Precision spectral estimation at sub-Hz frequencies: Closed-form posteriors and Bayesian noise projection

Lorenzo Sala[1,2] and Stefano Vitale[1]

[1]*Dipartimento di Fisica, Università di Trento, 38123 Trento, Italy*
[2]*Trento Institute for Fundamental Physics and Application/INFN, 38123 Trento, Italy*

(Received 15 April 2026; accepted 16 July 2026; published 9 September 2026)

We consider the problem of estimating cross-spectral quantities in the low-frequency regime, where long observation times limit averaging over large ensembles of periodograms, thereby preventing the use of approximate Gaussian statistics. This case is relevant for precision low-frequency gravitational experiments such as LISA and LISA Pathfinder. We present a Bayesian method for estimating spectral quantities in multivariate Gaussian time series. The approach, based on periodograms and Wishart statistics, yields closed-form expressions at any given frequency for the marginal posterior distributions of the individual power spectral densities, the pairwise coherence, and the multiple coherence, as well as for the joint posterior distribution of the full cross-spectral density matrix. In the context of noise projection—where one series is modeled as a linear combination of filtered versions of the others, plus a background component—the method also provides closed-form posteriors for both the susceptibilities, i.e., the filter transfer functions, and the power spectral density of the background. We apply the method to data from the LISA Pathfinder mission, showing effective decorrelation of temperature-induced acceleration noise and reliable estimation of its coupling coefficient.

DOI: 10.1103/hwmv-xt1k

I. INTRODUCTION

Spectral analysis based on Welch's method [1,2] is the standard approach for noise characterization. In this framework, power spectral densities (PSD) and cross-spectral densities (CPSD) of multivariate noise time series are estimated from properly normalized discrete Fourier transforms, known as periodograms.

To improve the precision of these estimates and to quantify their uncertainty, Welch's method partitions the time series into M equal length, possibly overlapping segments, yielding M independent (or very weakly correlated) periodogram samples. In the usual frequentist approach, PSD and CPSD are obtained by averaging these samples, with uncertainties scaling with the standard deviation of the mean. This procedure relies on the central limit theorem, which ensures that the averaged estimates rapidly approach Gaussian statistics, enabling the construction of confidence intervals.

However, spectral resolution decreases with decreasing M, and in many applications, particularly those at very low frequency, M is inevitably small, sometimes even $M = 1$. As M decreases, the Gaussian approximation for the statistics of the average becomes increasingly inaccurate, leading, for instance, to paradoxical outcomes such as non-negligible probabilities for negative PSD values. At $M = 1$, the frequentist approach becomes entirely inapplicable. By contrast, a Bayesian framework does not suffer from these limitations and remains well defined even in this regime.

In the first part of this paper, we derive the posterior distributions for many spectral quantities, under a Bayesian framework, focusing in particular on the case where just a few periodograms are available, that is, far from Gaussianity.

In the second part of this paper, starting from Sec. IV, we address a closely related problem: noise projection and time-series decorrelation. Specifically, we consider the case in which one time series is modeled as a linear combination of filtered versions of other series, plus an additional background component. We develop a Bayesian method to infer both the residual noise properties (after decorrelation) and the coupling coefficients. Particular attention is devoted to the underlying probability distributions, which can significantly deviate from Gaussianity when only a small number of periodograms is available. We provide a detailed technical discussion of the method, validate it through numerical simulations, and demonstrate its application to real data.

We encountered a need for these methods while analyzing data from the LISA Pathfinder (LPF) mission [3–5]. The mission's objective was to precisely measure the noise

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

spectrum of force disturbances acting on two nominally freely falling test masses in space, reaching acceleration levels as low as a few $\mathrm{fm\,s^{-2}/Hz^{1/2}}$, and frequencies down to approximately 20 μHz. Unlike typical spectral estimation applications, where the goal is to extract a signal from a noisy stochastic process, LPF aimed to measure the noise itself with the highest possible accuracy, particularly at the lowest achievable frequencies.

While reviewing the literature for a consistent and practical Bayesian approach suited to our needs, we found that the fundamental principles had long been established. However, we could not find a detailed, practical method applicable to our data processing, leading us to develop one independently.

We applied this method to the data analysis of LPF in Refs. [3,6], and briefly summarized its main features in the Appendices of the second paper. Here, we provide a detailed description of the method, discussing its foundations, deriving key procedures rigorously, and presenting quantitative evidence of its validity through numerical simulations.

The paper is organized as follows: in Sec. II we define the key experimental quantities of multivariate time series and derive their likelihood under the Gaussian data hypothesis; in Sec. III we build the Bayesian posteriors for all the related spectral quantities. In Sec. IV we discuss the case of noise projection, that is, the case where one series is modeled as a linear combination of filtered versions of the others, plus a background component of which one wants to estimate the spectrum; in Sec. V we apply our method to simulations and real data. Finally, in Sec. VI we give some concluding remarks. The method presented in this paper is available as a PYTHON package at [7].

II. PERIODOGRAMS, KEY PERIODOGRAM FUNCTIONS, AND THEIR LIKELIHOOD

In this section, we recall a few basic concepts and results that we will use in our Bayesian inference method. We assume that data are Gaussian for the purpose of the our analysis, something we found to be quantitatively true in the case of LPF [3].

A. Basic definitions and nomenclature

In our approach, we assume that we have acquired, synchronously and with sampling time T, the time series of p real, stationary, Gaussian, zero-mean stochastic processes. We call $x_i[n]$, with $1 \le i \le p$, the sample of the ith series taken at time $t = nT$.

The joint statistics of these samples is fully contained in the mean values of their products

$$\langle x_i[n] x_j[m] \rangle = \int_{-\infty}^{\infty} \Sigma_{i,j}(f) e^{i2\pi(m-n)fT} df, \tag{1}$$

and then in the Hermitian positive definite matrix $\mathbf{\Sigma}(f)$, with elements $\Sigma_{i,j}(f)$, which, by definition, is the joint two-sided power CPSD matrix of the p stochastic processes at frequency f. Note that the diagonal element $\Sigma_{i,i}(f)$ is the PSD of the process $x_i(t)$, while the off-diagonal $\Sigma_{i,j}(f)$ is the pair CPSD of $x_i(t)$ and $x_j(t)$.

Our goal is to infer each element of $\mathbf{\Sigma}(f)$ independently at each frequency, without assuming any functional dependence on f.

The main tool for such an inference is the periodogram $X_i[k]$ calculated over an N-long segment of the multivariate time series. $X_i[k]$ is defined as

$$X_i[k] = \sqrt{\frac{T}{N}} \sum_{n=0}^{N-1} x_i[n] w[n] e^{-2\pi i k n/N}, \tag{2}$$

with $0 \le k \le N-1$ an integer, and $w[n]$ the coefficients of a suitable tapering window. Since $x_i[N-k] = x_i^*[k]$, only the first $\lfloor N/2 \rfloor + 1$ of these coefficients carry independent information. $X_i[k]$ is Gaussian, complex, and zero-mean. Key for the inference are the complex mean values

$$\langle X_i[k] X_j^*[k'] \rangle = \frac{T}{N} \int_{-\infty}^{\infty} \Sigma_{i,j}(f) \tilde{w}\left(\frac{2\pi}{N}k - 2\pi f T\right) \times \tilde{w}^*\left(\frac{2\pi}{N}k' - 2\pi f T\right) df, \tag{3}$$

with $\tilde{w}(\phi) = \sum_{n=0}^{N-1} w[n] e^{-i\phi n}$ the Fourier sequence transform of the tapering window $w[n]$. If, ideally, one could choose $w[n]$ such that

$$\tilde{w}\left(\frac{2\pi}{N}k - 2\pi f T\right) \tilde{w}^*\left(\frac{2\pi}{N}k' - 2\pi f T\right) = \frac{N}{T} \delta_{kk'} \delta\left(\frac{2\pi}{N}k - 2\pi f T\right), \tag{4}$$

with $\delta_{kk'}$ the Kronecker delta of k and k', and $\delta(\phi)$ the Dirac delta of ϕ, then $X_i[k] X_j^*[k]$ would be an unbiased estimator of $\Sigma_{ij}(f = k/(NT))$, and $X_i[k]$ would be independent of $X_j^*[k']$ if $k \ne k'$.

In reality, $\tilde{w}(\phi)$ is a 2π-periodic function with a central lobe at $\phi = 0$, and a sequence of strongly suppressed side lobes. We assume that the aliasing deriving from the periodicity of $\tilde{w}(\phi)$ has been made negligible by properly choosing a short enough sampling time T.[1]

[1]If this assumption does not hold, then $\Sigma_{i,j}$ in Eq. (3) should instead be interpreted as the discrete-time CPSD of the discrete stochastic process obtained by sampling the original continuous-time process. All statistical conclusions presented in the remainder of the paper remain unchanged. However, if these discrete-time quantities are used to estimate their continuous-time counterparts, the resulting estimates will be systematically biased due to aliasing.

The width of the central lobe depends on the choice of the window, but is always of the form $\pm m(2\pi/N)$, with m a small integer. Thus, if $|k-k'|>m$, then $X_i[k]$ and $X_j^*[k']$ may be treated as independent within a reasonable accuracy.[2] Then, from Eq. (3),

$$\langle X_i[k]X_j^*[k]\rangle \simeq \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \Sigma_{i,j}(f)G\left(f-\frac{k}{NT}\right)df \tag{5}$$

with $G(f)=|\tilde{w}(2\pi fT)|^2$. Note that $w[n]$ is always normalized such that $\int_{-\frac{1}{2T}}^{\frac{1}{2T}} G(f-\frac{k}{NT})df=1$. Thus, $X_i[k]X_j^*[k]$ is an estimator of $\Sigma_{i,j}(f)$ averaged over the band $f=(k\pm m)/(NT)$.

From now on, unless otherwise specified, we indicate with $\mathbf{\Sigma}(f)$ this averaged version of the CPSD matrix.

As usual in statistics, the precision of the estimator can be increased by averaging over repeated measurements. To this aim, Welch's method prescribes to split the available time series into M segments,[3] each of length N, average over them, and use the *observed* CPSD matrix $\mathbf{\Pi}[k]$, with elements

$$\Pi_{ij}[k]=\frac{1}{M}\sum_{\ell=1}^{M} X_{i,(\ell)}[k]X_{j,(\ell)}^*[k], \tag{6}$$

as an estimator of $\mathbf{\Sigma}(f=k/(NT))$.

Thus, in summary, the matrices $\mathbf{\Pi}[k]$, with $k\in[0,2m,4m,6m,\ldots,\lfloor N/4m\rfloor]$ are independent estimators of the matrices $\mathbf{\Sigma}(f)$ with $f\in[0,2m,4m,6m,\ldots,\lfloor N/4m\rfloor]/(NT)$, with a spectral resolution of $\pm m/(NT)$.

One can show that $\mathbf{\Pi}(k)$, which is Hermitian, is positive definite only if $M\geq p$. As the positive definiteness is mandatory for an estimator of $\mathbf{\Sigma}(f=k/(NT))$, the minimum number of periodograms one should average on is p.

Some common applications deviate from the spectral estimator with evenly spaced frequencies described above [3,8,9]. In those applications, at each frequency of interest f, one adjusts M, and then N, according to some averaging optimization criterion, and picks just one matrix $\mathbf{\Pi}[k]$, with k selected such that $f=k/(NT)$. $\mathbf{\Pi}[k]$ is then an estimate of the CPSD at the frequency f. This procedure is repeated at each frequency of interest.

As N depends on the frequency, the width of the spectral window is no longer frequency-independent, as it is in the case of uniform spacing. As a consequence, the independence of the CPSD estimators at different frequencies must, in principle, be established frequency by frequency. An assessment of the methods in Refs. [3,9] shows that, for at least some choices of parameters, the estimators at different frequencies can be considered practically independent. We will use these methods in several examples presented in the remainder of the paper.

