LIGO SURF 2021 Interim Report 1: Marginalizing over noise properties in signal reconstruction

Cailin Plunkett

Advisors: Katerina Chatziioannou, Sophie Hourihane
(Dated: July 2 2021)

The traditional gravitational wave parameter estimation process relies on sequential estimation of noise properties and binary parameters. Using new capabilities of the `BayesWave` algorithm and recent developments in noise uncertainty modeling, we will simultaneously estimate the noise and binary parameters, which will mitigate the assumption of known noise variance in the fitting process. We will do so using both the wavelet-based and template-based models available in `BayesWave`. We will quantify any differences between these methods on parameter recovery and analyze the impact for astrophysical inference.

I. BACKGROUND

Gravitational wave (GW) data analysis requires models of both the genuine GW signal and the frequency-dependent noise in the raw data. Accurate parameter estimation of black hole and neutron star properties from compact binary coalescence (CBC) signals depends on the robustness of both of these models [1]. While creating waveform templates by numerically solving Einstein's equations has been the subject of many research operations over the last decades [2], noise models have not been traditionally given the same amount of attention.

The traditional parameter estimation process uses sequential estimation of the noise properties and the binary parameters. First, the noise is modeled using one of several methods, such as a periodogram- or Welch-based approach that averages the power spectrum of the data from many segments around, but not including, the segment containing the signal [3]. The resultant noise PSD is given to `LALInference` (LI), the primary parameter estimation pipeline used by the LIGO and Virgo collaborations [4]. LI and its successor `Bilby` are template-based GW searches that use a provided fixed noise model in their Bayesian estimation of binary parameters.

Consequently, analyses of CBC signals make three explicit assumptions about the noise properties: first, that the noise is Gaussian; second, that it is stationary in time; and third, that its frequency-dependent variance is known [3]. In practice, all three assumptions break down. The third is invoked in the sequential estimation of noise and parameters, which we will address by simultaneously inferring noise properties and binary parameters.

Crucial to parameter estimation is the likelihood function, $\mathcal{L}(d|h')$, which computes the probability density of measuring the detector data d under the condition of a true GW signal h'. The log-likelihood, after invoking the stationarity assumption, reduces to

$$\mathcal{L}(d|h') = -2\sum_{i}^{N/2} \frac{\tilde{r}_i \tilde{r}_i^*}{TS_n(f_i)} + \text{const}, \tag{1}$$

where $r = d - h'$ are the residuals, the tilde denotes the frequency domain, the star the complex conjugate, i iterates over frequency bins, and N is the number of time samples, equal to the sampling rate times the duration T. The details of the derivation are beyond the scope of this report, but are provided in Veitch *et al.* [4] and Chatziioannou *et al.* [3]. The key aspect is that the likelihood is explicitly dependent on $S_n(f)$, the power spectral density of the noise. As parameter estimation relies on a noise-weighted inner product, properly characterizing the noise is necessary. Chatziioannou *et al.* [3] provides and compares two methods of computing $S_n(f)$ to robustly estimate the noise variance in GW data.

The first method is that of the periodogram-based approach described above, also called an "off-source" approach as it does not use data containing the detected signal. The "on-source" spectral estimation method, in contrast, uses only data containing the signal. The second method uses the `BayesLine` (BL) algorithm [5], which is integrated into the broader `BayesWave` (BW) algorithm [1], a variable dimension, parameterized model to separate transient GW signals from detector noise. BL estimates the noise power spectral density (PSD) as a sum of spline and Lorentzians, selecting model parameters via a Markov Chain Monte Carlo method.

Chatziioannou *et al.* [3] found the "on-source" spectral estimation method to produce whitened data more consistent with a Gaussian likelihood. In addition, both methods for estimating $S_n(f)$ were tested on simulated CBC signals injected into observational data from the Advanced gravitational-wave detector network. Quantitative differences between the resultant parameter estimations demonstrated the importance of the chosen model for the noise variance and further confirmed the comparative strength of the on-source method. Fig. 3 depicts example power spectra with the data in light gray, recovered signal in color, and the noise PSD and associated uncertainty in black. This plot comes from an analysis I ran on GW150914.

