Gravitational-wave geodesy: Defining false alarm probabilities with respect to correlated noise

Kamiel Janssens,[1,2] Thomas A. Callister,[3] Nelson Christensen,[2] Michael W. Coughlin,[4] Ioannis Michaloliakos,[5] Jishnu Suresh,[6,7] and Nick van Remortel[1]

[1]*Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium*
[2]*Artemis, Université Côte d'Azur, Observatoire Côte d'Azur, Nice, France*
[3]*Center for Computational Astrophysics, Flatiron Institute, New York, New York 10010, USA*
[4]*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[5]*University of Florida, Gainesville, Florida 32611, USA*
[6]*Institute for Cosmic Ray Research, The University of Tokyo, Kashiwa City, Chiba 277-8582, Japan*
[7]*Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

(Received 7 December 2021; accepted 2 March 2022; published 5 April 2022)

Future searches for a gravitational-wave background using Earth-based gravitational-wave detectors might be impacted by correlated noise sources. A well-known example are the Schumann resonances, which are extensively studied in the context of searches for a gravitational-wave background. Earlier work has shown that a technique termed "gravitational-wave geodesy" can be used to generically differentiate observations of a gravitational-wave background from signals due to correlated terrestrial effects, requiring true observations to be consistent with the known geometry of our detector network. The key result of this test is a Bayes factor between the hypotheses that a candidate signal is astrophysical or terrestrial in origin. Here, we further formalize the geodesy test, mapping distributions of false-alarm and false-acceptance probabilities to quantify the degree with which a given Bayes factor will boost or diminish our confidence in an apparent detection of the gravitational-wave background. To define the false alarm probability of a given Bayes factor, we must have knowledge of our null hypothesis: the space of all possible correlated terrestrial signals. Since we do not have this knowledge, we instead construct a generic space of smooth functions in the frequency domain using Gaussian processes, which we tailor to be conservative. This enables us to use draws from our Gaussian processes as a proxy for all possible nonastrophysical signals. During the O2 observing run, the LIGO and Virgo collaborations observed an SNR $= 1.25$ excess in their search for an isotropic gravitational-wave background. To demonstrate the utility of gravitational-wave geodesy, we apply the method to the observed cross-correlated data.

DOI: 10.1103/PhysRevD.105.082001

I. INTRODUCTION

Since the first detection of gravitational waves (GWs) in 2015 [1], the Advanced LIGO [2], Advanced Virgo [3], and KAGRA [4] (LVK) collaborations have announced many more binary mergers [5–8]. These include binary black hole mergers, binary neutron star mergers, as well as neutron star–black hole mergers. In total, 90 observations were reported by the LVK collaborations in GWTC-1 [5], GWTC-2 [6], and GWTC-3 [8], and some lower-significance events in GWTC-2.1 [9].

Gravitational waves from the mergers of most binary systems at cosmological distances are too weak to be individually detected. However, the superposition of these signals forms a gravitational-wave background (GWB) [10–15]. The LVK collaborations have conducted searches for both an isotropic [16] and anisotropic [17] gravitational-wave background, but no such signal has yet been detected by the advanced detectors. Predictions based on the merger rate and mass distribution of compact binaries indicate that the GWB from the superposition of these events might be detected by LIGO, Virgo, and KAGRA during forthcoming observing runs [16,18]. Apart from the GWB from unresolved binary mergers, many other astrophysical and cosmological signals could contribute to a GWB [19].

As the GWB is too weak to be observed in a single detector, we rely instead on the cross correlation of strain data from multiple detectors [19,20]. If we assume that the noise present at different detector sites is uncorrelated, any excess correlation between the strain measured in two detectors must be due to an astrophysical gravitational-wave signal. However, it is not the case that gravitational-wave detectors measure strictly independent noise realizations. While many sources of terrestrial noise are indeed local, there are known sources that are correlated on global scales, introducing nonastrophysical correlations in the LIGO-Hanford, LIGO-Livingston, Virgo, and KAGRA

interferometers, despite their separation by several thousand kilometers. Some known examples of such noise sources are Schumann resonances [21,22] and the synchronization of on-site electronics to Global Positioning System (GPS) time [23]. Schumann resonances are electromagnetic excitations in the cavity formed by the Earth's surface and the ionosphere, sourced by lightning strikes across the globe [21,22]. They are expected to couple magnetically to the interferometers and induce a correlated signal of terrestrial origin [24,25].

The existence of known sources of correlated noise raises an important question: if we detect evidence for a correlated signal between two (or more) GW interferometers, how can we be confident that this source is of astrophysical origin rather than terrestrial? Until now, several methods have been or are being developed to help differentiate whether a correlated signal is due to gravitational waves or terrestrial sources. For Schumann resonances, in particular, methods have been investigated to directly measure and remove their effect by applying Wiener filtering [24–27] and more recently to incorporate Schumann resonances and a GW signal in one consistent Bayesian parameter estimation framework [28].

A complementary method, gravitational-wave geodesy (GW-geodesy), was previously proposed [29]. In GW-geodesy, the geometry of an interferometer network (the relative distances and orientations of component detectors) is reverse engineered from an observed GWB. This forms the basis for a consistency check that a true astrophysical signal must pass. A true GW signal must yield results consistent with the *known* geometry of our baseline, while the same is not the case for other sources. In the first implementation of the GW-geodesy framework, it was shown that the method can successfully differentiate an isotropic GWB coming from unresolved binary mergers from correlated terrestrial noise due to Schumann resonances or the synchronization of electronics to GPS time [29].

The output of the GW-geodesy test is a Bayes factor between two hypotheses: (*i*) a tentative detection yields consistency with our known baseline geometry and is therefore astrophysical, or (*ii*) the signal prefers an unphysical geometry and is hence nonastrophysical in origin. This Bayes factor acts as a secondary test statistic, independent of and complementary to the signal-to-noise ratio (SNR) with which we observe a given signal. Namely, this Bayes factor [which will be defined in Eq. (7) of Sec. II] makes a statement on how well the data fits the model, that is, the true geometry combined with a GW signal, whereas the SNR [which will be defined in Eq. (12) of Sec. IV] is a measure of the data's deviation from uncorrelated Gaussian noise. One might compare the Bayes factor we construct here with the χ^2 statistic in searches for compact binary coalescences [30] (which measures how well a template fits a signal) or a Bayesian coherence ratio [31] (which quantifies the self-consistency of a signal observed with multiple detectors). Like our geodesy Bayes factor, these statistics capture additional information beyond the overall amplitude of a signal, and are critical in determining the astrophysical significance of an apparent detection.

