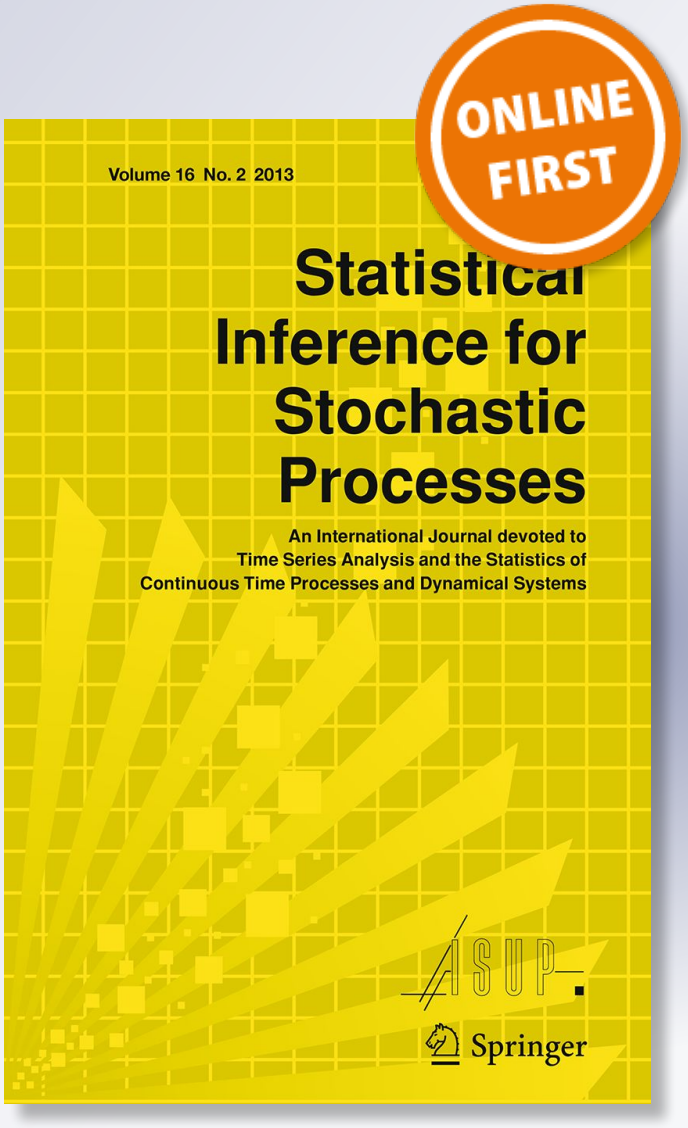
Two-step wavelet-based estimation for Gaussian mixed fractional processes

Patrice Abry, Gustavo Didier & Hui Li

Statistical Inference for Stochastic Processes
An International Journal devoted to Time Series Analysis and the Statistics of Continuous Time Processes and Dynamical Systems

ISSN 1387-0874

Stat Inference Stoch Process
DOI 10.1007/s11203-018-9190-z



Springer

Statistical Inference for Stochastic Processes
https://doi.org/10.1007/s11203-018-9190-z

Two-step wavelet-based estimation for Gaussian mixed fractional processes

Patrice Abry[1] · **Gustavo Didier**[2] · **Hui Li**[2]

Received: 10 January 2018 / Accepted: 16 August 2018

Abstract
A Gaussian mixed fractional process $\{Y(t)\}_{t\in\mathbb{R}} = \{PX(t)\}_{t\in\mathbb{R}}$ is a multivariate stochastic process obtained by pre-multiplying a vector of independent, Gaussian fractional process entries X by a nonsingular matrix P. It is interpreted that Y is observable, while X is a hidden process occurring in an (unknown) system of coordinates P. Mixed processes naturally arise as approximations to solutions of physically relevant classes of multivariate fractional stochastic differential equations under aggregation. We propose a semiparametric two-step wavelet-based method for estimating both the demixing matrix P^{-1} and the memory parameters of X. The asymptotic normality of the estimator is established both in continuous and discrete time. Monte Carlo experiments show that the estimator is accurate over finite samples, while being very computationally efficient. As an application, we model a bivariate time series of annual tree ring width measurements.

Keywords Multivariate stochastic process · Fractional stochastic process · Operator self-similarity · Demixing · Wavelets

Mathematics Subject Classfication Primary: 62M10 · 60G18 · 42C40

Patrice Abry was partially supported by Grant ANR-16-CE33-0020 MultiFracs. Gustavo Didier was partially supported by the prime Award No. W911NF-14-1-0475 from the Biomathematics subdivision of the Army Research Office, USA. Gustavo Didier's long term visits to ENS de Lyon were supported by the school.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11203-018-9190-z) contains supplementary material, which is available to authorized users.

✉ Gustavo Didier
gdidier@tulane.edu

Patrice Abry
patrice.abry@ens-lyon.fr

Hui Li
hli15@tulane.edu

[1] Laboratoire de Physique, ENS de Lyon, CNRS, Université de Lyon, Université Claude Bernard, 69342 Lyon, France

[2] Mathematics Department, Tulane University, 6823 St. Charles Avenue, New Orleans, LA 70118, USA

1 Introduction

A $\mathbb{R}^n$-valued multivariate stochastic process is called *mixed* when it has the form

$$\{Y(t)\}_{t\in\mathbb{R}} = \{PX(t)\}_{t\in\mathbb{R}}, \tag{1.1}$$

where P is a square, nonsingular real matrix and

$$\{X(t)\}_{t\in\mathbb{R}} = \{(X_1(t), \ldots, X_n(t))^T\}_{t\in\mathbb{R}} \tag{1.2}$$

is a vector of independent processes. The process $Y = \{Y(t)\}_{t\in\mathbb{R}}$ is assumed observable. The process $X = \{X(t)\}_{t\in\mathbb{R}}$ can be interpreted either as a hidden process whose components get scrambled by a mixing matrix parameter P, or as one occuring in a different system of coordinates. A *wavelet* is a unit $L^2(\mathbb{R})$-norm function that annihilates a certain number of polynomials [see (2.4)]. In this paper, we propose a new semiparametric, wavelet-based statistical method for a subclass of processes of the form (1.1), namely, we assume the hidden process X is component-wise Gaussian, has stationary increments of any order (including zero), and has *fractional* memory, namely, the covariance function of its increments has hyperbolic decay. Our perspective is that of *scaling analysis*, i.e., of retrieving the scaling laws governing the behavior of the observable Y (Mandelbrot and Ness 1968; Flandrin 1992; Wornell and Oppenheim 1992).

As a modeling framework, processes such as (1.1) bring together the literatures on fractional stochastic processes and on blind source separation in signal processing. The former is well-established for the univariate case, both for Gaussian and non-Gaussian instances (e.g., Taqqu 1975, 1979; Dobrushin and Major 1979; Granger and Joyeux 1980; Hosking 1981; Fox and Taqqu 1986; Dahlhaus 1989; Beran 1994; Robinson 1995a, b; Stoev et al. 2002; Bardet and Tudor 2010; Beran et al. 2013; Bardet and Tudor 2014; Clausel et al. 2014a, b). Partly motivated by several applications such as in signal processing, neuroscience, economics and finance, and Internet traffic, the volume of contributions on multivariate fractional processes has been expanding at a fast pace (Hosoya 1996, 1997; Lobato 1997; Marinucci and Robinson 2000; Shimotsu 2007; Becker-Kern and Pap 2008; Nielsen 2011; Sela and Hurvich 2012; Kechagias and Pipiras 2015a, b, in the time and Fourier domains, and Wendt et al. 2009; Amblard and Coeurjolly 2011; Amblard et al. 2012; Coeurjolly et al. 2013; Achard and Gannaz 2016, in the wavelet domain). On the other hand, the literature on blind source separation is well-established for ARMA-like signals, often driven by applications such as in acoustics or image analysis (e.g., Belouchrani et al. 1997; Cardoso 1998; Pham and Cardoso 2001; Moreau 2001; Yeredor 2002; Parra and Sajda 2003; Stone 2004; Ziehe et al. 2004; Choi et al. 2005; O'Grady et al. 2005; Fevotte and Godsill 2006; Li et al. 2009; Comon and Jutten 2010).

The main motivation for this paper is to provide statistically accurate and computationally efficient estimation methodology for a physically relevant subclass of multivariate fractional processes. There are at least two reasons to consider (Gaussian) mixed fractional models.

First, it was recently shown that processes of the form (1.1) naturally arise as approximations to solutions of physically comprehensive classes of multivariate fractional stochastic differential equations (SDEs) under aggregation (Tsai et al. 2018). In other words, if we digitalize one such solution $\{Y(t)\}_{t\geq 0}$ by aggregation over intervals of length $\triangle$, i.e.,

$$Y_s^{\triangle} = \int_{(s-1)\triangle}^{s\triangle} Y(u)du, \tag{1.3}$$

then, for large $\triangle$ and some matrix of coordinates P,

$$Y_s^{\triangle} \approx P X_s, \quad s \in \mathbb{N} \cup \{0\} \tag{1.4}$$

(for more details, see Sect. 2.2). It is well known that many real data sets—e.g., tree ring widths, river flows, rainfall or network traffic—are obtained by means of aggregation over a certain time interval, which is indicative of the usefulness of (1.1). Since the latter is a substantially more parsimonious model than a general multivariate fractional process, it is natural to investigate its suitability for scientific or technological applications where dimension plays some role. This motivation is especially significant in the modern era of "data deluge" or availability of large and complex multidimensional data sets (e.g., Bell et al. 2009; Ball and Brunner 2010; Chen and Zhang 2014).

Second, multivariate fractional processes of the form (1.1) are also closely related to the so-named operator self-similar (o.s.s.) random processes and fields. A $\mathbb{R}^n$-valued o.s.s. process Y satisfies the (operator) scaling relation

$$\{Y(ct)\}_{t\in\mathbb{R}} \stackrel{\mathcal{L}}{=} \{c^H Y(t)\}_{t\in\mathbb{R}}, \quad c > 0, \tag{1.5}$$

for a (Hurst) square matrix exponent H, where c^H is defined by the matrix exponential $\exp\{\log c\ H\} = \sum_{k=0}^{\infty} (\log c\ H)^k/k!$ and $\stackrel{\mathcal{L}}{=}$ denotes the equality of finite dimensional distributions. The literature on operator self-similarity started in the early 1980s (Laha and Rohatgi 1981; Hudson and Mason 1982) and significantly expanded in the past two decades (e.g., Maejima and Mason 1994; Mason and Xiao 2002; Biermé et al. 2007; Xiao 2009; Guo et al. 2009; Didier and Pipiras 2011, 2012; Clausel and Vedel 2011, 2013; Li and Xiao 2011; Dogan et al. 2014; Puplinskaitė and Surgailis 2015; Didier et al. 2017, 2018; Abry and Didier 2018a). In the context of o.s.s. and related processes, the estimation of the matrix P is itself of great interest, since it makes up the system of coordinates of the Hurst matrix (see Example 2.2). However, numerically speaking, parametric M-estimation of the whole matrix H is notoriously intricate and calls for *ad hoc* optimization methods even in low dimension (c.f. Frecon et al. 2016).

One key statistical challenge in the scaling analysis of Y is to retrieve the fractional information (on memory parameters or the related Hurst exponents) contained in X. From a mathematical standpoint, the effect of mixing scaling laws can be illustrated by means of the expression for the spectral density f_Y of the mixed process Y. For example, fractional Brownian motion (fBm) is the only Gaussian, self-similar, stationary increment stochastic process (e.g., Embrechts and Maejima 2002; Taqqu 2003), and its increment process is called fractional Gaussian noise (fGn). Suppose the hidden process X is a bivariate vector of independent fGn entries with memory parameters $-1/2 < \delta_1 < \delta_2 < 1/2$. For $x \neq 0$, each entry of the (continuous time) spectral density function of the observable Y takes the form

$$f_Y(x)_{i_1 i_2} = \left|\frac{e^{\mathbf{i}x} - 1}{\mathbf{i}x}\right|^2 \{c_{1,i_1 i_2}|x|^{-2\delta_1} + c_{12,i_1 i_2}|x|^{-(\delta_1+\delta_2)} + c_{2,i_1 i_2}|x|^{-2\delta_2}\}, \quad i_1, i_2 = 1, 2, \tag{1.6}$$

for so-named amplitude coefficients $c_\bullet$ [n.b.: (1.6) can be obtained from (2.22)]. The univariate-inspired approach of setting up a Fourier-domain log-regression has to cope with the double-sided challenge of mixed power laws, namely, what can be called the *dominance* and *amplitude effects*. On one hand, under mild assumptions on the amplitude coefficients in (1.6), the dominant power law $|x|^{-2\delta_2}$ always prevails around the origin of the spectrum. On the other hand, and paradoxically, even if the estimation of δ_2 is the target, the

magnitude of the amplitude coefficients themselves can arbitrarily bias the estimate over finite samples by masking the dominant power law (c.f. Abry and Didier 2018b, Introduction).

From a different perspective that is not pursued in this paper, under assumptions there are linear combinations of the entries of Y that reduce entrywise fractional memory. These instances form *cointegrated* systems. One reason to study the latter is that, in general, whether or not a cointegration relation is present influences the asymptotic distributions of estimators. This is a well-studied subject in the econometric literature (see, for instance, Marinucci and Robinson 2001; Robinson and Yajima 2002; Robinson 2008; Hualde and Robinson 2010; Nielsen and Frederiksen 2011; Shimotsu 2012).

The two-step nature of the proposed wavelet-based methodology reflects a combination of elements from the two largely separate research traditions of blind source separation (demixing) and fractional modeling (scaling analysis). More precisely, the method can be summed up as follows.