With either method to produce the PSD, the standard process remains to feed the computed median PSD into LI to estimate binary parameters. A recent development in the capability of BW, described in detail in Chatzi-

ioannou *et al.* [6], enables it to compute binary and noise parameters in concert. The simultaneous likelihood estimation mitigates the third assumption explained above and incorporates uncertainty in the PSD. As of yet, the method of marginalizing over the noise properties in parameter estimation have not been applied to actual CBC events, only to injected signals. As such, exploring the impact of these methods is an active area of research to which we aim to contribute.

II. PROJECT OBJECTIVES

We will study the effect of including uncertainty in noise on parameter estimation of the confirmed LIGO and Virgo CBC events. We will use BW to simultaneously model the noise $S_n(f)$, obtained using the on-source estimation method, and the signal. We will compare those results to ones obtained using the sequential estimation method, which first models only the noise, then extracts waveform parameters assuming that fixed noise model. We will quantify any differences.

Specifically, I will run two analyses on each GW signal. On one hand, I will run the "traditional" signal estimation method: sequential computation of first the noise PSD, which we use as a fixed input to estimate the signal properties. To do so, I will first run the data through BW on `cleanOnly` mode with `bayesLine` turned on. This cleaning mode coupled with BL estimates the noise PSD, the fixed output of which we will feed into another BW run, with the `signalOnly` model and BL turned off. The `signalOnly` model forces the algorithm to find a signal in the data.

The second analysis is the simultaneous estimation method. I will run BW on `signalOnly` mode with `bayesLine` turned on. This computes, in one step, the signal properties and noise PSD with uncertainty, mitigating the known-variance assumption. Table I schematically shows the difference between these methods.

The main BW pipeline uses wavelets to reconstruct the GW signals, rather than templates, like `Bilby`. This enables it to find weakly modeled GW signals as well as well-modeled events, like CBCs. Because this method does not use templates, it does not return CBC parameters like masses and spins for CBC events, but wavelet parameters, such as amplitude and frequency. A new capability of BW is the CBC model, which is a template-based search. I will also compare the sequential and simultaneous waveform estimation methods for this model to illuminate the quantitative differences between the methods. This model will output time-domain waveforms and power spectra like the wavelet model, but will also return credible regions for binary component masses. Fig. 1, which is Figure 6 from Abbott *et al.* [7] is the type of plot we aim to produce and compare: it shows the 90% credible regions for all CBC candidate events in GWTC-2 in M-q space, with events published prior to that study highlighted. I will quantify differences in these parameter estimates when we model the noise and signal sequentially versus in concert by using the mass posteriors given by the CBC model.

We will draw CBC events from the Gravitational-Wave Transient Catalog for the first and second observing runs (O1, O2) as well as the first half of the third observing run (O3a) of the advanced gravitational-wave detector network (Abbott *et al.* 7, Abbott *et al.* 8). The network comprises the two Advanced LIGO detectors and, since August 1, 2017, the Advanced VIRGO detector. In total, the combined catalogs contain over 60 confident and candidate CBC events.

New Method	Traditional Method
signalOnly bayesLine	cleanOnly bayesLine ↓ PSD ↓ signalOnly PSD
⇓ Waveform	⇓ Waveform

TABLE I. Visual representation of the signal estimation methods I will be comparing.

III. PROGRESS

Thus far, I have run the two `signalOnly` analyses on GW150914, one of the loudest and most well-studied GW events. The time domain waveforms and power spectra for each analysis method are shown in Figs. 2 - 5. Note that in the following plots, the labels "fixed" and "simult" are used to refer to the sequential and simultaneous estimation methods, respectively, as the former uses a fixed input PSD. I have started full-length `cbcOnly` runs but have not completed their analysis. The power spectrum for a shorter run on GW150914 with a fixed PSD and some different parameters, is shown in Fig. 9. Even on a shorter run, the templated search produces much cleaner and less uncertain results, as is expected.

Figs. 6 and 7 overplot the two estimation methods in the time and frequency domains, respectively. The waveforms and uncertainties almost perfectly overlap in both, indicating near-perfect agreement between the two methods for GW150914.

Lastly, Fig. 8 shows the overlapping histograms of all the recovered waveform moments for each method. Again, there is close agreement between the two methods for this signal.

