# Parameter estimation with the current generation of phenomenological waveform models applied to the black hole mergers of GWTC-1

Maite Mateu-Lucena,[1]* Sascha Husa,[1] Marta Colleoni,[1] Héctor Estellés,[1] Cecilio García-Quirós,[1,2] David Keitel,[1] Maria de Lluc Planas[1] and Antoni Ramos-Buades[1,3]

[1]*Departament de Física, Universitat de les Illes Balears, IAC3 – IEEC Crta. Valldemossa km 7.5, E-07122 Palma, Spain*
[2]*Laboratoire Astroparticule et Cosmologie, 10 Rue Alice Domon, E-75013, Paris, France*
[3]*Max Planck Institut für Gravitationsphysik (Albert Einstein Institut), Am Mühlenberg 1, Potsdam, Germany*

23 September 2022

## ABSTRACT

We consider the ten confidently detected gravitational-wave signals in the GWTC-1 catalog which are consistent with mergers of binary black hole systems, and perform a thorough parameter estimation re-analysis. This is made possible by using computationally efficient waveform models of the current (fourth) generation of the IMRPhenom family of phenomenological waveform models, which consists of the IMRPhenomX frequency-domain modelsand the IMRPhenomT time-domain models.The analysis is performed with both precessing and non-precessing waveform models with and without subdominant spherical harmonic modes. Results for all events are validated with convergence tests, discussing in particular the events GW170729 and GW151226. For the latter and the other two lowest-mass events, we also compare results between two independent sampling codes, Bilby and LALInference. We find overall consistent results with the original GWTC-1 results, with all Jensen-Shannon divergences between the previous results using IMRPhenomPv2 and our default IMRPhenomXPHM posteriors below 0.045 bits, but we also discuss cases where including subdominant harmonics and/or precession influences the posteriors.

**Key words:** gravitational waves – black hole mergers – methods: data analysis

## 1 INTRODUCTION

There has been rapid progress in gravitational-wave (GW) astronomy over recent years in developing both waveform models and parameter estimation techniques. Still it remains challenging to achieve robust inference results on real data, due to the complexity of fully modelling the physics of binary black hole (BBH) mergers and also the current limitations in sampling techniques for Bayesian inference used in this relatively new field. This makes it interesting to revisit older detections with the latest toolkit at our disposal, and important to check if robust results can be obtained. Revisiting the 10 BBH events from the first Gravitational-Wave Transient Catalog (GWTC-1, Abbott et al. 2019b), our purpose with this paper is threefold: First, updating the waveform models used for the analysis to the latest generation of phenomenological models and adding subdominant harmonics will allow us to sharpen parameter estimation results. Second, the computational efficiency of these models allows us to perform careful studies of convergence, and to validate key results by comparing different sampling methods and priors. Third, we will gain insight into effects of waveform systematics on the posterior distributions and provide a "stress-test" for the new waveform model families.

A first comprehensive analysis of detections from the first two observing runs of the advanced GW detector network was given in

Abbott et al. (2019b). This analysis was based on waveform models that describe only the dominant quadrupole content of the signals and assume quasi-circularity (i.e. orbital eccentricity is neglected). Bayesian parameter estimation was performed with the LALInference (Veitch et al. 2015) code and the posterior distributions were constructed from a combination of results with the IMRPhenomPv2 and SEOBNRv3 waveform models

These models are members of the two principal families of waveform models which have been used for parameter estimation by the LIGO–Virgo(–KAGRA) collaboration (LVC/LVK), and specifically in catalog papers (Abbott et al. 2019b, 2021c,a,b): IMRPhenom and SEOBNR. For details, references, and a brief summary of the broader context of waveform models see Sec. 2.2 and Table 1.

The IMRPhenomPv2 posteriors have later been reproduced in Romero-Shaw et al. (2020) with the Bilby code (Ashton et al. 2019) for GW parameter estimation, using the nested sampling (Skilling 2006) algorithm dynesty (Speagle 2020). For a study on the effect of waveform systematics on the first detected event, GW150914, see Abbott et al. (2017a). This study concluded that for signals with higher signal-to-noise ratio (SNR) than GW150914, or in other regions of the binary parameter space (lower masses, larger mass ratios, or higher spins), we expect that systematic errors in current waveform models may impact GW measurements, making more accurate models desirable for future observations.

Some limited analysis with models that contain subdominant harmonics has also been performed: Most notably the event GW170729

* E-mail: mt.mateu@uib.es

has been studied with various waveform models (Chatziioannou et al. 2019), some of which contain subdominant harmonics. For the GWTC-1 catalog (Abbott et al. 2019b) all BBH events have been cross-checked with the RapidPE algorithm (Pankow et al. 2015; Lange et al. 2018) and a waveform bank composed of numerical relativity (NR) simulations supplemented by waveforms from the NRSur7dq2 model (Blackman et al. 2017). The calibration region of NRSur7dq2 is restricted to mass ratios between 1 and 5, and dimensionless spins up to 0.8. For RapidPE results no detailed discussion of results has been provided, but the posterior samples have been released publicly (Abbott et al. 2019a). A similar study has been performed also with the RIFT code (Lange et al. 2018; Healy et al. 2020), which like RapidPE is based on a likelihood interpolation method, and a catalog of NR waveforms. A study of higher modes, without precession, in GWTC-1 based on the approximate method of likelihood reweighting was presented in Payne et al. (2019).

Here we present our analysis of the ten confidently detected BBH signals in GWTC-1, which has been the first to be performed based on state-of-the art methods in the following sense: We use waveform models where at least for the non-precessing sector all spherical harmonics have been calibrated to NR; the waveform models are sufficiently computationally efficient to use wide priors and perform systematic studies that include convergence tests and comparisons of different samplers; and we use direct Bayesian sampling evaluating the likelihood with these waveform models, without extra approximations, such as recurring to a discrete template bank or likelihood reweighting without sampling the posterior for each waveform model. Analyses of these GWTC-1 events with precessing higher-mode models have also later been performed by Nitz et al. (2021) and by the LVC (Abbott et al. 2021a), finding broad agreement with our results, but using only one of the models (IMRPhenomXPHM) we use here and will discuss next. Two events for which the picture is more complex, GW151226 (see also Chia et al. 2022; Vajpeyi et al. 2022) and GW170729 (see also Chatziioannou et al. 2019), will be discussed further below after first summarizing our methodology.

The waveform models we use correspond to the current (fourth) generation of the IMRPhenom family of inspiral–merger–ringdown phenomenological waveform models. These models consist of the IMRPhenomX sub-family of frequency-domain models (Pratten et al. 2020; García-Quirós et al. 2020, 2021; Pratten et al. 2021), and the IMRPhenomT time-domain models (Estellés et al. 2021b; Estellés et al. 2020, 2021a). IMRPhenomX constitutes a thorough upgrade of the previous generation of frequency-domain phenomenological waveform models based on IMRPhenomD (Husa et al. 2016; Khan et al. 2016; Hannam et al. 2014; Bohé et al. 2016; London et al. 2018; Khan et al. 2019, 2020), which included the first models in the IMRPhenom family to describe subdominant harmonics (IMRPhenomHM) and both subdominant harmonics and precession (IMRPhenomPv3HM). IMRPhenomT offers advantages in the description of precession, in particular concerning the merger and ringdown, as described in detail in Estellés et al. (2021a) and briefly summarized in Sec. 2.2.

Our main results are obtained with the parallel Bilby (Ashton et al. 2019; Smith et al. 2020) sampling code. In addition, for the three lowest-mass events we also cross-check with the independent LALInference code (Veitch et al. 2015), and demonstrate good agreement. We have added these cross-checks, which confirm our results, after the initial versions of Nitz et al. (2021) and Chia et al. (2022) appeared on the arXiv and showed tension for the event GW151226. In particular, the first version of Chia et al. (2022) found a bi-modal posterior with support for very unequal masses, in tension with both our results and with Nitz et al. (2021). However, the results in their

final journal version (using an updated method and priors) have significantly reduced this tension, with remaining differences consistent with those in priors. A later study by Vajpeyi et al. (2022) also studied possible multiple peaks and prior dependency for GW151226. We will discuss this case in detail in Sec. 3.2.2 and Appendix A. Another special case is GW170729, the highest-mass event of GWTC-1. We have also performed additional runs and comparisons with the literature for this event and discuss it in detail in Sec. 3.4.2.

The paper is organized as follows: We first present our methods in Sec. 2: our conventions, the observational data used, our waveform models, and the setup for running parameter estimation. In Sec. 3 we present our parameter estimation results, and we conclude in Sec. 4. Further details are presented in two appendices: In appendix A we present checks on the sampler convergence, and compare results for the three lowest mass events between the Bilby and LALInference samplers. Finally in appendix B we test aggressive settings of our multibanding algorithm to accelerate waveform evaluation.

## 2 METHODS

### 2.1 Notation and conventions

Masses refer to the source frame unless we note explicitly that we refer to the detector frame. Source frame masses are inferred assuming a standard cosmology (Ade et al. 2016) (see Appendix B of Abbott et al. (2019b)). Individual component masses are denoted by $m_i$, and the total mass is $M_T = m_1 + m_2$. The chirp mass is $\mathcal{M} = (m_1 m_2)^{3/5} M_T^{-1/5}$, and we define mass ratios as $q = m_2/m_1 \leq 1$ and $Q = m_1/m_2 \geq 1$.

We also report two effective spin parameters which are commonly used in waveform modelling and parameter estimation. The parameter $\chi_{\text{eff}}$ (Santamaría et al. 2010; Ajith et al. 2011) captures those dominant spin effects which do not depend on spin precession, and is defined as

$$\chi_{\text{eff}} = \frac{m_1 \chi_1 + m_2 \chi_2}{m_1 + m_2}, \tag{1}$$

where the $\chi_i$ are the projections of the dimensionless component spin vectors onto the orbital angular momentum. The effective spin precession parameter $\chi_{\text{p}}$ (Schmidt et al. 2015) is designed to capture the dominant effect of precession, and corresponds to an approximate average of the spin component in the precessing orbital plane over many precession cycles, see Schmidt et al. (2015) and Eq. (4.7) in Pratten et al. (2021). Both $\chi_{\text{eff}}$ and $\chi_{\text{p}}$ are dimensionless, and thus independent of the frame (source or detector).

When referring to specific spherical harmonic modes of the GW signal we will always consider pairs of both positive and negative modes, e.g. when we refer to the example list of multipoles $(l, m) = (2, \pm 2), (2, \pm 1)$ we will use the notation $(l, |m|) = (2, 2), (2, 1)$ or simply $(2, 2), (2, 1)$.

### 2.2 Waveform models used

Three main approaches have been used to construct waveform models for comparable-mass coalescences of quasi-circular BBHs, which have become standard tools of GW data analysis:

• The effective one-body (EOB) approach (Buonanno & Damour 1999) models the Hamiltonian and energy flux, which yield a system of ordinary differential equations. This is solved numerically, which is however computationally expensive. The Hamiltonian and flux, and any further quantities used in the construction of the waveform,

| Family | Full name | Precession | Multipoles $(\ell, |m|)$ included | Ref. |
|---|---|---|---|---|
| SEOBNR | SEOBNRv2 | × | (2, 2) | Taracchini et al. (2014) |
| | SEOBNRv3 | ✓ | (2, 2) | Babak et al. (2017) |
| | SEOBNRv4_ROM | × | (2, 2) | Bohé et al. (2017) |
| | SEOBNRv4HM_ROM | × | (2, 2), (2,1), (3, 3), (4, 4), (5,5) | Cotesta et al. (2018, 2020) |
| | SEOBNRv4P | ✓ | (2, 2), (2, 1) | Ossokine et al. (2020) |
| | SEOBNRv4PHM | ✓ | (2, 2), (2, 1), (3, 3), (4, 4), (5,5) | Ossokine et al. (2020) |
| IMRPhenom - 3$^{rd}$ Generation | IMRPhenomD | × | (2, 2) | Husa et al. (2016); Khan et al. (2016) |
| | IMRPhenomHM | × | (2, 2), (2, 1), (3, 3), (3, 2), (4,4), (4, 3) | London et al. (2018) |
| | IMRPhenomPv2 | ✓ | (2, 2) | Hannam et al. (2014); Bohé et al. (2016) |
| | IMRPhenomPv3 | ✓ | (2, 2) | Khan et al. (2019) |
| | IMRPhenomPv3HM | ✓ | (2, 2), (2, 1), (3, 3), (3, 2),(4, 4), (4, 3) | Khan et al. (2020) |
| IMRPhenomX | IMRPhenomXAS | × | (2, 2) | Pratten et al. (2020) |
| | IMRPhenomXHM | × | (2, 2), (2, 1), (3, 3), (3, 2), (4,4) | García-Quirós et al. (2020, 2021) |
| | IMRPhenomXP | ✓ | (2, 2) | Pratten et al. (2021) |
| | IMRPhenomXPHM | ✓ | (2, 2), (2, 1), (3, 3), (3, 2),(4, 4) | Pratten et al. (2021) |
| IMRPhenomT | IMRPhenomT | × | (2, 2) | Estellés et al. (2021b); Estellés et al. (2020) |
| | IMRPhenomTHM | × | (2, 2), (2, 1), (3, 3), (4,4), (5,5) | Estellés et al. (2020) |
| | IMRPhenomTP | ✓ | (2, 2) | Estellés et al. (2021b); Estellés et al. (2021a) |
| | IMRPhenomTPHM | ✓ | (2, 2), (2, 1), (3, 3),(4, 4), (5,5) | Estellés et al. (2021a) |

**Table 1.** The table lists waveform models from the IMRPhenom and SEOB families relevant to this paper. Only the latest generation of IMRPhenom waveforms has been used to obtain the results in this paper, however SEOB and older IMRPhenom waveforms have been used in relevant literature we compare to, and we list them for convenience of the reader.

For precessing models, the multipoles correspond to those in a co-precessing frame.

can be calibrated to NR simulations. Two families of such models have been developed: The SEOBNR family has been regularly used for parameter estimation by the LVC/LVK, and a list of models is given in Table 1. An alternative is the TEOBResumS model (Nagar et al. 2018).

• The IMRPhenom models describe the amplitude and phase of the spherical harmonic modes piecewise as closed-form expressions in the inspiral, merger and ringdown. These expressions have been calibrated to NR for the late phase of the coalescence, and for the third and current fourth generation to EOB for the early inspiral. These models are computationally much more efficient than EOB models, and can be used for Bayesian parameter estimation without further approximations.

• Another approach has been to construct numerical interpolants of discrete data sets; such types of models are often referred to as reduced-order models (ROMs) or surrogate models. This approach has been used to accelerate the evaluation of EOB models – e.g. the SEOBNRv4HM_ROM model (Cotesta et al. 2018), see Table 1 – or directly to interpolate data sets of NR waveforms (Blackman et al. 2017). In the latter case, validity is restricted to the parameter space where NR waveforms are available, with small extensions through extrapolations. Since SNRs for current ground-based detectors are relatively small, posteriors are typically wide, and for GW data analysis it is convenient to use EOB or IMRPhenom models, which can be used for larger regions of parameter space.

For a list of IMRPhenom and SEOB waveform models, which are most directly relevant for this paper, see Table 1.

In this work we present new parameter estimation results which we have obtained with the most recent waveform models of the IMRPHENOMX and IMRPHENOMT families, and compare with previous results from the GWTC-1 catalog (Abbott et al. 2019b), and also with Chatziioannou et al. (2019) and Payne et al. (2019) for GW170729.

Parameter estimation for the GWTC-1 catalog paper (Abbott et al. 2019b) has used two precessing waveform models: IMRPHENOMPv2 (Hannam et al. 2014; Bohé et al. 2016) is a predecessor of our

updated precessing phenomenological waveform models IMRPHE-NOMXPHM (Pratten et al. 2021) and IMRPHENOMTPHM (Estellés et al. 2021a)), and is based on the non-precessing IMRPHENOMD (Husa et al. 2016; Khan et al. 2016) model. SEOBNRv3 (Babak et al. 2017) is based on the non-precessing SEOBNRv2 (Taracchini et al. 2014) model and is a predecessor of SEOBNRv4PHM (Ossokine et al. 2020), constructed within the effective-one-body (EOB) framework (Buonanno & Damour 1999, 2000), which is based on solving a system of ordinary differential equations for the dynamics of the binary.

For the GW170729 event an analysis with higher mode models has already been performed (Chatziioannou et al. 2019), using the non-precessing IMRPHENOMHM (London et al. 2018) and SEOB-NRv4HM_ROM (Cotesta et al. 2018, 2020) models, with which we will also compare. The latter is an example of the reduced-order models (ROMs) which are often used to accelerate the evaluation of EOB waveforms. Preliminary analyses of GW170729 with our IMRPHE-NOMXAS and IMRPHENOMT models have already been presented in García-Quirós et al. (2020) and Estellés et al. (2020).

We now turn to describing the new IMRPHENOMX and IMR-PHENOMT families. Phenomenological frequency-domain models are constructed as piecewise closed-form expressions that are calibrated to post-Newtonian (PN) or EOB inspiral results and NR waveforms. Such explicit descriptions of the waveform can be evaluated very rapidly. The corner stone of the IMRPHENOMX family is IMRPHE-NOMXAS (Pratten et al. 2020), which models the $(\ell, |m|) = (2, 2)$ modes of non-precessing signals, and has been extended to include the $(\ell, |m|) = (2, 1), (3, 3), (3, 2), (4, 4)$ modes by IMRPHE-NOMXHM (García-Quirós et al. 2020, 2021). All the modes have been calibrated to around 500 numerical waveforms as reported in García-Quirós et al. (2021), whereas in the previous generation of IMRPHENOM waveforms only the dominant $(\ell, |m|) = (2, 2)$ modes have been calibrated to 19 NR waveforms (Husa et al. 2016; Khan et al. 2016).

IMRPHENOMXPHM (Pratten et al. 2021) augments IMRPHE-

NOMXHM to account for spin precession following previous phenomenological models (Schmidt et al. 2011, 2012; Hannam et al. 2014) in employing an approximate map between non-precessing and precessing waveforms, where the signal of the precessing system in a non-inertial co-precessing frame is identified with an appropriate non-precessing signal, and where the final spin is adjusted to account for the effect of spin components orthogonal to the instantaneous axis of the orbital plane. The non-precessing signal with adjusted final spin is then mapped to a precessing signal by employing a time-dependent rotation from the non-inertial co-precessing frame to the inertial frame where the signal is observed. This procedure is often referred to as "twisting up" in the literature. IMRPHENOMXPHM implements two approximate descriptions for the Euler angles that define the time-dependent rotation: an orbit-averaged effective single-spin description at the next-to-next-to-leading PN order (NNLO, Blanchet et al. 2011; Marsat et al. 2013), or an approximation based on a multiple scale analysis of the PN equations (Chatziioannou et al. 2013), which is the default choice. For a detailed discussion of the procedure and the conventions employed to describe precessing waveforms in the frequency domain see Pratten et al. (2021).

IMRPHENOMXPHM also allows to choose among different approximations for the final spin of the remnant black hole. Version 0 is equivalent to the estimate implemented in IMRPHENOMPv2, where the in-plane spin contribution is captured by the effective precession spin $\chi_p$. Version 1 is equivalent to version 0, with the replacement of $\chi_p$ by the x-component of the spin, $\chi_x$. Versions 2 and 3 attempt in different ways to incorporate the full double-spin effects, with the latter relying on precession-averaged quantities that arise in the context of the MSA formalism (Chatziioannou et al. 2013). Version 3 is the default in the LALSuite (LIGO Scientific Collaboration 2020) implementation. We refer the reader to Sec. IV D of Pratten et al. (2021) for a thorough discussion of these options. The IMRPHENOMT family of models (Estellés et al. 2021b; Estellés et al. 2020, 2021a) extends to the time domain the methods that have been used to construct the frequency-domain model IMRPHENOMXHM and calibrate it to a catalog of numerical waveforms. The main motivation for developing a time domain model is that the morphology of the waveform is often simpler to understand in the time domain, in particular for the late inspiral, merger, and ringdown, where rapid changes are not "smeared out" by a Fourier transform. This allows a cleaner separation of inspiral, merger and ringdown, which can be valuable for tests of general relativity, this is however not directly relevant for the present paper. Precession is again described by the "twisting up" approach (Estellés et al. 2021b; Estellés et al. 2021a), however the time domain model indeed allows some key improvements over IMRPHENOMXPHM: First, in order to obtain explicit expressions for the spherical harmonic modes of the precessing frequency-domain models, the stationary phase approximation (SPA) is used to compute approximate Fourier transforms. This also makes it difficult to incorporate analytical knowledge about the ringdown frequencies in the ringdown portion of a precessing waveform. This is however simple in the time domain, where the ringdown waveform can be simply modelled as a damped sinusoid, with the complex ringdown frequency determined by the final spin and mass. This allows a simple analytical approximation to the Euler angles during ringdown, based on the effective precessional motion observed in NR signals (O'Shaughnessy et al. 2013), which can be derived analytically in the small opening angle limit. (A derivation is presented in Estellés et al. (2021b), see also Marsat & Baker (2018).) In consequence, the merger and ringdown of precessing waveforms is typically more accurately described in our time-domain model.

