

Mind the gap: addressing data gaps and assessing noise mismodeling in LISA

Ollie Burke,^{1,2,3,4,*} Sylvain Marsat,^{2,†} Jonathan R. Gair,^{3,4} and Michael L. Katz^{3,5}

¹*School of Physics and Astronomy, University of Glasgow, University Avenue, Glasgow, G12 8QQ, UK*

²*Laboratoire des 2 Infinis - Toulouse (L2IT-IN2P3),*

Université de Toulouse, CNRS, UPS, F-31062 Toulouse Cedex 9, France

³*Max Planck Institute for Gravitational Physics (Albert Einstein Institute),
Am Mühlenberg 1, Potsdam-Golm 14476, Germany*

⁴*School of Mathematics, University of Edinburgh, James Clerk Maxwell Building,
Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK*

⁵*NASA Marshall Space Flight Center, Huntsville, Alabama 35811, USA*

Due to the sheer complexity of the Laser Interferometer Space Antenna (LISA) space mission, data gaps arising from instrumental irregularities and/or scheduled maintenance are unavoidable. Focusing on merger-dominated massive black hole binary signals, we test the appropriateness of the Whittle-likelihood on gapped data in a variety of cases. From first principles, we derive the likelihood valid for gapped data in both the time and frequency domains. Cheap-to-evaluate proxies to p-p plots are derived based on a Fisher-based formalism, and verified through Bayesian techniques. Our tools allow to predict the altered variance in the parameter estimates that arises from noise mismodeling, as well as the information loss represented by the broadening of the posteriors. The result of noise mismodeling with gaps is sensitive to the characteristics of the noise model, with strong low-frequency (red) noise and strong high-frequency (blue) noise giving statistically significant fluctuations in recovered parameters. We demonstrate that the introduction of a tapering window reduces statistical inconsistency errors, at the cost of less precise parameter estimates. We also show that the assumption of independence between inter-gap segments appears to be a fair approximation even if the data set is inherently coherent. However, if one instead assumes fictitious correlations in the data stream, when the data segments are actually independent, then the resultant parameter recoveries could be inconsistent with the true parameters. The theoretical and numerical practices that are presented in this work could readily be incorporated into global-fit pipelines operating on gapped data.

CONTENTS

I. Introduction	2	V. Waveform and noise models	18
A. Conventions	4	A. MBHB signals and analysis setup	18
II. The noise process	5	B. LISA response and TDI	20
A. Gaussian noise	5	C. Noise models	21
B. Stationary noise	6	VI. Analysis of the noise covariance matrices	22
III. Direct modeling of data gaps	7	A. Circulant and Toeplitz processes	22
A. Analysis set-up	7	B. Noise covariance matrices with gaps	22
B. The windowing procedure	8	C. Pseudo-inverses with gated functions	24
C. Computing the windowed covariance	10	D. Pseudo-inverses with tapering functions	25
D. Signal-to-noise ratio	12	VII. Results	26
E. Fisher matrices	12	A. Mismodeling Whittle on gapped data	27
IV. Quantifying noise mismodeling errors	14	B. Parameter estimation examples	30
A. Effects of mismodeling the noise process	14	C. Influence of the tapering length	36
B. Coherent Whittle-likelihood	15	D. Influence of the cadence	36
C. Segmented Whittle-likelihood	16	E. Mismodeling data segment independence	38
D. Mismodeling measures	17	F. Mismodeling the noise PSD at low frequencies	40
E. Computing the parameter bias covariance	18	G. Mismodeling the stationary process as circulant	41
		VIII. Conclusions and outlook	42
		A. Conclusions	42
		B. Future work and outlook	43
		IX. Acknowledgements	44

* ollie.burke@glasgow.ac.uk

† sylvain.marsat@l2it.in2p3.fr

A. Discrete Fourier transform of the time-domain covariance	44
1. The circulant case	44
2. The Toeplitz case	45
B. Fourier-domain noise generation in the circulant case	46
C. Likelihood with gaps using a permutation and masks	47
D. Degenerate gated covariance matrices in time and frequency domain	48
E. Details on the tapering window	48
References	48

I. INTRODUCTION

On the 20th of June, 2017, the space-based gravitational wave (GW) observatory, the Laser Interferometer Space Antenna (LISA) [1], was selected as the third large-class mission under the European Space Agency’s (ESA) Cosmic Vision program [1]. The LISA space-mission was then adopted by ESA on the 25th of January, 2024, with launch expected to be in ~ 2034 [2]. The goal of LISA is to observe GWs in the rich $f \in [0.1 \text{ mHz}, 0.1 \text{ Hz}]$ frequency range. GWs at these frequencies are generated by massive black hole binaries (MBHBs) [3], extreme mass-ratio inspirals (EMRIs) [4, 5], ultra compact binaries (UCBs) [6] and stochastic backgrounds [7–13]. In contrast to current ground-based detectors, the LISA instrument is expected to be signal dominated, resulting in a strong cocktail of millions of GW sources of multiple source types all overlapping in both time and frequency. In order to develop and test pipelines for successful source extraction from LISA data, the community have developed “LISA Data Challenges” of increasing complexity [14–16]. Recently, there have been major advancements in both computing and LISA data analysis, that have given rise to proposed solutions to the so-called “global-fit” problem, defined to be the search and characterization of all resolvable GWs buried within the LISA data stream [17–20]. Although impressive – marking a major milestone in LISA data analysis – so far none of these group’s proposed global-fit solutions were conducted with *realistic* LISA data. Indeed, it was assumed that the underlying noise statistics were Gaussian and (weakly-)stationary, facilitating use of rapid-to-evaluate likelihood models in parameter-inference. In reality, however, it is well known by the community that realistic LISA data will contain noise-transients such as instrumental artifacts (glitches), and be interrupted

via data gaps [21, 22]. The LISA Pathfinder mission, which demonstrated some of the LISA technology, exhibited such non-stationary features [23]. The *Spritz* data challenge [24] was built by the community to test source extraction pipelines, with an underlying Gaussian noise model with unknown spectral density with data corrupted via instrumental artifacts in the form of gaps and glitches. To date and to our knowledge, there has only been a single attempt at the *Spritz* data challenge given by [25, 26]. In their analysis, they assumed parametric models to detect and subtract glitches while treating gaps in the data using windowing functions. As it stands, the windowing procedure will induce non-stationary features in the underlying noise properties. This renders usual statistical models utilized for inference of data statistically inconsistent with the data stream itself, and it is therefore important to assess the impact of this mismodeling.

Probabilistic models that describe the data stream must be consistent with the data generating process itself. Since the noise is the only probabilistic quantity that defines the data set, the likelihood used to analyze the data is determined by the model assumed for the underlying noise process. It is often assumed within GW astronomy that the noise process is both Gaussian and stationary, the latter feature giving rise to a *Toeplitz* structure in the time-domain (TD) covariance matrix. An additional simplifying assumption is further placed on the noise process where not only is it Toeplitz, but also *circulant* – infinite duration or periodic if finite. As demonstrated by Whittle [27], in the stationary and circulant assumptions, the TD noise-covariance matrix can be diagonalized in a Fourier basis, resulting in a likelihood that can be calculated in $\mathcal{O}(N \log_2 N)$ operations. The Whittle likelihood is ubiquitous in GW astronomy. Since parameter estimation (PE) schemes usually require many evaluations of a likelihood function, it is clear why one would want to conduct analysis in the frequency-domain (FD) assuming these (restrictive) conditions are met.

In practice, the assumption that the noise process is circulant is clearly unrealistic for true LISA data: no real data stream is periodic and the LISA dataset will be signal-rich on a finite duration. Attempts to perform TD GW data analyses for stationary non-circulant noise are scarce. One example is given by recent analyses of black hole ring down signals [28–32]. In the latter case, the short-data segments containing the ring down signal would suffer from edge effects when analyzed in the frequency-domain (FD), which could contaminate the final results through spectral leakage [33]. Another related problem is given by premerger analyses for MBHB signals, for which a generalization of whitening and matched filtering has been proposed in [34]. In general, TD analyses conducted on large scale data sets are usually computationally infeasible, simply due to the

computational cost of computing the likelihood, which at best is an $\mathcal{O}(N^2)$ operation, since the covariance matrix entering the likelihood is not diagonal.

In the construction of LISA, instrumentalists will try to build an instrument such that its noise properties remain as close to stationary as possible. However, the instrument will not be perfect and we must be ready to account for non-stationary features of the noise when they are present. Examples could be environmental: such as impacts from micro-meteorites that collide with the space-crafts, that will form non-GW induced perturbations to the test masses; or instrumental: failure to discharge the on-board test masses due to charged particles hitting the test-masses. Such artifacts were observed in the LISA pathfinder mission [22, 35, 36]. These noise-transients are called glitches, which, if not accounted for, could feature in PE schemes as biases to recovered parameters [37]. Such loud noise-transients could be subtracted from the data stream [38–40], or simply *masked*, rendering the glitched data segment unusable.

In this paper, we will focus on instrumental artifacts that result in total losses of data, known as *gaps* in the data stream. Alongside gaps introduced by glitch-masking, LISA will experience instrumental gaps falling into two categories – *scheduled* and *unscheduled* gaps. Scheduled gaps are breaks in the data stream where routine maintenance is performed in order to achieve maximal sensitivity of the instrument. Examples being antenna re-pointing (gap duration ~ 3 hours every ~ 14 days), tilt-to-length coupling constant estimation (gap duration ~ 2 days with frequency $\sim$ four times per year) and point-ahead angle mechanism (PAAM) adjustments (three times per day, lasting ~ 100 seconds each). Unscheduled gaps could be more dangerous for LISA data analysis due to their unknown duration and frequency. Examples of unscheduled gaps include instrumental malfunctions, such as collisions between micro-meteorites and the craft that may result in the total loss of the gravitational reference sensor (approximately $\sim 30/\text{yr}$ resulting in ~ 1 days worth of lost information for *each* collision), to minor outages on the instrument, potentially lasting on the order of ~ 10 seconds. The lost data types with gap frequency/duration are listed in [41]. It is clear from this discussion that both the frequency and cumulative duration of gaps is large, and so must be accounted for in PE pipelines.

To date there have been a number of studies that have focused on gaps and non-stationary noise in general. On the topic of gaps, there are two families of studies – the first being data augmentation (gap filling) [42–45] and the other the windowing procedure [39, 46–48]. Data augmentation uses Bayesian (and recently auto-encoder [45]) techniques to *fill in* the data during the gap segment, usually learning the behavior of the noise through prior observations. The parameters of the sig-

nal are then recovered from the filled-in data set, facilitating use of the rapid-to-evaluate Whittle-likelihood assuming that the original noise process was Gaussian, stationary and circulant. However, the procedure used to fill in the gaps is a computationally expensive procedure and may be inaccurate if the instrument suffers significant changes of state as a result of the gap.

Another, albeit simpler approach is the windowing procedure. The windowing procedure does not attempt to impute the data in the gap – a binary mask or smooth window function is applied to the data stream over the gaps. This allows for a coherent data set to be analyzed, at the cost of losing out on the stationary nature of the noise as discussed in [39, 47, 48]. Applying a window function to a stationary noise process violates the Toeplitz condition, rendering the overall noise process *non-stationary* making the Whittle-likelihood inconsistent with data generation scheme. A number of studies have applied the windowing procedure, but have not accounted for the change in the underlying statistics of the noise process. Our work presented here corrects for this, providing methodology on how to build probabilistic models that describe the data stream under a variety of gap scenarios in both the TD and FD.

One of the first studies using the windowing procedure to investigate the impact of gaps on LISA-based parameter precision studies was performed by Carré and Porter in [46]. With focus on UCBs, they simulated the effect of gaps using the windowing procedure. They analyzed the impact of leakage and performed a monte-carlo parameter precision study on a large number of UCBs with gaps. Their results concluded that gaps in the data stream can lead to loss in precision on parameter estimates. An analysis on MBHBs was performed in [49], which studied the effect of gaps on the inspiral/merger phase. They concluded that the impact of gaps grew in severity the closer the gap was to the merger with parameter precision estimates degrading dramatically. In [44, 45], the data augmentation method was applied on MBHBs and reached similar conclusions. In each study [44, 46, 49] gaps were simulated by windowing the data set to zero before and after the gap segment but did not account for the induced correlations between the noise components in the FD. A result of this mis-modeling, as shown in [48], is that not only is the resultant posterior (co)variance incorrect but the (co)variance of the posterior scattering can be under/over-estimated.

We believe that the work here forms the most general treatment of gaps present in the literature to date. Our methodology can be applied to *any* family of gaps with any duration/frequency and with *any* underlying assumptions about the behavior of the noise across the gap, from coherence to independence. We carefully analyze the windowing procedure and demonstrate how inference can be performed on gapped time-series in both the TD and FD, highlighting advantages of each.

Starting from first principles in the TD, we will derive expressions for the signal-to-noise ratio, Fisher matrix and likelihood in the presence of missing data. All of our results are translated into the FD via linear algebra. We will then extend the results from [48] to derive a Fisher-based approximation to calculate: (1) the expectation of the posterior scattering caused by noise mismodeling and (2) the posterior thinning/widening as a consequence of mismodeling.

The paper is organized as follows: In Sec. IA, we will outline conventions and set up notation for our linear algebra formulation of the gaps problem. In Sec. IIA, we will outline the likelihoods for Gaussian noise in both the TD and FD. In Sec. IIB, we outline what it means for a stochastic process to be Toeplitz and circulant, deriving the familiar form of the Whittle-likelihood. Sec. III outlines methodology on how to approach gaps: with the marginalization procedure outlined in Sec. IIIA and the windowing procedure (the two proved to be equivalent) in Sec. IIIB. We outline efficient numerical schemes to compute the TD and FD covariance matrices in the presence of gaps in Sec. IIIC. We then derive expressions for the signal-to-noise ratio and Fisher-matrix in Sec. IIID and Sec. IIIE respectively. Sec. IV outlines the metrics we will use to assess mismodeling for the gapped noise procedure with any model-covariance matrix in either the time or frequency domain. Specifically, in Sec. IVA we derive Fisher-based statistics that are used to assess mismodeling data gaps, with Sec. IVB and Sec. IVC focusing on *assuming coherence* and *incoherence* between gapped segments. Sec. IVD derives expressions that can compare the model posterior variance to the true posterior variance Eq. (92) and to the expectation of the scatter of the noise fluctuations due to noise mismodeling given by Eq. (93). In Sec. IVE, we discuss how the metrics defined by Eqs. (92–93) are computed in practice. We outline our signal generation, gap placement and noise models in Sec. VA. We perform an analysis on circulant, Toeplitz and gated noise covariance matrices in Sec. VIA and Sec. VIB respectively, and in the Sec. VIC and Sec. VID we discuss numerical routines on how to handle the degenerate matrices via computing the Moore-Penrose pseudo-inverse. Our results section VII attempts to answer the following questions:

- [Sec. VIIA]: For various realistic noise models with different behavior at low and high frequencies, is it reasonable to assume the Whittle-likelihood in the presence of gaps?
- [Sec. VIIC]: What is the impact of the tapering on both time and frequency domains? Especially during critical observations such as during the merger phase of a MBHB?
- [Sec. VIID]: Is parameter inference sensitive to

the sampling rate of gapped data?

- [Sec. VII E]: What are the consequences of assuming dependence between gap segments when they are truly independent? And vice versa?
- [Sec. VII F]: Since low frequency noise is challenging to estimate – what is the impact of mismodeling the low-frequency content of the noise process?
- [Sec. VII G]: What are the consequences of assuming that the underlying noise process is circulant (Whittle-like) when the time-series is of finite duration (Toeplitz)?

Since our results are based on a Fisher-matrix approximation, we verify a number of our results using Bayesian techniques in section Sec. VIIB. Our conclusions and scope for future work are presented in Sec. VIIIA and Sec. VIIIB respectively.

We understand that this paper is long and rather technical. For a busy reader, we recommend that they skip straight to a brief summary of our results that can be found in the conclusion Sec. VIIIA. The main results are tabulated in Tabs. IV, VII, XI respectively. We use a color scheme of green as acceptable, and any other color as unacceptable for PE.

A. Conventions

Time-domain signals $x(t) \approx \{x(t_i)\}_{i=0}^{N-1}$ are sampled at discrete times $t_i \in [0, \Delta t, \dots, (N-1)\Delta t]$ with a uniform sampling interval $\Delta t = t_{i+1} - t_i$ and with N the size of the TD data segment, of duration $T = N\Delta t$. To facilitate usage of the fast Fourier transform (FFT), it is customary to use a segment length that is an integer power of two in length, so that $N = 2^J$ for $J \in \mathbb{N}$. We denote the continuous Fourier transform (CFT) through

$$\tilde{x}(f) = \mathcal{F}[x(t)] = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt, \quad (1)$$

with corresponding inverse

$$x(t) = \mathcal{F}^{-1}[\tilde{x}(f)] = \int_{-\infty}^{\infty} \tilde{x}(f) e^{2\pi i f t} df. \quad (2)$$

For real signals, $\tilde{x}(-f) = \tilde{x}(f)^*$, which implies that the negative frequencies are related to the positive frequencies by conjugation.

In the discrete domain, $\mathbf{x}$ will represent the data vector with time interval Δt , so that $\mathbf{x}_i = x(t_i)$, while $\tilde{\mathbf{x}}$ will represent the FD data with frequency interval $\Delta f = 1/T = 1/(N\Delta t)$. Note that we index the FD data from 0 to $N-1$, with the second half

of the vector corresponding to the negative frequencies: for $i = 0, \dots, N/2 - 1$, $\tilde{\mathbf{x}}_i = \tilde{x}(i\Delta f)$, while for $i = N/2, \dots, N - 1$, $\tilde{\mathbf{x}}_i = \tilde{x}((i - N)\Delta f)$.

We will use the discrete time Fourier transform (DFT) for $N \in 2\mathbb{Z}^+$

$$\tilde{\mathbf{x}}_j = \Delta t \sum_{i=0}^{N-1} \mathbf{x}_i e^{-2\pi i f_j t_i}, \quad (3)$$

with the corresponding inverse discrete Fourier transform (IDFT) given by

$$\mathbf{x}_i = \Delta f \sum_{j=0}^{N-1} \tilde{\mathbf{x}}_j e^{2\pi i f_j t_i}. \quad (4)$$

Here, $t_i = i\Delta t$, $f_j = j\Delta f$, and $t_i f_j = ij/N$ since $\Delta t \Delta f = 1/N$.

Another useful point of view on the DFT is to represent it as a linear algebra transformation:

$$\tilde{\mathbf{x}} = \Delta t \sqrt{N} \mathbf{P} \mathbf{x}, \quad (5a)$$

$$\mathbf{x} = \frac{1}{\Delta t \sqrt{N}} \mathbf{P}^\dagger \tilde{\mathbf{x}}, \quad (5b)$$

where we introduced the DFT transition matrix $\mathbf{P}$, built from powers of the N -th root of unity $\omega \equiv e^{2i\pi/N}$:

$$\mathbf{P}_{jk} = \frac{1}{\sqrt{N}} \omega^{-jk}. \quad (6)$$

The transition matrix $\mathbf{P}$ is unitary ($\mathbf{P}^\dagger = \mathbf{P}^{-1}$, where † stands for the hermitian conjugate) and symmetric ($\mathbf{P}^T = \mathbf{P}$). In our notations, $\mathbf{P}$ and $\mathbf{P}^\dagger$ are built to be dimensionless with a symmetric normalization, while $\mathbf{x}$, $\tilde{\mathbf{x}}$ have the appropriate normalization for data in the TD or FD. This linear algebra point of view is useful for derivations; it is never used for practical computations as matrix-vector multiplications cost $\mathcal{O}(N^2)$ operations while the DFT can be computed in $\mathcal{O}(N \log_2 N)$ thanks to the Fast Fourier Transform (FFT) algorithm.

The ensemble average of a continuous (ergodic) process $X(t)$ is denoted by $\Sigma(t, t') = \langle X(t) X^*(t') \rangle$. In a discretized format, we have $\Sigma = \mathbb{E}_{\mathbf{x}}[\mathbf{x} \mathbf{x}^\dagger] \approx \langle \mathbf{x} \mathbf{x}^\dagger \rangle$. Here $\mathbb{E}_{\mathbf{x}}[\cdot]$ represents an expectation with respect to the data generating process that determines $\mathbf{x}$. Parameter sets are denoted $\boldsymbol{\theta}$, with elements $\theta^a \in \boldsymbol{\theta}$ and parameter derivatives given by $\partial_a := \partial/\partial\theta^a$ for $a = 1, \dots, d$, with d the dimension of the parameter space.

II. THE NOISE PROCESS

A. Gaussian noise

The output data stream of a GW detector is assumed to be a superposition of noise $n(t)$ and a collection of

GW signals belonging to different source types. For simplicity, we will assume that there is a single GW signal $h(t; \boldsymbol{\theta}_0)$ with source parameters $\boldsymbol{\theta}_0$ buried in noise $n(t)$

$$d(t) = h(t; \boldsymbol{\theta}_0) + n(t). \quad (7)$$

The data stream (7) contains two key quantities that impact successful extraction and parameter estimation of sources in GW astronomy. The first is a deterministic GW signal, for which accurate (and efficient) model templates can be used to cross-correlate with the data $d(t)$, in order to pick up observational features of the signal. The second is probabilistic detector noise: a culmination of instrumental and environmental fluctuations that perturb the GW detector. Since the noise is the probabilistic feature, it is the noise distribution that determines the likelihood required to perform inference on the signal $h(t; \boldsymbol{\theta})$.

We will always assume that the noise $n(t)$ is a Gaussian time-series, with zero mean and unspecified covariance. Note that this assumption excludes transient disturbances such as instrumental glitches, known to be common in GW detectors, which depart from being Gaussian.

In discretized notation, we will write for TD data vectors of length N

$$\mathbf{d} = \mathbf{h}(\boldsymbol{\theta}_0) + \mathbf{n}, \quad (8)$$

and we will represent the noise as following a multivariate normal distribution $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the TD noise covariance matrix of shape $N \times N$ with

$$\boldsymbol{\Sigma} = \langle \mathbf{n} \mathbf{n}^T \rangle. \quad (9)$$

The likelihood for the data stream given signal parameters $\boldsymbol{\theta}$ is given by $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = p(\mathbf{n} = \mathbf{d} - \mathbf{h}(\boldsymbol{\theta}))$, where $p(\mathbf{n})$ is the Gaussian probability density for the noise, which gives:

$$\ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \ln \det \boldsymbol{\Sigma} - \frac{N}{2} \ln(2\pi), \quad (10)$$

where we use the notation $\mathbf{x} = \mathbf{d} - \mathbf{h}(\boldsymbol{\theta}) = \mathbf{n}$ for the residual. We will ignore the terms after the first in Eq. (10), as they appear as normalization constants when considering the likelihood as a function of $\boldsymbol{\theta}$; they would need to be considered e.g. for noise estimation.

Since the DFT is a unitary transform of the TD data, the FD likelihood preserves the same Gaussian structure. Defining the FD covariance as

$$\tilde{\boldsymbol{\Sigma}} = \langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^\dagger \rangle, \quad (11)$$

from Eq. (5a), one can show that the FD covariance is related to the TD covariance via the DFT matrix $\mathbf{P}$

$$\tilde{\boldsymbol{\Sigma}} = N \Delta t^2 \mathbf{P} \boldsymbol{\Sigma} \mathbf{P}^\dagger. \quad (12)$$

The TD covariance matrix is given by

$$\mathbf{\Sigma} = \frac{1}{N\Delta t^2} \mathbf{P}^\dagger \tilde{\mathbf{\Sigma}} \mathbf{P}. \quad (13)$$

Using Eqs. (5b) and (13), the FD likelihood reads

$$\ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \tilde{\mathbf{x}}^\dagger \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{x}} - \frac{1}{2} \ln \det \left(\frac{\Delta f}{\Delta t} \tilde{\mathbf{\Sigma}} \right) - \frac{N}{2} \ln(2\pi), \quad (14)$$

where we used the unitary character of $\mathbf{P}$ and $N = 1/(\Delta f \Delta t)$ to rewrite the determinant term.

In this section we have restricted ourselves to one data stream but we will generalize our results to the full set of A , E and T time-delay-interferometry (TDI) streams that LISA will produce in Sec. V.

B. Stationary noise

For real ergodic stationary processes, averaging a stochastic process over time through expectation $\mathbb{E}_n$ is equivalent to an ensemble average $\langle \cdot \rangle$. For now, we make the assumption that $n(t)$ is a Gaussian and stationary process implying that neither the mean or variance change over time and the auto-covariance function C_n only depends on the lag τ [50]

$$C_n(\tau) = \langle n(t)n(t+\tau) \rangle. \quad (15)$$

From this auto-covariance, the noise Power Spectral Density (PSD) S_n can be defined as

$$\frac{1}{2} S_n(f) = \int d\tau e^{-2i\pi f\tau} C_n(\tau), \quad (16)$$

with the factor 2 being a matter of convention, for S_n seen as a 1-sided PSD over positive frequencies. The PSD is real and positive: $S_n(f) \geq 0$.

Using the continuous Fourier transform (1), it can be shown that the ensemble average between the noise at two different frequencies f, f' of either sign is related to the one-sided Power Spectral Density (PSD) through [51, 52]

$$\Sigma_n(f, f') = \langle \tilde{n}(f) \tilde{n}^*(f') \rangle = \frac{1}{2} \delta(f - f') S_n(f), \quad (17a)$$

$$R_n(f, f') = \langle \tilde{n}(f) \tilde{n}(f') \rangle = \frac{1}{2} \delta(f + f') S_n(f). \quad (17b)$$

Here Eqs. (17a) and (17b) are a statement that, for positive frequencies $f > 0$ and $f' > 0$, the frequency components among stationary noise components are uncorrelated. Further, $R_n(f, f') = 0$ in that case and the variance of the real and imaginary components of the noise process $\tilde{n}(f)$ are non-zero and equal.

We now turn to the discrete domain in order to derive the usual Whittle-likelihood used commonly throughout

GW astronomy. The direct translation of the stationarity of the noise is that there exists a vector $\boldsymbol{\sigma}$ representing the auto-covariance $C_n(\tau)$, such that

$$\Sigma_{ij} = \sigma_{|i-j|}, \quad |i-j| = 0, \dots, N-1. \quad (18)$$

The covariance matrix above has a *Toeplitz* structure: entries of the covariance are constant along each diagonal of the matrix. The matrix must satisfy additional constraints arising from the fact that $C_n(\tau)$ must be a valid covariance function that can be written as the inverse Fourier transform of a real and positive spectral density.

The Whittle-likelihood is obtained by imposing a further assumption on the covariance, that of a *circulant* structure:

$$\sigma_{N-i} = \sigma_i, \quad i = 1, \dots, N-1. \quad (19)$$

This means that the covariance entries are not only constants along diagonals, but now these diagonals wrap around modulo N . In terms of physical interpretation, the Toeplitz structure is appropriate for an observation of a much longer time series, while the circulant model assumes that the observed data represents the entirety of the stochastic process, by enforcing periodicity (which is never physical). In practice, one often considers a circulant process defined on a time interval T much longer than the observed data, which is then a snapshot of that circulant process with a Toeplitz covariance. The choice of the circulant length T will affect the lowest frequency $1/T$ accessible to the process, and will in principle change the low-frequency content of the autocorrelation function – although we do not explore this in the present paper.

In App. A1, we demonstrate by direct calculation that (with no summation over repeated indices)

$$(\mathbf{P}\mathbf{\Sigma}\mathbf{P}^\dagger)_{ij} = \sqrt{N} \delta_{ij} (\mathbf{P}\boldsymbol{\sigma})_i, \quad (20)$$

where we have imposed the circulant condition in Eq. (19). One can now relate the PSD $S_n(f)$ to the Fourier transform of the auto-covariance function

$$\frac{1}{2} S_n(f_j) = \sqrt{N} \Delta t (\mathbf{P}\boldsymbol{\sigma})_j = \tilde{\boldsymbol{\sigma}}_j. \quad (21)$$

Allowing one to read off the elements of $\tilde{\boldsymbol{\sigma}}_j$.

$$\tilde{\boldsymbol{\sigma}}_j = \begin{cases} \frac{1}{2} S_n(j\Delta f) & \text{for } j = 0, \dots, N/2 - 1, \\ \frac{1}{2} S_n((j-N)\Delta f) & \text{for } j = N/2, \dots, N-1, \end{cases} \quad (22)$$

with the components of $\tilde{\boldsymbol{\sigma}}_j$ being real and positive, and showing the same symmetry as Eq. (19). The property $\tilde{\boldsymbol{\sigma}}_j \in \mathbb{R}$ is automatic for a circulant structure, while the condition $\tilde{\boldsymbol{\sigma}}_j \geq 0$ can be seen as a condition for $\boldsymbol{\sigma}$ to represent an admissible autocorrelation function,

since the covariance Σ must be positive (as does its FD counterpart $\tilde{\Sigma}$). In practice, we typically build the time-domain σ from the IFT of a positive PSD, not the other way around. Finally, the use of Eqs. (20), (21) and (22), one obtains the familiar FD diagonal matrix valid for both stationary and circulant noise

$$\tilde{\Sigma} = N\Delta t^2(\mathbf{P}\Sigma\mathbf{P}^\dagger) = N\Delta t \text{Diag}(\tilde{\sigma}) = \text{Diag}\left(\frac{S_n}{2\Delta f}\right), \quad (23)$$

where the values of the PSD along the diagonal are organized according to Eq. (22). This is the discrete equivalent to the continuous (and infinite duration) expression Eq. (17a).

Note however that we have to impose the circulant condition to arrive at this diagonal structure, which is never realistic in practice: our data segment is an excerpt from a continuous physical process that, even if stationary, has no reason to be periodic. As discussed in [53], Toeplitz and circulant matrices are asymptotically equivalent, implying that the Toeplitz matrices are asymptotically diagonalizable in a Fourier basis. This means that the Whittle-based likelihood better approximates the exact TD likelihood in the limit of an infinite observing time. We must therefore see the Whittle-likelihood as an approximation in that sense. This *Toeplitz* versus *circulant* problem is usually alleviated by considering data segments that are long enough to encompass the relevant signal with some margin, and by applying a smooth taper at each end of the data, enforcing periodicity of the time-segment. We will discuss circulant and Toeplitz processes in more details in Sec. VI, and investigate mis-modeling of such processes in the results Sec. VIIG.

From the diagonal structure given by Eq. (23), we can now derive the usual Whittle log-likelihood used commonly throughout GW astronomy. The matrix $\tilde{\Sigma}^{-1}$ is diagonal and readily inverted, and we can gather positive ($j = 1, \dots, N/2 - 1$) and negative ($j = N/2 + 1, \dots, N - 1$) frequency contributions. The FD likelihood (14) becomes, with $S_n^j \equiv S_n(j\Delta f)$,

$$\begin{aligned} \ln \mathcal{L}(\mathbf{d}|\theta) = & -2\Delta f \sum_{j=1}^{N/2-1} \frac{|\tilde{h}_j(\theta) - \tilde{\mathbf{d}}_j|^2}{S_n^j} - \sum_{j=1}^{N/2-1} \ln \frac{S_n^j}{\Delta t} \\ & - \frac{\Delta f}{S_n^0} |\tilde{h}_0(\theta) - \tilde{\mathbf{d}}_0|^2 - \frac{\Delta f}{S_n^{N/2}} |\tilde{h}_{N/2}(\theta) - \tilde{\mathbf{d}}_{N/2}|^2 \\ & - \frac{1}{2} \ln \frac{S_n^0}{\Delta t} - \frac{1}{2} \ln \frac{S_n^{N/2}}{\Delta t} - \frac{N}{2} \ln 2\pi. \end{aligned} \quad (24)$$

This is the Whittle-likelihood in discretized form, where we kept all the additive constants in $\ln \mathcal{L}$. When performing a Bayesian PE of the source parameters with a known PSD, the terms depending only on S_n are normalization constants that can be ignored. Note also

that the terms at $j = 0$ and $j = N/2$ are not relevant for detectors having a finite frequency range of sensitivity. This expression (24) is efficient to compute, with $\mathcal{O}(N)$ operations for the sum and $\mathcal{O}(N \log_2 N)$ for the FFT.

In continuous notations, introducing the usual noise-weighted inner product

$$(a|b) = 4\text{Re} \int_{f>0} df \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}, \quad (25)$$

the result (24) above is the discretized version of

$$\ln \mathcal{L}_{\text{Whittle}}(\mathbf{d}|\theta) = -\frac{1}{2}(h(\theta) - \mathbf{d}|h(\theta) - \mathbf{d}), \quad (26)$$

up to an additive constant incorporating PSD-dependent terms as an overall normalization.

The Whittle-likelihood (24) is suitable for parameter inference provided the noise process is both Gaussian and circulant. As we will see, the impact of gaps destroys the circulant (and Toeplitz) nature of the TD noise covariance matrix, making the effective overall noise process as non-stationary.

