Investigating the source of inferred parameters' discrepancies between LIGO detectors in GW231123 Interim Report 2

Krzysztof Król and Mentor: Sophie Bini

ABSTRACT

GW231123 is a short duration gravitational wave (GW) signal, consistent with a binary black hole merger with a total mass of 190-265 $M_\odot$. It is the most massive binary black hole confidently observed to date. Both components are highly spinning and likely have masses in the mass gap caused by pair-instability supernova processes. The event is challenging to analyze because of its very short duration and limited accuracy of waveform models for such an extreme system. Additionally, significant differences between inferred posteriors for total mass and spins arise when parameter estimation is performed using LIGO Hanford-only and LIGO Livingston-only data, raising concerns about the presence of spurious transient noise overlapping with the GW signal. In this project, we quantify these differences and investigate their causes through simulated signals similar to GW231123.

1. INTRODUCTION

GW231123 is a transient gravitational wave (GW) event observed on 23 November 2023 by the two Advanced LIGO detectors (Virgo was offline at the time of detection). The signal is consistent with the coalescence of a binary black hole system with total mass of between 190-265 $M_\odot$, which makes it the most massive merger observed to date with high confidence (The LIGO Scientific Collaboration et al. 2025).

GW231123 might have very important astrophysical implications. Due to pair-instability supernova process (Farmer et al. 2019), it is believed that black holes with masses between 60 $M_\odot$ and 130 $M_\odot$ cannot form as a result of standard stellar collapse. In the case of the previous most-massive merger, GW190521 (Abbott et al. 2020a,b), the primary black hole is believed to lie in the mass gap. For GW231123, the primary is within or above the mass gap produced by the pair-instability supernova process, while the secondary is within the mass gap with a probability of 83%. Further, GW231123 is the second event (after GW190521) observed that produced a remnant which would be classified as an intermediate mass black hole (a black hole with mass between $10^2 M_\odot$ and $10^5 M_\odot$).

Inferring the properties of the source of this exceptional event is particularly challenging due to its high mass. Its signal's frequency is noticeably lower than other events, peaking at close to 50 Hz. The event is also much shorter, with only a few observable cycles before merger and the entire event being shorter than 0.2 seconds.

Following parameter estimation (PE), differences between LIGO Livingston-only and LIGO Hanford-only results were noticed. In particular, the median total mass and the median spin precession differ by 2 sigma. In this project, we quantify these differences and investigate their possible sources.

2. METHODS

2.1. *Binary black hole coalescence parameters*

The main difficulty in parameter estimation from binary black hole GW signal comes from the multidimensionality of the data. There are 15 total parameters that describe a binary black hole waveform (for a quasicircular system), namely:

- masses of the two black holes m_1 and m_2 ,
- spins $\vec{\chi_1}$ and $\vec{\chi_2}$, expressed by the dimensionless spin vector $\vec{\chi_i} = \vec{S_i}c/(Gm_i^2)$ $(0 \leq |\vec{\chi_i}| < 1)$, where $\vec{S_i}$ is the angular moment of the black hole,
- luminosity distance D_L,
- inclination angle ι (angle between observer's line of sight and orbital plane of the system),
- coalescence time t_0 and phase φ,
- polarization angle ψ,
- source sky position θ and ϕ.

Due to the possibility of more precise measurement we also define chirp mass $\mathcal{M}$:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

for $m_1 > m_2$. Beyond that, in describing the system as a whole, it is often useful to introduce the spin parameters χ_eff and χ_p, which are the effective aligned spin and effective precessing spin. They are defined as (Ng et al. 2018; Thomas et al. 2021):

$$\chi_{\rm eff} = \frac{\vec{S_1}/m_1 + \vec{S_2}/m_2}{m_1 + m_2} \cdot \vec{L}$$

and

$$\chi_{\rm p} = \frac{\max(A_1|\vec{S_1}|,\ A_2|\vec{S_2}|)}{A_1 m_1^2},$$

where $A_1 = 2 + 3/2q$ and $A_2 = 2 + 3q/2$. $\chi_{\rm eff}$ is bounded by $[-1, 1]$ and $\chi_{\rm p}$ by $[0, 1]$ to avoid naked singularities. Furthermore, $\chi_{\rm eff}$ is a constant of motion up to at least the second post-Newtonian order (Racine 2008; Blanchet 2014) and typically can be measured more precisely than the individual spins.

2.2. *Bayesian statistics*

To infer the properties of GW signals, we use Bayesian statistics. Bayesian statistics is based on an interpretation of probability, where, unlike in the frequentist approach, probability expresses a degree of belief. It is achieved by combining prior knowledge with new data to compute new probabilities ("posterior") according to Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $P(A)$ is the prior probability, $P(B)$ is the probability of new observations, $P(B|A)$ the likelihood, and $P(A|B)$ the posterior probability. In the context of parameter estimation, Bayes' theorem can be rewritten as

$$p(\theta|d) = \frac{L(d|\theta)\pi(\theta)}{\int L(d|\theta)\pi(\theta)\,\mathrm{d}\theta}$$

where observations are denoted by d, unknown parameters are denoted by θ, L is the likelihood function and π is the prior probability density function (Christensen & Meyer 2022). The denominator serves as a normalization constant.

2.3. *Data preprocessing and sampling settings*

We perform our PE using whitened, bandpassed data. A glitch in LIGO Hanford occurred 1.7-1.1 s before the event in a frequency range between 15-30 Hz. It was modeled along with the waveform using BayesWave (Cornish & Littenberg 2015; Cornish et al. 2021; Chatziioannou et al. 2021). Another glitch occurred in LIGO Livingston 3.0-2.0 s in a frequency range between 10-20 Hz. This glitch, due to low frequency and the extended time between the event and the glitch, was not removed from the LIGO Livingston strain data.

We perform PE using `bilby` (Ashton et al. 2019), a Python library for GW-related Bayesian inference. It allows one to perform parameter estimation for a GW signal using nested sampling (Ashton et al. 2019). We are using the same power spectral density (PSD), likelihood, and sampler settings as for the GWTC-4 catalog.

2.4. *Comparing posterior distributions*

To understand the differences between posterior distributions resulting from different PE runs, we will both be visually examining the corner plots containing relevant parameters and trying to quantify them. To do the latter, we will use the Jensen-Shannon (JS) divergence between two posterior distributions Q_1 and Q_2, which is defined as

$$JS(Q_1||Q_2) = \frac{1}{2}D(Q_1||M) + \frac{1}{2}D(Q_2||M),$$

where $M = \frac{1}{2}(Q_1 + Q_2)$ is a mixture distribution of Q_1 and Q_2, and $D(Q_1||Q_2)$ is the Kullback-Leibler divergence defined as

$$D(Q_1||Q_2) = \sum_{x \in X} Q_1(x) \ln \frac{Q_1(x)}{Q_2(x)}.$$

We will use the convention in which the logarithm base in Kullback-Leibler divergence is e to stay consistent with the values calculated by the package `PESummary` that we used for part of our data visualization (Hoy & Raymond 2021). Then the JS divergence is a real number in the range $[0, \ln 2]$ ($\ln 2 \approx 0.69$) for discrete probability distributions (as is the case for numerically calculated posterior distributions), where $JS = 0$ if and only if the two distributions are identical.

