LIGO SURF Interim Report 2: Marginalizing over noise properties in parameter estimation

Cailin Plunkett

Advisors: Katerina Chatziioannou, Sophie Hourihane
(Dated: July 30 2021)

The traditional gravitational wave (GW) 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 simultaneously estimate the noise and binary properties, which mitigates the assumption that the noise variance is known during the fitting process. We do so using both the wavelet- and template-based models available in `BayesWave`. Initial results on GW150914 and injected signals suggest the methods produce signal reconstructions and posterior parameter distributions that agree to within uncertainty. Future work intends to repeat the analyses on more injected signals and all real events in the second Gravitational-Wave Transient Catalog (GWTC-2) to identify whether the method used to handle the noise systematically impacts GW 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 (PE) 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 PE 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

$$\mathcal{L}(d|h') = -2\sum_{i}^{N/2} \frac{\tilde{r}_i \tilde{r}_i^*}{T S_n(f_i)} + \text{const}, \tag{1}$$

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 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 PE results demonstrated the importance of the chosen noise model and further confirmed the comparative strength of the on-source method.

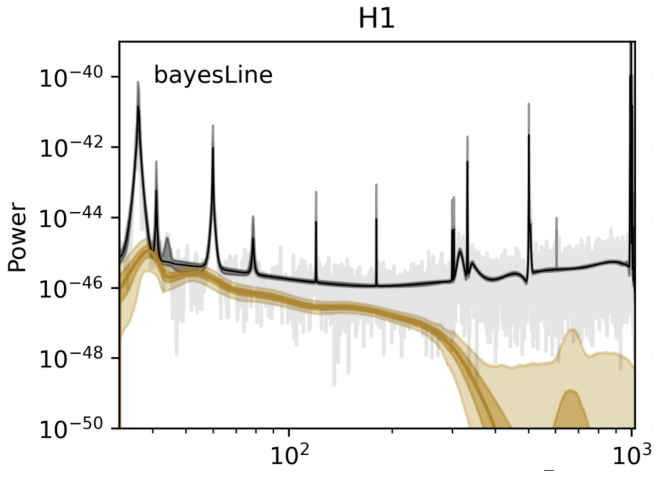

FIG. 1. Example power spectrum for GW150914. The recovered signal is in color, the noise PSD with uncertainty is in black, and the data are in light gray.

Fig. 1 depicts example power spectra with the on-source spectral estimation method. The data are in light gray, the recovered signal is in color, and the PSD with uncertainty is in black.

The differences between the off-source and on-source (BL) PSDs have shown the impact of the method used to compute the PSD. However, with either method, the standard process remains to feed the computed PSD median into LI to estimate binary parameters. A recent development in the capability of BW, described in detail in Chatziioannou *et al.* [6], uses BL to compute binary and noise parameters in concert. This mitigates the third assumption explained above by directly incorporating PSD uncertainty into the signal estimation. This method of marginalizing over uncertainty in PE has been applied to injected signals; the intent of this project is to now apply this method to real CBC events.

II. PROJECT OBJECTIVES

We are studying the effect of including uncertainty in noise on parameter estimation for the confirmed LIGO and Virgo CBC events.

We use BW to first model the noise $S_n(f)$ then use that result as an input to model the GW signal. We then compare that output to one obtained using the simultaneous estimation method, using `bayesLine` to estimate the PSD and signal in one step. Moreover, we complete each analysis on two models available in the code: the signal and CBC models.

The signal model uses sine-Gaussian wavelets to reconstruct the GW signals, rather than templates, like `Bilby`. This flexible approach enables it to find weakly modeled GW signals as well as well-modeled events, like CBCs. However, because this method does not use templates, it cannot return CBC parameters like masses and spins for CBC events, but wavelet parameters, such as amplitude and frequency. The CBC model is a new capability of BW introduced in Chatziioannou *et al.* [6] within the context of simultaneous glitch subtraction and parameter estimation. The CBC model is a template-based search, which, along with time-domain waveforms and power spectra, returns posteriors for GW parameters. Fig. 2, which is Figure 6 from Abbott *et al.* [7] is the type of plot we aim to produce using these posteriors and compare between the methods: it shows the 90% credible regions for all CBC candidate events in GWTC-2 in total mass M-mass ratio q space, with events published prior to that study highlighted.

