF/SAR further and dramatically reduces the variability within presumably homogeneous zones, inducing strong spatial coherence, which reinforces the contrast between regions of different multifractalities and visually sharpens their borders. Overall, these results indicate that, for the estimation of c_2 , the regression based estimation is outperformed by the Bayesian approach, especially GMRF/SAR, which manages to reproduce the spatial and spectral structures of the image texture with less variability. This is further assessed in the next paragraph via a discussion on the discriminative power of the different methods.

Discriminative power. As for c_1 , we compared the different methods by analysing the histograms of the estimates c_2 for subband $k_\lambda = 114$ and the associated Fisher linear discriminant criterion, which are displayed in Fig. 4.8. First, histograms (a) of the estimates of c_2 clearly indicate the significantly larger spatial variability (variance) of LF_w as compared to Bayesian approaches. Second, the Fisher linear discriminant criterion (b) [DHS12, Ch. 3.8] of the different methods indicates that the estimates obtained with GMRF/SAR and IG/N have a significantly superior discriminative power than those obtained with LF_w . Finally, the reference-free image quality indicators [BM12] of the different methods are 19.8 and 17.6 for GMRF/SAR and IG/N but a considerably smaller value of 10 for LF . This significant difference strengthens the visual inspection-based conclusions of improved spatial coherence for Bayesian estimators described above.

c) Reflectivity vs. c_1, c_2

Finally, we qualitatively compared the reflectivity and the maps of c_1 and c_2 estimates provided by the different procedures. Undeniably certain structures are visible both on the reflectivity and the log-cumulants, regardless of the estimation procedure. For instance, the city in the left bottom corner yields clusters associated with strongly negative c_2 in Fig. 4.7 and $c_1 \sim 1.2$ in Fig. 4.5. It is however interesting to note that the maps of c_1 and c_2 estimates do not identically reproduce the original

reflectivity image. This is particularly true for the Bayesian estimates, for which the reduced variability enables the identification of distinct regions in the image, notably for c_2 . For instance, observe in Fig. 4.7 that the forested area in the right half of the image is homogeneous (with $c_2 \sim -0.05$) in the spectral band $k_\lambda = 114$, while it is composed of a background (where $c_2 \sim -0.05$) and scattered clusters (where $c_2 \sim -0.1$) in the spectral band $k_\lambda = 87$, which could arguably indicate a physical change at this location (e.g., a lower tree density). Such a behavior is not directly noticeable when inspecting the reflectivity maps.

This observation provides further evidence that the log-cumulants capture complementary information with respect to the traditional moments. Viewed as a function of space, they can potentially provide local texture features that could be combined with spectral features to perform tasks such as classification or segmentation. In that line of thought, the multivariate Bayesian approaches (GMRF/SAR, SAR/SAR or a different combination) are promising as they yield estimates with improved contrast between regions of different textures.

4.5 Conclusions and perspectives

Conclusions. In this chapter, we proposed and investigated the use of the log-cumulants, in particular c_1 and c_2 , for the extraction of spatial/textural information in hyperspectral images. The investigation was based on experiments conducted on a real-world hyperspectral image. Results reported in this chapter enabled us to illustrate that the Bayesian methodology introduced in Chapters 2 and 3 is operational and relevant for the multifractal analysis of real data. Moreover a qualitative analysis of the results indicated that the log-cumulants c_1 and c_2 captured spatial information different from that given by the average and STD of reflectivity, suggesting that they could be used in a complementary manner with standard spectral features to perform tasks such as data-mining, segmentation or classification.

Perspectives. A direct continuation of this work would be the definition and the study of procedures allowing the systematic use of MFA in hyperspectral processing (cf. Chapter 5). Another

interesting lead to explore would be the investigation of the potential benefits of the proposed methodology for other applications where MFA is commonly used. Notably, in biomedical applications such as functional magnetic resonance imaging (fMRI), acquired data are multivariate and the sample size of each component can be limited. For this kind of dataset, the multivariate Bayesian methodology could enhance the quality of the estimation of multifractal parameters.

4.6 Conclusions et perspectives (in French)

Conclusions. Dans ce chapitre, nous avons proposé et étudié l'utilisation des log-cumulants, en particulier c_1 et c_2 , pour l'extraction d'information spatiale/texturale dans le contexte de l'imagerie hyperspectrale. Cette étude a été basée sur des expériences numériques conduites sur un jeu de données hyperspectrales réelles. Les résultats contenus dans ce chapitre nous ont permis d'illustrer que la méthodologie bayésienne introduite dans les Chapitres 2 et 3 était applicable pour l'analyse de données réelles. D'autre part, une analyse qualitative des résultats a indiqué que les log-cumulants c_1 et c_2 capturaient de l'information spatiale différente par rapport à la réflexivité de l'image. Ces résultats ont ainsi suggéré que ces paramètres pourraient être utilisés en complément de l'information spectrale pour accomplir des tâches telles que l'exploration de données, la segmentation ou encore la classification.

Perspectives. Une continuation directe du travail présenté dans ce chapitre serait la définition et l'étude de procédures permettant l'utilisation systématique de l'analyse multifractale pour le traitement de données hyperspectrales (voir Chapitre 5). Une autre piste intéressante à explorer serait l'étude des potentiels apports de notre méthodologie pour d'autres domaines d'application où l'analyse multifractale est communément utilisée. Notamment, dans les applications biomédicales comme l'imagerie par résonance magnétique fonctionnelle, les données recueillis sont par nature multivariées et la taille des échantillons est limitée. Pour ce type de données, la méthodologie bayésienne développée dans cette thèse pourrait améliorer la qualité de l'estimation des paramètres multifractaux.

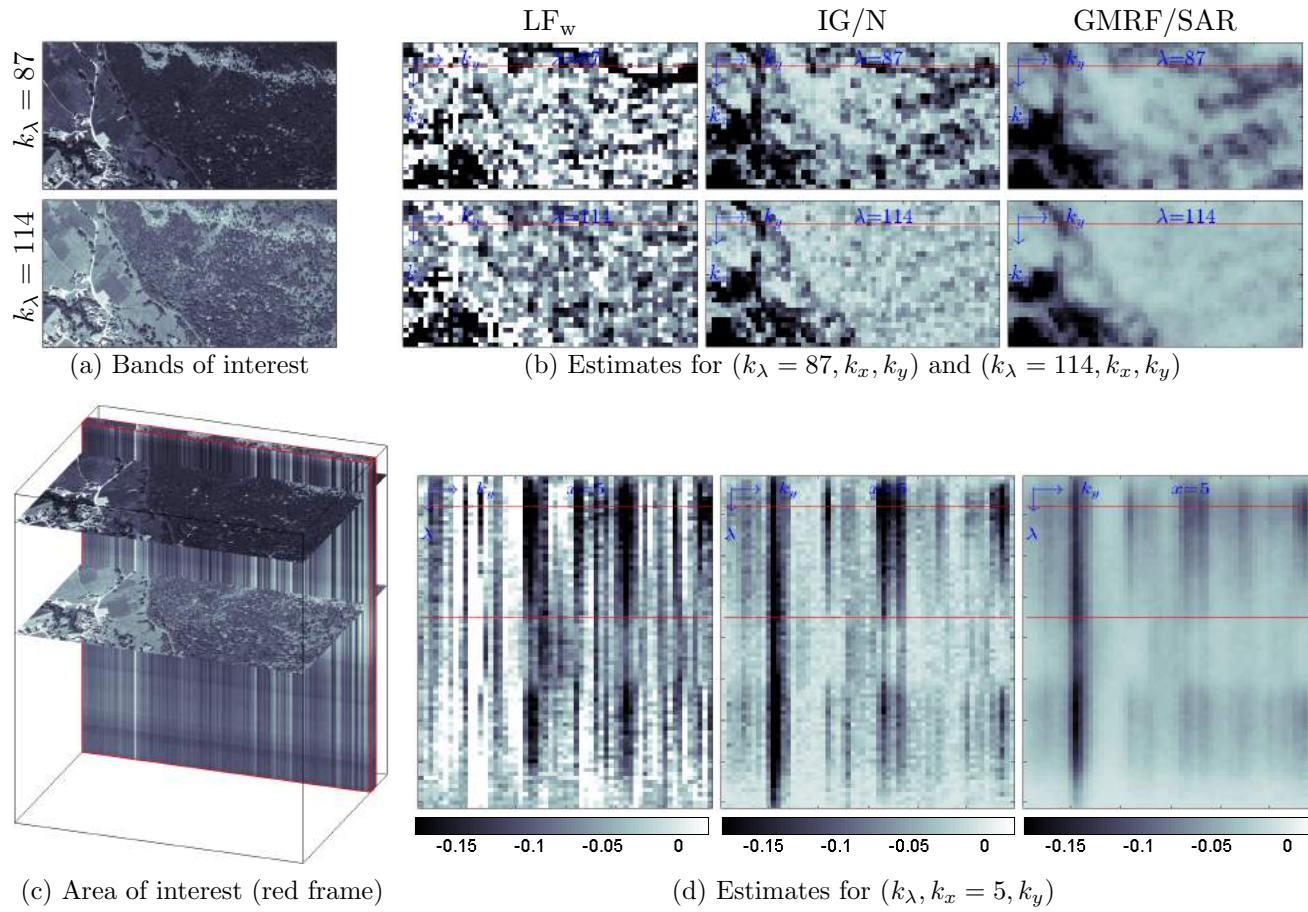


Figure 4.7: Estimate \hat{c}_2 of the hyperspectral image (bands 80 – 160) from the Madonna project.

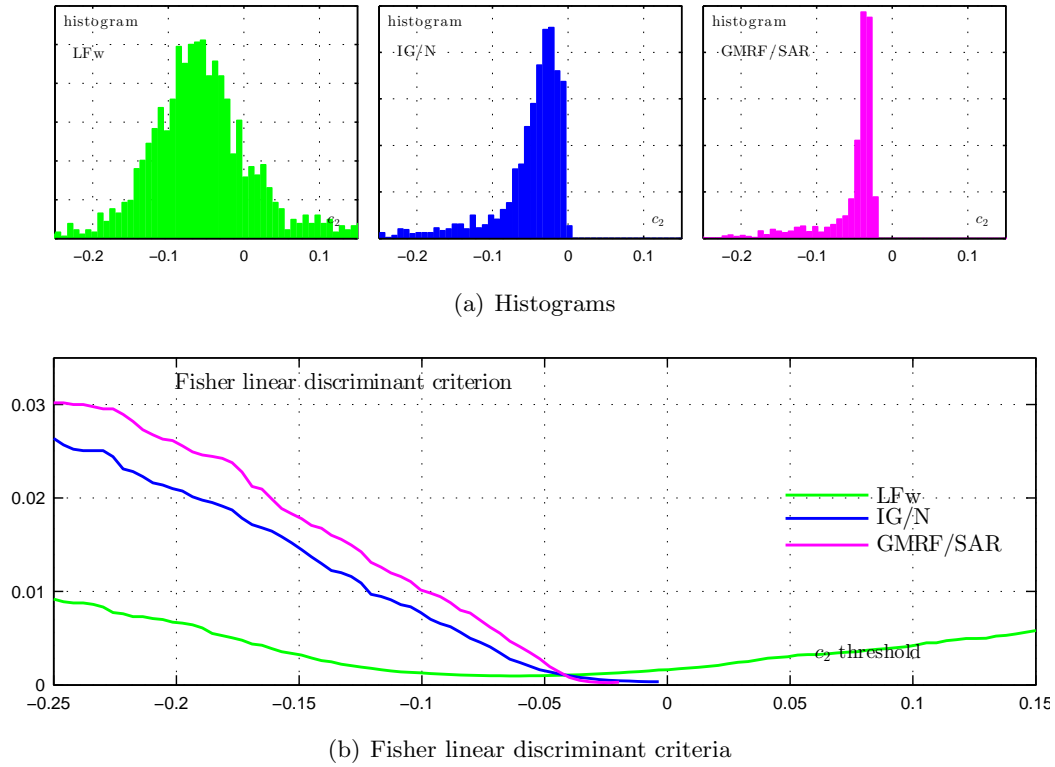


Figure 4.8: Discrimination power of estimation procedures for c_2 : histograms (a) and Fisher linear discriminant criteria (b) for estimators of c_2 obtained by LF_w (—), IG/N (—) and GMRF/SAR (—) for subband $k_\lambda = 114$.