Due to different PE seeds affecting different runs, noise and numerical estimations of posterior distributions we expect non-zero JS divergences even for two different PE runs performed on the same data. Typically, two distributions with $JS < 0.035$ would be considered to be in good agreement (Abbott et al. 2019) (we converted the given threshold by multiplying it by $\ln 2$ due to convention difference). Due to the extreme nature of the event analyzed in this project we expect that adopting a different threshold might be sensible; we will address that in Section 4.2.

3. SOURCE PROPERTIES OF GW231123

3.1. *Waveform approximant choice and ASD difference quantification*

The LIGO Scientific Collaboration et al. (2025) found that the NRSur7dq4 (Varma et al. 2019) waveform model is on average more accurate than every other waveform model tested (IMRPhenomXPHM, IMRPhenomXO4a, IMRPhenomTPHM, SEOBNRv5PHM) for events similar to GW231123. We first

attempted using NRSur7dq4 to perform PE and for further studies, but quickly discovered that its computational cost along with current LIGO Data Grid's load significantly slowed us down. We therefore decided to use IMRPhenomXPHM (Pratten et al. 2021), which offers close to 5× computational speedup.

In making this decision we inspected results from PE using both models and discovered a significant difference in the average modeled waveform's ASD between the two models. We plot this in Figure 1 and Figure 2 for NRSur7dq4 and IMRPhenomXPHM, respectively. Significant differences in LIGO Hanford and LIGO Livingston's sensitivities can immediately be noticed, especially in the 20-50 Hz frequency range. In particular, at 20 Hz LIGO Livingston's ASD value is almost two times lower than LIGO Hanford's ASD.

We acknowledge the fact that using the quicker model might influence the results we achieve and note that further investigation of this difference might be necessary in the future. It is also worth noting that the difference in strain between detectors is much more pronounced in the NRSur7dq4 model, and it further increases LIGO Livingston's expected higher SNR.

We later prepared further visualizations of the data. We plotted the whitened strain along with the maximum likelihood IMRPhenomXPHM waveforms resulting from full, Hanford-only and Livingston-only PE in Figure 3. There is only a slight difference between the LIGO Livingston-only estimated waveform and estimated waveform using both detectors, which can be expected considering that the two-detectors inference is driven by LIGO Livingston due to LIGO Livingston's higher SNR. We see, however, significant differences between the waveform resulting from LIGO Hanford-only PE and the other two runs. In particular, there is significant discrepancy between the calculated waveform and LIGO-Livingston strain data observed in the inspiral phase. We do acknowledge, however, that the poor ability to constrain the sky position based on data from one detector might be a significant driver of this difference and it is not unique for GW231123.

3.2. *Comparing PE results*

We present PE posteriors for selected parameters for the main compact binary coalescence parameters from LIGO Hanford-only and LIGO Livingston-only data in Figure 4. We also present pairwise Jensen-Shannon divergences for selected parameters from the three PE runs in Table 1. The estimated values from PE with 90% confidence intervals are shown in Table 2.

Those results by themselves are very interesting. The median masses differ noticeably (total mass values of $M^{\rm H} = 330^{+36}_{-55}\ M_\odot$ vs. $M^{\rm L} = 286^{+20}_{-26}\ M_\odot$), but the differences in luminosity distance (and inclination, which is coupled with it) and spins are the most striking. LIGO Hanford-only data favors medium values of $\chi^{\rm H}_{\rm eff} = 0.48^{+0.23}_{-0.50}$ and $\chi^{\rm H}_{\rm p} = 0.45^{+0.35}_{-0.26}$, while the LIGO Livingston-only data favor low $\chi^{\rm L}_{\rm eff} = 0.03^{+0.15}_{-0.22}$ and very high precession with $\chi^{\rm H}_{\rm p} = 0.72^{+0.21}_{-0.23}$. Typically, the luminosity distance and the inclination can be degenerate, but this degeneracy can be lifted in precessing systems (Usman et al. 2019) – such as GW231123 if $\chi_{\rm p}$ really has such an extreme value. Using Hanford data results in $\iota^{\rm H} = 1.53^{+1.35}_{-1.27}$ and $D^{\rm H}_{\rm L} = 4064^{+2905}_{-2758}$ Mpc, while Livingston data constrains those much better at $\iota^{\rm L} = 1.47^{+0.69}_{-0.55}$ $D^{\rm L}_{\rm L} = 568^{+390}_{-271}$ Mpc. We consider these differences significant and we believe their further investigation is warranted.

Table 1. Jensen–Shannon divergences of selected parameters between detector combinations.

Parameter	L1 vs. H1	H1 vs. H1L1	L1 vs. H1L1
redshift z	0.6093	0.5831	0.1916
luminosity distance $D_{\rm L}$	0.609	0.584	0.191
$\theta_{\rm JN}$	0.498	0.523	0.011
$\chi_{\rm eff}$	0.477	0.463	0.005
phase	0.467	0.430	0.015
primary mass	0.376	0.394	0.006
total mass M	0.373	0.369	0.009
inclination ι	0.352	0.312	0.008
$\chi_{\rm p}$	0.236	0.263	0.003
secondary mass	0.098	0.137	0.014
χ_1	0.020	0.014	0.008
χ_2	0.003	0.010	0.018

3.3. *Low frequency part of the spectrum*

We have also ran PE on data from both detectors in a different frequency range of 10-448 Hz instead of the standard 20-448 Hz. Typically including data from this low frequency region does not improve the fit because of low sensitivity of the detectors. However, it could be theorized that including this part of the spectrum is helpful in GW231123's case, as it is a significantly more massive event and therefore is detected at lower frequencies. After inspecting the results and calculating JS divergences between posteriors from two runs we concluded that changing the lower limit of frequency did not impact the posterior distributions in a way that would be distinguishable from random seed change in differ-

