

Quasinormal modes from numerical relativity with Bayesian inference

Richard Dyer^{1,*} and Christopher J. Moore^{1,2,3,†}

¹*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*

²*Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK*

³*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*

(Dated: October 15, 2025)

Numerical relativity (NR) enables the study of physics in strong and dynamical gravitational fields and provides predictions for the gravitational-wave signals produced by merging black holes. Despite the impressive accuracy of modern codes, the resulting waveforms inevitably contain numerical uncertainties. Quantifying these uncertainties is important, especially for studies probing subdominant or nonlinear effects around the merger and ringdown. This paper describes a flexible Gaussian-process model for the numerical uncertainties in all the spherical-harmonic waveform modes across a state-of-the-art catalog of NR waveforms and a highly efficient procedure for sampling the posteriors of quasinormal mode models without the need for expensive Markov chain Monte Carlo. The Gaussian-process model is used to define a likelihood function which allows many Bayesian data analysis techniques - already widely used in the analysis of experimental gravitational wave data - to be applied to NR waveforms as well. The efficacy of this approach is demonstrated by applying it to the analysis of quasinormal modes in Cauchy-characteristic evolved waveforms.

I. INTRODUCTION

The final stage of a black hole (BH) merger, known as the ringdown, is associated with the remnant object settling into its final Kerr state. During this stage, perturbation theory predicts the existence of damped sinusoidal oscillations, known as quasinormal modes (QNMs), at specific frequencies as a prominent component of the gravitational wave (GW) signal. These QNM frequencies contain information about the remnant BH and provide a promising avenue for testing general relativity with strong and dynamical gravitational fields, including the ‘no-hair’ theorem [1–4].

The ground-based GW observatories LIGO [5] and Virgo [6] have now detected several tens of high-mass binary BH mergers [7], with many more expected in the upcoming observing run [8]. The fundamental QNM is confidently identified in many of these events [9–11], with signs of additional QNMs [12], possibly including overtones [13–15], in several. These observations have already been used to test the BH no-hair [16] and area theorems [17–19].

BH ringdown can also be studied using numerical relativity (NR) by fitting QNM models to the ringdown waveform. This can be done either by analyzing the GW signal extracted in a particular viewing direction, by fitting to individual spherical harmonic modes (it is common to look at just the (2, 2) spherical mode, for example, see [20]), or by fitting multiple modes simultaneously [21–23] and effectively averaging over viewing directions. Traditionally, this is done using linear least-squares fitting (for example, see [20, 24, 25]), treating the QNM

amplitudes as free parameters. Other techniques, such as nonlinear extensions to least-squares, have also been proposed. However, these approaches face known challenges, including ambiguity in selecting the appropriate QNM content and the risk of overfitting [20, 23, 25, 26]. Moreover, least-squares fitting implicitly assumes stationary white noise and yields only point estimates (without uncertainties) for the QNM model parameters. Another alternative are rational filters [27], however while these can identify QNMs, they do not give their amplitudes.

As NR simulations and waveform extraction techniques improve, and as GW catalogs expand, there is a growing demand for accurate and physically motivated approaches to QNM modeling. A natural progression is a Bayesian approach, which accounts for fit uncertainties via a posterior distribution over the QNM parameters. There has already been one attempt at performing Bayesian QNM fits to NR [28] and this paper introduces a new framework which extends this in several ways. We introduce a new physically-motivated Gaussian process (GP) model which is trained on an NR catalog. Additionally, the QNM model is expressed in a way which (with a natural choice of prior) makes the posterior a multivariate normal distribution which can be efficiently sampled. This removes the need for Markov chain Monte Carlo methods which do not scale well to models involving many QNMs. This makes the computational cost of the new Bayesian approach comparable to that of least-squares fitting, therefore making it a viable alternative and supports scalable analyses across large catalogs in the future.

The Bayesian approach naturally yields posterior distributions over the QNM parameters, as opposed to just point estimates obtained from least-squares fitting; this is a key advantage of the new Bayesian approach. The posterior provides the foundation for assigning a measure of significance to the various QNMs in the model and

* richard.dyer@ast.cam.ac.uk

† christopher.moore@ast.cam.ac.uk

for performing a selection between models using different numbers or types of QNMs [29]. This is advantageous because point estimates alone cannot determine whether or not the NR data supports including a specific QNM in the model. This is expected to be increasingly important in future studies focusing on subdominant and higher order effects in the ringdown such as nonlinear quadratic QNMs [25, 26] and power-law tails [30].

This paper uses the public catalog of 13 Cauchy-Characteristic Evolved (CCE) waveforms produced by the Spectral Einstein Code (SpEC) [31, 32]. These waveforms more accurately extract the GW information at future null infinity than extrapolated methods [32]. This information is contained in the Bondi news function $\mathcal{N}$ (although the strain h or the Weyl curvature scalar Ψ_4 can also be used).

A Gaussian process (GP) is used to model the numerical uncertainties in these waveforms. The model is trained on the full catalog of CCE simulation residuals to create a bespoke kernel which is used to define the Bayesian likelihood. This provides a physically-motivated, flexible, and conservative estimate of the waveform uncertainty in the catalog.

This is a methods paper, in which the Bayesian QNM inference procedure and Gaussian process are described. This new approach is demonstrated by applying the techniques to two simple models. Results from these analyses showcase the potential of this methodology and provide a case study for the significance and posterior predictive check which are also introduced.

II. QNM MODEL AND LIKELIHOOD FUNCTION

The model described in this section can be applied to any of the GW quantities h , $\mathcal{N}$, or Ψ_4 . For concreteness, all quantities in this section are written in terms of h .

The complex GW strain, $h = h_+ - ih_\times$ can be expanded in (spin-weight) spherical harmonics, ${}_{-2}Y^{\ell m}$ as

$$rh(t, \theta, \phi) = M \sum_{\beta} h^{\beta}(t) {}_{-2}Y^{\beta}(\theta, \phi), \quad (1)$$

where $\beta = (\ell, m)$ denotes the pair of spherical harmonic indices and the sum extends over $\ell \geq 2$ and $|m| \leq \ell$. The harmonic modes $h^{\beta}(t)$ are obtained as an output of NR simulations (see Sec. III).

The QNM model for the GW strain consists of a sum of damped sinusoids. Following the notation of Ref. [24], this model is written as

$$rh(t, \theta, \phi) = M \sum_{\alpha} C_{\alpha} e^{-i\omega_{\alpha}(t-t_0)} {}_{-2}S_{\alpha}(\chi; \theta, \phi), \quad (2)$$

where $\alpha = (\ell, m, n, p)$ denotes the quadruple of spheroidal harmonic indices that uniquely identifies a QNM and the sum extends over $\ell \geq 2$, $|m| \leq \ell$, $n \geq 0$, and $p \in \{-, +\}$. The quantity t_0 is the ringdown start time and the model

in Eq. 2 is understood to apply for $t > t_0$. In practice, it is also necessary to choose an end time T for the analysis, so $t_0 \leq t \leq T + t_0$. This QNM model has any free amplitude parameters C_{α} . Optionally, the remnant BH's mass and spin, M_f and χ_f , can be included as free parameters. The start time t_0 is treated as fixed.

When the model described in Eq. 2 is fit to the strain, the resulting amplitudes are given in the so-called ‘strain-domain’. The Bondi news, which is the first derivative with respect to time of the strain, can also be fit using this model with the amplitudes in the ‘news-domain’. The same applies to the Weyl curvature scalar, which is the second derivative with respect to time of the strain. The various amplitudes are related as follows,

$$C_{\alpha}[h] = iC_{\alpha}[\mathcal{N}]/\omega_{\alpha} = -C_{\alpha}[\Psi_4]/\omega_{\alpha}^2, \quad (3)$$

where square brackets have been used to indicate the respective domain of the amplitude. The units of $C_{\alpha}[h]$, $C_{\alpha}[\mathcal{N}]$, and $C_{\alpha}[\Psi_4]$ are 1, M^{-1} , and M^{-2} respectively.

The spherical and spheroidal harmonics are related by the mode-mixing coefficients μ_{α}^{β} (see, for example, Eq. 4 of Ref. [24]) which allows the QNM model in Eq. 2 to be re-expanded in the spherical harmonic basis;

$$rh(t, \theta, \phi) = \sum_{\beta} rh^{\beta}(t) {}_{-2}Y^{\beta}(\theta, \phi), \quad (4)$$

where $rh^{\beta}(t) = M \sum_{\alpha} \mu_{\alpha}^{\beta} C_{\alpha} e^{-i\omega_{\alpha}(t-t_0)}.$

The μ_{α}^{β} coefficients and ω_{α} frequencies were calculated as functions of M_f and χ_f using the `qnm` package [33].

Both the NR data and QNM model are discretely sampled time series evaluated at sample times t_a . Indices $a, b, \dots \in \{1, 2, \dots N\}$ label points in the time series. If the numerical uncertainty on the waveform modes h^{β} is assumed to be independent and distributed as a zero-mean Gaussian in each spherical harmonic mode, then the log-likelihood function is given by

$$\log P(\mathbf{h}|\boldsymbol{\theta}) = \sum_{\beta} \frac{-1}{2} \langle h^{\beta} - h^{\beta} | h^{\beta} - h^{\beta} \rangle_{\beta}, \quad (5)$$

where the data are the NR waveform modes (denoted collectively by $\mathbf{h}$) and the model parameters $\boldsymbol{\theta}$ are the complex mode amplitudes C_{α} and (possibly) M_f and χ_f .

In Eq. 5, the angle brackets with a subscript β denote an inner product between two functions of time, $f(t)$ and $g(t)$ (for each spherical β mode) discretely sampled at the times t_a . This inner product is given by

$$\langle f|g \rangle_{\beta} = \text{Re} \left\{ \sum_{a,b} f_a^* g_b \left((\mathbf{K}^{\beta})^{-1} \right)_{ab} \right\}, \quad (6)$$

where $\mathbf{K}^{\beta}$ is the covariance matrix in each mode and a star $*$ denotes complex conjugation. This covariance is designed to model the numerical uncertainty in each spherical waveform mode. This is modeled using a GP

that is trained on a public catalog of waveforms that are described in Secs. III and IV.

The inner product in Eq. 6 explicitly depends on the choice of sample times, t_a ; these are usually chosen to be regularly spaced in the interval $[t_0, T + t_0]$ with some sampling frequency $1/\Delta t$. As the sampling frequency is increased and the sample points become dense in $[t_0, T + t_0]$ this sum converges to the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ inner product $\langle f|g \rangle_{\mathcal{H}_k}$ [34]. This is verified in Appendix F.

If the components of all the covariance matrices $\mathbf{K}_\beta$ are taken to be $\sigma^2 \delta_{ij}$ (i.e. white noise, with the same σ for each β mode) and the maximum likelihood value of θ is used as a point estimator for the parameters of the QNM model, then this reduces to the least-squares minimization methods that have been used almost exclusively so far in the literature. This paper improves on this in two ways. Firstly, instead of using a white-noise covariance, we introduce a physically-motivated GP model for the uncertainty in the NR waveforms and numerically train this model on a NR waveform catalog. Secondly, instead of using a point estimate for the QNM model parameters, we place a prior on θ and analytically sample from the resulting posterior.

Expressing the complex mode amplitudes in terms of their real and imaginary parts, $C_\alpha = \text{Re } C_\alpha + i \text{Im } C_\alpha$, and treating these as the $(2N)$ parameters, the QNM model in Eq. 2 is fully linear. Additionally, if the remnant BH mass and spin, M_f and χ_f , are to be included as free parameters then we linearize the model by Taylor expanding about a chosen reference point. This reference point is taken by performing a linear least-squares fit of the QNM model to the NR data, using the fixed asymptotic Bondi data (ABD) values of the mass and spin. The best fitting values for the amplitudes, in addition to the ABD mass and spin, are taken as the reference parameters and are denoted θ_* . The best fitting model evaluated at these parameters is denoted $H_*^\beta(t)$. Alternatively, a non-linear least-squares fit can be used to obtain mass and spin reference values, in addition to the amplitudes. This method is slower and does not significantly improve the fits. This is discussed further in App. A.

The QNM model in Eq. 2 is expanded to first order in changes in the model parameters about the chosen reference. In other words, the model is put into the form

$$h^\beta(t) = H_*^\beta(t) + \sum_\mu (\theta^\mu - \theta_*^\mu) h_\mu^\beta(t). \quad (7)$$

The quantities $h_\mu^\beta(t)$, sometimes called the model matrix, are built from the first derivatives of the model in Eq. 2 evaluated at the reference. Most of the differentiation involved in constructing the model matrix can be done analytically. The exceptions are the derivatives of the QNM frequencies ω_α and the mode-mixing coefficients μ_α^β with respect to the remnant BH spin χ_f ; these derivatives were taken numerically using finite differences.

The linearization of the QNM model in the mass and spin parameters necessarily introduces an additional

source of error. However, this is small because typically fitting to state-of-the-art NR waveforms leads to small uncertainties in these parameters. The typical maximum fractional error introduced in the QNM frequencies is less than $\sim 10^{-7}$. Quantities like ω_α and μ_α^β that depend on these parameters vary only slightly across the width of the typical set of the posterior and this is expected to be well approximated by a linear expansion about a suitably chosen reference point. The accuracy of the linear approximation is discussed further below and is quantified in App. A.