Chapter 5

Conclusions and future work

Context

Over the last decades, MFA has matured into a standard signal and image processing tool, enabling notably texture analysis to be envisaged via the study of the local regularity fluctuations of image amplitudes. Although successfully used in a large range of applications, MFA remains subject to two major practical limitations. First, current benchmark procedures for the estimation of multifractal parameters provide poor performance, in particular for small images, for which the variance of estimators is too large for being used in practical applications. Second, all current existing estimation procedures are conceptually limited to the independent processing of individual images, while in a increasing number of applications, the data to be analysed are multivariate. The primary objective of this thesis was to introduce new methodological solutions that address these practical limitations. To overcome these difficulties, this thesis formulated a novel Bayesian approach for multifractal parameter estimation. Moreover, the use of MFA in general and of these novel methods in particular in the context of hyperspectral imaging was investigated.

Conclusions

Chapter 1 summarized the key concepts of MFA and of the wavelet leader multifractal formalism. Within this multifractal formalism, the multifractal properties of an image can be rephrased in terms of coefficients c_p , called the log-cumulants. In particular, the two first log-cumulants c_1 and c_2 respectively capture the average regularity and the degree of the fluctuations of the regularity, and their estimation constitute the focus of this thesis. The current benchmark estimation procedures

were recalled and existing alternative procedures were reviewed with a discussion on their respective limitations. Finally, Chapter 1 briefly described a representative selection of scaling processes used throughout this thesis to validate the proposed estimation methods.

Chapter 2 introduced a Bayesian approach for the estimation of c_1 and c_2 for one single image. The key ingredient was the design of an empirical yet generic statistical model for the log-leaders, consisting of multivariate Gaussian distributions whose mean and covariance structure were respectively parametrized by the quantities of interest c_1 and c_2 . To bypass the numerically problematic computation of the likelihood associated with this model, the latter was evaluated in the spectral domain via a suitable Whittle approximation. Building on the statistical model and its approximation, a Bayesian model was introduced via the specification of prior distributions for the unknown parameters. Using a data augmentation scheme, we proposed an alternative Bayesian model, which, unlike the previous model, enabled the design of conjugate priors. For both Bayesian models, the associated Bayesian estimators were computed via appropriate MCMC algorithms. Estimation performance was numerically assessed using a large number of multifractal processes for several image sizes. The proposed estimation procedure enabled us to significantly improve the estimation of c_2 by dividing the root-mean square error values by up to 4 for MMC processes, and up to 10 for fBms when compared to the current benchmark estimator. Finally, it was observed that the proposed approach yielded no improvement for the estimation of c_1 when non-informative priors were considered.

Chapter 3 proposed to formulate the multifractal analysis of multivariate images in a Bayesian framework via the design of hierarchical models. These hierarchical models were built on the statistical model introduced in Chapter 2 and incorporated multivariate priors encoding a smoothness assumption on the dynamics of the collections of log-cumulants c_1 and c_2 . To describe the collection of log-cumulants c_2 , we investigated two different probabilistic models, namely GMRF and SAR priors, while for the collection of log-cumulants c_1 , we were naturally led to consider SAR priors. For all hierarchical models, parameter inference on the associated posterior distribution was achieved via suitable MCMC algorithms. We assessed the performance of the different smoothing priors and of the different combinations of priors on c_1 and c_2 by conducting numerical experiments on multivariate

datasets composed of synthetic multifractal processes with prescribed evolutions of c_1 and c_2 . Results demonstrated the benefits of the multivariate Bayesian approach with estimation performance significantly improved compared to univariate formulations.

Chapter 4 investigated the potential use and benefits of MFA and of the proposed Bayesian multifractal methodology in the context of hyperspectral imaging. Numerical experiments were conducted on a real-world hyperspectral image. They indicated that the multifractal parameters captured textural information different from mere reflectivity average or standard deviation. The reported results suggested that MFA is a promising tool for the extraction of spatial/textural features that could be incorporated in the processing of hyperspectral images. Moreover, they indicated that the proposed Bayesian multifractal methodology yielded robust estimation procedures that are operational for the analysis of real-world data. These procedure were notably shown to provide more relevant estimates than the benchmark linear regression based estimation procedure.

Perspectives and future work

At the end of this work, perspectives can be formulated at different levels.

Model developments

The statistical model proposed in Chapter 2 assumed that log-leaders at different scales are independent. This assumption is however not realistic as inter-scale dependence of wavelet coefficients (and a fortiori of the associated (log)-leaders) is commonly reported in the literature (see, e.g., [SS02, LM01, BS99] and references therein). Directions for future work include investigations on the inter-scale dependence of log-leaders and the design of appropriate models accommodating for it (see [ABM98] for results obtained for 1D random wavelet cascades, a specific instance of MMC processes). Such developments may yield further improvements in the estimation of multifractal parameters.

Another interesting lead to explore is the enhancement of the statistical model for the incorporation of additional log-cumulants c_p , $p \geq 3$. In particular, the log-cumulants c_3 and c_4 , respectively linked to the skewness and the kurtosis of the distribution of the log-leaders (cf. (1.15)), provide

information on the asymmetry and the kurtosis of the multifractal spectrum. Formulating their estimation in a Bayesian framework will require the design of a statistical model for log-leaders including higher order cumulants as parameters and, consequently, 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., [GVB07, CBP08, NO12]).

In Chapter 3, we addressed the multifractal analysis of multivariate images via the design of Bayesian hierarchical models incorporating multivariate priors. We focused in particular on smoothing priors to enforce a smooth estimation of the multifractal parameters. Future work needs to include extensions of the methodology with the study of priors accounting for other assumptions. For instance, total variation priors (see, e.g., [SC03, BMK09]) could be considered to induce estimates with piece-wise constant evolutions. In a different vein, it would be interesting to consider Bayesian classification or segmentation via the use of Potts models (cf., e.g., [Fig05, RVF⁺11, EBDT13, PDBT13]).

In this line of thought, the segmentation of a heterogeneous texture could be conducted via a decomposition of the image into rectangular patches followed by a joint estimation of associated multifractal parameters. However, this patch-based approach would be relevant only for large images as the number of patches would be limited for small images, preventing hence the accurate identification of the regions of different texture. One interesting development would consist of moving from a segmentation at the patch level to a segmentation at the pixel level. In that regards, the texture segmentation approach proposed in [Văc14] could fit here since this contribution dwells on a texture model in the Fourier domain as in the present thesis (via the Whittle approximation). The heterogeneous texture is therein seen as the sum of homogeneous textures each affected by a binary mask to be estimated. This approach would bypass the use of patches but at the price of higher computational complexity.

Computational and algorithmic developments

In Chapters 2 and 3, we performed parameter inference via MCMC algorithms. However, the large number of iterations required to achieve good approximations of Bayesian estimators can become computationally problematic for large datasets. An important continuation of this thesis would lie

in the development of efficient inference algorithms with low computational cost. Since the models introduced in Chapters 2 and 3 (with GMRF priors) involve latent variables, it would be interesting to consider the expectation-maximization (EM) algorithm ([DLR77]) and variational Bayesian (VB) methods ([BBB⁺03, Gri10]) as alternatives. Preliminary results for the EM algorithm are reported in Appendix E. Another possibility would be to recast the parameter inference in an optimization framework and then resort to more efficient optimization schemes (see, e.g., [BV04, GMW81, BPC⁺11]).

In Chapter 3, the intractability of the normalizing constant of GMRF priors prevented the estimation of the associated hyperparameters via a fully Bayesian approach. One obvious direction for future work is the investigation of methods addressing the automatic adjustment of GMRF hyperparameters. There are in the literature methods that are dedicated to inference on models with intractable normalizing constants. Among them, we mention, e.g., empirical Bayes based methods as in [PWAT14], in which unknown hyperparameters are replaced by point estimates computed from the observed data, approximate likelihood methods based on pseudo-likelihood ([Bes75]) and contrastive divergence ([Hin02]), and approaches based on extrapolation from precomputed values of $C(\mathbf{a}_i)$ for discrete sets of \mathbf{a}_i ([RVF⁺11, RICV09]).

Application-oriented developments

Results reported in Chapter 4 constitute a preliminary illustration of the application of the proposed Bayesian methodology on real-world data. Its practical use on a larger scale will demand further investigation.

From a methodological point of view, model selection/validation procedures are needed to, for instance, assess whether the log-leaders associated with the image under analysis are well described by the proposed statistical model, and, in the same spirit, to determine the range of scales over which the model holds. This line of research constitutes a challenging perspective 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).

Regarding more specifically the context of HS imaging, the work presented in Chapter 4 leads us to conclude that multifractal attributes could be considered for the extraction of textural information.

The systematic use of multifractal analysis in HS processing remains to be defined and tested on other datasets. In that regards, different prospects need be explored, such as the investigation of different multifractal formalisms (wavelet coefficients vs. wavelet leaders) or of different attributes (c_p , $p \geq 3$), the incorporation of multifractal attributes in segmentation/unmixing algorithms and the comparison with established methods.

Open issue

Finally, a more long-term work would consist of changing the paradigm used for the multifractal analysis of multivariate images via the joint estimation of multifractal parameters, as it has been achieved in Chapter 3, towards a multivariate multifractal analysis in the spirit of appropriate multivariate extensions of the multifractal spectrum $D(h)$. This constitutes a largely open perspective for which many aspects remain to be defined and studied, both from a theoretical and practical point of view. Notably, this would imply the definition of a meaningful and practically useful multifractal spectrum for multivariate data with many components (as opposed to bi-variate definitions in [HJK⁺86, SSLB96, MSKF90]). Similarly, a multivariate multifractal formalism, potentially relying on appropriate multi-component multiresolution coefficients, needs to be defined. Moreover, this would require the design of synthetic model processes with known multivariate multifractal spectra (and the corresponding synthesis procedures) in order to numerically assess the practical analysis tools.

Contexte (in French)

Au cours de ces dernières décennies, l'analyse multifractale a mûri comme un outil standard du traitement des signaux et des images, permettant notamment d'envisager l'analyse de textures via l'étude des fluctuations de la régularité locale de l'amplitude des images. Bien que mise à profit dans un grand nombre d'applications, l'analyse multifractale est soumise à deux limitations pratiques majeures. Premièrement, les procédures d'estimation de référence pour les paramètres multifractaux offrent de faibles performances d'estimation, en particulier pour les petites tailles d'image, pour lesquelles la variance des estimateurs est très élevée. Deuxièmement, toutes les procédures d'estimation actuelles sont conceptuellement limitées au traitement indépendant d'images, alors que dans un nombre croissant d'applications, les données à analyser sont multivariées. L'objectif central de cette thèse était d'introduire de nouvelles solutions méthodologiques abordant ces limitations pratiques. Pour pallier ces limitations, cette thèse a introduit une approche bayésienne originale pour l'estimation des paramètres multifractaux. De plus, l'utilisation de l'analyse multifractale en générale et plus particulièrement des méthodes proposées a été étudiée dans le contexte de l'imagerie hyperspectrale.

