

Comparing gravitational waveform models for binary black hole mergers through a hypermodels approach

Anna Puecher^{1,2}, Anuradha Samajdar^{2,3}, Gregory Ashton⁴, Chris Van Den Broeck^{1,2}, and Tim Dietrich^{3,5}

¹*Nikhef – National Institute for Subatomic Physics,
Science Park 105, 1098 XG Amsterdam, Netherlands*

²*Institute for Gravitational and Subatomic Physics (GRASP),
Utrecht University, Princetonplein 1, 3584 CC Utrecht, Netherlands*

³*Institut für Physik und Astronomie, Universität Potsdam,
Haus 28, Karl-Liebknecht-Str. 24/25, 14476, Potsdam, Germany*

⁴*Royal Holloway, University of London, London TW20 0EX, United Kingdom and*

⁵*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany*

(Dated: October 6, 2023)

The inference of source parameters from gravitational-wave signals relies on theoretical models that describe the emitted waveform. Different model assumptions on which the computation of these models is based could lead to biases in the analysis of gravitational-wave data. In this work, we sample directly on four state-of-the-art binary black hole waveform models from different families, in order to investigate these systematic biases from the 13 heaviest gravitational-wave sources with moderate to high signal-to-noise ratios in the third Gravitational-Wave Transient Catalog (GWTC-3). All models include spin-precession as well as higher-order modes. Using the “hypermodels” technique, we treat the waveform models as one of the sampled parameters, therefore directly getting the odds ratio of one waveform model over another from a single parameter estimation run. From the joint odds ratio over all 13 sources, we find the model `NRSur7dq4` to be favoured over `SEOBNRv4PHM`, with an odds ratio of 29.43; `IMRPhenomXPHM` and `IMRPhenomTPHM` have an odds ratio, respectively, of 4.70 and 5.09 over `SEOBNRv4PHM`. However, this result is mainly determined by three events that show a strong preference for some of the models and that are all affected by possible data quality issues. If we do not consider these potentially problematic events, the odds ratio do not exhibit a significant preference for any of the models. Although further work studying a larger set of signals will be needed for robust quantitative results, the presented method highlights one possible avenue for future waveform model development.

I. INTRODUCTION

Following the direct detection of gravitational waves (GWs) in 2015 [1, 2], we have direct access and constraints on strong-field gravity [3].

With close to 90 significant observations of binary black hole (BBH) mergers [8], hyperparameters characterizing population models [9] as well as more stringent bounds on strong-field gravity parameters from combining multiple events [10] have been estimated. Ongoing and future observing runs of the LIGO-Scientific, Virgo, and KAGRA collaborations will operate at higher sensitivities and enable us to see many more events. However, as the statistical biases reduce through improved detector sensitivities and by combining multiple events, the systematic effects from the GW models employed to analyze our data will start dominating. Several studies have been made to expose this problem with future-generation detectors, e.g. Ref. [11].

Typically, GW source properties are inferred by analyzing the data with multiple waveform models where the estimates broadly agree. This serves as a consistency test between different models developed employing different techniques. Separate analyses are therefore performed on a single event to obtain estimates of the same. However, while individual sources may be consistent, combining the data may expose a bias or preference for one model over another. In this work, we in-

fer the parameter properties of the 13 heaviest and significant BBH observations by Advanced LIGO [12] and Advanced Virgo [13] in GWTC-3 [8] and quantify the preference for one waveform model over another from the combined GW data. The choice of events is determined by the fact that, for one of the models employed, the region of validity covers only high values of the binary’s total mass; moreover, the shorter duration of signals produced by high-mass systems reduces the computational cost of the analysis. Reference [14] has looked at a very similar problem from a technical point of view, performing a joint Bayesian analysis with three different models on a large set of simulated events, showing consistent results with the ones obtained via a Bayesian model averaging method, and with a significant gain in terms of computational cost. However, the analyzed signals were all simulations apart from one real GW event, `GW200129_065458`, also included in our suite of events. The focus of our work is instead on real events, with the goal to investigate possible systematic biases caused by the different waveform approximants. We employ four waveform models: `NRSur7dq4` [15], `IMRPhenomXPHM` [16], `IMRPhenomTPHM` [17], and `SEOBNRv4PHM` [18]. In Ref. [19], all the events in GWTC-3 are analyzed with the `NRSur7dq4` model, finding, in some cases, different results with respect to the ones obtained with the `IMRPhenomXPHM` and `SEOBNRv4PHM` models in the LVK analyses.

For our study, we focus on the method introduced in Ref. [20], henceforth referred to as *hypermodels*. The purpose of our study is to obtain a quantitative measure of selection, in this case by using the *odds ratio*, between one waveform and another from a combination of GW events.

We outline our analysis method in Sec. II, by giving an overview of the models used in Sec. II A and the inference techniques in Sec. II B. We summarize our results on the individual events and the combined analyses in Sec. III. We conclude in Sec. IV.

II. METHODS

A. Waveform models

We consider for our analysis four state-of-the-art BBH waveforms, all including precession [21, 22] and higher-order modes [23]. The employed models are briefly described below.

1. *NRSur7dq4*

NRSur7dq4 [15] is a time-domain surrogate model that extends the previous *NRSur7dq2* [24] to higher values of mass ratio. Surrogate models [25, 26] are constructed by interpolating over a set of precomputed waveforms, in this case numerical-relativity (NR) waveforms built over the parameter space for precessing BBH systems. This approach produces very accurate waveforms, since it does not rely on any approximation, except for the numerical discretization in the simulations. However, due to the computational cost of NR simulations, only a limited parameter space region can be covered. In particular, the *NRSur7dq4* model is valid for mass ratio values up to $q \leq 6$ and for total mass values $M \gtrsim 66 M_\odot$ (cf. Fig. 9 in Ref.[15] for the precise range of validity as a function of the system’s mass ratio). *NRSur7dq4* includes all the subdominant modes up to $\ell \leq 4$.

2. *SEOBNRv4PHM*

SEOBNRv4PHM [18] is a time-domain, effective-one-body precessing waveform built from the aligned-spin model in Ref. [27]. The effective-one-body formalism (EOB) [28, 29] maps the dynamics of two bodies into the dynamic of a reduced-mass body moving in a deformed metric. The gravitational waveforms computed with this approach are accurate but slow to generate. For *SEOBNRv4PHM*, the precessing sector is calibrated to NR simulations. It includes the subdominant harmonics $(\ell, |m|) = (2, 1), (3, 3), (4, 4), (5, 5)$, and it is valid for mass ratio values in the range $1 \leq q \leq 50$.

3. *IMRPhenomXPHM*

IMRPhenomXPHM [16] is a phenomenological, frequency-domain approximant based on the non-precessing *IMRPhenomXHM* model [30], and constructed via the so-called “twisting-up” procedure [21, 31, 32], which allows to map non-precessing systems to precessing ones. Phenomenological models [33, 34] are built from piecewise closed-form expressions, which make them computationally cheap. The model is constructed separately for the three different inspiral, intermediate, and ringdown regions. In the inspiral and ringdown region, the information available from the post-Newtonian expansion or black hole perturbation theory is integrated with some phenomenological coefficients calibrated to NR simulations, while the intermediate region is fully calibrated to NR waveforms. This approximant includes the subdominant modes $(\ell, |m|) = (2, 2), (2, 1), (3, 3), (3, 2), (4, 4)$, which are calibrated to NR waveforms individually. The model is valid for spins magnitude up to 0.99 and $q \leq 1000$ (while its recommended usage region is $q \leq 20$, due to its calibration to NR simulations).

4. *IMRPhenomTPHM*

This approximant also belongs to the family of phenomenological models, but it is built in the time domain. Although working in the frequency domain offers an additional speed-up when computing the noise-weighted inner products, a time-domain model allows a direct description of the system’s dynamics. *IMRPhenomTPHM* [17] is built from the non-precessing model *IMRPhenomTHM* [35] via the “twisting-up” procedure, which is however different to the procedure applied in the frequency domain. This model includes the subdominant harmonics $(\ell, |m|) = (2, 1), (3, 3), (4, 4), (5, 5)$. The parameter range of validity is defined by: $m_2 \geq 0.5 M_\odot$, with m_2 being the secondary mass, and spin magnitude $|\chi_{1,2}| \leq 0.99$ for $q \leq 200$ (while its recommended usage region is $q \leq 20$, due to its calibration to NR simulations).

B. Bayesian framework

Analyzing GW signals in a Bayesian framework allows both inference of the source parameters and a comparison between different possible models describing the GW waveform. The source parameters $\vec{\theta}$ can be recovered from the detector data d evaluating the posterior $p(\vec{\theta}|d, \Omega)$, where Ω is the waveform model. In this context, Bayes theorem reads

$$p(\vec{\theta}|d, \Omega) = \frac{p(d|\vec{\theta}, \Omega)p(\vec{\theta}|\Omega)}{p(d|\Omega)}, \quad (1)$$

where $p(d|\vec{\theta}, \Omega)$ represents the *likelihood* of observing the data d given the model Ω and the specific set of pa-

parameters $\vec{\theta}$, and $p(\vec{\theta}|\Omega)$ the *prior probability density*. We employed the same default priors used in the parameter estimation analysis for these events in the LVK catalog papers [8, 36], adjusting them as follows in order to respect the region of validity of all the four approximants considered: $q \leq 6$, $\chi_{1,2} \leq 0.99$, $m_2 \geq 0.5M_\odot$. For some events, we also adjust the prior on chirp mass to ensure $\mathcal{M}_c \geq 26M_\odot$, to allow for the validity of NRSur7dq4 in the entire region of the prior volume. The denominator in Eq. 1 is the *evidence* for the model Ω , and is determined by the requirement that the posterior distribution must be normalized

$$p(d|\Omega) = \int d\vec{\theta} p(d|\vec{\theta}, \Omega) p(\vec{\theta}|\Omega). \quad (2)$$

The evidence allows us to compare different models, say Ω_A and Ω_B , computing the *odds ratio*

$$\mathcal{O}_B^A = \frac{p(\Omega_A|d)}{p(\Omega_B|d)} = \underbrace{\frac{p(d|\Omega_A)}{p(d|\Omega_B)}}_{\mathcal{B}_B^A} \underbrace{\frac{p(\Omega_A)}{p(\Omega_B)}}_{\pi_B^A} = \mathcal{B}_B^A \times \pi_B^A, \quad (3)$$

where the *Bayes factor* \mathcal{B}_B^A is the ratio of the evidence for the two models given the data, and π_B^A is usually set to 1, meaning that we do not have any a-priori preference for one of the models.