Initially, flat priors were used on all model parameters, $\pi(\theta) = \text{const.}$ Hereafter, this is referred to as ‘Prior 1’. (It would also be possible to use a multivariate Gaussian prior on θ without significantly changing any of the following conclusions.) Because the model is linear in all of its parameters, with this choice of prior the posterior is exactly a multivariate Gaussian;

$$\begin{aligned} \theta|h &\sim \mathcal{N}(m^\mu, (\Gamma^{-1})^{\mu\nu}), \\ \text{where } m^\mu &= \theta_*^\mu + (\Gamma^{-1})^{\mu\rho} \sum_\beta \langle h_\rho^\beta | d^\beta - H_*^\beta \rangle_\beta, \\ \text{and } \Gamma_{\mu\nu} &= \sum_\beta \langle h_\mu^\beta | h_\nu^\beta \rangle_\beta. \end{aligned} \quad (8)$$

This multivariate Gaussian posterior can be sampled extremely efficiently. First, the posterior covariance $\Gamma_{\mu\nu}$ (numerically equal to the Fisher information matrix) is evaluated and inverted using its eigenvalue/vector decomposition (computed using `scipy.linalg.eigh`). For reasons of numerical stability, it is necessary to regularize some of the very small eigenvalues of the Fisher matrix; any eigenvalues smaller than a specified tolerance value $\epsilon = 10^{-10}$ were replaced with the value ϵ . The mean vector m^μ and covariance matrix Γ^{-1} were then calculated and the posterior in Eq. 8 is then sampled analytically.

The ability to sample the multivariate Gaussian posterior in Eq. 8 analytically is key to the efficiency of the method. Moreover, the posterior only has a Gaussian form because of the choice of prior described above: a flat prior on the real and imaginary parts of the amplitudes. It is sometimes desirable to switch to using a prior that is instead flat on the mode amplitudes $|C_\alpha|$ and flat and periodic in the phase angles $\arg(C_\alpha)$. Hereafter, this choice is referred to as ‘Prior 2’. Sampling directly with this prior would be much less efficient because the posterior would no longer be a multivariate Gaussian. Therefore, if samples with Prior-2 posterior are needed, these are obtained by first sampling the Prior-1 posterior, as described above, and then using importance sampling to reweight the samples to the new prior. The importance sampling weights are given by

$$w_n = \prod_\alpha \frac{1}{|C_{\alpha,n}|}, \text{ for } n = 1, 2, \dots, N_{\text{samples}}, \quad (9)$$

where n indexes the posterior samples.

Care must be taken when using importance sampling that the effective number of independent samples doesn’t

drop too low. In the present case, this is liable to happen when there are a large number of QNMs included in the model with small amplitudes (see Eq. 9). The effective number of independent samples is quantified using [35]

$$n_{\text{eff}} = \frac{(\sum_n w_n)^2}{\sum_n w_n^2} \leq N_{\text{samples}}. \quad (10)$$

A major benefit of sampling from the full Bayesian posterior on the QNM mode parameters, as compared with earlier approaches that relied on point estimates, is that the extent of the posterior can be used to quantify the support for the point $C_\alpha = 0$ for any given mode α . This point corresponds to that QNM not being included in the model. This allows us to assign a well-motivated Bayesian measure of significance S_α to each QNM in the fit. This significance is closely related to the Bayes factor, or evidence ratio, between two QNM models with and without the α QNM included (but otherwise with identical QNM content). The precise definition of the significance and its relationship to the Bayes factor is explained in App. B.

III. CCE WAVEFORM CATALOG

The waveforms used in this study come from the public catalog [36] of 13 Spectral Einstein Code (SpEC) waveforms [37] extracted using SpECTRE CCE [31, 32, 38]. The simulations were transformed into the superrest frame $300M$ after the peak strain using `scri` [39–42] and shifted in time so that the peak of the $\ell = m = 2$ mode strain occurs at $t = 0$ [22, 43, 44].

In principle, the methods described here could be applied to either the strain $\mathfrak{h}$, the news $\mathcal{N}$, or the Weyl curvature scalar ψ_4 (see App. K). Typically, fits are done to the strain, however after performing the supertranslation on the simulations, there can be some residual error which manifests as the strain not perfectly decaying to zero. These imperfections are not present in the news, which always decays to zero [45]. Further details of this are discussed in App. J. Therefore in this study we choose to work with the news and report amplitudes in the news-domain.

For each simulation, the highest resolution level (L5) extracted at the second smallest worldtube radius was used as the preferred waveform.

Information about the uncertainty in the NR waveform can be found in the residuals $\mathfrak{r}^\beta(t)$, defined as the difference from the second-highest resolution level (L4) at the same worldtube radius. This gives an estimate of the uncertainty on the second-highest resolution level, however here it is used as a *conservative* estimate for the uncertainty on the highest resolution waveform. Full details on the processing of the numerical waveforms are given in App. C.

These waveform residuals are used to train a GP model for the uncertainty in the preferred waveform; this is described in Sec. IV.

IV. GP MODEL FOR THE WAVEFORM UNCERTAINTIES

The purpose of this section is to describe the model for the uncertainty in NR merger and ringdown waveforms. This is used to define the QNM likelihood described in Sec. II. The model describes the uncertainty in the real and imaginary parts of the waveform in each spherical harmonic mode β in each simulation i of the catalog independently. The model is taken to be a Gaussian Process (GP) with a physically-motivated choice for the kernel that includes a number of free parameters. The values of these free parameters are then learned by maximizing the likelihood of the GP model evaluated on the training data which is taken to be the residuals $\mathfrak{r}_i^\beta(t)$ for each waveform mode of each simulation in the catalog. The resulting model for the uncertainty is illustrated in Fig. 1 where it is plotted alongside some of the residuals that were used in the training. The rest of this section describes the details of this GP model.

Examination of the residuals shows that they share many properties with the waveforms themselves. The functions $\mathfrak{r}_i^\beta(t)$ are smooth and vary on timescales comparable to the GW period. The amplitude of the residuals also tracks the relevant quantity (either the strain, news, or curvature scalar), with the largest residuals found in the $(\ell, m) = (2, 2)$ mode around merger. The residuals also show common behavior across all the simulations in the catalog. This is exploited here by pooling kernel GP parameters to model the waveform uncertainty, with kernel properties inferred by training the Gaussian Process on the full catalog, resulting in a set of GP parameters which are learned from the waveforms used.

After experimenting with a number of options, the following kernel, hereafter the *standard kernel* (or simply GP for short), was chosen to model the numerical uncertainty in the waveforms,

$$\begin{aligned} k_i^\beta(t, t') &= k_{\text{Stationary}}(t, t'; P_i^\beta) \\ &\times \text{SMOOTHMIN}\left(\sigma_{0,i}^\beta e^{-t/\tau_i^\beta}, \sigma_{\text{max},i}^\beta; s\right) \\ &\times \text{SMOOTHMIN}\left(\sigma_{0,i}^\beta e^{-t'/\tau_i^\beta}, \sigma_{\text{max},i}^\beta; s\right). \end{aligned} \quad (11)$$

The first term in Eq. 11 is a stationary kernel, which is intended to capture the fact that the residuals vary on a predictable timescale, P_i^β related to the real part of the fundamental QNM frequency.

One possible choice for $k_{\text{Stationary}}(t, t'; P_i^\beta)$ would be the squared-exponential (SE) kernel, which would capture the periodicity of the residual. However, this results in a dense matrix with entries that become very small compared to the diagonal values at large $|t - t'|$. This can slow computations and increase numerical instability. For this reason, we instead use a $q = 3$ Wendland covariance function [34] (see App. H for details). This mimics the SE kernel but with compact support leading to a band structure in the covariance matrix. We find this choice improves numerical stability when inverting

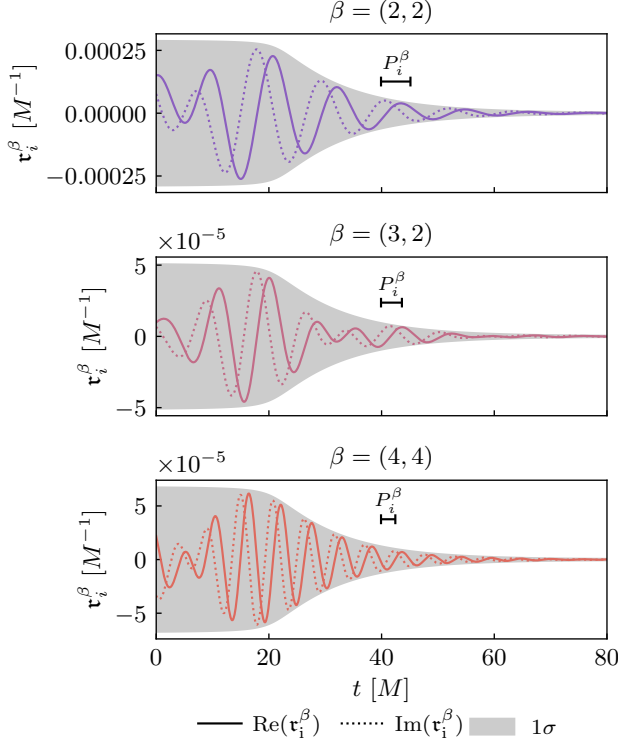


FIG. 1. These plots illustrate the standard GP model (Eq. 11) for the uncertainty in selected waveform modes ($\beta = (2, 2)$, $(3, 2)$, and $(4, 4)$ in the top middle and bottom panels respectively) of simulation $i = 0001$. The colored curves show the real and imaginary parts of the residuals in the Bondi news, $\tau_i^\beta(t)$. The shaded regions show the amplitude of GP model for the waveform uncertainties. What cannot be seen from the shaded regions is the timescale over which the GP model is correlated; this is illustrated with the horizontal bars on each panel, scaled to the squared-exponential timescale as described in the text.

the covariance matrix. In principle the sparse structure of the matrix can also speed up computations when using the appropriate solver functions, however we do not take advantage of this here.

The timescale P_i^β is closely related to period of the GW oscillations and is therefore expected to be close to the real part of the frequency ω_α of the fundamental ($n = 0$) prograde QNM with the same ℓ and m . Therefore, this timescale is taken to be

$$P_i^\beta = \frac{2\pi\mu}{\text{Re}\{\omega_{\ell m 0+}(M_{f,i}, \chi_{f,i})\}}. \quad (12)$$

The QNM frequencies are functions of the remnant mass and spin, $M_{f,i}$ and $\chi_{f,i}$, for each simulation i in the catalog which are known from the NR simulation; specifically, these values were obtained from the ABD object in the `scri` package. The pooled GP parameter μ is introduced to adjust this estimate for the timescale and is pooled across all spherical harmonic modes in all simulations in the catalog.

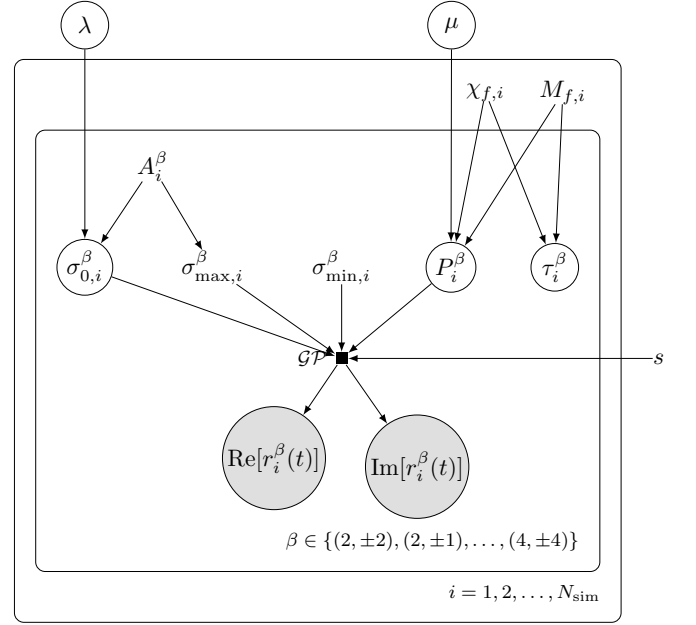


FIG. 2. PGM of the GP model for the NR waveform uncertainty described in the main text. The observed variables (shaded circles) are the real and imaginary parts of the waveform residuals in each simulation in the catalog (indexed by i) and in each spherical harmonic model (indexed by β). The residuals are the differences between the two highest resolution simulation available in the catalog. The parameters of the model are shown in empty circles. The values of the two pooled parameters in the top row are ultimately to be inferred from the NR catalog. The other latent parameters are determined from the pooled parameters and from the various fixed quantities which are shown without circles.

The timescale in the Wendland covariance function can be related to the timescale in the squared-exponential function by equating their second derivatives at the origin. This gives SE Timescale = $\sqrt{5/72} P_i^\beta$. This timescale is plotted using the horizontal bars in Fig. 1.

The next terms in Eq. 11 control the time dependence of the typical amplitude of the residuals. The function,

$$\sigma_{0,i}^\beta \exp\left(-\frac{t}{\tau_i^\beta}\right), \quad (13)$$

imposes the expected decay during the ringdown. The scale $\sigma_{0,i}^\beta$ controls the amplitude of the residuals and is estimated using the maximum amplitude A_i^β of the numerical residuals,

$$\sigma_{0,i}^\beta = \lambda A_i^\beta, \quad \text{where } A_i^\beta = \max_t |\tau_i^\beta(t)|. \quad (14)$$

The pooled GP parameter λ tunes this estimate.

The parameter τ_i^β controls the rate of the exponential decay and is estimated using the imaginary part of the fundamental prograde QNM with the same ℓ and m ,

$$\tau_i^\beta = \frac{-1}{\text{Im}\{\omega_{\ell m 0+}(M_{f,i}, \chi_{f,i})\}}. \quad (15)$$

TABLE I. The spherical harmonic modes for each simulation that were used for training the GP model. We focus on the loudest modes $\beta = (2, \pm 2), (2, \pm 1), (3, \pm 3), (3, \pm 2), (4, \pm 4)$ with some omissions on symmetry grounds. For the non-precessing simulations 0001-0007 and 0010-0012 the modes satisfy $h_{\ell,m} = h_{\ell,-m}^*$, hence the $m < 0$ modes are excluded as they carry no extra information. For the equal-mass aligned-spin simulations 0001-0004 (and for the ‘superkick’ simulation 0009) the odd m modes are suppressed and were excluded.

