Technical Note	LIGO-T2400223–v1	2024/10/22

Investigating methods of fitting quasinormal-modes in numerical-relativity ringdown signals

Erin Coleman

Distribution of this document:

AIC, ISC

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW22-295
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

http://www.ligo.caltech.edu/

1 Introduction

Gravitational waves (GWs) are perturbations in spacetime caused by compact objects moving at relativistic speeds. They are now routinely observed with interferometry by GW observatories around the world [1]. One major source of GWs is binary black-hole (BH) mergers. The final stage of a BH merger is known as the ringdown.

The ringdown occurs because a BH merger results in a single perturbed BH. This perturbed BH radiates gravitational energy to settle into a more stable state. The ringdown can be decomposed into quasinormal modes (QNMs), labelled with indices (ℓ, m, n). Here, ℓ and m are the usual angular indices, and n is an additional "overtone" number (which will be of particular interest in this project). Each QNM has angular dependence given by spin-weighted spheroidal harmonics and a complex frequency $\omega_{\ell m n} = 2\pi f_{\ell m n} - i/\tau_{\ell m n}$. These modes are called QNMs because, unlike normal modes, they decay over time. The real component of $\omega_{\ell m n}$ indicates the oscillatory frequency of that particular mode, and the imaginary component indicates how quickly it decays [2].

Within general relativity (GR) GWs have two possible polarizations, denoted by h_+ and $h_\times$. It is convenient to describe the overall strain in both polarizations as a complex number $h = h_+ - ih_\times$. Setting (without loss of generality) the ringdown start time $t_0 = 0$, the overall ringdown waveform at time $t > 0$ is given by

$$h(t, \theta, \varphi) = h_+(t, \theta, \varphi) - ih_\times(t, \theta, \varphi) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{n=0}^{\infty} {}_{-2}S_{\ell m n}(\theta, \varphi) C_{\ell m n} e^{-i\omega_{\ell m n} t} \qquad (1)$$

where ${}_{-2}S_{\ell m n}$ are the spheroidal harmonics and $C_{\ell m n} = A_{\ell m n} e^{i\phi_{\ell m n}}$ are complex amplitudes. Eq. 1 is valid in the rest-frame of the remnant BH, with its spin vector aligned with the positive z-axis (the angles θ and φ represent the polar and azimuthal angles in this frame, as in the typical formulation of spherical coordinates).

In this project we will be interested in applying Eq. 1 to numerical relativity (NR) simulations of merging BHs, with the aim of measuring the ringdown QNM content. However, in NR h is written as a sum of modes

$$h(t, \theta, \varphi) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} h_{\ell m}(t) {}_{-2}Y_{\ell m}(\theta, \varphi) \qquad (2)$$

over the spin-weighted spherical harmonics ${}_{-2}Y_{\ell m}$ instead of the spheroidal harmonics ${}_{-2}S_{\ell m n}$. In this basis, every spherical-harmonic mode has contributions from every QNM of the same m. This is an effect known as mode mixing [3]. In addition to this, the frame used by NR doesn't necessarily match the ringdown frame required by Eq. 1; typically, NR simulations place the initial BHs of the binary along the x axis, orbiting one another in the $x - y$ plane, and the z-axis points in the direction of the initial orbital angular momentum vector. The remnant BH isn't necessarily at rest in this frame, and its spin vector could in-principle point in any direction (particularly if the initial BH spins are not aligned with the orbital angular momentum, which causes an effect known as precession). To mitigate these issues with the remnant frame, we have worked with the aligned-spin CCE waveforms available in the SXS catalog [4, 5]. These were mapped to the remnant ("superrest") frame via the `scri` [6] Python package.

If a QNM can be recovered from a real GW signal, the resultant final BH mass M_f and dimensionless spin χ_f can be calculated. The No-Hair Theorem, which is a significant theorem in GR, states that these properties are the only properties of an astrophysical BH [7]. Thus, making measurements of multiple QNMs in a GW ringdown signal is an important test of GR; we can measure the frequencies of all QNMs received, and then determine whether they are consistent with a single value of χ_f and M_f. This technique is known as BH spectroscopy [8, 9].

We can study the ringdown of NR waveforms to inform our analysis of real GW data. A 2019 study by Giesler et al. [10] focused on fitting overtones of the fundamental $\ell = m = 2$, $n = 0$ QNM to the $\ell = m = 2$ spherical-harmonic mode h_{22}. Surprisingly, it was found that overtones up to $n = 7$ could be fit to the waveform, stretching back to the time of peak strain [10]. However, the same study found that the overtone amplitudes did not all decay as expected when fitting the ringdown at different times. This suggests that care should be taken when performing fits of overtone amplitudes to NR. This was the goal of a more recent study by Clarke et al. [11], who fitted QNM overtone amplitudes using a Bayesian analysis in an attempt to account for uncertainty in the content of the NR waveform. This approach also allows one to classify tone amplitudes as stably recovered, unstably recovered, or unresolved depending on the consistency of the amplitude and phase of the recovered tone with fits performed at different times in the waveform.

In this report, we employ two methods to fit overtone amplitudes and phases: least-squares fitting and Bayesian inference. Let $f(x, \vec{\theta})$ be a model of a system, where $\vec{\theta}$ represents the model parameters. Least-squares fits minimize the sum of the squares of the residuals for each data point. For a set of data points $d = (x_0, y_0), (x_1, y_1), ...(x_k, y_k)$ and a model $f(x, \vec{\theta})$, the least-squares fit finds the optimal set of parameters $\vec{\theta}$ by minimizing the sum

$$\sum_{i=0}^{k} (f(x_i, \vec{\theta}) - y_i)^2.$$

Throughout this work, we use the `qnmfits` [12, 13] Python package to perform least-squares fits.

Instead of giving one best-fit value, Bayesian inference gives a probability (posterior) distribution on the model parameters. This approach has numerous benefits, including accounting for uncertainty or noise in the data and providing information on correlations between model parameters (not just the maximum likelihood parameter value). In Bayesian inference we rely on application of Bayes' Theorem, which is as follows:

$$P(\vec{\theta}|d) = \frac{P(d|\vec{\theta})P(\vec{\theta})}{P(d)}.$$

The probability of the value of the parameters $\vec{\theta}$ given the data d (the posterior distribution) is equal to the likelihood of observing d given $\vec{\theta}$ times the prior distribution of $\vec{\theta}$, divided by the evidence $P(d)$. Since the evidence is independent of $\vec{\theta}$, it acts as a constant normalization which we can disregard. Bayesian inference relies on generating samples of the posterior distribution. Samples are taken using a Markov Chain Monte Carlo (MCMC) sampler, which starts from a prior distribution of parameter values and makes random proposals.

The probability distribution that each sample drawn from is dependent on the previous sample in the Markov Chain, and the probability that a proposal is accepted is related to the likelihood function. The chain converges to a set of samples of the true posterior distribution, which can then be used for analysis.

In section 2 we perform least-squares fits to determine the ringdown start time using two different metrics – the mismatch and accuracy of remnant mass-spin recovery. In section 3 we measure the impact of perturbing frequencies of each overtone from the GR value (e.g. the potential utility of each overtone for tests of GR). Section 4 details the issues with least-square fits for measuring amplitudes and introduces Bayesian inference to determine amplitude stability and resolvability.

