Impact of Gaussian noise on the inference of GW231123-like signals

Krzysztof Król

Statistics Journal Club
7.01.2026

Parameter estimation (PE)

To infer the properties of gravitational-wave (GW) signal sources we employ Bayesian statistics based on the Bayes theorem.

$$P(\text{prior}|\text{data}) = \frac{P(\text{data}|\text{prior})P(\text{prior})}{P(\text{data})}$$

We perform PE using *bilby*, a library for Bayesian inference that samples the parameter space using nested sampling. We use the same settings and similar priors to other LVK analyses.

Nested sampling approaches this problem by starting from the full prior volume and progressively concentrating on regions of higher likelihood. At each step, points with low evidence are replaced by new samples drawn from the prior subject to a minimum likelihood requirement, gradually “nesting” into the most probable regions of parameter space. This makes nested sampling particularly effective for exploring complex, high-dimensional posteriors such as those encountered in gravitational-wave analyses.

Parameter estimation (PE) and the parameter space we explore

A quasicircular binary black hole merger can be parameterized using 15 parameters:

- masses of the two objects,
- spins of the two objects (they are vectors - each spin consists of 3 parameters),
- luminosity distance,
- inclination, polarization, phase,
- coalescence time,
- source sky position.

According to GR, the spin of a black hole cannot exceed Gm^2/c to avoid naked singularities. We therefore introduce the dimensionless spin parameter $\chi = Sc/(Gm^2)$.

Bilby uses generated waveforms for each sample to compute the likelihood.

The PE waveforms cannot be computed using numerical relativity (NR). Generating a single NR waveforms takes an order of ~100,000 core-hours, making their use in PE inaccessible. Instead, waveform approximants are used.

In our work we focused on waveform approximants from 2 families - NR surrogate and phenomenological models. I will present mostly results regarding the NR surrogate model.

Waveform approximants

NRSur7dq4 (NR surrogate model family)

NR surrogate models provide fast and accurate waveform predictions by interpolating between a discrete set of simulations. The waveforms are decomposed into physically motivated components, and modeling techniques are used to build smooth interpolants across parameter space.

NRSur7dq4 is calibrated on 1528 precessing binary black hole simulations with mass ratios >0.25 and spin magnitudes up to 0.8, enabling efficient evaluation of high-fidelity waveforms.

This is the approximant I'll focus on in this presentation.

IMRPhenomXPHM (phenomenological model family)

Phenomenological waveform models describe the inspiral, merger, and ringdown by fitting analytic expressions directly to hybrid waveforms constructed from post-Newtonian theory, effective-one-body models, and numerical relativity. The **IMRPhenomXPHM** model includes spin precession and higher-order multipoles, calibrated to 461 NR simulations. This provides a computationally efficient yet accurate waveform approximant.

The waveform construction and the parameter space we explore

The waveform approximant is used to generate the complex strain, dependent only on 7 of the intrinsic parameters:

- spins
- mass ratio

The strain then is scaled by the Christodoulou mass of the system and detector response is calculated using the extrinsic parameters. Finally, the waveform gets bandpassed and whitened.

Other than the 15 parameters mentioned earlier:

- masses of the two objects,
- spins of the two objects,
- luminosity distance,
- inclination, polarization, phase,
- coalescence time,
- source sky position.

We can also define other, derived parameters, which could be easier to measure or have a more intuitive impact on the waveform.

Important spin parameters we infer

Effective aligned spin parameter

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

bounded by [-1, 1] to avoid naked singularities.

It is a constant of motion up to at least the second post-Newtonian order. It should be easier to reliably measure than individual spins.

Effective precessing spin parameter

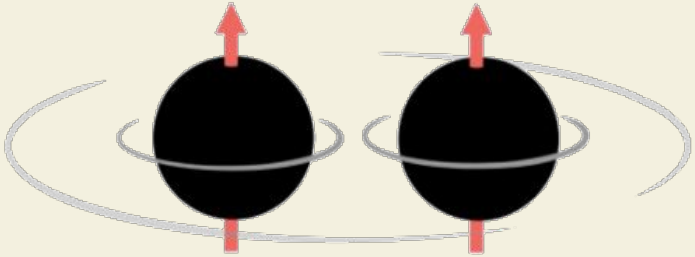
$$\chi_{\rm p} = \frac{\max(A_1|\vec{S_1}|,\ A_2|\vec{S_2}|)}{A_1 m_1^2},$$

$$\text{where } A_1 = 2 + 3/2q \text{ and } A_2 = 2 + 3q/2$$

bounded by [0, 1] to avoid naked singularities.

We take m1 > m2, q in [0, 1]. Constraining it in a compact binary coalescence is challenging.

Visualizing the spins



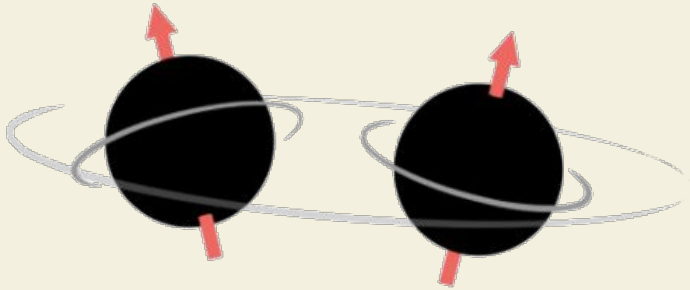
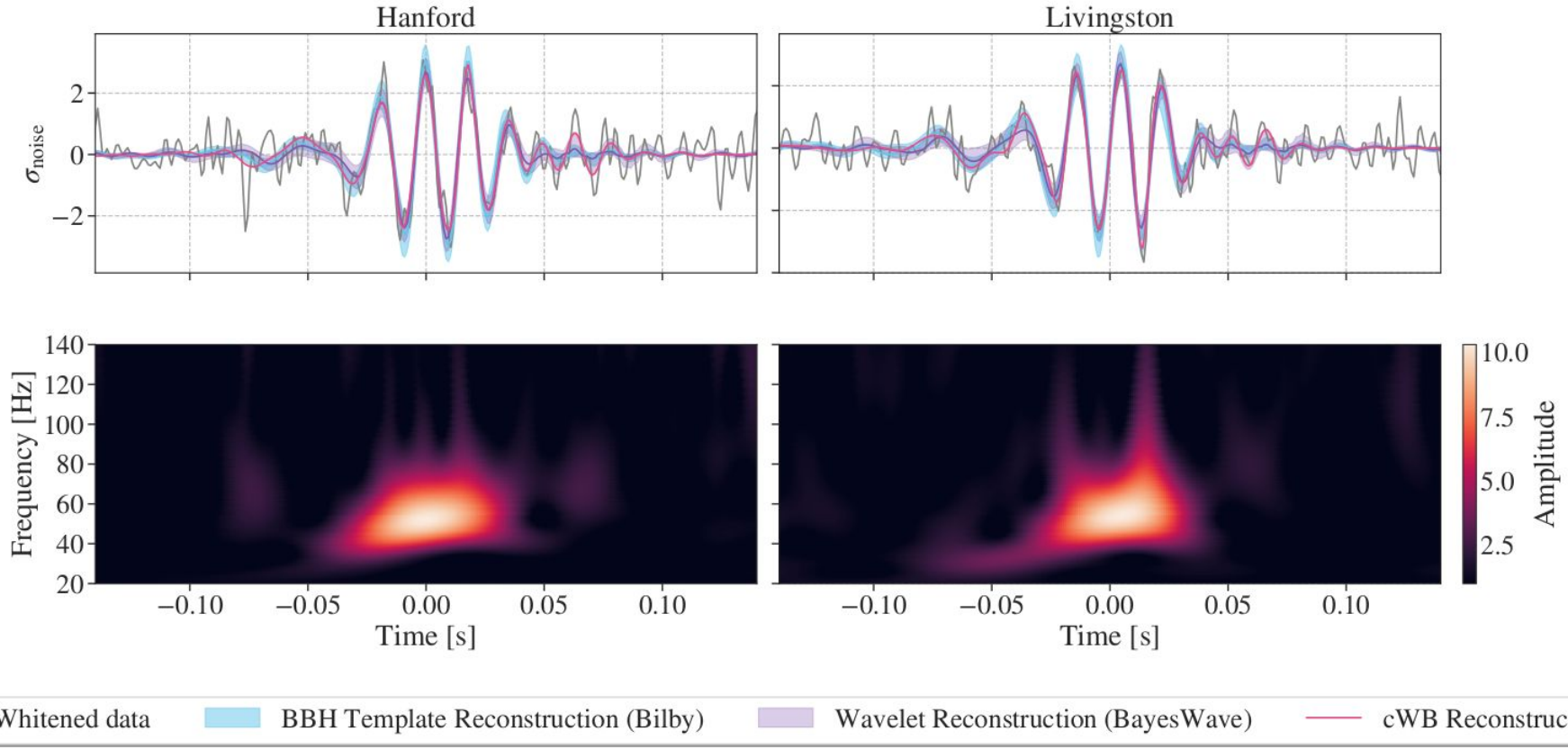


