LIGO-T2000384-v4

Information on self-gating of $h(t)$ used in O3 continuous-wave searches

John Zweizig (LIGO Laboratory), Keith Riles (University of Michigan)

Version 4 – January 3, 2021

I. INTRODUCTION

The LIGO strain data known as $h(t)$ from both the Hanford (H1) and Livingston (L1) interferometers was subject to relatively frequent and loud low-frequency glitches during the O3 observation run. The origin of those most of those glitches remains unknown as of this writing. Unlike glitches seen in previous LIGO data runs, these artifacts were loud enough to affect quite appreciably the PSD estimations for discrete time intervals as long as those used in creating "short Fourier transforms" (SFTs) for many continuous-wave (CW) searches, ranging from a half hour to two hours.

To avoid the degradation in CW sensitivity potentially arising from these noise artifacts, we implemented an *ad hoc* but effective "self-gating" algorithm to create new H1 and L1 $h(t)$ data streams suitable for CW analyses. The self-gated $h(t)$ streams were made available for direct use in time-domain searches. In addition, Tukey-windowed and Hann-windowed SFTs were generated for use in Fourier-domain searches.

This document describes the procedures and parameter-tuning followed in generating the $h(t)$ streams and SFTs. It also discuses deadtime considerations which led to the recommended exclusion of certain SFTs from analysis and a particular self-gated artifact that led to another recommended exclusion period. Auxiliary files provide lists of observing-mode segments (integer GPS start & stop times), lists of gates (sub-integer GPS start times and durations), lists of all SFT start / stop times for 1800-second and 7200-second SFTs and lists of SFTs we recommend be excluded from ordinary CW analyses. Note that the starts of the application of gates generated in the frames used by CW analyses preceded the nominal start times by one time sample (1/16384 s = 61.0 μs).

II. THE SELF-GATING PROCEDURE

Instead of excising the glitches precisely by trying to determine their shape and amplitude, then carrying out a subtraction, we chose a simpler gating procedure in which the short time intervals containing the glitches were zeroed out, preceded and followed by 0.25-second transitions. Hence the windows were equivalent to multiplying the original $h(t)$ stream by inverse Tukey windows containing identified glitches, where the windows begin at unity, drop smoothly to zero prior to each 1/16-second interval containing a triggering glitch and rise back to unity afterward. Figure 1 illustrates the window application schematically using simulated Gaussian noise with a brief superposed loud sine-Gaussian glitch. Figure 2 shows an actual loud (and long) L1 glitch example from early in the O3a run (O3a = April 1, 2019 to October 1, 2019; O3b = November 1, 2019 to March 27, 2020 in the time domain. Figure 3 shows a Fourier-domain representation of the same glitch.

While the gating was carried out directly on calibrated $h(t)$ data, the trigger for defining when a gate was needed was based on the online whitened channel [L1,H1]:GDS-DELTAL_EXTERNAL_DQ, which is a partly whitened version of $h(t)$ appropriate for use in the interferometer differental arm servo loops. A band-limited RMS was computed for two distinct bands of H1 and L1 data: 25-50 Hz and 70-110 Hz. Trigger thresholds are shown in Table I. Arriving at these threshold choices was an iterative and *ad hoc* process, based on subjective tradeoffs between livetime loss and spectral distortion. The tuning of the choices was especially difficult for H1 because of non-stationarity in the early months of O3. In addition, the low-band trigger was not applied during the first ∼two weeks of the H1 data, prior to the moving calibration lines to below 20 Hz. Figures 4-5 show examples of RMS strain histograms computed over 1/16-second intervals for a sample 1-day interval late in O3a: GPS [1253343600-1253430000].

For the initial O3a gating carried out in winter 2020, the gating procedure was applied in 200,000-second chunks with output strain written to 16-second frame files. For the fall 2020 gating of O3a and O3b data that included an output chanel based on the 60-Hz-subtracted $h(t)$ stream, the output strains were written to 4096-second frame files for better I/O efficiency in downstream analysis, using 102,400-second chunks. Histograms and time series of band-limited rms strain were produced as part of the processing, along with log files of gate GPS start times and durations with 1/16-second discreteness. Tables II-III summarize the locations of the input and output files used in the gating processes. The strain input channel names to which gating was applied are [L1,H1]:DCS-CALIB_STRAIN_C01 and [L1,H1]:DCS-CALIB_STRAIN_CLEAN_SUB60HZ_C01; the strain output channel names are [L1,H1]:DCS-GATED_STRAIN_C01

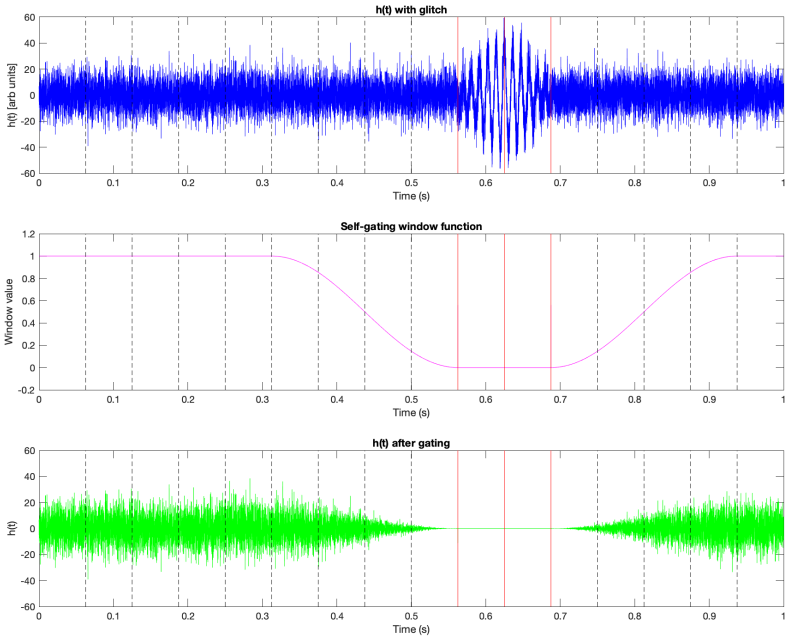

FIG. 1. Illustration of the inverse Tukey window multiplication. There are 0.25-second transitions before and after each contiguous 1/16-second segments with triggering clitches. The schematic diagram shows an example with two 1/16-second segments containing such a glitch.

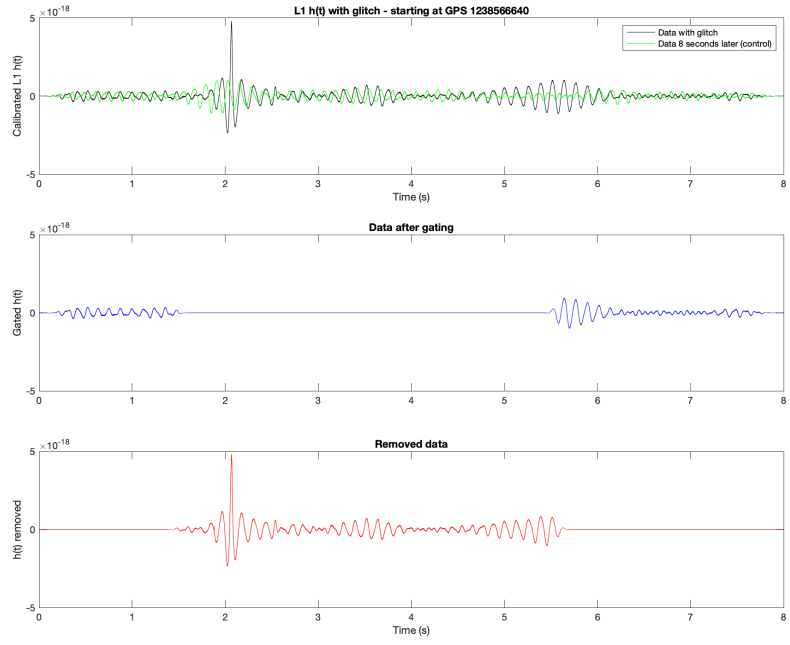

FIG. 2. Example drawn from early O3 L1 data of a loud glitch being removed via the inverse Tukey window gating.