2 Ringdown start time beyond $N = 7$

In the first part of this project, we studied least-squares fits of the ringdown of binary BH mergers. It has been demonstrated in Giesler et al. that a sum of overtones up to $n = 7$ are highly effective at modeling the ringdown as far back as the peak of the strain. Here we are interested in going beyond $n = 7$ to see how far this description can be pushed. We are also interested in what exactly is gained with the inclusion of each additional overtone (that is, we are interested in the "usefulness" of each overtone). Fitting additional overtones allows an earlier fit to the signal, which is desirable because the amplitudes of QNMs decay exponentially, so earlier times have a higher signal to noise ratio. We have run fits on the dominant $\ell = m = 2$ spherical harmonic mode for a set of CCE simulated waveforms[1] for BBH mergers with spins aligned along the z-axis.

2.1 Waveform mismatch

As a first metric to quantify the ringdown start time for a given number of overtones we will use the mismatch. Let $t_0^{N,\mathcal{M}}$ be the earliest time at which a fit with N overtones in addition to the fundamental mode is a good fit, as quantified by the mismatch. To determine $t_0^{N,\mathcal{M}}$ for a particular signal, we generated a mismatch curve for the signal. The mismatch $\mathcal{M}$ between two signals h_1 and h_2 is defined to be

$$\mathcal{M} = 1 - \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}} \tag{3}$$

where the inner product $\langle h_1 | h_2 \rangle$ is defined as

$$\langle h_1 | h_2 \rangle = \mathrm{Re} \left[\int_{t_0}^{T} h_1(t) h_2^*(t) \mathrm{d}t \right]. \tag{4}$$

The start and end times of the integral, t_0 and T, will be for us the ringdown start time and some sufficiently late time when the ringdown has decayed. In Fig. 1 we plot $\mathcal{M}$ vs t_0 for CCE waveform SXS_BBH_ExtCCE_0001 (CCE:01) with $T = t_0 + 100M$ for $N = 0$ through $N = 21$.

---

[1]The IDs of the CCE waveforms used were 0001, 0002, 0003, 0004, 0005, 0006, 0007, 0010, 0011, and 0012.
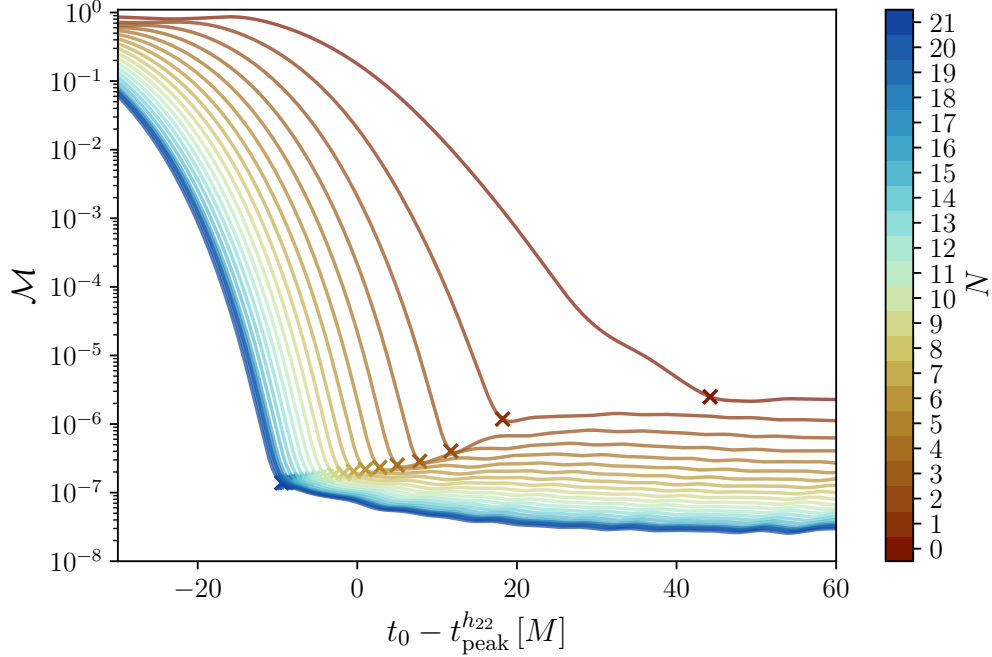
Figure 1: The mismatch vs ringdown start time curve for fits to the h_{22} mode of CCE:01 with a model consisting of N overtones of the fundamental $(2, 2, 0)$ QNM. Crosses mark the location of $t_0^{N,\mathcal{M}}$ for each N value.

Beyond $N = 21$ it becomes difficult to calculate QNM frequencies, but, as will be detailed later, 21 overtones is sufficient. We used the knee-finding Python package `kneed` [14, 15] to find the knee of the mismatch curve for each fit. This is the value of t_0 at which the mismatch curve flattens out and stops decreasing, which means that it is the earliest time at which a model with N overtones achieves its best fit. This is an indicator of the optimal ringdown start time for a given model; as will be explored later, this flattening-out of the mismatch curve roughly corresponds to where the overtones start to decay below some "noise" floor (where we use the word noise to refer to any unmodeled feature, which could be a numerical noise floor, NR error, or an unmodeled QNM). Thus, each knee is at $t_0^{N,\mathcal{M}}$. As can be seen in Fig. 1, each additional overtone allows the model to be fitted to the CCE waveform earlier, although the additional signal time gained by adding overtones decreases with N. This figure is an extension of work in Giesler et al. [10], and we note that based on the mismatch alone there is no clear indication that the maximum number of overtones is seven.

Note that although $t_0^{N,\mathcal{M}}$ is the earliest time at which a fit with N overtones is a good fit, that fit is not necessarily physically meaningful (each additional overtone in the model introduces new parameters and more model freedom, so it is expected that a better match can be achieved). For early enough values of t, the GW signal is from the binary BH merger or even inspiral, not the ringdown, so ringdown models should not be applied.

As can be seen in Fig. 1, each additional overtone seems to have "diminishing returns" in terms of the mismatch improvement. To quantify this we have defined one measure of usefulness to be the difference $\Delta t_0^{N,\mathcal{M}} = t_0^{N,\mathcal{M}} - t_0^{N-1,\mathcal{M}}$. Note that $\Delta t_0^{N,\mathcal{M}}$ is not defined for $N < 1$. This is a measure of how much additional signal can be fitted by adding the
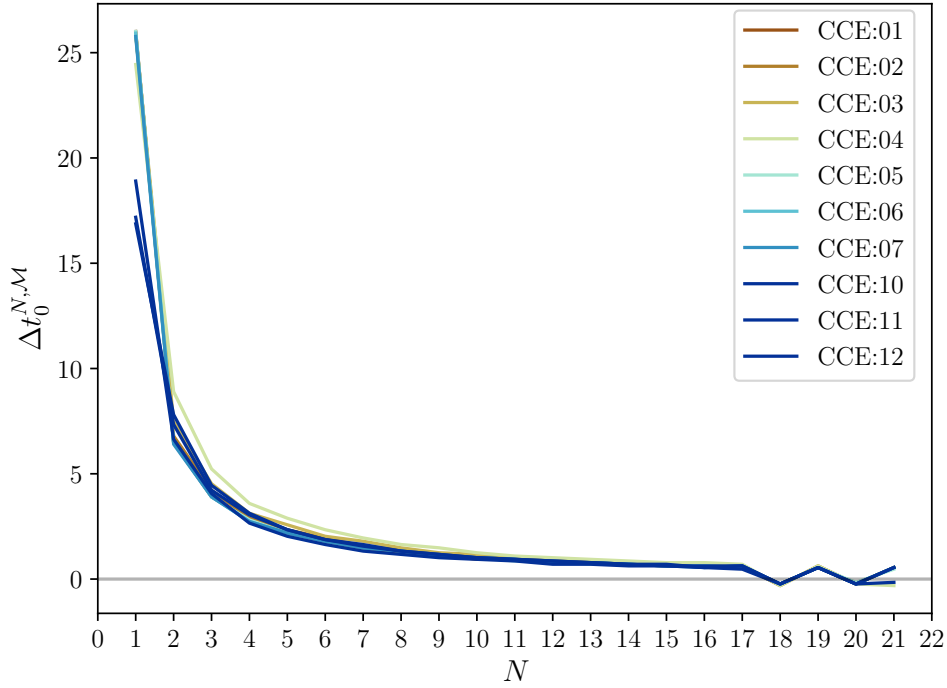
Figure 2: $\Delta t_0^{N,\mathcal{M}}$ vs N for the ten different CCE waveforms studied in this work.

