Identifying and Addressing Nonstationary LISA Noise

Matthew C. Edwards[1,2], Patricio Maturana-Russel[3], Renate
Meyer[1], Jonathan Gair[2,4], Natalia Korsakova[5], Nelson Christensen[5]

[1] *Department of Statistics, University of Auckland, Auckland, New Zealand*
[2] *School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom*
[3] *Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[4] *Albert Einstein Institute, Max Planck Institute for Gravitational Physics, Potsdam, Germany*
[5] *Université Côte d'Azur, Observatoire de Côte d'Azur, CNRS, Artemis, Nice, France*

We anticipate noise from the Laser Interferometer Space Antenna (LISA) will exhibit nonstationarities throughout the duration of its mission due to factors such as antenna repointing, cyclostationarities from spacecraft motion, and glitches as highlighted by LISA Pathfinder. In this paper, we use a surrogate data approach to test the stationarity of a time series, with the goal of identifying noise nonstationarities in the future LISA mission. This will be necessary for determining how often the LISA noise power spectral density (PSD) will need to be updated for parameter estimation routines. We conduct a thorough simulation study illustrating the power/size of various versions of the hypothesis tests, and then apply these approaches to differential acceleration measurements from LISA Pathfinder. We also develop a data analysis strategy for addressing nonstationarities in the LISA PSD, where we update the noise PSD over time, while simultaneously conducting parameter estimation, with a focus on planned data gaps. We show that assuming stationarity when noise is nonstationary leads to systematic biases and large posterior variances in parameter estimates for galactic white dwarf binary gravitational wave signals.

I. INTRODUCTION

The Laser Interferometer Space Antenna (LISA) is a planned space-based gravitational wave (GW) mission with an expected launch in 2034 led by the European Space Agency (ESA) [5]. The aim of this mission is to observe GW signals in the millihertz band which among others include astrophysical objects such as galactic white dwarf binaries [19], massive and supermassive black hole binaries [44], and extreme mass ratio inspirals (EMRIs) [21]. LISA will consist of a set of three spacecrafts arranged into an "equilateral" triangle, each separated by $L = 2.5 \times 10^6$ km connected with a laser link. The LISA constellation will cartwheel in an Earth-trailing heliocentric orbit around the Sun at an angle of 20 degrees between the Sun and Earth.

We expect LISA noise will be nonstationary in numerous ways. For example, as the spacecrafts will not always be able to point in the same direction towards Earth for us to receive data, there will be planned communication interruptions (gaps), where the antennae will be repointed to adjust the beam [13, 19]. This means physically moving the antennae, which will create noise. Another subtle effect of the repointing is that the distribution of mass near the test mass will change, which might affect the gravity gradient noise, leading to a change in acceleration noise [12, 36]. Controls may need to actively hold the proof mass using electrostatic actuation, which may lead to charging of the proof mass, and a change in the state of the noise [11, 14, 34].

Cyclostationarities are also expected in LISA, for example, due to the cartwheeling motion and orbits of the satellites. As LISA does not have uniform sensitivity in the sky and is more sensitive in the direction perpendicular to the plane of the constellation, there will be higher amplitude confusion noise when pointing to the line of sight of the galactic centre as this is where a large amount of galactic white dwarf binaries are located [29]. In addition, LISA has a periodic orbit around the Sun, pseudo-periodic solar activity can lead to cyclostationary noise [3, 4].

LISA Pathfinder (LPF) was an ESA satellite whose goal was to demonstrate the technology for the future LISA mission [7]. Glitches in differential acceleration measurements Δg have been analyzed in previous studies, occurring at a rate of one glitch per two days [8, 9]. As LISA will have a similar architecture to LPF, we expect glitches as another form of nonstationarity in the future mission [38].

To understand exactly what it means to have nonstationary noise, first we must discuss precisely what a stationary process is. A (weakly) stationary time

series $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ is a stochastic process that has constant and finite mean and variance over time, i.e.,

$$\mathbb{E}[Y_t] = \mu < \infty,$$
$$\mathrm{Var}[Y_t] = \sigma^2 < \infty,$$

for all t, and an autocovariance function $\gamma(.)$ that depends only on the time lag s [18]. That is, for a zero-mean weakly stationary process, the autocovariance function has the form

$$\gamma(s) = \mathbb{E}[Y_t Y_{t+s}], \quad \forall t,$$

where $\mathbb{E}[.]$ is the expected value operator, and t represents time. Note that the PSD function is the Fourier transform of the autocovariance function.

Nonstationarities in a time series can therefore come in the form of a trend, heteroskedasticity, or time-varying autocorrelations (or PSDs). One can also consider amplitude modulation (AM) and frequency modulation (FM) to be forms of nonstationarity. In this paper, we are interested in a time-varying PSD structure, where we want to identify and handle this type of nonstationarity. To this end, we propose two hypothesis tests to identify whether a time series is stationary in terms of its PSD, which will be described in Sections II C and II D. Further, we have developed an analysis strategy for dealing with nonstationary LISA noise, where we update the estimate of the noise PSD over time, rather than fixing it and assuming stationarity. It is worth noting that in the context of Laser Interferometer Gravitational-Wave Observatory (LIGO) data analysis, fluctuations in the PSD have systematically biased parameter estimates [1, 2, 16]. Here, we are particularly interested in the gap problem [13, 19], where we believe satellite repointing could temporarily change the noise structure of the LISA satellites.

Common approaches to testing the stationarity of a time series are the so-called *unit root tests*, including the Augmented Dickey-Fuller (ADF) test [43], Phillips-Perron (PP) test [33], and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [28]. Unit root tests have been noted in the GW literature by Romano and Cornish [39] to not be of particular value as GW noise generally exhibits high autocorrelation with roots close to the unit circle, which, as Müller [31] outlines, is undesirable.

There are also the *spectral analysis tests* that consider evolutionary (or time-varying) spectral estimates using time-frequency representations of the data. The most notable of these are the wavelet test of von Sachs and Neumann [42], where the authors propose using Haar wavelets of time-varying periodograms to test for covariance stationarity, and the Priestly-Subba Rao test [35] which tests the uniformity of a set of evolutionary spectra at different time intervals, and is similar to a two-factor analysis of variance (ANOVA). Another useful variant of the wavelet test was later proposed by Nason et al [32]. In the context of GW data analysis for LIGO and Virgo, Abbott et al [2] applied a a scaleogram test of stationarity using an Anderson-Darling test [6].

Various *resampling* approaches for testing the stationarity of a time series have also been introduced, where (usually) no parametric assumptions are made about the distribution of the test statistic under the null hypothesis. One such approach by Swanepoel and Van Wyk [45] uses a modification of the bootstrap of Efron [25] to test the equality of two spectral densities from two independent time series. Dette and Paparoditis [22] use a frequency-domain bootstrap to more generally test the equality of two or more spectral densities.

Our tests fall into the lesser-known *surrogate data* tests which were first introduced by Theiler et al [46] for testing non-linearities in time series, and later adapted by Xiao et al [50] and Borgnat and Flandrin [17] for testing stationarity. These tests are nonparametric in nature, where the original data are resampled to create stationary surrogates with the same periodogram. A version of the multitaper spectrogram of Thomson [47] with Hermite (rather than Slepian) window functions (as discussed by Bayram and Baraniuk [15]) is computed, where the estimated spectrum in each time segment is compared to a time-averaged spectrum using a distance measure, typically a combination of the Kullback-Leibler divergence and the log spectral deviation. The test statistic for these tests are the sample variance of these distances and a null distribution of test statistics (which usually looks like a Gamma distribution in shape) can be generated by replicating this on a large number of surrogates.