While these Bayes factors can be qualitatively interpreted, until now we have not been able to assign a precise *statistical significance* to a given Bayes factor. Ideally, we could assign any given Bayes factor a false alarm probability (FAP) and a false dismissal probability (FDP). The FAP indicates how often one might accidentally confirm a terrestrial signal based on this test, whereas the FDP indicates how often we accidentally dismiss a real signal. In this work, we quantity these FAPs and FDPs, allowing us to answer the crucial question: *how likely is an observed correlated signal with a given SNR and geodesy Bayes factor to be of astrophysical origin, rather than a yet-unidentified source of terrestrial correlation?* In this fashion, we can utilize GW-geodesy not only as a tool with which to reject terrestrial signals, but also as one to bolster our confidence in a real gravitational-wave background detection.

To be able to construct FAPs and FDPs, we first need a proxy for unknown correlated terrestrial signals, which have the possibility of contaminating the isotropic stochastic search. This is, by definition, challenging: we cannot know the nature of unknown contaminants. We therefore instead utilize Gaussian processes to represent random and *a priori* unknown contaminants, defining FAPs and FDPs over the space of continuous cross-correlation functions that LVK detectors might measure.

In Sec. II we cover some mathematical concepts of stochastic searches and the GW-geodesy framework that are crucial elements in this work. In Sec. III we introduce Gaussian processes, and we optimize the model parameters to create a very conservative scenario. In Sec. IV we demonstrate how the Gaussian processes can be used within the framework of GW-geodesy and why it can become a powerful tool to help in validating future observations of an isotropic stochastic background. As part of this demonstration, we apply our framework in Sec. IV C to the SNR $= 1.25$ excess observed for a $2/3$ power law by the LIGO and Virgo collaborations during their second observing run (O2). In Sec. V we discuss the tool together with an outlook on possible future additions or improvements.

II. GRAVITATIONAL-WAVE GEODESY

We often characterize the GWB in terms of its energy-density spectrum $\Omega(f)$ [Eq. (1)], expressed as the energy density $d\rho_{\mathrm{GW}}$ of GWs per logarithmic frequency interval $d\ln(f)$. $\Omega(f)$ is made dimensionless by dividing by the Universe's critical energy density $\rho_c = 3H_0^2c^2/(8\pi G)$, where H_0 is the Hubble constant, c is the speed of light, and G is Newton's constant [20,32]:

$$\Omega(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm GW}}{d\ln(f)}. \quad (1)$$

To measure the energy density of the GWB, one computes the cross-correlation spectrum $\hat{C}(f)$ between two gravitational-wave observatories. If $\tilde{s}_I(f)$ is the measured (Fourier-domain) strain of observatory I and ΔT is the duration of the analyzed data, one can express $\hat{C}(f)$ as

$$\hat{C}(f) = \frac{1}{\Delta T}\frac{20\pi^2}{3H_0^2} f^3 \mathrm{Re}[\tilde{s}_1^*(f)\tilde{s}_2(f)]. \quad (2)$$

The normalization of $\hat{C}(f)$ is chosen such that its expectation value is given by [32]

$$\langle \hat{C}(f)\rangle = \gamma(f)\Omega(f). \quad (3)$$

$\gamma(f)$ is the normalized overlap reduction function, which encodes the imprint of the detectors' baseline geometry (location and relative orientations) on the observed correlations. For two laser interferometers, the normalized overlap reduction is given by [33]

$$\gamma(f) = \frac{5}{8\pi}\sum_A \int_{\rm Sky} F_1^A(\hat{\mathbf{n}})F_2^A(\hat{\mathbf{n}})e^{2\pi i f \Delta\mathbf{x}\cdot\hat{\mathbf{n}}/c} d\hat{\mathbf{n}}. \quad (4)$$

$F_I^A(\hat{\mathbf{n}})$ is the antenna response of detector I to GWs with polarization A. Δx represents the separation vector between the two detectors, whereas $\hat{\mathbf{n}}$ indicates the sky direction. One sums over all tensor polarizations ("plus" and "cross") and integrates over all sky directions. By virtue of the leading factor $\frac{5}{8\pi}$, colocated and coaligned detectors will have $\gamma(f) = 1$ for all frequencies.

If we assume that we are in the weak-signal limit (which is a valid assumption since the GWB has yet to be observed), the covariance of $\hat{C}(f)$ at two different frequencies f and f' is given by $\langle \hat{C}(f)\hat{C}(f')\rangle = \delta(f-f')\sigma^2(f)$, where $\sigma^2(f)$ is given by [20,32]

$$\sigma^2(f) = \frac{1}{\Delta T}\left(\frac{10\pi^2}{3H_0^2}\right)^2 f^6 P_1(f)P_2(f), \quad (5)$$

and $P_I(f)$ is the noise power spectral density of detector I.

Traditionally, searches for a GWB assume a power law of the form

$$\Omega(f) = \Omega_{\rm ref}\left(\frac{f}{f_{\rm ref}}\right)^{\alpha}. \quad (6)$$

A power-law index $\alpha = 0$ is expected from several cosmological sources of a GWB, while $\alpha = 2/3$ is expected from individually unresolved binary coalescence events. A GWB with $\alpha = 3$ could be produced by supernovae [19,34].

Given the detection of a gravitational-wave background, we could seek to infer $\Omega_{\rm ref}$ and α. Additionally, however, the dependence of our measured cross-correlation spectrum $\hat{C}(f)$ on the overlap reduction function means that the GWB could be used to infer the geometry of our detector network itself. This fact forms the basis of GW-geodesy: a true GWB should yield an inferred geometry consistent with the true geometry of our detector network. An isotropic astrophysical/cosmological GWB must be consistent with the expected functional form of our baseline's overlap reduction function. Correlated terrestrial noise sources, on the other hand, do not necessarily need to follow the behavior of the overlap reduction function. Thus, there is no reason why non-GW correlated noise sources would prefer the true geometry over any random geometry [29].

We formalize this test by defining the following two hypotheses:

(1) Hypothesis $\mathcal{H}_\gamma$: The measured cross-correlation is consistent with the *true* baseline geometry and overlap reduction function.
(2) Hypothesis $\mathcal{H}_{\rm Free}$: The measured cross-correlation spectrum is consistent with an *unphysical* baseline geometry. Under this hypothesis, we treat our detector positions and orientations as free variables to be inferred from the data, allowing them to range (unphysically) across the surface of the Earth.

To compare the hypotheses $\mathcal{H}_\gamma$ and $\mathcal{H}_{\rm Free}$, one can construct a Bayes factor $\mathcal{B}$ between these hypotheses to establish which model is favored by the cross-correlated data $\hat{C}$,

$$\mathcal{B} = \frac{p(\hat{C}|\mathcal{H}_\gamma)}{p(\hat{C}|\mathcal{H}_{\rm Free})}, \quad (7)$$

where $p(\hat{C}|\mathcal{H}_\gamma)$ and $p(\hat{C}|\mathcal{H}_{\rm Free})$ are the probabilities of finding the observed cross correlation [as defined in Eq. (2)] given hypothesis $\mathcal{H}_\gamma$ and $\mathcal{H}_{\rm Free}$, respectively. Because $\mathcal{H}_{\rm Free}$ is a more complex model, it will be penalized by the Bayesian "Occam's factor." Therefore, an isotropic astrophysical signal will be consistent with both $\mathcal{H}_\gamma$ and $\mathcal{H}_{\rm Free}$, but will favor $\mathcal{H}_\gamma$ since it is the simpler hypothesis. Non-GW correlated noise sources have *a priori* no preference for $\mathcal{H}_\gamma$ and therefore will be better fit by the additional degrees of freedom provided by nonphysical geometries, leading to favoring the $\mathcal{H}_{\rm Free}$ hypothesis.