Conclusions (in French)

Le Chapitre 1 s'est intéressé à récapituler les concepts clés de l'analyse multifractale et du formalisme multifractal construit sur les coefficients dominants. Dans ce formalisme, les propriétés multifractales d'une image peuvent être reformulées en termes de coefficients c_p , appelés log-cumulants. En particulier, les deux premiers log-cumulants c_1 et c_2 quantifient respectivement la régularité moyenne et le degré des fluctuations de la régularité locale. Leur estimation est au cœur de cette thèse. Les procédures d'estimation de référence ont été rappelées et les procédures alternatives ont été revues, avec notamment une discussion sur leurs limitations respectives. Enfin, le Chapitre 1 a brièvement décrit une sélection de processus multifractaux de référence utilisés tout au long de cette thèse pour valider de nouvelles procédures d'estimation.

Le Chapitre 2 a introduit une approche bayésienne pour l'estimation de c_1 et c_2 pour une seule

image. L'ingrédient clé de cette approche a été la conception d'un modèle statistique empirique mais générique pour le logarithme des coefficients dominants. Ce modèle consiste à modéliser le logarithme des coefficients dominants par des variables gaussiennes distribuées selon une loi normale multivariée dont la moyenne et la covariance sont respectivement paramétrisées par les quantités d'intérêt, à savoir c_1 et c_2 . Pour contourner l'évaluation problématique de la vraisemblance associée à ce modèle, cette dernière a été calculée dans le domaine spectral via une approximation de Whittle. A partir de ce modèle statistique et de son approximation numérique, un modèle bayésien a été introduit via la spécification de lois *a priori* sur les paramètres inconnus. En utilisant une méthode d'augmentation des données, nous avons de plus introduit un modèle bayésien alternatif, pour lequel, et contrairement au modèle précédent, il est possible de spécifier des lois *a priori* conjuguées. Pour ces deux modèles bayésiens, les estimateurs bayésiens correspondants ont été calculés via des méthodes de Monte Carlo par chaînes de Markov. Les performances d'estimation ont été évaluées numériquement à partir de simulations numériques effectuées sur un grand nombre de processus multifractaux de référence. La procédure d'estimation proposée dans ce chapitre nous a permis d'améliorer significativement l'estimation de c_2 en divisant les valeurs de la racine carrée de l'erreur quadratique moyenne des estimateurs par 4 pour les processus construits sur des cascades multiplicatives et par 10 pour le mouvement brownien fractionnaire par rapport à l'estimateur de référence. Pour l'estimation de c_1 , il a été constaté que la procédure d'estimation n'apporte pas d'amélioration lorsque des lois *a priori* non informatives sont utilisées.

Le Chapitre 3 a proposé de formuler l'analyse multifractale d'images multivariées dans un cadre bayésien via la conception de modèles hiérarchiques. Ces modèles hiérarchiques sont construits sur le modèle statistique introduit dans le Chapitre 2 et incorporent en plus des lois *a priori* multivariées formalisant mathématiquement l'hypothèse d'une évolution lente des propriétés multifractales au sein du jeu de données. Pour décrire la collection des log-cumulants c_2 associée à des images multivariées, nous avons étudié deux modèles probabilistes, à savoir des champs gamma markoviens (GMRF) et des modèles d'auto-regressions simultanées (SAR), alors que pour la collection des log-cumulants c_1 , nous avons été naturellement amenés à considérer des modèles SAR. Pour tous les modèles hiérarchiques, les estimateurs bayésiens associés ont été calculés via des méthodes de Monte Carlo par chaînes

de Markov adaptées. Nous avons évalué les performances associées aux différentes lois *a priori* (et à leurs combinaisons) en effectuant des simulations numériques sur des jeux de données composés d'images multifractales synthétiques pour lesquelles nous contrôlions l'évolution des paramètres c_1 et c_2 . Les résultats rapportés dans ce chapitre ont démontré les bénéfices de notre approche bayésienne multivariée, aboutissant à des performances d'estimation significativement améliorées par rapport aux formulations univariées.

Le Chapitre 4 a étudié les potentiels bénéfices de l'utilisation de l'analyse multifractale et de la méthodologie bayésienne d'analyse multifractale introduite dans cette thèse dans le contexte de l'imagerie hyperspectrale. Des expériences numériques ont été conduites sur un jeu de données hyperspectrales réelles. Ces expériences ont indiqué que les paramètres multifractaux permettaient de capturer une information texturale différente de la simple réflectivité. Les résultats rapportés dans ce chapitre ont ainsi suggéré que l'analyse multifractale constitue un outil prometteur pour l'extraction de d'attributs spatiaux et texturaux, qui pourraient potentiellement être inclus dans le traitement d'images hyperspectrales. De plus, ces résultats ont indiqué que la méthodologie bayésienne d'analyse multifractale permet d'obtenir des procédures d'estimation robustes applicables à des données réelles. Il a notamment été montré que ces procédures permettent une estimation des paramètres multifractaux plus pertinente que les procédures de référence.

Appendices

Appendix A

Model fit for synthetic data

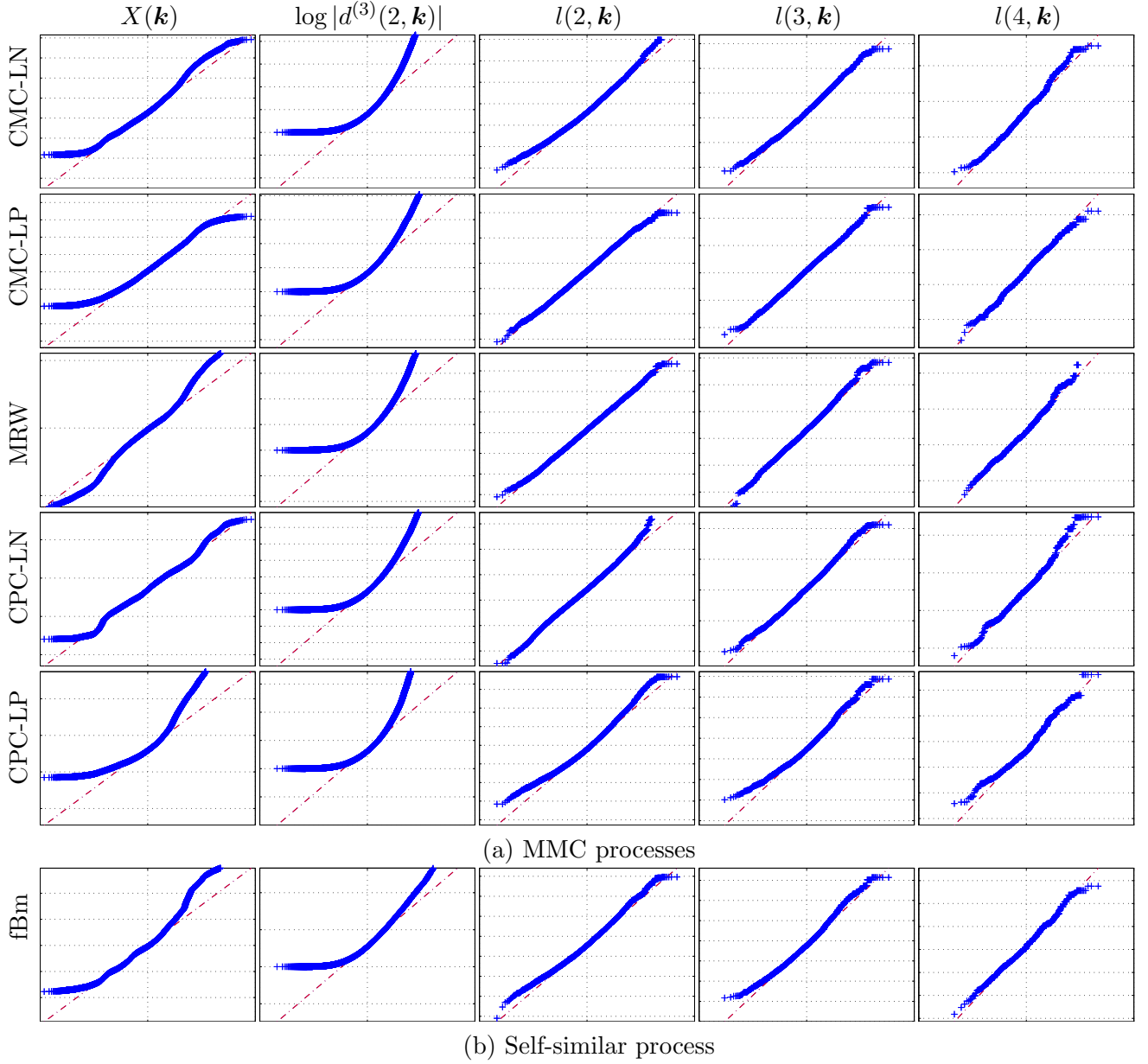


Figure A.1: Quantile-quantile plots of the empirical distributions (+) of the process $X(\mathbf{k})$, the log-wavelet coefficients $\log |d^{(3)}(2, \mathbf{k})|$ and the log-leaders $l(j, \mathbf{k})$, $j = 2, 3, 4$ against standard normal distribution (---) for a selection of MMC processes ($c_2 = -0.04$) (a) and self-similar processes ($H = 0.7$) (b) of size $2^8 \times 2^8$. The closer to the red dashed line, the better the fit to the normal distribution.

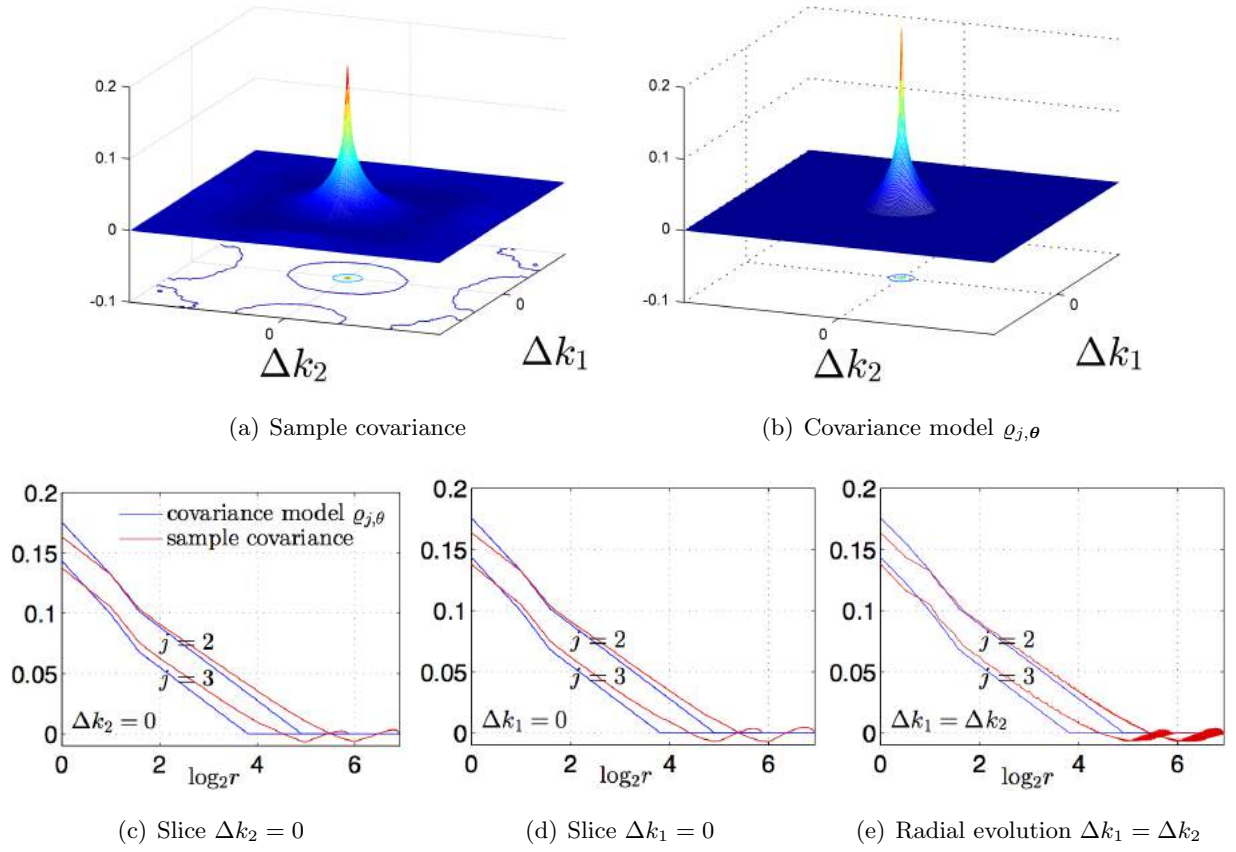


Figure A.2: Fitting between the sample covariance (a), averaged on 100 realizations of 2D MRW ($[N, c_2] = [2^9, -0.04]$), and the parametric covariance (b) at scale $j = 2$; (c), (d) and (e) compare the model (blue) and the sample covariance (red) for different directions at two scales $j = 2, 3$.