The posterior probability density and the evidence can be estimated with stochastic sampling methods. In particular, here we employ the *hypermodels* approach introduced in Ref. [20], with a Metropolis-Hastings MCMC algorithm [37, 38], based on the implementation of the BILBY-MCMC sampler [39].

C. Hypermodels

The waveform model Ω employed during the sampling is substituted with a hypermodel $\Omega = \{\Omega_0, \Omega_1, \dots, \Omega_{n-1}\}$, with n being the number of models we want to study. The parameter space investigated by the sampler, therefore, becomes $\{\vec{\theta}, \omega\}$, where $\vec{\theta}$ are the usual source parameters, while ω is a categorical parameter $\omega \in [0, 1, \dots, n-1]$ representing the waveform approximant. We define a mapping between the value of the parameter ω and a specific waveform approximant, so that at each iteration the sampler picks a value of $\{\vec{\theta}, \omega\}$ and generates the waveform with parameters $\vec{\theta}$ and the approximant corresponding to ω . We employ a uninformative prior $\pi(\omega) = 1/n$, which translates into a prior odds $\pi_B^A = 1$ for all the combinations of models considered. Among the final N posterior samples, we can distinguish the samples for each waveform ℓ from the value of the ω parameter. If n_ℓ is the number of samples for the ℓ -th approximant, its probability with respect to the other waveforms is given by $p_\ell = n_\ell/N$. The odds ratio between two models $\omega = A$ and $\omega = B$ is computed as

$$\mathcal{O}_B^A = \frac{p_A}{p_B} = \frac{n_A}{n_B}, \quad (4)$$

The error on $p_{A,B}$ is given by the variance of a Poisson process, yielding $\sigma_{p_A, p_B}^2 = p_{A,B}/N$; thus, propagating the uncertainty, and ignoring any correlation, the variance for the odds ratio \mathcal{O}_B^A is given by

$$\sigma_{\mathcal{O}_B^A}^2 \approx \frac{(\mathcal{O}_B^A)^2}{N} \left(\frac{1}{p_A} + \frac{1}{p_B} \right). \quad (5)$$

III. RESULTS

We analyze 13 events of GWTC-3, using the data available on GWOSC [40, 41], focusing on the ones with the highest total mass ($M > 59.4M_\odot$), and with moderate to high signal-to-noise ratios (SNRs). If $h(\vec{\theta})$ is the GW signal, with $\vec{\theta}$ the binary's parameters, the optimal SNR is defined as $\langle h(\vec{\theta})|h(\vec{\theta}) \rangle^{1/2}$. In particular, we consider events with a network SNR $\rho_{\text{net}} \geq \sqrt{N_d} \times 8^2$, where N_d is the number of interferometers detecting the event, corresponding to at least a signal-to-noise ratio 8 per detector. The waveform models employed include higher-order modes, and we used the ones that are available for all models: $(\ell, m) = (2, 1), (3, 3), (4, 4), (2, -2), (2, -1), (3, -3), (4, -4)$. For SEOBNRv4PHM, the sampling rate must ensure that the Nyquist frequency is larger than the ringdown frequency. For most events, this means that the required sampling rate was higher than the one used for the LVK catalog papers [8, 36]; therefore we estimated the events' power spectral densities (PSDs) in the needed frequency range, using *BayesLine* [42] and the same settings as in Ref. [8, 36]. The new PSDs are released together with this paper [43].

For our analysis runs, the waveform templates are generated starting at $f_{\text{low}} = 20$ Hz, the same lower frequency used for the analysis, for all the models apart from SEOBNRv4PHM, for which the waveform needs to be generated starting from lower frequencies to have all the modes in band [44].

The detector frame masses and spins estimated with the various models are reported in Table I and Table II, respectively, while Table III shows the median value, along with its 90% confidence interval, of the distribution of the logarithm of the samples' likelihood, $\log \mathcal{L}$. In general, we expect that higher values of $\log \mathcal{L}$ correspond to a higher probability for a given model. However, Table III reports the median of the recovered $\log \mathcal{L}$ distribution, therefore, since the shape of the distribution will affect the median value, in some cases the model with the largest $\log \mathcal{L}$ value might not correspond to the model with the largest probability.

Regarding the spins, information is reported through the *effective inspiral spin*

$$\chi_{\text{eff}} = \frac{(m_1 \chi_{1,\parallel} + m_2 \chi_{2,\parallel})}{M}, \quad (6)$$

Event	$\mathcal{M}_c [M_\odot]$					q				
	NRSur	SEOB	IMRX	IMRT	Combined	NRSur	SEOB	IMRX	IMRT	Combined
GW150914	$31.0^{+1.1}_{-1.2}$	$30.6^{+1.6}_{-1.5}$	$30.6^{+1.3}_{-1.6}$	$31.1^{+1.2}_{-1.2}$	$30.9^{+1.3}_{-1.5}$	$0.9^{+0.1}_{-0.2}$	$0.9^{+0.1}_{-0.2}$	$0.9^{+0.1}_{-0.2}$	$0.9^{+0.1}_{-0.2}$	$0.9^{+0.1}_{-0.2}$
GW190519_153544	$66.4^{+6.7}_{-11.6}$	$66.2^{+8.1}_{-12.0}$	$64.6^{+7.8}_{-10.6}$	$67.5^{+7.4}_{-12.5}$	$65.7^{+7.8}_{-11.4}$	$0.6^{+0.3}_{-0.2}$	$0.6^{+0.2}_{-0.2}$	$0.6^{+0.3}_{-0.2}$	$0.6^{+0.2}_{-0.2}$	$0.6^{+0.3}_{-0.2}$
GW190521_074359	$40.4^{+2.0}_{-3.2}$	$40.7^{+2.8}_{-2.7}$	$39.4^{+2.3}_{-2.4}$	$40.8^{+1.9}_{-3.0}$	$40.4^{+2.5}_{-2.9}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$
GW190620_030421	$58.6^{+7.2}_{-10.9}$	$60.3^{+9.8}_{-10.3}$	$58.9^{+9.2}_{-12.9}$	$60.1^{+6.6}_{-11.0}$	$59.5^{+8.2}_{-11.3}$	$0.7^{+0.3}_{-0.3}$	$0.7^{+0.3}_{-0.3}$	$0.6^{+0.3}_{-0.3}$	$0.7^{+0.3}_{-0.3}$	$0.7^{+0.3}_{-0.3}$
GW190630_185205	$29.5^{+1.5}_{-1.8}$	$29.3^{+1.9}_{-1.9}$	$29.4^{+1.7}_{-1.6}$	$29.6^{+1.6}_{-1.8}$	$29.5^{+1.6}_{-1.8}$	$0.7^{+0.3}_{-0.2}$	$0.6^{+0.3}_{-0.2}$	$0.7^{+0.3}_{-0.2}$	$0.7^{+0.3}_{-0.2}$	$0.7^{+0.3}_{-0.2}$
GW190910_112807	$43.3^{+3.6}_{-3.7}$	$43.3^{+3.9}_{-3.7}$	$43.2^{+4.1}_{-4.2}$	$43.5^{+3.9}_{-3.5}$	$43.3^{+3.9}_{-3.8}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$
GW191222_033537	$52.6^{+5.4}_{-6.2}$	$51.6^{+7.3}_{-6.6}$	$51.0^{+6.6}_{-7.0}$	$52.8^{+5.6}_{-5.9}$	$52.2^{+6.1}_{-6.6}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$
GW200112_155838	$33.8^{+2.5}_{-1.9}$	$34.1^{+3.4}_{-2.5}$	$33.8^{+2.6}_{-2.3}$	$34.0^{+2.7}_{-2.0}$	$33.9^{+2.8}_{-2.1}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$
GW200224_222234	$40.3^{+3.9}_{-3.8}$	$40.7^{+3.7}_{-3.8}$	$40.6^{+3.1}_{-3.7}$	$40.3^{+4.4}_{-3.8}$	$40.5^{+3.6}_{-3.8}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$
GW200311_115853	$32.7^{+2.6}_{-2.9}$	$32.6^{+2.8}_{-2.6}$	$32.4^{+2.6}_{-2.7}$	$33.1^{+2.9}_{-3.2}$	$32.6^{+2.8}_{-2.8}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$	$0.8^{+0.2}_{-0.3}$
GW190521	$112.8^{+12.1}_{-13.2}$	$119.3^{+18.9}_{-16.9}$	$104.5^{+16.9}_{-14.4}$	$114.5^{+18.7}_{-14.8}$	$114.5^{+18.9}_{-15.4}$	$0.8^{+0.1}_{-0.3}$	$0.7^{+0.2}_{-0.2}$	$0.7^{+0.3}_{-0.1}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$
GW191109_010717	$60.3^{+5.6}_{-9.4}$	$62.2^{+9.1}_{-7.5}$	$59.4^{+13.5}_{-8.4}$	$66.3^{+6.8}_{-8.4}$	$62.9^{+9.0}_{-8.2}$	$0.7^{+0.2}_{-0.3}$	$0.7^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.8^{+0.2}_{-0.2}$	$0.7^{+0.2}_{-0.3}$
GW200129_065458	$29.9^{+2.5}_{-1.5}$	$31.6^{+0.8}_{-1.3}$	$31.7^{+2.3}_{-3.1}$	$31.4^{+1.8}_{-1.8}$	$30.9^{+2.8}_{-2.4}$	$0.5^{+0.4}_{-0.1}$	$0.8^{+0.2}_{-0.4}$	$0.7^{+0.3}_{-0.3}$	$0.8^{+0.1}_{-0.2}$	$0.6^{+0.4}_{-0.2}$