IMRPHENOMTPHM also offers an improvement of the inspiral description by departing from the previous approach of using closed-form expressions for the Euler angles that describe the precession of the orbital plane, and by default a numerical integration of the equations for the spin dynamics is used, which can be carried out without significant increase of computational cost (Estellés et al. 2021a). For a more detailed discussion of the differences between IMRPHENOMXPHM and IMRPHENOMTPHM regarding the treatment of precession see Estellés et al. (2021c).

For lower-mass events, despite the gain in accuracy of describing precession, there is however some loss of accuracy for the phase of the dominant harmonic, which is constructed based on a calibrated extension of the TaylorT3 PN approximant (Blanchet 2006). The choice of this approximant was motivated by the explicit time dependence of the orbital frequency as a closed-form expression, but it is known (Buonanno et al. 2009) to be less accurate than other numerical PN approximants, and in consequence turns out to be more difficult to calibrate to NR. Resolving this weakness of the underlying "carrier phase" in IMRPHENOMT will be the goal of future work, and will allow to use this model also for low mass events. As discussed above, the primary motivation for the IMRPHENOMT family was to improve the treatment of the late part of the waveform, which is more relevant for higher-mass systems. We thus perform parameter estimation with the IMRPHENOMT family only for higher-mass events.

For IMRPHENOMXPHM we have implemented the "multibanding" method (Vinciguerra et al. 2017; García-Quirós et al. 2021) to accelerate waveform evaluation, based on interpolation and the choice of an accuracy parameter. In appendix B we discuss parameter estimation results with a more aggressive choice of multibanding settings, as would be appropriate for accelerated parameter estimation runs, e.g. for the initial exploration of newly detected events.

### 2.3 Methodology for Parameter estimation

#### 2.3.1 Data

We use public GW strain data from the Gravitational Wave Open Science Center (GWOSC, Vallisneri et al. 2015; LIGO Scientific Collaboration, Virgo Collaboration 2019a), and power spectral densities (PSDs, LIGO Scientific Collaboration, Virgo Collaboration 2019b) and calibration uncertainties (LIGO Scientific Collaboration, Virgo Collaboration 2019c) included in the GWOSC release for GWTC-1. The Virgo detector officially joined the detector network on August 1, 2017. However, Virgo data for the event GW170729 is also publicly available from GWOSC, and following the example of Abbott et al. (2019b) and Chatziioannou et al. (2019) we have used Virgo data also to analyse this event, in addition to the later events, but with the exception of GW170823, which was only observed by the LIGO detectors.

Table 2 lists the trigger times corresponding to the GW events, indicates the detector data sets we use for each event, the duration of data we analyze, the minimal frequency we use to compute the likelihood function (and also as a reference frequency to define the GW phase and, for precessing systems, the component spins), and the sampling rate used. All of these parameters correspond to the choices made in Abbott et al. (2019b). The sampling rate used in Abbott et al. (2019b) is 2048 Hz, corresponding to a Nyquist frequency of 1024 Hz. We have used this as our default setting to facilitate comparisons, together with GWOSC data sets sampled at 4 KHz. For high-mass events this is sufficient. For the three low-mass events GW151012, GW151226 and GW170608 however, the $|m| = 3, 4$ harmonics can reach higher frequencies than 1024 Hz, and we have

| Event | Trigger time | Duration | Low freq. | Sampling rate |
|---|---|---|---|---|
| GW150914 | 1126259462.4 | 8 | 20 | 2048 |
| GW151012 | 1128678900.4 | 8 | 20 | 2048/4096 |
| GW151226 | 1135136350.6 | 8 | 20 | 2048/4096 |
| GW170104 | 1167559936.6 | 4 | 20 | 2048 |
| GW170608 | 1180922494.5 | 16 | 20 | 2048/4096 |
| GW170729* | 1185389807.3 | 4 | 20 | 2048 |
| GW170809* | 1186302519.7 | 4 | 20 | 2048 |
| GW170814* | 1186741861.5 | 4 | 20 | 2048 |
| GW170818* | 1187058327.1 | 4 | 16 | 2048 |
| GW170823 | 1187529256.5 | 4 | 10 | 2048 |

**Table 2.** Data settings of each event from Abbott et al. (2019b) where * indicates that the three LIGO and Virgo detectors were taking data and all three data sets were used for our analysis. Note that for GW151012, GW151226 and GW170608 we extended the analysis also for a higher sampling rate.

performed additional parameter estimation runs with a sampling rate of 4096 Hz and the GWOSC 16 kHz data sets (Nyquist frequency of 2048 Hz). For these analyses we have computed PSDs with the BayesWave code (Cornish & Littenberg 2015; LIGO Scientific Collaboration, Virgo Collaboration 2021), which has been used by the LVC to produce the PSDs provided in the GWOSC data release. We have checked that we can reproduce the public PSDs when setting the sampling rate to 2048 Hz and using the 16 kHz data sets, and then used the same settings to produce PSDs at twice the sampling rate. These agree well with the PSDs using a lower sampling rate up to approximately 1 kHz; see Fig. 1 for the example of data around the GW151226 event. We also extend the calibration envelope files to 2048 Hz. In LIGO Scientific Collaboration, Virgo Collaboration (2018), we can find the plots for the magnitude and phase quantities of the O2 events. For GW170608, we extracted the data from 20 Hz to 2048 Hz from these plots. However, for the O1 events such plots are not publicly available. For that reason, we performed a linear extrapolation in frequency of the calibration uncertainties from 1024 Hz to 2048 Hz with the public data available in LIGO Scientific Collaboration, Virgo Collaboration (2019c).

### 2.3.2 Bayesian sampling algorithms

For our standard runs we use the parallel Bilby (Ashton et al. 2019; Smith et al. 2020) implementation of the `dynesty` (Speagle 2020) algorithm, based on nested sampling (Skilling 2006). In Appendix A we also compare with results obtained with the LALInference code (Veitch et al. 2015). For both Bilby and LALInference we sample the component masses in terms of the mass ratio and chirp mass. We also use the default settings of the Bilby code apart from the following choices: we fix the minimal (`walks`) and maximal (`maxmcmc`) number of Markov Chain Monte Carlo (MCMC) steps to 200 and 15000 respectively. We vary the number of nested sampling live points, setting `nlive` = 512, 1024, 2048, and 4096 for the three lowest mass events, and we vary the number of autocorrelation times to use before accepting a point, `nact` = 10, 30, 50, in order to study the convergence of results. As explained in appendix A, we choose `nlive` = 2048 and `nact` = 30 as the standard settings to report our main results. In order to speed up calculations we use distance marginalization as described in Thrane & Talbot (2019).

In this paper we only discuss the most relevant results, however the complete posterior data sets are included in our Zenodo data release (Mateu-Lucena et al. 2021).

### 2.3.3 Priors

In this paper we analyze the posterior probability densities of the source parameters $\theta$ from compact binary merger signals, computed using Bayes' theorem,

$$p(\theta|d, \mathcal{H}) = \frac{\mathcal{L}(d|\theta, \mathcal{H})\pi(\theta|\mathcal{H})}{\mathcal{Z}(d|\mathcal{H})} \qquad (2)$$

where $\mathcal{L}(d|\theta, \mathcal{H})$ is the likelihood, $\pi(\theta|\mathcal{H})$ the prior and $\mathcal{Z}(d|\mathcal{H})$ the evidence for a given model ($\mathcal{H}$) and data ($d$).

The prior incorporates previous knowledge about the source parameters. The intention of our prior choices is to be uninformative, however, ambiguities arise when it is not possible to choose a natural quantity in which to prescribe a flat prior. For several quantities no problems arise: We use an isotropic prior for the sky location, the prior for the inclination angle is uniform in its cosine, and the priors of the polarization angle and phase of coalescence are uniform. For precessing spins, we use uniform distributions in the dimensionless component spin magnitudes and isotropically distributed spin tilts, as also used in Abbott et al. (2019b) and Romero-Shaw et al. (2020). We allow dimensionless component spin magnitudes up to 0.99.

The distance prior used for the GWTC-1 catalog (Abbott et al. 2019b) is uniform in luminosity distance volume (described by a second-order power law, $\pi(d_L) \propto d_L^2$), which distributes mergers uniformly through a stationary Euclidean universe and does not take the expansion of the universe into account. The effect of the expansion of the universe can be neglected for small redshifts, however for larger redshifts it is more appropriate to use a prior that factors in our knowledge about cosmological expansion. (See the discussions in Appendix C of Abbott et al. (2021c) and section 3.2.3 of Romero-Shaw et al. (2020).) We will use the simpler prior "uniform in luminosity distance volume" for all events, however for the furthest events (GW170729 and GW170823) we also compare with a prior uniform in the comoving volume and source-frame time (`UniformComovingVolume` prior in the Bilby code) using the Planck2015 cosmology (Ade et al. 2016), following the Bilby catalog (Romero-Shaw et al. 2020).

Regarding the masses, the GWCT-1 catalog applies the LALInference (Veitch et al. 2015) prior, which is uniform in the component masses $m_1$ and $m_2$ with cuts in $\mathcal{M}_c$ and $q$. The Bilby re-analysis of GWTC-1 (Romero-Shaw et al. 2020) uses a prior that is flat in chirp mass and mass ratio, but then the samples are reweighted to the LALInference choice using the Jacobian given in Eq. (21) of Veitch et al. (2015). We follow this approach here for our default runs for all events, and present posteriors reweighted for a prior that is flat in component masses, consistent with Abbott et al. (2019b) and Romero-Shaw et al. (2020). A very different choice has been used recently in Nitz & Capano (2021) to re-analyse the signal GW190521, which is consistent with the creation of an intermediate-mass black hole in a heavy black hole merger. Using a prior that is flat in mass ratio $Q \geq 1$, they found a multi-modal source-frame total mass posterior. We analyse this event in a parallel paper (Estellés et al. 2021c). with improved waveform models and confirm the multi-modality, although we also find significant differences with respect to Nitz & Capano (2021). In order to check whether similar multi-modalities appear for the most massive events in GWTC-1, we rerun GW170729 with a prior that is flat in mass ratio $Q \geq 1$ and *detector-frame* total mass (a prior that is flat in source-frame masses adds complications to using distance marginalization, which we use to reduce the computational cost).

We also select GW170729 (see Sec. 3.4.2) and the three lowest-mass events (see appendix A) to compare results with the new prior implemented in Bilby that applies the uniform prior in com-
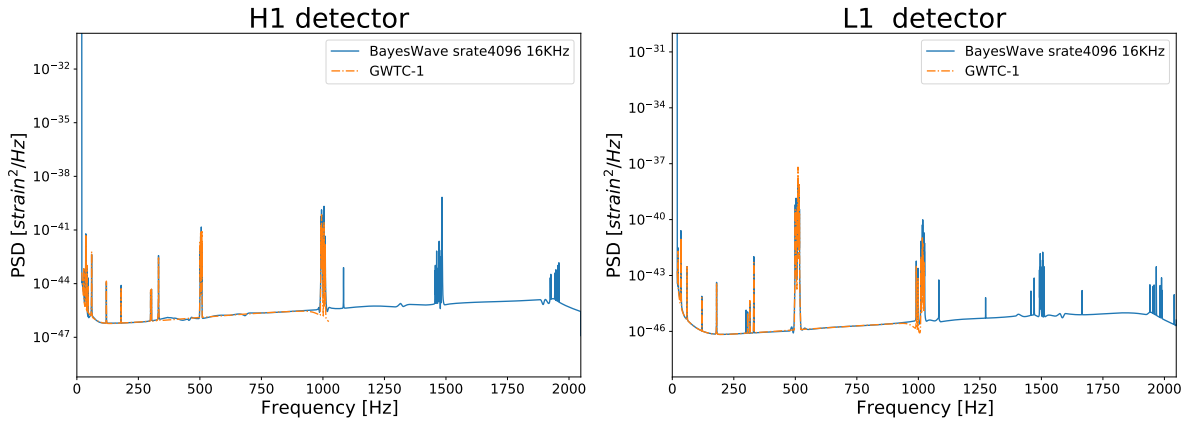
**Figure 1.** Comparison of PSDs of the two LIGO interferometers (H1 and L1) for GW151226 between the public data files provided in the 16 kHz GWOSC data release with a sampling rate of 2048 Hz (labeled with `GWTC-1`), and our PSDs produced with the `BayesWave` code, using a sampling rate of 4096 Hz. One can see that both PSD data sets are very similar up to a frequency of almost 1 kHz, where artefacts show up in the public PSD. These artefacts (decrease of the PSD) are similar to the high-frequency artefacts (near 2 kHz) that can be seen in the PSDs we have produced with BayesWave.

ponent masses while the mass ratio and chirp mass are sampled (by using the two classes `UniformInComponentsChirpMass` and `UniformInComponentsMassRatio` in the Bilby code), using the same component mass prior choices as the LALInference code.

The limits of the mass and luminosity distance priors vary between different events. We take these limits from the LALInference configuration files of the runs from Abbott et al. (2019b) and we check that the inferred parameters do not rail against the prior bounds.

### 2.3.4 Management of parameter estimation runs

In total we have performed 364 runs with parallel Bilby (Smith et al. 2020). In order to simplify the management of runs, and to avoid human errors when editing configuration files for each run, we have developed a `Python` code to generate configuration files from simpler descriptions which we call pseudo-config files. These pseudo-config files allow us to loop over parameters such as the waveform model, `nlive` or `nact`. Furthermore, we use this `Python` code to submit several runs to the `slurm` (Yoo et al. 2003) queuing system with one command, and to track version numbers of the codes we use when the jobs start running.

## 3 RESULTS

### 3.1 Summary of the black hole mergers in GWTC-1

Table 3 contains a summary of our results using precessing waveform models, with error estimates obtained from 90% posterior credible intervals, including for comparison also the results obtained for GWTC-1 (Abbott et al. 2019b) with IMRPHENOMPv2. Our new results are in general within the error estimates of the IMRPHENOMPv2 results, as can be expected from the initial analysis with subdominant harmonics performed for the GWTC-1 catalog, see appendix B of Abbott et al. (2019b): There all BBH events have been cross-checked with the RapidPE algorithm (Pankow et al. 2015; Lange et al. 2018) and a waveform catalog of NR simulations supplemented by waveforms from the NRSUR7DQ2 model (Blackman et al. 2017); samples were released with Abbott et al. (2019a). NRSUR7DQ2 is however restricted to mass ratios $q \leq 2$ and dimensionless $\chi_i \leq 0.8$. A "modest influence on the interpretation of observations" below the statistical

measurement uncertainty is found, and for GW170729 a Bayes factor of "approximately 1.4 for higher modes versus a pure quadrupole model", which we compare with our results below.

However, the use of multi-mode waveforms in full Bayesian parameter estimation is now standard in GW data analysis, and it is thus interesting to update the results of GWTC-1 to the methods used for O3 events (Abbott et al. 2021a,b), and to compare results. Differences between our default runs using IMRPHENOMXPHM and the IMRPHENOMPv2 results from the GWTC-1 release are shown in Fig. 2 in terms of the Jensen-Shannon (JS) divergence, for a definition see e.g. Eq. (B1) of Abbott et al. (2019b). The JS divergence takes values between 0 and 1 bits, where 0 means that there is no difference between two distributions, and 1 means that both distributions have a maximum divergence. The posteriors for the model with higher modes and precession are generally close to the results obtained with IMRPHENOMPv2– all divergence values between both posterior distributions are smaller than 0.045, which is smaller than the divergence between IMRPHENOMPv2 and SEOBNRv3 runs performed in the GWTC-1 paper, where the largest value is for GW151226 and the effective spin $\chi_{\rm eff}$, $JS_{\chi_{\rm eff}} = 0.14$ bits, see Fig. 16 of Abbott et al. (2019b).

Among the extrinsic parameters, the sky location (declination $\delta$ and right ascension $\alpha$) does not show significant changes, while the distance $D_L$ and the angle $\theta_{JN}$ between total angular momentum and the line of sight have the largest divergences.

An overview of recovered values for the effective spin and mass ratio for the different events is shown in Fig. 3. One of the trends that can be observed (see also the detailed results in Table 3) is that the effective spins $\chi_{\rm eff}$ inferred with IMRPHENOMXPHM have typically increased over the values inferred with IMRPHENOMPv2, although they are still consistent within the error estimates for all events. The event least following this trend is the highest-mass event GW170729, which is also one of the two events where $\chi_{\rm eff}$ has support only for positive values (GW151226 being the other).

In Table 4, we show the matched-filter SNRs obtained for runs with different models, as well as the Bayes factors between the signal and noise hypotheses for each of these runs, and compare these Bayes factors against our default IMRPHENOMXPHM runs. Bayes factors and SNRs are in general very consistent for different waveform models, confirming the expectation that neither precession, higher

