The Role of Data Quality in S5 Burst Analyses

Lindy Blackburn[1] for the LIGO Scientific Collaboration

[1] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

1: Overview

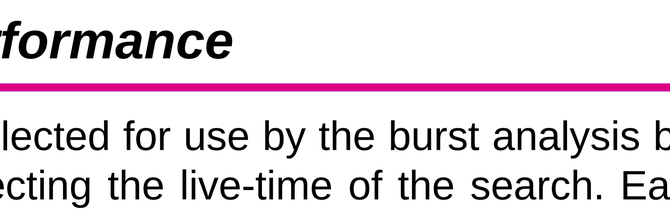
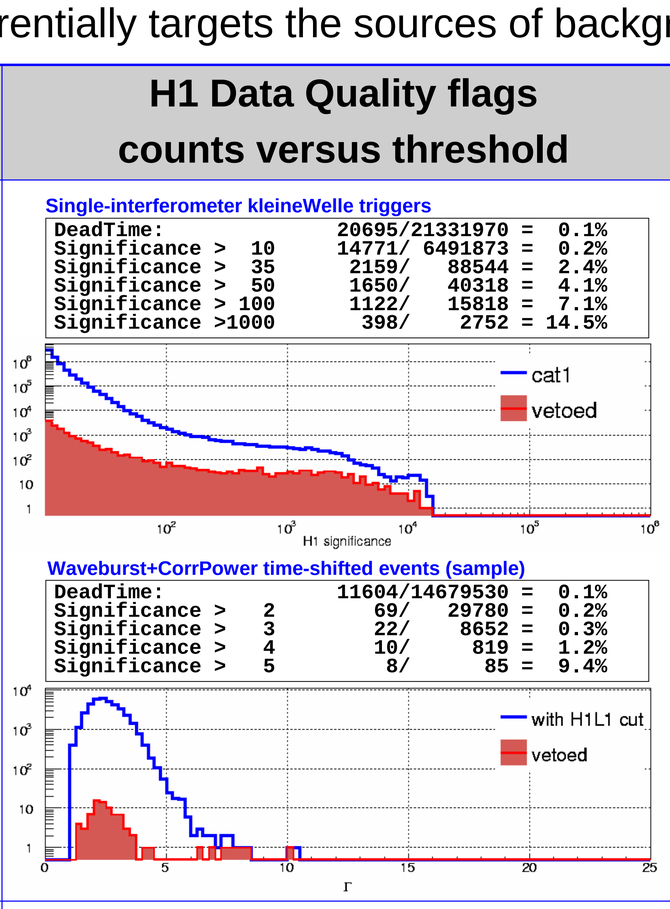
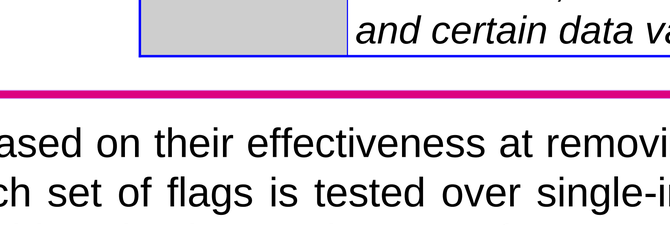
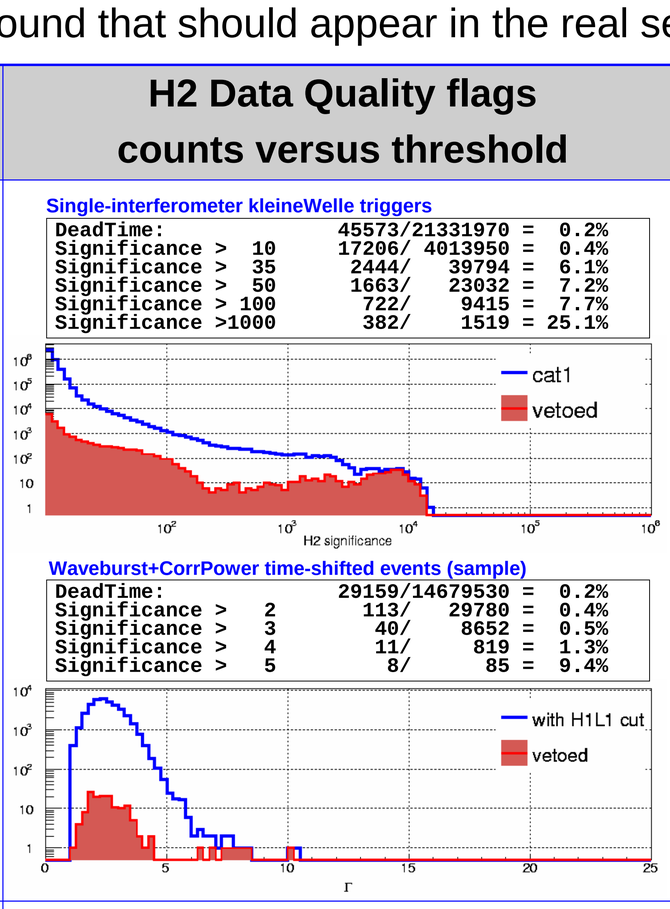
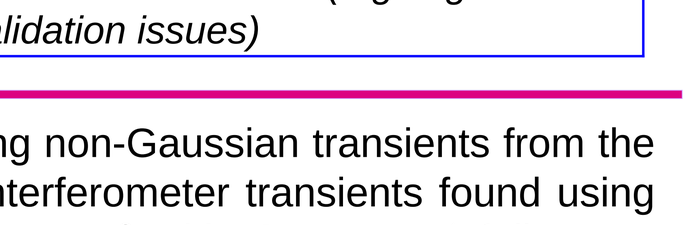
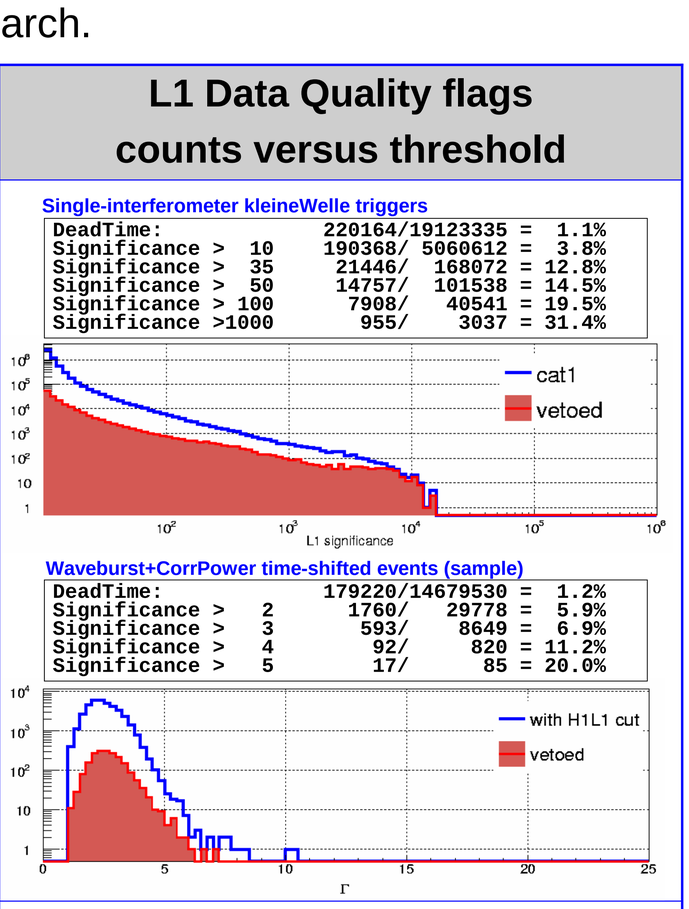
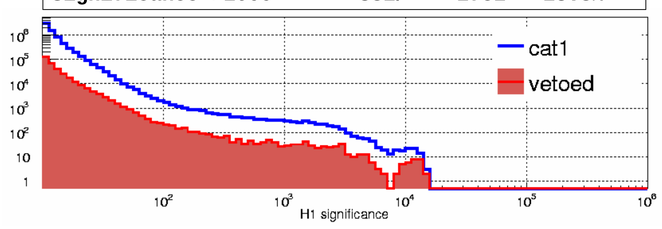
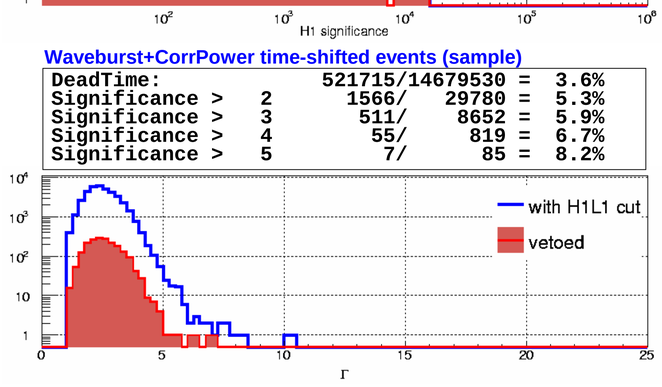
