Hurst multimodality detection based on large wavelet random matrices

Oliver Orejola, Gustavo Didier
Mathematics Department,
Tulane University, New Orleans, USA
{oorejola,gdidier}@tulane.edu

Patrice Abry
Université de Lyon, ENS de Lyon
CNRS, Lab. de Physique, Lyon (FR)
patrice.abry@ens-lyon.fr

Herwig Wendt
Université de Toulouse
CNRS, IRIT, Toulouse (FR)
herwig.wendt@irit.fr

***Abstract*—In the modern world, systems are routinely monitored by multiple sensors, generating "Big Data" in the form of a large collection of time series. In this paper, we put forward a statistical methodology for detecting multimodality in the distribution of Hurst exponents in high-dimensional fractal systems. The methodology relies on the analysis of the distribution of the log-eigenvalues of large wavelet random matrices. Depending on the presence of a single or many Hurst exponents, we show that the wavelet empirical log-spectral distribution displays one or many modes, respectively, in the threefold limit as dimension, sample size and scale go to infinity. This allows for the construction of a unimodality test for the Hurst exponent distribution. Monte Carlo simulations show that the proposed methodology attains satisfactory power for realistic sample sizes.**

***Index Terms*—self-similarity, operator fractional Brownian motion, wavelets, random matrices, high dimensions**

I. Introduction

Context: scale invariance. *Scale invariance* has been observed in signals coming from a wide range and variety of contexts in physics and engineering, see, e.g., [1], [2]. A signal X is called scale invariant, or fractal, when its temporal dynamics lack a characteristic scale. The focus of the analysis is thus on identifying the *scaling exponents* that relate the continuum of scales [3]. A cornerstone model of scale invariance is *self-similarity*: X is called self-similar when its finite-dimensional distributions (f.d.d.) are invariant under suitable time scaling, i.e., $\{X(t)\}_{t\in\mathbb{R}} \overset{\text{f.d.d.}}{=} \{a^H X(t/a)\}_{t\in\mathbb{R}}$, $a>0$, where the scaling exponent $0<H<1$ is called the Hurst parameter. The most prominent example is fractional Brownian motion (fBm), the only Gaussian, self-similar process with stationary increments [4]. Estimation of the Hurst parameter H plays a key role in signal processing tasks such as characterization, diagnosis, classification and detection. The so-named wavelet transform provides the analytical basis for well-established estimation methodologies for H [5].

Challenge: high-dimensional time series. The modeling of self-similarity in applications has remained so far based on the univariate fBm model. Yet, in the modern world of "Big Data," a plethora of sensors monitor natural and artificial systems, generating large data sets in the form of several joint time series. This can be seen in many fields of application. In neuroscience, for example, the number of macroscopic brain activity time series ranges from hundreds (MEG data) to several tens of thousands (fMRI data) [6]. Likewise, in climate studies, dealing with large numbers of measured components has become standard [7]. For such high-dimensional data, a multitude of scaling laws – i.e., *Hurst multimodality* – implies different large scale behavior of the system along possibly non-canonical coordinate axes. Ignoring Hurst multimodality in data may lead to arbitrarily large *estimation biases* due to the so-named dominance and amplitude effects (see [8], [9]).

Related work: self-similarity in high dimensions. A recently proposed multivariate extension of fBm, *operator fractional Brownian motion* (ofBm), [10]–[12] frames the modeling of scale invariance in a high-dimensional time series setting [8]. The so-named *wavelet eigenanalysis* methodology [8], [9] was shown to lead to efficient and robust estimation of Hurst exponents in both multivariate (fixed dimensions) and high-dimensional (the dimension $p(n)$ of the measurements grows as a function of the time series size n) settings [9], [13], [14]. Use of this model in applications immediately leads to a first critical question: *how many different scaling laws exist in the possibly very large number of time series?* In the statistical signal processing literature, the problem of identifying the number and properties of sources in multivariate or high-dimensional noisy signals has been studied for decades [15]–[18]. Examples of the proposed techniques include principal component analysis, factor analysis and sparse graphical Gaussian models [19]. Nevertheless, there has been a paucity of estimation methodologies for both *high-dimensional* and *scale invariant* signals; see *a contrario* [20] or [21], proposing a bootstrap-based method for counting the number of distinct Hurst exponents based on wavelet eigenanalysis, yet in a multivariate context with modest dimension. A key related difficulty is the study of random matrices under dependence, as they emerge in wavelet eigenanalysis, which is still a very active area of research [22], [23]. This is even more so in regards to the presence of fractional memory [24].