| Approx. | $q$ | $m_1/M_\odot$ | $m_2/M_\odot$ | $M_T/M_\odot$ | $\mathcal{M}/M_\odot$ | $M_f/M_\odot$ | $\chi_f$ | $\chi_{\text{eff}}$ | $\chi_p$ | $D_L/\text{Mpc}$ | $\theta_{JN}/rad$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **GW150914** | | | | | | | | | | | |
| Pv2 | $0.86^{+0.11}_{-0.16}$ | $35.3^{+3.7}_{-2.4}$ | $30.3^{+2.3}_{-3.4}$ | $65.7^{+2.7}_{-2.5}$ | $28.4^{+1.2}_{-1.1}$ | $61.6^{+2.7}_{-2.6}$ | $0.80^{+0.12}_{-0.08}$ | $-0.03^{+0.08}_{-0.09}$ | $0.39^{+0.38}_{-0.27}$ | $450^{+120}_{-140}$ | $2.72^{+0.27}_{-0.52}$ |
| XP | $0.85^{+0.12}_{-0.16}$ | $35.6^{+3.7}_{-2.6}$ | $30.0^{+2.5}_{-3.4}$ | $65.7^{+2.6}_{-2.6}$ | $28.4^{+1.2}_{-1.2}$ | $61.7^{+2.7}_{-2.8}$ | $0.80^{+0.11}_{-0.08}$ | $-0.02^{+0.08}_{-0.11}$ | $0.42^{+0.36}_{-0.29}$ | $440^{+140}_{-140}$ | $2.63^{+0.32}_{-0.54}$ |
| XPHM | $0.87^{+0.11}_{-0.15}$ | $35.0^{+3.5}_{-2.3}$ | $30.2^{+2.3}_{-3.2}$ | $65.2^{+2.5}_{-2.4}$ | $28.2^{+1.1}_{-1.1}$ | $61.2^{+2.4}_{-2.6}$ | $0.81^{+0.10}_{-0.08}$ | $-0.02^{+0.08}_{-0.10}$ | $0.44^{+0.34}_{-0.30}$ | $490^{+120}_{-120}$ | $2.76^{+0.24}_{-0.33}$ |
| TPHM | $0.88^{+0.10}_{-0.15}$ | $35.0^{+3.4}_{-2.1}$ | $30.7^{+2.1}_{-2.9}$ | $65.7^{+2.7}_{-2.3}$ | $28.4^{+1.1}_{-1.0}$ | $61.7^{+2.6}_{-2.4}$ | $0.80^{+0.10}_{-0.08}$ | $-0.01^{+0.09}_{-0.09}$ | $0.43^{+0.34}_{-0.30}$ | $480^{+120}_{-130}$ | $2.78^{+0.23}_{-0.37}$ |
| **GW151012** | | | | | | | | | | | |
| Pv2 | $0.61^{+0.30}_{-0.30}$ | $22.6^{+9.6}_{-4.3}$ | $13.8^{+3.2}_{-3.9}$ | $36.9^{+5.9}_{-3.0}$ | $15.2^{+1.2}_{-1.1}$ | $34.9^{+6.2}_{-3.0}$ | $0.77^{+0.13}_{-0.12}$ | $0.04^{+0.20}_{-0.15}$ | $0.33^{+0.35}_{-0.22}$ | $1090^{+390}_{-400}$ | $1.99^{+0.84}_{-1.64}$ |
| XP | $0.65^{+0.27}_{-0.31}$ | $22.0^{+8.9}_{-3.8}$ | $14.1^{+2.9}_{-3.8}$ | $36.6^{+5.4}_{-2.8}$ | $15.2^{+1.2}_{-0.9}$ | $34.4^{+5.6}_{-2.8}$ | $0.79^{+0.13}_{-0.11}$ | $0.05^{+0.21}_{-0.14}$ | $0.38^{+0.36}_{-0.25}$ | $1120^{+400}_{-380}$ | $1.84^{+0.96}_{-1.49}$ |
| XPHM | $0.57^{+0.32}_{-0.28}$ | $24.2^{+10.5}_{-5.1}$ | $13.6^{+3.5}_{-4.0}$ | $38.3^{+6.9}_{-3.7}$ | $15.5^{+1.5}_{-1.1}$ | $36.2^{+7.2}_{-3.7}$ | $0.79^{+0.13}_{-0.12}$ | $0.12^{+0.21}_{-0.17}$ | $0.36^{+0.37}_{-0.23}$ | $1060^{+460}_{-400}$ | $1.66^{+1.08}_{-1.25}$ |
| **GW151226** | | | | | | | | | | | |
| Pv2 | $0.54^{+0.30}_{-0.25}$ | $14.0^{+5.6}_{-2.9}$ | $7.6^{+1.8}_{-1.8}$ | $21.6^{+3.7}_{-1.3}$ | $8.9^{+0.3}_{-0.2}$ | $20.3^{+3.9}_{-1.4}$ | $0.84^{+0.12}_{-0.09}$ | $0.20^{+0.13}_{-0.06}$ | $0.47^{+0.32}_{-0.26}$ | $440^{+150}_{-160}$ | $1.05^{+1.66}_{-0.66}$ |
| XP | $0.62^{+0.30}_{-0.30}$ | $13.0^{+5.5}_{-2.3}$ | $8.0^{+1.7}_{-2.1}$ | $21.2^{+3.3}_{-1.0}$ | $8.8^{+0.3}_{-0.2}$ | $19.8^{+3.5}_{-1.1}$ | $0.84^{+0.11}_{-0.09}$ | $0.18^{+0.13}_{-0.06}$ | $0.48^{+0.32}_{-0.26}$ | $460^{+150}_{-160}$ | $1.01^{+1.71}_{-0.62}$ |
| XPHM | $0.52^{+0.34}_{-0.25}$ | $14.2^{+6.2}_{-3.2}$ | $7.4^{+2.0}_{-1.9}$ | $21.7^{+4.3}_{-1.4}$ | $8.8^{+0.2}_{-0.2}$ | $20.3^{+4.3}_{-1.4}$ | $0.88^{+0.14}_{-0.11}$ | $0.20^{+0.13}_{-0.07}$ | $0.57^{+0.28}_{-0.31}$ | $470^{+130}_{-160}$ | $0.79^{+1.92}_{-0.44}$ |
| **GW170104** | | | | | | | | | | | |
| Pv2 | $0.63^{+0.26}_{-0.19}$ | $31.2^{+5.7}_{-4.8}$ | $19.8^{+4.1}_{-3.7}$ | $51.1^{+4.1}_{-3.3}$ | $21.4^{+1.6}_{-1.4}$ | $48.2^{+4.1}_{-3.2}$ | $0.79^{+0.12}_{-0.09}$ | $-0.05^{+0.13}_{-0.15}$ | $0.38^{+0.32}_{-0.23}$ | $990^{+350}_{-350}$ | $1.10^{+1.59}_{-0.70}$ |
| XP | $0.67^{+0.25}_{-0.20}$ | $30.4^{+5.5}_{-4.4}$ | $20.3^{+4.0}_{-3.9}$ | $51.0^{+3.9}_{-3.2}$ | $21.4^{+1.6}_{-1.4}$ | $47.9^{+3.8}_{-3.2}$ | $0.81^{+0.11}_{-0.10}$ | $-0.02^{+0.13}_{-0.15}$ | $0.43^{+0.33}_{-0.27}$ | $1040^{+350}_{-370}$ | $1.19^{+1.51}_{-0.78}$ |
| XPHM | $0.71^{+0.22}_{-0.22}$ | $29.3^{+5.0}_{-3.7}$ | $20.8^{+3.6}_{-4.1}$ | $50.1^{+3.7}_{-3.0}$ | $21.2^{+1.7}_{-1.4}$ | $47.1^{+3.5}_{-2.9}$ | $0.81^{+0.11}_{-0.09}$ | $-0.02^{+0.13}_{-0.15}$ | $0.43^{+0.33}_{-0.28}$ | $1120^{+340}_{-390}$ | $0.89^{+1.87}_{-0.57}$ |
| **GW170104** | | | | | | | | | | | |
| Pv2 | $0.68^{+0.25}_{-0.29}$ | $11.0^{+3.9}_{-1.6}$ | $7.6^{+1.2}_{-1.7}$ | $18.7^{+2.1}_{-0.6}$ | $7.9^{+0.1}_{-0.1}$ | $17.6^{+2.3}_{-0.7}$ | $0.78^{+0.11}_{-0.08}$ | $0.04^{+0.14}_{-0.05}$ | $0.34^{+0.33}_{-0.22}$ | $330^{+100}_{-100}$ | $2.42^{+0.45}_{-1.95}$ |
| XP | $0.72^{+0.23}_{-0.29}$ | $10.8^{+3.4}_{-1.4}$ | $7.7^{+1.1}_{-1.6}$ | $18.6^{+1.7}_{-0.5}$ | $7.9^{+0.1}_{-0.1}$ | $17.5^{+1.8}_{-0.6}$ | $0.77^{+0.12}_{-0.07}$ | $0.04^{+0.11}_{-0.04}$ | $0.34^{+0.37}_{-0.22}$ | $330^{+100}_{-100}$ | $2.43^{+0.44}_{-1.96}$ |
| XPHM | $0.72^{+0.22}_{-0.28}$ | $10.8^{+3.2}_{-1.4}$ | $7.8^{+1.1}_{-1.6}$ | $18.6^{+1.5}_{-0.5}$ | $7.9^{+0.2}_{-0.1}$ | $17.5^{+1.6}_{-0.6}$ | $0.78^{+0.11}_{-0.07}$ | $0.05^{+0.11}_{-0.04}$ | $0.35^{+0.36}_{-0.23}$ | $350^{+100}_{-100}$ | $2.41^{+0.46}_{-1.99}$ |
| **GW170729** | | | | | | | | | | | |
| Pv2 | $0.63^{+0.28}_{-0.22}$ | $51.2^{+12.3}_{-9.1}$ | $32.4^{+7.7}_{-7.7}$ | $84.0^{+11.1}_{-8.7}$ | $35.0^{+4.6}_{-2.9}$ | $78.3^{+10.3}_{-7.7}$ | $0.87^{+0.08}_{-0.11}$ | $0.36^{+0.15}_{-0.21}$ | $0.44^{+0.28}_{-0.24}$ | $2840^{+1060}_{-1070}$ | $0.93^{+1.82}_{-0.64}$ |
| XP | $0.62^{+0.28}_{-0.21}$ | $51.7^{+11.4}_{-9.4}$ | $31.8^{+8.1}_{-7.4}$ | $83.7^{+10.5}_{-8.4}$ | $34.7^{+4.6}_{-3.6}$ | $77.9^{+9.8}_{-7.7}$ | $0.89^{+0.08}_{-0.10}$ | $0.35^{+0.15}_{-0.19}$ | $0.50^{+0.28}_{-0.27}$ | $2890^{+1060}_{-1070}$ | $0.94^{+1.82}_{-0.65}$ |
| TP | $0.6^{+0.29}_{-0.22}$ | $53.2^{+14.9}_{-10.4}$ | $31.8^{+7.8}_{-7.1}$ | $85.8^{+12.1}_{-9.3}$ | $35.4^{+4.2}_{-3.4}$ | $79.9^{+11.3}_{-8.5}$ | $0.89^{+0.09}_{-0.10}$ | $0.36^{+0.13}_{-0.16}$ | $0.50^{+0.25}_{-0.25}$ | $2870^{+1020}_{-990}$ | $0.91^{+1.83}_{-0.60}$ |
| XPHM | $0.58^{+0.29}_{-0.19}$ | $52.5^{+9.2}_{-9.5}$ | $30.5^{+8.4}_{-8.0}$ | $82.6^{+9.7}_{-9.3}$ | $34.1^{+4.6}_{-4.0}$ | $77.5^{+8.8}_{-7.5}$ | $0.86^{+0.09}_{-0.15}$ | $0.30^{+0.17}_{-0.25}$ | $0.42^{+0.33}_{-0.26}$ | $2750^{+1230}_{-1120}$ | $1.05^{+1.68}_{-0.69}$ |
| TPHM | $0.52^{+0.27}_{-0.18}$ | $57.3^{+12.0}_{-10.9}$ | $29.4^{+8.8}_{-7.2}$ | $87.1^{+11.0}_{-9.3}$ | $35.1^{+4.5}_{-3.9}$ | $81.7^{+9.9}_{-8.5}$ | $0.87^{+0.10}_{-0.15}$ | $0.33^{+0.18}_{-0.24}$ | $0.43^{+0.29}_{-0.25}$ | $2490^{+1160}_{-880}$ | $1.08^{+1.62}_{-0.71}$ |
| **GW170809** | | | | | | | | | | | |
| Pv2 | $0.68^{+0.24}_{-0.20}$ | $34.9^{+6.9}_{-5.1}$ | $23.7^{+4.2}_{-4.1}$ | $58.8^{+4.4}_{-3.2}$ | $24.8^{+1.6}_{-1.3}$ | $55.3^{+4.3}_{-3.2}$ | $0.80^{+0.12}_{-0.09}$ | $0.06^{+0.14}_{-0.12}$ | $0.38^{+0.35}_{-0.25}$ | $1030^{+240}_{-300}$ | $2.57^{+0.34}_{-0.48}$ |
| XP | $0.69^{+0.24}_{-0.21}$ | $34.8^{+6.8}_{-5.0}$ | $23.9^{+4.1}_{-4.3}$ | $58.9^{+4.3}_{-3.3}$ | $24.9^{+1.7}_{-1.3}$ | $55.3^{+4.2}_{-3.2}$ | $0.81^{+0.12}_{-0.10}$ | $0.08^{+0.13}_{-0.13}$ | $0.44^{+0.36}_{-0.29}$ | $1050^{+250}_{-320}$ | $2.53^{+0.37}_{-0.47}$ |
| XPHM | $0.72^{+0.22}_{-0.22}$ | $33.8^{+6.5}_{-4.4}$ | $24.1^{+3.9}_{-4.4}$ | $58.1^{+3.9}_{-3.0}$ | $24.6^{+1.5}_{-1.2}$ | $54.5^{+3.7}_{-2.9}$ | $0.82^{+0.12}_{-0.10}$ | $0.08^{+0.14}_{-0.13}$ | $0.46^{+0.35}_{-0.30}$ | $1130^{+240}_{-270}$ | $2.64^{+0.31}_{-0.41}$ |
| TPHM | $0.78^{+0.17}_{-0.22}$ | $32.5^{+5.7}_{-3.5}$ | $25.3^{+3.1}_{-4.2}$ | $58.0^{+3.7}_{-2.9}$ | $24.8^{+1.5}_{-1.2}$ | $54.5^{+3.5}_{-2.8}$ | $0.79^{+0.10}_{-0.08}$ | $0.09^{+0.14}_{-0.13}$ | $0.37^{+0.32}_{-0.24}$ | $1090^{+240}_{-290}$ | $2.64^{+0.31}_{-0.43}$ |
| **GW170814** | | | | | | | | | | | |
| Pv2 | $0.84^{+0.13}_{-0.19}$ | $30.5^{+4.1}_{-2.4}$ | $25.4^{+2.3}_{-3.1}$ | $56.0^{+2.7}_{-2.2}$ | $24.2^{+1.1}_{-0.9}$ | $52.3^{+2.5}_{-2.2}$ | $0.85^{+0.09}_{-0.11}$ | $0.07^{+0.10}_{-0.09}$ | $0.57^{+0.30}_{-0.36}$ | $590^{+120}_{-160}$ | $0.69^{+1.70}_{-0.39}$ |
| XP | $0.83^{+0.13}_{-0.18}$ | $30.3^{+3.9}_{-2.4}$ | $25.2^{+2.2}_{-3.1}$ | $55.5^{+2.6}_{-2.1}$ | $23.9^{+1.1}_{-0.9}$ | $52.1^{+2.5}_{-2.1}$ | $0.80^{+0.10}_{-0.08}$ | $0.06^{+0.10}_{-0.09}$ | $0.40^{+0.35}_{-0.27}$ | $610^{+130}_{-200}$ | $0.68^{+1.75}_{-0.42}$ |
| XPHM | $0.82^{+0.14}_{-0.18}$ | $30.6^{+4.0}_{-2.7}$ | $24.8^{+2.4}_{-3.2}$ | $55.5^{+2.5}_{-2.2}$ | $23.9^{+1.1}_{-0.9}$ | $52.1^{+2.4}_{-2.1}$ | $0.79^{+0.10}_{-0.08}$ | $0.06^{+0.10}_{-0.09}$ | $0.38^{+0.35}_{-0.26}$ | $610^{+140}_{-190}$ | $0.69^{+1.77}_{-0.40}$ |
| XPHM* | $0.83^{+0.13}_{-0.18}$ | $30.5^{+3.8}_{-2.5}$ | $25.3^{+2.3}_{-3.1}$ | $55.8^{+2.6}_{-2.2}$ | $24.0^{+1.1}_{-0.9}$ | $52.3^{+2.4}_{-2.1}$ | $0.81^{+0.10}_{-0.09}$ | $0.08^{+0.10}_{-0.09}$ | $0.44^{+0.33}_{-0.3}$ | $610^{+130}_{-170}$ | $0.67^{+1.68}_{-0.36}$ |
| TPHM | $0.77^{+0.18}_{-0.19}$ | $31.8^{+5.1}_{-3.2}$ | $24.6^{+2.8}_{-3.4}$ | $56.7^{+2.8}_{-2.4}$ | $24.3^{+1.0}_{-0.9}$ | $52.8^{+2.5}_{-2.1}$ | $0.86^{+0.09}_{-0.11}$ | $0.12^{+0.10}_{-0.10}$ | $0.58^{+0.28}_{-0.35}$ | $630^{+100}_{-140}$ | $0.60^{+1.80}_{-0.32}$ |
| **GW170818** | | | | | | | | | | | |
| Pv2 | $0.75^{+0.19}_{-0.20}$ | $35.4^{+5.5}_{-3.9}$ | $26.5^{+3.6}_{-4.2}$ | $62.0^{+3.7}_{-3.1}$ | $26.5^{+1.5}_{-1.3}$ | $57.8^{+3.8}_{-3.1}$ | $0.85^{+0.09}_{-0.11}$ | $-0.11^{+0.14}_{-0.16}$ | $0.56^{+0.27}_{-0.32}$ | $1050^{+310}_{-280}$ | $2.44^{+0.35}_{-0.35}$ |
| XP | $0.78^{+0.17}_{-0.21}$ | $34.9^{+5.8}_{-3.6}$ | $27.2^{+3.4}_{-4.2}$ | $62.3^{+3.9}_{-3.3}$ | $26.7^{+1.6}_{-1.4}$ | $58.1^{+3.8}_{-3.3}$ | $0.86^{+0.10}_{-0.11}$ | $-0.05^{+0.14}_{-0.16}$ | $0.59^{+0.28}_{-0.36}$ | $1130^{+320}_{-330}$ | $2.46^{+0.37}_{-0.37}$ |
| XPHM | $0.78^{+0.17}_{-0.19}$ | $34.8^{+5.3}_{-3.7}$ | $26.9^{+3.3}_{-4.0}$ | $61.8^{+3.7}_{-3.2}$ | $26.5^{+1.6}_{-1.3}$ | $57.7^{+3.6}_{-3.1}$ | $0.85^{+0.10}_{-0.11}$ | $-0.05^{+0.13}_{-0.16}$ | $0.55^{+0.31}_{-0.34}$ | $1190^{+320}_{-340}$ | $2.51^{+0.35}_{-0.37}$ |
| XPHM* | $0.76^{+0.19}_{-0.19}$ | $35.1^{+5.4}_{-3.9}$ | $26.4^{+3.6}_{-4.1}$ | $61.6^{+3.6}_{-3.1}$ | $26.3^{+1.6}_{-1.3}$ | $57.5^{+3.6}_{-3.2}$ | $0.85^{+0.09}_{-0.10}$ | $-0.09^{+0.15}_{-0.17}$ | $0.57^{+0.28}_{-0.32}$ | $1130^{+330}_{-310}$ | $2.48^{+0.35}_{-0.34}$ |
| TPHM | $0.77^{+0.18}_{-0.20}$ | $35.1^{+5.8}_{-3.8}$ | $26.9^{+3.4}_{-4.0}$ | $62.3^{+3.6}_{-3.1}$ | $26.6^{+1.5}_{-1.3}$ | $58.3^{+3.5}_{-3.1}$ | $0.84^{+0.10}_{-0.10}$ | $-0.06^{+0.14}_{-0.15}$ | $0.52^{+0.30}_{-0.31}$ | $1090^{+340}_{-320}$ | $2.47^{+0.36}_{-0.36}$ |
| **GW170823** | | | | | | | | | | | |
| Pv2 | $0.74^{+0.20}_{-0.23}$ | $39.3^{+8.2}_{-5.5}$ | $28.9^{+5.0}_{-5.7}$ | $68.2^{+7.7}_{-5.3}$ | $29.0^{+3.2}_{-2.4}$ | $63.9^{+7.2}_{-5.4}$ | $0.83^{+0.11}_{-0.10}$ | $0.06^{+0.16}_{-0.16}$ | $0.47^{+0.34}_{-0.30}$ | $1940^{+650}_{-710}$ | $1.53^{+1.24}_{-1.16}$ |
| XP | $0.73^{+0.21}_{-0.25}$ | $39.2^{+8.2}_{-5.6}$ | $28.4^{+5.1}_{-6.4}$ | $67.5^{+7.4}_{-5.8}$ | $28.6^{+3.2}_{-2.3}$ | $63.1^{+6.9}_{-5.7}$ | $0.84^{+0.11}_{-0.11}$ | $0.06^{+0.16}_{-0.17}$ | $0.52^{+0.33}_{-0.33}$ | $1930^{+670}_{-700}$ | $1.68^{+1.06}_{-1.27}$ |
| XPHM | $0.77^{+0.18}_{-0.24}$ | $37.7^{+7.2}_{-4.9}$ | $28.5^{+4.7}_{-5.7}$ | $66.1^{+6.9}_{-5.2}$ | $28.2^{+3.0}_{-2.3}$ | $61.8^{+6.4}_{-4.9}$ | $0.84^{+0.10}_{-0.11}$ | $0.07^{+0.16}_{-0.17}$ | $0.52^{+0.32}_{-0.32}$ | $2170^{+650}_{-720}$ | $1.28^{+1.53}_{-0.96}$ |
| TPHM | $0.79^{+0.17}_{-0.23}$ | $37.8^{+7.2}_{-4.7}$ | $29.2^{+4.4}_{-5.2}$ | $67.0^{+7.0}_{-5.3}$ | $28.6^{+3.0}_{-2.3}$ | $62.8^{+6.5}_{-5.0}$ | $0.81^{+0.10}_{-0.09}$ | $0.07^{+0.17}_{-0.16}$ | $0.43^{+0.34}_{-0.27}$ | $2050^{+640}_{-700}$ | $1.80^{+1.03}_{-1.49}$ |

**Table 3.** Inferred source parameter values of the first 10 BBH detections, given as posterior median values with 90% credible intervals. Masses correspond to the source frame. We compare the public IMRPhenomPv2 (Pv2) results (Abbott et al. 2019b) with those obtained with the new generation of phenomenological waveform models. IMRPhenomXP (XP) and the default version of IMRPhenomXPHM (XPHM) are listed for all events, IMRPhenomTPHM (TPHM) is included for high-mass events, IMRPhenomTP (TP) for GW170729 and an alternative version of IMRPhenomXPHM (NNLO angles and final spin version 2) is included for GW170814 and GW170818, indicated by a *.
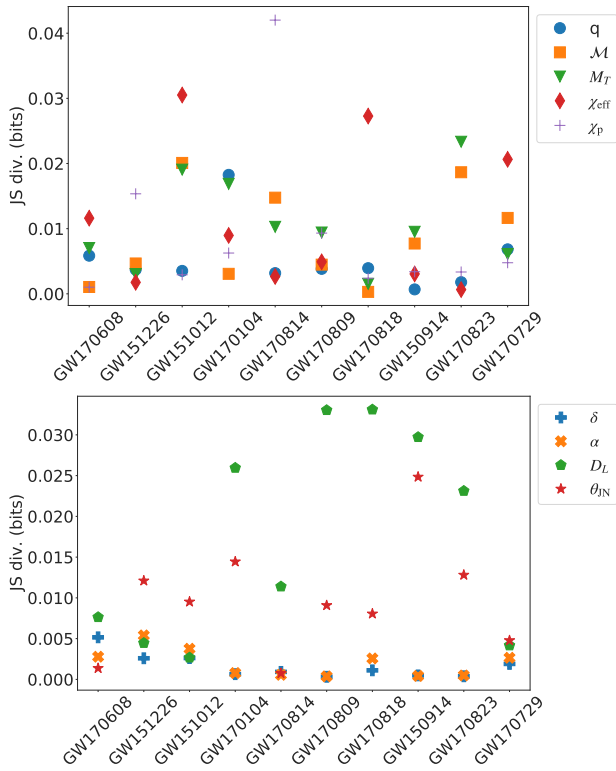
**Figure 2.** Jensen-Shannon divergences (JS div) between IMRPHENOMPv2 (Abbott et al. 2019b) and IMRPHENOMXPHM results of each event, which are ordered from left to right with increasing chirp mass $\mathcal{M}$. Top panel: intrinsic parameters. Bottom panel: extrinsic parameters.

mode effects, nor general waveform systematics effects can be clearly identified for any of the GWTC-1 events.

In the rest of this section, we will discuss all ten BBH events from GWTC-1 individually, grouping them into low-mass (GW151012, GW151226, GW170608), medium-mass (GW170104 and GW170814) and high-mass (GW150914, GW170729, GW170809, GW170818, GW170823) events. Note that we just make this distinction for convenience, and the only events that clearly stand out in terms of total mass are the two lightest ones (GW170608 and GW151226) and the heaviest (GW170729).

