Mitigating the effects of instrumental artifacts on source localizations

Maggie C. Huber
University of Michigan, Ann Arbor

Mentor: Derek Davis
LIGO, California Institute of Technology, Pasadena

First Interim Report
LIGO Caltech SURF Program 2021

Instrumental artifacts which materialize as glitches in strain data can overlap with gravitational wave detections and significantly impair the accuracy of sky localizations of compact binary coalescence (CBC) signals. To mitigate the effect of glitches, we are developing a method that applies a reweighting formula to the signal-to-noise ratio (SNR) of a signal. From tests on 1500 simulated signals, we determined that reweighting the SNR timeseries is able to improve the accuracy over zeroing out bad data. When we repeated this process for raw data with a simulated glitch, the reweighting formula likewise improves upon removing the data alone. In this report we discuss our results and future goals for the development of our new method to handle instrumental artifacts.

I. INTRODUCTION

Detection of gravitational waves requires extreme sensitivity to changes in length on the order of 10^{-18} m [1]. The level of strain sensitivity renders LIGO detectors susceptible to noise transients (also called glitches), which are bursts of excess power in the detector. Often, what causes these glitches is difficult to determine. They can be the result of either external environmental or internal instrumental interactions that alter the actual strain. Glitches are more likely to overlap with gravitational wave (GW) events that occur for a longer period, such as binary neutron star (BNS) events. As detection of GW events from BNS mergers become increasingly frequent [2], we expect to see more instances of noise transients overlapping with GW signals as seen in the case of BNS merger GW170817 [3].

This is problematic for many reasons, especially because noise transients diminish the accuracy of rapid sky localization and by extension all parameter estimation. BNS mergers are events that produce electromagnetic (EM) radiation requiring rapid, accurate followup observations. In order to gain useful and accurate astrophysical information from a GW event, it is important that glitches are mitigated in a way that minimizes bias in localization measurements.

There are multiple approaches one can use to try and address this. For GW170817, the effects of the noise transient were mitigated by applying a window function to remove it. Additionally, the glitch waveform was reconstructed with an analytic model that could be subtracted from the data [3], as shown in Figure 1. This method was ad hoc in nature. Different approaches are necessary to find a generalized solution for all GW events.

Window functions such as the one used for GW170817 gradually remove bad data to avoid discontinuity. However, they can introduce excess power leakage from the spectral lines in the power spectral density (PSD) of the detector. An alternative to window functions is inpainting [4], where the effects of discontinuities are calculated and subtracted. The end result is a gate that only masks bad seconds of data and has no affect on the data surrounding the inpainted hole.

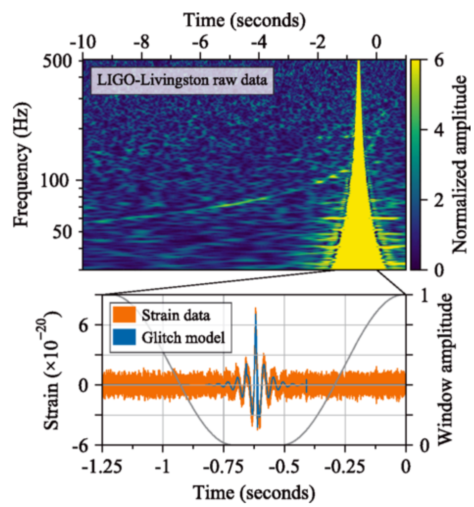

FIG. 1. *Top panel*: Time-frequency plot of LIGO Livingston data for GW170817 with the glitch present. *Bottom panel*: Strain data of the glitch, with a grey window function used to zero it out. The reproduced model of the glitch is shown with the blue curve. Replicated from [3].

When we inpaint a hole in GW data, we lose information from the signal and bias the sky localization. This effect is less noticeable when the fraction of data removed is less than $\lesssim 5\%$ of the total signal duration. For larger inpainting widths, this can add a noticeable bias to the sky localization. To ensure EM followup of events with data removed is as accurate as possible, it is necessary to correct for the effect of gating.

We used the BAYESTAR [5] algorithm in PyCBC to create our sky localizations. BAYESTAR is a rapid sky localization which uses Bayesian inference over Markov chain Monte Carlo (MCMC) methods. It takes a likelihood function and a well-defined parameter space to rapidly infer the location of GW signals on the sky.

To correct our BAYESTAR skymaps for inpainting bias, we developed an algorithm to reweight the signal-to-noise ratio (SNR) timeseries. Our goal is for recover the correct error and improve the accuracy of the localization.

In this report we will explain how our method works, how we tested it, our results, and what we are currently working to accomplish.

II. REWEIGHTING

The localization parameters of a GW signal are the time delay, phase and the SNR. The presence of an inpainted hole in the data will cause the SNR timeseries to deviate. We start by assuming we have a known signal template which is input into the algorithm. The SNR remaining after inpainting is given by [4]

$$\lambda_{hole}(t_0, h) \approx \frac{(|h_w|^2 \circledast \mathbb{1}_{valid})(t_0)}{\sum_t |h_w(t)|^2} \ , \tag{1}$$

where t_0 is the merger time, h_w is the whitened waveform, and $\mathbb{1}_{valid}$ returns zero for a data point in the inpainted hole and one otherwise. The equation convolves h_w with $\mathbb{1}_{valid}$, which we compute with a fast Fourier transform (FFT) as allowed by the convolution theorem.