N^{th} overtone. This quantity is never expected to be negative, because with each additional overtone we are only adding model freedom so the fit should never get worse (the lower-overtone models can be recovered by setting amplitudes of higher overtones to zero). Fig. 2 shows $\Delta t_0^{N,\mathcal{M}}$ as a function of N for ten CCE waveforms. For all waveforms studied, $\Delta t_0^{N,\mathcal{M}}$ decreases rapidly toward zero – this suggests there is a limit to how early overtones can fit the waveform. And, surprisingly, at $N = 18$, every waveform studied had $\Delta t_0^{N,\mathcal{M}} < 0$. This suggests that at $N = 18$ overtones, we begin to run up against the limit of the NR simulation precision, and that the maximum usable number of overtones is 17.

2.2 Injection study

Although the qualitative shape of the mismatch curves in Fig. 1 make sense, we were interested in the exact reason the curves level-out at $\mathcal{M} \sim 10^{-6}$. This will be particularly relevant when we move onto Bayesian inference, since the mismatch floor should be an indication of the "noise" in the simulation. To try to investigate this further, we created an injection waveform.

As mentioned in the introduction, because the ringdown is naturally decomposed into spheroidal harmonics each spherical harmonic mode has contributions from all other QNMs of the same m (mode mixing). The QNM with the largest expected contribution to the $(2,2,0)$ mode is the $(3,2,0)$ mode. We fitted CCE:01 with the QNMs $(2,2,0)$ through $(2,2,17)$ and $(3,2,0)$ at $t_0 = t_0^{17,\mathcal{M}}$. Then, we created a pure-QNM waveform using the fitted amplitudes of each
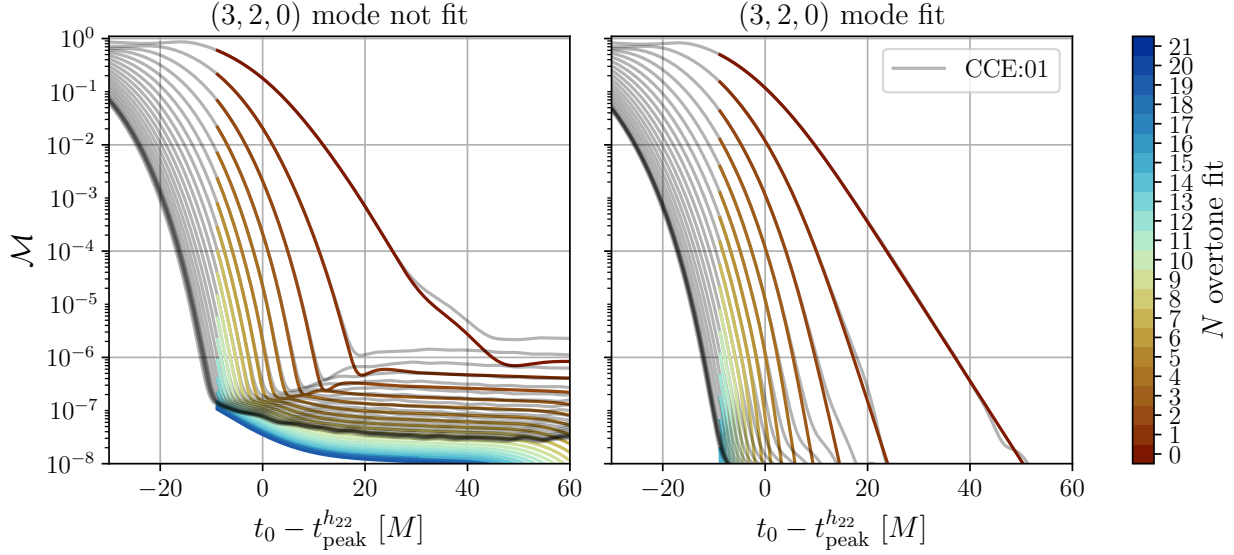
Figure 3: Plots of the mismatch between the injection waveform for fits of 0 through 21 overtones of the $(2, 2, 0)$ QNM. The gray lines show the mismatch curves for CCE:01 and the colored lines show mismatch curves for an injection waveform. The left panel shows fits where the $(3, 2, 0)$ QNM was not included, and the right panel shows fits where the $(3, 2, 0)$ mode was included.

mode, with the sum

$$h_I(t) = \left(\sum_{n=0}^{17} C_{22n} e^{-i\omega_{22n}(t-t_0)} \right) + C_{320} e^{-i\omega_{320}(t-t_0)}.$$

We then fitted models containing up to $N = 20$ overtones of the $(2, 2, 0)$ QNM to the injected pure-QNM waveform, without including the $(3, 2, 0)$ QNM in the model.

As shown in Fig. 3, we were able to reproduce the knee behavior seen in Fig. 1. When we fitted models which *did* contain the $(3, 2, 0)$ mode, the mismatch between the fit and the signal did not have a knee where it ceased to decrease. It decreased until it reached the numerical noise floor. This is a an important step in understanding the behavior of overtones: we know that the $(3, 2, 0)$ mode is influential in the mismatch leveling off after a knee. Although the location of the knees is expected to be related to when the highest overtone in the model decays below the $(3, 2, 0)$ mode, a clear relationship was hard to establish. This may be because the mismatch is an integral over the full waveform from $t_0^{\mathcal{M}}$ onwards, so relating its behaviour to a ratio of amplitudes at t_0 is an unreasonable goal (it is perhaps more related to a ratio of power in the modes).

Previous research on fitting overtones of the ringdown fit a maximum of seven overtones [10]. However, there is no particular reason to stop at seven overtones. We have shown that, for these waveforms, the maximum number of overtones that can be fit is 17. The mismatch is a mathematical metric for evaluating how well a model fits the simulated data, but does not describe whether a particular model is in physical agreement with the simulation. We will now introduce a more physical metric for goodness of fit and compare the optimal ringdown start times generated by both methods.

2.3 Remnant mass-spin recovery

Giesler et al. [10] introduced a metric ϵ which measures the physical agreement between a fit and a NR simulation. The No-Hair Theorem states that the mass M_f and dimensionless spin χ_f of the final BH completely determine the QNM frequencies. Thus, we can vary the QNM frequencies of a model by varying the model's mass and spin. For a model with N overtones and start time t_0, we minimize the mismatch $\mathcal{M}$ between the model and the signal by varying M_f and χ_f. In this work, we use `minimize` from `scipy.optimize` to perform this calculation. Then, δM_f and $\delta \chi_f$ are the differences between the simulation final mass and spin and the final mass and spin which minimize $\mathcal{M}$.

Then,

$$\epsilon = \sqrt{(\delta M_f/M)^2 + (\delta \chi_f)^2}.$$

We plot ϵ vs t_0 for CCE:01 in Fig. 4.