- ($S1$) **demixing step (change of coordinates)** generate an estimator $\widehat{P^{-1}}$ of the demixing matrix P^{-1} by jointly diagonalizing two wavelet variance matrices [i.e., $W(2^j)$ at two different octaves j; see (2.7)] of the mixed process Y;
- ($S2$) **memory parameter estimation step** estimate the memory parameters $\delta_1, \ldots, \delta_n$ by performing a univariate wavelet log-regression on each entry of the demixed process $\widehat{X} = \widehat{P^{-1}}Y$ (Veitch and Abry 1999; Moulines et al. 2007a, b, 2008).

(c.f. Sect. 3.1, which contains an overview of the method). In the preliminary study Didier et al. (2015), presented without proofs, a demixing estimator is proposed for P that draws upon the joint diagonalization of sample covariance matrices. However, it is well known that covariance matrices are sensitive to contamination by trends, that they can be strongly dependent, and that estimators based on sample covariance matrices may be asymptotically non-Gaussian (see Breuer and Major 1983, Theorem 1, Coeurjolly 2001, Proposition 1, or Pipiras and Taqqu 2017, Chapter 5). Among other well-documented benefits, a wavelet framework tackles all these issues simultaneously, as we now briefly describe.

- (i) *Computational efficiency* the computational complexity of the wavelet transform can be even lower than that of the Fourier transform. For finite filters such as the Daubechies, it is of the order $O(\nu)$, where ν is the sample size (see Daubechies 1992, or Mallat 1999, p. 259).
- (ii) *Robustness with respect to contamination by trends* since wavelet filters contain an embedded differencing operator, they are insensitive to polynomial trends of order lower than the chosen number of vanishing moments N_ψ—see (2.4) (Flandrin 1992; Abry and Veitch 1998; Craigmile et al. 2005 or Percival and Walden 2006, Section 9.4). For the same reason, for a large enough N_ψ, wavelet coefficients $\{D(2^j, k)\}_{k\in\mathbb{Z}} \in \mathbb{R}^n$ are stationary in the shift parameter k at every octave j [see (2.5) and Remark 2.4]. This makes wavelets a natural theoretical framework for stochastic systems with stationary increments of arbitrary order.
- (iii) *Quasi-decorrelation* the wavelet transform approximately decorrelates a wide range of stochastic processes (Masry 1993; Bardet 2002). This paves the way for the construction of log-linear regression- or approximate maximum likelihood-based estimators with good finite-sample properties (e.g., Veitch and Abry 1999, or Percival and Walden 2006, Section 9.3). Basing step ($S1$) on wavelet variance matrices of Y ensures that the demixing estimator $\widehat{P^{-1}}$ is consistent and asymptotically normal even for relatively small values of N_ψ (Theorem 3.1). In addition, the estimator of the vector of memory

parameters generated at step ($S2$) is also consistent and jointly asymptotically normal (Theorem 3.2).

Mathematically, we provide two extensions of Theorems 3.1 and 3.2. First, with a view toward hypothesis testing, the consistency and asymptotic normality of the estimators generated at both steps ($S1$) and ($S2$) are shown to hold under mild assumptions even in the presence of equal memory parameters (Corollary 3.1). Second, under the more realistic assumption that Y in (1.1) is observed in discrete time, the asymptotic properties of the proposed estimators do not qualitatively change (Theorems C.2, C.3 and Corollary C.1).

We stress that, entrywise, the hidden process X is *not* assumed to be exactly self-similar, namely, it does *not* necessarily satisfy relation (1.5) for scalar memory parameters [see (2.13), (2.14) and (2.18) and the discussion in Example 2.2].

We conducted broad Monte Carlo experiments for instances where X is made up of independent fBm/fGn components. In dimension 4, the results show that the performance of the proposed two-step estimation method is similar to that of semiparametric univariate estimators of memory parameters over finite samples. Comparison with fully parametric Whittle-type maximum likelihood estimation helps to show that the two-step method's performance over finite samples is quite accurate for semiparametric estimation in terms of mean squared error even for relatively small sample sizes (of the order 2^{10}). Moreover, the method bears the advantage of being computationally very fast, a key feature in the modeling of multivariate systems.

Mixed fractional models were used in the work of Tsai et al. (2018) to fit data from robotics. We use the two-step method to model a bivariate data set from bristlecone pine tree rings from California. The two panels in Fig. 1 illustrate the effect of the demixing (change of coordinates) step ($S1$), followed by scaling analysis. On the left, demixing causes the scaling curves to decouple, which is evidence of distinct, hidden power laws in data. The right panel displays the sample wavelet coherence function, akin to the Fourier domain coherence function. Demixing turns significant correlations into nearly zero ones for most octaves. This analysis shows that there is, indeed, evidence that a fractional SDE is a good approximation to the underlying physical process (for details, see Sect. 5).

This research leads to a number of interesting questions. First, how often can real world multivariate fractional phenomena be reasonably modeled as an aggregation over the solution of a fractional SDE? Driven by computational concerns, this is of special interest when some degree of robustness vis-à-vis dimension is required. Second, with the purpose of modeling non-Gaussian systems, what is the physical analogue of a fractional Gaussian measure-driven SDE? Third, are there other classes of multivariate fractional (in particular, o.s.s.) processes for which estimation of the coordinate system can be done in a computationally efficient way? Fourth, can results on mixed fractional processes be used to inspire new developments in the modeling of high-dimensional fractional systems, namely, the situation where the dimension n itself is comparable to the sample size ν?

This paper is organized as follows. In Sect. 2, we lay out the notation, assumptions and theoretical background of the paper. In Sect. 3, we construct the two-step wavelet-based method and establish its asymptotic properties (Section C of the supplementary material file Abry et al. (2018) contains the analogous results for discrete time measurements). Section 4 contains all Monte Carlo studies. In Sect. 5, we analyze and model the aforementioned tree ring data set. All proofs can be found in Abry et al. (2018), Online Appendix Sections A–E, together with auxiliary results.

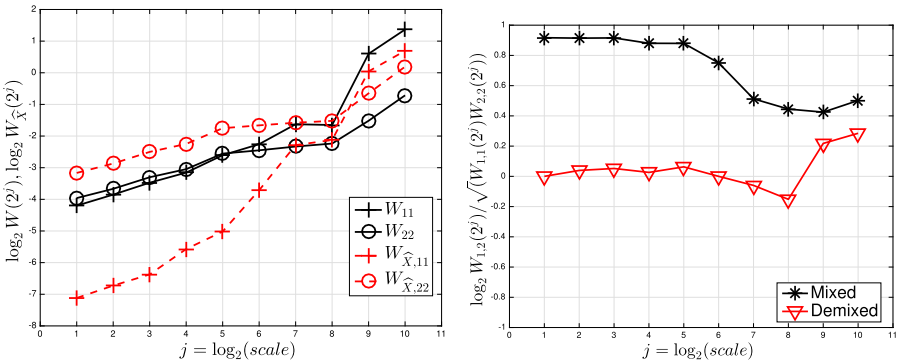


Fig. 1 Left: $\log_2 \widetilde{W}(2^j)_{ii}$ (wavelet variances) versus j for bivariate tree ring data. Before the demixing step ($S1$) (black), both functions $\log_2 \widetilde{W}(2^j)_{11}$ and $\log_2 \widetilde{W}(2^j)_{22}$ show scaling behavior with similar memory parameter values clearly departing from 0 in the stationary range $(-1/2, 1/2)$. This confirms the presence of fractional memory. After the demixing step ($S1$), the functions $\log_2 \widetilde{W}(2^j)_{11}$ and $\log_2 \widetilde{W}(2^j)_{22}$ still display scaling behavior, yet with quite distinct memory parameters, and clearly departing from 0. Right: wavelet coherence function. Before the demixing step ($S1$) (black), the wavelet coherence function shows significant (and nearly equivalent) correlations across all scales. After the demixing step ($S1$) (red), it shows nearly zero correlation at all scales, which is evidence of successful demixing. (Color figure online)

2 Preliminaries

Before developing the statistical methodology in Sect. 3, in this section we set the notation and some basic definitions (Sect. 2.1), we recap the connection between aggregation and mixed processes (Sect. 2.2) and express the framework of assumptions on the observable process Y and on the wavelet basis for continuous time measurements (Sect. 2.3).

2.1 Notation and definitions

The dimension of the mixed process Y is denoted by $n \geq 2$ throughout the paper.

We shall use the following matrix notation. $M(m, n, \mathbb{R})$ is the vector space of all $m \times n$ real-valued matrices, whereas $M(n, \mathbb{R})$ is a shorthand for $M(n, n, \mathbb{R})$. $GL(n, \mathbb{R})$ is the set of invertible matrices (general linear group), $O(n)$ is the set of matrices O such that $OO^* = I = O^*O$ (orthogonal group), where * represents the matrix adjoint and ^{T} is reserved for vector transpose. $\mathcal{S}(n, \mathbb{R})$, $\mathcal{S}_{\geq 0}(n, \mathbb{R})$ and $\mathcal{S}_{>0}(n, \mathbb{R})$ are, respectively, the space of symmetric, the cone of symmetric positive semidefinite and the cone of symmetric positive definite matrices. The symbol $\mathbf{0}$ represents a vector or matrix of zeroes. A block-diagonal matrix with main diagonal blocks $\mathcal{P}_1, \ldots, \mathcal{P}_n$ or m times repeated diagonal block $\mathcal{P}$ is represented by

$$\operatorname{diag}(\mathcal{P}_1, \ldots, \mathcal{P}_n), \quad \operatorname{diag}_m(\mathcal{P}), \tag{2.1}$$

respectively. The symbol $\|\cdot\|$ represents a generic matrix or vector norm. The ℓ_p entrywise norm of the matrix A is denoted by

$$\|A\|_{\ell_p} = \|(a_{i_1,i_2})_{\substack{i_1=1,\ldots,m \\ i_2=1,\ldots,n}}\|_{\ell_p} = \Big(\sum_{i_1=1}^{m}\sum_{i_2=1}^{n} |a_{i_1,i_2}|^p\Big)^{1/p}. \tag{2.2}$$

The Fourier transform of any function $f \in L^2(\mathbb{R})$ is defined by

$$\widehat{f}(x) = \int_{\mathbb{R}} f(t) e^{-\mathbf{i}xt} dt,$$

where equality holds in the $L^2(\mathbb{R})$ sense. For $S = (s_{i_1,i_2})_{i_1,i_2=1,\ldots,n} \in M(n, \mathbb{R})$, let

$$\begin{aligned} \mathrm{vec}_{\mathcal{S}}(S) &= (s_{11}, s_{21}, \ldots, s_{n1}, s_{22}, s_{32}, \ldots, s_{n2}, \ldots, s_{nn}), \\ \mathrm{vec}_{\mathcal{D}}(S) = (s_{11}, s_{22}, \ldots, s_{nn}), \quad \mathrm{vec}(S) &= (s_{11}, \ldots, s_{n1}, s_{12}, \ldots, s_{n2}, \ldots, s_{nn}). \end{aligned} \tag{2.3}$$

In other words, the operator $\mathrm{vec}_{\mathcal{S}}(\cdot)$ vectorizes the lower triangular entries of S, $\mathrm{vec}_{\mathcal{D}}(\cdot)$ vectorizes the diagonal entries of S, and $\mathrm{vec}(\cdot)$ vectorizes all the entries of S. Note that the expressions in (2.3) are defined as row vectors; this will make the notation simpler in several statements. When establishing bounds, C denotes a positive constant whose value can change from one inequality to the next.

A function ψ is called a wavelet when it satisfies

$$\int_{\mathbb{R}} \psi^2(t) dt = 1, \quad \int_{\mathbb{R}} t^q \psi(t) dt = 0, \quad q = 0, 1, \ldots, N_\psi - 1, \quad \int_{\mathbb{R}} t^{N_\psi} \psi(t) dt \neq 0, \tag{2.4}$$

for some number $N_\psi \in \mathbb{N}$ of so-named vanishing moments (Daubechies 1992). For any stochastic process Y taking values in $\mathbb{R}^n$, the vector wavelet transform of Y is naturally defined as

$$\mathbb{R}^n \ni D(2^j, k) = \int_{\mathbb{R}} 2^{-j/2} \psi(2^{-j} t - k) Y(t) dt, \quad j \in \mathbb{N} \cup \{0\}, \quad k \in \mathbb{Z}, \tag{2.5}$$

provided the integral in (2.5) exists in an appropriate sense. Under conditions [see (3.1)], the wavelet spectrum (variance) at scale j is the positive semidefinite matrix

$$\mathbb{E} D(2^j, k) D(2^j, k)^* = \mathbb{E} D(2^j, 0) D(2^j, 0)^* =: \mathbb{E} W(2^j), \tag{2.6}$$

regardless of k, and its natural estimator, the sample wavelet variance, is the random matrix

$$W(2^j) = \frac{1}{K_j} \sum_{k=1}^{K_j} D(2^j, k) D(2^j, k)^*, \quad K_j = \frac{\nu}{2^j}, \quad j = j_1, \ldots, j_m, \tag{2.7}$$

for a total of

$$\nu \text{ available (wavelet) data points.} \tag{2.8}$$

2.2 Aggregation and mixed processes

Recent work (Chan and Tsai 2010; Tsai et al. 2018) has established the connection between aggregation procedures and the emergence of mixed processes. We sketch the basic idea for the reader's convenience. Many Langevin-type SDE systems have been proposed with the purpose of modeling anomalous diffusion and non-Markovian dynamics (see, for instance, Mason and Weitz 1995; Kou 2008; Didier et al. 2012; Lysy et al. 2016). A parsimonious multivariate fractional SDE model is given by

$$dY(t) = \Phi Y(t) dt + \Sigma d B_{\mathbf{h}}(t), \quad t \geq 0, \quad -\Phi, \Sigma \in \mathcal{S}_{>0}(n, \mathbb{R}), \tag{2.9}$$

where $B_{\mathbf{h}}(t) = (B_{h_1}(t), \ldots, B_{h_n}(t))^T$ is a vector of independent fBm entries. In (2.9), the Hurst parameters satisfy the relations

$$h_i = \delta_i - \frac{1}{2}, \quad \frac{1}{2} < \delta_i < \frac{3}{2}, \quad i = 1, \ldots, n, \tag{2.10}$$

where $\delta_i, i = 1, \ldots, n$ are the associated memory parameters. The solution $\{Y(t)\}_{t\geq 0}$ of (2.9) exists a.s. and generalizes the univariate fractional Ornstein–Uhlenbeck process (c.f. Cheridito et al. 2003; Prakasa Rao 2010). Consider the case where the continuous time process $\{Y(t)\}_{t\geq 0}$ is digitalized by aggregation over intervals of length $\triangle$ as in expression (1.3). Then, as $\triangle \to \infty$,

$$-\operatorname{diag}(\triangle^{-h_1}, \ldots, \triangle^{-h_n})\Sigma^{-1}\Phi Y_s^{\triangle} \xrightarrow{\mathcal{L}} (B_{h_1}(s) - B_{h_1}(s-1), \ldots, B_{h_n}(s) - B_{h_n}(s-1))^T, \tag{2.11}$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence of the finite dimensional distributions (Tsai et al. 2018, p. 5). Therefore, for large $\triangle$, the aggregate process $Y_s^{\triangle}$ can be approximated by the mixed process

$$\widetilde{Y}_s := PX_s, \quad s \in \mathbb{N} \cup \{0\}. \tag{2.12}$$

Recall that fGn is the increment process of fBm. In (2.12), X_s is a vector of n independent fGn entries with Hurst parameters (2.10) and $P = -\Phi^{-1}\Sigma\operatorname{diag}(\triangle^{h_1}, \ldots, \triangle^{h_n})$. Note that the process (2.12) is a particular case of (1.1), with the latter restricted to discrete time.