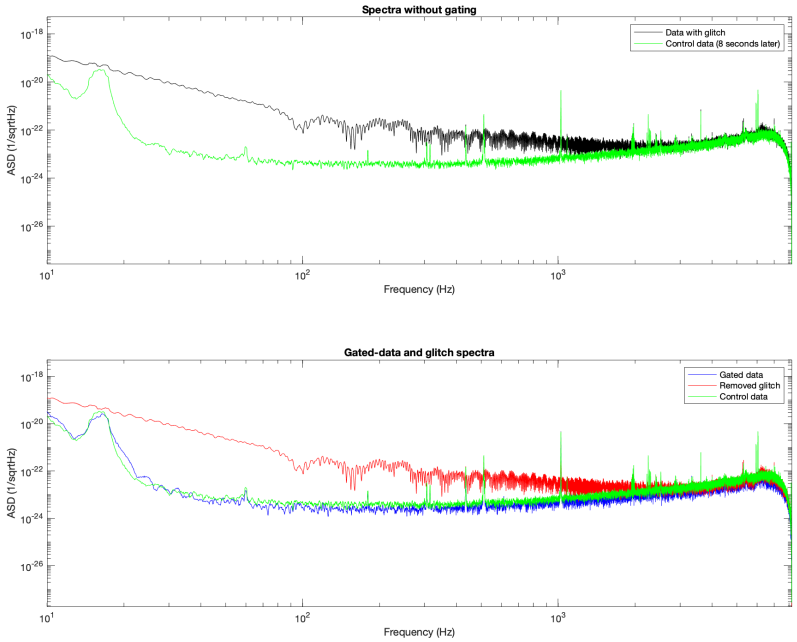

FIG. 3. Fourier-domain representation of the data shown in Figure 2.

and [L1,H1]:DCS-GATED_STRAIN_SUB60HZ_C01. The output frames also contain copies of the input strain channels, along with auxiliary data quality channels.

Band-limited RMS thresholds		
	L1	H1
25-50 Hz	2.0×10^{-7}	5.0×10^{-7}
70-110 Hz	2.2×10^{-7}	3.0×10^{-7}

TABLE I. Thresholds on band-limited RMS in the whitened strain channels [L1,H1]:GDS-DELTAL_EXTERNAL_DQ for the two bands (25-50 Hz and 70-110 Hz) used to trigger gates. The 25-50 Hz trigger was not used for H1 data in O3a prior to GPS 1239462000 (April 16, 2019) because of calibration lines present in the trigger band until that time. The 25-50 Hz trigger was not used for L1 data in O3b prior to GPS 1258700800 (November 25, 2019).

Input and output files in self-gating		
	L1	H1
Input frames	/hdfs/frames/O3/raw/L1/	/hdfs/frames/O3/raw/H1/
Input frames	/hdfs/frames/O3/hoft_C01/L1/L1/	/hdfs/frames/O3/hoft_C01/H1/
Output frames	/hdfs/user/jzweizig/frames/O3a/gated_hoft/L1/	/hdfs/user/jzweizig/frames/O3a/gated_hoft/H1v2/
Output log files	/home/jzweizig/public_html/gate-review/logs/L1/	/home/jzweizig/public_html/gate-review/logs/H1v2/
Output graph files	/home/jzweizig/public_html/gate-review/plots/L1/	/home/jzweizig/public_html/gate-review/plots/H1v2/

TABLE II. Locations on the Caltech cluster of input and output files used in the inital O3a self-gating process.

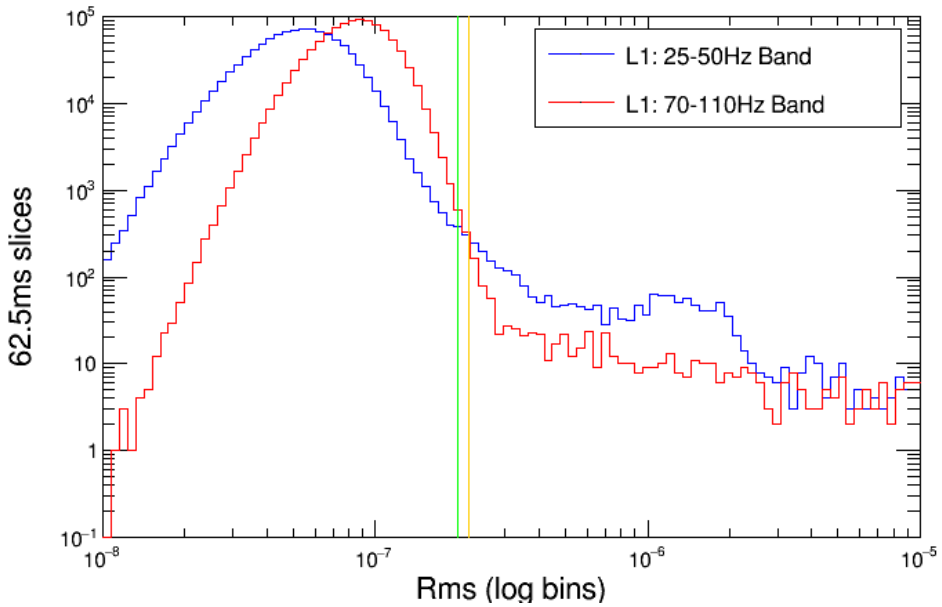

FIG. 4. Histograms of band-limited-RMS of the whitened strain channel L1:GDS-DELTAL_EXTERNAL_DQ for a sample 1-day interval late in O3a GPS [1253343600-1253430000]. The green and orange vertical lines show the trigger thresholds used for the 25-50 Hz and 70-110 Hz bands, respectively.