To obtain a PSD to use as a fixed input to both the signal and CBC models, I first run the data through BW with the `cleanOnly` flag. The cleaning mode, coupled with BL, estimates the noise PSD with uncertainty. I then take the median PSD from that run as the input to signal and CBC runs. For the simultaneous method, I run each model with BL.

The four main analyses I am running on each GW signal are the signal model with a fixed PSD (signal+fixed), the signal model with BL (signal+BL), the CBC model with a fixed PSD (CBC+fixed), and the CBC model with BL (CBC+BL). Table I schematically shows the difference between these methods using the run setting names for the signal 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 am comparing.

III. PROGRESS

I have run many versions of these four analyses (five, if one includes the cleaning mode) on GW150914 with slightly varying run settings in order to understand the

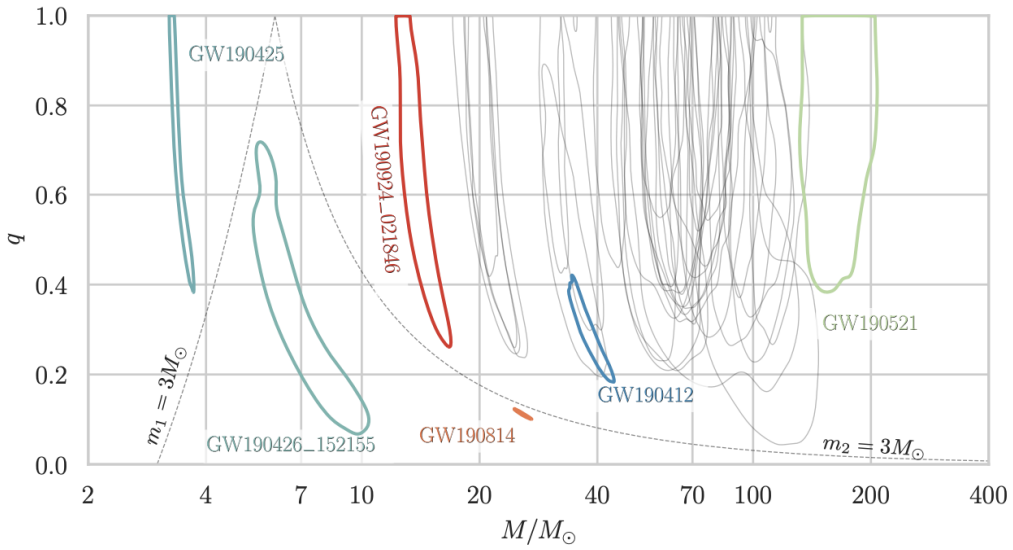

FIG. 2. 90% credible regions for all candidate events in total mass M and mass ratio q space, with previously published events highlighted. Dashed lines delineate where one of the objects can have a mass $< 3\,M_\odot$.

impact of each setting on the signal reconstructions and to troubleshoot inconsistent results. I have also completed several runs on GW150914-like (in terms of chirp mass and SNR) injections to further investigate issues found with the GW150914 runs. Note that all figures below, unless otherwise noted, depict data for the Hanford interferometer (H1). Limiting to one interferometer reduces the amount of content to be presented. Moreover, as the observed issues are more noticeable in Hanford than Livingston, we show those results.

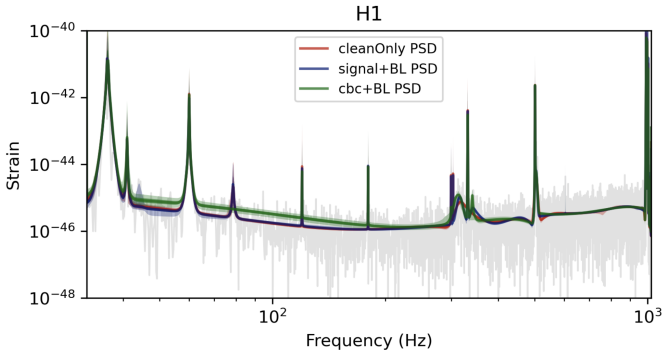

FIG. 3. PSDs from cleanOnly, signal+BL, and CBC+BL runs with 50 and 90% CIs. The residual data from the signal run is in gray.

