Quantitative analysis of LISA pathfinder test-mass noise

Luigi Ferraioli,[*] Giuseppe Congedo, Mauro Hueller, and Stefano Vitale

University of Trento and INFN, via Sommarive 14, 38123 Povo (Trento), Italy

Martin Hewitson

Albert-Einstein-Institut, Max-Planck-Institut fuer Gravitationsphysik und Universität Hannover,
Callinstrasse 38, 30167 Hannover, Germany

Miquel Nofrarias

Institut de Ciències de l'Espai, (CSIC-IEEC), Facultat de Ciències, Campus UAB, Torre C-5,
08193 Bellaterra, Spain

Michele Armano

SRE-OD ESAC, European Space Agency, Camino bajo del Castillo s/n, Urbanización Villafranca del Castillo,
Villanueva de la Cañada, 28692 Madrid, Spain

LISA Pathfinder (LPF) is a mission aiming to test the critical technology for the forthcoming space-based gravitational-wave detectors. The main scientific objective of the LPF mission is to demonstrate test masses free falling with residual accelerations below 3×10^{-14} m s$^{-2}/\sqrt{\text{Hz}}$ at 1 mHz. Reaching such an ambitious target will require a significant amount of system optimization and characterization, which will in turn require accurate and quantitative noise analysis procedures. In this paper, we discuss two main problems associated with the analysis of the data from LPF: i) excess noise detection and ii) noise parameter identification. The mission is focused on the low-frequency region ([0.1, 10] mHz) of the available signal spectrum. In such a region, the signal is dominated by the force noise acting on test masses. At the same time, the mission duration is limited to 90 days and typical data segments will be 24 hours in length. Considering those constraints, noise analysis is expected to deal with a limited amount of non-Gaussian data, since the spectrum statistics will be far from Gaussian and the lowest available frequency is limited by the data length. In this paper, we analyze the details of the expected statistics for spectral data and develop two suitable excess noise estimators. One is based on the statistical properties of the integrated spectrum, the other is based on the Kolmogorov-Smirnov test. The sensitivity of the estimators is discussed theoretically for independent data, then the algorithms are tested on LPF synthetic data. The test on realistic LPF data allows the effect of spectral data correlations on the efficiency of the different noise excess estimators to be highlighted. It also reveals the versatility of the Kolmogorov-Smirnov approach, which can be adapted to provide reasonable results on correlated data from a modified version of the standard equations for the inversion of the test statistic. Closely related to excess noise detection, the problem of noise parameter identification in non-Gaussian data is approached in two ways. One procedure is based on the maximum-likelihood estimator and another is based on the Kolmogorov-Smirnov goodness-of-fit estimator. Both approaches provide unbiased and accurate results for noise parameter estimation and demonstrate superior performance with respect to standard weighted least squares and Huber's norm. We also discuss the advantages of the Kolmogorov-Smirnov formalism for the estimation of confidence intervals of parameter values in correlated data.

I. INTRODUCTION

LISA Pathfinder (LPF), a European Space Agency mission, will be used to characterize and analyze all possible sources of disturbance which perturb free-falling test masses from their geodesic motion. The system is composed of a single spacecraft (SC) enclosing a scientific payload, the LISA Technology Package (LTP), which is composed of two test masses (TMs) whose position is sensed by an interferometer. The spacecraft cannot simultaneously follow both masses, therefore the trajectory of only one test mass is used as a drag-free reference along the measurement axis. In order to prevent the trajectories of the test masses from diverging, the second test mass is capacitively actuated to follow the first (free-falling) TM. In the main science operating mode, the first interferometer channel measures the displacement of the SC relative to the

_____
[*]Present address: APC, UMR 7164, Université Paris 7 Denis Diderot, 10 rue Alice Domon et Léonie Duquet, 75025 Paris Cedex 13, France.
luigi@science.unitn.it

free-falling TM. The second interferometer channel (the differential channel) measures the relative displacement between the two TMs.

LPF is a controlled system which can only be fully assessed during flight operation, therefore a considerable number of experiments will be devoted to the identification of the details of the dynamics of the system. A dynamical model of LPF is built in advance on the basis of physical considerations and from the results of test campaigns. The dynamical model is parametric so that it can be updated on the basis of the experiments that will be conducted during mission operations. The overall aim of the process is to reach the best free-fall quality (below 3×10^{-14} m s$^{-2}/\sqrt{\text{Hz}}$ at frequencies around 1 mHz) in a step-by-step procedure in which the result of the previous experiment is used to adjust the detailed configuration of the following experiments [1–4].

Such a demanding program requires daily analysis of the instrument signals constrained by two major factors: i) the amount of available data is tightly constrained by LTP mission duration (90 days), the telemetry bandwidth, and the length of each data segment (typically 24 h); ii) the scientific interest is mainly focused on the analysis of noise sources which act directly on the TMs since that should provide a baseline reference for the forthcoming space-based gravitational-waves observatories [5–8]. The direct forces on the TMs are expected to dominate the instrument output in the frequency range [0.1, 10] mHz. Sample power spectra are typically calculated with Welch's averaging periodogram method [9]. In order to keep enough frequency resolution at low frequencies, the sample power spectra cannot be averaged more than a few times (we average 4 times in the present paper); this results in highly non-Gaussian data for which we are developing dedicated techniques. In particular, this paper aims to propose a solution for two major data analysis challenges encountered in LPF: i) Different measurements of the same physical quantity can exhibit different noise content if they are performed under slightly different environmental conditions. The objective of LISA Pathfinder data analysis during operations will be to discover such differences, understand their origin, and adjust spacecraft physical parameters accordingly. Such a problem requires reliable excess noise detection procedures which have to be based on solid statistical considerations. ii) Along with the demonstration of unprecedented test-mass free fall, LPF will provide a model for the expected test-mass force noise for future space-based gravitational-wave detectors. In order to do this, we need to be able to match an analytical model to a noisy power spectral density measurement. The quality of the match must be statistically quantified. Both data analysis problems deal with sample spectra and the corresponding statistical properties. Section II reports on the properties of different experimental procedures for the detection of noise excess. In particular, we considered

two cases where the noise excess is evaluated with respect to reference data or a reference model. The accuracy of the methods is theoretically analyzed for the case of broadband and band-limited excess noise. In Sec. III, the problem of noise parameter estimation for non-Gaussian data is explored and an algorithm based on maximum likelihood is derived. In parallel, the application of the Kolmogorov-Smirnov formalism to the construction of a goodness-of-fit estimator is discussed. Section IV reports briefly on the extension of the analysis procedures to the case of nonstationary noise and time-frequency investigations. In Sec. V, we provide an application of the developed algorithms to synthetic data with LPF-like qualities. This allows us to shed light on the effects of data correlations on the accuracy of the developed excess noise estimators. The analysis procedures are available as MATLAB tools in the framework of the LTPDA Toolbox [10–12], an object-oriented MATLAB Toolbox for advanced data analysis.

II. QUANTITATIVE DETECTION OF NOISE POWER VARIATIONS

The problem of the detection of noise power variation in consecutive measurements can be formulated in two different ways: i) two different measurements of the power are compared; ii) the different measurements of the power are compared with a reference model. The problem in the first case is of general character and can be applied to a wide range of experiments; the second case, on the other hand, assumes that a reference model for the noise power is known and data must be compared against the given model in order to understand if the system is performing under known conditions. The latter case is likely to be the scenario for LPF operations. The quantity typically used for the detection of noise power variations is the total energy content of the data series. It is defined as $\mathcal{E} = T\sum_i |x_i|^2$. Unfortunately, $\mathcal{E}$ provides a poor estimator for two reasons: i) as soon as the data series $x_0, \ldots, x_{N-1}$ departs from zero-mean Gaussian white noise, the statistic of $\mathcal{E}$ becomes ill defined and the definition of a confidence interval becomes cumbersome; ii) $\mathcal{E}$ provides global information as it is not sensitive to noise changes in a given frequency band. While the first problem could be overcome (without little difficulty) by a numerical identification of the expected statistic, the second problem suggests that a spectral-based estimator would provide supplementary information which could be fundamental in discriminating different noise sources.

A. Detection of noise variations with a model

In the case that a reference model is available, the detection of excess noise in the spectral domain can be effectively implemented with a test on the normalized Welch's overlapped segment averaging (WOSA) spectrum $R_{\text{WOSA}}(f_k)$ defined in Eq. (A9). In particular, the test can pursue two different philosophies of which one aims to test a global scalar indicator of the properties of the data and

another aims to test the details of the statistical distribution of the data.

A sensible estimator for the first approach is provided by the integral of the normalized spectrum (IR) which, in the discrete case, can be written

$$\text{IR} = \sum_{k=1}^{N_f} R_{\text{WOSA}}(f_k). \tag{1}$$

In the simplifying assumption of independent spectral data, the statistic of each element of $R_{\text{WOSA}}(f_k)$ is described by a gamma distribution as defined by Eq. (A8) with $\delta = 1/N_s$ and $h = N_s$. The sum over the different values at the frequencies f_k is still a gamma distribution with $\delta = 1/N_s$ and $h = N_s N_f$. The expectation value for IR is easily obtained as $E[\text{IR}] = \delta h = N_f$.