Goal, contributions and outline. The goal of the present work is to study, in the context of high-dimensional asymptotic limits (as the sample size, the number of components and the scale go to infinity at a fixed rate), the properties of the statistical distribution of the multivariate eigenwavelet-based

G.D.'s long term visits to ENS de Lyon were supported by the school, the CNRS and the Simons Foundation collaboration grant #714014. H.W. supported by ANR-18-CE45-0007 MUTATION, France.

estimation of the vector of Hurst exponents, both theoretically and practically. Notably, aiming to test the equality of all Hurst exponents, the uni- vs. multi-modality of these distributions is carefully addressed. To that end, the definitions and properties of ofBm and the high-dimensional model are summarized in Sections II-A and II-B, while multivariate eigenwavelet estimation procedures are briefly sketched in Section II-C. Section III describes the core contribution of this work. It explains, first, how large random matrix theory leads to the modification of the high-dimensional limit formulation by incorporating the analysis of scales and establishes, second, the high-dimensional asymptotic behavior of the distribution of the estimator of the vector of Hurst exponents, thus permitting the discussion of uni- vs. multimodality. Furthermore, these theoretical asymptotic limits are tested in practice from extensive Monte Carlo simulations using synthetic ofBm sample paths, see Section IV. First, the uni- or multimodality of the distributions of estimators are shown to be practically observed with data of moderate and realistic sizes (see Section IV-B). Second, elaborating on the Hartigans' dip test statistic [25], we devise a procedure used to test unimodality in the distributions of the estimated Hurst exponents, and assess its practical power and relevance (see Section IV-B). Reported results indicate that the proposed test has satisfactory performance and can be readily applied to real-world high-dimensional data, including for sample sizes typical in neuroscience (cf. [26]).

II. Self-similarity analysis and modeling

A. Operator fractional Brownian motion

Operator fractional Brownian motion (ofBm) is a canonical model for multidimensional scale invariant structures in real-world data. We briefly recall its definition and some properties (see [11] for the general definition and properties of ofBm).

Let $\underline{B}_{\underline{H},\Sigma}(t) = \big(B_{H_1}(t),\ldots,B_{H_p}(t)\big)_{t\in\mathbb{R}}$ denote a collection of p possibly correlated fBm components defined by their individual self-similarity exponents $\underline{H} = (H_1,\ldots,H_p)$, $0 < H_1 \leq \ldots \leq H_p < 1$. Let Σ be a pointwise covariance matrix with entries $(\Sigma)_{\ell,\ell'} = \sigma_\ell\sigma_{\ell'}\rho_{\ell,\ell'}$, where σ_ℓ^2 are the variances of the components and $\rho_{\ell,\ell'}$ their (pairwise) correlation coefficients. We define ofBm as the stochastic process $\underline{B}_{P,\underline{H},\Sigma}(t) := P\underline{B}_{\underline{H},\Sigma}(t)$, where P is a real-valued, $p\times p$ invertible matrix that mixes the components (changes the scaling coordinates) of $\underline{B}_{\underline{H},\Sigma}(t)$. OfBm consists of a multivariate Gaussian self-similar process with stationary increments. Moreover, it satisfies the (operator) self-similarity relation

$$\{\underline{B}_{P,\underline{H},\Sigma}(t)\}_{t\in\mathbb{R}} \overset{\text{f.d.d.}}{=} \{a^{\underline{\underline{H}}}\underline{B}_{P,\underline{H},\Sigma}(t/a))\}_{t\in\mathbb{R}}, \qquad (1)$$

$\forall a > 0$. In (1), the matrix (Hurst) exponent is given by $\underline{\underline{H}} = P\mathrm{diag}(\underline{H})P^{-1}$, and $a^{\underline{\underline{H}}} := \sum_{k=0}^{+\infty}\log^k(a)\underline{\underline{H}}^k/k!$, where $\overset{\text{f.d.d.}}{=}$ stands for the equality of finite-dimensional distributions.