### 3.2 Low-mass events

To leading order in the post-Newtonian expansion the duration $\Delta T$ of the inspiral starting at a frequency $f_0$ is

$$\Delta T = \frac{5}{256 f_0 \eta \pi^{8/3}} \left( \frac{f_0 GM}{c^3} \right)^{-5/3},$$

where $M$ is the total mass of the system, $\eta = m_1 m_2/(m_1 + m_2)^2$ is the symmetric mass ratio, $G$ is Newton's constant, and $c$ is the speed of light. The observable part of the waveform is thus significantly longer for lower-mass systems.

In general, the chirp mass is the best measured intrinsic parameter (Cutler & Flanagan 1994), since it is the quantity that appears in the PN waveform at leading order (Blanchet 2006).

In Fig. 4, we compare several parameter estimation runs for GW151012, GW151226 and GW170608 with the standard settings mentioned in Sec. 2.3.2: using distance marginalization, `nlive = 2048` and `nact = 30`. We compare the whole IMRPHE-

**Table 4.** Network matched-filter SNRs with 90% credible intervals and $\log_{10}$ signal-to-noise Bayes factors for runs with the IMRPHENOMX and IMRPHENOMT families, where XPHM* denotes the version with NNLO Euler angles and final spin option 2. The last column shows the Bayes factor against our default IMRPHENOMXPHM runs.

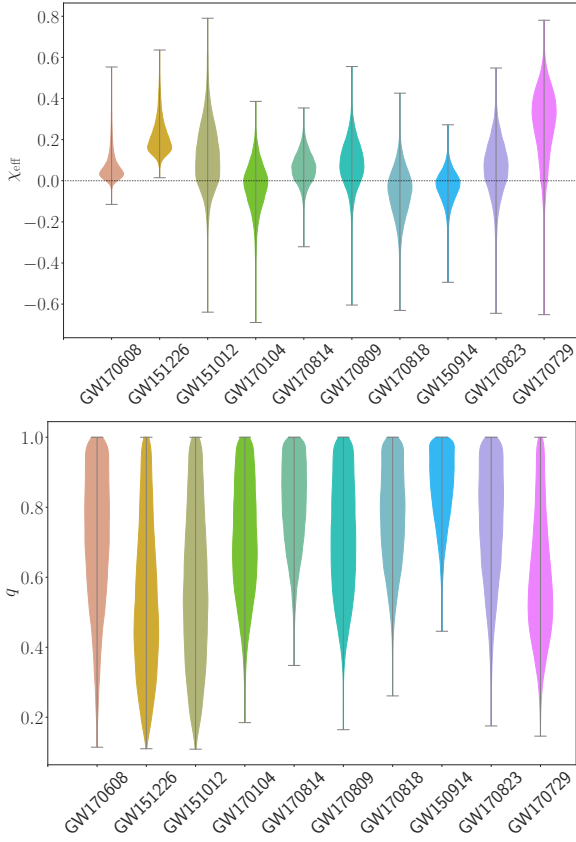| Event | Approx. | SNR | $\log_{10} \mathcal{BF}$ | $\mathcal{BF}_{\text{XPHM}}$ |
|---|---|---|---|---|
| GW150914 | XAS | $25.1^{+0.1}_{-0.1}$ | $124.07 \pm 0.07$ | $0.99^{+0.24}_{-0.19}$ |
| | XHM | $25.1^{+0.1}_{-0.1}$ | $124.08 \pm 0.07$ | $1.01^{+0.24}_{-0.20}$ |
| | THM | $25.1^{+0.1}_{-0.1}$ | $123.83 \pm 0.07$ | $0.57^{+0.14}_{-0.11}$ |
| | XP | $25.1^{+0.1}_{-0.1}$ | $124.05 \pm 0.07$ | $0.95^{+0.25}_{-0.19}$ |
| | TPHM | $25.1^{+0.1}_{-0.1}$ | $123.86 \pm 0.07$ | $0.61^{+0.15}_{-0.12}$ |
| | XPHM | $25.1^{+0.1}_{-0.1}$ | $124.07 \pm 0.07$ | $1.00$ |
| GW151012 | XAS | $9.2^{+0.2}_{-0.3}$ | $10.13 \pm 0.06$ | $0.58^{+0.12}_{-0.10}$ |
| | XHM | $9.4^{+0.3}_{-0.4}$ | $10.27 \pm 0.06$ | $0.81^{+0.16}_{-0.14}$ |
| | THM | $9.4^{+0.3}_{-0.4}$ | $10.37 \pm 0.06$ | $1.02^{+0.20}_{-0.17}$ |
| | XP | $9.2^{+0.2}_{-0.3}$ | $10.25 \pm 0.06$ | $0.77^{+0.15}_{-0.13}$ |
| | XPHM | $9.3^{+0.3}_{-0.4}$ | $10.36 \pm 0.06$ | $1.00$ |
| GW151226 | XAS | $12.4^{+0.2}_{-0.2}$ | $22.04 \pm 0.06$ | $0.26^{+0.06}_{-0.05}$ |
| | XHM | $12.4^{+0.2}_{-0.3}$ | $22.03 \pm 0.06$ | $0.26^{+0.06}_{-0.05}$ |
| | THM | $12.4^{+0.2}_{-0.3}$ | $22.22 \pm 0.06$ | $0.40^{+0.10}_{-0.08}$ |
| | XP | $12.5^{+0.2}_{-0.3}$ | $22.60 \pm 0.07$ | $0.95^{+0.23}_{-0.18}$ |
| | XPHM | $12.6^{+0.3}_{-0.3}$ | $22.62 \pm 0.07$ | $1.00$ |
| GW170104 | XAS | $13.9^{+0.1}_{-0.2}$ | $32.10 \pm 0.06$ | $0.79^{+0.17}_{-0.14}$ |
| | XHM | $13.9^{+0.1}_{-0.2}$ | $32.11 \pm 0.06$ | $0.81^{+0.18}_{-0.14}$ |
| | THM | $13.9^{+0.1}_{-0.2}$ | $31.95 \pm 0.06$ | $0.57^{+0.12}_{-0.10}$ |
| | XP | $13.9^{+0.1}_{-0.2}$ | $32.28 \pm 0.06$ | $1.21^{+0.26}_{-0.21}$ |
| | XPHM | $14.0^{+0.2}_{-0.2}$ | $32.20 \pm 0.06$ | $1.00$ |
| GW170608 | XAS | $15.5^{+0.1}_{-0.2}$ | $39.27 \pm 0.07$ | $0.93^{+0.23}_{-0.18}$ |
| | XHM | $15.5^{+0.1}_{-0.2}$ | $39.24 \pm 0.07$ | $0.86^{+0.21}_{-0.17}$ |
| | THM | $15.5^{+0.1}_{-0.2}$ | $39.33 \pm 0.07$ | $1.06^{+0.26}_{-0.21}$ |
| | XP | $15.5^{+0.1}_{-0.2}$ | $39.29 \pm 0.07$ | $0.96^{+0.24}_{-0.19}$ |
| | XPHM | $15.5^{+0.2}_{-0.2}$ | $39.31 \pm 0.07$ | $1.00$ |
| GW170729 | XAS | $10.7^{+0.3}_{-0.3}$ | $15.97 \pm 0.06$ | $0.64^{+0.14}_{-0.11}$ |
| | T | $10.7^{+0.3}_{-0.3}$ | $16.12 \pm 0.06$ | $0.93^{+0.20}_{-0.16}$ |
| | XHM | $10.9^{+0.3}_{-0.4}$ | $16.32 \pm 0.06$ | $1.47^{+0.31}_{-0.26}$ |
| | THM | $10.9^{+0.3}_{-0.4}$ | $16.35 \pm 0.06$ | $1.56^{+0.34}_{-0.28}$ |
| | XP | $10.7^{+0.3}_{-0.3}$ | $16.03 \pm 0.06$ | $0.74^{+0.16}_{-0.13}$ |
| | TP | $10.8^{+0.3}_{-0.3}$ | $16.13 \pm 0.06$ | $0.94^{+0.20}_{-0.17}$ |
| | TPHM | $10.9^{+0.3}_{-0.4}$ | $16.39 \pm 0.06$ | $2.16^{+0.47}_{-0.39}$ |
| | XPHM | $10.8^{+0.3}_{-0.3}$ | $16.16 \pm 0.06$ | $1.00$ |
| GW170809 | XAS | $12.6^{+0.2}_{-0.2}$ | $24.12 \pm 0.06$ | $0.97^{+0.21}_{-0.17}$ |
| | XHM | $12.6^{+0.2}_{-0.2}$ | $24.10 \pm 0.06$ | $0.93^{+0.21}_{-0.17}$ |
| | THM | $12.6^{+0.2}_{-0.2}$ | $23.95 \pm 0.06$ | $0.66^{+0.15}_{-0.12}$ |
| | XP | $12.6^{+0.2}_{-0.2}$ | $24.13 \pm 0.06$ | $1.00^{+0.22}_{-0.18}$ |
| | TPHM | $12.6^{+0.2}_{-0.2}$ | $23.79 \pm 0.06$ | $0.45^{+0.10}_{-0.08}$ |
| | XPHM | $12.6^{+0.2}_{-0.3}$ | $24.13 \pm 0.06$ | $1.00$ |
| GW170814 | XAS | $17.4^{+0.1}_{-0.2}$ | $53.73 \pm 0.07$ | $1.38^{+0.33}_{-0.27}$ |
| | XHM | $17.5^{+0.1}_{-0.2}$ | $53.69 \pm 0.07$ | $1.25^{+0.30}_{-0.24}$ |
| | THM | $17.4^{+0.1}_{-0.2}$ | $53.57 \pm 0.07$ | $0.95^{+0.23}_{-0.18}$ |
| | XP | $17.4^{+0.1}_{-0.2}$ | $53.62 \pm 0.07$ | $1.06^{+0.26}_{-0.21}$ |
| | TPHM | $17.5^{+0.2}_{-0.2}$ | $54.14 \pm 0.07$ | $3.48^{+0.85}_{-0.68}$ |
| | XPHM* | $17.5^{+0.1}_{-0.2}$ | $53.89 \pm 0.07$ | $1.97^{+0.48}_{-0.39}$ |
| | XPHM | $17.5^{+0.1}_{-0.2}$ | $53.60 \pm 0.07$ | $1.00$ |
| GW170818 | XAS | $11.7^{+0.2}_{-0.2}$ | $19.91 \pm 0.06$ | $0.25^{+0.06}_{-0.05}$ |
| | XHM | $11.8^{+0.2}_{-0.2}$ | $20.11 \pm 0.06$ | $0.39^{+0.09}_{-0.07}$ |
| | THM | $11.8^{+0.2}_{-0.3}$ | $19.97 \pm 0.06$ | $0.29^{+0.07}_{-0.05}$ |
| | XP | $11.9^{+0.2}_{-0.3}$ | $20.27 \pm 0.07$ | $0.57^{+0.13}_{-0.11}$ |
| | TPHM | $11.9^{+0.2}_{-0.3}$ | $20.32 \pm 0.06$ | $0.63^{+0.15}_{-0.12}$ |
| | XPHM* | $12.0^{+0.2}_{-0.3}$ | $20.73 \pm 0.07$ | $1.64^{+0.39}_{-0.31}$ |
| | XPHM | $11.9^{+0.2}_{-0.3}$ | $20.51 \pm 0.06$ | $1.00$ |
| GW170823 | XAS | $12.0^{+0.1}_{-0.1}$ | $22.91 \pm 0.05$ | $0.78^{+0.15}_{-0.13}$ |
| | XHM | $12.0^{+0.1}_{-0.2}$ | $22.90 \pm 0.06$ | $0.77^{+0.15}_{-0.13}$ |
| | THM | $12.0^{+0.1}_{-0.2}$ | $22.81 \pm 0.06$ | $0.63^{+0.12}_{-0.10}$ |
| | XP | $12.0^{+0.1}_{-0.2}$ | $23.07 \pm 0.05$ | $1.13^{+0.22}_{-0.19}$ |
| | TPHM | $12.0^{+0.2}_{-0.2}$ | $22.73 \pm 0.06$ | $0.52^{+0.10}_{-0.09}$ |
| | XPHM | $12.0^{+0.2}_{-0.2}$ | $23.01 \pm 0.06$ | $1.00$ |

**Figure 3.** Violin plots of the IMRPHENOMXPHM default results for $\chi_{\rm eff}$ and q, where the short horizontal lines indicate the 90% credible intervals. The events are sorted by the $\mathcal{M}$ values. Plots created with the PESummary code (Hoy & Raymond 2020).

| | GW151226 | GW170814 | GW170818 |
|---|---|---|---|
| XP vs XAS | $3.63^{+0.86}_{-0.69}$ | $0.77^{+0.18}_{-0.15}$ | $2.27^{+0.53}_{-0.43}$ |
| XPHM vs XHM | $3.85^{+0.92}_{-0.74}$ | $0.80^{+0.19}_{-0.16}$ | $2.54^{+0.60}_{-0.48}$ |
| TPHM vs THM | - | $3.66^{+0.89}_{-0.71}$ | $2.22^{+0.52}_{-0.42}$ |
| TPHM vs XPHM | - | $3.48^{+0.85}_{-0.68}$ | $0.63^{+0.15}_{-0.12}$ |
| TPHM vs XPHM* | - | $1.76^{+0.43}_{-0.35}$ | $0.38^{+0.09}_{-0.07}$ |
| XPHM* vs XHM | - | $1.58^{+0.38}_{-0.31}$ | $4.18^{+0.98}_{-0.80}$ |
| XPHM* vs XPHM | - | $1.97^{+0.48}_{-0.39}$ | $1.64^{+0.39}_{-0.31}$ |

**Table 5.** Comparison of Bayes factors between precessing and aligned-spin IMRPHENOMX and IMRPHENOMT waveform models for GW151226, GW170814 and GW170818.

ferences in the mass posteriors between our results and those of Nitz et al. (2021). Nitz et al. (2021) recover slightly more unequal masses posteriors in their results but consistent with our own results.

From our analysis, Bayes factor differences for different waveform models are within the error estimates, except for a small suppression of IMRPHENOMXAS and IMRPHENOMXP in comparison to IMR-PHENOMXPHM of $0.58^{+0.12}_{-0.10}$ and $0.77^{+0.15}_{-0.13}$ respectively. As can be seen in Fig. 4, the models with higher modes shift the mass ratio posterior toward more unequal masses. Runs with our two different precession versions (NNLO and MSA angles) and with different final spin descriptions show very consistent posterior distributions and Bayes factors that are consistent within the error estimates.

### 3.2.2 GW151226

This event shows support only for positive values of $\chi_{\rm eff}$, and very mild support for precession, with a Bayes factor of $3.85^{+0.92}_{-0.74}$ between IMRPHENOMXHM and IMRPHENOMXPHM in favor of precession, see Table 5. Some differences are seen in Fig. 4 regarding the mass ratio posterior, and in particular IMRPHENOMTHM shifts the mass ratio support toward more equal masses, compared with the other models. Fig. 5 also shows that adding higher mode content shifts the precession parameter $\chi_{\rm p}$ toward larger values, but again all results are consistent within error estimates. See also Appendix A for further consistency tests and comparisons with other results on this event.

### 3.2.3 GW170608

GW170608 is the event with the lowest mass, the closest luminosity distance, and with the highest SNR of the low-mass events. Bayes factors for all waveform models are consistent within the error estimates, and posterior distributions are visually very similar, including for all versions of IMRPHENOMXPHM which we have run (MLA and MSA Euler angles and the four final spin versions for each angle description). JS divergences between the various posteriors back up this finding, with the divergences for all pairs of waveforms and most parameters below 0.02. Only the divergences for spin magnitudes between precessing and aligned-spin waveforms, which are expected to differ more, reach higher values (up to 0.16 between IMRPHENOMXPHM and IMRPHENOMTHM).

### 3.3 Medium-mass events

Fig. 6 provides a comparison of posteriors for several key quantities for the two medium-mass events GW170104 and GW170814. For these, all different models recover essentially the same total mass and $\chi_{\rm eff}$ posterior distributions. We discuss further details below.

NOMX family (the aligned-spin, precessing, dominant and subdominant mode models), the more recent aligned-spin time domain model IMRPHENOMTHM with higher-modes content, and the previous IMRPHENOMPv2 results taken from the fist catalog (Abbott et al. 2019b). As mentioned in Sec. 2.3.1, for these low-mass events we use a sampling rate of 4096 Hz and the GWOSC 16 kHz data sets, and PSDs and calibration envelopes files which go up to 2048 Hz. All models show good agreement in total mass and $\chi_{\rm eff}$. GW151012 and GW151226 are the events with the largest uncertainty in mass ratio.

We note that the initial preprint version of Chia et al. (2022) and also Nitz et al. (2021) have reported results for GW151226 that are in tension with ours, including multi-modal posteriors. In order to further increase the confidence in our results we have therefore also performed tests with a different sampling method in Bilby and with the independent LALInference code for the three lowest mass events. These checks confirm our results as discussed in Appendix A. We also show how lowering the sampling rate can result in bi-modal posteriors. As discussed in the appendix, the remaining differences between our results and the published results of Chia et al. (2022), as well as the results of Nitz et al. (2021), seem to be largely consistent with differences in prior choices.

### 3.2.1 GW151012

This is the low-mass event with the smallest SNR ($9.3^{+0.3}_{-0.4}$ for IMRPHENOMXPHM) and Bayes factor, see Table 4. Apart from GW151226, this is the only event where we find some visual dif-
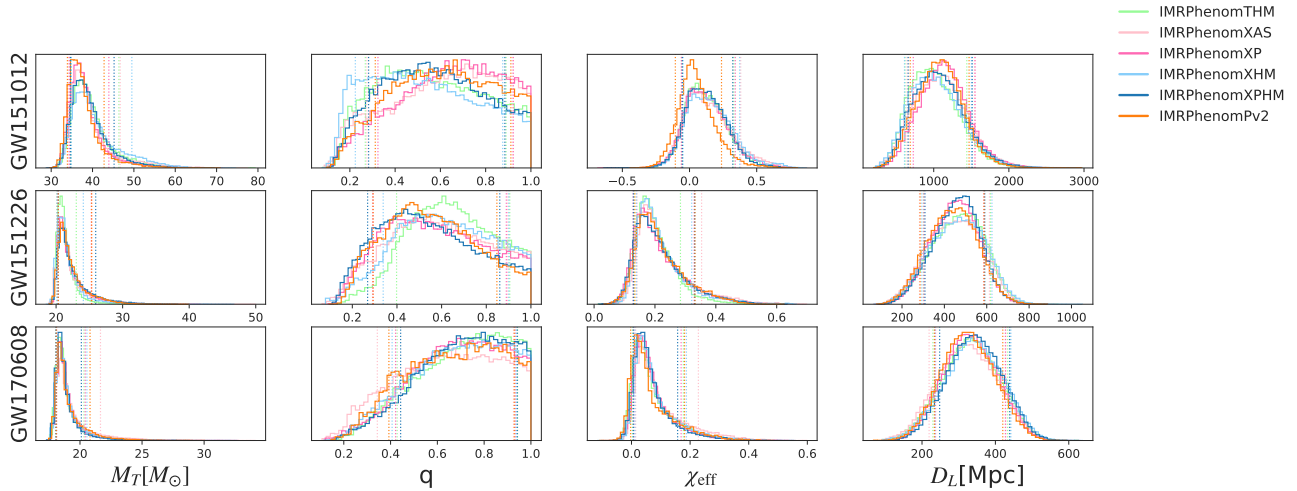
**Figure 4.** Posterior distributions for total mass, mass ratio, effective spin and luminosity distance for all three low-mass events and different waveform models.
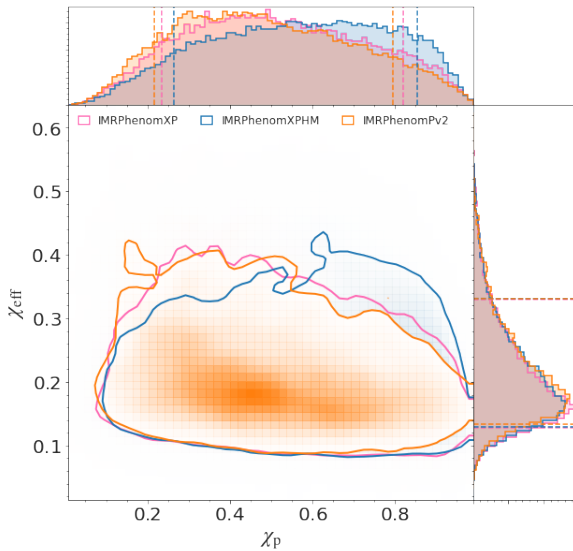


**Figure 5.** Joint posterior distributions for $\chi_p$ and $\chi_{eff}$ of GW151226, estimated with IMRPHENOMXP (pink), IMRPHENOMXPHM (blue) and IMRPHENOMPv2 (orange) (Abbott et al. 2019b). Here and in similar figures throughout the paper, the central panel shows the 2D joint posterior with contours marking 90% credible intervals, while the smaller panels on top and to the right show the corresponding 1D distributions for the individual parameters, with the 90% credible interval indicated by the dashed lines.

### 3.3.1 GW170104

This is the first event of the O2 observing run, and the fourth loudest BBH event at an SNR of $14.0^{+0.2}_{-0.2}$ for all the waveform models that we have considered. Bayes factors for different models are consistent within error estimates, apart from the Bayes factor of IMRPHE-NOMTHM, which is suppressed relative to IMRPHENOMXPHM by $0.57^{+0.12}_{-0.10}$, see Table 4. (IMRPHENOMTPHM was only run for higher-mass events.) The higher-mode models shift the posterior marginally toward more equal masses and larger distance. We also find that IMRPHENOMXPHM versions give consistent results regarding posterior distributions and Bayes factors, and we do not list these results separately but rather refer to our data release for full details.