The natural estimator for the second approach is provided by the empirical cumulative distribution function (ECDF). The ECDF for a data series can be defined as $F_N(x) = z/N$, where z is the number of observations reporting a value less than or equal to x. x denotes the values taken by the data, in our case $R_{\text{WOSA}}(f_k)$ with $k = 1, \ldots, N_f$. The ECDF can be tested against the theoretical expectation provided by Eq. (A8). If the model well represents the given sample spectrum, then $R_{\text{WOSA}}(f_k)$ is distributed according to the expected gamma distribution. Alternatively, if the sample spectrum contains excess noise with respect to the reference model, then the distribution of the normalized WOSA spectrum will be different from the one reported in Eq. (A8). The hypothesis that the two distributions are equal can be tested if a "distance" between the ECDF and the theoretical reference, F_Γ, is defined as

$$d_K(x) = |F_N(x) - F_\Gamma(x)|, \tag{2}$$

with $d_K(x)$ assuming values on the interval [0, 1] and $K = N_f$. Kolmogorov found that the statistical properties of

$$d_K = \max[d_K(x)] \tag{3}$$

are independent from the specific distributions under test [13,14]. This property qualifies d_K as an excellent candidate for the construction of a general test for cumulative distribution functions; the limiting statistic for d_K was identified by Kolmogorov himself and then inverted by Smirnov [15,16], who provided an analytical expression for the calculation of d_K as a function of the significance level. General details about the Kolmogorov-Smirnov test (KS test) are reported in Appendix B.

It is interesting to calculate the expected sensitivity for the two estimators. IR is expected to be distributed as a gamma distribution (at least when the model and the data are in agreement); the corresponding cumulative distribution function (CDF) is provided by the incomplete gamma function $\mathcal{P}(h, x/\delta)$ [17]. The CDF assumes values in the interval [0, 1] as it defines the probability associated with

the observation x. The inverse of the CDF provides the critical values, x_ρ, associated with a given probability, $\rho \in [0, 1]$. The confidence range for a probability, ρ, for a gamma-distributed variable can then be defined by the boundaries $x_{\text{lw}} = \mathcal{P}^{-1}(\alpha/2)$ and $x_{\text{up}} = \mathcal{P}^{-1}(1 - \alpha/2)$, with $\alpha = 1 - \rho$. We assert that the measured sample spectrum is compatible with the reference model if $x_{\text{lw}} \leq \text{IR} \leq x_{\text{up}}$ for the given probability, ρ, or significance level, α.

If the noise excess is provided by a scale factor, γ, which affects the noise on the complete band of frequencies, the expected value for IR changes to $E(\text{IR}) = \gamma N_f$. Therefore, the detection threshold for γ is fixed by the interval $x_{\text{lw}}/N_f \leq \gamma \leq x_{\text{up}}/N_f$. In other words, the IR can detect a noise difference with respect to the reference model only if $\gamma < x_{\text{lw}}/N_f$ or if $\gamma > x_{\text{up}}/N_f$. If γ is nonzero only in a restricted band of frequencies $[f_a, f_b]$, then the expectation value for IR changes to $E(\text{IR}) = N_f - N_{ab} + N_{ab}\gamma$. In this case, IR can detect the noise difference only if $\gamma < (x_{\text{lw}} - N_f + N_{ab})/N_{ab}$ or $\gamma > (x_{\text{up}} - N_f + N_{ab})/N_{ab}$. N_{ab} is the number of frequency points in the interval $[f_a, f_b]$.

In the case of the Kolmogorov-Smirnov (KS) estimator, the ECDF of the WOSA normalized spectrum is compared with the theoretical expectation $\mathcal{P}(N_s, xN_s)$. In the case that the noise excess is simply a constant scale factor, γ, over the whole frequency band, the expected CDF for the normalized WOSA spectrum is $\mathcal{P}(N_s, xN_s/\gamma)$. The expected value for the KS-test variable is then written as

$$d_K(\gamma) = \max \left| \mathcal{P}\left(N_s, \frac{xN_s}{\gamma}\right) - \mathcal{P}(N_s, xN_s) \right|. \tag{4}$$

Once a significance value α is provided, the corresponding critical value, $d_K(\alpha)$, can be calculated from the equations for the inversion of the limiting CDF for d_K [15,16]. The two distributions in Eq. (4) are incompatible at the given significance level if $d_K(\gamma) > d_K(\alpha)$. This defines the detection threshold for γ. If γ is nonzero only in a restricted band of frequencies $[f_a, f_b]$, then the expected distribution for the normalized WOSA estimator is difficult to calculate. Nevertheless, the detection threshold for the KS estimator can be calculated numerically from synthetic data.

In Fig. 1, the ranges of nondetectability of the KS and IR estimators are reported as a function of N_f for three different values of the WOSA averages N_s. Data refer to the case that the scale factor, γ, extends over the complete frequency band. As can be clearly seen, the IR estimator always has a better sensitivity than the KS estimator. The sensitivity for the case of a band-limited excess noise is reported in Fig. 2 as a function of the ratio N_{ab}/N_f. As in the previous case, the IR estimator provides a better sensitivity with respect to the KS estimator. The difference is particularly relevant for $N_{ab}/N_f < 0.07$; below such values, the sensitivity of the KS estimator becomes poor. Both in Fig. 1 and in Fig. 2, the confidence level for γ detection is fixed at 95%.
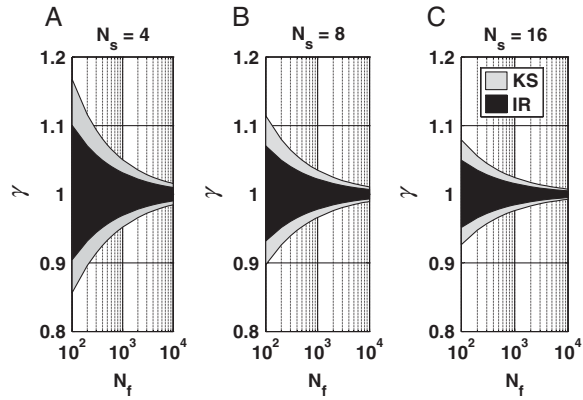
FIG. 1. Nondetection ranges for IR and KS estimators. Noise excess is assumed as a constant multiplicative factor γ extending along the whole band of frequencies. Darkened intervals define the thresholds for γ detection; if γ is larger or smaller than the shaded values, it can be detected with a confidence of 95%. N_s refers to WOSA averages. N_f is the number of frequency points in the spectrum.
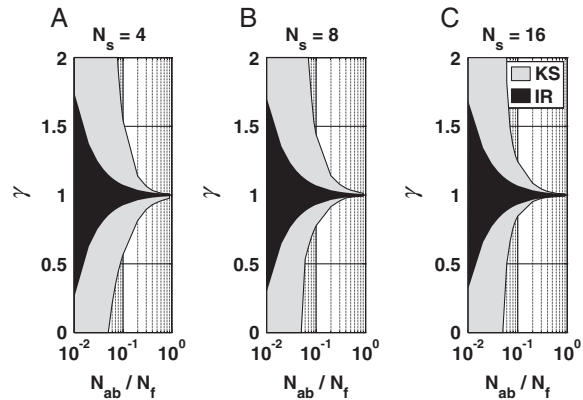

FIG. 2. Nondetection ranges for IR and KS estimators in the case that noise excess coefficient γ is different from zero in a restricted band $[f_a, f_b]$. Data are presented as a function of the ratio N_{ab}/N_f where N_{ab} is the number of frequency points in the interval $[f_a, f_b]$ and N_f is the total number of frequency points in the spectrum. N_s refers to WOSA averages. Confidence level for excess detection is fixed to 95%.

It is worth noting that the KS algorithm can be used on time series to quantitatively assess departures from a given distribution (e.g., Gaussian). Once the ECDF for the data is calculated, it can be compared with the expected distribution by calculating $d_K(x)$ from Eq. (2). Since the distribution of $d_K(x)$ coefficients is known, it is straightforward to set a confidence threshold. The procedure can be applied even in presence of correlations thanks to the generalizations discussed in Sec. V B and in Appendix B.

B. Detection of noise variations without a model

In the case that an excess of noise has to be detected by comparing different measurements, the Parseval's theorem

[9] $[\mathcal{E} = \sum_{k=0}^{N/2} P(f_k)]$ suggests that the sum of the elements of a sample spectrum in a given frequency band $\mathcal{E}_{ab} = \sum_{k=a}^{b} P(f_k)$ could provide a sensitive estimator for noise power variations. Such an estimator would be loosely equivalent to the IR discussed above, except that its statistic is hard to determine in a typical experimental situation. The statistic of $P(f_k)$ at each frequency, f_k, is defined by Eq. (A4). Therefore, in the case of non-white-noise, its parameters depend on $S(f_k)$ and the statistic is different at different frequencies. Thus, the statistic of $\mathcal{E}_{ab}$ is not easily known.

An interesting alternative to $\mathcal{E}_{ab}$ is provided by the KS estimator. The hypothesis that two data series $\{x_{n_1}\}$ and $\{x_{n_2}\}$ (of length N_1 and N_2, respectively) have the same limiting cumulative distribution function $F(x)$ can be tested if a distance in the ECDFs space is defined as

$$d_K(x) = |F_{n_1}(x) - F_{n_2}(x)|, \quad (5)$$

with $d_K(x)$ defined on the interval [0, 1] and $K = (N_1 N_2)/(N_1 + N_2)$ [14]. Also in this case, the statistical properties of

$$d_K = \max[d_K(x)] \quad (6)$$

are independent from the distributions of x_{n_1} and x_{n_2}. The same equations used in the case of the comparison with a given model can now be used for the inversion of the cumulative statistic of d_K [18]. Considering the simplifying assumption of independent spectral data, the sensitivity of the KS estimator at a given significance level can be calculated in analogy to what was discussed in the previous paragraph.