TABLE I. Median values and their 5% and 95% quantiles from the probability density functions of mass parameters, chirp mass \mathcal{M}_c and mass ratio q , for the different models' posteriors and for the combined one.

with $\chi_{1,\parallel}, \chi_{2,\parallel}$ being the spin components parallel to the angular momentum, and the *effective precession spin*

$$\chi_p = \max \left\{ \chi_{1,\perp}, \frac{q(4q+3)}{4+3q} \chi_{2,\perp} \right\}, \quad (7)$$

where $\chi_{1,\perp}, \chi_{2,\perp}$ are the spin components perpendicular to the angular momentum. Figure 1 shows the posterior probability density of \mathcal{M}_c , q , χ_{eff} , and χ_p for all the events, comparing the posteriors recovered with the different waveform models. We can usually place only weak constraints on χ_p . Thus, its posterior distribution is heavily affected by the prior one, which in turn is determined by the source parameters χ_1, χ_2 , and q , and peaks at non-zero values of χ_p also in the absence of precession. Therefore, recovering a non-zero value of χ_p does not constitute sufficient evidence of precession, but we need to check if the posterior distribution is significantly different from the prior one. This is evaluated through the Jensen-Shannon (JS) divergence [45], which estimates the difference between two probability distributions p_1 and p_2 as

$$D_{\text{JS}} = \frac{1}{2} \left[\sum_x p_1(x) \log \left(\frac{p_1(x)}{m(x)} \right) + \sum_x p_2(x) \log \left(\frac{p_2(x)}{m(x)} \right) \right], \quad (8)$$

with $m(x) = 0.5(p_1(x) + p_2(x))$. Table IV shows the JS divergence values for χ_p posteriors with respect to their

prior distribution, $D_{\text{JS}}^{\chi_p, \text{prior}}$. We also compare our results with the ones from LVK analyses in Table V, where the difference between the posterior distributions is again evaluated as a JS divergence. Furthermore, the probabilities recovered for each model, together with their errors, are reported in Table VI for all the events analyzed.

A. Single events

In this section, we comment on the individual event recoveries with the different waveform models.

1. GW150914

For this event, the parameters and the log-likelihoods recovered are consistent for all four models. The recovered values for the source parameters can be found in Table I-II, and are consistent with the LVK results in [36, 46], as shown in Table V. The probabilities for each approximant are reported in Table VI, where we see a slight preference for the IMRPhenomTPHM model.

Event	χ_{eff}					χ_{p}				
	NRSur	SEOB	IMRX	IMRT	Combined	NRSur	SEOB	IMRX	IMRT	Combined
GW150914	$-0.02^{+0.09}_{-0.11}$	$-0.03^{+0.11}_{-0.12}$	$-0.04^{+0.10}_{-0.14}$	$-0.01^{+0.09}_{-0.10}$	$-0.02^{+0.10}_{-0.12}$	$0.35^{+0.44}_{-0.27}$	$0.33^{+0.43}_{-0.25}$	$0.50^{+0.39}_{-0.39}$	$0.39^{+0.42}_{-0.31}$	$0.39^{+0.44}_{-0.31}$
GW190519_153544	$0.31^{+0.20}_{-0.23}$	$0.34^{+0.21}_{-0.26}$	$0.33^{+0.19}_{-0.26}$	$0.31^{+0.21}_{-0.26}$	$0.33^{+0.20}_{-0.25}$	$0.50^{+0.33}_{-0.32}$	$0.45^{+0.35}_{-0.27}$	$0.47^{+0.36}_{-0.29}$	$0.52^{+0.34}_{-0.34}$	$0.48^{+0.35}_{-0.30}$
GW190521_074359	$0.12^{+0.11}_{-0.13}$	$0.15^{+0.11}_{-0.12}$	$0.08^{+0.12}_{-0.11}$	$0.16^{+0.10}_{-0.14}$	$0.14^{+0.11}_{-0.14}$	$0.44^{+0.34}_{-0.31}$	$0.43^{+0.37}_{-0.29}$	$0.32^{+0.39}_{-0.25}$	$0.45^{+0.36}_{-0.31}$	$0.42^{+0.37}_{-0.30}$
GW190620_030421	$0.32^{+0.22}_{-0.25}$	$0.39^{+0.20}_{-0.22}$	$0.35^{+0.20}_{-0.28}$	$0.37^{+0.19}_{-0.23}$	$0.35^{+0.21}_{-0.25}$	$0.51^{+0.35}_{-0.33}$	$0.46^{+0.35}_{-0.30}$	$0.54^{+0.35}_{-0.36}$	$0.46^{+0.33}_{-0.29}$	$0.49^{+0.35}_{-0.32}$
GW190630_185205	$0.10^{+0.13}_{-0.14}$	$0.10^{+0.14}_{-0.14}$	$0.09^{+0.13}_{-0.13}$	$0.11^{+0.14}_{-0.15}$	$0.10^{+0.13}_{-0.14}$	$0.34^{+0.40}_{-0.25}$	$0.30^{+0.35}_{-0.22}$	$0.30^{+0.38}_{-0.23}$	$0.31^{+0.34}_{-0.23}$	$0.31^{+0.37}_{-0.23}$
GW190910_112807	$-0.02^{+0.17}_{-0.18}$	$0.00^{+0.16}_{-0.20}$	$-0.01^{+0.17}_{-0.20}$	$0.00^{+0.18}_{-0.18}$	$-0.01^{+0.17}_{-0.19}$	$0.43^{+0.42}_{-0.34}$	$0.39^{+0.39}_{-0.32}$	$0.39^{+0.45}_{-0.31}$	$0.40^{+0.42}_{-0.32}$	$0.41^{+0.42}_{-0.32}$
GW191222_033537	$-0.03^{+0.19}_{-0.22}$	$-0.01^{+0.20}_{-0.25}$	$-0.05^{+0.19}_{-0.24}$	$-0.02^{+0.19}_{-0.20}$	$-0.02^{+0.19}_{-0.23}$	$0.41^{+0.44}_{-0.32}$	$0.41^{+0.43}_{-0.32}$	$0.40^{+0.42}_{-0.30}$	$0.42^{+0.43}_{-0.33}$	$0.41^{+0.44}_{-0.32}$
GW200112_155838	$0.04^{+0.15}_{-0.13}$	$0.07^{+0.17}_{-0.15}$	$0.05^{+0.14}_{-0.15}$	$0.06^{+0.16}_{-0.13}$	$0.06^{+0.16}_{-0.14}$	$0.36^{+0.42}_{-0.28}$	$0.35^{+0.41}_{-0.28}$	$0.39^{+0.45}_{-0.30}$	$0.36^{+0.41}_{-0.28}$	$0.36^{+0.43}_{-0.28}$
GW200224_222234	$0.09^{+0.17}_{-0.15}$	$0.11^{+0.14}_{-0.16}$	$0.10^{+0.14}_{-0.16}$	$0.11^{+0.17}_{-0.16}$	$0.10^{+0.15}_{-0.16}$	$0.43^{+0.41}_{-0.34}$	$0.39^{+0.42}_{-0.30}$	$0.48^{+0.39}_{-0.35}$	$0.38^{+0.41}_{-0.30}$	$0.44^{+0.41}_{-0.33}$
GW200311_115853	$-0.02^{+0.16}_{-0.19}$	$-0.01^{+0.15}_{-0.19}$	$-0.04^{+0.16}_{-0.19}$	$0.01^{+0.17}_{-0.21}$	$-0.02^{+0.16}_{-0.19}$	$0.44^{+0.39}_{-0.34}$	$0.40^{+0.43}_{-0.31}$	$0.49^{+0.39}_{-0.37}$	$0.48^{+0.40}_{-0.37}$	$0.46^{+0.41}_{-0.35}$
GW190521	$-0.14^{+0.35}_{-0.37}$	$0.07^{+0.32}_{-0.38}$	$-0.08^{+0.35}_{-0.46}$	$-0.17^{+0.38}_{-0.31}$	$-0.10^{+0.39}_{-0.38}$	$0.75^{+0.20}_{-0.35}$	$0.71^{+0.24}_{-0.37}$	$0.49^{+0.35}_{-0.34}$	$0.76^{+0.19}_{-0.33}$	$0.73^{+0.22}_{-0.37}$
GW191109_010717	$-0.42^{+0.29}_{-0.27}$	$-0.32^{+0.38}_{-0.26}$	$-0.33^{+0.59}_{-0.33}$	$-0.24^{+0.25}_{-0.28}$	$-0.31^{+0.36}_{-0.28}$	$0.60^{+0.29}_{-0.26}$	$0.74^{+0.22}_{-0.36}$	$0.60^{+0.31}_{-0.35}$	$0.85^{+0.12}_{-0.33}$	$0.75^{+0.21}_{-0.37}$
GW200129_065458	$-0.01^{+0.14}_{-0.11}$	$0.07^{+0.09}_{-0.04}$	$0.10^{+0.15}_{-0.18}$	$0.07^{+0.13}_{-0.13}$	$0.04^{+0.18}_{-0.16}$	$0.86^{+0.12}_{-0.35}$	$0.28^{+0.52}_{-0.13}$	$0.82^{+0.15}_{-0.39}$	$0.48^{+0.38}_{-0.34}$	$0.83^{+0.14}_{-0.41}$

TABLE II. Median values and their 5% and 95% quantiles from the probability density functions of spin parameters, effective-spin χ_{eff} and spin-precessing parameter χ_{p} , for the different models' posteriors and for the combined one.

2. *GW190519_153544*

In this case, data show a preference for **IMRPhenomXPHM** (see Table VI), although parameter estimates and log-likelihood values are consistent for all the models. As shown in Table II, we find support for positive, non-zero values of χ_{eff} . This is consistent with the results reported in Ref. [36].