In this paper, we propose two variants on the surrogate data testing of Xiao et al [50] and Borgnat and Flandrin [17]. We consider an autoregressive spectrogram where each short-time segment uses a

frequentist autoregressive (AR) estimate of its spectrum, with order selected based on the Akaike information criterion (AIC). In the first variant, we can either compute the Kolmogorov-Smirnov statistic or the Kullback-Leibler distance to measure the distance between local spectra of short time segments and the global spectrum. A test statistic is then computed as the sample variance of these distances and we use surrogates to populate the sampling distribution of this test statistic under the null hypothesis of stationarity. Large variability in the distances of the original time series would provide evidence against stationarity. As a novel alternative, we fit a least squares regression line to the cumulative median of Euclidean distances between columns in the AR spectrogram. The slope of this line is used as a test statistic and surrogates are again used to generate the null distribution. Here, if a time series is stationary, we would expect the PSD in neighbouring segments of the spectrogram to be similar over time, meaning the median of Euclidean distances should fluctuate around a constant. A non-zero slope would then provide evidence against the stationarity hypothesis. In both variants, empirical percentiles are used to create a critical value that is used as a rejection threshold.

We introduce these hypothesis tests to be used as a tool for future LISA data analysis, with the overall goal of determining how often we should update the noise PSD. Once this is decided, parameter estimation routines can be implemented. In this paper, we propose the use of a blocked Metropolis-within-Gibbs sampler to simultaneously estimate the parameters of a galactic white-dwarf binary gravitational wave signal and estimating the noise PSD before and after a planned data gap. We show that this model leads to improved mean squared errors of our parameter estimates than when assuming the PSD does not change.

The paper is structured as follows. In Section II, we introduce the notion of surrogate data testing, defining two specific hypothesis tests to be used in the future LISA mission. We then conduct a simulation study to demonstrate the power of these tests, and then apply the tests to differential acceleration measurements from LPF to highlight nonstationarities in that data. In Section III, we introduce our data analysis strategy for handling nonstationary LISA noise. We inject a galactic white-dwarf binary GW signal in nonstationary noise and im-

plement a blocked Metropolis-within-Gibbs sampler. We mimic what we believe could happen to LISA noise when repointing satellites during planned gaps, and demonstrate that our model yields lower mean squared errors for all signal parameters than under an incorrect assumption of stationarity. We then give concluding remarks in Section IV.

II. IDENTIFYING NONSTATIONARY NOISE

A. Stationary Surrogates

Surrogate data testing was originally proposed by Theiler et al [46] for testing non-linearities in time series, and later adapted by Xiao et al [50] and Borgnat and Flandrin [17] for testing stationarity. The main idea here is that one can create stationary "surrogates" of a (potentially nonstationary) time series by directly manipulating the data in the frequency-domain, preserving the second-order statistics, but randomizing higher order statistics. In this way, we can generate a stationary surrogate of a time series that has the same empirical spectrum (periodogram) as the original time series.

First, Fourier transform the time series $Y(t)$ using

$$\tilde{Y}(\omega) = \int e^{-2\pi i t \omega} Y(t) dt,$$

to get a frequency-domain representation (where t represents time and ω represents frequency), and express this in polar coordinates such that

$$\tilde{Y}(\omega) = A(\omega)e^{i\varphi(\omega)},$$

where $A(\omega) = |\tilde{Y}(\omega)|$ is the magnitude vector and $\varphi(\omega) = \arg\left(\tilde{Y}(\omega)\right)$ is the phase vector.

Keeping the magnitude vector $A(\omega)$ fixed, we replace the phase vector $\varphi(\omega)$ by a new phase vector $\varphi^*(\omega)$ that is populated by iid Uniform$[-\pi, \pi]$ random variables. We now have a randomized frequency-domain representation of the surrogate $\tilde{Y}^*(\omega) = A(\omega)e^{i\varphi^*(\omega)}$ which is inverse Fourier transformed to give a time-domain representation of the surrogate:

$$Y^*(t) = \frac{1}{n} \int e^{2\pi i t \omega} \tilde{Y}^*(\omega) d\omega.$$

Let $(\omega_0, \omega_1, \ldots, \omega_{n/2-1}, \omega_{n/2})$ be the positive Fourier frequencies. We only randomize the phase for $\omega_1, \omega_2, \ldots, \omega_{n/2-1}$ because ω_0 and $\omega_{n/2}$ are always real-valued with zero phase, and the negative Fourier coefficients are complex conjugates of the positive Fourier coefficients for the inverse Fourier transform to be real-valued, meaning $\varphi(-\omega) = -\varphi(\omega)$.

Surrogates are extremely useful for testing stationarity as they not only have the same periodogram as the original data (which may or may not be stationary), but they are stationary themselves, meaning if one can compute a test statistic that can distinguish the null hypothesis (stationary) from the alternative hypothesis (nonstationary), it is straightforward to generate the sampling distribution of the test statistic but computing the test statistic on a large number of surrogates. We now focus our attention on useful test statistics based on the autoregressive spectrogram.

B. Autoregressive Spectrogram

The spectrogram is the most fundamental tool used in time-frequency analysis. It contains at each column an approximation of the PSD function for consecutive time intervals. Thus, it allows us to assess the evolution of this function over time. It is computed as follows. First compute the short-time Fourier transform (STFT),

$$\tilde{Y}(\omega, T) = \int \mathrm{e}^{-2\pi \mathrm{i} t \omega} Y(t) W(t - T) \mathrm{d}t,$$

where $W(.)$ is a window function of duration T of the window. Then take the squared modulus of each segment. This amounts to computing the periodogram of short windowed segments of the data, which may or may not be overlapping in time.

It is well-known in the time series literature that the periodogram is an asymptotically unbiased estimator of the spectral density function, but it is not a consistent estimator. This has lead to a large amount of literature on periodogram smoothing to reduce the variance.

The most popular parametric approach is to fit an autoregressive model where the order chosen by AIC. In this paper, we use an AR estimate of the spectrum for each segment of the spectrogram rather than using the raw periodogram. Although there

are more sophisticated approaches to spectrum estimation that perhaps do not rely on parametric assumptions (see for example Choudhuri et al [20], Edwards et al [24], Kirch et al [27], Maturana-Russel and Meyer [30] for novel Bayesian approaches), we use the frequentist AR method for the sake of computational speed and ease.

For the remainder of the paper, when computing the AR spectrogram, we utilize the Tukey window with tapering coefficient equal to $(1 - \mathrm{Overlap})/10$.

C. Variance of Local Contrast (VOCAL) Test

In this section, we describe the first of two surrogate tests, which we call the Variance of Local Contrast (VOCAL) Test. As with any hypothesis test, we need to first define a test statistic that can distinguish between the null hypothesis and alternative hypothesis.

First consider the original time series and find its AR spectrogram. We need to contrast local features in the spectrogram with the global spectrum by computing a *local contrast* for each time segment (column) in the spectrogram. This is computed as

$$c_l = \kappa(\hat{f}_l, \hat{f}), \quad l = 1, 2, \ldots, L,$$

where L is the number of time segments (columns) in the spectrogram, $\hat{f}_l$ is the estimated (local) PSD of the lth time segment of the spectrogram, $\hat{f}$ is the estimated (global) PSD of the entire time series (estimated using the same AR routine in the spectrogram), and κ is a suitable spectral distance,

In this paper we use as our local contrasts either the Kolmogorov-Smirnov (KS) statistic

$$\kappa(f_1, f_2) = \sup_{\omega} |F_1(\omega) - F_2(\omega)|,$$

where F_1 and F_2 are standardized empirical cumulative distribution functions (ECDFs) computed by normalizing the estimated PSDs f_1 and f_2, and taking their cumulative sums, or we use the symmetric Kullback-Leibler (KL) divergence

$$\kappa(f_1, f_2) = \frac{1}{2} \int \left(f_1(\omega) - f_2(\omega) \right) \log \frac{f_1(\omega)}{f_2(\omega)} \mathrm{d}\omega,$$

where f_1 and f_2 are normalized PSDs.

Fluctuations in the local contrasts can be used to distinguish between stationarity and nonstationarity as we would expect very little variability in the local contrasts if a time series was stationary and more variability if the time series was nonstationary. To this end, we use the sample variance of local contrasts as the test statistic for this test, i.e.,

$$V = \text{Var}(\mathbf{c}),$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_L)$.

