Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: [http://oatao.univ-toulouse.fr/](http://oatao.univ-toulouse.fr/)

Eprints ID: 17036

The contribution was presented at IWSSIP 2016:

[http://iwssip.stuba.sk/](http://iwssip.stuba.sk/)


Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr
A Bayesian Approach for the Multifractal Analysis of Spatio-Temporal Data

S. Combreselle, H. Wendt, J.-Y. Tourneret
IRIT - ENSEEIHT, CNRS,
Univ. of Toulouse, France
firstname.lastname@irit.fr

Y. Altmann, S. McLaughlin
School of Engineering, Heriot-Watt Univ.,
Edinburgh, Scotland
initial.lastname@hw.ac.uk

P. Abry
CNRS, Physics Dept.,
ENS Lyon, France
patrice.abry@ens-lyon.fr

Abstract—Multifractal (MF) analysis enables the theoretical study of scale invariance models and their practical assessment via wavelet leaders. Yet, the accurate estimation of MF parameters remains a challenging task. For a range of applications, notably biomedical, the performance can potentially be improved by taking advantage of the multivariate nature of data. However, this has barely been considered in the context of MF analysis. This paper proposes a Bayesian model that enables the joint estimation of MF parameters for multivariate time series. It builds on a recently introduced statistical model for leaders and is formulated using a 3D gamma Markov random field joint prior for the MF parameters of the voxels of spatio-temporal data, represented as a multivariate time series, that counteracts the statistical variability induced by small sample size. Numerical simulations indicate that the proposed Bayesian estimator significantly outperforms current state-of-the-art algorithms.

I. CONTEXT, RELATED WORK AND CONTRIBUTIONS

Context. Multifractal analysis is a widely used signal processing tool and enables the study of the scale invariance properties of data. It has been successfully used in a large variety of applications, ranging from biomedical [1], [2], physics [3] to finance [4] and Internet [5], cf., e.g., [6] for a review. Scale invariance implies that the dynamics of a time series $X(t)$ are driven by a large continuum of time scales instead of only a few characteristic scales. This translates into power law behaviors of the time averages of well chosen multisolution quantities $T_X(j, k)$ of $X$ (i.e., quantities that depend jointly on scale $2^j$ and time instance $k$) over a large range of scales $2^j$

$$S(q, j) = \frac{1}{n_j} \sum_k |T_X(j, k)|^q \approx (2^j)^\zeta(q), \quad 2^j_1 \leq 2^j \leq 2^j_2$$

where $n_j$ is the number of $T_X(j, k)$ at scale $j$. In this work, wavelet leaders $l(j, k)$ are used as multisolution quantities, which can be shown to be well suited for this purpose [6], [7] (and are defined in Section II). Given a time series $X$, the goal is the estimation of the so-called scaling exponents $\zeta(q)$ of the power law in (1), which fully characterize the scale invariance properties of $X$. Notably, the scaling exponents permit discrimination between the two fundamental classes of scale invariance models: self-similar processes, which are obtained by additive construction mechanisms and characterized by a linear function $\zeta(q) = qH$ [8]; multifractal multiplicative cascades (MMC), which have a multiplicative structure and yield a strictly concave function $\zeta(q)$ [4]. In order to understand the construction mechanisms underlying data, it is crucial to decide which model better fits the data in applications. This decision can be reached by considering the polynomial development $\zeta(q) = \sum_{m \geq 1} c_m q^m / m!$ at $q = 0$. It can be shown that the coefficient $c_2$, called the multifractality parameter, is strictly negative for MMC but equals zero for self-similar processes, cf., e.g., [7], [9]. Therefore, the estimation of $c_2$ is central in multifractal analysis.

Estimation of $c_2$. The multifractality parameter $c_2$ can be directly linked to the variance of the logarithm of $l(j, k)$ [9]

$$C_2(j) = \text{Var} \ln l(j, \cdot) = c_0^2 + c_2 \ln 2^j.$$  

This motivates estimation of $c_2$ as a linear regression of the sample variance $\text{Var} \ln l(j, \cdot)$ of log-leaders with respect to scale $j$