III. GAUSSIAN PROCESSES

If, in the future, we compute a geodesy Bayes factor $\mathcal{B}$ associated with a candidate gravitational-wave background, we do not yet know exactly how we should quantitatively interpret this result. In order to fully understand the statistical significance of a particular Bayes factor, we need to know how often it arises simply by chance.

This poses a dilemma, however. To quantify a false-alarm probability, we would need to know all possible terrestrial signals, which we do not. So instead, we can ask a similar question: how often do particular Bayes factors arise from the generic space of smooth functions in the frequency domain? This latter question can be answered using Gaussian processes.

Gaussian processes are very flexible and are frequently used for model fitting or model predictions based on a data set [35]. In this work, we use the Gaussian processes to produce a distribution of functions with a given mean and variance. We consider each draw from this distribution to be a possible realization of a correlated terrestrial signal.

The two main inputs of the Gaussian process are a covariance matrix, often called the *kernel*, and the mean. In this work, we use the Gaussian processes to generate possible realizations of $\langle \hat{C}(f) \rangle$. The probability density function for the draws is given in Eq. (8). This is the probability to draw a certain cross-correlation spectrum $\mathbf{C}(\mathbf{f})$, from the function space governed by the covariance matrix Σ [35],

$$p(\mathbf{C}(\mathbf{f})|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \cdot \times \exp\left(-\frac{1}{2}(\mathbf{C}(\mathbf{f})-\mu)^T\Sigma^{-1}(\mathbf{C}(\mathbf{f})-\mu)\right). \quad (8)$$

In Eq. (8), n is the dimension of the covariance matrix Σ.

We generate distributions of cross-correlation spectra with zero mean and consider one of the most commonly used covariance matrices: the so-called *squared exponential* (SE). In this case, the covariance between cross-correlation values measured at frequencies f_i and f_j is given by [35]

$$\Sigma_{\mathrm{SE},ij}(\sigma, l) = \sigma^2 \exp\left(-\frac{(f_i - f_j)^2}{2l^2}\right). \quad (9)$$

σ^2 is the variance at a single frequency, and while this parameter can be used to scale the signal strength with respect to observations, it will have no impact on the spectral *shape* of the draws from the distribution. l is the characteristic length scale over which our $\mathbf{C}(\mathbf{f})$ measurements are correlated.

In order to utilize Eqs. (8) and (9), we still need to choose a length-scale parameter l. We will tune this parameter by deliberately targeting the most conservative scenario, in which the Gaussian process yields cross-correlation spectra that look (on average) most like a proper astrophysical/cosmological signal. In doing so, we will always obtain conservative estimates of the false-alarm probability for a given geodesy Bayes factor.

To select the most conservative value of the length-scale parameter, we will maximize the probability for our *target astrophysical signal* itself to be drawn from the Gaussian process. The cross-correlation spectrum we expect from a gravitational-wave background is $\mathbf{C}(\mathbf{f}) = \gamma_{\mathrm{HL}} \cdot \Omega_{\mathrm{ref}} (\frac{f}{f_{\mathrm{ref}}})^{\alpha} := \mathbf{C}_{\alpha}(\mathrm{f})$, with $\alpha = 0, 2/3$, or 3, where we approximate the mean to be zero. These are the power laws typically looked for by the LIGO, Virgo, and KAGRA collaborations [16]. Note that we have chosen the Hanford-Livingston (HL) baseline as our observing baseline. This baseline is the most sensitive for observing an isotropic GWB, due to their better sensitivity compared to Virgo, but more importantly because their overlap reduction function is considerably larger. The log-likelihood of this signal under our Gaussian process is

$$\ln(\mathcal{L}) = \ln p(\mathbf{C}_{\alpha}(\mathrm{f})|\mu = \mathbf{0}, \mathbf{\Sigma}) = -\frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}\mathbf{C}_{\alpha}(\mathrm{f})^{\mathbf{T}}\mathbf{\Sigma^{-1}}\mathbf{C}_{\alpha}(\mathrm{f}) + \text{constants}. \quad (10)$$

As we will later freely adjust the overall amplitude of our Gaussian process draws in order to vary their SNRs, we will optimize only the length parameter l, fixing the overall covariance to $\sigma^2 = 1$. Accordingly, for consistency we normalize our target astrophysical signal via

$$\tilde{\mathbf{C}}_{\alpha}(f) = \frac{\mathbf{C}_{\alpha}(\mathbf{f})}{\sqrt{\mathrm{Var}\mathbf{C}_{\alpha}(\mathbf{f})}}, \quad (11)$$

choosing the l that maximizes $\ln p(\tilde{\mathbf{C}}_{\alpha}(f)|\mu = 0, \Sigma)$.

In practice, the inversion of Σ is unstable, with a determinant that is nearly zero. To increase stability, we add a diagonal term to the covariance matrix: $\Sigma \to \Sigma + \epsilon I$, where I is the identity matrix and ϵ is a small dimensionless constant, here chosen to be 10^{-3}. This diagonal term can be interpreted as a noisy observation of our function $\mathbf{C}_{\alpha}(\mathrm{f})$.

We optimize the kernel parameter l using the formalism described above in the following parameter range: $l \in [10^{-4}\ \mathrm{Hz}, 10^{4}\ \mathrm{Hz}]$. The optimal length-scale parameters to mimic a power law with slope $\alpha = 0, 2/3$, and 3 are, respectively, $l = 47.65$, 33.18, and 29.20 Hz, determined using a 0.01 Hz frequency resolution for l.

In Fig. 1 we show several random draws from the Gaussian process described by the SE kernel, optimized to most closely match an $\alpha = 2/3$ astrophysical signal, as well as the signals we expect to observe from unresolved binary mergers and terrestrial Schumann resonances. Here we have chosen $\alpha = 2/3$ since this gravitational-wave background is the signal expected to be first observable by Advanced LIGO and Advanced Virgo.