3. *GW190521*

GW190521 is the most massive event detected so far, and one among the ones with the strongest signature of higher-order modes in the signal [47, 48]. The consequently high values needed for the prior on chirp mass, combined with the employed prior on mass ratio, cause potential issues with the **IMRPhenomTPHM** model since the computed peak frequency for the $\ell, m = (2, 2)$ mode might be below the 20 Hz low-frequency cutoff used for our analysis. To avoid this issue, for this event, we adjust the prior on mass ratio such that $q \leq 2$. The recovered values for mass and spin parameters are reported in Table I and Table II respectively. They are consistent with the results in Ref. [36], where, however, only the **IMRPhenomXPHM** and **SEOBNRv4PHM** ap-

proximants were used¹, and with the **NRSur7dq4** results first shown in the discovery paper [47]. We find evidence of precession for the **NRSur7dq4**, **SEOBNRv4PHM**, and **IMRPhenomTPHM** models, cf. Table IV. The probabilities for the different approximants are shown in Table VI: the **IMRPhenomTPHM** model is slightly favored over the other ones, while **IMRPhenomXPHM** is strongly disfavored. Interestingly, these findings are consistent with the fact that **NRSur7dq4**, **SEOBNRv4PHM**, and **IMRPhenomTPHM** models include precession in the description of the ringdown, while **IMRPhenomXPHM** does not; cf. [17, 18]. Nevertheless, the extremely short duration of this event and the lack of the inspiral part of the signal make it difficult to draw clear conclusions.

Many works investigated this event from different perspectives and explored the possible processes that lead to the formation of such a system. One of the most investigated hypotheses is the presence of eccentricity [49–51], which could mimic precession [52, 53]. Multiple alternative scenarios that could lead to the emission of this signal have been proposed, like dynamical capture in hyperbolic orbits [54], a primordial BH merger [55], and a

¹ In Ref. [36], further analyses computed the precession SNR to be too small to claim the presence of strong evidence for precession.

Event	log \mathcal{L}				
	NRSur	SEOBS	IMRX	IMRT	Combined
GW150914	322.2 ^{+2.7} _{-4.3}	321.6 ^{+2.5} _{-4.1}	322.2 ^{+2.8} _{-4.0}	322.4^{+2.6}_{-4.4}	322.2 ^{+2.7} _{-4.3}
GW190519_153544	114.6 ^{+3.7} _{-4.9}	115.4^{+3.7}_{-5.3}	115.1 ^{+3.3} _{-5.1}	114.6 ^{+3.2} _{-5.2}	115.0 ^{+3.5} _{-5.1}
GW190521_074359	320.0 ^{+3.5} _{-4.8}	321.3^{+3.2}_{-5.1}	319.7 ^{+3.1} _{-4.4}	320.6 ^{+3.4} _{-4.6}	320.6 ^{+3.5} _{-4.8}
GW190620_030421	64.1 ^{+3.9} _{-5.3}	64.0 ^{+4.1} _{-5.6}	63.7 ^{+3.6} _{-5.4}	64.2^{+3.8}_{-5.6}	64.0 ^{+3.9} _{-5.5}
GW190630_185205	117.7^{+3.1}_{-5.1}	116.8 ^{+3.2} _{-5.1}	116.9 ^{+3.1} _{-4.9}	117.7 ^{+3.1} _{-5.0}	117.4 ^{+3.2} _{-5.1}
GW190910_112807	90.5 ^{+3.3} _{-4.6}	90.8^{+3.9}_{-4.6}	90.4 ^{+3.1} _{-4.5}	90.4 ^{+3.7} _{-4.5}	90.5 ^{+3.5} _{-4.6}
GW191222_033537	70.0 ^{+2.5} _{-4.1}	69.5 ^{+2.5} _{-4.0}	69.3 ^{+2.5} _{-4.0}	70.1^{+2.5}_{-4.1}	69.8 ^{+2.6} _{-4.1}
GW200112_155838	166.2 ^{+2.9} _{-4.4}	165.5 ^{+2.8} _{-4.6}	165.6 ^{+2.7} _{-4.4}	166.4^{+2.9}_{-4.4}	166.0 ^{+2.9} _{-4.5}
GW200224_222234	188.1 ^{+3.6} _{-4.5}	188.0 ^{+2.7} _{-4.4}	188.6^{+3.3}_{-4.6}	187.4 ^{+2.7} _{-4.5}	188.1 ^{+3.3} _{-4.5}
GW200311_115853	145.4 ^{+2.7} _{-4.2}	146.0 ^{+2.6} _{-4.2}	146.2^{+2.5}_{-4.3}	145.6 ^{+2.8} _{-4.2}	145.9 ^{+2.7} _{-4.3}
GW190521	88.0 ^{+4.2} _{-5.6}	87.4 ^{+4.2} _{-5.4}	83.6 ^{+4.3} _{-4.3}	88.4^{+3.6}_{-5.5}	87.8 ^{+4.1} _{-5.8}
GW191109_010717	133.3 ^{+3.9} _{-6.2}	136.4^{+5.6}_{-6.9}	132.2 ^{+6.9} _{-6.6}	135.9 ^{+5.4} _{-6.7}	135.8 ^{+5.9} _{-6.9}
GW200129_065458	347.2^{+4.4}_{-7.1}	341.0 ^{+2.6} _{-3.8}	345.3 ^{+4.7} _{-6.4}	341.1 ^{+5.3} _{-4.6}	346.1 ^{+4.8} _{-7.0}

TABLE III. Median values and their 5% and 95% quantiles from the probability density functions of the recovered log \mathcal{L} with the different models and for the combined results. For each event, the highest value of log \mathcal{L} is marked in bold.

high-mass BH-disk system [56]. In Ref. [57], an analysis of this event with a population-based prior led to the conclusion that neither of the component masses lies in the pair-instability supernova mass gap. In Ref. [58], the use of a high-mass prior showed the possibility of GW190521 being an intermediate-mass-ratio BBH merger. However, a further investigation carried out in Ref. [59], where different precession prescriptions and higher-order-mode contents are investigated with the IMRPhenomXPHM and IMRPhenomTPHM models, showed that, despite the presence of a multimodal likelihood for the mass ratio parameter, the peaks are characterized by very different probabilities. The parameters recovered by our analysis are consistent with both the IMRPhenomXPHM and IMRPhenomTPHM results in Ref. [59], when using models with the same settings.

4. GW190521_074359

This event shows a preference for the SEOBNRv4PHM approximant, although the recovered parameters and log-likelihood values are similar for all four models. Also in this case, our results are consistent with the ones in the LVK papers [36], and we find no evidence of precession.

5. GW190620_030421

The LVK studies report this source as a BBH binary with high effective spin χ_{eff} . In our re-analysis, we find all the waveform families to perform comparably and return consistent estimates of parameters as well as the values of log-likelihood. Moreover, the existing LVK analyses on this event with IMRPhenomXPHM and SEOBNRv4PHM return consistent results with ours. We also find support for positive values of χ_{eff} , as shown in Table II. The estimates of intrinsic parameters from different models are consistent with each other, however, from the values of posterior probability (see Table VI), NRSur7dq4 seems to be the most favored.

6. GW190630_185205

We find consistent estimates of parameters and log-likelihoods among all models compared, and no evidence for spin. Among the four models considered, NRSur7dq4 and IMRPhenomTPHM seem to be most preferred by the data, with almost the same probability.

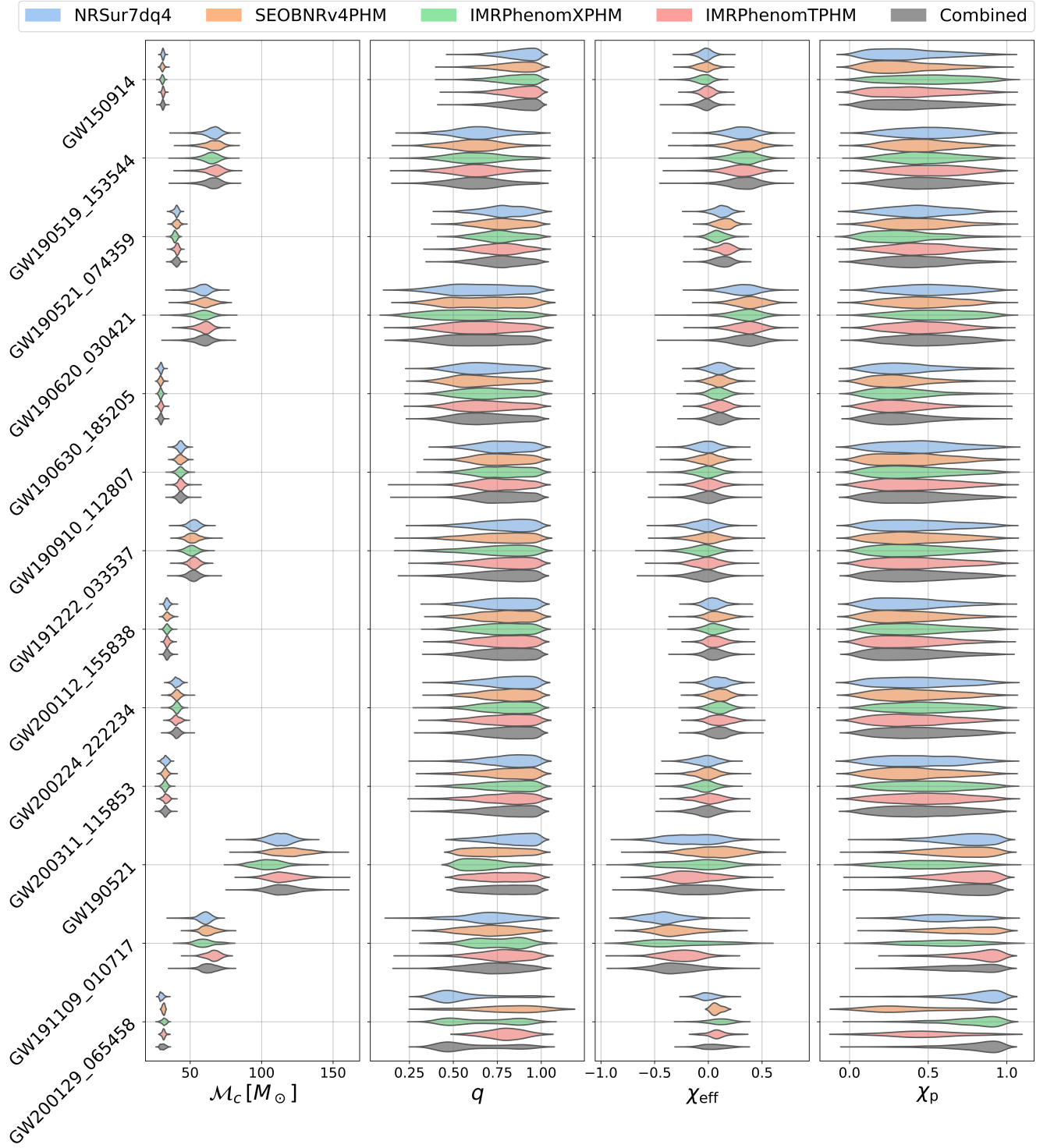


FIG. 1. Posterior probability densities for \mathcal{M}_c , q , χ_{eff} , and χ_p as recovered with the different waveform approximants and for the combined posterior, for all the events analyzed.