2.3 Assumptions

Throughout the paper, unless otherwise stated, we assume the observable process has the mixed form $Y = PX$ for some coordinates matrix P, where X is made up of entrywise fractional processes with stationary increments of any order, including zero.

The precise statements appear in assumptions $(A1)$, $(A2)$ and $(A3)$ below. They describe, respectively, the covariance structure of the hidden process X in the Fourier domain, and regularity conditions on P and on the high frequency behavior of the components of X.

Assumption $(A1)$: the observed process has the mixed form (1.1), where $X_i, i = 1, \ldots, n$, in (1.2) is either a N_ith ($N_i \geq 1$) order covariance stationary process with harmonizable representation

$$\{X_i(t)\}_{t\in\mathbb{R}} = \left\{ \int_{\mathbb{R}} \frac{e^{\mathbf{i}tx} - \sum_{l=0}^{N_i-1} \frac{1}{\ell!}(\mathbf{i}tx)^{\ell}}{(\mathbf{i}x)^{N_i}} |x|^{-(\delta_i - N_i)} g_i(x) \widetilde{B}(dx) \right\}_{t\in\mathbb{R}}, \quad N_i - 1/2 \leq \delta_i < N_i + 1/2, \tag{2.13}$$

or a covariance stationary process (i.e., $N_i = 0$) with harmonizable representation

$$\{X_i(t)\}_{t\in\mathbb{R}} = \left\{ \int_{\mathbb{R}} e^{\mathbf{i}tx} \frac{e^{\mathbf{i}x} - 1}{\mathbf{i}x} |x|^{-\delta_i} g_i(x) \widetilde{B}(dx) \right\}_{t\in\mathbb{R}}, \quad -1/2 \leq \delta_i < 1/2. \tag{2.14}$$

By convention, the so-named memory parameters are ordered as

$$-1/2 \leq \delta_1 < \delta_2 < \ldots < \delta_n. \tag{2.15}$$

In (2.13) and (2.14), $\widetilde{B}(dx)$ is a Gaussian random measure satisfying $\widetilde{B}(-dx) = \overline{\widetilde{B}(dx)}$ and $\mathbb{E}|\widetilde{B}(dx)|^2 = dx$.

Assumption $(A2)$:

$$P \in GL(n, \mathbb{R}), \quad \|\mathbf{p}_{\cdot\ell}\| = 1, \tag{2.16}$$

where $\mathbf{p}_{\cdot\ell}$, $\ell = 1, \ldots, n$, are the column vectors of P. Moreover,

$$p_{i(\ell),\ell} > 0, \quad \ell = 1, \ldots, n, \tag{2.17}$$

where $i(\ell) = \min\{i = 1, \ldots, n : p_{i\ell} > 0\}$.

ASSUMPTION $(A3)$: the $\mathbb{C}$-valued functions $g_i(x)$ in (2.13) and (2.14) are bounded and, for some small $\eta > 0$, satisfy

$$||g_i(x)|^2 - |g_i(0)|^2| < L|x|^\beta, \quad |g_i(0)|, L > 0, \quad i = 1, \ldots, n, \tag{2.18}$$

for any $x \in (-\eta, \eta)$. In (2.18), $\beta \in (1, 2]$ and satisfies

$$\beta + 1 < 2\delta_1 + 2\alpha \tag{2.19}$$

for some

$$\alpha > 1. \tag{2.20}$$

The Gaussian scheme represented by assumptions $(A1–3)$ is quite general. Mathematically speaking, it is closely related to the widely used framework constructed in Moulines et al. (2007a, b, 2008), except that it is defined in continuous time (n.b.: the extension to discrete time is provided in Section C of the supplementary file Abry et al. 2018). Assumption $(A1)$ expresses the entrywise processes $X_i, i = 1, \ldots, n$, in terms of Cramér-Wold-type stochastic integral representations (see Brockwell and Davis 1991, Section 4.8; Yaglom 1987, Section 24.3). Relations (2.13) and (2.14) are equivalent, respectively, to the covariance functions

$$\begin{aligned}&\mathbb{E}X_i(s)X_i(t)\\ &= \int_{\mathbb{R}} \Big(\frac{e^{\mathbf{i}sx} - \sum_{\ell=0}^{N_i-1} \frac{1}{\ell!}(\mathbf{i}sx)^\ell}{(\mathbf{i}x)^{N_i}}\Big)\overline{\Big(\frac{e^{\mathbf{i}tx} - \sum_{\ell=0}^{N_i-1} \frac{1}{\ell!}(\mathbf{i}tx)^\ell}{(\mathbf{i}x)^{N_i}}\Big)}|x|^{-2(\delta_i - N_i)}|g_i(x)|^2 dx\end{aligned} \tag{2.21}$$

and

$$\mathbb{E}X_i(s)X_i(t) = \int_{\mathbb{R}} e^{\mathbf{i}(s-t)x} \left|\frac{e^{\mathbf{i}x} - 1}{\mathbf{i}x}\right|^2 |x|^{-2\delta_i}|g_i(x)|^2 dx \tag{2.22}$$

for $s, t \in \mathbb{R}$ and $i = 1, \ldots, n$. Note that, in (2.15), one incurs no loss of generality by assuming that the memory parameters are disposed in ascending order, since a permutation matrix can always be incorporated into the mixing (coordinates) matrix P. Assumption $(A2)$ simply states that the columns of P are linearly independent unit vectors and sets an orientation for each of the latter. Assuming P is invertible, the observable process $Y(t) = PX(t)$ is well-defined and identifiable without (2.16) or (2.17). However, the latter conditions are necessary for the consistent estimation of a demixing matrix (see Definition 3.1). Assumption $(A3)$ is typical in a semiparametric estimation setting (e.g., Robinson 1995a; Moulines and Soulier 2003). Note that (2.18) only describes the behavior of the high frequency functions $g_i(x)$, $i = 1, \ldots, n$, around the origin, where larger values of β correspond to greater smoothness.

Example 2.1 For a fixed $i = 1, \ldots, n$, a choice of triple $(\delta_i, N_i, g_i(x))$ determines entry X_i of the hidden process X. For example, for a fixed memory parameter value δ_i, a fBm, a fGn (in continuous time), or a fractional Ornstein–Uhlenbeck process corresponds, respectively, to the instances

$$g_i(x) \equiv C \quad (N_i = 1), \quad g_i(x) \equiv C \quad (N_i = 0), \quad \text{or}$$
$$g_i(x) = \frac{\mathbf{i}x}{e^{\mathbf{i}x} - 1}\frac{C}{\lambda + \mathbf{i}x} \quad (N_i = 0)$$

for some $\lambda > 0$. The instance $N_i = 0$ and $g_i(x) = \frac{C\mathbf{i}x}{e^{\mathbf{i}x}-1}|x|^{\delta_i}(1-e^{-\mathbf{i}x})^{-\delta_i}1_{[-\pi,\pi)}$ corresponds, in discrete time, to FARIMA$(0, \delta_i, 0)$ (e.g., Taqqu 2003, Section 6). In terms of the covariance functions (2.21) and (2.22), these four instances can be expressed as

$$\mathbb{E}X_i(s)X_i(t) = C^2 \int_{\mathbb{R}} \left\{ \begin{array}{cc} (e^{\mathbf{i}sx}-1)(e^{-\mathbf{i}tx}-1)|x|^{-2}\ |x|^{-2(\delta_i-1)} & (\text{fBm}) \\ e^{\mathbf{i}(s-t)x}|e^{\mathbf{i}x}-1|^2|x|^{-2}\ |x|^{-2\delta_i} & (\text{fGn}) \\ e^{\mathbf{i}(s-t)x}|x|^{-2\delta_i}|\lambda^2+x^2|^{-1} & (\text{fOU}) \\ e^{\mathbf{i}(s-t)x}|1-e^{-\mathbf{i}x}|^{-2\delta_i}1_{[-\pi,\pi)}(x) & (\text{FARIMA}(0,\delta_i,0)) \end{array} \right\} dx.$$

Example 2.2 If the high frequency functions $g_i(x)$ are constant and $N_i - 1/2 < \delta_i < N_i + 1/2$, $N_i \geq 1$, $i = 1 \ldots, n$, then the observed process Y satisfies the operator self-similarity property (1.5), with Hurst matrix

$$H = P\,\text{diag}\Big(\delta_1 - \frac{1}{2}, \ldots, \delta_n - \frac{1}{2}\Big)P^{-1}. \tag{2.23}$$

If, in particular, $N_i = 1, i = 1, \ldots, n$, then Y is an instance of operator fractional Brownian motion, namely, a Gaussian, operator self-similar, stationary increment process (Mason and Xiao 2002; Didier and Pipiras 2011, 2012).

Remark 2.1 For a fixed i, in the boundary cases $\delta_i = N_i - 1/2$ the finiteness of second moments in (2.13) and (2.14) implies that the high frequency function $g_i(x)$ must decay fast enough as $x \to \infty$ so as to make up for the lack of integrability of the power law. On the range $\delta_i < -1/2$, see Remark C.2 in Section C.

Remark 2.2 The assumption (2.15) that memory parameters are pairwise distinct is lifted in Corollary 3.1.

Remark 2.3 Mathematically speaking, it is natural to ask how useful it is to consider the model (1.1) with a full rank matrix $P \in M(m, n, \mathbb{R})$, where $m \neq n$. However, both cases $m > n$ and $m < n$ fall outside the scope of this paper. When $m > n$, the observed process Y is improper, namely, its finite dimensional distributions are contained in a proper subspace of $\mathbb{R}^n$ for some $t \neq 0$ (even if, in addition, the high frequency functions $g_i, i = 1, \ldots, n$, are constant, Y cannot be operator self-similar: see Example 2.2 or Hudson and Mason 1982). When $m < n$, the spectral densities involved are potentially much more complicated, with sums of power laws. Either situation demands the construction of particular methods.

In Sect. 3, we implicitly make the following assumptions on the underlying wavelet basis. Hence, we omit them in statements.

ASSUMPTION $(W1)$: $\psi \in L^1(\mathbb{R})$ is a wavelet function satisfying (2.4) for

$$N_\psi \geq N_n + 1, \tag{2.24}$$

where N_n is as in (2.13) or (2.14).

ASSUMPTION $(W2)$:

$$\text{supp}(\psi) \text{ is compact.} \tag{2.25}$$

ASSUMPTION $(W3)$: For $\alpha > 1$ as in (2.20),

$$\sup_{x\in\mathbb{R}} |\widehat{\psi}(x)|(1+|x|)^\alpha < \infty. \tag{2.26}$$

Conditions (2.24) and (2.25) imply that $\widehat{\psi}(x)$ exists, is everywhere differentiable and its first $N_\psi - 1$ derivatives are zero at $x = 0$ (see Mallat, Theorem 7.4). This implies, using a Taylor expansion, that

$$|\widehat{\psi}^{(\ell)}(x)| = O(|x|^{N_\psi - \ell}), \quad \ell = 0, \ldots, N_\psi, \quad x \to 0. \tag{2.27}$$

Condition (2.26), in turn, implies that ψ is continuous (see Mallat 1999, Theorem 6.1).

Example 2.3 Assumptions $(W1–W3)$ are generally satisfied, for example, by a Daubechies wavelet function (see Daubechies 1992; Moulines et al. 2008, p. 1927, or Mallat 1999). If ψ is a Daubechies wavelet with N_ψ vanishing moments, $\text{supp}(\psi) = [0, 2N_\psi - 1]$ (see Mallat 1999, Proposition 7.4).

Remark 2.4 Assumption $(W1)$ requires using a number of vanishing moments N_ψ larger than the unknown integration order N_n. In practice, though, the latter parameter is rarely greater than 2, so the requirement is easily met even for low values of N_ψ.

Remark 2.5 Section C, on measurements in discrete time, requires a slightly different set of assumptions on the wavelet basis.

3 Wavelet-based estimation (continuous time)

In this section, we construct the two-step wavelet-based estimator. We assume a continuous time sample path of Y is available, since discrete time measurements do not qualitatively change the nature of the results, as shown in Section C.