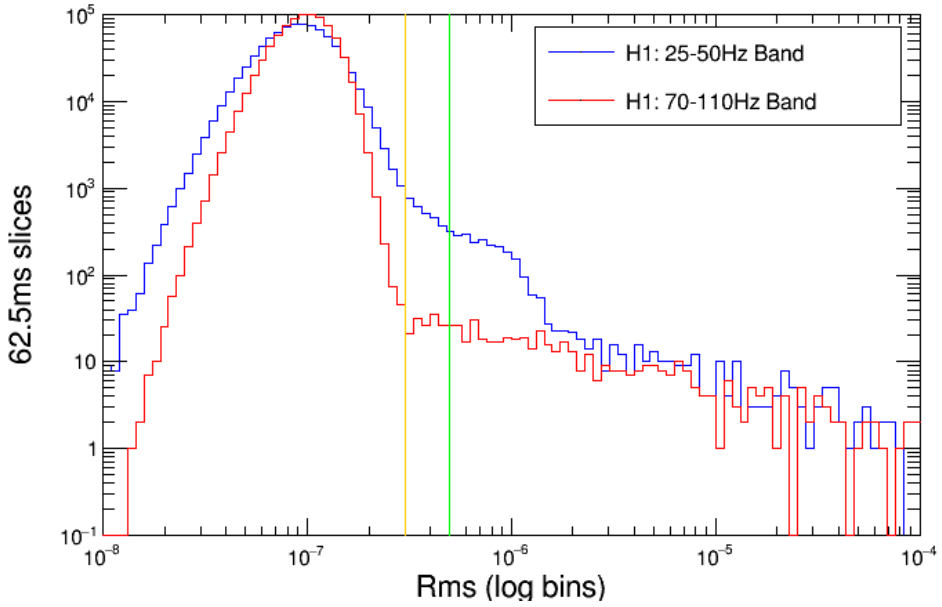

FIG. 5. Histograms of band-limited-RMS of the whitened strain channel H1:GDS-DELTAL_EXTERNAL_DQ for a sample 1-day interval late in O3a GPS [1253343600-1253430000] The green and orange vertical lines show the trigger thresholds used for the 25-50 Hz and 70-110 Hz bands, respectively.

Input and output files in self-gating		
	L1	H1
Input frames	/hdfs/frames/O3/raw/L1/	/hdfs/frames/O3/raw/H1/
Input frames	/hdfs/frames/O3/hoft_C01/L1/	/hdfs/frames/O3/hoft_C01/H1/
Input frames	/hdfs/frames/O3/hoft_C01_clean_sub60Hz/L1/	/hdfs/frames/O3/hoft_C01_clean_sub60Hz/H1/
Output frames	/hdfs/user/jzweizig/frames/O3a/gated_hoft/L1/	/hdfs/user/jzweizig/frames/O3a/gated_hoft/H1/
Output frames	/hdfs/user/jzweizig/frames/O3a/gated_clean_hoft/L1/	/hdfs/user/jzweizig/frames/O3a/gated_clean_hoft/H1/
Output frames	/hdfs/user/jzweizig/frames/O3b/gated_hoft/L1/	/hdfs/user/jzweizig/frames/O3b/gated_hoft/H1/
Output frames	/hdfs/user/jzweizig/frames/O3b/gated_clean_hoft/L1/	/hdfs/user/jzweizig/frames/O3b/gated_clean_hoft/H1/
Output logs	/home/jzweizig/public_html/gate-review/logs/L1/	/home/jzweizig/public_html/gate-review/logs/H1/
Output graphs	/home/jzweizig/public_html/gate-review/plots/L1/	/home/jzweizig/public_html/gate-review/plots/H1/

TABLE III. Locations on the Caltech cluster of input and output files used in the O3a+O3b self-gating process, applied to the C01 and C01 60-Hz-subtracted data streams.

III. GENERATION OF SFTS AND DEADTIME MEASUREMENTS

Most CW analyses begin with "Short Fourier Transforms" (SFTs) created from half-hour or longer contiguous time intervals. The self-gated frames were used to produce self-gated SFTs for use in searches below ∼500 Hz, where the loud glitches have the greatest effect on power spectral estimates. Including especially long gates or a large number of short gates leads to significant spectral noise biasing, if uncorrected.

As part of the iterative process of determining trigger thresholds, especially long single gates (>30 seconds) were flagged and excluded from the contiguous segments used for SFT generation. In addition, once the SFTs were generated (using the MakeSFTDAG script [1]), the list of generated gate windows was used to estimate deadtime in each SFT. Those SFTs with excessive total deadtime (>30 seconds for 1800s SFTs, > 2 minutes for 7200s SFTs) were flagged via a change of the file suffix from ".sft" to ".sftbad". Hence they are still available to analyses, but do not get used by default. Table IV summarizes the deadtime measurements for 1800s and 7200s SFTs for H1 and L1, for all SFTs, those marked with excessive gate time and those remaining.

Deadtimes in SFTs during O3	
L1 7200s	Sum of gates over all 2624 SFTs with any gate sum = 185958.1875, average gate sum = 70.9 (0.98%)
	Sum of gates over the 298 SFTs with high gate sum = 98568.5625, average gate sum = 330.8 (4.59%)
	Sum of gates over the 2326 SFTs with low gate sum = 87389.6250, average gate sum = 37.6 (0.52%)
L1 1800s	Sum of gates over all 11665 SFTs with any gate sum = 213133.6875, average gate sum = 18.3 (1.01%)
	Sum of gates over the 1367 SFTs with high gate sum = 127104.8750, average gate sum = 93.0 (5.17%)
	Sum of gates over the 10298 SFTs with low gate sum = 86028.8125, average gate sum = 8.4 (0.46%)
H1 7200s	Sum of gates over all 2519 SFTs with any gate sum = 77010.4375, average gate sum = 30.6 (0.42%)
	Sum of gates over the 69 SFTs with high gate sum = 28092.6250, average gate sum = 407.1 (5.65%)
	Sum of gates over the 2450 SFTs with low gate sum = 48917.8125, average gate sum = 20.0 (0.28%)
H1 1800s	Sum of gates over all 11327 SFTs with any gate sum = 88709.3750, average gate sum = 7.8 (0.43%)
	Sum of gates over the 303 SFTs with high gate sum = 34965.1250, average gate sum = 115.4 (6.41%)
	Sum of gates over the 11024 SFTs with low gate sum = 53744.2500, average gate sum = 4.9 (0.27%)

TABLE IV. Deadtimes over the full-O3 run for all SFTs generated from the gated, 60-Hz-cleaned data, for those with excessive gate times and for those remaining.

In addition, spectral investigations of a 1/120-Hz L1 spectral comb near calibration and other loud lines led to the identification of an interval (April 20-25, 2019) during which a Nikon camera was left in a state in which a mechanical shutter clicked every two minutes. SFTs produced from data during this interval were given the suffix ".sftnikon", again still available for analysis, but not used by default.

IV. SPECTRAL COMPARISONS

To assess the potential benefits from self-gating, O3a-averaged spectra were computed for L1 and H1 1800s and 7200s SFTs. Both arithmetic and weighted averages were computed, where the weighting used was inverse power weighting. Figures 6-9 show comparisons over the six-month data set.

In each figure, the top panel shows four different spectral estimates. The large gap between the arithmetic and weighted averages for ungated data below 500 Hz is due primarily to the large glitches. The inverse-noise weighting allows SFTs with few or no loud glitches to dominate the PSD, bringing it closer to what it would be in the absence of the glitches. The gated-data spectra show what is achieved by removing the glitches, and as expected, the difference between arithmetic and weighted averages is much smaller.

In the bottom panel of each figure is the ratio of the ungated spectra to the gated spectra, using both arithmetic and weighted averaging. That ratio can be quite large for the arithmetic average and still appreciable for weighted averaging. Since most CW search pipelines use some form of inverse noise weighting, it is the ratio of weighted averages that is of greater interest[1].