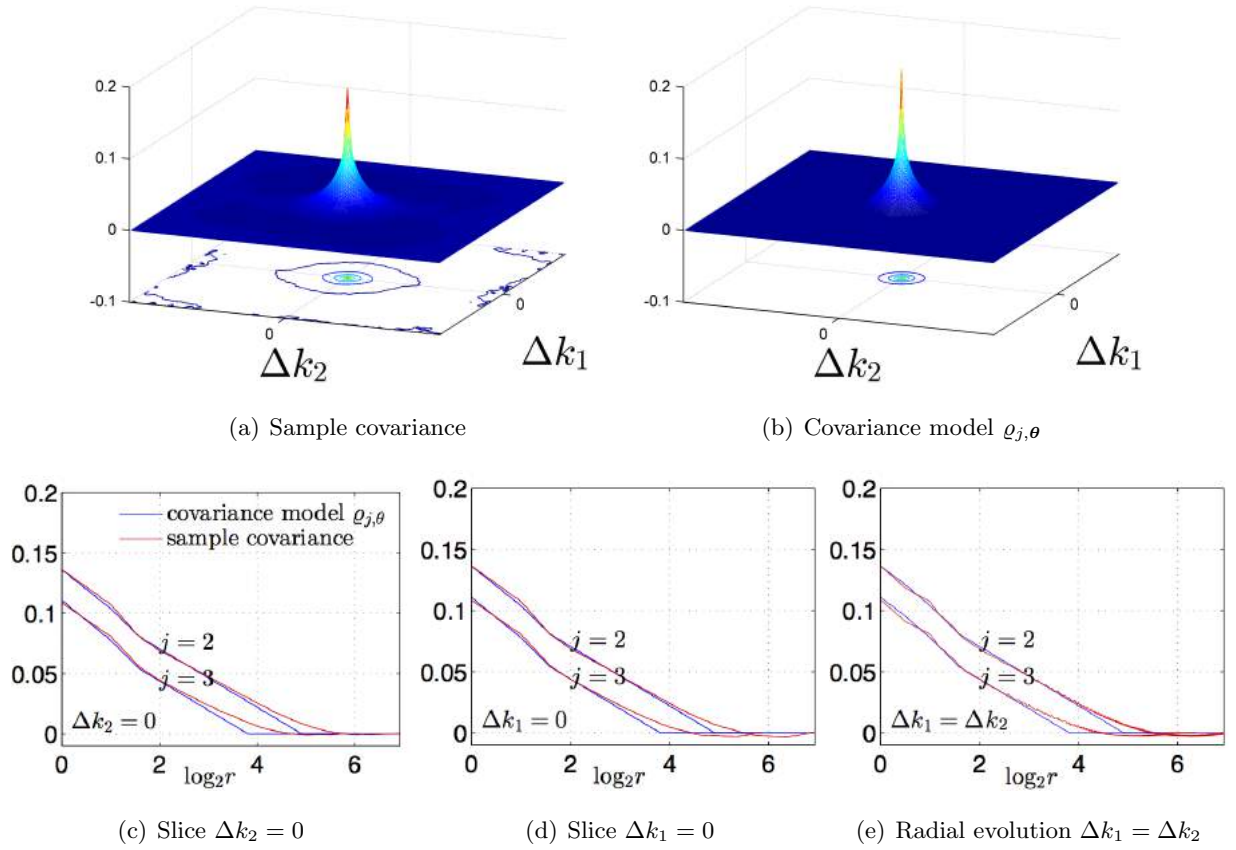


Figure A.3: Fitting between the sample covariance (a), averaged on 100 realizations of 2D CPC-LN ($[N, c_2] = [2^9, -0.04]$), and the parametric covariance (b) at scale $j = 2$; (c), (d) and (e) compare the model (blue) and the sample covariance (red) for different directions at two scales $j = 2, 3$.

Appendix B

Marginalization of the latent variables

In this appendix, we show that the spectral likelihood (2.55)

$$p(\mathbf{y}|\boldsymbol{\lambda}) \propto (\det \boldsymbol{\Gamma}_v)^{-1} \exp\left(-\mathbf{y}_\emptyset^H \boldsymbol{\Gamma}_v^{-1} \mathbf{y}_\emptyset\right) \times (\det \boldsymbol{\Gamma}_{v,0})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_0 - \mathbf{X}\mathbf{w})^T \boldsymbol{\Gamma}_{v,0}^{-1}(\mathbf{y}_0 - \mathbf{X}\mathbf{w})\right) \quad (\text{B.1})$$

is the marginal likelihood of the augmented likelihood (2.58)

$$p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}) \propto v_1^{-N_\emptyset} \exp\left(-\frac{1}{v_1}(\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset)^H \tilde{\mathbf{F}}_1^{-1}(\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset)\right) v_2^{-N_\emptyset} \exp\left(-\frac{1}{v_2} \boldsymbol{\mu}_\emptyset^H \tilde{\mathbf{F}}_2^{-1} \boldsymbol{\mu}_\emptyset\right) \times v_1^{-N_0/2} \exp\left(-\frac{1}{2v_1}(\mathbf{y}_0 - \boldsymbol{\mu}_0)^T \tilde{\mathbf{F}}_{0,1}^{-1}(\mathbf{y}_0 - \boldsymbol{\mu}_0)\right) \times v_2^{-N_0/2} \exp\left(-\frac{1}{2v_2}(\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})^T \tilde{\mathbf{F}}_{0,2}^{-1}(\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})\right) \quad (\text{B.2})$$

when latent variables $\boldsymbol{\mu} = \{\boldsymbol{\mu}_\emptyset, \boldsymbol{\mu}_0\}$ are integrated out of the model (2.56), i.e., we prove that

$$p(\mathbf{y}|\boldsymbol{\lambda}) \propto \int p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}) d\boldsymbol{\mu}. \quad (\text{B.3})$$

To begin with, note that since $p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda})$ is separable in $(\mathbf{y}_\emptyset, \boldsymbol{\mu}_\emptyset)$ and $(\mathbf{y}_0, \boldsymbol{\mu}_0)$, we can separately show that

- (i) the likelihood associated with the model $\mathbf{y}_0|\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{X}\mathbf{w}, v_1\tilde{\mathbf{F}}_{0,1} + v_2\tilde{\mathbf{F}}_{0,2})$ is the marginal likelihood of the augmented likelihood associated with the augmented model

$$\mathbf{y}_0|\boldsymbol{\lambda}, \boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, v_1\tilde{\mathbf{F}}_{0,1}), \quad \boldsymbol{\mu}_0|\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{X}\mathbf{w}, v_2\tilde{\mathbf{F}}_{0,2}) \quad (\text{B.4})$$

- (ii) the likelihood associated with the model $\mathbf{y}_\emptyset|\boldsymbol{\lambda} \sim \mathcal{CN}(\mathbf{0}, v_1\tilde{\mathbf{F}}_1 + v_2\tilde{\mathbf{F}}_2)$ is the marginal likelihood of the augmented likelihood associated with the augmented model

$$\mathbf{y}_\emptyset|\boldsymbol{\lambda}, \boldsymbol{\mu}_\emptyset \sim \mathcal{CN}(\boldsymbol{\mu}_\emptyset, v_1\tilde{\mathbf{F}}_1), \quad \boldsymbol{\mu}_\emptyset|\boldsymbol{\lambda} \sim \mathcal{CN}(\mathbf{0}, v_2\tilde{\mathbf{F}}_2). \quad (\text{B.5})$$

To prove (i), note that since $\tilde{\mathbf{F}}_{0,i}$ is diagonal, we can work with \mathbf{y}_0 element-wise. Let y_0 denote one element of \mathbf{y}_0 , say $y_0 = (\mathbf{y}_0)_k$, and, correspondingly, $f_{0,i} = (\tilde{\mathbf{F}}_{0,i})_{k,k}$ and $x_0 = (\mathbf{X}\mathbf{w})_k$. The likelihood

$$p(y_0|\mathbf{v}, x_0) \propto (f_{0,1}v_1 + f_{0,2}v_2)^{-\frac{1}{2}} \exp\left(-\frac{(y_0 - x_0)^2}{2(f_{0,1}v_1 + f_{0,2}v_2)}\right) \quad (\text{B.6})$$

associated with the Gaussian model

$$y_0|\mathbf{v}, x_0 \sim \mathcal{N}(x_0, v_1f_{0,1} + v_2f_{0,2}) \quad (\text{B.7})$$

is recovered by integrating the latent variable μ_0 out of the following generative model

$$y|\mathbf{v}, x_0, \mu_0 \sim \mathcal{N}(\mu_0, v_1f_{0,1}), \quad \mu_0|\mathbf{v}, x_0 \sim \mathcal{N}(x_0, v_2f_{0,2}) \quad (\text{B.8})$$

associated with the augmented likelihood

$$p(y_0, \mu_0|\mathbf{v}, x_0) \propto \exp\left(-\frac{1}{2v_1f_{0,1}}(y_0 - \mu_0)^2\right) \exp\left(-\frac{1}{2v_2f_{0,2}}(\mu_0 - x_0)^2\right). \quad (\text{B.9})$$

Indeed, by denoting $\tilde{y}_0 = y_0 - x_0$ and using the change of variable $\mu_0 - x_0 \rightarrow \mu_0$, we have the following development

$$\begin{aligned} \int_{-\infty}^{\infty} p(y_0, \mu_0|\mathbf{v}, x_0) d\mu_0 &\propto \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2v_1f_{0,1}}(\tilde{y}_0 - \mu_0)^2\right) \exp\left(-\frac{1}{2v_2f_{0,2}}\mu_0^2\right) d\mu_0 \\ &\propto \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\mu_0^2 \frac{f_{0,1}v_1 + f_{0,2}v_2}{f_{0,1}v_1f_{0,2}v_2} - 2\tilde{y}_0\mu_0 \frac{1}{f_{0,1}v_1} + \tilde{y}_0^2 \frac{1}{f_{0,1}v_1}\right)\right) d\mu_0 \\ &\propto \int_{-\infty}^{\infty} \exp\left(-\frac{f_{0,1}v_1 + f_{0,2}v_2}{2f_{0,1}v_1f_{0,2}v_2}\left(\mu_0^2 - 2\tilde{y}_0\mu_0 \frac{f_{0,2}v_2}{f_{0,1}v_1 + f_{0,2}v_2} + \tilde{y}_0^2 \frac{f_{0,2}v_2}{f_{0,1}v_1 + f_{0,2}v_2}\right)\right) d\mu_0 \\ &\propto \exp\left(\frac{\tilde{y}_0^2}{f_{0,1}v_1 + f_{0,2}v_2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{f_{0,1}v_1 + f_{0,2}v_2}{2f_{0,1}v_1f_{0,2}v_2}\left(\mu_0 - \tilde{y}_0 \frac{f_{0,2}v_2}{f_{0,1}v_1 + f_{0,2}v_2}\right)^2\right) d\mu_0 \\ &\propto \exp\left(-\frac{(y_0 - x_0)^2}{2(f_{0,1}v_1 + f_{0,2}v_2)}\right) \\ &\propto p(y_0|\mathbf{v}, x_0) \end{aligned}$$

which proves (i). To prove (ii), since $\tilde{\mathbf{F}}_i$ is also diagonal, we can use the same argument with manipulations on the complex Gaussian distribution.