In Fig. 3, the calculated interval for nondetection is reported for the case that the excess, γ, is extending over the whole frequency band and for the case that $\gamma \neq 0$ in $[f_a, f_b]$.
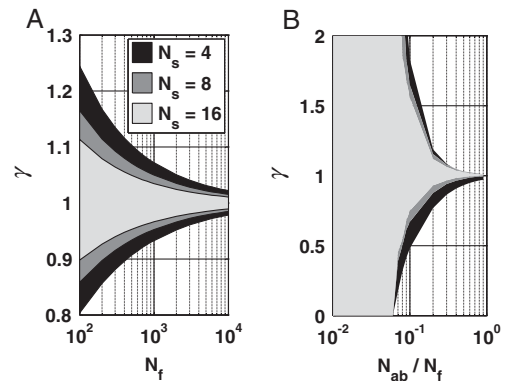

FIG. 3. Nondetection ranges for the KS estimator for the comparison of two sample spectra. (A) Noise excess coefficient $\gamma \neq 0$ on the full frequency band. N_s is the number of WOSA averages. N_f is the number of frequency points in the sample spectrum. (B) $\gamma \neq 0$ only in a restricted interval of frequencies $[f_a, f_b]$. N_{ab} is the number of points in the interval.

III. NOISE MODEL IDENTIFICATION

Closely related to the problem of excess noise detection, the problem of noise model identification is one of the principal scientific objectives of the LPF mission. Of particular interest is the identification of a model for the force noise acting on the TMs, which can be used as a baseline for the forthcoming space-based gravitational-wave observatories.

The main constraints on the identification of force noise on the TMs in LPF are fixed by i) the limited data series length, typically 24 h; ii) TMs force noise is dominating the output signal below 10 mHz; iii) force noise data that are not directly accessible since the system measures and reports TM displacement, requiring that force noise on TMs be reconstructed by a numerical procedure [19,20].

The result of the combination of the first two constraints is that the number of segment averages in the WOSA procedure for sample spectrum estimation should be taken as low as possible so as to have a reasonable number of frequency points in the range $f \in [0.1, 10]$ mHz. As a consequence, the distribution of the WOSA spectrum strongly departs from a Gaussian distribution [21], meaning that the classical least-squares minimization procedure for parameter estimation is not well conditioned and a full maximum-likelihood procedure is required.

A. Maximum-likelihood parameter estimation

If we replace $S(f_k)$ in Eq. (A9) with a parametric model for the spectrum, $S(f_k; \theta_1, \ldots, \theta_H)$, the normalized WOSA spectrum becomes parametric, $R_{\text{WOSA}}(f_k; \theta_1, \ldots, \theta_H)$, and can be used for the estimation of noise model parameters $\{\theta_1, \ldots, \theta_H\}$. In this section, we develop the likelihood formalism for the simple case that the noise model is a function of a single parameter, $S(f_k; \theta) = \theta S_0(f_k)$. This allows us not only to find a sensible goodness-of-fit estimator for $R_{\text{WOSA}}(f_k; \theta)$ that can be used also in the case of multiple parameters, but also to place the excess noise estimator, IR, in a more solid theoretical framework.

Indicating with θ_{TRUE} the "true" value for the θ parameter, $R_{\text{WOSA}}(f_k; \theta)$ can be rewritten as

$$R_{\text{WOSA}}(f_k; \phi) = \phi \frac{P_{\text{WOSA}}(f_k)}{\theta_{\text{TRUE}} S_0(f_k)} = \phi R_{\text{WOSA}}^{\text{true}}(f_k). \quad (7)$$

Here, $\phi = \theta_{\text{true}}/\theta$. Assuming that $\theta_{\text{TRUE}} S_0(f_k)$ correctly reproduces the expected value for the spectrum, the distribution of $R_{\text{WOSA}}^{\text{true}}(f_k)$ is reported in Eq. (A8) with $\delta = 1/N_s$ and $h = N_s$. The distribution of the samples $R_{\text{WOSA}}(f_k; \phi)$ is then

$$f(y; h, \delta, \phi) = \frac{e^{-((y)/(\phi\delta))} y^{h-1}}{(\phi\delta)^h \Gamma(h)}. \quad (8)$$

Under the simplifying assumption that the values of $R_{\text{WOSA}}(f_k; \phi)$ are independent for different f_k, the likelihood function can be written as

$$\mathcal{L}(h, \delta, \phi) = \prod_k f(y_k; h, \delta, \phi) \Delta y. \quad (9)$$

Here, Δy is a constant term required to have a finite probability from the probability distribution function $f(y_k; h, \delta, \phi)$. y_k are observed samples corresponding to $R_{\text{WOSA}}(f_k; \phi)$. It is typically more convenient to work with the natural logarithm of the likelihood function $l(h, \delta, \phi) = \ln \mathcal{L}(h, \delta, \phi)$.

$$l(h, \delta, \phi) \sim (h-1) \sum_k \ln y_k - \frac{1}{\phi\delta} \sum_k y_k - N_f h \ln\phi. \quad (10)$$

N_f is the total number of frequency samples. Taking the first derivative with respect to ϕ and equating to 0 we find the maximum-likelihood estimator for the parameter ϕ:

$$\Lambda = \frac{1}{\phi} \sum_k R_{\text{WOSA}}(f_k; \phi) - N_f. \quad (11)$$

The value of ϕ at which $\Lambda = 0$ corresponds to the maximum-likelihood estimation for the parameter. The Λ estimator is unbiased, since, remembering that $\phi = \theta_{\text{true}}/\theta$, it can be verified that $\lim_{\theta \to \theta_{\text{true}}} E[\Lambda] = 0$. For the practical purpose, we can find the zero crossing of the reduced estimator

$$\tilde{\Lambda} = \sum_k R_{\text{WOSA}}(f_k; \theta) - N_f, \quad (12)$$

which is crossing zero at the same θ value of Λ, since $\phi \to 1$ when $\theta \to \theta_{\text{true}}$. It is worth noting that $\tilde{\Lambda}$ is practically the IR excess noise estimator with the expectation value N_f subtracted. It is then worth noting that $\tilde{\Lambda}$ can be used not only in the simple case of one parameter but it can also be applied in the general case of a model $S(f_k; \theta_1, \ldots, \theta_H)$ since the condition $\tilde{\Lambda} \to 0$ when $\{\theta_1, \ldots, \theta_H\} \to \{\theta_{1\text{true}}, \ldots, \theta_{H\text{true}}\}$ is always satisfied.

B. Kolmogorov-Smirnov parameter estimation

As discussed in Sec. II A, the KS estimator can be used as an alternative to a maximum-likelihood procedure for parameter estimation.

Thanks to the statistical properties of $R_{\text{WOSA}}(f_k; \theta)$, the closer θ is to θ_{true}, the better the distribution of $R_{\text{WOSA}}(f_k; \theta)$ is described by Eq. (A8) with $\delta = 1/N_s$ and $h = N_s$. Therefore, the Kolmogorov-Smirnov distance parameter provides an effective goodness-of-fit estimator:

$$d_K(\theta) = \max|F_R(x; \theta) - \mathcal{P}(N_s, xN_s)|. \quad (13)$$

$F_R(\alpha)$ is the ECDF for the current $R_{\text{WOSA}}(f_k; \theta)$ estimate; $\mathcal{P}(N_s, xN_s)$ is the limiting distribution for $\theta \to \theta_{\text{true}}$.

KS estimation for θ is obtained by the minimization of $d_K(\theta)$ with respect to θ. A confidence range for the parameter estimation can be readily defined from the non-rejection region of the KS test at a given significance level. In practice, having defined a significance level, the corresponding critical values of the KS statistic, $d_K(\alpha)$, can be calculated with standard equations [16] or Monte Carlo simulations in the case of correlated data. The values of $\bar{\theta}$ for which $d_K(\bar{\theta}) = d_K(\alpha)$ provide the boundary for the confidence range at the given significance. The KS statistic can also be successfully applied to multiparameter identification, since the convergence of $F_R(\alpha)$ to $\mathcal{P}(N_s, xN_s)$ is always verified when $\{\theta_1, \ldots, \theta_H\} \rightarrow \{\theta_{1\text{true}}, \ldots, \theta_{H\text{true}}\}$.

IV. ANALYSIS OF NONSTATIONARY NOISE

The implementation of noise analysis procedures was so far discussed in the context of stationary or *pseudostationary* noise [22]. In the case of truly nonstationary noise, the spectral content of a time series is investigated by time-frequency analysis techniques which include the spectrogram and the wavelet transform. The spectrogram is estimated by the square modulus of the short-time Fourier transform of the data [23]. It provides a direct extension of the sample spectrum formalism to nonstationary time series. Given a data series of N samples, a fraction of length $Q < N$ is windowed and the Fourier transform is applied. Then, the data window is time shifted and the process is repeated. The calculation of the spectrogram is based on data windowing and the application of the Fourier transform, therefore the considerations noted in the previous sections for stationary noise can be applied directly to the spectrogram analysis of nonstationary noise.