We can then generate the sampling distribution of this test statistic under the null hypothesis by repeating this same process on stationary surrogate data. That is, for each surrogate (indexed by $s = 1, 2, \ldots, S$, for large S) compute the AR spectrogram, the local contrasts $\mathbf{c}_s$, and finally the test statistic to give us

$$V_0(s) = \text{Var}(\mathbf{c}_s), \quad s = 1, 2, \ldots, S,$$

where $\mathbf{c}_s = (c_{s,1}, c_{s,2}, \ldots, c_{s,L})$.

The hypothesis test can then be formalized by considering where V lies in the distribution of V_0. Let

$$H_0 : V < \gamma \quad \text{(Stationary)},$$
$$H_1 : V \geq \gamma \quad \text{(Nonstationary)},$$

where γ is the critical value chosen such that

$$p(V_0 \leq \gamma) = 1 - \alpha,$$

where α is the rejection threshold. Thus for an $\alpha = 0.05$ rejection threshold, γ is computed as the 95% percentile of V_0. Alternatively, an approximate p-value can be computed by

$$\frac{1}{S} \sum_{s=1}^{S} I_{\{V_0(s) \geq V\}},$$

where I is an indicator function. Note that this is a one-sided test.

The precision to which the p-value can be computed depends on the number of surrogates generated. For example, if $S = 1,000$, the p-value can be computed to three decimal places, and if $S = 10,000$, the p-value can be computed to four decimal places.

As an illustrative example of the test, consider the autoregressive (AR) model, defined as:

$$Y_t = \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t,$$

where p is the order, $(\varphi_1, \ldots, \varphi_p)$ are the model parameters, and $\varepsilon_t \sim \text{N}(0, \sigma^2)$ for all t is the white noise innovation process.

Consider the case where we have a length $n = 2^{13}$ time series generated from an AR(2) with parameters (0.9, -0.9), and we concatenate this with a length $n = 2^{13}$ time series generated from an AR(1) with parameter 0.9, each with standard normal innovations, as illustrated in Figure 1.
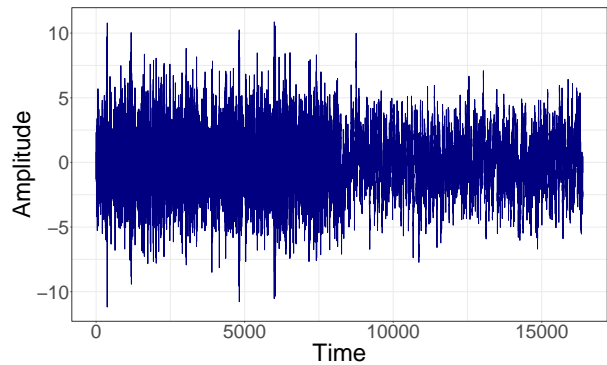

FIG. 1: Time series containing 2^{13} realizations from an AR(2) with parameters $(0.9, -0.9)$ and 2^{13} realizations from an AR(1) with parameter 0.9. Each series uses $\text{N}(0, 1)$ innovations.

Setting the overlap to 75% and window length to 2^{10}, the associated AR spectrogram can be seen in Figure 2. Notice how the spectrum changes around halfway through the time series.
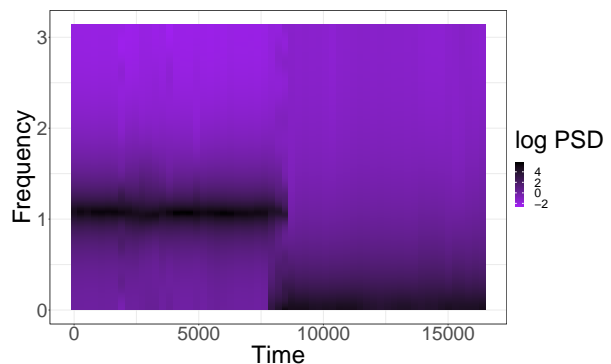

FIG. 2: AR spectrogram from the time series presented in Figure 1. Notice the abrupt change in PSD structure at the halfway point.

We now generate 1,000 surrogates. One example

of a surrogate of our original time series can be seen in Figure 3 and its associated AR spectrogram can be seen in Figure 4.
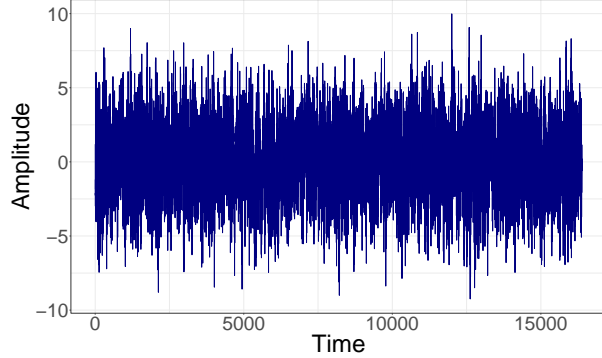

FIG. 3: One example of stationary surrogate data based on the time series presented in Figure 1.
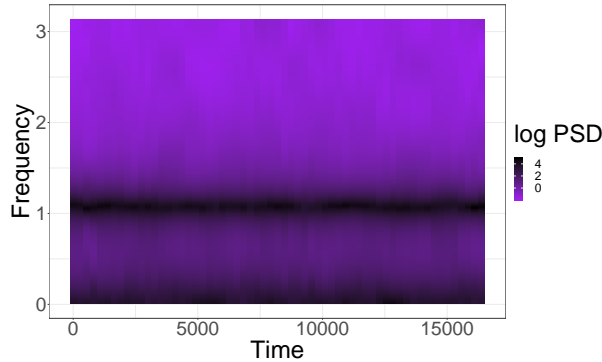

FIG. 4: AR spectrogram from the stationary surrogate data presented in Figure 3.

Using the KS statistic as the local contrast, we can generate the test statistic V from the original data, and the empirical sampling distribution of the test statistic using $(V_0(1), V_0(2), \ldots, V_0(S))$. Using a 5% rejection threshold, we compute the 95% percentile of the empirical sampling distribution. This is illustrated in Figure 5. As the test statistic V is greater than the 95% percentile of the empirical sampling distribution, we reject the null hypothesis of stationarity.
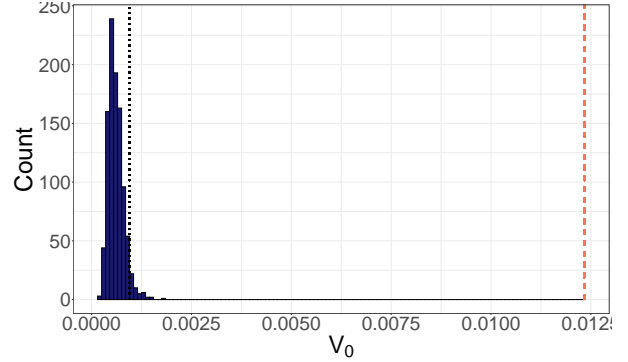

FIG. 5: Empirical sampling distribution of test statistic (variance of local contrasts computed using the KS statistic). The dotted black line is γ (the 95% percentile of this null distribution) and the dashed pink line is the test statistic V from the original time series.

D. Slope of Median Euclidean Distance (SOMED) Test

For our second surrogate test, we compare the *Euclidean distances* between the estimated PSD functions over time, i.e., a comparison between the columns of the spectrogram. If a time series is stationary, each column in the spectrogram should look approximately similar over time (see e.g., Figure 4). Consequently, a sequence of consecutive distances should fluctuate around a constant. We propose to test stationarity by testing the significance of the slope in a simple linear regression model fitted to these distances.

First, we calculate the AR spectrogram. This conforms a matrix $(r \times m)$ where the rows and columns stand for the energy or power at a particular frequency and the time intervals, respectively. Then, we calculate the Euclidean distance of each column with respect to the other ones, that is

$$d_{ij} = \sqrt{\sum_{k=1}^{r} (Y_{ki} - Y_{kj})^2},$$

where $\mathbf{Y}_i = (Y_{1i}, \ldots, Y_{ki}, \ldots, Y_{ri})^\top$ is the ith column of the spectrogram for $i = 1, \ldots, m$. The distances d compound a symmetric matrix $\mathbf{D}$ which has a vector of zeros in its diagonal.