$$\hat{c}_2 = (\ln 2)^{-1} \sum_{j=j_1}^{j_2} w_j \text{Var} \ln l(j, \cdot)$$

where $w_j$ are appropriate regression weights [7]. The estimator (3) is widely used but is known to yield poor performance for small sample size [10], [11]. Alternative estimators have been described in, e.g., [12], [13], but they make assumptions, (e.g. fully parametric model, specific multifractal process), that are often too restrictive in real-world applications. More recently, Bayesian estimators for $c_2$ have been proposed in [10], [11]. Their advantage lies in the use of a semi-parametric model for the statistics of the log-leaders that is generically valid for MMC processes and induces considerable performance gains when compared to (3). These gains were obtained at the price of increased computational cost since the Bayesian inference was achieved by a Markov chain Monte Carlo (MCMC) algorithm with a Metropolis-Hastings within Gibbs (MHG) sampler. A significantly more efficient algorithm was obtained very recently in [14] by considering a data augmented formulation of the Bayesian model. None of these developments addressed the estimation of $c_2$ for multivariate data.

Contributions. This paper devises a Bayesian procedure for the joint estimation of $c_2$ associated with multivariate time series registered on a volume (voxels) that makes use of the dependence of neighboring voxels in order to improve estimation accuracy. The algorithm combines the statistical model introduced in [10], [11] with the data augmentation strategy proposed for images in [14] (summarized in Section II). The
key contribution (described in Section III) resides in the design of an appropriate joint prior for the multifractality parameters for voxels. It is characterized by its number of vanishing moments and is defined as the largest wavelet coefficient within $\Delta k$ and where $c_2$ is associated with voxels of multivariate time series (cf., Section IV).

II. STATISTICAL MODEL FOR LOG-LEADERS

A. Direct statistical model in the time-domain

Wavelet leaders. A mother wavelet $\psi_0(t)$ is a reference pattern that has narrow supports in the time and frequency domains. It is chosen so that the collection $\{\psi_{j,k}(t) \equiv 2^{-j/2}\psi_0(2^{-j}t-k), j, k \in \mathbb{N}\}$ forms a basis of $L^2(\mathbb{R})$ and is characterized by its number of vanishing moments $N_0 \geq 1$ $(\forall k = 0, 1, \ldots, N_0 - 1, \int_k 2^j \psi_0(t)dt \equiv 0$ and $\int_k 2^j \psi_0(t)dt \neq 0$. The ($L^1$-normalized) discrete wavelet transform coefficients of $X$ are defined as $d_X(j,k) \equiv \langle X, 2^{-j/2}\psi_{j,k}\rangle$, cf., [16] for details. Let $\lambda_{j,k} = [k2^j, (k+1)/2^j)$ denote the dyadic interval of size $2^j$ and $3\lambda_{j,k}$ the union of $\lambda_{j,k}$ with its 2 neighbors. The wavelet leaders are defined as the largest wavelet coefficient within $3\lambda_{j,k}$ over all finer scales [6, 7]

$$l(j,k) = \sup_{X \in 3\lambda_{j,k}} |d_X(X)|.$$  (4)

Statistical model. Denote as $\ell_j$ the vector of the log-leaders $\ell(j,\cdot) \equiv \ln l(j,\cdot)$ at scale $j$ after mean subtraction (since it conveys no information on $c_2$) and $\ell \equiv [\ell_{1T}, \ldots, \ell_{2T}]^T$. The statistics of $\ell_j$ of MMC based processes can be well approximated by a multivariate Gaussian distribution whose covariance $C_j((k, \Delta k) \equiv \text{Cov}(\ell(j,k), \ell(j,k+\Delta k))$ is [10]

$$C_j(k, \Delta k) \approx \delta_j^0(\Delta k; \theta) + \delta_j^1(\Delta k; \theta)$$  (5)

where $\theta = (c_2, c_2^0)$, $\delta_j^0(r; \theta) \equiv c_2 \ln |4|/|n_{ij}| |I_{(3, n_{ij}/4)}(r)|$, $\delta_j^1(r; \theta) \equiv (\ln n_{ij}) |I_{(3, \theta)} - c_2 \ln 2^j| + c_2^0 + c_2^0 ln 2^j |I_{(0, 0)}(r)|$ and $I_{(\cdot, \cdot)}$ is the indicator function of the set $A$. Assuming independence between $\ell_j$ at different scales $j$ leads to the following likelihood for $\ell$

$$p(\ell|\theta) = \prod_{j=1}^{2^j} |\Sigma_{j,\theta}|^{-1/2} \exp \left(-\frac{1}{2} \ell^T \Sigma_{j,\theta}^{-1} \ell \right)$$  (6)

where the matrices $\Sigma_{j,\theta}$ are defined element-wise by (5), $\cdot$ denotes the determinant and $T$ the transpose operator.