Since the short-time Fourier transform has the same resolution across the time-frequency plane, it is often preferable to resort to the wavelet transform. Wavelet transform is a decomposition of the time series over time-frequency elements that are obtained by scaling and translating a mother function $\psi \in \mathbf{L}^2(\mathbb{R})$:

$$\psi_{u,s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t - u}{s}\right). \tag{14}$$

The function $\psi(t)$ has zero average and the wavelet elements $\psi_{u,s}$ are normalized to 1. The wavelet transform of a function $f(t)$ is then defined as

$$Wf(u, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t - u}{s}\right) dt. \tag{15}$$

In the discrete case, the result of the wavelet transform on a time series is an array of coefficients $w_{u,s}$ where u is the time index and s is the scale index which is associated to a given frequency band [24]. In the case of uncorrelated Gaussian noise, the distribution of the coefficients $w_{u,s}$ is still Gaussian with a certain amount of correlation introduced by the convolutionlike transform operation [24]. In such a favorable situation, the extension of the method

described in the stationary case appears straightforward. In particular, Kolmogorov-Smirnov procedures are excellent candidates since their robustness to correlation and the possibility of extending KS distance definition to a two-dimensional space [25,26].

V. APPLICATION TO LPF SYNTHETIC DATA

In this section, the different procedures for excess noise detection and parameter estimation are applied to synthetic LPF data. This provides not only an interesting framework for testing their accuracy and precision, but it also helps to clarify the role of correlations among spectral data.

A. Synthetic data and noise projection

LPF provides two output channels along the principal measurement axis which are sensing the displacement of the SC relative to the free-falling TM and the relative displacement between the TMs.

From the knowledge of the displacement signals and a linear model for the system dynamics, an effective force per unit mass, a_{eff}, acting on the TMs can be extracted. a_{eff} is the combination of the true force per unit mass acting on TMs and a projected interferometer readout noise. Details of the calculation are reported in Appendix D.

Following the same scheme, a given prediction for the different input noise sources can be projected to a prediction for the power spectrum of a_{eff}, as reported in Fig. 4. Here, the model for the spectrum of force noise acting on TMs and the model for the spectrum of readout noise are projected into a model for the power spectrum of a_{eff} (TMs + Readout in the figure notation), which represents this paper's baseline for the spectrum of a_{eff}.

In Fig. 4, we also report the project specifications for LPF and the expected noise spectrum for a_{eff} in a worst-case scenario. In our baseline, we assumed a reduced force
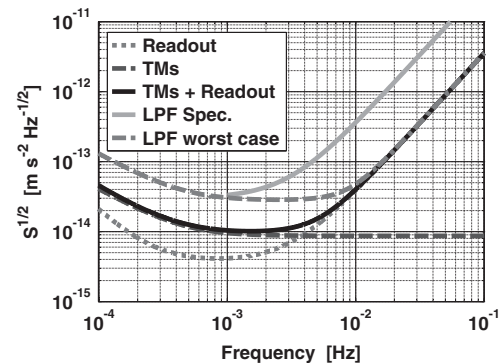

FIG. 4. Projection of the spectrum of a_{eff}. "Readout" is the projection of the readout noise to a_{eff}. "TMs" is the projection of the force noise on the test masses to a_{eff}. "TMs + Readout" is the complete noise projection for a_{eff}; it represents the baseline noise level assumed in the present paper. "LPF worst case" refers to a worst-case scenario for a_{eff} and "LPF Spec." corresponds to the mission specifications.

noise on the TMs compared to the worst case but choose to keep the worst case for the readout noise. This was done in order to represent one of the possible scenarios (not the best one) that can be experienced during the mission.

The model assumed for the force noise on the TMs is characterized by a low-frequency $1/f^2$ behavior and a flat part for $f > 1$ mHz. The model can be written as $S_{TM}(f) = \theta S_{TM}^0(f)$. θ is an adjustable parameter which assumes values $\theta = 1$ for the worst-case scenario and $\theta = 0.1$ for our baseline model. It is worth noting that $S_{TM}(f)$ is projected (together with the readout noise model) through LPF dynamics in order to obtain the expected noise spectrum for a_{eff} which we indicate with $S_a(f)$. $S_a(f)$ with $\theta = 0.1$ corresponds to "TMs + Readout" in Fig. 4; $S_a(f)$ with $\theta = 1$ corresponds, instead, to the "LPF worst case."

B. Excess noise detection

A change of the noise level on $S_{TM}(f)$ ($\theta \neq 0.1$) produces a variation of the energy content of a_{eff}. Such variation, which may be "improperly" identified as excess noise, can be detected with the procedures defined in Sec. II. In particular, we tested the detection of excess noise between two data series and between a data series and a reference model. Synthetic data were produced according to the following procedure:

(1) Different models for $S_a(f)$ are produced changing θ around the reference value $\theta = 0.1$. Readout noise level is kept fixed.
(2) Corresponding noise time series for a_{eff} are generated using the procedure reported in [27]. The time series are 24 h long and have a sampling frequency of 1 Hz.

(3) Sample spectra are calculated for each series with the WOSA algorithm. We chose a Blackman-Harris data window, 50% segment overlap, and number of segments averages $N_s = 4$.
(4) The analysis is restricted to the frequency interval [0.1, 10] mHz since, as can be seen from Fig. 4, it represents the region in which the force noise on the TMs dominates $S_a(f)$.

Spectral data are tested for excess noise. We used the KS algorithm [Eq. (6)] in the case of the test of two data series. The data series for $\theta \neq 0.1$ are compared against the reference series with $\theta = 0.1$. In the case of the test of a data series against a model, both the KS algorithm [Eq. (3)] and the IR algorithm [Eq. (1)] are used. The reference model is the projected $S_a(f)$ for $\theta = 0.1$. The results of the tests are summarized in Table I. KS critical values $d_K(\alpha)$ and IR confidence intervals are calculated for a significance level $\alpha = 0.05$ which corresponds to a 95% confidence.

Each value of θ corresponds to a value of the in-band energy content $E(\theta)$ of $S_a(f)$ in the analyzed frequency band. We report in Table I the relative change in energy $\Delta E/E$ corresponding to a relative change in θ. The quantity $\Delta E/E$ plays the same role of the parameter γ in Figs. 1 and 3, even though the two quantities are not completely comparable since γ in Fig. 1 assumes independence of the data. Spectral data are correlated among different frequency values because of two effects [9,28]: i) data windowing which corresponds to a convolution in the frequency domain of the window function with the sample spectrum; ii) WOSA overlapped segment averaging. The first effect is unavoidable; the second effect,

TABLE I. Detection of noise differences in the frequency band [0.1, 10] mHz. The symbol ✓ indicates compatibility between tested objects. The symbol × is instead used for indicating test rejection. θ is an adjustable parameter which assumes value $\theta = 0.1$ for our baseline model. Different values of θ correspond to different values of the in-band energy content $E(\theta)$ of $S_a(f)$. We reported here relative values with respect to reference. Details on $d_K^{eff}(\alpha)$, $d_K^{MC}(\alpha)$, and MC confidence interval for IR estimator can be found in Appendices B and C, respectively.

| | | KS vs data | | | KS vs model | | | | IR | | |
| | | | $d_K(\alpha)$ | $d_K^{MC}(\alpha)$ | | $d_K(\alpha)$ | $d_K^{eff}(\alpha)$ | $d_K^{MC}(\alpha)$ | | Confidence Int. | MC Confidence Int. |
$\frac{\Delta E}{E}$	$\frac{\Delta\theta}{\theta}$	d_K	0.1030	0.1006	d_K	0.0730	0.0982	0.0969	$\sum_k R_{WOSA}(f_k)$	[323.14, 359.33]	[311.33, 372.55]
−0.14	−0.7	0.2352	×	×	0.2547	×	×	×	250.74	×	×
−0.08	−0.4	0.1424	×	×	0.1103	×	×	×	303.72	×	×
−0.06	−0.3	0.1806	×	×	0.1747	×	×	×	274.54	×	×
−0.05	−0.25	0.1740	×	×	0.1685	×	×	×	294.42	×	×
−0.04	−0.2	0.0887	✓	✓	0.0696	✓	✓	✓	317.10	×	✓
−0.03	−0.15	0.0906	✓	✓	0.0433	✓	✓	✓	324.87	✓	✓
−0.01	−0.05	0.0771	✓	✓	0.0302	✓	✓	✓	339.47	✓	✓
0.01	0.05	0.0551	✓	✓	0.0539	✓	✓	✓	352.26	✓	✓
0.03	0.15	0.0629	✓	✓	0.0370	✓	✓	✓	345.38	✓	✓
0.04	0.2	0.0603	✓	✓	0.0635	✓	✓	✓	357.14	✓	✓
0.05	0.25	0.0519	✓	✓	0.0818	×	✓	✓	369.05	×	✓
0.06	0.3	0.0857	✓	✓	0.1253	×	×	×	388.01	×	×
0.08	0.4	0.1103	×	×	0.1137	×	×	×	402.14	×	×
0.14	0.7	0.1520	×	×	0.2423	×	×	×	459.64	×	×

instead, can be attenuated by a proper choice of the segment overlap. It can be demonstrated [29] that for a Blackmann-Harris window the effect is negligible with 50% overlap.