Since $\mathbf{D}$ is symmetric, we discard the upper triangular part and calculate the median of each row, which generates a sequence $\mathbf{v} = (v_2, \ldots, v_m)$, where v_i is the median of the Euclidean distances of the estimated PSD for the ith time interval (column in the spectrogram matrix) with respect to all the estimated PSD of the previous time intervals, i.e., it is a cumulative median. Since the first v_i values embody a few comparisons that tend to generate low discrepancies, these can be discarded, for instance, the first 10% of the sequence.

If the time series is stationary, we would expect a similar PSD across time. In other words, the cumulative median of the Euclidean distances should fluctuate around a constant, which can be tested evaluating the slope of a fitted simple linear regression model. Thus, we fit a linear model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where the responses are the sequence $\mathbf{v}$ and the explanatory variables points in time. We assume that the errors ε_i are independent and identically distributed with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$. If the estimated slope is zero it means that the time series is stationary, otherwise the time series is nonstationary. We assess this assumption of the time series through the following hypotheses:

$$H_0 : \beta_1 = 0 \quad \text{(Stationary)}$$
$$H_1 : \beta_1 \neq 0 \quad \text{(Nonstationary)}.$$

The null hypothesis establishes that the sequence of medians $\mathbf{v}$ does not change over time or equivalently the PSD functions do not vary significantly over time, showing the stationarity of the time series.

To test H_0, we compare the slope estimated from the original data $\widehat{\beta}$ with the empirical distribution of the slopes estimated from surrogate data sets $\widehat{\boldsymbol{\beta}}_S = (\widehat{\beta}_1, \ldots, \widehat{\beta}_S)$, i.e., under the null hypothesis that assumes stationarity. Then, the p-value is calculated by

$$\frac{1}{S} \sum_{s=1}^{S} \left(I_{\{-|\widehat{\beta}| > \widehat{\beta}_s\}} + I_{\{|\widehat{\beta}| < \widehat{\beta}_s\}} \right),$$

where I is an indicator function.

This test also has the potential of detecting glitches using conventional statistical techniques used to detect outliers in linear regression models. This can be assessed by analyzing the cumulative median values of the original data set.

Consider the AR spectrogram used in Section II C. The nonstationary design of this process can be clearly noted in the spectrogram displayed in Figure 2. The two PSDs corresponding to the AR(2) and AR(1) processes have their peaks at different frequencies. This difference is also clear in the comparison of the Euclidean distances displayed in Figure 6. The discrepancy in the PSD estimates is represented in the magnitude of the distances which conform a block in the lower-right part.
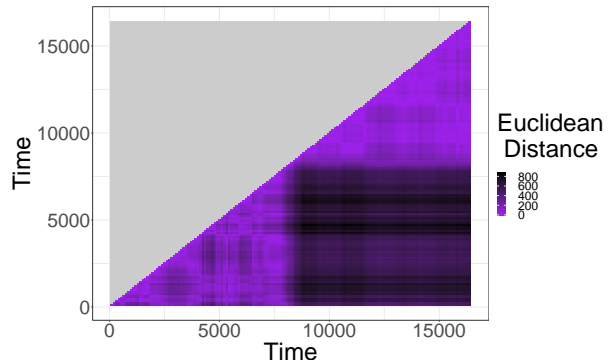

FIG. 6: Euclidean distances for the spectrogram displayed in Figure 2.

The medians of the Euclidean distances of a specific time interval in Figure 6 with respect to its previous intervals are displayed in Figure 7. It can be noticed the design of the process: the first half is centred below the second one. The slope of the simple linear model is evidently non zero. The discrepancy of the PSD estimates do not seem to fluctuate randomly around a constant, which is evidence in favour of the nonstationary nature of the process. Comparing this slope with the empirical distribution of the slopes calculated from the surrogate data sets we get a p-value of 0.000. The SOMED test rejects the null hypothesis, identifying successfully this data set as nonstationary.

E. Testing Simulated Data

We now apply the surrogate tests to simulated AR data (with standard white noise innovations) and compute power or size for different scenarios. Consider a length $n = 2^{12}$ time series $\mathbf{Y}$ that is split in half into two length $n/2 = 2^{11}$ time series $\mathbf{Y}_1$ and $\mathbf{Y}_2$. For the following three scenarios, let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have the:
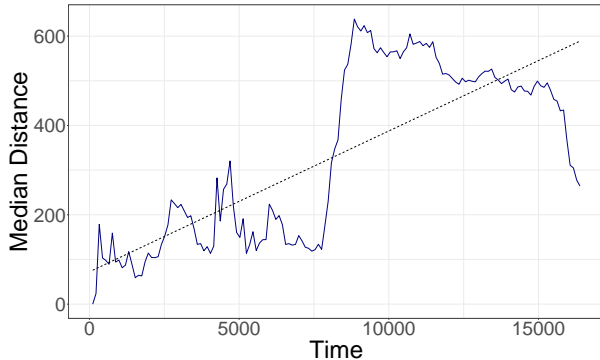
FIG. 7: Median of the Euclidean distances for each column of Figure 6. The dashed line stands for a simple linear model.

1. Same dependence structure;

2. Different dependence structure;

3. Similar dependence structure.

In Scenario 1, we consider a time series with the same dependence structure (and therefore PSD) throughout its duration. Let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be generated from an AR(1) with parameter 0.9. In this scenario, we show that both tests yield small Type I Errors, and fail to reject the null hypothesis of stationarity the vast majority of times.

In Scenario 2, we look at an extreme example, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have vastly different dependence structures. Let $\mathbf{Y}_1$ be generated from an AR(2) with parameters (0.9, -0.9) and $\mathbf{Y}_2$ be generated from an AR(1) with parameter 0.9. Here, we demonstrate that both methods reject the null hypothesis of stationarity, with high power.

In Scenario 3, we let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have very similar (but not equivalent) dependence structures. Let $\mathbf{Y}_1$ come from an AR(1) with parameter 0.8 and $\mathbf{Y}_2$ come from an AR(1) with parameter 0.9.

Finally we add a fourth scenario:

4. Time-varying dependence structure.

We use a time-varying autoregressive model (TVAR), where coefficients vary linearly from -0.6 to 0.6. Here, we demonstrate that both approaches reject the stationarity hypothesis when the spectrum is time-varying, with high power.

For each scenario we generate a time series, compute its AR spectrogram, and test statistic. We then create 1,000 stationary surrogates, compute their AR spectrograms and test statistics and compare the observed test statistic against the sampling distribution of test statistics. If the observed test statistic is in the tails of the distribution, this gives us evidence against the stationarity hypothesis. Specifically, we use the 95% percentile as the critical value for the one-sided VOCAL tests (i.e., a p-value of < 0.05), and p-value of < 0.05 for the two-sided SOMED test.

The AR spectrograms are generated using a window length of 2^9, and overlap of 75%. We conduct both the VOCAL and the SOMED hypothesis tests, and consider both the KS and the KL variants on the VOCAL test.

We replicate each simulation 1,000 times and report the size or power of each test, at the 5% significance level, where the size of a test is the probability of falsely rejecting the null hypothesis when it is true (or the probability of making a Type I Error), and the power of a test is the probability of correctly rejecting the null hypothesis when it is false (or one minus the probability of making a Type II Error). Our results are presented in Table I.

TABLE I: Test size (probability of falsely rejecting H_0 when it is true) for Scenario 1, and test power (probability of correctly rejecting H_0 when it is false) for Scenarios 2, 3, and 4.