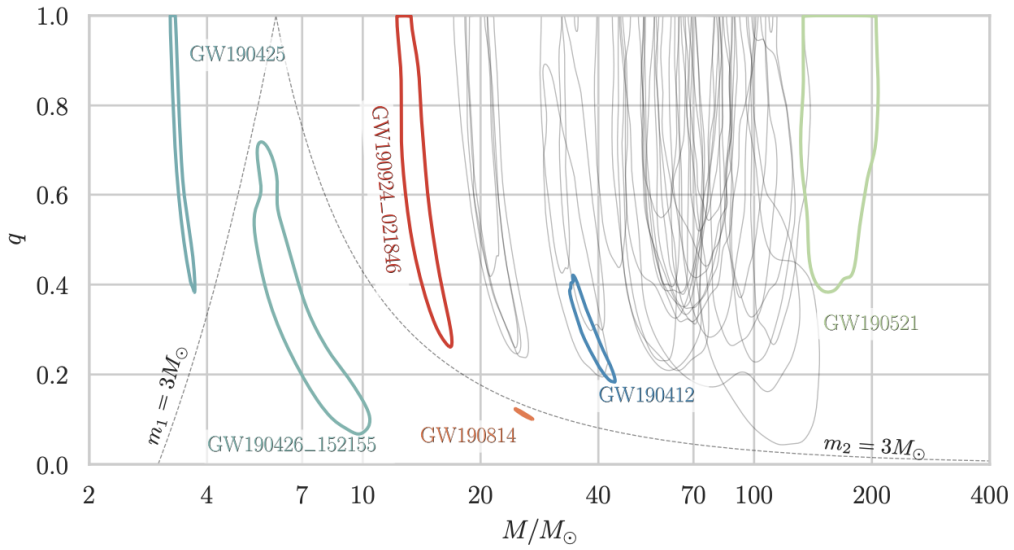

FIG. 1. 90% credible regions for all candidate events in total mass M and mass ratio q space, with previously published events highlighted. Dashed lines delineate where one of the objects can have a mass $< 3\,M_\odot$.

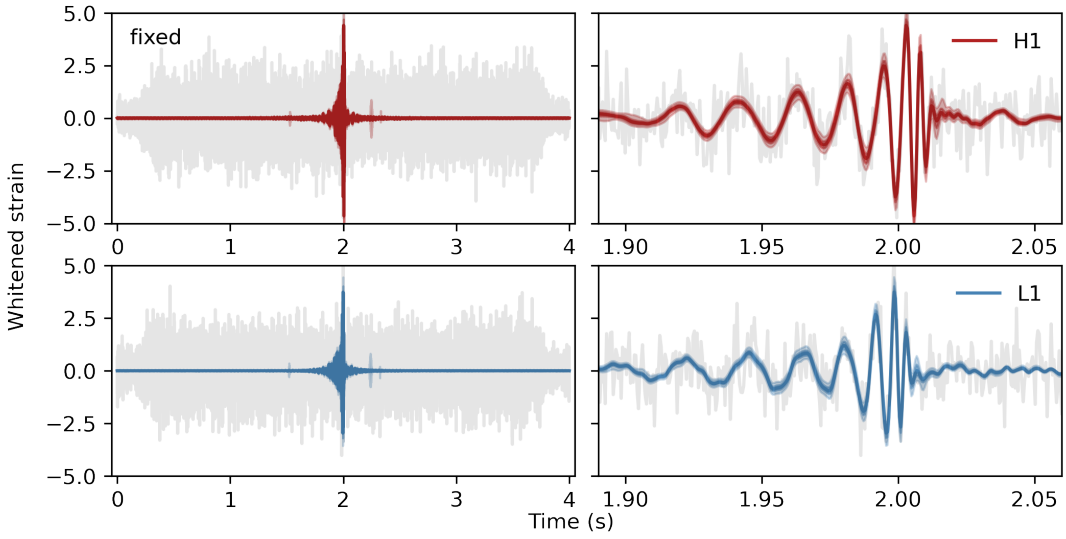

FIG. 2. Time domain waveform for the sequential estimation method, for each detector, at two levels of resolution.