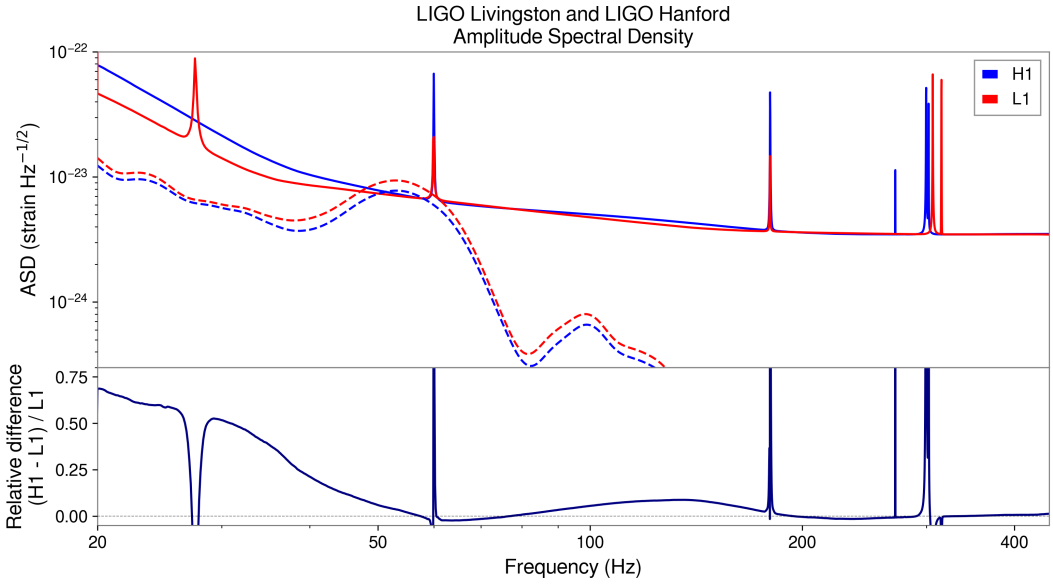

Figure 1. Top plot: ASD of LIGO Livingston (red), LIGO Hanford (blue) and of the average waveform resulting from NRSUR7DQ4 PE posterior samples for both of the interferometers (average waveform resulting from LIGO Hanford+LIGO Livingston PE posteriors). Bottom plot: relative difference in ASD of the two interferometers defined as $(\mathrm{ASD}_{\mathrm{H_1}} - \mathrm{ASD}_{\mathrm{L_1}})/\mathrm{ASD}_{\mathrm{L_1}}$.

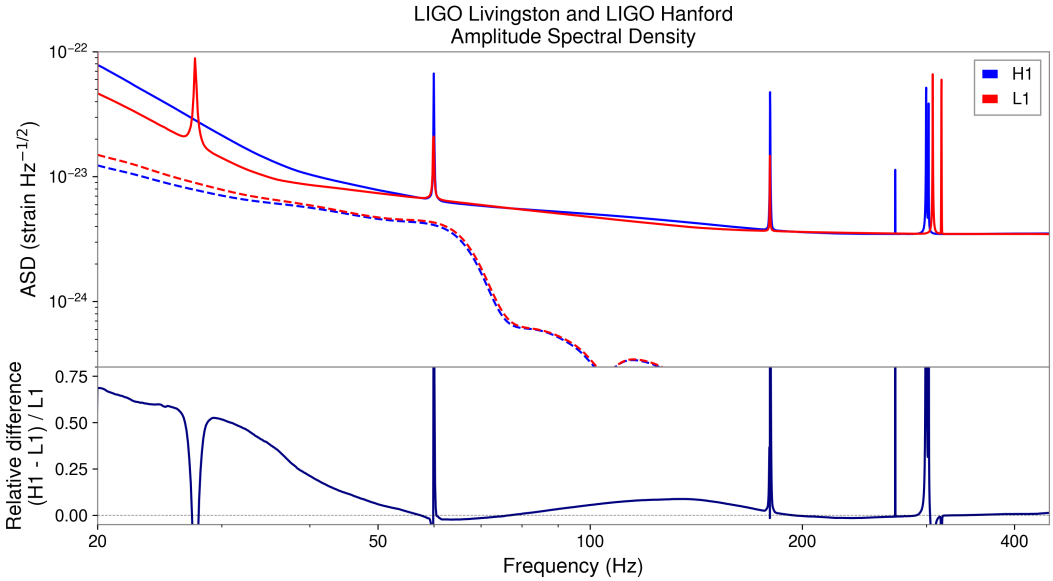

Figure 2. Top plot: ASD of LIGO Livingston (red), LIGO Hanford (blue) and of the average waveform resulting from IMRPHENOMXPHM PE posterior samples for both of the interferometers (average waveform resulting from LIGO Hanford+LIGO Livingston PE posteriors). Bottom plot: relative difference in ASD of the two interferometers defined as $(\mathrm{ASD}_{\mathrm{H_1}} - \mathrm{ASD}_{\mathrm{L_1}})/\mathrm{ASD}_{\mathrm{L_1}}$.

ent PE runs; therefore, we will not include the lower frequency range in our further analysis.

3.4. *GW231123's most likely parameters*

Throughout this report, as noted previously, we will use the results obtained from IMRPHENOMXPHM PE for constistency. However, we acknowledge that new work has emerged in the field that proposes a new method to incorporate the precision of the model in the inference of the GW parameters (Hoy et al. 2025). We decide not to incorporate this method because its applicability to an event as extreme as GW231123 would need to be studied first.

Table 2. Posterior values of selected parameters from 3 PE runs for GW231123 with 90% confidence intervals. Masses are given in source reference frame. Optimal SNR refers to optimal SNR for given detector/network.

Parameter	H1	L1	H1+L1
luminosity distance $D_{\rm L}$ (Mpc)	4064^{+2905}_{-2758}	568^{+390}_{-271}	851^{+349}_{-293}
redshift z	$0.658^{+0.363}_{-0.407}$	$0.118^{+0.072}_{-0.054}$	$0.171^{+0.062}_{-0.055}$
χ_1	$0.79^{+0.18}_{-0.40}$	$0.75^{+0.20}_{-0.26}$	$0.79^{+0.18}_{-0.26}$
χ_2	$0.60^{+0.36}_{-0.53}$	$0.54^{+0.40}_{-0.48}$	$0.68^{+0.28}_{-0.55}$
$\chi_{\rm eff}$	$0.48^{+0.23}_{-0.50}$	$0.03^{+0.15}_{-0.22}$	$0.05^{+0.16}_{-0.21}$
$\chi_{\rm p}$	$0.45^{+0.35}_{-0.26}$	$0.72^{+0.21}_{-0.23}$	$0.74^{+0.19}_{-0.24}$
inclination ι (rad)	$1.53^{+1.35}_{-1.27}$	$1.47^{+0.69}_{-0.55}$	$1.41^{+0.80}_{-0.56}$
log likelihood	$90.2^{+3.0}_{-4.6}$	$149.4^{+3.7}_{-5.0}$	$244.9^{+3.7}_{-5.4}$
optimal SNR	$13.4^{+1.7}_{-1.7}$	$17.5^{+1.7}_{-1.7}$	$22.2^{+1.6}_{-1.7}$
primary mass ($M_\odot$)	128^{+38}_{-31}	157^{+14}_{-13}	150^{+13}_{-12}
secondary mass ($M_\odot$)	69^{+31}_{-26}	98^{+22}_{-27}	94^{+19}_{-20}
total mass ($M_\odot$)	330^{+36}_{-55}	286^{+20}_{-26}	285^{+21}_{-21}
mass ratio	$0.55^{+0.35}_{-0.28}$	$0.62^{+0.15}_{-0.17}$	$0.63^{+0.12}_{-0.13}$
geocentric time since 1384782888 (s)	$0.6018^{+0.0240}_{-0.0217}$	$0.6020^{+0.0258}_{-0.0166}$	$0.6268^{+0.0026}_{-0.0035}$
phase (rad)	$2.89^{+2.57}_{-2.10}$	$5.10^{+0.43}_{-0.50}$	$5.03^{+0.60}_{-0.58}$
$\theta_{\rm JN}$ (rad)	$1.56^{+1.33}_{-1.30}$	$1.56^{+0.46}_{-0.40}$	$1.54^{+0.41}_{-0.37}$