Scenario	KS	KL	SOMED
1	0.036	0.048	0.046
2	1.000	1.000	1.000
3	0.794	0.739	0.962
4	1.000	1.000	0.999

We see that when $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have the same PSD, all tests have a very small test size and that there is less than a 5% chance of making a Type I error. For the extreme case where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have very different PSDs, all tests give us power 1, which means there is zero chance of making a Type II error. In the case where we have similar but not equivalent PSDs, all tests reject the null hypothesis the majority of the time and the SOMED test works particularly well, which is remarkable considering how similar the $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are. When we have a time-varying PSD,

we again have high power. All of these results give us great confidence that the surrogate tests are performing as required.

F.   LISA Pathfinder

We now demonstrate that our surrogate tests can detect nonstationarities in the clean (Level 3) Δg data from the noise runs of LPF. These data have been corrected for the acceleration coming from centrifugal force, acceleration on the x-axis coming from the spacecraft motion along other degrees of freedom, and spurious acceleration noise from the digital to analog converter of the capacitive actuation and Euler force. Details can be found in the technical note on the LPF data archive [10].

We analyze segments from two separate noise runs. These have the following starting times and lengths:

1. 2016-04-03 14:55:00 UTC for 12 days, 16 hours, 29 minutes, 59.40 seconds. We refer to this data set as the *Glitch Data Set*.

2. 2017-02-13 07:55:00 UTC for 18 days, 13 hours, 59 minutes, 59.40 seconds. We refer to this data set as the *Amplitude Modulation (AM) Data Set*.

The LPF data are originally sampled at a rate of 10 Hz (with sample interval $\Delta_t = 0.1$ s). For the Glitch Data Set, we downsampled the data to 0.2 Hz ($\Delta_t = 5$ s) to obtain a Nyquist frequency of 0.1 Hz (but first Tukey windowing with parameter 0.01, then applying a low-pass Butterworth filter of order 4 and critical frequency 0.1 Hz to avoid aliasing issues). The frequency range of interest for most GW signals detectable by LISA is $[10^{-4}, 10^{-1}]$ Hz. To resolve the lowest frequency in this band, the shortest (base 2) time series we can analyze is $n = 2^{11}$. We therefore split the data into non-overlapping segments of length $n = 2^{11}$ to speed up computations.

It is important to note that in the mean sense of stationarity, once filtered and downsampled, the Glitch Data Set is nonstationary, as there is a trend. We therefore remove this trend piecewise linearly for each non-overlapping segment, and we focus our attention on the question of whether LPF noise is nonstationary in terms of its autocovariance function, or equivalently its PSD.

For the AM Data Set, we take the Level 3 data without any additional preprocessing. We examine the first four hours of this data set.

1.   Glitch Data Set

Here, we analyze the Glitch Data Set for four different cases. These are:

1. The full time series (see Figure 8).

2. A segment with a large glitch at the end of the time series (see Figure 9).

3. A segment with a large glitch not at the end of the time series (see Figure 10).

4. A stationary segment with no glitches present (see Figure 11).

For the following surrogate tests, we compute an AR spectrogram with no overlap and window length 2^9 for Case 1, and 2^7 for Cases 2–4. 1,000 surrogates are then used to generate the sampling distribution of the test statistics.

The full downsampled, filtered, and piecewise linear detrended data can be seen in Figure 8. This data set is full of transient, high amplitude "glitches".
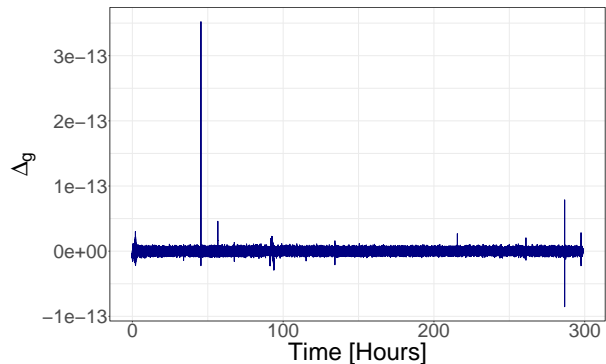

FIG. 8: Δg LPF data from the Glitch Data Set.

When considering the full data set, we report a p-value of 0.001 and 0.000 for the KS and KL variants of the VOCAL test respectively, and 0.001 for the SOMED test. These results indicate that all of the surrogate tests provide evidence against the notion of stationarity, which we attribute to the glitches.

Now consider the case where we look at a segment of the data set where the largest glitch is present. We can see in Figure 8 that the largest glitch in the time series is somewhere around 45 hours into data collection (in the 15th segment from preprocessing). We zoom on this segment (of length $n = 2^{11}$) and its neighbouring earlier (14th) segment in Figure 9.
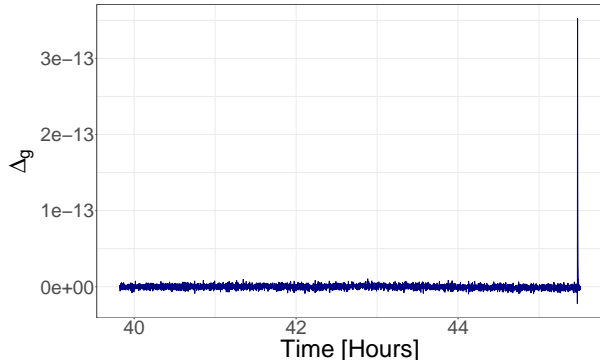

FIG. 9: The 14th and 15th length $n = 2^{11}$ segments from the Glitch Data Set. There is a noticeably large glitch at the end of the displayed time series.

When analyzing the time series in Figure 9, where the glitch is at the end of the time series, we report a p-value of 0.001 for the KS variant of the VO-CAL test, 0.000 for the KL variant of the VOCAL test, and 0.002 for the SOMED test, all providing very strong evidence against the notion of stationarity. We attribute this nonstationarity to the glitch present in the data set.

The glitch at the end of the times series causes naturally a large Euclidean distance for the last interval in comparison to the previous ones in the SOMED test case. This is reflected in the estimated simple regression model. The glitch has a leverage effect in the estimated slope, which results in the rejection of the null hypothesis.

When the large glitch is not at the end of the time series as in Figure 10, the KS and KL variants of the VOCAL test both yield p-values of 0.000, meaning we have very strong evidence against stationarity. However, for the SOMED test, we report a p-value of 0.701, which means we are not rejecting the notion of stationarity here.

Unlike the previous case, the glitch is relatively in the middle of the sequence, which results in a large value in one of the central cumulative medians of the
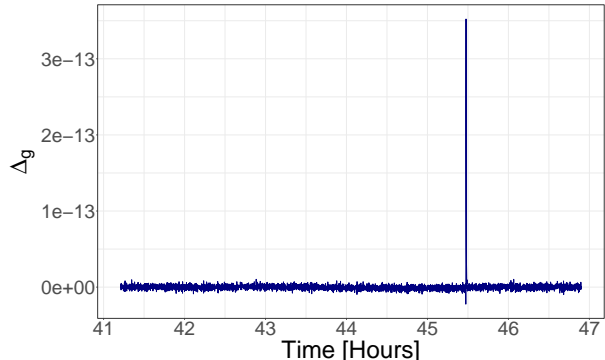

FIG. 10: Same data as in Figure 9 but translated so that the glitch occurs 75% of the way through the time series.

Euclidean distances in the SOMED test case. This large value has a null effect on the estimated slope of the linear model due to its position. Thus, the method fails wrongly to reject the null hypothesis. However, this large value can be visualized via the Cook's distance, a measure of the impact of a single observation in the parameter estimates. In this case, the interval that contains the glitch has a Cook's distance value of 0.39, which is extremely close to the cut point given by the rule of thumb 0.4, and it is quite different from the rest of the Cook's distance values, which have a median of 0.014 and standard deviation of 0.070. Even though the SOMED test fails to reject the stationary hypothesis in this case, the glitch can be detected and thus the validity of the conclusions based on this test can be questioned. This procedure can be applied to other similar situations.

For Case 4 where the data looks stationary, we report the following p-values: 0.836 and 0.198 for the KS and KL variants of the VOCAL test respectively, and 0.702 for the SOMED test. All three do not reject the null hypothesis, meaning we have no evidence against stationarity for this segment of data.