### 3.3.2 GW170814

GW170814 is the second loudest BBH event with an SNR of $17.5^{+0.1}_{-0.2}$, and a well recovered sky position due to being observed by both LIGO (Hanford and Livingston) detectors and the Virgo detector. Indeed GW170814 was the first BBH published as a coincident observation between the three LIGO–Virgo detectors (Abbott et al. 2017b).

Apart from the effective precession parameter $\chi_p$, the posterior distributions for different waveform models are rather similar for other quantities, with only marginal differences for masses and mass ratio. Just as GW151226, and to a lesser degree GW170818, GW170814 does however show marginal support for precession, as expressed by the Bayes factors listed in Table 5, and it is an interesting case from the point of view of waveform systematics. GW170814 shows the largest value of the JS divergence when comparing our IM-RPHENOMXPHM results to the GWTC-1 IMRPHENOMPv2 results, corresponding to $JS_{\chi_p} = 0.042$ for the effective in-plane spin $\chi_p$. The default version of IMRPHENOMXPHM in fact recovers a lower value of $\chi_p$ than IMRPHENOMPv2, see Table 3. However it turns out that when using the NNLO Euler angles rather than the MSA angles for IMRPHENOMXPHM, a larger value of $\chi_p = 0.44^{+0.33}_{-0.30}$ is recovered, and also a larger Bayes factor of $1.97^{+0.48}_{-0.39}$ with respect to the default IMRPHENOMXPHM version (see Tables 3 and 4). Indeed both IMRPHENOMXAS and IMRPHENOMXHM recover marginally better Bayes factors than the default version of IMRPHENOMXPHM. Also IMRPHENOMTPHM has a higher Bayes factor, $3.48^{+0.85}_{-0.68}$ against the default version of IMRPHENOMXPHM and $1.76^{+0.43}_{-0.35}$ against the NNLO angles version of IMRPHENOMXPHM. IMRPHENOMTPHM also recovers a higher value of $\chi_p = 0.58^{+0.28}_{-0.35}$, very close to the value for IMRPHENOMPv2, even though the non-precessing time-domain model IMRPHENOMTHM is marginally disfavored when comparing to IMRPHENOMXHM, with a Bayes factor ratio of about 1.3 in Table 4. These results suggest that this event is a case where the default choice for modelling precession in IMRPHENOMXPHM, the MSA description of the precession Euler angles, is less accurate than the NNLO description, and thus an improved precession treatment for the frequency domain is needed to obtain robust results for this event. Furthermore IMRPHENOMTPHM might recover an even higher Bayes factor, were it not for the shortcomings of the model in the inspiral, already in the absence of precession, and it will thus be interesting
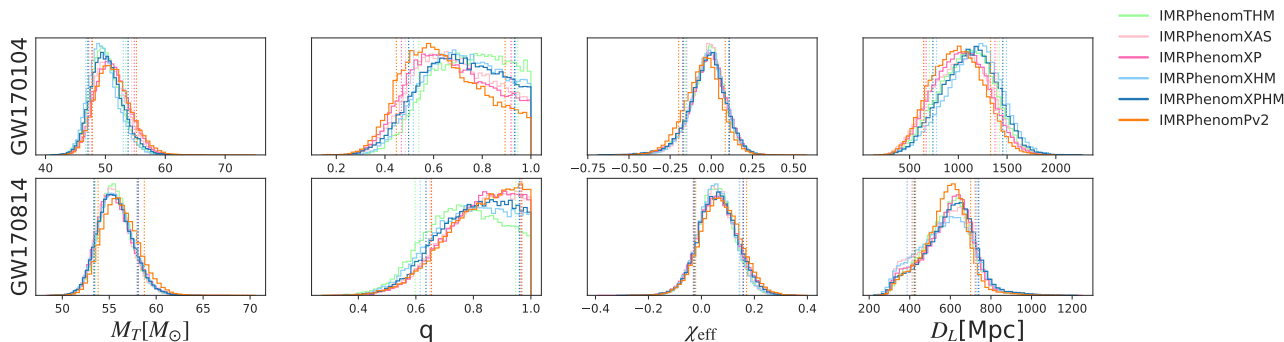
**Figure 6.** Posterior distributions for total mass, mass ratio, effective spin and luminosity distance are compared for the medium-mass events and different waveform models.

to re-analyze this event again in the future with improved waveform models.

Comparisons of posterior distributions for the mass ratio, total mass, $\chi_{\mathrm{eff}}$ and $\chi_{\mathrm{p}}$ for this event using our different precessing waveform models are shown in Fig. 7.

### 3.4 High-mass events

Fig. 8 shows a comparison of posteriors for several key quantities. For high mass events, we also add the IMRPHENOMTPHM waveform model to our comparisons.

#### 3.4.1 GW150914

GW150914 was the first detection of a GW signal (Abbott et al. 2016) and has the highest SNR of events consistent with BBH mergers to date (in the O1, O2 (Abbott et al. 2019b) and O3a (Abbott et al. 2021c) observing runs).

Bayes factors in Table 4 are consistent between different models within error estimates, apart from IMRPHENOMTHM and IMRPHE-NOMTPHM, which are disfavored by factors of $0.57^{+0.14}_{-0.11}$ (THM) and $0.61^{+0.15}_{-0.12}$ (TPHM) with respect to the default IMRPHENOMXPHM version. Posterior results show only very marginal deviations, see Fig. 8 and Tables 3, 4. The same holds for all the eight versions of IMRPHENOMXPHM we have tested (NNLO and MSA Euler angles and four final spin versions for each angle description): the default version of IMRPHENOMXPHM has the largest Bayes factor, and the most disfavored version of IMRPHENOMXPHM (NNLO angles with final spin version FS2) has a Bayes factor of $0.69^{+0.17}_{-0.13}$ with respect to the default version, which is still broadly consistent within the error estimates. Posteriors between different IMRPHENOMXPHM versions are in general very similar and without noteworthy changes.

#### 3.4.2 GW170729

This is the most massive and most distant BBH merger detected during O1 and O2, with substantial support for unequal masses (Abbott et al. 2019b). In consequence, a first set of parameter estimation runs with different waveform models with higher-mode content was performed in Chatziioannou et al. (2019). It is thus particularly interesting to compare our results with the IMRPHENOMX and IMRPHE-NOMT families against these earlier results, and also against those obtained by Payne et al. (2019) through approximate reweighting of IMRPHENOMD results to the NRHYBSUR3DQ8 target waveform. In addition, we test how posterior distributions for this event change

with different mass and distance priors. GW170729 is also interesting since its recovered effective spin is positive within error estimates. Unfortunately the SNR is relatively low, at $10.8^{+0.3}_{-0.3}$ for IMRPHE-NOMXPHM and $10.9^{+0.3}_{-0.4}$ for IMRPHENOMTPHM. Only GW151012 has lower SNR within our set of ten events. Consequently, posterior distributions for GW170729 are relatively broad. We find that all eight versions of IMRPHENOMXPHM yield Bayes factors and posterior distributions that are consistent within error estimates.

The more massive component of the binary, with central values just above 50, $M_\odot$ from various posteriors, see Table 3, is situated near the lower end of the pair-instability supernova (PISN) mass gap. That hypothetical gap is placed between approximately 52 and 133 solar masses by Woosley (2017), with the edges however uncertain – see Woosley & Heger (2021) for a more recent overview. IMRPHENOMTPHM, which recovers the largest Bayes factor, also recovers the highest value for the larger mass, $m_1 = 57.3^{12.0}_{-10.9}\, M_\odot$, potentially situating it beyond the onset of the mass gap.

*a. Higher-mode content:* As demonstrated by the Bayes factors in Table 6, we find weak evidence of higher-order mode content in this event.

These Bayes factors in favour of higher modes are larger (in the range 1.9–2.9 including uncertainties) for aligned-spin IMRPHE-NOMX and for precessing IMRPHENOMT models, but below 2 for precessing IMRPHENOMX and aligned-spin IMRPHENOMT models. This different behaviour of the two model families is consistent with IMRPHENOMX having a more accurate aligned-spin sector but IMR-PHENOMT providing improved treatment of precession as described in Sec. 2.2. In either case, our Bayes factors are lower than the value of 5.1 found by Chatziioannou et al. (2019) between the older aligned-spin models IMRPHENOMHM and IMRPHENOMD, but the aligned-spin IMRPHENOMX and precessing IMRPHENOMT values are slightly larger than the value of 1.4 obtained by Abbott et al. (2019b) using the RapidPE method (Pankow et al. 2015; Lange et al. 2018) with a catalog of NR and NRSUR7DQ2 (Blackman et al. 2017) waveforms. We thus find that when using improved waveform models the Bayes factors are more consistent with the RapidPE method than what was found by Chatziioannou et al. (2019), and the Bayes factor in favor of higher-mode content is lower.

In Fig. 9 we see that IMRPHENOMPv2 and the corresponding newer precessing model for the dominant quadrupole, IMRPHENOMXP, give very similar results, while the non-precessing multi-mode waveform models recover lower mass ratio and higher luminosity distance, as was also observed by Chatziioannou et al. (2019) and Payne et al. (2019). The non-precessing likelihood-reweighted results of Payne et al. (2019) (not plotted) are in very good agreement with our IM-
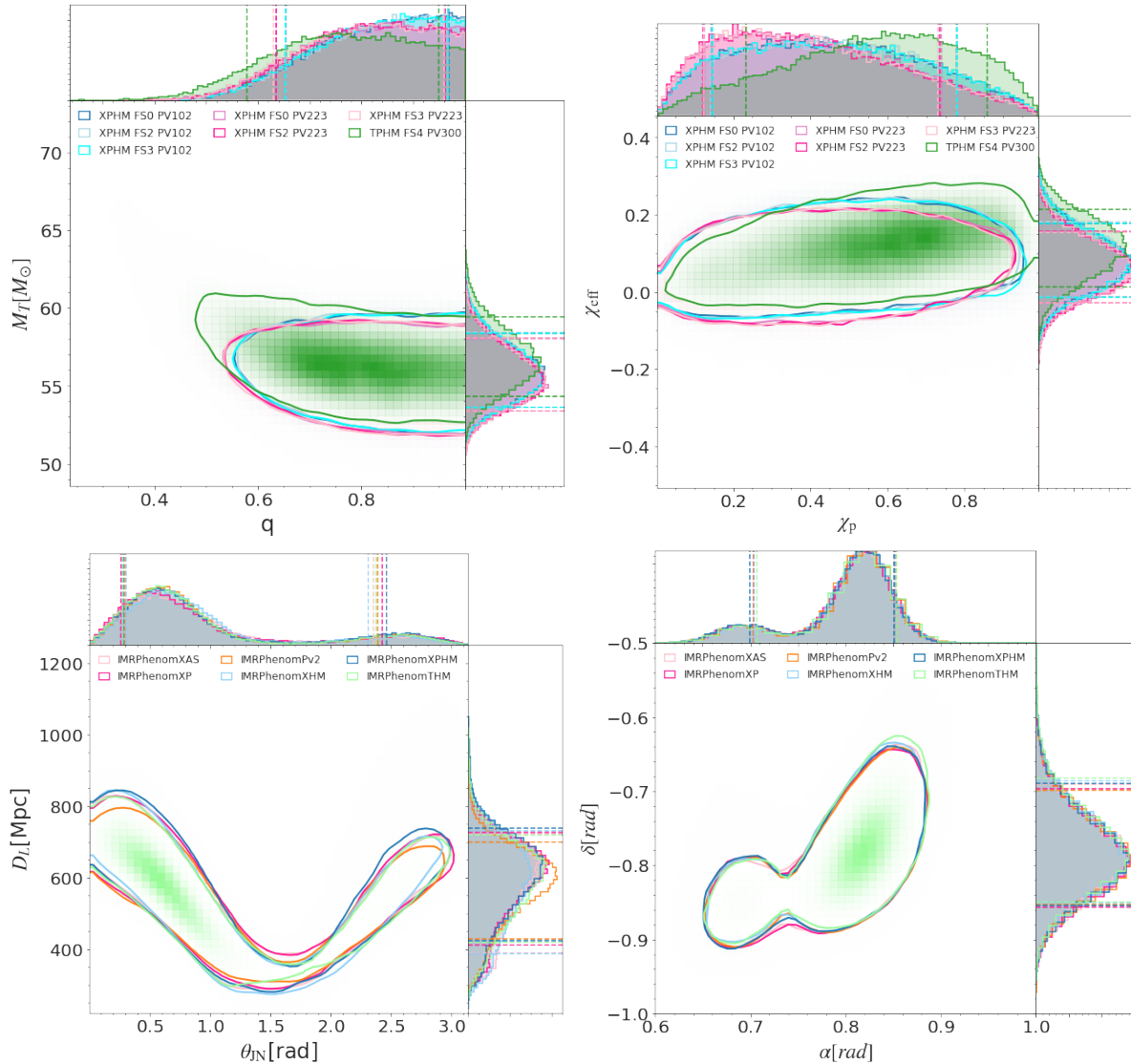
**Figure 7.** Comparison of posteriors for GW170814, indicating 90% credible intervals. Upper panel: total mass, mass ratio, and spin parameters for the MSA (blue) and NNLO (pink) versions of IMRPHENOMXPHM and for IMRPHENOMTPHM (green). Lower panel: distance and inclination (left) and sky position (right) are shown for the dominant-mode models IMRPHENOMXAS, IMRPHENOMXP (light and dark pink) and IMRPHENOMPv2 (orange), and the multi-mode models IMRPHENOMXHM, IMRPHENOMXPHM (light and dark blue) and IMRPHENOMTHM (green).

RPHENOMXHM and IMRPHENOMTHM posteriors for most parameters, with the peak for $\chi_{\mathrm{eff}}$ shifted down towards 0 but its posterior still well consistent within 90% credible intervals. Adding not only higher modes but precession, we obtain more evidence for unequal masses for IMRPHENOMTPHM, as seen in Fig. 10.

In Figs. 9 and 10 we also directly compare the frequency and time domain families IMRPHENOMX and IMRPHENOMT. We can see that the distribution of likelihood values across the posterior for IMR-PHENOMTPHM is shifted toward larger values, and indeed we obtain a Bayes factor of $2.16^{+0.47}_{-0.39}$ in favor of IMRPHENOMTPHM when compared to IMRPHENOMXPHM, while the aligned-spin versions, IMRPHENOMTHM and IMRPHENOMXHM, give much more similar results.

*b.    Comparison of different mass and distance priors:* Given that GW170729 is the furthest event of the first catalog and has support for unequal masses, we compare the different distance and mass prior choices explained in Sec. 2.3.3, in order to see their influence

| Hypotheses | Model properties | PhenomX | PhenomT |
|---|---|---|---|
| HM vs $\ell = 2 = |m|$ | aligned | $2.27^{+0.48}_{-0.40}$ | $1.69^{+0.36}_{-0.30}$ |
| | precessing | $1.35^{+0.29}_{-0.24}$ | $2.31^{+0.50}_{-0.41}$ |

**Table 6.** Comparison of Bayes factors between models of the IMRPHENOMX and IMRPHENOMT families concerning the hypothesis that the signal of GW170729 contains higher modes, for the precessing and non-precessing waveform models. The IMRPHENOMX aligned check compares IMRPHENOMXHM against IMRPHENOMXAS, the IMRPHENOMX precessing check compares IMRPHENOMXPHM against IMRPHENOMXP, the IMRPHENOMT aligned check compares IMRPHENOMTHM against IMRPHENOMT, and the IMRPHENOMT precessing check compares IMRPHENOMTPHM against IMRPHENOMTP.

on the posterior distributions. For this study we run IMRPHENOMX-PHM using `nlive = 2048` and `nact = 10`. Fig. 11 shows that there is a shift to lower distance and masses when we change from the power-law distance prior to the prior that is uniform in comoving
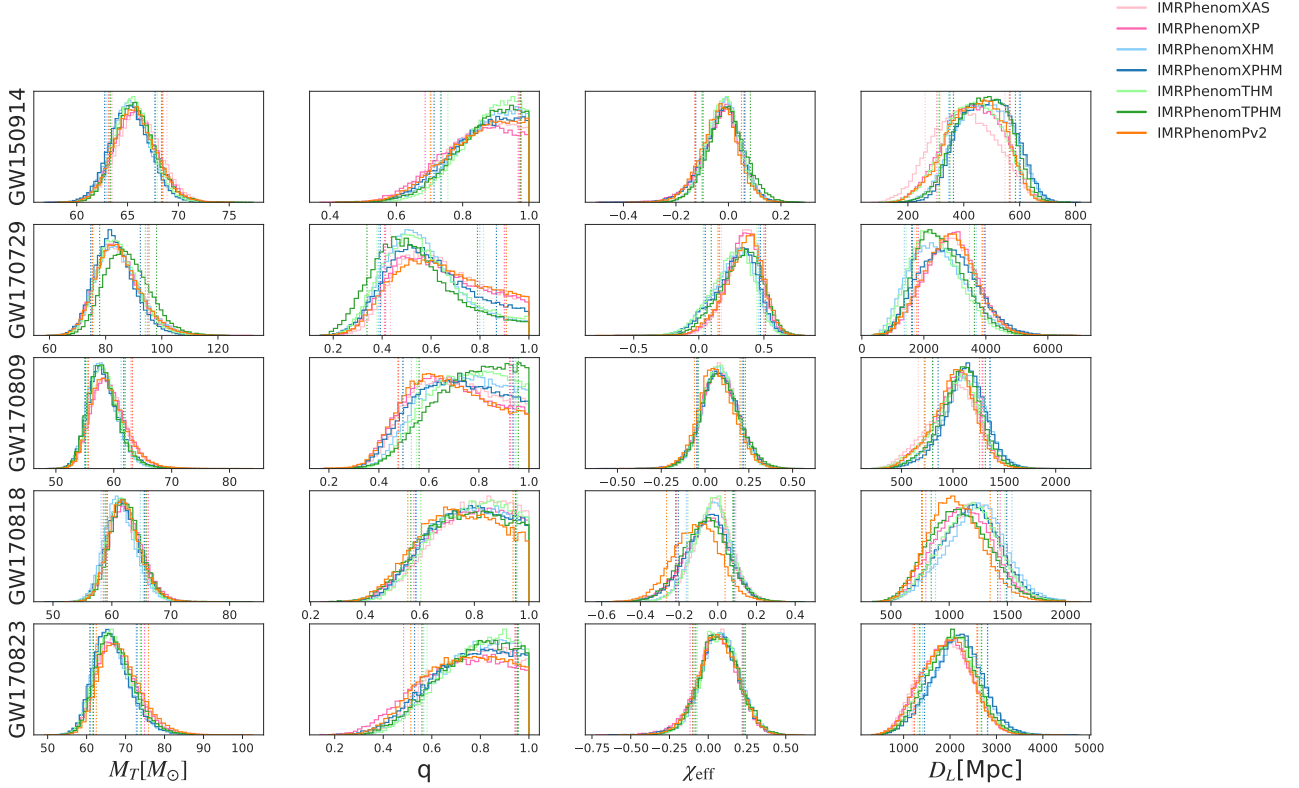
**Figure 8.** Posteriors for total mass, mass ratio, $\chi_{\mathrm{eff}}$ and luminosity distance are shown for high-mass events and different waveform models.

source-frame volume. As discussed in Sec. 2.3.3, the standard power-law distance prior distributes the mergers uniformly throughout an Euclidean universe, which is only a good approximation at relatively close distances (small redshifts). However, at larger distances, it is not a good approximation, and including proper cosmological information in the prior is required, which the comoving volume prior achieves better (Abbott et al. 2021c; Romero-Shaw et al. 2020). On the other hand, varying the mass priors does not affect the posteriors noticeably. This is as expected because masses are more strongly constrained by the data, hence the prior has a smaller influence on the posterior – at least for the GWTC-1 events.

In order to quantify how much the posteriors change, we compute the JS divergence between the results. In Table 7 we show, for every possible combination of results, which parameter has the highest divergence. If we compare the results that have different mass priors, all divergences are below 0.003, meaning that the posterior results only vary very minimally. Comparing different distance priors we obtain higher divergences, where the luminosity distance is the parameter with the largest values, $JS_{d_L} = 0.015$ and 0.012.

### 3.4.3 GW170809

The posterior for this event is consistent with vanishing spins (both $\chi_{\mathrm{eff}}$ and $\chi_p$), and $\theta_{JN}$ is relatively well measured to correspond to a face-off orientation. The masses are slightly more unequal than for GW150914 and the total mass is slightly smaller. The event is however located at approximately twice the distance, corresponding to only about half the SNR of $12.6^{+0.2}_{-0.3}$ for IMRPhenomXPHM. Similar to GW150914, the Bayes factors for the MSA-angle based versions of IMRPhenomXPHM, IMRPhenomXHM and IMRPhenomXAS are

also consistent within error estimates, while the NNLO-based versions of IMRPhenomTHM and even more of IMRPhenomTPHM are disfavored. MSA versions recover slightly higher $\chi_p$ than the NNLO-based versions, otherwise we find no visible differences in posteriors between IMRPhenomXPHM versions.