The Schumann resonance spectrum used here is identical to the spectrum used in the first description of GW-geodesy [29]. This spectrum is obtained by replacing the strain data s_1 and s_2 in Eq. (2) with data from magnetometers measuring the Schumann resonances in a magnetically quiet location, the Hylaty station in Poland, as reported in Ref. [27]. We furthermore assume that the transfer function

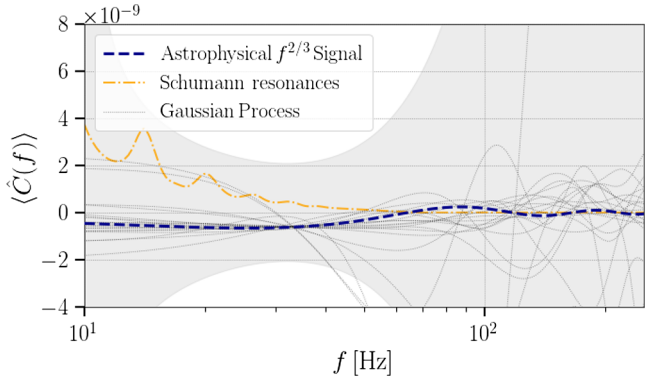

FIG. 1. Expected cross-correlation for an astrophysical background from unresolved binary systems (blue/bold), Schumann resonances (orange), and 20 random draws from the SE kernel (gray). The kernel parameter used for the Gaussian process instantiations is the optimal value for mimicking a 2/3-power-law signal, as described in the text. The injected signal strength yields a SNR = 3 after a 3-year long observation at LIGO's design sensitivity.

linking environmental magnetic fields to gravitational-wave interferometers is a power law that declines as $\propto f^{-2}$ in frequency [29]. An important assumption about our correlated magnetic spectrum is that it is positive since the frequency-dependent sign of the cross power between the two LIGO sites is unknown. More detailed studies in the future could investigate the effect when relaxing this assumption. Two models have been developed modeling a Schumann resonance: a simple analytical model [36], and a more complex model taking the anisotropic character of lightning strikes sourcing the Schumann resonances into account [37]. These analytical models give a mathematical framework to investigate the departure of the assumption of a fully positive magnetic cross spectrum as a function of the magnetic phase angle between the two sites [36] or the global lightning density map [37].

This figure illustrates that the SE kernel with optimized parameters is able to produce cross-correlation curves that are, on average, functionally similar to the true-signal curve. As discussed above, this behavior is deliberately conservative, maximizing the probability that our Gaussian process will spuriously yield cross-correlation measurements with favorable geodesy Bayes factors. As already was stated in the initial implementation of the GW-geodesy framework, having similar zero crossings might play a crucial role in mimicking a signal [29]. Everywhere in this analysis we investigate the GW-geodesy framework in the frequency range of 10–250 Hz with a frequency resolution of 0.24 Hz. The lower limit is based on the lowest frequency that can be observed by current ground-based interferometers. Because of the combined effects of the overlap reduction function and the reduced sensitivity at higher frequencies, there is a negligible gain when going to higher frequencies compared to the computational costs. For example, when looking for an isotropic power-law signal using the HL baseline with a power-law slope of $\alpha = 0$, $2/3$, and 3, 99% of the sensitivity during the latest O3 run was respectively contained below 76.1, 90.2, and 282.8 Hz [16].

In the rest of this paper, when we mention a Gaussian process signal, we are referring to a Gaussian process with a SE kernel and a length-scale parameter optimized for the relevant power-law slope α. Also, when mentioning "unknown correlated noise sources," we refer to this Gaussian process signal, which in this study serves as a conservative proxy for these unknown correlated noise sources.

IV. FALSE-ALARM PROBABILITIES AND DETECTION CONFIDENCE

A. Simulations

Using our Gaussian process machinery, we will explore the false-alarm probabilities and statistical significance associated with geodesy Bayes factors. We will make use of three different sources of cross correlation: a power-law signal with slope α, magnetic Schumann resonances, and a proxy for unknown terrestrial cross correlation mimicking an α-power-law signal using Gaussian processes. As a proof of concept we will start by investigating Bayes factors given by signals with SNR = 3 after 3 years of observation at LIGO's design sensitivity [38]. This specific signal-to-noise ratio is chosen as this is the fiducial value when one might first claim evidence for a detected GWB.

The signal-to-noise ratio of our model signal $\mathbf{C}_\alpha(\mathbf{f})$ is given by

$$\mathrm{SNR}_\alpha = \frac{\int_f \hat{C}(f)\mathbf{C}_\alpha(\mathbf{f}) \times \sigma_\alpha^2/\sigma^2}{\sigma_\alpha}, \quad \text{with}$$
$$\sigma_\alpha^2 = \left(\int_f \frac{\mathbf{C}_\alpha{}^2(\mathbf{f})}{\sigma^2}\right)^{-1}. \tag{12}$$

In what follows we will drop the subscript α and refer to the signal-to-noise ratio SNR_α as just SNR. One can compute the *expected* SNR or the *observed* SNR. In the former case $\hat{C}(f)$ is the injected signal, whereas in the latter case Gaussian-distributed noise consistent with Eq. (5) has been added to each frequency bin of $\hat{C}(f)$. When performing an injection for a certain power law with slope $\alpha = 0, 2/3, 3$, the signal is injected with an expected SNR_α of the desired strength. Note that in the cases of the Schumann signal and the Gaussian process signal, sometimes a power law with a different slope might be recovered with a higher SNR: $\mathrm{SNR}_{\tilde{\alpha}} > \mathrm{SNR}_\alpha$, where $\tilde{\alpha} \neq \alpha$.

Note that the SNR defined in Eq. (12) can be positive (if the data well match our signal model, that is, the true HL overlap reduction function with a power-law GW signal) or

negative (if data anticorrelate with our signal model). As we are concerned primarily with signals that we might *mistake* as astrophysical, we only investigate those that yield positive SNRs as observed by the HL baseline. Therefore, if the SNR of a Gaussian process draw is negative at the target α (0, 2/3 or 3), it is rejected and a new signal is simulated. This happens, on average, 50% of the time. The rejection of this signal is chosen to match a realistic experimental condition. In the case of a detection, one would like to apply this tool mainly to a positive SNR detection, whereas a negative SNR detection would immediately be categorized as unphysical.

For our three signal classes, we simulate and analyze 5000 simulated cross-correlation spectra. In the case of the astrophysical power law and Schumann resonances, this involves generating 5000 distinct Gaussian noise realizations that are added to the fixed underlying models. Under our Gaussian process, meanwhile, each trial involves a random draw from our Gaussian process (restricted to positive SNR) and a randomly generated noise spectrum. For every injection, we use PyMultiNest to compute Bayesian evidences [39], using 2000 live points. PyMultiNest is a PYTHON interface for MultiNest [40,41], which is an implementation of the nested sampling algorithm [42,43].