Before getting into detailed definitions and discussions, we briefly motivate the two-step method in Sect. 3.1. Sections 3.2 and 3.3 contain the main mathematical results of the paper, namely, the construction of the demixing step $(S1)$ and the post-demixing memory parameter estimation step $(S2)$, respectively. Note that $(S1)$ only involves wavelet analysis at fixed scales, whereas $(S2)$ generally requires taking a coarse scale limit $a(\nu)2^j \to \infty$, due to the lack of exact self-similarity in (2.13) and (2.14).

To develop the two-step method, we need some basic and asymptotic properties of the wavelet transform and variance of the process Y at fixed scales. Under assumptions $(A1–3)$ and $(W1–3)$, the wavelet domain process $\{D(2^j, k)\}_{k\in\mathbb{Z}}$ is stationary in k, namely,

$$\{D(2^j, k+h)\}_{k\in\mathbb{Z}} \overset{\mathcal{L}}{=} \{D(2^j, k)\}_{k\in\mathbb{Z}}, \quad h \in \mathbb{Z} \tag{3.1}$$

[see Proposition B.1, $(P2)$]. Consequently, the wavelet variance matrices $\mathbb{E}W(2^j)$ and $W(2^j)$, as defined by (2.6) and (2.7), respectively, do not depend on k. In addition, the vectorized sample wavelet variance $\text{vec}_{\mathcal{S}} W(2^j)$ at a fixed set of octaves is asymptotically normal (Theorem B.1).

3.1 Overview of the two-step method

We can express the expected value of the sample wavelet variance matrix as

$$\mathbb{E}W(2^j) = P\mathcal{E}(2^j)^{1/2}\text{diag}(2^{2j\delta_1}, \ldots, 2^{2j\delta_n})\mathcal{E}(2^j)^{1/2}P^*, \quad j = j_1, \ldots, j_m, \tag{3.2}$$

where $\mathcal{E}(2^j)$ is a diagonal matrix-valued auxiliary function [see Proposition B.1, $(P3)$, expressions (B.1) and (B.5)]. Consider the matrix

$$B = \Pi\text{diag}(2^{-j_1\delta_1}, \ldots, 2^{-j_1\delta_n})\mathcal{E}(2^{j_1})^{-1/2}P^{-1}, \tag{3.3}$$

where Π is (any) diagonal matrix whose nonzero entries are ± 1. Note that B jointly diagonalizes all matrices (3.2), namely,

$$B\mathbb{E}W(2^j)B^* \text{ is diagonal for any } j. \tag{3.4}$$

On the other hand, by expressions (3.2) at scale $a(\nu)2^j$ and (3.3),

$$B\,\mathbb{E}W(a(\nu)2^j)\,B^* \sim \Pi \mathrm{diag}(2^{-j_1\delta_1},\ldots,2^{-j_1\delta_n})\mathcal{E}(2^{j_1})^{-1/2}\mathcal{E}_g^{1/2} \cdot \mathrm{diag}\Big((a(\nu)2^j)^{2\delta_1},\ldots,(a(\nu)2^j)^{2\delta_n}\Big)\mathcal{E}_g^{1/2}\mathcal{E}(2^{j_1})^{-1/2}\mathrm{diag}(2^{-j_1\delta_1},\ldots,2^{-j_1\delta_n})\Pi, \quad (3.5)$$

as $\nu \to \infty$. In (3.5), we use the property that

$$\mathcal{E}(a(\nu)2^j) \to \mathcal{E}_g, \quad \nu \to \infty, \quad (3.6)$$

for a diagonal matrix of constants $\mathcal{E}_g$ depending on $g_1,\ldots,g_n$ [see Proposition B.1, $(P3)$]. Due to demixing, all matrices on the right-hand side of the asymptotic relation (3.5) are diagonal. Moreover, only one is a function of the asymptotic scale $a(\nu)2^j$, namely,

$$\mathrm{diag}\Big((a(\nu)2^j)^{2\delta_1},\ldots,(a(\nu)2^j)^{2\delta_n}\Big).$$

Therefore, up to a known constant, the memory parameters $\delta_1,\ldots,\delta_n$ appear as the *slope* of the main diagonal entries of (3.5) on a log-log scale.

With the random matrices $W(\cdot)$ in place of $\mathbb{E}W(\cdot)$, this reasoning holds up to stochastic error. Therefore, using the (random) sample wavelet matrices $W(a(n)2^j)$, we can construct a two-step estimation method consisting of, first, jointly diagonalizing the latter matrices, and, second, running a log-regression procedure on the diagonalized matrices.

3.2 Wavelet-based demixing (step (*S*1))

The joint diagonalization of two matrices is a well-known problem. For the case of symmetric matrices, its description and full characterization can be stated as follows [see Theorem 4.5.17, (b), in Horn and Johnson (1985)]. Suppose C_0 and C_1 are symmetric and C_0 is nonsingular. Then, there are a nonsingular $S \in M(n,\mathbb{R})$ and complex diagonal matrices Λ_0 and Λ_1 such that

$$C_0 = S\Lambda_0 S^*, \quad C_1 = S\Lambda_1 S^*, \quad (3.7)$$

if and only if the matrix $C_0^{-1}C_1$ is diagonalizable (in its Jordan form). In light of this, we can cast a joint diagonalization algorithm in the form of pseudocode.

Pseudocode for exact joint diagonalization (EJD)

Input: C_0, C_1 are symmetric matrices and the former is positive definite;
Step 1: set $W = C_0^{-1/2}$ so that $C_0^{-1} = W^*W$;
Step 2: compute $Q \in O(n)$ (orthogonal matrix) in the spectral decomposition $WC_1W^* = Q^*D_1Q$;
Step 3: compute the demixing matrix $B := QW$;
Step 4: stop and exit.

(n.b.: **Steps 1–4** of the EJD algorithm should not be confused with steps $(S1)$ and $(S2)$ of the proposed wavelet-based estimation method).

Example 3.1 In view of (3.2), it is clear that $C_0 = \mathbb{E}W(2^{J_1})$, $C_1 = \mathbb{E}W(2^{J_2})$, $J_1 < J_2$, can be jointly diagonalized, where the underlying process is defined in (1.1) under the assumptions $(A1-2)$. In addition,

$$C_0^{-1}C_1 = (P^*)^{-1}\Big(\mathrm{diag}(2^{2(J_2-J_1)\,\delta_1},\ldots,2^{2(J_2-J_1)\,\delta_n})\mathcal{E}(2^{J_1})^{-1}\mathcal{E}(2^{J_2})\Big)P^*.$$

This expression constitutes a diagonal Jordan decomposition, whence (3.7) holds.

The proposed wavelet-based estimator $\widehat{B}_\nu$ of a demixing matrix is defined next.

Definition 3.1 (*(S1) demixing step*) Consider two octaves $0 \leq J_1 < J_2$ for which $\mathcal{E}(2^{J_1})$, $\mathcal{E}(2^{J_2})$ have strictly positive diagonal entries and

$$\text{diag}(2^{2(J_2-J_1)\delta_1}, \ldots, 2^{2(J_2-J_1)\delta_n})\mathcal{E}(2^{J_1})^{-1}\mathcal{E}(2^{J_2}) \quad \text{has pairwise distinct diagonal entries.} \tag{3.8}$$

For $\nu \in \mathbb{N}$, the wavelet-based demixing estimator $\widehat{B}_\nu$ is the output of the EJD algorithm when setting

$$C_0 = W(2^{J_1}) \text{ and } C_1 = W(2^{J_2}). \tag{3.9}$$

In the following theorem, we establish the consistency and asymptotic normality of the estimator $\widehat{B}_\nu$. This involves characterizing the set of solutions provided by the EJD algorithm, which includes a class of non-identifiability factors denoted by $\mathcal{I}$ in the theorem.

Theorem 3.1 *For $j \in \mathbb{N}$, let $\mathcal{E}(2^j)$ be as in (B.5). Also let*

$$\mathcal{I} = \{\Pi \in M(n, \mathbb{R}) : \Pi \textit{ has the form } diag(\pm 1, \ldots, \pm 1)\}. \tag{3.10}$$

(i) *Then,*

$$\mathcal{M}_{EJD} = \{\Pi\, diag(2^{-J_1\delta_1}, \ldots, 2^{-J_1\delta_n})\mathcal{E}(2^{J_1})^{-1/2}P^{-1}, \Pi \in \mathcal{I}\} \tag{3.11}$$

is the set of matrix solutions produced by the EJD algorithm when setting

$$C_0 = \mathbb{E}W(2^{J_1}) \textit{ and } C_1 = \mathbb{E}W(2^{J_2}); \tag{3.12}$$

(ii) *in addition, assume condition (3.8) holds and suppose the (random) matrix $Q^* = \widehat{Q}^* \in O(n)$ from **Step 2** of the EJD algorithm is chosen as to satisfy condition (2.17). Then, there is $\Pi \in \mathcal{I}$ such that the obtained estimator sequence $\{\widehat{B}_\nu\}_{\nu\in\mathbb{N}}$ satisfies*

$$\widehat{B}_\nu \xrightarrow{P} \Pi\, diag(2^{-J_1\delta_1}, \ldots, 2^{-J_1\delta_n})\mathcal{E}(2^{J_1})^{-1/2}P^{-1}, \quad \nu \to \infty; \tag{3.13}$$

(iii) *an estimator sequence $\{\widehat{B}_\nu\}_{\nu\in\mathbb{N}}$ as described in (ii) satisfies*

$$\sqrt{\nu}(vec(\widehat{B}_\nu - \Pi\, diag(2^{-J_1\delta_1}, \ldots, 2^{-J_1\delta_n})\mathcal{E}(2^{J_1})^{-1/2}P^{-1}))^T \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_F(J_1, J_2)) \tag{3.14}$$

for some matrix $\Pi \in \mathcal{I}$, where the asymptotic covariance matrix $\Sigma_F(J_1, J_2)$ is a function of F, and F is defined in Theorem B.1, with $m = 2$.

Remark 3.1 Note that, for $J_1 < J_2$ and under (2.15), condition (3.8) always holds for large enough J_1, J_2. This is a consequence of Proposition B.1, $(P4)$, and of the fact that $\mathcal{E}(2^{J_1})^{-1}\mathcal{E}(2^{J_2}) \to I$ as $J_1, J_2 \to \infty$.

Remark 3.2 By (3.13), the sequence $\widehat{B}_\nu^{-1}$ has a limit in probability of the form $B^{-1} := P\text{diag}(\beta_1, \ldots, \beta_n)$, $|\beta_i| \neq 0$, $i = 1, \ldots, n$, i.e., involving a non-identifiability factor post-multiplying the mixing matrix P. However, note that $\Delta := P\text{diag}(\delta_1, \ldots, \delta_n)P^{-1} = B^{-1}\text{diag}(\delta_1, \ldots, \delta_n)B$, i.e., the columns of B^{-1} consist of (non-unit) eigenvectors of the memory matrix Δ [c.f. (B.4)]. Consequently, $\widehat{\Delta} := \widehat{B}_\nu^{-1}\text{diag}(\widehat{\delta}_1, \ldots, \widehat{\delta}_n)\widehat{B}_\nu$ is a natural estimator of the latter, where $\widehat{\delta}_1, \ldots, \widehat{\delta}_n$ are univariate (e.g., wavelet-based) estimators of the individual memory parameters obtained from the demixed process.

Producing a direct estimator of P is straightforward. Just normalize each column of the matrix estimator $\widehat{B}_\nu^{-1}$ and multiply it by -1 if necessary to arrive at a matrix $\widehat{P}$ with positive diagonal entries (cf. (2.16)). This procedure is used in Sect. 4.

Remark 3.3 More precisely, the covariance matrix in the limit (3.14) can be written as $\Sigma_F(J_1, J_2) = A_3 \Sigma_2 A_3^*$, where Σ_2 and A_3 are given by expressions (B.50) and (B.51), respectively. It is clear that the expression for $\Sigma_F(J_1, J_2)$ is quite intricate, and the construction of theoretical confidence intervals is a matter for future investigation (c.f. Wendt et al. 2017).

3.3 Wavelet-based estimation of memory parameters after demixing/changing the coordinates (step (*S2*))

Starting with the output of step (*S*1) (Sect. 3.2), let $\widehat{B}_\nu$ be the demixing matrix described in (3.13). Then, the demixed process is defined by

$$\widehat{X}(t) := \widehat{B}_\nu Y(t), \quad t \in \mathbb{R}, \tag{3.15}$$

of which only ν (wavelet) data points are available [c.f. (2.8)]. Establishing a semiparametric estimator requires taking coarse scale limits as the sample size ν goes to infinity. So, let $a(\nu)$ be a divergent dyadic sequence that is slow by comparison to ν [see (3.23)]. For $j \in \mathbb{N}$, let

$$W_{\widehat{X}}(a(\nu)2^j), \quad \mathbb{E}W_X(a(\nu)2^j), \tag{3.16}$$

be the sample wavelet variance of $\widehat{X}$ and the wavelet variance of the hidden process X, respectively, at scale $a(\nu)2^j$. We are now in a position to define an estimator of the vector of memory parameters $\boldsymbol{\delta}^T = (\delta_1, \ldots, \delta_n)$ of the hidden process X. The estimator stems from log-linear weighted least squares regressions on each main diagonal entry of the matrices $W_{\widehat{X}}(a(\nu)2^j)$ for different j.