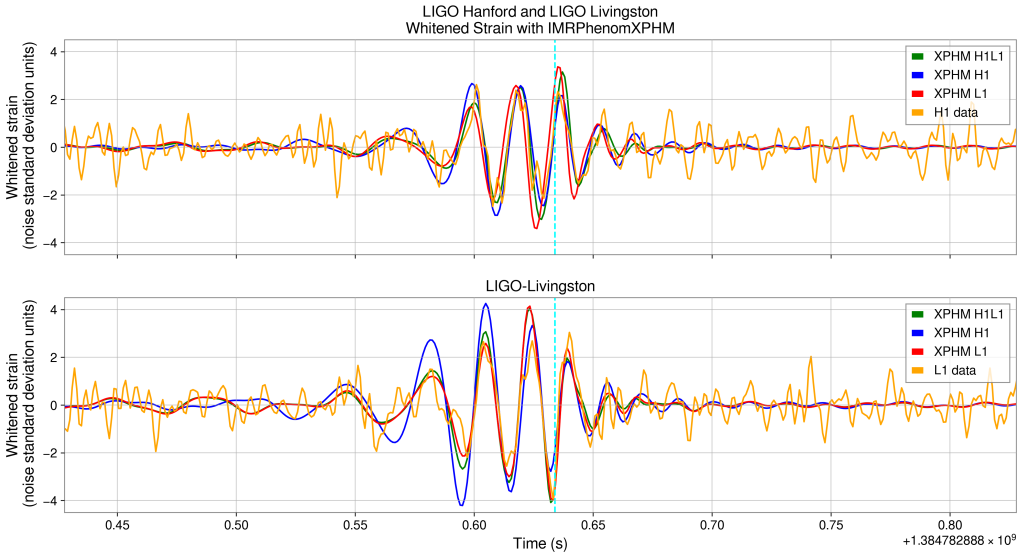

Figure 3. Whitened strain bandpassed to the frequency range [20Hz, 256Hz] (orange) plotted with best fitting IMRPhenomXPHM waveform models from LIGO Hanford+LIGO Livingston (green), LIGO Hanford-only (blue) and LIGO Livingston-only (red) PE for both LIGO interferometers. The y-axis is in noise standard deviation units.

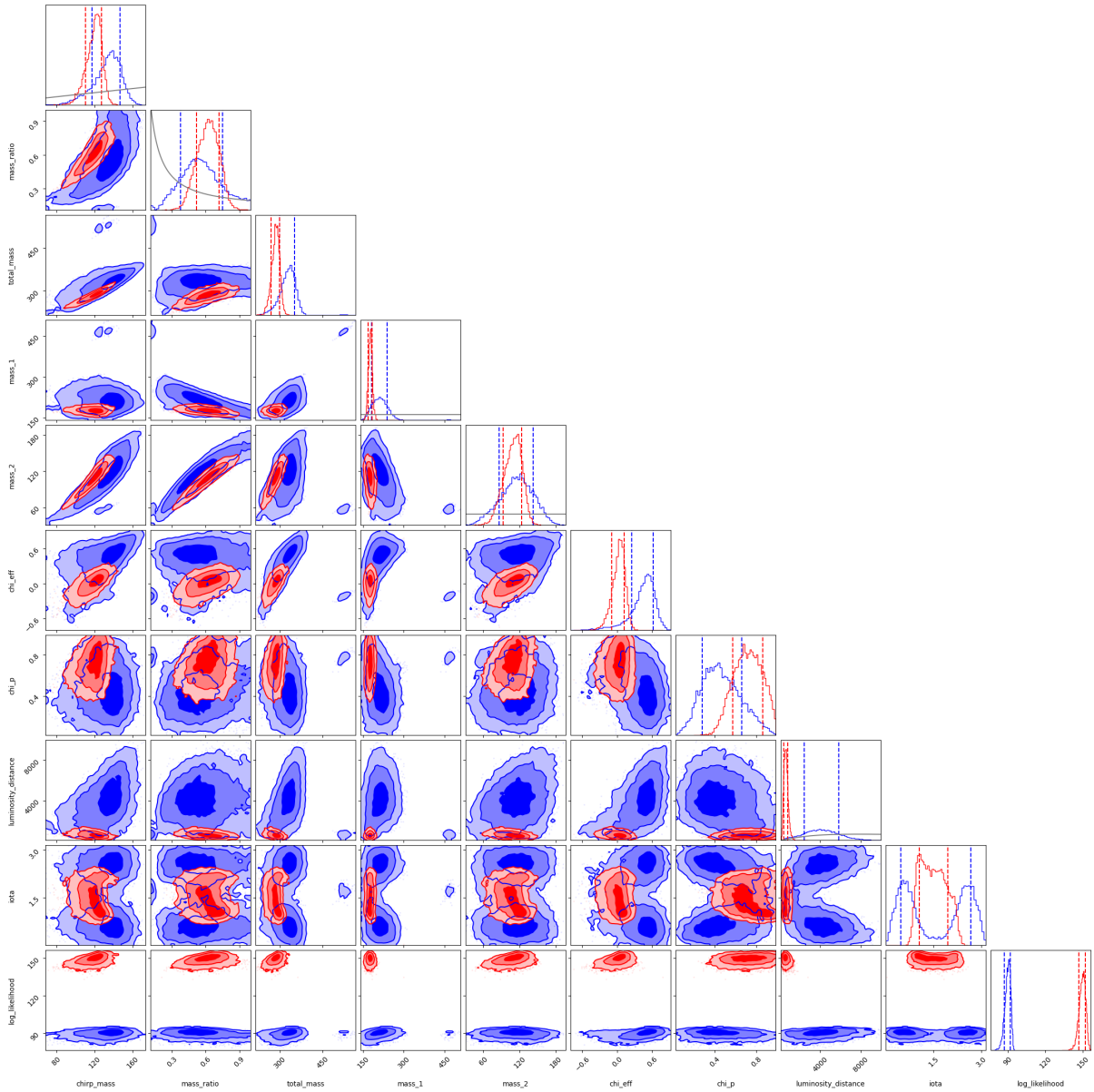

Figure 4. Comparison of LIGO Hanford-only (blue) and LIGO Livingston-only (red) parameter estimation results. Gray line on histograms indicates prior.

4. INJECTION STUDY

4.1. *Injection setup*

A numerical relativity (NR) waveform has only 9 intrinsic parameters associated with it: ratio of Christodoulou masses q, dimensionless spins $\vec{\chi_1}, \vec{\chi_2}$ (which also dictate the values of $\chi_{\rm eff}, \chi_{\rm p}$), eccentricity e (which will be very close to 0 for all waveforms we might examine), and mean anomaly at reference time ℓ (Scheel et al. 2025; Boyle et al. 2019). The system's masses can be scaled, and all extrinsic parameters can be varied.