B. High-dimensional model

Let $\pi(dH)$ be a discrete distribution of Hurst exponents with (ordered) support $\{H_1,\ldots,H_m\}$, $m\in\mathbb{N}$. Given a vector $\underline{H}\in(0,1)^p$ of i.i.d. samples from $\pi(dH)$, the process

$$Y(t) := \underline{B}_{P,\underline{H},\Sigma} \qquad (2)$$

as defined in Section II-A is, *conditionally on* $\underline{H}$, a p-variate ofBm with Hurst matrix $\underline{\underline{H}}$. Given a time series $\{Y(t)\}_{t=1,\ldots,n}$, we further assume

$$p = p(n) \to \infty \text{ as } n\to\infty,$$

under which (2) is a *high-dimensional model* and has $p = p(n)\to\infty$ Hurst exponents (on models of the general form (2) under weak dependence, see, for instance, [27]).

C. Scaling in the wavelet domain

Multivariate wavelet transform. Let ψ be a mother wavelet, i.e., a real-valued function such that $\int_{\mathbb{R}}\psi^2(t)dt = 1$. For all $k,j\in\mathbb{Z}$, the multivariate DWT of $\{Y(t)\}_{t\in\mathbb{R}}$ is defined as $D_Y(2^j,k) := (D_{Y_1}(2^j,k),\ldots,D_{Y_p}(2^j,k))$, where $D_{Y_\ell}(2^j,k) := \langle 2^{-j/2}\psi(2^{-j}t-k)|Y_\ell(t)\rangle \in \mathbb{R}$ for $\ell\in\{1,\ldots,p\}$. For a detailed introduction to wavelet transforms, see [28]. It can be shown that the wavelet coefficients $\{D_Y(2^j,k)\}_{k\in Z}$ of p-variate ofBm $Y = \underline{B}_{P,\underline{H},\Sigma}$ satisfy, for every fixed octave j, the operator self-similarity relation [8], [9]

$$\{D_Y(2^j,k)\}_{k\in\mathbb{Z}} \overset{\text{f.d.d.}}{=} \{2^{j(\underline{\underline{H}}+\frac{1}{2}I)}D_Y(1,k)\}_{k\in\mathbb{Z}}. \qquad (3)$$

Wavelet random matrices and high-dimensional eigenanalysis. Given *any* p-variate process $\{Y(t)\}_{t\in\mathbb{R}}$ (in particular, model (2)), the sample wavelet spectrum (variance) at octave $j = j_1,\ldots,j_2$ is given by the $p\times p$ wavelet random matrices

$$S_Y(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D_Y(2^j,k)D_Y(2^j,k)^* \qquad (4)$$

where n is the time series (sample) size and $n_j\simeq n/2^j$ is the number of wavelet coefficients available at scale 2^j. It has been shown and discussed [8], [9] that, in general, estimation based on the entrywise multiscale behavior of $S_Y(2^j)$ is arbitrarily biased and effectively meaningless. Instead, it was proposed in [8], [9] that the estimation of the multivariate Hurst exponents should be based on the eigenvalues

$$\lambda_1(2^j),\ldots,\lambda_p(2^j)$$

of the random matrix $S_Y(2^j)$ as in (4). Notably, it was shown that the Hurst exponents H_ℓ can be efficiently estimated by means of the *weighted* wavelet log-eigenvalues

$$\widehat{H}_\ell = \Big(\sum_{j=j_1}^{j_2} w_j\log_2\lambda_\ell(2^j)\Big)\Big/2 - \frac{1}{2}, \quad \ell = 1,\ldots,p, \qquad (5)$$

where $\sum_j w_j = 0$ and $\sum_j jw_j = 1$. In fact, (5) has good statistical performance in both Gaussian and non-Gaussian frameworks [8], [9], [13], [14].