One can see from comparing Figure 6 to Figure 7 and from comparing Figure 8 to Figure 9 that larger improvements in weighted-average spectra are seen from gating for the 7200s SFTs than for the 1800s SFTs. That greater improvement is expected because longer SFTs mean that fewer of them are free of loud glitches.

One can also see significant dips in the ratio at very strong lines, as expected, but if one looks closely, artifacts can be observed too *in the vicinity* of strong lines in which the ratio falls below unity. This odd effect is an artifact of the inverse Tukey-window gate, which abruptly (0.25-second transition) turns off the the strong lines and then turns them back on again. The Fourier transform of these abrupt turn-offs/turn-ons creates significant spectral leakage near the line frequenciess. Since the gated data is the linear difference between the original data and what is excised, that leakage affects the retained data. Figure 10 shows a toy example comparing the ASD for a pure 60.00028-Hz signal (half-FFT-bin offset) in a half-hour data stretch with that for the same signal with a 5-second inverse-Tukey-window gate imposed about 500 seconds into the interval, along with the ASD of the excised data. Substantial spectral leakage is apparent. Figure 11 shows a horizontal zoom of the 60-Hz power main region for both the spectral curves and derived ratios for the actual H1 data (zoom of Figure 8). As a result of this spectral leakage, one must exercise care in CW searches near very strong lines. Similar behavior is observed for the O3b data set.

V. HARDWARE INJECTION RECOVERY VALIDATION

To verify that the gating procedure has not interfered with CW signal detection, the CW "hardware injections" were recovered using a fully coherent $\mathcal{F}$-statistic procedure [3, 4] using the 1800s SFTs and assuming the exactly known simulated sky positions and frequency / spindown parameters. Of the 18 injections performed in the O3 run in each interferometer's data, 15 were for isolated stars with frequencies below 2000 Hz (see Table V for injection parameters). One of those was at an injection frequency of about 12.34 Hz which proved unrecoverable in the high noise background below 20 Hz. Of the remaining 14 injections, all were recoverable in both the original ungated and the regenerated gated data. Figures 12- 13 show the computed $\mathcal{F}$-statistic values and recovered strain amplitudes h_0 for both interferometers. One sees that for injections above 500 Hz, there is no appreciable difference, but that for lower frequencies, one sees louder $\mathcal{F}$-statistic values for gated data and recovered h_0 amplitudes with smaller uncertainties and central values that agree better, on the whole, with the true injected values.

[1] Note that the PowerFlux search pipeline [2] uses a more aggressive noise weighting (inverse power squared) and is somewhat less susceptible to the glitches.

Injection	f_0 (Hz)	$\dot{f}$ (Hz/s)	A_+	$A_\times$	h_0	$\cos(\iota)$	ψ (rad)	Dec (rad)	RA (rad)
0	265.58	-4.15e-12	5.00e-26	4.87e-26	6.12e-26	0.795	0.770	-0.981	1.249
1	848.94	-3.00e-10	3.33e-25	2.54e-25	5.47e-25	0.464	0.356	-0.514	0.653
2	575.16	-1.37e-13	7.07e-26	-7.05e-26	7.59e-26	-0.929	-0.222	0.060	3.757
3	108.86	-1.46e-17	6.53e-26	-1.05e-26	1.30e-25	-0.081	0.444	-0.584	3.113
4	1391.01	-2.54e-08	5.75e-25	2.96e-25	1.07e-24	0.277	-0.648	-0.218	4.887
5	52.81	-4.03e-18	2.42e-25	1.85e-25	3.99e-25	0.463	-0.364	-1.463	5.282
6	145.50	-6.73e-09	1.97e-25	-5.91e-26	3.84e-25	-0.154	0.471	-1.142	6.261
7	1220.44	-1.12e-09	1.30e-25	1.25e-25	1.65e-25	0.757	0.512	-0.357	3.900
8	190.17	-8.65e-09	6.55e-26	9.62e-27	1.30e-25	0.074	0.170	-0.583	6.133
9	763.85	-1.45e-17	8.96e-26	-8.02e-26	1.30e-25	-0.619	-0.009	1.321	3.471
10	26.33	-8.50e-11	6.19e-25	-6.19e-25	6.26e-25	-0.988	0.615	0.748	3.867
11	31.42	-5.07e-13	1.76e-25	-1.04e-25	3.17e-25	-0.329	0.412	-1.017	4.976
12	37.85	-6.25e-09	1.32e-25	1.16e-26	2.63e-25	0.044	-0.068	-0.296	5.792
13	12.43	-1.00e-11	1.32e-24	0.00e+00	2.65e-24	0.000	0.000	0.250	0.250
14	1991.09	-1.00e-12	9.13e-25	0.00e+00	1.83e-24	0.000	1.000	-0.250	5.250

TABLE V. Parameters of 15 hardware injections for isolated stars below 2000 Hz. Reference GPS times for injections 0-9: 751680013; injections 10-15: 930582085; injections 16-17: 1230336018.

[1] G. Mendell, python script used to generate SFTs using Condor for general use in CW searches.
[2] B. Abbott *et al.*, *Phys. Rev.* D **77** 022001 (2008).
[3] P. Jaranowski, A. Królak and B.F. Schutz, *Phys. Rev.* D **58** 063001 (1998).
[4] R. Prix, LIGO Technical Report T0900149-v6, August 2018; `https://dcc.ligo.org/LIGO-T0900149`.

