Application of gating to stochastic searches in O3

Andrew Matas,[1,*] Irina Dvorkin,[2,†] Alba Romero,[3,‡] and Tania Regimbau[4,§]

[1] *Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476 Potsdam, Germany*

[2] *Institut dAstrophysique de Paris, Sorbonne Université & CNRS, UMR 7095, 98 bis bd Arago, F-75014 Paris, France*

[3] *Institut de Física dAltes Energies (IFAE), Barcelona Institute of Science and Technology, E-08193 Barcelona, Spain*

[4] *LAPP,CNRS, 9 Chemin de Bellevue, 74941 Annecy-le-Vieux, France*

(Dated: January 18, 2021)

Due to a high rate of loud glitches in the LIGO detectors in O3, the stochastic and continuous-wave searches pre-processed the strain data by applying an inverse Tukey window to times where the whitened strain RMS was large in the 25-50 Hz or 70-110 Hz frequency bands. In this short technical document, we detail the results of several tests that show this procedure does not introduce any bias into the stochastic search. First, we present the results of a mock data challenge to demonstrate the analysis of gated data are consistent with the analysis of data without gating, running on simulated data with and without a signal present. We then consider the effect on window factors that appear in the analysis. Finally, we show that spectral artifacts introduced by the analysis are minimal, and that compact binary signals are not systematically removed by the gating procedure applied. Collectively, these results give us confidence that it is safe to apply the stochastic search to gated data in O3.

I. INTRODUCTION

A large rate of very loud glitches is present in O3. If left unchecked, these glitches have a pernicious effect on long-duration stochastic searches. One large glitch can corrupt the estimate of the power spectral density (PSD) and cross spectral density (CSD) estimation from an entire 192 s segment. We show an example glitch in Figure 1 in the time domain, as well a PSD for a segment containing the glitch. As in the main text, the PSD is estimated using a Welch estimator of 50% Hann-windowed segments of duration 32s. The standard non-stationarity cut, which measures the difference in σ between adjacent segments, resulted in losing $> 50\%$ of the data. Furthermore, inverse noise weighting is not sufficient to remove the effect of these glitches, even if we are willing to essentially remove a large fraction of data. The reason is that because adjacent segments are used to estimate the PSD, a glitch in a given segment can cause the CSD to be biased, while still having a normal-looking PSD. This leads to segments with very large signal-to-noise ratio (SNR).

Therefore in the O3 stochastic analysis [1], we applied the analysis to a data set which was preconditioned by applying an inverse Tukey window to stretches of data when the RMS motion of the whitened strain channel exceeds a threshold in the 25-50 Hz or 70-110 Hz frequency bands; the precise definition of the thresholds and channels used is given in [2]. Tukey windows have been regularly applied to GW data analysis at least since the work of [3], and have been found to avoid artifacts in PSD estimation due to spectral leakage. The construction of this data set, and applications to continuous wave (CW) searches are described in [2]. In Fig. 1, we also show the effect of this solution. The time series with the glitch removed, and the resulting PSD is shown in the right hand panel. Because the glitch is a relatively small fraction of a total 192 s segment, we don't expect the lost time to have a significant impact on the search, and additionally we expect to gain back a large fraction of the time lost to a naive application of the delta sigma cut.

Nevertheless, one may have a few concerns about applying this procedure to strain data before a cross-correlation stochastic search. First, one should prove whether the effect of analyzing data with zeros really is small. Second, one may worry about new lines being introduced. Finally, one can worry if individual gravitational-wave events are removed by the gating procedure, effectively reducing the stochastic background.

In this work we present the results of a software injection study to test the performance of the stochastic search with gating present.

II. INJECTION TEST

We produced a set of Gaussian noise frames, spanning the interval Apr 01 2019 15:00:00 UTC - Jul 29 2019 03:26:40 UTC, colored using a smoothed version of a representative O3 PSD. Since we are testing the effect of gating on the search, and not the effectiveness of a gating algorithm to remove glitches, we do not inject any non-Gaussian features into the noise. We conditioned the data in the same way as in the main text. To save computational cost, we downsampled the frames to 1024 Hz and used a maximum frequency of 400 Hz. Since the isotropic search for an $\alpha = 0$ power law background gets most of its sensitivity from below 100 Hz, this change does not result in a substantial effect on the results pre-

* andrew.matas@aei.mpg.de
† dvorkin@iap.fr
‡ aromero@ifae.es
§ tania.regimbau@lapp.in2p3.fr

Run	Gating applied?	Injection	$Y/10^{-8}$	$\sigma/10^{-8}$	SNR
1	No	No	0.301	1.05	0.29
2	Yes	No	0.296	1.04	0.29
3	No	SNR=3	2.26	1.05	2.22
4	Yes	SNR=3	2.33	1.04	2.24

TABLE I. Results from MDC described in the main text. We ran the cross-correlation search over data with only Gaussian noise present, and with noise plus an SNR=3 stochastic background. We ran the analysis with and without the gates from O3 applied. We find that the results with gating and without gating are consistent to well within 1-sigma, in other words the effect of the gating is consistent with a different noise realization.

sented here. We do not expect the conclusions to depend on the power law used for the injection.