Definition 3.2 (*(S2) Memory parameter estimation step*) Let

$$W_{\widehat{X}}(\cdot)_{ii'}, \quad \mathbb{E}W_X(\cdot)_{ii'}, \quad i, i' = 1, \ldots, n, \tag{3.17}$$

be the (i, i')th entries of the matrices $W_{\widehat{X}}(\cdot)$ and $\mathbb{E}W_X(\cdot)$, respectively. Consider the regression weight vectors

$$\mathbf{w}^i = (w_1^i, \ldots, w_m^i)^T, \tag{3.18}$$

where

$$\sum_{\ell=1}^m w_\ell^i = 0, \quad 2\sum_{\ell=1}^m j_\ell w_\ell^i = 1, \quad i = 1, \ldots, n. \tag{3.19}$$

The wavelet-based estimator of the memory parameters $\delta_1, \ldots, \delta_n$ in (2.15) is obtained by log-regressing the main diagonal terms $W_X(a(\nu)2^j)_{ii}$ on the scale indices $a(\nu)2^j$, $j = j_1, \ldots, j_m$, i.e.,

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\delta}_1 \\ \vdots \\ \widehat{\delta}_n \end{pmatrix} := \begin{pmatrix} \sum_{\ell=1}^m w_\ell^1 \log_2 W_{\widehat{X}}(a(\nu)2^{j_\ell})_{11} \\ \vdots \\ \sum_{\ell=1}^m w_\ell^n \log_2 W_{\widehat{X}}(a(\nu)2^{j_\ell})_{nn} \end{pmatrix}. \tag{3.20}$$

Remark 3.4 The choice of regression weights (3.18) and (3.19) can be made as to minimize bias and compensate for the increasing variance of sample wavelet variances for larger octaves j. From weighted linear regression and some approximations, a common choice is

$$w_j = \frac{1}{\varsigma_j^2} \frac{j\mathcal{S} - \mathcal{S}_j}{\mathcal{S}\mathcal{S}_{jj} - \mathcal{S}_j^2}, \quad j = j_1, \ldots, j_m, \tag{3.21}$$

where

$$\mathcal{S} = \sum_{j=j_1}^{j_m} \frac{1}{\varsigma_j^2}, \quad \mathcal{S}_j = \sum_{j=j_1}^{j_m} \frac{j}{\varsigma_j^2}, \quad \mathcal{S}_{jj} = \sum_{j=j_1}^{j_m} \frac{j^2}{\varsigma_j^2}. \tag{3.22}$$

In (3.21) and (3.22), based on approximations to the distribution of univariate sample wavelet variances, one further sets $\varsigma_j^2 = \zeta(2, n_j/2)/\log_2^2$, where $\zeta(\cdot,\cdot)$ is a generalized Riemann zeta function [see the discussions in Veitch and Abry (1999, pp. 881–882), Abry et al. (2000, p. 73) and Stoev et al. (2002, p. 1880)].

The consistency and asymptotic normality of the estimator $\widehat{\boldsymbol{\delta}}$ are established in the following theorem.

Theorem 3.2 *Let $\widehat{\boldsymbol{\delta}}^T = (\widehat{\delta}_1, \ldots, \widehat{\delta}_n)$ be the estimator defined by (3.20). Let $a(\nu)$ be a dyadic sequence such that*

$$\frac{a(\nu)}{\nu} + \frac{\nu}{a(\nu)^{1+2\beta}} \to 0, \quad \nu \to \infty, \tag{3.23}$$

where β satisfies (2.19). Then, $\widehat{\boldsymbol{\delta}}$ is consistent, namely,

$$\widehat{\boldsymbol{\delta}} \xrightarrow{P} \boldsymbol{\delta}, \quad \nu \to \infty. \tag{3.24}$$

Moreover,

$$\sqrt{\frac{\nu}{a(\nu)}} \left[\begin{pmatrix} \widehat{\delta}_1 \\ \vdots \\ \widehat{\delta}_n \end{pmatrix} - \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix} \right] \xrightarrow{d} \mathcal{N}(0, \mathcal{W}), \quad \nu \to \infty. \tag{3.25}$$

In (3.25), the asymptotic covariance matrix is given by

$$\mathcal{W} = diag((\mathbf{w}^1)^T V(\delta_1)\mathbf{w}^1, \ldots, (\mathbf{w}^n)^T V(\delta_n)\mathbf{w}^n),$$

the weight vectors $\mathbf{w}^i$, $i = 1, \ldots, n$ satisfy (3.19), and the matrix $V(\delta) = \{V_{k_1,k_2}(\delta)\}_{k_1,k_2=1,\ldots,m}$ is defined entrywise by

$$V_{k_1,k_2}(\delta) = \frac{4\pi b_{j_{k_1},j_{k_2}}^{4\delta-1}}{2^{2(j_{k_1}+j_{k_2})\delta} K^2(\delta)} \int_{\mathbb{R}} x^{-4\delta} \left| \widehat{\psi}\left(\frac{2^{j_{k_1}} x}{b_{j_{k_1},j_{k_2}}}\right) \right|^2 \left| \widehat{\psi}\left(\frac{2^{j_{k_2}} x}{b_{j_{k_1},j_{k_2}}}\right) \right|^2 dx, \tag{3.26}$$

where $K(\delta) = \int_{\mathbb{R}} |\widehat{\psi}(x)|^2 |x|^{-2\delta} dx$ and $b_{j_{k_1},j_{k_2}} = \gcd(2^{j_{k_1}}, 2^{j_{k_2}})$.

One consequence of Theorem 3.2 is that the individual memory estimators $\widehat{\delta}_1, \ldots, \widehat{\delta}_n$ are asymptotically independent. In fact, the joint asymptotic distribution of $\widehat{\boldsymbol{\delta}}$, estimated from the demixed process $\widehat{X}$, is equal to that of the joint entrywise wavelet-based estimators of $\delta_1, \ldots, \delta_n$ obtained from the hidden process X (see Remark B.1). In other words, asymptotically, the demixing step ($S1$) washes out the effect of the mixing (coordinates) matrix P on the estimation procedure.

Note that removing the condition (2.15) can alter the limits (3.25) (see Remark B.2).

Example 3.2 In the case where Y has stationary increments (see Example 2.2), together Theorems 3.1 and 3.2 imply that the whole Hurst matrix parameter (coordinates matrix and Hurst eigenvalues) can be estimated.

Remark 3.5 In practice, the choice of $a(\nu)$ involves a statistical compromise. A large value of $a(\nu)$ with respect to ν implies a relatively small bias, but also a relatively large variance. Simulation results suggest the ratio $\nu/a(\nu)2^j$ should be no less than 2^3.

For hypothesis testing, we need to consider the case where some, or all, memory parameters $\delta_1, \ldots, \delta_n$ are equal (c.f. Hualde 2013). In light of Remark B.2, this requires making some change to the assumptions. However, to attain consistency and asymptotic normality in steps $(S1)$ and $(S2)$, it suffices to add minor constraints on the high frequency functions $g_i(x)$, $i = 1, \ldots, n$, and hence replace $(A1)$ and $(A3)$ with the following assumptions.

ASSUMPTION $(A1')$: the observed process Y has the mixed form (1.1), where each component $X_i, i = 1, \ldots, n$, of the hidden process in (1.2) has the form (2.13) or (2.14), and the memory parameters can be ordered as

$$-1/2 \leq \delta_1 = \ldots = \delta_{n_1} < \delta_{n_1+1} = \ldots = \delta_{n_2} < \ldots < \delta_{n_p+1} = \ldots = \delta_n.$$

ASSUMPTION $(A3')$: In addition to satisfying $(A3)$, the high frequency functions $g_i(x)$, $i = 1, \ldots, n$, are such that the matrix $\mathrm{diag}(2^{2(J_2-J_1)\delta_1}, \ldots, 2^{2(J_2-J_1)\delta_n})\mathcal{E}(2^{J_1})^{-1}\mathcal{E}(2^{J_2})$ has pairwise distinct diagonal entries.

Corollary 3.1 *Suppose the mixed process Y satisfies assumptions $(A1')$, $(A2)$ and $(A3')$. Then, the conclusions of Theorems 3.1 and 3.2 hold.*

4 Monte Carlo studies

To study the statistical performance of the two-step wavelet-based method over finite samples, we assumed the hidden process X is made up of 4 independent fBm components observed in discrete time. For notational simplicity, write $X := B_{\mathbf{h}}$, $Y := B_H$ (see Example 2.2). In this case, recall that relation (2.23) holds between the memory parameters and the individual Hurst exponents, namely,

$$h_i = \delta_i - \frac{1}{2} \in (0, 1), \quad i = 1, \ldots, n.$$

We simulated $R = 10,000$ sample paths of sizes ranging from $n = 2^{10}$ to 2^{20} (results are reported for the smallest and largest sample size only) with individual Hurst parameters $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and mixing matrix

$$P = \begin{pmatrix} 0.6834 & -0.7142 & 0.6960 & -0.1165 \\ -0.0096 & 0.4539 & -0.0908 & 0.7740 \\ 0.4771 & -0.2345 & 0.3359 & -0.4243 \\ 0.5525 & -0.4784 & -0.6281 & 0.4553 \end{pmatrix} \tag{4.1}$$

(see also Remark 4.2 on the choice of P). The entrywise Hurst exponents are denoted by $h_{X,i}$, $h_{Y,i}$, $i = 1, \ldots, n$, whereas $h_{\widetilde{X},i}$, $i = 1, \ldots, n$, denote the Hurst exponents of the demixed sequence $\widetilde{X} = \widehat{P^{-1}}Y$ for normalized demixing matrix estimates $\widehat{P^{-1}}$.

The results consist of comparisons of the Monte Carlo log-averages of the sample wavelet variance $\langle \log_2 \widetilde{W}_X(2^j)_{ii} \rangle$, $\langle \log_2 \widetilde{W}_Y(2^j)_{ii} \rangle$ and $\langle \log_2 \widetilde{W}_{\widetilde{X}}(2^j)_{ii} \rangle$ ($\langle \cdot \rangle$ represents a Monte Carlo average) for each of the $n = 4$ components for the sample sizes 2^{20} and 2^{10} (Figs. 2, 5); boxplots for $\widehat{h}_{X,i} - h_i$, $\widehat{h}_{Y,i} - h_i$ and $\widehat{h}_{\widetilde{X},i} - h_i$, $i = 1, 2, 3, 4$ (Figs. 3, 6); and boxplots for the 16 entries of $\widehat{P^{-1}}P - I$ (Figs. 4, 7). Following the procedure described in Remark 3.2, the columns of $\widehat{P}$ were adjusted as to eliminate the non-identifiability factor. In all cases, the sample wavelet variance matrices were computed based on Daubechies wavelet filters with $N_\psi = 2$ vanishing moments. Using a different wavelet with $N_\psi \geq 2$ leads to similar conclusions.

In Figs. 2 and 5, as expected for the mixed data Y all components of $\langle \log_2 \widetilde{W}_Y(2^j)_{ii} \rangle$ display patent departures from the original data $\langle \log_2 \widetilde{W}_X(2^j)_{ii} \rangle$. After demixing, all components of $\langle \log_2 \widetilde{W}_{\widetilde{X}}(2^j)_{ii} \rangle$ remarkably superimpose those of $\langle \log_2 \widetilde{W}_X(2^j)_{ii} \rangle$, with the possible exception of a few coarse scales for $h = 0.2$ and 0.4. In addition, the boxplots in Figs. 3 and 6 show that the Monte Carlo distributions for $\widehat{h}_{\widetilde{X},i} - h_i$ resemble those of $\widehat{h}_{X,i} - h_i$, which illustrates the successful demixing of Y. Figures 4 and 7 further indicate that $\widehat{P^{-1}}$ is very well estimated with negligible biases. In all comparisons, as expected the observed estimator properties improve significantly when passing from the relatively small sample size 2^{10} to the large sample size 2^{20}. In addition, simulation results not displayed also show that the standard deviation of the estimates decreases with the sample size according to the scaling ratio $C/\sqrt{\nu}$ for some $C > 0$, as anticipated.

Remark 4.1 Theorems 3.2 and C.2 (for continuous and discrete time, respectively) leave open the question of how to optimally choose the octaves $J_1 < J_2$. For multiple choices of wavelet octaves, namely, $J_1 = 1$ [which involves the largest number of sum terms in (2.7)] and $J_2 = 2, \ldots, 6$, Table 1 shows the performance of the individual Hurst exponents' estimators in terms of Monte Carlo bias, standard deviation and (square root) mean squared error. For sample sizes 2^{20} and 2^{10}, the results indicate that for low values of the Hurst exponents, the use of two widely separated wavelet octaves produces better results in terms of mean squared error, whereas for large values of the Hurst exponents the choice of octaves has little impact on the estimation.

Remark 4.2 Simulation studies not included show that the choice of the mixing matrix (4.1) does not substantially affect the finite sample results. Moreover, the demixing estimator in step $(S1)$ is very robust with respect to the condition number of the mixing matrix P. The distributions of the estimated scalar Hurst eigenvalues after demixing are barely affected for condition numbers of the order of at least 10^5.