IV. SUMMARY AND PLANNED WORK

In summary, I will run four analyses on GWTC-2 signals. Two will use `BayesWave`'s `signalOnly` model, which is a weakly modeled search that uses wavelets to model non-Gaussian features in the data, and the other two will use its `cbcOnly` model, which is a template-based search. In each case, the two analyses will be the sequential and simultaneous estimation methods described previously. I have completed the two `signalOnly` analyses on GW150914 and the `cbcOnly` analyses are in progress.

These analyses are the "prototypes" for the work I will be doing for the rest of the summer. I will continue similar runs and analyses on confirmed GW signals from GWTC-2. I will perform these methods of signal reconstruction and statistically compare the results.

[1] N. J. Cornish and T. B. Littenberg, Classical and Quantum Gravity **32**, 135012 (2015).

[2] L. Blanchet, Living Reviews in Relativity **17**, 2 (2014),

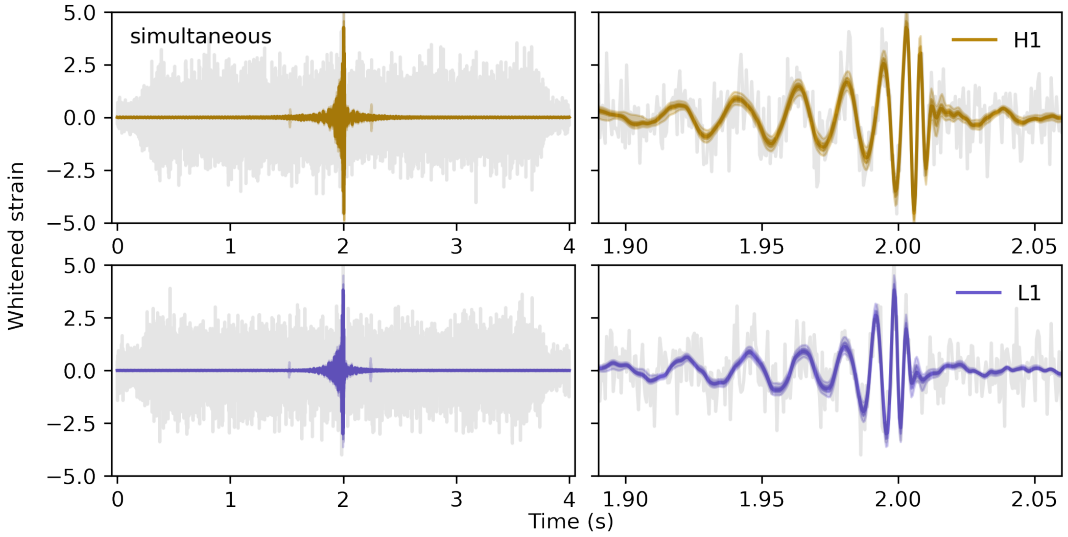

FIG. 3. Same as Fig. 2 for the simultaneous method.

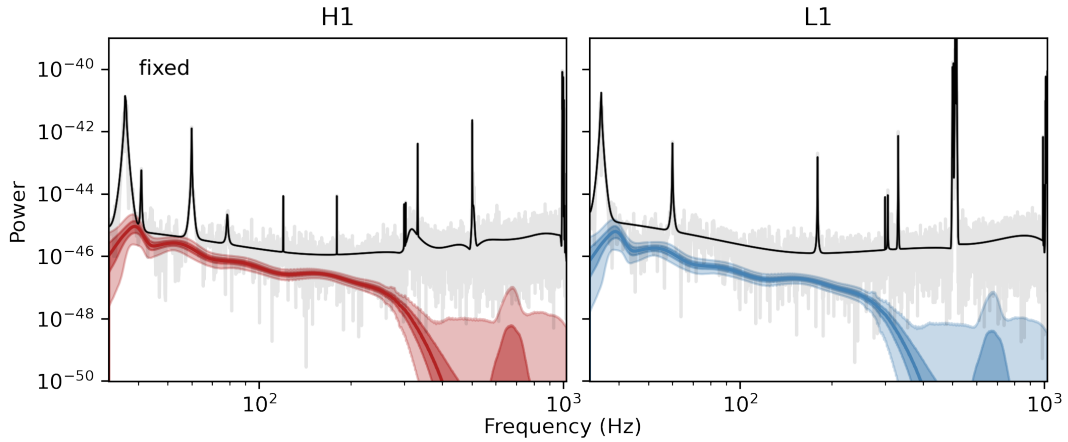

FIG. 4. Power spectra for the sequential estimation method. The data are in gray, the noise PSD in black, and the signals in color.