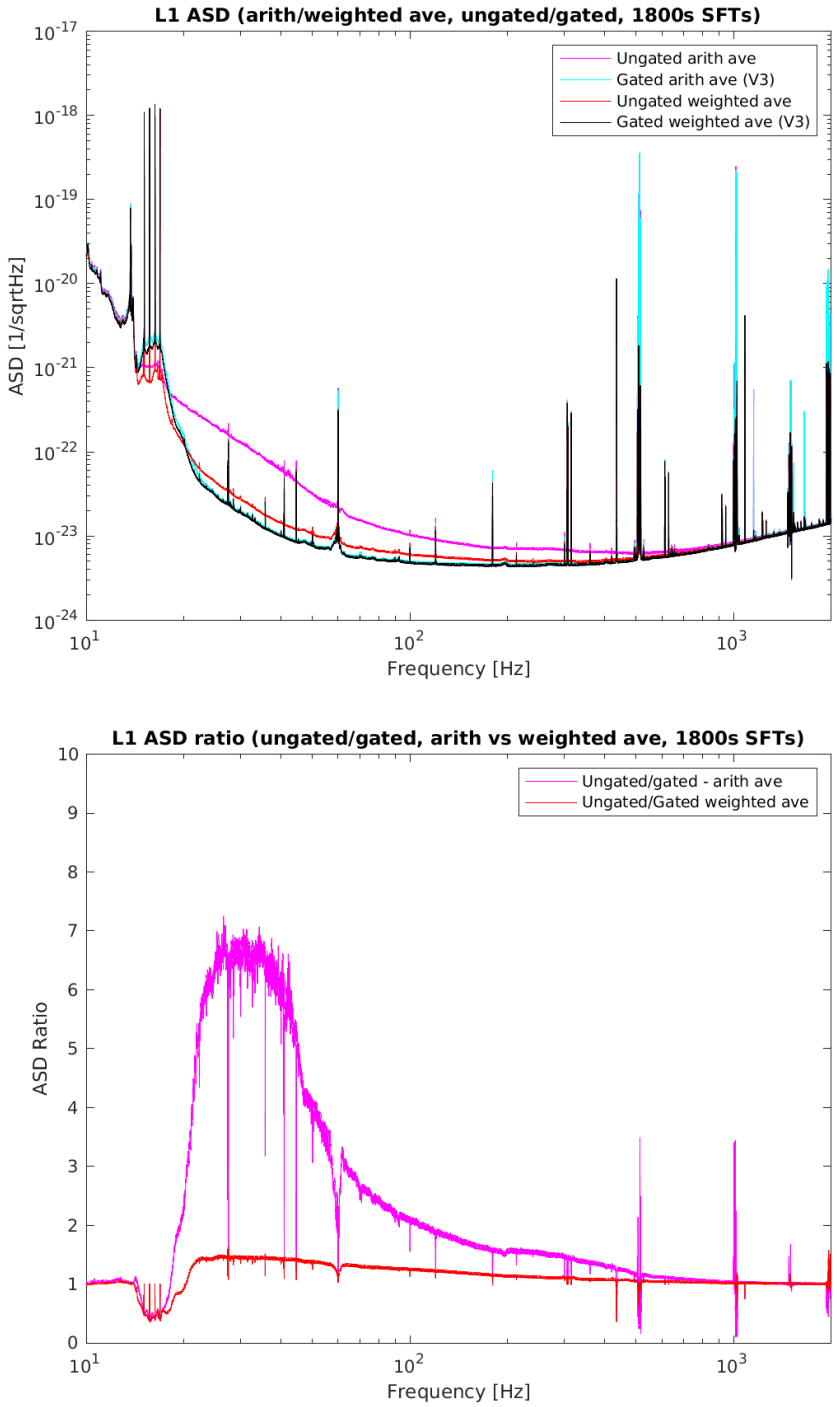

FIG. 6. *Top panel:* Arithmetic and weighted averages of 1800s L1 SFTs over the six months of O3a for both SFTs generated from ungated data and from gated data. *Bottom panel:* Ratios of the ungated to gated spectra for both arithmetic and weighted averaging.

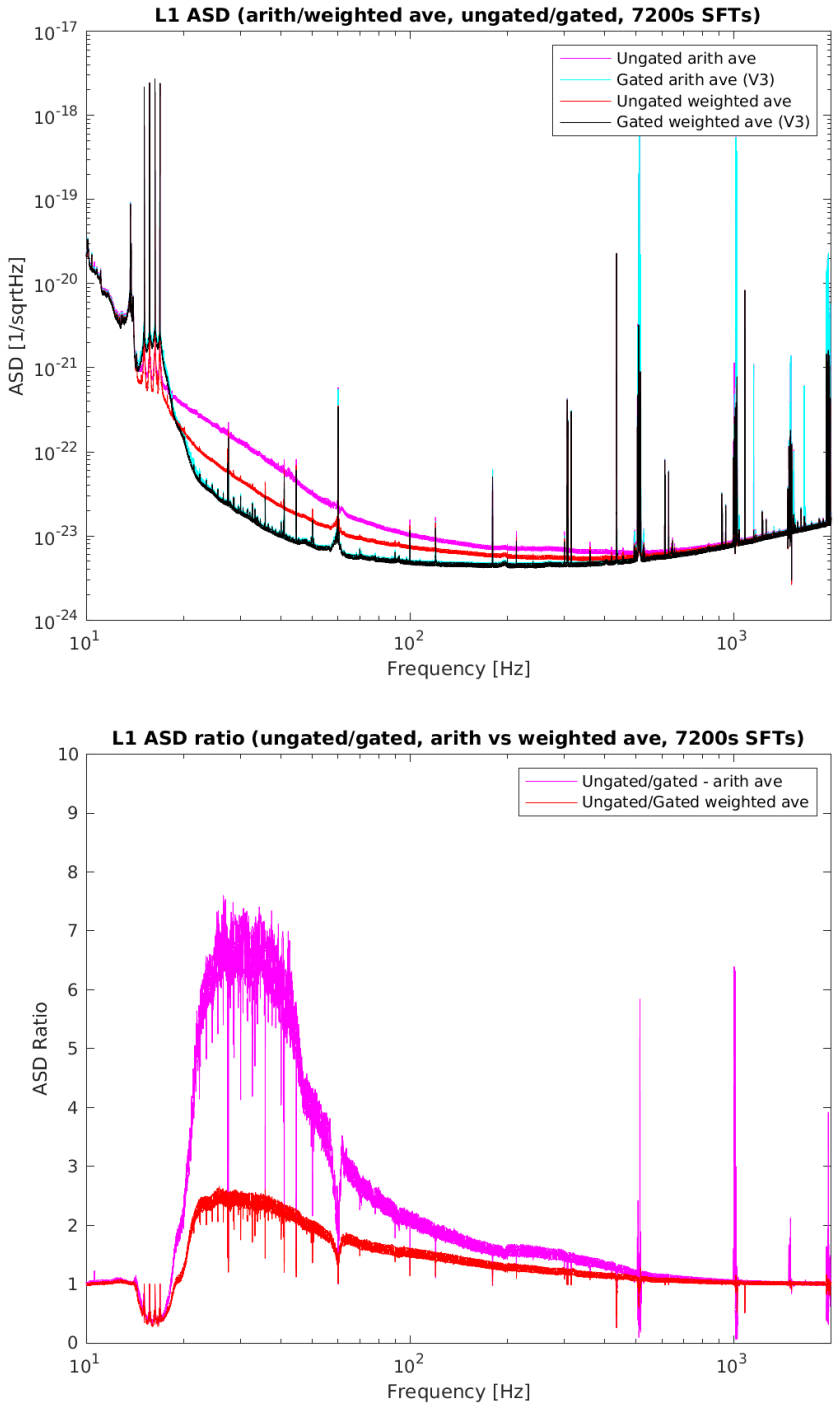

FIG. 7. *Top panel:* Arithmetic and weighted averages of 7200s L1 SFTs over the six months of O3a for both SFTs generated from ungated data and from gated data. *Bottom panel:* Ratios of the ungated to gated spectra for both arithmetic and weighted averaging.