III. DIRECT MODELING OF DATA GAPS

A. Analysis set-up

We will now present the *direct* approach to data gaps, where we marginalize over the missing data, while retaining statistical consistency. We will assume that we know the noise covariance of the underlying stochastic process representing the instrumental noise, and we treat the missing data entries as random variables that will be marginalized over, according to the joint probability distribution for the full process. We will also relate this *direct* approach to the *windowing* approach. We recall that, throughout this paper, we keep the assumption that the noise process is Gaussian.

Representing gaps can be done most easily in the TD and in the language of linear algebra. We present in App. C a derivation for a generic gap configuration, but for simplicity we will consider here a setting with a single gap inside the data set. For illustrative purposes, consider such a data set $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$ of length $N = N_0 + N_1 + N_2$ with corresponding covariance matrix in block matrix form

$$\Sigma = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} & \Sigma_{02} \\ \Sigma_{01}^T & \Sigma_{11} & \Sigma_{12} \\ \Sigma_{02}^T & \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}. \quad (27)$$

Now consider a gap in the data set, which we will represent by setting $\mathbf{x}_1 = 0$, with observed data given by $\mathbf{x}_0$ and $\mathbf{x}_2$. In the *direct* approach to data gaps, we

marginalize over the missing data. The marginal joint probability distribution for the observed noise $\mathbf{n}_0$ and $\mathbf{n}_2$ is

$$p(\mathbf{n}_0, \mathbf{n}_2) = \int d\mathbf{n}_1 p(\mathbf{n}_0, \mathbf{n}_1, \mathbf{n}_2), \quad (28)$$

where the joint probability $p(\mathbf{n}_0, \mathbf{n}_1, \mathbf{n}_2)$ is described by the full covariance Σ . Marginalizing a multidimensional Gaussian distribution is done by directly truncating its covariance, resulting in a $N_0 + N_2$ dimensional matrix

$$\Sigma^{\text{marg}} = \begin{pmatrix} \Sigma_{00} & \Sigma_{02} \\ \Sigma_{02}^T & \Sigma_{22} \end{pmatrix}, \quad (29)$$

and the likelihood therefore features the inverse of the $(N_0 + N_2) \times (N_0 + N_2)$ matrix above.

Assuming that Σ_{00} and the Shur complement of Σ_{00} , i.e., $\mathbf{S} = (\Sigma_{22} - \Sigma_{02}^T \Sigma_{00}^{-1} \Sigma_{02})$ are invertible, the usual 2×2 block-inverse for a symmetric matrix reads [54]

$$(\Sigma^{\text{marg}})^{-1} = \begin{pmatrix} \mathbf{T}_{00} & \mathbf{T}_{02} \\ \mathbf{T}_{02}^T & \mathbf{T}_{22} \end{pmatrix}, \quad (30a)$$

$$\mathbf{T}_{00} = \Sigma_{00}^{-1} + \Sigma_{00}^{-1} \Sigma_{02} \mathbf{S}^{-1} \Sigma_{02}^T \Sigma_{00}^{-1}, \quad (30b)$$

$$\mathbf{T}_{02} = -\Sigma_{00}^{-1} \Sigma_{02} \mathbf{S}^{-1}, \quad (30c)$$

$$\mathbf{T}_{22} = \mathbf{S}^{-1}, \quad (30d)$$

which can be used to obtain

$$\ln \mathcal{L}_{\text{gap}}(\mathbf{x}|\boldsymbol{\theta}) = -\frac{1}{2} \mathbf{x}_0^T \mathbf{T}_{00} \mathbf{x}_0 - \mathbf{x}_0^T \mathbf{T}_{02} \mathbf{x}_2 \quad (31)$$

$$- \frac{1}{2} \mathbf{x}_2^T \mathbf{T}_{22} \mathbf{x}_2 + \text{const.}, \quad (32)$$

where as usual we use $\mathbf{x}_j = \mathbf{d}_j - \mathbf{h}_j(\boldsymbol{\theta})$. We see that this direct approach is straightforward in the TD for Gaussian noise: we simply have to replace the full covariance with a sub-covariance. However, we still need to draw a connection to the FD which requires N -sized data segment, and to the windowing approach.

B. The windowing procedure

Introducing a window (or gating) function $w(t)$ can be represented by an $N \times N$ diagonal matrix $\mathbf{W} = \text{Diag}(w(i\Delta t), i = 0, \dots, N-1)$. In the case where $w(t)$ is everywhere non-zero, $\mathbf{W}$ is invertible. This is the case where w is a simple deterministic modulation, with no loss of information. Assuming the data analyst is given the modulated output $\mathbf{W}\mathbf{x}$, they can still rewrite the likelihood using

$$\mathbf{x}^T \Sigma^{-1} \mathbf{x} = (\mathbf{W}\mathbf{x})^T (\mathbf{W}\Sigma\mathbf{W})^{-1} (\mathbf{W}\mathbf{x}) \quad (\text{if } w \neq 0).$$

Note that the covariance for the modulated data $\mathbf{W}\Sigma\mathbf{W}$ will lose its Toeplitz structure in general; even though

the underlying $\mathbf{x}$ is drawn from a stationary process, such modulated data is manifestly not stationary anymore – but a modulation can be undone through inference schemes. For studies on non-stationary noise given modulated noise processes, we refer the reader to [48] for further discussion.

Data gaps cannot be undone as they erase information, and they can be represented by setting to zero the corresponding entries in $\mathbf{W}$ such that $t \in \text{gap}$. While we could have w implemented via a smooth taper around the gaps with $w(t) \in [0, 1]$, here we will simply consider a gating¹ window of the form

$$\mathbf{W} = \begin{pmatrix} \mathbb{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{1} \end{pmatrix}, \quad (33)$$

and we have

$$\mathbf{W}\Sigma\mathbf{W} = \begin{pmatrix} \Sigma_{00} & \mathbf{0} & \Sigma_{02} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Sigma_{02}^T & \mathbf{0} & \Sigma_{22} \end{pmatrix}, \quad (34)$$

which is non-invertible, representing the erasure of information by the introduction of gaps.

However, the notion of Moore-Penrose pseudo-inverse [55, 56] will allow us to find a direct connection to the result in Eq. (28) for the likelihood marginalized over missing data values. For any matrix $\mathbf{A}$ (with no prerequisites, $\mathbf{A}$ can even be non-square), there exists a unique pseudo-inverse $\mathbf{A}^+$ satisfying the four properties

$$\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+, \quad (35a)$$

$$\mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A}, \quad (35b)$$

$$(\mathbf{A} \mathbf{A}^+)^{\dagger} = \mathbf{A} \mathbf{A}^+, \quad (35c)$$

$$(\mathbf{A}^+ \mathbf{A})^{\dagger} = \mathbf{A}^+ \mathbf{A}. \quad (35d)$$

The pseudo-inverse coincides with the normal inverse when the matrix is invertible. The following properties of the pseudo-inverse, that can be checked from the definition, will be useful for us: if $\mathbf{B}$ is unitary ($\mathbf{B}^{\dagger} \mathbf{B} = \mathbb{1}$), $(\mathbf{A} \mathbf{B})^+ = \mathbf{B}^+ \mathbf{A}^+ = \mathbf{B}^{\dagger} \mathbf{A}^+$ and similarly $(\mathbf{B} \mathbf{A})^+ = \mathbf{A}^+ \mathbf{B}^+ = \mathbf{A}^+ \mathbf{B}^{\dagger}$. Scalar multiplication works as for a normal inverse: $(\alpha \mathbf{A})^+ = \mathbf{A}^+ / \alpha$ for $\alpha \in \mathbb{C}$, $\alpha \neq 0$. Finally, for a block-diagonal matrix, the pseudo-inverse respects the block-diagonal structure. In practice, one general algorithm to compute the pseudo-inverse [57] consists in first computing the singular value decomposition (SVD) of the matrix, identifying the zero

¹ Whenever we refer to a window as a gating window we are referring to a rectangular window where $w(t) = 1$ outside the gap and $w(t) = 0$ inside the gap.

singular values (SVs) and setting their reciprocal to zero when approximating the matrix inverse. This will be discussed in more detail in Sec. VIC.

For the 3×3 block structure that we have, the pseudo-inverse can be calculated directly through repeated application of the 2×2 block inverse formula, or checked a posteriori, resulting in

$$(\mathbf{W}\Sigma\mathbf{W})^+ = \begin{pmatrix} \mathbf{T}_{00} & \mathbf{0} & \mathbf{T}_{02} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{02}^T & \mathbf{0} & \mathbf{T}_{22} \end{pmatrix}, \quad (36)$$

with block matrices given as blocks of the 2×2 block-inverse in Eq. (30). Notice that the block diagonal structure has been preserved. We can verify that

$$(\mathbf{W}\Sigma\mathbf{W})(\mathbf{W}\Sigma\mathbf{W})^+ = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad (37)$$

showing that $(\mathbf{W}\Sigma\mathbf{W})^+$ is indeed the unique pseudo-inverse of $(\mathbf{W}\Sigma\mathbf{W})$, satisfying all four properties in Eq. (35). The likelihood for gated data in the TD reads

$$\begin{aligned} \ln \mathcal{L}_{\text{gap}} &= (\mathbf{W}\mathbf{x})^T (\mathbf{W}\Sigma\mathbf{W})^+ (\mathbf{W}\mathbf{x}) \\ &= -\frac{1}{2} \mathbf{x}_0^T \mathbf{T}_{00} \mathbf{x}_0 - \mathbf{x}_0^T \mathbf{T}_{02} \mathbf{x}_2 - \frac{1}{2} \mathbf{x}_2^T \mathbf{T}_{22} \mathbf{x}_2, \end{aligned} \quad (38)$$

which is precisely the same form as the TD likelihood Eq. (32) where we marginalized out the data corresponding to the gap segment $\mathbf{x}_1$. This is an important point: it shows that marginalizing the data (via truncating the noise covariance matrix) is equivalent to treating the data coherently while filling the gap segment with zeros. The operation on the space of vectors via $(\mathbf{W}\Sigma\mathbf{W})^+$ or $(\Sigma^{\text{marg}})^{-1}$ is equivalent in all cases.

We arrive at the gated likelihood

$$\ln \mathcal{L}_{\text{gate}}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \mathbf{x}^T (\mathbf{W}\Sigma\mathbf{W})^+ \mathbf{x}, \quad (39a)$$

$$= -\frac{1}{2} (\mathbf{W}\mathbf{x})^T (\mathbf{W}\Sigma\mathbf{W})^+ (\mathbf{W}\mathbf{x}), \quad (39b)$$

with $\mathbf{x} = \mathbf{d} - \mathbf{h}$. In the above, the data set is $\mathbf{d} = (\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2)$ and the model template is $\mathbf{h} = (\mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_2)$; note here that $\mathbf{d}_1, \mathbf{h}_1$ do not contribute to the likelihood. Assuming that the correct pseudo-inverse has been computed, then its presence will *annihilate* any data within the gap segment.

This is a point warranting more discussion. The data stream of LISA may be a collection of processed data products via first-stage L0/L1 data pipelines, which will result in time-ordered data sets [2]. The missing data segments are segments of data that do not exist between these data sets. Glitches (or instrumental/environmental noise transients) may contaminate the data

stream so violently that they may need to be masked and thus removed entirely. Throughout this section we have made the convenient choice to fill in those missing data segments with zeros in order to “connect” data products $\mathbf{d}_0$ and $\mathbf{d}_2$ allowing for a *coherent* data set $\mathbf{d} = (\mathbf{d}_0, \mathbf{0}, \mathbf{d}_2)$ to be analyzed. Understand that this is already placing an assumption that the segments are correlated, which may not be the case in practice. This will be investigated in a later Sec. VII E. We have shown that provided the covariance matrix of the noise accounts for missing-data as zeros, then the content of the data stream is irrelevant during the gated segment. For instance, in terms of the input gated data $\mathbf{W}\mathbf{d} = (\mathbf{d}_0, \mathbf{0}, \mathbf{d}_2)$, then Eq. (39b) reads equivalently:

$$\ln \mathcal{L}_{\text{gate}}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} (\mathbf{W}\mathbf{d} - \mathbf{h})^T (\mathbf{W}\Sigma\mathbf{W})^+ (\mathbf{W}\mathbf{d} - \mathbf{h}), \quad (40)$$

ignoring additive constants. Thus, the content of the data and model template during the gap segment are erased entirely. In reality, the data stream during the gated segment will not exist and could be represented via NaNs or $\mathbf{d}_1 = 0$, the point here is that the choice would not matter. For inference purposes, we could use the full model template, fill the gap segment with zeros or even arbitrary values and always get the same result for the likelihood.

Note that we were assuming a gating window above, with $w(t) = 0$ or $w(t) = 1$. For a more general tapering window with $w(t) \in [0, 1]$ outside the gap, we can repeat the same derivation with

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_2 \end{pmatrix}, \quad (41)$$

$$\mathbf{W}\Sigma\mathbf{W} = \begin{pmatrix} \mathbf{W}_0 \Sigma_{00} \mathbf{W}_0 & \mathbf{0} & \mathbf{W}_0 \Sigma_{02} \mathbf{W}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_2 \Sigma_{02}^T \mathbf{W}_0 & \mathbf{0} & \mathbf{W}_2 \Sigma_{22} \mathbf{W}_2 \end{pmatrix}, \quad (42)$$

with $\mathbf{W}_0$ and $\mathbf{W}_2$ invertible, we would obtain the same result as in Eq. (39b). Note that when the window is not a gate, we need to modify Eq. (40) to also apply the tapering outside of the gaps to the templates $\mathbf{h}$.

We conclude our time-domain prescription of data gaps with a particularly counter-intuitive point about gaps with smooth tapers applied pre and post-gap segment. For the smooth tapering matrix defined in Eq. (41), it's easy to show that

$$(\mathbf{W}\mathbf{W}^+) = (\mathbf{W}^+\mathbf{W}) = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad (43)$$

leading to the observation that

$$\ln \mathcal{L}_{\text{gap}} \propto -\frac{1}{2}(\mathbf{W}\mathbf{x})^T(\mathbf{W}\Sigma\mathbf{W})^+(\mathbf{W}\mathbf{x}) \quad (44)$$

$$= (\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3) \begin{pmatrix} \Sigma_{00} & \mathbf{0} & \Sigma_{02} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Sigma_{02}^T & \mathbf{0} & \Sigma_{22} \end{pmatrix}^+ \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{pmatrix} \quad (45)$$

$$\equiv \ln \mathcal{L}_{\text{gate}}. \quad (46)$$

The above argument proves that there is no loss of statistical consistency between tapered and gated processes. Assuming that the pseudo-inverse has been correctly computed, the tapering scheme will result in *no loss of information* of the signal. From the time-domain perspective, there is little reason to taper your data since the gated and tapered process will result in the same likelihood calculation. However, tapering may prove advantageous in the FD in order to mitigate edge-effects as a result of working with finite-duration time-series.

We will now translate this gap treatment to the FD, using the version of the formalism that uses windowed N -size data. Starting from Eq. (39), using the DFT definitions in Eqs. (5a), we have

$$\mathbf{W}\mathbf{x} = \frac{1}{\sqrt{N\Delta t}} \mathbf{P}^\dagger \widetilde{\mathbf{W}}\mathbf{x}, \quad (47)$$

for $\widetilde{\mathbf{W}}\mathbf{x} = \tilde{\mathbf{W}}\tilde{\mathbf{x}}$. Since $\mathbf{P}$ is unitary, we can absorb it inside the pseudo-inverse as

$$\mathbf{P}(\mathbf{W}\Sigma\mathbf{W})^+ \mathbf{P}^\dagger = (\mathbf{P}\mathbf{W}\Sigma\mathbf{W}\mathbf{P}^\dagger)^+. \quad (48)$$

Using Eq. (12) for the FD covariance matrix, and introducing the notation (note that this is not the diagonal matrix built from the vector $\tilde{\mathbf{w}}_j$)

$$\tilde{\mathbf{W}} = \mathbf{P}\mathbf{W}\mathbf{P}^\dagger, \quad (49)$$

we can further rewrite the above as

$$\frac{1}{N\Delta t^2} (\mathbf{P}\mathbf{W}\Sigma\mathbf{W}\mathbf{P}^\dagger)^+ = (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})^+. \quad (50)$$

We finally obtain the FD likelihood in the gap case, marginalized over missing data,

$$\ln \mathcal{L}_{\text{gap}}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} (\widetilde{\mathbf{W}}\mathbf{x})^\dagger (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})^+ (\widetilde{\mathbf{W}}\mathbf{x}), \quad (51)$$

with $\mathbf{x} = \mathbf{d} - \mathbf{h}(\boldsymbol{\theta})$ and ignoring additive constants.

Coming back to the implementation choice, we have a counter-intuitive freedom in how we choose to fill the missing data values for the template. It is natural to fill in the model template $\mathbf{h}$ with zeros, as the most convenient choice. For particular GW model templates, however, we can have situations where the windowing

would strongly affect the shape of the Fourier transform of the signal. This is already the case for a MBHB, but it would be much worse for a quasi-monochromatic GB signal. If a smooth taper were applied to the TD GB signal, the GBs FD spectrum would no longer be compact. According to our argument, however, we can choose to not apply the window to the template, allowing us to use existing codes producing FD templates with no modification. This implementation would read:

$$\ln \mathcal{L}_{\text{gate}}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} (\widetilde{\mathbf{W}}\mathbf{d} - \tilde{\mathbf{h}})^\dagger (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})^+ (\widetilde{\mathbf{W}}\mathbf{d} - \tilde{\mathbf{h}}), \quad (52)$$

which may be a more favorable likelihood model to compute for models built directly in the FD [58–61]. For smooth tapers, it is necessary to include the window in the model templates $\widetilde{\mathbf{W}}\mathbf{h}$ in Eq.(52).

C. Computing the windowed covariance

If one were to compute the FD covariance with gated data and transform to the TD covariance, then one is making the hidden assumption that the data is just missing observations in the middle of some continuous process. In other words, the two data segments are *dependent* and coherent with one another. In this section, we will provide algorithms that can be used to compute the noise covariance matrix assuming coherence between observed data segments, which translates both in the noise level being assumed to be the same before and after the gap, and in the presence of off-diagonal blocks in the TD covariance.

In the above, we found that the likelihood can be written down in terms of the covariance matrices for the windowed noise process in the time or FD,

$$\begin{aligned} \langle (\mathbf{W}\mathbf{n})(\mathbf{W}\mathbf{n})^T \rangle &= \mathbf{W}\Sigma\mathbf{W}, \\ \langle (\widetilde{\mathbf{W}}\mathbf{n})(\widetilde{\mathbf{W}}\mathbf{n})^\dagger \rangle &= \tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}}. \end{aligned} \quad (53)$$

We will now present how these matrices can be computed in practice.

In the TD, the computation of the windowed covariance is straightforward. In functional notations, we have

$$\langle w(t)n(t)w(t')n(t') \rangle = w(t)w(t')C_n(t, t'), \quad (54)$$

and the translation of the above direct product in discrete notations is, for $i, j = 0, \dots, N-1$,

$$(\mathbf{W}\Sigma\mathbf{W})_{ij} = \mathbf{w}_i \mathbf{w}_j \Sigma_{ij}, \quad (55)$$

where $\mathbf{w}$ is the vector of values for the window.

To build the circulant TD noise covariance matrix, an inverse discrete-Fourier transform of the PSD $S_n(f)$ is

computed, which then forms the first row of the matrix. As the matrix is circulant, one can easily construct the matrix via Eq. (19). Constructing the Toeplitz matrix is slightly more involved, since Eq. (16) is a definition that assumes infinite observation times. In practice, we construct the inverse discrete-Fourier transform of the PSD $S_n(f)$ assuming twice the observation time. Half of the time-array is chosen to be the first row of the matrix and, applying Eq. (18), one can construct the resultant Toeplitz matrix. When gaps are present, one simply truncates the rows and columns corresponding to the time-indices where the gap is present (see Eq. (34)). Note that this process is *not invariant* to the choice of the length of the extended circulant process, of which our Toeplitz process is a snapshot. Choosing more than twice the target length would change the low-frequency content of the autocorrelation function. We did not explore this aspect in the present study.

In the FD, the product with the window becomes a convolution, according to

$$\tilde{w}\tilde{n}(f) = \int du \tilde{w}(u)\tilde{n}(f-u). \quad (56)$$

Using Eq. (17b) for stationary noise, the functional FD covariance becomes [39, 47]

$$\begin{aligned} \Sigma_{wn}(f, f') &= \langle \tilde{w}\tilde{n}(f)\tilde{w}\tilde{n}(f')^* \rangle \\ &= \frac{1}{2} \int du S_n(u)\tilde{w}(f-u)\tilde{w}(f'-u)^*. \end{aligned} \quad (57)$$

In linear algebra notations, this result will translate into a discrete convolution. First, from our definition for $\tilde{W}$,

$$\tilde{W}_{ij} = (\mathbf{P}\mathbf{W}\mathbf{P}^\dagger)_{ij} = \frac{1}{N} \sum_k \omega^{k(i-j)} \tilde{w}_k. \quad (58)$$

For discrete convolutions, it is useful to introduce periodic indices, to be interpreted modulo N . We will denote them with an overline, $\tilde{w}_{\bar{k}} = \tilde{w}_{k \bmod N}$. Using these notation allows to simplify intermediate sums while ignoring edge effects on indices. Using the diagonal structure for the FD covariance given by Eq. (23) in the circulant case, we obtain the convolution

$$(\tilde{W}\tilde{\Sigma}\tilde{W})_{ij} = \frac{\Delta f}{2} \sum_{k=0}^{N-1} S_n^k \tilde{w}_{\bar{i-k}} \tilde{w}_{\bar{j-k}}^*. \quad (59)$$

The above matrix can be computed more efficiently than with the apparent $\mathcal{O}(N^3)$ of this formula, by using the FFT for convolutions. Namely, for each periodic diagonal $j-i=p \bmod N$, if we define the vector $\mathbf{v}_i^{(p)} = \tilde{w}_i \tilde{w}_{i+p}^*$, then the computation for this diagonal can be seen as the computation of the convolution $\sum S_n^k \mathbf{v}_{i-k}^{(p)}$. The latter convolution can be computed by taking the IFFT of the product of the two vector's FFT,

for an overall cost of $\mathcal{O}(N \log_2 N)$. This needs to be done for each diagonal, which gives the overall scaling $\mathcal{O}(N^2 \log_2 N)$ or computing the full matrix².

Computing the pseudo-inverse of that matrix, however, requires computing its SVD in $\mathcal{O}(N^3)$, as we will describe in Sec. VI C. Below in Alg. 1 we provide pseudo-code that can be used to compute the FD noise covariance matrix assuming a Circulant and Gaussian noise process. When building the frequency-domain gated covariance matrix for Toeplitz processes, we found it simpler to construct the time-domain covariance Σ^{toe} and compute $\tilde{\Sigma} = (N\Delta t)^{-2} \mathbf{P}\Sigma^{\text{toe}}\mathbf{P}^\dagger$. Analytical formulae for the FD variate of the Toeplitz matrix are given in App. A 2, Eq. (A14). The TD and FD covariances for both Circulant and Toeplitz processes will be illustrated and discussed in Sec. VI.

Algorithm1 Fourier Domain Noise Covariance

Require: S_n PSD values of length $N/2+1$, $\mathbf{w}$ window array in TD of length N and Δt the sampling interval.
Ensure: $\tilde{\Sigma}$ (Fourier-domain covariance matrix, complex array)
 $N \leftarrow \text{length}(\mathbf{w})$
 $\Delta f \leftarrow 1/(N\Delta t)$
 Compute Fourier transform: $\tilde{\mathbf{w}} \leftarrow \Delta t \cdot \text{FFT}(\mathbf{w})$
 Initialize $\mathbf{0} \leftarrow \mathbf{u} \in \mathbb{C}^{N \times 1}$
 Complex $\tilde{\Sigma} \leftarrow$ Zero Matrix of size $N \times N$.
 $\mathbf{u}[0 : N/2 + 1] \leftarrow S_n$,
 $\mathbf{u}[N/2 + 1 : N] \leftarrow S_n[1 : -1]$ with S_n reversed
for diagonal $\in \{0, \dots, N/2\}$ **do**
 $\mathbf{w}_{\text{shift}} \leftarrow \text{permute}(\mathbf{w}, \text{-diagonal})^*$
 $\mathbf{v} \leftarrow \mathbf{w} \cdot \mathbf{w}_{\text{shift}}$
 $\hat{\mathbf{u}} \leftarrow \text{FFT}(\mathbf{u})$
 $\hat{\mathbf{v}} \leftarrow \text{FFT}(\mathbf{v})$
 $D \leftarrow (\Delta f/2) \cdot \text{IFFT}(\hat{\mathbf{u}} \cdot \hat{\mathbf{v}})$
 Set diagonal of $\tilde{\Sigma} \leftarrow D$
 if diagonal $\neq (N/2)$ **then**
 Assign D^* to mirrored diagonal entries
 end if
end for
return $\tilde{\Sigma}$

² Similarly, given knowledge of the TD noise covariance matrix the act of $\tilde{\Sigma} \propto \mathbf{P}\Sigma\mathbf{P}^\dagger$ is equivalent to taking an FFT on the *columns* of the matrix Σ (pre-multiplying by $\mathbf{P}$) and then an IFFT on the post-matrix rows (equivalent to post-multiplying by $\mathbf{P}^\dagger = \mathbf{P}^{-1}$). The cost of this operation scales like $\mathcal{O}(N^2 \log_2 N)$ for data lengths a power of two. The algorithm in the text computes each diagonal separately, which could be advantageous for band-dominated matrices, should we truncate the matrix to a preset number of diagonals.

D. Signal-to-noise ratio

In this section, we rewrite the usual definitions for the Wiener filtering and Signal-to-Noise Ratio (SNR) in the TD, using the same linear algebra notations as in previous sections, and adapt the derivation to the presence of gaps in the data.

For a given filter $k(t)$, we consider the L^2 inner product of this filter with the data stream $\int dt k(t)d(t)$. We introduce the “signal” and “noise” as the expectation value of the inner product when a signal $h(t; \theta)$ is present in the data, and its standard deviation when the signal is absent (the expectation value of the noise being zero), respectively. The signal-to-noise ratio (SNR) is then

$$\left(\frac{S}{N}\right) = \frac{\int_{-\infty}^{\infty} dt k(t)h(t; \theta)}{\sqrt{\left\langle \left(\int_{-\infty}^{\infty} dt k(t)n(t)\right)^2 \right\rangle}}, \quad (60)$$

and we wish to optimize this SNR by best matching the chosen filter k to the target signal $h(t; \theta)$, assumed to be known.

In linear algebra notations, the filter and waveform templates become vectors $\mathbf{k}, \mathbf{h} \in \mathbb{R}^N$ (we omit the θ -dependency in $\mathbf{h}$) and the SNR is

$$\left(\frac{S}{N}\right) = \frac{\mathbf{k}^T \mathbf{h}}{\sqrt{\left\langle (\mathbf{k}^T \mathbf{n})^2 \right\rangle}} = \frac{\mathbf{k}^T \mathbf{h}}{\sqrt{\mathbf{k}^T \boldsymbol{\Sigma} \mathbf{k}}}, \quad (61)$$

where we used the TD noise covariance matrix in Eq. (9) to rewrite the denominator.

For $\boldsymbol{\Sigma}$ symmetric and positive definite, there exists precisely one positive definite and symmetric square-root matrix³ $\boldsymbol{\Sigma}^{1/2}$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{1/2}$ [62]. Denoting the inverse square root as $\boldsymbol{\Sigma}^{-1/2}$ and using the Cauchy-Schwarz inequality,

$$|\mathbf{k}^T \mathbf{h}| = \left| \left(\boldsymbol{\Sigma}^{1/2} \mathbf{k} \right)^T \left(\boldsymbol{\Sigma}^{-1/2} \mathbf{h} \right) \right| \leq \sqrt{\mathbf{k}^T \boldsymbol{\Sigma} \mathbf{k}} \sqrt{\mathbf{h}^T \boldsymbol{\Sigma}^{-1} \mathbf{h}}, \quad (62)$$

with the bound being saturated for the two vectors being collinear, that is to say for $\boldsymbol{\Sigma}^{1/2} \mathbf{k} \propto \boldsymbol{\Sigma}^{-1/2} \mathbf{h}$, or equivalently for

$$\mathbf{k} \propto \boldsymbol{\Sigma}^{-1} \mathbf{h}. \quad (63)$$

Coming back to the expression of the SNR (61), this gives the optimal square SNR as

$$\rho_{\text{opt}}^2 = \left(\frac{S}{N} \right)_{\text{opt}}^2 = \mathbf{h}^T \boldsymbol{\Sigma}^{-1} \mathbf{h}. \quad (64)$$

³ This would still hold with positive-semi-definite matrices.

Working in the FD rather than in the TD, The expressions Eqs (5b) and (13) allow us to write Eq. (64) in the form

$$\rho_{\text{opt}}^2 = \tilde{\mathbf{h}}^\dagger \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{h}}. \quad (65)$$

For stationary noise (and in the circulant case), $\tilde{\boldsymbol{\Sigma}}$ is diagonal and this inner product becomes a simple sum, as in Eq. (24), which is the discrete equivalent of the usual expression $\rho_{\text{opt}}^2 = (h|h)$ using the functional inner product defined by Eq. (25).

Now, considering the case of data gaps, we can define the observed data stream via $\mathbf{W}\mathbf{h}$ and marginalised TD covariance matrix $\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}$. The SNR is then

$$\left(\frac{S}{N}\right) = \frac{\mathbf{k}^T \mathbf{W}\mathbf{h}}{\sqrt{\left\langle (\mathbf{k}^T \mathbf{W}\mathbf{n})^2 \right\rangle}} = \frac{\mathbf{k}^T \mathbf{W}\mathbf{h}}{\sqrt{\mathbf{k}^T \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{k}}}. \quad (66)$$

As in the case of the likelihood, we can either work with an explicit expression of the pseudo-inverse in a 3×3 block structure as in Sec. III, or use the reordering notations of App. C, and obtain for the optimal SNR

$$(\rho_{\text{opt}}^{\text{gap}})^2 = (\mathbf{W}\mathbf{h})^T (\mathbf{W}\boldsymbol{\Sigma}\mathbf{W})^+ (\mathbf{W}\mathbf{h}). \quad (67)$$

In the above, the factor $\mathbf{W}$ in $\mathbf{W}\mathbf{h}$ is only mandatory if the window is not a pure gating window but includes a smooth tapering on the segments outside the gaps. We remind the reader that the presence of $(\mathbf{W}\boldsymbol{\Sigma}\mathbf{W})^+$ kills all imputed information within the gap segment, implying that we are free to use the non-gated templates for SNR computations. Similarly, the string of equalities leading to the statement in Eq. (46) applies here, showing that the SNR of the gated signal is identical to the SNR of the tapered process. Hence the only portion of the signal that results in information loss is only the segment of zero data – during the gap. This phenomena will be verified in Fig. 12 found in Sec. VID.

We will see in later sections that tapered waveforms with assumed Whittle-based diagonal covariances could significantly impact the SNR of the source, especially when information-rich content of the signal is being erased.

The translation to the FD is the same as for the likelihood, and reads

$$(\rho_{\text{opt}}^{\text{gap}})^2 = (\widetilde{\mathbf{W}\mathbf{h}})^\dagger (\widetilde{\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}})^+ (\widetilde{\mathbf{W}\mathbf{h}}). \quad (68)$$

E. Fisher matrices

The Fisher matrix (FM) formalism is a useful tool for approximate PE, and we will use it extensively in the rest of the paper. We repeat here the derivation of the consequences of the linearized signal approximation, adapted to the case of data with gaps.

We start by writing the linearized signal approximation for the Whittle-likelihood (26) and with the functional overlap (25), for simplicity of notation before adapting to the gap case. We expand the signal at linear order in deviations from the true parameters θ_0 as

$$h(\theta) = h(\theta_0) + \Delta\theta^a \partial_a h(\theta_0) + \mathcal{O}(\Delta\theta^2). \quad (69)$$

Given a noise realization so that $\mathbf{d} = \mathbf{h} + \mathbf{n}$ and in the absence of waveform uncertainties the likelihood becomes a quadratic form [63] in $\Delta\theta$:

$$\begin{aligned} \ln \mathcal{L} = & -\frac{1}{2}(\Delta\theta^a - \Delta\hat{\theta}_{\text{bf}}^a) \Gamma_{ab} (\Delta\theta^b - \Delta\hat{\theta}_{\text{bf}}^b) \\ & -\frac{1}{2}(n|n) - (\partial_a h|n)(\Gamma^{-1})_{ab} (\partial_b h|n) \\ & + \mathcal{O}(\Delta\theta^3) \end{aligned} \quad (70)$$

with quantities

$$\Gamma_{ab} = (\partial_a h|\partial_b h), \quad (71)$$

$$\Delta\hat{\theta}_{\text{bf}}^a = (\Gamma^{-1})_{ab} (\partial_b h|n). \quad (72)$$

Notice that the first term in Eq. (70) is of the same form of a multivariate-Gaussian distribution with parameter covariance matrix given by Γ^{-1} centered on “best-fit” parameters $\Delta\hat{\theta}_{\text{bf}}^a$. Direct calculation⁴ shows that

$$\begin{aligned} \mathbb{E}_{\mathbf{n}} [-\partial_a \partial_b \log p(\theta|d)] &= \langle (\partial_a \partial_b h|n) \rangle + \langle (\partial_a h|\partial_b h) \rangle \\ &= (\partial_a h|\partial_b h) \\ &= \Gamma_{ab} \end{aligned}$$

allowing us to identify the matrix Γ in Eq. (71) as the expectation of the observed information matrix, usually called the FM.