Whittle approximation. The likelihood (6) is problematic to evaluate numerically since it requires the computation of the matrix inverses $\Sigma_{j,\theta}^{-1}$. Thus, it has been proposed in [11] to approximate (6) with the asymptotic Whittle likelihood [17]

$$p_W(\ell|\theta) = |\Gamma_{\theta}|^{-1} \exp \left(-\frac{1}{2} \ell^T \Gamma_{\theta}^{-1} \ell \right),$$  (7)

$$y \triangleq [y_{1T}, \ldots, y_{2T}]^T,$$  (8)

where $\Gamma_{\theta} \triangleq c_2^2 F + c_2^2 G$ is an $N_Y \times N_Y$ diagonal covariance matrix, with $N_Y = \text{card}(y)$, $F \triangleq \text{diag}(y)$, $G \triangleq |I_{(\cdot, \cdot)}|$. The diagonal elements of $\Gamma_{\theta}$ correspond to the discretized spectral densities $c_2 f_{j}(m) + c_2^0 g_{j}(m)$ associated with the model (5), for the positive frequencies $\omega_m = 2\pi m/\sqrt{T}$, $m \in \mathbb{N}^+$. Here, $y_j \equiv \mathcal{F}(\ell_j)$ is the periodogram of $\ell_j$, where the operator $\mathcal{F}(\cdot)$ computes and vectorizes the discrete Fourier transform coefficients for $\omega_m$, $m \in \mathbb{N}^+$ and $H$ is the conjugate transpose operator. Note that $f_{j}$ and $g_{j}$ do not depend on $\theta$ and can be precomputed (and stored) using the fast Fourier transform.

B. Data augmented statistical model in the Fourier domain

The parameters $\theta$ are encoded in $\Sigma_{j,\theta}$, and their conditional distributions are not standard. Sampling the posterior distribution with an MCMC method would hence require accept/reject procedures [10, 11]. A more efficient algorithm can be obtained by interpreting (7) as a statistical model for the Fourier coefficients $y$ [14]. Assuming that $\Gamma_{\theta}$ is positive definite, (7) amounts to modeling $y$ by a random vector with a centered circular-symmetric complex Gaussian distribution $\mathcal{CN}(0, \Gamma_{\theta})$, hence to the use of the likelihood

$$p(y|\theta) = |\Gamma_{\theta}|^{-1} \exp \left(-\frac{1}{2} y^H \Gamma_{\theta}^{-1} y \right).$$  (9)

Reparametrization. The matrix $\Gamma_{\theta}$ is positive definite as long as the parameters $\theta = (c_2, c_2^0)$ belong to the admissible set

$$A = \{ \theta \in \mathbb{R}_+^2 | c_2^0 f(m) + c_2^2 g(m) > 0, m = 1, \ldots, N_Y \}.$$  (10)

Since $\forall m, c_2^0 g(m) > 0$ (while $c_2 f(m) < 0$), (9) can be mapped onto independent positivity constraints by a one-to-one transformation from $\theta \in A$ to $\nu \in \mathbb{R}_+^2$ defined as $\theta \mapsto \nu = (v_1, v_2) \equiv (-c_2^0 f(m)/\gamma + c_2)$, where $\gamma = \sup_{m} f(m)/g(m)$ [14]. Consequently, (8) can be expressed using $\nu \in \mathbb{R}_+^2$ as

$$p(y|\nu) \propto |\Gamma_{\nu}|^{-1} \exp \left(-\nu^H \Gamma_{\nu}^{-1} y \right).$$  (11)

Data augmentation. One can now introduce an $N_Y \times 1$ vector of latent variables $\mu$ that enables us to augment (10) using the model $y|\nu, v_2 \sim \mathcal{CN}(\nu, v_2 \tilde{G}), \nu|v_1 \sim \mathcal{CN}(0, v_1 \tilde{F})$, which is associated with the extended likelihood [14]

$$p(y, \mu|\nu, v_1 \sim v_2^{-N_Y} \exp \left(-\nu^{-1} (y - \mu)^H \tilde{G}^{-1} (y - \mu) \right) \times v_1^{-N_Y} \exp \left(-\nu^{-1} \mu^H \tilde{F}^{-1} \mu \right).$$  (12)

Simple calculations show that (11) leads to standard conditional distributions when inverse-gamma priors are used for $v_i$, $i = 1, 2$, and that (10) is recovered by marginalization of (11) with respect to $\mu$. 