Just as models with more overtones have a low mismatch at earlier times, models with more overtones have a low ϵ at earlier times. Let $t_0^{N,\epsilon}$ be the earliest time at which $\epsilon(t_0)$ reaches a local minimum. As can be seen in Fig. 4, the curve has several local minima, but we choose the most prominent one.[2] This metric differs from $t_0^{N,\mathcal{M}}$ because it determines the earliest time at which a model physically matches a NR simulation.

Now, let $\Delta t_0^{N,\epsilon} = t_0^{N,\epsilon} - t_0^{N-1,\epsilon}$. This is another measure of "usefulness" for each additional

---

[2]Peak prominence was determined using Scipy's `signal.peak_prominence` method [16]. This algorithm determines the height of a peak relative to the minimum height in the largest interval which contains the peak and no higher peaks.
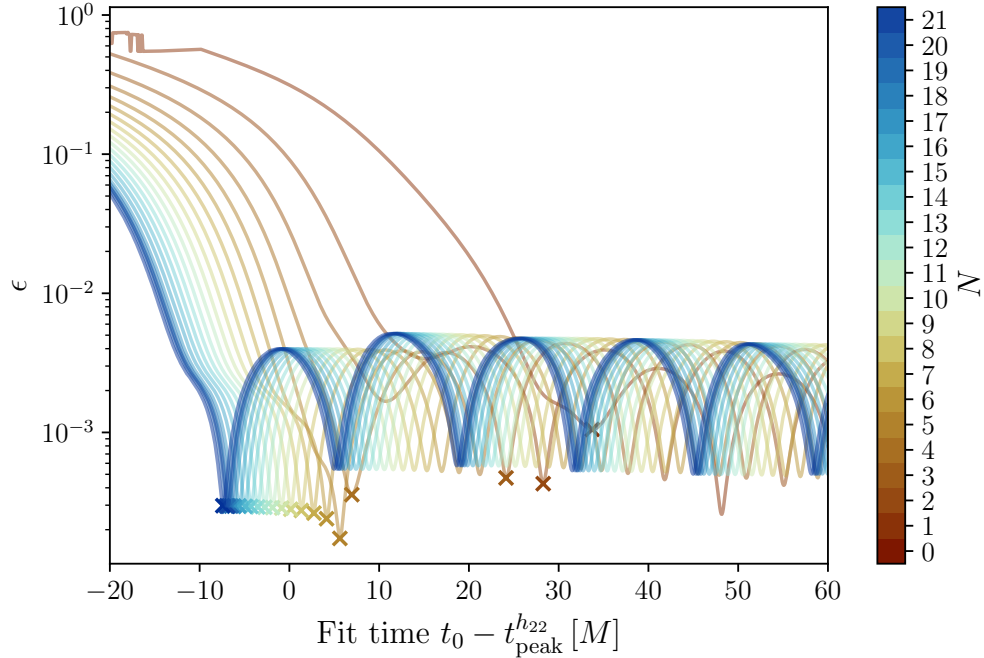

Figure 4: A plot showing the value of ϵ as a function of t_0 for CCE:01. Crosses mark the location of $t_0^{N,\epsilon}$.
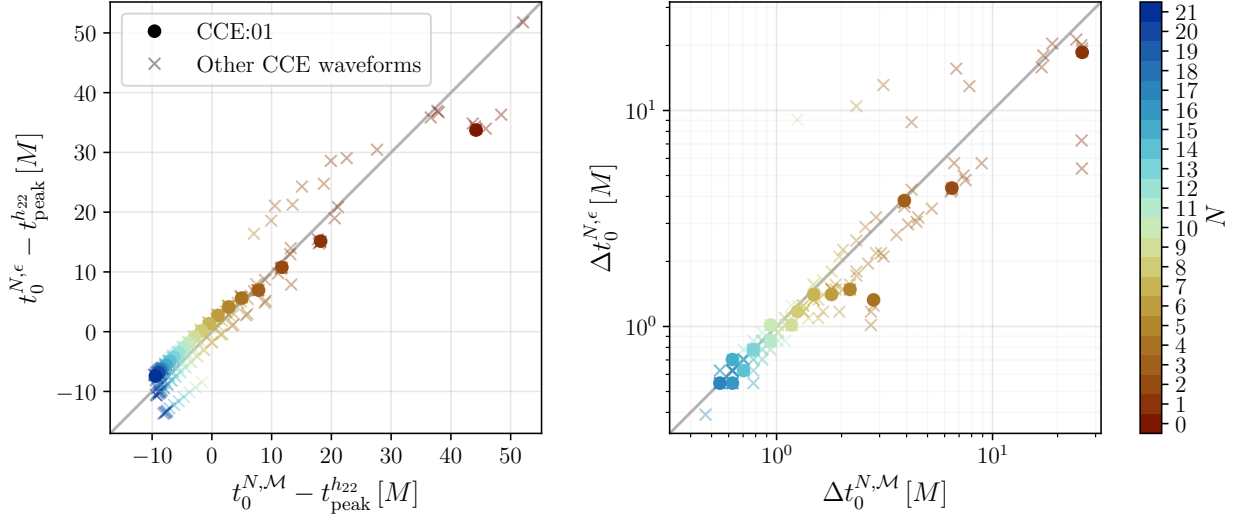
Figure 5: A plot showing $t_0^{N,\mathcal{M}}$ vs $t_0^{N,\epsilon}$ and $\Delta t_0^{N,\mathcal{M}}$ vs $\Delta t_0^{N,\epsilon}$ for the ten CCE waveforms used in this analysis for fits with $N \in [0, 21]$ overtones. Each point represents one fit, with color indicating the number of overtones used. The two times $t_0^{N,\mathcal{M}}$ and $t_0^{N,\epsilon}$ are correlated, especially for higher values of N.

overtone. The greater the value of $\Delta t_0^{N,\epsilon}$ is, the more value (e.g. earlier fitting time and higher SNR) is gained by adding the Nth overtone. We calculate $\Delta t_0^{N,\epsilon} = t_0^{N,\epsilon} - t_0^{N-1,\epsilon}$ in the same way as $\Delta t_0^{N,\mathcal{M}}$. These two quantities are very different measurements, but $\Delta t_0^{N,\mathcal{M}}$ and $\Delta t_0^{N,\epsilon}$ are highly correlated, especially for higher numbers of overtones (see Fig. 5). This suggests that they provide reasonable ringdown start time estimates. The plot shows a general downward trend in $\Delta t_0^{N,\epsilon}$, similar to Fig. 2, although some CCE waveforms significantly deviate from this trend. Also note that both methods can't seem to fit before $-10M$ relative to the time of peak strain.

So far, we have defined two definitions of usefulness: $\Delta t_0^{N,\mathcal{M}}$ and $\Delta t_0^{N,\epsilon}$. Both measures show diminishing returns for each additional overtone fitted, although the decline in $\Delta t_0^{N,\epsilon}$ is less uniform.