In Fig. 8 one can see that the waveforms with subdominant modes have more support for equal masses. This effect is strongest for IMRPhenomTPHM, which is however disfavored by the Bayes factor. In addition, the higher-mode models marginally shift the total mass to lower values, and the luminosity distance to higher values.

Using the RapidPE method with a catalog of higher-mode NR waveforms, Abbott et al. (2019b) found a revised $\chi_{\mathrm{eff}}$ distribution which is symmetric about a median value of zero, while our higher-mode models show no such effect.

### 3.4.4 GW170818

GW170718 is a heavy event most consistent with a negative effective spin, $\chi_{\mathrm{eff}} = -0.11^{+0.14}_{-0.16}$ for IMRPhenomPv2 (Abbott et al. 2019b). However, in our analysis, using the IMRPhenomX and IMRPhenomT families, the effective spin parameter is shifted toward zero, as can be seen in Fig. 12. Unfortunately, this is also the third faintest event, which limits the information that can be extracted, compared to other events. There is weak evidence for precession, with a Bayes factor for XPHM against XHM of $2.54^{+0.60}_{-0.48}$, see Table 5.

The IMRPhenomX and IMRPhenomT families show good consistency in Figs. 8 and 12, and a comparison of posteriors for different IMRPhenomXPHM versions with the results for IMRPhenomPv2 is shown in Fig. 13. IMRPhenomXPHM with NNLO Euler angles obtains a weakly higher Bayes factor than the default version, by a
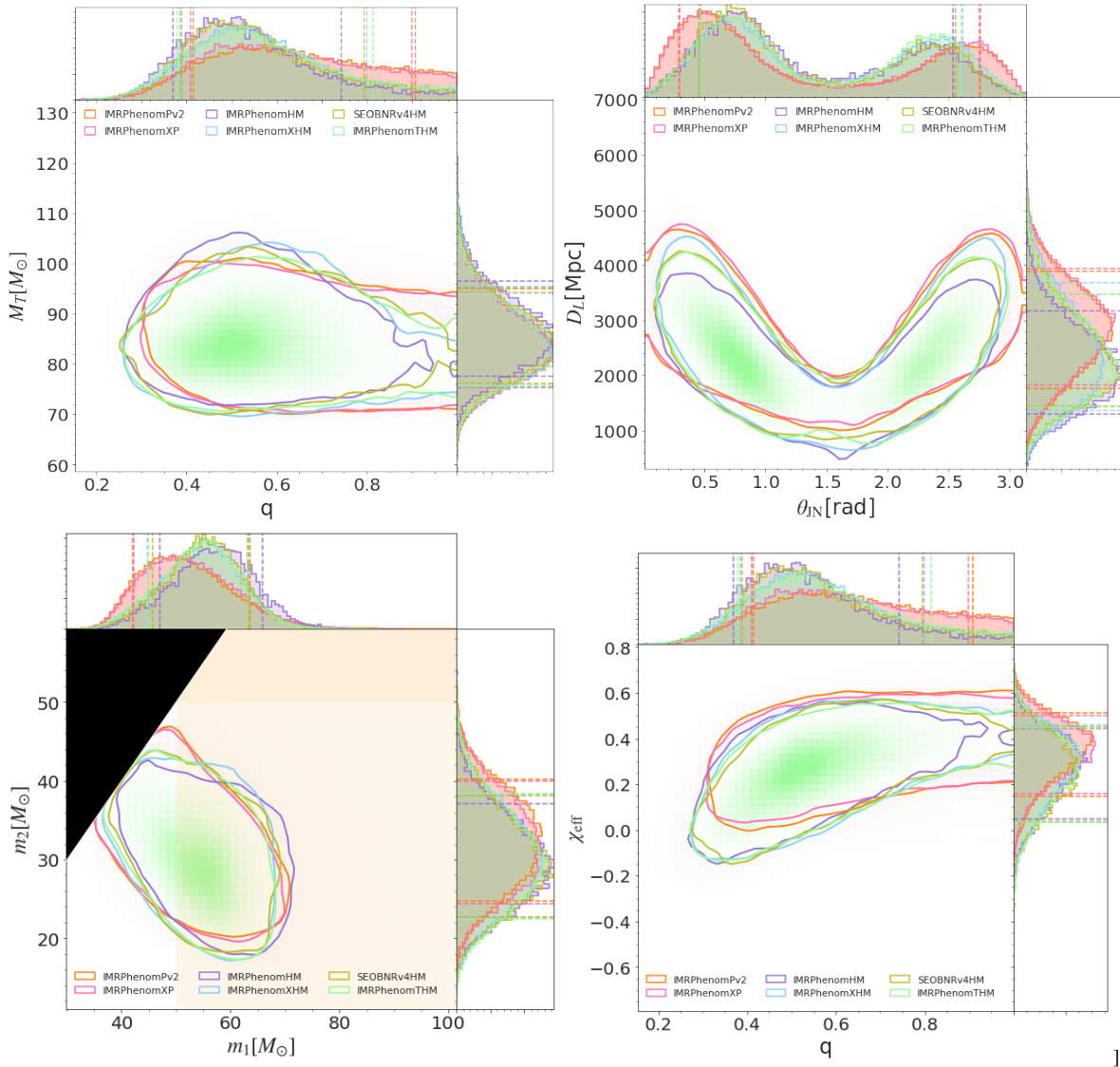
**Figure 9.** Posterior distributions of GW170729 for the total mass, mass ratio, luminosity distance, inclination, component masses, mass ratio and effective spin, using the dominant-mode models IMRPHENOMPv2 (orange, Abbott et al. 2019b) and IMRPHENOMXP (light pink), the subdominant-mode models IMRPHENOMHM (purple), SEOBNRv4HM (olive green, Chatziioannou et al. 2019), IMRPHENOMXHM (light blue) and IMRPHENOMTHM (light green). All contours are at 90% credible intervals.

| | PowerLaw - FlatCompMass | UnifComovVo - FlatCompMass | PowerLaw - UniCompMass | PowerLaw - UniformQ |
|---|---|---|---|---|
| PowerLaw - FlatCompMass | - | $JS_{D_L} = 0.0152$ | $JS_{ra} = 0.0029$ | $JS_{\mathcal{M}} = 0.0009$ |
| UnifComovVo - FlatCompMass | | - | $JS_{D_L} = 0.0128$ | $JS_{D_L} = 0.0119$ |
| PowerLaw - UniCompMass | | | - | $JS_{ra} = 0.0030$ |
| PowerLaw - UniformQ | | | | - |

**Table 7.** Maximum Jensen-Shannon (JS) divergence values between the posterior distributions for GW170729 estimated with IMRPHENOMXPHM using different priors in distance and masses, as explained in Sec. 2.3.3. `PowerLaw` refers to a power-law distance prior. `UnifComovVo` denotes a distance prior that is uniform in comoving volume. `FlatCompMass` is used for a prior uniform in $q \leq 1$ and $\mathcal{M}_c$, with samples afterwards reweighted to a flat prior in component masses. `UniCompMass` denotes directly using a flat prior in component masses while still sampling in $q$ and $\mathcal{M}_c$. Finally, `UniformQ` is a prior flat in mass ratio $Q \geq 1$.

factor $1.64^{+0.39}_{-0.31}$ (see Tables 4 and 5), and the corresponding posteriors for $\chi_{\text{eff}}$ and the luminosity distance are more similar to the IMRPHENOMPv2 results.

### 3.4.5 GW170823

GW170823 is the second most massive and second most distant event. The SNR is relatively low at $12.0^{+0.1}_{-0.2}$ for IMRPHENOMXPHM, limiting the information that can be extracted. Indeed, in Fig. 8
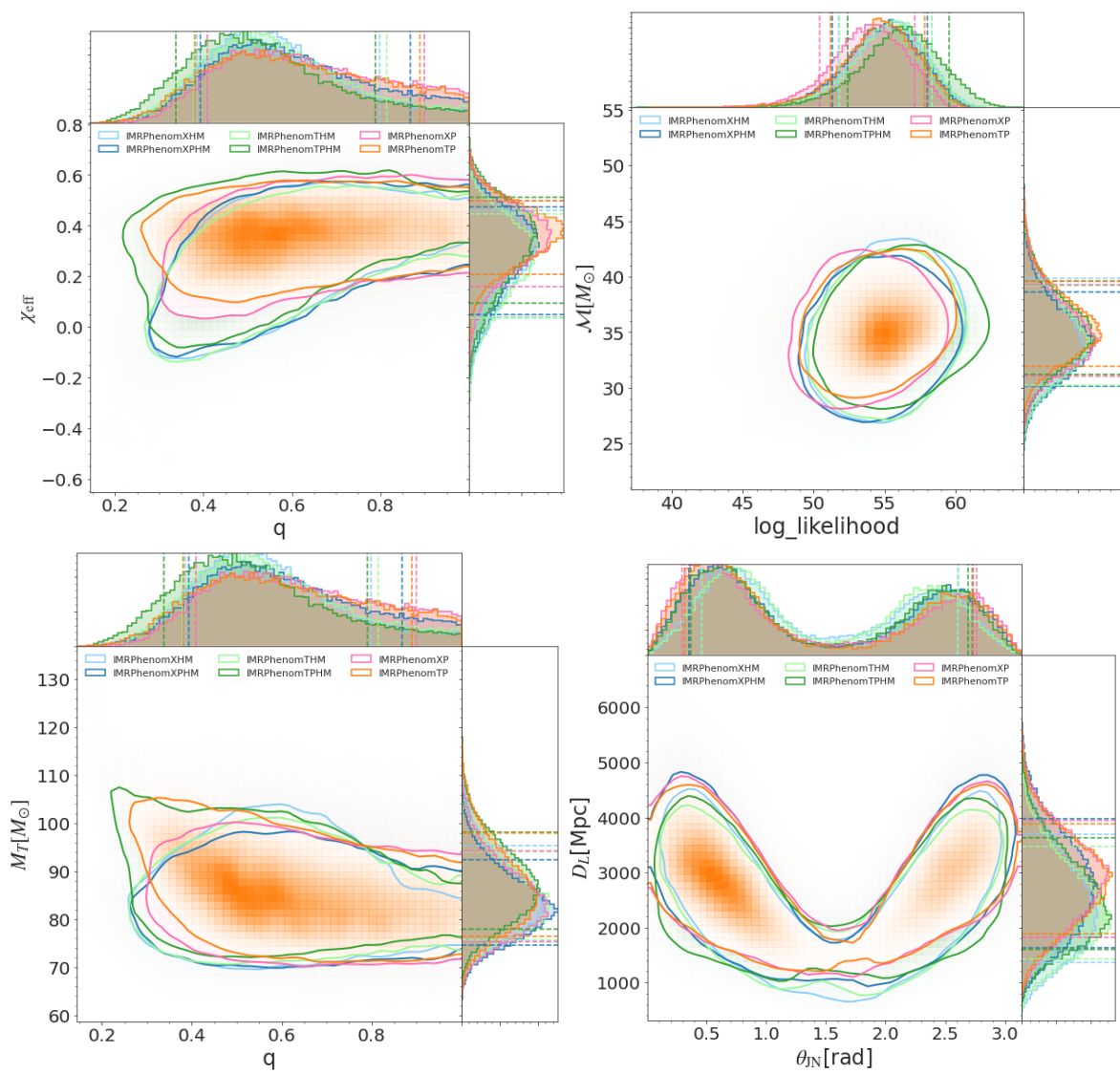
**Figure 10.** Posterior distributions of GW170729 for the mass ratio, total mass, chirp mass, effective spin parameter, luminosity distance, inclination and log likelihood using the dominant-mode models IMRPʜᴇɴᴏᴍXP and IMRPʜᴇɴᴏᴍTP (light pink and orange respectively) and the subdominant-mode models IMRPʜᴇɴᴏᴍXHM (light blue), IMRPʜᴇɴᴏᴍXPHM (dark blue) and IMRPʜᴇɴᴏᴍTHM (light green), IMRPʜᴇɴᴏᴍTPHM (dark green). All contours are at 90% credible intervals.

we find very good agreement between the posteriors obtained with different waveform models.

Due to the large distance we compare two distance priors, uniform in comoving volume and proportional to the luminosity distance squared, as we do in Sec. 3.4.2 for GW170729. As expected, we find that the differences between both priors are smaller than for GW170729. The luminosity distance is the parameter with the highest JS divergence when comparing the two distance prior choices; however, all divergence values are lower than 0.007. The impact of the distance prior is very minor, as illustrated by Fig. 14, where we compare the runs with different priors all using IMRPʜᴇɴᴏᴍXPHM.

We find the Bayes factors between different IMRPʜᴇɴᴏᴍXPHM versions to be consistent within error estimates, and find no notable differences in the respective posterior distributions for the source parameters.

## 4 CONCLUSIONS

In this paper we have obtained parameter estimation results for the BBH events in GWTC-1 with the new fourth generation of phenomenological waveform models. This has served as a stress test before using these models for new events observed in the second half of the third observing run (O3b), as presented in Abbott et al. (2021b), and in a reanalysis of O1, O2 and O3a events as part of Abbott et al. (2021a). Our results overall confirm the expectations that subdominant harmonics and waveform systematics only have minor consequences for parameter estimation results for these ten events, and no major problems have been encountered using the IMRPʜᴇɴᴏᴍX and IMRPʜᴇɴᴏᴍT families.

However, we have put particular emphasis on providing a more in-depth study of this set of events than previous works, considering different waveform models along with tests of convergence and comparisons of different sampling algorithms. This has been particularly important for the event GW151226, where the initial arXiv version
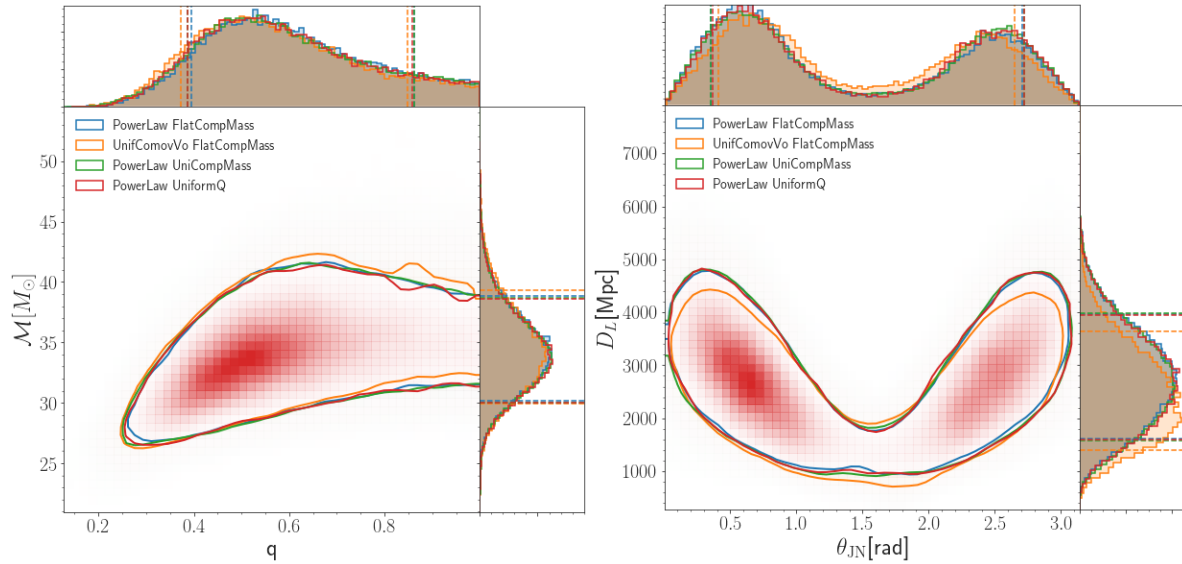
**Figure 11.** Comparison of posterior distributions for GW170729 estimated with IMRPhenomXPHM using different distance and mass priors. Here the string PowerLaw denotes our standard power-law distance prior. UnifComovVo denotes a distance prior that is uniform in comoving volume. FlatCompMass runs use a prior uniform in $q \leq 1$ and $\mathcal{M}_c$, with samples afterwards reweighted to a flat prior in component masses. UniCompMass denotes the more recent method to directly ensure a flat prior in component masses while still sampling in $q$ and $\mathcal{M}_c$. Finally, UniformQ denotes a flat prior in mass ratio $Q \geq 1$. All these priors are described in Sec. 2.3.3. All contours are at 90% credible intervals.
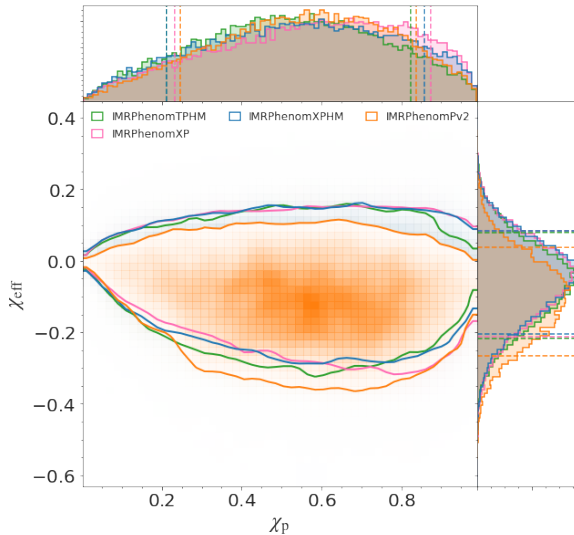


**Figure 12.** Comparison of posterior distributions for $\chi_p$ and $\chi_{eff}$ of GW170818, estimated with IMRPhenomTPHM (green), IMRPhenomXP (pink), IMRPhenomXPHM (blue) and IMRPhenomPv2 (orange, from Abbott et al. (2019b)). 90% credible intervals are indicated.

of Chia et al. (2022) had suggested that other results, including our analysis, may be inaccurate. But various recent publications (Nitz & Capano 2021; Vajpeyi et al. 2022), as well as the final journal version of Chia et al. (2022), now find broadly consistent results: the event is compatible with a broader range of mass ratios than reported in the initial LVC studies, and while results by different groups disagree in the relative posterior weight assigned to more or less unequal configurations, consensus appears to have been reached that these remaining differences are consistent with differences in priors.

For the other GWTC-1 event where results have been less stable in past studies, the heavy GW170729, we find consistent results with

previous, more limited higher-modes analyses by Chatziioannou et al. (2019) and Payne et al. (2019). Results between different waveforms are consistent at 90% credible level, but the IMRPhenomTPHM model tends to prefer higher primary masses, making it more likely to be a signal from a binary containing a black hole in the PISN mass gap. Still, the uncertainties both on the GW parameter estimation and the astrophysical predictions are too large to make a decisive claim on this interpretation.

Given the high computational cost of Bayesian parameter estimation, not many other GW events have been analysed in similar detail. In two parallel papers (Colleoni et al. 2021; Estellés et al. 2021c)[1] Our re-analysis shows excellent agreement between the latest generations of non-precessing waveform models, where subdominant harmonics are calibrated to NR simulations, and broad consistency between different precessing models with subdominant harmonics. In particular we find good agreement between our IMRPhenomX and IMRPhenomT families. GW190521 is the shortest signal detected so far (only approximately 0.1 seconds long), and much more challenging: due to the lack of information in such a short signal, waveforms that correspond to very different sources fit the signal well, and the posterior distribution is multi-modal. We can however show good consistency between different sampling codes – LALInference (Veitch et al. 2015) and Bilby (Ashton et al. 2019; Smith et al. 2020) – and different choices of mass priors. We also confirm our expectation that for events where only the merger and ringdown are observable, our description of precession in the

---

[1] Note that we have updated our results for GW190412 to include the precessing time domain model IMRPhenomTPHM in Estellés et al. (2021a).

we have provided even more detailed analyses of two selected events from the third observing run O3: GW190412 (Abbott et al. 2021c, 2020a,b) and GW190521 (Abbott et al. 2020b,c). The results found in these two papers complement our findings in the present paper:

GW190412 was the first event where subdominant spherical harmonic modes could be clearly identified.
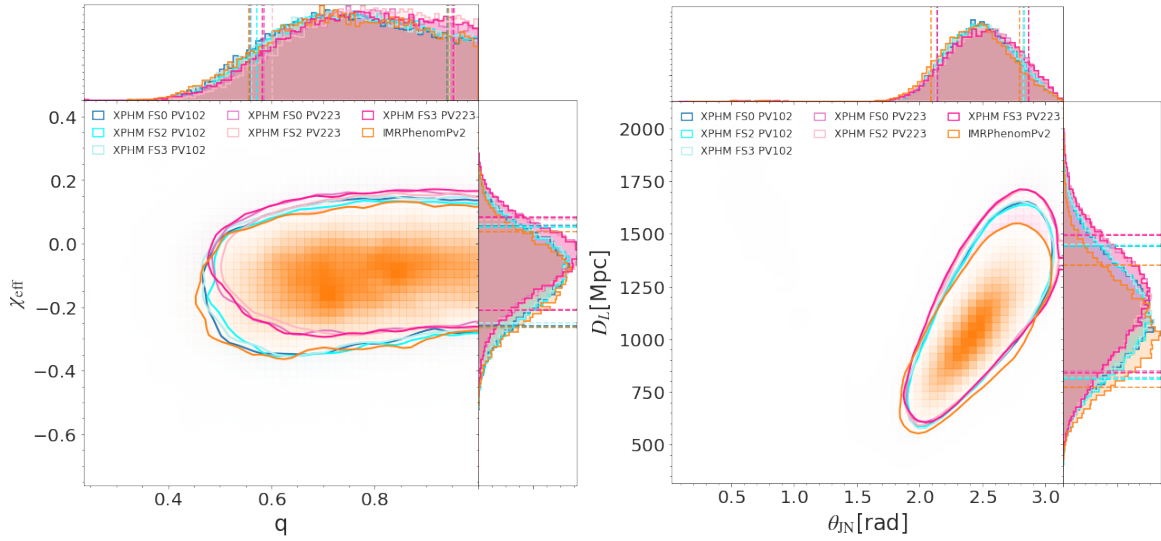
**Figure 13.** Comparison of posterior distributions of GW170818 for the mass ratio, the effective spin parameter, the luminosity distance and the inclination using the MSA (blues) and NNLO (pinks) IMRPHENOMXPHM versions and IMRPHENOMPv2 (orange) results from Abbott et al. (2019b). 90% credible intervals are indicated.
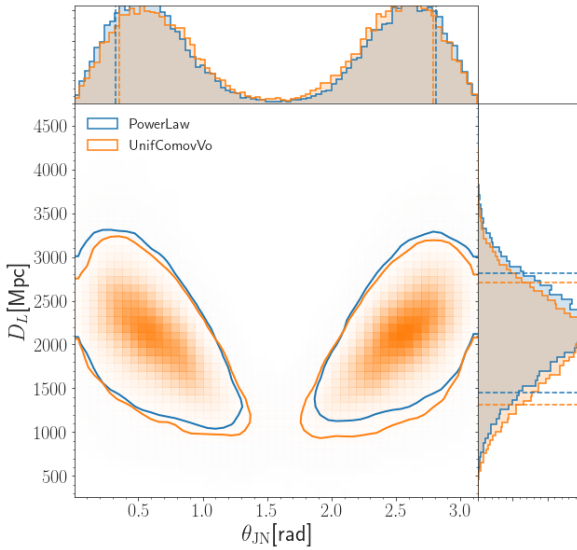


**Figure 14.** Comparison of the posterior distributions for GW170823 estimated with IMRPHENOMXPHM using different distance priors (proportional to the luminosity distance squared and uniform in the comoving volume and source frame time, labeled by `PowerLaw` and `UnifComovVo` respectively).

time domain (IMRPHENOMTPHM) improves significantly over our frequency-domain model (IMRPHENOMXPHM).