After we inpaint and apply Equation 1, we multiply a normalization factor to the PSD and SNR timeseries. When a portion of a signal is inpainted, we effectively decrease the sensitivity of our measurement. Renormalizing the PSD corrects the error for our BAYESTAR localization. We are in progress of determining the optimal normalization factor to get the most accurate error measurent. For the results shown in this report, we used a factor corresponding to the maximum SNR value of the timeseries calculated in Equation 1.

There are multiple advantages of reweighting the SNR timeseries. The algorithm is independent of where the data is inpainted and the waveform template, taking them as inputs. Reweighting is also deterministic - the calculation is the same for any variation of the input parameters. It is also instantaneous to compute, typically taking less than a second. These benefits render this method conducive to rapid and accurate sky localization of GW events in real time, even in the presence of glitches.

III. METHODS

To assess the performance of our reweighting algorithm, we simulated 1500 compact binary coalescence (CBC) signals for testing. We used a gate width of 1024 ms starting 64 ms from coincident time, set both masses to 10 $M_\odot$, and used a distance range of 10-400 Mpc. We then filtered the signal template list to include what we would expect to detect by applying an SNR threshold of 10. We chose these signal parameters corresponding to what we expected to be the most biased by this method. If a test runs successfully, we verify that it is likely to work with most other cases.

To run our tests, we inject the simulated signals into the background PSD from the LIGO Hanford (H1) and Livingston (L1) detectors and get the raw SNR timeseries. We then get the SNR timeseries from using the inpainting function alone, then both inpainting and reweighting. We create an XML file which is put into BAYESTAR to localize all three cases. To see how the method performs, we obtain the credible region of the true location, total searched area, the area of the 90 percent credible region, and the overlap of inpainting and reweighting skymaps with the raw skymap.

After testing cases for raw data without a glitch, we wanted to see if we could create a glitch that biased the skymap and recover the source location by reweighting. We injected a sine-gaussian wavelet with a frequency of 80 Hz and strain of 2.5×10^{-21} m. We then created skymaps using the same method as the data without a glitch and obtained the same metrics from BAYESTAR to see if we corrected the glitch and inpainting bias.

IV. RESULTS

A. Raw data without a glitch

Various metrics from BAYESTAR allow us to determine how reweighting compares to inpainting alone. We primarily use probability-probability (P-P) plots showing the credible region of the true source location vs. the fraction of total simulated signals (Figure 2). Ideally, the distribution on a P-P plot is linear with a slope of one. Due to an internal factor in BAYESTAR to normalize the plot in GstLAL, the raw data without the glitch lies above the diagonal. This distribution above the diagonal is therefore ideal for our plots made using the PyCBC pipeline.

One implication from the plot we created is that inpainting a hole in the data will bias the skymap and report an incorrect error. When we reweight the SNR timeseries, the error is recovered and the skymap is more likely to return a localization that contains the source.

When we find the correct normalization to apply to the PSD, the reweighted P-P plot will line up with the raw data without a glitch and we will be accounting for the increase in error after inpainting.

To check if the skymap shows an accurate credibe region, we create a histogram of the total searched area in degrees (Figure 3). The searched area we refer to is the area of the credible region housing the true source

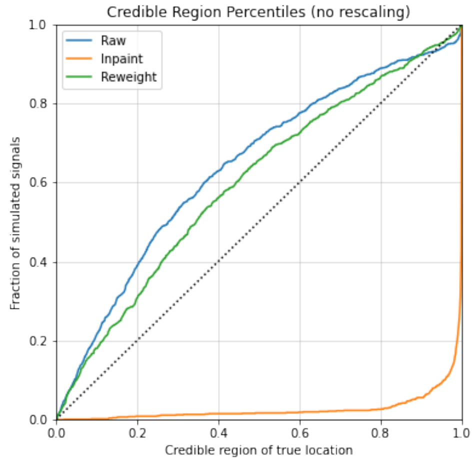

FIG. 2. P-P plot showing the credible region returned by BAYESTAR of the true source location. The raw data has no glitch. Due to the normalization factor in BAYESTAR meant for the GstLAL pipeline, we can see the raw and reweighted lines are overestimating the error. This causes the lines to inflate above the diagonal. From this plot we can determine that inpainting a hole in the data causes the error to be significantly underestimated, and reweighting brings the error much closer to accurate.