3   Frequency perturbations

Another method of defining the usefulness of an overtone is by how much it influences the fit overall. A major goal of GW science is to test GR, so the impact of deviations from GR in each overtone is an important metric. This was calculated by introducing a variation in frequency δ. Within GR, the frequency $\omega_{\ell mn}$ of the (ℓ, m, n) QNM is completely determined by the mass and dimensionless spin of the final BH. To introduce deviations from GR in one mode, the QNM frequency for a particular mode $(2, 2, \tilde{n})$ were modified by a factor of $1 + \delta$, such that $\tilde{\omega}_{\ell m \tilde{n}} = (1 + \delta)\omega_{\ell m \tilde{n}}^{\text{GR}}$. Then a least-squares fit was performed on the phases and amplitudes of the set of modes with frequencies $\{\omega_{220}, \omega_{221}, \ldots, \tilde{\omega}_{22\tilde{n}}, \ldots, \omega_{22N}\}$. ϵ was calculated for this fit, and plotted vs $\tilde{n}$ in Fig. 6. The uncertainty shown in this figure was calculated by finding difference in remnant mass and spin between the two highest resolution
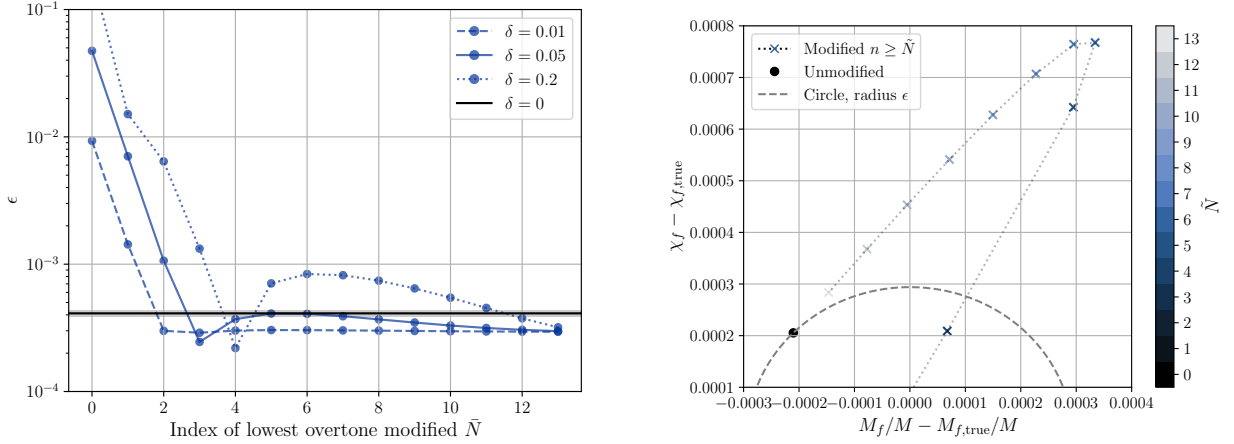
Figure 6: *Left:* A plot showing the deviation ϵ from the true M_f and χ_f for different size frequency modifications δ for CCE:01 in fits with $N = 13$. In this plot, QNMs with $n \geq \bar{N}$ were modified. The behavior of the other CCE waveforms used in this study is similar. *Right:* A plot showing the difference between best-fit M_f and χ_f values and the true M_f and χ_f for fits with 13 overtones with $\delta = 0.2$, where each cross represents the best fit mass-spin pair for a different modified overtone. This plot is an alternate visualization of the dotted line from the left panel. For reference, we show a circle centered on the true remnant mass and spin with a radius equal to the value of ϵ obtained via a 13-overtone fit with unmodified frequencies. Crosses inside the circle indicate that a particular modified fit provided more accurate mass and spin than the unmodified fit.

versions of the CCE waveform, and calculating $\epsilon_{\mathrm{simulation}} = \sqrt{\Delta(M_f/M)^2 + (\Delta\chi)^2}$.

We found that each successive overtone has a smaller effect on the best-fit mass-spin of the system.

This result also provides insight into a result from Giesler et al. [10]. That study found that when all overtones are modified together, as the number of overtones increases, the value of ϵ decreases, suggesting that each overtone added does indeed improve the match between signal and model (see their Fig. 7). We repeated this analysis, and found that the increase in ϵ was largely driven by the modification to the least-damped overtone (in this case, the (2,2,1) mode). Consequently, this result does not clearly demonstrate the impact of modifications of the higher overtones. This effect can be seen in Fig. 7, where a similar result is obtained by either modifying the $n = 1$ mode only (orange line), or all modes with $n \geq 1$ (blue line, as in Giesler et al.).

The NR waveforms used in this analysis were created using numerical solutions to the equations of GR, so the $\delta = 0$ case should yield the best results. However, we find that unmodified modes perform worse than modified modes for low values of δ for $N > 4$. Thus, caution must be used when applying this type of analysis, as the results may not be precise.

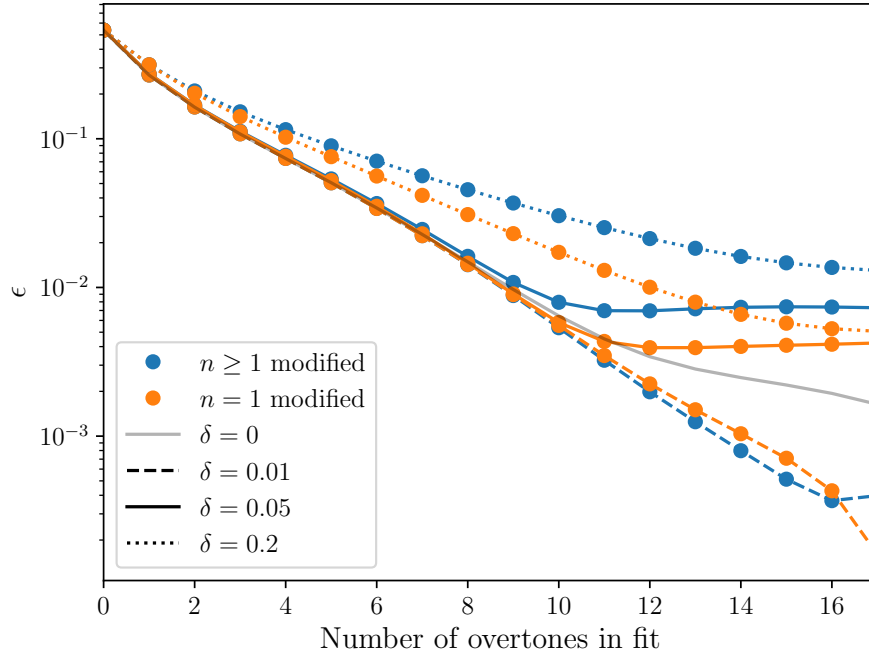Figure 7: ϵ for fits to CCE:01 with varying numbers of overtones and modifications to the GR frequency of one or more overtones. In blue are fits which modify the frequency of only the $(2, 2, 1)$ mode, and in orange are fits which modify the frequencies of all overtones of the $(2, 2, 0)$ mode.

4 Amplitude stability and resolvability

We are also interested in the stability of the amplitudes of overtones for different fitting times. When a fit is performed at time t_0, it returns the amplitude of each mode at t_0. To compare amplitudes, we adjust the amplitude to be the amplitude at some time t_{ref}. This is possible because each QNM decays at a known rate. Throughout this work, we set $t_{\text{ref}} = 0 = t_{\text{peak}}^{h_{22}}$.

Then, we can rescale $A_{\ell mn}(t_0)$ to the reference time t_{ref} using the relationship

$$A_{\ell mn}(t_{\text{ref}}) = A_{\ell mn}(t_0) \exp\left(\frac{t_0 - t_{\text{ref}}}{\tau_{\ell mn}(M_f, \chi_f)}\right)$$

where $\tau_{\ell mn}(M_f, \chi_f)$ is the decay time of the mode. $\tau_{\ell mn}$ is related to the complex frequency of the (ℓ, m, n) QNM $\omega_{\ell mn}$ by the equation

$$\tau_{\ell mn} = -\frac{1}{\text{Im}[\omega_{\ell mn}]}.$$

Now that we can compare the amplitudes of QNMs fitted at different times, we measure the stability of the fitted amplitudes. Let $A_{\ell mn}^N$ be the amplitude of the (ℓ, m, n) mode in a fit with N overtones, fitted at $t_0 = t_0^{N,\mathcal{M}}$. We plot $A_{\ell mn}^N$ vs N for up to seven overtones in Fig. 8. It is possible to plot $A_{\ell mn}^N$ for larger numbers of overtones, but the plot becomes very large.