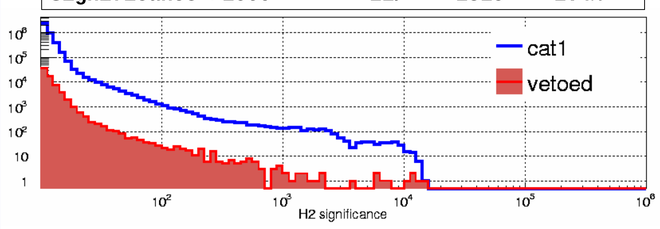
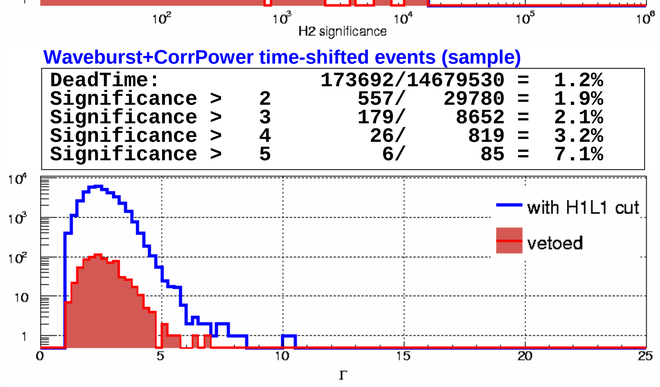
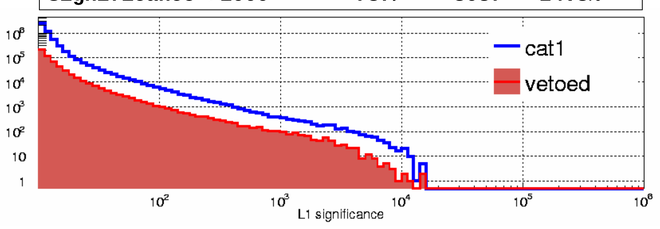
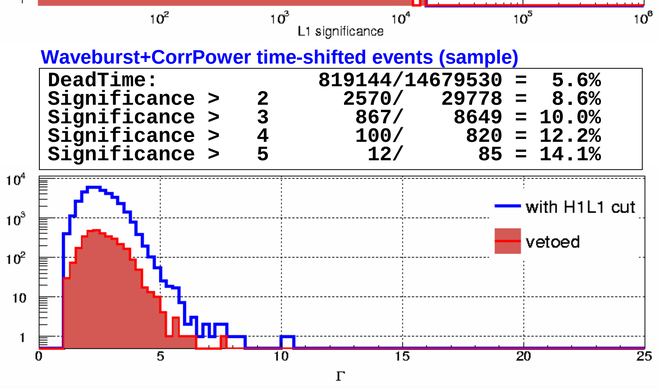
Data Quality (DQ) flags identify epochs in Science data which may have a negative impact on the analyses due to errors in data acquisition, poor sensitivity, excessive contribution to the false event rate, or general untrustworthiness of data. The intervals constructed target known problems with the instrument or environmental conditions. Individual Data Quality flags are evaluated by their effectiveness at removing single-interferometer and coincident noise transients, particularly the loudest ones, from the gravitational-wave data streams. To make sure Data Quality flags remain independent of the presence of a true gravitational wave, we check that they are not triggered by hardware injections. The LSC burst group, like the inspiral group[1], makes use of tiered system of Data Quality flags which are applied depending on the type of analysis.