A. PSD Problems

As Fig. 3 depicts, the CBC+BL PSD is much higher in the 40-200 Hz range—the region of most interest—than a PSD from a cleaning run or a signal+BL run. This remained true for every cleaning and signal+BL run we performed, indicating a difference in the CBC+BL PSD calculation. This is a problem: while we expect slightly different signal reconstructions, perhaps with more uncertainty, the PSD should be roughly the same with each method: it certainly should not be far outside the 90% CI.

In an attempt to address this problem, we took a multi-step approach. First, we started many `cleanOnly` runs with both versions of the code (the master branch and CBC branch) using different random seeds for the MCMC sampling. As Fig. 4 shows, there are regions in which several of the runs did not agree to within uncertainty, particularly between 60 and 100 Hz. The runs seemed to cluster near one of two maxima in likelihood. While this may seem to suggest a convergence issue, with four million iterations, that should not be a problem. Additionally, setting the random seed did not make the run reproducible, which seems to be due to slight rounding differences that propagate.

The next item to check was the cleaning phases of the CBC runs: fixed and BL runs complete a cleaning phase to start, the PSD from which BL uses as a starting point (fixed PSD runs benefit from the subtracted glitches but do not use the cleaning PSD). This revealed another problem. The CBC+BL cleaning PSDs were drastically different than all of the others. Many plots later, we discovered the root: the flag `waveletFmin=150` was set in the CBC+BL run. As the CBC model does not use wavelets, this affected the cleaning phase but should not have affected the main run. Converged runs should give the same result regardless of the cleaning phase, but in order to ensure observed differences are due to the method used for the PSD and not the parameters of the cleaning

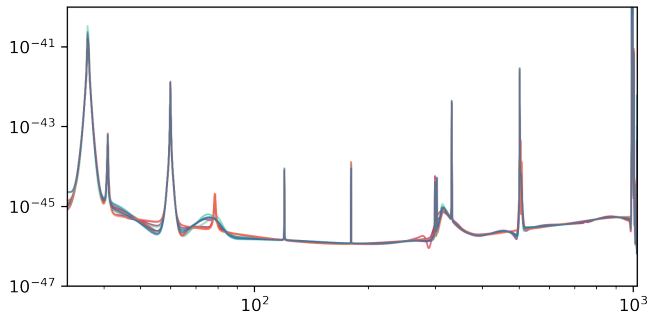

FIG. 4. Median PSDs from cleaning runs with various starting seeds on CBC branch (cool colors) and master branch (warm colors). There is disagreement particularly between 60 and 90 Hz. There is some apparent branch dependence, with the master branch tending to include a Lorentzian and the CBC branch tending to omit it.

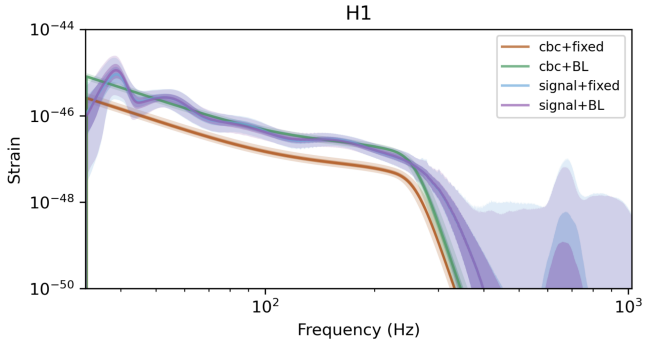

FIG. 5. Spectra for signal+fixed, signal+BL, CBC+fixed, and CBC+BL runs. The CBC+fixed run is much lower in power than the other three, which agree. The signal runs cut off at a higher frequency than the CBC+BL run.

phase, this error required restarting all runs from scratch.