The FM $\Gamma \sim \rho^2$ and $\Delta\hat{\theta}_{\text{bf}}^a \sim \rho^{-1}$, meaning that the first term in (70) is an $\mathcal{O}(\rho)$ quantity whereas the last two terms are $\mathcal{O}(1)$ quantities. This reasoning implies that in the limit of high signal-to-noise ratio, those latter terms can be neglected and the likelihood can be approximated as a multi-variate Gaussian distribution with parameter covariance matrix given by Γ^{-1} and maximum likelihood estimate given by $\Delta\hat{\theta}_{\text{bf}}^a$ [65–67]. These best-fit parameters are unbiased $\langle \Delta\hat{\theta}_{\text{bf}} \rangle = 0$, and with variance given by the Fisher covariance itself,

$$\begin{aligned} \langle \Delta\hat{\theta}_{\text{bf}}^a \Delta\hat{\theta}_{\text{bf}}^b \rangle &= (\Gamma^{-1})_{ac} (\Gamma^{-1})_{bd} \langle (\partial_c h|n)(n|\partial_d h) \rangle \\ &= (\Gamma^{-1})_{ac} (\Gamma^{-1})_{bd} \Gamma_{cd} \\ &= (\Gamma^{-1})_{ab}, \end{aligned} \quad (73)$$

where we used $\langle (a|n)(n|b) \rangle = (a|b)$, easily derived using Eq. (17a). We will refer to these “best-fit” quantities $\Delta\hat{\theta}_{\text{bf}}^a$ as “noise biases”⁵, in contrast to the *Cutler-Vallisneri* biases that arise in the context of waveform systematics [67].

Coming back to the gap case, in linear algebra notations, we start from the likelihood in Eq. (39) and apply the same linearization of the signal around the true parameters. We obtain again a quadratic form, with the new FM

$$\Gamma_{ab}^{\text{gap}} = (\mathbf{W} \partial_a \mathbf{h})^T (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \partial_b \mathbf{h}), \quad (74)$$

and best-fit parameters

$$\Delta\hat{\theta}_{\text{bf,gap}}^a = (\Gamma_{\text{gap}}^{-1})_{ab} (\mathbf{W} \partial_b \mathbf{h})^T (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \mathbf{n}), \quad (75)$$

which provide an unbiased estimator for the parameters with variance

$$\begin{aligned} \langle \Delta\hat{\theta}_{\text{bf,gap}}^a \Delta\hat{\theta}_{\text{bf,gap}}^b \rangle &= (\Gamma_{\text{gap}}^{-1})_{ac} (\Gamma_{\text{gap}}^{-1})_{bd} \langle (\mathbf{W} \partial_c \mathbf{h})^T (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \mathbf{n}) (\mathbf{W} \mathbf{n})^T (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \partial_d \mathbf{h}) \rangle \\ &= (\Gamma_{\text{gap}}^{-1})_{ac} (\Gamma_{\text{gap}}^{-1})_{bd} (\mathbf{W} \partial_c \mathbf{h})^T (\mathbf{W} \Sigma \mathbf{W})^+ \mathbf{W} \Sigma \mathbf{W} (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \partial_d \mathbf{h}) \\ &= (\Gamma_{\text{gap}}^{-1})_{ac} (\Gamma_{\text{gap}}^{-1})_{bd} (\mathbf{W} \partial_c \mathbf{h})^T (\mathbf{W} \Sigma \mathbf{W})^+ (\mathbf{W} \partial_d \mathbf{h}) \\ &= (\Gamma_{\text{gap}}^{-1})_{ac} (\Gamma_{\text{gap}}^{-1})_{bd} \Gamma_{cd}^{\text{gap}} \\ &= (\Gamma_{\text{gap}}^{-1})_{ab}, \end{aligned} \quad (76)$$

⁴ Under specific regularity conditions (see page 64 in [64]), it can be shown that $\mathbb{E}_{\mathbf{n}} [-\partial_a \partial_b \log p(\theta|d)] = \mathbb{E}_{\mathbf{n}} [\partial_a \log p(\theta|d) \partial_b \log p(\theta|d)]$ with latter expression the formal definition of the FM. A short calculation shows that $\Gamma_{ab} = \mathbb{E}_{\mathbf{n}} [\partial_a \log p(\theta|d) \partial_b \log p(\theta|d)] = (\partial_a h|\partial_b h)$.

⁵ The quantities $\Delta\hat{\theta}_{\text{bf}}^a$ are not biases in the statistical sense, since

$\langle \Delta\hat{\theta}_{\text{bf}}^a \rangle = 0$. However, as LISA will observe a single noise realization that in principle will “shift” recovered parameters away from the true parameters, these noise fluctuations represent the “bias” in individual observations.

where we have used $\langle (\mathbf{W}\mathbf{n})(\mathbf{W}\mathbf{n})^T \rangle = \mathbf{W}\Sigma\mathbf{W}$ and one of the defining properties of the pseudo-inverse, $\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+$. The equivalent expressions in the FD are obtained with the same rewriting that we used for the likelihood, using the unitarity of $\mathbf{P}$ to absorb it under the pseudo-inverse. This results in

$$\tilde{\mathbf{\Gamma}}_{ab}^{\text{gap}} = \left(\widetilde{\mathbf{W}\partial_a \mathbf{h}} \right)^\dagger (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})^+ \left(\widetilde{\mathbf{W}\partial_b \mathbf{h}} \right), \quad (77)$$

and

$$\Delta \hat{\theta}_{\text{bf,gap}}^a = \left(\tilde{\mathbf{\Gamma}}_{\text{gap}}^{-1} \right)_{ab} \left(\widetilde{\mathbf{W}\partial_b \mathbf{h}} \right)^\dagger (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})^+ \left(\widetilde{\mathbf{W}\mathbf{n}} \right). \quad (78)$$

IV. QUANTIFYING NOISE MISMODELING ERRORS

Having explored how to write down the correct likelihood, SNR and FM in the presence of data gaps (and possibly of a tapering window), we will now present tools to assess the impact that mismodeling the statistics of the noise could have on PE.

A. Effects of mismodeling the noise process

In all mismodeling cases that we will consider, the approximate model for the likelihood $\mathcal{L}'$ will remain Gaussian-like, in the sense that it will possess the structure

$$\ln \mathcal{L}'(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} (\mathbf{W}\mathbf{x})^T (\Sigma')^+ (\mathbf{W}\mathbf{x}). \quad (79)$$

We have written the model's inverse covariance $(\Sigma')^+$ generically as a pseudo-inverse and in the TD, but in applications we might be using an invertible covariance (e.g. when using the Whittle-likelihood), and we might

be working in the FD. To connect to the correct modeling case, in this notation, this inverse covariance absorbs the effect of windows: $(\Sigma')_{\text{correct}}^+ = (\mathbf{W}\Sigma\mathbf{W})^+$.

Note that we do not consider the case where we apply the window to the data but not the template, as we were able to do in Eq. (52) when using the correct covariance with the gating window. This would complicate notations and also presumably represent an additional source of errors whenever using an approximate inverse covariance; in all applications, we will apply the same window (gating or tapering) to the data stream and to the templates, $\mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{h}(\boldsymbol{\theta}) - \mathbf{W}\mathbf{d}$.

The *wrong* likelihood defined in Eq. (79) does not encode the correct statistics for the windowed noise process. Since it retains the usual quadratic structure, we can repeat the steps leading to the derivation of the FM and noise bias. Using the linearized signal approximation, we obtain the FM

$$\mathbf{\Gamma}'_{ab} = (\mathbf{W}\partial_a \mathbf{h})^T (\Sigma')^+ (\mathbf{W}\partial_b \mathbf{h}), \quad (80)$$

that will differ in general from the correct FM $\mathbf{\Gamma}_{ab}$. We also obtain the following noise bias, for a given noise realization $\mathbf{n}$,

$$\left(\Delta \hat{\theta}_{\text{bf}}^a \right)' = (\mathbf{\Gamma}')_{ab}^{-1} (\mathbf{W}\partial_b \mathbf{h})^T (\Sigma')^+ (\mathbf{W}\mathbf{n}). \quad (81)$$

These best-fit parameters should not be interpreted as giving only the best-fit in L^2 sense of the template $\mathbf{h}(\boldsymbol{\theta})$ to the data $\mathbf{d}$: the model's pseudo-inverse covariance $(\Sigma')^+$ intervenes in the computation. Inaccuracies in this covariance will impact the best-fit solution.

Over many realisations of the noise, the best-fit parameters are still unbiased by construction, with $\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \rangle = 0$, but they will have a variance that does not match the expected parameter covariance represented by the FM and encoded in the posterior distributions,

$$\begin{aligned} \left\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \left(\Delta \hat{\theta}_{\text{bf}}^b \right)' \right\rangle &= (\mathbf{\Gamma}')_{ac}^{-1} (\mathbf{\Gamma}')_{bd}^{-1} \left\langle (\mathbf{W}\partial_c \mathbf{h})^T (\Sigma')^+ (\mathbf{W}\mathbf{n}) (\mathbf{W}\mathbf{n})^T (\Sigma')^+ (\mathbf{W}\partial_d \mathbf{h}) \right\rangle \\ &= (\mathbf{\Gamma}')_{ac}^{-1} (\mathbf{\Gamma}')_{bd}^{-1} (\mathbf{W}\partial_c \mathbf{h})^T (\Sigma')^+ \mathbf{W}\Sigma\mathbf{W} (\Sigma')^+ (\mathbf{W}\partial_d \mathbf{h}), \end{aligned} \quad (82)$$

where the inner matrix products do not simplify further, in general $(\Sigma')^+ \mathbf{W}\Sigma\mathbf{W} (\Sigma')^+ \neq (\mathbf{W}\Sigma\mathbf{W})^+$, so that this *parameter bias covariance* will not match the model's Fisher covariance (nor the correct Fisher covariance):

$$\left\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \left(\Delta \hat{\theta}_{\text{bf}}^b \right)' \right\rangle \neq (\mathbf{\Gamma}')_{ab}^{-1}. \quad (83)$$

Only in the case where $\Sigma' \equiv \mathbf{W}\Sigma\mathbf{W}$, i.e., where the model noise covariance matrix matches the noise covariance matrix of the underlying noise process, would we achieve a simplification and thus equality in Eq. (83). In the FD,

Eq. (82) reads

$$\left\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \left(\Delta \hat{\theta}_{\text{bf}}^b \right)' \right\rangle = \left(\tilde{\Gamma}' \right)_{ac}^{-1} \left(\tilde{\Gamma}' \right)_{bd}^{-1} \left(\widetilde{\mathbf{W} \partial_c \mathbf{h}} \right)^\dagger \left(\tilde{\Sigma}' \right)^+ \tilde{\mathbf{W}} \tilde{\Sigma} \tilde{\mathbf{W}} \left(\tilde{\Sigma}' \right)^+ \left(\widetilde{\mathbf{W} \partial_d \mathbf{h}} \right), \quad (84)$$

with $\tilde{\Gamma}$ given by the Whittle-based FM computed in the FD. We remark that Eq. (84) was derived in [48] that explored mismodeling non-stationary processes in ground-based detectors as stationary in the context of modulations and bursts, but not gaps.

B. Coherent Whittle-likelihood

In practice, as discussed in Sec. III B, for data with gaps one could make the choice to treat the data segment as a whole, using the Whittle-likelihood for that full segment, while simply filling the missing data in the gap with zeroes. We can either use a gating window w_g , or apply a tapering window w_t smoothing each side of the gap. This approach has been used in a number of works [25, 39, 44, 45, 47–49, 68]. Thus the approximate likelihood, denoted with a prime, would read

$$\ln \mathcal{L}'(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} (\mathbf{W} \mathbf{x})^T \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbf{x}), \quad (85)$$

with $\mathbf{x} = \mathbf{h}(\boldsymbol{\theta}) - \mathbf{d}$. In practice, this allows for an efficient computation in the FD,

$$\ln \mathcal{L}'(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \left(\widetilde{\mathbf{W} \mathbf{x}} \right)^\dagger \tilde{\Sigma}^{-1} \left(\widetilde{\mathbf{W} \mathbf{x}} \right), \quad (86)$$

where $\tilde{\Sigma}$ is diagonal, as for the Whittle-likelihood, with components given by (23), and where the window would be applied in the TD to the residual $\mathbf{x}$ before taking an FFT. We could instead use the Toeplitz form of Eq. (85) and transform the likelihood into the FD. A major disadvantage would be that the favored diagonal structure

of the FD likelihood would be lost (a result of assuming Toeplitz, but not circulant) and this rarely, if not ever done in practice. For this reason, we will use the Circulant condition to facilitate the rapid-to-evaluate Whittle-likelihood. Later in results Sec. VII we will test mismodeling errors when assuming the underlying process is circulant when the true noise-process is Toeplitz.

We can apply the results of Sec. IV A with the replacement $(\boldsymbol{\Sigma}')^+ = \boldsymbol{\Sigma}^{-1}$, and distinguish two cases:

- $\mathbf{W} = \mathbf{W}_g$, a gating window, meaning we directly fill the gaps with zeros and leave the rest of the data untouched;
- $\mathbf{W} = \mathbf{W}_t$, a smooth tapering window: we taper each gap edge using the Planck window (E1) with a lobe length that we will allow to vary from 1 minute to 2 hours.

The naive gating approach will give us a worst-case estimate of the impact of the gap. The tapering approach will reduce the impact of the mismodeling, but will also erase some of the information conveyed by the data inside the tapering lobes, depending on the lobe length⁶.

The results of Sec. IV A give in this context

$$\mathbf{\Gamma}'_{ab} = (\mathbf{W} \partial_a \mathbf{h})^T \boldsymbol{\Sigma}^{-1} (\mathbf{W} \partial_b \mathbf{h}), \quad (87a)$$

$$\left(\Delta \hat{\theta}_{\text{bf}}^a \right)' = (\mathbf{\Gamma}')_{ab}^{-1} (\mathbf{W} \partial_b \mathbf{h})^T \boldsymbol{\Sigma}^{-1} (\mathbf{W} \mathbf{n}), \quad (87b)$$

$$\left\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \left(\Delta \hat{\theta}_{\text{bf}}^b \right)' \right\rangle = (\mathbf{\Gamma}')_{ac}^{-1} (\mathbf{\Gamma}')_{bd}^{-1} (\mathbf{W} \partial_c \mathbf{h})^T \boldsymbol{\Sigma}^{-1} \mathbf{W} \boldsymbol{\Sigma} \mathbf{W} \boldsymbol{\Sigma}^{-1} (\mathbf{W} \partial_d \mathbf{h}). \quad (87c)$$

Eqs. (87c) and (87a) will be used extensively in the results Sec. VII A, namely to test the appropriateness of the Whittle-likelihood on gated and tapered data for PE purposes.

⁶ The loss of information during the tapering scheme is a feature of mismodeling the data using the Whittle-like circulant covariance matrix. If the model covariance were consistent with the tapered noise, no information would be lost. See Eq. (46) and surrounding text for a reminder.

C. Segmented Whittle-likelihood

The previous Whittle model for the likelihood in Sec. IV B treats the data segment as a whole (and thus coherent). Another approach would be to segment the data, treating each data segment between gaps as independent from the others, and using the Whittle-likelihood for stationary (and circulant) noise on each of these data segments. We will also use the same noise model before and after the gap segments. Assuming incoherence between data segments gives an advantage of working with data chunks of limited size where the data gaps simply amount to ignoring certain segments. The caveats are that:

- One neglects all noise correlations between separate segments. As we will see in details in Sec. VII E, this loss of information is in fact very minor.
- One uses the Whittle-likelihood on short segments, which comes with errors in itself at the segment edges. As in the previous section, this can be alleviated by tapering, but tapering itself erases some information contained in the data.
- One will observe a loss of resolvable frequencies, impacting sensitivity for low-frequency gravitational-wave sources. Assuming LISA is sensitive to sources for $f \in (10^{-4}, 10^{-1})$, this becomes problematic for data segments of length $T \lesssim 3$ hours.

Note that this approach is perhaps the closest to short-time Fourier transform or time-frequency methods.

In our gap-in-the-middle notations, this approach is

equivalent to taking

$$(\Sigma')^+ = \begin{pmatrix} (\Sigma_{00}^{\text{circ}})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\Sigma_{22}^{\text{circ}})^{-1} \end{pmatrix}, \quad (88)$$

where $\Sigma_{11}^{\text{circ}}, \Sigma_{33}^{\text{circ}}$ are both circulant matrices built from the same PSD (assumed to be known), for the segment lengths N_1 and N_3 . Note that these block matrices are not sub-matrices of Σ , i.e. $\Sigma_{11}^{\text{circ}} \neq \Sigma_{11}$, even if Σ is itself Toeplitz or circulant. To reduce edge effects, windows $\mathbf{W}_1, \mathbf{W}_3$ are used to taper both ends of each segment.

In a more general notation for K segments indexed by $I = 1, \dots, K$, the model likelihood is then a sum over the segments that are not part of the set of gaps,

$$\ln \mathcal{L}'(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2} \sum_{I \notin \text{gaps}} (\mathbf{W}_I \mathbf{x}_I)^T (\Sigma_{II}^{\text{circ}})^{-1} (\mathbf{W}_I \mathbf{x}_I), \quad (89a)$$

$$= -\frac{1}{2} \sum_{I \notin \text{gaps}} \left(\widetilde{\mathbf{W}_I \mathbf{x}_I} \right)^\dagger (\tilde{\Sigma}_{II}^{\text{circ}})^{-1} \left(\widetilde{\mathbf{W}_I \mathbf{x}_I} \right), \quad (89b)$$

with $\mathbf{x}_I = \mathbf{h}_I(\boldsymbol{\theta}) - \mathbf{d}_I$ the residual on the segment I . In the FD, each $\tilde{\Sigma}_{II}^{\text{circ}}$ is diagonal and the FFT can be used; the length of each segment N_I is decided by the gap locations and will not be a power of 2 in general. This will have an impact on speed of the discrete Fourier transform, making it at worst an $\mathcal{O}(N^2)$ operation rather than a $\mathcal{O}(N \log_2 N)$ operation.

The results of Sec. IV A for the model's FM and noise bias also become sums over I , in the same fashion:

$$\mathbf{\Gamma}'_{ab} = \sum_{I \notin \text{gaps}} (\mathbf{W}_I \partial_a \mathbf{h}_I)^T (\Sigma_{II}^{\text{circ}})^{-1} (\mathbf{W}_I \partial_b \mathbf{h}_I), \quad (90a)$$

$$(\Delta \hat{\boldsymbol{\theta}}_{\text{bf}}^a)' = (\mathbf{\Gamma}')_{ab}^{-1} \sum_{I \notin \text{gaps}} (\mathbf{W}_I \partial_b \mathbf{h}_I)^T (\Sigma_{II}^{\text{circ}})^{-1} (\mathbf{W}_I \mathbf{n}_I), \quad (90b)$$

The bias covariance, however, is a double sum over all segments that are observed:

$$\begin{aligned} \left\langle \left(\Delta \hat{\boldsymbol{\theta}}_{\text{bf}}^a \right)' \left(\Delta \hat{\boldsymbol{\theta}}_{\text{bf}}^b \right)' \right\rangle &= (\mathbf{\Gamma}')_{ac}^{-1} (\mathbf{\Gamma}')_{bd}^{-1} \sum_{I \notin \text{gaps}} \sum_{J \notin \text{gaps}} \left\langle (\mathbf{W}_I \partial_c \mathbf{h}_I)^T (\Sigma_{II}^{\text{circ}})^{-1} (\mathbf{W}_I \mathbf{n}_I) (\mathbf{W}_J \mathbf{n}_J)^T (\Sigma_{JJ}^{\text{circ}})^{-1} (\mathbf{W}_J \partial_d \mathbf{h}_J) \right\rangle \\ &= (\mathbf{\Gamma}')_{ac}^{-1} (\mathbf{\Gamma}')_{bd}^{-1} \sum_{I \notin \text{gaps}} \sum_{J \notin \text{gaps}} (\mathbf{W}_I \partial_c \mathbf{h}_I)^T (\Sigma_{II}^{\text{circ}})^{-1} \mathbf{W}_I \Sigma_{IJ} \mathbf{W}_J (\Sigma_{JJ}^{\text{circ}})^{-1} (\mathbf{W}_J \partial_d \mathbf{h}_J). \end{aligned} \quad (91)$$

Indeed, even if our model assumes segment indepen-

dence, cross-terms between segments appear because

the true underlying noise process $\mathbf{n}$ will have in general correlations between different segments, represented by Σ_{IJ} with $I \neq J$.

D. Mismodeling measures

Recall that our framework is different from mismodeling induced by systematic errors in the waveform models [69]. In the latter case, the part of the shift between best-fit and true parameters due to waveform errors is deterministic, although it will vary from source to source. Here, our biases in Eq. (81) are zero-mean, but have the wrong statistics (82).

In our case, we have two types of mismodeling errors introduced by our use of the wrong statistical model for the likelihood:

- A mismatch between the width of the posteriors obtained with the wrong model likelihood and correct model likelihood.
- A mismatch between the scatter of best-fit parameters when varying the noise realization and the width of the modelled posteriors.

The first type of error motivates the introduction of *noise mismodeling posterior width ratio* measuring the ratio of 1σ errors between the incorrect and correct analyses, as approximated in the Fisher approach:

$$\Xi_a = \frac{\sqrt{(\mathbf{\Gamma}')_{aa}^{-1}}}{\sqrt{(\mathbf{\Gamma})_{aa}^{-1}}}, \quad (92)$$

where $(\mathbf{\Gamma})_{aa}^{-1}$ is the correct Fisher covariance. The effect measured posterior width ratio is qualitatively similar (although not identical, as correlations could change) to a loss of SNR. For instance, we might want to alleviate the effect of gaps by introducing an aggressive tapering window around them, but this would come at the expense of a loss of information reflected by values $\Xi > 1$.

The second type of errors lead us to introduce the *noise-mismodeling posterior scatter-to-width ratio*

$$\Upsilon_a = \frac{\sqrt{\left\langle \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \left(\Delta \hat{\theta}_{\text{bf}}^a \right)' \right\rangle}}{\sqrt{(\mathbf{\Gamma}')_{aa}^{-1}}}. \quad (93)$$

We will sometimes refer to Eq. (93) in short by the *scatter-to-width ratio*. When forming the product $\Xi_a \Upsilon_a$, the approximate FM $\mathbf{\Gamma}'$ cancels out and we end up with

a ratio. The numerator is the variance of an unbiased estimator of the posterior co-variance, while the denominator is the FM for the true noise process, $\mathbf{\Gamma}$; this means that the Cramer-Rao inequality [63, 70, 71] applies and ensures that

$$\Xi_a \Upsilon_a \geq 1. \quad (94)$$

The scatter-to-width ratio Υ_a will be the most important quantity we will compute. It is a measure of the statistical consistency of the use of an incorrect statistical model for the noise, and provides a quick-to-compute proxy for a probability-probability (PP) plot. The numerator measures the scatter of best-fit parameters, while the denominator measures the width of the posterior, both obtained with the same approximate likelihood. A value $\Upsilon_a > 1$ indicates that the scatter in the posterior mode is larger than the posterior width; therefore, for most noise realizations, the true parameters will be excluded by the posteriors, leading to statistically significant random biases. For a concrete example of a case where $\Upsilon_a > 1$, see Fig. 14 and Fig. 17 below in the results section Sec. VII. A value $\Upsilon_a \simeq 1$ is the target, it would indicate that the bias scatter and posterior width are comparable, see Fig. 15 for an example. When using the correct likelihood, we have exactly $\Upsilon_a = 1$, as shown by Eq. (76).

We will use these Υ_a for diagonal elements of the covariance that are scaled by the approximate Fisher covariance so they can be understood to represent “number of sigmas” with respect to the approximate posterior. We could also compare off-diagonal components (without the square roots), representing correlations between parameters. In practice we found that the values of Υ_a and Ξ_a for different physical parameters θ^a showed the same overall trends, and in tables we will show simply their average over parameters

$$\bar{\Upsilon} \equiv \frac{1}{d} \sum_{a=1}^d \Upsilon_a, \quad \bar{\Xi} \equiv \frac{1}{d} \sum_{a=1}^d \Xi_a. \quad (95)$$

An important remark in contrasting the roles of the posterior width ratio Ξ_a and scatter-to-width ratio Υ_a is the following: what happens if one uses a wrong noise model that under- or over-estimates the overall noise level, while having the correct correlations? This would translate as $\Sigma' = \lambda^2 \Sigma$ (or for the PSD, $S'_n = \lambda^2 S_n$) for some factor $\lambda > 0$ (quadratic for the covariance, linear for the noise). Fisher covariances would scale as $(\mathbf{\Gamma}')^{-1} = \lambda^2 \mathbf{\Gamma}^{-1}$, marking a widening in the approximate posterior. However, noise biases (81) are insensitive to λ ; these max-likelihood biases are sensitive to the noise spectral shape, not to the noise level (similarly, in the context of waveform modeling errors, Cutler-Vallisneri biases are SNR-independent). We would therefore have $\Xi_a = \lambda$, $\Upsilon_a = 1/\lambda$, and $\Xi_a \Upsilon_a = 1$,

saturating the bound in Eq. (94). These scalings are in line with intuition: for $\lambda > 1$, we overestimate the noise, the posteriors would be too wide, encompassing the true parameters too often; for $\lambda < 1$, we underestimate the noise, the posteriors are too narrow and they often exclude the true parameters. As useful as this example is, we stress that in the rest of the paper, we focus on the mismodeling of noise correlations, not of the overall noise level. We will see examples where $\Xi_a \simeq 1$ but $\Upsilon_a > 1$ and conversely. We present examples in Tables IV and V, computed via the Whittle-likelihood to analyze gated data for a noise process with a large red-noise (low frequency) component.

E. Computing the parameter bias covariance

In general, we will be in a situation where computations with the correct likelihood are expensive, while using the incorrect model likelihood is much cheaper. While the ratio of Fisher covariances Ξ_a can perhaps be approximated by the ratio of SNRs, the scatter-to-width ratio Υ_a is the most important and informative quantity and requires computing the parameter bias covariance defined by Eq. (82). Different approaches can be used to compute it in practice:

- A *direct computation* of Eq. (82), which requires in general to compute the $N \times N$ matrix $(\Sigma')^+ \mathbf{W} \Sigma \mathbf{W} (\Sigma')^+$, which can be challenging for large N .
- A *Monte-Carlo computation* of the variance, by simulating many realizations of the noise $\mathbf{n}$ and using the noise biases in Eq. (81) for estimating the biases, using only the model likelihood.
- Full Bayesian PE using a set of noise realizations, producing directly max-likelihood parameters as best-fit parameters (without the linearized signal approximation).

In the second approach, we bypass the limitations imposed by working with $N \times N$ matrices where all that is required is the ability simulate noise realizations $\mathbf{n}$ according to the correct covariance Σ . The computation of each noise bias itself then only involves the *model* likelihood, not the correct one. If we can also compute the associated Fisher covariance, this means that we can compute Υ_a , with only the cheaper model likelihood; by contrast, Ξ_a requires the Fisher covariance for the correct likelihood.

Full Bayesian PE sampling is expensive in itself, potentially also for the model likelihood, and can be used as a check of the linear signal approximation estimate for biases. In our analysis, we will verify through multiple PE simulations that the three approaches are consistent

System	$M_z/M_\odot$	q	χ_1	χ_2	z	t_c	ι	φ	λ_L	β_L	ψ_L
M3e7	3×10^7	2	0.5	0.5	1	0	$\pi/3$	1.1	0.8	0.3	1.7
M3e6	3×10^6										

Table I. Physical parameters of our two MBHB systems: redshifted mass M_z , mass ratio q , aligned spin components χ_1, χ_2 , redshift z , coalescence time t_c , inclination ι , azimuthal phase φ , LISA-frame sky longitude and latitude λ_L, β_L , LISA-frame polarization ψ_L .

with each other. This is demonstrated in Fig. 17. If doing a series of Bayesian PE runs, we would be close to producing a PP-plot – with the difference that a PP-plot requires in principle to randomize also the input parameters of the signal, according to the priors used.

It is important to emphasize that the computation of the parameter bias covariance (82) does not rely on any assumption about the underlying noise process: the noise $\mathbf{n}$ can also be *non-Gaussian* [72]. The assumption that enters the derivation is that the model likelihood itself is Gaussian. In the direct computation approach, we only need access to the variance of the noise process (before applying windows) $\langle \mathbf{n} \mathbf{n}^T \rangle$. In the Monte-Carlo approach, we only need to be able to simulate multiple draws of $\mathbf{n}$. We do remark however that, when $\mathbf{n}$ is non-Gaussian, the distribution of the estimator itself may no longer be Gaussian so its statistical properties may not be entirely represented by the covariance matrix.

Finally, although we focus here on non-stationarity in the data stream in the form of data gaps, we stress that the tools described above could apply to other forms of mismodeling: mismodeling of the PSD itself, non-stationarities in the form of a modulated galactic binary foreground or evolving instrumental PSDs, and non-Gaussian features such as non-Gaussianity in the foreground or instrumental glitches.

V. WAVEFORM AND NOISE MODELS

A. MBHB signals and analysis setup

We apply our methods to the case of MBHB observations by LISA. At high masses, these signals are short and dominated by the merger. We will consider two MBHB systems at redshift $z = 1$, with redshifted masses $M = 3 \times 10^7 M_\odot$ and $M = 3 \times 10^6 M_\odot$; their parameters, otherwise identical, are given in Tab. I. The lighter signal will extend to higher frequencies and allows us to explore the importance of the frequency content of the signal itself. Focusing on high-mass systems allows us to capture most of the SNR (SNR values are given in Tab. III) in a short data segment where we can apply the full treatment of gaps by marginalization described

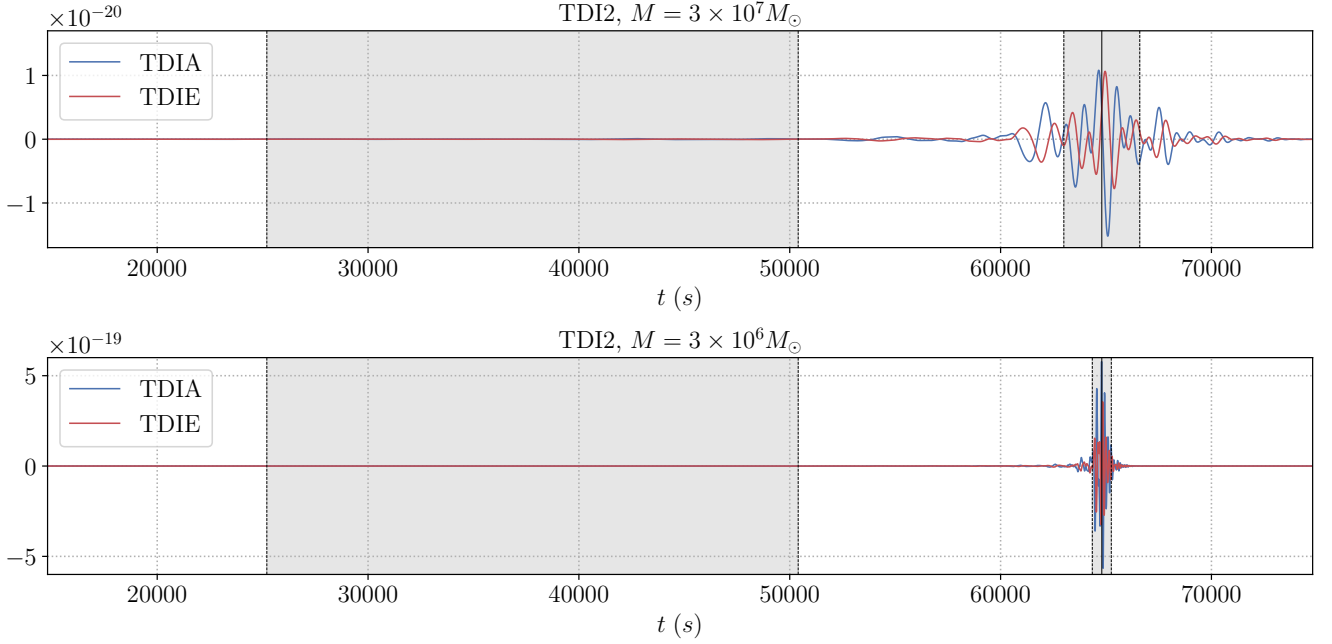


Figure 1. **(Top Panel):** We present the M3e7 case with parameters given in the top row of Tab. I. The full extent of the data is $[0, 81920]$ s and the greyed out regions indicate the gap placements (Given in Tab. II) we will consider throughout this work. The blue/red is the second generation TDI variable A/E given by the response of the instrument to the incoming MBH radiation. **(Bottom Panel):** The same as the top panel, but with MBH/gap parameters given in the bottom row of Tab. I and Tab. II respectively.