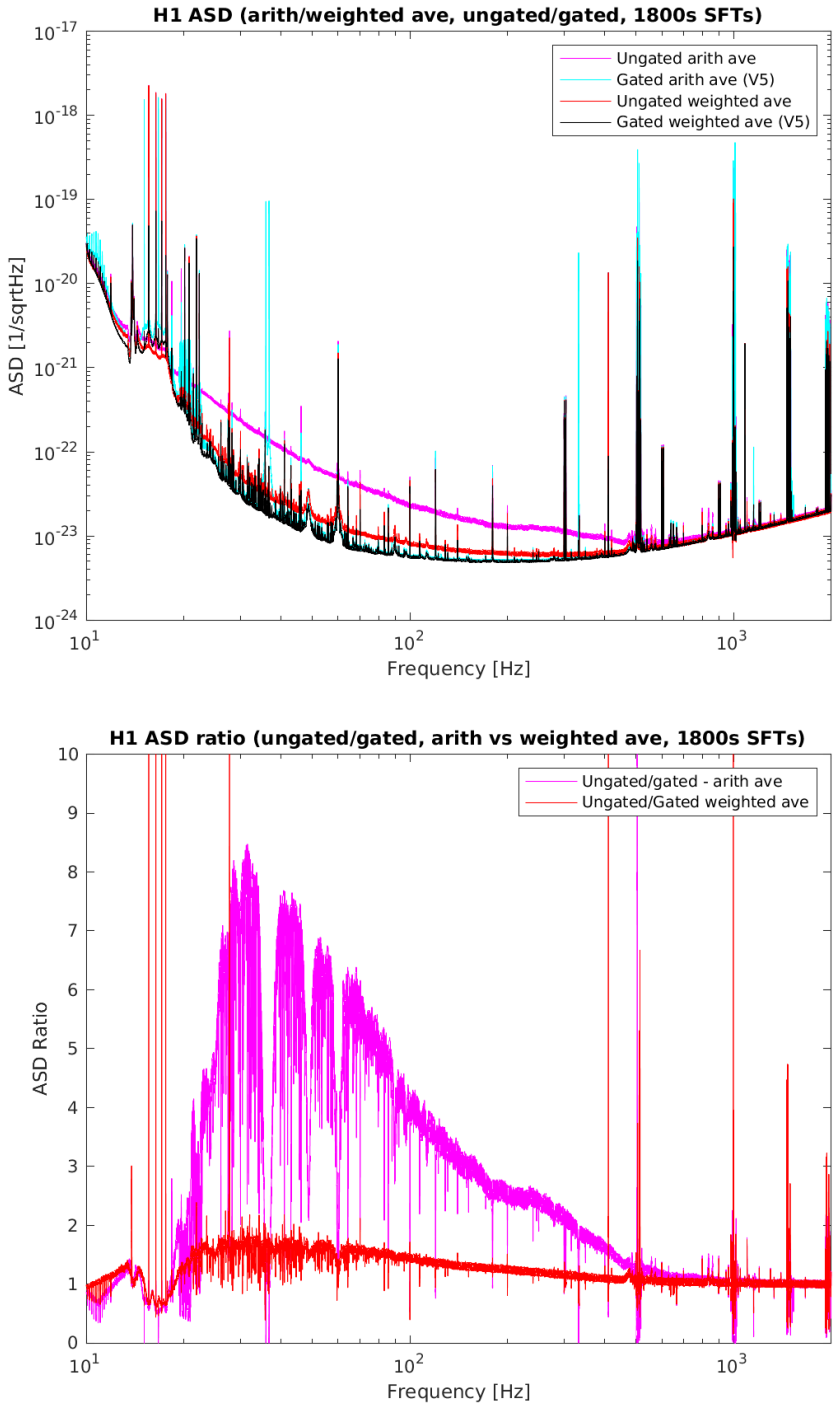

FIG. 8. *Top panel:* Arithmetic and weighted averages of 1800s H1 SFTs over the six months of O3a for both SFTs generated from ungated data and from gated data. *Bottom panel:* Ratios of the ungated to gated spectra for both arithmetic and weighted averaging.

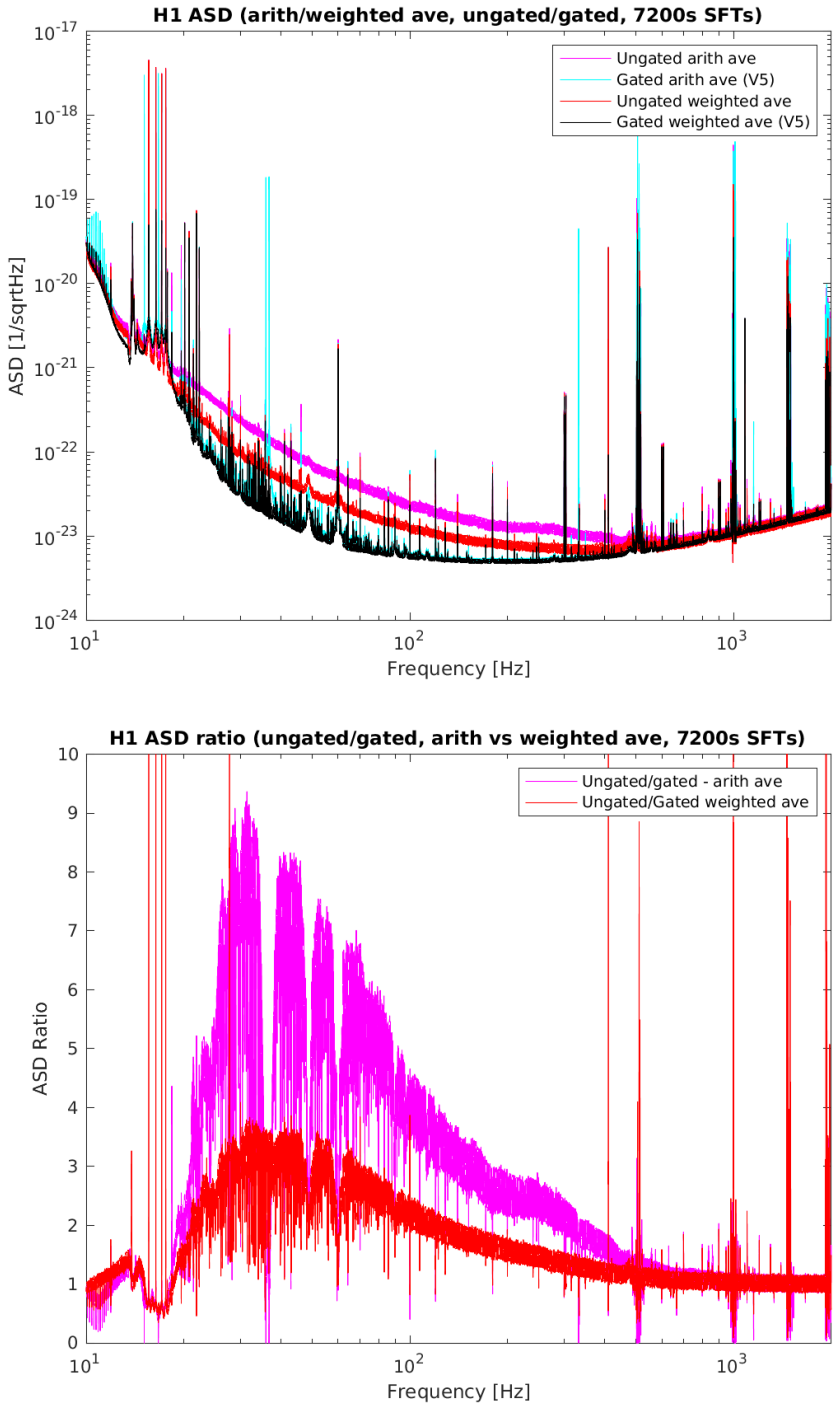

FIG. 9. Arithmetic and weighted averages of 7200s H1 SFTs over the six months of O3a for both SFTs generated from ungated data and from gated data.