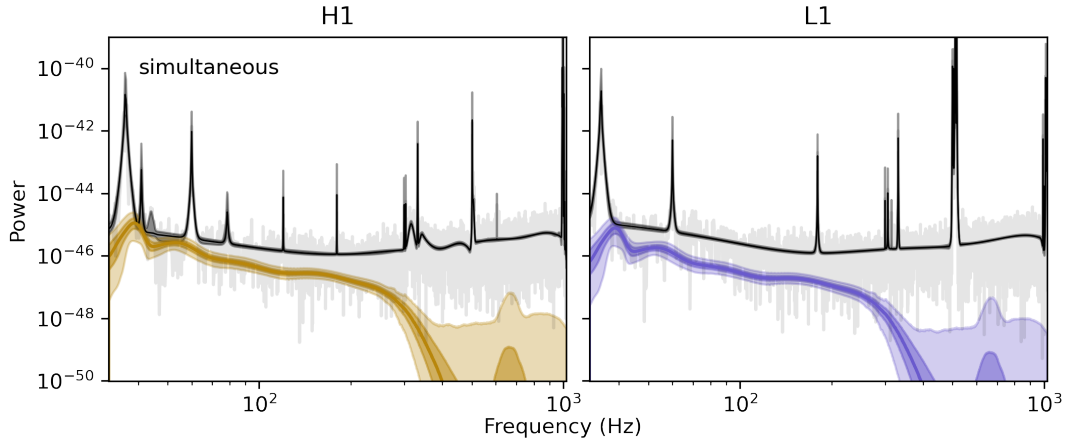

FIG. 5. Same as Fig. 4 for the simultaneous method; the uncertainty in the noise PSD is in dark gray.

arXiv:1310.1528 [gr-qc].

[3] K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M.

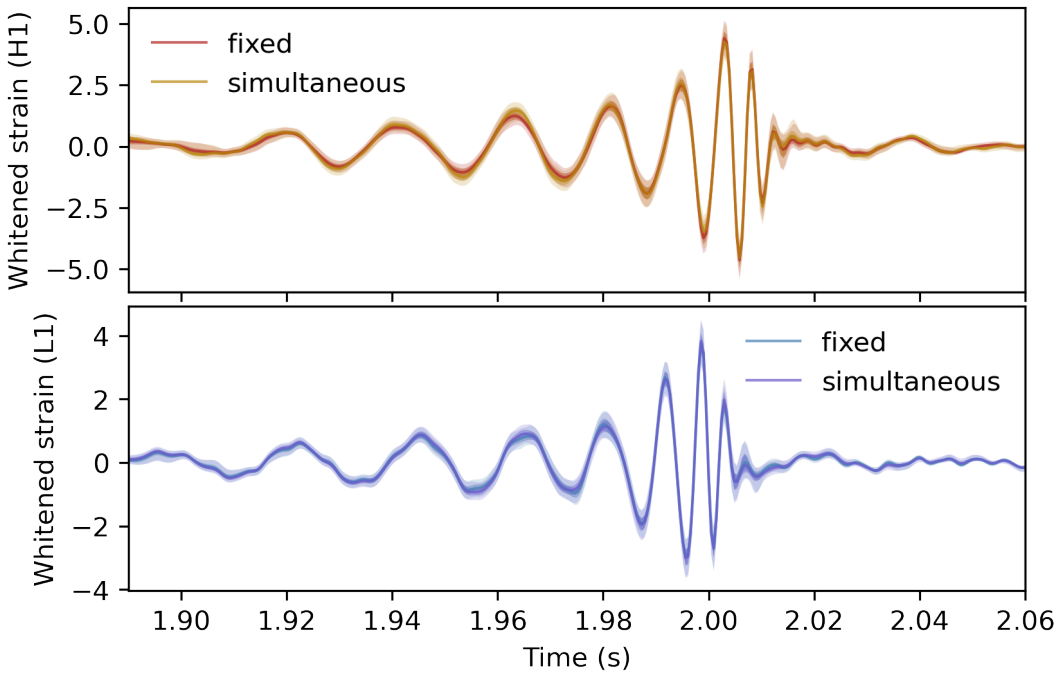

FIG. 6. Time-domain waveforms in each detector with the two analysis methods both plotted.