To gauge the comparative statistical and computational performance of the two-step method, maximum likelihood (ML)-type estimation is a natural choice due to its wide applicability and well-known asymptotic properties. Notwithstanding its usually excellent statistical performance, ML estimation is notoriously slow (e.g., see Beran 1994, Chapter 5, Craigmile et al. 2005; see also the broad discussion in Caragea and Smith (2007) and references therein). For example, evaluation of the multivariate Gaussian likelihood involves calculating the inverse and determinant of a $n\nu \times n\nu$ matrix, which according to most commonly used algorithms, takes $O(\nu^3)$ steps. For many data sets in the modern era of "Big Data", ν is of the order of many thousands or more (see, for instance, Fontugne et al. 2017 on Internet traffic). This makes such computation prohibitively time consuming, especially in the presence of fractional memory. Various approximate likelihood methods have been proposed that are more efficient to compute, while not sacrificing much statistical efficiency. Some of these methods exploit fast transforms of the data such as the fast Fourier transform. One popular method of this kind is the Whittle estimator (e.g., Palma 2007, Section 4.4), which requires reexpressing the likelihood function in the Fourier domain and using some approximations. More details about multivariate Whittle-type estimation can be found, for instance, in Hosoya (1996, 1997) and Robinson (2008). In our Monte Carlo experiments, we compared the statistical and computational finite sample performances of two-step wavelet-based and Whittle-type ML estimation for mixed bivariate (operator) fractional Gaussian noise. As a representative case, we picked the second instance in Table 2 in Tsai et al. (2018). In other words, we assumed X in (1.1) is a vector of two independent fGn entries with Hurst

Table 1 Choice of scales 10,000 Monte Carlo runs, sample sizes 2^{20} and 2^{10}, $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$

h	J_1, J_2	$\widehat{h}$ (2^{20})	$Bias$	SD	$\sqrt{\text{MSE}}$	$\widehat{h}$ (2^{10})	$Bias$	SD	$\sqrt{\text{MSE}}$
0.20	1,2	0.25	0.05	0.04	0.06	0.33	0.13	0.10	0.15
	1, 3	0.22	0.02	0.03	0.04	0.26	0.06	0.08	0.10
	1, 4	0.22	0.02	0.03	0.03	0.24	0.04	0.08	0.09
	1, 5	0.21	0.01	0.02	0.03	0.23	0.03	0.08	0.09
	1, 6	0.21	0.01	0.02	0.03	0.22	0.02	0.09	0.09
0.40	1, 2	0.40	−0.00	0.02	0.02	0.45	0.05	0.08	0.10
	1, 3	0.40	−0.00	0.01	0.02	0.41	0.01	0.07	0.07
	1, 4	0.39	−0.01	0.01	0.02	0.40	0.00	0.07	0.07
	1, 5	0.40	−0.00	0.01	0.01	0.40	0.00	0.07	0.07
	1, 6	0.39	−0.01	0.01	0.01	0.40	−0.00	0.07	0.07
0.60	1, 2	0.60	−0.00	0.01	0.02	0.60	−0.00	0.07	0.07
	1, 3	0.59	−0.01	0.01	0.02	0.59	−0.01	0.06	0.06
	1, 4	0.59	−0.01	0.01	0.02	0.58	−0.02	0.07	0.07
	1, 5	0.59	−0.01	0.01	0.02	0.58	−0.02	0.08	0.08
	1, 6	0.59	−0.01	0.01	0.02	0.58	−0.02	0.07	0.07
0.80	1, 2	0.79	−0.01	0.01	0.02	0.76	−0.04	0.07	0.08
	1, 3	0.79	−0.01	0.01	0.02	0.78	−0.02	0.07	0.07
	1, 4	0.79	−0.01	0.01	0.02	0.77	−0.03	0.07	0.07
	1, 5	0.79	−0.01	0.01	0.02	0.77	−0.03	0.07	0.07
	1, 6	0.79	−0.01	0.01	0.02	0.77	−0.03	0.07	0.07

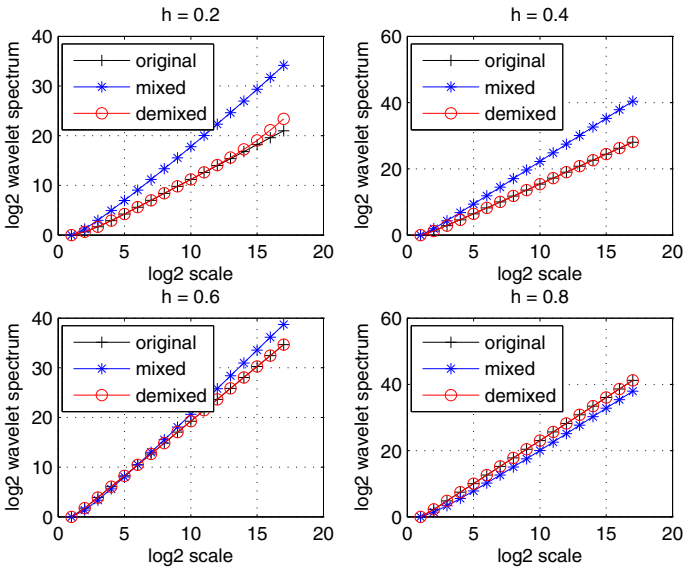


Fig. 2 Scaling: $\log W_{\cdot,\cdot}(2^j)$ versus j for each of the $n = 4$ components based on the wavelet variance scales 2^1 and 2^2. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{20}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

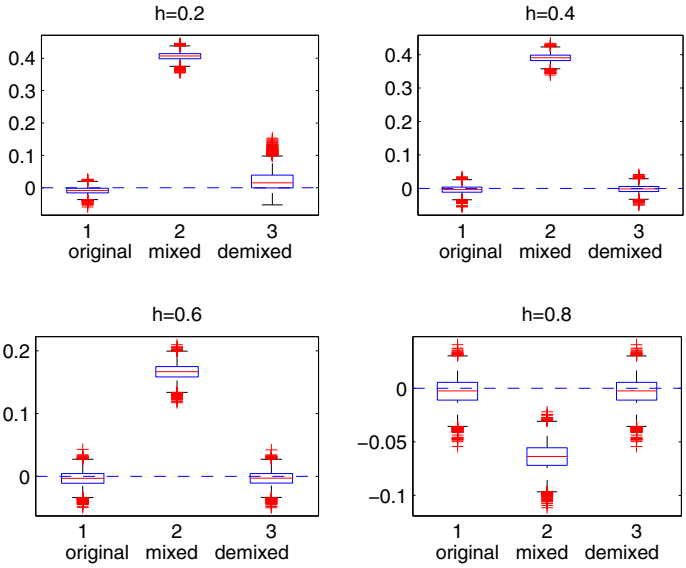


Fig. 3 Boxplots based on the wavelet variance scales 2^1 and 2^2 for $i = 1, 2, 3, 4$, $\widehat{h}_{X,i} - h_i$ (hidden, left), $\widehat{h}_{Y,i} - h_i$ (mixed, middle) and $\widehat{h}_{\widetilde{X},i} - h_i$ (demixed, right), for each of the $n = 4$ components, sorted by ascending order in terms of h. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{20}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

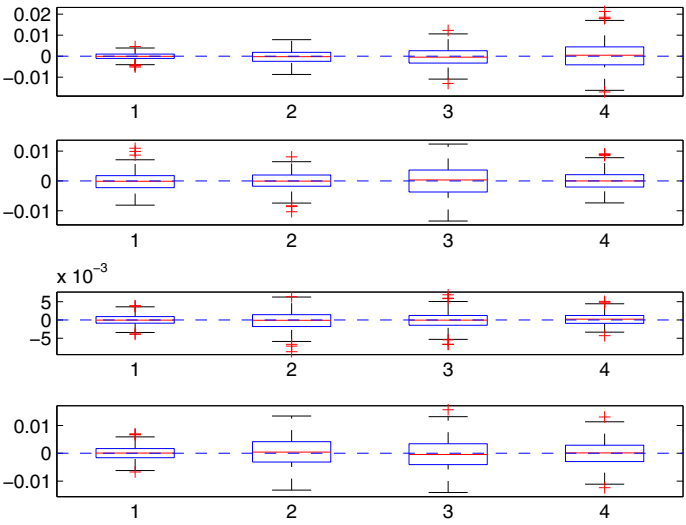


Fig. 4 Boxplots based on the wavelet variance scales 2^1 and 2^2 for the 16 entries of $\widehat{P^{-1}}P - I$. The (i_1, i_2)th boxplot denotes the (i_1, i_2)th entry of $\widehat{P^{-1}}P - I$. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{20}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

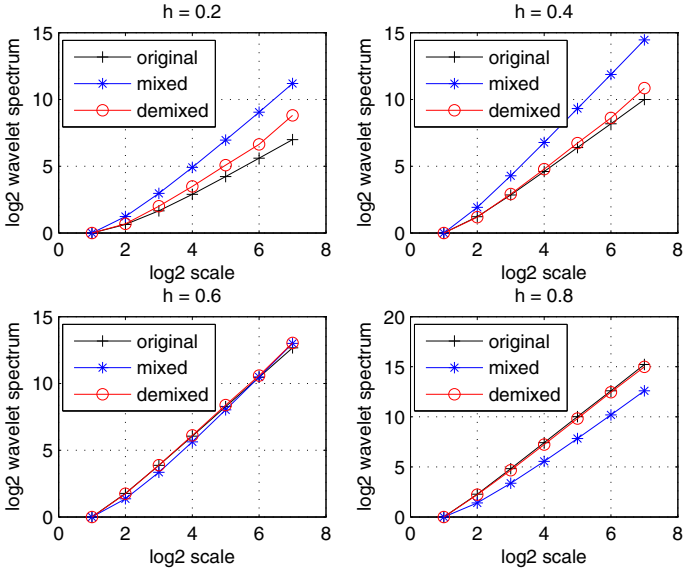


Fig. 5 Scaling: $\log W_{\cdot,\cdot}(2^j)$ vs. j for each of the $n = 4$ components based on the wavelet variance scales 2^1 and 2^2. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{10}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

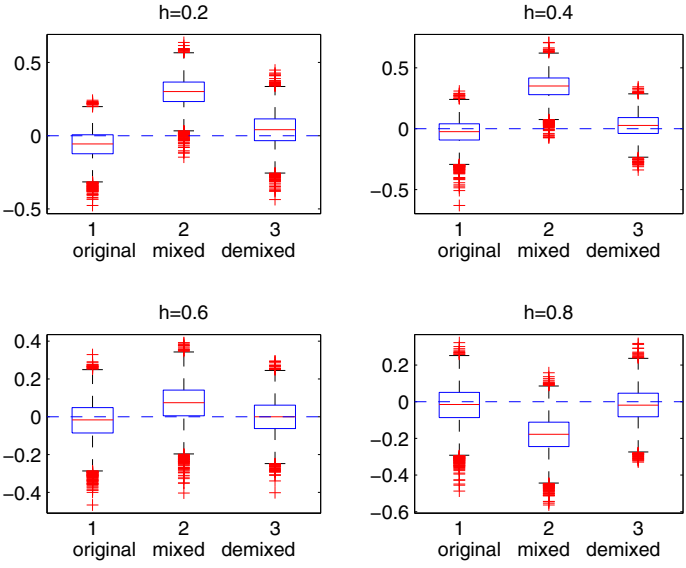


Fig. 6 Boxplots based on the wavelet variance scales 2^1 and 2^2 for $i = 1, 2, 3, 4$, $\widehat{h}_{X,i} - h_i$ (hidden, left), $\widehat{h}_{Y,i} - h_i$ (mixed, middle) and $\widehat{h}_{\widetilde{X},i} - h_i$ (demixed, right), for each of the $n = 4$ components, sorted by ascending order in terms of h. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{10}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

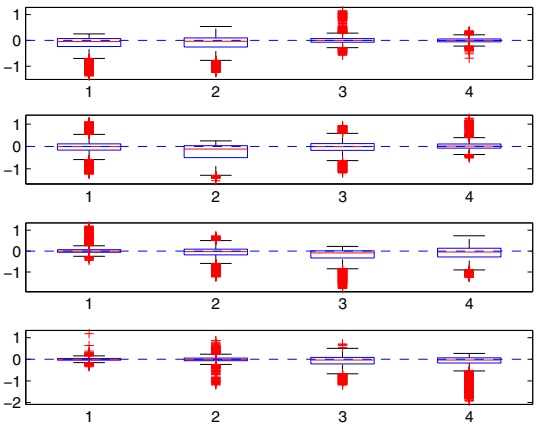


Fig. 7 Boxplots based on the wavelet variance scales 2^1 and 2^2 for the 16 entries of $\widehat{P^{-1}P - I}$. The (i_1, i_2)th boxplot denotes the (i_1, i_2)th entry of $\widehat{P^{-1}P - I}$. The plots were produced by means of 10,000 Monte Carlo runs of sample size 2^{10}, with parameter values $\mathbf{h} = (0.2, 0.4, 0.6, 0.8)$ and $N_\psi = 2$

parameter values $h_1 = 0.40$, $h_2 = 0.85$ and chose the change of coordinates matrix P as to match the matrix A in Section 2 of Tsai et al. (2018), i.e.,

$$A = \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}, \tag{4.2}$$

where $A = P\text{diag}(e(h_1), e(h_2))$ for functions $e(h_i) := \{\Gamma(2h_i + 1)\sin(\pi h_i)/2\pi\}^{1/2}$, $i = 1, 2$. In light of the results in Tsai et al. (2018) for fully parametric (hence, more accurate) estimation, Table 2 shows that the two-step method performs well, MSE-wise, for a *semi*parametric estimator especially given the relatively small sample size 2^{10}. In particular, the individual Hurst parameter estimation performance is similar to that of semiparametric estimation for univariate time series.

To shed light on computational aspects, we implemented Whittle-ML estimation in Matlab. We used the function `fminsearch.m` to minimize the approximate log-likelihood function with respect to the parameters h_1, h_2 and A [see Tsai et al. 2018, Expression (8)]. In Table 3, computational times are compared. It can be seen that the computational time per realization of Whittle-ML grows rapidly as a function of the path size ν, and the ratio between computational times for the two methods grows exponentially fast. Hence, in spite of the Whittle approximation, ML-type estimation demands further efforts to ease the computational burden of minimization with respect to $O(n^2)$ unknown parameters. By contrast, the computational robustness of the two-step method with respect to the sample path size is striking.

5 Application

As an application, we fit a bivariate series of annual tree ring measurements from bristlecone pine trees in California. The data can be found in the Time Series Data Library, which is available on the website DataMarket (https://datamarket.com/data/list/?q=provider:tsdl). The so-named White Mountain and Methuselah pine tree data sets are provided by C. W. Ferguson, E. Schulman and H. C. Fritts, and by D. A. Graybill, respectively.

