Data Quality Studies

Methods and Milestones

J. McIver for the
LIGO Scientific Collaboration
and the Virgo Collaboration

Outline

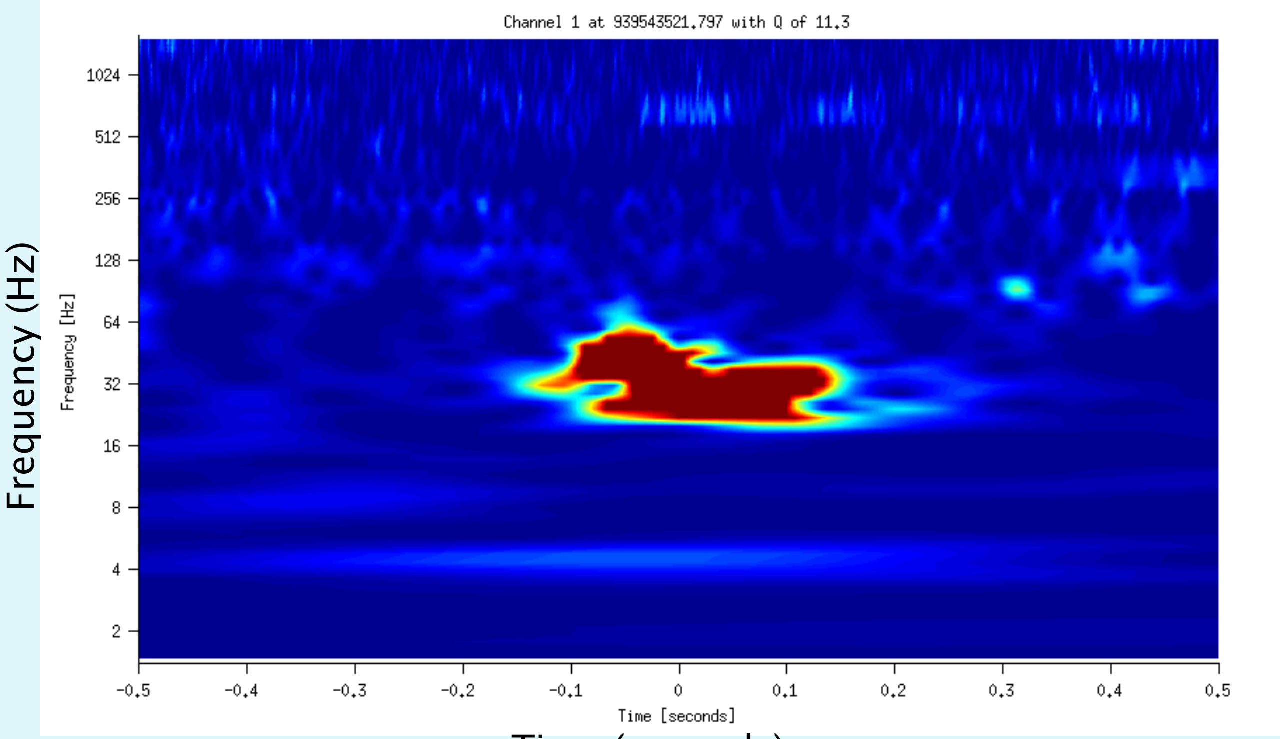
- An introduction to “glitches”
  - How loud are they and how often do we see them?
  - How do they limit the searches?
- Methods used to abate glitches
  - Hunting down the source
  - Data quality vetoes
    - Virgo
    - eLIGO
- Milestones in enhanced detector data quality
- The advanced detector era

Introduction to Glitches

- What is a "glitch"?

Gravitational wave data channel

- Noise transient
- Characterized by
  - SNR (signal to noise ratio)
  - Central freq
  - Shape in spectrogram plot

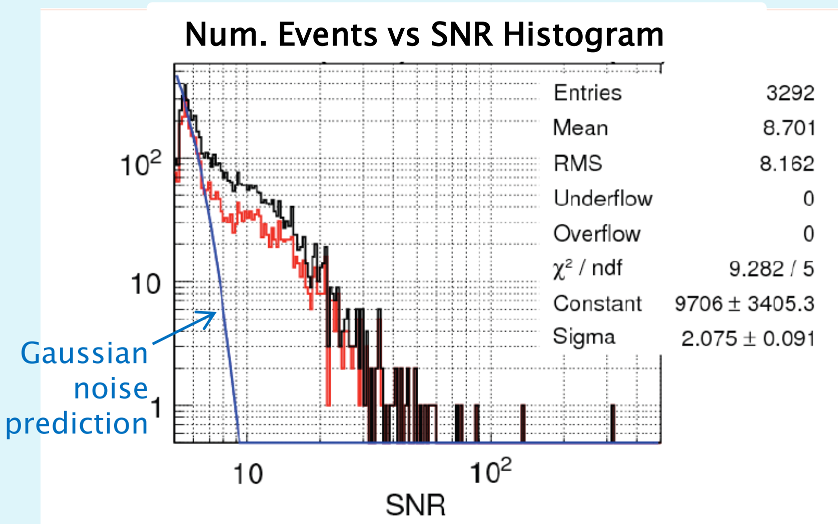


Time (seconds)