image credit: Eliot Finch

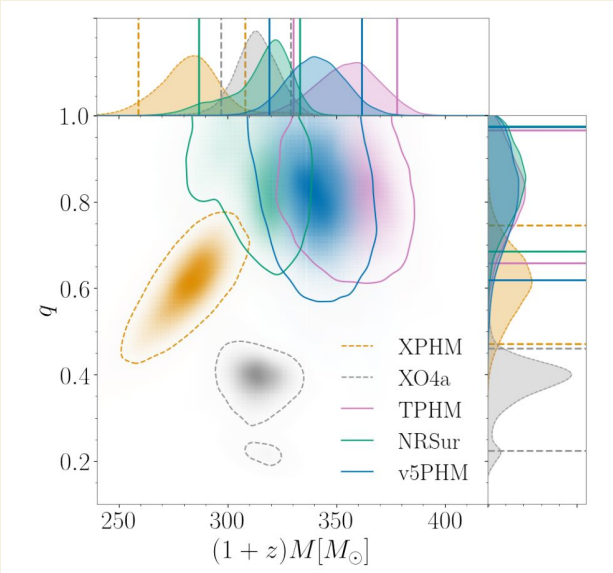
aligned spins:
higher effective aligned spin parameter
lower effective precessing spin parameter
negligible effect of higher order modes

misaligned spins:
lower effective aligned spin parameter
higher effective precessing spin parameter
significant presence of higher order modes

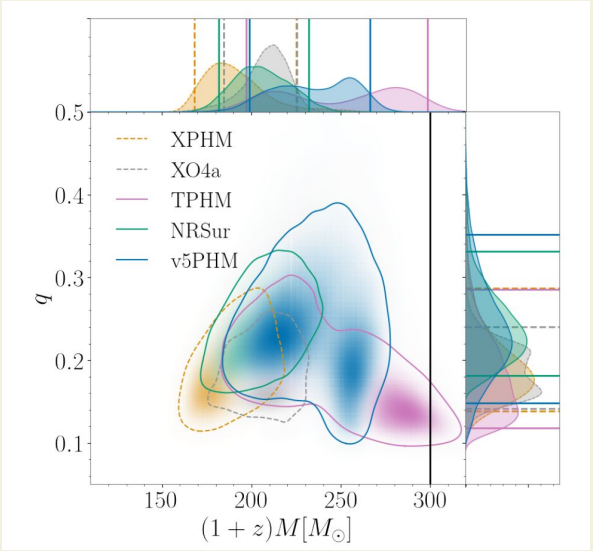


GW231123 as observed by LIGO Hanford (H1) and LIGO Livingston (L1). Virgo was offline at the time of detection. Time measured relative to 13:54:30 UTC, 23rd November 2023. Figure source: LVK Collaboration, 2025

This event is extremely challenging to analyze with current techniques. Not only was the detected signal very short, but also posteriors inferred using different waveform approximants show significant waveform systematics.



Systematics observed in GW231123

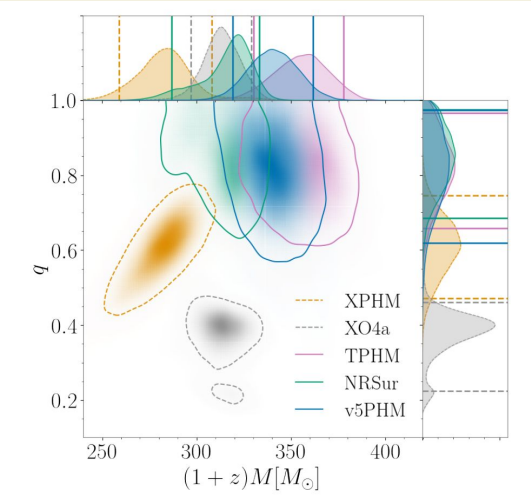


Inference on NR simulation (SXS:BBH:4030) (q=1, high spins)

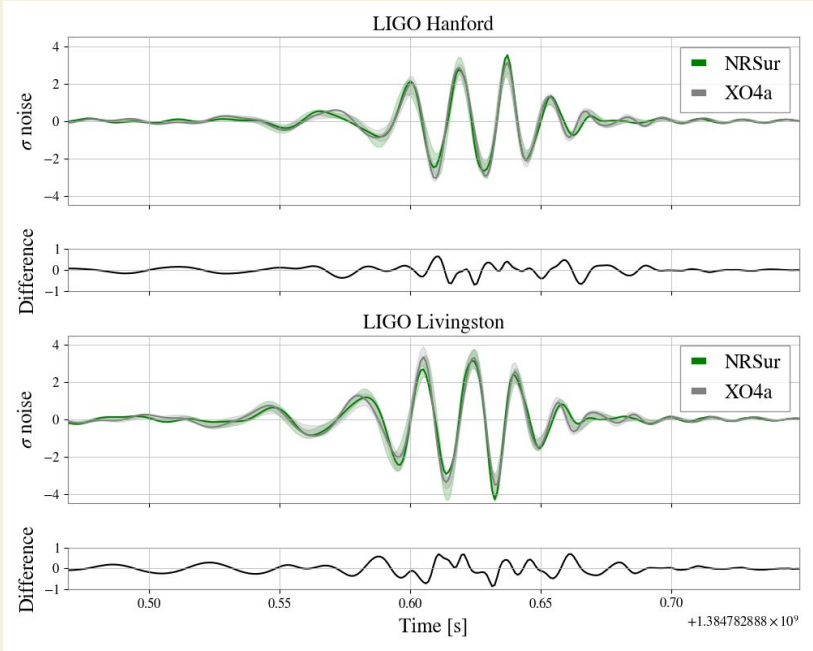
Systematics in measurement of detector-frame mass using different waveform approximants (left). The LVK paper was unable to reproduce these systematics in inference of NR simulations (right).
Figures from the LVK paper on GW231123

Can we reproduce waveform systematics in absence of noise?

We inject NRSur maximum-likelihood waveform in zero-noise and recover with 3 waveform models