Simulation ID	GP Training Modes
0001-0004	(2, 2), (3, 2), (4, 4)
0005-0007, 0010-0012	(2, 2), (2, 1), (3, 3), (3, 2), (4, 4)
0008, 0013	(2, ± 2), (2, ± 1), (3, ± 2), (3, ± 3), (4, ± 4)
0009	(2, ± 2), (3, ± 2), (4, ± 4)

The exponential decay only occurs during the ring-down. During the merger, the amplitude of the residuals is expected to be roughly constant with time. This is achieved in our model using the SMOOTHMIN function which limits the maximum value of output to

$$\sigma_{\max,i}^{\beta} = \nu A_i^{\beta} \quad (16)$$

at early times, where we fix $\nu = 1.1$. At intermediate times, the function smoothly transitions to the exponential decay with the smoothness controlled by the parameter s which was fixed to 10^{-3} . Further details about the SMOOTHMIN function are given in App. D.

At very late times, the residuals are expected to, again, become roughly constant with time when the signal has decayed below the level of the numerical noise floor of the simulation. To model this, and to prevent issues relating to vanishing numerical uncertainties and ensure numerical stability we add a small jitter term along the diagonal of covariance matrix that is scaled relative to the average amplitude of the late-time residuals $\sigma_{\min,i}^{\beta}$,

$$\sigma_{\min,i}^{\beta} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} |\mathbf{r}_i^{\beta}(t)| dt \quad (17)$$

$$k_{\text{jitter}}(t_a, t_b) = (\epsilon \sigma_{\min,i}^{\beta})^2 \delta_{ab}, \quad (18)$$

where $t_2 = 300$, $t_1 = 250$. We fix $\epsilon = 1 \times 10^2$, which prevents any eigenvalues of the covariance matrix becoming negative and gives a conservative estimate of the late-time uncertainty. The jitter cannot be seen in Fig. 1 because it is too small. However, it can be seen clearly in Fig. 14 in App. E.

In total, the standard GP kernel, for a spherical harmonic mode of a specific NR simulation, is determined by six parameters: $\sigma_{0,i}^{\beta}$, $\sigma_{\max,i}^{\beta}$, $\sigma_{\min,i}^{\beta}$, P_i^{β} , τ_i^{β} , and s . An i index indicates a parameter that depends on the NR simulation being studied, while a β index indicates a parameter that depends on the spherical harmonic mode. These parameters are in turn either fixed or controlled by the pooled GP parameters λ and μ which are pooled

across all modes of all simulations. The kernel also depends on several quantities that are taken directly from the highest resolution (L5) NR simulation: $M_{f,i}$, $\chi_{f,i}$, or the residual of the L5 and L4 simulations: A_i^{β} and $\sigma_{\min,i}^{\beta}$. The dependencies of all the parameters in this model for the waveform uncertainty are illustrated in the probabilistic graphical model (PGM) in Fig. 2.

The kernel is used to build the covariance matrix, $\mathbf{K}^{\beta}$, for each simulation i by evaluating the kernel at all pairs of times in the waveform time series;

$$\left(\mathbf{K}_i^{\beta}\right)_{ab} = k_i^{\beta}(t_a, t_b). \quad (19)$$

The pooled GP kernel parameters $\psi = \{\lambda, \mu\}$ are learned from the public catalog of $N_{\text{sim}} = 13$ CCE simulations by maximizing the GP log-likelihood across all the waveform modes in all of these simulations. The real and imaginary parts of the uncertainty on each mode of each simulation are assumed to be independent. The GP log-likelihood is [34]

$$\log P(\{\mathbf{r}_i^{\beta}\}|\psi) = \frac{-1}{2} \sum_{\mathbf{X}, i, \beta} \left[(\mathbf{X} \mathbf{r}_i^{\beta})^T (\mathbf{K}_i^{\beta})^{-1} (\mathbf{X} \mathbf{r}_i^{\beta}) + \log(|\mathbf{K}_i^{\beta}|) + n \log(2\pi) \right], \quad (20)$$

where $\mathbf{X} \in \{\text{Re}, \text{Im}\}$ denotes the real or imaginary part and $i = 1, \dots, N_{\text{sim}}$.

In practice, when training the GP using Eq. 20, only modes $\beta \in \{(2, \pm 2), (2, \pm 1), (3, \pm 3), (3, \pm 2), (4, \pm 4)\}$ were used. These were chosen because they are typically expected to be the loudest. Using this restricted set of modes reduces the cost of evaluating Eq. 20, thereby speeding up the training. We also deliberately exclude $m = 0$ modes which have previously been shown to behave differently (e.g. see [24]) and we suspect may need to be treated independently from the $m \neq 0$ modes when determining their numerical uncertainty contribution. For systems with certain symmetries the number of modes was further restricted: e.g. equal-mass, aligned-spin, only the even m modes were included. For details of the exact modes used, see Table I. Although the model is only trained on modes with $\ell \leq 4$, the GP model nevertheless provides a good description of the numerical uncertainties in higher-order modes (with $m \neq 0$) as well. This is demonstrated up to $\ell \leq 7$ in Fig. 14 of App. E.

The negative log-likelihood was minimized using `scipy.optimize.minimize` employing the Nelder-Mead method. (Maximizing the likelihood implicitly assumes flat priors on all of the ψ parameters in Eq. 21.) In the GP case, the model was trained on the residual time spanning $20M \leq t \leq 80M$ during which the model can efficiently tune to the periodicity and amplitude of the residual during ringdown. In the WN case, the residual was trained on the range $0 \leq t \leq 200M$ so as to also account for the early- and late-time data. The pooled GP parameters were allowed to vary in the following ranges:

$\lambda \in [0.01, 50]$ and $\mu \in [0.1, 10]$. This training yielded the following values for the GP parameters:

$$\psi_{\text{GP}} = (\lambda, \mu) = (6.92, 1.68). \quad (21)$$

We note here that the GP model for the waveform uncertainty and the resulting posteriors on the QNM parameters are not sensitive to the choice of the sampling frequency. This is demonstrated in App. F.

V. COMPARISON OF GP UNCERTAINTY MODELS

The standard GP kernel described in Sec. IV provides a well-motivated description of the NR waveform noise, with parameters that are based on physically-relevant features in the ringdown. However, many choices went into the design of this model and these are clearly far from being unique. In order to validate the model and to demonstrate both that it can not be significantly improved upon by introducing additional flexibility with more free pooling parameters and, conversely, that it is not already too flexible and is overfitting the training data, it was compared to several other kernel models. This section presents the results comparing the standard GP model described above to two alternatives: one simpler model (the white noise kernel, or WN) and one more complicated model (the ‘complicated’ kernel, or GPc).

First, the simple, white-noise kernel function is defined as

$$k_i^\beta(t_a, t_b) = \sigma_{0,i}^\beta \delta_{ab}, \quad \text{where} \quad \sigma_{0,i}^\beta = \lambda A_i^\beta, \quad (22)$$

where λ is the sole pooled parameter that, after training, takes the value

$$\psi_{\text{WN}} = (\lambda) = (0.20). \quad (23)$$

This kernel model is extremely simple with just a single pooled parameter (the PGM for this model is not shown). If Fig. 1 were replotted using this WN kernel then the width of the shaded regions would be large and constant for all times.

In the unique case of a white noise kernel, which has a diagonal covariance matrix, the inner product defined in Eq. 6 must be modified by multiplying by a factor of the time step used Δt . This accounts for scaling that occurs as the density of times used to construct the matrix is changed, and ensures the inner product properly converges to an integral in the limit as $\Delta t \rightarrow 0$.

Another reason for considering the WN kernel is that it is closely related to the least-squares fits that have been widely used previously and therefore makes a good comparison. Specifically, for fixed M_f and χ_f , the maximum *a posteriori* parameter estimator for the QNM parameters θ obtained using a likelihood defined using covariance matrices $\mathbf{K}^\beta$ built from this WN kernel, and using Prior 1, are exactly the same parameters that would be found in the usual least-squares fit.

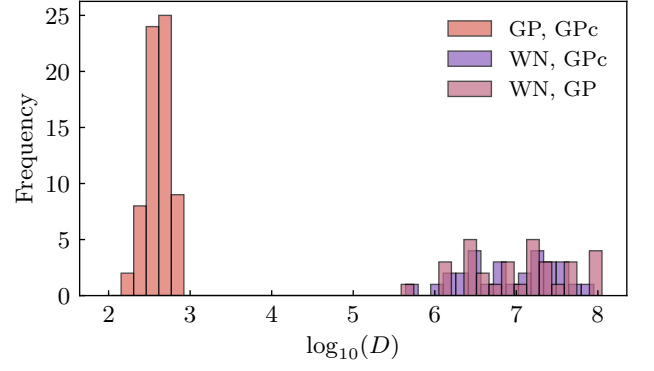


FIG. 3. The symmetrized KLD, D , (in bits) between the Gaussian distributions defined using the three kernels across the full catalog and all spherical modes used in training. The difference between the standard kernel (GP) or complicated kernel (GPc) and the white noise (WN) kernel are significant, however those between the GP kernel and the GPc are localized at smaller values, indicating that they are more similar.

The more complex kernel function (GPc), uses a flexible mixture of a stationary and periodic kernel. This introduces additional freedom into the kernel that can be used to model the oscillatory part of the residual. The time-dependent model for the amplitude is the same as in the standard GP case. The kernel is defined as

$$\begin{aligned} k_i^\beta(t, t') = & (a k_{\text{Stationary}}(t, t'; p_i^\beta) \\ & + (1-a) k_{\text{Periodic}}(t, t'; p_i^\beta, l_i^\beta)) \\ & \times \text{SMOOTHMIN}(\sigma_{0,i}^\beta e^{-t/\tau_i^\beta}, \sigma_{\text{max},i}^\beta; s) \\ & \times \text{SMOOTHMIN}(\sigma_{0,i}^\beta e^{-t'/\tau_i^\beta}, \sigma_{\text{max},i}^\beta; s). \end{aligned} \quad (24)$$

where the periodic kernel and its parameters are

$$k_{\text{Periodic}}(t, t'; p_i^\beta, l_i^\beta) = \exp \left(\frac{-2 \sin^2 \left(\frac{\pi |t-t'|}{p_i^\beta} \right)}{(l_i^\beta)^2} \right), \quad (25)$$

where

$$p_i^\beta = \frac{2\pi\eta}{\text{Re}\{\omega_{\ell m 0+}(M_{f,i}, \chi_{f,i})\}}, \quad (26)$$

$$l_i^\beta = \frac{-\kappa}{\text{Im}\{\omega_{\ell m 0+}(M_{f,i}, \chi_{f,i})\}}, \quad (27)$$

and η and κ are new pooled parameters. The final pooled parameter $0 < a < 1$ controls the relative weight given to the squared-exponential and periodic components of the kernel. Flat priors were used on all three of the new pooled parameters and they were allowed to vary in the following ranges: $\eta \in [0.1, 10]$, $\kappa \in [0.1, 10]$, and $a \in [0, 1]$.

In total, this complex GPc kernel includes three additional parameters compared to the standard GP kernel described above: a , η , and κ . The structure of this

more complicated kernel model is illustrated in the PGM shown in App. G. Training this GPc model yielded the following parameter values,

$$\psi_{\text{GPc}} = (\lambda, \mu, \eta, \kappa, a) = (8.29, 1.96, 1.00, 0.42, 0.12). \quad (28)$$

The three kernel models considered in this paper hereafter will be referred to as the ‘simple’ white-noise (WN) kernel, (Eq. 22), the ‘standard’ kernel (or GP, Eq. 11) and the ‘complicated’ kernel, (or GPc, Eq. 24). The maximum GP log-likelihood (Eq. 20) values (computed over the time range $0M \leq t \leq 200M$) were 4.31×10^6 , 7.34×10^6 and 7.35×10^6 for the white-noise (WN), standard (GP), and complex (GPc) kernels respectively. This suggests the white-noise kernel is a significantly worse description of the data, and the complex kernel offers only a small improvement over the standard kernel. If Fig. 1 were replotted using this GPc kernel then there would be no noticeable differences in the shaded regions.

The kernels can also be compared using the symmetrized Kullback–Leibler Divergence (KLD) to measure the distances between zero-mean Gaussian distributions with covariance matrices constructed from the three kernels. This is plotted in Fig. 3 where it can be seen that that the simple (WN) kernel differs significantly from the standard kernel. The standard and complex kernels are much more similar, as suggested by their similar log-likelihood values. However, the complex (GPc) kernel requires an additional three parameters to fit the data which is not penalized when comparing maximum likelihood values. We find that the standard (GP) kernel is a significant improvement over the WN kernel and is sufficiently flexible to model the waveform uncertainty without overfitting; hereafter, the standard (GP) kernel is used for all calculations unless explicitly stated otherwise.

VI. QNM INFERENCE RESULTS

As an example of the new Bayesian approach to modeling the merger and ringdown, this section shows the results of some QNM fits to an example NR waveform. It should be stressed that the purpose of the results in this section is to demonstrate the new method by reproducing known results; readers interested in the new physical insights that can be gained using the new methods are referred to the accompanying paper Ref. [29].