2. AM Data Set

We see cyclostationary behaviour in the LPF data. This is highlighted in the AM Data Set, which is illustrated in Figure 12.
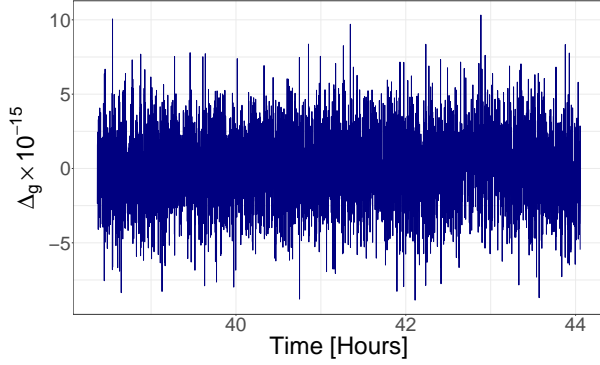
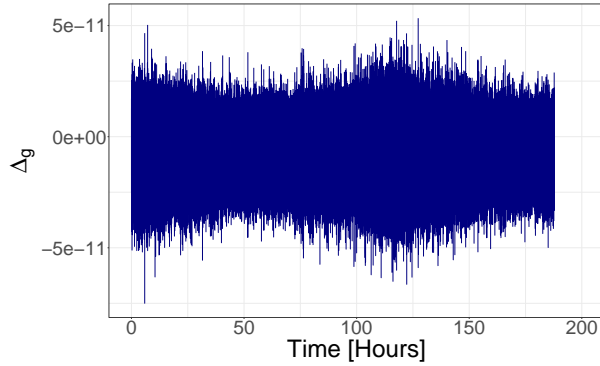FIG. 11: Stationary segment of the Glitch Data Set occurring before the large glitch in Figures 9 and 10.

FIG. 12: Δg data from the AM Data Set.

For all of the surrogate tests, we compute an AR spectrogram with no overlap and window length 2^9. Using 1,000 surrogates to generate the sampling distribution of the test statistics, we report a p-value of 0.008 for the KS variant of the VOCAL test, 0.000 for the KL variant of the VOCAL test, and 0.000 for the SOMED test, all providing very strong evidence against the notion of stationarity.

III. ADDRESSING NONSTATIONARY NOISE

Once we know how often to update the LISA noise PSD (using the hypothesis tests defined in Section II C and Section II D, or similar), we can develop a LISA data analysis strategy. Here we describe a

parameter estimation routine for one non-chirping galactic binary GW signal, where we simultaneously estimate signal parameters and update the LISA noise PSD over time to take into account the time-varying nature of the noise. We include a planned gap in the data stream and use different noise structures before and after the gap to mimic what we expect to happen to LISA noise due to antenna re-pointing. We also highlight the impact that nonstationary noise has on parameter estimates when noise is assumed to be stationary.

A. Galactic White Dwarf Binary Gravitational Wave Signal Model

We assume the low frequency approximation to the LISA response as described by Carré and Porter [19]. We define the GW strain in one TDI channel as

$$h(t) = h_+(t)F^+(t) + h_\times(t)F^\times(t),$$

where the GW polarisations are defined as

$$h_+(t) = A_0 \left(1 + \cos^2 \iota\right) \cos\left(\Phi\left(t\right) + \varphi_0\right),$$
$$h_\times(t) = -2A_0 \cos \iota \sin\left(\Phi\left(t\right) + \varphi_0\right),$$

for a non-chirping galactic white dwarf binary. Here, A_0 is the amplitude, ι is the inclination angle between the orbital plane of the source and the observer, φ_0 is the initial phase, and $\Phi(t)$ is the time-dependent phase, which for a circular orbit, is defined as

$$\Phi(t) = 2\pi\omega_0 \left(t + R_\oplus \sin\theta \cos\left(2\pi\omega_m t - \phi\right)\right),$$

where ω_0 is the monochromatic frequency, ω_m is the LISA modulation frequency (defined as the reciprocal of the number of seconds in a year), $R_\oplus$ is the time light takes to travel one astronomical unit, and (θ, ϕ) is the sky location of the source.

Using the definitions of Rubbo et al [41], the antenna beam factors are

$$F^+(t) = \frac{1}{2} \left(\cos\left(2\psi\right) D^+\left(t\right) - \sin\left(2\psi\right) D^\times\left(t\right)\right),$$
$$F^\times(t) = \frac{1}{2} \left(\sin\left(2\psi\right) D^+\left(t\right) + \cos\left(2\psi\right) D^\times\left(t\right)\right),$$

where

$$D^+(t) = \frac{\sqrt{3}}{64}\bigg(-36\sin^2(\theta)\sin(2\alpha(t) - 2\lambda)$$

$$+ (3 + \cos(2\theta))\bigg(\cos(2\phi)\bigg(9\sin(2\lambda)$$

$$-\sin(4\alpha(t) - 2\lambda)\bigg) + 2\sin(2\phi)$$

$$\bigg(\cos(4\alpha(t) - 2\lambda) - 9\cos(2\lambda)\bigg)\bigg)$$

$$-4\sqrt{3}\sin(2\theta)\bigg(\sin(3\alpha(t) - 2\lambda - \phi)$$

$$-3\sin(\alpha(t) - 2\lambda + \phi)\bigg)\bigg),$$

$$D^\times(t) = \frac{1}{16}\bigg(\sqrt{3}\cos(\theta)\bigg(9\cos(2\lambda - 2\phi)$$

$$-\cos(4\alpha(t) - 2\lambda - 2\phi)\bigg)$$

$$-6\sin(\theta)\bigg(\cos(3\alpha(t) - 2\lambda - \phi)$$

$$+3\cos(\alpha(t) - 2\lambda + \phi)\bigg)\bigg),$$

and $\alpha(t) = 2\pi\frac{t}{T} + \kappa$ is the orbital phase of the centre of mass of the constellation, where T is the number of seconds in a year (though in this study, we increase the orbital modulation so that T is the number of seconds in a day for computational reasons), and $\kappa = 0$ is the initial ecliptic longitude.

The parameters we are interested in estimating are amplitude A_0, monochromatic frequency ω_0, initial phase φ_0, and inclination ι. All other parameters, e.g., sky location (θ, ϕ), GW polarization angle ψ, and initial ecliptic longitude κ, are fixed. To this end, we place the following noninformative priors on the signal parameters:

$$A_0 \sim \text{Uniform}[0, \infty),$$
$$\cos\varphi_0 \sim \text{Uniform}[-1, 1],$$
$$\cos\iota \sim \text{Uniform}[-1, 1],$$
$$\omega_0 \sim \text{Uniform}[0.0001, 0.0191].$$

Although data will eventually be analyzed in the three TDI channels A, E, and T [48] (where T is the noise-only channel containing no signal information), for simplicity, we will only consider the A channel, meaning we set TDI channel angle $\lambda = 0$.

B. Bayesian Nonparametric Noise Model

To model the noise PSD, we use the Bayesian nonparametric B-spline prior introduced by Edwards et al [24]. The B-spline prior has the following representation as a mixture of B-spline densities:

$$s_r(x; k, \mathbf{w}_k, \boldsymbol{\xi}) = \sum_{j=1}^{k} w_{j,k} b_{j,r}(x; \boldsymbol{\xi}),$$

where $b_{j,r}(.)$ is the jth B-spline density of fixed degree r, k is the number of B-spline densities in the mixture, $\mathbf{w}_k = (w_{1,k}, \ldots, w_{k,k})$ is the weight vector, and $\boldsymbol{\xi}$ is the nondecreasing knot sequence.

The noise PSD $f(.)$ is then modelled as as follows:

$$f(\pi x) = \tau \times s_r(x; k, G, H), \quad x \in [0, 1],$$

where the mixture weights and knot differences are induced by CDFs G and H respectively, each on $[0, 1]$, and $\tau = \int_0^1 f(\pi x)\mathrm{d}x$ is the normalization constant.