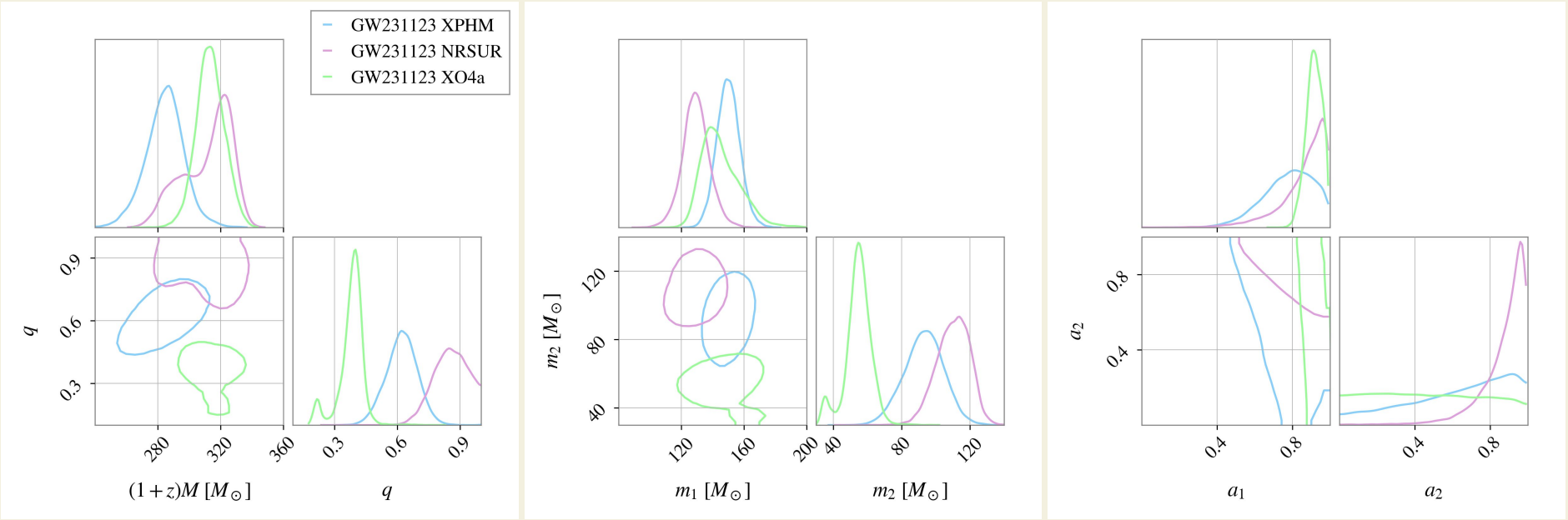


Systematics observed in GW231123

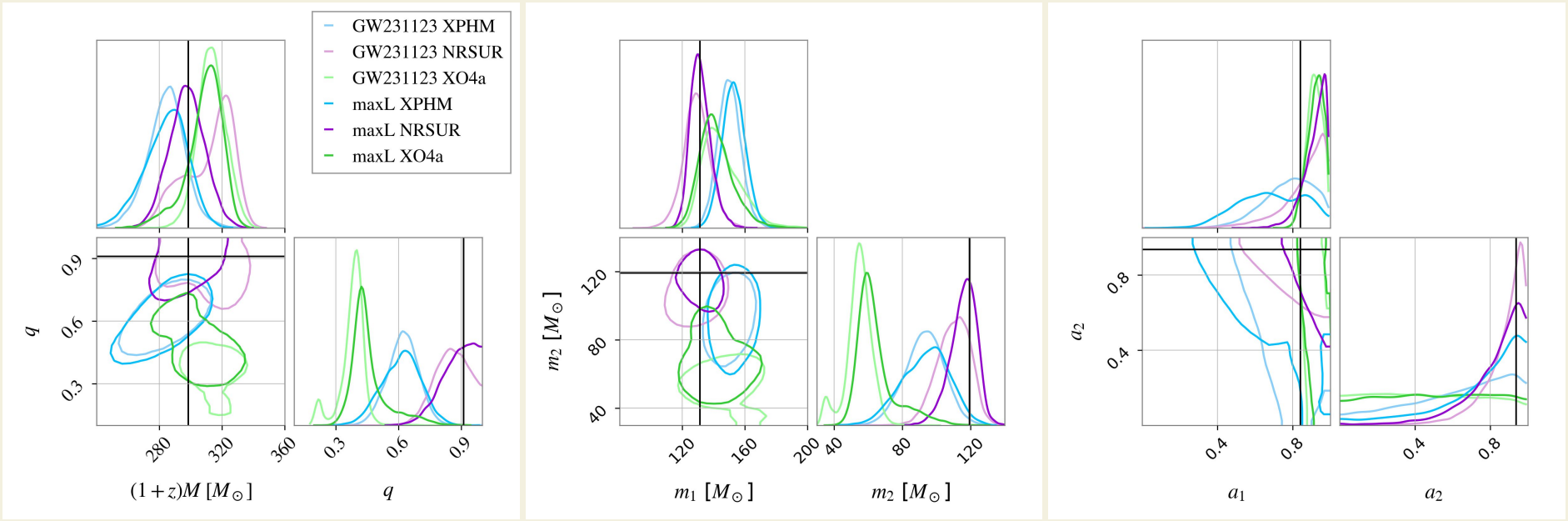


NRSur and **XO4a** maximum - likelihood waveforms

Comparison between GW231123 results and NRsur maximum-likelihood simulation in zero-noise



Comparison between GW231123 results and NRsur maximum-likelihood simulation in zero-noise



We can reproduce the waveform systematics observed in GW231123 in zero-noise

The Collaboration paper on GW231123 notes that there is a significant discrepancy between inferred posteriors from a LIGO Hanford-only run and a LIGO Livingston-only run.

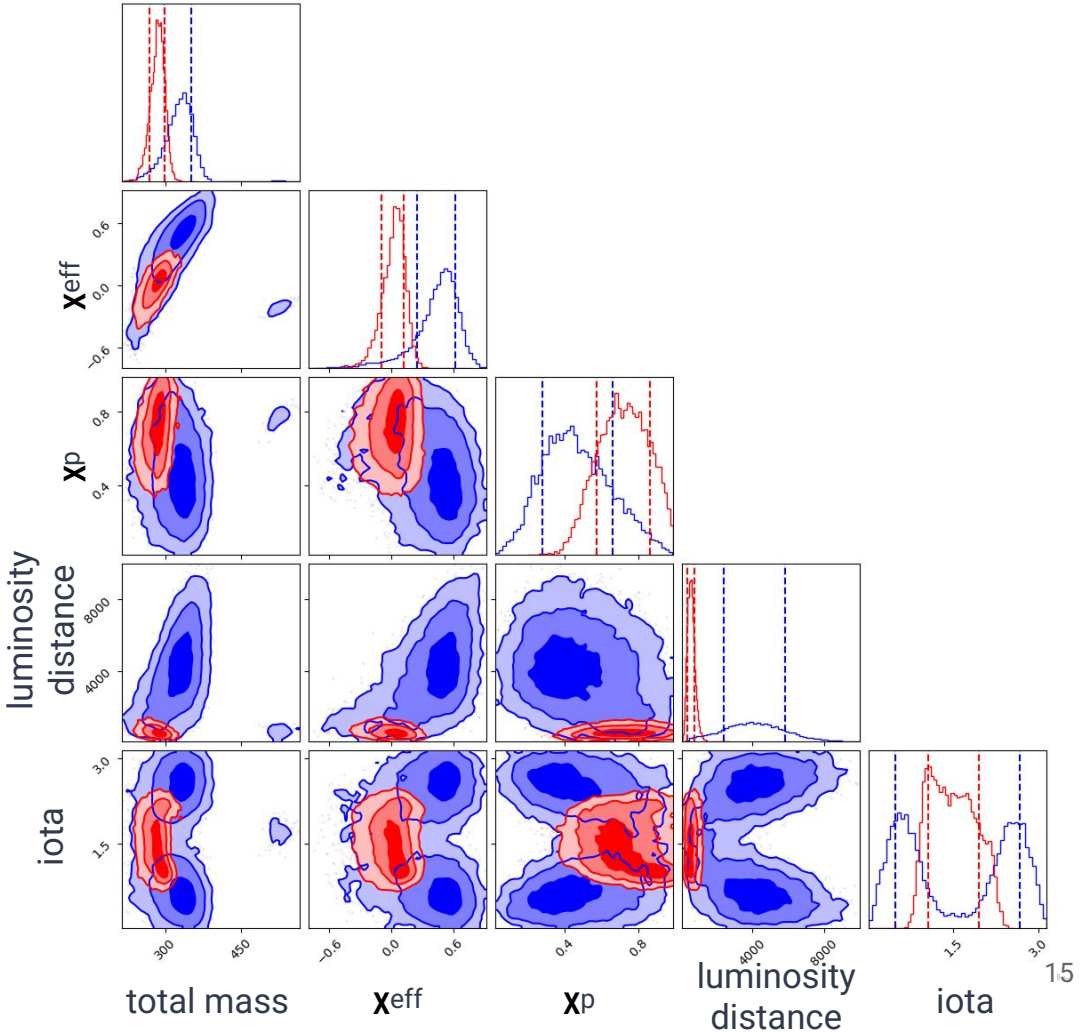
This was the most important part of my project. Let’s take a closer look.