III. High-dimensional limits

A. High dimensions: Wavelet log-eigenvalue distribution in the three-way limit

As indicated in the Introduction, the present work addresses high-dimensional limits. In contradistinction to the classical setting where the sample size goes to infinity for fixed number of components, i.e., $n\to+\infty$, $p = p_0$, here we consider

high-dimensional settings, where both the sample size and the number of components go jointly to infinity, i.e., $n, p \to +\infty$. In the context of sample covariance matrices, this is more sharply characterized by the two-way limit $\lim_{n\to\infty} p/n = c \in [0, \infty)$ [29].

In the context of multivariate Hurst exponent estimation using wavelet-based linear regressions, as in (5), when both $n, p \to +\infty$, thorough analysis of scale invariant systems further implies that the range of scales where linear regressions are performed should also grow towards infinity: $(j_1, j_2) \to +\infty$. Technically, in such settings, the high-dimensional behavior analyzed by means of large random matrices actually leads us to consider the *three-way* limit [9], [30]

$$n, p, j \to +\infty, \ \frac{p}{n/2^j} \to c \in [0, \infty), \tag{6}$$

a non-trivial contribution of the present work, both theoretically and numerically.

B. High dimensions: Multimodality of Hurst exponents

Mathematical results in [9], [30] imply that, in the three-way limit (6) and under (2), the eigenvalues of the sample wavelet spectrum behave as

$$\log_2 \lambda_\ell(2^j) = C_\ell + j \cdot (2H_\ell + 1) + o_{\mathbb{P}}(1), \tag{7}$$

$\ell = 1, \ldots, p$, for some constant C_ℓ, where $o_{\mathbb{P}}(1)$ vanishes in probability. In fact, eigenvalue scaling relations of the type (7) are shown to hold for high-dimensional signal-plus-noise instances, broader than the noise-free model (2) [9], [30].

Further, assume measurements given by the high-dimensional model (2). In light of (7), we obtain the key result of this paper.

Fundamental property of weighted log-eigenvalues of wavelet random matrices: *the distribution of the $p = p(n)$ estimates $\widehat{H}_\ell$ resulting from weighted log-eigenvalues of wavelet random matrices as defined in* (5) *satisfies, for small $\varepsilon > 0$ and any $q = 1, \ldots, m$,* as $n \to +\infty$,

$$\frac{\#\big\{\ell = 1, ..., p(n) : \widehat{H}_\ell \in (H_q - \varepsilon, H_q + \varepsilon)\big\}}{p(n)} \xrightarrow{\mathbb{P}} \pi(H_q). \tag{8}$$

Also, and importantly, we note that the convergence (8) of the wavelet empirical (log)spectral distribution to $\pi(dH)$ takes place on the condition that $n_j = n/2^j$ is large enough w.r.t. $p = p(n)$ so as to prevent significant bias stemming from *eigenvalue repulsion* (see [31]–[33]), or for the random matrices (4) to have *deficient rank* (which yields some negatively infinite log-eigenvalues).

This key theoretical result is further inspected practically in Section IV-B and is illustrated in Fig. 1.

IV. Practical performance assessment

A. Monte Carlo simulation setting

To assess the practical relevance of the theoretical results stated in Section III, we make use of Monte Carlo simulations based on 1000 of independent realizations of p-variate measurements $\underline{B}_{P,\underline{H},I} = P\underline{B}_{I,\underline{H},I}$, where P is a randomly chosen orthogonal matrix and $\{\underline{B}_{I,\underline{H},I}(t)\}_{t\in\mathbb{R}}$ made up of