Appendix C

Appendix of Chapter 3

C.1 Derivates of the potential energy

The HMC algorithm introduced in Chapter 3 requires the computation of the derivatives of the potential energy (3.34). We show here that these derivatives can be calculated analytically. To begin with, when injecting (3.30) in (3.34), the potential energy directly reads

$$E(\mathbf{v}_i) = \sum_{\mathbf{k}} \left(\sum_{j=j_1}^{j_2} \left(\sum_{\mathbf{m}} \log \phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}}) + \frac{|y_{\ell_{j,\mathbf{k}}}(\omega_{\mathbf{m}})|^2}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}})} \right) + \frac{\log \phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_0)}{2} + \frac{(y_{\ell_{j,\mathbf{k}}}(\omega_0) - m_j \mu_{j,\mathbf{w}_{\mathbf{k}}})^2}{2\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_0)} \right) + \sum_{d \in \xi} \frac{\|\mathbf{D}^{(d)} \mathbf{v}_i\|^2}{2\epsilon_i^{(d)}}$$

where $\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}}) = v_{1,\mathbf{k}} \tilde{f}_{1,j}(\omega_{\mathbf{m}}) + v_{2,\mathbf{k}} \tilde{f}_{2,j}(\omega_{\mathbf{m}})$ with functions $\tilde{f}_{i,j}$ defined in Eq. (2.39). Then, by making use of the two following intermediate derivatives

$$\begin{aligned} \text{(i)} \quad & \frac{\partial \log \phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}})}{\partial v_{i,\mathbf{k}}} = \frac{\tilde{f}_{i,j}(\omega_{\mathbf{m}})}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}})} \\ \text{(ii)} \quad & \frac{\partial \|\mathbf{D}^{(d)} \mathbf{v}_i\|^2}{\partial v_i} = 2\mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{v}_i \end{aligned}$$

the partial derivative of $E(\mathbf{v}_i)$ with respect to $v_{i,\mathbf{k}}$ can be expressed as

$$\begin{aligned} \frac{\partial E(\mathbf{v}_i)}{\partial v_{i,\mathbf{k}}} &= \sum_{j=j_1}^{j_2} \left(\sum_{\mathbf{m}} \frac{\tilde{f}_{i,j}(\omega_{\mathbf{m}})}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}})} \left(1 - \frac{|y_{\ell_{j,\mathbf{k}}}(\omega_{\mathbf{m}})|^2}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_{\mathbf{m}})} \right) + \frac{\tilde{f}_{i,j}(\omega_0)}{2\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_0)} \left(1 - \frac{(y_{\ell_{j,\mathbf{k}}}(\omega_0) - m_j \mu_{j,\mathbf{w}_{\mathbf{k}}})^2}{\phi_{j,\mathbf{v}_{\mathbf{k}}}(\omega_0)} \right) \right) \\ &+ \sum_{d \in \xi} \left[\frac{\mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{v}_i}{\epsilon_i^{(d)}} \right]_{\tau(\mathbf{k})} \end{aligned} \quad (\text{C.1})$$

where we recall that $\tau(\mathbf{k})$ is the lexicographic ordering operator.

C.2 Sampling of \mathbf{W}

In this section, we derive the conditional posterior distributions (3.50) and (3.51) that are required for the sampling of \mathbf{W} in Section 3.6.2. To do so, we first discuss the conditional prior distribution of elements in \mathbf{W} induced by the prior (3.44).

C.2.1 Conditional prior distribution

Temporal/spectral or spatial decomposition

To simplify, let us first consider the case where ξ reduces to a singleton, i.e., $d = 1$ or $d = 2$. In this case, the prior (3.44) can be developed as follows

$$p(\mathbf{W}|\tilde{\epsilon}) \propto \exp \left(-\frac{1}{2} \left[\frac{1}{\tilde{\epsilon}_1^{(d)}} \mathbf{w}_1^T \mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{w}_1 + \frac{1}{\tilde{\epsilon}_2^{(d)}} \mathbf{w}_2^T \mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{w}_2 \right] \right). \quad (\text{C.2})$$

We now try to find the distribution of the element $\mathbf{w}_{\underline{k}} = [w_{1,\underline{k}}, w_{2,\underline{k}}]^T$ conditionally on other elements $\mathbf{w}_{\underline{k}' \neq \underline{k}}$. By using the symmetry of $\mathbf{D}^{(d)T} \mathbf{D}^{(d)}$ and the notation $\delta_{k_1, k_2}^{(d)}$, which stands here for the element in the k_1 -th row and k_2 -th column of the matrix $\mathbf{D}^{(d)T} \mathbf{D}^{(d)}$, we have the following equation

$$\mathbf{w}_i^T \mathbf{D}^{(d)T} \mathbf{D}^{(d)} \mathbf{w}_i = \delta_{\tau(\underline{k}), \tau(\underline{k})}^{(d)} w_{i,\underline{k}}^2 + 2 \sum_{\underline{k}' \neq \underline{k}} \delta_{\tau(\underline{k}), \tau(\underline{k}')}^{(d)} w_{i,\underline{k}} w_{i,\underline{k}'} + \dots \quad (\text{C.3})$$

which, when injected in (C.2), leads to

$$p(\mathbf{W}|\tilde{\epsilon}) \propto \exp \left(-\frac{1}{2} \left[\delta_{\tau(\underline{k}), \tau(\underline{k})}^{(d)} \mathbf{w}_{\underline{k}}^T \boldsymbol{\Sigma}^{(d)-1} \mathbf{w}_{\underline{k}} + 2 \sum_{\underline{k}' \neq \underline{k}} \delta_{\tau(\underline{k}), \tau(\underline{k}')}^{(d)} \mathbf{w}_{\underline{k}}^T \boldsymbol{\Sigma}^{(d)-1} \mathbf{w}_{\underline{k}'} + \dots \right] \right) \quad (\text{C.4})$$

with

$$\boldsymbol{\Sigma}^{(d)} = \text{diag}(\tilde{\epsilon}_1^{(d)}, \tilde{\epsilon}_2^{(d)}). \quad (\text{C.5})$$

Therefore, it directly reads that

$$\mathbf{w}_{\underline{k}} | \mathbf{w}_{\underline{k}' \neq \underline{k}}, \tilde{\epsilon} \sim \mathcal{N} \left(-\frac{1}{\delta_{\tau(\underline{k}), \tau(\underline{k})}^{(d)}} \sum_{\underline{k}' \neq \underline{k}} \delta_{\tau(\underline{k}), \tau(\underline{k}')}^{(d)} \mathbf{w}_{\underline{k}'}, \frac{\boldsymbol{\Sigma}^{(d)}}{\delta_{\tau(\underline{k}), \tau(\underline{k})}^{(d)}} \right). \quad (\text{C.6})$$

Spatio-temporal/spectral decomposition

We now consider the prior distribution when $\xi = \{1, 2\}$, which can be seen as the product of the prior

distribution when $\xi = \{1\}$ and the one when $\xi = \{2\}$. To that end, we recall that the product of two multivariate Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ is a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with [PP⁺08]

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2) \quad (\text{C.7a})$$

$$\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}. \quad (\text{C.7b})$$

Therefore, by identifying

$$\boldsymbol{\mu}_d = -\frac{1}{\delta_{\tau(\underline{\mathbf{k}}), \tau(\underline{\mathbf{k}})}^{(d)}} \sum_{\underline{\mathbf{k}}' \neq \underline{\mathbf{k}}} \delta_{\tau(\underline{\mathbf{k}}), \tau(\underline{\mathbf{k}}')}^{(d)} \mathbf{w}_{\underline{\mathbf{k}}'} \quad (\text{C.8a})$$

$$\boldsymbol{\Sigma}_d = \frac{\boldsymbol{\Sigma}^{(d)}}{\delta_{\tau(\underline{\mathbf{k}}), \tau(\underline{\mathbf{k}})}^{(d)}} \quad (\text{C.8b})$$

in Eq. (C.7) for $d = 1, 2$, we directly find Eqs. (3.49) and (3.48), i.e.,

$$\mathbf{w}_{\underline{\mathbf{k}}} | \mathbf{w}_{\underline{\mathbf{k}}' \neq \underline{\mathbf{k}}}, \tilde{\boldsymbol{\epsilon}} \sim \mathcal{N}(\mathbf{w}_{0, \underline{\mathbf{k}}}, \boldsymbol{\Sigma}_{0, \underline{\mathbf{k}}})$$

with

$$\begin{aligned} \mathbf{w}_{0, \underline{\mathbf{k}}} &= \boldsymbol{\Sigma}_{0, \underline{\mathbf{k}}} \sum_{d \in \xi} \sum_{\underline{\mathbf{k}}' \neq \underline{\mathbf{k}}} -\delta_{\tau(\underline{\mathbf{k}}), \tau(\underline{\mathbf{k}}')}^{(d)} \boldsymbol{\Sigma}^{(d)-1} \mathbf{w}_{\underline{\mathbf{k}}'} \\ \boldsymbol{\Sigma}_{0, \underline{\mathbf{k}}} &= \left(\sum_{d \in \xi} \delta_{\tau(\underline{\mathbf{k}}), \tau(\underline{\mathbf{k}})}^{(d)} \boldsymbol{\Sigma}^{(d)-1} \right)^{-1}. \end{aligned}$$

Note that the formulations above still hold for $\xi = \{1\}$ and $\xi = \{2\}$.

C.2.2 Conditional posterior distribution

In a similar fashion, Eqs. (3.50) and (3.51) are derived from the property on the product of two Gaussian distributions. In particular, Eq. (3.50) is found by identifying

$$\begin{aligned} \boldsymbol{\mu}_1 &= \mathbf{w}_{0, \underline{\mathbf{k}}} \text{ and } \boldsymbol{\mu}_2 = \boldsymbol{\mu}_{0, \underline{\mathbf{k}}} \\ \boldsymbol{\Sigma}_1 &= \boldsymbol{\Sigma}_{0, \underline{\mathbf{k}}} \text{ and } \boldsymbol{\Sigma}_2 = \left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_{2, \underline{\mathbf{k}}}} \right)^{-1} \end{aligned}$$

in (C.7), while (3.51) is found by identifying

$$\begin{aligned}\boldsymbol{\mu}_1 &= \mathbf{w}_{0,\underline{\mathbf{k}}} \text{ and } \boldsymbol{\mu}_2 = \mathbf{y}_{0,\underline{\mathbf{k}}} \\ \boldsymbol{\Sigma}_1 &= \boldsymbol{\Sigma}_{0,\underline{\mathbf{k}}} \text{ and } \boldsymbol{\Sigma}_2 = \left(\mathbf{X}^T \boldsymbol{\Gamma}_{\mathbf{v}_{\underline{\mathbf{k}},0}}^{-1} \mathbf{X} \right)^{-1}.\end{aligned}$$

Appendix D

Whittle approximation for time series

In this appendix, we briefly describe the statistical model for the collection of log-leaders $\log L(j, k)$ of time series. This model uses as its main ingredients of the model proposed in [WDTA13] and is based on two complementary developments. First, unlike [WDTA13], the mean of the log-leaders is included in the model, which enables the estimation of c_1 . Second, a Whittle approximation is devised to enable the efficient evaluation of the associated likelihood in the spectral model. With the resulting model replacing (2.12), all developments of Chapters 2 and 3 carry over to time series, *mutatis mutandis*.