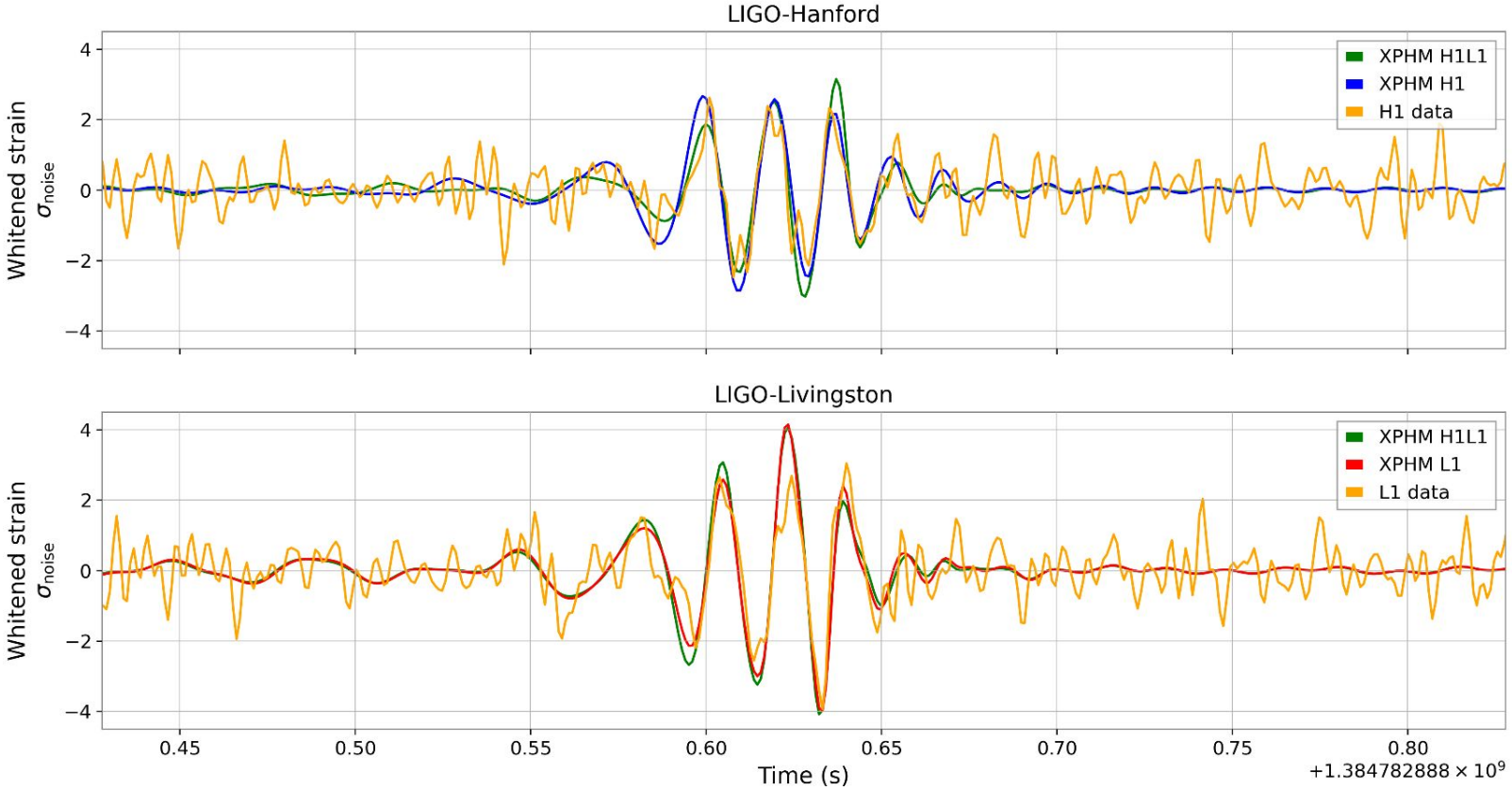
Results of Hanford-only & Livingston-only runs

Hanford is blue, Livingston is red.

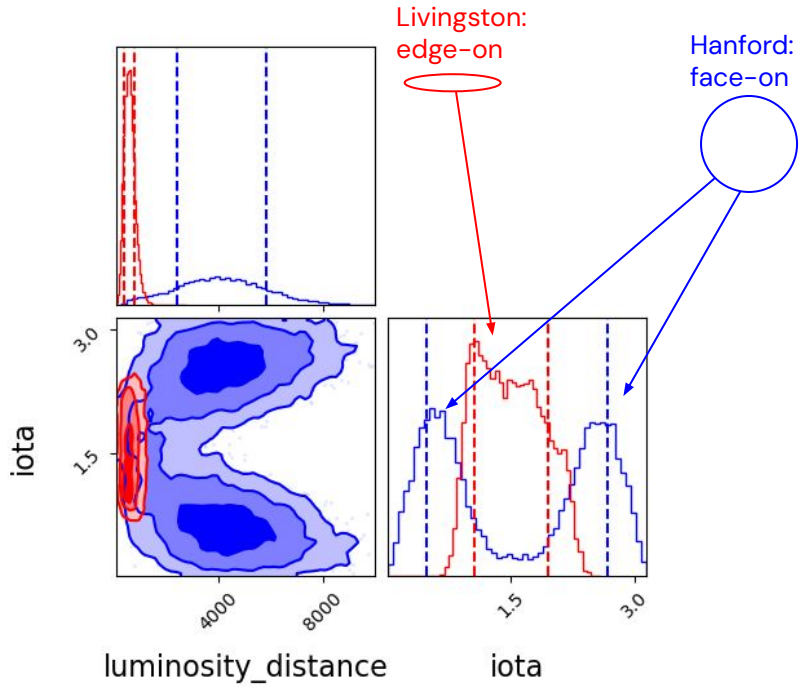
Corner plot showing estimated posterior probability distributions for selected parameters showing the most significant differences between detectors.

Results show that GW231123 is the most massive event ever observed, but constraining its parameters *with certainty* is challenging.




LIGO-Hanford
Whitened strain
σ_{noise}
XPHM H1L1
XPHM H1
H1 data
LIGO-Livingston
Whitened strain
σ_{noise}
XPHM H1L1
XPHM L1
L1 data
Time (s)
+1.384782888 × 10^9