System	T	Δt	N	t_0	$\Delta T_{\text{merger}}^{\text{gap}}$	$\Delta T_{\text{inspiral}}^{\text{gap}}$	Merger gap	Inspirational gap
M3e7	81920s	10s	8192	64800s	1hr	7hr	$[63000\text{s}, 66600\text{s}]$	$[25200\text{s}, 50400\text{s}]$
M3e6					15min		$[64350\text{s}, 65250\text{s}]$	

Table II. Segment length and gap configurations for our two MBHB systems. Note that the merger gap has a different length for the two masses: 1hr for M3e7 and 15min for M3e6.

SNR	No gap	Merger gap	Inspirational gap
M3e7	2238	1372	2234
M3e6	4811	1361	4807

Table III. SNR of our MBHB systems in different gap scenarios.

in Sec. III, which requires to work with $N \times N$ matrices. This full treatment will give us a crucial point of reference to assess modeling errors with the tools of Sec. IV.

We consider different gap scenarios, described in full in Tab. II. In order to gauge the importance of gap location, we introduce either a 7h gap during the inspiral, which represents perhaps the most realistic instrumental setting (particularly if the gap is planned), or a gap centered right on the merger, which represents the worst-case scenario of an unlucky unplanned gap; note however that the most massive signals are short and

vulnerable even to planned gaps in the absence of an advance detection. For the gap at merger, we choose a *different* duration of 1h for $M = 3 \times 10^7 M_\odot$ and 15min for $M = 3 \times 10^6 M_\odot$; the merger signal is longer at high masses, and this choice is done in order to arrive at roughly the same SNR after introducing the gap. The SNR values obtained with and without the gaps are listed in III. For tapering, we employ a Planck window as described in App. E.

The waveforms themselves are produced with the FD waveform model **IMRPhenomHM** [73] with higher harmonics $\{(2,2), (2,1), (3,3), (3,2), (4,4), (4,3)\}$. We restrict the parameter space to spins that are aligned with the angular momentum. Although more recent and more accurate models are available (e.g. [74]), we do not expect the waveform choice to have an impact on the noise mis-modeling errors (beyond qualitative differences like the inclusion of higher modes, as they change the morphology of the posteriors, see [75] for more details). We do

not consider waveform systematics in the present study, and the waveform model is identical for simulated data and analysis templates.

In our analysis, we generate the MBHB GW signals and compute the LISA response in the FD using `lisabeta` [75]. We approximate LISA orbits as having constant and equal armlengths. Again, we do not expect that more realistic LISA orbits and response would change our results about the effects of gaps specifically, as the same approximations enter the simulations and the templates.

Bayesian parameter estimation is performed using the parallel-tempered code `ptemcee` [76], that is based on the ensemble sampling technique [77, 78].

B. LISA response and TDI

We will give a minimal account of the Time-Delay Interferometry (TDI) observables that form the LISA data stream (see [2] for an introduction to the LISA mission), as we wish to highlight later on the importance of the noise spectral content, from red noise to blue noise. We use the notations of [75]. In the equal-armlength approximation, the one-arm laser frequency shift observables, $y_{slr} = (\nu_s - \nu_r)/\nu$, measured between emitting spacecraft s and receiving spacecraft r along the link l take the form [79] (setting $c = 1$):

$$y_{slr} = \frac{1}{2} \frac{n_l \otimes n_l}{1 - k \cdot n_l} : [H(t - L - k \cdot p_s) - H(t - k \cdot p_r)] , \quad (96)$$

where k is the GW propagation unit vector, n_l is the unit vector from s to r along the LISA arm, p_s, p_r are the spacecraft positions at some time t , and H is the GW in 3×3 matrix form (see the notations of [75]). The notation $\otimes, :$ indicates that H is contracted twice with n_l .

TDI is a construction that allows to reproduce in post-processing interferometric configurations allowing to reduce the otherwise dominant laser noise, and is crucial to enable LISA observations ([80–86], see [87] for a review). Multiple variables can be constructed, see e.g. [86]. First-generation TDI cancels laser-noise for static, unequal-arms. Second-generation TDI is necessary to provide the needed cancellations for realistic orbits, with armlengths changing over time [83–86]. For recent work on TDI including higher-generation TDI, we refer the reader to Refs. [88–91].

The full TDI expressions simplify drastically in the approximation of equal-armlengths, treating all delays as equal on both ways along each arm. There is then a single delay in the construction, $L/c \simeq 8$ s for $L = 2.5$ Gm. With the notation $y_{sr,nL} = y_{sr}(t - nL)$ (dropping the middle index), the first-generation and second-

generation TDI Michelson observables X_1, X_2 read:

$$\begin{aligned} X_0 &= y_{31} + y_{13,L} - (y_{21} + y_{12,L}) , \\ X_1 &= X_0 - X_{0,2L} , \\ X_2 &= X_1 - X_{1,4L} . \end{aligned} \quad (97)$$

Other Michelson variables Y, Z are obtained by cyclic permutations. Here, X_0 could be seen as a 0-th generation Michelson TDI. Geometrically (see Fig. 3 of [86]), it would correspond to a single pass along each arm; such a combination does not cancel laser noise for unequal arms.

Uncorrelated combinations A, E and T [92] can be built from X, Y, Z (of any generation) as

$$\begin{aligned} A &= \frac{1}{\sqrt{2}} (Z - X) , \\ E &= \frac{1}{\sqrt{6}} (X - 2Y + Z) , \\ T &= \frac{1}{\sqrt{3}} (X + Y + Z) . \end{aligned} \quad (98)$$

These combinations present the advantage of featuring uncorrelated instrumental noise, allowing to treat these three channels as three independent detectors. This is only true for equal armlengths, and we would have to consider correlations across channels in any realistic setting for the instrument; we will retain this approximation for simplicity here. The channel T is strongly suppressed at low frequencies, where only the two channels A, E survive.

As illustrated by (97), successive TDI generations correspond in first approximation to discrete derivatives of the data stream. In the FD, delays are simple phase factors and these finite differences take simple expressions. In [75], we introduced a notation for rescaled TDI variables, factoring out these frequency-dependent terms as

$$\tilde{a}, \tilde{e} = \frac{e^{-2i\pi fL}}{i\sqrt{2}\sin(2\pi fL)} \times \tilde{A}, \tilde{E} , \quad (99)$$

and a different rescaling for $\tilde{T}$, which is unnecessary for this work.

In our study, we will use these different TDI generations to explore the importance of red and blue noise; a derivative changes the spectrum of a variable by one power of f , and second-generation TDI is therefore a process that has a blue tilt compared to first-generation. Our nomenclature for tables and figures is:

- TDI0: channels A_0, E_0 given by the combinations (98) from X_0, Y_0, Z_0 given in (97); they differ by a constant factor from the variables given in (99) as $\tilde{a}, \tilde{e} = -\sqrt{2}\tilde{A}_0, \tilde{E}_0$ (while $\tilde{t}$ of [75] is not the same as $\tilde{T}_0$);

- TDI1: channels A_1, E_1 given by the combinations (98) from X_1, Y_1, Z_1 given in (97);
- TDI2: channels A_2, E_2 given by the combinations (98) from X_2, Y_2, Z_2 given in (97).

We will also ignore the T channel entirely for our low-frequency signals. Assuming independence across the A, E TDI channels (of any generation), the log-likelihood becomes a sum over the two channels. For instance, (39) becomes

$$\ln \mathcal{L}_{\text{gap}} = -\frac{1}{2} \sum_{C=A,E} (\mathbf{W} \mathbf{x}_C)^T (\mathbf{W} \Sigma_C \mathbf{W})^+ (\mathbf{W} \mathbf{x}_C), \quad (100)$$

with in fact an identical noise covariance for the two channels, $\Sigma_A = \Sigma_E$. Similarly, all expressions of Secs. III and IV for the SNR, Fisher matrices, noise biases, with and without mismodeling, are similarly given with sums $\sum_{A,E}(\dots)$.

Strictly speaking, gaps would differently affect the different TDI generations. As shown by Eq. (97), the TDI

construction involves delayed combinations of the base observables y_{sr} . This leads to (a priori) a loss of information at the edges of the missing base data, leading to a wider data gap, an effect that is worse for higher-order TDI constructions. For TDI2, the maximum delay that intervenes in (97) is $7L/c \simeq 58\text{s}$ (1 for TDI0, +2 for TDI1, +4 for TDI2). In the gap case, this leads to the SNR being slightly different for the three TDI versions. We leave these differences aside in the present study, and the gaps are treated as covering the same time interval in all TDI generations. We refer the reader to Ref. [91] for more discussions related to how gaps impact TDI variables for various generations.

C. Noise models

Throughout our analysis we will focus on three noise curves each with different properties. We will use the **SciRDV1** noise curve given in the LISA mission requirements document [93]. We will define the instrument related single-link optical metrology noise P_{OMS} and test mass acceleration noise P_{acc} by (restoring factors of c)

$$P_{\text{OMS}}(f) = (15 \times 10^{-12} \text{m})^2 \left(1 + \left(\frac{2 \text{mHz}}{f} \right)^4 \right) \left(\frac{2\pi f}{c} \right)^2 \text{Hz}^{-1} \quad (101)$$

$$P_{\text{acc}}(f) = (3 \times 10^{-15} \text{m s}^{-2})^2 \left(1 + \left(\frac{0.4 \text{mHz}}{f} \right)^2 \right) \left(1 + \left(\frac{f}{8 \text{mHz}} \right)^4 \right) \left(\frac{1}{2\pi c f} \right)^2 \text{Hz}^{-1} \quad (102)$$

The noise curves for the A and E channels in each TDI configuration, TDI0, TDI1 and TDI2 respectively, are given by

$$S_n^{A_0, E_0}(f) = 2 [2(3 + 2 \cos x + \cos 2x) P_{\text{acc}}(f) + (2 + \cos x) P_{\text{OMS}}(f)] , \quad (103a)$$

$$S_n^{A_1, E_1}(f) = 4 \sin^2 x S_n^{A_0, E_0}(f) , \quad (103b)$$

$$S_n^{A_2, E_2}(f) = 4 \sin^2 2x S_n^{A_1, E_1}(f) . \quad (103c)$$

where $x = 2\pi f L/c$, $L = 2.5 \text{ Gm}$ the length of the LISA arms and c the speed of light. We also account for the presence of the galactic foreground $S_c(f)$, which is folded into $S_n(f) \mapsto S_n(f) + S_c(f)$, according to [94, 95]. In this work, we assume a LISA observation duration $T_{\text{obs}} = 4\text{yr}$ to set the level of this foreground.

With cadence $\Delta t = 10$ seconds over a ~ 22.475 hour long interval, plots for the different noise curves are given in Figs. (103a - 103c).

Noise described by the S_n^{TDI0} PSD, yields a powerful red-noise component at low frequencies, proportional to $\sim 1/f^4$. Time-domain noise generated from S_n^{TDI0} would describe low frequency oscillations of the arms with noise amplitudes many orders of magnitude higher than noise generated from the TDI1 and TDI2 noise curves respectively. Similarly, TDI1 noise

curve $S_n^{\text{TDI1}}(f)$ yields a non-trivial red noise component $\sim 1/f^2$. The TDI2 noise curve S_n^{TDI2} , in contrast to both TDI0/1 curves, is a white noise process for low frequencies $f \rightarrow 0$. Each process demonstrates the same blue noise behavior in general for high frequencies $\sim f^2$. In contrast to TDI0, both TDI1 and TDI2 have “zero-crossing” behavior that increase in number for higher frequencies. In [75], TDI0 was constructed for the sake of convenience to avoid such zero-crossings resulting in 0/0 numerical instabilities. For more discussion, please

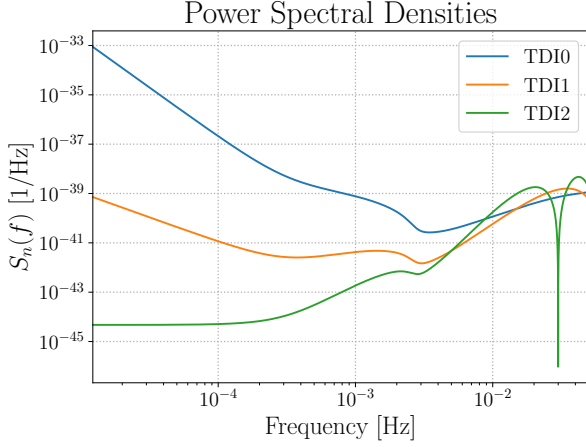


Figure 2. The blue, orange and green curves depict the behavior of the noise curves given by $S_n^{\text{TDI0}}(f)$ (103a), $S_n^{\text{TDI1}}(f)$ (103b) and $S_n^{\text{TDI2}}(f)$ (103c) respectively. The white dwarf background is present and it has been assumed that LISA has been operational for one year.

refer to Eqs.(29–33) in [75].

VI. ANALYSIS OF THE NOISE COVARIANCE MATRICES

With noise models described by Eqs. (103a–103c), gap structure in Tab.II and tapering function (E1), and with formalisms developed in Sections III A and III C, we are now in a position to numerically investigate the impact of gaps on the noise covariance matrix in both the time and FD. Sec. VIA illustrates the differences between circulant and Toeplitz structures. Sec. VIB demonstrates the impact of gaps on the covariance matrices, manifesting as a breakdown of the TD Toeplitz structure leading to an overall *non-stationary* process. In Sec. VIC, we outline our numerical schemes detailing how we compute pseudo-inverses in practice. Throughout this section, we will consider an observation time of $T \sim 22.435$ hrs with cadence $\Delta t = 10$ s resulting in a data stream of length $N = 8192$ with lowest resolvable frequency $\Delta f \sim 1/T \approx 10^{-5}$ Hz.

A. Circulant and Toeplitz processes

Plots of the circulant and Toeplitz matrices are given respectively in Figs. (3–4). In Fig. 3, the matrices are circulant (19), obeying not only the Toeplitz rule of constant entries along each diagonal, with all entries deduced from the first row of the matrix, but also showing a circular symmetry: each row is permuted by one element to the right with respect to the previous row,

wrapping around the edge. The first row is also symmetric with respect to its middle point. Notice that the matrices in Fig. 4 maintain the Toeplitz property in Def. (18) but lack circular symmetry: different rows are not related by a cyclic permutation. There are clear differences in the TD noise covariance matrices between the two cases. The FD can shed further light on the differences between the two stationary processes. We can compute

$$\tilde{\Sigma}^{\text{circ}} = N\Delta t^2(\mathbf{P}\Sigma^{\text{circ}}\mathbf{P}^\dagger) = \text{Diagonal Matrix}, \quad (104)$$

$$\tilde{\Sigma}^{\text{toe}} = N\Delta t^2(\mathbf{P}\Sigma^{\text{toe}}\mathbf{P}^\dagger) = \text{Banded Matrix}. \quad (105)$$

The matrix $\tilde{\Sigma}^{\text{circ}}$ is a diagonal matrix with elements given by the noise curves Eqs.(103a–103c). Converting the Toeplitz TD noise covariance matrices to their FD analogs $\tilde{\Sigma}^{\text{toe}}$ results in a *banded matrix* with sub-leading off-diagonal elements. This result was expected given the calculation present in App. A 1 leading to Eq. (A14). A plot of the main diagonal of the circulant matrix $\tilde{\Sigma}^{\text{circ}}$ compared to the leading and sub-leading diagonals of $\tilde{\Sigma}^{\text{toe}}$ is given in Fig. 5. For TDI1, it appears that the assumption of a circulant process when the underlying process is Toeplitz appears reasonable for our range of observable frequencies. At high frequencies one obtains a near perfect match between the Toeplitz and circulant based FD matrices, with minor deviations at low frequency. The leading diagonals are subdominant, with power decreasing monotonically. For TDI0, the noise is (significantly) under-estimated which would result in tighter posteriors and significantly larger bias-scatters with respect to the true posterior. In the context of Sec. IV D, we would expect the scatter-to-width ratios $\Upsilon_a \gg 1$. We find the same result for TDI2, but perhaps to a slightly lesser degree than TDI0. We remark that this is consistent with our later findings in Tab.XI found in the results Sec. VII, where we perform parameter inference assuming the underlying noise process as circulant when in reality it is Toeplitz.

In our work, unless stated otherwise, we will assume that the underlying noise process is both stationary, circulant and gaps modeled as missing data. This is to remain in line with the general community that use the Whittle-likelihood on the usual assumption that the noise is circulant, and it allows us to isolate the effects of gap mismodeling from the effects of Toeplitz/circulant mismodeling.

B. Noise covariance matrices with gaps

The left-most panels in Figs. (6 – 8) visually demonstrate the impact of a single gap on the TD noise covariance matrix with noise curves TDI0/1/2 respectively. Notice that both the circulant and Toeplitz structure,

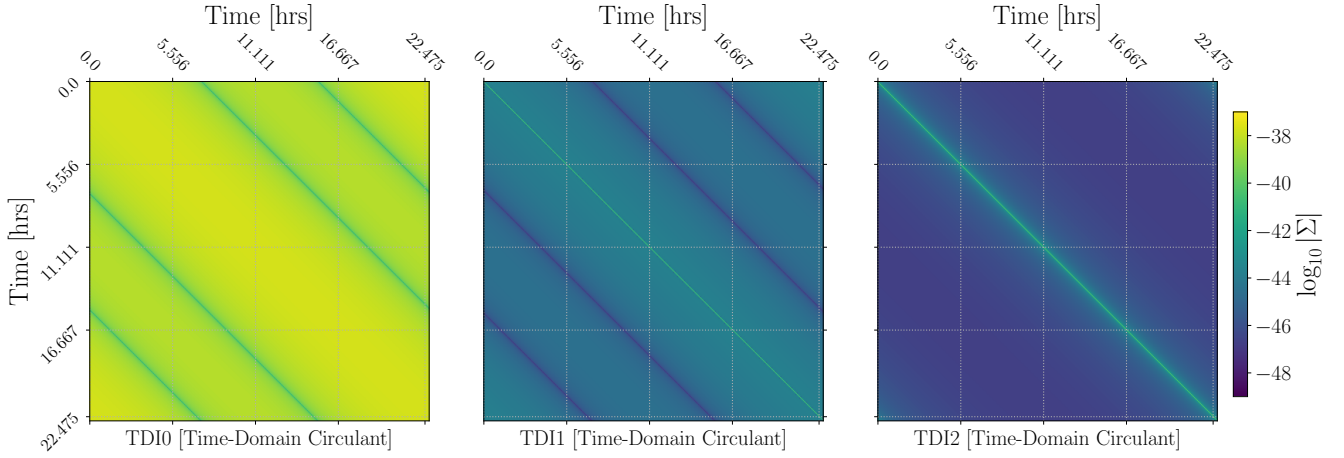


Figure 3. (Left to right panels:) Time-domain and *circulant* noise covariance matrices for TDI0/1/2 respectively.

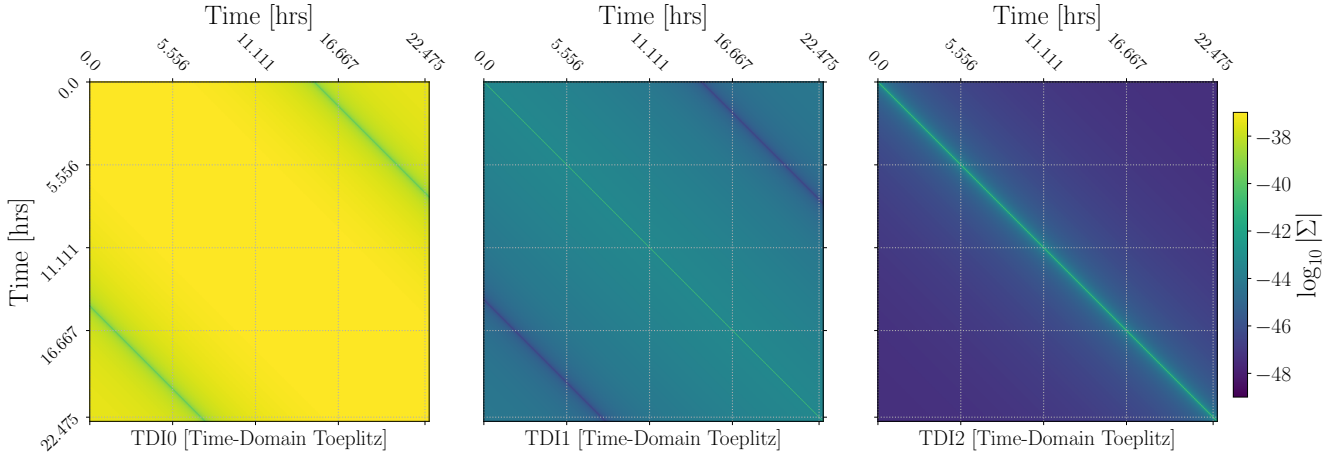


Figure 4. (Left to right panels:) Time-domain and *Toeplitz* noise covariance matrices for TDI0/1/2 respectively.

as depicted via Fig. 3, is now lost, and instead a rank deficient matrix (see Eq.(34)) features. The individual diagonal sub-matrices each have a Toeplitz structure and are *not* circulant. The off-diagonal sub-matrices represent correlations of the noise process pre and post gap segment. If these off-diagonal matrices were neglected, then the analyst would be making the assumption that the noise pre and post gap is independent. This was mentioned in Sec. IV C, where the analyst may decide to make the approximation that the individual diagonal sub-matrices are circulant. Notice, that since the circulant property is lost in the individual main diagonal sub-matrices, their FD covariances cease to be diagonal, and the Whittle-likelihood loses validity. These noise mis-modeling investigations will be discussed in more detail in the results Sec. VIIE.

The TD noise covariance matrices can also be expressed in the FD. For TDI0, the middle panel of

Fig. 6 is the FD version of the TD gated noise covariance matrix given by the left panel of Fig. 6, computed via Alg. 1. The excess power at low frequencies $f \approx 10^{-5}$ Hz, given by the steep rise of the red-noise dominated noise curve, correlates strongly with higher frequencies as $f \rightarrow 0.05$ Hz. Introducing a taper with lobes of length 30 minutes mitigates the leakage effects, restoring the banded nature of the FD noise covariance matrix. The correlations between low and high frequencies are still present and this is due to the enormous red-noise component of the noise-curve for TDI0. For TDI1, the middle and right panels of Fig. 7 demonstrate to a slightly lesser degree the effects of leakage, with tapering focusing the power more along the main diagonals. The same for TDI2 is observed in Fig. 8, but with the additional feature of correlated zero-crossings between frequency components. In the limit as $\Delta t \rightarrow 0^+$ (higher Nyquist frequency) and/or infinite observation

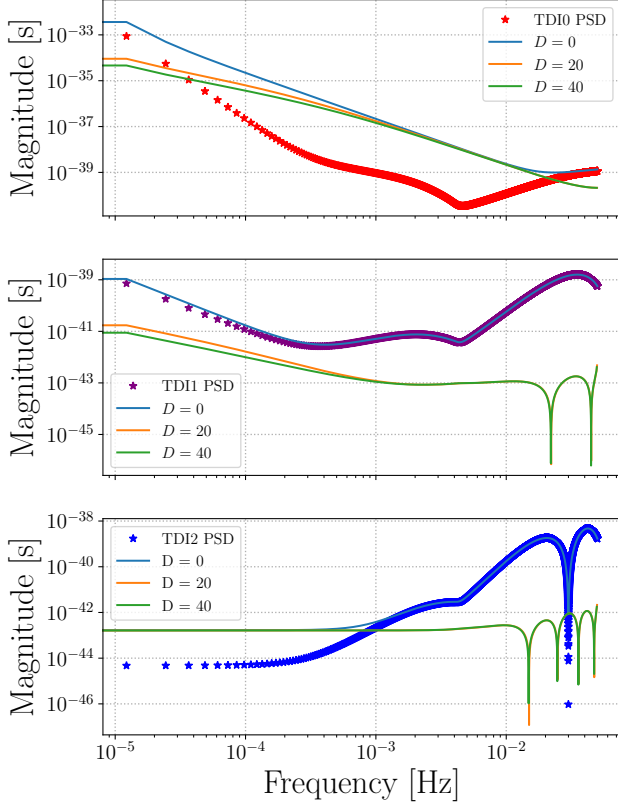


Figure 5. **(Top plot):** Comparison between (in red) TDI0 PSD, the main diagonal of circulant noise process, with the main diagonal ($D = 0$, blue) and sub-leading diagonals ($D = \{20, 40\}$, orange and green respectively) of the same noise curve but given by a Toeplitz process $\hat{\Sigma}^{\text{toe}}$. **(Middle and bottom plot):** The same as the top plot but for (in purple) TDI1 noise curves and (in blue) TDI2 noise curves respectively. mismodeling the Toeplitz process as a circulant process (most common in practice) appears to be worse for TDI0 and TDI2 than the TDI1 noise curve. The Toeplitz TDI1 process looks to be well approximated by the corresponding circulant process.

times similar features would be observed in TDI1, but not TDI0.

The tapering scheme, in theory, should not impact observables involving inner products with the noise covariance matrices in the FD assuming the model covariance matrix is consistent with the true noise covariance matrix. The likelihood (and SNR) should be independent of the tapering scheme and information lost should only happen during the gated segment. The band-diagonal nature of the FD noise covariance matrices for TDI1 and TDI2 could be exploited to accelerate likelihood computations. However, we will see in the next Sec. [VIC](#) that the introduction of a smooth taper makes the computation of the pseudo-inverse challenging.

C. Pseudo-inverses with gated functions

The gated TD noise-covariance matrix has a block-like structure, which facilitates a relatively simple computation of the pseudo-inverse. The gated FD noise-covariance matrix has no obvious structure to take advantage of. Below we will outline a numerical scheme that allows for an *exact* calculation of the pseudo-inverse for FD noise covariance matrices assuming a simple gating function. The results are then extended to windows with non-trivial lobes in Sec. [VID](#).

In App. [D](#) we prove that the number of zero-eigenvalues in the FD gated noise-covariance matrix is equivalent to the number of zero rows (or columns) in the TD gated covariance matrix. Following [\[57\]](#), one can compute the Singular Value Decomposition (SVD) $\hat{\Sigma}^{\text{gate}} = U\Lambda V^\dagger$ to obtain the pseudo-inverse as $\hat{\Sigma}^+ = V\Lambda^+U^\dagger$. Since the matrix $\hat{\Sigma}^{\text{gate}}$ is hermitian symmetric, the number of zero-eigenvalues are equivalent to the number of zero-singular values in the matrix $\mathbf{S}$. Fig. [9](#) shows a plot of the normalized singular matrix $\mathbf{S}/\max(\mathbf{S})$ computed via the SVD of $\hat{\Sigma}^{\text{gate}}$ for each of the TDI0/1/2 noise curves using a rectangular window. The parameters of the rectangular window are given by the last column of Tab. [II](#). On all panels in Fig. [9](#), the red curves are the SVs that correspond to the gap. The blue curves are the SVs that point towards data points that exist between gap segments. The top plot is for TDI0, where we observe a steep drop in the SVs that represents the strong red-noise component of the noise curve. This feature is not present in TDI1 or TDI2. We understand from previous arguments that the number of zero-SVs are precisely the number of time-bins where the window function is zero. In practice we set these zero-SVs as infinite so they will not contribute in matrix calculations using the inverse Λ^+ . The routine outlined here will always calculate the *unique* pseudo-inverse for the FD covariance matrix for simple gating functions.

Computing the pseudo-inverse in the TD is trivial in our set up. The pseudo-inverse of the TD noise covariance matrix is simply the inverse of the individual block Toeplitz matrices as discussed in section [III A](#). For a generic gap configuration, the same method applies after a reordering of the matrix, as shown in App. [C](#). This becomes more challenging if a smooth taper is applied, since the action of the taper increases the condition number of the matrix. One could apply a similar logic of the discussion surrounding [\(106\)](#) for the TD matrix to compute the pseudo-inverse. Whenever calculations are conducted in the TD, we will never taper the data to ensure that our results are as numerically stable as possible.

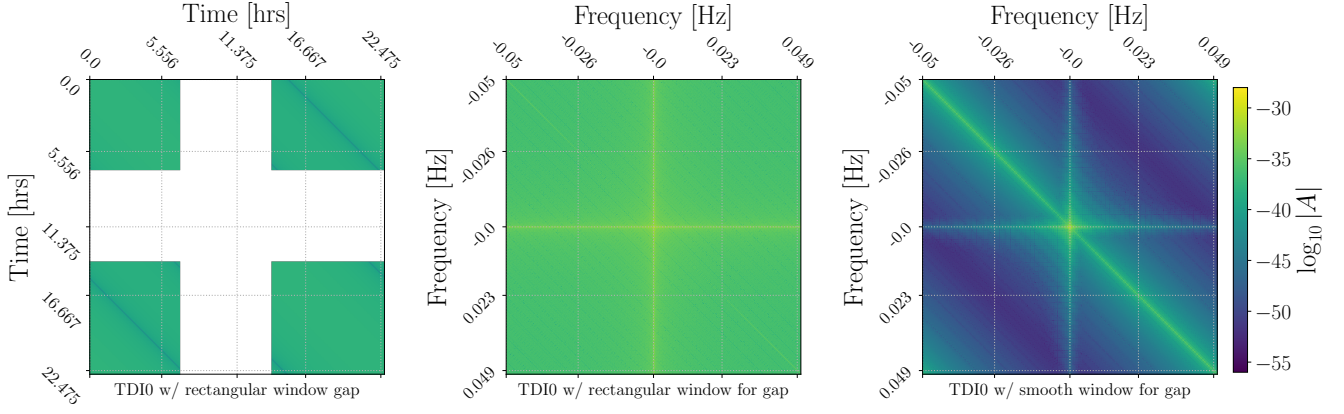


Figure 6. **(Left panel:)** Time-domain gated noise covariance matrix for a circulant process with a TDI0 noise curve. Here the gap is represented via zeros in the noise covariance matrix. **(Middle panel:)** The FD representation of the left-most panel. **(Right panel:)** The FD noise-covariance matrix of the left-most panel but with a smooth 30 minute taper at each endpoint of the gap segment. Notice that the powerful red-noise component of TDI0 yields significant leakage (and thus correlations) amongst the frequency components of the middle panel. The right-most panel has precisely the same power as the middle panel, but more concentrated along the diagonals (and $0 - f$ frequency correlations).

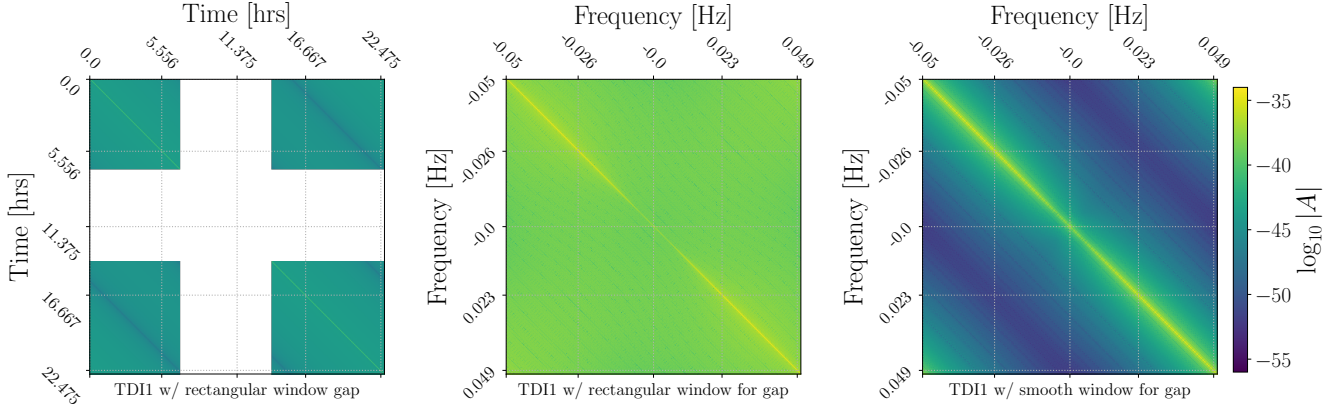


Figure 7. The same set up as Fig. 6 but for TDI1. The middle panel, in contrast, retains diagonal dominance but much of the power leaks out into neighboring frequency bins. This effect is reduced significantly if a 30 minute taper is applied as shown by the right-most panel.