2: Data Quality Categories

Category 1	Obvious Data Quality cuts. These times are not processed by the search pipelines. *(e.g. calibration problems, test injections, photodiode saturations)*
Category 2	"Unconditional" post-processing cuts: data is unreliable and there is an established one-on-one correlation with loud transients. *(e.g. saturations in the alignment control system, glitches in the power mains)*
Category 3	"Conditional" post-processing cuts, for upper limit: statistical correlation to loud transients. We may still look for detection candidates at these times, exerting caution when establishing detection confidence. *(e.g. train/seismic flags, 1 minute pre-lock-loss, "dips" of light stored in the arm cavities)*
Category 4	Advisory flags: no clear evidence of direct correlation to loud transients, but if we find a detection candidate at these times, we need to exert caution *(e.g. high wind and certain data validation issues)*

3: Data Quality Performance

Data Quality flags are selected for use by the burst analysis based on their effectiveness at removing non-Gaussian transients from the data while minimally effecting the live-time of the search. Each set of flags is tested over single-interferometer transients found using *kleineWelle*[2] as well as a sample of time-shifted triple-coincident background events from *WaveBurst*[3] with *CorrPower*[4] follow-up. Single-interferometer analysis provides the best statistics and a clear picture of what happens at each instrument, while the time-shifted coincident analysis preferentially targets the sources of background that should appear in the real search.