Many tree ring data sets exhibit long range dependence properties (Tsai and Chan 2005; Bai and Taqqu 2018). Annual tree ring width measurements can be modeled as aggregates of the underlying continuous time growth rate process over time intervals between two

Table 2 Monte Carlo biases, standard deviations and (square root) mean squared errors of the two-step method over 10,000 replications with sample size $\nu = 2^{10}$ from the two-step wavelet-based method for the parameters $h_1 = 0.40$, $h_2 = 0.85$ and $A = (a_{ij})_{i,j=1,2}$ as in (4.2), $j_1 = 1$, $j_2 = 5$, $N_\psi = 1$

Parameter	Bias	SD	$\sqrt{\text{MSE}}$
h_1	0.0166	0.0495	0.0522
h_2	− 0.0077	0.0415	0.0422
a_{11}	− 0.0124	0.0932	0.0940
a_{12}	− 0.0031	0.0976	0.0976
a_{21}	0.0194	0.0635	0.0664
a_{22}	− 0.0169	0.1018	0.1032

Table 3 Computational performance. Whittle-type ML (using Matlab's `fminsearch.m`) and two-step wavelet based methods, dimension $n = 2$

Sample path size	Time in seconds (per realization)		Time ratio (ML/two-step wavelet)
	ML	Two-step wavelet	
2^8	7.6	0.0570	133.3
2^{10}	91.2	0.0631	1445.3
2^{12}	1209.5	0.0815	14,840.0
2^{14}	18376.0	0.1243	147,840.0

consecutive sampling time points. Assuming reasonable physical models, the latter, in turn, can be approximated by a mixed fractional process, as explained in Sect. 2.2. Although the full data set covers the period 5142 BC–1962 AD, we focus instead on the subperiod 4141 BC–1962 AD, since preliminary wavelet-based analysis revealed stationarity in the latter. The time series are displayed in Fig. 8, top plots.

Data analysis is conducted both in the time and wavelet domains. In the time domain, we examine the data by means of sample autocorrelation and cross-correlation functions (ACFs and CCFs, respectively). It is well known that spurious cross-correlation may occur as a result of the presence of fractional memory in each time series; thus, CCFs are only displayed for pre-whitened data following the standard procedure described in Cryer and Chan (2008, Section 11.3). In the wavelet domain, we show main diagonal wavelet scaling plots $\log_2 \widetilde{W}(2^j)_{11}$ and $\log_2 \widetilde{W}(2^j)_{22}$ [see (C.4)] as functions of $\log_2 2^j = j$, as well as the so-named sample wavelet coherence function $\widehat{w}_{12}(2^j)$, $j = j_1, \ldots, j_m$. The latter is a wavelet version of the CCF and can also be used to check the cross-correlation in bivariate data. For each j, the associated term is defined by

$$\widehat{w}_{12}(2^j) = \widetilde{W}(2^j)_{12} \Big/ \sqrt{\widetilde{W}(2^j)_{11}\widetilde{W}(2^j)_{22}}$$

(see Whitcher et al. 2000).

The sample ACF for each time series, shown on the lower panel in Fig. 8, suggest that the time series have long memory. This is confirmed by wavelet analysis, as displayed in Fig. 1 (left plot). Indeed, both $\log_2 \widetilde{W}(2^j)_{11}$ and $\log_2 \widetilde{W}(2^j)_{22}$ suggest scaling behavior with Hurst parameters that clearly depart from $1/2$, i.e., long memory, where, in this case,

$$h_i = \delta_i + \frac{1}{2} \in (0, 1), \quad i = 1, 2.$$

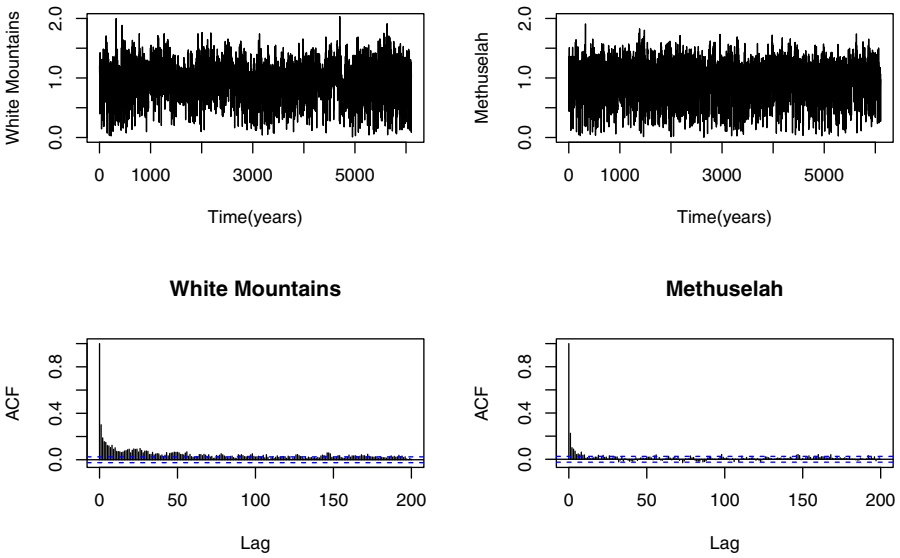


Fig. 8 Upper: time series plots of tree ring measurements. Lower: sample autocorrelations of tree ring measurements

Moreover, the fact that both curves resemble each other (namely, close Hurst parameter values) can be explained as the preponderance of one of the two underlying scaling laws (see the discussion in the Introduction). The upper panel in Fig. 9 displays the CCF. It reveals that the sequences are contemporaneously strongly correlated but not cross-correlated at any nonzero lag values. This is confirmed by the wavelet coherence function (Fig. 1, right plot), which shows significant and nearly constant correlation across all scales.

The demixing step ($S1$) of the proposed wavelet-based method yields the following estimated demixing matrix

$$\widehat{P^{-1}} = \begin{pmatrix} 0.9112 & -0.7827 \\ 0.1467 & 1.1922 \end{pmatrix}.$$

Demixed ring tree time series are computed by applying $\widehat{P^{-1}}$ to the original data. Inspection of the CCF for the demixed tree ring data reveals that the proposed wavelet-based method successfully decorrelated the data (lower panel in Fig. 9). This is further confirmed by the wavelet coherence function (Fig. 1, right plot), which evidences near zero correlations at all scales but a few of the coarsest. In addition, both functions $\log_2 \widetilde{W}(2^j)_{11}$ and $\log_2 \widetilde{W}(2^j)_{22}$ (for demixed data) still display scaling behavior. However, the Hurst exponents seem quite distinct and bounded away from $1/2$. This is confirmed by the proposed estimation method. After demixing, the memory parameter estimation step ($S2$) yields the estimates $\widehat{h}_1 = 0.65$, $\widehat{h}_2 = 0.93$ (using scales (j_1, j_2)=(3,7)), and $\widehat{h}_1 = 0.65$, $\widehat{h}_2 = 0.96$ (using scales (j_1, j_2)=(3,9)). In other words, there is little sensitivity of the parameter estimates to the choice of octave range.

Table 4 displays the Monte Carlo mean and standard deviation of $\widehat{h_1 - h_2} = \widehat{h}_1 - \widehat{h}_2$ assuming $h_1 = h_2 = h$ for some values of h within the range of interest. The difference between the estimated Hurst parameters for the demixed tree ring data is $|0.65 - 0.96| > 2.575 \times \widehat{\mathrm{sd}}(\widehat{h}_1 - \widehat{h}_2)$, with the statistic $\widehat{h}_1 - \widehat{h}_2$ lying far outside the 99% confidence interval (see Table 4). In other words, there is evidence for the hypothesis $h_1 \neq h_2$ in the demixed

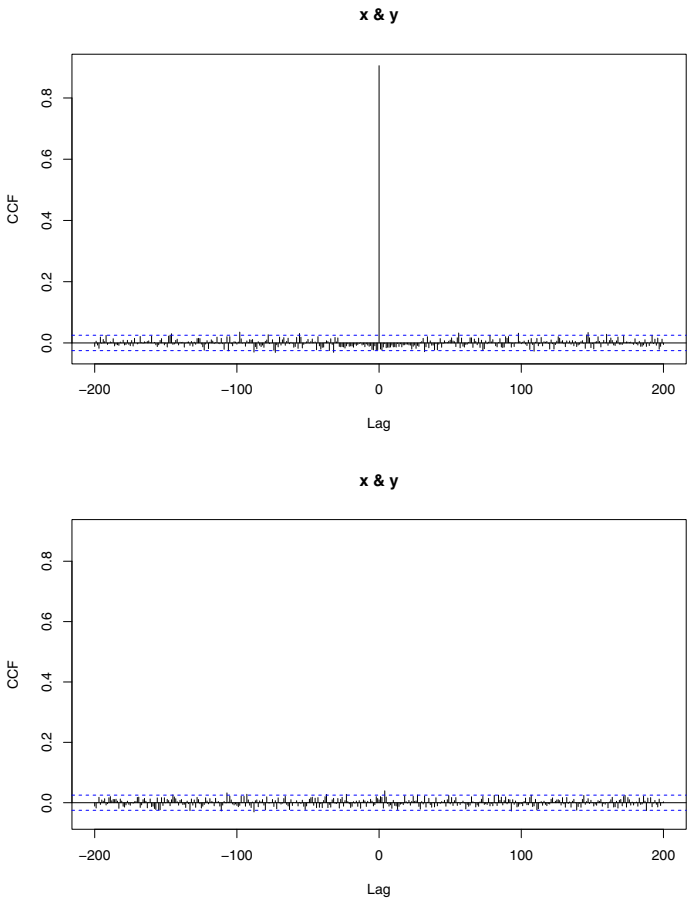


Fig. 9 Upper: sample cross-correlation between (pre-whitened) tree ring measurements. Lower: sample cross-correlation of the demixed (and post-demixing pre-whitened) data. The dashed lines correspond to the threshold $\pm 1.96/\sqrt{v}$ at 5% significant level

Table 4 Monte Carlo standard deviation for the statistic $\widehat{h}_1 - \widehat{h}_2$ assuming $h_1 = h_2 = h$. Simulations were conducted with $(j_1, j_2) = (3, 9)$, $N_\psi = 1$, sample size = 6000, number of Monte Carlo runs = 10,000

h	$\widehat{\mathrm{sd}}(\widehat{h}_1 - \widehat{h}_2)$
0.7	0.0335
0.8	0.0338
0.9	0.0342

ring tree data. Note that this could not have been detected had we skipped step $(S1)$, i.e., if Hurst exponent estimation had been conducted directly on the original data.

6 Conclusion

A Gaussian mixed fractional process $\{Y(t)\}_{t\in\mathbb{R}} = \{PX(t)\}_{t\in\mathbb{R}}$ is a multivariate stochastic process obtained by pre-multiplying a vector of independent, Gaussian fractional process

entries X by a nonsingular matrix P. It is interpreted that Y is observable, while X is a hidden process occurring in an (unknown) system of coordinates P. Mixed processes naturally arise as approximations to solutions of physically relevant classes of multivariate fractional SDEs under aggregation. We construct a semiparametric two-step wavelet-based method for estimating both the demixing matrix P^{-1} and the memory parameters of X. The asymptotic normality of the estimators is established both in continuous and discrete time. Monte Carlo experiments show that the finite sample estimation performance is accurate over finite samples, while being very computationally efficient. As an application, we model a bivariate time series of annual tree ring width measurements.

The research in this paper leads to a number of intriguing questions: (i) can other real world multivariate fractional phenomena be reasonably modeled as an aggregation over the solution of a fractional SDE? If so, as shown in this paper, parameter estimation can be carried out in a computationally faster and more dimension-robust way than by a general methodology for multivariate fractional processes; (ii) what is the natural physical non-Gaussian analogue of a fractional Gaussian measure-driven SDE?; (iii) are there other classes of multivariate fractional (in particular, o.s.s.) processes for which Hurst eigenvector estimation can be done in a computationally efficient way?; (iv) can results on mixed fractional processes be used to inspire new developments in the modeling of high-dimensional fractional systems?