D. Pseudo-inverses with tapering functions

a tolerance $0 \leq \epsilon \ll 1$, such that

$$\Lambda_i^+ \mapsto \begin{cases} \Lambda_i^{-1} & \mathbf{S}_i / \max(\mathbf{S}) > \epsilon \\ 0 & \text{otherwise.} \end{cases} \quad (106)$$

It was seen in the right-most panels of Figs. (6–8), that there was clear impact of tapering on the structure of the FD noise-covariance matrix. The presence of a taper marginally complicates the computation of the pseudo-inverse, since there is now a smooth drop towards the SVs that cause the bad behavior of the noise covariance matrix. As such, there is no clear criterion on how many SVs to remove when smooth tapers are introduced (see [47] for more discussion). To get by this, we normalize the SVs by $\mathbf{S} / \max(\mathbf{S}) \in [1, 0)$ and define

The ideal choice of ϵ would be small enough such that maximal information during the tapering scheme is retained *and* the matrix is well behaved. In this section, we will demonstrate the implications for the FD analysis if a smooth taper were applied to the TD data. We refer the reader to Fig. 11, which illustrates our tapering scheme unique to this section. We will study a gap at merger elapsing one hour, with one window function a simple binary mask (gate) and the other a smooth taper which lobe lengths ~ 2 hours. The num-

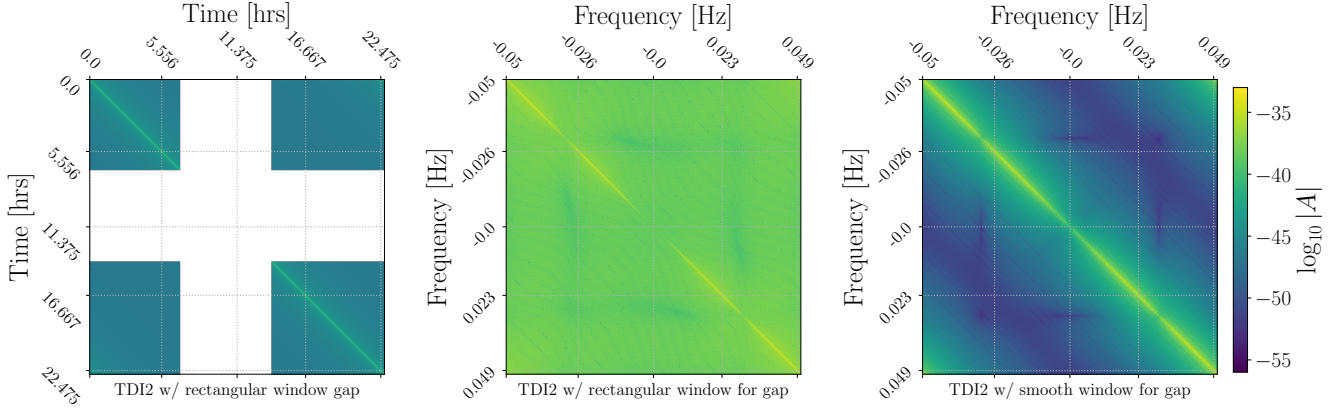


Figure 8. The same set up as Fig. 6 but for TDI2. The zero crossings of the PSD (as shown in Fig. 2) are seen along the main diagonal of the middle panel and correlations with that zero crossing are shown by the fainter patches in both the middle and right panels. Like the right-most panels of Fig. 6 and Fig. 7, the leakage effects are reduced significantly only if a smooth taper is applied.

ber of zero-data is identical between the two windowing cases. The blue stars in Fig. 10 are the result of FD SNR computations via Eq.(68) for different choices of small ϵ when computing the pseudo-inverse using the regularization procedure outlined in Eq. (106). The red horizontal dashed line is the SNR of the gated MBH signal given by the blue curve in Fig. 11. The green horizontal dashed line is the SNR of the tapered MBH when using a diagonal (Whittle) based noise covariance matrix. In the limit as $\epsilon \rightarrow 0^+$, the FD SNR tends towards the gated SNR, implying that *all* of the information is retained during the tapering scheme. Notice that the SNR assuming a Whittle-based covariance, with taper applied to the MBH, is significantly less than the SNRs computed using correct pseudo-inverses that correctly account for the windowed noise process. To put this into perspective, the SNR of the yellow curve in Fig. 11 is *equivalent* to the SNR of the blue curve in Fig. 11, thanks to the appropriate incorporation of the window in the covariance itself. If the window was not incorporated into the noise covariance matrix, then this situation is equivalent to analyzing a signal that tapers to zero whilst the noise does not, yielding a lower SNR. This validates the claim made around Eq. (42), that information is only lost during the gated segment of the data stream. Assuming the pseudo-inverse is correctly calculated, information should be retained during the smooth tapering procedure.

To close this section, we plot the tapered FD covariance matrix and corresponding pseudo-inverse (with regularization $\epsilon = 10^{-13}$) in the left and right panels of Fig. 12. To add to the middle and right panels of Fig.(6–8), we observe that the gentler the tapering scheme the more the power is concentrated along the main diagonals. The right-most plot is a plot of the pseudo-inverse,

possessing little exploitable structure aside from the fact that the matrix is Hermitian. The pseudo-inverse is a dense matrix, making FD likelihood computations expensive $\sim \mathcal{O}(N^2)$. We are hopeful that there will exist methods that can improve the cost of computing the likelihood in the FD through, perhaps, exploiting band-diagonal nature of $(\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})$ and accelerating using parallelization techniques on Graphical Processing Units (GPUs). Performance studies of the TD and FD likelihoods are a topic we did not explore in great detail, and we will leave this matter for future work.

Since there are no obvious cost reductions (yet) for analyses conducted in the FD, we will use the TD methods described in this paper to compute quantities presented in the results Sec. VII. We have checked that calculations (involving $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{E}}$ as defined in Eq. (95) remain unchanged whether computations are performed in either the FD or TD.

VII. RESULTS

The first instance of mismodeling that we will consider is the case of the improper use of the Whittle-likelihood, appropriate for Gaussian stationary noise, to analyze a data stream where missing data breaks the stationary feature of the stochastic process.

The advantage of using the Whittle-likelihood is obvious from the computational standpoint, allowing for $\mathcal{O}(N \log_2 N)$ likelihood evaluations. Searching for sources in LISA could be done using cheap Whittle-based likelihoods, before further refinements are made with the full, correct but more expensive likelihoods. We want therefore to ascertain what would be the consequences of using the Whittle-likelihood despite the

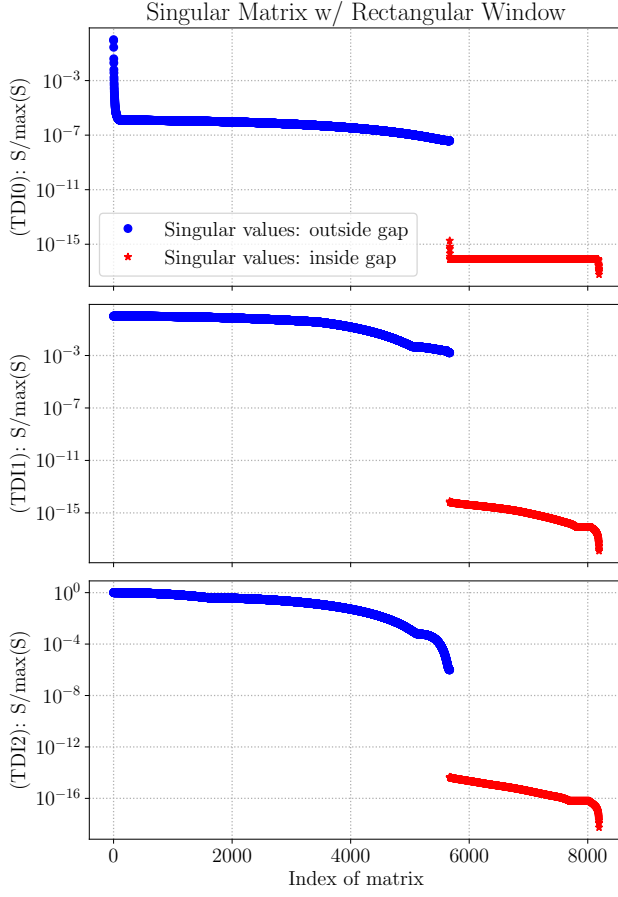


Figure 9. **(Top plot to bottom plot):** Normalised SVs of the FD noise covariance matrix for TDI0/1/2 respectively. In red/blue are the SVs that are retained/discarded. Since we use a rectangular window, we know precisely that the zero-SVs correspond to the number of zeros in the window function w_{rect} .

presence of gaps in the data, and whether introducing a smooth taper around gaps can alleviate noise mismodeling errors.

Here, we will assume that we know exactly the PSD and the covariance Σ of the underlying noise process. We also assume that the noise process is coherent over the full interval, and the gap acts as a mask, causing data to be missing without any other impact on its statistics. We will come back to the case where the true noise process is independent before and after the gap in Sec. VIIE. We will also assume that the Σ covariance has a circulant structure, and not only Toeplitz, on the full segment:

$$\Sigma = \Sigma_{\text{circ}} \quad (107)$$

This is not physical, as no real data stream will have this property; we make this choice so that the discrete

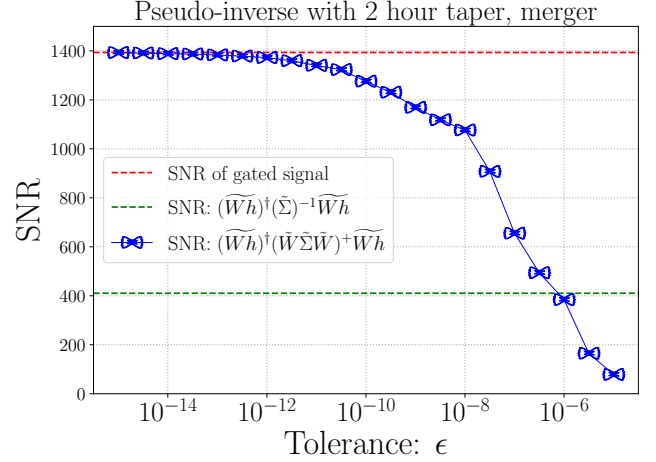


Figure 10. **(Red horizontal dashed line):** The SNR of the M3e7 MBH with one hour gap at merger using a gating function, with pseudo-inverse computed using only the (blue) SVs retained in the bottom panel of Fig. 9. **(Green horizontal dashed line):** The SNR of the MBH using a smooth taper with 2 hour length lobes with Whittle-based noise covariance. **(Blue farfalle):** The SNR computations now using a smooth tapering function (see the yellow tapered MBH signal in Fig. 11) with pseudo-inverse computed with increasing accuracy as $\epsilon \rightarrow 0$.

Whittle-likelihood (24) is exact before gaps are introduced, in order to isolate other sources of mismodeling from the Toeplitz/circulant mismodeling, to which we will come back in Sec. VIIG.

A. Mismodeling Whittle on gapped data

In order to assess the impact of noise mismodeling in the presence of data gaps with a Whittle-based approximation, we start by computing the mismodeling ratios defined in Eq. (92) and Eq. (93), for the different systems and TDI/gap configurations described in Sec. V. Results are reported for Υ in Tab. IV and for Ξ in Tab. V with layout of the tables described in Tab. IV. With independent codes developed specific to the FD, we have verified results found in the the top row of Tab. IV and Tab. V.

The columns in those tables give the result for the different generations TDI2, TDI1 and TDI0 as introduced in Sec. V. The expected outcome of the L0/L1 noise reduction pipeline will be TDI2 [2], but comparing with noise processes corresponding to TDI1 and TDI0 allows us to understand the effect of the shape of the PSD, from blue-noise dominated for TDI2, to featuring both blue and red noise for TDI1, and to strongly red-noise dominated for TDI0 (as illustrated in Fig. 2). In each case we show the results for both gap configurations, during the

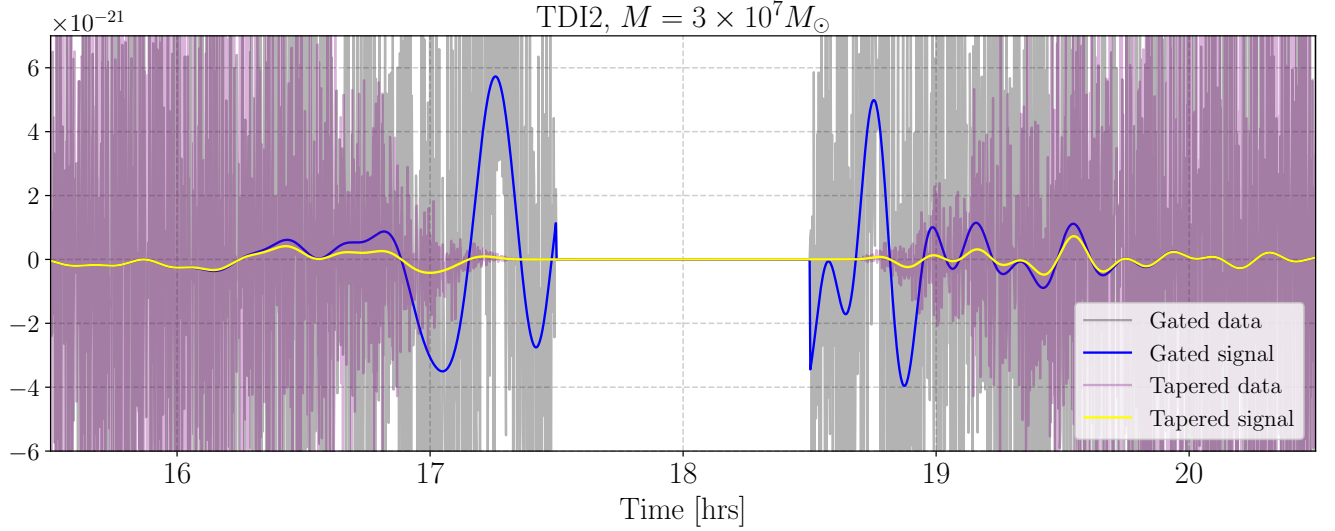


Figure 11. **(The Blue/Yellow solid curves):** The gated/tapered MBH signal for the M3e7 case with a gap at merger lasting 1 hour. **(The purple/black curves):** The gated/tapered data stream consisting of noise and the MBH signals. The SNR of the gated signal is ~ 1397 , whereas the SNR of the tapered signal is ~ 1393 using a pseudo-inverse regularized according to Eq.(106) with $\epsilon = 10^{-13}$. Assuming accurate computation of the pseudo-inverse of the noise covariance matrix, the SNR should be *invariant* to the choice of tapering function with only loss of SNR occurring during the gated segment. In theory, albeit very counter-intuitive, the SNR of the orange and blue curves above are identical, provided the noise covariance includes the effect of the tapering.

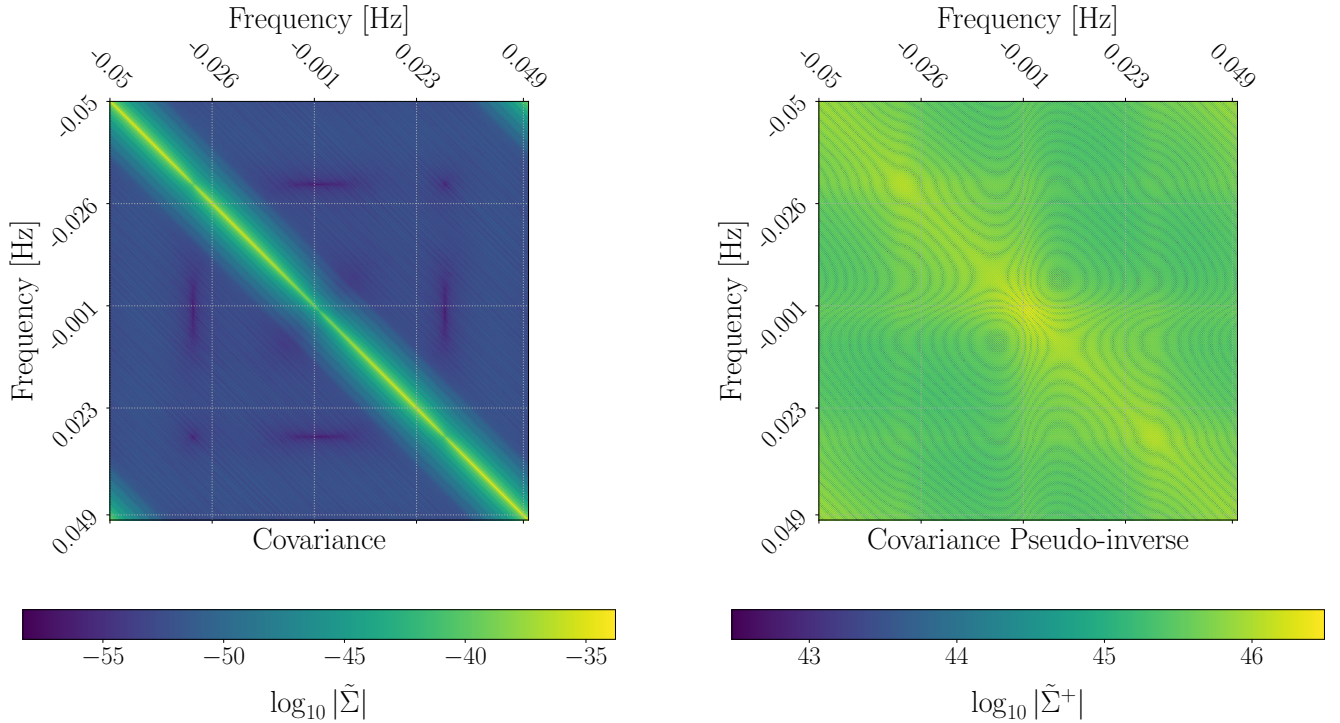


Figure 12. **(Left Panel):** The FD tapered noise covariance matrix with 1 hour gap at merger with a very smooth ~ 2 hours worth of lobes. Notice that the matrix appears to be band-diagonal, with all power concentrated along those specific diagonals. **(Right Panel):** The *pseudo-inverse* of the matrix presented in the left panel, for a regularization choice of $\epsilon = 10^{-13}$. Aside from hermitian symmetry and psychedelic patterns, there is little exploitable structure to be had in the pseudo-inverse.

Mismodeling $\bar{\Upsilon}$ data gaps		TDI2				TDI1				TDI0			
		Merger		Insp.		Merger		Insp.		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6
Data	Model												
Coherent, gap	Whittle, gated	12.2	40.5	1.3	1.4	1.2	1.6	1.0	1.0	39.7	38.0	21.9	6.4
Coherent, gap	Whittle, taper 10min	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.0	8.6	3.2	5.5	1.6
Coherent, gap	Whittle, taper 30min	1.0	0.8	1.0	1.0	1.0	0.8	1.0	1.0	6.1	2.0	3.0	1.1
Coherent, gap	Seg. Whittle, taper 10min	1.1	0.9	1.0	1.0	1.0	0.9	1.0	1.0	10.7	2.9	6.5	1.9
Coherent, gap	Seg. Whittle, taper 30min	1.0	0.8	1.0	1.0	1.0	0.8	1.0	1.0	6.6	2.0	3.6	1.4

Table IV. **(From left to right):** The “Data” column describes how the data stream was generated and should be regarded as the truth. The “model” column represents the approximate likelihood used in the model. The next three main columns represent which TDI process is used, TDI2/1/0 respectively. The sub-columns distinguish between a gap placement during the merger or inspiral, and between high and low-mass MBHBs with $M = 3e6$ or $M = 3e7$. We refer to Fig. 1 for signal illustrations and gap placement. **(From top to bottom):** “Data” is always the same here: the data is first treated as coherent (and circulant) over the full segment, and then a gap is introduced. “Model”: “Whittle” implies that the full-segment Whittle covariance is used as in IV B, effectively ignoring the gap in the covariance, while either gating or tapering the data on 10, and 30 minutes respectively. “Seg. Whittle” means that a segmented Whittle likelihood is used, treating data segments before and after the gap as independent, as in Sec. IV C; tapering is applied at both ends of each segment. **(Table entries):** individual numerical results are the scatter-to-width ratio $\bar{\Upsilon}$ (93), averaged over parameters as in (95). $\bar{\Upsilon}$ measures statistical inconsistency, $\bar{\Upsilon} \gg 1$ shows posteriors that are scattered from the truth much more than they are wide. The colors green/yellow/orange/red indicate the magnitude of the deviation from 1 on a scale of acceptable (green) to unacceptable (red).

Mismodeling $\bar{\Xi}$ data gaps		TDI2				TDI1				TDI0			
		Merger		Insp.		Merger		Insp.		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6
Data	Model												
Coherent, gap	Whittle, gated	0.8	0.5	1.0	1.0	1.0	0.9	1.0	1.0	0.9	0.9	1.0	1.0
Coherent, gap	Whittle, taper 10min	1.0	1.7	1.0	1.0	1.1	1.8	1.0	1.0	1.1	1.7	1.0	1.0
Coherent, gap	Whittle, taper 30min	1.2	3.9	1.0	1.0	1.3	3.9	1.0	1.0	1.3	3.7	1.0	1.0
Coherent, gap	Seg. Whittle, taper 10min	1.0	1.8	1.0	1.0	1.1	1.8	1.0	1.0	1.1	1.8	1.0	1.0
Coherent, gap	Seg. Whittle, taper 30min	1.2	4.0	1.0	1.0	1.3	3.9	1.0	1.0	1.3	3.8	1.0	1.0

Table V. Identical layout as Tab. IV except that $\bar{\Xi}$ is shown, representing the averaged ratio between the true and model posteriors widths, in accordance with Eqs. (92) and (95).

inspiral or right at merger, and for both MBHB masses, $M = 3 \times 10^7 M_\odot$ (M3e7) and $M = 3 \times 10^6 M_\odot$ (M3e6).

The first three lines of Tab. IV and Tab. V show the result for the approach of Sec. IV B, where we use a Whittle-likelihood for the full segment. The first line shows the result without a taper (the window is a direct gating function, 1 or 0), the second and third line for a tapering window as defined in Eq. (E1), with lobe lengths of 10 minutes and 30 minutes respectively.

We first observe that $\bar{\Upsilon}$ is always much larger than 1 for TDI0, catastrophically so as $\bar{\Upsilon}$ is counted in units of σ . Tapering helps in reducing these values significantly but does not eliminate the problem, with a slightly better performance of the longer tapering lobes. We attribute this to the fact that TDI0 is very red-noise dominated, with long-range correlations encoded in the high

values of the PSD at low frequencies, which make the effect of gap mismodeling much more pronounced. The values of $\bar{\Xi}$, however, are close to 1 for the gap in the inspiral in all cases, while raising above 1 for the gap at merger when tapering is introduced.

At this point, we wish to raise a word of caution. In GW astronomy, it is common to use sky-averaged sensitivity curves [96] for SNR computations and simplified parameter estimation purposes with Fisher matrices. Sensitivity curves incorporating the transfer function into the PSDs can result in large red-noise components, worse than our TDI0 example. For such a dominant red-noise process, in early explorations we found scatter-to-width ratios that were very severe ($\bar{\Upsilon} \gg 1$), but ratios of posteriors widths that were mild ($\bar{\Xi} \sim 1$). This indicates that exploratory studies investigating SNRs and

Fisher Matrices may be fine but PE studies with simulated noise would become problematic when using sensitivity curve-based noise, with a poor representation of the impact of gaps compared to realistic TDI2 data.

Focusing on the gap at merger, we see that treating the gap through a naive gating window (with no taper) gives unfavorable Υ values for TDI2. The gating window case is not as bad for TDI1, with a difference that may be attributed to the stronger blue noise component of TDI2. However, introducing tapering significantly limits the issue, with Υ close to 1 for the high-mass system M3e7, while the lower-mass system M3e6 has values below 1 indicating that posteriors are too broad with respect to their scatter. The values of Ξ are lower than 1 for the gating window, indicating underestimated posterior widths. With tapering, the values of Ξ rise above 1, particularly for the M3e6 case, indicating that tapering, if it limits statistical inconsistency by decreasing Υ , worsens information loss by increasing Ξ .

The gap in the inspiral, by contrast, induces much weaker mismodeling errors for TDI2 and TDI1, with values for both Υ and Ξ that are close to 1 in all cases, except when using a gating window in the TDI2 case for Υ . This is an important and positive result: the inspiral gap represents a more typical configuration, while the gap at merger represents a worst-case scenario.

The last two lines of Tab. IV and Tab. V correspond to the segmented Whittle approach of Sec. IV C. Although they are two parts of a coherent noise process realization, the data segments before and after the gap are treated as independent, and we use a Whittle-likelihood on each of those segments, with a tapering at each end of the segments. The two lines show the results for a lobe length of 10 minutes and 30 minutes respectively.

The behavior of this segmented Whittle model is very similar to the full-data Whittle model across all configurations, with the same qualitative behavior in both Υ and Ξ when using the same tapering lengths. This points towards a limited, or even negligible, impact of correlations across the gaps, with independence being a safe approximation to make – we will directly test for this in Sec. VII E below.

B. Parameter estimation examples

The mismodeling ratios Υ and Ξ are useful to paint a qualitative picture of mismodeling errors in a synthetic manner, but they are based on the FM approach, relying on a linearized signal approximation around the true parameters. It is therefore important to validate this approach with full Bayesian PE runs, which will also give us a better illustration of the full posteriors than the Gaussian approximation with a Fisher based covariance.

The PE runs are performed with the TD implemen-

tation of the likelihoods, using the prescription outlined in in Sec. III A with likelihood given by Eq.(40). We will only present results for the TDI2 case, with a gap at merger, for both masses $M = 3 \times 10^7 M_\odot$ and $M = 3 \times 10^6 M_\odot$. The noise realization $\mathbf{n}$ is kept the same in all simulations. The posterior samples obtained span all 11 physical parameters, but to save space we will only show corner plots for two subsets: the intrinsic parameters $\{\mathcal{M}_c, q, \chi_+, \chi_-\}$, and the most important extrinsic parameters $\{D, \iota, \lambda_L, \beta_L\}$ (see the caption of Tab. I). We found that the correlations in the full corner plots are comparable to those shown in these reduced plots. In all corner plots, we show the result of the linear signal approximation estimate for the max-likelihood bias obtained with each likelihood model.

We start by showing in Fig. 13 posteriors for the case where we do not introduce any gap (where the Whittle-likelihood is valid), and for the statistically correct treatment of the merger gap described by the gated likelihood in Eq. (39). This comparison demonstrates how much information is erased by the gap, and we see that the introduction of the gap significantly broadens the posteriors, in line with the degradation in SNR given by Tab. III. The estimates for the best-fit parameters given by (75) are in good agreement with the posteriors.

Next, Fig. 14 compares the correct *gated* likelihood to the Whittle-likelihood for the full segment described in Sec. IV B, with a direct gating window. As predicted by Tab. IV and Tab. V, we see that the width of the Whittle posterior is comparable to the correct one for M3e7, and significantly smaller for M3e6, while their center are moved away to the point of excluding the true parameters with great significance. Importantly, we find that the noise biases as estimated by (81), if not exact, are reasonably accurate in evaluating the magnitude of the bias. This is an important check as the validity of the Fisher approach, which is intrinsically only a local approximation to the posterior distribution.

Fig. 15 shows the same comparison, but this time introducing a smooth tapering window with a lobe length of 10 minutes. As predicted by Tab IV and Tab. V, for the heavy system M3e7, the introduction of this taper gives a posterior that is in good agreement with the correct posterior, both in centering and width. For the M3e6 system, however, the posteriors are significantly broadened. They are still compatible with the true parameters, with slightly larger biases; in accordance with the expected value $\bar{\Upsilon} = 0.9$ in Tab. IV, the broadening of the posteriors dominates over the slight change in biases, and the max-likelihood parameters are actually closer to the true parameters in units of σ -deviations. We also notice that posteriors in extrinsic parameters are non-Gaussian.

Finally, Fig. 16 compares the correct likelihood to the segmented Whittle-likelihood as described in Sec. IV C,

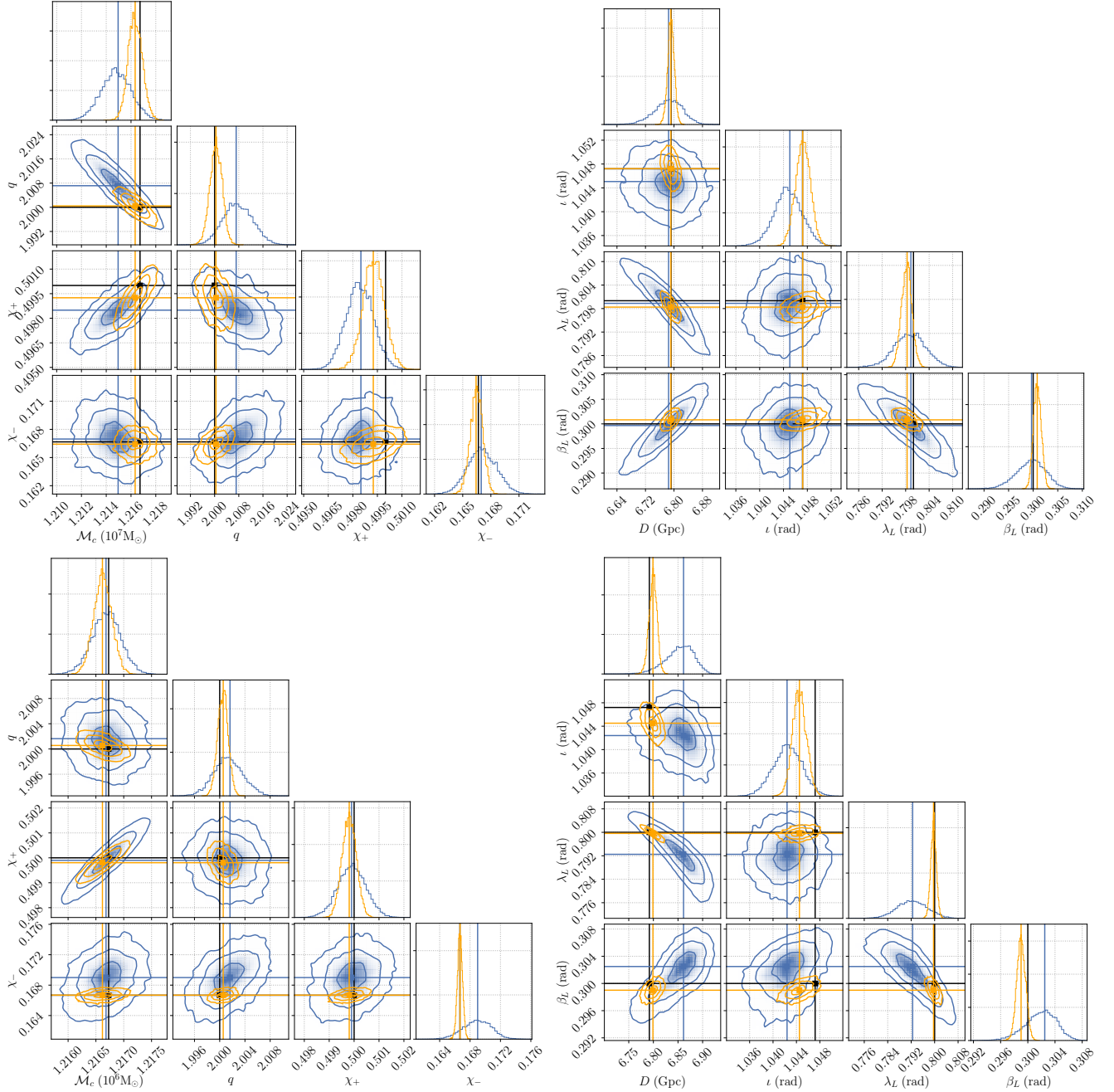


Figure 13. **(Top Row/Bottom Row)** Parameter estimation results for the case M3e7/M3e6, gap at merger. The inference is done on the full parameter space but we show only sub-corner plots for the intrinsic parameters on the left and extrinsic parameters on the right. The black vertical and horizontal lines indicate the true parameters. The yellow posteriors show for reference the result of parameter estimation on a data stream without any gap using the Whittle-based likelihood (which is then the correct likelihood by assumption). The blue posteriors were generated by performing inference on a data stream with gaps during the merger phase when using a model that is statistically consistent with the data generation process, as described in Sec. III. The blue and yellow lines are the corresponding Fisher-based predictions of the MLE.