Fig. 8 shows that the amplitudes of higher overtones are more unstable than the amplitude of lower overtones. In general, this plot also suggests that the amplitude of the nth overtone decreases as more overtones are added, although the magnitude of each decrease gets smaller.

However, least-squares fitting can overfit to numerical noise in the data. In particular, least-squares fits of a QNM after it has decayed below the noise floor of a simulation attempt to fit that QNM to the noise, rather than recognizing that it has an amplitude consistent with zero. Bayesian inference, in contrast to least-square fitting, allows us to incorporate a notion of uncertainty into the fits and overcome this problem.

We used the Python module `emcee` [17] to create posterior distributions for QNM amplitudes and phases. This sampler is an affine invariant MCMC sampler. Since the $(3, 2, 0)$ mode dominates the effective noise floor, as seen in Fig. 3, we wanted to use the amplitude of the $(3, 2, 0)$ mode to set the scale for the noise amplitude for a Bayesian analysis of the data.

We tested the sampler's limits using zero-noise injected waveforms, to determine the reliability of sampled posterior distributions. We created injections with a varying number of overtones, then sampled the posterior distributions for each mode's amplitude and phase using a fixed signal noise floor σ. We assumed that the amplitude of the noise is time independent and Gaussian, so the likelihood of a data point d_i if there is a signal present with parameters $\vec{\theta}$ is

$$\mathcal{L}(d_i|\vec{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{d_i - h_i(\vec{\theta})}{\sigma}\right)^2\right].$$
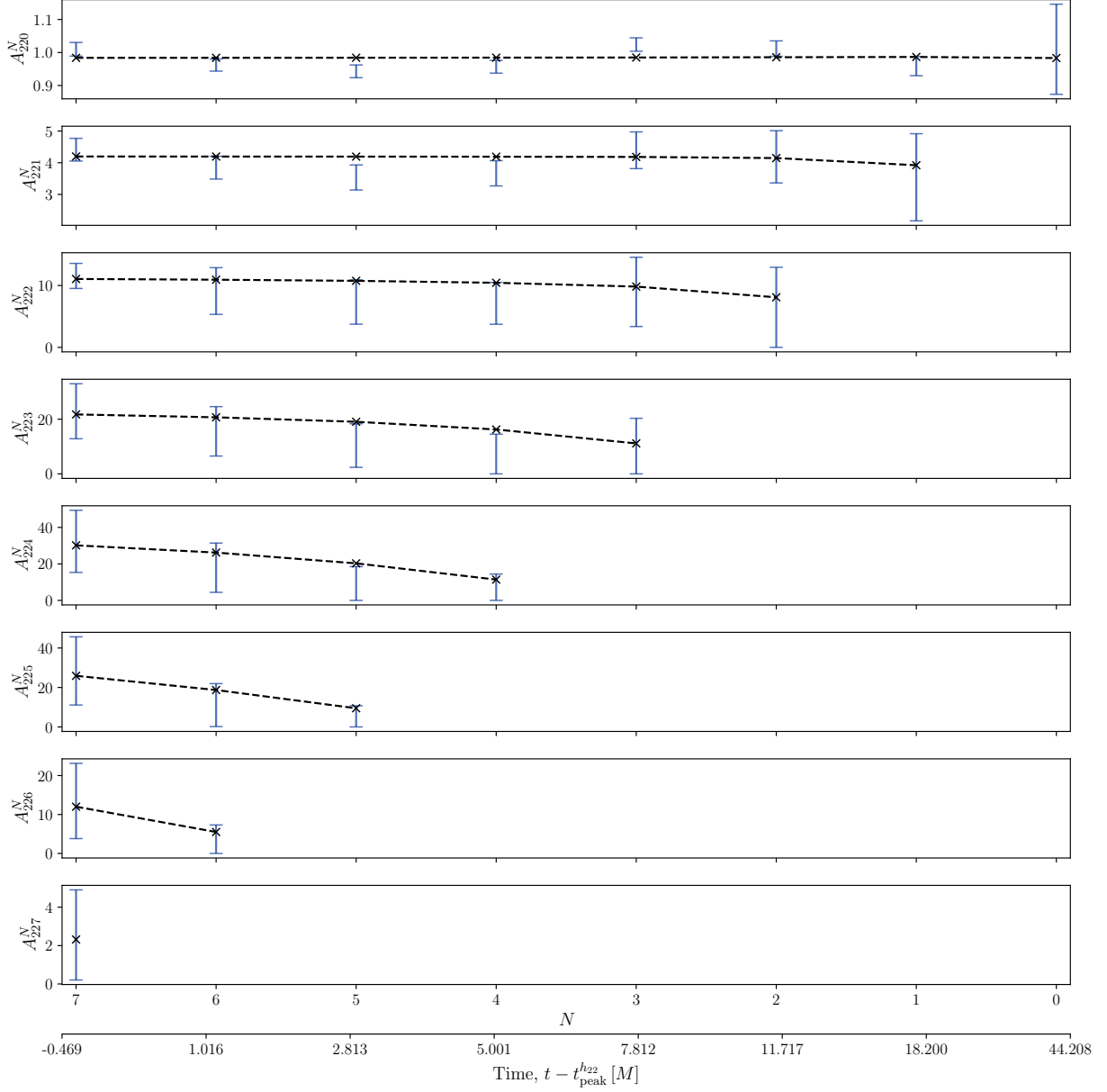
Figure 8: A plot showing the amplitude of each QNM for a fit performed at $t_0^{N,\mathcal{M}}$, adjusted to $t_{\mathrm{ref}} = 0$, as a function of N. Crosses show the least-squares best fit amplitude, and error bars show the 90% credible interval for amplitude found using Bayesian analysis.
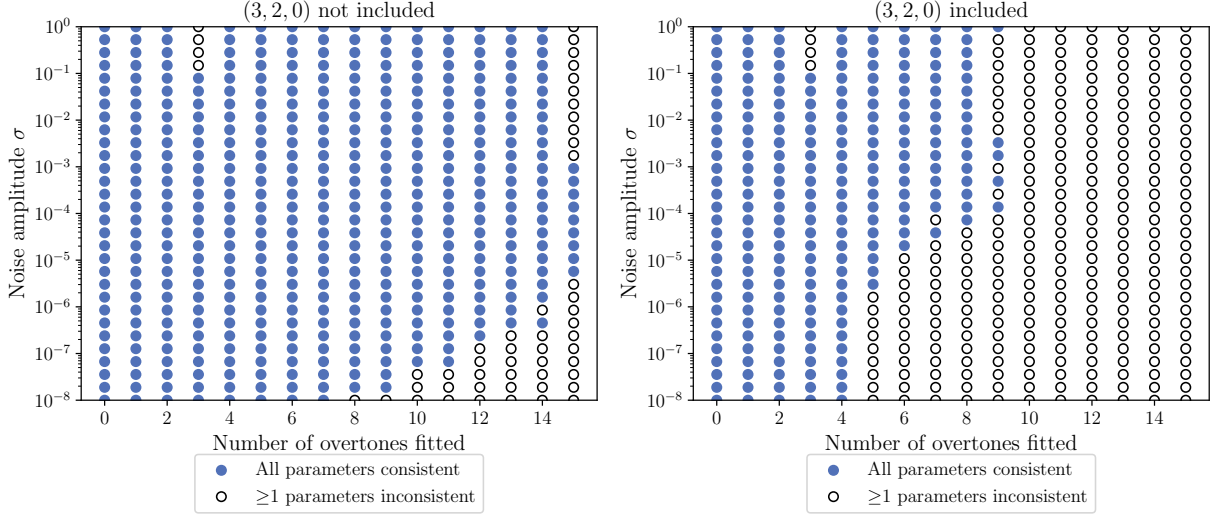
Figure 9: The range of noise amplitudes and number of overtones for which `emcee` could successfully recover all parameters correctly at the 90% confidence level. *Left*: Consistency of sampled posteriors for injected waveforms without the $(3, 2, 0)$ mode. All QNMs injected were modeled. *Right*: Consistency of sampled posteriors for injected waveforms which included the $(3, 2, 0)$ mode. Although the $(3, 2, 0)$ mode was included in the injected waveform, it was not modelled.