Since the standard statistics for the estimators [$d_K(\alpha)$ for KS and confidence interval for IR] are calculated in Sec. II with the assumption of data independence, the standard critical values, $d_K(\alpha)$, for the KS estimators and the confidence intervals for the IR estimator can be applied only if the correlations among data are negligible. If this is not the case, the effective statistic of the estimators can be numerically calculated with Monte Carlo simulations. The corresponding results of a Monte Carlo (MC) simulation with $N_{MC} = 5000$ realizations of the reference data series are indicated in Table I with the suffix MC. The symbol ✓ is used to indicate that the spectral data for the corresponding $\Delta E/E$ is compatible with the reference. On the contrary, the symbol × indicates a rejection.

Observing the test results reported in columns 7 and 9 and in columns 11 and 12, it is readily seen that correlations among data play a role. In the case of the KS test, the comparison of the data with $d_K(\alpha)$ determines a rejection for $\Delta E/E = 0.05$. On the other hand, the same value is accepted when Monte Carlo result $d_K^{MC}(\alpha)$ is used. In the case of the IR estimator, the comparison with the standard confidence interval leads to the rejection in correspondence with $\Delta E/E = 0.05$ and $\Delta E/E = -0.04$. Such values are instead considered compatible by the Monte Carlo confidence interval. These results provide us with clear information that the presence of correlations among data has affected the tests statistics and therefore the standard equations, assuming independence among data, are not usable in this situation.

Looking at the results for the KS test between two data series, we discover that $d_K(\alpha)$ and $d_K^{MC}(\alpha)$ are practically equal. In fact, the results for the two corresponding columns of Table I (columns 4 and 5) are in perfect agreement. This is the practical result of one of the most interesting properties of the KS test. The test is based on Eq. (6). It states that the two empirical cumulative distributions under test have the same limiting CDF. Since the sample spectra in our test are calculated following the same WOSA procedure, the degree of correlation among different frequency points is the same, therefore the test statistic is not spoiled.

Comparing $d_K(\alpha)$ and $d_K^{MC}(\alpha)$ of columns 7 and 9, we see that the effect of correlation is to increase the maximum expected spread between ECDF and limiting CDF. Therefore, the effect of data correlation is to distort the expected statistic for d_K. The "distortion" of d_K statistic can be taken into account if an effective value for the parameter K is introduced. In the case of the comparison of the ECDF for correlated data against a theoretical CDF, the application of the standard values for d_K, where $K = N_f$, N_f being the number of data elements, leads to

a statistically unfair test. We then discovered that test fairness can be recovered if an effective value for K is used rather than the standard $K = N_f$. In particular, for spectral data produced with the WOSA method, Blackmann-Harris window, $N_s = 4$ averages on 50% overlapped segments, we obtained $K_{eff} = \beta N_f$ with $\beta = 0.55$ for a significance level $\alpha = 0.05$. It is worth noting that the value of β is independent from the number of data points considered, and from the spectral shape, provided that the different shapes have reasonably comparable smoothness on a frequency interval comparable with the width of the first lobe of the data window. As an example, the value of $\beta = 0.55$ is valid for LPF-like data and for white-noise equivalently. The requirement on the smoothness of the spectrum is connected to the expression of the correlations introduced by data windowing. It can be demonstrated that if the spectrum can be assumed constant in a frequency range of the order of the width of the first lobe of data window [9], then the correlations are independent from the particular shape of the spectrum and are determined only by the window function. For such class of spectra, we expect the same value for β once the required significance level is fixed.

C. Noise model identification

The problem of noise parameter identification is strictly connected to the problem of excess noise detection by a comparison of a data series with a reference model. While in excess noise detection different data series are compared with a given reference, in parameter estimation different realizations of a parametric model are compared with a data set in order to find the best fit. Because of this, the same algorithms (i.e. KS and IR) can be applied to the solution of the two problems. Precision and accuracy in the estimation of the parameter, θ, controlling the excess force noise on the TMs are tested with a Monte Carlo simulation over $N_{MC} = 5 \times 10^3$ realizations of the same process. The data series reproduce a_{eff} corresponding to the reference value $\theta = 0.1$. Sample spectra are calculated with the procedure described above and the analysis is restricted to the frequency range [0.1, 10] mHz. For each realization, data are compared with a bank of models obtained by the projection of TMs force noise $S_{TM}(f; \theta) = \theta S_{TM}^0(f)$ and readout noise into $S_a(f; \theta)$ for different values of θ around the reference value. The KS and IR estimators are calculated for each element of the model bank, in particular, it is simpler to analyze the results in terms of $\tilde{IR} = |IR - N_f|$. Both KS and $\tilde{IR}$ are expected to have a minimum corresponding to the best estimate for the parameter θ. The two methods proposed here are compared with the performance of a classical weighted least-squares method, which works by minimizing the mean squared error $MSE = \sum_{f_k}((P_{WOSA}(f_k) - S_a(f_k; \theta))/S_a(f_k; \theta))^2$, and a Huber's norm estimator (details are reported in Appendix E). The results of the analysis are reported in
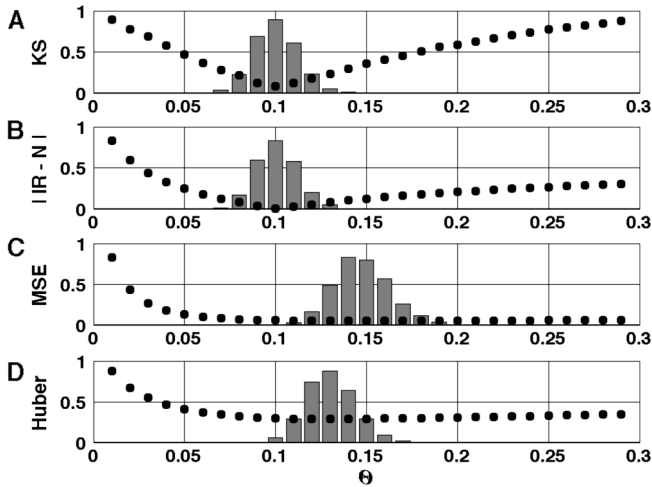
FIG. 5. KS, IR, MSE, and Huber's norm accuracy and precision in the estimation of the force-per-unit-of-mass noise parameter θ. Corresponding histograms report the result of $N_{\mathrm{MC}} = 5000$ MC realizations of the parameter search on a grid of template models. We also show (black dots) the evolution along the model grid of the four estimators for a particular set of data taken from the N_{MC} available.

Fig. 5, where the histograms of the best-fit values over N_{MC} realizations are reported for the four procedures. In the same figure, we also show the evolution along the model grid of the four estimators for a particular set of data from the available N_{MC}.

The distributions for the best-fit parameter are reasonably symmetric for all the estimators; mean values and sample standard deviations are respectively $\theta_{\mathrm{KS}} = 0.100$, $\sigma_{\mathrm{KS}} = 0.012$, $\theta_{\mathrm{IR}} = 0.100$, $\sigma_{\mathrm{IR}} = 0.012$, $\theta_{\mathrm{MSE}} = 0.148$, $\sigma_{\mathrm{MSE}} = 0.016$, $\theta_{\mathrm{Huber}} = 0.130$, and $\sigma_{\mathrm{Huber}} = 0.014$. From the analysis of the Monte Carlo results, it is readily seen that both the KS and the $\tilde{\mathrm{IR}}$ estimators provide equivalently precise and accurate results. On the other hand, the MSE algorithm provides a poor estimation, both from the accuracy and from the precision point of view. The best estimation for the parameter is $\theta_{\mathrm{MSE}} = 0.148$, which is strongly biased with respect to the reference value of $\theta = 0.1$. Also, the distribution of the parameter values is wider than those obtained from the KS and $\tilde{\mathrm{IR}}$ estimators. Huber's norm estimator, with the chosen parameter $c = 0.05$ E, performs better than MSE but the result $\theta_{\mathrm{Huber}} = 0.130$ is still far from the true value.

It is worth discussing the effect of correlations among spectral data on the estimation of the parameter θ. As can be observed from the results of the MC simulation, the accuracy of the estimators is not affected by correlations; the results for KS and $\tilde{\mathrm{IR}}$ estimators are practically indistinguishable. Some problems with $\tilde{\mathrm{IR}}$ can arise from data correlations when the confidence interval for a single estimation is required. As discussed in the previous paragraph, correlations modify the statistic of the IR estimator

making it impossible to easily calculate confidence intervals from the standard equations. The statistic of the KS estimator is affected by correlations too, but we have demonstrated that accurate critical values $d_K(\alpha)$ can be recovered if an effective value for K is used, $K_{\mathrm{eff}} = \beta N_f$. β is a shape parameter depending only on the correlations in the spectral data and on the required significance level. It can be calculated for a specific spectrum (e.g., white noise) and effectively extended to a wide family of spectral shapes. Using the value $\beta = 0.65$ reported in Appendix B, we obtain $d_K^{\mathrm{eff}}(\alpha) \approx 0.064$ for $\alpha = 1 - 0.68\%$. The intersection of such a value with the curve of d_K as a function of θ reported as black dots in Fig. 5(a) provides a 68% confidence interval for the single estimation. In this specific case, such an interval is $\theta \in [0.086, 0.114]$.