III. BAYESIAN MODEL FOR MULTIVARIATE TIME SERIES

A. Likelihood

Based on the likelihood (11) for one single time series $X(t)$, we now design a joint Bayesian model for the analysis of multivariate time series. Let $X_m$, $m \triangleq (m_1,m_2,m_3)$, $m_d = 1,\ldots,M_d$, denote $M_1 \times M_2 \times M_3$ discrete time series (voxels) of length $N$ (as illustrated in Fig. 1). Denote as $y_m$, $\mu_m$ and $\nu_m$ the Fourier coefficients, latent variables and parameter vector associated with $X_m$ and as $Y \triangleq \{y_m\}$, $M \triangleq \{\mu_m\}$, and $V \triangleq \{V_1, V_2\}$ (where $V_i \triangleq \{v_i,m\}$, $i = 1,2$) the corresponding collections for all voxels $X_m$. Assuming independence between the vectors $y_m$, the joint likelihood of $Y$ can be written as

$$p(Y, M | V) \propto \prod_m p(y_m, \mu_m | v_m).$$

(12)

B. Gamma Markov random field prior

Inverse-gamma distributions $IG(\alpha_{i,m},\beta_{i,m})$ are conjugate priors for the parameters $v_i,m$ in (12), and we propose to specify $(\alpha_{i,m},\beta_{i,m})$ such that the resulting prior for $V_i$ is a hidden GMRF [15]. A GMRF makes use of a set of positive priors for the parameters $\beta_{i,m}$, and $V \triangleq \{V_1, V_2\}$ (where $V_i \triangleq \{v_i,m\}$, $i = 1,2$) the corresponding collections for all voxels $X_m$. Assuming independence between the vectors $y_m$, the joint likelihood of $Y$ can be written as

$$p(Y, M | V) \propto \prod_m p(y_m, \mu_m | v_m).$$

(12)

C. Posterior distribution and Bayesian estimators

Under the assumption of prior independence between $(V_1, Z_1)$ and $(V_2, M, Z_2)$, the joint posterior distribution associated with the proposed model is obtained as

$$p(V, Z, M | Y, \rho_1, \rho_2) \propto p(Y | V_2, M) p(M | V_1) \times p(V_1, Z_1 | \rho_1) p(V_2, Z_2 | \rho_2)$$

(14)

using Bayes’ theorem. To infer the parameters of interest $V_i$, we consider the marginal posterior mean (minimum mean square error) estimator, denoted MMSE, which is defined as

$$V_i^{\text{MMSE}} \triangleq \mathbb{E}[V_i | Y, \rho_1],$$

where the expectation is taken with respect to the marginal posterior density $p(V_i | Y, \rho_1)$. The direct computation of $V_i^{\text{MMSE}}$ is not tractable since it requires integrating the posterior (14) over the variables $Z$ and $M$ and computation of the expectation. Instead, by considering a Gibbs sampler (GS) generating samples $(|V_i^{(q)}|, M^{(q)}, |Z_i^{(q)}|)_{q=1}^{N_{mc}}$ that are asymptotically distributed according to (14), it can be approximated as [18]

$$V_i^{\text{MMSE}} \approx (N_{mc} - N_{bi})^{-1} \sum_{q=N_{bi}}^{N_{mc}} V_i^{(q)}$$

(15)

TABLE I

<table>
<thead>
<tr>
<th>Estimation performance for 100 independent realizations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>101.26</td>
</tr>
<tr>
<td>0.0819</td>
</tr>
</tbody>
</table>