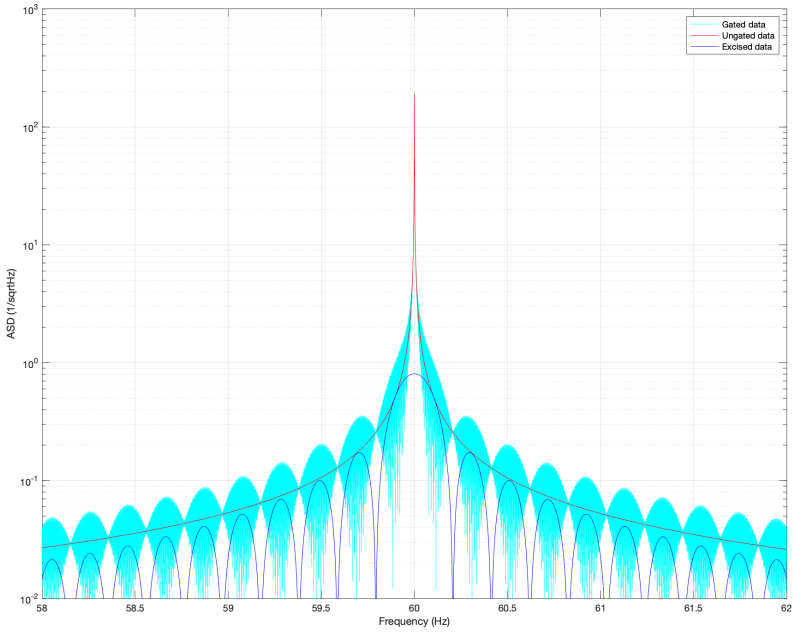

FIG. 10. Illustration with toy signal of the effect of a 5-second inverse-Tukey-window gate on a monochromatic signal of frequency 60.00028 Hz. The ASD of the gated data inherits spectral leakage from the subtraction of the excised data's spectrum.

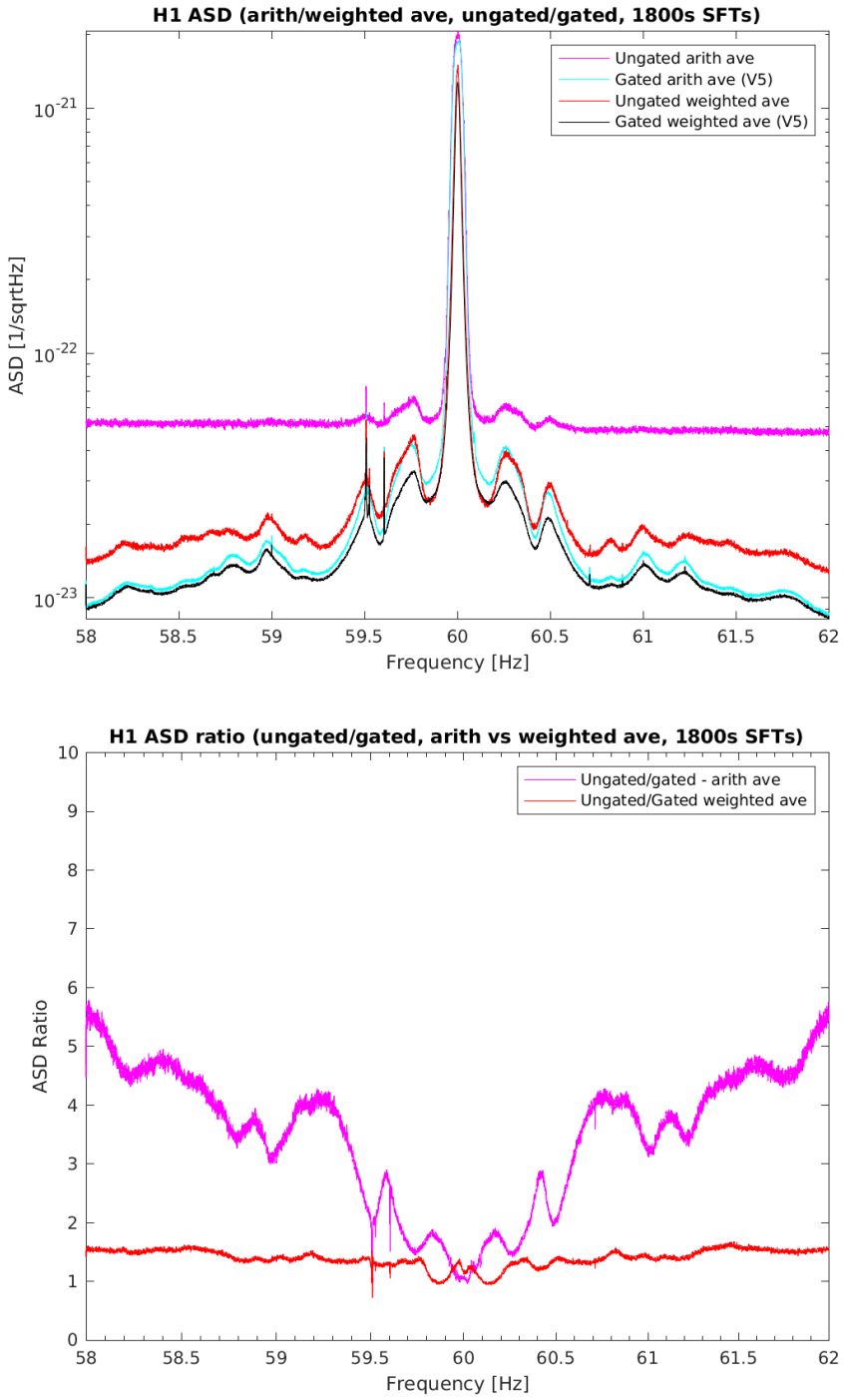

FIG. 11. *Top panel:* Zoom-in of the 60 Hz power mains for H1 O3a data for ungated and ungated spectral averages. *Bottom panel:* Resulting ratios of ungated to gated spectra. Note that spectral leakage from the inverse Tukey-window gate leads to regions where the ratio falls slightly below unity.