References

Abry P, Didier G (2018a) Wavelet eigenvalue regression for n-variate operator fractional Brownian motion. J Multivar Anal 168:75–104

Abry P, Didier G (2018b) Wavelet estimation for operator fractional Brownian motion. Bernoulli 24(2):895–928

Abry P, Veitch D (1998) Wavelet analysis of long-range dependent traffic. IEEE Trans Inf Theory 44:2–15

Abry P, Flandrin P, Taqqu MS, Veitch D (2000) Wavelets for the analysis, estimation and synthesis of scaling data. In: Park K, Willinger W (eds) Self-similar network traffic and performance evaluation. Wiley, New York, pp 39–88

Abry P, Didier G, Li H (2018) Two-step wavelet-based estimation for Gaussian mixed fractional processes: supplementary material. Stat Inference Stoch Process 1–32

Achard S, Gannaz I (2016) Multivariate wavelet Whittle estimation in long-range dependence. J Time Ser Anal 37(4):476–512

Amblard P-O, Coeurjolly J-F (2011) Identification of the multivariate fractional Brownian motion. IEEE Trans Signal Process 59(11):5152–5168

Amblard P, Coeurjolly J-F, Lavancier F, Philippe A (2012) Basic properties of the multivariate fractional Brownian motion. Bull Soc Math Fr Sémin Congr 28:65–87

Bai S, Taqqu M (2018) How the instability of ranks under long memory affects large-sample inference. Stat Sci 33(1):96–116

Ball NM, Brunner RJ (2010) Data mining and machine learning in astronomy. Int J Mod Phys D 19(07):1049–1106

Bardet J-M (2002) Statistical study of the wavelet analysis of fractional Brownian motion. IEEE Trans Inf Theory 48(4):991–999

Bardet J-M, Tudor C (2010) A wavelet analysis of the Rosenblatt process: chaos expansion and estimation of the self-similarity parameter. Stoch Process Appl 120(12):2331–2362

Bardet J-M, Tudor C (2014) Asymptotic behavior of the Whittle estimator for the increments of a Rosenblatt process. J Multivar Anal 131:1–16

Becker-Kern P, Pap G (2008) Parameter estimation of selfsimilarity exponents. J Multivar Anal 99:117–140

Bell G, Hey T, Szalay A (2009) Beyond the data deluge. Science 323(5919):1297–1298

Belouchrani A, Abed-Meraim K, Cardoso J-F, Moulines E (1997) A blind source separation technique using second-order statistics. IEEE Trans Signal Process 45(2):434–444

Beran J (1994) Statistics for long-memory processes. CRC Press, Boca Raton

Beran J, Feng Y, Ghosh S, Kulik R (2013) Long memory processes—probabilistic properties and statistical models. Springer, Heidelberg

Biermé H, Meerschaert MM, Scheffler H-P (2007) Operator scaling stable random fields. Stoch Process Appl 117(3):312–332

Breuer P, Major P (1983) Central limit theorems for non-linear functionals of Gaussian fields. J Multivar Anal 13(3):425–441

Brockwell P, Davis R (1991) Time series: theory and methods, 2nd edn. Springer, Berlin

Caragea PC, Smith RL (2007) Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. J Multivar Anal 98(7):1417–1440

Cardoso J-F (1998) Blind signal separation: statistical principles. Proc IEEE 86(10):2009–2025

Chan K, Tsai K (2010) Inference of bivariate long-memory aggregate time series, pp 1–17. http://public.econ.duke.edu/brossi/NBERNSF/Tsai.pdf

Chen CLP, Zhang C-Y (2014) Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf Sci 275:314–347

Cheridito P, Kawaguchi H, Maejima M (2003) Fractional Ornstein–Uhlenbeck processes. Electron J Probab 8(3):1–14

Choi S, Cichocki A, Park H-M, Lee S-Y (2005) Blind source separation and independent component analysis: a review. Neural Inf Process Lett Rev 6(1):1–57

Clausel M, Vedel B (2011) Explicit construction of operator scaling Gaussian random fields. Fractals 19(01):101–111

Clausel M, Vedel B (2013) An optimality result about sample path properties of operator scaling Gaussian random fields. In: Proceedings of the "XIème Colloque Franco-Roumain de Mathématiques Appliquées, vol 4, no LXII, pp 375–409

Clausel M, Roueff F, Taqqu MS, Tudor C (2014a) Asymptotic behavior of the quadratic variation of the sum of two Hermite processes of consecutive orders. Stoch Process Appl 124(7):2517–2541

Clausel M, Roueff F, Taqqu M S, Tudor C (2014b) Wavelet estimation of the long memory parameter for Hermite polynomial of Gaussian processes. ESAIM Probab Stat 18:42–76

Coeurjolly J-F (2001) Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. Stat Inference Stoch Process 4(2):199–227

Coeurjolly J-F, Amblard P-O, Achard S (2013) Wavelet analysis of the multivariate fractional Brownian motion. ESAIM Probab Stat 17:592–604

Comon P, Jutten C (2010) Handbook of blind source separation: independent component analysis and applications. Academic Press, Cambridge

Craigmile P, Guttorp P, Percival D (2005) Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes. IEEE Trans Signal Process 53(8):3151–3161

Cryer J, Chan K (2008) Time series analysis with application in R. Springer, Berlin

Dahlhaus R (1989) Efficient parameter estimation for self-similar processes. Ann Stat 17(4):1749–1766

Daubechies I (1992) Ten lectures on wavelets, vol 61. Society for Industrial and Applied Mathematics, Philadelphia

Didier G, Pipiras V (2011) Integral representations and properties of operator fractional Brownian motions. Bernoulli 17(1):1–33

Didier G, Pipiras V (2012) Exponents, symmetry groups and classification of operator fractional Brownian motions. J Theor Probab 25:353–395

Didier G, McKinley SA, Hill DB, Fricks J (2012) Statistical challenges in microrheology. J Time Ser Anal 33(5):724–743

Didier G, Helgason H, Abry P (2015) Demixing multivariate-operator selfsimilar processes. In: IEEE international conference on acoustics. Speech and signal, processing (ICASSP), Brisbane, Australia, pp 1–5

Didier G, Meerschaert MM, Pipiras V (2017) Exponents of operator self-similar random fields. J Math Anal Appl 448(2):1450–1466

Didier G, Meerschaert MM, Pipiras V (2018) Domain and range symmetries of operator fractional Brownian fields. Stoch Process Appl 128(1):39–78

Dobrushin R, Major P (1979) Non-central limit theorems for non-linear functional of Gaussian fields. Probab Theory Relat Fields 50(1):27–52

Dogan M, Van Dam R, Liu G, Meerschaert MM, Butler JJ, Bohling G, Benson DA, Hyndman DW (2014) Predicting flow and transport in highly heterogeneous alluvial aquifers. Geophys Res Lett 41(21):7560–7565

Embrechts P, Maejima M (2002) Selfsimilar processes, princeton series in applied mathematics. Princeton University Press, Princeton

Fevotte C, Godsill S-J (2006) A Bayesian approach for blind separation of sparse sources. IEEE Trans Audio Speech Lang Process 14(6):2174–2188

Flandrin P (1992) Wavelet analysis and synthesis of fractional Brownian motion. IEEE Trans Inf Theory 38:910–917
Fontugne R, Abry P, Fukuda K, Veitch D, Cho K, Borgnat P, Wendt H (2017) Scaling in internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections. IEEE/ACM Trans Netw 25(4):1–14
Fox R, Taqqu M (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann Stat 14(2):517–532
Frecon J, Didier G, Pustelnik N, Abry P (2016) Non-linear wavelet regression and branch and bound minimization for the full identification of bivariate operator fractional Brownian motion. IEEE Trans Signal Proc 64(15):4040–4049
Granger C, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. J Time Ser Anal 1(1):15–29
Guo H, Lim CY, Meerschaert MM (2009) Local Whittle estimator for anisotropic random fields. J Multivar Anal 100(5):993–1028
Horn R, Johnson C (1985) Matrix analysis. Cambridge University Press, Cambridge
Hosking JRM (1981) Fractional differencing. Biometrika 68(1):165–176
Hosoya Y (1996) The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence. J Econom 73:217–236
Hosoya Y (1997) A limit theory for long-range dependence and statistical inference on related models. Ann Stat 25(1):105–137
Hualde J (2013) A simple test for the equality of integration orders. Econ Lett 119(3):233–237
Hualde J, Robinson P (2010) Semiparametric inference in multivariate fractionally cointegrated systems. J Econom 157(2):492–511
Hudson W, Mason J (1982) Operator-self-similar processes in a finite-dimensional space. Trans Am Math Soc 273(1):281–297
Kechagias S, Pipiras V (2015a) Definitions and representations of multivariate long-range dependent time series. J Time Ser Anal 36(1):1–25
Kechagias S, Pipiras V (2015b) Identification, estimation and applications of a bivariate long-range dependent times series model with general phase 1–25. https://pipiras.web.unc.edu/files/2018/04/kechagias-pipiras-2018-mlrd-phase.pdf
Kou S (2008) Stochastic modeling in nanoscale biophysics: subdiffusion within proteins. Ann Appl Stat 2(2):501–535
Laha RG, Rohatgi VK (1981) Operator self-similar stochastic processes in $\mathbf{R}^d$. Stoch Process Appl 12(1):73–84
Li Y, Xiao Y (2011) Multivariate operator-self-similar random fields. Stoch Process Appl 121(6):1178–1200
Li Y, Adali T, Wang W, Calhoun V (2009) Joint blind source separation by multiset canonical correlation analysis. IEEE Trans Signal Process 57(10):3918–3929
Lobato I (1997) Consistency of the averaged cross-periodogram in long memory series. J Time Ser Anal 18(2):137–155
Lysy M, Pillai N, Hill DB, Forest MG, Mellnik J, Vasquez P, McKinley SA (2016) Model comparison and assessment for single particle tracking in biological fluids. J Am Stat Assoc 111(516):1413–1426
Maejima M, Mason J (1994) Operator-self-similar stable processes. Stoch Process Appl 54:139–163
Mallat S (1999) A wavelet tour of signal processing. Academic Press, Cambridge
Mandelbrot B, Van Ness J (1968) Fractional Brownian motions, fractional noises and applications. SIAM Rev 10(4):422–437
Marinucci D, Robinson PM (2000) Weak convergence of multivariate fractional processes. Stoch Process Appl 86(1):103–120
Marinucci D, Robinson PM (2001) Semiparametric fractional cointegration analysis. J Econom 105(1):225–247
Mason J, Xiao Y (2002) Sample path properties of operator-self-similiar Gaussian random fields. Theory Probab Appl 46(1):58–78
Mason TG, Weitz DA (1995) Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids. Phys Rev Lett 74(7):1250
Masry E (1993) The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. IEEE Trans Inf Theory 39(1):260–264
Moreau E (2001) A generalization of joint-diagonalization criteria for source separation. IEEE Trans Signal Process 49(3):530–541
Moulines E, Soulier P (2003) Semiparametric spectral estimation for fractional processes. In: Theory and applications of long-range dependence, Birkhäuser, pp 251–301

Moulines E, Roueff F, Taqqu M (2007a) Central limit theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. Fractals 15(4):301–313
Moulines E, Roueff F, Taqqu M (2007b) On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. J Time Ser Anal 28(2):155–187
Moulines E, Roueff F, Taqqu M (2008) A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. Ann Stat 36(4):1925–1956
Nielsen FS (2011) Local Whittle estimation of multi-variate fractionally integrated processes. J Time Ser Anal 32(3):317–335
Nielsen M, Frederiksen P (2011) Fully modified narrow-band least squares estimation of weak fractional cointegration. Econom J 14(1):77–120
O'Grady P, Pearlmutter B, Rickard S (2005) Survey of sparse and non-sparse methods in source separation. Int J Imaging Syst Technol 15(1):18–33
Palma W (2007) Long-memory time series: theory and methods, vol 662. Wiley, Hoboken
Parra L, Sajda P (2003) Blind source separation via generalized eigenvalue decomposition. J Mach Learn Res 4:1261–1269
Percival D B, Walden A (2006) Wavelet methods for time series analysis, vol 4. Cambridge University Press, Cambridge
Pham D-T, Cardoso J-F (2001) Blind separation of instantaneous mixtures of nonstationary sources. IEEE Trans Signal Process 49(9):1837–1848
Pipiras V, Taqqu MS (2017) Long-range dependence and self-similarity. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge
Prakasa Rao BLS (2010) Statistical inference for fractional diffusion processes. Wiley series in probability and statistics. Wiley, Hoboken
Puplinskaitė D, Surgailis D (2015) Scaling transition for long-range dependent Gaussian random fields. Stoch Process Appl 125(6):2256–2271
Robinson P (1995a) Gaussian semiparametric estimation of long range dependence. Ann Stat 23(5):1630–1661
Robinson P (1995b) Log-periodogram regression of time series with long range dependence. Ann Stat 23(3):1048–1072
Robinson P (2008) Multivariate local Whittle estimation in stationary systems. Ann Stat 36(5):2508–2530
Robinson PM, Yajima Y (2002) Determination of cointegrating rank in fractional systems. J Econom 106(2):217–241
Sela R, Hurvich C (2012) The averaged periodogram estimator for a power law in coherency. J Time Ser Anal 33(2):340–363
Shimotsu K (2007) Gaussian semiparametric estimation of multivariate fractionally integrated processes. J Econom 137:277–310
Shimotsu K (2012) Exact local Whittle estimation of fractionally cointegrated systems. J Econom 169(2):266–278
Stoev S, Pipiras V, Taqqu M (2002) Estimation of the self-similarity parameter in linear fractional stable motion. Signal Process 82:1873–1901
Stone J (2004) Independent component analysis: a tutorial introduction. MIT Press, Cambridge
Taqqu MS (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. Probab Theory Relat Fields 31(4):287–302
Taqqu MS (1979) Convergence of integrated processes of arbitrary Hermite rank. Probab Theory Relat Fields 50(1):53–83
Taqqu MS (2003) Fractional Brownian motion and long range dependence. In: Doukhan P, Oppenheim G, Taqqu MS (eds) Theory and applications of long-range dependence. Birkhäuser, Boston, pp 5–38
Tsai H, Chan K (2005) Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. J Time Ser Anal 26(5):691–713
Tsai H, Rachinger H, Chan K-S (2018) Inference of bivariate long-memory aggregate time series. Stat Sin 28(1):399–421
Veitch D, Abry P (1999) A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Trans Inf Theory 45(3):878–897
Wendt H, Scherrer A, Abry P, Achard S (2009) Testing fractal connectivity in multivariate long memory processes. In: IEEE international conference on acoustics, speech and signal processing, Taipei, Taiwan, April 19–24, 2009, pp 2913–2916
Wendt H, Didier G, Combrexelle S, Abry P (2017) Multivariate Hadamard self-similarity: testing fractal connectivity. Phys D 356–357:1–36
Whitcher B, Guttorp P, Percival DB (2000) Wavelet analysis of covariance with application to atmospheric time series. J Geophys Res 105(D11):14941–14962

Wornell G, Oppenheim A (1992) Estimation of fractal signals from noisy measurements using wavelets. IEEE Trans Signal Process 40(3):611–623

Xiao Y (2009) Sample path properties of anisotropic Gaussian random fields. In: Khoshnevisan D, Rassoul-Agha F (eds) A minicourse on stochastic partial differential equations. Springer, Berlin, pp 145–212

Yaglom A M (1987) Correlation theory of stationary and related random functions I: basic results. Springer, Berlin

Yeredor A (2002) Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. IEEE Trans Signal Process 50(7):1545–1553

Ziehe A, Laskov P, Nolte G, Müller K-R (2004) A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. J Mach Learn Res 5:777–800