We have prepared an injection study setup based on an NR waveform for a system similar to GW231123. The first waveform we examine is calculated from SXS:050 from the Simulating eXtreme Spacetimes catalog (Scheel et al. 2025). It is a binary black hole coalescence of two black holes with equal detector frame masses of 150 $M_\odot$ each, $\chi_{\rm eff} = 0.159$ and $\chi_{\rm p} = 0.940$ at luminosity distance $D_{\rm L} = 4043$ Mpc, with an orientation such that $\iota \approx 0$, $\varphi \approx 2$ and $\psi \approx 2.4$. The optimal SNR of the injected signal is 12.35 for LIGO Hanford and 17.16 for LIGO Livingston, similar to GW231123. Examining the NR result will ensure we use realistic, high-fidelity waveforms without introducing errors associated with the use of waveform approximants; it does, however, severely limit the parameter space that can be explored as there exists a limited number of NR waveforms.

4.2. *Zero-noise injections*

We ran zero-noise SXS050 PE runs using LIGO Hanford-only, LIGO Livingston-only, and LIGO Hanford with LIGO Livingston's PSD-only data. The posteriors for selected parameters are shown in Figure 5. As can be seen, the masses are inferred correctly, but the spins aren't well constrained. Although there is support for high spin magnitude, there is also a long tail in χ_1 and especially in χ_2. $\chi_{\rm eff}$ is overestimated and $\chi_{\rm p}$ is not measured, even though the system is highly precessing. This behavior is not unexpected for very massive systems (Xu & Hamilton 2023). Beyond that, the injected value of ι isn't recovered. We don't see significant differences between the two LIGO Hanford-only and the LIGO Livingston-only runs, only that posteriors of the LIGO Hanford-only run with higher SNR are narrower.

We note that the JS divergence of Hanford-only and Livingston-only posteriors is higher than the standard threshold of 0.035 for $M, \mathcal{M}, \chi_{\rm eff}$. We do not find that to be concerning and associate the higher divergences with short duration of the event amplifying the uncertainties associated with posterior estimation. Based on this observation, we propose to interpret two distributions with $JS < 0.06$ (in natural logarithm convention) as distributions in good agreement for GW231123-related injections.

The choice of values of inclination, phase and polarization is expected to be of crucial importance on measurability of $\chi_{\rm p}$ (Miller et al. 2025). Unfortunately, the results of Miller et al. (2025) cannot be directly applied to the system we are injecting and studying the measurability of spin precession is beyond the scope of this project. Beyond that, the values of those three angles are not well constrained for GW231123 (The LIGO Scientific Collaboration et al. 2025). We therefore arbitrarily chose the inclination angle and later compared our results with zero noise injections with different values of the inclination angle ($\iota \approx 0$ and $\iota \approx 1.4$) and different phase values ($\varphi \approx 0.5$, $\varphi \approx 2$ and $\varphi \approx 5$), for a total of 6 combinations. We decided not to vary the polarization, as Miller et al. (2025) found it less important than the other two parameters mentioned. We found differences in posteriors of some parameters between different values of those angles, but we did not notice a significant difference in $\chi_{\rm p}$ posteriors. We therefore decided to only keep investigating our initial choice, so $\iota \approx 0, \varphi \approx 2, \psi \approx 2.4$.

Interestingly, we found noticeable differences between posteriors for $\chi_{\rm eff}$, M and q between different orientations of the orbit, in particular emerging systematic underestimation/overestimation of these parameters for certain orientations. We believe such behavior should not occur for zero noise injections, especially considering the fact that intrinsic parameters' impact on the waveform is not model-dependent, but gets calculated based on the orbit orientation. We believe that it is probably related to the very short inspiral that we observe that makes inference of certain parameters much more challenging. For now we decided to focus on other aspects of our analyses though, so we are not currently exploring that any further.

4.3. *Injections with Gaussian noise*

We injected SXS050 signal into different Gaussian noise realizations using the same PSD as for GW231123 and performed PE of the injected signal as described in Section 2, separately for LIGO Hanford and LIGO Livingston. We present the obtained posteriors for $\chi_{\rm eff}, \chi_{\rm p}, M$ in the Appendix in Figures A1, A2, A3.

Analyzing the plots and JS divergences allows us to notice that for some seeds behavior similar to GW231123 emerges; a limited number of seeds has JS divergence equal to or higher than GW231123 ($JS_{\rm GW}$) for $\chi_{\rm eff}$, $\chi_{\rm p}$ or M. A noticeable fraction of seeds exhibits comparable difference between posteriors to the event and most of the injected seeds have JS divergence

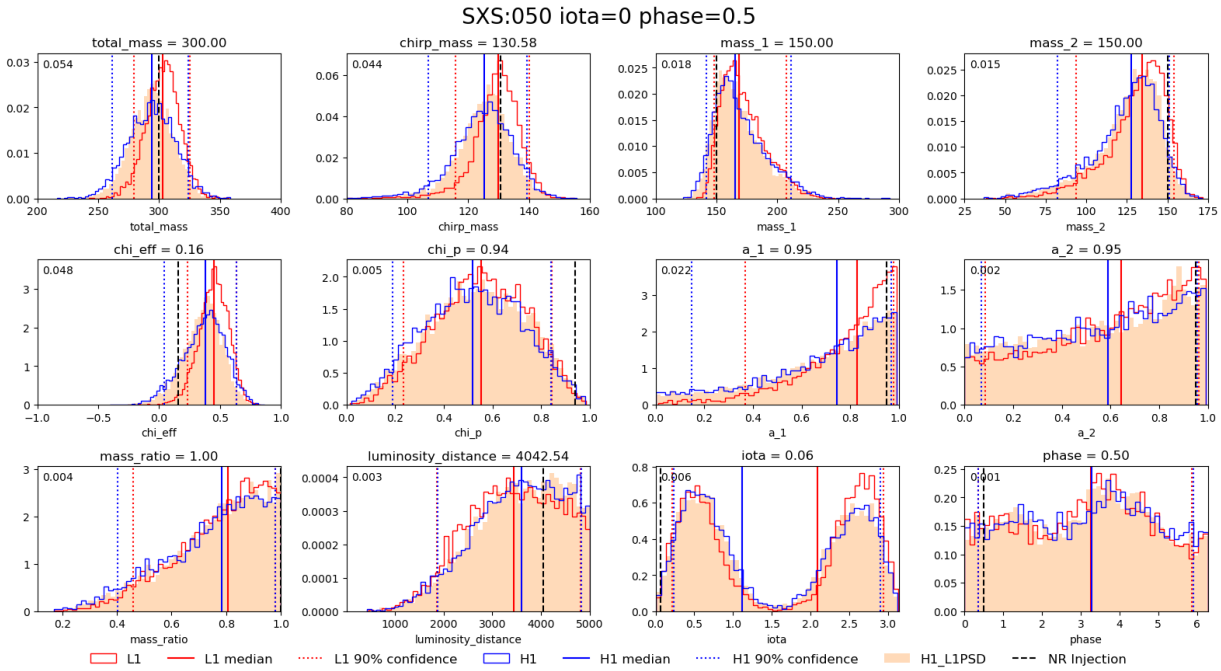

Figure 5. Comparison of LIGO Hanford-only (blue), LIGO Livingston-only (red) and LIGO Hanford with LIGO Livingston's PSD (solid peachpuff) IMRPhenomXPHM parameter estimation results for a zero-noise SXS:050 injection. Injected values are marked with a dotted black line. Jensen-Shannon divergence between LIGO Hanford-only and LIGO Livingston-only posterior is shown in the upper left corner for each injection.

over our established threshold of 0.06 for $\chi_{\rm eff}$ and M, as shown in Table 3. For a more detailed description, refer to Figures A1, A2, A3 and their captions.