Some functions of $\mathbf{\Pi}[k]$ that we will also use in the following are

(i) the measured magnitude-squared cross-coherence (MSC), between the ith and jth time series

$$|\hat{\rho}_{ij}[k]|^2=\left|\frac{\Pi_{ij}[k]}{\Pi_{ii}^{1/2}[k]\Pi_{jj}^{1/2}[k]}\right|^2; \tag{7}$$

a useful diagnostics of the possible linear correlation between the two underlying processes;

(ii) the multiple coherence [10], a useful generalization of $|\hat{\rho}_{ij}[k]|^2$ to the case of multiple series,

$$\hat{R}^2[k]=1-(\Pi_{11}[k]\Pi_{11}^{-1}[k])^{-1}, \tag{8}$$

with $\Pi_{i,j}^{-1}$ the elements of the inverse $\mathbf{\Pi}^{-1}$ of $\mathbf{\Pi}$. This is used as a diagnostic of how much of the noise power in $x_1[n]$ is due to its correlation to the remaining series; and

(iii) The Schur complement of any sub-block $\boldsymbol{C}$ in the decomposition of the Hermitian matrix $\mathbf{\Pi}[k]$ as

$$\mathbf{\Pi}[k]=\begin{pmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{B}^\dagger & \boldsymbol{C}\end{pmatrix}. \tag{9}$$

Here $\boldsymbol{A}$ is a $q\times q$ matrix, $\boldsymbol{C}$ is $r\times r$, $\boldsymbol{B}$ is $q\times r$, and $q+r=p$. The Schur complement of the block $\boldsymbol{C}$ in $\mathbf{\Pi}[k]$, is

$$\mathbf{\Pi}[k]/\boldsymbol{C}\equiv\boldsymbol{A}-\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\dagger. \tag{10}$$

We discuss in the following section the sampling distributions of all these quantities.

B. Sampling distribution of the CPSD matrix

Reference [10] shows that the joint sampling distribution of the elements of the matrix $\boldsymbol{W}=M\mathbf{\Pi}$, conditional to the theoretical CPSD matrix $\mathbf{\Sigma}$, is a complex Wishart distribution, with probability density function (PDF)

$$p(\boldsymbol{W}|\mathbf{\Sigma},M)=\frac{|\boldsymbol{W}|^{M-p}}{\tilde{\Gamma}_p(M)|\mathbf{\Sigma}|^M}\,\mathrm{etr}[-\mathbf{\Sigma}^{-1}\boldsymbol{W}]. \tag{11}$$

Here, $|\cdot|$ is the determinant, etr the exponential trace $\mathrm{etr}(\cdot)=\exp(\mathrm{tr}(\cdot))$, and $\tilde{\Gamma}_p(M)$ is the multivariate complex gamma function

[2] In this paper, when showing plots with spaced frequency points, we always choose k and k' so that adjacent frequencies exhibit a controlled correlation in the range of 10% to 30%, while correlations between second-nearest neighbors are negligible.

[3] These segments do not need to be disjoint. It has been shown that, as the window $w[n]$ tapers their ends, some overlap between adjoining segments does not significantly change the statistical properties of the derived quantities with respect to the case of disjoint segments [1].

$$\tilde{\Gamma}_p(M) = \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^{p} \Gamma(M-i+1).$$

Note that we have dropped, for clarity, the explicit dependence of all quantities on frequency.

We denote this distribution[4] with $\mathcal{CW}(\mathbf{\Sigma}, M)$. As expected, $\mathcal{CW}(\mathbf{\Sigma}, M)$ is defined only if $M \geq p$, that is, if $\boldsymbol{W}$ is positive definite.

We use the conditional probability in Eq. (11) as the likelihood function for the inference of $\mathbf{\Sigma}$.

C. Sampling distribution of derived quantities

The complex Wishart distribution describes the joint probability of all the elements of the matrix $\boldsymbol{W}$. Starting from that, and employing its mathematical properties [10,11], we give the sampling distributions of some derived quantities that we will use as likelihood functions for the Bayesian inference of the relative quantities.

a. Power spectral density. For $p = 1$, that is, in the case of a single univariate stochastic process, let W, Π, and S denote the only elements of $\mathbf{W}$, $\mathbf{\Pi}$, and $\mathbf{\Sigma}$, respectively. Since $\mathbf{W} = M\mathbf{\Pi}$, it follows that $W = M\Pi$. The PDF in Eq. (11) then reduces to

$$p(M\Pi|S) = \frac{(M\Pi)^{M-1}}{\Gamma(M)S^M} e^{-M\Pi/S.} \tag{12}$$

This means that $\Pi \sim \Gamma(M, S/M)$, with $\Gamma(M, S/M)$ the gamma distribution with shape parameter M and scale parameter S/M. Equivalently, Eq. (12) implies that $2M\Pi/S$ is chi-square distributed with $2M$ degrees of freedom. Note that this result is also obtained by calculating the marginal distribution of any of the diagonal elements of $\mathbf{\Pi}$ from the joint PDF in Eq. (11).

b. Magnitude squared coherence. Defining the theoretical ρ_{ij}

$$\rho_{ij} = \frac{\Sigma_{i,j}}{\Sigma_{i,i}^{1/2}\Sigma_{j,j}^{1/2}}, \tag{13}$$

the sampling distribution of the MSC, $|\hat{\rho}_{ij}|^2$ is [10,12]

$$p(|\hat{\rho}|^2||\rho|^2) = (M-1)(1-|\hat{\rho}|^2)^{M-2}(1-|\rho|^2)^M \times {}_2F_1(M, M, 1, |\hat{\rho}|^2|\rho|^2), \tag{14}$$

where ${}_2F_1$ represents Gauss' hypergeometric function.

Note that this distribution only holds for $M > 1$ as, notoriously, when $M = 1$, $|\hat{\rho}|^2 = 1$ holds exactly. More in general, for low values of M, $p(|\hat{\rho}|^2||\rho|^2)$ carries a significant bias toward $|\hat{\rho}|^2 > |\rho|^2$.

c. Multiple coherence. Similarly to the case of MSC, defining the theoretical R^2

$$R^2 = 1 - (\Sigma_{11}\Sigma_{11}^{-1})^{-1}, \tag{15}$$

with $\Sigma_{i,j}^{-1}$ the elements of $\mathbf{\Sigma}^{-1}$, the PDF of the sample multiple coherence $\hat{R}^2$ is [10]

$$p(\hat{R}^2|R^2) = \frac{\Gamma(M)}{\Gamma(p-1)\Gamma(M-p+1)} (\hat{R}^2)^{p-2}(1-\hat{R}^2)^{M-p} \times (1-R^2)^M {}_2F_1(M, M, p-1, R^2\hat{R}^2), \tag{16}$$

with $M \geq p$. In the 2-D case, the multiple coherence and the MSC coincide.

d. Schur complement Finally, if $\boldsymbol{W}$ is decomposed as

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\dagger & \boldsymbol{C} \end{pmatrix}, \tag{17}$$

with $\boldsymbol{C}$ an $r \times r$ matrix, then the Schur complement of $\boldsymbol{C}$,

$$\boldsymbol{W}/\boldsymbol{C} \equiv \boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^\dagger \tag{18}$$

is distributed as $\mathcal{CW}(\mathbf{\Sigma}/\boldsymbol{S}, M-r)$, where $\mathbf{\Sigma}/\boldsymbol{S}$ is the Schur complement of the $r \times r$ block in $\mathbf{\Sigma}$ [[13,14], p. 539]. This result holds only if $M > r$.

III. BAYESIAN INFERENCE FOR SPECTRAL QUANTITIES

We now use the results from the previous section to perform Bayesian inference of the theoretical distribution underlying a set of observed spectral quantities.

Our starting point is the likelihood in Eq. (11), which, when multiplied by an appropriate prior distribution $p(\mathbf{\Sigma})$, yields the Bayesian posterior for the theoretical CPSD matrix $\mathbf{\Sigma}$. Since $\mathbf{\Sigma}$ is the only free parameter in the sample distribution, this posterior fully captures the statistical information of the stochastic processes under investigation.

The key step in this approach is selecting a suitable prior. Before addressing the general case for $p > 1$, we begin with the simpler, yet illuminating case of $p = 1$, the inference of the PSD of a single stochastic process, which provides valuable insight for the general case.

A. Inference of the PSD for a single stochastic process

When $p = 1$, Eq. (11) becomes Eq. (12), and to build a posterior for the PSD S we need a prior $p(S)$.

We have considered three options.

(1) The uniform, noninformative prior $p(S) = \Theta(S)$, with $\Theta(S)$ the Heaviside theta function. With this choice, the posterior distribution of S conditional on

[4]We use the symbolic expression $a \sim A$ to indicate that a random variable a is distributed according to a distribution A. Thus, $\boldsymbol{W} \sim \mathcal{CW}(\mathbf{\Sigma}, M)$.

the observation of Π is

$$S|\Pi \sim \text{invΓ}(M-1, M\Pi), \tag{19}$$

with invΓ the inverse gamma distribution. This posterior is only defined for $M > 1$.

(2) Jeffreys noninformative prior [15]. Calculating the Fisher information $\mathcal{I}(S)$ from Eq. (12), as prescribed by Jeffreys formula, we get $p(S) \propto \sqrt{\mathcal{I}(S)} \propto 1/S$, for $S \geq 0$.

As $p(\log(s)) = Sp(S)$, Jeffreys prior is uniform as a function of $\log(s)$ and corresponds then to a complete lack of prior knowledge even on the order of magnitude of S, a rather realistic description of the situation in most cases of noise calibration.

Note that the main property of Jeffreys prior is the invariance under reparametrization. Thus, the switch $S \to \log(S)$ does not change the prior probability.

With Jeffreys prior

$$S|\Pi \sim \text{invΓ}(M, M\Pi). \tag{20}$$

(3) For comparison we have also considered a prior $p(S) = 1/S^2$ that yields the posterior

$$S|\Pi \sim \text{invΓ}(M+1, M\Pi) \tag{21}$$

The three posteriors above carry some bias. To quantify, it is useful to calculate the posterior predictive distribution of a further observation $\tilde{\Pi}$, conditional on the past observation Π, the PDF of which is, by definition

$$p(\tilde{\Pi}|\Pi) = \int_0^\infty p(\tilde{\Pi}|S)p(S|\Pi)dS. \tag{22}$$

The calculation gives $\tilde{\Pi}|\Pi \sim \beta'(M, \tilde{M}, 1, \Pi)$ with β' the beta prime distribution. The integer $\tilde{M}$ is $\tilde{M} = M-1$, $\tilde{M} = M$, $\tilde{M} = M+1$ for the flat, Jeffreys and $1/S^2$ priors respectively. It seems reasonable that a posterior with minimum bias should assign equal or similar probabilities to future observations larger than the past observation, $\tilde{\Pi} \geq \Pi$, and to those smaller $\tilde{\Pi} \leq \Pi$. This means that the cumulative distribution function (CDF) $c(\tilde{\Pi}|\Pi)$ should obey $c(\tilde{\Pi} = \Pi|\Pi) \simeq 1$. In Fig. 1, we plot $c(\tilde{\Pi} = \Pi|\Pi)$ as a function of M for the three different priors. The figure clearly shows that, within this definition of bias, the only unbiased choice is Jeffreys prior.

In conclusion, given that Jeffreys prior is unbiased, defined down to $M = 1$, invariant under reparametrization, and based on a very realistic assumption about the lack of prior knowledge on the order of magnitude of S, we definitely adopt it as the preferred choice. As a consequence, we adopt the posterior for S in Eq. (20). A plot of the PDF of this posterior for a few choices of M is shown in Fig. 2.

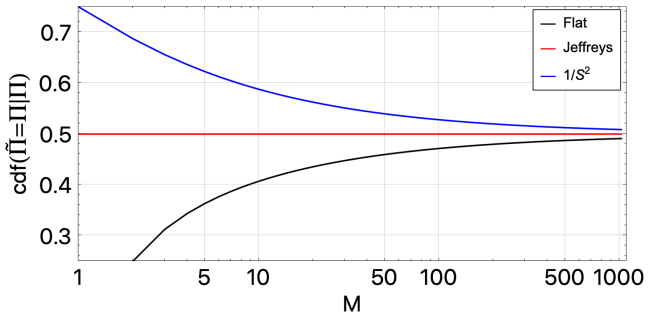


FIG. 1. CDF of the posterior predictive distribution of a future observation $\tilde{\Pi}$, conditional on the past one Π, for the three prior options discussed in the text: flat, Jeffreys, and $1/S^2$. The function is calculated at $\tilde{\Pi} = \Pi$ and plotted as a function of the number of averaged periodograms M.