Statistical model. At each scale j , the distribution of the collection of log-leaders $\{\log L(j, k)\}_{k=1}^{n_j}$ of 1D MMC processes is modeled by a multivariate Gaussian distribution with the following mean

$$\mathbb{E}[\log L(j, k)] = \mu_{j, \mathbf{w}} = c_1^0 + j c_1 \log 2 \quad (\text{D.1})$$

where \mathbf{w} is defined in (2.2), and with a variance-covariance $\varrho_{j, \boldsymbol{\theta}}(\Delta k)$ defined as follows [WDTA13]

$$\varrho_{j, \boldsymbol{\theta}}(\Delta k) = \begin{cases} \varrho_{j, \boldsymbol{\theta}}^S(\Delta k) & 0 \leq |\Delta k| < 3 \\ \varrho_{j, \boldsymbol{\theta}}^L(\Delta k) & 3 \leq |\Delta k| \end{cases} \quad (\text{D.2})$$

with

$$\varrho_{j, \boldsymbol{\theta}}^L(\Delta k) = \max(0, \Delta \varrho + c_2(\log |\Delta k| + \log 2^j)) \quad (\text{D.3})$$

and

$$\varrho_{j, \boldsymbol{\theta}}^S(\Delta k) = \frac{\log(|\Delta k| + 1)}{\log 4} (\varrho_{j, \boldsymbol{\theta}}^L(3) - (c_2^0 + c_2 \log 2^j)) + c_2^0 + c_2 \log 2^j \quad (\text{D.4})$$

where $\boldsymbol{\theta}$ is defined in (2.3). The parameter $\Delta \varrho$ in (D.3) was originally set using the heuristic condition $\varrho_{j, \boldsymbol{\theta}}^S(\lfloor n_j/5 \rfloor) = 0$ in [WDTA13]. It is also possible to include it in the vector of unknown parameters

in order to devise an estimation procedure for the integral scale, another parameter of interest in multifractal models [Man74, Fri95]. This has been studied in detail in [CWA⁺15]. In what follows, $\Delta\varrho$ is a constant.

By denoting again ℓ_j the vector gathering the log-leaders at scale j (organized in lexicographic order) the likelihood of ℓ_j with respect to $\lambda = \{\theta, \mathbf{w}\}$ is given by

$$p(\ell_j|\lambda) \propto (\det \Sigma_{j,\theta})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\ell_j - \mu_{j,\mathbf{w}})^T \Sigma_{j,\theta}^{-1} (\ell_j - \mu_{j,\mathbf{w}})\right) \quad (\text{D.5})$$

where $\Sigma_{j,\theta}$ corresponds here to the $n_j \times n_j$ covariance matrix whose entries are induced by the 1D parametric covariance function model (D.2) and $\mu_{j,\mathbf{w}} = \mu_{j,\mathbf{w}} \mathbf{1}_{n_j}$. Finally, assuming independence between log-leaders at different scales j , the joint likelihood for the vector $\ell = [\ell_{j_1}^T, \dots, \ell_{j_2}^T]^T$ gathering the log-leaders at scales $j = j_1, \dots, j_2$ reads

$$p(\ell|\lambda) \propto \prod_{j=j_1}^{j_2} (\det \Sigma_{j,\theta})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\ell_j - \mu_{j,\mathbf{w}})^T \Sigma_{j,\theta}^{-1} (\ell_j - \mu_{j,\mathbf{w}})\right). \quad (\text{D.6})$$

Whittle approximation. The Whittle approximation of the scale-wise Gaussian likelihood (D.5) is given by [Whi53, Pri81]

$$\begin{aligned} p_W(\ell_j|\lambda) &= \prod_{m=1}^{\lfloor \frac{n_j}{2} \rfloor} (\phi_{j,\theta}(\omega_m))^{-1} \exp\left(-\frac{y_{\ell_j}^*(\omega_m) y_{\ell_j}(\omega_m)}{\phi_{j,\theta}(\omega_m)}\right) \\ &\times (\phi_{j,\theta}(\omega_0))^{-\frac{1}{2}} \exp\left(\frac{(y_{\ell_j}(\omega_0) - n_j \mu_{j,\mathbf{w}})^2}{2\phi_{j,\theta}(\omega_0)}\right) \end{aligned} \quad (\text{D.7})$$

where $y_{\ell_j}(\omega_m)$ is the discrete Fourier transform (DFT) coefficient of $\{\log L(j, k)\}_{k=1}^{n_j}$ at frequency $\omega_m = 2\pi m/n_j$. A closed-form of the spectral density $\phi_{j,\theta}(\omega)$ associated with the 1D covariance function $\varrho_{j,\theta}(\Delta k)$ in (D.7) cannot be derived. It is here computed using a discrete Fourier transform as follows

$$\phi_{j,\theta}^D(\omega_m) = \sum_{k \in \llbracket -n_j, n_j \rrbracket} \varrho_{j,\theta}(\Delta k) \exp(-ik\omega_m). \quad (\text{D.8})$$

From (D.8), a decomposition analogous to (2.32) can be derived, which enables us to use the reparametrization (2.35) and the data augmentation scheme (2.56). This gives rise to a fast implementation of the estimation procedure introduced in [WDTA13] (see [CWA⁺15] for more details).

Moreover, this implies that the multivariate Bayesian methodology introduced in Chapter 3 directly carries over to time series. This has been reported in [CWT⁺16b, CWA⁺16a].

Appendix E

Expectation-maximization

The expectation-maximization (EM) algorithm is an iterative scheme that can be used for performing maximum likelihood or maximum a posterior estimation for augmented models. The EM algorithm generates a sequence of estimates that increases the marginal likelihood (posterior) at each iteration and converges to a local maximum [DLR77]. In this appendix we devise an EM algorithm for the approximation of the MAP estimator (2.79) based on the posterior (2.60) (cf. Chapter 2). Given an initial guess $\boldsymbol{\lambda}^{(0)}$, this EM algorithm iterates over two steps that are detailed below.

Expectation. The expectation step (E-step) consists in computing the expectation of the log-augmented posterior $\log p(\boldsymbol{\lambda}, \boldsymbol{\mu}|\mathbf{y})$ with respect to the conditional distribution of the latent variable $\boldsymbol{\mu}$ given the current estimate $\boldsymbol{\lambda}^{(p)}$, which reads here

$$\begin{aligned} Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^{(p)}) &= \mathbb{E}_{p(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\lambda}^{(p)})}[\log p(\boldsymbol{\lambda}, \boldsymbol{\mu}|\mathbf{y})] \\ &= \mathbb{E}_{p(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\lambda}^{(p)})}[\log p(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda})] + \log p(\boldsymbol{\lambda}). \end{aligned} \quad (\text{E.1})$$

As in detailed in the TSG algorithm, the conditional distribution $p(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\lambda}^{(p)})$ can be shown to be separable as

$$p(\boldsymbol{\mu}|\mathbf{y}, \boldsymbol{\lambda}^{(p)}) \propto p(\boldsymbol{\mu}_\emptyset|\mathbf{y}, \boldsymbol{\lambda}^{(p)}) \times p(\boldsymbol{\mu}_0|\mathbf{y}, \boldsymbol{\lambda}^{(p)}) \quad (\text{E.2})$$

with $p(\boldsymbol{\mu}_\emptyset|\mathbf{y}, \boldsymbol{\lambda}^{(p)})$ and $p(\boldsymbol{\mu}_0|\mathbf{y}, \boldsymbol{\lambda}^{(p)})$ respectively given by

$$p(\boldsymbol{\mu}_\emptyset|\mathbf{y}, \boldsymbol{\lambda}^{(p)}) = p^{(t)}(\boldsymbol{\mu}_\emptyset) \sim \mathcal{CN}(\tilde{\mathbf{x}}_{\mathbf{v}^{(p)}}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}^{(p)}}) \quad (\text{E.3})$$

and

$$p(\boldsymbol{\mu}_0|\mathbf{y}, \boldsymbol{\lambda}^{(p)}) = p^{(t)}(\boldsymbol{\mu}_0) \sim \mathcal{N}(\tilde{\mathbf{x}}_{0, \mathbf{v}^{(p)}, \mathbf{w}^{(p)}}, \tilde{\boldsymbol{\Sigma}}_{0, \mathbf{v}^{(p)}}) \quad (\text{E.4})$$

where the mean vector and the covariance matrix of the Gaussian distributions above are defined in (2.69), (2.73), (2.70) and (2.74). Moreover, since the distribution $p(\mathbf{y}, \boldsymbol{\mu} | \boldsymbol{\lambda})$ is from an exponential family, the expectation in (E.1) is straightforward to evaluate and boils down to computing the sufficient statistics of the latent variables $\boldsymbol{\mu}_\emptyset$ and $\boldsymbol{\mu}_0$ as follows

$$\begin{aligned} -\mathbb{E}_{p(\boldsymbol{\mu} | \mathbf{y}, \boldsymbol{\lambda}^{(p)})} [\log p(\mathbf{y}, \boldsymbol{\mu} | \boldsymbol{\lambda})] &= N_\emptyset \log v_1 + \frac{N_0}{2} \log v_1 + N_\emptyset \log v_2 + \frac{N_0}{2} \log v_2 \\ &\quad + \frac{1}{v_1} \mathbb{E}_{p^{(t)}(\boldsymbol{\mu}_\emptyset)} [(\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset)^H \tilde{\mathbf{F}}_1^{-1} (\mathbf{y}_\emptyset - \boldsymbol{\mu}_\emptyset)] + \frac{1}{v_2} \mathbb{E}_{p^{(t)}(\boldsymbol{\mu}_\emptyset)} [\boldsymbol{\mu}_\emptyset^H \tilde{\mathbf{F}}_2^{-1} \boldsymbol{\mu}_\emptyset] \\ &\quad + \frac{1}{2v_1} \mathbb{E}_{p^{(t)}(\boldsymbol{\mu}_0)} [(\mathbf{y}_0 - \boldsymbol{\mu}_0)^T \tilde{\mathbf{F}}_{0,1}^{-1} (\mathbf{y}_0 - \boldsymbol{\mu}_0)] + \frac{1}{2v_2} \mathbb{E}_{p^{(t)}(\boldsymbol{\mu}_0)} [(\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})^T \tilde{\mathbf{F}}_{0,2}^{-1} (\boldsymbol{\mu}_0 - \mathbf{X}\mathbf{w})]. \end{aligned} \quad (\text{E.5})$$

It comes that, up to an additive constant, $-Q(\mathbf{v}; \mathbf{v}^{(p)})$ can be developed into

$$\begin{aligned} -Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^{(p)}) &= (N_\emptyset + \frac{N_0}{2} + \alpha_1) \log v_1 + (N_\emptyset + \frac{N_0}{2} + \alpha_2) \log v_2 + \frac{1}{2} (\mathbf{w}_0 - \mathbf{w})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w}_0 - \mathbf{w}) \\ &\quad + \frac{\text{Tr}(\tilde{\mathbf{F}}_1^{-1} \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}^{(p)}}) + \|\mathbf{y}_\emptyset - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_1^{-1}}}{v_1} + \frac{\text{Tr}(\tilde{\mathbf{F}}_{0,1}^{-1} \tilde{\boldsymbol{\Sigma}}_{0,\mathbf{v}^{(p)}}) + \|\mathbf{y}_0 - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_{0,1}^{-1}}}{2v_1} \\ &\quad + \frac{\text{Tr}(\tilde{\mathbf{F}}_2^{-1} \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}^{(p)}}) + \|\tilde{\mathbf{x}}_{\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_2^{-1}}}{v_2} + \frac{\text{Tr}(\tilde{\mathbf{F}}_{0,2}^{-1} \tilde{\boldsymbol{\Sigma}}_{0,\mathbf{v}^{(p)}}) + \|\mathbf{X}\mathbf{w} - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_{0,2}^{-1}}}{2v_2} \end{aligned} \quad (\text{E.6})$$

where $\text{Tr}(\cdot)$ is the trace operator.