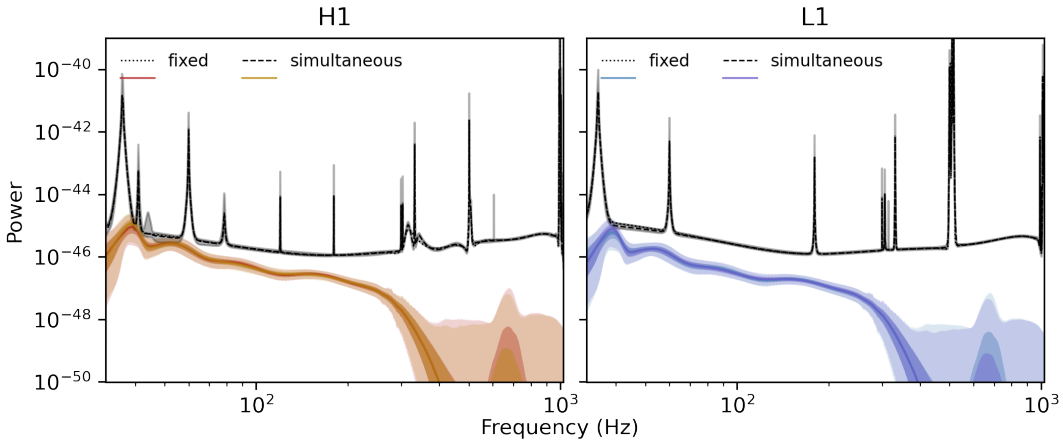

FIG. 7. Same as Fig. 6 for the power spectra.

Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, Physical Review D **100** (2019), 10.1103/physrevd.100.104004.

[4] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C. J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[5] T. B. Littenberg and N. J. Cornish, Physical Review D **91** (2015), 10.1103/physrevd.91.084034.

[6] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, Phys. Rev. D **103**, 044013 (2021), arXiv:2101.01200 [gr-qc].

[7] R. Abbott, T. D. Abbott, and S. e. a. Abraham, arXiv e-prints , arXiv:2010.14527 (2020), arXiv:2010.14527 [gr-qc].

[8] B. Abbott, R. Abbott, and T. e. a. Abbott, Physical Review X **9** (2019), 10.1103/physrevx.9.031040.