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

D. Gibbs sampler

Here, the GS consists of successively drawing samples from the conditional distributions that are associated with the posterior (14) [18]. Simple calculations lead to

$$p(\mu | Y, V, \mu_0) \sim CN\left(\mu_0, \sum_{v \in V} \mu_0 + 1 \right)$$

(16a)

$$p(\mu | Y, M, Z_i) \sim IG(N_0 + \alpha_1, \beta_1 + \beta_i)$$

(16b)

$$p(\mu | Z_i) \sim IG(\alpha_1, \gamma_i)$$

(16c)

where the subscript $m$ has been omitted for notational convenience and where $\sum_{l} = ||\mu||_{\ell^1}, \sum_{l} = ||y\mu||_{\ell^1}$ with $||x||_{\ell^1} \triangleq \sum_{\ell} x_{\ell}$. All conditionals (16a–16c) are standard laws that can be used efficiently, without MHG steps. Finally, note that when the parameters $v_i,m$ are assumed to be independent and have $IG(c_i,d_i)$ priors instead of (13) (i.e., no smooth spatial evolution is assumed), a Bayesian model is obtained that can also be sampled using the GS steps (16a–16b), with $\alpha_{i,m} = c_i$ and $\beta_{i,m} = d_i$.

IV. NUMERICAL EXPERIMENTS

We compare the performance of the proposed estimator (denoted as GMRF) with its counterpart with an IG prior (denoted as IG) and with the linear regression estimator (3) (denoted as LF, with weights as in [7]) by applying it to 100 independent realizations of a cube of $32^3$ voxels of length $N = 512$. Each voxel is an independent realization of MRW, with prescribed values $c_2 \in \{-0.01, -0.03, -0.06\}$ (indicated as green, yellow, dark blue, respectively); the slices correspond to those analyzed in Fig. 2. Single realizations of time series corresponding to 3 voxels with different value of $c_2$ (right panel).

![Fig. 1. Illustration of the cube of 32 x 32 x 32 voxels of time series (left panel) with prescribed multifractal properties $c_2 \in \{-0.01, -0.03, -0.06\}$ (indicated as green, yellow, dark blue, respectively); the slices correspond to those analyzed in Fig. 2. Single realizations of time series corresponding to 3 voxels with different value of $c_2$ (right panel).](image-url)
(H = 0.72 for the results presented below). The regularization parameters have been fixed a priori using cross-validation.

**Illustration for a single realization.** Fig. 2 displays estimates obtained for one single realization using LF, IG and GMRF (2nd to 4th row, respectively) for the slices $z = 16$, $z = 10$ and $z = 23$ shown in Fig. 1 and yields the following conclusions. First, the LF estimator completely fails to reveal the existence of two zones of voxels with constant $c_2 \in \{-0.03, -0.06\}$ in the background of voxels with $c_2 = -0.01$. The IG estimator improves estimation accuracy with respect to the LF such that the three groups of voxels can be evidenced visually, but its variability is too large to permit accurate identification of the voxels sharing the same value for $c_2$. In contrast, the GMRF estimator yields excellent estimates that accurately capture the geometry of the three zones of voxels and the corresponding values for $c_2$.

**Performance.** The estimation performance for $c_2$ is quantified via the bias, standard deviation and root mean squared error defined as $b = E[c_2] - c_2$, $\text{std} = (\text{Var}(c_2))^{1/2}$ and $\text{rmse} = (b^2 + \text{std}^2)^{1/2}$, respectively. The results are given in Table I and confirm the above conclusions. While IG reduces std values to $\frac{1}{3}$ of those of LF, GMRF further and dramatically reduces std values to $\frac{1}{10}$ of those of LF, which clearly demonstrates the effectiveness of the proposed joint Bayesian model. The bias is found to be small but non-negligible for all three methods (largest for LF and smallest for IG). As a result, gains in rmse values for GMRF are smaller but still significant (rmse values of GMRF are one order of magnitude below those of LF).

**V. CONCLUSIONS**

This paper has proposed a Bayesian procedure for the joint estimation of $c_2$ for spatio-temporal data (voxels). The Bayesian model is composed of a data-augmented Whittle likelihood for log-leaders and a GMRF joint prior for the multifractality parameters for voxels and yields a significant improvement in estimation performance for each voxel (by more than one order of magnitude when compared to linear regression). Moreover, it is designed in such a way that the associated estimators can be approximated efficiently by an MCMC algorithm. The proposed joint estimator enables, for the first time, the reliable assessment of small differences of $c_2$ for voxels with as little as $N = 512$ samples. Future work will include incorporation of the regularization parameters $\rho_k$ in the model and allowing them to differ for each voxel in order to permit abrupt changes in the data while maintaining strong smoothing within zones of voxels with constant $c_2$.

**REFERENCES**