As an example, we consider fitting an overtone QNM model to just the $\beta = (2, 2)$ mode of the SXS:CCE:0001 simulation. The QNM overtone model includes the following QNMs: $(2, 2, n, +)$ for $n \in \{0, 1, \dots, 6\}$. The remnant BH mass and spin were also included as free parameters. This model was chosen because it is well understood, having been previously investigated by many authors using least-squares methods (see, for example, [20, 22, 46–48]) and also using QNM filters [27]. However, the methods described here provide additional in-

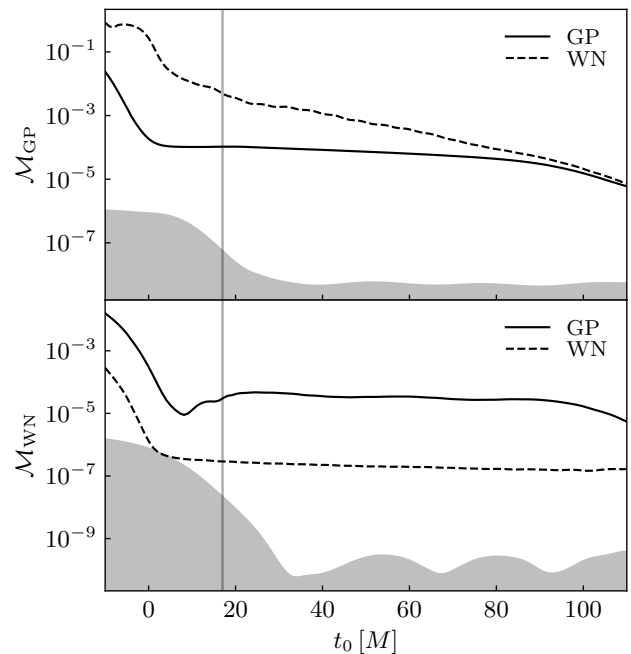


FIG. 4. The mismatch of the MAP QNM model compared to the NR data as a function of the ringdown start time for Model 1. The top panel gives the mismatch, calculated with respect to an inverse covariance matrix computed using the standard kernel function ($\mathcal{M}_{\text{GP}}$) while the lower panel shows the typical white-noise mismatch $\mathcal{M}_{\text{WN}}$. The vertical black line indicates $t_0 = 17M$ which is used as the reference start time in later plots. In both panels the gray shaded regions gives an indication of the noise floor of the simulation and show the mismatch between the NR waveforms at the two highest resolution levels.

sight by allowing the study of the full posterior distribution of the QNM model parameters. We call this overtone model ‘Model 1’. For comparison, we also include results for this overtone model plus the next leading QNM that mixes into the $(2, 2)$ spherical mode, $\alpha = (3, 2, 0, +)$. We call this ‘Model 2’. App. I also contains some results with this model.

For most of the results shown here, the standard GP model (Eq. 11) was used to model the uncertainty in the NR waveform. For comparison, selected results obtained using the white-noise model (Eq. 22) are also shown. Bayesian fits were performed for ringdown start times in the range $-10M < t_0 < 110M$. The results of these fits are illustrated in Figs. 4 to 6. On these figures a vertical line indicates a start time of $t_0 = 17M$. This start time was previously found to minimize the angle-averaged mismatch for this simulation; see Table 1 of Ref. [24]. Subsequent fits are performed at this value of t_0 . Throughout all these plots a consistent color scheme is used to distinguish all the different QNMs in the model and the different line styles are consistently used to distinguish results obtained with the different kernels.

Firstly, the overall quality of the model fit to the data

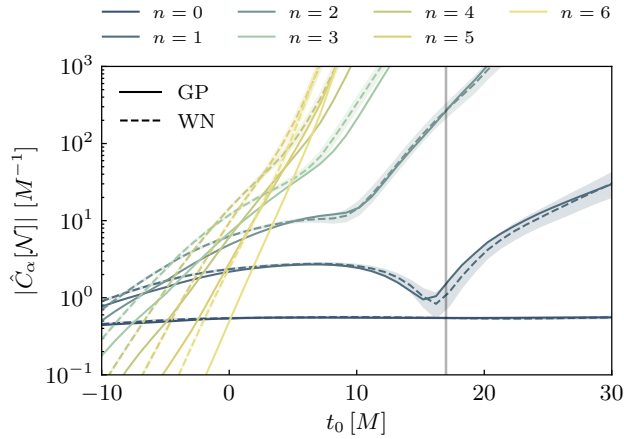


FIG. 5. The QNM decay-corrected amplitudes $|\hat{C}_\alpha|$ as a function of the ringdown start time for Model 1. Solid (dashed) lines show the results obtained using the GP (WN) covariance and the colors distinguish the QNM overtones. The median value is plotted and the shaded regions show the 50% confidence intervals for the Prior 1 posterior on the WN kernel. The interval is too small to be seen for the GP kernel, so has been omitted. The logarithmic scale used on this plot to show all the overtones gives the misleading impression that the width of the amplitude posteriors is constant at late times; this is related to the behavior of Prior 1 at small amplitudes and is discussed in the main text.

is assessed for Model 1. This is done by taking the maximum *a posteriori* (MAP) estimate for the QNM parameters and computing the mismatch between a QNM model with these parameters and the NR waveform. The mismatch between two signals a and b

$$\mathcal{M}(a, b) = 1 - \frac{|\langle a|b \rangle|}{\sqrt{\langle a|a \rangle \langle b|b \rangle}}, \quad (29)$$

can be computed using any suitable inner product $\langle \cdot | \cdot \rangle$. Here, two inner products are considered: a white-noise inner product (denoted $\mathcal{M}_{\text{WN}}$) and an inner product computed with respect to the inverse covariance matrix defined using the GP model described above (denoted $\mathcal{M}_{\text{GP}}$). These mismatches are plotted in Fig. 4.

Note, that if the quality of the fit is assessed using the simple WN mismatch $\mathcal{M}_{\text{WN}}$, then the standard GP results appear to perform worse than the simple WN results. This is inevitable because maximizing the QNM likelihood $P(\mathbf{h}|\boldsymbol{\theta})$ is, by definition, related to minimizing the mismatch. Similarly, the standard GP results appear to perform better when judging by the $\mathcal{M}_{\text{GP}}$ mismatch. Furthermore, in both cases, the mismatch is determined solely using the MAP value, rather than taking advantage of the full posterior distribution. Therefore, the waveform mismatch (with either definition) is not the right metric by which to judge the quality of fit. As an alternative, Sec. VII defines a measure of the quality of the fits based on a Bayesian posterior predictive check (PPC).

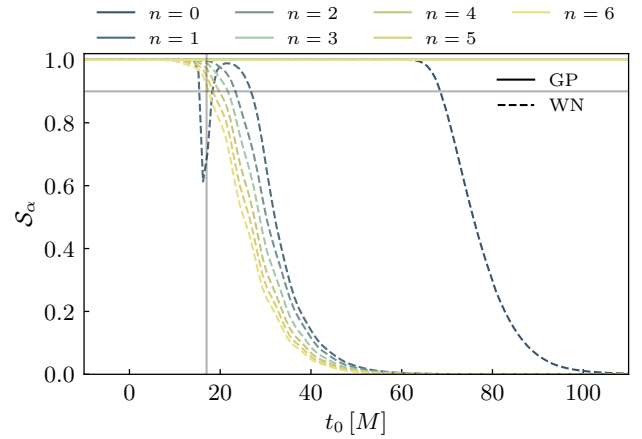


FIG. 6. The significance of each QNM in the model S_α as a function of the ringdown start time for Model 1. Solid (dashed) lines show the results obtained using the standard (white-noise) covariance and colors are used to distinguish the different QNMs. The vertical line indicates the reference start time $t_0 = 17M$ and the horizontal line indicates a typical significance threshold of $S_\alpha = 0.9$, which is approximately equivalent to the 2-dimensional complex amplitude posterior excluding the zero point at two standard deviations.

Secondly, the QNM mode amplitudes obtained in these fits are examined. Figure 5 shows the decay-corrected¹ amplitudes obtained from both the GP and WN fits. As expected, both the WN and GP versions of the fit obtain essentially identical values for the amplitude of the fundamental mode which is exceptionally stable across the entire range of start times considered. The amplitudes of the first few overtones show some stability for a smaller range of start times before starting to drift upwards; this behavior has been extensively studied in the literature and is connected to faster decay of the higher overtones which certainly means they are not expected to be detectable at late times. The high-order overtone amplitudes show no stability with varying start time.

The shaded confidence intervals for the amplitudes in Fig. 5 reveal a drawback of our chosen prior (Prior 1) which is flat on the real and imaginary parts of the QNM amplitudes. This prior assigns no prior probability to a zero value of the amplitude $|C_\alpha| = 0$. Consequently, looking at Fig. 5 would appear to suggest that all of the QNMs have amplitude posteriors that confidently exclude zero at all start times. This is definitely not the case. One solution to this problem would be to switch to Prior 2 (flat prior on $|C_\alpha|$) and this is explored below (see Fig.

¹ With our definition of QNM model (see Eq. 2) the amplitudes $|C_\alpha|$ obtained at later start times decay with t_0 . The decay-corrected amplitude $|\hat{C}_\alpha| = |C_\alpha \exp(-i\omega_\alpha t_0)|$ corrects for the expected exponential decay by referring the amplitude back to a chosen reference time which is here taken to be $t = 0$.

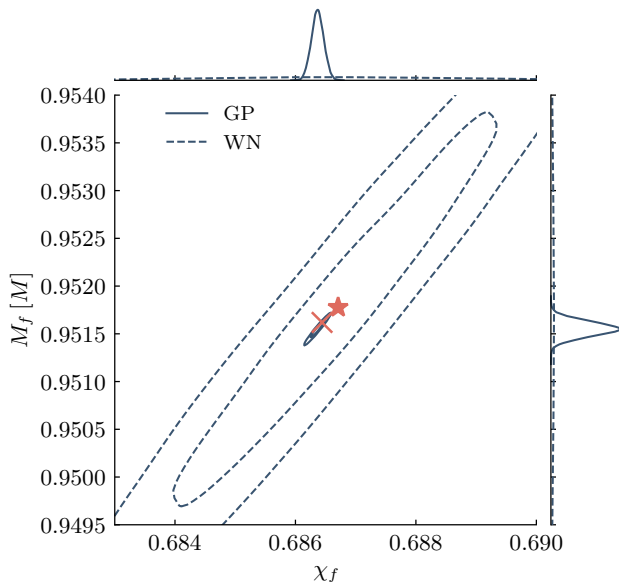


FIG. 7. A joint plot of the two-dimensional marginalized posterior on the remnant mass and spin parameters for QNM fits performed at a start time $t_0 = 17M$ using Model 2. In the main panel 50% and 90% isoprobability contours are shown both for the posterior obtained using the standard GP kernel (solid lines) and the WN kernel (dashed lines). The cross marks the remnant parameters obtained from the ABD in the NR simulation. The star marks the remnant parameters obtained from the nonlinear least-squares model fit performed prior to the Bayesian inference. Either point could be used as the reference value about which to expand the linear approximation. Throughout this paper we use the ABD value. The posterior obtained using the standard GP noise model is much more constraining than that obtained using the WN model.

8). However, first we show how to define a measure of significance for each QNM using the posteriors already obtained with Prior 1.

Given a two-dimensional marginalized posterior on the real and imaginary part of a QNM amplitude, the significance of that QNM can be defined by how confidently the posterior excludes the point $C_\alpha = 0$. This significance $\mathcal{S}_\alpha$ is defined in App. B. The significance varies in the range $0 \leq \mathcal{S}_\alpha \leq 1$ with higher values corresponding to QNMs that are confidently detected in the NR data.

Figure 6 plots the significance $\mathcal{S}_\alpha$ for each QNM in the model as a function of the start time for both GP and WN kernels. At early times, all the QNM amplitude posteriors are peaked confidently away from zero with high significance, $\mathcal{S}_\alpha \approx 1$. At late start times the significance of the QNMs drop off in a predictable order, with the higher order overtones decaying first and the fundamental mode surviving longest. The significance values defined using the GP kernel remain higher than those obtained using the WN kernel for longer; this is related to the fact the GP posteriors are narrower (see, for example, Fig. 8) and much more constraining than the WN

posteriors. At around $t_0 = 16M$, the significance of the second WN overtone fluctuates. This indicates that the amplitude posterior moves closer to zero – corresponding to a lower significance – and then further away as t_0 changes.

The behavior is investigated further in Fig. 9 and App. I, where it is demonstrated that this may relate to the omission of QNMs from the model, in particular, the absence of the $(3, 2, 0, +)$ mode. Hereafter, fits are performed using Model 2, which includes this mode.

Figure 7 shows the two-dimensional marginalized posterior on the mass and spin at a fixed value of $t_0 = 17M$. In both cases, the parameters are well-constrained, which is important because it means the model described in Sec. II is likely to be accurate. This is quantified in App. A.

Figure 8 shows example posteriors on the complex amplitudes of selected QNMs for $t_0 = 17M$. For a full set of posterior distributions for every parameter, using the strain, news, and curvature scalar see Appendix K. At this start time the fundamental mode has a high significance, $\mathcal{S}_\alpha$, and its posterior is away from zero. The second WN overtone significance has started to decrease at this start time and its amplitude is marginally consistent with zero. For all the QNM amplitudes, the posteriors obtained with the GP kernel are significantly more constraining (i.e. narrower) than those obtained with the WN kernel.

The inset plots in Fig. 8 show how the posteriors change when they are reweighted to prior 2. This has the effect of decreasing the amplitude of these modes, particularly higher overtones. This decreases the significance of the higher overtones.

In Fig. 9 we revisit the issue of the fluctuating significance values observed in Fig. 6. As the QNM amplitudes C_α decay toward zero, they typically trace out characteristic exponential spirals on the complex plane; the phase advances linearly with start time while the modulus decays exponentially. However, some deviations from this expected behavior are observed with the amplitudes moving ‘in and out’ over time. These movements are associated with the fluctuating significance values observed above. The reason for these features is not fully understood but is connected to the fact that Model 1 is incomplete and in particular does not include the $\alpha = (3, 2, 0, +)$ mode which is known to be important. To demonstrate this, App. I shows a version of Fig. 6 for Model 2, where $(3, 2, 0, +)$ has been included.

VII. POSTERIOR PREDICTIVE CHECKING

In Sec. VI it was argued that the mismatch (see Eq. 29 and Fig. 4) is not the right metric by which to judge the quality of the QNM model fit to the NR data. In a Bayesian context, the appropriate tool for assessing the quality of fit is a posterior predictive check (PPC).