Note that at low values of M the PDF is rather skew, with rather asymmetric equal probability tails around the median. This shows that a naive use of Gaussian statistics may become highly inaccurate.

To further illustrate this point, we compare the predictions of the posterior in Eq. (20) to those of the simplified, Gaussian-based frequentist method, which defines equal-tail credible intervals for S as $S \in (\Pi \pm k s_\Pi/\sqrt{M})$, where s_Π is the periodogram sample standard deviation. The parameter k determines the likelihood $\ell(k)$ of the interval. Since this method is based on Gaussian statistics, it yields $\ell(1) \approx 0.68$, $\ell(2) \approx 0.95$, and $\ell(3) \approx 0.997$. We also include in the comparison a variant of this method that accounts for the fact that, in Gaussian statistics, $t = (S_t - \Pi)/(s_\Pi/\sqrt{M})$ follows a Student's t-distribution with $M-1$ degrees of freedom. Thus, in the calculation of the equal-tail credible intervals, it would be more accurate to replace $-k$ by the $\ell/2$-quantile of the Student-t distribution with $M-1$ degrees of freedom, and $+k$ by the $(1-\ell/2)$-quantile.

To perform the comparison, we have done a simulation. Each trial of this simulation consists of the following steps.

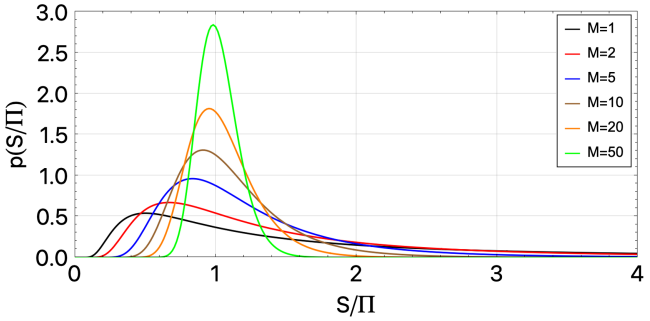


FIG. 2. Plot of $p(S|\Pi)$ for $S|\Pi \sim \text{invΓ}(M, M\Pi)$, for a few different values of M. For the sake of clarity, the PDF is for the ratio S/Π.

(i) We extract M samples Π_i from a $\Gamma(1,1)$ distribution, thus simulating M periodograms of a process with true PSD $S_{\text{true}} = 1$.
(ii) From the samples above, we calculate the sample mean Π and standard deviation s_Π. From these we calculate the credible intervals with likelihoods $\ell(1)$, $\ell(2)$ and $\ell(3)$. We do this for all three methods: the direct frequentist method, the Student-t variant, and the Bayesian posterior in Eq. (20).
(iii) We check which, if any, of these 9 intervals contains the true value $S_{\text{true}} = 1$.

By performing a large number of trials, we estimate the probability p_{miss} that the true value $S = 1$ is not included within each estimated credible interval, and we compare it with the estimated likelihood of this same event $\ell_{\text{miss}} = 1 - \ell$. A consistent estimator should have $p_{\text{miss}}/\ell_{\text{miss}} \simeq 1$. The results of this simulation are shown in Fig. 3.

The figure clearly shows that while the Bayesian estimate is consistent and unbiased, the frequentist method may have a probability of missing the true value significantly exceeding the estimated likelihood. The effect increases at low M and may become rather large for tails beyond the $\ell(1)$ threshold even at $M \simeq 50$.

The effect is because the use of Gaussian statistics predicts a credible interval significantly narrower than that predicted by the correct invΓ one. Thus for the same random sample, the true value may belong to the latter, but fall outside the former.

B. Inference of the entire CPSD matrix

In the general case $p > 1$ some difficulty with the choice of the proper prior for $\mathbf{\Sigma}$ makes the spectral inference more complex. Let us start with the basic choice $p(\mathbf{\Sigma}) = 1$ on the positive definite complex matrices domain. With such a choice

$$\mathbf{\Sigma}|\boldsymbol{W} \sim \mathcal{CW}^{-1}(\boldsymbol{W}, M - p), \tag{23}$$

with $\mathcal{CW}^{-1}$ the complex inverse Wishart distribution [16].

This distribution has a few problems. First, it carries some bias. Though defining what bias is for a matrix distribution may be difficult, it is worth inspecting the posterior predictive distribution of a future observation $\tilde{\boldsymbol{W}}$ conditional on the observation of $\boldsymbol{W}$.

We find that $\boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}}$ is $\boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}} \sim \mathbb{C}\mathrm{B}_p^{II}(M, M - p)$ with $\mathbb{C}\mathrm{B}_p^{II}(a, b)$ the matrix-variate type-2 complex Beta distribution [17].

Reference [17] shows that $\langle \boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}} \rangle = \mathbf{I}_p M/(M - 2p)$ with $\mathbf{I}_p$ the $p \times p$ identity matrix, while for an unbiased estimation one would expect $\langle \boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}} \rangle = \mathbf{I}_p$. Note that this mean value bias depends on the number of series considered together and becomes infinite when $M = 2p$.

That such dependence of the bias on p is paradoxical is well illustrated by the marginal distribution of the diagonal elements. Indeed, the marginal distribution of Σ_{ii}, that is the estimate of the PSD of the ith time series S_i, can be calculated [16] to be $S_i|\Pi_{ii} = \Sigma_{ii}|\Pi_{ii} \sim \mathrm{invΓ}(M - 2p + 1, M\Pi_{ii})$, a distribution only defined for $M \geq 2p$, and different from that one gets by considering the ith series alone, $S_i|\Pi \sim \mathrm{invΓ}(M - 1, M\Pi_{ii})$.

Thus, just assuming there are other $p - 1$ series that may be correlated with the one under study would change the inferred posterior for S_i, and would induce a bias increasing with p.

The situation is slightly better for Jeffreys prior. This can be calculated to be [18] $p(\mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-p}$ yielding

$$\mathbf{\Sigma}|\boldsymbol{W} \sim \mathcal{CW}^{-1}(\boldsymbol{W}, M). \tag{24}$$

With this choice, $\langle \boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}} \rangle = \mathbf{I}_p M/(M - p)$. Still $S_i|\Pi = \Sigma_{ii}|\Pi \sim \mathrm{invΓ}(M - p + 1, M\Pi_{ii})$, instead of the almost unbiased posterior $S_i|\Pi \sim \mathrm{invΓ}(M, M\Pi_{ii})$ that one gets from Jeffreys prior in the $p = 1$ case. Thus, the bias is somewhat reduced but still paradoxically depends on p.

The prior that gives the marginal posterior $\Sigma_{ii}|\Pi \sim \mathrm{invΓ}(M, M\Pi_{ii})$, consistent with that estimated from the ith series alone and Jeffreys prior, is $p(\mathbf{\Sigma}) = |\mathbf{\Sigma}|^{-2p+1}$. For this

$$\mathbf{\Sigma}|\boldsymbol{W} \sim \mathcal{CW}^{-1}(\boldsymbol{W}, M + p - 1). \tag{25}$$

Such prior falls within the class on noninformative priors for $\boldsymbol{Q} = \mathbf{\Sigma}^{-1}$ discussed in [18], $p(\boldsymbol{Q}) \propto |\boldsymbol{Q}|^{-K}\mathrm{etr}[-\boldsymbol{Q}\mathbf{\Lambda}]$. Indeed, remembering that the Jacobian of the transformation $\boldsymbol{Q} \to \mathbf{\Sigma}$ is $|\mathbf{\Sigma}|^{-2p}$ [16], this prior corresponds to $K = -1$ and $\mathbf{\Lambda} = 0$, that is $p(\boldsymbol{Q}) \propto |\boldsymbol{Q}|^{-1}$.

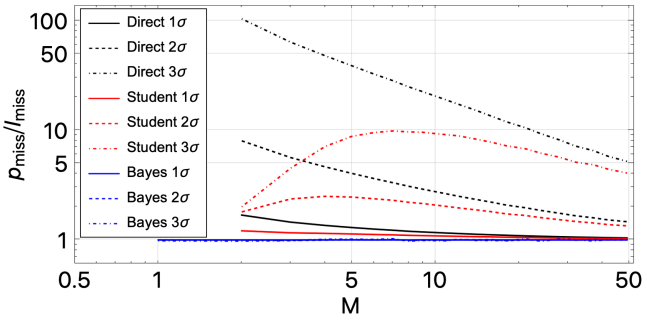


FIG. 3. Comparison of the PSD prediction accuracy of the direct frequentist method, its Student-t variant, and the Bayesian posterior in Eq. (20). The plots show the probability p_{miss} that the prediction misses the true value, divided by the estimated likelihood of that same event. For each method the simulation has been repeated for equal tail credible intervals with likelihood $\ell(1) \approx 0.68(1\sigma)$, $\ell(2) \approx 0.95(2\sigma)$, and $\ell(3) \approx 0.997(3\sigma)$ and as a function of the number M of periodograms in the available sample. The plots for the Bayesian case are barely distinguishable as they are all superimposed on each other at $\simeq 1$, regardless of the value of M.

Note that for the posterior in Eq. (25) $\langle \boldsymbol{W}^{-1} \cdot \tilde{\boldsymbol{W}} \rangle = \mathbf{I}_p M/(M-1)$. As the distribution only holds for $M > 1$, and actually the entire multidimensional Bayesian inference only holds for $M \geq p$, the bias remains smaller than $p/(p-1)$ and independent of p.

Based on the discussion above, we recommend the $p(\boldsymbol{Q}) \propto |\boldsymbol{Q}|^{-1}$ prior when in need of inferring the whole CPSD at a given frequency.

When only particular functions of the CPSD are needed, as in some of the following sections, the proper priors will be formulated in terms of those functions and not of the whole CPSD.

C. Inference of the MSC

We use Eq. (14) to derive the posterior distribution for the theoretical MSC $|\rho|^2$ (implicitly intended as $|\rho_{ij}|^2$). We adopt a constant prior (over $1 \geq |\rho|^2 \geq 0$) for $|\rho|^2$, as it has a closed form which is not far from Jeffreys one (see the end of this subsection for details):

$$p(|\rho|^2||\hat{\rho}|^2) = (M+1)(1-|\rho|^2)^M (1-|\hat{\rho}|^2)^{M-2} \times \frac{{}_2F_1(M,M,1,|\hat{\rho}|^2|\rho|^2)}{{}_2F_1(2,2,2+M,|\hat{\rho}|^2)}. \quad (26)$$

The MSC is mostly used, within noise characterization, as a diagnostic for the existence of linear correlation between two processes. To illustrate to what extent such a diagnostic parameter is effective, we plot in Fig. 4 the $\ell(2)(\simeq 0.95)$ likelihood, equal tail credible interval, and the median predicted by the posterior in Eq. (26). We do that as a function of both $|\hat{\rho}|^2$ and M.

The figure shows that one reaches a reasonable confidence that some correlation exists between the two processes, only when both M and $|\hat{\rho}|^2$ are large enough. For instance, this confidence is never reached for $M = 2$,

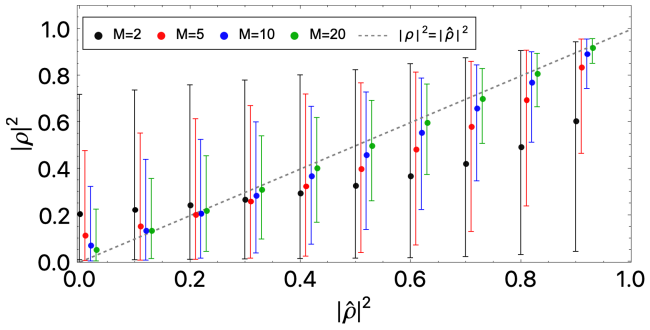


FIG. 4. The equal tail, $\ell(2)(\simeq 0.95)$ likelihood credible intervals for MSC (error bars) as a function of the observed value of $|\hat{\rho}|^2$ and of the number of averaged periodograms M. We use a flat prior on $|\rho|^2$. The central dots are the values of the median. The dashed line $|\rho|^2 = |\hat{\rho}|^2$ is given for reference. For the sake of clarity, for different values of M we plot $|\rho|^2$ at slightly shifted values of $|\hat{\rho}|^2$.