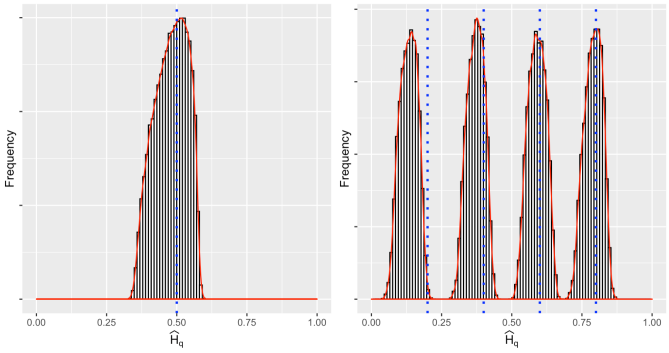

Fig. 1. **Distribution of weighted wavelet log-eigenvalues $\widehat{H_q}$ in the three-way limit $\lim_{n\to\infty} p\ 2^j/n$.** (Left) $p-$variate ofBm with Hurst matrix $\underline{\underline{H}} = P\mathrm{diag}(0.5, \ldots, 0.5)P^{-1} \in \mathbb{R}^{p^2}$. (Right) $p-$variate ofBm with Hurst matrix $\underline{\underline{H}} = P\mathrm{diag}(H_1, \ldots, H_p)P^{-1} \in \mathbb{R}^{p^2}$, where $H_1, \ldots, H_p$ are sampled from $\pi(dH)$ supported on $\{0.2, 0.4, 0.6, 0.8\}$. Red curves are numerical best fits. A unimodal (left) or a four-modal (right) distribution emerges in the wavelet log-eigenspectrum, as opposed to the traditional Marčenko-Pastur-based laws in the eigenspectrum of sample covariance matrices.

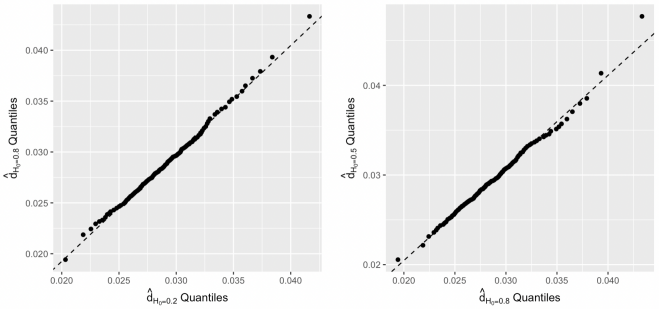

Fig. 2. **Quantile-quantile plots for $\widehat{d}_{H_0}$.** The plots compare the quantiles of the distributions of $\widehat{d}_{H_0}$, for different pairs of H_0: (Left) $H_0 = 0.8$ vs. $H_0 = 0.2$; (Right) $H_0 = 0.8$ vs. $H_0 = 0.5$; and show that the (non-Gaussian) distribution of $\widehat{d}_{H_0}$ is independent of H_0.

independent univariate fBms with $n = 2^{14}$ (such sample sizes are realistic, for example, in the context of the analysis of infraslow brain activity [26]). The univariate fBms are generated using the **R** package `somebm` (see [34]). The p-dimensional vector $\underline{H}$ is obtained by drawing p i.i.d. samples from Unif$(H_1, \ldots, H_m)$, for each realization independently. Here, we limit ourselves to the consideration of $m = 1, 2, 4,$ and 6. The wavelet transformation is generated by means of Mallat's algorithm based on a Daubechies filter with $N_\psi = 2$. In our simulations, we set the dimension to $p = 2^6$. The range of regression scales is chosen to be $j_1 = 2$ to $j_2 = 5$. Hence, $p/n_j < 1$ for $j = j_1, \ldots, j_2$, implying that all the wavelet random matrices used in the regression procedure have full rank.

B. Multimodality in practice

To illustrate the practical relevance of the high-dimensional asymptotic limits in (7) and (8), Fig. 1 provides an example of the high-dimensional behavior of the wavelet log-eigenvalue distribution. In the plots, a unimodal (left) or a four-modal (right) distribution strikingly emerges in the wavelet

eigenspectrum depending on the presence of one or four modes, respectively, in the support of $\pi(dH)$. To put this remarkable result in context, consider the traditional random matrix theoretic framework of sample covariance matrices. In this case, as $p/n \to c \in [0, \infty)$ and under unimodality, we observe a combination (a *free convolution*; see [35]) between a Marčenko-Pastur law and the spectral density of the univariate stochastic process whose independent copies form the sample covariance matrix. In particular, the Hurst exponent only appears implicitly in the limit eigenspectrum distribution, which is not available in closed form [24].