PPCs involve simulating multiple realisations of the

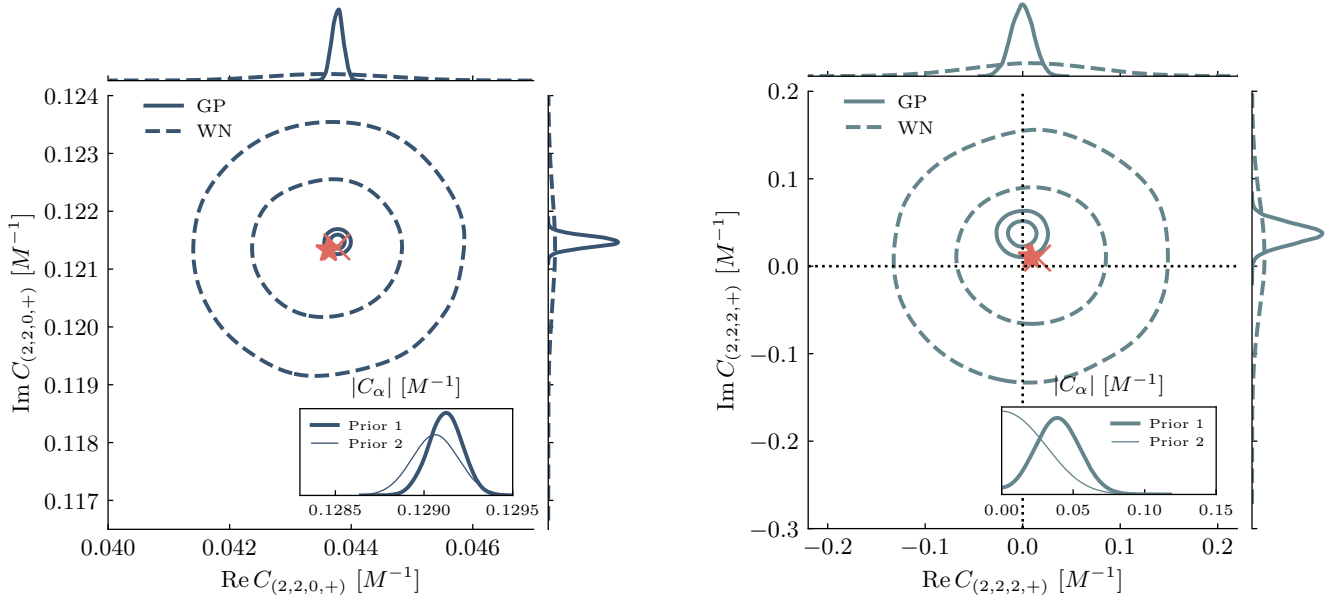


FIG. 8. A joint plot of the two-dimensional marginalized posterior on the real and imaginary parts of selected QNM amplitudes $|C_\alpha|$ for QNM fits performed at a start time $t_0 = 17M$ using Model 2. The left-hand plot shows the fundamental QNM $\alpha = (2, 2, 0, +)$ while the right-hand panel shows the $n = 2$ overtone $\alpha = (2, 2, 2, +)$. In the main panels 50% and 90% isoprobability contours are shown both for the posterior obtained using the standard GP kernel (solid lines) and the WN kernel (dashed lines). These plots were made using $N_{\text{samples}} = 10^4$. The star markers indicate the point estimates for the amplitudes obtained from the nonlinear least-squares model fit performed prior to the Bayesian inference. The cross marks the linear least-squares fit, with mass and spin fixed to their ABD values. The posterior on amplitude of the fundamental QNM (left panel) confidently excludes zero. The fundamental mode has an extremely high significance of $\mathcal{S}_{(2,2,0,+)} \approx 1$ for both the standard GP and WN kernel posteriors. However, the $n = 2$ overtone amplitude posterior does not confidently exclude the zero point indicated by the horizontal and vertical dotted lines on the right-hand plot. The overtone has significance $\mathcal{S}_{(2,2,2,+)} = 0.9758$ ($\mathcal{S}_{(2,2,2,+)} = 0.0292$) for the GP (WN) posterior. The inset plots show the posteriors on the amplitudes $|C_\alpha|$ of these modes for both the default prior (Prior 1, thick line, flat on $\text{Re } C_\alpha$ and $\text{Im } C_\alpha$) and reweighted to the alternative prior (Prior 2, thin line, flat on $|C_\alpha|$). The samples used to build the inset KDEs have been symmetrized about zero to enforce $|C_\alpha| \geq 0$. These plots were made using $N_{\text{samples}} = 1 \times 10^7$ samples to ensure a sufficiently high effective sample number. After reweighting, the effective number of samples is $n_{\text{eff}} = 1.5 \times 10^3$. The prior reweighting only has a noticeable effect on modes which already have a significance and for these modes the effect of switching to Prior 2 is to further suppress the mode amplitude.

data under the fitted model and comparing to the observed data. In the present context, this is achieved by drawing QNM model parameters from the posterior $\theta \sim P(\theta|\mathbf{h})$, computing the QNM model $h^\beta(t; \theta)$ in Eq. 4, and forming the residuals with the NR waveform data,

$$R^\beta(t_a; \theta) \equiv \mathbf{h}^\beta(t_a) - h^\beta(t_a; \theta), \quad (30)$$

where, for consistency, we continue to use strain notation. However, as before, this procedure is equally valid in the news and Weyl curvature scalar domains. We then ask: are the model-data residuals consistent with draws from the GP that models the numerical noise?

This question is answered by performing a hypothesis test using the L^2 norm of the residuals as a test statistic,

$$\xi^2 \equiv \|R\|_{L^2}^2 = \sum_\beta \sum_a \left[\text{Re}(R_a^\beta)^2 + \text{Im}(R_a^\beta)^2 \right]. \quad (31)$$

This statistic is used to assess the quality of fit; we use the symbol ξ to avoid confusion with the standard chi-squared distribution and because the symbol χ is already

used to denote the dimensionless BH spin. The null hypothesis asserts that the residuals are distributed according to the GP used to model the numerical noise,

$$R(t) \sim \mathcal{GP}(0, k). \quad (32)$$

(Hereafter the spherical harmonic β index is suppressed in our notation for clarity.) It is necessary to know the sampling distribution of ξ^2 under the null hypothesis.

The L^2 norm follows a generalized type of chi-squared distribution. In order to show this we use the eigenfunction description of GPs [34]. Mercer's theorem allows the kernel to be written as an infinite sum of orthonormal eigenfunctions,

$$k(t, t') = \sum_i \lambda_i \phi_i(t) \phi_i(t'). \quad (33)$$

Random functions $f(t) \sim \mathcal{GP}(0, k)$ can also be expressed

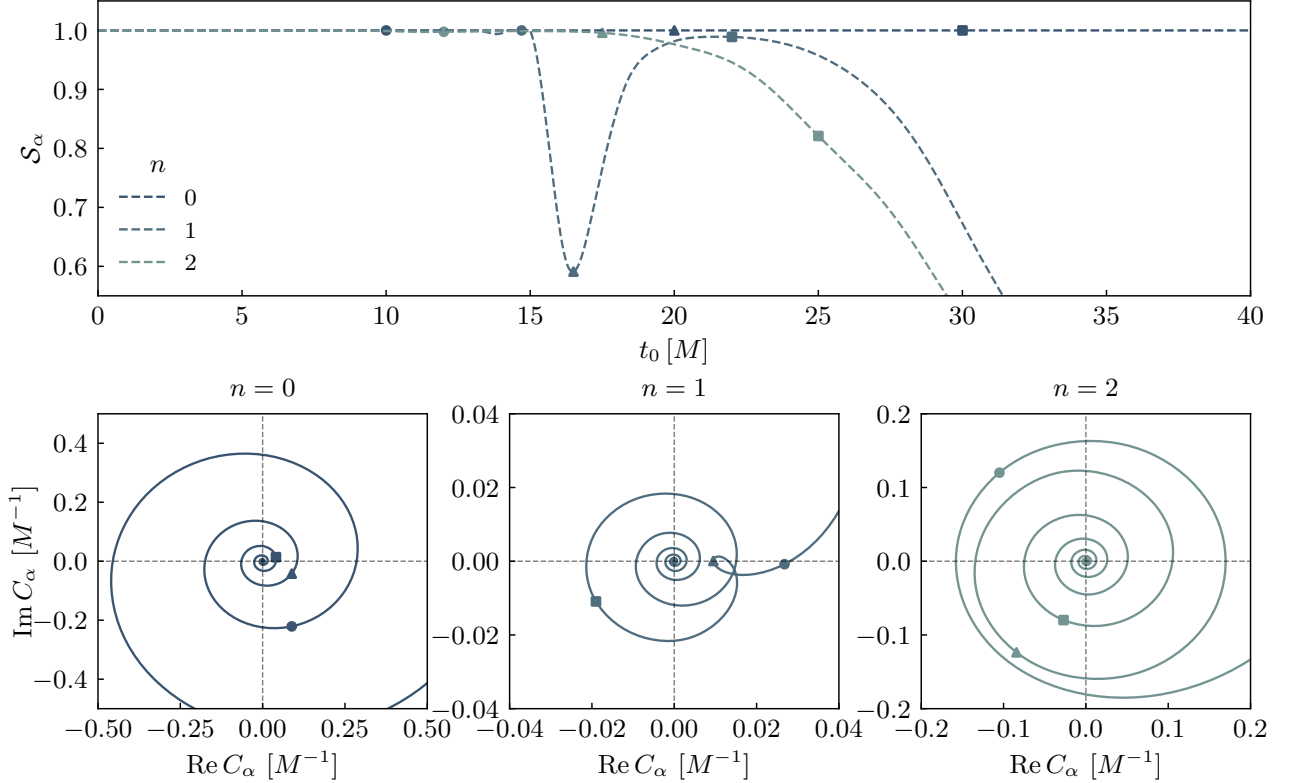


FIG. 9. The significance $\mathcal{S}_\alpha$ curves as functions of start time t_0 in Fig. 6 show oscillatory behavior at early times. This is investigated further here for the three QNMs with $n = 0, 1$ and 2 using the WN posterior. The significance curves for these modes are reproduced in the top panel. The bottom panels show the QNM amplitudes C_α in the complex plane as functions of t_0 . In both panels, the curves have been smoothed using a cubic spline interpolation. The fundamental mode shows the characteristic exponential spiral, corresponding to the predicted decay of the amplitude over the start times considered. The first overtone, which displays the instability, leaves the predicted exponential decay and continues on a new spiral. The behavior here is less regular and the curve self intersects as the amplitude $|C_\alpha|$ increases and decreases. This is the reason for the behavior seen in the top panel and can be compared to the second overtone, which displays a more regular decrease in significance. The three markers on each curve indicate specific start times around important features of the curves to enable comparisons between the different panels. Animated versions of this figure can be viewed online by clicking here: [🔗](#) [49].

in this basis using the Karhunen–Loève expansion,

$$f(t) = \sum_i \sqrt{\lambda_i} Z_i \phi_i(t), \quad (34)$$

where the coefficients are standard normal variates, $Z_i \sim \mathcal{N}(0, 1)$. Computing the norm gives

$$\|f\|_{L^2}^2 = \int_{t_0}^T dt \left(\sum_i \sqrt{\lambda_i} Z_i \phi_i(t) \right)^2 = \sum_i \lambda_i Z_i^2, \quad (35)$$

where the second step has used the orthonormality property of the kernel eigenfunctions. This shows that the squared norm follows a type of generalized chi-squared distribution with the eigenvalues λ_i playing the role of weights. (Recall, for standard chi-squared random variable with ν degrees of freedom the sum in Eq. 35 would contain only ν terms all equally weighted with $\lambda_i = 1$.)

The eigenvalues λ_i can be computed numerically from the Gram matrix $\mathbf{K}_{ab} = k(t_a, t_b)$ on the discretely sampled on the discrete grid of times t_a .

Figure 10 shows the results of a PPC for the $i = 0001$ simulation modeled using the standard GP noise kernel and Model 2, at a range of start times. The model-data residuals were calculated for 1000 sets of QNM parameters drawn from the posterior distribution. For each set of residuals the statistic ξ^2 was calculated and the distribution of these values is shown in the vertical shaded bands in the bottom panel. These observed values are then compared to the generalized chi-squared sampling distribution in Eq. 35; this is shown using KDEs in the bottom panel. Note that this sampling distribution depends somewhat on the choice of t_0 through its dependence on the eigenvalues of $\mathbf{K}$. A p -value-type quantity can then be defined by computing the cumulative distribution function (CDF) of the sampling distribution at the observed value of the statistic. Large values of the CDF indicate that the model lies in the right-hand tail of the distribution and provides a poor fit to the data. Conversely, small values of the CDF indicate that the

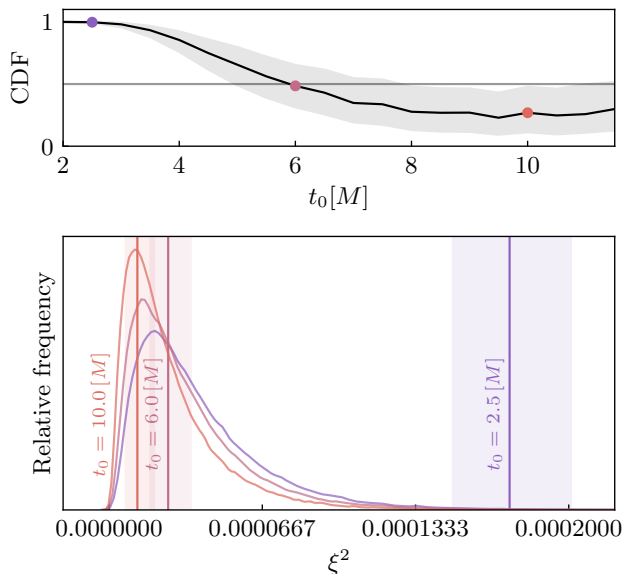


FIG. 10. *Top panel:* The proportion of the cumulative distribution function (CDF) of the reference generalized chi-squared distribution ξ^2 which falls to the left of the samples taken at a given t_0 . The shaded region gives the 50% interval. Colored points indicate the selected t_0 values plotted in the lower panel, moving left to right with increasing t_0 . *Bottom panel:* The median values and 50% widths of the residual-squared values obtained from 1000 samples of the posterior fits at values of t_0 . The reference generalized chi-squared distributions for the selected t_0 values are plotted as KDEs in the background.

model lies in the left-hand tail of the distribution and is overfitting the data. CDF values $\sim 1/2$ indicate a good fit. The results in Fig. 10 show that the combined QNM and GP noise model underfit at early times ($t_0 \lesssim 5M$) and begin to overfit at late times ($t_0 \gtrsim 8M$).