Event	NRSur	SEOB	IMRX	IMRT	Combined
GW150914	0.008	0.010	0.050	0.017	0.015
GW190519_153544	0.010	0.017	0.010	0.011	0.011
GW190521_074359	0.037	0.029	0.027	0.029	0.029
GW190620_030421	0.010	0.006	0.016	0.012	0.006
GW190630_185205	0.030	0.067	0.049	0.065	0.050
GW190910_112807	0.023	0.012	0.009	0.014	0.014
GW191222_033537	0.012	0.011	0.014	0.012	0.011
GW200112_155838	0.012	0.015	0.011	0.014	0.012
GW200224_222234	0.008	0.011	0.024	0.010	0.010
GW200311_115853	0.026	0.018	0.041	0.038	0.031
<i>GW190521</i>	<i>0.243</i>	<i>0.158</i>	<i>0.007</i>	<i>0.264</i>	<i>0.202</i>
<i>GW191109_010717</i>	<i>0.095</i>	<i>0.227</i>	<i>0.070</i>	<i>0.422</i>	<i>0.243</i>
<i>GW200129_065458</i>	<i>0.459</i>	<i>0.005</i>	<i>0.330</i>	<i>0.051</i>	<i>0.378</i>

TABLE IV. $D_{JS}^{\chi_p, \text{prior}}$ values in bit computed between the posterior of χ_p obtained with our analysis and the prior distribution conditioned to χ_{eff} , for the posteriors recovered with the different waveforms and the combined one. Events for which we find values of χ_p significantly different from the prior are marked in italic.

7. *GW190910_112807*

This event again returns very consistent estimates of log-likelihoods and intrinsic parameters among the different models. In particular, we find no evidence for spins. From the values of posterior probabilities supported by all waveforms, we also note that the data have an almost equal preference for all models.

8. *GW191109_010717*

We find SEOBNRv4PHM to be the most favored model, as reported in Table VI. We also recover a high probability for IMRPhenomTPHM, while NRSur7dq4 and IMRPhenomXPHM are strongly disfavored. We find evidence of non-zero χ_p with both SEOBNRv4PHM and IMRPhenomTPHM, but not for the other two models, as shown in Table II and IV. For all models, we find significant support for negative values of χ_{eff} , confirming the results in Refs. [8, 60]. In the latter, the possibility of formation by dynamical capture for the binary generating this event is discussed. However, GW191109_010717 was among the O3 events that required data mitiga-

tion due to the presence of glitches. In particular, GW191109_010717 was affected by a glitch in both the detectors online at the time of the event, in the frequency range 25-45 Hz for Hanford and 20-32 Hz for Livingston. As shown in Ref. [61], different deglitching procedures influence the posteriors obtained for both χ_{eff} and χ_p . In particular, if the Livingston data are analyzed only for frequencies larger than 40 Hz, the support for negative χ_{eff} disappears. However, this result is not sufficient to label the negative support of χ_{eff} as a noise artifact, since most of the spin information comes from low frequencies, and, being GW191109_010717 already a signal with a short inspiral, removing the low-frequency part discards most of the information, yielding non-informative results. The presence of glitches overlapping a significant part of the inspiral for both the detectors is also regarded as the most likely cause for deviations from general relativity found for this event by some LVK pipelines [10].

9. *GW191222_033537*

Although the returned parameter estimates, as well as log-likelihood values, are quite similar, IMRPhenomTPHM seems to be the most favored model (see Table VI), while the least favored model is IMRPhenomXPHM. We find no evidence for spins.

10. *GW200112_155838*

We recover similar probabilities for all the approximants, with SEOBNRv4PHM slightly disfavored and IMRPhenomTPHM slightly favored, as shown in Table VI. Consistently, we find no significant difference between the recovered parameters and log-likelihoods for the different waveforms. The IMRPhenomXPHM and SEOBNRv4PHM posteriors estimated by our study are consistent with the LVK ones.

11. *GW200129_065458*

We find a strong preference for NRSur7dq4 and IMRPhenomXPHM, while the probability for SEOBNRv4PHM and IMRPhenomTPHM is close to zero. This discrepancy is reflected in the posteriors of χ_p , with NRSur7dq4 and IMRPhenomXPHM finding strong evidence for high χ_p values, cf. Table IV, while for the other two models results are dominated by the prior. This is consistent with what was found in the LVK GWTC-3 analysis [8], where IMRPhenomXPHM recovers χ_p and SEOBNRv4PHM does not. In Ref. [62], strong evidence for precession was found when analyzing this event with the NRSur7dq4 model. For this event, precession was measured also in Ref. [63], where the recoil velocity was also estimated. The main difference between these two works and the LVK analysis [8], which did not find conclusive evidence of preces-

Event	IMRPhenomXPHM				SEOBNRv4PHM			
	$D_{JS}^{\chi_p}$	$D_{JS}^{\chi_{\text{eff}}}$	$D_{JS}^{\mathcal{M}_c}$	D_{JS}^q	$D_{JS}^{\chi_p}$	$D_{JS}^{\chi_{\text{eff}}}$	$D_{JS}^{\mathcal{M}_c}$	D_{JS}^q
GW150914	0.006	0.001	0.007	0.005	0.005	0.076	0.032	0.052
GW190519_153544	0.002	0.001	0.001	0.001	0.011	0.002	0.014	0.007
GW190521_074359	0.007	0.004	0.005	0.001	0.024	0.033	0.005	0.007
GW190620_030421	0.005	0.001	0.007	0.001	0.001	0.001	0.006	0.001
GW190630_185205	0.002	0.001	0.004	0.002	0.010	0.022	0.025	0.015
GW190910_112807	0.002	0.000	0.002	0.001	0.007	0.002	0.015	0.009
GW200112_155838	0.002	0.007	0.004	0.003	0.009	0.018	0.014	0.026
GW200224_222234	0.001	0.003	0.006	0.001	0.004	0.016	0.018	0.006
GW200311_115853	0.002	0.001	0.001	0.001	0.006	0.006	0.015	0.014
GW190521	0.019	0.003	0.066	0.075	0.020	0.003	0.018	0.035
GW191109_010717	0.024	0.006	0.012	0.006	0.029	0.011	0.016	0.007
GW200129_065458	0.003	0.010	0.005	0.008	0.139	0.046	0.137	0.141

TABLE V. Values of Jensen-Shannon divergence for χ_p , χ_{eff} , \mathcal{M}_c , and q , computed between the posteriors recovered by our analysis and the LVK ones [8, 36] for the available waveforms, IMRPhenomXPHM and SEOBNRv4PHM.

Event	NRSur	SEOB	IMRX	IMRT
GW150914	27.55 ± 0.7	16.22 ± 0.8	23.34 ± 0.7	32.88 ± 0.7
GW190519_153544	20.82 ± 0.6	20.95 ± 0.6	40.87 ± 0.5	17.35 ± 0.6
GW190521_074359	14.76 ± 1.2	40.50 ± 1.0	17.53 ± 1.2	27.22 ± 1.1
GW190620_030421	32.98 ± 0.6	19.48 ± 0.6	20.22 ± 0.6	27.32 ± 0.6
GW190630_185205	33.79 ± 0.6	15.36 ± 0.6	18.90 ± 0.6	31.95 ± 0.6
GW190910_112807	22.86 ± 0.6	25.92 ± 0.6	27.85 ± 0.6	23.37 ± 0.6
GW191222_033537	28.11 ± 0.5	20.58 ± 0.6	18.78 ± 0.6	32.53 ± 0.5
GW200112_155838	30.56 ± 0.6	15.61 ± 0.6	19.82 ± 0.6	34.01 ± 0.5
GW200224_222234	21.82 ± 0.6	23.39 ± 0.6	40.43 ± 0.5	14.36 ± 0.7
GW200311_115853	15.68 ± 0.6	27.70 ± 0.6	35.69 ± 0.6	20.93 ± 0.6
<i>GW190521</i>	<i>31.78 ± 0.6</i>	<i>26.39 ± 0.6</i>	<i>4.60 ± 0.7</i>	<i>37.23 ± 0.5</i>
<i>GW191109_010717</i>	<i>7.54 ± 1.6</i>	<i>62.29 ± 1.0</i>	<i>5.06 ± 1.7</i>	<i>25.11 ± 1.5</i>
<i>GW200129_065458</i>	<i>46.94 ± 1.4</i>	<i>0.66^{+1.9}_{-0.66}</i>	<i>51.14 ± 1.3</i>	<i>1.25^{+1.9}_{-1.25}</i>

TABLE VI. Probability percentages, including errors, for each model in the different events. Events that strongly favor or disfavor some of the models are marked in italic.