VI. CONCLUSIONS

The problem of excess noise detection and noise parameter estimation for non-Gaussian data is analyzed in the framework of the LISA Pathfinder mission. Excess noise detection can be approached in two ways. In one way, the noise content of a data series is compared with a reference data series; in the other way, the noise content of the data is compared with a reference model. In the first case, simple estimators like the total energy content in a data series are not suitable for formulating quantitative statements on a solid statistical basis. As an alternative, a KS estimator is proposed and successfully applied to LPF synthetic data. The KS estimator has the advantage of being independent of the statistical properties of the data under test. It is demonstrated in the paper that such a convenient property makes the estimator robust to correlations among spectral data. Two different estimators are investigated for the problem of comparing a data series to a model for excess noise detection. One estimator (IR) is based on the integral of the normalized WOSA spectrum; the other is a KS estimator for the comparison of an empirical cumulative distribution function with a limiting theoretical function. Despite the fact that the IR estimator proves to have better sensitivity on independent data, the versatility of the KS estimator is highly advantageous on correlated data. Since the statistics of the estimators are distorted by the presence of correlations among test data, the standard and simple procedures for the determination of the confidence intervals, based on the inversion of the limiting distribution function, provide inaccurate results. While, in the case of the IR estimator, the problem can be overcome only with dedicated Monte Carlo simulations, we demonstrated that the introduction of a shape parameter allows us to use standard equations to calculate proper boundaries for KS confidence intervals. The shape parameter depends on the required significance level and on the data correlations. It does not depend on the number of data considered for the test. Since correlations among spectral data are mainly introduced by the windowing process,

the shape parameter is fixed for a wide family of spectral shapes. For example, synthetic LPF data share the same shape parameter with white noise.

Closely related to noise excess detection, the problem of noise parameter identification is analyzed with a maximum-likelihood approach which, in the particular case of linear dependence on a single parameter, provides an algorithm analogous to the IR estimator used for excess noise detection. A KS algorithm was proposed as an alternative to the IR algorithm and the accuracy and precision of both were tested with a Monte Carlo simulation on LPF synthetic data. Both the IR and KS estimators were demonstrated to give equivalently good results, even though the capability of the KS to handle data correlation is a clear advantage for the definition of a confidence interval for the estimated noise parameter.

Data analysis procedures introduced in this paper are easily extended to the vast context of time-frequency analysis of nonstationary noise. The KS algorithm can be applied effectively both to spectrogram and wavelets coefficients provided that correlations among data are taken into account. The KS algorithm has the advantage that can be generalized to two dimensions thus allowing to extend the analysis to the time-frequency plane.

APPENDIX A: STATISTIC OF THE SAMPLE SPECTRUM

In the case of a discrete, real-valued stationary process $\{x_h\}$, the continuous spectral density function is defined as

$$S(f) = T \sum_{h=-\infty}^{\infty} s_h e^{-\iota 2\pi f h T}. \tag{A1}$$

Here, s_h is the autocovariance sequence of the process $\{x_h\}$, T is the sampling period and f is the frequency expressed in Hz. f is defined in the range $|f| \leq f_{Nq} \equiv 1/2T$ and f_{Nq} is known as the Nyquist frequency. In the case of a finite representation $x_0, \ldots, x_{N-1}$ of the discrete process $\{x_h\}$, the approximation to the spectral density function is provided by the sample spectrum

$$\tilde{P}(f) = \frac{T}{N} \left| \sum_{h=0}^{N-1} x_h e^{-\iota 2\pi f h T} \right|^2. \tag{A2}$$

f in this case is also defined on the interval $[-f_{Nq}, f_{Nq}]$. If the sample spectrum is calculated on the grid of Fourier frequencies ($f_k = k/(NT)$, $|k| \leq N/2$) then it corresponds to the squared modulus of the discrete Fourier transform of the data sequence $x_0, \ldots, x_{N-1}$. In practical applications, only the positive frequency part of the spectrum is considered, and the one-sided sample spectrum is defined as $P(f_k) = 2\tilde{P}(f_k)$ with $k = 0, 1, \ldots, N/2$. In the rest of the paper, the one-sided sample spectrum will be simply named the sample spectrum.

If the data series $x_0, \ldots, x_{N-1}$ is Gaussian distributed and the elements x_j are independent, then the Fourier transform produces a complex series $X(f_k)$ whose elements are approximately independent and whose real and imaginary parts are Gaussian distributed. The term $|X(f_k)|^2 = |\text{Re}[X(f_k)]|^2 + |\text{Im}[X(f_k)]|^2$ is then the sum of two independent variables distributed as a χ^2_ν where ν is the number of degrees of freedom of the distribution ($\nu = 1$ in this case). If the correlations among the elements of the data series $x_0, \ldots, x_{N-1}$ are nonvanishing, the statistical properties of the sample spectrum, in the simplifying assumption of independent $P(f_k)$ elements, can be calculated from the case of a χ^2_ν distributed variable y multiplied by a constant $z = \lambda y$. The characteristic function for z is

$$\phi_z(t) = E[e^{\iota t \lambda y}] = (1 - 2\iota t \lambda)^{-(\nu/2)}. \tag{A3}$$

Here, $E[\ ]$ indicates the expected value. The inverse Fourier transform of $\phi_z(t)$ provides the probability density function for z:

$$\mathcal{F}^{-1}[\phi_z(t)] = \frac{e^{-((z)/(2\lambda))} z^{((\nu)/(2)-1)}}{(2\lambda)^{(\nu/2)} \Gamma(\frac{\nu}{2})}. \tag{A4}$$

This is a gamma distribution, $f(z; k, \theta)$, with $k = \nu/2$ and $\theta = 2\lambda$. $\Gamma(k)$ is the gamma function. In the case of the sample spectrum at a given frequency, $z = P(f_k)$ and $\lambda = E[P(f_k)]/\nu \approx S(f)/\nu$ [30]. It is useful, for the statistical analysis of the spectrum, to introduce the normalized sample spectrum

$$R(f_k) = \nu \frac{P(f_k)}{S(f)}, \tag{A5}$$

which, at each frequency f_k, is distributed as χ^2_ν.

The sample spectrum in Eq. (A2) can also be seen as a special case of the power spectral density, $S(f)$, when the infinite data series $\{x_h\}$ is chopped by a square data window of length N. This operation introduces a considerable amount of spectral leakage because of the convolution with the frequency response of the square data window [9,31]. Therefore, it is common practice to multiply the time series $x_0, \ldots, x_{N-1}$ with a more performant data window, which increases the accuracy of the sample spectrum in the case of processes with a high dynamic range. The application of a data window introduces correlations among different elements of the sample spectrum. Such correlations affect the statistics of the spectrum, resulting in a change in the probability distribution of the sample spectrum. An analytical treatment of the spectrum statistics under such conditions is cumbersome, and it is easier to numerically evaluate (with a Monte Carlo simulation) the statistics of the sample spectrum for the case of interest.

In order to improve the variance properties of the sample spectrum, the WOSA method is applied [9]. The data series $x_0, \ldots, x_{N-1}$ is divided in overlapping windowed segments. The estimates of the sample spectrum of each segment are then averaged. The practice of averaging

overlapping segments can modify the expected statistics of the spectrum since the data in different segments can be correlated. Then, even in the simplifying assumption of vanishing spectral correlations from windowing and overlapping, the averaging process changes the statistics of the estimated sample spectrum. In the assumption of vanishing window and overlap correlations, the statistic of the WOSA spectrum corresponds to the average of N_s gamma-distributed variables

$$P_{\text{WOSA}}(f_k) = \frac{1}{N_s} \sum_{j=1}^{N_s} P_j(f_k). \qquad (A6)$$

The critical function for the sum is $\phi_{P_{\text{WOSA}}}(t) = \prod_j \phi_j(t)$, where $\phi_j(t)$ is the critical function for P_j/N_s. Thanks to Eq. (A3),

$$\phi_{P_{\text{WOSA}}}(t) = \left(1 - \frac{\iota t \theta}{N_s}\right)^{-kN_s}, \qquad (A7)$$

where $\theta = 2S(f_k)/\nu$ and $k = \nu/2$. The inverse Fourier transform of $\phi_{P_{\text{WOSA}}}(t)$ provides the probability density function for the WOSA spectrum

$$f_{\text{WOSA}}(z; h, \delta) = \frac{z^{(h-1)}e^{-(z)/(\delta)}}{\delta^h \Gamma(h)}, \qquad (A8)$$

with $\delta = S(f_k)/N_s$ and $h = N_s$. Again, it is useful to define a normalized WOSA spectrum as

$$R_{\text{WOSA}}(f_k) = \frac{P_{\text{WOSA}}(f_k)}{S(f_k)}, \qquad (A9)$$

which is gamma distributed [Eq. (A8)] with $\delta = 1/N_s$ and $h = N_s$.

APPENDIX B: KOLMOGOROV—SMIRNOV TEST

Kolmogorov—Smirnov is a well-known test for distributions [13–16,32,33]. An ECDF is tested against a continuous theoretical model or, alternatively, two ECDFs are tested with the hypothesis that they share the same limiting cumulative distribution function. Indicating with $f(x)$ the probability density function associated with a given random process X, the corresponding CDF is defined as

$$F(x) = \text{Prob}[X \le x] = \int_{-\infty}^x f(u)du. \qquad (B1)$$

Given a particular realization of the random process X,

$$X_N = \{x_1, \dots, x_N\}, \qquad (B2)$$

ECDF is written as $F_N(x) = k/N$ where k is the number of observations which is smaller or equal to x.