Meanwhile, the GWTC-1 events studied in this paper suggest that for lower-mass signals there is no simple preference for either the IMRPHENOMX or IMRPHENOMT family: The IMRPHENOMTHM and IMRPHENOMTPHM models are typically characterized by lower Bayes factors when compared to their IMRPHENOMX counterparts. The reason is likely that for lower-mass events the higher accuracy of the phase of the $(2, 2)$ mode for IMRPHENOMX makes the frequency-domain model more accurate than the current time-domain version. The exception is GW170814, which has a total mass of about 56 solar masses, but does show some support for precession, see the Table

5 of Bayes factors. In this case IMRPHENOMTPHM is also favored over IMRPHENOMXPHM.

Given that the IMRPHENOMX models are both computationally cheaper and simpler to use (time-domain models require a careful choice of starting frequency to ensure consistent start times between different modes and a careful treatment of Fourier transforms when using matched filtering in the frequency domain), use of the IMRPHENOMX family is preferred for lower-mass events. Still, the example of GW170814 suggests that a comparison with IMRPHENOMTPHM may provide further insight into such events when support for spin precession is found.

For higher masses, preference tilts toward IMRPHENOMTPHM, as the accuracy of the phasing during inspiral is less relevant. For GW170729, which has a median value of the total mass in the detector frame above 120 solar masses ($122.7^{+14.3}_{-15.7}\, M_\odot$ for the IMRPHENOMXPHM default run), both IMRPHENOMXPHM and IMRPHENOMTPHM provide consistent results, but support for IMRPHENOMTPHM is marginally larger than for IMRPHENOMXPHM. For GW170729 the better accuracy of the inspiral phasing for IMRPHENOMX is much less relevant, while the treatment of precession in the time-domain model is superior. For GW170729 the advantage of IMRPHENOMTPHM is however still rather small, while for the even heavier GW190521 we find (Estellés et al. 2021c) that IMRPHENOMTPHM gives significantly better results.

Finally, the computational efficiency of the IMRPHENOMX and IMRPHENOMT models has allowed us to also systematically test the influence of variations in priors and sampler settings. We have verified that changes in the mass and distance priors only have very marginal influence on results. Perhaps most interestingly, we have checked that, choosing a prior that emphasizes large mass ratios, we have not found any multi-modal posteriors for any of the GWTC-1 events, as has been the case for GW190521 (Nitz & Capano 2021; Estellés et al. 2021c).

In summary, we consider that detailed studies of waveform systematics and parameter estimation robustness, as performed in this paper, are essential for the robust interpretation of GW detections. Along with providing publicly available data sets that allow to repro-

duce an analysis, such investigations are an important step forward for the reliable astrophysical interpretation of GWs.

## DATA AVAILABILITY

In order to complement this manuscript, we provide a data release (Mateu-Lucena et al. 2021) which contains all the posterior samples and Bilby configuration files developed in this work. It also contains PSDs and calibration envelope files for the events where publicly released datasets from GWOSC (LIGO Scientific Collaboration, Virgo Collaboration 2019a) are not sufficient.

## REFERENCES

Abbott B. P., et al., 2016, Phys. Rev. Lett., 116, 061102
Abbott B. P., et al., 2017a, Class. Quant. Grav., 34, 104002
Abbott B. P., et al., 2017b, Phys. Rev. Lett., 119, 141101
Abbott B. P., et al., 2019a, Technical Report LIGO-P1900124, Supplementary parameter estimation sample release for GWTC-1: Comparisons to NR, https://dcc.ligo.org/LIGO-P1900124/public. LIGO Project, https://dcc.ligo.org/LIGO-P1900124/public
Abbott B. P., et al., 2019b, Phys. Rev., X9, 031040
Abbott R., et al., 2020a, Phys. Rev. D, 102, 043015
Abbott R., et al., 2020b, Phys. Rev. Lett., 125, 101102
Abbott R., et al., 2020c, Astrophys. J. Lett., 900, L13
Abbott R., et al., 2021a, preprint (arXiv:2108.01045)
Abbott R., et al., 2021b, preprint (arXiv:2111.03606)
Abbott R., et al., 2021c, Phys. Rev. X, 11, 021053
Ade P. A. R., et al., 2016, A&A, 594, A13
Ajith P., et al., 2011, Phys. Rev. Lett., 106, 241101
Ashton G., et al., 2019, Astrophys. J. Suppl., 241, 27
Babak S., Taracchini A., Buonanno A., 2017, Phys. Rev. D, 95, 024010
Blackman J., et al., 2017, Phys. Rev. D, 96, 024058
Blanchet L., 2006, Living Rev. Relativ., 9, 4
Blanchet L., Buonanno A., Faye G., 2011, Phys. Rev. D, 84, 064041
Bohé A., Hannam M., Husa S., Ohme F., Puerrer M., Schmidt P., 2016, Technical Report LIGO-T1500602, PhenomPv2 - Technical Notes for LAL Implementation, https://dcc.ligo.org/LIGO-T1500602. LIGO Project, https://dcc.ligo.org/LIGO-T1500602
Bohé A., et al., 2017, Phys. Rev. D, 95, 044028
Buonanno A., Damour T., 1999, Phys. Rev., D59, 084006
Buonanno A., Damour T., 2000, Phys. Rev., D62, 064015
Buonanno A., Iyer B., Ochsner E., Pan Y., Sathyaprakash B. S., 2009, Phys. Rev. D, 80, 084043
Chatziioannou K., Klein A., Yunes N., Cornish N., 2013, Phys. Rev. D, 88, 063011
Chatziioannou K., et al., 2019, Phys. Rev. D, 100, 104015
Chia H. S., Olsen S., Roulet J., Dai L., Venumadhav T., Zackay B., Zaldarriaga M., 2022, Phys. Rev. D, 106, 024009
Colleoni M., Mateu-Lucena M., Estellés H., García-Quirós C., Keitel D., Pratten G., Ramos-Buades A., Husa S., 2021, Phys. Rev. D, 103, 024029
Cornish N. J., Littenberg T. B., 2015, Class. Quant. Grav, 32
Cotesta R., Buonanno A., Bohé A., Taracchini A., Hinder I., Ossokine S., 2018, Phys. Rev. D, 98, 084028
Cotesta R., Marsat S., Pürrer M., 2020, Phys. Rev. D, 101, 124040
Cutler C., Flanagan E. E., 1994, Phys. Rev. D, 49, 2658
Estellés H., Husa S., Colleoni M., Keitel D., Mateu-Lucena M., García-Quirós C., Ramos-Buades A., Borchers A., 2020, preprint (arXiv:2012.11923)
Estellés H., Colleoni M., García-Quirós C., Husa S., Keitel D., Mateu-Lucena M., Planas M. L., Ramos-Buades A., 2021a, preprint (arXiv:2105.05872)
Estellés H., Ramos-Buades A., Husa S., García-Quirós C., Colleoni M., Haegel L., Jaume R., 2021b, Phys. Rev. D, 103, 0124060
Estellés H., et al., 2021c, ApJ, 924, 70
García-Quirós C., Colleoni M., Husa S., Estellés H., Pratten G., Ramos-Buades A., Mateu-Lucena M., Jaume R., 2020, Phys. Rev. D, 102, 064002
García-Quirós C., Husa S., Mateu-Lucena M., Borchers A., 2021, Class. Quant. Grav., 38, 015006
Hannam M., Schmidt P., Bohé A., Haegel L., Husa S., et al., 2014, Phys.Rev.Lett., 113, 151101
Healy J., Lousto C. O., Lange J., O'Shaughnessy R., 2020, Phys. Rev. D, 102, 124053
Hoy C., Raymond V., 2020, preprint (arXiv:2006.06639)
Husa S., Khan S., Hannam M., Pürrer M., Ohme F., Jiménez Forteza X., Bohé A., 2016, Phys. Rev. D, 93, 044006
Khan S., Husa S., Hannam M., Ohme F., Pürrer M., Jiménez Forteza X., Bohé A., 2016, Phys. Rev. D, 93, 044007
Khan S., Chatziioannou K., Hannam M., Ohme F., 2019, Phys. Rev. D, 100, 024059
Khan S., Ohme F., Chatziioannou K., Hannam M., 2020, Phys. Rev. D, 101, 024056
LIGO Scientific Collaboration 2020, LIGO Algorithm Library - LAL-Suite, free software (GPL), https://doi.org/10.7935/GT1W-FZ16, doi:10.7935/GT1W-FZ16
LIGO Scientific Collaboration, Virgo Collaboration 2018, O2 C02 Calibration Uncertainty Results Review/Update/Summary (https://dcc.ligo.org/LIGO-G1800319/public), https://dcc.

ligo.org/public/0150/G1800319/001/G1800319-v1_O2C02_UncertaintyReview.pdf

LIGO Scientific Collaboration, Virgo Collaboration 2019c, Calibration uncertainty envelope release for GWTC-1 (https://dcc.ligo.org/LIGO-P1900040/public), https://dcc.ligo.org/LIGO-P1900040/public

LIGO Scientific Collaboration, Virgo Collaboration 2019a, Gravitational Wave Open Science Center, https://www.gw-openscience.org

LIGO Scientific Collaboration, Virgo Collaboration 2019b, Power Spectral Densities (PSD) release for GWTC-1 (https://dcc.ligo.org/LIGO-P1900011/public), https://dcc.ligo.org/LIGO-P1900011/public

LIGO Scientific Collaboration, Virgo Collaboration 2021, BayesWave Software, https://git.ligo.org/lscsoft/bayeswave

Lange J., O'Shaughnessy R., Rizzo M., 2018, preprint (arXiv:1805.10457)

London L., et al., 2018, Phys. Rev. Lett., 120, 161102

Marsat S., Baker J. G., 2018, preprint (arXiv:1806.10734)

Marsat S., Blanchet L., Bohe A., Faye G., 2013. (arXiv:1312.5375)

Mateu-Lucena M., Husa S., Colleoni M., Estellés H., García-Quirós C., Keitel D., Planas M. L., Ramos-Buades A., 2021, Data release for the paper "Adding harmonics to the interpretation of the black hole mergers of GWTC-1", doi:10.5281/zenodo.4740753

Nagar A., et al., 2018, Phys. Rev. D, 98, 104052

Nitz A. H., Capano C. D., 2021, ApJ. Lett., 907, L9

Nitz A. H., Capano C. D., Kumar S., Wang Y.-F., Kastha S., Schäfer M., Dhurkunde R., Cabero M., 2021, ApJ., 922, 76

O'Shaughnessy R., London L., Healy J., Shoemaker D., 2013, Phys. Rev. D, 87, 044038

Ossokine S., et al., 2020, Phys. Rev. D, 102, 044055

Pankow C., Brady P., Ochsner E., O'Shaughnessy R., 2015, Phys. Rev. D, 92, 023002

Payne E., Talbot C., Thrane E., 2019, Phys. Rev. D, 100, 123017

Pratten G., Husa S., Garcia-Quiros C., Colleoni M., Ramos-Buades A., Estelles H., Jaume R., 2020, Phys. Rev. D, 102, 064001

Pratten G., et al., 2021, Phys. Rev. D, 103, 104056

Romero-Shaw I. M., et al., 2020, Mon. Not. Roy. Astron. Soc., 499, 3295

Roulet J., Olsen S., Mushkin J., Islam T., Venumadhav T., Zackay B., Zaldarriaga M., 2022, preprint (arXiv:2207.03508)

Santamaría L., et al., 2010, Phys. Rev. D, 82, 064016

Schmidt P., Hannam M., Husa S., Ajith P., 2011, Phys. Rev. D, 84, 024046

Schmidt P., Hannam M., Husa S., 2012, Phys. Rev. D, 86, 104063

Schmidt P., Ohme F., Hannam M., 2015, Phys. Rev. D, 91, 024043

Skilling J., 2006, Bayesian Analysis, 1, 833

Smith R. J. E., Ashton G., Vajpeyi A., Talbot C., 2020, Mon. Not. Roy. Astron. Soc., 498, 4492–4502

Speagle J. S., 2020, Mon. Not. Roy. Astron. Soc., 493, 3132–3158

Taracchini A., et al., 2014, Phys. Rev. D, 89, 061502

Thrane E., Talbot C., 2019, Pub. Astron. Soc. Aust., 36, e010

Vajpeyi A., Smith R., Thrane E., 2022, preprint (arXiv:2203.13406)

Vallisneri M., Kanner J., Williams R., Weinstein A., Stephens B., 2015, J. Phys. Conf. Ser., 610, 012021

Veitch J., et al., 2015, Phys. Rev. D, 91, 042003

Venumadhav T., Zackay B., Roulet J., Dai L., Zaldarriaga M., 2019, Phys. Rev. D, 100, 023011

Vinciguerra S., Veitch J., Mandel I., 2017, Class. Quant. Grav., 34, 115006

Woosley S. E., 2017, ApJ., 836, 244

Woosley S. E., Heger A., 2021, ApJL., 912, L31

Yoo A. B., Jette M. A., Grondona M., 2003, in Feitelson D., Rudolph L., Schwiegelshohn U., eds, Job Scheduling Strategies for Parallel Processing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 44–60

## APPENDIX A: CONVERGENCE TEST AND COMPARISON OF SAMPLING METHODS

In order to set up appropriate sampler settings, we performed a convergence study for all the different events and waveform approx-

imants. That convergence test consists in comparing several sampler settings from low to high resolution in order to get the most converged result but with an affordable computational cost. In our case we test two different numbers of live points for the Nested Sampling algorithm, nlive = 512 and 2048, and also the number of autocorrelation times: nact = 10, 30, and for completeness we also test nact = 50 for IMRP$_{\text{HENOM}}$XHM and IMRP$_{\text{HENOM}}$THM.

For all the BBH events where we performed the convergence test, we get the same conclusion: default settings of nlive = 2048 and nact = 30 result in sufficiently accurate posteriors at moderate computational cost. In Fig. A1 we show a comparison for GW170729 using different waveform models and sampler settings. Using nlive = 512 both the histograms and the 2D plots show fluctuations, indicating insufficient convergence. However, when we increase nlive to 2048, the posterior distributions are much smoother. Comparing the number of autocorrelation times when we use a large number of nlive, changing the nact value does not make an important difference.

We can get a more quantitative perspective from Table A1, where we show the maximum JS divergence values corresponding to the posteriors of Fig. A1. As commented above, the highest differences appear when we change from low to high nlive, but not when we change the nact configuration. We decided to use nlive = 2048 and nact = 30 as the default setting because if we compare with the same run but using nact = 50, we obtain an indistinguishable distribution: the divergences between both posterior distributions are smaller than 0.001 bits.

In addition we perform further tests for the three lowest-mass events, motivated by the results in Nitz et al. (2021) and in particular the initial arXiv version of Chia et al. (2022) for GW151226, which are, or were, to some degree in tension with our results.

As mentioned in the introduction, a large part of this tension has been resolved with improvements in the sampling technique employed by Chia et al. (2022), as they report in their final journal version. A consensus seems to now have emerged that the event is consistent with a broader range of mass ratios than found by initial LVC studies, but no clear preference for unequal masses persists, and that remaining differences seem to be largely consistent with differences in prior choices. In addition, we note that the results of Nitz et al. (2021) have been obtained at half our sampling rate (i.e. only 2048 Hz, according to the data release of Nitz et al. 2021), and the discrepant results in the initial arXiv version of Chia et al. (2022) had been obtained at an even lower sampling rate of 1024 Hz (using the sampling code described in Venumadhav et al. 2019), while for the final results (using the updated method from Roulet et al. 2022) they used 2048 Hz. See also the discussion of this point in the note at the end of the journal version of Chia et al. (2022), phrased in terms of the Nyquist frequency (half the sampling rate).

A reduced sampling rate implies a lower cutoff frequency corresponding to the Nyquist frequency, and will reduce the higher-mode content. While the effect on the recovered SNR is very small, we find a significant effect on the posterior distributions of some quantities, as we will discuss below.

We first compare the results for this event from our standard Bilby runs with alternative sampling choices: As previously shown in Fig. 11 for GW170729, we again compare our standard method of reweighting the standard-prior runs to a prior that is flat in component masses, against directly sampling with the alternative flat-in-component-masses prior (using the two Bilby classes UniformInComponentsChirpMass and UniformInComponentsMassRatio, as discussed in Sec. 2.3.3). Furthermore we compare with results for the LALInference code

(Veitch et al. 2015), which is part of the LALSuite (LIGO Scientific Collaboration 2020) package for GW data analysis, using its implementation of parallel-tempered Markov Chain Monte Carlo (MCMC) sampling. LALInference samples in mass ratio and chirp mass, reweighting to a prior that is flat in component masses as described in Veitch et al. (2015). We use essentially standard LAL-Inference settings with eight temperatures, but a large number of 60 independent chains in order to have a broad distribution of chains covering the parameter space. These runs employ the same dataset, PSDs and calibration envelopes as the Bilby runs, discussed in Sec. 3.2, but we do not employ the distance marginalization used for our Bilby runs.

In Fig. A2 we show a comparison between different runs for GW151226, employing two different sampling rates (our default 4096 Hz and a lower one at 2048 Hz), the two different mass priors, and different sampling settings (`nlive` of 2048 and 4096 at fixed `nact`=30), all with Parallel Bilby, as well as the standard LALInference run as described above, and the results reported in Nitz et al. (2021) (results data for Chia et al. (2022) have not been released). We observe that while all our runs are consistent, employing the Bilby prior defined by the `UniformInComponentsChirpMass` and `UniformInComponentsMassRatio` classes (labelled "UniComp-Mass" in the plot) reproduces LALInference results best, as expected, meaning that our results are well converged. For this event we also observe that differences with respect to the results reported in Nitz et al. (2021) are broadly consistent with the differences we observe between different sampling rates in our runs.

We also perform runs that aim to reproduce the bimodal results of the initial arXiv version of Chia et al. (2022) with LALInference, using their initial sampling rate of 1024 Hz. In Fig. A3 we can observe that indeed reducing our sampling rate by a factor of four we find a bimodality in the mass ratio distribution, and increased support for smaller mass ratio $q$ (we can not show direct comparisons since the initial posterior data of Chia et al. (2022) have not been released). In this figure we also plot the results of a LALInference run with the standard 4096 Hz sampling rate but with a prior restricted to $q \leq 0.4$, aimed at improving the sampling of more unequal mass ratios. We find that this restricted run shows very good agreement with that using the standard unrestricted prior, enhancing the posterior region corresponding to higher spins, but remaining consistent with the low-q portion of the standard run and not producing an additional peak. We thus conclude that the bimodality initially found by Chia et al. (2022) could at least partially be explained by the reduction of the higher-mode content when a low sampling rate is used that will cut off part of the merger–ringdown signal. However, while in the final journal version of Chia et al. (2022) they also increased their sampling rate (to 2048 Hz, half ours), they report that the dominant effect in changing their results to the new, less discrepant posteriors was in improving their relative binning sampling algorithm. For a full explanation of the results for GW151226 found by other studies it would also be interesting to compare results using the different PSD estimates employed, but the PSD used in Chia et al. (2022) is not publicly available. (The PSD used in Nitz et al. (2021) can in principle be reproduced from configuration files in their data release.)