We place the following *a priori* independent priors on the noise PSD model parameters (k, G, H, τ):

$$p(k) \propto \exp\{-\theta k^2\},$$
$$G \sim \text{DP}(G_0, M_G),$$
$$H \sim \text{DP}(H_0, M_H),$$
$$\tau \sim \text{IG}(\alpha, \beta),$$

where DP represents a Dirichlet process, IG is the inverse-gamma distribution, θ is a smoothing coefficient, G_0 and H_0 are base measures, and M_G and M_H are concentration parameters.

Finally, the joint prior is updated by the commonly used Whittle likelihood [49] to yield a pseudo-posterior. For more details, such as implementation, we refer the reader to Edwards et al [24].

This is in essence a blocked Metropolis-within-Gibbs sampler similar to Edwards et al [23], where we sequentially sample the signal parameters given the noise parameters, and then the noise parameters given the signal parameters.

C. Example

Consider the simple case where we have 48 hours of data from the A TDI LISA channel, and there

is one planned outage at 22 hours for a duration of four hours due to antenna repointing. Assume this antenna repointing changes the noise structure. Whether this is realistic is yet to be determined.

We generate a (non-chirping) galactic white-dwarf binary signal with the following parameters to be estimated:

$$A_0 = 1 \times 10^{-21}$$
$$\omega_0 = 0.005$$
$$\varphi_0 = 3\pi/4$$
$$\iota = \pi/2.$$

We fix the sky location $(\theta = \pi/4, \psi = \pi/4)$ and GW polarization angle $\phi = 0$. Let TDI channel angle $\lambda = 0$ as we only consider the A channel. We set the sample interval to $\Delta_t = 10$ s, yielding a Nyquist frequency of $\omega_* = 0.05$ Hz.

The noise for this example is created as follows. Before the gap, an AR(2) process with parameters $(0.9, -0.9)$ and standard deviation 1×10^{-22} is generated. After the gap, an AR(1) process with parameter 0.9 and standard deviation 4×10^{-22} is generated. The increase in the variance of noise and the change in the autocovariance structure during the second half is our attempt at simulating a change in noise structure due to the repointing of antennae. This noise setup yields an overall signal-to-noise ratio (SNR) of $\varrho \approx 50$ (when considering both noise segments).

We add this noise to the generated GW signal and remove the middle four hours of the data to create a gap. We then multiply the data by a Tukey-type window, where we taper off any data to zero where there is a gap, with a chosen taper parameter of $r = 0.1$. Note that this Tukey-type window will be applied to all galactic white-dwarf binary signals proposed during the MCMC algorithm to ensure gaps are in the correct place in the signal model.

A realization of this data setup can be seen in Figure 13.

We implement the noise model in two different ways. In the first approach, we assume the stationary model, where the PSD is assumed to stay constant throughout the entire time series. In the second approach, we do not assume stationarity, which allows us to model the PSD before and after the gap differently if they are in fact different (which they are in this example). We then compare bias and
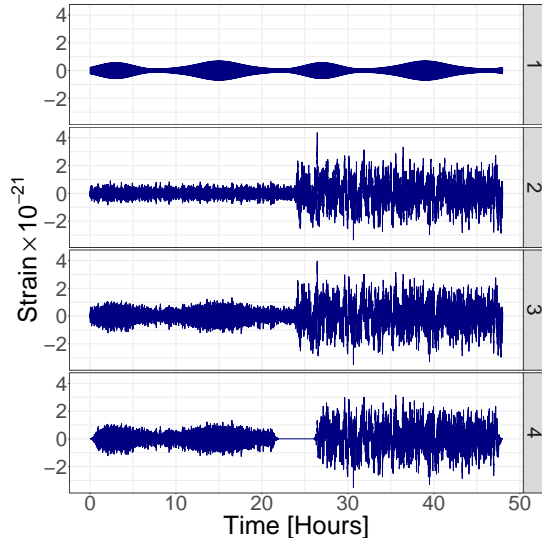

FIG. 13: Panel 1: Non-chirping galactic white-dwarf binary GW signal. Panel 2: AR(2) noise in first half, AR(1) noise in second half. Panel 3: Data (signal plus noise) with four hour gap in the middle. Panel 4: Data multiplied by Tukey-type window (with $r = 0.1$).

posterior variance between the two approaches.

D. Results

We run both algorithms for 100,000 iterations, with a burn-in of 50,000 and thinning factor of 5. We also use an adaptive proposal for each signal parameter described by Roberts and Rosenthal [37]. That is, for each parameter, we use a standard Metropolis step with Normal proposal centred on the previous value, and variance that is automatically tuned to achieve a desired acceptance rate of 0.44.

Figure 14 compares the posterior densities of our parameter vector $(A_0, \omega_0, \varphi_0, \iota)$ when we assume the nonstationary model versus the stationary model.

We can see that the bias (posterior median minus truth) and posterior variance are smaller under the nonstationarity assumption. This is summarized in Table II as the mean squared error (MSE), which is computed by:

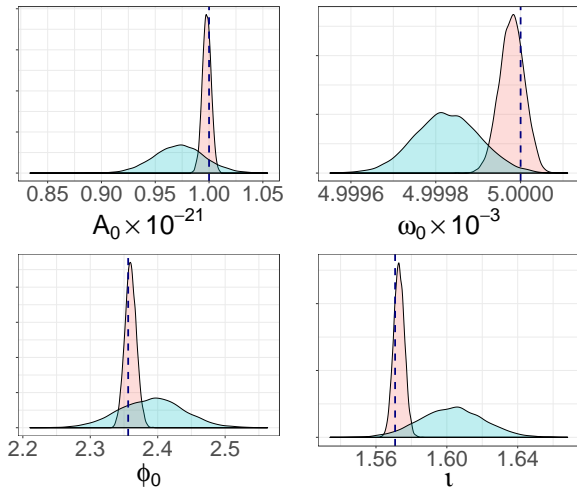$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2,$$

FIG. 14: Posterior densities of parameters of interest using the nonstationary noise assumption (pink) versus the stationary noise assumption (blue).

for each parameter θ.

TABLE II: Mean squared errors for parameters of interest under the stationary and nonstationary noise assumptions.

	Stationary	Nonstationary
A_0	1.314×10^{-45}	2.086×10^{-47}
ω_0	3.633×10^{-14}	1.256×10^{-15}
φ_0	3.527×10^{-3}	8.557×10^{-5}
ι	1.369×10^{-3}	1.424×10^{-5}

We see that when we correctly use a nonstationary noise model, the MSE for our parameter estimates are one or two orders better than when using an incorrect stationary noise model. These results indicate that it is extremely important and necessary to update the noise PSD when noise is nonstationary, otherwise we run the risk of introducing systematic biases into our astrophysical parameter estimates.

IV. DISCUSSION

In this paper, we have discussed methods to identify and address nonstationary noise in the future LISA mission. We demonstrated the usefulness of the lesser-known surrogate tests for assessing the stationarity of a time series, introducing a novel variant in the form of the SOMED test. We applied the surrogate tests to real LPF data and showed that certain segments are nonstationary in nature, due to glitches, and amplitude modulations. As the architecture of LISA will share many similarities to LPF, we see this as an important first step in understanding the stationarity/nonstationarity of LISA data.

We introduced a Bayesian semiparametric framework for conducting parameter estimation when there is nonstationary noise as a result of antenna repointing. We highlighted the risk of assuming a stationary noise model in this situation, as it leads to systematic biases in astrophysical parameter estimates, as well as larger posterior variances.

An interesting alternative framework for modelling noise could be to modify the time-varying spectrum estimation regime of Rosen et al [40], which utilizes reversible jump MCMC [26] to determine the number of locally stationary segments in a time series. One could use a blocked Metropolis-within-Gibbs sampler similar to the one introduced in this paper to model signal parameters given noise parameters and vice versa. This is one avenue we aim to explore in a future paper.

Another future initiative includes investigating the impact of planned data gaps and nonstationary noise on EMRI GW signals, particularly those arising from near-extremal black holes.

in this manuscript.