The model corresponding to our hypothesis $\mathcal{H}_\gamma$ has two free parameters: the reference amplitude $\Omega_{\rm ref}$ of the signal at $f_{\rm ref} = 25$ Hz, and the power-law slope α. We use a log-uniform prior for the reference amplitude between 10^{-12} and 10^{-6}. For α we use a Gaussian prior with a standard deviation of 3.5. Our alternative hypothesis $\mathcal{H}_{\rm Free}$ has three additional free parameters: the distance between the two interferometers Δx, and the rotation angles of the interferometers ϕ_1 and ϕ_2 [29]. We use uniform priors on ϕ_1 and ϕ_2 (0, 2π). Furthermore we use a uniform prior on $\cos\theta$, where $\Delta x = 2R_{\rm Earth}\sin\theta/2$. This corresponds to a prior on the distance between the detectors: $\mathrm{p}(\Delta x) \propto \Delta x$. These priors are chosen to be consistent with earlier work [29].

Figure 2 represents the log-Bayes distribution for the different signal models, assuming a SNR $= 3$ and $\alpha = 2/3$. The lower bin in the histograms also includes all simulations with a recovered log-Bayes factor smaller than -6. The smallest log-Bayes factor for an injection with a Schumann signal is approximately -23, whereas this is approximately -6×10^7 in the case of the Gaussian process signal.

First, we notice that the overlap of the histograms of the Gaussian process and the power-law signals is significantly larger than the overlap between the Schumann signal and the power-law signal. This shows that the selection of the Gaussian process parameters in Sec. III is successful and indeed yields a conservative condition, where the Gaussian process is able to mimic the power-law signal. However, it also shows what is possibly one of the weaknesses of the current implementation of the tool: our Gaussian process might be overly conservative as it is able to very well mimic the $\alpha = 2/3$ power law, and therefore to a large extent yield similar log-Bayes factors.

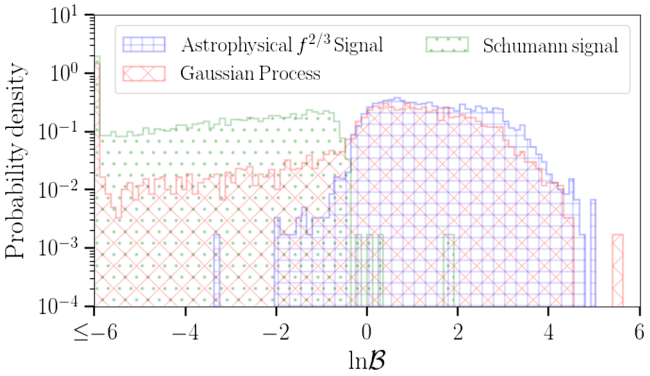

FIG. 2. Probability density of $\ln\mathcal{B}$ for an astrophysical $\alpha = 2/3$ power-law signal, random draws from a Gaussian process using our optimized (most conservative) SE kernel, and a Schumann signal. The lowest bin contains all $\ln\mathcal{B} \leq -6$. For each signal type, 5000 injections were performed with an injection strength of SNR $= 3$, when recovered with an $\alpha = 2/3$ signal model.

On the other hand, whereas power-law signals yield a handful of mildly negative log-Bayes values, the Gaussian processes give an extended tail towards negative log-Bayes factors, which will further grow as we increase the SNR of our injections. This illustrates the intrinsic random nature of the Gaussian process. Despite the process being able to produce signals mimicking the power-law signal, at the same time, other types of signals are produced that are not properly described by a power law.

Assuming the Gaussian process is a (very) conservative estimate of a terrestrial contamination for isotropic GWB searches, we can construct an upper bound on the FAP and detection probability associated with our geodesy Bayes factors. The FAP is the probability with which our terrestrial signal (for which the Gaussian process is our proxy) gives a geodesy Bayes factor as high as or higher than the Bayes factor we recover from our actual data. The detection probability gives the probability to detect the signal at a certain FAP/log-Bayes factor. The detection probability and FDP are linked to each other by FDP $=$ $1 -$ det. $prob.$, where the FDP is a figure of merit of the probability to wrongly reject the signal model.

For future observations, we would like the FAP to be as small as possible (unlikely to be a false signal) and a large detection probability (likely to be a true signal). Generally, before analyzing the data, one chooses a FAP considered to be the largest allowed value, such as 5% or 1%. Given a signal injected with a SNR $= 3$, the log-Bayes factors and detection probability are shown in Table I for a FAP of 5% and 1%. We also show the results when looking for a power-law signal with slope $\alpha = 0$ or 3. We see, for example, that given an apparent detection of the gravitational-wave background with SNR $= 3$ under an $\alpha = 2/3$

TABLE I. Log-Bayes factor and detection probabilities matching a FAP of 1% and 5% comparing a α-power-law GW signal and a Gaussian process signal. For each signal type, 5000 injections were performed with an injection strength of SNR = 3.

	$\alpha = 0$		$\alpha = 2/3$		$\alpha = 3$	
FAP	$\ln \mathcal{B}$	det. prob.	$\ln \mathcal{B}$	det. prob.	$\ln \mathcal{B}$	det. prob.
5.0%	2.95	6.92%	2.94	10.56%	3.42	15.44%
1.0%	3.68	0.76%	3.72	1.78%	4.04	8.10%

TABLE II. Log-Bayes factor and detection probabilities matching a FAP of 1% and 5% comparing a α-power-law GW signal and a Gaussian process signal. For each signal type, 5000 injections were performed with an injection strength of SNR = 5.

	$\alpha = 0$		$\alpha = 2/3$		$\alpha = 3$	
FAP	$\ln \mathcal{B}$	det. prob.	$\ln \mathcal{B}$	det. prob.	$\ln \mathcal{B}$	det. prob.
5.0%	3.54	6.78%	3.67	15.54%	3.91	67.08%
1.0%	4.12	0.62%	4.30	1.80%	4.71	41.40%

model, there is no more than a 5% chance that a log-Bayes factor $\ln \mathcal{B} = 2.94$ would arise by chance from a non-astrophysical signal.

We notice that for a given FAP, the detection probability becomes higher for power-law signals with a steeper slope (larger α). However, all reported detection probabilities are very small. This is linked to our Gaussian process being (overly) conservative and very good at mimicking the GW power-law signals. This was already clear from Fig. 2 for $\alpha = 2/3$.

Table II shows the same for an injected signal with SNR = 5 when looking for a power law with slope $\alpha = 0, 2/3$, or 3. The detection probability for a flat and 2/3 power law remains very small. Note that for a flat power law the detection probability even decreases slightly when going from SNR = 3 to SNR = 5. On the other hand, for an $\alpha = 3$ power law there is a drastic increase in detection probability when going from SNR = 3 to SNR = 5.

Table III, for comparison, shows log-Bayes factors and detection probabilities when one compares the power-law signals with a Schumann signal. For a fixed FAP of 1%, the detection probability is >99%, meaning the GW-geodesy tool is very effective at differentiating a Schumann signal from a power-law signal at a SNR of 3.

TABLE III. Log-Bayes factor and detection probabilities matching a FAP of 1% and 5% comparing a α-power-law GW signal and a Schumann signal. For each signal type, 5000 injections were performed with an injection strength of SNR = 3.