A more physical explanation - higher order modes

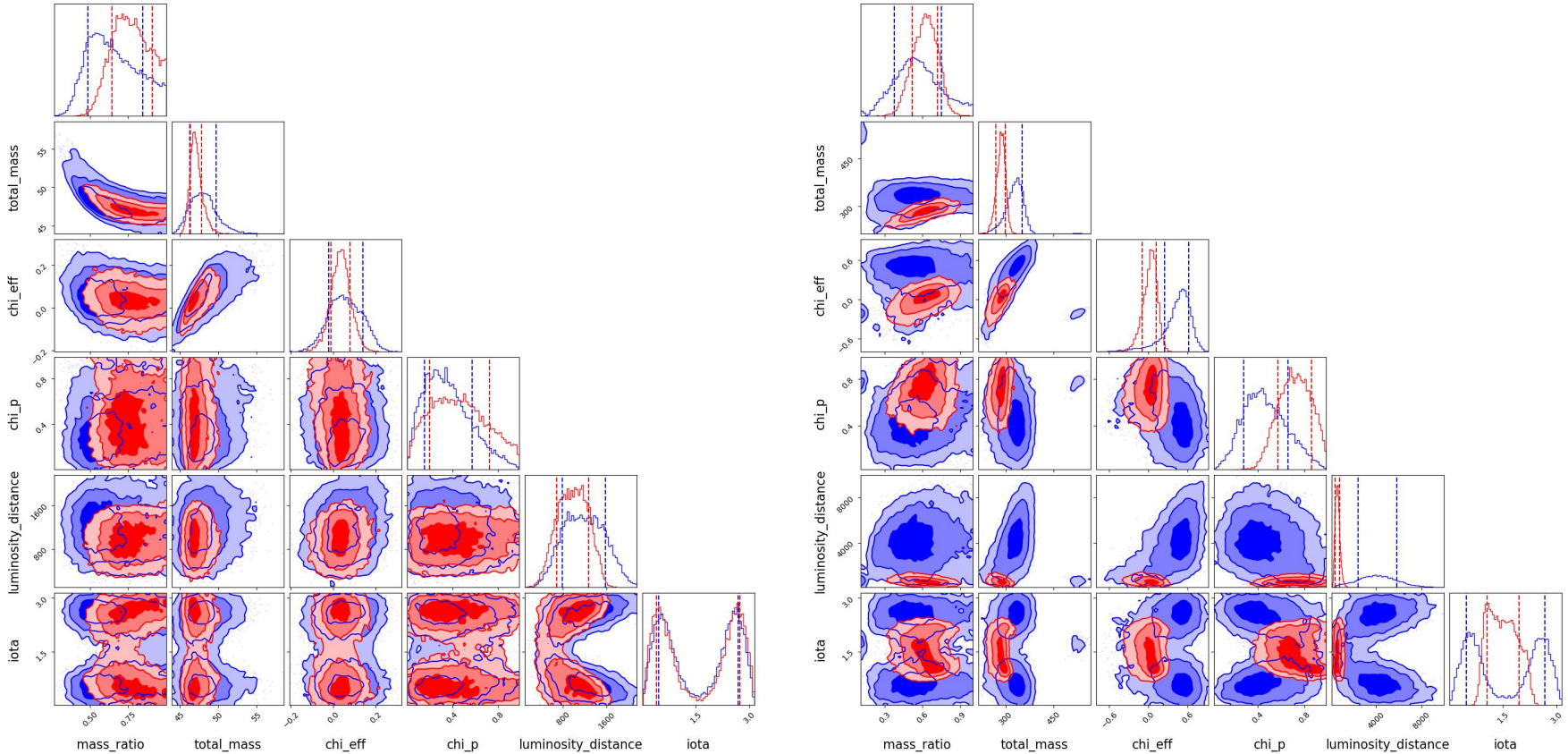


Higher order modes allow to resolve the inclination angle-luminosity distance degeneracy, allowing for distance measurement - even using a single detector.

H1-only and L1-only parameter estimation (PE) lead to very different measurements of SNR for higher order modes. The maximum likelihood values for event observed by Hanford are:

H1-only SNR:	**L1-only SNR:**
(2, 2) - 13.3	(2, 2) - 14.9
(3, 3) - 2.0	(3, 3) - 3.0
other modes < 1	(3, 2) - 2.9
	(2, 1) - 2.8
	(4, 4) - 2.6
combined - 14.35	**combined - 14.29**

How do GW231123's posteriors compare to other events? Is the behavior we see normal?
Let's look at GW230927be - a "standard" event from O4a.



GW230927be (left) and GW231123 (right) posterior comparison for selected parameters.

Let's quantify the differences: Jensen-Shannon divergence

JS is bounded by [0, 1].

JS=0 iff distributions are identical.

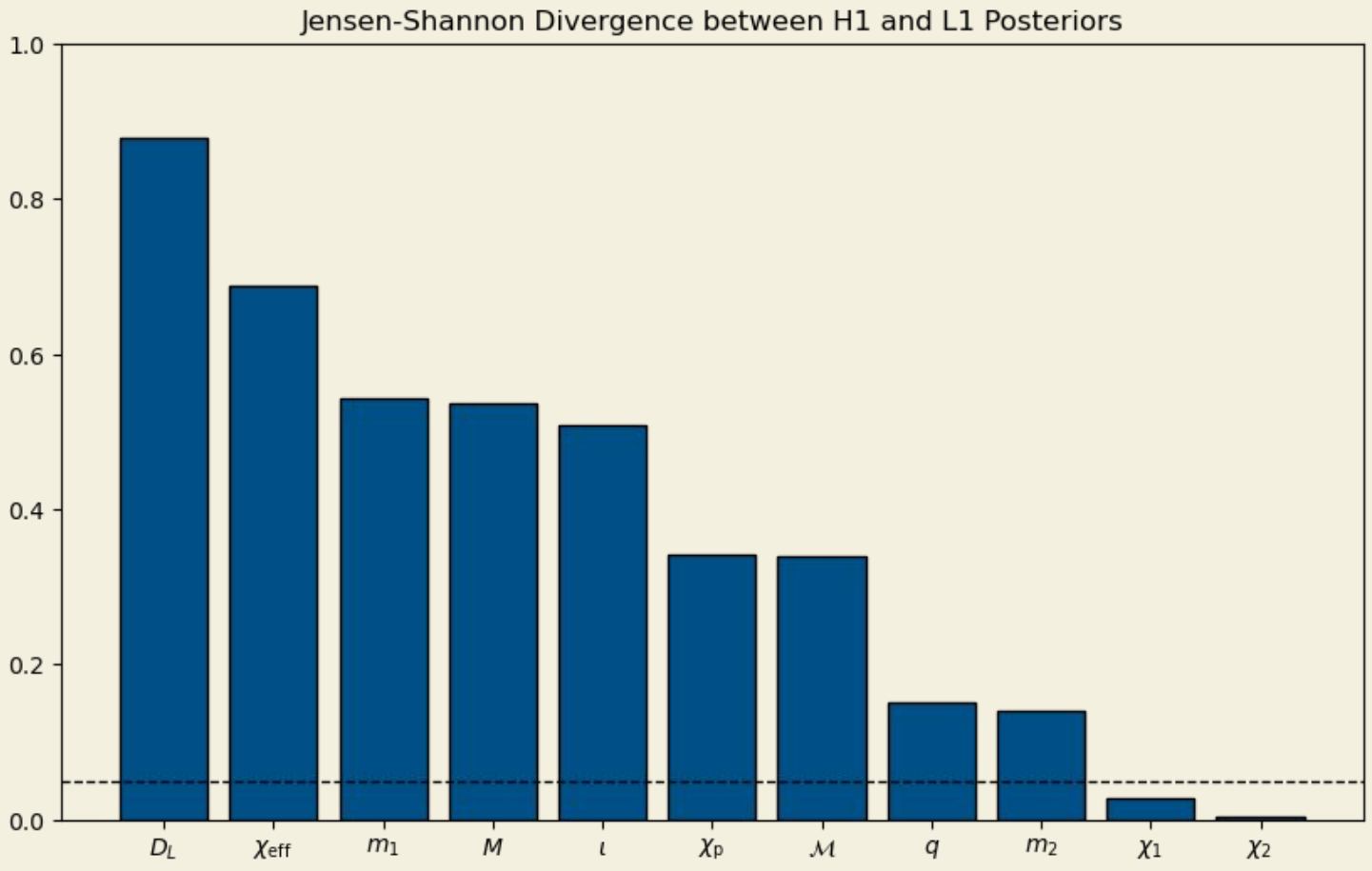
Due to PE seeds, etc. JS~0.05 is typically considered the "significance" threshold.

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

$$M = \frac{1}{2}(Q_1 + Q_2)$$

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

JS - Jensen-Shannon divergence
Q - distributions
D - Kullback-Leibler divergence
M - mixture distribution (convex combination)
we use $\log_2$



Jensen-Shannon divergences between LIGO Hanford-only run and LIGO Livingston-only run for selected parameters. Approximate “significance threshold” of 0.05 is marked with a dashed line.

What could be the source of these differences?

We can think of a couple possible explanations.

1. **Available waveform approximants fail to correctly model waveforms in this region of the parameter space.**