[1] Aasi J, et al (2013) Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. Physical Review D 88:062,003

[2] Abbott BP, et al (2020) A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals. Class Quant Grav 37(5):055,002, doi:10.1088/1361-6382/ab685e, 1908.11170

[3] Adams MR, Cornish NJ (2010) Discriminating between a Stochastic Gravitational Wave Background and Instrument Noise. Phys Rev D82:022,002, doi: 10.1103/PhysRevD.82.022002, 1002.1291

[4] Adams MR, Cornish NJ (2014) Detecting a Stochastic Gravitational Wave Background in the presence of a Galactic Foreground and Instrument Noise. Phys Rev D89(2):022,001, doi: 10.1103/PhysRevD.89.022001, 1307.4116

[5] Amaro-Seoane P, et al (2017) Laser interferometer space antenna. 1702.00786

[6] Anderson TW, Darling DA (1954) A test of goodness of fit. Journal of the American Statistical Association 49:765–769

[7] Armano M, et al (2016) Sub-Femto-g free fall for space-based gravitational wave observatories: LISA Pathfinder results. Physical Review Letters 116:231,101

[8] Armano M, et al (2016) Sub-femto-g free fall for space-based gravitational wave observatories: Lisa pathfinder results. Phys Rev Lett 116:231,101, doi:10.1103/PhysRevLett.116.231101, URL https://link.aps.org/doi/10.1103/PhysRevLett.116.231101

[9] Armano M, et al (2018) Beyond the required lisa free-fall performance: New lisa pathfinder results down to 20 μHz. Phys Rev Lett 120:061,101, doi:10.1103/PhysRevLett.120.061101, URL https://link.aps.org/doi/10.1103/PhysRevLett.120.061101

[10] Armano M, et al (2019) Delta g release notes. URL http://lpf.esac.esa.int/lpfsa/aio/data-action?ProductType=DOCUMENT&DOCUMENT.DOCUMENT_OID=11850

[11] Armano M, et al (2019) Lisa pathfinder. 1903.08924

[12] Armano M, et al (2019) Lisa pathfinder micronewton cold gas thrusters: In-flight characterization. Phys Rev D 99:122,003, doi: 10.1103/PhysRevD.99.122003, URL https://link.aps.org/doi/10.1103/PhysRevD.99.122003

[13] Baghi Q, Thorpe JI, Slutsky J, Baker J, Dal Canton T, Korsokova N, Karnesis N (2019) Gravitational-wave parameter estimation with gaps in LISA: a Bayesian data augmentation method. Pre-print arXiv:1907.04747 [gr-qc]

[14] Baker J, et al (2019) Space Based Gravitational Wave Astronomy Beyond LISA 1907.11305

[15] Bayram M, Baraniuk R (2000) Multiple window time-varying spectrum estimation. In: Fitzgerald WJ, et al (eds) Nonstationary Signal Processing, Cambridge University Press, Cambridge, pp 292–316

[16] Biscoveanu S, Haster CJ, Vitale S, Davies J (2020) Quantifying the effect of power spectral density uncertainty on gravitational-wave parameter estimation for compact binary sources. arXiv:200405149

[17] Borgnat P, Flandrin P (2009) Stationarization via surrogates. Journal of Statistical Mechanics: Theory and Experiment p P01001

[18] Brockwell PJ, Davis RA (1991) Time Series: Theory and Methods, 2nd edn. Springer, New York

[19] Carré J, Porter EK (2010) The effect of data gaps on LISA galactic binary parameter estimation. Pre-print arXiv:1010.1641 [gr-qc]

[20] Choudhuri N, Ghosal S, Roy A (2004) Bayesian estimation of the spectral density of a time series. Journal of the American Statistical Association 99(468):1050–1059

[21] Chua AJK, Moore CJ, Gair JR (2017) The Fast and the Fiducial: Augmented kludge waveforms for detecting extreme-mass-ratio inspirals. Physical Review D 96:044,005

[22] Dette H, Paparoditis E (2007) Testing equality of spectral densities. Technical Report 2007, 29

[23] Edwards MC, Meyer R, Christensen N (2015) Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis. Physical Review D 92:064,011

[24] Edwards MC, Meyer R, Christensen N (2019) Bayesian nonparametric spectral density estimation using B-spline priors. Statistics and Computing 29:67–78

[25] Efron B (1979) Bootstrap methods: Another look at the Jackknife. The Annals of Statistics 7:1–26

[26] Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732

[27] Kirch C, Edwards MC, Meier A, Meyer R (2018) Beyond Whittle: Nonparametric correction of a

parametric likelihood with a focus on Bayesian time series analysis. Bayesian Analysis Advanced publication:1–37

[28] Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54:159–178

[29] Lamberts A, Blunt S, Littenberg TB, Garrison-Kimmel S, Kupfer T, Sanderson RE (2019) Predicting the LISA white dwarf binary population in the Milky Way with cosmological simulations. Mon Not Roy Astron Soc 490(4):5888–5903, doi: 10.1093/mnras/stz2834, 1907.00014

[30] Maturana-Russel P, Meyer R (2019) Spectral density estimation using P-spline priors. arXiv:190501832 [statME]

[31] Müller UK (2005) Size and power of tests of stationarity in highly autocorrelated time series. Journal of Econometrics 128:195–213

[32] Nason GP, von Sachs R, Kroisandt G (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. Journal of Royal Statistical Society B 6:271–292

[33] Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75:335–346

[34] Pollack SE, Turner MD, Schlamminger S, Hagedorn CA, Gundlach JH (2010) Charge Management for Gravitational Wave Observatories using UV LEDs. Phys Rev D81:021,101, doi: 10.1103/PhysRevD.81.021101, 0912.1769

[35] Priestley MB, Subba Rao T (1969) A test for non-stationarity of time-series. Journal of the Royal Statistical Society Series B (Methodological) 31:140–149

[36] Purdue P, Larson SL (2007) Spurious acceleration noise in spaceborne gravitational wave interferometers. Classical and Quantum Gravity 24(23):5869–5887, doi:10.1088/0264-9381/24/23/010, URL https://doi.org/10.1088%2F0264-9381%2F24%2F23%2F010

[37] Roberts GO, Rosenthal JS (2009) Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18:349–367

[38] Robson T, Cornish NJ (2019) Detecting Gravitational Wave Bursts with LISA in the presence of Instrumental Glitches. Phys Rev D99(2):024,019, doi: 10.1103/PhysRevD.99.024019, 1811.04490

[39] Romano JD, Cornish NJ (2017) Detection methods for stochastic gravitational-wave backgrounds: a unified treatment. Living Reviews in Relativity 20:2, doi:https://doi.org/10.1007/s41114-017-0004-1

[40] Rosen O, Wood S, Roy A (2012) AdaptSpec: Adaptive spectral density estimation for nonstationary time series. Journal of the American Statistical Association 107:1575–1589

[41] Rubbo LJ, Cornish NJ, Poujade O (2004) Forward modeling of space-borne gravitational wave detectors. Physical Review D 69:082,003

[42] von Sachs R, Neumann MH (1998) A wavelet-based test for stationarity. Journal of Time Series Analysis 21 (5):597–613

[43] Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71:599–607

[44] Sesana A, Haardt F, Madau P, Volonteri M (2005) The gravitational wave signal from massive black hole binaries and its contribution to the LISA data stream. The Astrophysical Journal 623:23–30

[45] Swanepoel JWH, Van Wyk JWJ (1986) The comparison of two spectral density functions using the bootstrap. Journal of Statistical Computation and Simulation 24:271–282

[46] Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58:77–94

[47] Thomson DJ (1982) Spectrum estimation and harmonic analysis. Proceedings of the IEEE 70 (9):1055–1096

[48] Tinto M, Dhurandhar SV (2005) Time-delay interferometry. Living Reviews in Relativity 8(4)

[49] Whittle P (1957) Curve and periodogram smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 19:38–63

[50] Xiao J, Borgnat P, Flandrin P (2007) Testing stationarity with time-frequency surrogates. Conference Proceedings EUSIPCO-2007, Poznan, Poland pp 2020–2024