		$\alpha = 0$	$\alpha = 2/3$	$\alpha = 3$
$\ln \mathcal{B}$	FAP		det. prob.	
−0.82	5.0%	99.72%	99.64%	99.60%
−0.55	1.0%	99.44%	99.28%	99.28%

As searches for the gravitational-wave background accumulate SNR slowly over the course of months to years, the above scenario in which a candidate signal has a moderate SNR = 3 represents a realistic situation in which we will first need to use the Geodesy test. It is instructive, however, to more broadly investigate how the distributions of log-Bayes factors evolve as a function of SNR. To do this, we simulate signals with strengths logarithmically spaced between SNR = 0.1 and SNR = 100. Here we only look at the $\alpha = 2/3$ case, injecting astrophysical $\alpha = 2/3$ power laws and drawing random "terrestrial" signals from our Gaussian process optimized to this same power-law form.

As mentioned above, the free parameters are the reference signal strength $\Omega_{\rm ref}$ at 25 Hz and the power-law slope α in the case of our hypothesis $\mathcal{H}_\gamma$. For the hypothesis $\mathcal{H}_{\rm Free}$, the set of five free parameters consists of $\Omega_{\rm ref}$, α, Δx, ϕ_1, and ϕ_2. Note that for our astrophysical signal, even though we inject $\alpha = 2/3$, our best-fit α may well differ due to different noise instantiations, whereas for the Gaussian process realizations α will adjust to best fit the random signal with a power law.

In Fig. 3 the median of the α posterior is shown with respect to the SNR at which this signal would be observed by the HL baseline. Note that both α and SNR are calculated from the posterior consistent with the HL baseline. Although this might be disfavored with respect to the posteriors from the random baseline, we are interested in how the HL baseline would observe such a signal.

For an extremely small SNR (<1), α is closely centered around 0 for both the true power-law GW signal as well as the simulated Gaussian process signal. These signals are so weak that whatever we observe is dominated by the Gaussian background of our search. With respect to itself this Gaussian background has by definition no power-law slope and therefore matches $\alpha = 0$.

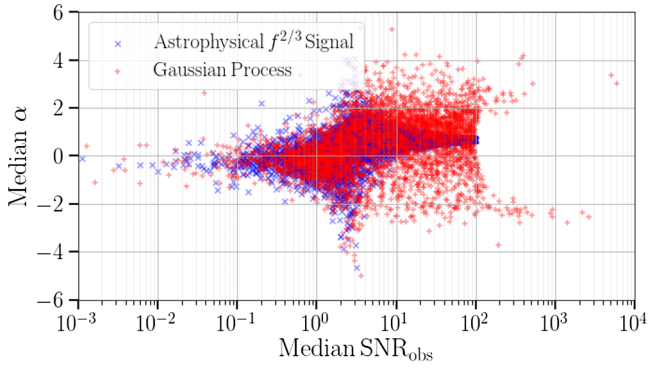

FIG. 3. The median of the posterior of the power-law slope α is represented against the median of the SNR-posterior, assuming the signal has been observed by the HL baseline. Each point in this scatter plot represents one injected signal. For each signal type, 5000 injections were performed with a logarithmically spaced injection strength between SNR $= 0.1$ and SNR $= 100$.

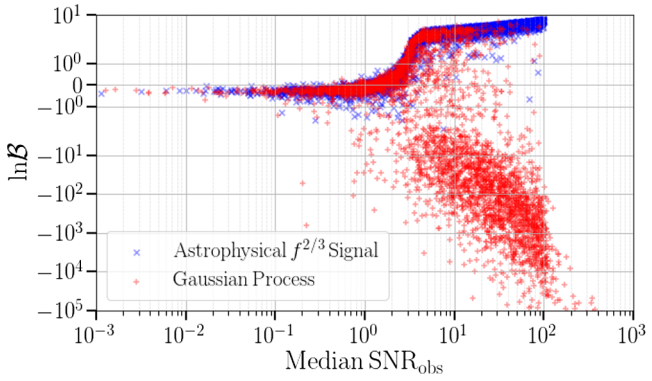

FIG. 4. The log-Bayes factor comparing a $2/3$-power-law signal with a Gaussian process is represented against the median of the SNR posterior, assuming the signal has been observed by the HL baseline. There are 40 events with $\ln\mathcal{B} < 10^{-5}$, which are not shown in this figure. The smallest $\ln\mathcal{B} = -5.3 \times 10^{7}$. Each point in this scatter plot represents one injected signal. For each signal type, 5000 injections were performed with a logarithmically spaced injection strength between SNR $= 0.1$ and SNR $= 100$.

When our signal has a strength of order SNR $= 1$, the retrieved values of α are still centered around zero, but the variance increases. The posterior-median α spans a range from large negative to large positive power-law slopes. We start to see some excess, but given the weak strength of the signal the randomness of the Gaussian background can drive the large negative or positive α.

The behavior for power-law GW signals and the Gaussian process signals starts to differ from SNR ~ 10. In the case of a power-law signal, the variation in α drastically decreases and the center of the retrieved values shifts from 0 to the real value: $2/3$. This kind of behavior is not present in the case of the Gaussian process signal, which remains centered around zero with a large variation.

This seems to indicate that with a large enough SNR, we can make a clear distinction between a power-law GW signal and some unknown (terrestrial) correlated noise sources, simulated by the Gaussian processes.

To further strengthen this statement, Fig. 4 shows the log-Bayes factor between $\mathcal{H}_\gamma$ and $\mathcal{H}_{\mathrm{Free}}$ as a function of the observed SNR. Weak signals are on average unable to differentiate between $\mathcal{H}_\gamma$ and $\mathcal{H}_{\mathrm{Free}}$, leading to log-Bayes factors ~ 0. When the observed SNR reaches values ~ 1, the log-Bayes factor of the $2/3$-power-law signals starts to prefer positive values, with increasing log-Bayes factor for increasing SNR. This is consistent with both our expectations for a true signal as well as the earlier results of the GW-geodesy tool [29].

At the same time, the data from the Gaussian process starts to separate into two categories when the SNR reaches order one. One population of signals starts to prefer negative log-Bayes factors, preferring $\mathcal{H}_{\mathrm{Free}}$ over $\mathcal{H}_\gamma$. These are the signals that look nothing like a power-law signal as observed by the HL baseline. However, there is also a population of signals that starts to prefer positive log-Bayes factors. These are the signals that succeeded in mimicking the HL baseline $2/3$ power-law signal to a (very) good extent. This behavior is expected because we purposely chose our Gaussian process parameters to have this kind of behavior, which is the key ingredient in creating conservative estimates for the FAP. However, as the observed SNR keeps increasing the population of signals with positive log-Bayes factors decreases at the cost of the population with negative log-Bayes factors. Although at high SNR the probability for a Gaussian process to have high log-Bayes factors is very small, it is still nonzero.

Figures 3 and 4 enable us to make statements on the distinctive character of our tool to differentiate a power-law GW signal and some unknown (terrestrial) correlated noise sources, simulated by the Gaussian process.