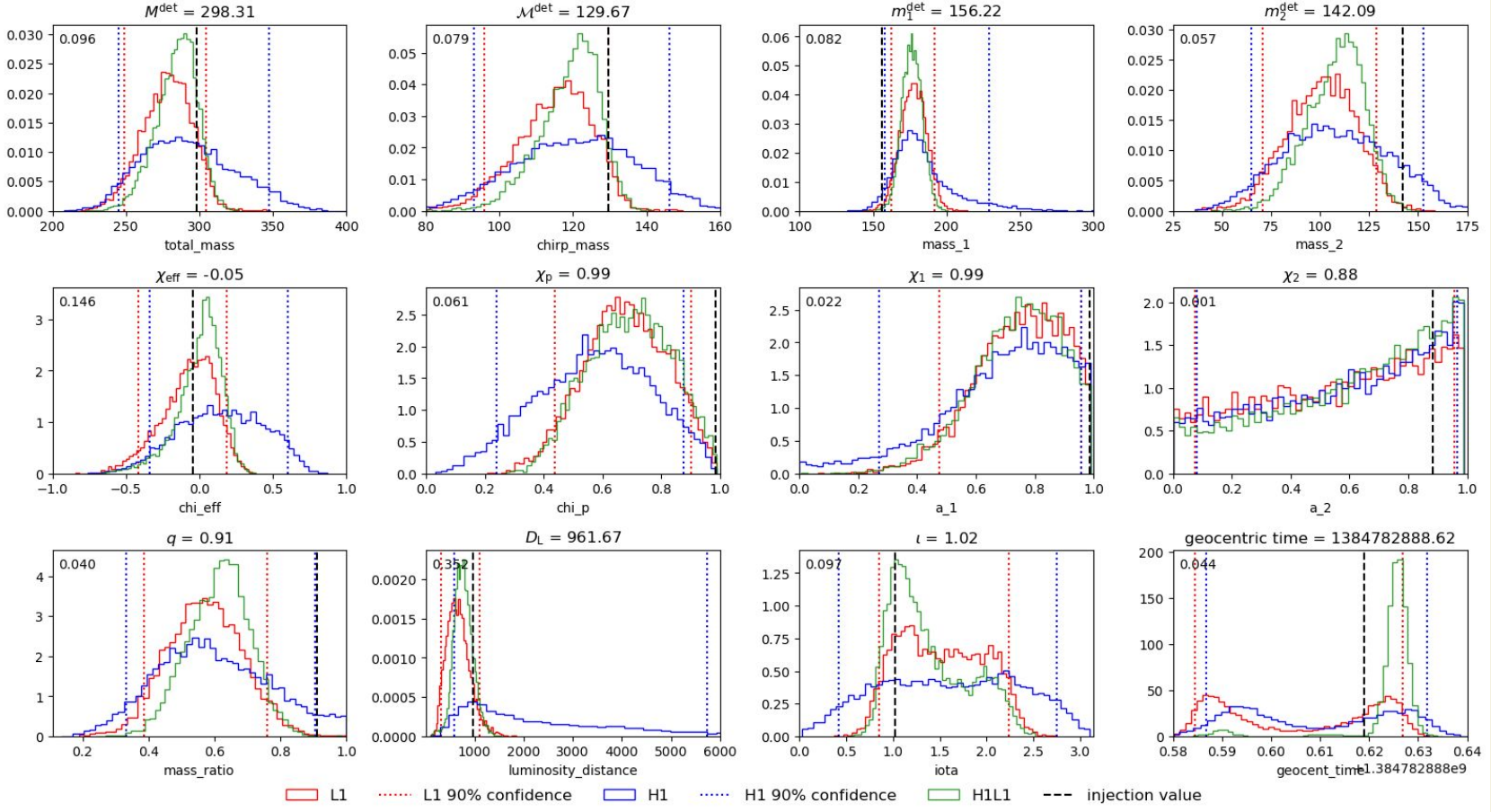
We know they perform poorly from the GW231123 paper, but we don't know if this is the sole reason.

2. **Certain Gaussian noise realizations could bias the parameter estimation.**

That can be checked by performing injections with Gaussian noise.

3. **Non-Gaussian noise (glitch) coincident with the signal.**

We hope this is not the case - it would be very challenging to deal with.

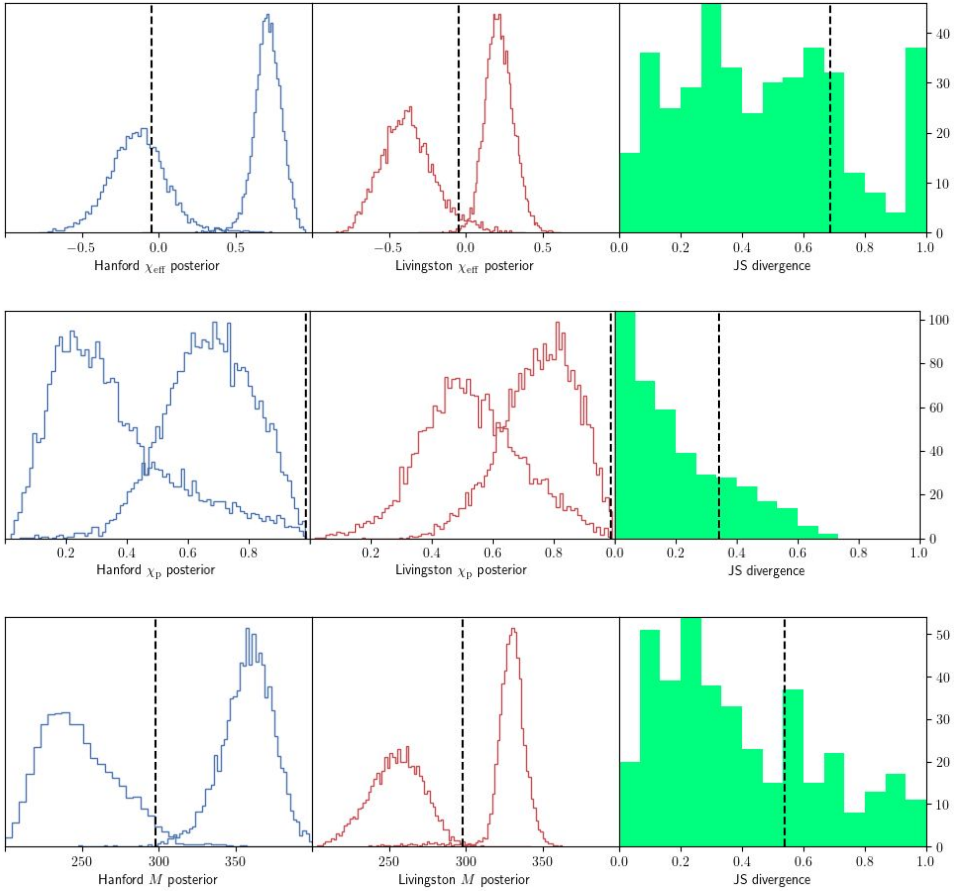


Injecting NRSur7dq4 maximum likelihood waveform into zero-noise and retrieving with NRSur7dq4 does not lead to significantly different posteriors from the runs, only to wider/narrower posteriors.

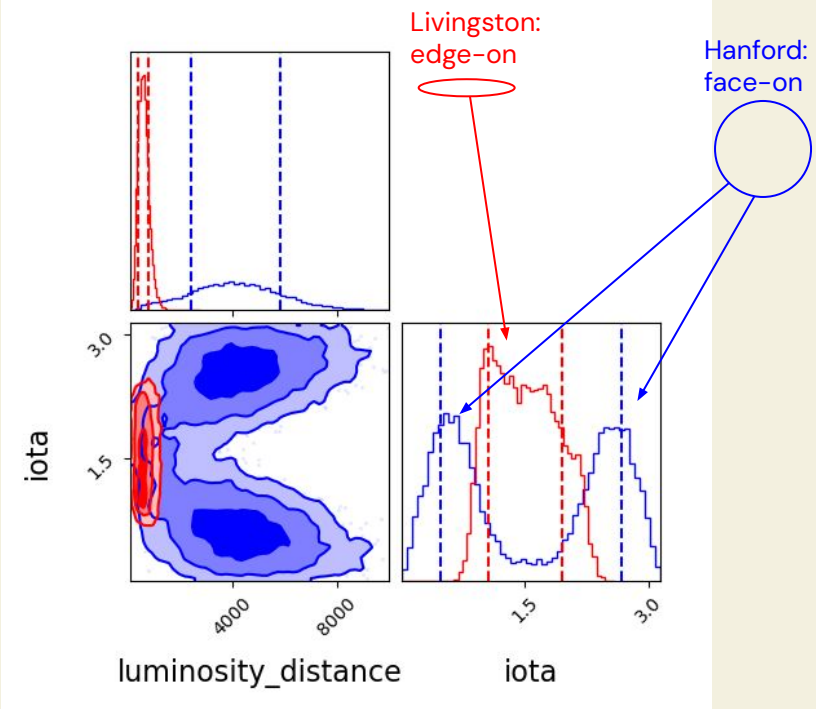
We ran maxL injections to check the impact of Gaussian noise