We ran the stochastic pipeline on these data, in two passes.

- First, we applied the cross correlation analysis directly to the data.
- Second, we applied the actual set of O3 gates to the data [1], and ran the search.

Since the underlying Gaussian noise realization is the same, we can directly compare the results with and without gating. This comparison is not possible with real data, because it is not possible to perfectly subtract the glitches from the original data stream.

In the first two rows of Table I, we show a comparison between the two analyses of the data. In the top two panels of Fig. 2, we show the comparison of 74000 individual segments, with gating and without gating applied. We see there is no net shift to larger or negative values in the point estimate distribution.

We then injected an SNR=3 background with a flat spectrum $\alpha = 0$ and repeated this experiment. The results are present in the bottom two panels of Fig. 2 and the bottom two rows of Tab. I. In both cases we see that the signal is recovered within 1-sigma uncertainty. The difference between the gating and non-gating runs is smaller than the variation expected from a different noise realization.

We see in both cases that when gating is applied, point estimate shifts well within $1-\sigma$, and the error bar shifts by about 1%. This shows that there is only a small effect on the search due to gating.

III. WINDOW FACTORS

The standard search applies a window factor to account for the effect of using 50% overlapping Hann windows. Applying gating effectively changes the windows used for cross correlation and power spectral density estimation. In the stochastic search, there are three relevant window factors that naturally appear

$$\overline{w_1 w_2} = \frac{1}{N}\sum_{j=1}^{N} w_1(j) w_2(j) = 0.375 \tag{1}$$

$$\overline{(w_1 w_2)^2} = \frac{1}{N}\sum_{j=1}^{N} w_1(j)^2 w_2(j)^2 = 0.2734 \tag{2}$$

$$\overline{(w_1 w_2)^2}_{\rm ovl} = \frac{1}{N}\sum_{j=1}^{N/2} w_1(j)^2 w_2(j+N/2)^2 = 0.0234 \tag{3}$$

where N is the number of time samples in each window. For $192s$ segments and $1/32$ Hz frequency resolution and a sampling rate of 1024 Hz, this yields $N = 32768$. The right hand side of the above equations shows the values for 50% overlapping Hann (sine) windows with this value of N. The Hann window is given by $w(j) = \sin\left(\frac{2\pi j}{N}\right)$ for $0 \le j \le N$, and 0 otherwise.

We show the distribution of $\overline{w_1 w_2}$ in the data analyzed in Fig. 3. The average is about 2% different from the value we would obtain if no gating were applied. We find a similar result for $\overline{(w_1 w_2)^2}$ and $\overline{(w_1 w_2)^2}_{\rm ovl}$. Therefore we conclude there is a minimal impact on the analysis.

A more complete treatment would analyze each segment with a different window factor. However, this quickly becomes quite complex. As an example, consider the standard welch PSD estimate

$$P_1(f) = \frac{1}{\overline{w_1^2}}\frac{1}{N_{\rm seg}}\sum_{i=1}^{N_{\rm seg}} P_j(f) \tag{4}$$

where $N_{\rm seg} = 2(2T\Delta f - 1)$ is the number of segments used to estimate the PSD. This should be replaced with a weighted average

$$P(f) = \frac{1}{N_{\rm seg}}\sum_{i=1}^{N_{\rm seg}} P_j(f)\frac{1}{w_{1,j}(f)^2} \tag{5}$$

Given that these corrections are quite small, we do not attempt to implement this procedure in the stochastic search. However if large glitches continue to present a problem in future observing runs, it may be interesting to develop a fully self-consistent method for handling the window factors.

IV. COMPACT BINARY SIGNALS

We found that none of the events detected in GWTC-2 [4] were gated. This is not surprising, since the thresholds for the gates are very large compared to what is expected for a CBC signal. Since the majority of CBCs that will comprise the stochastic background are undetected, we do not expect that transient signals comprising

[1] We only included gates when the detectors were in science mode; this removes several gates when the observation intent bit was set to on while the detector was not locked.