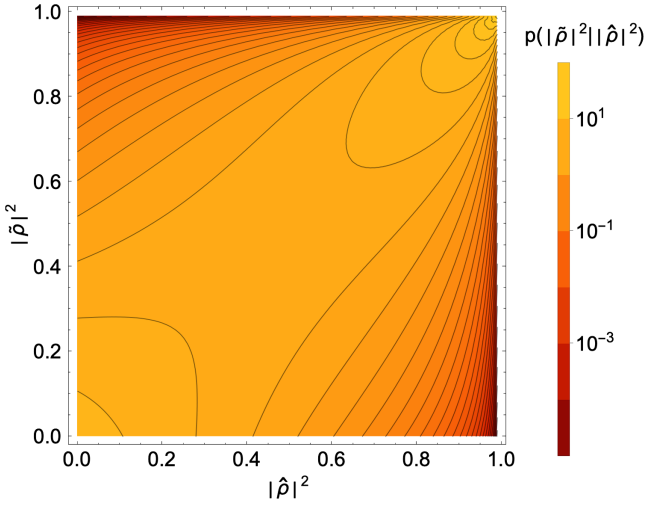


FIG. 5. Contour plot of the probability density function $p(|\tilde{\rho}|^2||\hat{\rho}|^2)$ of the posterior predictive distribution of a future observation $|\tilde{\rho}|^2$ conditional on the past observation $|\hat{\rho}|^2$. The calculation is for $M = 5$.

only if $|\hat{\rho}|^2 \gtrsim 0.6$ for $M = 5$, and even for $M = 20$, one would require $|\hat{\rho}|^2 \gtrsim 0.2$.

Note that the values of the median are always found below the "unbiased" line $|\rho|^2 = |\hat{\rho}|^2$. This bias is only apparent, and in reality, it compensates for the already mentioned significant bias of the sample distribution toward high values.

To check this, we have numerically calculated the posterior predictive distribution of a future observation $|\tilde{\rho}|^2$ conditional on the past observation $|\hat{\rho}|^2$. We give in Fig. 5 a contour plot of $p(|\tilde{\rho}|^2||\hat{\rho}|^2)$ for $M = 5$ that is clearly symmetric around the line $|\tilde{\rho}|^2 = |\hat{\rho}|^2$, thus showing the lack of real bias of our posterior.

To conclude the discussion on the MSC, as anticipated, we show that a flat prior on $|\rho|^2$ does not differ much from the noninformative Jeffreys one. We evaluate numerically Jeffreys prior $p(|\rho|^2)$, built from the Fisher information as $p(|\rho|^2) \propto |\mathcal{I}(|\rho|^2)|^{1/2}$. In Fig. 6, we show the comparison with the flat prior, for a few values of M and $|\hat{\rho}|^2$.

D. Inference of R^2

The case of R^2 is very similar to the one in the previous section. The case with Jeffreys prior is not that far from the flat prior, and can be calculated numerically. We give the closed-form posterior with the flat prior

$$p(R^2|\hat{R}^2) = (M+1)(1-R^2)^M \times \frac{{}_2F_1(M,M,p-1,\hat{R}^2R^2)}{{}_pF_q\left(\begin{matrix}(1,M,M)\\(M+2,p-1)\end{matrix};\hat{R}^2\right)}, \quad (27)$$

where ${}_pF_q$ is the generalized hypergeometric function.

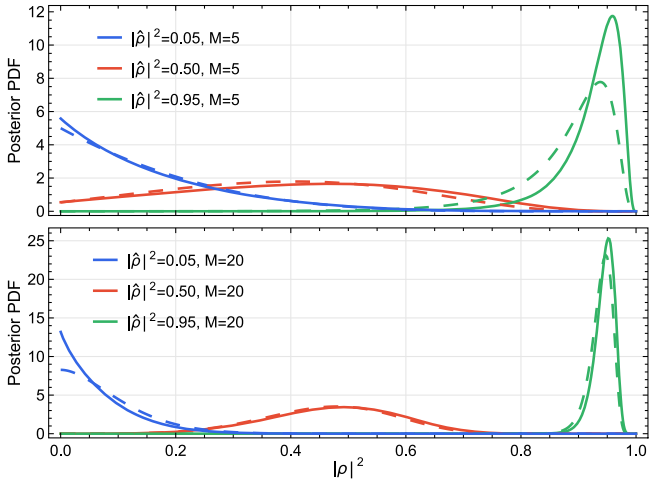


FIG. 6. Comparison between posteriors with Jeffreys prior (solid lines) and the flat prior (dashed lines), with $M = 5$ in the upper panel and $M = 20$ in the lower panel. Different colors represent different experimental MSCs, $|\hat{\rho}|^2$, as in the legend.

As said, R^2 is a generalization of MSC for the case $p > 2$. We will show [Eqs. (32), (33)] that, if $x_1[n]$ is a linear combination of the remaining $p-1$ processes, plus a residual one, then R^2 measures the fraction of the total PSD of $x_1[n]$ that is due to the linear combination of the other processes. So ideally, for completely uncorrelated processes, $R^2 = 0$, while, in the opposite case of negligible residual, $R^2 = 1$.

Similarly to what we have done for the MSC, to get a sense of the effectiveness of this measure, we plot in Fig. 7 the $\ell(2)(\simeq 0.95)$ likelihood, equal tail credible interval, and the median predicted by the posterior in Eq. (27). We do that as a function of both $|\hat{R}|^2$ and M in the case $p = 5$. The plot shows that also $\hat{R}^2$ carries a very significant bias

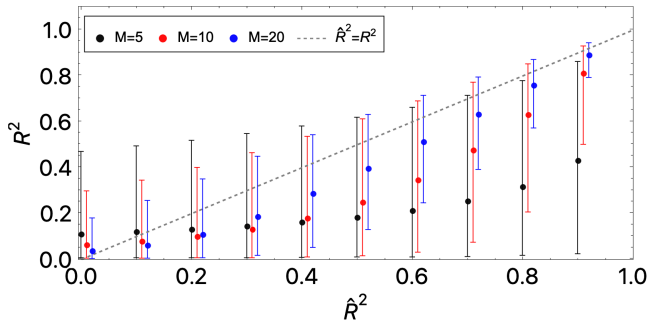


FIG. 7. The equal tail, $\ell(2)(\simeq 0.95.5)$ likelihood credible intervals for the multiple coherence R^2 (error bars) as a function of the observed value of $\hat{R}^2$ and of the number of averaged periodograms M. We use a flat prior on R^2. The calculation is for $p = 5$-variate stochastic process. The central dots are the values of the median. The dashed line $R^2 = \hat{R}^2$ is given for reference. For the sake of clarity, for different values of M we plot R^2 at slightly shifted values of $\hat{R}^2$.

toward higher values, and the posterior compensates for such large bias. Again, to conclude that a significant fraction of the noise power in $x_1[n]$ is contributed by the part correlated with the remaining series, one needs comparatively large values both of M and of $\hat{R}^2$.

IV. NOISE PROJECTION AND TIME SERIES DECORRELATION

We now consider the case where the p-variate stochastic process consists of a "main" process $x(t)$ and $r = p-1$ "disturbances" y_i, with $1 \le i \le r$, modeled as

$$x(t) = x_0(t) + \sum_{i=1}^{r} \int_{-\infty}^{+\infty} \alpha_i(t-t') y_i(t') \mathrm{d}t'. \quad (28)$$

Here, $x_0(t)$—the "residual"—is a process independent of the $y_i(t)$'s. We refer to the Fourier transforms of the functions $\alpha_i(t)$, denoted by $\alpha_i(f)$, as the "susceptibilities."

Our goal is to estimate the PSD $S_{x_0x_0}$ of the residual $x_0(t)$, and the susceptibilities $\alpha_i(f)$. This task, often referred to as noise projection or noise decorrelation, is common in what is known as noise hunting, which involves identifying the source of noise in the main data series of a physical apparatus by examining correlations with other independently measured disturbances that may have coupled into the primary measurement. The noise hunting carried out for LPF was no exception, and it required us to develop the approach described below.

We consider two cases:

(1) When $\alpha_i(t)$ is a general function. This allows estimation of both $\alpha_i(f)$ and $S_{x_0x_0}(f)$ independently at each frequency.

(2) When $\alpha_i(t) = \alpha_i \delta(t)$, with α_i constant. In this case, $\alpha_i(f) = \alpha_i$ becomes frequency independent, and thus a global parameter of the estimate, while $S_{x_0x_0}(f)$ remains dependent on frequency.

It is convenient for the rest of the discussion to make the following block partition of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \left(\begin{array}{c|c} S_{xx} & \boldsymbol{S}_{xy} \\ \hline \boldsymbol{S}_{xy}^{\dagger} & \boldsymbol{S}_{yy} \end{array}\right), \quad (29)$$

where S_{xx} is just the PSD of $x(t)$, $\boldsymbol{S}_{yy}$ is the $r \times r$ CPSD matrix of the disturbances, and $\boldsymbol{S}_{xy}$ is the $1 \times r$ vector of the CPSD between $x(t)$ and all the disturbances $y_i(t)$. Note that, for clarity, we have omitted the explicit dependence of all quantities on frequency. We continue to do so in the rest.

Within the model in Eq. (28),

$$\begin{aligned} S_{xx} &= S_{x_0x_0} + \sum_{i,j=1}^{r} \alpha_i \alpha_j^* S_{y_i,y_j} \\ &= S_{x_0x_0} + \boldsymbol{\alpha} \cdot \boldsymbol{S}_{yy} \cdot \boldsymbol{\alpha}^{\dagger}, \end{aligned} \quad (30)$$

where we have introduced the r-long vector $\boldsymbol{\alpha}$ with components α_i. Furthermore,

$$\boldsymbol{S}_{xy} = \sum_{i=1}^{r} \alpha_i S_{y_i,y_j} = \boldsymbol{\alpha} \cdot \boldsymbol{S}_{yy}, \tag{31}$$

that is $\boldsymbol{\alpha} = \boldsymbol{S}_{xy} \cdot \boldsymbol{S}_{yy}^{-1}$, a relation that will be useful in the following.[5]

Note that $S_{x_0x_0}$ is the Schur complement $\boldsymbol{\Sigma}/\boldsymbol{S}_{yy}$ of $\boldsymbol{S}_{yy}$ in the matrix $\boldsymbol{\Sigma}$

$$S_{x_0x_0} = S_{xx} - \boldsymbol{S}_{xy} \cdot \boldsymbol{S}_{yy} \cdot \boldsymbol{S}_{xy}^{\dagger} = \boldsymbol{\Sigma}/\boldsymbol{S}_{yy} = 1/(\boldsymbol{\Sigma}^{-1})_{11} \tag{32}$$

and that the multiple coherence is

$$R^2 = 1 - \frac{S_{x_0x_0}}{S_{xx}}, \tag{33}$$

i.e., the fraction of the PSD of $x(t)$ contributed by the disturbances, a useful quantity one wants to estimate.

A. Inference of susceptibilities, residuals, and CPSD of disturbances in the general case

Our starting point is a key reparametrization of the sample distribution in Eq. (11). For the sake of such reparametrization, we need to introduce, in analogy with Eq. (29), the block partition of $\boldsymbol{W}$ and $\boldsymbol{\Pi}$

$$\boldsymbol{W} = \left(\begin{array}{c|c} W_{xx} & \boldsymbol{W}_{xy} \\ \hline \boldsymbol{W}_{xy}^{\dagger} & \boldsymbol{W}_{yy} \end{array}\right) = M \left(\begin{array}{c|c} \Pi_{xx} & \boldsymbol{\Pi}_{xy} \\ \hline \boldsymbol{\Pi}_{xy}^{\dagger} & \boldsymbol{\Pi}_{yy} \end{array}\right). \tag{34}$$

Two functions of $\boldsymbol{W}$ that we also need in the following are

(1) The "observed" residual noise PSD Π_0 that we define from

$$\Pi_0 = \frac{1}{M-r} \times \frac{1}{(\boldsymbol{W}^{-1})_{xx}} \equiv \frac{1}{M-r} \times W_0, \tag{35}$$

with $(\boldsymbol{W}^{-1})_{xx}$ the upper-left 1×1 block of $\boldsymbol{W}^{-1}$.