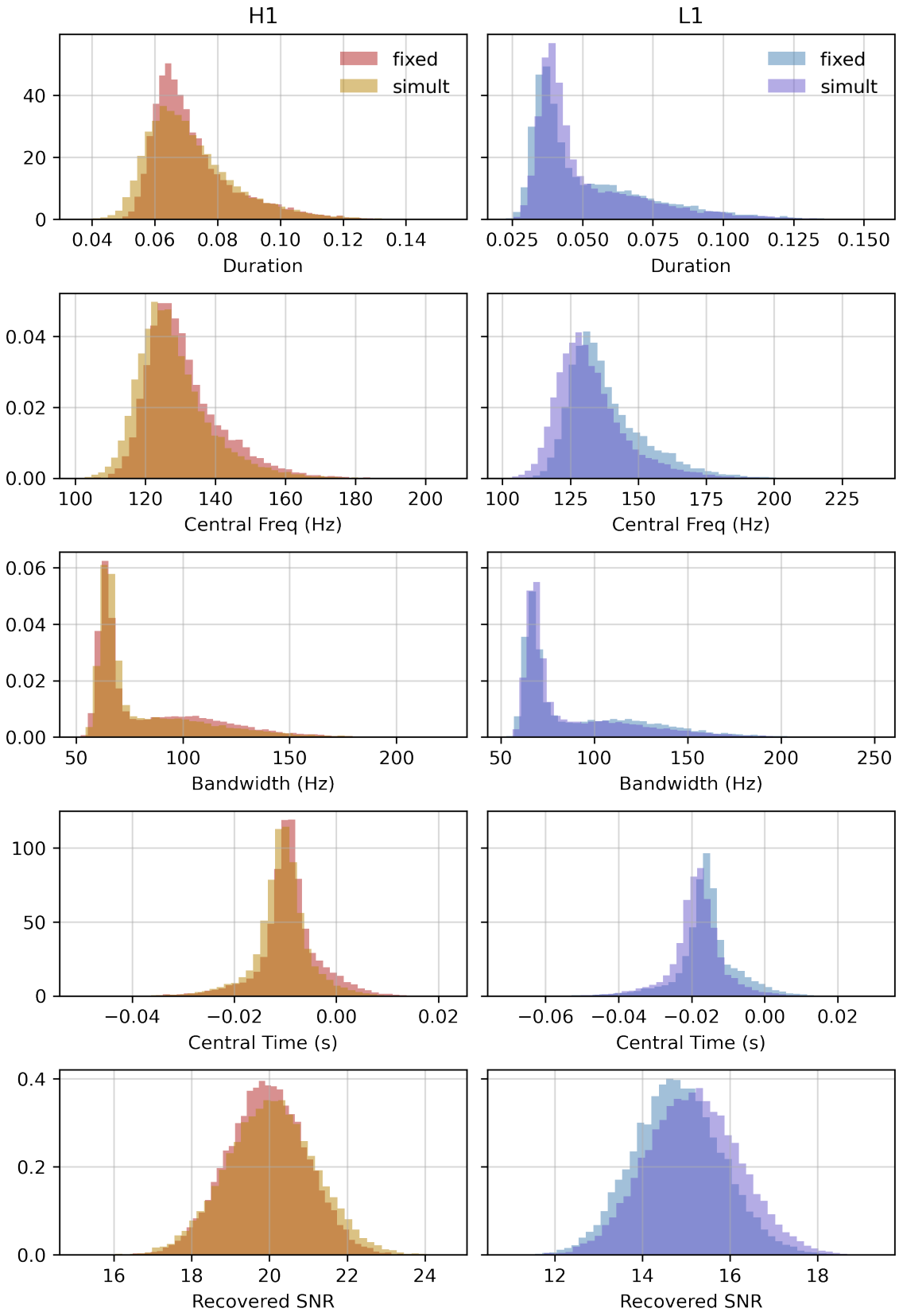

FIG. 8. Recovered waveform moments for each detector. In each subplot, the distributions for each analysis are overplotted.

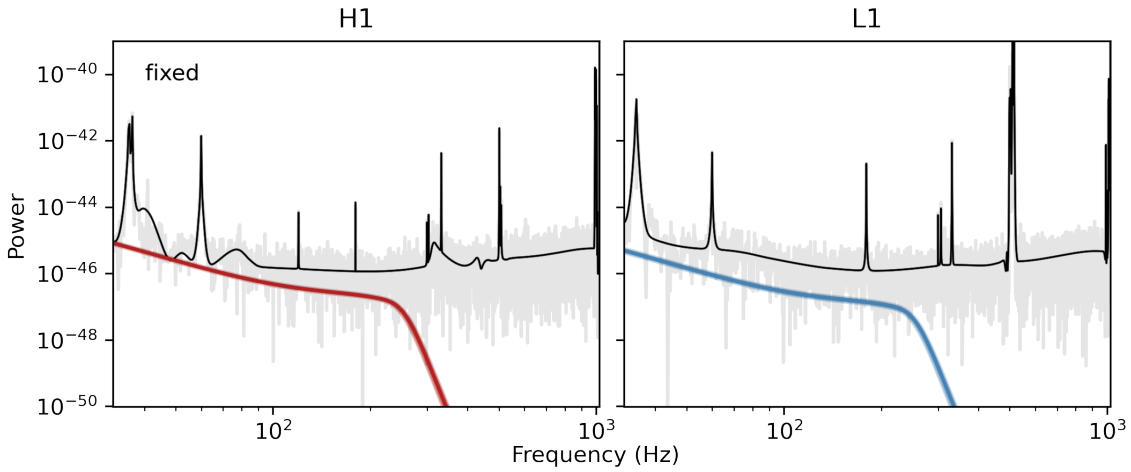

FIG. 9. Power spectrum for a CBC model run with a fixed PSD on GW150914.