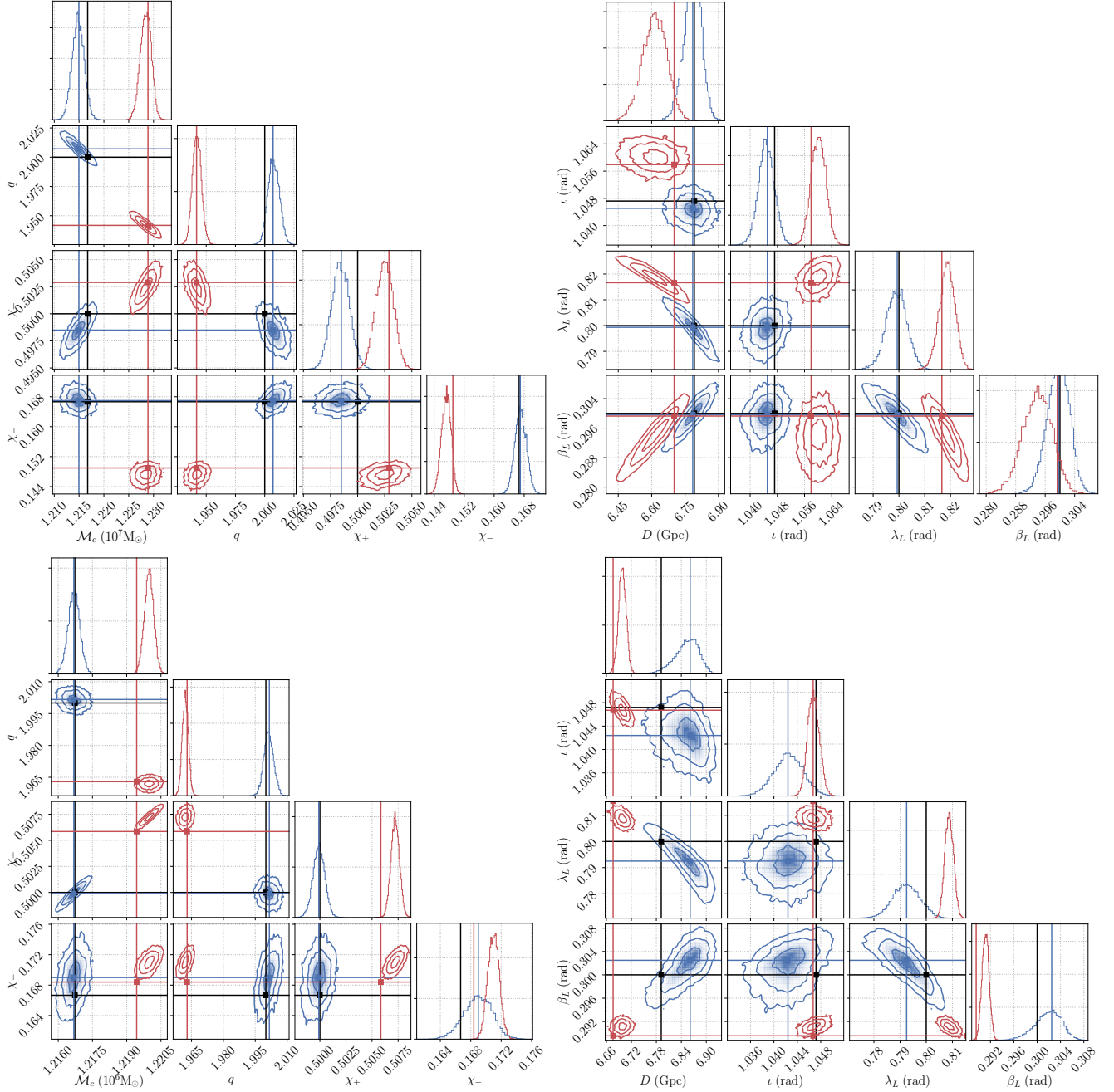


Figure 14. A similar configuration to figure 13. The blue posteriors are the result of parameter estimation on a data stream with gaps using the correct gated likelihood. The red posteriors were generated by performing inference on a data stream with gaps during the merger phase when gating the data (no tapering) and using the Whittle-likelihood of Sec. IV B, which is statistically inconsistent with the data generating process.

with the same tapering lobe lengths equal to 10 minutes. We find comparable results to Fig. 15, in accordance to Tab. IV and Tab. V, with good results for M3e7 and similarly broadened posteriors for M3e6. We verify again that biases seem well estimated by the formula given by Eq. (90).

The full PE runs that have just been discussed show a good level of agreement between the measured noise biases and (90) and the true max-likelihood parameters found in PE. In order to validate the synthetic Fisher-based error measures Υ and Ξ , we repeated this check multiple times.

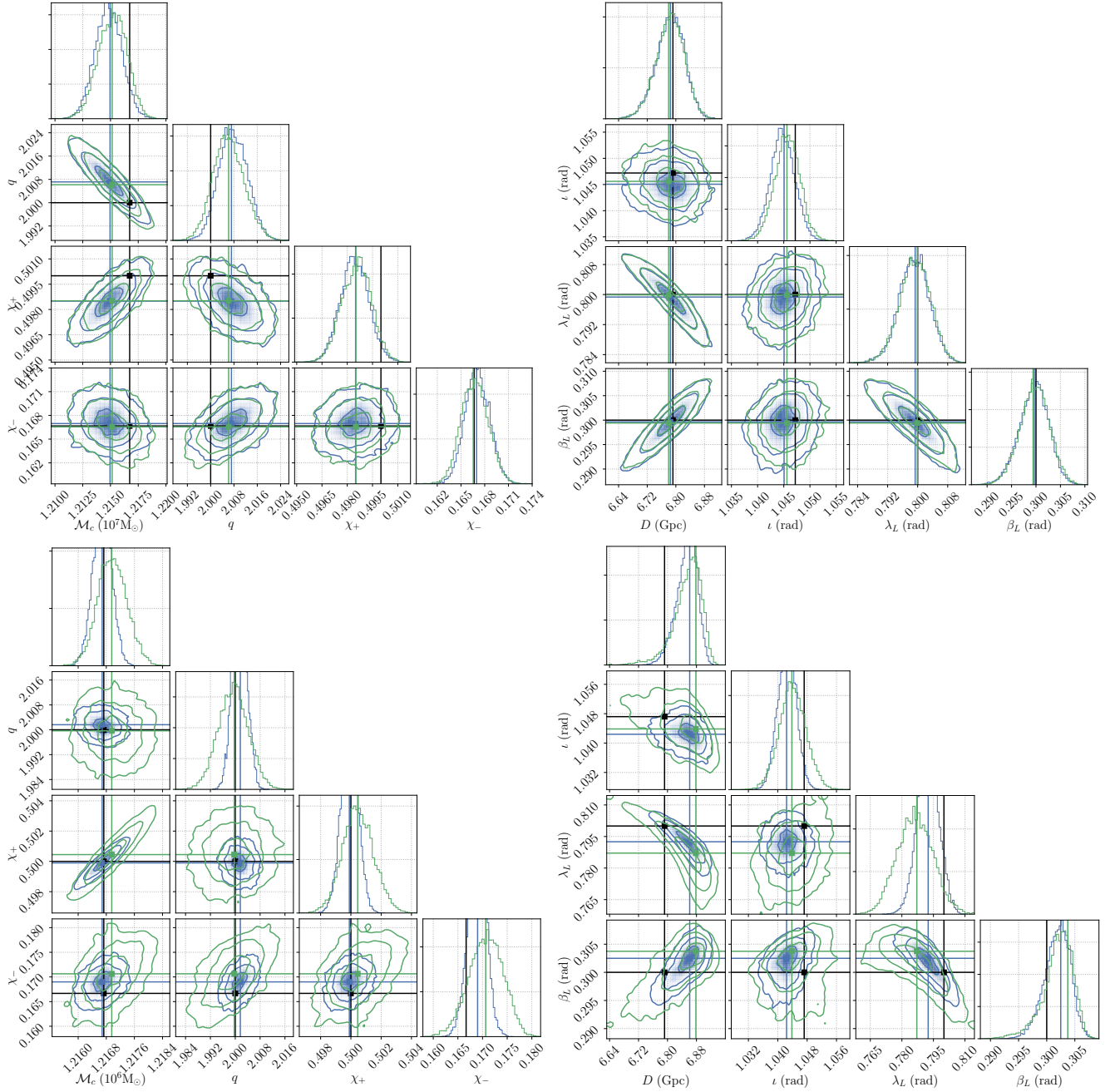


Figure 15. The same set up as figure 14 but applying a 10 minute taper before and after the gap segment when using the approximate Whittle-likelihood of Sec. IV B, represented by the green posterior. The blue posterior, as before, represents a truthful result where the gated data stream is inferred with a likelihood built from a gated covariance.

For this set of PE runs, performed 350 random draws of the noise realization $\mathbf{n}$, focusing on the Whittle case with a simple gating window, for TDI2 and M3e7 with a merger gap (first line, leftmost entry in Tab. IV and Tab. V). This choice allows us to validate these noise bias estimates in a regime where the parameter shifts are multiple units of σ , while the Fisher approach relies

on a local approximation to the likelihood. We expect the linear signal approximation bias estimates to work better in the other cases with tapering, as they show more moderate biases.

The results are shown as a corner plot in Fig. 17. Note that this is not a corner plot for posterior samples. Instead, this is showing a (multidimensional) histogram of

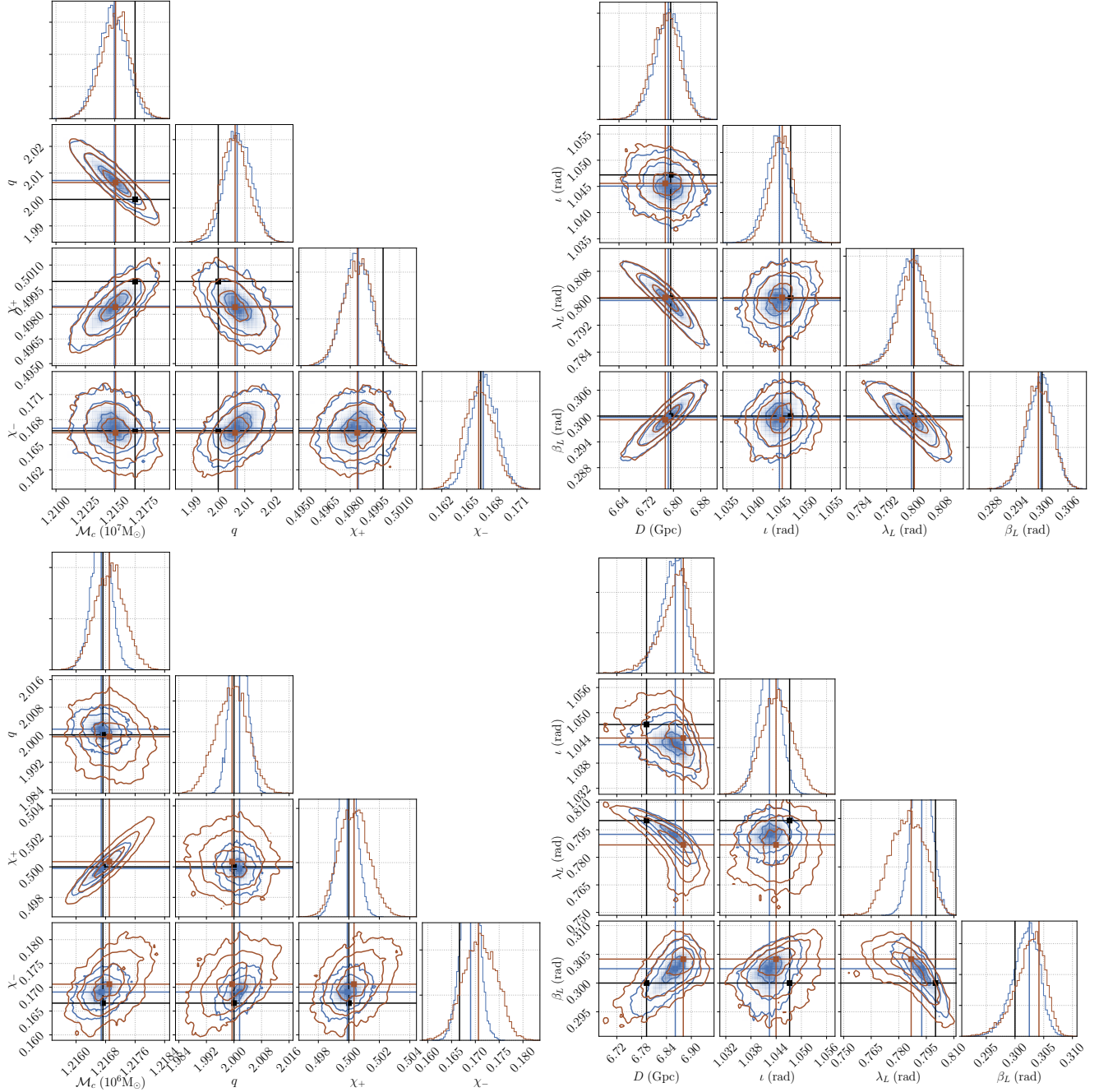


Figure 16. The same set up as Fig.15 but where the model likelihood, represented by the brown posterior, is instead the segmented Whittle model of Sec. IV C, and the independence between segments is assumed. A smoothing taper of 10 minutes is applied to both ends of each segment. The blue posterior, as before, represents the truthful situation where a gated likelihood is used to perform inference on a gated data stream.

max-likelihood values extracted from the PE runs, estimates for the max-likelihood calculated according to Eq. (87b), along with ellipses corresponding to a Gaussian distribution using the direct expression given by Eq. (87c) for the bias covariance. The comparison be-

tween the ellipses and the contours for the biases is a pure consistency check, relying only on the Gaussian nature of the noise. The comparison between the contours for the biases and the contours for the max-likelihood results from PE is a stronger check, verifying that the

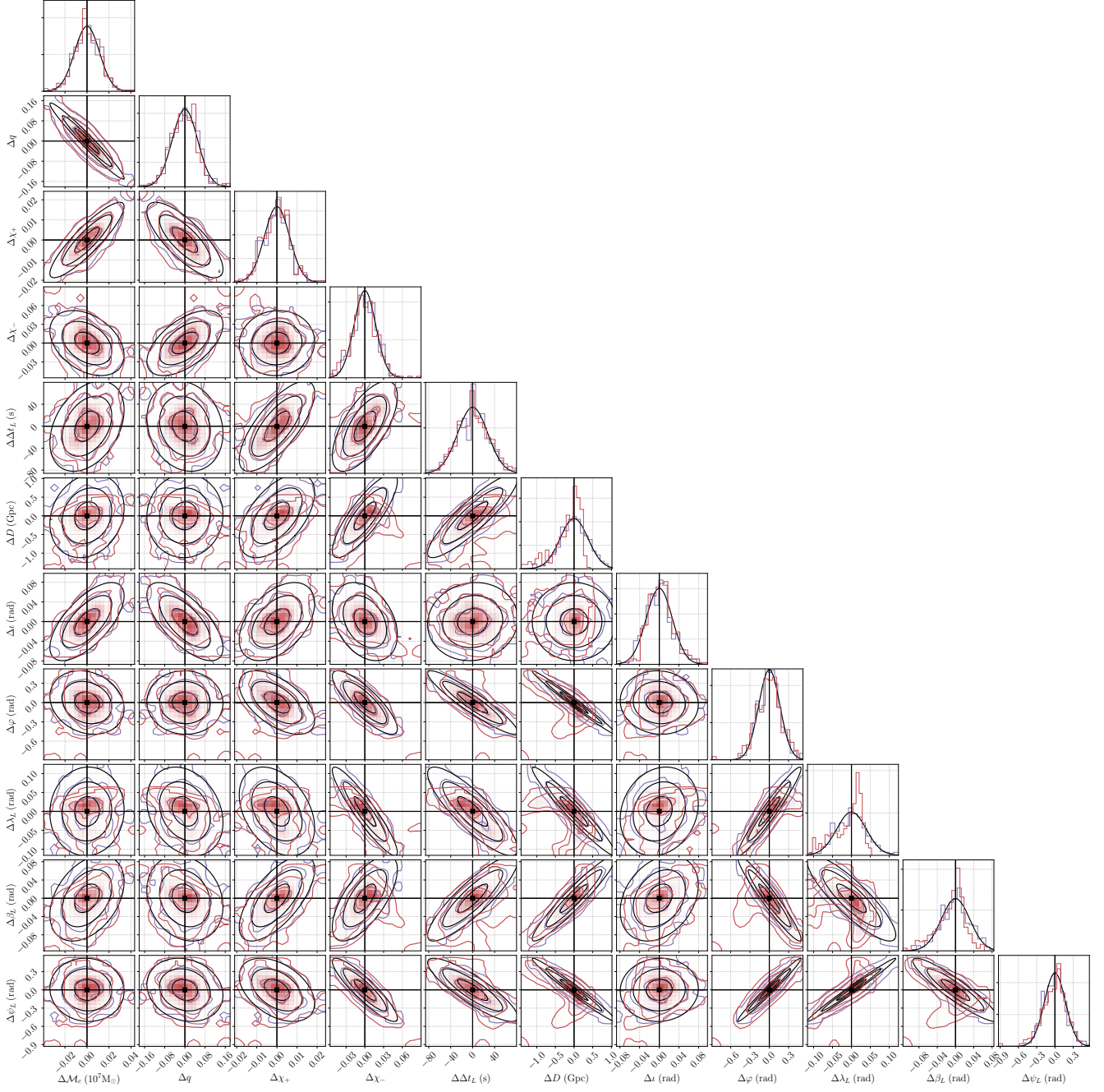


Figure 17. In this figure, which is not a corner plot for a posterior, we are checking the agreement between the Fisher-based estimation for max-likelihood estimates and max-likelihood parameters found in full PE runs. We use the TDI2 noise curve with MBH primary mass $M = 3e7$, with a gap during the merger, and consider the Whittle likelihood with gated data (the configuration of Fig. 14). In red we plot the histogram of max-likelihood values obtained from full PE using 350 different noise realisations. In blue we plot the histogram of max-likelihood values calculated through Eq.(87b). Finally, the black curve is a Gaussian distribution, centered on the true parameter with parameter covariance matrix equivalent to Eq.(87c). Note that the scale is different from the previous corner plots: the width shown here is set by the scale of the posterior *scatter* of Fig. 14, which is much larger than the posterior *width* for this case where $\Upsilon \gg 1$.

Fisher results are accurate enough even though individual posteriors might be non-Gaussian. We find an overall good agreement, with some overall noise that is due to the limited number of simulations (350). We see some deviations in some parameters, for instance D and λ_L , that show a clear deviation between the two distributions. Given that the gated Whittle has large biases and plays the role of a stress-test here, we conclude that the estimates shown in Tab. IV and Tab. V are at least qualitatively reliable.

C. Influence of the tapering length

The effect of introducing a smooth tapering window is crucial, in both reducing the mismodeling ratio Υ (improving the statistical consistency of posteriors) at the price of increasing the mismodeling ratio Ξ (broadening the posteriors and erasing information). We found a strong dependency of the results on the chosen tapering length, and we explore this dependency in Fig. 18 for the full-segment Whittle-likelihood of Sec. IV B and Fig. 19 for the segmented Whittle-likelihood of Sec. IV C.

In both these figures, we show the dependency of the SNR, Ξ and Υ on the tapering length, exploring lengths chosen to be powers-of-2 in minutes from 1 to 128. We focus on the TDI2 case, with a gap at merger. We recall that the gap length itself is different for the two masses, of 1 hour for M3e7 and of 15 minutes for M3e6. We also show individual curves for the different parameters Ξ_a , Υ_a , illustrating that the trends are similar across parameters. We also show the reference value of the SNR obtained with the correct treatment of the gap.

Both the Whittle and the segmented Whittle-likelihood show the same qualitative trends. We see that increasing the lobe length washes away information, as expected: this is evident in both SNR and Ξ . There is a marked difference between the M3e7 (1-hour gap, lower-frequency signal) and M3e6 (15-minutes gap, higher-frequency signal) cases: Ξ degrades much faster for increasing lobe length for M3e6, while for M3e7 there is an extended range where $\Xi \simeq 1$. For the shortest lobe lengths, in both cases, we obtain an overestimated SNR and values $\Xi < 1$. The behavior of Υ , with an explosion towards $\Upsilon > 1$, particularly pronounced for the low-mass case M3e6, and values $\Upsilon \lesssim 1$ for longer tapering lobes.

The results for the Whittle and the segmented Whittle-likelihood are almost identical for both the SNR and Ξ . The behavior in Υ is also close, with one noticeable difference for short tapering lengths in the M3e6 case, where the segmented Whittle model seems to perform better with a more limited increase.

Overall, Fig. 18 and Fig. 19 indicate the existence of an optimal tapering length, which is in line with intuition: the tapering window must alleviate the statisti-

cal inconsistencies concentrated near the edges of the gap, while avoiding erasing too much of the information conveyed by the signal there. We find indeed that Ξ increases, Υ decreases, for an increasing lobe length. In our case, it seems that the value of 10 minutes shown in Tab. IV and Tab. V is close to optimal (while a bit high for the M3e6 case).

Our tools allow to look systematically for an optimal tapering length, optimizing for both Υ and Ξ simultaneously, or possibly optimizing their product $\Xi\Upsilon$ which is bounded from below by 1 and whose dependency on tapering length has a characteristic U-shape. This could be done quite fast when working with Fisher matrices and linear signal approximation biases, without the need for full PE runs.

More explorations would be needed to understand the dependency of the optimal tapering length: does it depend primarily on the length of the gap? On the morphology of the signal, in particular its high-frequency content? How much does this optimum vary for different signals? We leave these questions for future work.

D. Influence of the cadence

We conclude this exploration of gap mismodeling with an important observation. The frequency resolution in the discrete FD is $\Delta f = 1/T$, where T is the total duration of the data. On the other hand, the cadence or time step Δt drives the frequency extent, through $f_{\text{Nyquist}} = 1/(2\Delta t)$. The lowest and highest frequencies present in the data are therefore driven by $1/T$ and $1/(2\Delta t)$ respectively. Considering the noise PSDs illustrated in Fig. 2, we can see that for a red process T will affect the importance of the red noise, while for a blue process Δt will affect the importance of the blue noise.

All our results presented above were produced with a cadence $\Delta t = 10\text{s}$ (while the L1 TDI data is rather expected to use $\Delta t = 5\text{s}$ – this factor of 2 in the data size N eases our computations). Our most massive system, with $M = 3 \times 10^7 M_\odot$, is so massive that a Nyquist frequency of 5mHz (a cadence of $\Delta t = 100\text{s}$) would be sufficient to resolve it.

In Tab. VI, we contrast the mismodeling ratio Υ that we would obtain with a cadence of $\Delta t = 10\text{s}$ or $\Delta t = 100\text{s}$, for the gated Whittle-likelihood. We see that changing the cadence would strongly change the mismodeling ratio for TDI2, reducing it from ~ 12 to ~ 1.4 . We attribute this to the fact that TDI2 is a blue process, and the smaller sampling rate reduces the relative importance of the high-frequency noise. This change would have a moderate effect in the TDI1 case, and a worsening effect in the TDI0 case, mostly likely indicative of a breakdown in the linear signal approximation, since for large values of Υ the best-fit param-

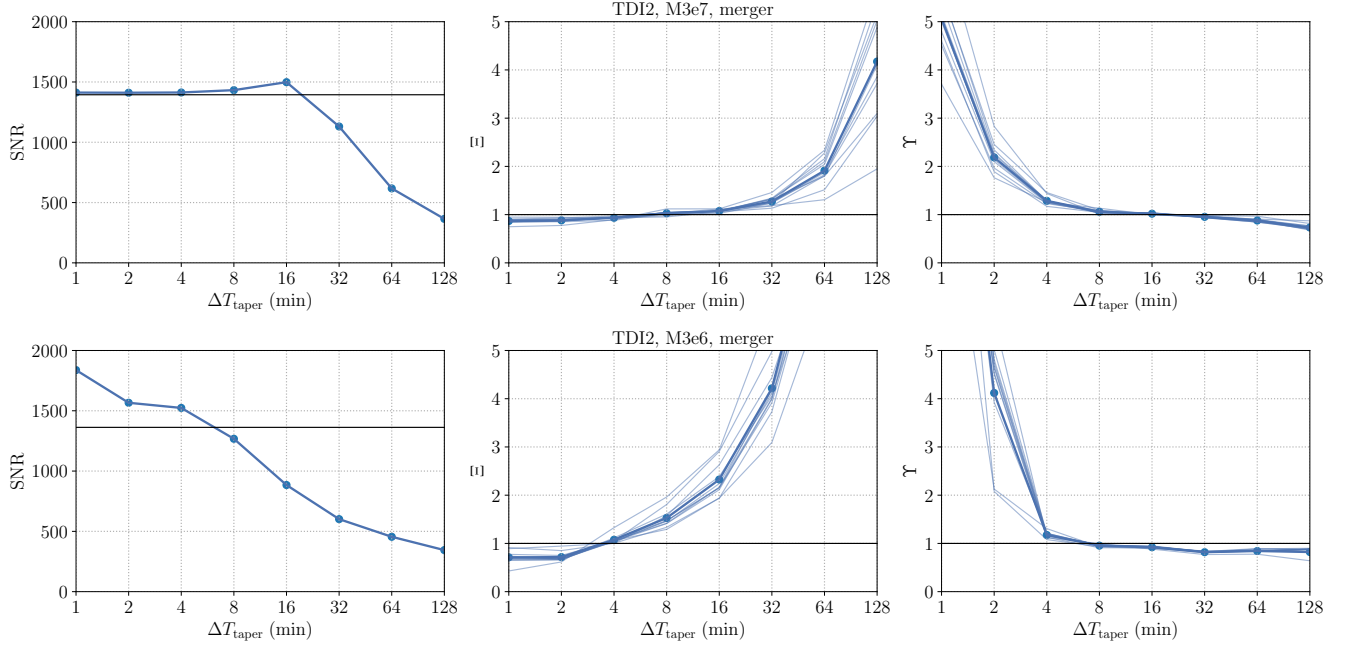


Figure 18. **(Top row from left to right):** Respective SNR, Ξ and Υ computations as a function of the lobe-length in minutes for the $M = 3e7$ case. The horizontal black line represents the true values with model TD noise covariance matrix consistent with the gated TD noise-covariance matrix. The blue dots are computations for SNR, Υ and Ξ using a Whittle-based covariance matrix assuming dependence between segments. The fainter blue lines are the individual parameter Υ_a and Ξ_a values. **(Bottom row from left to right):** The same set up as the top row, but with the M3e6 case.

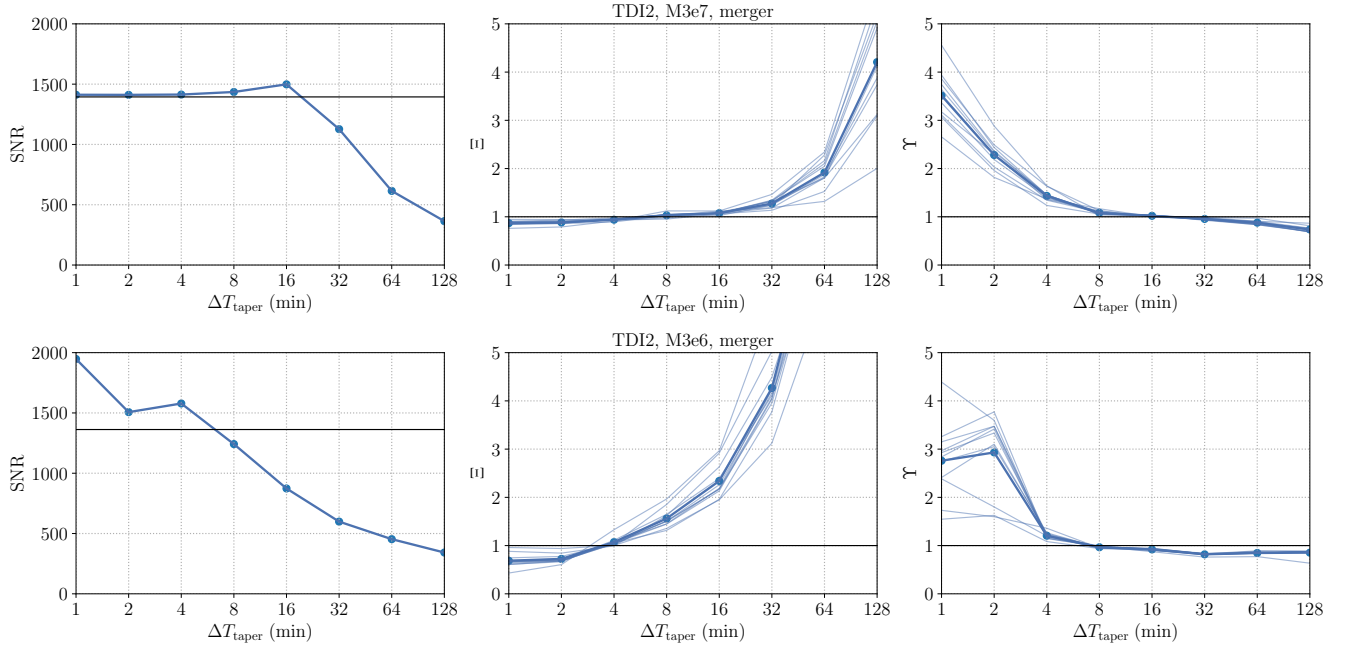


Figure 19. The same set up as figure 18 except with model covariance matrix replaced by the Segmented-Whittle covariance (assuming independence between gap segments).

Mismodeling $\bar{\gamma}$ data gaps: effect of cadence – M3e7 merger		TDI2		TDI1		TDI0	
Data	Model	$dt = 10s$	$dt = 100s$	$dt = 10s$	$dt = 100s$	$dt = 10s$	$dt = 100s$
Coherent, gap	Whittle, gated	12.2	1.4	1.2	1.0	39.7	74.6

Table VI. A similar set up to Tab. IV but only focusing on the M3e7 case with gap at merger, using a model covariance matrix consistent with Whittle and gating the data. The two variables changed in our experiments here are the cadence $\Delta t \in [10, 100]$ seconds and noise curve TDI2/1/0 respectively from left to right.

ters are many σ 's away from the truth.

We conclude that the cadence can have an effect on the gap mismodeling that must not be overlooked. Direct downsampling of the data, prior to analysis, is an operation that could be implemented easily, although only relevant for low-frequency signals. Our results might change when using the more realistic cadence of $\Delta t = 5s$. The effect of the total length on red noise would also need to be explored, which we leave for future work.

E. Mismodeling data segment independence

Data segment independence might be an important question for LISA data analysis, as interruptions in the data stream might be of different nature. At one extreme, for missing data, the data gap acts as a mask and the underlying unmasked noise process $\mathbf{n}$ does not lose coherence and is not affected by the gap at all. At another extreme, the data gap corresponds to a major disturbance of the instrument, and the data before and after the gap are best represented as pertaining to two separate, statistically independent experiments. Note that the latter case is a hard limit for data imputation methods [43–45, 97], as they cannot invent data joining two independent processes in a consistent way. For a real instrument, the truth might lie somewhere in-between these two extremes.

More generally, we might want to cut the data stream in different independent pieces for computational reasons, in order to keep the length of each data sub-segment low. By construction, under the independence assumption, the covariance matrix becomes block-diagonal, with blocks that are much more manageable than the full $N \times N$ matrix. Another motivation can be to improve the stationary nature on each sub-segment by considering a different PSD on each, better resolving PSD drifts and slow variations in the GB foreground.

Therefore, we investigate the mismodeling errors induced by both scenarios: assuming that the data segments are independent when they are in fact segments from a coherent stochastic process; and assuming that the data segments are coherent when they are in fact

independent. Here we want to isolate the effect of segment independence from the effect of gap mismodeling. To this end, we do not use the Whittle-likelihood but use the correct gap treatment of Sec. III B.

In the first scenario, where the data is coherent but the model independent, we have for the data and model respectively

$$\begin{aligned} \langle (\mathbf{W}\mathbf{n})(\mathbf{W}\mathbf{n})^T \rangle &= \mathbf{W}\Sigma\mathbf{W} = \begin{pmatrix} \Sigma_{00} & \mathbf{0} & \Sigma_{02} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Sigma_{02}^T & \mathbf{0} & \Sigma_{22} \end{pmatrix}, \\ (\Sigma')^+ &= \begin{pmatrix} \Sigma_{00}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{22}^{-1} \end{pmatrix}. \end{aligned} \quad (108)$$

We remark here that the configurations presented in Eq. (108) are subtly different from the model specified by Eq. (88), in Sec. IV C. Notice that the individual blocks in Σ_{00} and Σ_{22} are Toeplitz and, in the model assuming incoherence, we are using block inverses of those specific Toeplitz matrices. This isolates two sources of mismodeling that was present in Eq. (108), where one assumes the underlying process is *circulant* and the data segments between gaps are *independent*. In this case, we are mismodeling independence but correctly modeling the underlying process as Toeplitz. The model likelihood here Σ' , although modeling stationary noise, is *not* Whittle-based.

In the second scenario, where the data is independent but the model is coherent, we exchange the role of the two matrices for the data and model:

$$\begin{aligned} \langle (\mathbf{W}\mathbf{n})(\mathbf{W}\mathbf{n})^T \rangle &= \mathbf{W}\Sigma\mathbf{W} = \begin{pmatrix} \Sigma_{00} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{22} \end{pmatrix}, \\ (\Sigma')^+ &= \begin{pmatrix} \Sigma_{00} & \mathbf{0} & \Sigma_{02} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \Sigma_{02}^T & \mathbf{0} & \Sigma_{22} \end{pmatrix}^+. \end{aligned} \quad (109)$$

We extend the analysis by considering a data split, without any missing data between the two data segments. This situation is closest to the voluntary segmentation that one might implement for long signals for computational reasons [98–102]. It represents a useful check: the longer the gap is, the more justified it

Mismodeling Υ segment independence		TDI2				TDI1				TDI0			
		Merger		Insp.		Merger		Insp.		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6
Data	Model												
Coherent, gap	Incoherent, gap	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Incoherent, gap	Coherent, gap	1.2	1.1	1.1	1.5	1.0	1.0	1.0	1.0	26.2	8.1	13.4	14.2
Coherent, split	Incoherent, split	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Incoherent, split	Coherent, split	5.6	2.1	1.3	1.5	1.3	1.6	1.0	1.0	27.3	41.0	11.8	11.9

Table VII. The same set up as Tab. IV with different assumptions. The data generation process “Data” and the “Model” used for the likelihood all start from the same underlying circulant full-segment covariance and apply the correct treatment of the gap from Sec. III. For “Coherent” the covariance blocks describing correlations between before/after gap segments are kept, for “Incoherent” they are ignored (see Eqs. (108) and (109)). In the second part of the table, “Split” indicates that the data stream is directly split into two parts, with no gap in-between.