Given two data series X_N and Y_M, with N and M not necessarily equal, we can test if the two series are two particular realizations of the same random variable by analysis of their ECDFs. Under the hypothesis that the two data series come from the same distribution function,

Kolmogorov has demonstrated that the maximum distance between the two ECDFs has a limiting distribution which is independent from the statistical properties of the corresponding random variable. If the test is performed against a theoretical distribution, the distance is defined as

$$d_K = \max|F_N(x) - F(x)|. \qquad (B3)$$

In such a case, $K = N$. Alternatively, if the test is performed between two ECDFs, $K = (NM)/(N + M)$ and

$$d_K = \max|F_N(x) - F_M(x)|. \qquad (B4)$$

The test is defined as follows:
(1) In the case of the test on a single data series, the null hypothesis is that the data are realizations of a random variable which is distributed according to the given probability distribution. In the case of two data series, the null hypothesis is that the two data series are realizations of the same random variable, which means their ECDFs should tend to the same CDF. The test rejects or accepts the null hypothesis on the basis of the analysis of d_{nm}.
(2) A significance level α is defined as the probability that the test rejects the null hypothesis when it is indeed true.
(3) The test can be formulated in terms of critical values. The critical value for the test is the value of $d_K(\alpha)$ corresponding to the significance level. Then, if $d_K > d_K(\alpha)$, the null hypothesis is rejected.

KS theory was formulated for independent data sets and the available equations for critical values are valid only if this condition is satisfied [16]. The test can be formulated also in the presence of data correlation but the distortion to d_K statistic introduced should be taken into account. This is possible if an effective value for K is introduced as $K_{\text{eff}} = \beta K$, with β a shape parameter depending only on the data correlations and the required significance level. Alternatively, realistic critical values can be calculated with dedicated MC simulations [34]. The advantage of the shape parameter β is that it depends only on correlations and the required significance level. Therefore, it can be determined for a specific spectrum (e.g., white noise) and shared among a wide family of spectral shapes. Once β is known, it can be used to calculate critical values for correlated data using standard equations reported in the literature [16].

Focusing on the particular problem, we performed a Monte Carlo estimation of $d_K(\alpha)$ for WOSA spectra representative of LPF. The number of frequency data considered is $N_f = 341$, corresponding K values are $K = N_f$ in the case of the test against a theoretical distribution and $K = N_f/2$ in the case of the test between two ECDFs. The results are summarized in Table II. In the same table, we report the values of β that are required to obtain proper critical values from the standard equations in the case of the test against a theoretical distribution.

TABLE II.   Table of KS critical values for correlated spectral data. Critical values are calculated with a Monte Carlo simulation for different values of the significance level α. Data sets used are representative of the data analyzed in the current paper. We reported the values for testing an ECDF against a theoretical CDF ($K = N_f$) and the values for testing two ECDFs for the same limiting CDF ($K = N_f/2$). We also reported the values of the shape parameter beta that can be used to calculate proper critical values from standard equations in the case of correlated data. α refers to the significance level whereas $1 - \alpha$ is the corresponding confidence level for the test. $N_f = 341$.

α	$1 - \alpha$	$K = N_f$	β	$K = N_f/2$
0.32	0.68	0.0643	0.65	0.0723
0.10	0.90	0.0863	0.58	0.0910
0.05	0.95	0.0969	0.55	0.1006
0.01	0.99	0.1214	0.52	0.1191

TABLE III.   Table of confidence bounds for the IR estimator on correlated spectral data. The values are calculated with a Monte Carlo simulation for different values of the significance level α. Data sets used are representative of the data analyzed in the current paper. The number of frequency points is $N_f = 341$, corresponding to the number of available spectral data in the range [0.1, 10] mHz, for a time series 24 h long and number of averages for the WOSA estimator $N_s = 4$.

α	$1 - \alpha$	x_{lw}	x_{up}
0.32	0.68	325.88	357.41
0.10	0.90	316.33	367.28
0.05	0.95	311.33	372.55
0.01	0.99	300.48	380.69

APPENDIX C: IR TEST

The statistics of the IR excess noise estimator are numerically estimated with a Monte Carlo simulation on the spectral data used for the present paper. Data series are 24-hour-long synthetic reconstructions of the force per unit mass expected on the TMs. WOSA spectra are calculated with $N_s = 4$ averages on 50% overlapped segments which were windowed with the Blackman-Harris window. Results are reported in Table III.

APPENDIX D: CONVERSION OF DISPLACEMENT NOISE

LPF can be considered as a three-body controlled system composed of the two TMs and the SC. Its equations of motion along the measurement axis can be written as

$$m_1 \ddot{x}_1 + m_1 \ddot{x}_{sc} + m_1 \omega_1^2 x_1 = f_1,$$
$$m_2 \ddot{x}_2 + m_2 \ddot{x}_{sc} + m_2 \omega_2^2 x_2 = f_2 + f_{c2},$$
$$m_{sc} \ddot{x}_{sc} - m_1 \omega_1^2 x_1 - m_2 \omega_2^2 x_2 = f_{sc} - f_{c2} + f_{csc}.$$
(D1)

Here,

(i) x_1 and x_2 are TMs coordinates along the sensitive axis. They are relative coordinates in the SC reference frame.

(ii) x_{sc} is the absolute SC coordinate along the sensitive axis.

(iii) m_1, m_2, and m_{sc} are the masses of the two TMs and of the SC.

(iv) ω_1^2 and ω_2^2 are the parasitic stiffnesses coupling the TMs and the SC. The TMs are coupled to the spacecraft through the parasitic stiffness thus producing an oscillator-like equation of motion. The spacecraft at the same time experiences reaction forces given by $-m_1 \omega_1^2 x_1$ and $-m_2 \omega_2^2 x_2$.

(v) f_1, f_2, and f_{sc} are the forces acting on TMs and SC, respectively.

(vi) f_{c2} and f_{csc} are control forces on the second TM and the SC, respectively. Since f_{c2} is an internal force to the system, the SC experiences a reaction force $-f_{c2}$.

(vii) Dots over symbols represent time derivatives.

In the main LPF science operation mode, one TM (indicated here as TM_1) is in free fall and provides the reference for the other TM (TM_2) and the SC. In order to avoid unwanted drifting, both TM_2 and the SC are controlled to follow TM_1. It is worth noting that the system is, by construction, symmetric and the role of the two TMs can be inverted. In order to avoid confusion, we indicate with TM_1 the free-fall reference and with TM_2 the actuated TM.

Moving to the Laplace domain, substituting for the SC dynamics and substituting for the differential coordinate x_Δ of TM_2 with respect to TM_1, the Eq. (D1) can be rewritten as

$$s^2 x_1 \omega_1^2 \left(1 + \frac{m_1}{m_{sc}}\right) x_1 + \omega_2^2 \frac{m_2}{m_{sc}} x_1 + \omega_2^2 \frac{m_2}{m_{sc}} x_\Delta$$
$$= \frac{f_1}{m_1} - \frac{f_{sc}}{m_{sc}} + \frac{1}{m_{sc}} H_2 o_\Delta - \frac{1}{m_{sc}} H_{sc} o_1,$$
$$s^2 x_\Delta + (\omega_2^2 - \omega_1^2) x_1 + \omega_2^2 x_\Delta = \frac{f_2}{m_2} - \frac{f_1}{m_1} + \frac{1}{m_2} H_2 o_\Delta.$$
(D2)

Here,

(i) o_1 and o_Δ are output displacement signals as provided by the interferometer readout system. o_1 is the displacement between the SC and the TM_1. o_Δ is the displacement of TM_2 relative to TM_1.

(ii) H_2 and H_{sc} are transfer functions of the control systems on TM_2 and SC. The force applied by the controllers is calculated on the basis of the output displacement, therefore $f_{c2} = H_2 o_\Delta$ and $f_{csc} = H_{sc} o_1$.

Calculations can be more easily performed if we introduce a matrix notation:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_\Delta \end{pmatrix}, \tag{D3}$$

$$\mathbf{o} = \begin{pmatrix} o_1 \\ o_\Delta \end{pmatrix}, \tag{D4}$$

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_{\text{sc}} \end{pmatrix}, \tag{D5}$$

$$\mathbf{D} = \begin{pmatrix} s^2 + \omega_1^2 + \frac{m_1}{m_{\text{sc}}}\omega_1^2 + \frac{m_2}{m_{\text{sc}}}\omega_2^2 & \frac{m_2}{m_{\text{sc}}}\omega_2^2 \\ \omega_2^2 - \omega_1^2 & s^2 + \omega_2^2 \end{pmatrix}, \tag{D6}$$

$$\mathbf{G} = \begin{pmatrix} \frac{1}{m_1} & 0 & -\frac{1}{m_{\text{sc}}} \\ -\frac{1}{m_1} & \frac{1}{m_2} & 0 \end{pmatrix}, \tag{D7}$$

$$\mathbf{C} = \begin{pmatrix} -\frac{1}{m_{\text{sc}}} & \frac{1}{m_{\text{sc}}} \\ 0 & \frac{1}{m_2} \end{pmatrix}, \tag{D8}$$

$$\mathbf{H} = \begin{pmatrix} H_{\text{sc}} & 0 \\ 0 & H_2 \end{pmatrix}. \tag{D9}$$

With the notation introduced, the equation of motion can be rewritten:

$$\mathbf{D} \cdot \mathbf{x} = \mathbf{G} \cdot \mathbf{f} + \mathbf{C} \cdot \mathbf{H} \cdot \mathbf{o}. \tag{D10}$$

Then, we should consider that the output displacement $\mathbf{o}$ corresponds to the measurement of $\mathbf{x}$ provided by the sensing system:

$$\mathbf{o} = \mathbf{S} \cdot \mathbf{x} + \mathbf{o_{rn}}. \tag{D11}$$

Here, $\mathbf{S}$ is a 2×2 sensing matrix and $\mathbf{o_{rn}}$ is the readout noise.