Maximization. The maximization step (M-step) requires to update the parameter estimate by maximizing the function $Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^{(p)})$, i.e.,

$$\boldsymbol{\lambda}^{(p+1)} = \underset{\boldsymbol{\lambda}}{\text{argmax}} Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^{(p)}). \quad (\text{E.7})$$

The estimate $\boldsymbol{\lambda}^{(p+1)}$ is found by canceling the gradient of the function Q , which yields the following set of equations

$$\mathbf{w}^{(p+1)} = \left(\frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X}}{v_2^{(p+1)}} + \boldsymbol{\Sigma}^{-1} \right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \mathbf{w}_0 + \frac{\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}}{v_2^{(p+1)}} \right) \quad (\text{E.8})$$

$$v_1^{(p+1)} = \frac{2(\text{Tr}(\tilde{\mathbf{F}}_1^{-1} \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}^{(p)}}) + \|\mathbf{y}_\emptyset - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_1^{-1}} + \beta_1) + \text{Tr}(\tilde{\mathbf{F}}_{0,1}^{-1} \tilde{\boldsymbol{\Sigma}}_{0,\mathbf{v}^{(p)}}) + \|\mathbf{y}_0 - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_{0,1}^{-1}}}{2(N_\emptyset + \alpha_1 + 1) + N_0} \quad (\text{E.9})$$

$$v_2^{(p+1)} = \frac{2(\text{Tr}(\tilde{\mathbf{F}}_2^{-1} \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}^{(p)}}) + \|\tilde{\mathbf{x}}_{\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_2^{-1}} + \beta_2) + \text{Tr}(\tilde{\mathbf{F}}_{0,2}^{-1} \tilde{\boldsymbol{\Sigma}}_{0,\mathbf{v}^{(p)}}) + \|\mathbf{X}\mathbf{w}^{(p+1)} - \tilde{\mathbf{x}}_{0,\mathbf{v}^{(p)}}\|_{\tilde{\mathbf{F}}_{0,2}^{-1}}}{2(N_\emptyset + \alpha_2 + 1) + N_0}. \quad (\text{E.10})$$

The expressions for $\mathbf{w}^{(p+1)}$ and $v_2^{(p+1)}$ are coupled and intricate, and solving these equations is not straightforward¹. However, when a non informative Gaussian prior $\mathcal{N}(\mathbf{w}_0, \mathbf{\Sigma})$ with $\mathbf{w}_0 = [0, 0]^T$, $\mathbf{\Sigma} = \text{diag}(\sigma^2, \sigma^2)$ and $\sigma^2 \gg 1$ is assigned to \mathbf{w} (as it is the case in the numerical experiment in Section 2.6), one can reasonably ignore the contribution of the prior in (E.8) and update $\mathbf{w}^{(p+1)}$ with

$$\mathbf{w}^{(p+1)} \approx (\mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{F}}_{0,2}^{-1} \tilde{\mathbf{x}}_{0,v^{(p)}}. \quad (\text{E.11})$$

Computational cost reduction. Preliminary experiments have shown that the proposed EM algorithm (using the approximation (E.11)) succeeded in providing MAP estimates numerically very close to those obtained with MCMC algorithms studied in Chapter 2. Moreover, it has been observed that it yielded a computational cost close to the one of the linear regression based procedure, cf. Fig. E.1.

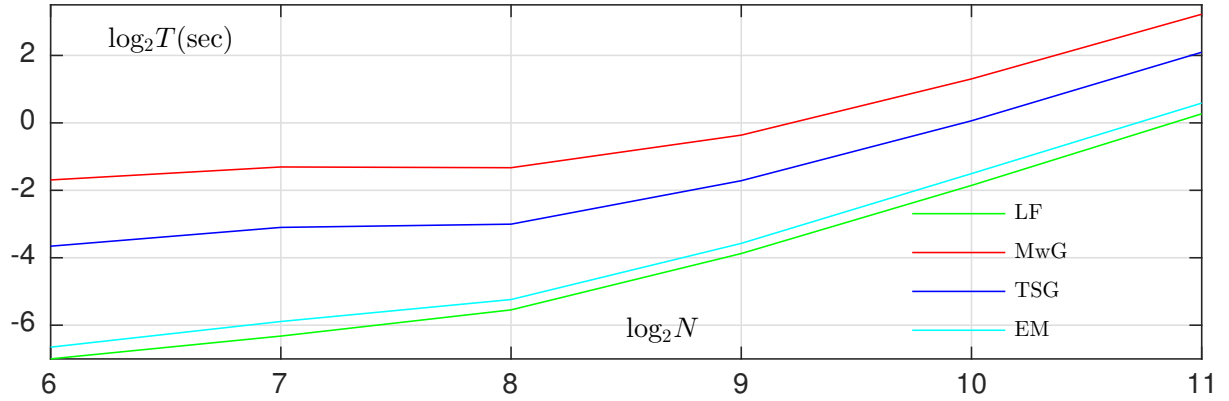


Figure E.1: Computational time T (in seconds) versus image size N with $j_1 = \{1, 1, 2, 2, 2\}$ and $j_2 = \log_2 N - 4$ for all methods (EM included).

¹When injecting (E.8) in (E.10), it can actually be shown that finding $v_2^{(p+1)}$ amounts to finding the roots of a fifth degree polynomial

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.
- [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. *J. Math. Phys.*, 39(8):4142–4164, 1998.
- [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 *Adv. Imag. Electr. Phys.*, volume 126, pages 1–98. Academic Press, 2003.
- [AFTV00] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88. 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.
- [AL95] V. V. Anh and K. E. Lunney. Parameter estimation of random fields with long-range dependence. *Math. Comput. Model.*, 21(9):67–77, 1995.
- [AMB02] R. J. Aspinall, W. A. Marcus, and J. W. Boardman. Considerations in collecting, processing, and analysing high spatial resolution hyperspectral data for environmental investigations. *Journal of Geographical Systems*, 4(1):15–29, 2002.
- [AMVA04] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali. *Two-Dimensional Wavelets and their Relatives*. Cambridge University Press, 2004.
- [APM15] Y. Altmann, M. Pereyra, and S. McLaughlin. Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis. *IEEE Trans. on Computational Imaging*, 1(3):174–185, 2015.
- [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.
- [BBB⁺03] J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian statistics*, 7:453–464, 2003.
- [BDM01] E. Bacry, J. Delour, and J. F. Muzy. Multifractal random walk. *Phys. Rev. E*, 64:026103, 2001.
- [Ber94] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, New York, 1994.
- [Bes75] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, pages 179–195, 1975.

- [BGJM11] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.
- [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.
- [BM02] J. Barral and B. Mandelbrot. Multifractal products of cylindrical pulses. *Probab. Theory Relat. Fields*, 124:409–430, 2002.
- [BM12] G. Blanchet and L. Moisan. An explicit sharpness index related to global phase coherence. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, pages 1065–1068, Kyoto, Japan, 2012.
- [BMK09] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian blind deconvolution using a total variation prior. *IEEE Trans. on Image Proces.*, 18(1):12–26, 2009.
- [BMR08] G. Blanchet, L. Moisan, and B. Rougé. Measuring the global phase coherence of an image. In *Proc. Int. Conf. on Image Processing (ICIP)*, pages 1176–1179, 2008.
- [BPC⁺11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [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. *J. Bone Miner. Res.*, 16(4):697–704, 2001.

- [BR12] D. R Brillinger and M. Rosenblatt. Asymptotic theory of estimates of k th-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.
- [BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [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.
- [CC96] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.
- [CE07] P. Campisi and K. Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2007.
- [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*, 68(3-4):387–400, 1993.
- [Cha97] R. E. Chandler. A spectral method for estimating parameters in rainfall models. *Bernoulli*, pages 301–322, 1997.
- [Cha07] P. Chainais. Infinitely divisible cascades to model the statistics of natural images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2105–2119, 2007.
- [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.

- [CRA05] P. Chainais, R. Riedi, and P. Abry. On non scale invariant infinitely divisible cascades. *IEEE Trans. Inform. Theory*, 51(3):1063–1083, 2005.
- [Cre15] N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.
- [CS99] M.-H. Chen and Q.-M. Shao. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999.
- [CVA⁺12] 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.
- [CVF⁺13] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu. Fast joint detection-estimation of evoked brain activity in event-related fmri using a variational approach. *IEEE Trans. Medical Imaging*, 32(5):821–837, 2013.
- [CWA⁺15] S. Combrexelle, H. Wendt, P. Abry, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret. A Bayesian approach for the joint estimation of the multifractality parameter and integral scale based on the whittle approximation. In *IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, Brisbane, Australia, April 2015.
- [CWA⁺16a] S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, and P. Abry. Bayesian estimation for the local assessment of the multifractality parameter of multivariate time series. In *Proc. European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, September 2016. to appear.
- [CWA⁺16b] S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, and P. Abry. A Bayesian framework for the multifractal analysis of images using data augmentation and a Whittle approximation. In *Proc. ICASSP*, Shanghai, China, 2016.
- [CWA⁺16c] S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, and P. Abry. Bayesian joint estimation of the multifractality parameter of image patches using

- gamma Markov random field priors. In *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Phoenix, AZ, USA, September 2016. to appear.
- [CWA⁺16d] S. Combrexelle, H. Wendt, Y. Altmann, J.-Y. Tournet, S. McLaughlin, and P. Abry. Multifractal analysis of multivariate images using gamma Markov random field priors. *To be submitted*, 2016.
- [CWD⁺15] S. Combrexelle, H. Wendt, N. Dobigeon, J.-Y. Tournet, 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. Tournet, 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. Tournet, N. Dobigeon, S. McLaughlin, and P. Abry. Estimation bayésienne locale du paramètre de multifractalité à l’aide d’un algorithme de Monte Carlo hamiltonien. In *Proc. GRETSI Symposium Signal and Image Processing*, Lyon, France, Sept. 2015.
- [CWT⁺15c] S. Combrexelle, H. Wendt, J.-Y. Tournet, S. McLaughlin, and P. Abry. Hyperspectral image analysis using multifractal attributes. In *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Tokyo, Japan, June 2015.
- [CWT⁺16a] S. Combrexelle, H. Wendt, J.-Y. Tournet, P. Abry, and S. McLaughlin. Bayesian multifractal analysis of multi-temporal images using smooth priors. In *Proc. IEEE Workshop Statistical Signal Proces. (SSP)*, Palma de Mallorca, Spain, June 2016.
- [CWT⁺16b] S. Combrexelle, H. Wendt, J.-Y. Tournet, Y. Altmann, S. McLaughlin, and P. Abry. A Bayesian approach for the multifractal analysis of spatio-temporal data. In *Proc.*

- Int. Conf. Systems, Signals and Image Proces. (IWSSIP)*, Bratislava, Slovakia, May 2016.
- [DC10] O. Dikmen and A.T. Cemgil. Gamma Markov random fields for audio source modeling. *IEEE Trans. Audio, Speech, and Language Proces.*, 18(3):589–601, March 2010.
- [DHS12] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.
- [DKPR87] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222, 1987.
- [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.
- [Don08] P. Dong. Fractal signatures for multiscale processing of hyperspectral image data. *Adv. Space Res.*, 41(11):1733 – 1743, 2008.
- [DOT03] P. Doukhan, G. Oppenheim, and M. S. Taqqu. *Theory and Applications of Long-Range Dependence*. Birkhäuser Boston, 2003.
- [EBDT13] O. Eches, J. A. Benediktsson, N. Dobigeon, and J.-Y. Tourneret. Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images. *IEEE Trans. on Image Proces.*, 22(1):5–16, 2013.