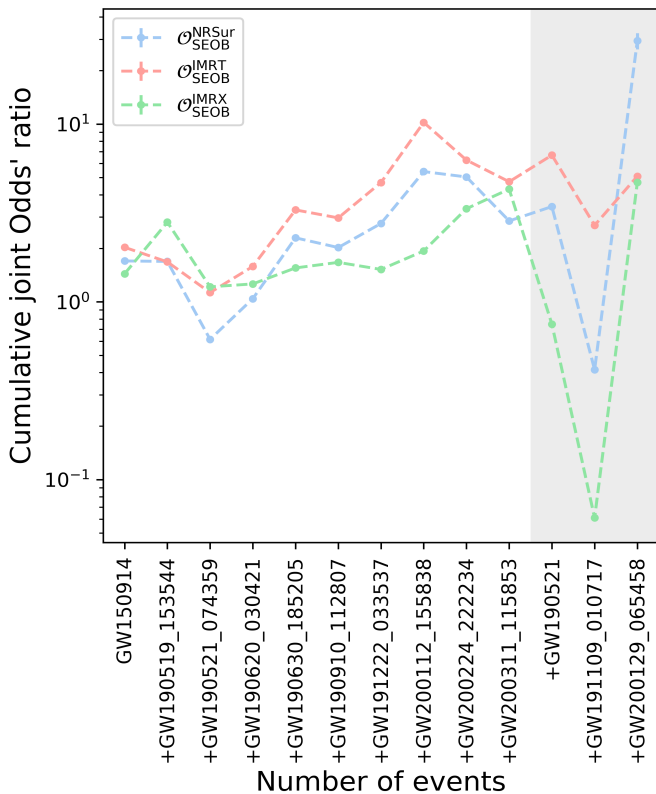


FIG. 2. Evolution of the joint odds ratio for each approximant, with respect to `SEOBNRv4PHM`, as events are added; for any one event shown on the x-axis, the joint odds ratio is calculated from all the events occurring to the left of that event. The events in the gray-shaded area are affected by possible data quality issues. Note that the symmetric error bars 1σ are included in the data points but too small to be discernible.

sion, is that in the latter data were analyzed only with the `IMRPhenomXPHM` and `SEOBNRv4PHM` approximants. In Refs. [62] and [63], the `NRSur7dq4` model was used, because, being generated from NR simulations, it is expected to be more accurate, as shown by the mismatch computation in Ref. [62]. However, in our study, we do not find an overall preference for `NRSur7dq4`. `GW200129` data were affected by a glitch overlapping the event in the Livingston detector [61], therefore, in our analysis, we used the deglitched data, as was done in Ref. [8]. Reference [64] explores the influence of data quality issues for this event, finding that the evidence for precession comes exclusively from the Livingston strain of data between 20–50 Hz, where such issues are present.

12. `GW200224_222234`

For this event the recovered parameters and log-likelihood values are consistent for the different waveforms. We find a slight preference for `IMRPhenomXPHM`, cf. Table VI. Our results for both `IMRPhenomXPHM` and

`SEOBNRv4PHM` are consistent with the `LVK` ones. We do not find support for precession.

13. `GW200311_115853`

Specific to this event, we find no evidence of spin and consistent source parameters and log-likelihood estimates among all models. However, `IMRPhenomXPHM` seems to be the most favored approximant by the event (cf. Table VI).

B. Combined events

Figure 3 shows the cumulative joint odds ratio as a function of the number of events, while Table VII reports the odds ratio values obtained by combining information from all the sources analyzed. We do not find a specific approximant being preferred or disfavored consistently for all the events. Combining results for all the 13 sources, the `NRSur7dq4` model results favored with respect to `SEOBNRv4PHM`, with an odds ratio of 29.43. However, this value is dominated by the results for `GW200129_065458`, and without this event the odds ratio becomes 0.46. This is unexpected, because `NRSur7dq4`, being fully informed by NR simulations, is assumed to be the most accurate model and therefore to describe the data best. Table VII shows also how odds ratios change with the three events with a strong preference for one of the models: while `GW200129_065458` is responsible for `NRSur7dq4` being favored over `SEOBNRv4PHM`, `GW191109_010717`, which instead finds a significant preference for `SEOBNRv4PHM` and `IMRPhenomTPHM`, balances this result; if we do not take `GW191109_010717` into account, `NRSur7dq4` and `IMRPhenomXPHM` are strongly favored over `SEOBNRv4PHM`, with an odds ratio of 243.31 and 57.84 respectively. In addition, without this event, $\mathcal{O}_{\text{SEOB}}^{\text{IMRT}} = 12.62$, and $\mathcal{O}_{\text{SEOB}}^{\text{NRSur}} = 19.27$. Similarly, the results from `GW190521` heavily influence the final odds ratio for `IMRPhenomXPHM`: if we do not include this event, we obtain $\mathcal{O}_{\text{SEOB}}^{\text{IMRX}} = 26.99$. Without these three sources, we find no significant preference for any of the models.

We look for possible trends for the preference of given approximants with respect to the binary parameters, which would point to the waveforms with the best description for specific regions of the parameter space. Figure 3 shows the probabilities recovered for the different models as a function of the source’s mass and spin parameters, and the network optimal matched-filter SNR, as computed by the parameter estimation analyses in the catalog papers [8, 36]. We do not find any trends with respect to the binary parameters or the signal SNR.

Interestingly, we find that for all the events that show a strong preference for one of the models, i.e., `GW190521`, `GW191109_010717`, and `GW200129_065458`, the preferred models are not the same, but in each case are the

	$\mathcal{O}_{\text{SEOB}}^{\text{NRSur}}$	$\mathcal{O}_{\text{SEOB}}^{\text{IMRX}}$	$\mathcal{O}_{\text{SEOB}}^{\text{IMRT}}$	$\mathcal{O}_{\text{IMRX}}^{\text{NRSur}}$	$\mathcal{O}_{\text{IMRT}}^{\text{NRSur}}$	$\mathcal{O}_{\text{IMRT}}^{\text{IMRX}}$
All events	29.43 ± 1.11	4.70 ± 0.07	5.09 ± 0.08	6.26 ± 0.11	5.78 ± 0.10	0.92 ± 0.01
No GW200129_065458	0.42 ± 0.00	0.06 ± 0.00	2.69 ± 0.03	6.82 ± 0.12	0.15 ± 0.00	0.02 ± 0.00
No GW190521	24.44 ± 0.84	26.99 ± 0.97	3.61 ± 0.05	0.91 ± 0.01	6.77 ± 0.12	7.48 ± 0.14
No GW191109_010717	243.31 ± 26.35	57.84 ± 3.05	12.62 ± 0.31	4.21 ± 0.06	19.27 ± 0.59	4.58 ± 0.07
Without all three	2.85 ± 0.03	4.30 ± 0.06	4.74 ± 0.07	0.66 ± 0.00	0.60 ± 0.00	0.91 ± 0.01

TABLE VII. Joint odds ratios including errors. We report results for all the events combined and results without the events that show a strong preference for some models.

ones that recover precession. This is particularly evident in the case of GW190521, where `IMRPhenomXPHM` does not recover evidence of precession and has a probability only of roughly 4%, while the other models, which show evidence supporting non-zero values of χ_p , have all a probability $\sim 30\%$. Although, as mentioned, the results for these events might be biased by their short duration or potential data quality issues, the fact that a given model recovers precession better than another one systematically implies a higher probability. Evidence for this behavior is supported by the fact that the preferred models are different for the three events, leaving the recovery of precession as the only element systematically connected to higher probability values.

IV. SUMMARY

We analyzed the 13 events with the highest mass and moderate to high SNR among the ones detected so far by Advanced LIGO and Advanced Virgo, using the “hypermodels” technique developed in Ref. [20]. This method allows us to sample directly over different waveform approximants, in order to determine which one is favored by the data. We analyzed data with four different approximants, all including precession and higher-order modes: `NRSur7dq4`, `SEOBNRv4PHM`, `IMRPhenomXPHM`, and `IMRPhenomTPHM`. For each event, we recover the source parameters, finding both mass and spin parameters to be in agreement with the LVK results, cf. Table V. For three events, GW191109_010717, GW200129_065458, and GW190521, we recover non-zero values for the effective precession spin parameter, with a distribution significantly different from the prior one. These events are also the ones for which we find a strong preference for some models over the other ones, although the preferred approximants are different. GW191109_010717 shows a strong preference for `SEOBNRv4PHM`, with `NRSur7dq4` and `IMRPhenomXPHM` being disfavored. On the other hand, for GW200129_065458, `NRSur7dq4` and `IMRPhenomXPHM` are strongly favored, and the probability for `SEOBNRv4PHM` and `IMRPhenomTPHM` is close to

zero. Finally, GW190521 recovers a very low probability, roughly 4%, for `IMRPhenomXPHM`, while the other models do not show significant differences among them. However, GW191109_010717 and GW200129_065458 data were affected by glitches [61], and the short duration of GW190521 implies that we could not see its inspiral phase; therefore, we cannot draw clear conclusions about these events. Nonetheless, we systematically find that the models recovering evidence for non-zero values of χ_p are the ones with the higher probabilities. For all the other events, we recover only slight preferences for a given approximant, with the recovered parameters’ posteriors and log-likelihoods being similar. Overall, we do not find one model to be consistently preferred over the others. This is unexpected, considering that we included `NRSur7dq4` in the analysis, which is predicted to be the most accurate model for high-mass signals, being interpolated from NR simulations. The odds ratios combined over all the sources show `NRSur7dq4` being favored over `SEOBNRv4PHM`, with $\mathcal{O}_{\text{SEOB}}^{\text{NRSur}} = 29.43$, while for `IMRPhenomXPHM` and `IMRPhenomTPHM` we find $\mathcal{O}_{\text{SEOB}}^{\text{IMRX}} = 4.70$ and $\mathcal{O}_{\text{SEOB}}^{\text{IMRT}} = 5.09$ respectively. However, this result is mostly determined by GW200129_065458, for which `SEOBNRv4PHM` and `IMRPhenomTPHM` probabilities are close to zero. If we remove this event from the combined odds ratio calculation, we obtain $\mathcal{O}_{\text{SEOB}}^{\text{NRSur}} = 0.42$. Finally, if we do not take into account the three sources favoring one of the approximants, we find no significant preference for any of the models.