VIII. DISCUSSION AND CONCLUDING REMARKS

In this work, a Bayesian framework for fitting QNMs to CCE waveforms has been introduced, incorporating a GP model to account for numerical uncertainty in the NR data. For a simple overtone model, the method reproduces many of the familiar results reported in earlier works.

The method can be applied to the strain, news, or the curvature scalar. When applied to the latest CCE waveforms, we show that the best results are obtained from the news or curvature scalar, owing to their decay to zero at late times.

The method is an improvement over least-squares fitting because it accurately accounts for noise, and provides posterior distributions rather than point estimates. Furthermore the physically-motivated noise model de-

scribed in the paper leads to significantly tighter posteriors than those obtained under as simple WN treatment, demonstrating the value of constructing a well-motivated model and training this on a catalog of waveforms.

We have shown that our particular choice of kernel is well-suited to capturing numerical uncertainty in the CCE waveforms. While already flexible, the kernel can be extended further. For example, it could also describe the inspiral and merger portions of the NR signal, or exploit larger training sets such as extrapolated simulations in the SXS catalog.

A particular strength of the approach is in its computational efficiency. The linearized structure of the model means that, in addition to the amplitudes, the mass and spin posteriors can be obtained extremely rapidly without reliance on expensive MCMC or other stochastic sampling methods. The method therefore scales efficiently to models with arbitrary numbers of QNMs, with a computational cost comparable to least-squares fitting. Flexibility in prior choice is also achieved, since alternative priors can be accommodated by rapid sample reweighting.

This Bayesian approach will also reduce ambiguity surrounding (sometimes contentious) claims of subdominant mode detections. We have outlined a new quantity called the significance, which gives the consistency of a mode amplitude with zero. The presence of specific QNMs in the ringdown can now be determined using this statistical measure.

Finally, a Bayesian PPC has been described, allowing assessment of whether a given QNM model under- or over-fits the data. This diagnostic offers a natural means of probing the QNM content of the ringdown and building more realistic models [29].

DATA AND CODE AVAILABILITY

The code developed to perform the Bayesian fits to NR waveforms has been made publicly available at this repository [\[50\]](#). All the results and plotting scripts have additionally been made publicly available at this repository [\[49\]](#). We are grateful that the NR data used in this study have already been made publicly available [36] by the SXS collaboration and this data was processed using the methods described in the main paper.

ACKNOWLEDGMENTS

We are grateful to Eliot Finch, Maximiliano Isi, Keefe Mitman, Erin Coleman, Teagan Clarke, Nicole Khusid, and Harrison Siegel for helpful discussion about this work.

-
- [1] E. Berti, V. Cardoso, and A. O. Starinets, *Classical and Quantum Gravity* **26**, 163001 (2009).
- [2] B. Carter, *Phys. Rev. Lett.* **26**, 331 (1971).
- [3] O. Dreyer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, *Classical and Quantum Gravity* **21**, 787–803 (2004).
- [4] E. Berti, V. Cardoso, G. Carullo, and J. Abedi, preprint (2025), arXiv:2505.23895 [gr-qc].
- [5] J. Aasi, B. P. Abbott, R. Abbott, and Abbott, *Classical and Quantum Gravity* **32**, 074001 (2015).
- [6] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, and N. Allemandou, *Classical and Quantum Gravity* **32**, 024001 (2014).
- [7] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, arXiv e-prints, arXiv:2508.18082 (2025), arXiv:2508.18082 [gr-qc].
- [8] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, and C. Adams, *Living Reviews in Relativity* **23** (2020), 10.1007/s41114-020-00026-9.
- [9] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, *et al.*, arXiv e-prints, arXiv:2112.06861 (2021), arXiv:2112.06861 [gr-qc].
- [10] The LIGO Scientific Collaboration, The Virgo Collaboration, The Kagra Collaboration, and Others, arXiv e-prints, arXiv:2509.07348 (2025), arXiv:2509.07348 [gr-qc].
- [11] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, arXiv e-prints, arXiv:2509.08099 (2025), arXiv:2509.08099 [gr-qc].
- [12] 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** (2023), 10.1103/physrevlett.131.221402.
- [13] R. Cotesta, G. Carullo, E. Berti, and V. Cardoso, *Physical Review Letters* **129** (2022), 10.1103/physrevlett.129.111102.
- [14] E. Finch and C. J. Moore, *Physical Review D* **106** (2022), 10.1103/physrevd.106.043005.
- [15] A. Correia, Y.-F. Wang, J. Westerweck, and C. D. Capano, “Low evidence for ringdown overtone in gw150914 when marginalizing over time and sky location uncertainty,” (2024), arXiv:2312.14118 [gr-qc].
- [16] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, *Physical Review Letters* **123** (2019), 10.1103/physrevlett.123.111102.
- [17] M. Cabero, C. D. Capano, O. Fischer-Birnholtz, B. Krishnan, A. B. Nielsen, A. H. Nitz, and C. M. Biwer, *Physical Review D* **97** (2018), 10.1103/physrevd.97.124069.
- [18] M. Isi, W. M. Farr, M. Giesler, M. A. Scheel, and S. A. Teukolsky, *Physical Review Letters* **127** (2021), 10.1103/physrevlett.127.011103.
- [19] A. G. Abac *et al.*, *Phys. Rev. Lett.* **135**, 111403 (2025), arXiv:2509.08054 [gr-qc].
- [20] M. Giesler, M. Isi, M. A. Scheel, and S. A. Teukolsky, *Physical Review X* **9** (2019), 10.1103/physrevx.9.041060.
- [21] G. B. Cook, *Physical Review D* **102** (2020), 10.1103/physrevd.102.024027.
- [22] L. Magaña Zertuche *et al.*, arXiv e-prints, arXiv:2110.15922 (2021), arXiv:2110.15922 [gr-qc].
- [23] L. M. Zertuche, L. Gao, E. Finch, and G. B. Cook, “Multimode ringdown modelling with **qnmfits** and **KerrRingdown**,” (2025), arXiv:2502.03155 [gr-qc].
- [24] R. Dyer and C. J. Moore, *Phys. Rev. D* **111**, 024002 (2025), arXiv:2410.13935 [gr-qc].
- [25] K. Mitman, M. Lagos, L. C. Stein, S. Ma, L. Hui, Y. Chen, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, M. A. Scheel, S. A. Teukolsky, W. Throwe, and N. L. Vu, *Physical Review Letters* **130** (2023), 10.1103/physrevlett.130.081402.
- [26] M. H.-Y. Cheung, V. Baibhav, E. Berti, V. Cardoso, G. Carullo, R. Cotesta, W. Del Pozzo, F. Duque, T. Helfer, E. Shukla, and K. W. K. Wong, *Phys. Rev. Lett.* **130**, 081401 (2023), arXiv:2208.07374 [gr-qc].
- [27] S. Ma, K. Mitman, L. Sun, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, W. Throwe, N. L. Vu, and Y. Chen, *Phys. Rev. D* **106**, 084036 (2022), arXiv:2207.10870 [gr-qc].
- [28] T. A. Clarke, M. Isi, P. D. Lasky, E. Thrane, M. Boyle, N. Deppe, L. E. Kidder, K. Mitman, J. Moxon, K. C. Nelli, W. Throwe, and N. L. Vu, *Physical Review D* **109** (2024), 10.1103/physrevd.109.124030.
- [29] R. Dyer and C. J. Moore, in prep (2025).
- [30] S. Ma, M. A. Scheel, J. Moxon, K. C. Nelli, N. Deppe, L. E. Kidder, W. Throwe, and N. L. Vu, arXiv e-prints, arXiv:2412.06906 (2024), arXiv:2412.06906 [gr-qc].
- [31] N. Deppe *et al.*, “**SpECTRE v2024.09.29**,” 10.5281/zenodo.13858965 (2024).
- [32] J. Moxon, M. A. Scheel, S. A. Teukolsky, N. Deppe, N. Vu, F. Hébert, L. E. Kidder, and W. Throwe, *Phys. Rev. D* **107**, 064013 (2023).
- [33] L. C. Stein, *J. Open Source Softw.* **4**, 1683 (2019), arXiv:1908.10377 [gr-qc].
- [34] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, 2006).
- [35] A. Kong, Technical Report (A Note on Importance Sampling using Standardized Weights) **348** (1992).
- [36] The SXS project, “SXS Ext-CCE Waveform Database,” (2024), https://data.black-holes.org/waveforms/extcce_catalog.html [Accessed: August 2024].
- [37] M. Boyle *et al.*, *Class. Quant. Grav.* **36**, 195006 (2019), arXiv:1904.04831 [gr-qc].
- [38] J. Moxon, M. A. Scheel, and S. A. Teukolsky, *Physical Review D* **102** (2020), 10.1103/physrevd.102.044052.
- [39] M. Boyle, D. Iozzo, and L. C. Stein, “**moble/scri: v1.2**,” (2020).
- [40] M. Boyle, *Phys. Rev. D* **87**, 104006 (2013).
- [41] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, “Gravitational-wave modes from precessing black-hole binaries,” (2014), arXiv:1409.4431 [gr-qc].
- [42] M. Boyle, *Phys. Rev. D* **93**, 084031 (2016).
- [43] K. Mitman, M. Boyle, L. C. Stein, N. Deppe, L. E. Kidder, J. Moxon, H. P. Pfeiffer, M. A. Scheel, S. A. Teukolsky, W. Throwe, and N. L. Vu, *Classical and Quantum Gravity* **41**, 223001 (2024), arXiv:2405.08868 [gr-qc].
- [44] K. Mitman *et al.*, *Phys. Rev. D* **106**, 084029 (2022), arXiv:2208.04356 [gr-qc].
- [45] K. Mitman, I. Pretto, H. Siegel, M. A. Scheel, S. A. Teukolsky, M. Boyle, N. Deppe, L. E. Kidder, J. Moxon, K. C. Nelli, W. Throwe, and N. L. Vu, arXiv e-prints, arXiv:2503.09678 (2025), arXiv:2503.09678 [gr-qc].
- [46] E. Finch and C. J. Moore, *Phys. Rev. D* **103**, 084048 (2021), arXiv:2102.07794 [gr-qc].

- [47] A. Dhani, Phys. Rev. D **103**, 104048 (2021), arXiv:2010.08602 [gr-qc].
- [48] X. J. Forteza and P. Mourier, Phys. Rev. D **104**, 124072 (2021), arXiv:2107.11829 [gr-qc].
- [49] R. Dyer and C. J. Moore, “Bayesian QNM Fits Methods Figures,” (2024), https://github.com/Richardvnd/bgp_methods [Accessed: August 2024].
- [50] R. Dyer and C. J. Moore, “Bayesian QNM Fits Code,” (2024), https://github.com/Richardvnd/bgp_qnm_fits [Accessed: August 2024].
- [51] H. Zhu, F. Pretorius, S. Ma, R. Owen, Y. Chen, N. Deppe, L. E. Kidder, M. Okounkova, H. P. Pfeiffer, M. A. Scheel, and L. C. Stein, “Imprints of changing mass and spin on black hole ringdown,” (2024), arXiv:2404.12424 [gr-qc].
- [52] N. Lu, S. Ma, O. J. Piccinni, L. Sun, and E. Finch, “Statistical identification of ringdown modes with rational filters,” (2025), arXiv:2505.18560 [gr-qc].
- [53] Keefe Mitman, (2025), private communication.

Appendix A: Accuracy of the linearized approximation in χ_f and M_f

The linearization in the remnant mass and spin parameters, M_f and χ_f , used in the QNM model (see Eq. 7) necessarily introduces some error. This error is expected to be small because the remnant parameters are typically extremely well constrained in a QNM fit of an NR waveform; see, for example, Fig. 7. The accuracy of this linear approximation is quantified by drawing samples M_f, χ_f from the two-dimensional posterior on the remnant parameters. For each of these posteriors samples any of the QNM frequencies ω_α or mode mixing coefficients μ_α^β can be calculated in two ways: (i) using the full nonlinear expressions for the Kerr metric (calculated using the `qnm` package [33]), or (ii) using the linearized expressions expanded about the reference parameters θ_* (see, e.g., Eq. 7). The fractional error between these two quantities can then be calculated and used to quantify the accuracy of the linear approximation. This was done for three quantities: the real and imaginary parts of the fundamental QNM frequency $\omega_{(2,2,0,+)}$ and the modulus of the most significant mode mixing coefficient $\mu_{(2,2,0,+)}^{(2,2)}$. The fractional error for a quantity Y is defined as

$$\frac{|Y_i(\text{linear approx.}) - Y_i(\text{nonlinear})|}{Y_i(\text{linear approx.})}, \quad (\text{A1})$$

and is computed for 10^4 posterior samples (indexed by i) drawn from the distributions shown in Fig. 7. This was done using both the GP and WN posterior distributions; the WN posterior covers a larger region of the mass-spin plane and therefore the typical errors in the linear approximation are larger in this case. This can be seen in the results in Fig. 11 where kernel density estimators (KDEs) are used to show the distributions of these fractional errors. The figure shows that when expanding about the ABD remnant parameters the fractional error in the linear approximation to the QNM model is

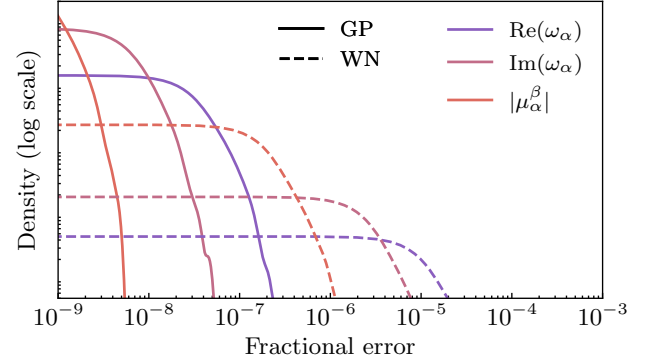


FIG. 11. KDEs for the fractional errors in selected quantities involved in the linearized approximation to the QNM model. The three quantities considered are the real and imaginary parts of ω_α with $\alpha = (2, 2, 0, +)$ and the modulus $|\mu_\alpha^\beta|$ of the primary mode mixing coefficient with $\beta = (2, 2)$. The solid (dashed) lines show the results for the posteriors obtained using the GP and WN kernels respectively.

less than around 1 part in $\sim 10^7$ across the typical set of the GP posterior. The approximation can be further improved by expanding about the nonlinear mass and spin parameters, rather than the ABD values, however such an improvement has a negligible effect on the overall fits.