Data Quality applied (after Category 1)	H1 Data Quality flags counts versus threshold	H2 Data Quality flags counts versus threshold	L1 Data Quality flags counts versus threshold
Category 2: •*saturations in the alignment control system* •*severe impulsive glitches in the power mains* •*uncertain calibration* •*large glitches in the thermal compensation system*	Single-interferometer kleineWelle triggers DeadTime: 20695/21331970 = 0.1% Significance > 10: 14771/6491873 = 0.2% Significance > 35: 2159/88544 = 2.4% Significance > 50: 1650/40318 = 4.1% Significance > 100: 1122/15818 = 7.1% Significance >1000: 398/2752 = 14.5%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 11604/14679530 = 0.1% Significance > 2: 69/29780 = 0.2% Significance > 3: 22/8652 = 0.3% Significance > 4: 10/819 = 1.2% Significance > 5: 8/85 = 9.4% 	Single-interferometer kleineWelle triggers DeadTime: 45573/21331970 = 0.2% Significance > 10: 17206/4013950 = 0.4% Significance > 35: 2444/39794 = 6.1% Significance > 50: 1663/23032 = 7.2% Significance > 100: 722/9415 = 7.7% Significance >1000: 382/1519 = 25.1%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 29159/14679530 = 0.2% Significance > 2: 113/29780 = 0.4% Significance > 3: 40/8652 = 0.5% Significance > 4: 11/819 = 1.3% Significance > 5: 8/85 = 9.4% 	Single-interferometer kleineWelle triggers DeadTime: 220164/19123335 = 1.1% Significance > 10: 190368/5060612 = 3.8% Significance > 35: 21446/168072 = 12.8% Significance > 50: 14757/101538 = 14.5% Significance > 100: 7908/40541 = 19.5% Significance >1000: 955/3037 = 31.4%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 179220/14679530 = 1.2% Significance > 2: 1760/29778 = 5.9% Significance > 3: 593/8649 = 6.9% Significance > 4: 92/820 = 11.2% Significance > 5: 17/85 = 20.0% 
Category 3 (no overlap in time with Category 2): •*120s prior to lock-loss* •*noise in power mains* •*transient drop in light stored in arm cavities* •*times when one Hanford instrument is unlocked* •*very poor sensitivity* •*severe seismic activity or wind speed* •*earthquakes and hurricanes*	Single-interferometer kleineWelle triggers DeadTime: 838153/21331970 = 3.9% Significance > 10: 327528/6491873 = 5.0% Significance > 35: 9685/88544 = 10.9% Significance > 50: 4878/40318 = 12.1% Significance > 100: 1978/15818 = 12.5% Significance >1000: 381/2752 = 13.8%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 521715/14679530 = 3.6% Significance > 2: 1566/29780 = 5.3% Significance > 3: 511/8652 = 5.9% Significance > 4: 55/819 = 6.7% Significance > 5: 7/85 = 8.2% 	Single-interferometer kleineWelle triggers DeadTime: 255527/21331970 = 1.2% Significance > 10: 70192/4013950 = 1.7% Significance > 35: 816/39794 = 2.1% Significance > 50: 489/23032 = 2.1% Significance > 100: 229/9415 = 2.4% Significance >1000: 21/1519 = 1.4%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 173692/14679530 = 1.2% Significance > 2: 557/29780 = 1.9% Significance > 3: 179/8652 = 2.1% Significance > 4: 26/819 = 3.2% Significance > 5: 6/85 = 7.1% 	Single-interferometer kleineWelle triggers DeadTime: 1124227/19123335 = 5.9% Significance > 10: 569721/5060612 = 11.3% Significance > 35: 31016/168072 = 18.5% Significance > 50: 19018/101538 = 18.7% Significance > 100: 8132/40541 = 20.1% Significance >1000: 737/3037 = 24.3%  Waveburst+CorrPower time-shifted events (sample) DeadTime: 819144/14679530 = 5.6% Significance > 2: 2570/29778 = 8.6% Significance > 3: 867/8649 = 10.0% Significance > 4: 100/820 = 12.2% Significance > 5: 12/85 = 14.1% 

4: Summary and Future Plans

Data Quality plays a critical role in making sure the background distribution for the burst analysis is well behaved and that a small amount of livetime does not ruin an upper limit or the confidence of a real event. With a long run such as S5, which included various periods of commissioning, there is an ongoing effort to understand the evolution of data quality issues throughout the run, and apply data quality cuts tailored to the specific state of the instrument at a given time. Also new data quality flags are continually being developed and tested.

[1] See accompanying GWDAW12 poster by Jake Slutsky, *Data Quality and Vetoes for the CBC analysis in LIGO's 5th Science Run*

[2] kleineWelle identifies excess signal power in the wavelet domain of a whitened data stream. It is used by the burst and inspiral analysis groups to find transients in the gravitational-wave and auxiliary data streams for data quality and veto work, and for general detector characterization, see: **L. Blackburn, *kleineWelle technical documentation*, http://www.ligo.caltech.edu/docs/T/T060221-00.pdf**

[3, 4] Waveburst identifies coincident and consistent excess power in the wavelet domain across a set of gravitational-wave data streams. CorrPower is a cross-correlation based test for waveform consistency which has been used to ultimately rank events by significance (Γ) of the correlation in the burst analyses. Waveburst+CorrPower has been used for the LIGO S2, S3, and S4 burst upper limit searches., see: **LIGO Scientific Collaboration, *Search for gravitational-wave bursts in LIGO data from the fourth science run*, Class. Quantum Grav. 24 (2007) 5343-5369. (arXiv:gr-qc/0704.0943)**