Since every data point is independent, the likelihood for a signal consisting of N data points is thus

$$\mathcal{L}(d|\vec{\theta}) = \prod_{i=0}^{N} \mathcal{L}(d_i|\vec{\theta}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left\{-\frac{1}{2}\sum_{i=0}^{N}\left[\frac{d_i - h_i(\vec{\theta})}{\sigma}\right]^2\right\}.$$

Additionally, we assume that the noise in the plus and cross polarizations are uncorrelated, so

$$\mathcal{L}_{\text{total}}(d|\vec{\theta}) = \mathcal{L}_+(d|\vec{\theta})\mathcal{L}_\times(d|\vec{\theta}).$$

We considered the samples for a parameter to be consistent with the injected value if the 90% credible range included the parameter's injected value. We tested a range of noise floor values σ and N values, and the results for the number of consistent parameters are shown in Fig. 9. Each fit with N overtones was performed at $t_0 = t_0^{N,\mathcal{M}}$. We find that the minimum sigma value required to consistently recover all parameters increases exponentially with the number of overtones.

We also tested the performance of `emcee` when unmodeled modes are present in data. We used `emcee` to perform fits on zero-noise injections where the $(3, 2, 0)$ mode was present, but unmodeled, similar to the fitting procedure described in Section 2. We found that in this case, the sampler performed slightly worse, but that there was no N dependency. For our sampler settings, the parameter values could be successfully recovered if $\sigma \lesssim 0.001$. For this waveform, this corresponds to approximately the amplitude of the $(3, 2, 0)$ mode.

Having probed the limits of the `emcee` sampler, we began to analyze real data to determine the consistency of amplitudes across fit times. As a conservative choice of σ, we chose σ
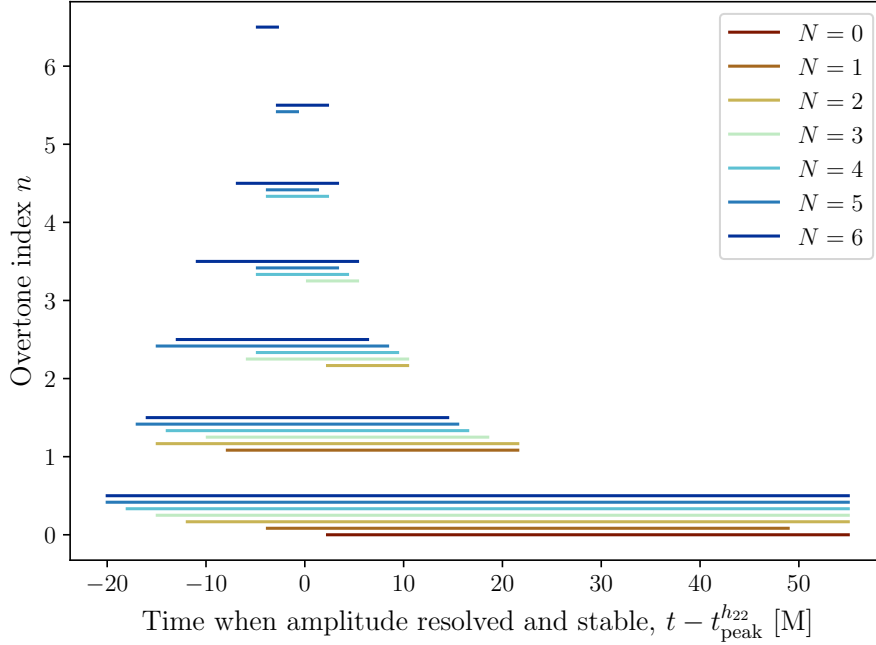
Figure 10: The times at which QNMs in several Bayesian fit were both stable and resolved.

such that $\sigma \approx 3A_{320}(t_0 = 0)$. The $(3, 2, 0)$ QNM decays very slowly relative to the $(2, 2, 0)$ QNM and its overtones, so its amplitude is essentially constant throughout the time period we are fitting on (approximately $-30M$ to $60M$). Unlike injections, NR does not include true mode amplitudes, so we require metrics relating to self-consistency. We adopt the two criteria detailed by Clarke et al. in [11]: resolvability and stability, using 90% credibility as our threshold. When a mode has decayed below the noise floor of the simulation (e.g. when its amplitude is less than the amplitude of the noise realization employed in the likelihood calculation of the Bayesian fit), its amplitude is consistent with zero. We considered the (ℓ, m, n) QNM to be resolved if the highest density 90% credible interval of the posterior distribution for $A_{\ell mn}$ did *not* include zero. Before a model with N modes is a good fit to the data, parameter values are unstable. We considered a QNM to be stable at a particular fit time if its highest-density 90% credible interval overlapped the highest-density 90% credible interval at all subsequent resolved times for both its phase and amplitude. See Fig. 10 for a plot showing the times at which any given mode were both stable and resolved. In their 2024 study, Clarke et al.[11] were able to simultaneously recover a maximum of three QNMS with nonzero amplitude.