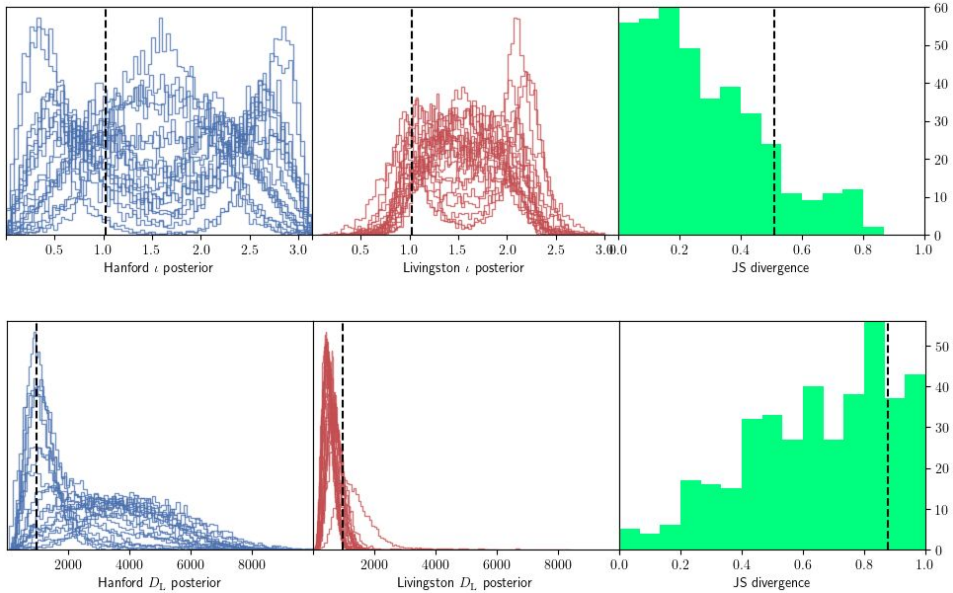
Maximum likelihood injections in Gaussian noise (with PSD identical to that calculated for the time of detection) should reproduce the behavior we see in GW231123

Left two histograms are extreme posteriors from H1-only and L1-only PE with injected value marked with black dashed line. Rightmost histograms show distribution of JS divergences of combinations of runs with black dashed line indicating the value for GW231123.



Inclination and distance





We see behavior similar to GW231123 for inclination in around 13% of the seeds and behavior similar for luminosity distance in 18% of the seeds.

Conclusion

We were able to reproduce waveform systematics similar to those observed for GW231123 in zero-noise injections.

We also found that Gaussian noise fluctuations can cause differences similar to the ones we see in GW231123's parameter estimation.

The fact that such differences have not emerged for other events suggest that Gaussian noise might impact shorter events to a larger degree.

We have corroborated these findings using IMRPhenomXPHM and XO4a waveform approximants.

Our findings demonstrate that the main findings of the LVK paper on GW231123, namely the astrophysically relevant properties of GW231123 (high mass and high spin magnitudes) inferred with NRSur7dq4, are robust.

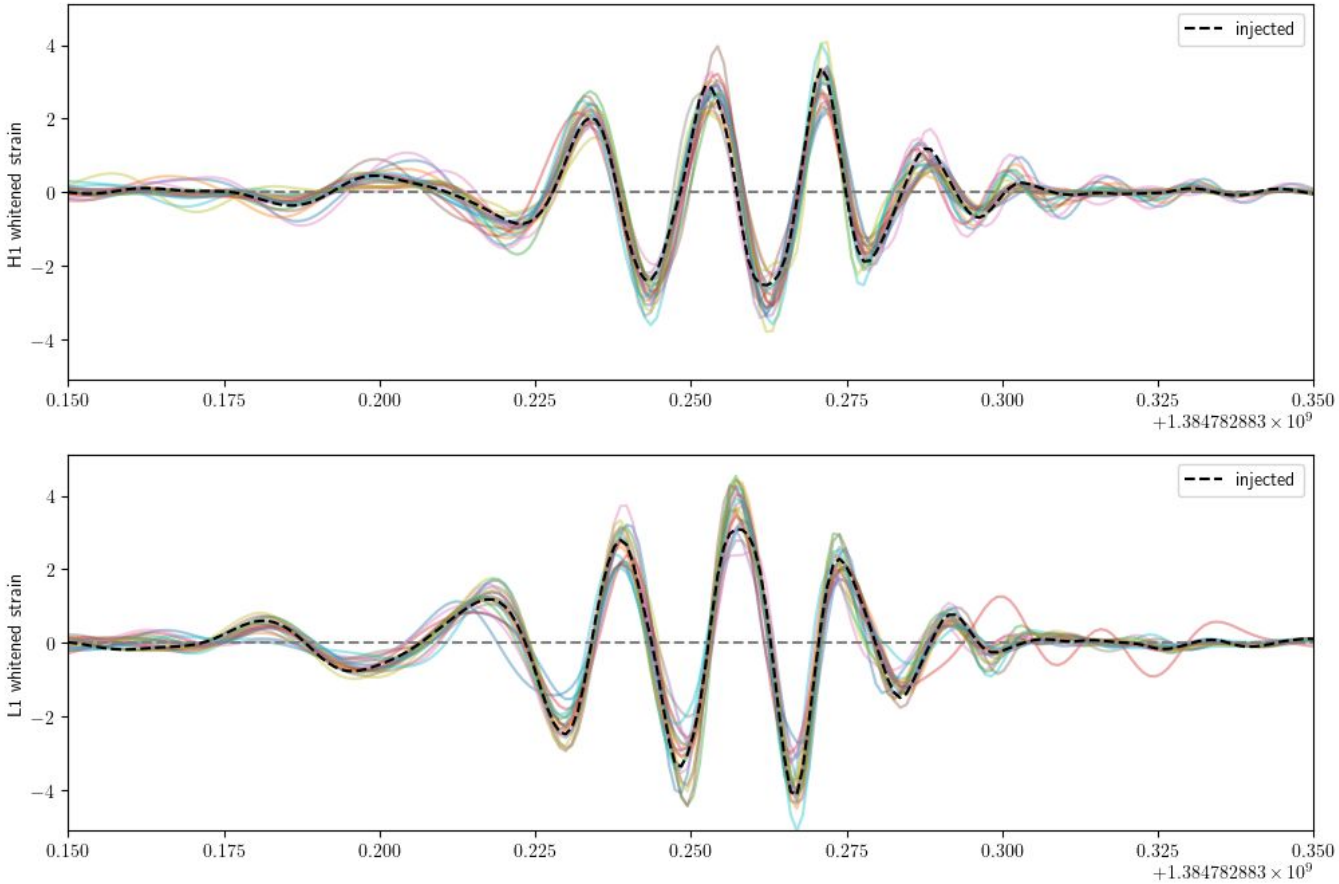
For such events, differences like those observed in GW231123 between single-detector posteriors are expected.

Bonus slides

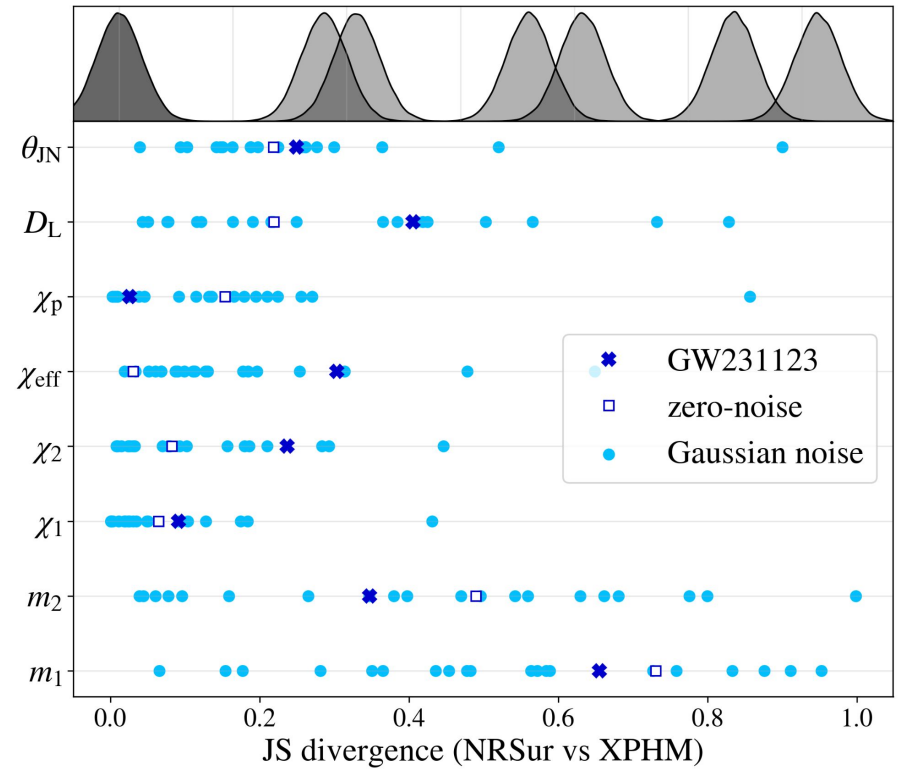
Parameter	Value
luminosity distance $D_{\rm L}$ (Mpc)	962
primary spin magnitude χ_1	0.988
secondary spin magnitude χ_2	0.883
eff. aligned spin parameter $\chi_{\rm eff}$	-0.047
eff. precessing spin parameter $\chi_{\rm p}$	0.986
inclination angle ι (rad)	1.03
source-frame primary mass ($M_\odot$)	131
source-frame secondary mass ($M_\odot$)	119
source-frame total mass ($M_\odot$)	250
detector-frame primary mass ($M_\odot$)	156
detector-frame secondary mass ($M_\odot$)	142
detector-frame total mass ($M_\odot$)	298
mass ratio q	0.91
geocentric time since 1384782888 (s)	0.619
phase φ (rad)	2.35
viewing angle $\theta_{\rm JN}$ (rad)	1.34
LIGO Hanford optimal SNR	13.6
LIGO Livingston optimal SNR	16.0