Table 3. Number of seeds whose LIGO Hanford-only and LIGO Livingston-only posteriors have certain values of JS divergences. $\rm JS_{GW}$ refers to JS divergence of the two posteriors from GW231123's IMRPhenomXPHM PE: 0.477 for $\chi_{\rm eff}$, 0.236 for $\chi_{\rm p}$, 0.373 for M, as given in 1. Specific seeds are given in captions of Figures A1, A2, A3.

Parameter	$\rm JS \geq JS_{GW}$	$\rm JS \approx JS_{GW}$	$\rm JS \geq 0.06$
$\chi_{\rm eff}$	1/20	3/20	14/20
$\chi_{\rm p}$	1/20	1/20	5/20
M	2/20	3/20	11/20

Importantly, we notice that changing the PSD for LIGO Hanford doesn't meaningfully impact the posteriors beyond making them narrower. That observation also holds for posteriors of other parameters. We therefore conclude that the difference in low-frequency sensitivity isn't responsible for the differences we observe for GW231123.

Our observations are consistent with previous work in the field for systems similar to GW190521, the previous most massive system (Xu & Hamilton 2023). We might also investigate in the future if a certain Gaussian noise realization could cause us to infer a non-zero spin precession for a system with no spin precession. Xu & Hamilton (2023) found that not to be the case, but we suspect verifying their observations for a more massive system might be warranted, as the smaller number of observable cycles in LIGO frequency band could theoretically make Gaussian noise effects more pronounced than for GW190521-like systems.

Furthermore, we notice that zero-noise injections as well as a large part of injections with Gaussian noise overestimate the value of $\chi_{\rm eff}$. We don't know what the reason behind that is and we suspect it could be due to waveform systematics, which will be discussed in more detail in Section 5.

5. CONCLUSIONS AND FUTURE WORK

We have rejected the original idea that differences in PSDs of the two LIGO detectors are what caused the discrepancies between posteriors inferred from GW231123 PE runs based on the results of our simulations. We have three hypotheses that could explain these discrepancies that we will continue investigating.

1. Available waveform approximants fail to correctly model waveforms in this region of the parameter space.
   From analysis done by The LIGO Scientific Col-

laboration et al. (2025) we know that the current waveform approximants do not model events with such high spins with the required accuracy. We are currently investigating these discrepancies in NRSur7dq4, IMRPhenomXPHM and comparing the generated waveforms to NR waveforms. Currently we are working on that, but we have ran into technical issue with calculating the waveforms.

2. Certain Gaussian noise realizations could bias the parameter estimation.
   We found this to be true based on our injections. However, we notice that the behavior seen in GW231123 is exhibited by a small fraction of the injections that we performed. We do not yet know whether Gaussian noise fully explains the differences seen in GW231123. We are exploring other injections.

3. Non-Gaussian noise (glitch) coincident with the signal.
   Analysis performed by The LIGO Scientific Collaboration et al. (2025) found no glitches coincident with the signal, only glitches occurring seconds before the event. We don't have a reliable way to conclude the existence of or deal with a low amplitude glitch, if one was coincident with GW231123 signal in either detector.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, Physical Review X, 9, 031040, doi: 10.1103/PhysRevX.9.031040

Abbott, R., et al. 2020a, PhRvL, 125, 101102, doi: 10.1103/PhysRevLett.125.101102

—. 2020b, ApJL, 900, L13, doi: 10.3847/2041-8213/aba493

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27, doi: 10.3847/1538-4365/ab06fc

Blanchet, L. 2014, Living Reviews in Relativity, 17, 2, doi: 10.12942/lrr-2014-2

Boyle, M., Hemberger, D., Iozzo, D. A. B., et al. 2019, Classical and Quantum Gravity, 36, 195006, doi: 10.1088/1361-6382/ab34e2

Chatziioannou, K., Cornish, N., Wijngaarden, M., & Littenberg, T. B. 2021, Phys. Rev. D, 103, 044013, doi: 10.1103/PhysRevD.103.044013

Christensen, N., & Meyer, R. 2022, Reviews of Modern Physics, 94, 025001, doi: 10.1103/RevModPhys.94.025001

Cornish, N. J., & Littenberg, T. B. 2015, Class. Quant. Grav., 32, 135012, doi: 10.1088/0264-9381/32/13/135012

Cornish, N. J., Littenberg, T. B., Bécsy, B., et al. 2021, Phys. Rev. D, 103, 044006, doi: 10.1103/PhysRevD.103.044006

Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, ApJ, 887, 53, doi: 10.3847/1538-4357/ab518b

Hoy, C., Akçay, S., Mac Uilliam, J., & Thompson, J. E. 2025, Nature Astronomy, doi: 10.1038/s41550-025-02579-7

Hoy, C., & Raymond, V. 2021, SoftwareX, 15, 100765, doi: 10.1016/j.softx.2021.100765

Miller, S. J., Isi, M., Chatziioannou, K., Varma, V., & Hourihane, S. 2025, arXiv e-prints, arXiv:2505.14573, doi: 10.48550/arXiv.2505.14573

Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. 2018, PhRvD, 98, 083007, doi: 10.1103/PhysRevD.98.083007

Pratten, G., García-Quirós, C., Colleoni, M., et al. 2021, PhRvD, 103, 104056, doi: 10.1103/PhysRevD.103.104056

Racine, É. 2008, PhRvD, 78, 044021, doi: 10.1103/PhysRevD.78.044021

Scheel, M. A., Boyle, M., Mitman, K., et al. 2025, arXiv e-prints, arXiv:2505.13378, doi: 10.48550/arXiv.2505.13378

The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. 2025, arXiv e-prints, arXiv:2507.08219, doi: 10.48550/arXiv.2507.08219

Thomas, L. M., Schmidt, P., & Pratten, G. 2021, Phys. Rev. D, 103, 083022, doi: 10.1103/PhysRevD.103.083022

Usman, S. A., Mills, J. C., & Fairhurst, S. 2019, ApJ, 877, 82, doi: 10.3847/1538-4357/ab0b3e

Varma, V., Field, S. E., Scheel, M. A., et al. 2019, Phys. Rev. Research., 1, 033015, doi: 10.1103/PhysRevResearch.1.033015