- [EC03] L. E. Eberly and G. Casella. Estimating Bayesian credible intervals. *Journal of statistical planning and inference*, 112(1):115–132, 2003.
- [FGK14] E. Foufoula-Georgiou and P. Kumar. *Wavelets in geophysics*, volume 4. Academic Press, 2014.
- [Fig05] M. A. T. Figueiredo. Bayesian image segmentation using Gaussian field priors. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 74–89. Springer, 2005.
- [Fla92] P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motions. *IEEE Trans. Inform. Theory*, 38:910–917, 1992.
- [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.
- [Fri95] U. Frisch. *Turbulence: the legacy of A.N. Kolmogorov*. Cambridge University Press, 1995.
- [FTB⁺13] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE*, 101(3):652–675, 2013.
- [Fue07] M. Fuentes. Approximate likelihood for large irregularly spaced spatial data. *J. Am. Statist. Assoc.*, 102:321–331, 2007.
- [GC11] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.
- [GDGF10] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

- [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.
- [GG84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, (6):721–741, 1984.
- [GMW81] P. E. Gill, W. Murray, and M. H. Wright. Practical optimization. 1981.
- [Goo63] N. R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Stat.*, 34(1):pp. 152–177, 1963.
- [Gri10] J. Grimmer. An introduction to Bayesian inference via variational approximations. *Political Analysis*, page mpq027, 2010.
- [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.
- [GVB07] H. W. Gómez, O. Venegas, and H. Bolfarine. Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18(4):395–407, 2007.
- [Har79] R. M. Haralick. Statistical and structural approaches to texture. *Proc. of the IEEE*, 67(5):786–804, 1979.
- [Hin02] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [HJK⁺86] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A*, 33:1141–1151, Feb 1986.

- [HMTS16] A. Halimi, C. Mailhes, J.-Y. Tourneret, and H. Snoussi. Bayesian estimation of smooth altimetric parameters: Application to conventional and delay/Doppler altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4):2207–2219, 2016.
- [HPP⁺96] J. M. Hausdorff, P. L. Purdon, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol. (1985)*, 80(5):1448–57, 1996.
- [Jaf97a] S. Jaffard. Multifractal formalism for functions part I: results valid for all functions. *SIAM J. Math. Anal.*, 28(4):944–970, 1997.
- [Jaf97b] S. Jaffard. Multifractal formalism for functions part II: self-similar functions. *SIAM Mathe. Anal.*, 28(4):971–998, 1997.
- [Jaf04] S. Jaffard. Wavelet techniques in multifractal analysis. In M. Lapidus and M. van Frankenhuysen, editors, *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, Proc. Symp. Pure Math.*, volume 72(2), pages 91–152. AMS, 2004.
- [JAW15] S. Jaffard, P. Abry, and H. Wendt. Irregularities and scaling in signal and image processing: Multifractal analysis. In M. Frame and N. Cohen, editors, *Benoit Mandelbrot: A Life in Many Dimensions*, pages 31–116. World scientific publishing, Singapore, 2015.
- [Jes12] J. Jesus. *Contributions to inference without likelihoods*. PhD thesis, UCL (University College London), 2012.
- [JLA06] S. Jaffard, B. Lashermes, and P. Abry. Wavelet leaders in multifractal analysis. In T. Qian, M. I. Vai, and X. Yuesheng, editors, *Wavelet Analysis and Applications*, page 219264. Cambridge, 2006.
- [JMS⁺14] C. R. Johnson, P. Messier, W.A. Sethares, A.G. Klein, C. Brown, A.H. Do, P. Klausmeyer, P. Abry, S. Jaffard, H. Wendt, S. Roux, N. Pustelnik, N. van Noord, L. van der Maaten, E. Potsma, J. Coddington, L.A. Daffner, H. Murata, H. Wilhelm, S. Wood,

- and M. Messier. Pursuing automated classification of historic photographic papers from raking light photomicrographs. *J. Amer. Inst. Conserv.*, 53(3):159–170, 2014.
- [JSM06] K. Jones-Smith and H. Mathur. Fractal analysis: Revisiting Pollock’s drip paintings. *Nature*, 444(7119):E9–E10, 2006.
- [KCC89] J. M. Keller, S. Chen, and R. M. Crownover. Texture description and segmentation through fractal geometry. *Comp. Vis., Graphics, and Image Proces.*, 45(2):150–166, 1989.
- [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.
- [Kol41] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. In *Dokl. Akad. Nauk SSSR*, volume 30, pages 301–305. JSTOR, 1941.
- [KP76] J.-P. Kahane and J. Peyriere. Sur certaines martingales de Benoît Mandelbrot. *Advances in mathematics*, 22(2):131–145, 1976.
- [KZKB⁺02] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1):87–114, 2002.
- [Lam62] J. Lamperti. Semi-stable stochastic processes. *Transactions of the American mathematical Society*, 104(1):62–78, 1962.
- [LB09] R. Lopes and N. Betrouni. Fractal and multifractal analysis: A review. *Medical Image Analysis*, 13:634–649, 2009.
- [LdABD⁺15] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J.-Y. Tourneret, M. A. Veganzones,

- G. Vivone, Q. Wei, and N. Yokoya. Hyperspectral pansharpening: A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46, Sept 2015.
- [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.
- [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.
- [LT05] S. B. Lowen and M. C. Teich. *Fractal-based point processes*. Wiley, 2005.
- [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. *J. Business & Economic Stat.*, 26:194–210, 2007.
- [Lux08] T. Lux. The Markov-switching multifractal model of asset returns. *J. Business & Economic Stat.*, 26(2):194–210, 2008.

- [LVMB92] J. Lévy-Véhel, P. Mignot, and J. Berroir. Multifractals, texture and image analysis. In *Proc. IEEE Conf. Comp. Vis. Pattern Recognition (CVPR)*, pages 661–664, Champaign, IL, USA, June 1992.
- [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*, 448:319–339, 2016.
- [Mal08] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition, 2008.
- [Man74] B. B. Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *J. Fluid Mech.*, 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.
- [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. *Physical review E*, 47(2):875, 1993.
- [MIV⁺08] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, 2008.
- [MLA⁺03] W. A. Marcus, C. J. Legleiter, R. J. Aspinall, J. W. Boardman, and R. L. Crabtree. High spatial resolution hyperspectral mapping of in-stream habitats, depths, and woody debris in mountain streams. *Geomorphology*, 55(1):363–380, 2003.
- [MMK06] R. Molina, J. Mateos, and A. K. Katsaggelos. Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans. Image Proces.*, 15(12):3715–3727, 2006.

- [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.
- [MSKF90] C. Meneveau, K.R. Sreenivasan, P. Kailasnath, and M. S. Fan. Joint multifractal measures: Theory and applications to turbulence. *Phys. Rev. A*, 41(2):894, 1990.
- [Nea10] R. M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162, 2010.
- [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.
- [OW00] M. Ossiander and E. C. Waymire. Statistical estimation for multiplicative cascades. *Ann. Stat.*, 28(6):1533–1560, 2000.
- [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.
- [PDBT13] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret. Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm. *IEEE Trans. Image Proces.*, 22(6):2385–2397, 2013.
- [PMdJF11] J. M. Prats-Montalbán, A. de Juan, and A. Ferrer. Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, 107(1):1–23, 2011.

- [Pou99] A. D. Poularikas. The Hankel transform. In A. D. Poularikas, editor, *The Handbook of formulas and Tables for Signal Processing*. CRC Press LLC, 1999.
- [PP⁺08] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.
- [PPLV02] B. Pesquet-Popescu and J. Lévy-Véhel. Stochastic fractal models for image processing. *IEEE Signal Proces. Mag.*, 19(5):48–62, 2002.
- [Pri81] M. B. Priestley. *Spectral analysis and time series*. Academic press, London, 1981.
- [PWAT14] M. Pereyra, N. Whiteley, C. Andrieu, and J.-Y. Tourneret. Maximum marginal likelihood estimation of the granularity coefficient of a Potts-Markov random field within an MCMC algorithm. In *Proc. IEEE Workshop on Statistical Signal Proces. (SSP)*, pages 121–124, June 2014.
- [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. *Eur. Phys. J. B*, 15(4):765–786, 2000.
- [RC96] T. Subba Rao and R.E. Chandler. A frequency domain approach for estimating parameters in point process models. In *Athens Conf. Applied Probability and Time Series Analysis*, pages 392–405. Springer, 1996.
- [RC98] C. P. Robert and D. Cellier. Convergence control of MCMC algorithms. In *Discretization and MCMC Convergence Assessment*, pages 27–46. Springer, 1998.
- [RC05] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, USA, 2005.

- [RICV09] L. Risser, J. Idier, P. Ciuciu, and T. Vincent. Fast bilinear extrapolation of 3D Ising field partition function, application to fMRI image analysis. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 833–836. IEEE, 2009.
- [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.
- [Rob95] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Stat.*, 23(5):1630–1661, 1995.
- [RV10] R. Robert and V. Vargas. Gaussian multiplicative chaos revisited. *Ann. Proba.*, 38(2):605–631, 2010.
- [RVF⁺11] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. Min-max extrapolation scheme for fast estimation of 3D potts field partition functions. application to the joint detection-estimation of brain activity in fMRI. *J. Signal Proces. Syst.*, 65(3):325–338, 2011.
- [SC03] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003.
- [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.
- [SS01] F. G. Schmitt and L. Seuront. Multifractal random walk in copepod behavior. *Physica A*, 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.

- [SSLB96] F. Schmitt, D. Schertzer, S. Lovejoy, and Y. Brunet. Multifractal temperature and flux of temperature variance in fully developed turbulence. *Europhys. Lett.*, 34(3):195, 1996.
- [ST94] G. Samoradnitsky and M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC press, 1994.
- [Ste02] M. L. Stein. Fast and exact simulation of fractional Brownian surfaces. *J. Comput. Graph. Statist.*, 11(3):587–599, 2002.
- [SUS14] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson. Hyperspectral unmixing with regularization. *IEEE Trans. Geoscience and Remote Sensing*, 52(11):6793–6806, 2014.
- [SXGL06] W. Sun, G. Xu, P. Gong, and S. Liang. Fractal analysis of remotely sensed images: A review of methods and applications. *Int. J. Remote Sens.*, 27:4963–4990, 2006.
- [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.
- [Văc14] C. P. Văcar. *Inversion for textured images : unsupervised myopic deconvolution, model selection, deconvolution-segmentation*. Theses, Université de Bordeaux, Sept. 2014.
- [VAT03] D. Veitch, P. Abry, and M. S. Taqqu. On the automatic selection of the onset of scales. *Fractals*, 11(04):377–390, 2003.
- [VCG08] T. Virtanen, A. T. Cemgil, and S. Godsill. Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In *IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP)*, pages 1825–1828. IEEE, 2008.
- [VR00] C. Velasco and P. M. Robinson. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Am. Statist. Assoc.*, 95(452):1229–1243, 2000.
- [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.
- [WDTA13] H. Wendt, N. Dobigeon, J.-Y. Tournet, and P. Abry. Bayesian estimation for the multifractality parameter. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, 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.
- [Whi54] P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–449, 1954.
- [WJA12] H. Wendt, S. Jaffard, and P. Abry. Multifractal analysis of self-similar processes. In *Proc. IEEE Workshop on Statistical Signal Proces. (SSP)*, pages 69–72, Ann Arbor, MI, USA, Aug. 2012.
- [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.
- [WT11] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688, 2011.
- [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.
- [Yag66] A. M. Yaglom. Effect of fluctuations in energy dissipation rate on the form of turbulence characteristics in the inertial subrange. *Dokl. Akad. Nauk. SSR*, 166:49–52, 1966.
- [YGJ12] J. Yin, C. Gao, and X. Jia. Using Hurst and Lyapunov exponent for hyperspectral image feature extraction. *IEEE Geoscience and remote sensing letters*, 9(4):705–709, 2012.
- [ZTPP10] M. Zortea, D. Tuia, F. Pacifici, and A. Plaza. Spectral-textural endmember extraction. In *Proc. IEEE Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Reykjavik, Iceland, June 2010.