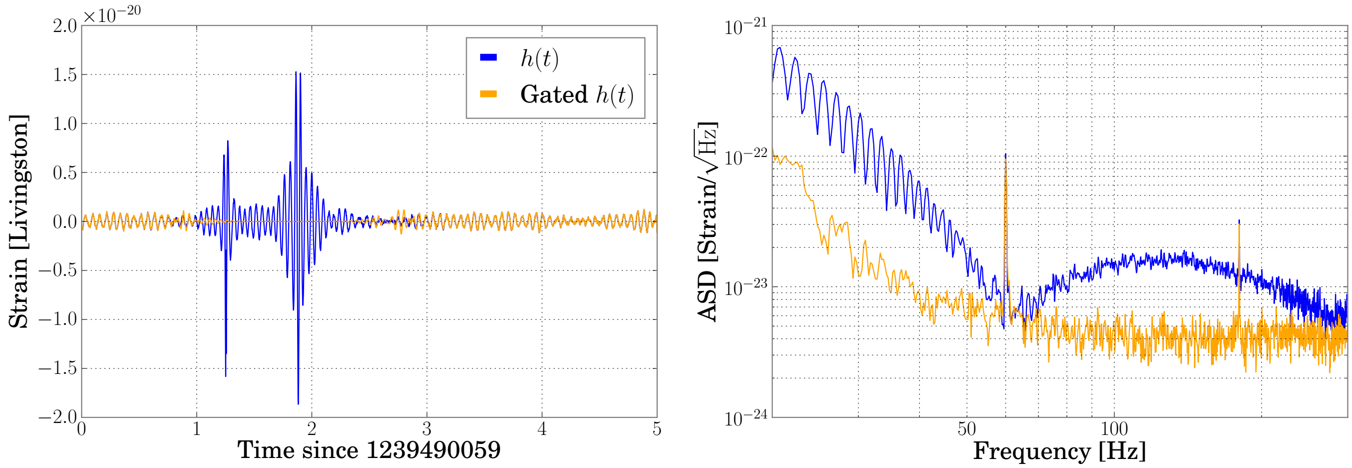

FIG. 1. Example of gating. Left panel: time series of 5s of Livingston data with and without gating applied, starting at April 16, 2019 22:47:21 UTC. A band-pass filter with knee frequencies 20 and 300 Hz has been applied to both time series. Right panel: the ASD for a 16s segment containing the data on the left panel, with and without gating applied. We see that even though the glitch is short compared to the duration of the segment, it leads to a large, broadband increase in the PSD. In the stochastic search, due to the use of adjacent segments to estimate the PSD, the standard non-stationarity cut used in past stochastic searches would remove up to three segments due to this single glitch. Therefore applying gating allows more data to be analyzed, while only removing a small amount of data.

the stochastic background will be affected by the gating procedure. Because of the lack of correlation between the times which are gated and the times containing a CBC signal, the stochastic search for a CBC background should then be unaffected.

V. CONCLUSION

We conclude that a significant bias is not introduced by gating in the stochastic searches. We also do not see evidence that the gated data removes GW signals. We therefore conclude that it is safe to run the stochastic searches on gated data. This procedure improves the stochastic search in O3; without gating, more than 50% of the data is removed by the non-stationary cut, whereas when gating is applied 17.9% is removed, which is comparable to the livetime lost in previous observing runs [5, 6].

ACKNOWLEDGMENTS

The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. Numerical computations were performed with the `numpy` [7] library, and plots with the `matplotlib` [8] library.

[1] `https://dcc.ligo.org/P2000314`.
[2] `https://dcc.ligo.org/T2000384/`.
[3] S. A. Usman *et al.*, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
[4] R. Abbott *et al.* (LIGO Scientific, Virgo), (2020), arXiv:2010.14527 [gr-qc].
[5] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **118**, 121101 (2017).
[6] B. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. D **100**, 061101 (2019), arXiv:1903.02886 [gr-qc].
[7] T. Oliphant, *Guide to NumPy* (2006).
[8] J. D. Hunter, Computing in Science & Engineering **9**, 90 (2007).