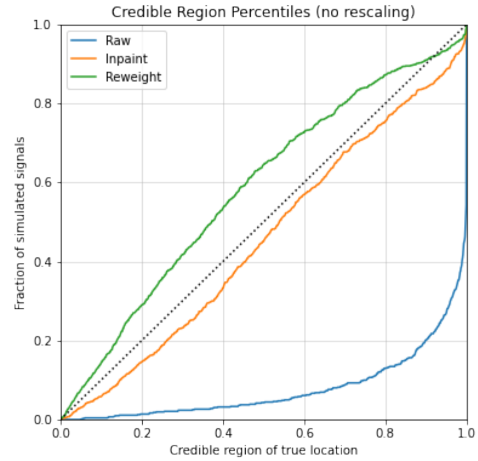

FIG. 4. P-P plot showing the credible region returned by BAYESTAR of the true source location. Raw data is with a glitch present. Due to the normalization factor in BAYESTAR meant for the GstLAL pipeline, we can see the raw and reweighted lines are overestimating the error. This causes the reweighted line to inflate above the diagonal. From this plot we can determine that a glitch in the data underestimates the error, which we are able to improve with inpainting and even more by reweighting the signal-to-noise ratio (SNR).

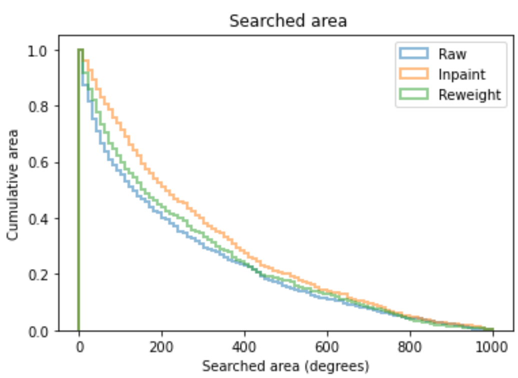

FIG. 3. Histogram showing the total searched area by BAYESTAR in degrees vs. the cumulative sum of signals. The raw data in blue shows the ideal distribution. Inpainting a hole causes the searched area to deviate and reweighting recovers the data, lying closer to the original.

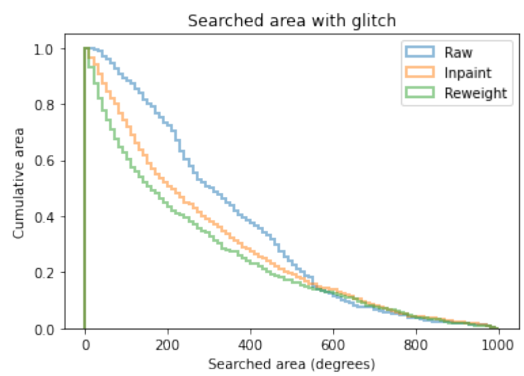

FIG. 5. Histogram showing the total searched area by BAYESTAR in degrees vs. the cumulative sum of signals. The raw data in blue shows the glitch impairs skymap accuracy. Inpainting a hole brings the searched area closer to the ideal distribution and and reweighting does slightly better.

location. Ideally, the cumulative area drops off faster as the searched area increases. This demonstrates that the resulting skymap predicts the source location to be in the lower credible region. Similar to the results of the P-P plots, the searched area histograms show reweighting the data gets the distribution closer to the original data.

B. Raw data with a glitch

For the data with a simulated glitch close to the time of merger, we created the same figures to determine if reweighting recovers a more accurate skymap than inpainting alone. For the P-P plot in Figure 4, the glitch biases the error estimate in BAYESTAR. Inpainting corrects for some of the error, and reweighting gets a more accurate estimate than inpainting.

The searched area plot in Figure 5 displays a similar behavior. We see that a glitch biases the accuracy of the skymap in the raw data, and it is recovered best by reweighting.

V. CONCLUSION

For upcoming LIGO observing runs, it is imperative that we have a way to to mitigate instrumental artifacts in the detector instantaneously. Quick and reliable sky localization of graviational wave signals allows us to expand the field of multi-messenger astrophysics.

We demonstrated that glitches and removing segments of a GW signal are sources of bias in sky localization and parameter estimation of a source. From our results we can see that reweighting the SNR timeseries was able to correct for this bias in both cases to return a localization that is more accurate than inpainting alone.

The next steps of the project will be repackaging and optimizing the code for our algotrithm. Ultimately, it should run faster and be less susceptible to random error. We will also be working with the normalization factor of the PSD to line up the reweighted error estimate with the raw estimate.

Overall, we are making progress towards developing a fast and reliable new method to mitigate the effect of glitches after we zero them from the data. We anticipate this method will be available for the next observing run and will be applicable to real GW strain data.

[1] B. P. Abbott et al. GW150914: The Advanced LIGO Detectors in the Era of First Discoveries. *Physical Review Letters*, 116(13):131103, April 2016.

[2] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 21(1):3, April 2018.

[3] B. P. Abbott et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[4] Barak Zackay, Tejaswi Venumadhav, Javier Roulet, Liang Dai, and Matias Zaldarriaga. Detecting Gravitational Waves in Data with Non-Gaussian Noise. *arXiv e-prints*, page arXiv:1908.05644, August 2019.

[5] Leo P. Singer and Larry R. Price. Rapid bayesian position reconstruction for gravitational-wave transients. *Physical Review D*, 93(2), Jan 2016.