Appendix B: QNM significance

A significant drawback of the least-squares approach to fitting QNMs is that it only provides a point estimate of the mode amplitudes and it is not possible to tell from this alone whether or not a particular QNM belongs in the model. There are several potential reasons why a QNM might not belong in the model. As QNMs decay, their amplitudes at late start times become comparable to numerical noise and they can no longer be confidently detected. Conversely, some modes (e.g. high-order overtones) have large amplitudes at early start times that destructively interfere with other modes, leaving their physical significance unclear. The problem of determining the QNM content has been addressed in a variety of *ad hoc* ways. For example, fits can be performed for a range of start times to check that the recovered amplitudes are stable [20, 51], the real and imaginary parts of a particular QNM frequency can be allowed to vary freely and be compared with predicted QNM frequencies [20, 25, 26], or loud QNMs can be filtered out to identify subdominant modes [27, 52].

A key advantage of the Bayesian approach is that it gives a posterior on the QNM amplitudes, not just a point estimate. The posterior allows us to assess if zero amplitude (i.e. no QNM) is consistent with the data. Furthermore, this can be done from a single fit, without the need to vary the start time to check for amplitude stability or introduce additional parameters as in a free frequency fit.

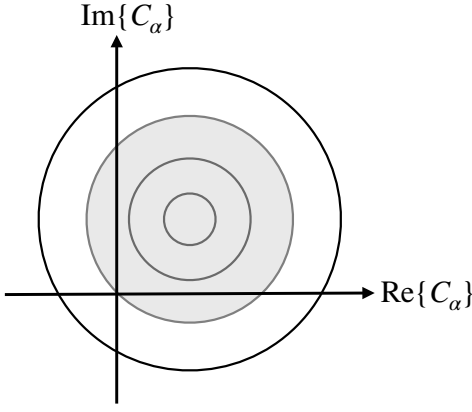


FIG. 12. Diagram illustrating the definition of the significance $\mathcal{S}_\alpha$ in Eq. B1. Lines are iso-probability contours of the 2D marginalized posterior on the real and imaginary parts of the QNM amplitude. The integral is over the shaded region.

For a particular QNM, indexed by α , we introduce the following measure for its significance ($0 \leq \mathcal{S}_\alpha \leq 1$);

$$\mathcal{S}_\alpha = 1 - \exp\left(-\frac{1}{2}|b_\alpha|^2\right), \quad (\text{B1})$$

where μ_α and Γ_α^{-1} are the mean vector and covariance matrix for the 2-dimensional marginalized posterior on the real and imaginary parts of the QNM amplitude, L_α is the Cholesky decomposition of Γ_α^{-1} (i.e. $\Gamma_\alpha^{-1} = L_\alpha \cdot L_\alpha^T$ with L_α lower triangular), and $b_\alpha = L_\alpha^{-1} \cdot \mu_\alpha$. High $\mathcal{S}_\alpha$ indicate a preference for the inclusion of the α QNM; conversely, low $\mathcal{S}_\alpha$ indicates a preference for omitting the mode. The definition of significance in Eq. B1 can be motivated in either a Bayesian or a frequentist way.

In a Bayesian spirit, compute the posterior odds ratio $\mathcal{O}_\alpha$ for a pair of models, with and without QNM α , but otherwise identical. A uniform prior is used on the real and imaginary parts of the QNM amplitude in the range $(-A_{\max}, A_{\max})$. As the models are nested, the evidence ratio can be computed using the Savage-Dickey density ratio. Assuming equal prior odds for the two models, the posterior odds ratio in the limit of large $A_{\max}$ is given by

$$\mathcal{O}_\alpha = \frac{\exp(-\frac{1}{2}|b_\alpha|^2)}{8\pi|L_\alpha|A_{\max}^2}, \quad (\text{B2})$$

As usual in Bayesian inference, the result depends on our prior; in this case, on $A_{\max}$. However, the power-law dependence of the result on $A_{\max}$ in the denominator is usually dominated by the exponential dependence on the data through b_α . Keeping only the exponential terms we arrive at the approximate result $\mathcal{O}_\alpha \propto 1 - \mathcal{S}_\alpha$.

In a frequentist spirit, the significance of the QNM α is defined using the number of standard deviations away from zero the (two-dimensional) posterior on the (complex) mode amplitude is peaked. This significance is defined as the fraction of the posterior mass associated with higher posterior density than the point of zero amplitude;

i.e. the integral of the posterior inside the iso-probability contour that passes through the zero amplitude point $C_\alpha = 0$ (see Fig. 12),

$$\mathcal{S}_\alpha = \int_{\{C_\alpha | P(C_\alpha|d) > P(0|d)\}} d(\text{Re } C_\alpha) d(\text{Im } C_\alpha) P(C_\alpha|d), \quad (\text{B3})$$

where $P(C_\alpha|d)$ is the 2D marginalized posterior on the complex amplitude C_α . Integrating the two-dimensional Gaussian posterior on the QNM amplitude with mean μ_α and covariance Γ_α^{-1} gives the result in Eq. B1.

For loud QNMs with $\mathcal{S}_\alpha \rightarrow 1$, Eq. B1 becomes numerically unstable. In this limit it is convenient to use

$$\log \mathcal{S}_\alpha \approx -\exp\left(-\frac{1}{2}|b_\alpha|^2\right). \quad (\text{B4})$$

Appendix C: NR waveform preprocessing

The results in this paper use the 13 CCE waveforms provided in the SXS catalog. For these simulations, preprocessing described in the main text is carried out. For each simulation, the highest resolution level (L5), extracted at the second smallest world-tube radius, is used as the preferred waveform. The same simulation, extracted at the second highest level (L4) is used as the second ‘best’ simulation for computing the residual. This was the simulation whose mismatch with the preferred waveform was minimum. Therefore it most closely resembles the preferred waveform up to differences due to numerical uncertainty.

The subsequent processing is also applicable to the extrapolated simulations in the main SXS catalog. To compute the residual, we first perform a time and a phase shift to align the two waveforms. The data is up-sampled using a cubic spline implemented in `scipy.interpolate.make_interp_spline`. The choice of interpolation timestep is taken from [24], chosen so that the change in mismatch due to rolling the wavefunction by a timestep has converged to a negligible value. The shift can then be computed by maximizing the overlap between the two waveforms. This was done efficiently using fast Fourier transform techniques.

To compute the residual, the time-shift corrected waveforms are re-interpolated onto the same (coarser) time grid and subtracted from each other;

$$\mathbf{r}_i^\beta = \mathbf{h}_i^\beta(\text{L5}) - \mathbf{h}_i^\beta(\text{L4}). \quad (\text{C1})$$

These residuals are used to train the Gaussian process model for the waveform uncertainties (Sec. IV).

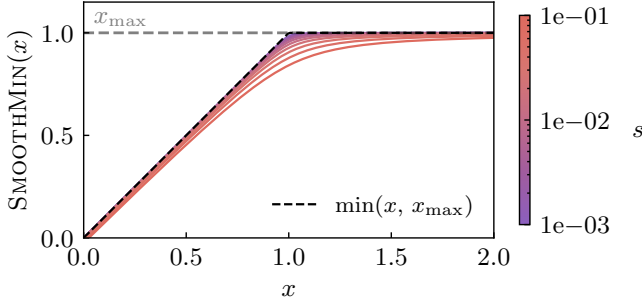


FIG. 13. The $\text{SMOOTHMIN}(x, x_{\max}; s)$ function (Eq. D1) with $x_{\max} = 1$ and $10^{-3} \leq s \leq 10^{-1}$. As $s \rightarrow 0$ it approaches the black dashed line with discontinuous first derivative.

Appendix D: The SmoothMin function

The SMOOTHMIN function used in the definition of the GP kernel functions in Eqs. 11 and 24 is defined as

$$\text{SMOOTHMIN}(x, x_{\max}; s) = \frac{1}{2} \left[x + x_{\max} \left(1 - \sqrt{\left(\frac{x}{x_{\max}} - 1 \right)^2 + s} \right) \right]. \quad (\text{D1})$$

This is a smooth version of the C^0 function $\min(x; x_{\max})$ with a parameter $s > 0$ that controls the smoothness. With this definition, in particular with $x_{\max}$ factored outside the parentheses containing the square root, the effect of the smoothing parameter is independent of the scale $x_{\max}$. The function is plotted in Fig. 13.

Appendix E: Regularization of the GP

Figure 14 shows the role of the jitter term on controlling the late-time form of the GP kernel. In particular, the figure demonstrates how the value of ϵ sets the height on the late-time uncertainty. A higher value of ϵ would increase the height of the flat region (which for $\epsilon = 10^2$ starts at $t \sim 120M$ in the (2,2) mode) this would consequently reduce the time at which this part of the model dominates. On the other hand, a smaller value of ϵ pushes the end of the exponential decay segment to later times. However numerical instabilities start to arise as the value becomes too small. A value of $\epsilon = 10^2$ was chosen to balance conservatively modeling the late-time uncertainty and ensuring numerical stability, while accurately describing the residual in this regime.

Appendix F: Sampling frequency checks

To ensure that the sampling densities used in training the GP model, and used to perform the fits, are not sensitive to the timestep Δt , we perform two stability tests. The first test shown in Fig. 15 gives the values of the

pooled parameters across a range of sampling time steps, demonstrating good stability for $\Delta t < 1$. The small amount of drift in these values we attribute to small numerical instabilities as the matrices become significantly larger.

In Fig. 16 we fix the pooled parameters to their values at $\Delta t = 0.1$, and perform fits on data sampled at a range of timesteps. Both the WN and GP kernels display relatively good stability across a range of values of Δt . In the GP case, there is some drift, which can be primarily attributed to the jitter term. As Δt decreases, the jitter has more of an impact on the eigenvalues of the covariance matrix. The mixing of a WN-like kernel with the GP kernel means the inner product used to compute the fisher matrix does not perfectly converge, resulting in a drift. As the jitter is necessary for numerical stability, this is unavoidable. Smaller values of ϵ will reduce this effect, but result in greater numerical instability.

While both tests vary the sampling frequency of the data, note that the timestep Δt used in training need not be the same as that used for fitting.

Appendix G: The ‘complicated’ GP kernel

The main GP model (Eq. 11) for the waveform uncertainty was illustrated using the PGM in Fig. 2 in the main body of the paper. Here the PGM for the complicated GP model (Eq. 24) is shown in Fig. 17.

Appendix H: The stationary kernel

The stationary kernel introduced in the main text in Sec. IV is defined to be the one-dimensional, $q = 3$ Wendland covariance function [34];

$$k_{\text{Stationary}}(t, t'; P_i^\beta) = k_{\text{Wendland}} \left(\frac{|t - t'|}{P_i^\beta} \right), \quad (\text{H1})$$

where

$$k_{\text{Wendland}}(\tau) = \frac{1}{15} (1 - \tau)_+^{j+3} \times \left((j^3 + 9j^2 + 23j + 15)\tau^3 + (6j^2 + 36j + 45)\tau^2 + (15j + 45)\tau + 15 \right), \quad (\text{H2})$$

and where $j = q + 1$ and $(x)_+ \equiv \max(0, x)$.

Appendix I: Including the (3,2,0,+) QNM

To highlight the limitations of the overtone model (Model 1) discussed in this paper, we reproduce Fig. 6 from the main text, but with the loudest $(\ell, m) \neq (2, 2)$

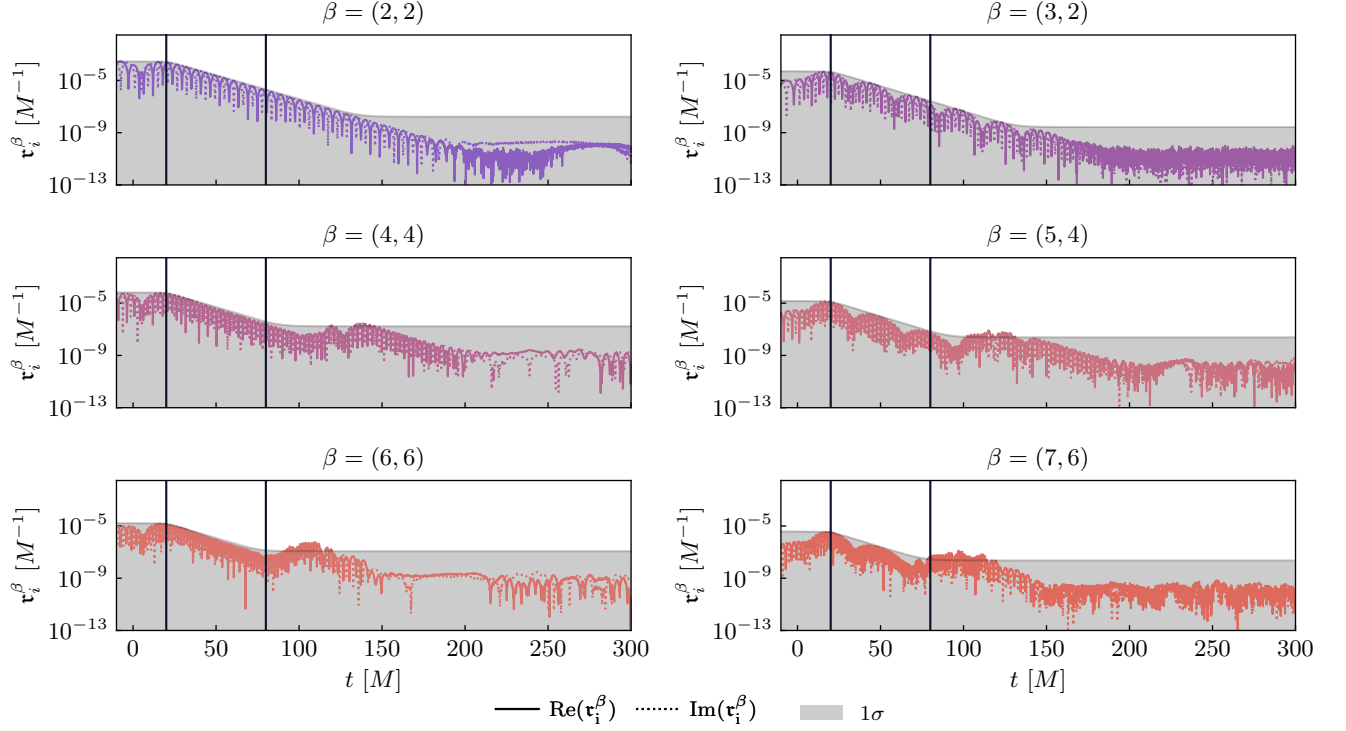


FIG. 14. A version of Fig. 1 from the main text plotted using a logarithmic y -axis scale, with a plot range extended to later times and with more spherical modes included. This highlights the role played by the regularizing ϵ parameter. The bump around $100 - 130 M$ that is visible in the $\ell \geq 4$ spherical modes is likely related to a re-gridding that occurs in the adaptive mesh refinement (AMR) at this time [53].