(2) The "observed" susceptibility vector

$$\boldsymbol{\alpha}_0 = \boldsymbol{W}_{xy} \cdot \boldsymbol{W}_{yy}^{-1}. \tag{36}$$

In the Appendix, where we show that the sample distribution in Eq. (11) can be reparametrized as

$$p(\boldsymbol{W}|S_{x_0x_0}, \boldsymbol{\alpha}, \boldsymbol{S}_{yy}) \propto \frac{1}{S_{x_0x_0}^{M}} \exp\left[-\frac{W_0}{S_{x_0x_0}}\right] \times \exp\left[-\frac{(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) \cdot \boldsymbol{W}_{yy} \cdot (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^{\dagger}}{S_{x_0x_0}}\right] \times \frac{1}{|\boldsymbol{S}_{yy}|^{M}} \mathrm{etr}[-\boldsymbol{S}_{yy}^{-1}\boldsymbol{W}_{yy}]. \tag{37}$$

Thus, the distribution splits into two independent parts, one depending on $S_{x_0x_0}$ and $\boldsymbol{\alpha}$ but not on $\boldsymbol{S}_{yy}$, and one that only depends on $\boldsymbol{S}_{yy}$. Therefore, if one selects a prior of the kind $p(S_{x_0x_0}, \boldsymbol{\alpha}) \times p(\boldsymbol{S}_{yy})$, then the posterior also splits into the product of the joint posterior for $S_{x_0x_0}$ and $\boldsymbol{\alpha}$, with the posterior for $\boldsymbol{S}_{yy}$ alone. The estimate of the latter reduce to the estimate of the CPSD that we have already treated. From now on we focus then on the estimate of $S_{x_0x_0}$ and $\boldsymbol{\alpha}$ only.

The most realistic, least informative prior for $S_{x_0x_0}$, as for all other PSDs we have met, is again $p(S_{x_0x_0}) \propto 1/S_{x_0x_0}$ independently of the value of $\boldsymbol{\alpha}$.

On the other hand, the components of $\boldsymbol{\alpha}$ are, in the language of statics, location parameters. If they can be assumed independent of each other, then, for each of them, the least informative prior is just $p(\alpha_i) = 1$.

From the consideration above, it follows that a sound noninformative joint prior for $S_{x_0x_0}$, and $\boldsymbol{\alpha}$ is $p(S_{x_0x_0}, \boldsymbol{\alpha}) = 1/S_{x_0x_0}$, with which their properly normalized joint posterior becomes

$$p(S_{x_0x_0}, \boldsymbol{\alpha}|\boldsymbol{W}) = \frac{(W_0)^{M-r}}{\Gamma(M-r) S_{x_0x_0}^{M-r+1}} \exp\left[-\frac{W_0}{S_{x_0x_0}}\right] \frac{|\frac{\boldsymbol{W}_{y,y}}{S_{x_0x_0}}|}{\pi^r} \exp \times \left[-(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) \cdot \frac{\boldsymbol{W}_{y,y}}{S_{x_0x_0}} \cdot (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^{\dagger}\right]. \tag{38}$$

This is equivalent to stating that

$$S_{x_0x_0}|\boldsymbol{W} \sim \mathrm{inv}\Gamma(M-r, W_0) \tag{39}$$

and

$$\boldsymbol{\alpha}|\boldsymbol{W}, S_{x_0x_0} \sim \mathcal{CN}(\boldsymbol{\alpha}_0, S_{x_0,x_0}\boldsymbol{W}_{yy}^{-1}), \tag{40}$$

with $\mathcal{CN}(S_{x_0,x_0}\boldsymbol{W}_{yy}^{-1}, \boldsymbol{\alpha}_0)$ the complex, circularly symmetric r-variate Gaussian distribution, with covariance matrix $S_{x_0,x_0}\boldsymbol{W}_{yy}^{-1}$ and mean value $\boldsymbol{\alpha}_0$.

Note that the marginal distribution of $S_{x_0x_0}$ is already given by Eq. (39) while that of $\boldsymbol{\alpha}$ may be obtained by integrating Eq. (38) over $S_{x_0x_0}$. By performing this integration, we get

$$\boldsymbol{\alpha} \sim ct_r(\boldsymbol{\alpha}_0, W_0\boldsymbol{W}_{yy}^{-1}, M-r), \tag{41}$$

the latest being the complex multivariate t-distribution for an r-long complex vector, with mean value $\boldsymbol{\alpha}_0$, scale factor

[5]This relation is exactly true only if, at a given frequency f, $\boldsymbol{S}_{xy}(f)$ and $\boldsymbol{S}_{yy}(f)$ are the exact values at f and not those smoothed over the spectral window [see Eq. (3)] that we are using here. This spectral smoothing may bias the estimation of $\boldsymbol{\alpha}(f)$ should it have a strong dependency on frequency. Such bias can be mitigated by properly reducing the width of the spectral window.

$W_0 \boldsymbol{W}_{yy}^{-1}$ and $M-r$ degrees of freedom. This means that the real and imaginary parts of $\boldsymbol{\alpha}$, recast into the $2r$-long real vector $\boldsymbol{\alpha}_R = \begin{pmatrix} \mathrm{Re}\,\boldsymbol{\alpha} \\ \mathrm{Im}\,\boldsymbol{\alpha} \end{pmatrix}$ follow a joint multivariate Student t_{2r} distribution [19]

$$\boldsymbol{\alpha}_R \sim t_{2r}(\boldsymbol{\alpha}_{0,R}, \boldsymbol{\Omega}, 2(M-r)), \tag{42}$$

with $2(M-r)$ degrees of freedom, mean value $\boldsymbol{\alpha}_{0,R} = \begin{pmatrix} \mathrm{Re}\,\boldsymbol{\alpha}_0 \\ \mathrm{Im}\,\boldsymbol{\alpha}_0 \end{pmatrix}$, and a scale matrix given by

$$\boldsymbol{\Omega} = \frac{1}{2}\Pi_0 \begin{pmatrix} \mathrm{Re}\boldsymbol{W}_{yy} & \mathrm{Im}\,\boldsymbol{W}_{yy} \\ -\mathrm{Im}\,\boldsymbol{W}_{yy} & \mathrm{Re}\boldsymbol{W}_{yy} \end{pmatrix}^{-1}. \tag{43}$$

From this joint marginal distribution, we also get the marginal distributions of the single components of $\boldsymbol{\alpha}_R$ that are univariate t distributions with $2(M-r)$ degrees of freedom [19], and scale parameter given by the corresponding element in $\boldsymbol{\Omega}$.

Note that the covariance of the elements of $\boldsymbol{\alpha}_R$, $((M-r)/(M-r-1))\boldsymbol{\Omega}$, decreases with decreasing Π_0, the PSD of residuals. This is expected as, for a given value of the total PSD, a small Π_0 implies a large contribution of the disturbances and then a large signal-to-noise ratio for the components of $\boldsymbol{\alpha}_R$. It is also straightforward to calculate that $\boldsymbol{\Omega} \propto M^{-1}$, so that this signal-to-noise ratio, as expected, also increases with increasing averaging.

To test the validity of our method in a controlled environment, we first studied a simulated case: we generated a set of time series with known correlation, and applied our method to retrieve the known background noise and susceptibilities. This is discussed in Sec. V A. We have also used the approach described in this section to decorrelate the effect of the temperature from the acceleration data series of LPF [3]: details on this are given in Sec. V C.

B. Inference in case of additional readout noise

The model discussed so far assumes the disturbances $y_i(t)$ are measured with negligible readout noise. It is therefore important to consider, before concluding this section, the consequences of applying the method when such noise is in reality not negligible.

Let us consider first the estimate of $S_{x_0x_0}$. Our method in reality estimates $1/\Sigma_{11}^{-1}$, whatever the detailed form of $\boldsymbol{\Sigma}$ is. Indeed our starting point is $1/W_{11}^{-1}$, whose sampling distribution is $1/W_{11}^{-1} \sim \Gamma(M-r, 1/\Sigma_{11}^{-1})$. From this, and the distribution in Eq. (20), one can derive the sampling distribution of our Bayesian estimate for S_{x_0,x_0}. We find that this distribution is $\beta'(M-r, M-r, 1/\Sigma_{11,\mathrm{true}}^{-1})$, a distribution whose median is equal to $1/\Sigma_{11}^{-1}$, and a relative uncertainty that only depends on $M-r$. This confirms that the methods gives an unbiased estimate of $1/\Sigma_{11}^{-1}$.

In the presence of readout noise, $1/\Sigma_{11}^{-1} \neq S_{x_0x_0}$, as the true form of $\boldsymbol{\Sigma}$ is not that in Eq. (29). Indeed, in the simplest model of additive noise, the measured disturbance is $y_i(t) + n_i(t)$, with $n_i(t)$ a zero-mean stationary process independent of all the y_i's. In this case, the lower diagonal block of $\boldsymbol{\Sigma}$ becomes $\boldsymbol{S}_{yy} + \boldsymbol{S}_n$ with $\boldsymbol{S}_n$ a diagonal matrix whose generic element S_{n_i,n_i} is the PSD of $n_i(t)$.

Working out the formula for $1/\Sigma_{11}^{-1}$ becomes particularly simple if also $\boldsymbol{S}_{yy}$ is diagonal, that is, if the disturbances are mutually uncorrelated. One can readily calculate that in this case

$$1/\Sigma_{11}^{-1} = S_{x_0x_0} + \sum_{i=1}^{r} |\alpha_i|^2 \frac{S_{y_i,y_i} S_{n_i n_i}}{S_{y_i,y_i} + S_{n_i n_i}}, \tag{44}$$

with $S_{y_iy_i}$ the PSD of $y_i(t)$. One can recognize that in the limit of dominant readout noise $1/\Sigma_{11}^{-1} \to S_{xx}$. In other words, a dominant readout noise, as expected, completely obscures any correlation between $x(t)$ and the y's.

Furthermore, within the same simplification of uncorrelated disturbances, the product $\sum_{j=2}^{p} \Sigma_{1,j}\Sigma_{j,k}^{-1}$, with $k > 1$ which, in the noiseless limit, is $(\boldsymbol{S}_{xy} \cdot \boldsymbol{S}_{yy}^{-1})_{k-1} = \alpha_{k-1}$, becomes instead

$$\sum_{j=2}^{p} \Sigma_{1,j}\Sigma_{j,k}^{-1} = \frac{S_{y_{k-1}y_{k-1}}}{S_{y_{k-1},y_{k-1}} + S_{n_{k-1}n_{k-1}}} \alpha_{k-1}. \tag{45}$$

Thus, in the presence of significant readout noise, our method overestimates the PSD of the residuals, underestimates the absolute value of the susceptibility, and should only be used for an upper limit on S_{x_0,x_0}.

C. The case of real frequency-independent susceptibilities

In many practical circumstances, $\boldsymbol{\alpha}$ may be in practice a real frequency-independent vector. If this is the case, $\boldsymbol{\alpha}$ becomes a common parameter in the sampling distribution of the $\boldsymbol{W}$'s at all frequencies. Thus, to build up a posterior for $\boldsymbol{\alpha}$, one needs to consider the joint likelihood of all the $\boldsymbol{W}$'s for a given value of $\boldsymbol{\alpha}$. We anticipate that, in this case, we do not get a closed form posterior distribution, but a useful form that can be integrated by using the Markov chain Monte Carlo (MCMC) approach.

To build this posterior, let us call $\boldsymbol{W}_i$ the sample at frequency f_i, obtained by averaging over M_i periodograms, and similarly let us indicate the corresponding theoretical quantities with $S_{x_0,x_0,i}$ and $\boldsymbol{\Sigma}_{yy,i}$.

Let us also assume that f_i and f_{i+i} are sufficiently far apart that $\boldsymbol{W}_i$ and $\boldsymbol{W}_{i+1}$ may be treated as independent, so that the likelihood of the $\boldsymbol{W}_i$ is just the product of their marginal likelihoods $\prod_{i=1}^{N_f} p(\boldsymbol{W}_i | S_{x_0x_0,i}, \boldsymbol{\alpha}, \boldsymbol{S}_{yy,i})$, with N_f the number of considered frequencies.