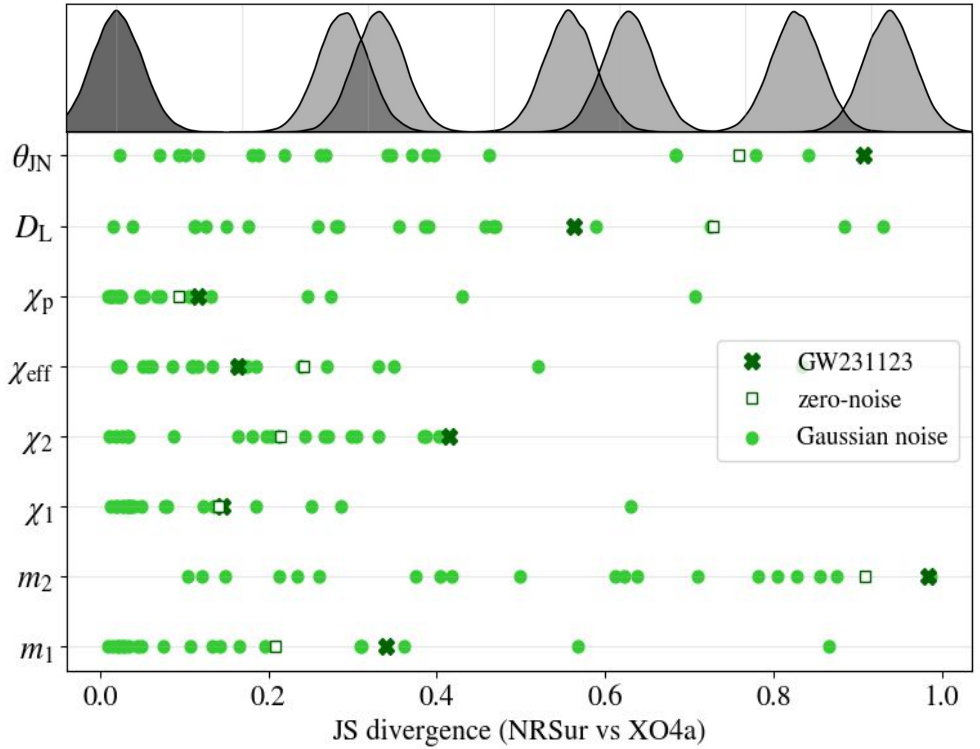
Parameter	JS $\geq$ JS$_{\rm GW231123}$
primary spin magnitude χ_1	0.81
secondary spin magnitude χ_2	0.73
detector-frame secondary mass	0.68
mass ratio q	0.63
detector-frame total mass	0.30
eff. precessing spin parameter $\chi_{\rm p}$	0.22
eff. aligned spin parameter $\chi_{\rm eff}$	0.20
luminosity distance $D_{\rm L}$	0.18
detector-frame primary mass	0.18
inclination angle ι	0.13
phase φ	0.06

Left: NRSur7dq4 maximum likelihood parameters. Right: Fraction of injection seed combinations with JS divergence higher or equal to that of GW231123.

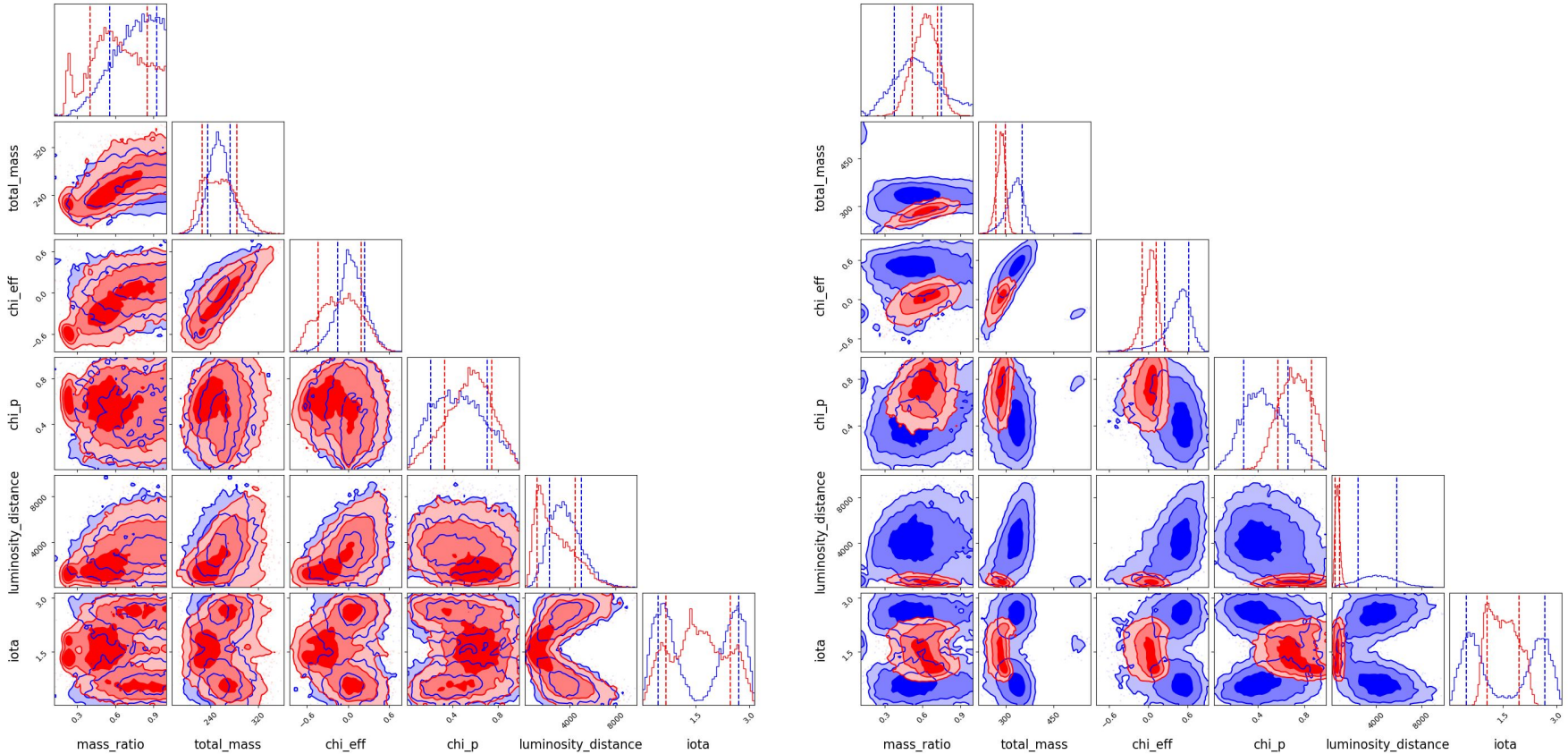


Maximum likelihood whitened waveforms from the 20 Gaussian noise injections in LIGO Hanford and 20 Gaussian noise injections in LIGO Livingston. The injected waveform is plotted with a dashed black line. The y-axis is in noise standard deviation units.





Gaussian noise shifts the posterior distributions. Such shifts might increase or reduce the waveform systematics.



GW190521 (left) and GW231123 (right) posterior comparison for selected parameters.