We find many strong correlations between fit parameters (see Fig. 11). This can be at least partially explained by the nature of the sum of QNMs. As n increases, the amplitude of the $(2, 2, n)$ mode increases, but each QNM is out of phase with the previous one, keeping the summed amplitude relatively small. This causes the amplitude and phase of the (ℓ, m, n) mode to show correlations with the amplitude and phase of the $(\ell, m, n-1)$ and $(\ell, m, n+1)$ QNMs. This effect is particularly strong for high n. This means that a sampler applied to this problem should be affine-invariant, or a transformation should be applied to produce symmetric posterior distributions for Hamiltonian Monte-Carlo samplers.
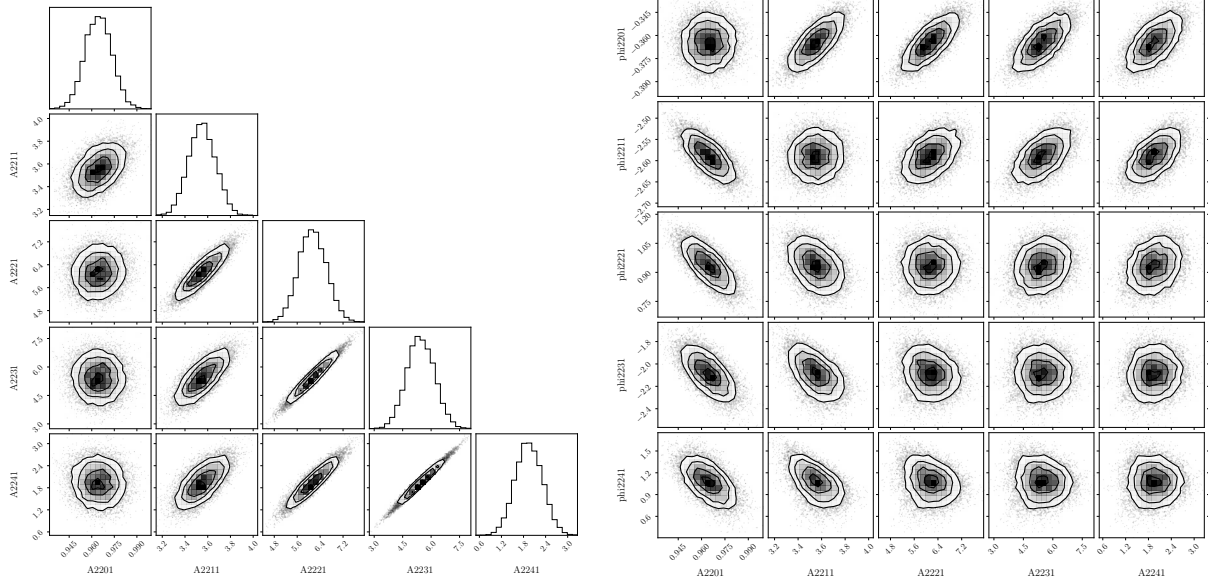
Figure 11: Posterior distributions from a Bayesian inference performed at $t = 0$ of CCE:01 with four overtones. *Left:* Posterior distributions of QNM amplitudes. Strong correlations are seen between the amplitudes of QNMs, especially for higher values of n. A similar correlation exists between phase posteriors. *Right:* The posteriors for QNM phases vs the posteriors for QNM amplitudes. We believe the correlations seen between the fundamental mode amplitude and higher-overtone phases (notably with ϕ_{224}) are due to a trickle-down effect from strong correlations in amplitude and phase in higher overtones.

Within the current constraints on noise threshold and number of overtones, the sampler has produced excellent results. At late times, when an overtone is expected to have decayed below the noise floor, the posterior shows that the amplitude of that overtone is most likely to be near-zero.

5   Conclusions

Valid ringdown start times can be found using mismatch and mass-spin recovery using least-squares fits for models containing up to 17 overtones. Beyond 17 overtones, we find that these fits do not perform well. This behavior is likely related to the precision limit of the NR simulations used. We also find that ringdown start times calculated using mismatch and mass-spin recovery are highly correlated, suggesting that when a model is physically valid, it is a good fit to the data. Our frequency perturbation work found that the contribution of each overtone to tests of GR decreases with overtone index, and that previous work on perturbed frequencies overlooked the dominance of the low-n overtones. In Bayesian fits of amplitude and phase, we employ two criteria to determine whether amplitude and phase posteriors indicate that a model with N overtones is valid at a given time: amplitude resolvability (whether a given QNM amplitude is consistent with zero at the 90% credible level) and amplitude and phase stability across fit times. These fits also revealed intriguing correlations between phase and amplitude posteriors, especially for high overtone index n.

6   Acknowledgements

References

[1] R. Abbott et al. "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run". In: *Phys. Rev. X* 13.4 (2023), p. 041039. DOI: 10.1103/PhysRevX.13.041039. arXiv: 2111.03606 [gr-qc].

[2] Emanuele Berti, Vitor Cardoso, and Andrei O. Starinets. "Quasinormal modes of black holes and black branes". In: *Class. Quant. Grav.* 26 (2009), p. 163001. DOI: 10.1088/0264-9381/26/16/163001. arXiv: 0905.2975 [gr-qc].

[3] Emanuele Berti and Antoine Klein. "Mixing of spherical and spheroidal modes in perturbed Kerr black holes". In: *Phys. Rev. D* 90.6 (2014), p. 064012. DOI: 10.1103/PhysRevD.90.064012. arXiv: 1408.1860 [gr-qc].

[4] Michael Boyle et al. "The SXS Collaboration catalog of binary black hole simulations". In: *Class. Quant. Grav.* 36.19 (2019), p. 195006. DOI: 10.1088/1361-6382/ab34e2. arXiv: 1904.04831 [gr-qc].

[5] SXS Collaboration. *The Ext-CCE Waveform Database.* https://data.black-holes.org/waveforms/extcce_catalog.html.

[6] Michael Boyle et al. *scri.* Version v2022.8.10. Feb. 2024. DOI: 10.5281/zenodo.10709136. URL: https://doi.org/10.5281/zenodo.10709136.

[7] Maximiliano Isi et al. "Testing the no-hair theorem with GW150914". In: *Phys. Rev. Lett.* 123.11 (2019), p. 111102. DOI: 10.1103/PhysRevLett.123.111102. arXiv: 1905.00869 [gr-qc].

[8] Olaf Dreyer et al. "Black hole spectroscopy: Testing general relativity through gravitational wave observations". In: *Class. Quant. Grav.* 21 (2004), pp. 787–804. DOI: 10.1088/0264-9381/21/4/003. arXiv: gr-qc/0309007.

[9] Gregorio Carullo, Walter Del Pozzo, and John Veitch. "Observational black hole spectroscopy: A time-domain multimode analysis of GW150914". In: *Physical Review D* 99.12 (June 2019). ISSN: 2470-0029. DOI: 10.1103/physrevd.99.123029. URL: http://dx.doi.org/10.1103/PhysRevD.99.123029.

[10] Matthew Giesler et al. "Black Hole Ringdown: The Importance of Overtones". In: *Phys. Rev. X* 9.4 (2019), p. 041060. DOI: 10.1103/PhysRevX.9.041060. arXiv: 1903.08284 [gr-qc].

[11] Teagan A. Clarke et al. "Toward a self-consistent framework for measuring black hole ringdowns". In: *Phys. Rev. D* 109.12 (2024), p. 124030. DOI: 10.1103/PhysRevD.109.124030. arXiv: 2402.02819 [gr-qc].

[12] Eliot Finch. *Black-hole Ringdown: Quasinormal Modes in Numerical-relativity Simulations and Gravitational-wave Observations.* https://etheses.bham.ac.uk/id/eprint/13992/. 2023.

[13] Eliot Finch and Christopher J. Moore. *qnmfits.* 2024. URL: https://github.com/eliotfinch/qnmfits.

[14] Ville Satopaa et al. "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior". In: *2011 31st International Conference on Distributed Computing Systems Workshops.* 2011, pp. 166–171. DOI: 10.1109/ICDCSW.2011.20.

[15] Kevin Arvai. *kneed.* Version v0.8.5. July 2023. DOI: 10.5281/zenodo.8127224. URL: https://doi.org/10.5281/zenodo.8127224.

[16] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[17] Daniel Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[18] Fabio Crameri. *Scientific colour maps.* Version 8.0.1. Oct. 2023. DOI: 10.5281/zenodo.8409685. URL: https://doi.org/10.5281/zenodo.8409685.

[19] Daniel Foreman-Mackey. "corner.py: Scatterplot matrices in Python". In: *The Journal of Open Source Software* 1.2 (June 2016), p. 24. DOI: 10.21105/joss.00024. URL: https://doi.org/10.21105/joss.00024.

[20] Leo C. Stein. "qnm: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients". In: *The Journal of Open Source Software* 4.42 (2019), p. 1683. DOI: 10.21105/joss.01683. arXiv: 1908.10377 [gr-qc].