B. Detection probability curve

We construct a detection probability curve, for the situation where $\alpha = 2/3$, where we show the behavior of the detection probability versus the SNR of the signal for several fixed false-alarm rates. Injections at five different SNRs are performed: 1.25, 3, 5, 10, and 20. At each SNR, we perform 5000 injections for both the conservative Gaussian process signal and an $\alpha = 2/3$ power-law signal. The result is shown in Fig. 5.

The results indicate that our Gaussian process is very conservative. Given a FAP of 1%, the detection probability at SNR $= 20$ does not even reach 40%. A detection probability of 50% and 100% is indicated by the black dashed curves. Given a FAP of 5%, a detection probability of 50% is reached above SNR $= 10$.

As shown in Table II, in the case of a power law with SNR $= 5$ and a steeper slope, e.g., $\alpha = 3$, a detection

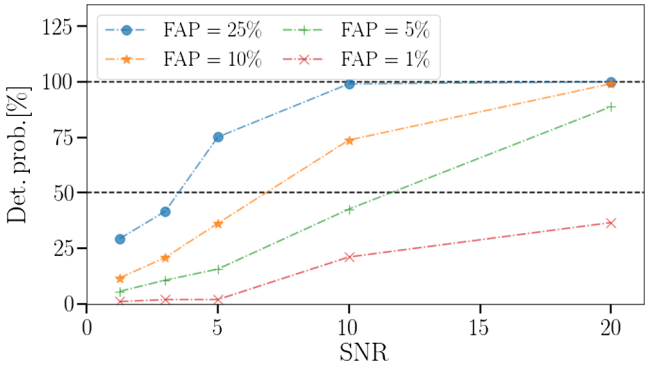

FIG. 5. Detection probability of a 2/3-power-law signal, for false-alarm probabilities of 1% (red), 5%(green), 10%(orange), and 25% (blue). The black dashed lines indicate a detection probability of 50% and 100%. Signals were injected at SNR = 1.25, 3, 5, 10, and 20. For each signal type and SNR, 5000 injections were performed.

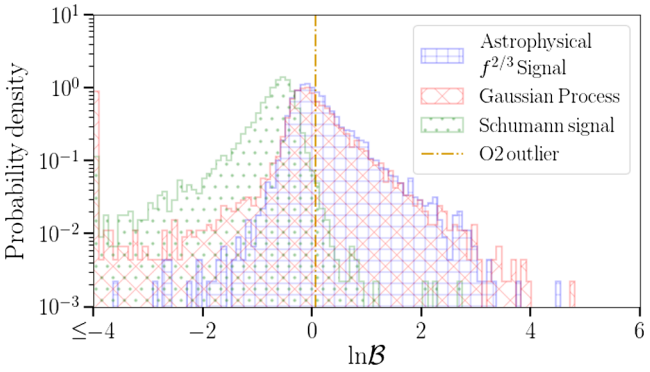

FIG. 6. Probability density of $\ln\mathcal{B}$ for an astrophysical $\alpha = 2/3$ power-law signal, random draws from a Gaussian process using our optimized (most conservative) SE kernel, and a Schumann signal. The lowest bin contains all $\ln\mathcal{B} \leq -4$. For each signal type, 5000 injections were performed with an injection strength of SNR = 1.25 The choice of $\alpha = 2/3$ and SNR = 1.25 matches the parameters of the observed outlier of the O2 results for an isotropic GWB [44].

probability curve of 41.40% (67.08%) is reached for a FAP of 1% (5%). This seems to indicate that even with our very conservative Gaussian process signal generation, a GWB signal with a steeper power-law slope is more easily distinguishable from correlated terrestrial noise, here modeled by the Gaussian process.

C. Application to a real-life scenario: O2 outlier

When analyzing the results of their O2 run, the LIGO and Virgo collaborations observed an excess of SNR = 1.25 for a power-law model with $\alpha = 2/3$, as well as $\alpha = 3$ [44]. At the time that the paper was published, it was stated that the low SNR excess was very likely due to random fluctuations in the data. This was confirmed by the lack of detection by the subsequent O3 results [16]. With the geodesy tool described in this paper we could, at the time of the O2 observation, have answered a complementary question: given that we have observed an excess with SNR = 1.25 for a power-law model with $\alpha = 2/3$, what is the probability that the observed signal is due to a source of correlated noise instead of gravitational waves? Although the O3 results have confirmed that the excess was just a random fluctuation, it is instructive to demonstrate how the geodesy tool could be used in the future. In our demonstration, we only investigate the excess given a power law with $\alpha = 2/3$, although one can easily apply the tool to the $\alpha = 3$ case as well.

In what follows we use the publicly available cross-correlation spectrum observed by LIGO during O2 [45] to compute the log-Bayes factor linked to this observation. To construct the FAP and detection probability, we perform injections of the conservative Gaussian process signal, Schumann resonances, and a 2/3-power-law signal with an observed injection strength of SNR = 1.25. All injections consist of 5000 samples and result in the distribution of log-Bayes factors shown in Fig. 6.

The log-Bayes factor for the observed signal of the O2 run by LIGO and Virgo was computed and found to be 0.063. Given the distributions from the simulations shown in Fig. 6, the observed signal is consistent with a FAP of $39.00\% \pm 0.02\%$ and a detection probability of $46.80\% \pm 0.02\%$. The high FAP does not give us enough confidence to prefer a gravitational-wave signal over a correlated noise source.

However, if one compares a gravitational-wave signal with a correlated signal coming from Schumann resonances a FAP of $1.00\% \pm 0.02\%$ is found, effectively ruling out Schumann resonances as a possible source with high confidence. This is consistent with projections showing that there was no significant magnetic coupling in the analysis for an isotropic GWB using O2 data [44].

V. CONCLUSION AND OUTLOOK

In this paper we presented a tool that requires the observed signal to be consistent with the geometry of the observing detectors. We used Gaussian processes as a conservative proxy for the unknown space of all terrestrial correlated signals that might impact stochastic searches. This enabled us to make quantitative statistical statements and false-alarm probabilities concerning the origin of the observed signal. The framework was applied to a SNR = 1.25 excess for a 2/3 power law, observed by the LIGO and Virgo collaborations during their second observing run. Based on this analysis, there was not enough evidence to prefer a GWB signal over terrestrial correlated noise. However, Schumann resonances were effectively ruled out as a possible source.

In this section we discuss how this tool can be used in the future, as well as the assumptions used in the current work and the possibilities for future improvements.