ACKNOWLEDGMENTS

We thank Vijay Varma for his help with the `NRSur7dq4` model. A.P. and C.V.D.B. are supported by the research programme of the Netherlands Organisation for Scientific Research (NWO). A.S. thanks the Alexander von Humboldt foundation in Germany for a Humboldt fellowship for postdoctoral researchers. This work was co-funded by the European Union (ERC, SMART, project number 101076369). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect

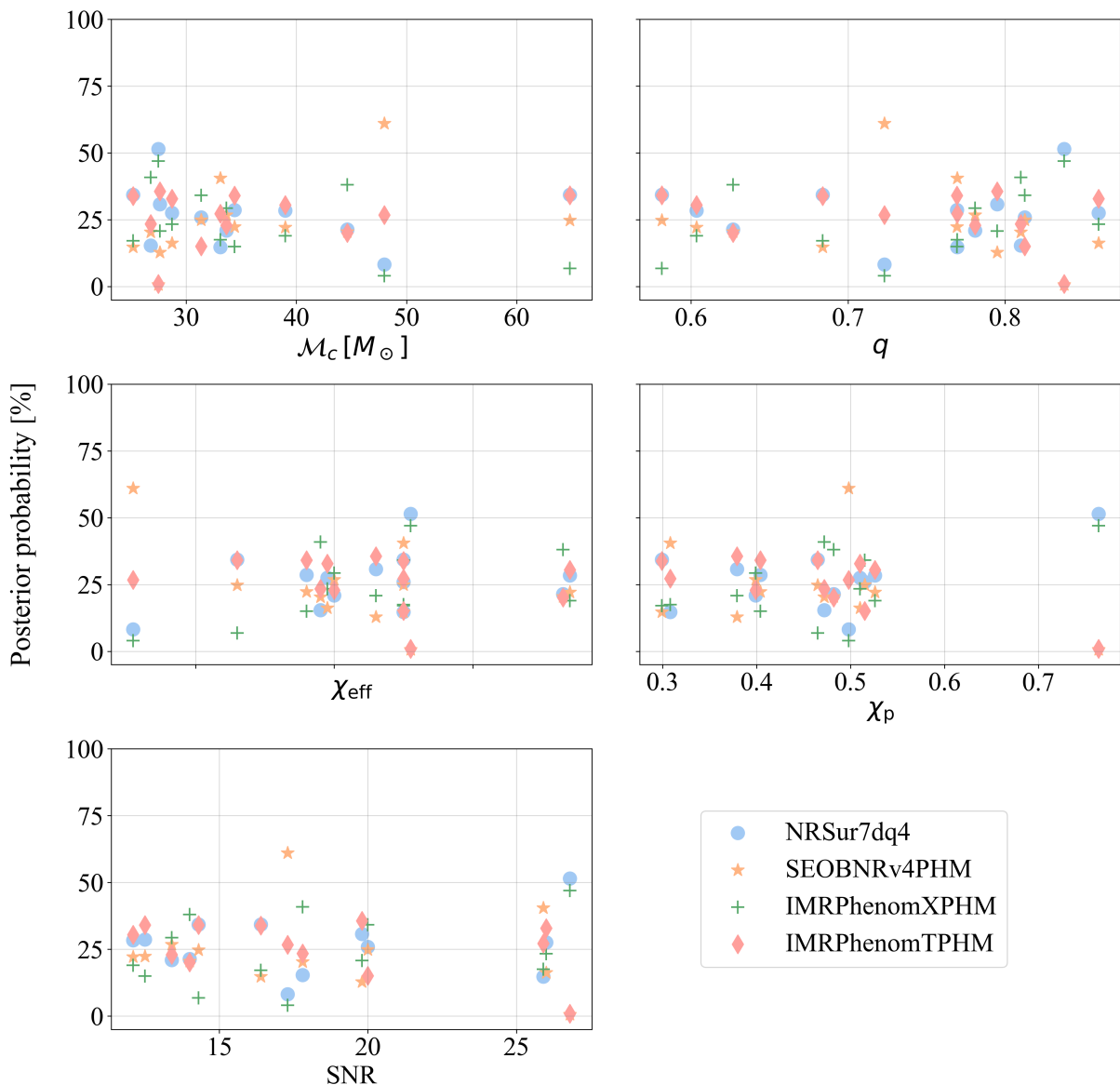


FIG. 3. Posterior probability for the different approximants as a function of the LVK estimated values of \mathcal{M}_c (top left panel), q (top right), χ_{eff} (middle left), χ_p (middle right), and SNR (bottom left panel).

those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

We are grateful for computational resources provided by Cardiff University, and funded by an STFC grant (ST/I006285/1) supporting UK Involvement in the Operation of Advanced LIGO. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO

Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support

for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

Appendix A: Injection runs

In addition to the analysis of real GW events, we want to prove in the following the validity of the method through an injection study. For this purpose, we perform a hypermodels analysis with the same waveform approximants and settings previously described, to analyze simulated signals in zero noise. The details of the injections are given in Table VIII. Injection 1 and injection 2 are produced using the maximum-likelihood parameters and approximants recovered from the analyses of GW190521 and GW191109_010717, respectively. Injection 3 and 4 are generated with the maximum-likelihood

parameters of GW200129_065458 using `IMRPhenomXPHM` and `NRSur7dq4`, which are the models with the highest recovered probability and likelihood, respectively. For the other injections we employed the maximum-likelihood mass values recovered for GW190519_153544, a fixed luminosity distance, and two different values of spin magnitudes and inclinations, considering injections both with `IMRPhenomXPHM` and `NRSur7dq4`. Figure 4 shows the probability density distributions of the recovered log-likelihoods for the different models, together with their percentage probabilities, including errors. In most cases we clearly recover the highest probability for the injected model. When the most favored model is not the injected one, however, the probability of the injected model is very close to the highest one. This is likely due to the fact that the two waveform descriptions are very similar, and the injected model is guaranteed to provide the best fit only at the injection point. To further understand why the injected model in some cases is not the most favored one, a detailed analysis of different ingredients for all employed waveform models would be required, which is however outside the scope of this paper. From the statistical point of view, the injection study indicates that our uncertainty on the odds might not measure the full uncertainty. A validation of the uncertainty estimates would need multiple runs on the same data set. We also note that, in order to validate the method, we performed these analyses in zero noise: in real-events analysis, the presence of noise and noise fluctuations will affect the differences between the evidences.

-
- [1] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
 - [2] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **6**, 041015 (2016), [Erratum: *Phys.Rev.X* **8**, 039903 (2018)], arXiv:1606.04856 [gr-qc].
 - [3] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 221101 (2016), [Erratum: *Phys.Rev.Lett.* **121**, 129902 (2018)], arXiv:1602.03841 [gr-qc].
 - [4] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
 - [5] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **9**, 011001 (2019), arXiv:1805.11579 [gr-qc].
 - [6] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi-GBM, INTEGRAL), *Astrophys. J. Lett.* **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].
 - [7] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **121**, 161101 (2018), arXiv:1805.11581 [gr-qc].
 - [8] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03606 [gr-qc].
 - [9] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03634 [astro-ph.HE].
 - [10] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].
 - [11] M. Pürrer and C.-J. Haster, *Phys. Rev. Res.* **2**, 023151 (2020), arXiv:1912.10055 [gr-qc].
 - [12] J. Aasi *et al.* (LIGO Scientific), *Class. Quant. Grav.* **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
 - [13] F. Acernese *et al.* (VIRGO), *Class. Quant. Grav.* **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
 - [14] C. Hoy, *Phys. Rev. D* **106**, 083003 (2022), arXiv:2208.00106 [gr-qc].
 - [15] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. Research*, **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].
 - [16] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].
 - [17] H. Estellés, M. Colleoni, C. García-Quirós, S. Husa, D. Keitel, M. Mateu-Lucena, M. d. L. Planas, and A. Ramos-Buades, *Phys. Rev. D* **105**, 084040 (2022), arXiv:2105.05872 [gr-qc].
 - [18] S. Ossokine *et al.*, *Phys. Rev. D* **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].
 - [19] T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O’Shaughnessy, and R. Smith, (2023), arXiv:2309.14473 [gr-qc].
 - [20] G. Ashton and T. Dietrich, *Nature Astron.* **6**, 961 (2022), arXiv:2111.09214 [gr-qc].
 - [21] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].