B. PSD Uncertainty

The overarching realm of my project is that of understanding how different noise models impact parameter estimation. The primary way in which we are probing this question is BW's own PE method with BL, which marginalizes over uncertainty in the noise PSD. However, the primary use of BW in the production LIGO parameter estimation pipeline is to subtract glitches and model the PSD that is given to `Bilby`. Since the cleaning mode that completes this process outputs the noise PSD with uncertainty, and every posterior sample is equally valid, another way of exploring the impact of noise models is to compare PE results when we use the median PSD—the standard—as opposed to other posterior samples. That is, we ask the extent to which using different PSDs, equally valid under the data, impacts PE.

To do so, we took the upper and lower 90% confidence interval bounds from our cleaning run, and completed the fixed analyses for these PSDs as well as the median.

C. Heterodyning

An interesting issue present in the CBC model runs is that the signal found by CBC+fixed has much lower power than CBC+BL. Shown in Fig. 5, the power from the signal model runs agrees with that of the CBC+BL run, suggesting it is the CBC+fixed run that is anomalous. As Fig. 6, the residual spectrogram for the CBC+fixed run, reveals, the run did not successfully subtract all of the signal power: there is a clear remaining signal in the residual. In searching for the cause, we started CBC+BL and CBC+fixed runs without heterodyning, a method used to speed up the computation process.

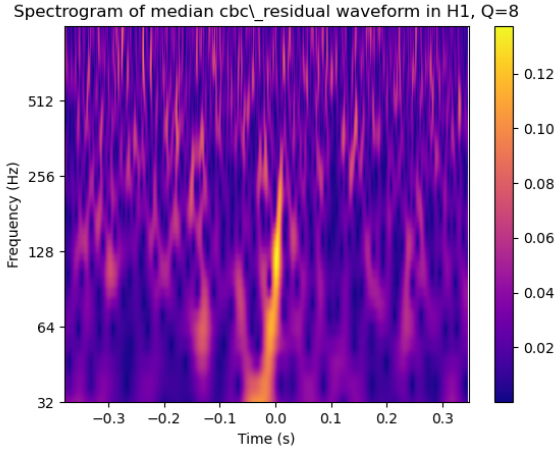

FIG. 6. Spectrogram of residual power from a CBC+fixed run in H1, with $Q = 8$. The remaining visible chirp shows the signal is not completely subtracted.

Heterodyning (HD) is a method of speeding up the likelihood calculations in GW Bayesian inference. For a full discussion of heterodyned likelihoods for GW Bayesian inference, the reader is directed to Cornish [9]. For the purposes of this report, it is sufficient to state that heterodyning is a shortcut approach to compute likelihoods that avoids calculating full waveforms, which greatly speeds up the computation. While the approximations are sufficient in most cases, this method can introduce errors.

As described above, the result of CBC+fixed is much lower in power than CBC+BL when we use heterodyning. However, without heterodyning, while the CBC+BL run is unchanged, the CBC+fixed run jumps up to the same power level, shown in Fig. 7. The residual spectrogram (not shown) also suggests a much better subtraction of the signal.

Moreover, when the low and high 90% CI PSDs from the cleaning run were used for the CBC+fixed runs, the same trend emerged. Fig. 8 show the six CBC+fixed

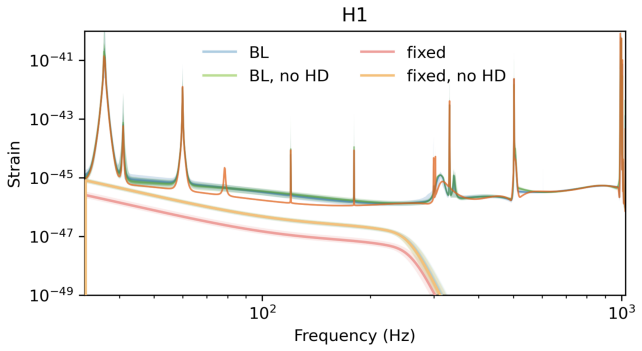

FIG. 7. Spectra of CBC+BL and CBC+fixed runs with and without heterodyning. The CBC+fixed run without heterodyning matches the power of the CBC+BL runs.