C. Unimodality testing

To go beyond visual evidence, we are now interested in practically testing multimodality vs. unimodality in the distribution of Hurst exponents. More precisely, this means testing the null hypothesis

$$\mathcal{H}_0 : \text{the distribution } \pi(dH) \text{ of Hurst exponents is unimodal,}$$

i.e., that the support of $\pi(dH)$ contains a single, unspecified value in $(0, 1)$. We put forward a test based on the wavelet empirical (log)spectral distribution, i.e., the distribution of the estimates $\widehat{H}_\ell$. Indeed, in light of (8), this distribution is expected to be *unimodal* under $\mathcal{H}_0$, and *multimodal* otherwise.

Unimodality test. With this in mind, we propose a unimodality testing methodology that builds upon the original idea behind Hartigan's dip test for multimodality (see [25]). Let

$$\widehat{F}(x) = \frac{1}{p} \sum_{\ell=1}^{p} 1_{\{\widehat{H}_\ell \leq x\}}, \quad x \in \mathbb{R},$$

be the empirical distribution function of the weighted wavelet log-eigenvalues

$$\widehat{H}_\ell, \quad \ell = 1, \ldots, p, \tag{9}$$

as in (5). As in [25], let $\mathcal{U}$ be the class of all unimodal distribution functions, and define the statistic

$$\widehat{d} = \inf_{G \in \mathcal{U}} \sup_{x \in \mathbb{R}} |\widehat{F}(x) - G(x)|, \tag{10}$$

that is, $\widehat{d}$ is computed based on picking the unimodal distribution that provides the best fit to the empirical distribution function (see also [25], Section 4, on the details of how to compute (10) in practice). The proposed test procedure is defined as

$$\text{reject } \mathcal{H}_0 \text{ when } \widehat{d} > d_\alpha, \tag{11}$$

where the size of the test $\alpha \in [0, 1]$ is defined as [36]

$$\alpha = \sup_{H_0 \in (0,1)} \mathbb{P}\big(\widehat{d} > d_\alpha \big| \text{supp } \pi(dH) = H_0\big). \tag{12}$$

The statistic $\widehat{d}$ as in (10) was computed using the **R** package `diptest` (see [37]); test significance was set to $\alpha = 0.05$.

Test statistics practical threshold. In regard to the computation of the rejection threshold d_α, note that the observations are assumed independent in the original dip test [25], a condition that does not hold for the weighted wavelet log-eigenvalues (9). For this reason, we simulate a large number of independent p-variate ofBms with Hurst matrix exponent $\underline{\underline{H}} = \text{diag}(H_0, \ldots, H_0)$, $H_0 \in (0, 1)$, calculate the corresponding test statistics (10) and use (12) to compute the rejection threshold d_α for fixed α.

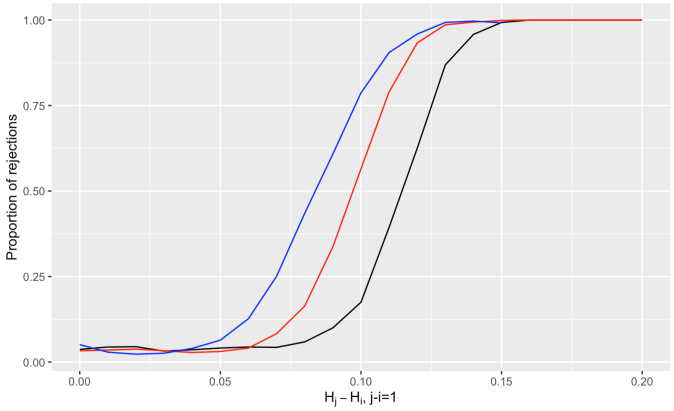

Fig. 3. **Test powers.** Proportions of rejections over 1000 Monte Carlo runs for several alternative hypothesis: (Black) supp $\pi(dH) = \{H_1, H_2\}$ (Red) supp $\pi(dH) = \{H_1, H_2, H_3, H_4\}$ (Blue) supp $\pi(dH) = \{H_1, ..., H_6\}$.

When computing the test statistic $\widehat{d} \equiv \widehat{d}_{H_0}$ under the null hypothesis $\mathcal{H}_0$, the sampling distribution of $\widehat{d}_{H_0}$ is found to be independent of H_0 (see Figure 2). This result is non-trivial, but it is in line with and corroborates previous studies of the properties of wavelet-based estimators for scale-free systems (e.g., [5]).