Xu, Y., & Hamilton, E. 2023, PhRvD, 107, 103049, doi: 10.1103/PhysRevD.107.103049

APPENDIX

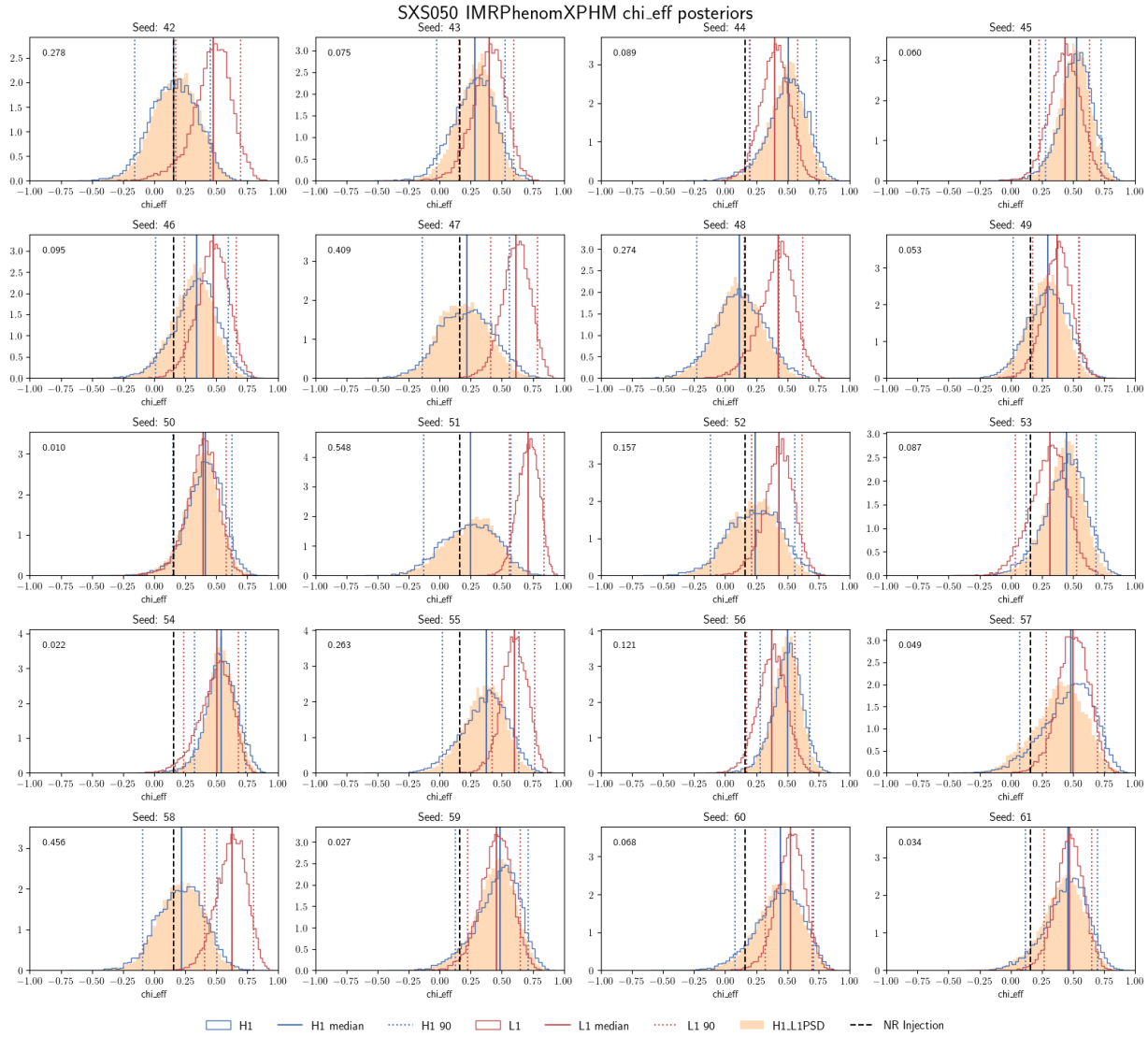

Figure A1. Comparison of LIGO Hanford-only (blue), LIGO Livingston-only (red) and LIGO Hanford with LIGO Livingston's PSD (solid peachpuff) IMRPhenomXPHM parameter estimation results for 20 SXS050 injections with Gaussian noise with different seeds. Injected value is marked with a dotted black line. Jensen-Shannon divergence between LIGO Hanford-only and LIGO Livingston-only posterior is shown in the upper left corner for each injection.
GW231123's $\chi_{\rm eff}$ JS divergence is 0.477. Seed 51's JS divergence is higher than that, besides that seeds 47 and 58 have JS divergences above 0.4. For seeds 51 and 58, in addition to high JS divergences, the two 90% confidence intervals don't overlap, similarly to GW231123.