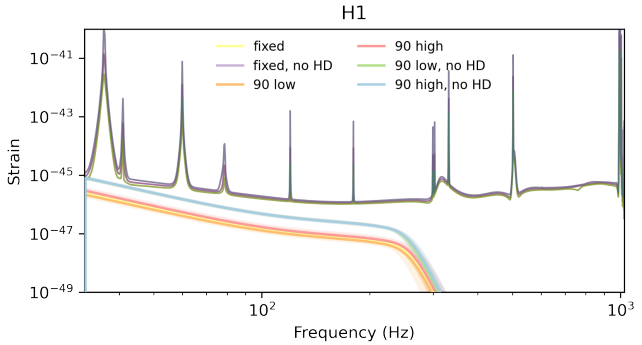

FIG. 8. Spectra of CBC+fixed runs with and without heterodyning using the median and 90% low and high CI PSDs from a cleaning run. The runs with heterodyning are all lower in power than those without heterodyning; those without heterodyning agree with each other and the CBC+BL runs.

runs with different PSDs with and without heterodyning. The runs without HD all agree in power, but those with HD are lower in power and scale in order of PSD height. We are unclear why heterodyning on the CBC+fixed run resulted in a posterior with low signal power and are investigating.

D. Parameter Estimation Results

The CBC model runs output CBC parameter posteriors as well as waveform models. Our eventual goal is to compare the PE results between different models across all of GWTC-2. With our runs thus far, we can compare between the models for GW150914.

Figs. 9 and 10 show posterior distributions for the primary mass $M_1(M_\odot)$ in the detector frame and the event distance in Mpc for real-data runs on GW150914. Observe that the CBC+fixed+HD model has a lower median M_1 and higher distance, which matches the expectation for a lower powered signal. While the medians are not identical, the histogrammed distributions show significant overlap. To understand whether there is systematic differences in the parameter estimation for these methods, we will need to complete these same analysis on a series of injections and real events.

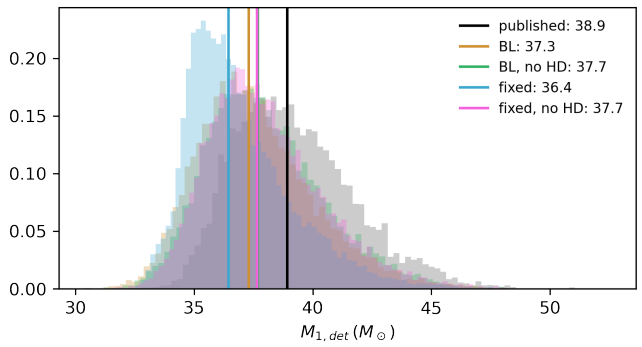

FIG. 9. Posterior distributions for the primary mass, $M_1(M_\odot)$, in the detector frame for several models with medians overplotted. The published result is in black. While the medians differ, all distributions overlap within their 90% CIs.

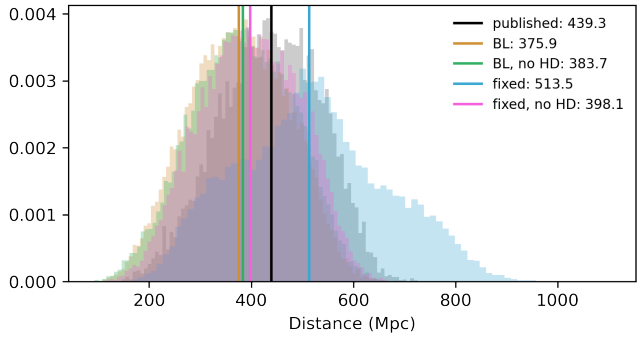

FIG. 10. Posterior distributions for the distance in Mpc for several models with medians overplotted. The published result is in black.

E. Injections

In searching for the cause of the higher CBC+BL PSD and low power CBC+fixed signal, we performed injections. We took the `LALSimAdLIGO` PSDs—the design sensitivity of the detectors—and created Gaussian noise with them. We then injected a GW150914-like signal into the data. The signal was not completely similar to GW150914; the distance, in particular, was much farther, but due to a lower noise threshold, the network SNR was in a similar range (25-30).