Mismodeling Ξ segment independence		TDI2				TDI1				TDI0			
		Merger		Insp.		Merger		Insp.		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6
Data	Model												
Coherent, gap	Incoherent, gap	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Incoherent, gap	Coherent, gap	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Coherent, split	Incoherent, split	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.0
Incoherent, split	Coherent, split	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.0

Table VIII. The same set up as Tab. VII, but computing Ξ instead.

will be to treat the data before/after the gap as independent, since correlations will be dominated by short timescales (the near-diagonal elements of the covariance matrix); a zero-length gap is therefore a worst-case scenario for ignoring correlations between segments present in the data. In this scenario, the gated covariance and the pseudo-inverse model covariance matrices are the same as in Eq. (108) and Eq. (109), but eliminating the middle line and column since there is no gap present.

We report the resulting mismodeling ratios in Tab. VII and Tab. VIII. The results for the case where the data is coherent (dependent) but the model incoherent (independent) show very little mismodeling in both Υ and Ξ , whether the gap is present or whether we simply split the data in two segments. We find small losses of SNR and deviation of Ξ from 1, but they are limited and below 1%. Our intuition would tell us that ignoring cross-block terms in the covariance matrix means that we choose to neglect some information by ignoring correlations. This test shows that the amount of information erased is negligible. By contrast, in the case where the data is incoherent but the model is coherent, while we still find $\Xi \simeq 1$, we also find $\Upsilon > 1$ in most cases. These values are very large for TDI0, but also deviate significantly from 1 for TDI2, and are worse for the case where the data is simply split in two. Here, our

intuition would say that we are analyzing the data while inventing correlations that are not here in reality. The presence of the gaps (putting some distance between the two segments) alleviates but does not eliminate the issue.

The near-absence of mismodeling errors when enforcing segment independence, even without gaps, at the expense of a negligible loss of information, is a validation of methods based on data segmentation. It is also consistent with the findings of Sec. VII A where the full-segment Whittle model and the segmented Whittle model were giving comparable results. On the other hand, the presence of mismodeling errors when enforcing data coherence on independent segments is a cautionary tale for treating instrument disturbances as missing data in a coherent process. In short, wrongly assuming coherence can be dangerous, assuming independence is safe.

We did not explore the impact of tapering windows in this experiment. Although the correct gap treatment does not require any tapering, in the present case the independence represents another source of mismodeling, and it is possible that tapering would alleviate the mismodeling in the case where the data is incoherent and the model coherent (at the expense of erasing some information, as always).

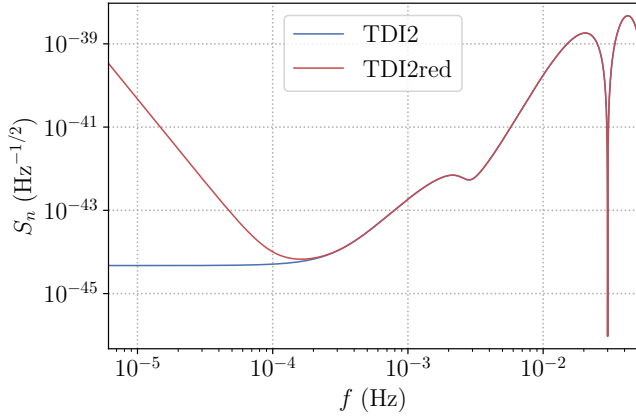


Figure 20. **(In blue):** The usual noise curve for TDI2 given by Eq. (103c). **(In red):** The noise curve for TDI2 but with a f^{-4} red noise degradation at low frequencies, as given in (110).

F. Mismodeling the noise PSD at low frequencies

We have seen in previous sessions that the frequency content of the noise process, from a red process like TDI0 to a blue process like TDI2, can strongly affect the importance of mismodeling errors. Another important point is the role of mismodeling the low-frequency noise.

In the standard setting of gravitational-wave data analysis, the stationarity justifies independence in the FD, and the bucket shape of the instrument's sensitivity justifies ignoring the low and high frequency regions where all signals would be drowned in noise. The lower frequencies, in particular, are a difficult limit to build a statistical understanding of any instrument: longer data stretches are needed to resolve lower frequencies, and stationarity makes less sense as an assumption. For instance, this is true of LIGO-Virgo instruments [Ollie: help] and also of LISA Pathfinder [21, 23, 35, 36]. PSD estimation itself is challenged at low frequencies, with the lack of data challenging statistical averaging procedures (such as Welch's) and resulting in a large uncertainty [103–106].

Yet the presence of gaps breaks the stationary feature of the noise in principle, and allows for correlations across the frequency band. Hence a natural question: do we need to model accurately the PSD at low frequencies to avoid introducing noise mismodeling errors?

We will address this question by focusing on the TDI2 case. As shown in Fig. 20, a simple noise model for TDI2, dominated by point-mass noise at low frequencies, gives white noise below $\sim 10^{-4}$ Hz. However, it is possible that other unmodeled sources of noise would start dominating there; and it is likely that the instrument models that we build on the ground would become more inaccurate in that range. We will consider

an alternative PSD with a f^{-4} red noise explosion below $\sim 10^{-4}$ Hz, following (see (102))

$$P_{\text{acc}}(f) \rightarrow P_{\text{acc}}(f) \left[1 + \left(\frac{f_0}{f} \right)^\alpha \right], \quad (110)$$

with the values $f_0 = 10^{-4}$ Hz, $\alpha = 4$. The corresponding PSD, that we will call TDI2_{red}, is shown in Fig 20. We stress that this model is *ad hoc* and not based on any detailed instrumental considerations. In the following, it will serve as a toy model to investigate the effect of mismodeling this type of additional red noise.

Assuming that the true PSD of the noise process is TDI2_{red} while the PSD TDI2 is used in the model, we compute mismodeling ratios Υ and Ξ in different gap scenarios and report the results in Tab. IX and Tab. X.

In the first scenario, we do not introduce any gap in the data stream. The two processes are stationary (and circulant) so the covariance is diagonal in the FD. We are still mismodeling the PSD at low frequencies, but even with the underestimation of the noise there the signal is drowned in noise, neutralizing mismodeling errors. We check that $\Upsilon = 1$ and $\Xi = 1$ in that case.

In the second scenario, we introduce a gap that is correctly modeled, while mismodeling the low- f PSD. That is to say, the model likelihood respects the construction of Sec. III B, but starting with a PSD that is TDI2 instead of TDI2_{red}. While we obtain $\Xi \simeq 1$, we get values $\Upsilon > 1$ in all cases, the worst being the M3e7 case with the gap at merger. This means that, due to the presence of gaps, the low- f PSD mismodeling contaminates the analysis, even though the gap themselves are properly accounted for.

In the third scenario (lines 3-5) we investigate the case where we both: (i) mismodel the PSD at low- f and (ii) mismodel the gap by using the Whittle-likelihood, as in Sec. IV A, with different tapering windows. With a simple gating window, we find large values $\Upsilon > 1$ and also values of $\Xi < 1$. Both issues are mitigated but not eliminated by introducing a tapering window. In particular, we find some mismodeling $\Upsilon > 1$ (but $\Xi \simeq 1$) even for the inspiral gap, which contrasts with Tab. IV.

These results call for caution when dealing with mismodeled low-frequency noise, even when correctly taking the gaps into account. However, we did not explore possible simple solutions. In the stationary case without gaps, applying a simple high-pass filter to the data eliminates the low-frequency content. One could construct a filtering procedure, properly generalized to work in the presence of gaps. We also note that the large uncertainty on the low- f part of the PSD should be marginalized over as part of the LISA global fit. We leave these questions for future work.

Mismodeling Υ low- f PSD		TDI2			
		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6
Data	Model				
PSD TDI2 _{red} , no gap	PSD TDI2, no gap	1.0	1.0	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2, gap	2.2	1.3	1.2	1.1
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, gated	13.4	40.9	1.6	1.4
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, taper 10min	3.6	1.2	1.3	1.1
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, taper 30min	2.2	1.0	1.2	1.1
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, gated	9.6	39.0	1.2	1.2
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, taper 10min	1.6	1.3	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, taper 30min	1.2	1.1	1.0	1.0

Table IX. The same set up as Tab. IV, except that we test for the mismodeling of noise processes with unknown large red noise components. Here TDI2_{red} is the TDI2 noise curve with an f^{-4} explosion at low frequencies given in Eq. (110) and shown in Fig. 20, while TDI2 is the usual noise curve that approaches a constant as $f \rightarrow 0^+$. In the first line, no gap is present and the only mismodeling is that of the PSD. In all other cases, the data features a gap. In the second line, the model likelihood incorporates the gap correctly according to Sec. III, while using the wrong PSD. In the Whittle case (lines 3-5), either gated or tapered, the same models are used as in Tab. IV, but this time the PSD is another source of mismodeling. We also show in the last three lines the result for a Whittle likelihood model where the PSD TDI2_{red} is correctly modeled.

Mismodeling Ξ low- f PSD		TDI2			
		Merger		Insp.	
		M3e7	M3e6	M3e7	M3e6
Data	Model				
PSD TDI2 _{red} , no gap	PSD TDI2, no gap	1.0	1.0	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2, gap	1.0	1.0	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, gated	0.8	0.5	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, taper 10min	1.0	1.7	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2, Whittle, taper 30min	1.2	3.9	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, gated	0.9	0.5	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, taper 10min	1.1	1.7	1.0	1.0
PSD TDI2 _{red} , gap	PSD TDI2 _{red} , Whittle, taper 30min	1.2	3.9	1.0	1.0

Table X. The same set up as Tab. IX, but computing Ξ instead.

G. Mismodeling the stationary process as circulant

Finally, we will close our mismodeling investigations with the a limited investigation of the effect of mismodeling a Toeplitz covariance matrix as a circulant matrix. As explained in Sec. IV A, the true covariance of the stationary noise process will have a Toeplitz structure. But the FD covariance is only diagonal when the covariance is circulant. In our previous tests, we have taken a circulant covariance over the full segment to be the true covariance, because we wanted to isolate the different sources of mismodeling. For a reminder, we refer the reader to figures 3 and 4 for a visual illustration of the differences between circulant and Toeplitz matrices respectively. It may also be worth reading the caption and the surrounding text of Fig.(5) of the potential impact

of mismodeling a Toeplitz process as circulant.

We now investigate the Toeplitz-circulant mismodeling errors more rigorously by considering a data stream with no gaps, computing mismodeling ratios when the true covariance is Toeplitz while the model likelihood is circulant (Whittle). The results for Υ are reported in Tab. XI, comparing two scenarios: in one, we use the raw data; in the other, we taper the data at each end of the data segment.

We can see that we find $\Upsilon > 1$ for TDI2, large values of Υ for TDI0, while $\Upsilon \simeq 1$ TDI1. Introducing a tapering window gives $\Upsilon \simeq 1$ for TDI2, and reduces but does not fully eliminate the mismodeling for TDI0. We do not present a table for Ξ since in all cases we find $\Xi \simeq 1$. The tapering introduced here is at the edges of the data segment, away from the merger that dominates the SNR.

Mismodeling $\bar{\Upsilon}$ Toeplitz/circulant		TDI2		TDI1		TDI0	
Data	Model	M3e7	M3e6	M3e7	M3e6	M3e7	M3e6
Toeplitz, no gap	Circulant, no gap, no taper	1.1	1.4	1.0	1.0	37.3	37.3
Toeplitz, no gap	Circulant, no gap, taper 1hr	1.0	1.0	1.0	1.0	2.1	1.3

Table XI. The same set up as Tab. IV, except computing mismodeling errors $\bar{\Upsilon}$ under the assumption that the underlying noise process is circulant when in reality it is Toeplitz. We either keep the data as it is or apply a taper of 1hr at both ends of the data segment. We do not incorporate data gaps in this specific study.

Note also that, when using a segmented Whittle-likelihood, the model covariance is circulant on each segment while the true covariance, taken to be circulant over the full range of the data, is not circulant on the segments. Yet, in Tab. IV and Tab. V, the segmented Whittle and full Whittle (i.e. coherent) give similar results, pointing towards a limited impact of this mismodeling.

We conclude that tapering the ends of the data seems to prevent mismodeling errors due to the use of a circulant structure, except for the red-noise dominated process TDI0. Note that when Whittle is used in practice, it is generally with a tapering window. Transient signals that are confined to the inside of the data interval will not be affected much by such a window tapering the end points of the data.

VIII. CONCLUSIONS AND OUTLOOK

A. Conclusions

The primary focus of this work is to understand the impact data gaps have on PE studies for GW signals in LISA data. Data gaps, whether planned or unplanned, are unavoidable and will have to be included in future LISA data analysis pipelines. In our work, limiting ourselves to a stationary Gaussian underlying noise process, we have explored how the presence of gaps breaks the stationary feature of the observed data.

Working in the TD, in Sec. III we presented a formalism describing the likelihood of the observed data obtained by marginalizing over unobserved data values, then translated the result to another point of view where the gap is represented by a gating window applied to the original-sized data. The use of the Moore-Penrose pseudo-inverse is crucial, allowing for a direct translation of all results to the discrete FD. The formalism allows for the introduction of smooth windows tapering each gap edge. This approach relies on using $N \times N$ covariance matrices and is limited to short data stretches, however it represents an exact treatment of gaps and can be used to assess mismodeling errors induced by other approximate treatments. The concepts within this pa-

per are robust to any family of data gaps LISA may expect – be them scheduled/unplanned gaps of any duration.

We developed a formalism to describe these mismodeling errors in Sec. IV, based on the FM approach. Mismodeling the noise process will lead to biases in the PE that, although zero-mean, will have the wrong statistics. For a generic mismodeled Gaussian likelihood, We introduce two complementary error measures: Ξ in Eq.(92) that measures the broadening of the posteriors, describing how much information is erased (notably, by the use of tapering windows), and Υ in Eq.(93), which measures the bias covariance (scatter of the max-likelihood) in units of the σ of the model posterior, describing the statistical inconsistency induced by the mismodeling. These error measures can be computed for any Gaussian model likelihood, as long as the true process can be simulated, and are much faster to evaluate than PP plots requiring full Bayesian PE.

We focused on MBHB signals in LISA, as described in Sec. V. We only considered a short data stretch of less than one day, which is enough to represent the merger of MBHB signals if their masses are large enough, and a limited array of cases, for a gap at merger or in the late inspiral, and for a heavier or lighter black hole. We investigate different TDI generations (TDI2, TDI1, TDI0), to explore the effect of the frequency content of the noise, from blue noise at high frequencies to red noise at low frequencies. In Sec. VI we compared covariance matrices for both circulant and Toeplitz structured covariance matrices, showing qualitatively their differences in both the TD and FD. We also demonstrated the impact of gaps with/without tapering on the covariance matrices, showing that hard cut-offs in the window function used to mimic gaps introduce highly non-trivial leakage in the FD noise-covariance matrix, giving both dense and degenerate matrices. Tapering concentrates the power more along the main diagonals, but degeneracy is still present. We then demonstrated a practical method to compute the pseudo-inverse of both the TD and FD noise covariance matrices by identifying the singular values corresponding to the gaps and excluding them.

We then applied our error measure formalism to different mismodeling cases in Sec VII, exploring the ef-

fects of mismodeling the gaps, data segment independence, and the low-frequency PSD, as well as a limited exploration of the mismodeling of the true Toeplitz covariance with a circulant covariance. We also ran a set of full PE runs validating our approach. A highlight of our findings would include:

- Assuming the correct pseudo-inverse for the gated covariance matrix has been calculated, the likelihood is invariant to whether the model templates are gated or not. See Eq. (40) and Eq. (52).
- There is *no* loss of information concerning the tapered signal assuming that the tapering scheme has been correctly incorporated into the pseudo-inverse of the windowed noise covariance matrix.
- There is a strong contrast between blue and red-noise processes; red noise is associated with long-lived correlations, and gap mismodeling is the worst in that case, but a naive gating window with no smooth taper also gives large errors for the blue process TDI2. This is demonstrated in Tab. IV A.
- A gap during the inspiral gives much smaller errors than a worst-case gap happening at merger for MBHs. This is consistent with the analysis presented in [26, 44, 45, 49].
- Introducing tapering windows reduces statistical inconsistency errors significantly, at the expense of broadened posteriors.
- Adjusting the tapering length makes a significant difference, with an exchange between statistical inconsistency and information erasure; our methods can be used to cheaply find an optimal tapering length.
- A segmented Whittle-likelihood, treating the data before/after gap as independent, gives results that are similar to a full-segment Whittle-likelihood.
- Treating data segments as independent when they are in reality correlated, even in the case of a data split (no in-between gap), gives little to no errors.
- Treating data segments as correlated when they are in reality independent leads to significant errors.
- mismodeling the low- f PSD by underestimating the red noise leads to significant errors, even when properly taking the gaps into account. Tapering mildly alleviates this issue, but does not completely remove it.
- Treating the data as circulant when it is in fact Toeplitz appears reasonable provided smooth tapers are applied to the data stream.

We believe that the methodology discussed in this paper is a suitable procedure for assessing noise mismodeling errors in presence of data gaps and may be essential in validating LISA global fit pipelines. When in doubt, we strongly suggest that you smoothly taper your data.

B. Future work and outlook

The present paper presents a preliminary exploration of gap mismodeling, calling for many extensions, some which could be achieved with the same formalism. One strong limitation of the results presented is that we only considered a short stretch of data, appropriate only for merger-dominated MBHB signals. This choice was made to allow for a complete exploration and thorough validation of the method, comparing with the exact treatment of the gaps requiring to work with $N \times N$ matrices. However, assessing the mismodeling error Υ , representing statistical inconsistency, is in fact only as expensive as working with the fast approximate likelihood, provided the true noise can be simulated (note that this is not true for Ξ , which compares the model Fisher covariance with the true model covariance). Having built confidence in the methods, they could be applied to longer signals: GBs, EMRIs, SBHBs. In particular, the segmented likelihood seems a promising avenue to explore and our method could assess the systematics introduced.

We also limited ourselves to a few examples, and a more systematic study of the role of the gap length, signal characteristics. Our method applies to any data plan, for instance for short repeated gaps.

Some of our formalism translates directly to other contexts, where instead of a gap we have an abrupt termination of the data. In ringdown analyses [28, 31, 107–109] the analyst purposely selects data that starts shortly after the amplitude peak, right where the signal is strongest. In pre-merger analyses of MBHBs [34], the observed data ends right where the signal is the loudest.

While we saw that the low-frequency noise can be responsible for noise mismodeling errors, we did not explore how filtering or whitening methods, that can readily eliminate low- f noise in the stationary case, can be adapted to the presence of gaps.

Beyond data gaps, the same mismodeling formalism could be used to assess other approximations and errors made in the covariance matrix. The mismodeling ratios that we introduced allow for a much faster computation than building PP plots, which requires repeating full PE runs for many noise realizations. The approximations that we could explore include non stationary in the true noise process, either in the form of instrumental PSD drifts or in the form of annual modulations

of the GB confusion foreground. Our formalism only requires that the model likelihood is Gaussian, and that the true noise can be simulated; as such, it could also be used to explore noise mismodeling errors caused by non-Gaussian features, such as distributions with heavier tails [110–112] or glitches [22, 113, 114].

Time-frequency methods using the short-time-fourier-transforms (STFTs) [115–117] or Wilson-Daubechies-Meyer (WDM) basis functions [118] appear to be growing in popularity. Within the WDM formalism, the window functions in both time and frequency are local, which have proved advantageous on data streams with mild non-stationary features (e.g., drifts to the PSD) [13, 119, 120]. We have demonstrated in our TD / FD analysis that it may be reasonable to assume independence between segments if the underlying data stream is dependent. The same could be said for analyses conducted in the time-frequency domain. However, we are certain that working in time-frequency would still be negatively impacted by noise processes with significant red noise. Similarly, time-frequency analyses are not invincible to artifacts arising from edge-effects when due to working with finite length time-series as discussed in [120]. This rich body of work will be left to future work.

IX. ACKNOWLEDGEMENTS

O.Burke and S.Marsat made equal contributions in the preparation of this work. They also acknowl-

edge support from the French space agency CNES in the framework of LISA. O.Burke also acknowledges support from the UK Space Agency with Grant No. ST/Y004906/1. O.Burke and S.Marsat thank John Baker and Christian Chapman-Bird for their careful reading of our manuscript. O.Burke shows immense appreciation towards collaborator and friend Lorenzo Speri for pushing him to complete this work. We express our deepest thanks towards Quentin Baghi, Avi Vajpeyi, Giorgio Mentasti, Matthew C. Edwards, Vasco Gennari, Antoine Petiteau, Natalia Korsakova, Michael J. Williams, Graham Woan, Peter Wolf, Jean Baptiste-Bayle and Olaf Hartwig, all for insightful discussions during the preparation of this work. This work has made use of `pastamarkers` [121].

Appendix A: Discrete Fourier transform of the time-domain covariance

1. The circulant case

We will now derive a FD equivalent of eq. (9), for the matrix $\langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^\dagger \rangle$. In our notations,

$$\tilde{\mathbf{n}} = \Delta t \sqrt{N} \mathbf{P} \mathbf{n}, \quad (\text{A1a})$$

$$\tilde{\mathbf{n}}^\dagger = \Delta t \sqrt{N} \mathbf{n}^T \mathbf{P}^\dagger, \quad (\text{A1b})$$

so that the FD analog of the covariance reads

$$\tilde{\Sigma} \equiv \langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^\dagger \rangle = N \Delta t^2 \mathbf{P} \langle \mathbf{n} \mathbf{n}^T \rangle \mathbf{P}^\dagger = N \Delta t^2 \mathbf{P} \Sigma \mathbf{P}^\dagger. \quad (\text{A2})$$

We can compute explicitly (with implicit labels on sums extending from 0 to $N - 1$):

$$(\mathbf{P} \Sigma \mathbf{P}^\dagger)_{ij} = \frac{1}{N} \sum_k \sum_l \omega^{-ik} \Sigma_{kl} \omega^{lj}, \quad (\text{A3a})$$

$$= \frac{1}{N} \sum_k \sum_l \omega^{-ik} \omega^{lj} \sigma_{l-k}, \quad (\text{A3b})$$

$$= \frac{1}{N} \sum_k \sum_{l=0}^{k-1} \omega^{-ik} \omega^{lj} \sigma_{k-l} + \frac{1}{N} \sum_k \sum_{l=k}^{N-1} \omega^{-ik} \omega^{lj} \sigma_{l-k}, \quad (\text{A3c})$$

$$= \frac{1}{N} \sum_k \sum_{m=1}^k \omega^{k(j-i)} \omega^{-mj} \sigma_m + \frac{1}{N} \sum_k \sum_{m=0}^{N-k-1} \omega^{k(j-i)} \omega^{mj} \sigma_m, \quad (\text{A3d})$$

where we used the Toeplitz structure, the symmetry of σ for negative indices, and changed variables to $m = k - l$ and $m = k + l$. Using the circulant structure, we can rewrite the first term as:

$$\frac{1}{N} \sum_k \sum_{m=1}^k \omega^{k(j-i)} \omega^{-mj} \sigma_m = \frac{1}{N} \sum_k \sum_{m=1}^k \omega^{k(j-i)} \omega^{-mj} \sigma_{N-m} = \frac{1}{N} \sum_k \sum_{m=N-k}^{N-1} \omega^{k(j-i)} \omega^{mj} \sigma_m, \quad (\text{A4})$$

where we used $\omega^N = 1$. We can now gather terms as

$$(P\Sigma P^\dagger)_{ij} = \left(\frac{1}{N} \sum_k \omega^{k(j-i)} \right) \sum_{m=0}^{N-1} \omega^{mj} \sigma_m = \delta_{ij} \sum_{m=0}^{N-1} \omega^{-mj} \sigma_m = \delta_{ij} \sqrt{N} (P\sigma)_i, \quad (\text{A5})$$

where we used the circulant condition again in the last equality to change the sign of the exponent.

We can also do a similar computation with the transpose of $\tilde{N}$ instead of its Hermitian conjugate, which will allow us to complete the description of the statistics of the real and imaginary parts of $\tilde{N}$. This is a repetition

of the computation of $P\Sigma P^\dagger$, changing simply the sign of the exponent of ω . Without detailing each step, we get (mind the special case $i = j = 0$)

$$(P\Sigma P^T)_{ij} = \left(\frac{1}{N} \sum_k \omega^{-k(i+j)} \right) \sum_{m=0}^{N-1} \omega^{-mj} \sigma_m = (\delta_{i,0} \delta_{j,0} + \delta_{N,i+j}) \sum_{m=0}^{N-1} \omega^{-mj} \sigma_m. \quad (\text{A6})$$

We arrive at (note that the last index is j and not i)

$$(P\Sigma P^T)_{ij} = \sqrt{N} (\delta_{i,0} \delta_{j,0} + \delta_{N,i+j}) (P\sigma)_j. \quad (\text{A7})$$

Thus, considering the $N \times N$ matrix form with four quadrants for positive and negative frequency entries, $P\Sigma P^\dagger$ has a diagonal structure through the positive-positive and negative-negative quadrants, while $P\Sigma P^T$ has an anti-diagonal-like structure $i + j = N$ (and not $i + j = N - 1$, as it would be for the main anti-diagonal) extending through the positive-negative and negative-positive frequency quadrants, plus an extra component at $i = j = 0$.

2. The Toeplitz case

We now investigate the case where the TD covariance Σ has a Toeplitz but not circulant structure.

First, we note that, if we build the auto-covariance as an IFT of the noise PSD, we will automatically get a

circulant structure: if $\tilde{\sigma}$ is built according to (22) and if $\sigma = P^\dagger \tilde{\sigma} / (\Delta t \sqrt{N})$, then we have explicitly

$$\sigma_i = \frac{1}{2N\Delta t} \left(S_n^0 - S_n^{N/2} + \sum_{k=1}^{N/2-1} (\omega^{ik} + \omega^{-ik}) S_n^k \right), \quad (\text{A8})$$

so that $\sigma_i \in \mathbb{R}$ and $\sigma_i = \sigma_{N-i}$.

This means that a non-circulant autocorrelation requires another generation procedure; one such way is to generate a circulant covariance of size $2N \times 2N$ with the IFT, then extract its first quadrant to obtain an $N \times N$ Toeplitz matrix.

Coming back to the computation of $\langle \tilde{n} \tilde{n}^\dagger \rangle$, we have to remember that we used the circulant condition twice. We will get an extra contribution from $\sigma_{N-m} \neq \sigma_m$ in (A4), and we have to stop before the last inequality in (A5). As a result,

$$(P\Sigma P^\dagger)_{ij} = \delta_{ij} \sum_m \omega^{im} \sigma_m + \Delta_{ij}, \quad (\text{A9})$$

where

$$\Delta_{ij} = \frac{1}{N} \sum_k \sum_{m=1}^k \omega^{k(j-i)} \omega^{-mj} (\sigma_m - \sigma_{N-m}) = \frac{1}{N} \sum_{m=1}^{N-1} \left(\sum_{k=m}^{N-1} \omega^{k(j-i)} \right) \omega^{-mj} (\sigma_m - \sigma_{N-m}). \quad (\text{A10})$$

For $i = j$, the sum in parentheses is $N - m$, and we have

$$\Delta_{ii} = \frac{1}{N} \sum_{m=1}^{N-1} (N - m) \omega^{-mi} (\sigma_m - \sigma_{N-m}) = \sum_{m=1}^{N-1} \omega^{-mi} \sigma_m - \frac{1}{N} \sum_{m=1}^{N-1} m (\omega^{mi} + \omega^{-mi}) \sigma_m \quad (\text{A11})$$

For $i \neq j$,

$$\sum_{k=m}^{N-1} \omega^{k(j-i)} = -\frac{1 - \omega^{m(j-i)}}{1 - \omega^{j-i}} \text{ for } i \neq j. \quad (\text{A12})$$

So, for $i \neq j$ we have

$$\Delta_{ij} = -\frac{1}{N} \sum_{m=1}^{N-1} \frac{\omega^{-mj} - \omega^{-mi}}{1 - \omega^{j-i}} (\sigma_m - \sigma_{N-m}), \quad (\text{A13a})$$

$$= \frac{1}{1 - \omega^{j-i}} \frac{1}{N} \sum_{m=0}^{N-1} (\omega^{-mi} - \omega^{-mj} - \omega^{mi} + \omega^{mj}) \sigma_m, \quad (\text{A13b})$$

which we can rewrite using $\mathbf{P}\boldsymbol{\sigma}$. Gathering the diagonal and off-diagonal terms, we obtain

$$(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\dagger)_{ij} = \delta_{ij}\boldsymbol{\Lambda}_i + \boldsymbol{\Omega}_{ij}, \quad (\text{A14a})$$

$$\boldsymbol{\Lambda}_i = \boldsymbol{\sigma}_0 + \frac{1}{N} \sum_{m=1}^{N-1} (N-m) (\omega^{mi} + \omega^{-mi}) \boldsymbol{\sigma}_m, \quad (\text{A14b})$$

$$\boldsymbol{\Omega}_{ij} = \frac{1}{\sqrt{N}} \frac{1}{1 - \omega^{j-i}} [(\mathbf{P}\boldsymbol{\sigma})_i - (\mathbf{P}\boldsymbol{\sigma})_i^* - (\mathbf{P}\boldsymbol{\sigma})_j + (\mathbf{P}\boldsymbol{\sigma})_j^*] \text{ for } i \neq j. \quad (\text{A14c})$$

In this rewriting, we can check that $\boldsymbol{\Lambda}_i \in \mathbb{R}$ and $\boldsymbol{\Omega}$ is Hermitian, $\boldsymbol{\Omega}^\dagger = \boldsymbol{\Omega}$.

This calculation can be repeated, with few changes, for the computation of $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T$. This is required if one needs a complete control on both real and imaginary parts in the FD; however, $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\dagger$ is enough to express the likelihood. The result reads

$$(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T)_{00} = \boldsymbol{\sigma}_0 + \frac{2}{N} \sum_{m=1}^{N-1} (N-m) \boldsymbol{\sigma}_m, \quad (\text{A15a})$$

$$(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T)_{i,N-i} = \boldsymbol{\sigma}_0 + \frac{1}{N} \sum_{m=1}^{N-1} (N-m) (\omega^{mi} + \omega^{-mi}) \boldsymbol{\sigma}_m \text{ for } i = 1, \dots, N-1, \quad (\text{A15b})$$

$$(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T)_{ij} = \frac{1}{\sqrt{N}} \frac{1}{1 - \omega^{-i-j}} [(\mathbf{P}\boldsymbol{\sigma})_i - (\mathbf{P}\boldsymbol{\sigma})_i^* + (\mathbf{P}\boldsymbol{\sigma})_j - (\mathbf{P}\boldsymbol{\sigma})_j^*] \text{ for } i+j \neq 0, N. \quad (\text{A15c})$$

Appendix B: Fourier-domain noise generation in the circulant case

We give here a detailed description of noise generation in the discrete FD, in the circulant case. This will constitute a discrete equivalent of the better-known functional relations (17).

We wish to draw a FD realization of the noise, $\tilde{\mathbf{n}}$, given a PSD $S_n(f)$. We first recall that, since $\mathbf{n} \in \mathbb{R}^N$, $\tilde{\mathbf{n}}$ is entirely represented by its $N/2 + 1$ components $\tilde{\mathbf{n}}_j$ for $j = 0, \dots, N/2$ included, with other components being then fixed as $\tilde{\mathbf{n}}_{N-j} = \tilde{\mathbf{n}}_j^*$ for $j = 1, \dots, N/2 - 1$. Moreover, $\tilde{\mathbf{n}}_0, \tilde{\mathbf{n}}_{N/2} \in \mathbb{R}$, which brings us back to N degrees of freedom.