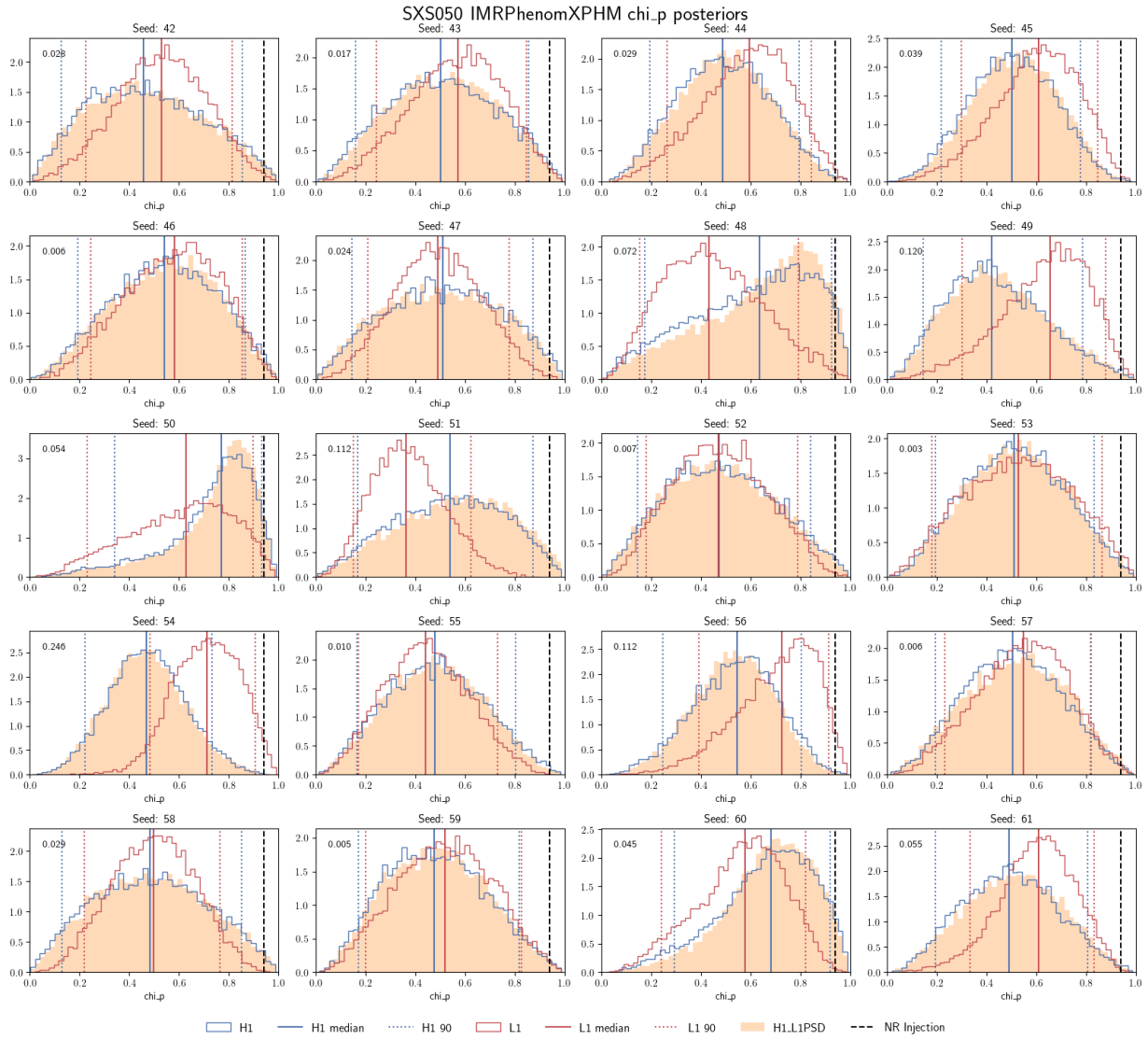

Figure A2. Comparison of LIGO Hanford-only (blue), LIGO Livingston-only (red) and LIGO Hanford with LIGO Livingston's PSD (solid peachpuff) IMRPhenomXPHM parameter estimation results for 20 SXS050 injections with Gaussian noise with different seeds. Injected value is marked with a dotted black line. Jensen-Shannon divergence between LIGO Hanford-only and LIGO Livingston-only posterior is shown in the upper left corner for each injection.
GW231123's $\chi_{\rm p}$ JS divergence is 0.236. Seed 54's JS divergence is higher than that and it is the only seed with the difference similar to the analyzed event. Seeds 48, 49, 51, 56 also exhibit noticeable differences between posteriors from the two analyses, but their JS divergences are lower (mainly due to how broad the posteriors are).

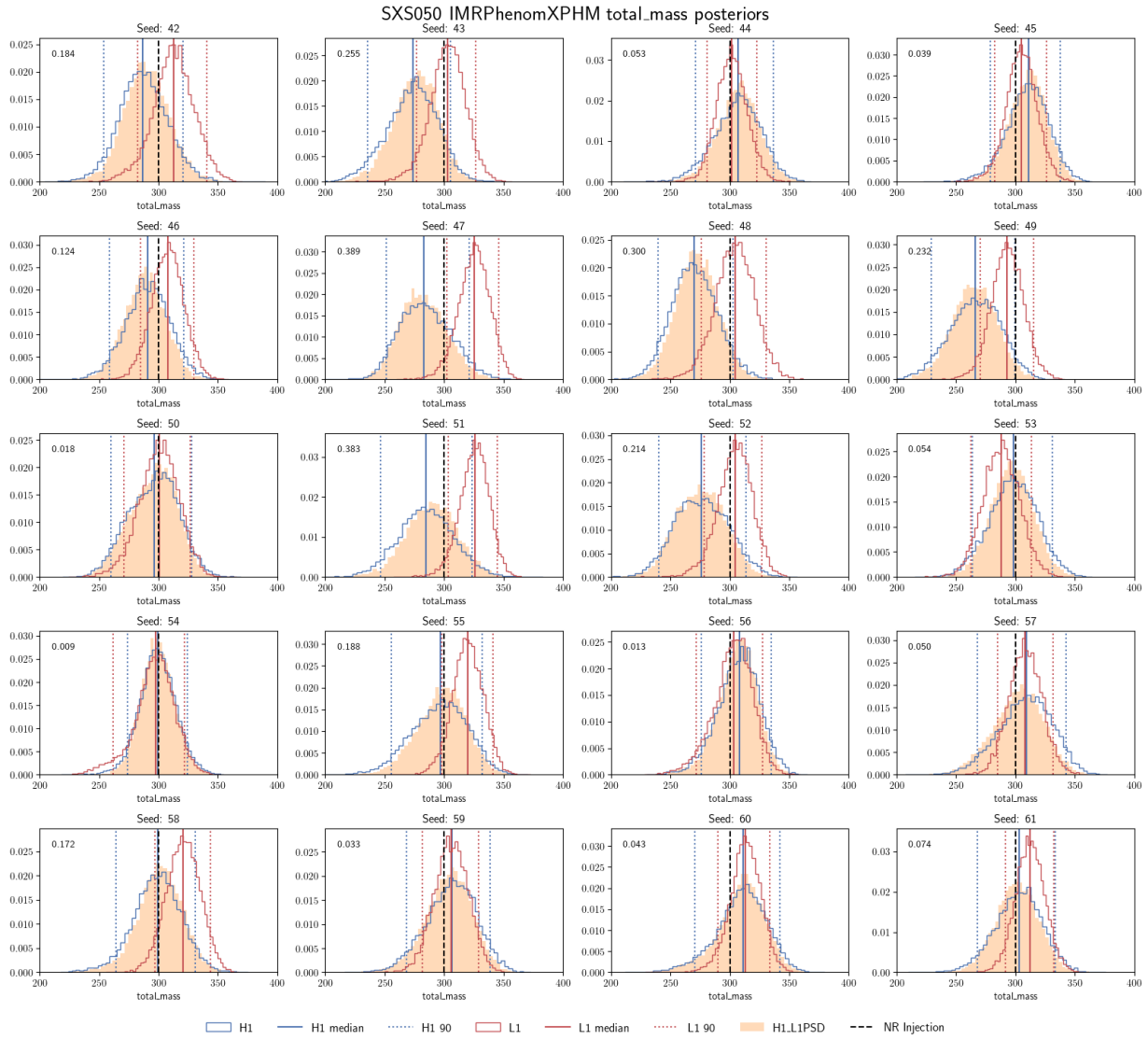

Figure A3. Comparison of LIGO Hanford-only (blue), LIGO Livingston-only (red) and LIGO Hanford with LIGO Livingston's PSD (solid peachpuff) IMRPhenomXPHM parameter estimation results for 20 SXS050 injections with Gaussian noise with different seeds. Injected value is marked with a dotted black line. Jensen-Shannon divergence between LIGO Hanford-only and LIGO Livingston-only posterior is shown in the upper left corner for each injection.
GW231123 M JS divergence is 0.373. Seeds 47 and 51 have higher JS divergences and their 90% confidence intervals barely overlap. Seeds 48 also ehxibits similar behavior, but with slightly smaller JS divergence.