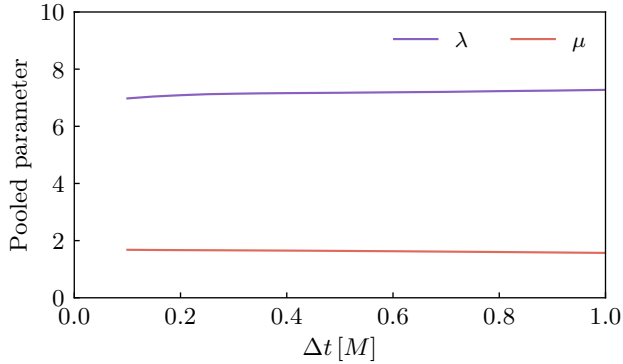


FIG. 15. The pooled parameter values as a function of timestep Δt . These values demonstrate good stability.

mode included. This the fundamental prograde QNM with $(\ell, m) = (3, 2)$ which appears in the $\beta = (2, 2)$ spherical harmonic strain as a result of mode mixing. This model is referred to as Model 2 in the main text.

Figure 18 displays the same characteristics as in Model 1, with overtones dropping in significance more rapidly with increasing overtone number. However, the spike in the significance at early start times has disappeared. We

also note that the $\alpha = (3, 2, 0, +)$ QNM is the second longest lived mode in the fit, and dominates over the overtones for a substantial range of start times. However, it is unlikely that the $(3, 2, 0, 1)$ mode is as significant in $(2, 2)$ as this figure alone would suggest. In practice, a multi-mode fit including the $(3, 2)$ mode would be required to give a more accurate representation of the mode's contribution to the ringdown.

Appendix J: Comparing residuals from the strain, news and curvature scalar

In this paper, we chose to work with the news rather than the strain, which is typically what is used in ringdown studies. This is due to the imperfections in the superrest frame transformation, which only appear for the strain, and result in a much larger late-time residual. This can be seen in Fig. 19. The imperfection inflates the value of $\sigma_{\min, i}^\beta$ compared to the news and curvature scalar which results in a greater estimate of the uncertainty and hence a much less well-constrained posterior.

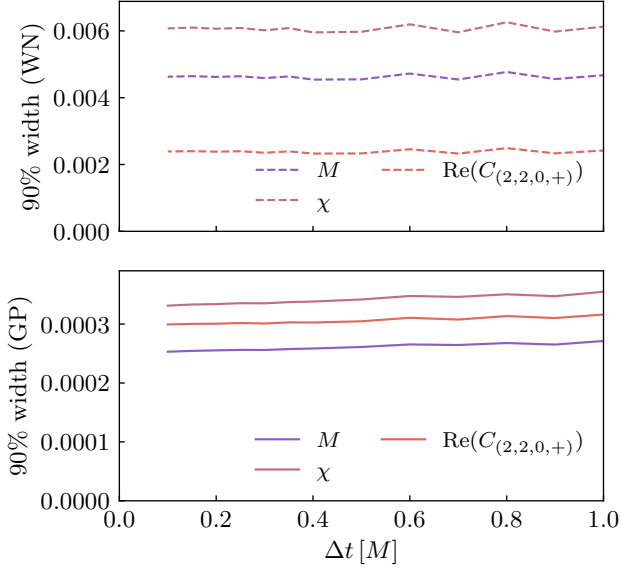


FIG. 16. Three model parameters as a function of timestep (with fixed kernel parameters). The WN parameters are stable across a range of Δt . The GP parameters are also relatively stable, with some deviation attributed to the necessary addition of jitter.

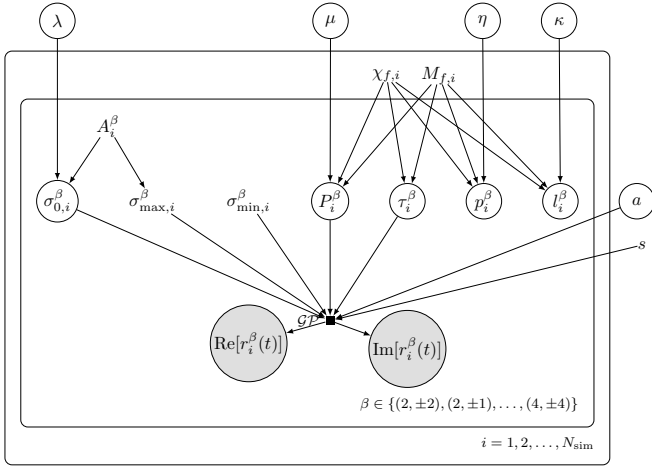


FIG. 17. PGM of the complicated GP model in Eq. 24 for the waveform uncertainty. This should be compared with the PGM for the standard GP kernel model in Eq. 11 shown in Fig. 2 of the main paper.

Appendix K: Comparing posteriors from fitting the strain, news and curvature scalar

The methods discussed in this paper are equally applicable to the strain, news, and curvature scalar. This is demonstrated in Fig. 20 where the full posterior distributions for each parameter in Model 2 at $t_0 = 17M$ are given for each of these quantities.

In order to compare the fits, the news and curvature

scalar amplitudes are converted to the strain using Eq. 3.

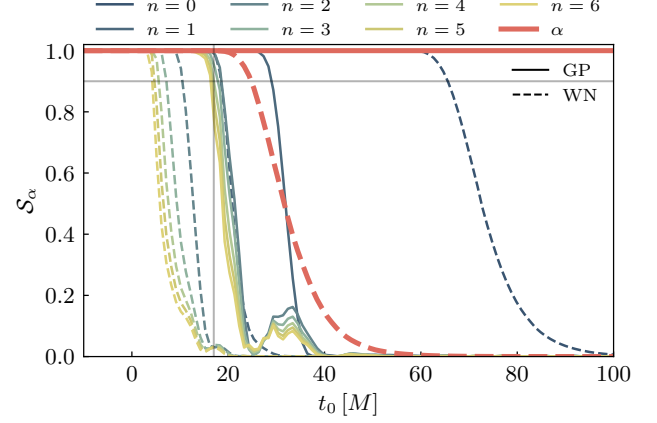


FIG. 18. A version of Fig. 6 for the QNM model including the $\alpha = (3, 2, 0, +)$ mode (i.e. Model 2). The significance of each QNM in the model S_α as a function of the ringdown start time. Solid (dashed) lines show the results obtained using the GP (white-noise) covariance and different colors are used to distinguish the different QNMs.

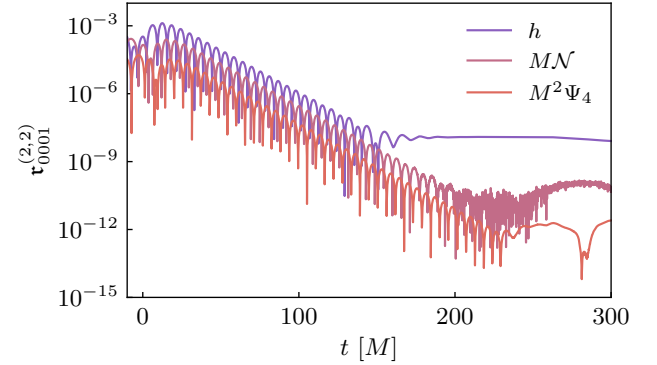


FIG. 19. A comparison of the residuals for the $\beta = (2, 2)$ mode in the $i = 0001$ simulation for the strain, news, and curvature scalar. The late-time residual of the strain is clearly much larger than the news and curvature scalar, resulting from imperfections in the superrest frame mapping.

The linear and non-linear least squares fits to the strain are also shown as vertical lines. In principle, these fits should yield consistent results and this is checked here.

In general, the MAP values are in broad agreement across the three quantities. However, the posterior widths are not consistent. This can be attributed to differences in the GP noise model and, in particular, the late-time jitter. For the strain, which has a larger late-time uncertainty, this jitter is significantly higher, which leads to the wider posterior (see App. J). The differences in the news and curvature scalar can also be attributed to differences in this part of the kernel, however they are in much better agreement. It is clear that the news noise model is much more constraining than the strain, which motivates using it for this work.

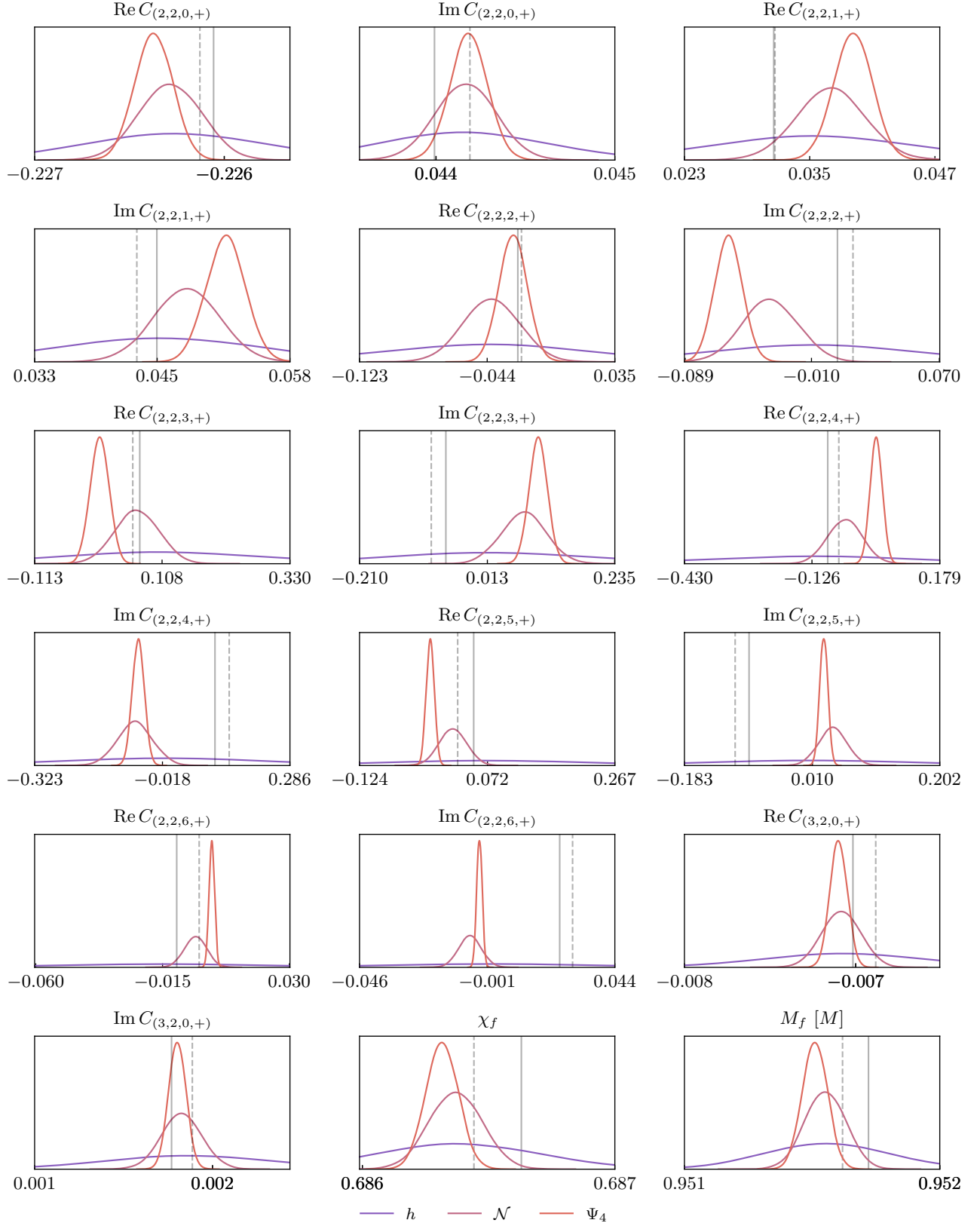


FIG. 20. The posterior distributions for every model parameter in Model 2, on the strain, news, and curvature scalar. The amplitudes have been converted into their strain-domain values. The bold vertical line show results obtained from a non-linear least-squares fit to the strain. The dashed lines show the reference ABD values of the mass and spin and, for the amplitudes, the linear least-squares fits for the strain.