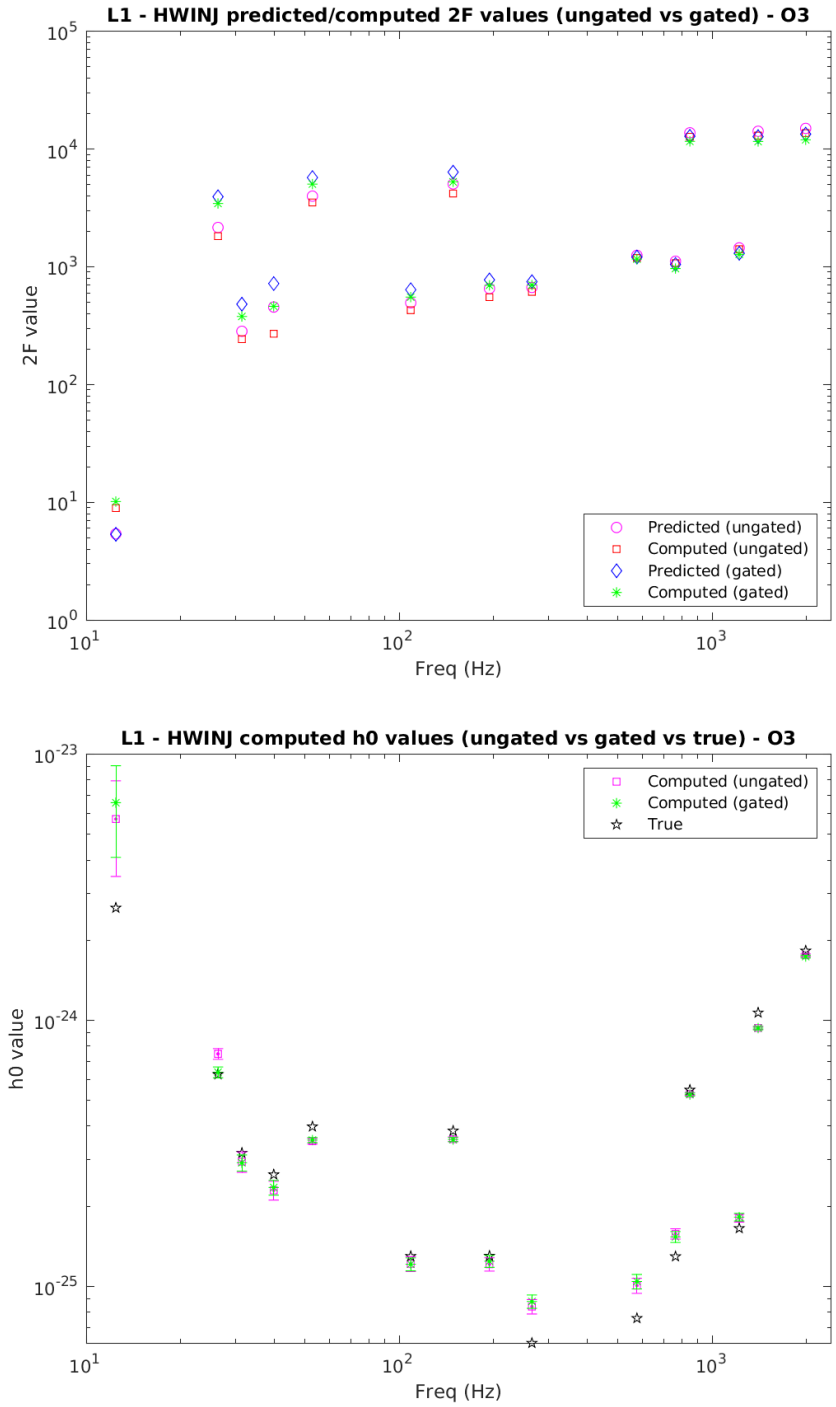

FIG. 12. *Top panel:* Computed and predicted $\mathcal{F}$-statistic values for 15 isolated-star hardware injections in L1 data *vs.* frequency for ungated and gated full-O3 data. Setting aside the 12.34 Hz injection into heavily contaminated data, one sees an increase in $\mathcal{F}$-statistic values for the gated data for frequencies below ∼500 Hz, where loud glitches inflate apparent detector noise in ungated data. *Bottom panel:* Corresponding reconstructed strain amplitude h_0 values estimated from the $\mathcal{F}$-statistic reconstruction. One sees, on the whole, smaller uncertainties and improved agreement between measured and true amplitudes for the gated SFTs.

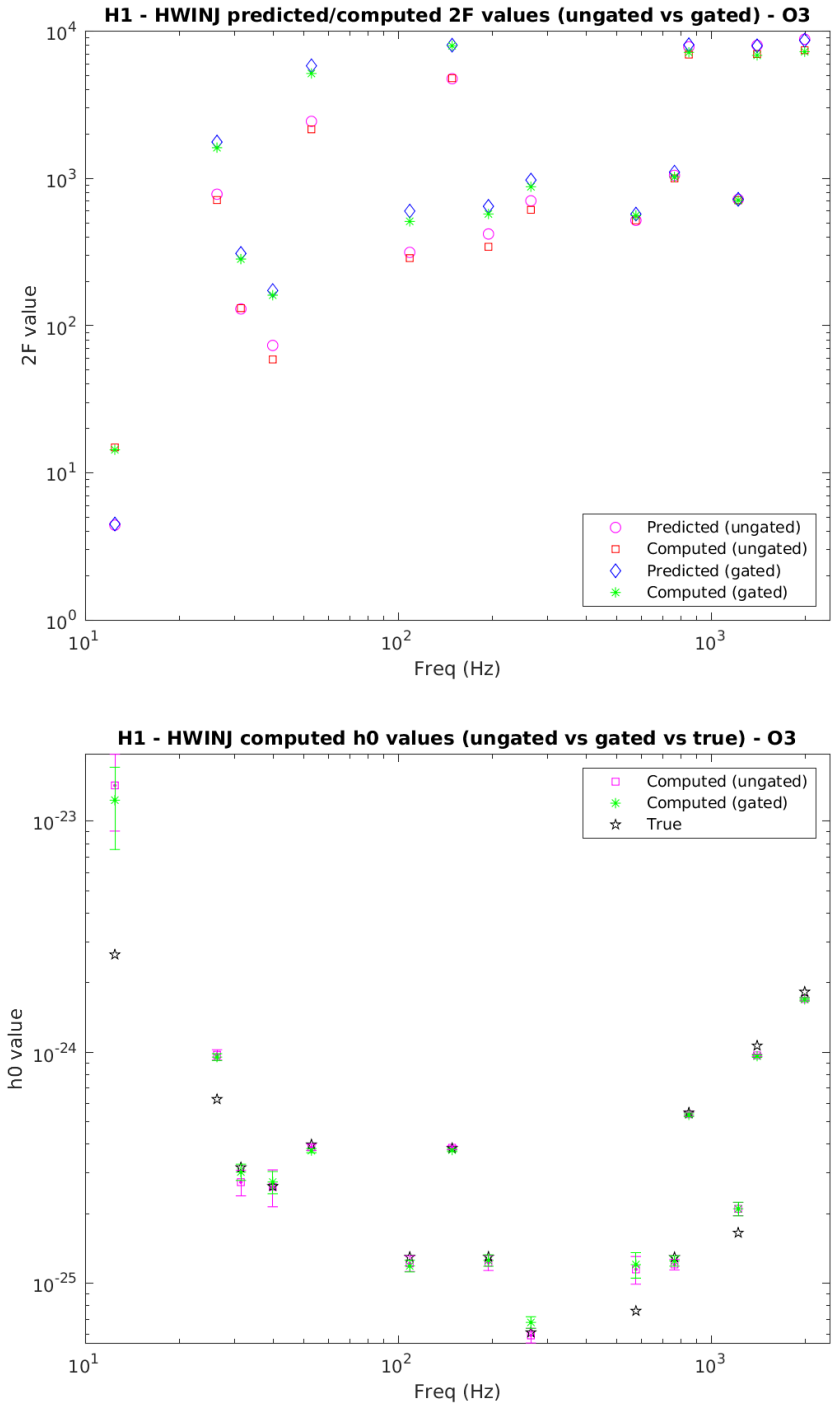

FIG. 13. *Top panel:* Computed and predicted $\mathcal{F}$-statistic values for 15 isolated-star hardware injections in H1 data *vs.* frequency for ungated and gated full-O3 data. Setting aside the 12.34 Hz injection into heavily contaminated data, one sees an increase in $\mathcal{F}$-statistic values for the gated data for frequencies below ∼500 Hz, where loud glitches inflate apparent detector noise in ungated data. *Bottom panel:* Corresponding reconstructed strain amplitude h_0 values estimated from the $\mathcal{F}$-statistic reconstruction. One sees, on the whole, smaller uncertainties and improved agreement between measured and true amplitudes for the gated SFTs.