In addition, we assume that $S_{x_0,x_0,i}$ and $\boldsymbol{\Sigma}_{yy,i}$ have independent prior distributions, that the prior for $S_{x_0,x_0,i}$

is, as before, $\propto 1/S_{x_0,x_0,i}$, and that the priors for the component of $\boldsymbol{\alpha}$ are, again as before, independent and uniform.

Using also the fact that, for real $\boldsymbol{\alpha}$, Eq. (A9) gives $W'_{x_0x_0,i} = W_{xx,i} - 2\boldsymbol{\alpha}\cdot\mathrm{Re}(\boldsymbol{W}_{xy,i}) + \boldsymbol{\alpha}\cdot\mathrm{Re}(\boldsymbol{W}_{yy,i})\cdot\boldsymbol{\alpha}$, we get for the logarithm Λ of the joint posterior of all parameters

$$\begin{aligned}\Lambda = &-\sum_{i=1}^{N_f}(M_i+1)\log(S_{x_0,x_0,i}) \\ &-\sum_{i=1}^{N_f}\frac{W_{xx,i}}{S_{x_0x_0,i}} + 2\boldsymbol{\alpha}\cdot\sum_{i=1}^{N_f}\frac{\mathrm{Re}(\boldsymbol{W}_{xy,i})}{S_{x_0x_0,i}} \\ &-\boldsymbol{\alpha}\cdot\left(\sum_{i=1}^{N_f}\frac{\mathrm{Re}(\boldsymbol{W}_{yy,i})}{S_{x_0x_0,i}}\right)\cdot\boldsymbol{\alpha}, \end{aligned} \tag{46}$$

plus an independent term that only contains $\boldsymbol{S}_{yy,i}$ and that is not used here.

The posterior in Eq. (46) can be used for the MCMC estimate of the parameter posterior distribution. We have used this method extensively within the data processing of LPF [3]. To test the validity of our method in a controlled environment, we applied it to the simulated case, as discussed in Sec. V B.

Before closing this section, it is worth noticing that one advantage of this simultaneous fit to the data at all frequencies, is that one can include also data at very low frequency where the condition $M_i \geq p$ may be violated. This is shown as follows.

First, the distribution of a singular $\boldsymbol{W}_i$ obtained from $M_i < p$ periodograms, is the singular complex Wishart distribution [20]

$$p(\boldsymbol{W}_i|\boldsymbol{\Sigma}_i) = \frac{\pi^{M_i(M_i-p)}|\boldsymbol{\Lambda}_i|^{M_i-p}}{\tilde{\Gamma}_p(M_i)|\boldsymbol{\Sigma}_i|^{M_i}}\mathrm{etr}[-\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_i], \tag{47}$$

with $|\boldsymbol{\Lambda}_i|$, the product of the nonzero eigenvalues of $\boldsymbol{W}_i$. The dependence of the likelihood in Eq. (47) on $\boldsymbol{\Sigma}$ and M is the same as that in Eq. (11). Thus the difference between the two likelihoods makes no difference for the derivation of the posterior. Furthermore, the stability of the posterior in Eq. (46), requires that the matrices $\sum_{i=1}^{N_f}\frac{\mathrm{Re}(\boldsymbol{W}_{xy,i})}{S_{x_0x_0,i}}$ and $\sum_{i=1}^{N_f}\frac{\mathrm{Re}(\boldsymbol{W}_{yy,i})}{S_{x_0x_0,i}}$ are full rank, not the individual $\boldsymbol{W}_i$. As the rank of a sum of positive semidefinite matrices is larger than or equal to the maximum rank of the terms in the sum [21], it suffices that just one of the $\boldsymbol{W}_i$ is of full rank to give full rank to both sums.

As, in practice, all the $\boldsymbol{W}_i$ above a certain frequency are full rank, the posterior is well defined even if a few terms at the lowest frequency have $M_i < p$.

V. TEST SIMULATIONS AND APPLICATION TO REAL DATA

A. Test simulation: Single frequency

We have tested our method in a controlled simulation environment. We have generated a time series $x(t) = x_0(t) + \sum_{i=1}^{3} n_i z_i(t)$, with all series Gaussian and zero-mean, and with n_i three real coefficients. We have also generated the three "observed" disturbances $y_i(t) = h(t) * z_i(t)$, with $h(t)$ the impulse response of a low-pass filter, and with $*$ indicating the time-convolution.

Within this simple model, the susceptibilities become $\alpha_i(f) = n_i/h(f)$ with $h(f)$ the frequency response of the filter, that is the Fourier transform of $h(t)$. The susceptibilities are then complex, frequency dependent, and noncausal.

For the simulation, we selected the PSD of the residuals x_0 to consist of a $\propto 1/f^2$ low frequency tail merging into a plateau extending up to some double-pole roll-off. More explicitly[6] (see Fig. 8)

$$\begin{aligned}S_{x_0}(f) = &\left|\frac{(1-e^{-2\pi f_1 T})}{1-e^{-2\pi f_1 T}e^{-i2\pi fT}}\frac{(1-e^{-2\pi f_2 T})}{1-e^{-2\pi f_2 T}e^{-i2\pi fT}}\right|^2 \\ &+\left|\frac{1-e^{-T/\tau}e^{-i2\pi f_0 T}}{1-e^{-T/\tau}e^{-i2\pi fT}}\right|^2, \end{aligned} \tag{48}$$

with $T = 1$ s the sampling time, $f_1 = 0.10$ Hz and $f_2 = 0.11$ Hz the two roll-off frequencies, and $f_0 = 1$ mHz the cross-over frequency between the tail and the plateau.

The disturbances are in the form $z_i(t) = c_{lf,i}z_{lf,i}(t) + c_{hf,i}z_{hf,i}(t) + c_i z_0(t)$, where all the time series on the right hand side are Gaussian, zero mean, and mutually independent, and the coefficient $c_{lf,i}$, $c_{hf,i}$ and c_i are real and randomly selected.

The $z_{lf,i}(t)$ and $z_0(t)$ share the same PSD

$$S_{lf}(f) = \left|\frac{1-e^{-T/\tau_1}e^{-i2\pi f_0 T}}{1-e^{-T/\tau_1}e^{-i2\pi fT}}\frac{1-e^{-T/\tau_2}e^{-i2\pi f_0 T}}{1-e^{-T/\tau_2}e^{-i2\pi fT}}\right|^2, \tag{49}$$

with $\tau_1 = 1.0\times10^5$ s and $\tau_2 = 1.1\times10^5$ s. For $f \gg 1/\tau_1, 1/\tau_2$ this PSD amounts to a $\propto 1/f^4$ low frequency tail with unit value at $f = f_0$.

The $z_{hf,i}(t)$ have PSD

$$S_{hf}(f) = \left|\frac{1-e^{-T/\tau}e^{-i2\pi f_0 T}}{1-e^{-T/\tau}e^{-i2\pi fT}}\right|^2, \tag{50}$$

again a $\propto 1/f^2$ tail with unit value at $f = f_0$. The presence of the shared series $z_0(t)$ induces correlation among the z's with CPSD $S_{z_i,z_j}(f) = c_i c_j S_{lf}(f)$.

[6]Note that in this section we show and calculate single-sided PSDs, as this is the standard practice. Discussion, results, and susceptibilities are not affected by this choice.

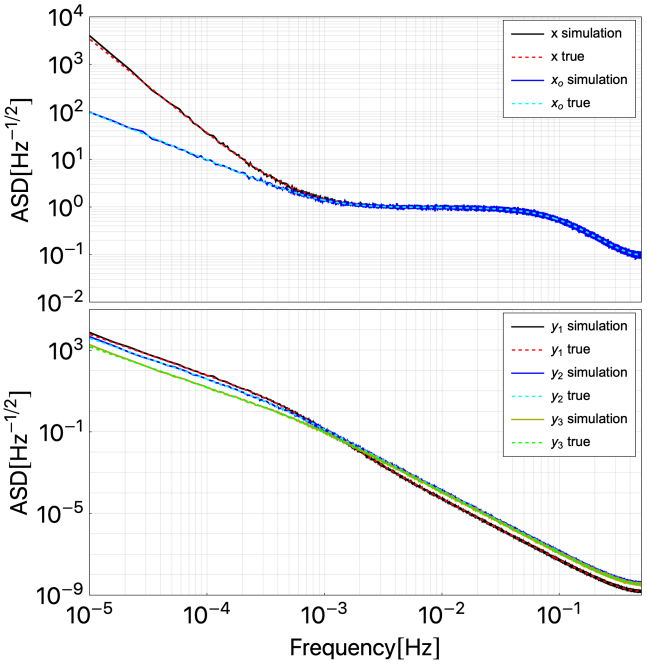


FIG. 8. ASD of time series used in the simulation (Welch periodograms). The dashed lines represent the "true values" that is those used to generate the simulated data that have been calculated from Eqs. (48) to (50), and from one random extraction of the set of numerical coefficients n_i, $c_{lf,i}$, $c_{hf,i}$ and c_i. The noisy lines are the averages of the estimated ASD from 100 different simulations generated from the true spectrum above. Time series are sampled with $T = 1$ and last 5×10^5 s. Frequency dependent data partition for periodogram calculation is performed according to the method of Ref. [9]. We used the Nuttall four-coefficient minimal-side-lobe spectral window [22].

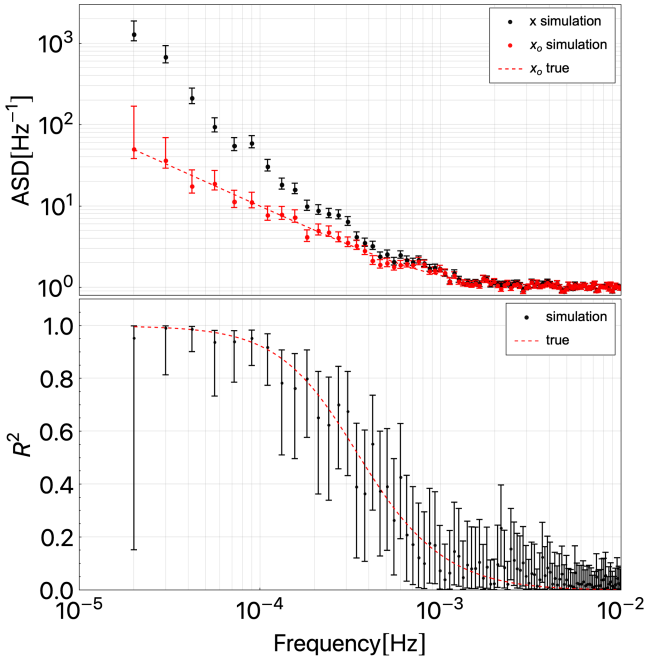


FIG. 9. Example of noise decorrelation for a single realization of the multivariate time series $\{x(t), y_1(t), y_2(t), y_3(t)\}$, generated with the same coefficients as in Fig. 8. Top panel: black points—posterior ASD of $x(t)$ from Eq. (20); red points—posterior ASD of $x_0(t)$ from Eq. (39). Points denote medians, with $\ell(1)$ (≃68.5%) equal-tail credible intervals. ASDs are computed with a Nuttall window; frequency spacing yields 10–30% correlation between adjacent estimates and negligible correlation second-nearest neighbors. The red dashed line shows the true value from Eq. (48). Bottom panel: black points—posterior of R^2 from Eq. (26), with medians and $\ell(2)$ (≃95.5%) symmetric credible intervals. The red dashed line shows the true value, $1 - S_{x_0}(f)/S_x(f)$.

All PSDs above must be intended to be zero for $|f| \geq 1/(2T)$. With this prescription, they can be read as discrete time Fourier transforms of the corresponding discrete time autocorrelation, and their shape allows a straightforward implementation as autoregressive moving average (ARMA) stochastic processes. In Fig. 8, we illustrate the ASD of the simulated time series. As a first, preliminary assessment, we have checked that all PSDs of the simulated data follow a gamma distribution, as stated in Eq. (12).