We will introduce random variables for the real and imaginary parts as $\tilde{\mathbf{n}}_j = A_j + iB_j$ for $j = 0, \dots, N/2$ included (we will come back later to B_0 and $B_{N/2}$). By assumption, A_j, B_j are Gaussian (as linear combinations of the Gaussian vector $\mathbf{n}$). Considering (23) and (A7) in the quadrant $\llbracket 0, N/2 \rrbracket^2$, we see that $\langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^\dagger \rangle$ has support on the diagonal, while $\langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^T \rangle$ only has support at $i = j = 0$ and $i = j = N/2$. Off-diagonal, we get

$$\langle (A_j + iB_j)(A_k \pm iB_k) \rangle = 0 \quad (j \neq k), \quad (\text{B1})$$

from which we deduce that A_j, B_j are independent from A_k, B_k when $j \neq k$. In turn, along the diagonal, for

$j = 1, \dots, N/2 - 1$, we get

$$\langle (A_j + iB_j)(A_j - iB_j) \rangle = \frac{S_n^j}{2\Delta f}, \quad (\text{B2a})$$

$$\langle (A_j + iB_j)(A_j + iB_j) \rangle = 0, \quad (j = 1, \dots, N/2 - 1) \quad (\text{B2b})$$

The real and imaginary part of the second relation give that $\langle A_j^2 \rangle = \langle B_j^2 \rangle$ and $\langle A_j B_j \rangle = 0$, so the Gaussian variables A_j, B_j are independent with the same variance. The first relation gives $\langle A_j^2 + B_j^2 \rangle = 2\langle A_j^2 \rangle = 2\langle B_j^2 \rangle = S_n^j/(2\Delta f)$, and we obtain that

$$\text{Re } \tilde{\mathbf{n}}_j, \text{Im } \tilde{\mathbf{n}}_j \sim \mathcal{N}\left(0, \frac{S_n(j\Delta f)}{4\Delta f}\right) \quad (j = 1, \dots, N/2 - 1), \quad (\text{B3})$$

and are independent both from each other and across frequency bins.

Coming back to the special cases $j = 0$ and $j = N/2$ with a non-zero component for $\langle \tilde{\mathbf{n}} \tilde{\mathbf{n}}^T \rangle$, we get

$$\langle (A_j + iB_j)(A_j - iB_j) \rangle = \frac{S_n^j}{2\Delta f}, \quad (\text{B4a})$$

$$\langle (A_j + iB_j)(A_j + iB_j) \rangle = \frac{S_n^j}{2\Delta f}, \quad (j = 0, N/2) \quad (\text{B4b})$$

The second equation gives $\langle A_j B_j \rangle = 0$ as before. We then get $\langle A_j^2 + B_j^2 \rangle = \langle A_j^2 - B_j^2 \rangle$, which shows that B_j

is identically zero, as it should be for $\tilde{\mathbf{n}}$ to be the DFT of a real vector. We arrive at

$$\text{Re } \tilde{\mathbf{n}}_j \sim \mathcal{N}\left(0, \frac{S_n(j\Delta f)}{2\Delta f}\right), \quad \text{Im } \tilde{\mathbf{n}}_j = 0 \quad (j = 0, N/2), \quad (\text{B5})$$

which completes the result. Thanks to independence, we can draw FD circulant noise with individual Gaussian draws.

Once noise is generated in the FD efficiently thanks to the independence property, it can be translated back to the TD via an IFFT. One efficient strategy to generate noise in the Toeplitz case is to artificially extend the Toeplitz covariance as a circulant covariance over an segment of length $2N$, draw an extended circulant noise realization in the FD, and perform an IFFT to obtain TD noise of length $2N$; the first half of this vector then gives an N -sized TD realization of the desired noise process with a Toeplitz covariance.

Appendix C: Likelihood with gaps using a permutation and masks

We can treat the general case of an arbitrary data plan, with multiple gaps of varying lengths, with the help of Boolean masks. We can define a mask as a boldface index $\mathbf{g}$ such that $\mathbf{x}_{\mathbf{g}}$ is the subvector of values inside gap segments, of length N_g . Similarly, for matrices $\Sigma_{-\mathbf{g}, -\mathbf{g}}$ is the $(N - N_g) \times (N - N_g)$ submatrix of values outside of the gap segments. The marginalization over elements of a Gaussian vector,

$$p(\mathbf{n}_{-\mathbf{g}}) = \int d\mathbf{n}_{\mathbf{g}} p(\mathbf{n}_{\mathbf{g}}, \mathbf{n}_{-\mathbf{g}}), \quad (\text{C1})$$

gives $p(\mathbf{n}_{-\mathbf{g}}) \sim \mathcal{N}(\mathbf{0}, \Sigma_{-\mathbf{g}, -\mathbf{g}})$. This gives the log-likelihood

$$\begin{aligned} \ln \mathcal{L}_{\text{gap}}(\mathbf{d}_{-\mathbf{g}} | \boldsymbol{\theta}) = & -\frac{1}{2} \mathbf{x}_{-\mathbf{g}}^T \Sigma_{-\mathbf{g}, -\mathbf{g}}^{-1} \mathbf{x}_{-\mathbf{g}} - \frac{1}{2} \det \Sigma_{-\mathbf{g}, -\mathbf{g}} \\ & - \frac{1}{2} (N - N_g) \ln 2\pi, \end{aligned} \quad (\text{C2})$$

We can reorder indices to put all indices pertaining to the gaps at the end, by introducing a permutation⁷ σ , such that

$$\begin{cases} \sigma(i) \leq N - N_g & \text{for } i \notin \text{gap segments} \\ \sigma(i) > N - N_g & \text{for } i \in \text{gap segments.} \end{cases} \quad (\text{C3})$$

The precise ordering within the gap and no-gap segments will not matter, for a simple connection with the

mask notation we can assume that we keep the relative ordering on each sub-segment. In our linear algebra notations, this permutation can be represented by a matrix $\mathbf{R}$ such that

$$\mathbf{R}_{ij} = \delta_{i\sigma(j)}. \quad (\text{C4})$$

A permutation matrix is automatically orthogonal (it preserves the scalar product), $\mathbf{R}^T \mathbf{R} = \mathbf{1}$. If $\mathbf{W} = \text{Diag}(w_g(i\Delta t), i = 0, \dots, N - 1)$ with w_g a gating window, we will define the *reordered gated covariance* as

$$\mathring{\Sigma} \equiv \mathbf{R} \mathbf{W} \Sigma \mathbf{W} \mathbf{R}^T, \quad (\text{C5})$$

and we can check explicitly, using (C4), that

$$\begin{aligned} \mathring{\Sigma}_{ij} &= \sum_k \sum_l \mathbf{R}_{ik} \mathbf{w}_k \Sigma_{kl} \mathbf{w}_l \mathbf{R}_{lj} \\ &= \sum_k \sum_l \delta_{i\sigma(k)} \mathbf{w}_k \Sigma_{kl} \mathbf{w}_l \delta_{l\sigma(j)} \\ &= \mathbf{w}_{\sigma^{-1}(i)} \Sigma_{\sigma^{-1}(i)\sigma^{-1}(j)} \mathbf{w}_{\sigma^{-1}(j)} \\ &= \begin{cases} \Sigma_{\sigma^{-1}(i)\sigma^{-1}(j)} & \text{if } i, j \leq N - N_g \\ 0 & \text{else } i \in \text{gap segments,} \end{cases} \end{aligned} \quad (\text{C6})$$

so that the reordered covariance has the block structure

$$\mathring{\Sigma} = \begin{pmatrix} \Sigma_{-\mathbf{g}, -\mathbf{g}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (\text{C7})$$

with pseudo-inverse

$$\mathring{\Sigma}^+ = \begin{pmatrix} (\Sigma_{-\mathbf{g}, -\mathbf{g}})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \quad (\text{C8})$$

Similarly, if we define reordered gated vectors

$$\mathring{\mathbf{x}} \equiv \mathbf{R} \mathbf{W} \mathbf{x} = \begin{pmatrix} \mathbf{x}_{-\mathbf{g}} \\ \mathbf{0} \end{pmatrix}, \quad (\text{C9})$$

the block structure shows directly that the marginalized likelihood can be written in terms of our reordered matrices and vectors as (ignoring constants)

$$\ln \mathcal{L}_{\text{gap}}(\mathbf{d}_{-\mathbf{g}} | \boldsymbol{\theta}) = -\frac{1}{2} \mathring{\mathbf{x}}^T \mathring{\Sigma}^+ \mathring{\mathbf{x}}. \quad (\text{C10})$$

Now, we can use the properties of the pseudo-inverse and the fact that $\mathbf{R}$ is orthogonal to rewrite

$$\begin{aligned} \mathring{\mathbf{x}}^T \mathring{\Sigma}^+ \mathring{\mathbf{x}} &= \mathbf{x}^T \mathbf{W} \mathbf{R}^T (\mathbf{R} \mathbf{W} \Sigma \mathbf{W} \mathbf{R}^T)^+ \mathbf{R} \mathbf{W} \mathbf{x} \\ &= \mathbf{x}^T \mathbf{W} \mathbf{R}^T \mathbf{R} (\mathbf{W} \Sigma \mathbf{W})^+ \mathbf{R}^T \mathbf{R} \mathbf{W} \mathbf{x} \\ &= (\mathbf{W} \mathbf{x})^T (\mathbf{W} \Sigma \mathbf{W})^+ \mathbf{W} \mathbf{x}. \end{aligned} \quad (\text{C11})$$

Thus, we arrive at our expression (39) for the likelihood in presence of gaps.

⁷ Not to be confused with the auto-covariance vector.

Note that we could also have used the re-ordered but not gated data vectors

$$\dot{\mathbf{x}} \equiv \mathbf{R}\mathbf{x} = \begin{pmatrix} \mathbf{x}_{-g} \\ \mathbf{x}_g \end{pmatrix}, \quad (\text{C12})$$

and the likelihood would retain the same expression thanks to the structure of $\dot{\Sigma}^+$:

$$\ln \mathcal{L}_{\text{gap}}(\mathbf{d}_{-g}|\boldsymbol{\theta}) = -\frac{1}{2} \dot{\mathbf{x}}^T \dot{\Sigma}^+ \dot{\mathbf{x}}. \quad (\text{C13})$$

Similarly to the previous case, we have

$$\dot{\mathbf{x}}^T \dot{\Sigma}^+ \dot{\mathbf{x}} = \mathbf{x}^T (\mathbf{W}\Sigma\mathbf{W})^+ \mathbf{x}, \quad (\text{C14})$$

in which we recognize the property that the likelihood is insensitive to the content of the residual vector inside the gap, so gating the residual is facultative.

Appendix D: Degenerate gated covariance matrices in time and frequency domain

Let $\mathbf{W}\Sigma\mathbf{W} = \Sigma^{\text{gap}} \in \mathbb{R}^{N \times N}$. From the rank-nullity theorem, we must have that $\text{rank}(\Sigma^{\text{gap}}) + \text{nullity}(\Sigma^{\text{gap}}) = N$. Making reference to App.C, Eq. (C7) there are an equivalent number of linearly dependent rows and columns in Σ^{gap} , implying that the dimension of the column space $\dim(\text{Col}(\Sigma^{\text{gap}})) = \text{rank}(\Sigma^{\text{gap}}) = N - N_g$, for N_g the number of zero columns of the matrix Σ^{gap} . We remark here that the swapping of rows and columns will not change the dimension of the column space. By the rank-nullity theorem, the dimension of the null space $\text{nullity}(\Sigma^{\text{gap}}) = N_g$, meaning that there are N_g vectors $\mathbf{v} \in \text{null}(\Sigma^{\text{gap}})$ such that $\Sigma^{\text{gap}}\mathbf{v} = \lambda\mathbf{v} = 0$. Hence there must exist N_g many zero eigenvalues in the matrix Σ^{gap} . We can translate

this into the FD via the unitary DFT matrix $\mathbf{P}$

$$\begin{aligned} \Sigma^{\text{gap}}\mathbf{v} &= 0 \\ \mathbf{P}[(\mathbf{W}\Sigma\mathbf{W})]\mathbf{v}\mathbf{P}^\dagger &= 0 \\ (\tilde{\mathbf{W}}\tilde{\Sigma}\tilde{\mathbf{W}})\tilde{\mathbf{u}} &= 0 \\ \tilde{\Sigma}^{\text{gap}}\tilde{\mathbf{u}} &= 0 \end{aligned}$$

implying that $\mathbf{P}\mathbf{v}\mathbf{P}^\dagger = \tilde{\mathbf{u}} \in \text{null}(\tilde{\Sigma}^{\text{gap}})$. Finally, noting that $\text{rank}(\Sigma^{\text{gap}}) = \text{rank}(\mathbf{P}^\dagger\tilde{\Sigma}^{\text{gap}}\mathbf{P}) = \text{rank}(\tilde{\Sigma}^{\text{gap}})$, which implies that $\text{nullity}(\Sigma^{\text{gap}}) = \text{nullity}(\tilde{\Sigma}^{\text{gap}})$, there must be N_g zero-eigenvalues in the FD gated covariance matrix $\tilde{\Sigma}^{\text{gap}}$. Finally, since $\tilde{\mathbf{u}} = \mathbf{P}\mathbf{v}\mathbf{P}^\dagger$, there exists a 1-1 correspondence between the zero-eigenvalues between the TD and FD covariance matrices. This means that the zero-eigenvalues present due to the gated segment in the TD matrix correspond to the zero-eigenvalues of the FD matrix. Finally, Since the matrix $\tilde{\Sigma}^{\text{gap}}$ is hermitian symmetric, the Singular Values (SVs) following a SVD are equivalent to the absolute values of the eigenvalues, i.e., $\mathbf{S}(\tilde{\Sigma}^{\text{gap}})_i = |\boldsymbol{\lambda}(\tilde{\Sigma}^{\text{gap}})_i|$. In conclusion, there are N_g many singular values of the FD gated noise covariance matrix.

Appendix E: Details on the tapering window

When smoothly tapering signal at the edges of a data gap or at the ends of the data segment, we will use a Planck window, as described in [122]. For a window $w(t)$ that is zero outside of $[t_i, t_f]$, smoothly transitions from 0 to 1 on $[t_i, t_i + \Delta t_i]$ and back from 1 to 0 on $[t_f - \Delta t_f, t_f]$, the expression is

$$w(t) = \begin{cases} 0 & \text{for } t \leq t_i \\ \frac{1}{1 + \exp\left[\frac{\Delta t_i}{t - t_i} + \frac{\Delta t_i}{t - (t_i + \Delta t_i)}\right]} & \text{for } t_i < t < t_i + \Delta t_i \\ 1 & \text{for } t_i + \Delta t_i \leq t \leq t_f - \Delta t_f \\ \frac{1}{1 + \exp\left[-\frac{\Delta t_f}{t - t_f} - \frac{\Delta t_f}{t - (t_f - \Delta t_f)}\right]} & \text{for } t_f - \Delta t_f < t < t_f \\ 0 & \text{for } t \geq t_f \end{cases} \quad (\text{E1})$$

For instance, multiplying the data with $1 - w$ would implement a gap on $[t_i + \Delta t_i, t_f - \Delta t_f]$ with tapering lobes $\Delta t_i, \Delta t_f$.

[1] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, *et al.*, arXiv preprint arXiv:1702.00786 (2017).

[2] M. Colpi, K. Danzmann, M. Hewitson, K. Holley-Bockelmann, P. Jetzer, G. Nelemans, A. Petiteau, D. Shoemaker, C. Sopena, R. Stebbins, *et al.*, arXiv preprint arXiv:2402.07571 (2024).

- [3] M. Colpi, K. Holley-Bockelmann, T. Bogdanovic, P. Natarajan, J. Bellovary, A. Sesana, M. Tremmel, J. Schnittman, J. Comerford, E. Barausse, E. Berti, M. Volonteri, F. Khan, S. McWilliams, S. Burke-Spolaor, J. Hazboun, J. Conklin, G. Mueller, and S. Larson, [Astro2020 science white paper: The gravitational wave view of massive black holes](#) (2019), [arXiv:1903.06867 \[astro-ph.GA\]](#).
- [4] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, *Physical Review D* **95**, 103012 (2017).
- [5] C. P. L. Berry, S. A. Hughes, C. F. Sopuerta, A. J. K. Chua, A. Heffernan, K. Holley-Bockelmann, D. P. Mihaylov, M. C. Miller, and A. Sesana, [The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy](#) (2019), [arXiv:1903.03686 \[astro-ph.HE\]](#).
- [6] T. B. Littenberg, K. Breivik, W. R. Brown, M. Eracleous, J. Hermes, K. Holley-Bockelmann, K. Kremer, T. Kupfer, and S. L. Larson, [arXiv preprint arXiv:1903.05583](#) (2019).
- [7] Z.-C. Chen, F. Huang, and Q.-G. Huang, *The Astrophysical Journal* **871**, 97 (2019).
- [8] M. Bonetti and A. Sesana, *Physical Review D* **102**, 103023 (2020).
- [9] F. Pozzoli, S. Babak, A. Sesana, M. Bonetti, and N. Karnesis, *Physical Review D* **108**, 103039 (2023).
- [10] A. J. Ruiter, K. Belczynski, M. Benacquista, S. L. Larson, and G. Williams, *The Astrophysical Journal* **717**, 1006 (2010).
- [11] G. Nelemans, *Classical and Quantum Gravity* **26**, 094030 (2009).
- [12] S. E. Timpano, L. J. Rubbo, and N. J. Cornish, *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **73**, 122001 (2006).
- [13] M. C. Digman and N. J. Cornish, *The Astrophysical Journal* **940**, 10 (2022).
- [14] B. S. M. L. Jeune., A. Petiteu, and M. Vallisneri, *LISA Data Challenge: Sangria*, Tech. Rep. LISA-LCST-SGS-MAN-001 (The LISA consortium, 2022) property of the LISA consortium.
- [15] K. A. Arnaud, G. Auger, S. Babak, J. G. Baker, M. J. Benacquista, E. Bloomer, D. A. Brown, J. Camp, J. Cannizzo, N. Christensen, *et al.*, *Classical and Quantum Gravity* **24**, S529 (2007).
- [16] K. A. Arnaud, S. Babak, J. G. Baker, M. J. Benacquista, N. J. Cornish, C. Cutler, L. Finn, S. Larson, T. Littenberg, E. Porter, *et al.*, *Classical and Quantum Gravity* **24**, S551 (2007).
- [17] T. B. Littenberg and N. J. Cornish, *Physical Review D* **107**, 063004 (2023).
- [18] S. H. Strub, L. Ferraioli, C. Schmelsbach, S. C. Stähler, and D. Giardini, *Physical Review D* **110**, 024005 (2024).
- [19] M. L. Katz, N. Karnesis, N. Korsakova, J. R. Gair, and N. Stergioulas, *Physical Review D* **111**, 024060 (2025).
- [20] S. Deng, S. Babak, M. L. Jeune, S. Marsat, É. Plagnol, and A. Sartirana, [arXiv preprint arXiv:2501.10277](#) (2025).
- [21] M. Armano, H. Audley, J. Baird, P. Binetruy, M. Born, D. Bortoluzzi, E. Castelli, A. Cavalleri, A. Cesarini, A. M. Cruise, *et al.*, *Monthly Notices of the Royal Astronomical Society* **494**, 3014 (2020).
- [22] Q. Baghi, N. Korsakova, J. Slutsky, E. Castelli, N. Karnesis, and J.-B. Bayle, *Physical Review D* **105**, [10.1103/physrevd.105.042002](#) (2022).
- [23] M. Armano, H. Audley, G. Auger, J. T. Baird, M. Bas-san, P. Binetruy, M. Born, D. Bortoluzzi, N. Brandt, M. Caleno, *et al.*, *Physical review letters* **116**, 231101 (2016).
- [24] E. Castelli, B. S. M. L. Jeune., N. Korsakova, Q. Baghi, *et al.*, *LISA Data Challenge: Spritz*, Tech. Rep. LISA-LCST-XXX-TN-001 (The LISA consortium, 2022) property of the LISA consortium.
- [25] E. Castelli, Q. Baghi, J. G. Baker, J. Slutsky, J. Bobin, N. Karnesis, A. Petiteau, O. Sauter, P. Wass, and W. J. Weber, [arXiv preprint arXiv:2411.13402](#) (2024).
- [26] E. Castelli, Q. Baghi, J. G. Baker, J. Slutsky, J. Bobin, N. Karnesis, A. Petiteau, O. Sauter, P. Wass, and W. J. Weber, [Extraction of gravitational wave signals from lisa data in the presence of artifacts](#) (2025), [arXiv:2411.13402 \[gr-qc\]](#).
- [27] P. Whittle, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **19**, 38 (1957).
- [28] H. Siegel, M. Isi, and W. M. Farr, [Analyzing black-hole ringdowns ii: data conditioning](#) (2024), [arXiv:2410.02704 \[gr-qc\]](#).
- [29] C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, Y.-F. Wang, A. B. Nielsen, and B. Krishnan, *Physical Review Letters* **131**, 221402 (2023).
- [30] C. Pitte, Q. Baghi, M. Besançon, and A. Petiteau, *Physical Review D* **110**, 104003 (2024).
- [31] M. Isi and W. M. Farr, [arXiv preprint arXiv:2107.05609](#) (2021).
- [32] G. Carullo, W. Del Pozzo, and J. Veitch, *Physical Review D* **99**, 123029 (2019).
- [33] F. J. Harris, *Proceedings of the IEEE* **66**, 51 (1978).
- [34] G. Cabourn Davies, I. Harry, M. J. Williams, D. Bandopadhyay, L. Barack, J.-B. Bayle, C. Hoy, A. Klein, H. Middleton, C. J. Moore, *et al.*, *Physical Review D* **111**, 043045 (2025).
- [35] M. Armano, H. Audley, J. Baird, P. Binetruy, M. Born, D. Bortoluzzi, E. Castelli, A. Cavalleri, A. Cesarini, A. Cruise, *et al.*, *Physical review letters* **120**, 061101 (2018).
- [36] M. Armano, H. Audley, J. Baird, P. Binetruy, M. Born, D. Bortoluzzi, E. Castelli, A. Cavalleri, A. Cesarini, V. Chiavegato, *et al.*, *Physical Review D* **106**, 062001 (2022).
- [37] A. Spadaro, R. Buscicchio, D. Vetrugno, A. Klein, D. Gerosa, S. Vitale, R. Dolesi, W. J. Weber, and M. Colpi, *Physical Review D* **108**, 123029 (2023).
- [38] M. Abu Thaher Chowdhury, [arXiv e-prints](#), [arXiv](#) (2024).
- [39] B. Zackay, T. Venumadhav, J. Roulet, L. Dai, and M. Zaldarriaga, *Physical Review D* **104**, [10.1103/physrevd.104.063034](#) (2021).
- [40] B. P. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X.

- Adhikari, V. B. Adya, *et al.*, Physical review letters **119**, 161101 (2017).
- [41] Q. Baghi, E. Castelli, A. Petiteau, M. Hewitson, *et al.*, *Data Analysis Robustness Report*, Tech. Rep. LISA-LCST-SGS-RP-004 (The LISA consortium, 2023) property of the LISA consortium.
- [42] Q. Baghi, J. I. Thorpe, J. Slutsky, J. Baker, T. D. Canton, N. Korsakova, and N. Karnesis, Physical Review D **100**, 10.1103/physrevd.100.022003 (2019).
- [43] A. Blelly, J. Bobin, and H. Moutarde, Sparse data inpainting for the recovery of galactic-binary gravitational wave signals from gapped data (2021), [arXiv:2104.05250 \[gr-qc\]](#).
- [44] L. Wang, H.-Y. Chen, X. Lyu, E.-K. Li, and Y.-M. Hu, The European Physical Journal C **85**, 120 (2025).
- [45] R. Mao, J. E. Lee, and M. C. Edwards, Physical Review D **111**, 024067 (2025).
- [46] J. Carré and E. K. Porter, arXiv preprint [arXiv:1010.1641](#) (2010).
- [47] C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith, Physical Review Research **3**, 043049 (2021).
- [48] O. Edy, A. Lundgren, and L. K. Nuttall, arXiv preprint [arXiv:2101.07743](#) (2021).
- [49] K. Dey, N. Karnesis, A. Toubiana, E. Barausse, N. Korsakova, Q. Baghi, and S. Basak, Physical Review D **104**, 044035 (2021).
- [50] C. Cutler and E. E. Flanagan, Physical Review D **49**, 2658 (1994).
- [51] N. Wiener *et al.*, Acta mathematica **55**, 117 (1930).
- [52] A. Khintchine, Mathematische Annalen **109**, 604 (1934).
- [53] R. M. Gray *et al.*, Foundations and Trends® in Communications and Information Theory **2**, 155 (2006).
- [54] T.-T. Lu and S.-H. Shiou, Computers & Mathematics with Applications **43**, 119 (2002).
- [55] E. H. Moore, Bulletin of the american mathematical society **26**, 294 (1920).
- [56] R. Penrose, in *Mathematical proceedings of the Cambridge philosophical society*, Vol. 51 (Cambridge University Press, 1955) pp. 406–413.
- [57] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing* (Cambridge university press, 2007).
- [58] L. Speri, M. L. Katz, A. J. Chua, S. A. Hughes, N. Warburton, J. E. Thompson, C. E. Chapman-Bird, and J. R. Gair, Frontiers in Applied Mathematics and Statistics **9**, 1266739 (2024).
- [59] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044007 (2016).
- [60] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044006 (2016).
- [61] R. Cotesta, S. Marsat, and M. Pürrer, Physical Review D **101**, 10.1103/physrevd.101.124040 (2020).
- [62] R. A. Horn and C. R. Johnson, *Matrix analysis* (Cambridge university press, 2012).
- [63] L. S. Finn, Phys. Rev. D **46**, 5236 (1992), [arXiv:gr-qc/9209010](#).
- [64] A. I. Burke and O. Burke, *Extreme precision and extreme complexity: source modelling and data analysis development for the laser interferometer space antenna*, Ph.D. thesis, Edinburgh U. (2021).
- [65] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4566 (1998), [arXiv:gr-qc/9710129](#).
- [66] M. Miller, Phys. Rev. D **71**, 104016 (2005).
- [67] C. Cutler and M. Vallisneri, Phys. Rev. D **76**, 104018 (2007), [arXiv:0707.2982 \[gr-qc\]](#).
- [68] E. K. Porter, arXiv preprint [arXiv:0910.0373](#) (2009).
- [69] M. Vallisneri, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA **91109** (2007).
- [70] C. R. Rao, in *Breakthroughs in Statistics: Foundations and basic theory* (Springer, 1992) pp. 235–247.
- [71] H. Cramér, *Mathematical methods of statistics*, Vol. 26 (Princeton university press, 1999).
- [72] A. Antonelli, O. Burke, and J. R. Gair, Monthly Notices of the Royal Astronomical Society **507**, 5069 (2021).
- [73] L. London, S. Khan, E. Fauchon-Jones, C. García, M. Hannam, S. Husa, X. Jiménez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, Phys. Rev. Lett. **120**, 161102 (2018), [arXiv:1708.00404 \[gr-qc\]](#).
- [74] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Physical Review D **102**, 064002 (2020).
- [75] S. Marsat, J. G. Baker, and T. Dal Canton, Physical Review D **103**, 083011 (2021).
- [76] W. D. Voudsen, W. M. Farr, and I. Mandel, Mon. Not. Roy. Astron. Soc. **455**, 1919 (2016), [arXiv:1501.05823](#).
- [77] J. Goodman and J. Weare, Commun. Appl. Math. Comput. Sc. **5**, 65 (2010).
- [78] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013), [arXiv:1202.3665 \[astro-ph.IM\]](#).
- [79] F. B. Estabrook and H. D. Wahlquist, Gen. Rel. Grav. **6**, 439 (1975).
- [80] J. Armstrong, F. Estabrook, and M. Tinto, The Astrophysical Journal **527**, 814 (1999).
- [81] M. Tinto and J. W. Armstrong, Phys. Rev. D **59**, 102003 (1999).
- [82] S. V. Dhurandhar, K. Rajesh Nayak, and J. Y. Vinet, Phys. Rev. D **65**, 102002 (2002), [arXiv:gr-qc/0112059](#).
- [83] N. J. Cornish and R. W. Hellings, Class. Quant. Grav. **20**, 4851 (2003), [arXiv:gr-qc/0306096](#).
- [84] D. A. Shaddock, M. Tinto, F. B. Estabrook, and J. W. Armstrong, Phys. Rev. D **68**, 061303 (2003), [arXiv:gr-qc/0307080](#).
- [85] M. Tinto, F. B. Estabrook, and J. W. Armstrong, Phys. Rev. D **69**, 082001 (2004), [arXiv:gr-qc/0310017](#).
- [86] M. Vallisneri, Physical Review D—Particles, Fields, Gravitation, and Cosmology **71**, 022001 (2005).
- [87] M. Tinto and S. V. Dhurandhar, Living Reviews in Relativity **17**, 1 (2014).
- [88] J.-B. Bayle, M. Lilley, A. Petiteau, and H. Halloin, Physical Review D **99**, 084023 (2019).
- [89] M. Tinto, S. Dhurandhar, and D. Malakar, Physical Review D **107**, 082001 (2023).
- [90] M. Tinto and S. Dhurandhar, Physical Review D **108**, 082003 (2023).

- [91] N. Houba, J.-B. Bayle, and M. Vallisneri, arXiv preprint arXiv:2412.20793 (2024).
- [92] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, *Phys. Rev. D* **66**, 122002 (2002), [arXiv:gr-qc/0209039](#).
- [93] LISA Science Study Team, *LISA Science Requirements Document, ESA-L3-EST-SCI-RS-001. Technical Report 1.0*, Tech. Rep. (ESA, 2018).
- [94] N. Karnesis, S. Babak, M. Pieroni, N. Cornish, and T. Littenberg, *Physical Review D* **104**, 043019 (2021).
- [95] S. Babak, A. Petiteau, and M. Hewitson, arXiv preprint arXiv:2108.01167 (2021), [arXiv:2108.01167 \[astro-ph.IM\]](#).
- [96] T. Robson, N. J. Cornish, and C. Liu, *Classical and Quantum Gravity* **36**, 105011 (2019).
- [97] Q. Baghi, J. I. Thorpe, J. Slutsky, J. Baker, T. D. Canton, N. Korsakova, and N. Karnesis, *Physical Review D* **100**, 022003 (2019).
- [98] J. R. Gair, L. Barack, T. Creighton, C. Cutler, S. L. Larson, E. S. Phinney, and M. Vallisneri, *Classical and Quantum Gravity* **21**, S1595–S1606 (2004).
- [99] P. Amaro-Seoane, J. R. Gair, M. Freitag, M. C. Miller, I. Mandel, C. J. Cutler, and S. Babak, *Classical and Quantum Gravity* **24**, R113 (2007).
- [100] C. Messenger, *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **84**, 083003 (2011).
- [101] D. Bandopadhyay and C. J. Moore, *Physical Review D* **108**, 084014 (2023).
- [102] D. Bandopadhyay and C. J. Moore, *Physical Review D* **110**, 103026 (2024).
- [103] S. Biscoveanu, C.-J. Haster, S. Vitale, and J. Davies, *Physical Review D* **102**, 023008 (2020).
- [104] C. Talbot and E. Thrane, *Physical Review Research* **2**, 043298 (2020).
- [105] M. C. Edwards, R. Meyer, and N. Christensen, *Statistics and Computing* **29**, 67 (2019).
- [106] Q. Baghi, G. Métris, J. Bergé, B. Christophe, P. Touboul, and M. Rodrigues, *Physical Review D* **93**, 122007 (2016).
- [107] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015), [arXiv:1409.7215 \[gr-qc\]](#).
- [108] S. Kastha, C. D. Capano, J. Westerweck, M. Cabero, B. Krishnan, and A. B. Nielsen, *Physical Review D* **105**, 064042 (2022).
- [109] V. Gennari, G. Carullo, and W. Del Pozzo, *The European Physical Journal C* **84**, 233 (2024).
- [110] N. Karnesis, A. Sasli, R. Buscicchio, and N. Stergioulas, *Physical Review D* **111**, 022005 (2025).
- [111] A. Sasli, N. Karnesis, and N. Stergioulas, *Physical Review D* **108**, 103005 (2023).
- [112] C. Röver, *Physical Review D—Particles, Fields, Gravitation, and Cosmology* **84**, 122004 (2011).
- [113] N. Houba, L. Ferraioli, and D. Giardini, *Physical Review D* **109**, [10.1103/physrevd.109.083027](#) (2024).
- [114] O. Sauter, P. Wass, W. Sanchez, and H. Inchauspé, [Maximum likelihood detection of instrumental glitches in lisa tdi data](#) (2025), [arXiv:2501.06315 \[gr-qc\]](#).
- [115] R. Tenorio and D. Gerosa, arXiv preprint arXiv:2502.11823 (2025), [arXiv:2502.11823 \[gr-qc\]](#).
- [116] J. R. Gair, I. Mandel, and L. Wen, *Classical and Quantum Gravity* **25**, 184031 (2008).
- [117] H. Kawahara, T. Kuroda, T. Takiwaki, K. Hayama, and K. Kotake, *The Astrophysical Journal* **867**, 126 (2018).
- [118] V. Nuccia, S. Klimenko, and G. Mitselmakher, in *Journal of Physics: Conference Series*, Vol. 363 (IOP Publishing, 2012) p. 012032.
- [119] M. C. Digman and N. J. Cornish, *Physical Review D* **108**, 023022 (2023).
- [120] N. J. Cornish, *Physical Review D* **102**, 124038 (2020).
- [121] PASTA Collaboration, [arXiv e-prints](#), [arXiv:2403.20314](#) (2024), [arXiv:2403.20314 \[astro-ph.IM\]](#).
- [122] D. J. A. McKechn, C. Robinson, and B. S. Sathyaprakash, *Class. Quant. Grav.* **27**, 084020 (2010), [arXiv:1003.2939 \[gr-qc\]](#).