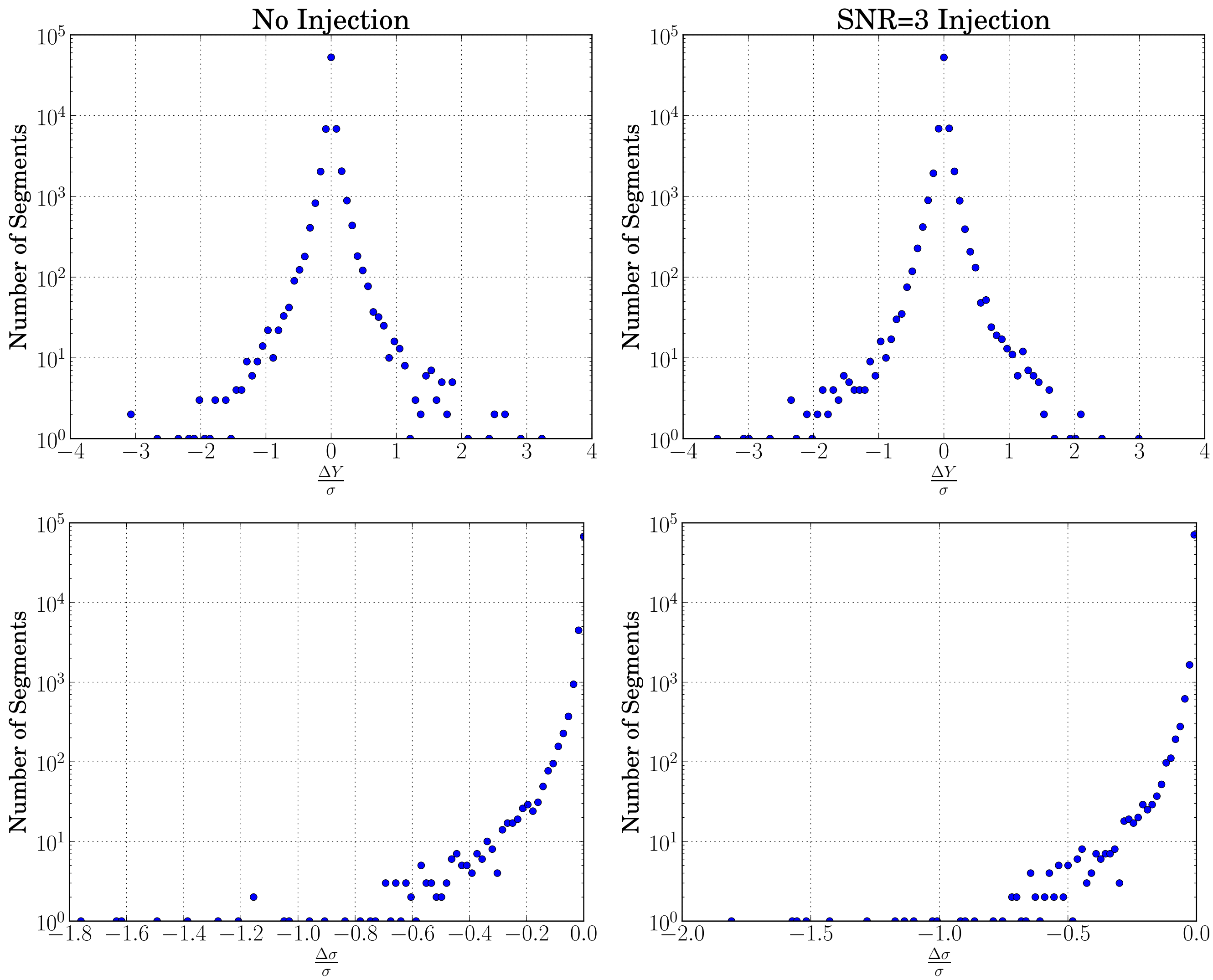

FIG. 2. Distribution of the difference between the point estimates (top panels) and error bars (bottom panels) obtained with gating and without gating for individual 192s segments, normalized by the error bar in each segment. In the left two panels we present the results for a Gaussian noise realization, while in the right two panels include an injection with SNR = 3. We see there is no systematic bias to larger or smaller values of the point estimates. The shifts are well within 1-σ.

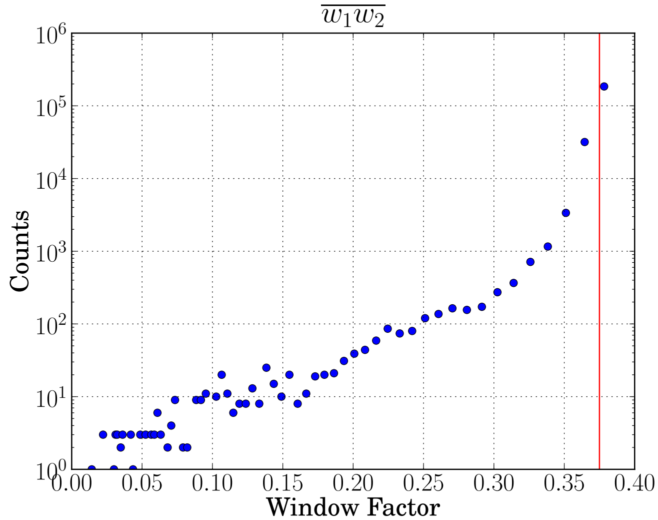

FIG. 3. Window factor comparison. We show the distribution of window factors $\overline{w_1 w_2}$ defined in the main text for the set of gates in O3 that we apply in the MDC. We see that the distribution is very sharply peaked around the value for Hann windows with no gating applied, although there is a tail that goes to smaller window factors due to segments with longer gates. The distributions for $\overline{(w_1 w_2)^2}$ and $\overline{(w_1 w_2)^2}_{\rm ovl}$ are qualitatively similar.