As for the filter $h(t)$, its transfer function is

$$h(f) = \frac{1 - e^{-2\pi T/\tau_a}}{1 - e^{-2\pi T/\tau_a} e^{-i2\pi T f}} \frac{1 - e^{-2\pi T/\tau_b}}{1 - e^{-2\pi T/\tau_b} e^{-i2\pi T f}}, \quad (51)$$

with $\tau_a = 2000$ s and $\tau_a = 2001$ s, and zero for $|f| \geq 1/(2T)$.

One can recognize the transfer function of a discrete-time two-pole infinite impulse response low-pass filter, which is easily implemented again as an ARMA filter on the discrete time series of the z's. This model possesses many features of a realistic situation, complex frequency dependence of PSD, cross-correlation among disturbances, high data dynamic range, complex susceptibility etc., to give a meaningful test of the method.

In Fig. 9 we apply the decorrelation method described in Sec. IV A on one example of a 5×10^5 s multivariate time series generated as described above.

The figure shows that the method, at least for this example, gives an unbiased estimate of the ASD of the residuals. The ASD indeed fluctuates, within the uncertainties predicted by the posterior in Eq. (20), around the true value in Eq. (48). The figure also shows, for reference, the estimate of the multiple coherence coefficient R^2, from the posterior in Eq. (27). The plot indicates that at $f \simeq 1$ mHz, where $M \simeq 30$, the method allows to detect a ≃10% contribution of the disturbances to the total PSD.

In Fig. 10 we show the estimate of the susceptibility $\alpha_1(f)$ for the same set of data used for Fig. 9 and compare it with the true value $n_1/h(f)$, with $h(f)$ from Eq. (51). The figure is limited to $f \simeq 1$ mHz, as above that frequency the susceptibility is in practice compatible with $\alpha_1 = 0$, as expected from the fact that the contribution of the

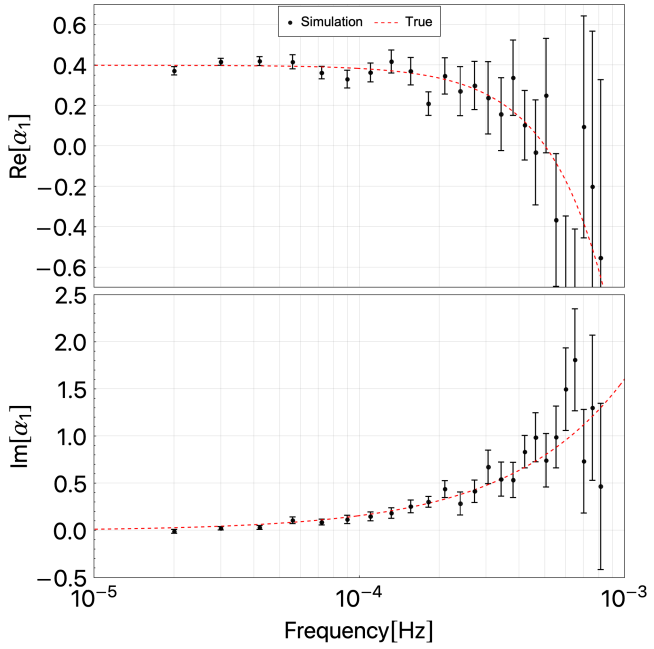


FIG. 10. Black data points: susceptibility $\alpha_1(f)$ of $x(t)$ to $y_1(t)$ for the example data of Fig. 9 from its marginal Student t-posterior. Dots represent the medians of the posteriors, while error bars delimit their symmetric-tail $\ell(1)$ credible intervals. Red dashed line: true susceptibility value $n_1/h(f)$, with $h(f)$ from Eq. (51).

disturbances to total power becomes undetectable. Within this frequency range the estimate appears unbiased and in agreement with the true value within the uncertainties predicted by the proper marginal t-distribution.

B. Test simulation: Frequency-independent susceptibilities

We also tested the method in Sec. IV C, applying it to simulations. We choose time series with the same properties as in Sec. V A, except that here the disturbances have not been filtered, that is, $y_i(t) = z_i(t)$. With this prescription, the susceptibilities are just the real, frequency-independent numbers $\alpha_i = n_i$. The result of such a simulation is presented in Figs. 11, 12. The figures show again that the method gives a consistent and unbiased estimate of all the $S_{x_0x_0,i}$ and all the α_i.

C. Application to real data

Finally, we apply the method developed in the previous sections to real data from the LPF mission. Over a run lasting 18.5 days, the relative acceleration of two test masses, Δg, and the average temperature were synchronously measured with a sampling time of 0.1 s. The long duration of the measurement allows us to probe frequencies as low as 3.75 μHz with $M = 2$. By selecting sufficiently spaced frequencies such that ASD estimates can be treated as statistically independent, we employ the single-frequency approach described in Sec. IV A.

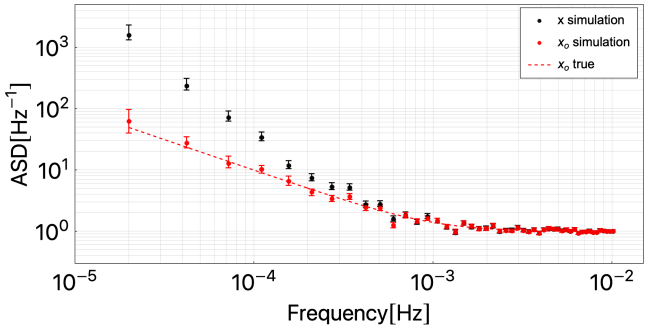


FIG. 11. Example of noise decorrelation for one sample of the multivariate time series $\{x(t), y_1(t), y_2(t), y_3(t)\}$ with real frequency independent susceptibilities. Meanings of quantities are the same as those in the upper panel of Fig. 9. The posterior for the ASD of $x_0(t)$ has been obtained with an MCMC integration of the posterior in Eq. (46). Data were taken at every other frequency of the data in Fig. 9, both to simplify the calculation and to ensure their mutual independence.

The results are shown in Fig. 13. At μHz frequencies, temperature-induced forces dominate the acceleration noise, and the decorrelation procedure proves highly effective. This, in turn, enables a reliable estimation of the susceptibility α, indicating that its imaginary part is consistent with zero, and providing a precise measurement of its real part.

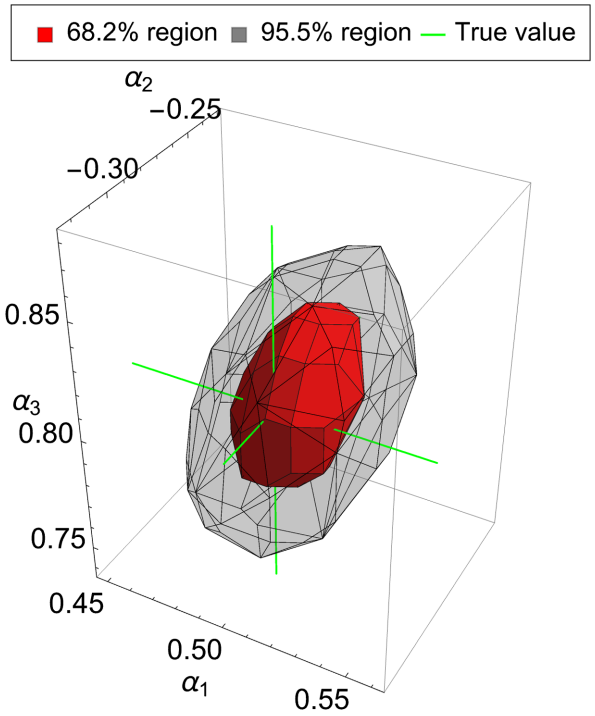


FIG. 12. The joint posterior for the three susceptibilities α_1, α_2, and α_3. The red surface delimits a credible region with $\simeq\ell(1)$ likelihood, while the cyan, semitransparent surface delimits a credible region with $\simeq\ell(2)$ likelihood. The green axes cross at the true value $\boldsymbol{\alpha} = \mathbf{n}$, with $\mathbf{n}$ the vector with components n_i used in the simulation.

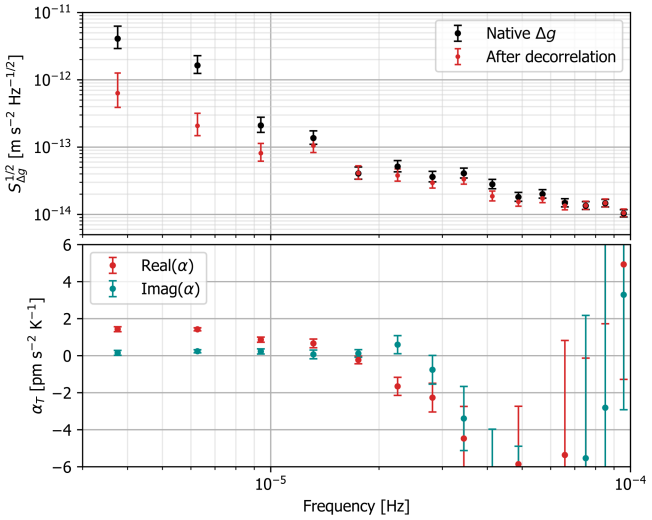


FIG. 13. Decorrelation of real data from the LPF mission. Top panel: noise decorrelation, native PSD and projected PSD after the application of our method. Bottom panel: estimate of the susceptibilities, real and imaginary parts.

VI. CONCLUSIONS

In conclusion, we have presented a set of Bayesian low-bias closed-form posteriors—based on simple and physically meaningful priors—for the most commonly estimated quantities at a given frequency in the spectral analysis of multivariate time series, and in particular in noise projection of physical instruments.

The distributions of some of these posteriors are available within the main software platforms, which makes the calculation of credible intervals and other statistical quantities particularly simple. For the others, we give the explicit form of the PDF that can be used to numerically calculate the relevant statistical quantities. For the reader's convenience, these posteriors are summarized in Table I.

For the case of noise projection, we have shown with simulations that the method is capable of retrieving, with negligible bias, a residual whose ASD is orders of magnitude smaller than the part due to the measured disturbances, and we have also investigated the robustness of the method in the presence of readout noise in the disturbance measurement.

Still on noise projection, in addition to the single frequency closed-form prior, we have also presented a simple likelihood to be used in multifrequency noise projection in case the susceptibilities may be confidently assumed to be real and frequency independent.

We want to stress again that these results originate from the experience of data processing at very low frequency, from μHz to Hz [3], in which, due to length of the required measurement time, only comparatively few periodograms are available, and should be particularly suitable for any situation in which a similar limitation in number of available

TABLE I. Summary of closed-form posteriors presented in this paper. For the meaning of the symbols, please refer to the section indicated in the rightmost column.