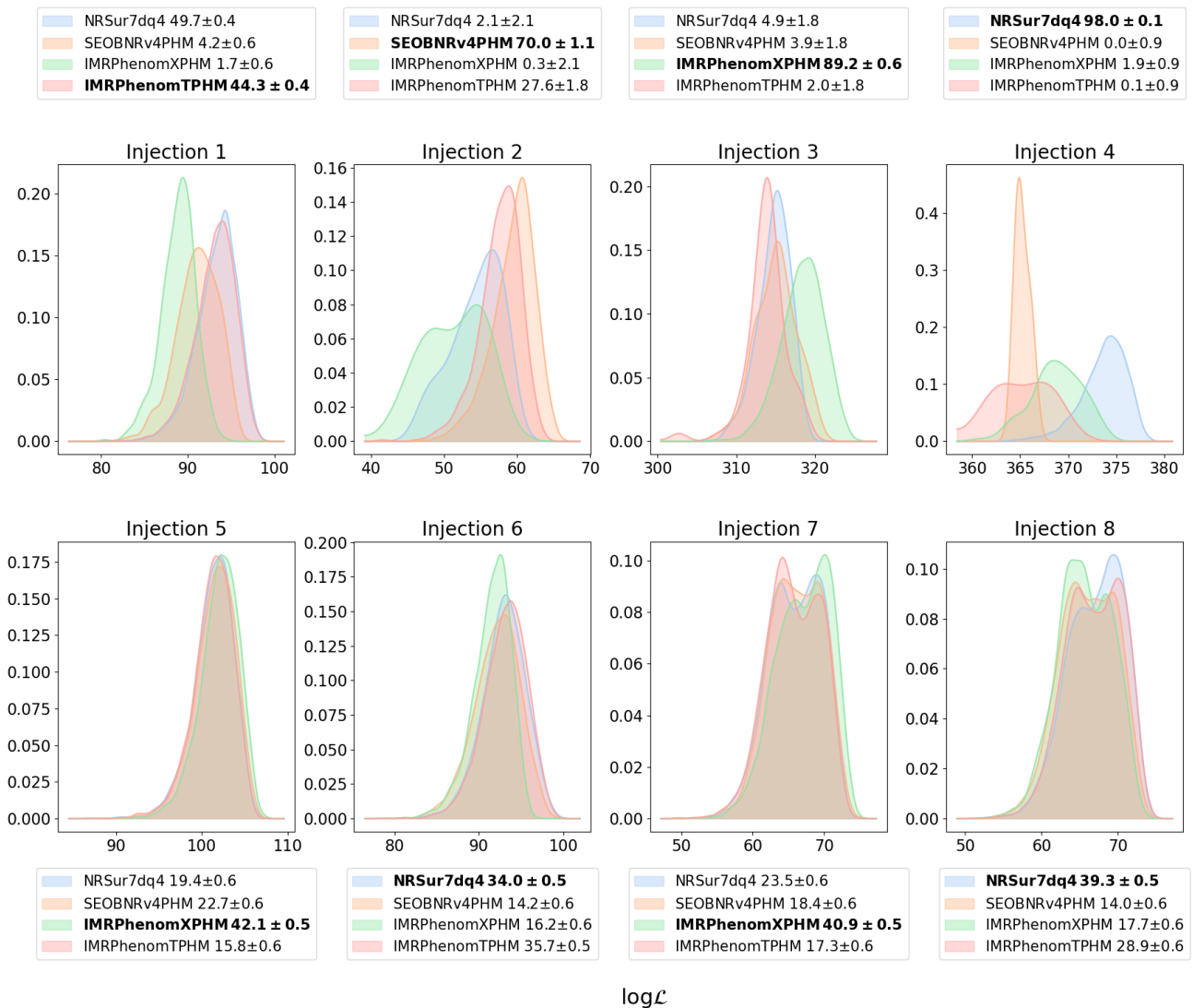


FIG. 4. Probability density distributions for $\log \mathcal{L}$ for the different models considered in the analysis. The legend reports the recovered probability percentages, including error; the model marked in bold is the one used for the injection.

- [22] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].
- [23] L. Blanchet, G. Faye, B. R. Iyer, and S. Sinha, Class. Quant. Grav. **25**, 165003 (2008), [Erratum: Class.Quant.Grav. 29, 239501 (2012)], arXiv:0802.1249 [gr-qc].
- [24] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, Phys. Rev. D **96**, 024058 (2017), arXiv:1705.07089 [gr-qc].
- [25] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014), arXiv:1308.3565 [gr-qc].
- [26] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Phys. Rev. Lett. **115**, 121102 (2015), arXiv:1502.07758 [gr-qc].
- [27] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, Phys. Rev. D **98**, 084028 (2018), arXiv:1803.10701 [gr-qc].
- [28] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2014), arXiv:1307.6232 [gr-qc].
- [29] S. Babak, A. Taracchini, and A. Buonanno, Phys. Rev. D **95**, 024010 (2017), arXiv:1607.05661 [gr-qc].
- [30] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].
- [31] P. Schmidt, M. Hannam, and S. Husa, Phys. Rev. D **86**, 104063 (2012), arXiv:1207.3088 [gr-qc].

	Model	$\mathcal{M}_c [M_\odot]$	q	a_1	θ_1 [rad]	a_2	θ_2 [rad]	D_L [Mpc]
Injection 1	IMRPhenomTPHM	108.79	0.92	0.97	2.59	0.93	1.66	2751.72
Injection 2	SEOBNRv4PHM	71.32	0.54	0.99	1.12	0.81	1.96	3488.44
Injection 3	IMRPhenomXPHM	28.94	0.42	0.88	1.55	0.73	1.95	1358.51
Injection 4	NRSur7dq4	28.94	0.42	0.88	1.55	0.73	1.95	1358.51
Injection 5	IMRPhenomXPHM	65.72	0.63	0.81	1.74	0.68	1.72	2000.0
Injection 6	NRSur7dq4	65.72	0.63	0.81	1.74	0.68	1.72	2000.0
Injection 7	IMRPhenomXPHM	65.72	0.63	0.64	0.0	0.58	0.0	2000.0
Injection 8	NRSur7dq4	65.72	0.63	0.64	0.0	0.58	0.0	2000.0

TABLE VIII. Approximant model and parameters used for injections; $a_{1,2}$ and $\theta_{1,2}$ represent the magnitude and tilt angle of the components' spins, while D_L is the luminosity distance.

- [32] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, *Phys. Rev. D* **84**, 024046 (2011), arXiv:1012.2879 [gr-qc].
- [33] P. Ajith *et al.*, *Phys. Rev. Lett.* **106**, 241101 (2011), arXiv:0909.2867 [gr-qc].
- [34] L. Santamaria *et al.*, *Phys. Rev. D* **82**, 064016 (2010), arXiv:1005.3306 [gr-qc].
- [35] H. Estellés, S. Husa, M. Colleoni, D. Keitel, M. Mateu-Lucena, C. García-Quirós, A. Ramos-Buades, and A. Borchers, *Phys. Rev. D* **105**, 084039 (2022), arXiv:2012.11923 [gr-qc].
- [36] R. Abbott *et al.* (LIGO Scientific, VIRGO), (2021), arXiv:2108.01045 [gr-qc].
- [37] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [38] W. K. Hastings, *Biometrika* **57**, 97 (1970).
- [39] G. Ashton and C. Talbot, *Mon. Not. Roy. Astron. Soc.* **507**, 2037 (2021), arXiv:2106.08730 [gr-qc].
- [40] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), *Astrophys. J. Suppl.* **267**, 29 (2023), arXiv:2302.03676 [gr-qc].
- [41] R. Abbott *et al.* (LIGO Scientific, Virgo), *SoftwareX* **13**, 100658 (2021), arXiv:1912.11716 [gr-qc].
- [42] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **91**, 084034 (2015), arXiv:1410.3852 [gr-qc].
- [43] A. Puecher, “Data release: Comparing gravitational waveform models for binary black hole mergers - a hypermodels approach,” (2023).
- [44] R. Cotesta, S. Marsat, and M. Pürrer, *Phys. Rev. D* **101**, 124040 (2020), arXiv:2003.12079 [gr-qc].
- [45] J. Lin, *IEEE Trans. Info. Theor.* **37**, 145 (1991).
- [46] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].
- [47] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].
- [48] R. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J. Lett.* **900**, L13 (2020), arXiv:2009.01190 [astro-ph.HE].
- [49] I. M. Romero-Shaw, P. D. Lasky, E. Thrane, and J. C. Bustillo, *Astrophys. J. Lett.* **903**, L5 (2020), arXiv:2009.04771 [astro-ph.HE].
- [50] V. Gayathri, J. Healy, J. Lange, B. O’Brien, M. Szczepanczyk, I. Bartos, M. Campanelli, S. Klimentko, C. O. Lousto, and R. O’Shaughnessy, *Nature Astron.* **6**, 344 (2022), arXiv:2009.05461 [astro-ph.HE].
- [51] V. Gayathri, J. Healy, J. Lange, B. O’Brien, M. Szczepanczyk, I. Bartos, M. Campanelli, S. Klimentko, C. O. Lousto, and R. O’Shaughnessy, *Astrophys. J. Lett.* **908**, L34 (2021), arXiv:2009.14247 [astro-ph.HE].
- [52] Y. Xu and E. Hamilton, *Phys. Rev. D* **107**, 103049 (2023), arXiv:2211.09561 [gr-qc].
- [53] J. Calderón Bustillo, N. Sanchis-Gual, A. Torres-Forné, and J. A. Font, *Phys. Rev. Lett.* **126**, 201101 (2021), arXiv:2009.01066 [gr-qc].
- [54] R. Gamba, M. Breschi, G. Carullo, S. Albanesi, P. Rettegno, S. Bernuzzi, and A. Nagar, *Nature Astron.* **7**, 11 (2023), arXiv:2106.05575 [gr-qc].
- [55] V. De Luca, V. Desjacques, G. Franciolini, P. Pani, and A. Riotto, *Phys. Rev. Lett.* **126**, 051101 (2021), arXiv:2009.01728 [astro-ph.CO].
- [56] M. Shibata, K. Kiuchi, S. Fujibayashi, and Y. Sekiguchi, *Phys. Rev. D* **103**, 063037 (2021), arXiv:2101.05440 [astro-ph.HE].
- [57] M. Fishbach and D. E. Holz, *Astrophys. J. Lett.* **904**, L26 (2020), arXiv:2009.05472 [astro-ph.HE].
- [58] A. H. Nitz and C. D. Capano, *Astrophys. J. Lett.* **907**, L9 (2021), arXiv:2010.12558 [astro-ph.HE].
- [59] H. Estellés *et al.*, *Astrophys. J.* **924**, 79 (2022), arXiv:2105.06360 [gr-qc].
- [60] R. C. Zhang, G. Fragione, C. Kimball, and V. Kalogera, (2023), arXiv:2302.07284 [astro-ph.HE].
- [61] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, *Class. Quant. Grav.* **39**, 245013 (2022), arXiv:2207.03429 [astro-ph.IM].
- [62] M. Hannam *et al.*, *Nature* **610**, 652 (2022), arXiv:2112.11300 [gr-qc].
- [63] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C.-J. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale, *Phys. Rev. Lett.* **128**, 191102 (2022), arXiv:2201.01302 [astro-ph.HE].

- [64] E. Payne, S. Hourihane, J. Golomb, R. Udall, R. Udall, D. Davis, and K. Chatziioannou, *Phys. Rev. D* **106**, 104017 (2022), arXiv:2206.11932 [gr-qc].