The primary use for this tool is to apply it to the observed cross correlation spectrum when (a hint of) a power-law isotropic GWB is observed. From analyzing the data, three estimated parameters will be needed as input for the GW-geodesy tool we are describing here: the power-law index α, the observed SNR, and the log-Bayes factor between $\mathcal{H}_{\text{Free}}$ and $\mathcal{H}_{\gamma}$ for the observed signal. The observed α will be used to reoptimize the kernel parameters to get the most conservative scenario for this specific power law. The SNR will dictate the injection strength of our data. This will lead to a figure equivalent to Fig. 2 in this work and log-Bayes factors linked to a certain FAP as in Table I. If the observed log-Bayes factor is larger than or equal to the log-Bayes factor linked to the desired $\text{FAP}_{\text{desired}}$ (fixed beforehand), we can state that the observed signal originates from gravitational waves instead of a source of correlated noise with a confidence of $1 - \text{FAP}_{\text{observation}}$.

In both its current and previous forms [29], the GW-geodesy tool can only be used to validate an isotropic GWB. We are currently working to extend this analysis for use with anisotropic GWB searches.

In this paper we only demonstrated the tool for the HL baseline, since this is currently the most sensitive detector pair for an isotropic GWB. However, it is very easy to apply this technique to any preferred detector pair. At some point the ever increasing detector sensitivities and long observation times observation times may enable the detection of a GWB by more than one detector pair. The current tool is able to make statements about all of the baselines separately, but one can imagine extending the tool to get one overall figure of merit to make statements about the entire detector network.

It is important to note that the framework demonstrated in this paper was only tested when there was either a power-law GW signal present or a globally coherent noise source. The separation between a true GW signal and correlated noise becomes less straightforward if they are both present at the same time with similar strengths. In the case of a known background, a technique (such as that proposed in Ref. [28]) could be used to search for both sources at the same time.

When applying this tool to a signal coming from Schumann resonances, we assumed the magnetic Schumann spectrum to have a positive cross power, since the (possibly frequency-dependent) sign is not known for the Hanford-Livingston baseline. In the future, a departure from this assumption could be investigated by relying on analytical models describing Schumann resonances and their global cross power [36,37].

Instead of only considering power-law signals, one could also investigate more complex models. In the earlier implementation of the GW-geodesy framework [29] it was shown that the tool is quite robust against modeling a broken power law with a single power law. This could mean that the framework is mainly sensitive to the zero crossings of the overlap reduction function.

ACKNOWLEDGMENTS

The authors acknowledge access to computational resources provided by the LIGO Laboratory supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This paper has been given LIGO DCC number P2100383 and Virgo TDS number VIR-1126A-21. This material is based upon work supported by Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO). K. J. is supported by FWO-Vlaanderen via Grant No. 11C5720N. M. W. C. acknowledges support from the National Science Foundation with Grants No. PHY-2010970 and No. OAC-2117997. I. M. was supported by the University of Florida through a Graduate Teaching Assistantship.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).
[2] J. Aasi *et al.*, Classical Quantum Gravity **32**, 074001 (2015).
[3] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[4] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (The KAGRA Collaboration), Phys. Rev. D **88**, 043007 (2013).
[5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).
[6] R. Abbott *et al.*, Phys. Rev. X **11**, 021053 (2021).
[7] R. Abbott, T. D. Abbott *et al.*, Astrophys. J. Lett. **915**, L5 (2021).
[8] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley *et al.* (The LIGO Scientific, the Virgo, the KAGRA Collaborations), arXiv:2111.03606.
[9] R. Abbott *et al.*, arXiv:2108.01045.
[10] T. Regimbau and V. Mandic, Classical Quantum Gravity **25**, 184018 (2008).
[11] P. A. Rosado, Phys. Rev. D **84**, 084004 (2011).
[12] X.-J. Zhu, E. Howell, T. Regimbau, D. Blair, and Z.-H. Zhu, Astrophys. J. **739**, 86 (2011).
[13] S. Marassi, R. Schneider, G. Corvino, V. Ferrari, and S. P. Zwart, Phys. Rev. D **84**, 124037 (2011).
[14] C. Wu, V. Mandic, and T. Regimbau, Phys. Rev. D **85**, 104024 (2012).

[15] X.-J. Zhu, E. J. Howell, D. G. Blair, and Z.-H. Zhu, Mon. Not. R. Astron. Soc. **431**, 882 (2013).
[16] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. D **104**, 022004 (2021).
[17] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. D **104**, 022005 (2021).
[18] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos *et al.*, Living Rev. Relativity **23**, 3 (2020).
[19] N. Christensen, Rep. Prog. Phys. **82**, 016903 (2019).
[20] J. D. Romano and N. J. Cornish, Living Rev. Relativity **20**, 2 (2017).
[21] W. O. Schumann, Z. Naturforsch. Teil A **7**, 149 (1952).
[22] W. O. Schumann, Z. Naturforsch. Teil A **7**, 250 (1952).
[23] P. B. Covas, A. Effler, E. Goetz, P. M. Meyers, A. Neunzert, M. Oliver, B. L. Pearlstone, V. J. Roma, R. M. S. Schofield, V. B. Adya *et al.*, Phys. Rev. D **97**, 082002 (2018).
[24] E. Thrane, N. Christensen, and R. M. S. Schofield, Phys. Rev. D **87**, 123009 (2013).
[25] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler, Phys. Rev. D **90**, 023013 (2014).
[26] M. W. Coughlin *et al.*, Classical Quantum Gravity **33**, 224003 (2016).
[27] M. W. Coughlin *et al.*, Phys. Rev. D **97**, 102007 (2018).
[28] P. M. Meyers, K. Martinovic, N. Christensen, and M. Sakellariadou, Phys. Rev. D **102**, 102005 (2020).
[29] T. A. Callister, M. W. Coughlin, and J. B. Kanner, Astrophys. J. **869**, L28 (2018).
[30] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Classical Quantum Gravity **37**, 055002 (2020).
[31] M. Isi, R. Smith, S. Vitale, T. J. Massinger, J. Kanner, and A. Vajpeyi, Phys. Rev. D **98**, 042007 (2018).
[32] B. Allen and J. D. Romano, Phys. Rev. D **59**, 102001 (1999).
[33] N. Christensen, Phys. Rev. D **46**, 5250 (1992).
[34] T. Regimbau, Res. Astron. Astrophys. **11**, 369 (2011).
[35] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, MA, 2006).
[36] Y. Himemoto and A. Taruya, Phys. Rev. D **96**, 022004 (2017).
[37] Y. Himemoto and A. Taruya, Phys. Rev. D **100**, 082001 (2019).
[38] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos *et al.*, Living Rev. Relativity **23**, 3 (2020).
[39] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, Astron. Astrophys. **564**, A125 (2014).
[40] F. Feroz and M. P. Hobson, Mon. Not. R. Astron. Soc. **384**, 449 (2008).
[41] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).
[42] J. Skilling, Bayesian Anal. **1**, 833 (2006).
[43] J. Skilling, AIP Conf. Proc. **735**, 395 (2004).
[44] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **100**, 061101 (2019).
[45] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Data for a search for the isotropic stochastic background using data from Advanced LIGO's second observing run, Technical Report No. LIGO-T1900058-v3, 2019.