Substituting Eq. (D11) into Eq. (D10), the dynamics can be rewritten in terms of the known output $\mathbf{o}$:

$$\mathbf{D} \cdot \mathbf{S}^{-1} \cdot \mathbf{o} - \mathbf{C} \cdot \mathbf{H} \cdot \mathbf{o} = \mathbf{G} \cdot \mathbf{f} + \mathbf{D} \cdot \mathbf{S}^{-1} \cdot \mathbf{o_{rn}}. \tag{D12}$$

The quantity $\mathbf{G} \cdot \mathbf{f}$ represents the force per unit mass acting on the test masses. Such a quantity is not directly known from the system output since the available signal is $\mathbf{o}$. From Eq. (D12), it is readily seen that an effective force per unit mass acting on the TMs can be reconstructed

from the knowledge of $\mathbf{o}$ if the force applied by the control system $(\mathbf{C} \cdot \mathbf{H} \cdot \mathbf{o})$ is calculated and subtracted from the reconstructed dynamics $\mathbf{D} \cdot \mathbf{S}^{-1} \cdot \mathbf{o}$. Therefore, the available quantity is $\mathbf{g} = \mathbf{G} \cdot \mathbf{f} + \mathbf{D} \cdot \mathbf{S}^{-1} \cdot \mathbf{o_{rn}}$ which contains the force per unit mass acting on the TMs corrupted by the readout noise of the system.

The expected values for $\mathbf{g}$ can be obtained thanks to the dynamics reported in Eq. (D12) once the expected values for $\mathbf{f}$ and $\mathbf{o_{rn}}$ are known. The same applies to the projection of the noise spectra for $\mathbf{f}$ and $\mathbf{o_{rn}}$ to $\mathbf{g}$.

APPENDIX E: HUBER'S NORM

Huber's norm [35] is a way to construct a goodness-of-fit estimator which is more robust to outliers and non-Gausianity than the standard MSE. The norm is constructed as $\sum_i \rho(r_i)$, where r_i are the residuals between a data series and a parametric model. The function $\rho(r_i)$ is defined as

$$\rho(r_i) = \begin{cases} \frac{1}{2}r_i^2 & \text{for } |r_i| < c \\ c|r_i| - \frac{1}{2}c^2 & \text{for } |r_i| \geq c. \end{cases} \tag{E1}$$

The value of the threshold constant c may depend on the given data set, therefore some efforts must be spent to select the value of c providing the most accurate results. Normalized residuals are defined as $r_i = (P_{\text{WOSA}}(f_i) - S_a(f_i; \theta))/S_a(f_i; \theta)$ in accordance to the MSE definition in Sec. V C. With such a definition, residuals are expected to be zero mean and unitary variance in correspondence of the true model $S_a(f_i; \theta)$. Different values of c were tested with a Monte Carlo estimation on the first 1000 data of the analysis reported in Sec. V C. Huber's norm is minimized for each MC iteration and the corresponding values of the parameter θ are stored. The histogram of θ shows a mean steadily fixed on 0.13 for values of c ranging from 0.001 to 0.1. Increasing the value of c, the distribution starts to shift toward the distribution obtained with MSE minimization. MSE and Huber distributions are practically undistinguishable for values $c > 2$. Since the true value for the parameter θ is set to 0.1, Huber's norm is performing better for values $c < 0.1$. It was then decided to use the value $c = 0.05$ for the final analysis reported in Fig. 5.

[1] F. Antonucci *et al.*, Classical Quantum Gravity **28**, 094002 (2011).

[2] F. Antonucci *et al.*, Classical Quantum Gravity **28**, 094006 (2011).

[3] F. Antonucci *et al.*, Classical Quantum Gravity **28**, 094001 (2011).

[4] M. Armano *et al.*, Classical Quantum Gravity **26**, 094001 (2009).

[5] LISA International Science Team, Report No. ESA/SRE (2011)3, 2011.

[6] J. Crowder and N. J. Cornish, Phys. Rev. D **72**, 083005 (2005).

[7] S. Kawamura *et al.*, Classical Quantum Gravity **23**, S125 (2006).

[8] W. Ni, S. Shiomi, and A. Liao, Classical Quantum Gravity **21**, S641 (2004).

[9] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications* (Cambridge University Press, Cambridge, England, 1993), p 290.

[10] M. Hewitson *et al.*, Classical Quantum Gravity **26**, 094003 (2009).

[11] MATLAB, http://www.mathworks.com.

[12] LTPDA: a MATLAB toolbox for accountable and reproducible data analysis, http://www.lisa.aei-hannover.de/ltpda.

[13] A. N. Kolmogorov, Giorn. Ist. Ital. Attuari **4**, 83 (1933).

[14] W. Feller, Ann. Math. Stat. **19**, 177 (1948).

[15] N. Smirnov, Ann. Math. Stat. **19**, 279 (1948).

[16] L. H. Miller, J. Am. Stat. Assoc. **51**, 111 (1956).

[17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), p. 260.

[18] As formulated by Eq. (6), the test is searching for differences in the noise content of the data series. It does not provide information about which series has the largest noise content. Such information can be recovered graphically with a distribution plot (for example, a quantile-quantile plot). Alternatively, a pure excess noise test for a data series, x_m, with respect to a reference series, x_n, can be formulated by substituting d_K with $d_K^+ = \max\{F_{n_2}(x) - F_{n_1}(x)\}$. The formalism used for the calculation of d_K and d_K^+ is the same, therefore the procedures described in this paper can be applied in either case without any significant modification.

[19] A. Monsky *et al.*, Classical Quantum Gravity **26**, 094004 (2009).

[20] L. Ferraioli, M. Hueller, and S. Vitale, Classical Quantum Gravity **26**, 094013 (2009).

[21] The distribution of the WOSA estimator for the sample spectrum is a gamma distribution, as reported in Eq. (A8). Such a distribution tends to a Gaussian distribution as the number of averages increases. The difference between the true distribution and the corresponding Gaussian can be quantified by the maximum distance among cumulative distribution functions, d. As an example, $d = 0.067$ for $N_s = 4$, $d = 0.03$ for $N_s = 16$, $d = 0.019$ for $N_s = 50$, $d = 0.013$ for $N_s = 100$, and $d = 0.0094$ for $N_s = 200$. It

is worth noting that since d is a difference in the space of the values of the CDF, it can assume values in [0, 1]. As can be seen, the distribution of the WOSA spectrum reasonably approaches a Gaussian only for N_s as large as 200.

[22] The term pseudostationary indicates a data series which is affected by slight nonstationarity of the kind that can be removed with mean subtraction or standard polynomial fit trend removal.

[23] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 2010), p. 714.

[24] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, 1998).

[25] J. A. Peacock, Mon. Not. R. Astron. Soc. **202**, 615 (1983) [http://adsabs.harvard.edu/abs/1983MNRAS.202..615P].

[26] G. Fasano and A. Franceschini, Mon. Not. R. Astron. Soc. **225**, 155 (1987) [http://adsabs.harvard.edu/abs/1987MNRAS.225..155F].

[27] L. Ferraioli *et al.*, Phys. Rev. D **82**, 042001 (2010).

[28] D. J. Thomson, Bell Syst. Tech. J. **56**, 1769 (1977) [http://www.alcatel-lucent.com/bstj/vol56-1977/bstj-vol56-issue09.html].

[29] If N_s is the number of averaging segments, N_f the length of each segment, and h the shift factor, which is indicating the number of data points two consecutive data segments are shifted by, then the correction to the variance of the averaged process is proportional to [9] $\Gamma(h) = \sum_{j=1}^{N_s-1} \sum_{t=1}^{N_f} h_t h_t^{jh}$. Here, h_t^{jh} are window elements shifted by a factor jh. With our choice of $N_s = 4$, a segment overlap of 50% and a total data length of 24 h, we find $\Gamma(h) = 2.7 \times 10^{-4}$.

[30] It is easy to verify that, in the case of $z = P(f_k)$ and $\lambda = E[P(f_k)]/\nu$, the probability density function $f(z; \nu, \lambda) = \frac{e^{-((z)/(2\lambda))} z^{((\nu)/(2)-1)}}{(2\lambda)^{(\nu)/(2)} \Gamma(\frac{\nu}{2})}$ correctly represents the statistic of the sample spectrum. $E[z] = \frac{\nu}{2} 2\lambda = E[P(f_k)]$ as expected for the spectrum [9]. $\mathrm{var}[z] = \frac{\nu}{2}(2\lambda)^2 = \frac{2}{\nu} E[P(f_k)] = E[P(f_k)]$ since $\nu = 2$ in the present case.

[31] F. J. Harris, Proc. IEEE **66**, 51 (1978).

[32] N. I. Fisher, Int. Stat. Rev. **51**, 25 (1983).

[33] M. B. Wilk and R. Gnanadesikan, Biometrika **55**, 1 (1968).

[34] M. S. Weiss, J. Am. Stat. Assoc. **73**, 872 (1978).

[35] P. J. Huber, Ann. Stat. **1**, 799 (1973).