To conclude the discussion of this event, the recent work by Vajpeyi et al. (2022) has directly compared the posterior weight between equal and unequal mass configurations, contributing to the consensus picture hinted at above: consistency with a range of mass ratios, no strong preference for one end of the scale, and some remaining dependence of results on prior choice.

For the other two low-mass events, similar consistency is achieved between our different runs. In particular, for GW170608 our runs with

a sampling rate of 4096 Hz show excellent agreement, for both mass priors in Parallel Bilby and for LALInference, while for GW151012 small differences are visible between the standard prior Bilby run and the alternative prior and LALInference runs. However, differences between our different runs for this event are much smaller than differences with the results reported in Nitz et al. (2021), as can be observed in Fig. A4. Despite the small differences, the overall self-consistency of our results for GW151012 and GW170608 suggests that the results presented in Nitz et al. (2021) cannot be explained by different choices of the prior and sampling rate, and further examination of the convergence of those results and the influence of the choice of PSD estimation would be needed to understand the discrepancies.

## APPENDIX B: MULTIBANDING

As mentioned in Sec. 2.2, in order to accelerate the evaluation of the waveform model, IMRPHENOMXHM and IMRPHENOMXPHM implement the multibanding interpolation method described for non-precessing systems in García-Quirós et al. (2021), and extended to precession in Pratten et al. (2021). The method uses interpolation from an appropriately chosen unequally spaced coarse frequency grid to the equally spaced fine grid used for data analysis. The grid spacing of the coarse grid is controlled by two threshold parameters, which are related to the local interpolation error for the phase and amplitude. The parameter controlling the accuracy of the non-precessing modes is called MB, and the parameter controlling the accuracy of the Euler angles evaluation used to construct the precessing waveform is PMB.

We want to quantify the loss of accuracy and gain in computational speed for runs with a very aggressive multibanding threshold. To do that, we perform a comparison between the default version of IMRPHENOMXPHM (MB = $10^{-3}$ and PMB = 0, i.e. no multibanding for the Euler angles) using `nlive` = 512, 2048 and `nact` = 10, 30 respectively, with two aggressive multibanding runs (MB = PMB = 0.1) using `nlive` = 512, 1024 and `nact` = 10. In Fig. B1 we show this comparison for GW170104, which has the lowest total mass from the medium-mass events group. Results for the other events are contained in our data release (Mateu-Lucena et al. 2021). We find, not surprisingly, that for the low resolution runs (`nlive` = 512 and `nact` = 10) convergence is poor with and without the aggressive multibanding threshold. However, when we increase the number of live points to 1024 and 2048, the posterior distributions become smoother even using aggressive multibanding, and agree well with the distributions obtained with less aggressive multibanding. Comparing the computational resources of the runs in Table B1, we can observe that with an aggressive multibanding threshold the cost of each likelihood evaluation is reduced by approximately half. Although using a more aggressive multibanding threshold is not as accurate as runs without this acceleration, due to the approximations used, we see that in our case the change of the posterior distributions is very small. Thus, aggressive multibanding is particularly appropriate for quick exploratory runs, e.g. to tune prior bounds for masses or distance.

This paper has been typeset from a TEX/LATEX file prepared by the author.
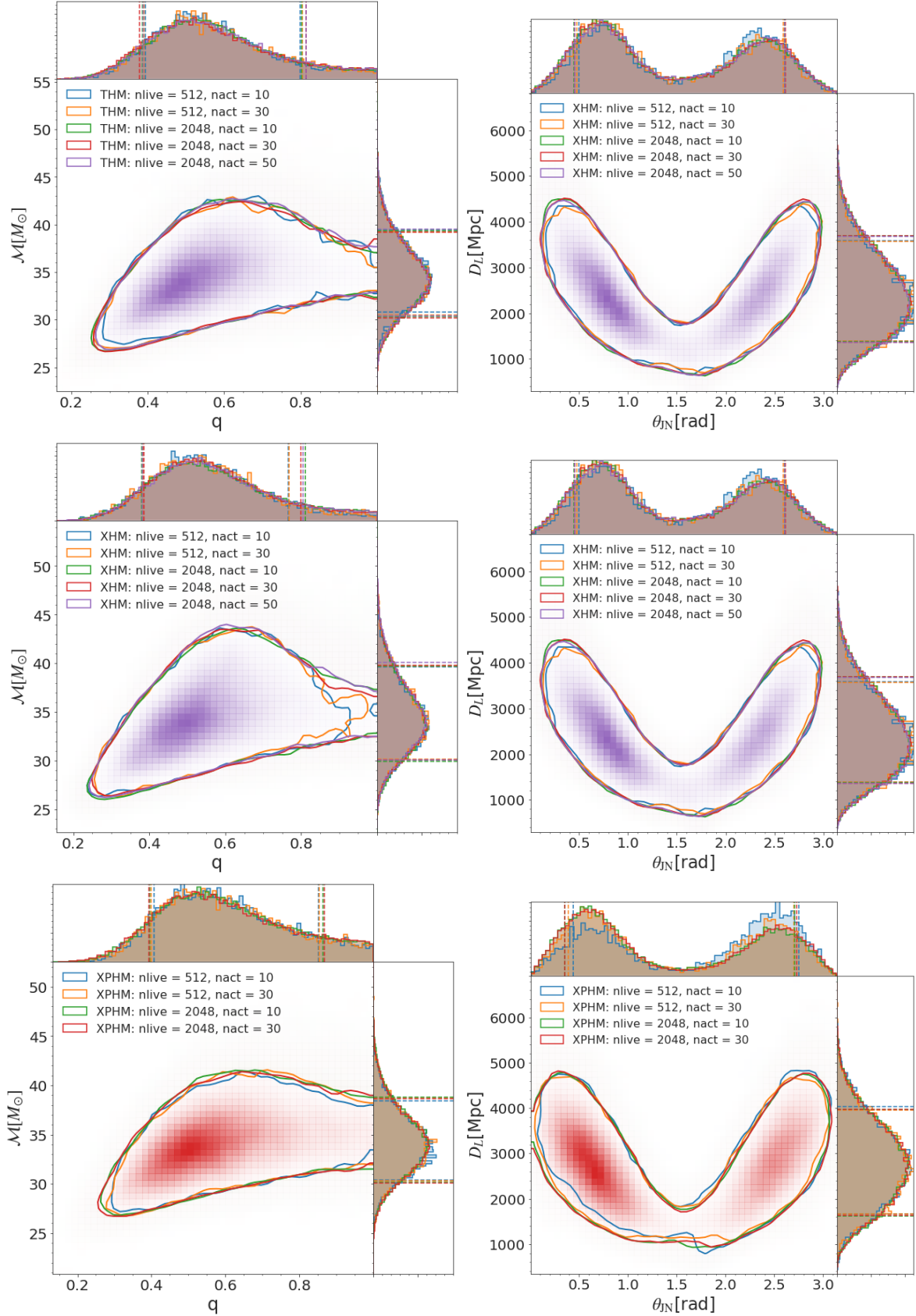
**Figure A1.** GW170729 convergence test comparing posterior distributions for the mass ratio, chirp mass, luminosity distance and inclination using IMRPHE-NOMTHM, IMRPHENOMXHM and IMRPHENOMXPHM (one per row) and with different sampler settings (`nlive` = 512, 2048 and `nact`=10,30,50) for each waveform, displaying 90% credible contours.

|  |  | N512 NA10 | N512 NA30 | N2048 NA10 | N2048 NA30 | N2048 NA50 |
|---|---|---|---|---|---|---|
| IMRPHENOMTHM | N512 NA10 | - | $JS_{t_1} = 0.006$ | $JS_{t_2} = 0.020$ | $JS_{t_2} = 0.019$ | $JS_{t_2} = 0.019$ |
|  | N512 NA30 |  | - | $JS_{t_2} = 0.023$ | $JS_{t_2} = 0.021$ | $JS_{t_1} = 0.024$ |
|  | N2048 NA10 |  |  | - | $JS_{ra} = 0.001$ | $JS_{ra} = 0.001$ |
|  | N2048 NA30 |  |  |  | - | $JS_{t_1} = 0.001$ |
|  | N2048 NA50 |  |  |  |  | - |
|  |  | N512 NA10 | N512 NA30 | N2048 NA10 | N2048 NA30 | N2048 NA50 |
| IMRPHENOMXHM | N512 NA10 | - | $JS_{ra} = 0.007$ | $JS_{t_2} = 0.019$ | $JS_{ra} = 0.024$ | $JS_{ra} = 0.025$ |
|  | N512 NA30 |  | - | $JS_{t_2} = 0.023$ | $JS_{t_2} = 0.023$ | $JS_{t_2} = 0.023$ |
|  | N2048 NA10 |  |  | - | $JS_{ra} = 0.005$ | $JS_{ra} = 0.004$ |
|  | N2048 NA30 |  |  |  | - | $JS_{ra} = 0.001$ |
|  | N2048 NA50 |  |  |  |  | - |
|  |  | N512 NA10 | N512 NA30 | N2048 NA10 | N2048 NA30 |  |
| IMRPHENOMXPHM | N512 NA10 | - | $JS_{ra} = 0.021$ | $JS_{ra} = 0.039$ | $JS_{ra} = 0.038$ |  |
|  | N512 NA30 |  | - | $JS_{ra} = 0.012$ | $JS_{ra} = 0.009$ |  |
|  | N2048 NA10 |  |  | - | $JS_{ra} = 0.001$ |  |
|  | N2048 NA30 |  |  |  | - |  |

**Table A1.** Maximum Jensen-Shannon (JS) divergence values between the posterior distributions for GW170729 estimated with IMRPHENOMTHM, IMRPHENOMXHM and IMRPHENOMXPHM using different sampler settings, where N is short for `nlive` and NA is short for `nact`.
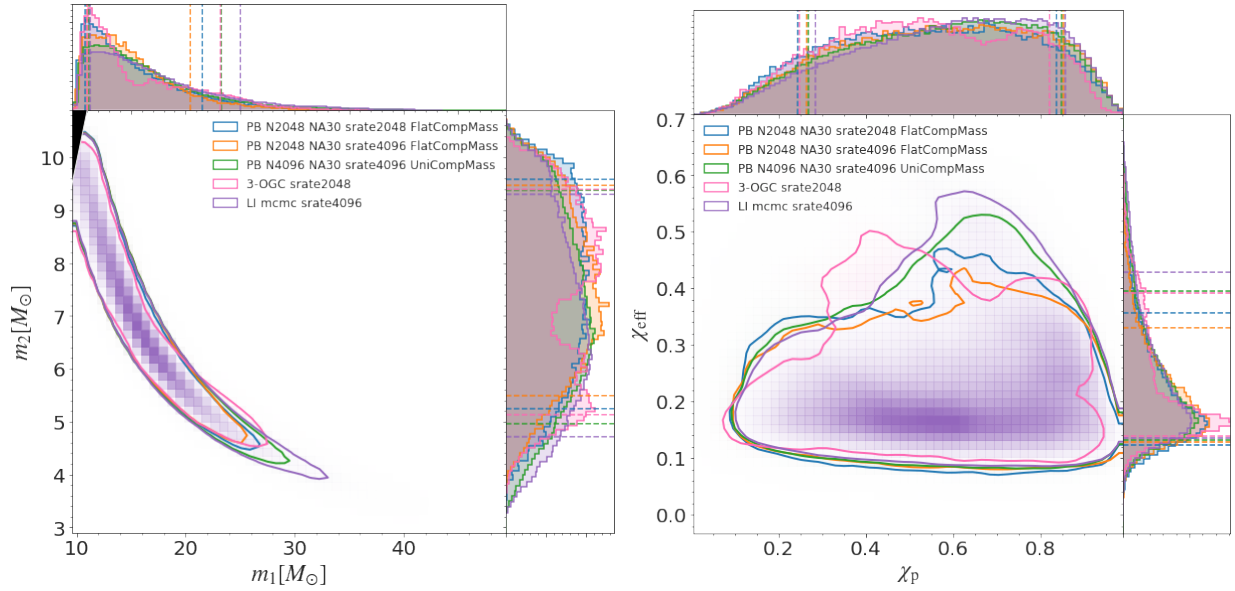


**Figure A2.** Comparison of the recovered posterior distributions for GW151226 for different values of the sampling rate (`srate`) and mass priors in Parallel Bilby (PB), a standard run in LALInference (`LI mcmc`), and the results reported for this event in Nitz et al. (2021) (denoted as 3-OGC in the legend). For brevity, `nlive` has been abbreviated as "N" and `nact` as "NA" in the legend. `FlatCompMass` refers to a run using a prior uniform in $q \leq 1$ and $\mathcal{M}_c$ and then reweighted to a flat prior in component masses. `UniCompMass` refers to directly using a uniform prior in component masses.

| nlive | nact | MB | PMB | CPU h | L.eval. | Cost/L.eval.[ms] |
|---|---|---|---|---|---|---|
| 2048 | 30 | $10^{-3}$ | 0 | 1765 | $2.24 \times 10^8$ | 28.30 |
| 512 | 10 | $10^{-3}$ | 0 | 154 | $1.67 \times 10^7$ | 33.11 |
| 1024 | 10 | $10^{-1}$ | $10^{-1}$ | 157 | $3.46 \times 10^7$ | 16.34 |
| 512 | 10 | $10^{-1}$ | $10^{-1}$ | 77 | $1.71 \times 10^7$ | 16.32 |

**Table B1.** Computational cost comparison between runs on GW170104 using IMRPHENOMXPHM with different multibanding thresholds and sampler settings, where MB and PMB correspond to the multibanding thresholds of the aligned-spin modes and the Euler angles evaluation, respectively. The number of likelihood evaluations and the mean cost of each evaluation in ms are also shown.
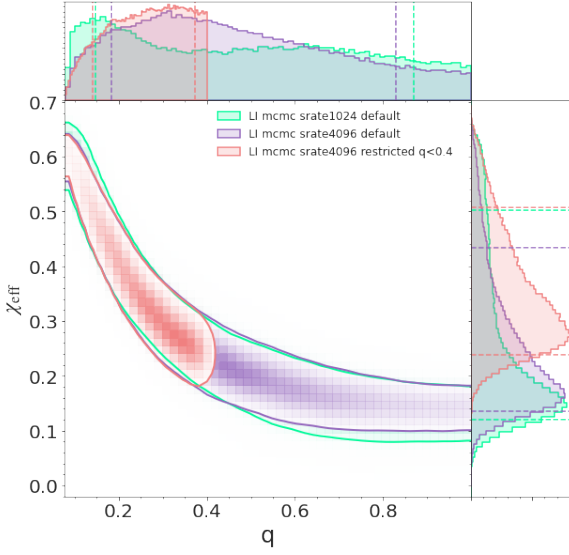
**Figure A3.** Comparison of the recovered posterior distributions for GW151226 in LALInference (LI mcmc) for two different values of the sampling rate (srate = 1024 and 4096 Hz) and two different mass-ratio prior choices (the default one and one restricted to $q \leq 0.4$). The mass-ratio histograms have been normalized to equal height for ease of visibility.
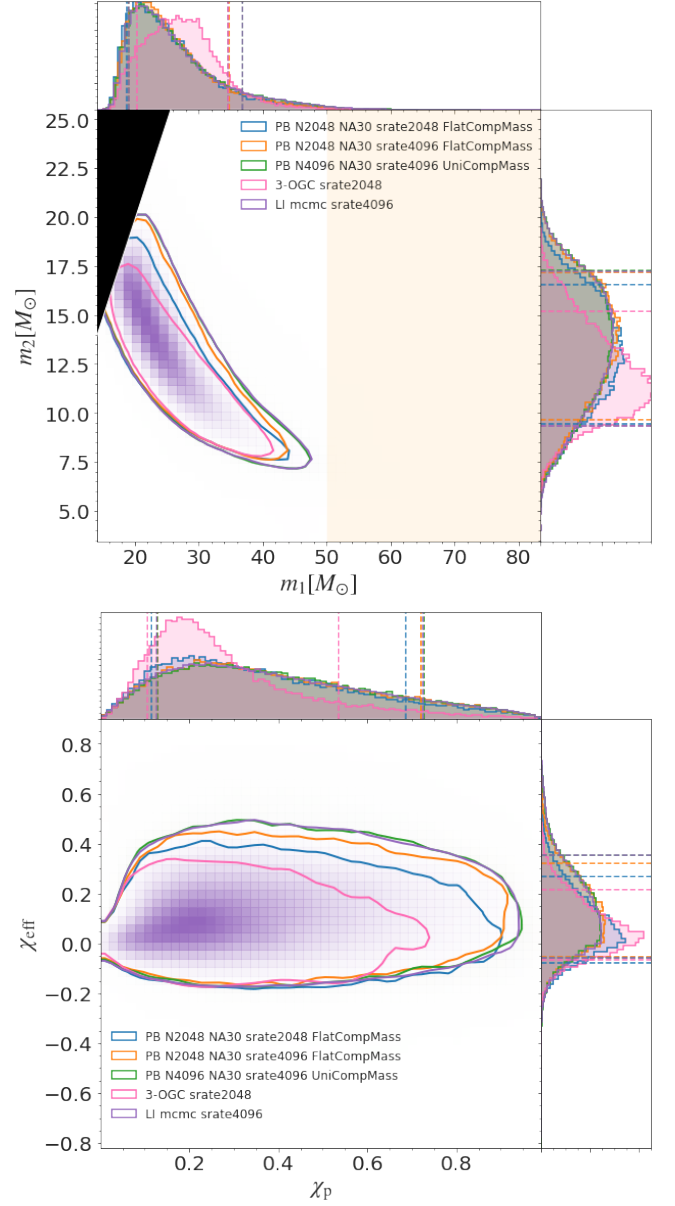


**Figure A4.** Comparison of the recovered posterior distributions for GW151012 for different values of the sampling rate (srate) and mass priors in Parallel Bilby (PB), a standard run in LALInference (LI mcmc), and the results reported for this event in Nitz et al. (2021) (denoted as 3-OGC in the legend). For brevity, nlive has been abbreviated as "N" and nact as "NA" in the legend. FlatCompMass refers to a run using a prior uniform in $q \leq 1$ and $\mathcal{M}_c$ and then reweighted to a flat prior in component masses. UniCompMass refers to directly using a uniform prior in component masses.
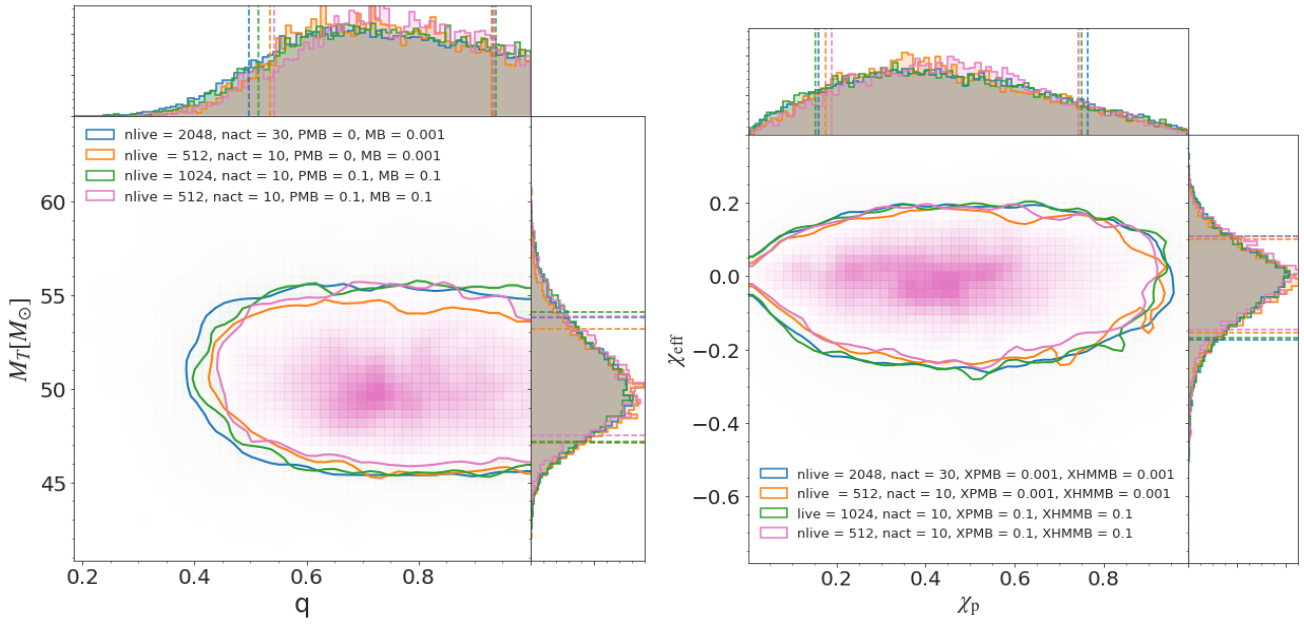
**Figure B1.** Comparison of the GW170104 posterior distributions for the total mass, mass ratio and effective spin obtained with IMRPʜᴇɴᴏᴍXPHM using different sampler settings (`nlive` and `nact`) and multibanding thresholds for the aligned-spin modes (XHMMB) and for the Euler angles (XPMB) where 0.001 is the multibanding default value.