Test performance. Results reported in Fig. 3 (leftmost point) show that, for different settings, the proposed unimodality test (11) has the prescribed size $\alpha = 0.05$ under $\mathcal{H}_0$. To study the power of the test, we first considered instances of the alternative hypothesis where $\pi(dH)$ is supported on the set $\{H_1, H_2\}$ of two distinct Hurst exponents. Namely, we compute the proportion of rejections for the pairs of values $H_1 = 0.6 - \Delta$, $H_2 = 0.6 + \Delta$, with $\Delta \in [0.0, 0.10]$. Figure 3 shows that the test attains power greater than 0.80 provided that $H_2 - H_1 \geq 0.125$.

Next, we repeated the analysis for instances of the alternative hypothesis where $\pi(dH)$ is supported on a set of 4 or 6 distinct Hurst parameters, respectively, reported in Figure 3 (Red) and (Blue). In both cases, the parameters $H_1, ..., H_m$ are initialized at 0.6. From there, the difference between adjacent Hurst parameters is increased so as to keep the parameters equidistant and centered around 0.6. As expected, the power of the test was shown to increase as a function of the distance between the Hurst exponents in both cases. Moreover, comparative inspection of the distinct lines in Fig. 3 reveals that, starting at the same value 0.6 under the null hypothesis, the speed of convergence of the power of the test to 1 is greater when the number of distinct Hurst parameters is larger.

Overall, these results confirm that the proposed test procedure is operational, has satisfactory performance and can be readily applied in the study of self-similarity in high-dimensions.

V. Conclusions and perspectives

A statistical methodology for detecting multimodality in the distribution of Hurst exponents in high-dimensional fractal systems has been devised. It relies on the analysis of the distribution of the log-eigenvalues of large wavelet random matrices in the threefold limit as dimension, sample size and scale go to infinity. Depending on the presence of a single or several modes in the distribution of Hurst exponents, the wavelet empirical (log)spectral distribution also displays one or several modes. A practical unimodality test is further proposed for the Hurst exponent distribution. Future work includes the construction of methodology for the identification of the distribution $\pi(dH)$ once unimodality has been rejected (cf. [21] in the multivariate context). Studying high-dimensional limits is not a purely mathematical issue; instead, it is absolutely crucial for practical use in applications. Indeed, multiscale analysis implies *by its own nature* that the effective sample size is not large compared to the number of components at coarse scales. Real data modeling also calls for the investigation of the more general case where the components of $X(t)$ in (2) are correlated. This requires specific mathematical efforts to be made in the future.

References

[1] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*, Springer Science & Business Media, 2006.

[2] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Info. Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[3] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Info. Theory*, vol. 38, pp. 910 – 917, March 1992.

[4] V. Pipiras and M. S. Taqqu, *Long-Range Dependence and Self-Similarity*, vol. 45, Cambridge University Press, 2017.

[5] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. Info. Theory*, vol. 45, no. 3, pp. 878–897, 1999.

[6] P. Ciuciu, G. Varoquaux, P. Abry, S. Sadaghiani, and A. Kleinschmidt, "Scale-free and multifractal time dynamics of fMRI signals during rest and task," *Front. Physiol.*, vol. 3, 2012.

[7] F. A. Isotta, C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassegues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, and M. Munari, "The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data," *Int. J. Climatol.*, vol. 34, no. 5, pp. 1657–1675, 2014.

[8] P. Abry and G. Didier, "Wavelet estimation for operator fractional Brownian motion," *Bernoulli*, vol. 24, no. 2, pp. 895–928, 2018.

[9] P. Abry and G. Didier, "Wavelet eigenvalue regression for n-variate operator fractional Brownian motion," *J. Multivar. Anal.*, vol. 168, pp. 75–104, November 2018.

[10] J. D. Mason and Y. Xiao, "Sample path properties of operator-self-similar Gaussian random fields," *Theory Probab. Appl.*, vol. 46, no. 1, pp. 58–78, 2002.