General Spectral Estimation				
Quantity	Symbol	Posterior	PDF	Section
Single series PSD	S	$S \sim \mathrm{inv}\Gamma(M, M\Pi)$	$\frac{e^{-\frac{M\Pi}{S}}(\frac{M\Pi}{S})^M}{S\Gamma(M)}$	Sec. III A
Multivariate series CPSD	$\boldsymbol{\Sigma}$	$\boldsymbol{\Sigma} \sim \mathcal{CW}^{-1}(\boldsymbol{W}, M+p-1)$	$\frac{\lvert\boldsymbol{W}\rvert^{M+p-1}\mathrm{etr}[-\boldsymbol{\Sigma}^{-1}\boldsymbol{W}]}{\tilde{\Gamma}_p(M-p+1)\lvert\boldsymbol{\Sigma}\rvert^{M+2p-1}}$	Sec. III B
Two-series MSC	$\lvert\rho\rvert^2$	⋯	(Flat prior)$(M+1)(1-\lvert\rho\rvert^2)^M$ $\times(1-\lvert\hat{\rho}\rvert^2)^{M-2}\frac{{}_2F_1(M,M,1,\lvert\hat{\rho}\rvert^2\lvert\rho\rvert^2)}{{}_2F_1(2,2,2+M,\lvert\hat{\rho}\rvert^2)}$	Sec. III C
Multiplecoherence	R^2	⋯	(Flat prior)$(M+1)(1-R^2)^M \times \times \frac{{}_2F_1(M,M,p-1,\hat{R}^2R^2)}{{}_pF_q\left(\begin{matrix}(1,M,M)\\(M+2,p-1)\end{matrix};\hat{R}^2\right)}$	Sec. III C

Noise Projection				
Quantity	Symbol	Posterior	PDF	Section
PSD of residual(marginal)	$S_{x_0x_0}$	$S_{x_0x_0} \sim \mathrm{inv}\Gamma(M-r, W_0)$	$\frac{e^{-\frac{W_0}{S}}(\frac{W_0}{S})^{(M-r)}}{S\Gamma(M-r)}$	Sec. IV
Susceptibilities(marginal)	$\boldsymbol{\alpha}_R$	$\boldsymbol{\alpha}_R \sim t_{2r}(\boldsymbol{\alpha}_{0,R}, \boldsymbol{\Omega}, 2(M-r))$	$\frac{\Gamma(M)}{\pi^r(2(M-r))^r\Gamma(M-r)\lvert\boldsymbol{\Omega}\rvert^{1/2}}$ $\times(1+\frac{(\boldsymbol{\alpha}_R-\boldsymbol{\alpha}_{0,R})\cdot\boldsymbol{\Omega}^{-1}\cdot(\boldsymbol{\alpha}_R-\boldsymbol{\alpha}_{0,R})}{2(M-r)})^{-M}$	Sec. IV
Susceptibilities(conditional to $S_{x_0x_0}$)	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}\lvert S_{x_0x_0} \sim \mathcal{CN}(\boldsymbol{\alpha}_0, S_{x_0x_0}\boldsymbol{W}_{yy}^{-1})$	$\frac{\lvert\frac{\boldsymbol{W}_{y,y}}{S_{x_0x_0}}\rvert}{\pi^r}e^{-(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\cdot\frac{\boldsymbol{W}_{y,y}}{S_{x_0x_0}}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^\dagger}$	Sec. IV

periodograms may occur. We have tested the method against controlled simulations, and provided an application to the LPF mission, in the decorrelation of force-inducing thermal effects from the observed acceleration data.

Finally, as for almost all commonly used results in data processing, also those presented here are based on Gaussian statistics of the time series under processing. The applicability of the result depends then on how much the data may be safely considered Gaussian.

ACKNOWLEDGMENTS

This work has been supported in part by Agenzia Spaziale Italiana (ASI), Project No. 2017-29-H.0-2020 "Attività per la fase A della missione LISA," and Project No. 2024-36-HH.0 "Attività per la fase B2/C della missione LISA." The authors would like to acknowledge various useful discussions with all the members of the Trento LISA group.

DATA AVAILABILITY

Some of the data that support the findings of this article are openly available [7]. Simulation data are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX: DERIVATION OF NOISE-PROJECTION POSTERIOR

Starting from the complex Wishart distribution Eq. (11), we derive the joint posterior of the decorrelation parameters in Eq. (38). This was synthetically discussed in [3], and we repeat it here, in more detail, for the reader's convenience.

We start by defining the block matrix $\boldsymbol{U}$

$$\boldsymbol{U} = \left(\begin{array}{c|c} 1 & \boldsymbol{\alpha} \\ \hline \mathbf{0} & \mathbf{I}_r \end{array}\right), \tag{A1}$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix.

$\boldsymbol{U}$ performs the linear transformation $x_0 \to x$, $y_i \to y_i$. Its inverse is obtained from $\boldsymbol{U}$, by simply replacing $\boldsymbol{\alpha}$ with $-\boldsymbol{\alpha}$.

A second important block matrix is,

$$\boldsymbol{\Sigma}' = \left(\begin{array}{c|c} S_{x_0x_0} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{S}_{yy} \end{array}\right). \tag{A2}$$

It is straightforward to show that

$$\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Sigma}'\boldsymbol{U}^\dagger, \tag{A3}$$

where $\boldsymbol{U}^\dagger$ is the conjugate transpose of matrix $\boldsymbol{U}$.

As the determinant of $\boldsymbol{U}$ is $|\boldsymbol{U}| = 1$, then

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}'| = S_{x_0x_0}|\boldsymbol{S}_{yy}|. \tag{A4}$$

The exponent in Eq. (11) can be rewritten in terms of $\boldsymbol{\Sigma}'$ as

$$\mathrm{etr}[-\boldsymbol{\Sigma}^{-1}\boldsymbol{W}] = \mathrm{etr}[-\boldsymbol{\Sigma}'^{-1}\boldsymbol{W}'], \tag{A5}$$

with $\boldsymbol{W}'$ defined as $\boldsymbol{W}' = \boldsymbol{U}^{-1}\boldsymbol{W}(\boldsymbol{U}^\dagger)^{-1}$.

As

$$\boldsymbol{\Sigma}'^{-1} = \left(\begin{array}{c|c} 1/S_{x_0x_0} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{S}_{yy}^{-1} \end{array}\right), \tag{A6}$$

by using the block decomposition

$$\boldsymbol{W}' = \left(\begin{array}{c|c} W'_{x_0x_0} & \boldsymbol{W}'_{x_0y} \\ \hline (\boldsymbol{W}'_{x_0y})^\dagger & \boldsymbol{W}'_{yy} \end{array}\right), \tag{A7}$$

we get

$$\begin{aligned} \mathrm{etr}[-\boldsymbol{\Sigma}'^{-1}\boldsymbol{W}'] &= e^{-\frac{W'_{x_0x_0}}{S_{x_0x_0}}}\mathrm{etr}[-\boldsymbol{S}_{yy}^{-1}\cdot\boldsymbol{W}'_{yy}] \\ &= e^{-\frac{W'_{x_0x_0}}{S_{x_0x_0}}}\mathrm{etr}[-\boldsymbol{S}_{yy}^{-1}\cdot\boldsymbol{W}_{yy}], \end{aligned} \tag{A8}$$

where, in the last term, we have used the straightforward result $\boldsymbol{W}'_{yy} = \boldsymbol{W}_{yy}$.

$W'_{x_0x_0}$ in Eq. (A8) may be written as

$$W'_{x_0x_0} = W_{xx} - \boldsymbol{\alpha}\cdot\boldsymbol{W}_{xy}^\dagger - \boldsymbol{W}_{xy}\cdot\boldsymbol{\alpha}^\dagger + \boldsymbol{\alpha}\cdot\boldsymbol{W}_{yy}\cdot\boldsymbol{\alpha}^\dagger. \tag{A9}$$

By introducing

$$\boldsymbol{\alpha}_0 = \boldsymbol{W}_{xy}\cdot\boldsymbol{W}_{yy}^{-1}, \tag{A10}$$

we get

$$\begin{aligned} \boldsymbol{\alpha}\cdot\boldsymbol{W}_{xy}^\dagger &= \boldsymbol{\alpha}\cdot\boldsymbol{W}_{yy}\cdot\boldsymbol{\alpha}_0^\dagger \\ \boldsymbol{W}_{xy}\cdot\boldsymbol{\alpha}^\dagger &= \boldsymbol{\alpha}_0\cdot\boldsymbol{W}_{yy}\cdot\boldsymbol{\alpha}^\dagger, \end{aligned} \tag{A11}$$

so that

$$\begin{aligned} W'_{x_0x_0} = W_{xx} &+ (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\cdot\boldsymbol{W}_{yy}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^\dagger \\ &- \boldsymbol{W}_{xy}\cdot\boldsymbol{W}_{yy}^{-1}\cdot\boldsymbol{W}_{x,y}^\dagger. \end{aligned} \tag{A12}$$

But

$$= W_{xx} - \boldsymbol{W}_{xy}\cdot\boldsymbol{W}_{yy}^{-1}\cdot\boldsymbol{W}_{x,y}^\dagger = \boldsymbol{W}/\boldsymbol{W}_{yy} = \frac{1}{(\boldsymbol{W}^{-1})_{1,1}}, \tag{A13}$$

thus finally

$$W'_{x_0x_0} = \frac{1}{(\boldsymbol{W}^{-1})_{1,1}} + (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\cdot\boldsymbol{W}_{yy}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^\dagger. \tag{A14}$$

As $W'_{x_0x_0}$ is independent of $\boldsymbol{S}_{yy}$, the probability density in Eq. (11) splits in the product of two pieces, one that contains only $\boldsymbol{\alpha}$ and $S_{x_0x_0}$, and one that contains only $\boldsymbol{S}_{yy}$.

Using Eq. (A14), and the definition in Eq. (35), we get

$$p(\boldsymbol{W}|S_{x_0x_0},\boldsymbol{\alpha},\boldsymbol{S}_{yy}) = \frac{|\boldsymbol{W}|^{M-p}}{\tilde{\Gamma}_p(M)}\frac{1}{S_{x_0x_0}^M} e^{\frac{(M-r)\Pi_0}{S_{x_0x_0}}} e^{-\frac{(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\cdot\boldsymbol{W}_{yy}\cdot(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^\dagger}{S_{x_0x_0}}} \frac{1}{|\boldsymbol{S}_{yy}|^M}\mathrm{etr}[-\boldsymbol{S}_{yy}^{-1}\boldsymbol{W}_{yy}]. \tag{A15}$$

This is used for Eq. (37).

[1] P. D. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. **15**, 70 (1967).

[2] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in Electrical Engineering: Communications and Signal Processing (Tata McGraw-Hill, 2002).

[3] M. Armano *et al.*, In-depth analysis of LISA Pathfinder performance results: Time evolution, noise projection, physical models, and implications for LISA, Phys. Rev. D **110**, 042004 (2024).

[4] M. Armano *et al.*, Beyond the required LISA free-fall performance: New LISA pathfinder results down to 20 μHz, Phys. Rev. Lett. **120**, 061101 (2018).

[5] M. Armano *et al.*, Sub-femto-*g* free fall for space-based gravitational wave observatories: LISA pathfinder results, Phys. Rev. Lett. **116**, 231101 (2016).

[6] M. Armano *et al.*, Nano-Newton electrostatic force actuators for femto-Newton-sensitive measurements: System performance test in the LISA Pathfinder mission, Phys. Rev. D **109**, 102009 (2024).

[7] L. Sala, Noisefinder: Spectral estimation and noise decorrelation, 10.5281/zenodo.22097647 (2026)

[8] M. Tröbs and G. Heinzel, Improved spectrum estimation from digitized time series on a logarithmic frequency axis, Measurement **39**, 120 (2006).

[9] S. Vitale, L. Sala, and D. Vetrugno, An optimised variant of the periodogram method for noise power spectral density estimation in LISA hardware testing, LISA-UTN-INST-TN-0035, University of Trento, 2025.

[10] N. R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (An Introduction). Ann. Math. Stat. **34**, 152 (1963).

[11] D. K. Nagar and A. K. Gupta, Expectations of functions of complex wishart matrix, Acta Appl. Math. **113**, 265 (2011).

[12] G. Carter, C. Knapp, and A. Nuttall, Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing, IEEE Trans. Audio Electroacoust. **21**, 337 (1973).

[13] D. V. Ouellette, Schur complements and statistics, Linear Algebra Appl. **36**, 187 (1981).

[14] C. R. Rao, editor, *Linear Statistical Inference and its Applications*, Wiley Series in Probability and Statistics (John Wiley & Sons, Inc., Hoboken, NJ, USA, 1973).

[15] H. Jeffreys, An invariant form for the prior probability in estimation problems, Proc. R. Soc. A **186**, 453 (1946).

[16] P. Shaman, The inverted complex Wishart distribution and its application to spectral estimation, J. Multivariate Anal. **10**, 51 (1980).

[17] A. M. Mathai, S. B. Provost, and H. J. Haubold, *Multivariate Statistical Analysis in the Real and Complex Domains* (Springer Nature, London, 2022).

[18] L. Svensson and M. Lundberg, On posterior distributions for signals in gaussian noise with unknown covariance matrix, IEEE Trans. Signal Process. **53**, 3554 (2005).

[19] E. Coornish, The multivariate t-distribution associated with a set of normal sample deviates, Aust. J. Phys. **7**, 532 (1954).

[20] T. Ratnarajah and R. Vaillancourt, Complex singular Wishart matrices and applications, Comput. Math. Appl. **50**, 399 (2005).

[21] Is the rank of the sum of two positive semidefinite matrices larger than their individual ranks? Mathematics Stack Exchange (2017), https://math.stackexchange.com/q/2153772,

[22] A. H. Nuttall, Some windows with very good sidelobe behavior, IEEE Trans. Acoust. Speech Signal Process. **29**, 84 (1981).