For three different trigger times, we performed `cleanOnly` runs with master and CBC, signal+BL on both branches and CBC+BL on the CBC branch. We then repeated all of those with a fixed PSD: the injected PSD itself. Additionally, we repeated both CBC runs without heterodyning.

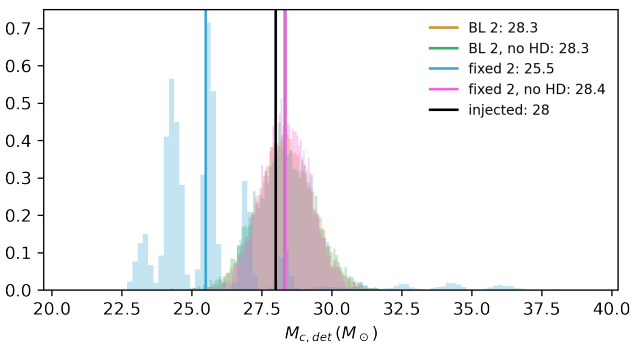

FIG. 11. Posterior distributions for the detector frame chirp mass, $M_c(M_\odot)$ on an injection for several models with medians overplotted. The injected value is in black.

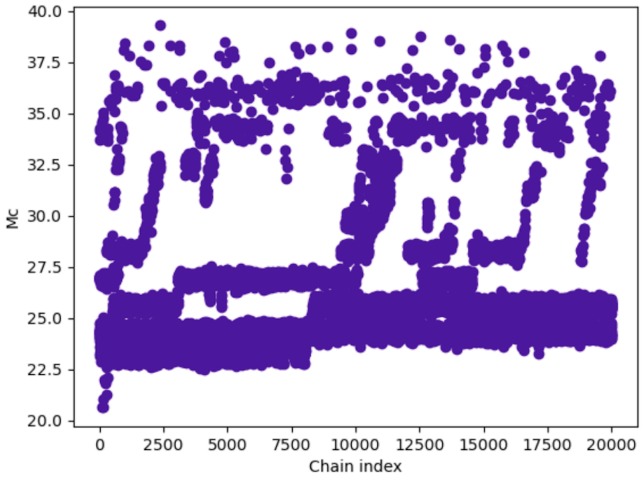

FIG. 12. Posterior chirp mass for each chain index for a CBC+fixed+HD run. The injected M_c was 28 $M_\odot$.

We observed similar phenomena with the injections as with the real data runs. First, the CBC+BL PSDs were higher in power than the actual PSD, cleaning PSDs, and signal+BL PSDs, particularly at low frequencies. One potential cause of this is the segment length. We used 4 seconds, which may not be long enough to get an accurate PSD estimation regardless of the number of iterations. We are repeating the runs with 8 and 16 second segment lengths, but those have not finished at the time of writing.

In addition, the CBC+fixed+HD run is lower in power than the other CBC runs. Moreover, the posterior distributions, particularly for chirp mass, show severe sampling issues. Fig. 11 shows the posterior distributions for the detector frame chirp mass. The fixed+HD posterior is multimodal and scarcely recovers the injected value. Fig. 12 plots the posterior value for each chain index, demonstrating problems with the MCMC sampling. We are currently exploring possible causes of this issue. The M_c posteriors for the CBC+fixed+HD real data runs also have some multimodality and sampling problems.

IV. PLANNED WORK

In order to move toward our goal of comparing PE results between the BL and fixed PSD methods, we plan to use BW runs with various settings on a series of injections to troubleshoot the issues described in detail above: the high CBC+BL PSDs and low CBC+fixed+HD signal reconstructions. First, we are repeating all the analyses on another injection file. Second, as mentioned, we are repeating the CBC+BL runs with 8 and 16s segment lengths. If those PSDs more in line with the cleaning and signal+BL PSDs, then we may be able to attribute that issue to insufficient segment length. We are drafting a plan to investigate the sampling issue with the CBC+fixed+HD runs.

[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), arXiv:1310.1528 [gr-qc].

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

[9] N. J. Cornish, arXiv e-prints , arXiv:1007.4820 (2010), arXiv:1007.4820 [gr-qc].