[11] G. Didier and V. Pipiras, "Integral representations and properties of operator fractional Brownian motions," *Bernoulli*, vol. 17, no. 1, pp. 1–33, 2011.

[12] G. Didier and V. Pipiras, "Exponents, symmetry groups and classification of operator fractional Brownian motions," *J. Theor. Probab.*, vol. 25, pp. 353–395, 2012.

[13] P. Abry, H. Wendt, and G. Didier, "Detecting and estimating multivariate self-similar sources in high-dimensional mixtures," in *IEEE Stat. Signal Process. Workshop*, 2019, pp. 1–5.

[14] B.C. Boniece, H. Wendt, G. Didier, and P. Abry, "On multivariate non-Gaussian scale invariance: fractional Lévy processes and wavelet estimation," in *Proc. Eur. Signal. Process. Conf. (EUSIPCO)*, 2019, pp. 1–5.

[15] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, 2010.

[16] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, 2004.

[17] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, 2001.

[18] J. P. C. L. Da Costa, A. Thakre, F. Roemer, and M. Haardt, "Comparison of model order selection techniques for high-resolution parameter estimation algorithms," in *Proc. 54th Intern. Scient. Colloq. (IWK'09)*, 2009.

[19] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 4, Prentice-Hall, New Jersey, 2014.

[20] R. Zhang, P. Robinson, and Q. Yao, "Identifying cointegration by eigenanalysis," *J. Am. Stat. Assoc.*, vol. 114, no. 526, pp. 916–927, 2018.

[21] C.-G. Lucas, P. Abry, H. Wendt, and G. Didier, "Counting the number of different scaling exponents in multivariate scale-free dynamics: Clustering by bootstrap in the wavelet domain," in To appear in *IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, 2022, pp. 1–5.

[22] H. Liu, A. Aue, and D. Paul, "On the Marčenko–Pastur law for linear time series," *Ann. Statist.*, vol. 43, no. 2, pp. 675–712, 2015.

[23] L. Erdős, T. Krüger, and D. Schröder, "Random matrices with slow correlation decay," in *Forum of Mathematics, Sigma*. Cambridge University Press, 2019, vol. 7.

[24] F. Merlevède and M. Peligrad, "On the empirical spectral distribution for matrices with long memory and independent rows," *Stochastic Process. Appl.*, vol. 126, no. 9, pp. 2734–2760, 2016.

[25] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *Ann. Statist.*, vol. 13, no. 1, pp. 70 – 84, 1985.

[26] D. La Rocca, H. Wendt, V. van Wassenhove, P. Ciuciu, and P. Abry, "Revisiting functional connectivity for infraslow scale-free brain dynamics using complex wavelets," *Frontiers in Physiology*, vol. 11, pp. 1651, 2021.

[27] J. Chang, B. Guo, and Q. Yao, "Principal component analysis for second-order stationary vector time series," *Ann. Statist.*, vol. 46, no. 5, pp. 2094–2124, 2018.

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

[29] Z. Bai and J. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2010.

[30] P. Abry, B. C. Boniece, G. Didier, and H. Wendt, "On high-dimensional wavelet eigenanalysis," *arXiv preprint arXiv:2102.05761*, 2021.

[31] T. Tao, *Topics in Random Matrix Theory*, vol. 132, American Mathematical Society, 2012.

[32] H. Wendt, P. Abry, and G. Didier, "Bootstrap-based bias reduction for the estimation of the self-similarity exponents of multivariate time series," in *IEEE Int. Conf. Acoust., Speech, and Signal Proces. (ICASSP)*, 2019, pp. 4988–4992.

[33] Abry P. Lucas, C.-G., H. Wendt, and G. Didier, "Bootstrap for testing the equality of selfsimilarity exponents across multivariate time series," in *Proc. Eur. Signal. Process. Conf. (EUSIPCO)*, 2021, pp. 1–5.

[34] J. Huang, *somebm: some Brownian motions simulation functions*, 2013, R package version 0.1.

[35] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010.

[36] E. L. Lehmann and J. Romano, *Testing Statistical Hypotheses*, Springer, 2005.

[37] M. Maechler, *diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected*, 2021, R package version 0.76-0.