Introduction to Glitches

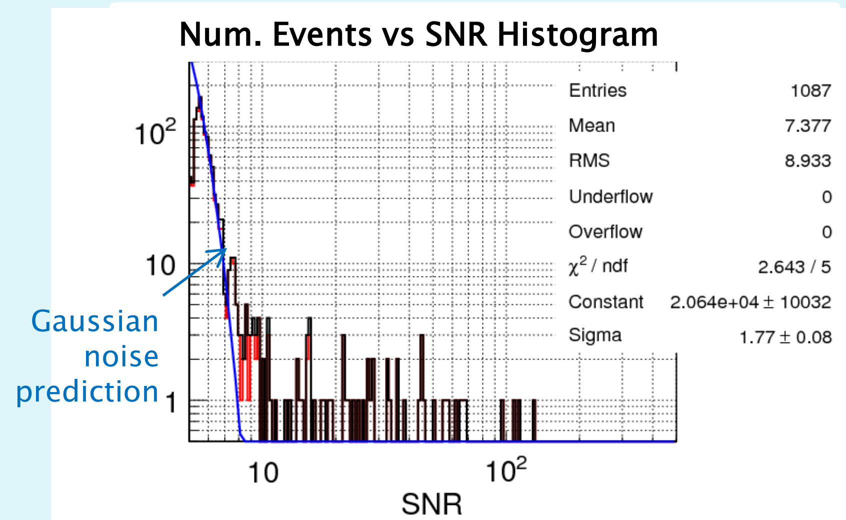
- How loud are they?
  - It depends on the weather, seismic noise, ocean waves, human movement, traffic, trains, instrumental/electronic malfunctions etc.

A “**bad**” day at Hanford during S6



S6D July 25 2010 LHO

A “**good**” day at Hanford during S6

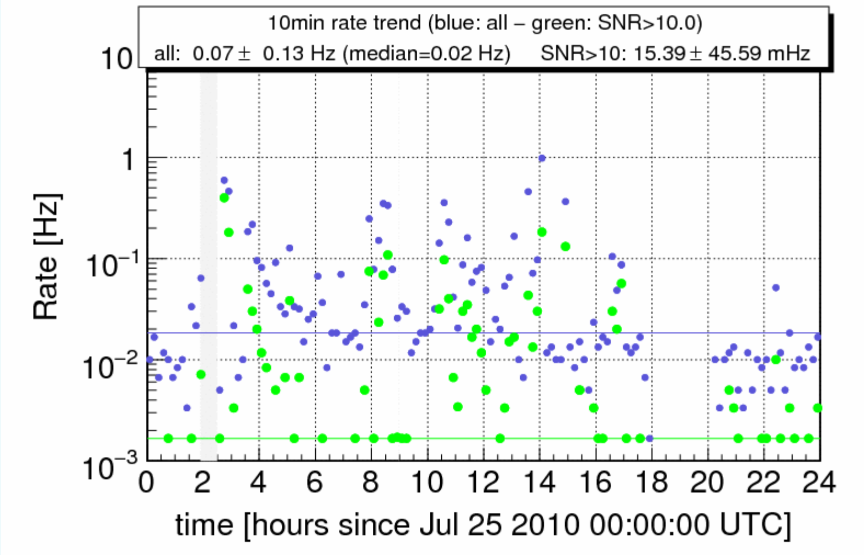


S6D Aug 21 2010 LHO

Introduction to Glitches

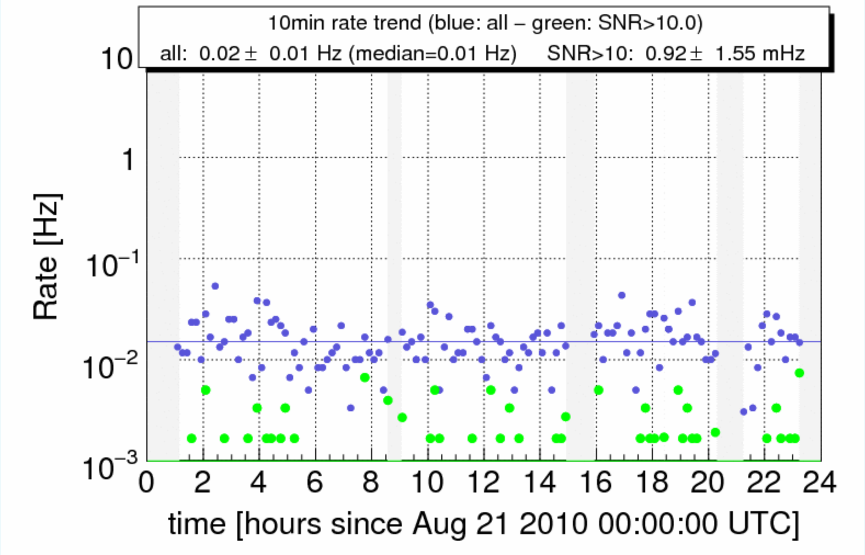
- How often do we see them?

A “**bad**” day at Hanford during S6



S6D July 25 2010 LHO

A “**good**” day at Hanford during S6



S6D Aug 21 2010 LHO

How do glitches limit the searches?

- Coherent or coincident searches are much less sensitive to glitches
  - Coincidence in time is the first line of defense against falsely claiming a detection
  - But, glitches are common enough that coincidence does still happen

How do glitches limit the searches?

- Low mass compact binary coalecence vs. burst searches
  - The low mass compact binary coalescence (CBC) search requires candidate signals to match a template of known waveforms as well as coincidence
    - Limited by loud glitches that blind the searches, and glitches that mimic the template forms
  - Searches for astrophysical “burst” sources require coincidence
    - Limited by frequent glitches (increased likelihood of coincidence)

Methods used to abate glitches

Google maps image of Hanford detector and surrounding highways

Hunting Glitches

- Glitch hunting targets:
  - Glitches that are having the most impact on the searches
  - Patterns, anomalies, or unusual behavior seen in glitches found by the searches
  - Glitches that cause lock loss

- Prominent sources of glitches could be:
  - a problem with the physical equipment
  - a digital signal processing artifact,
  - an environmental issue, etc.

- The optimal solution is to fix these instrumentally

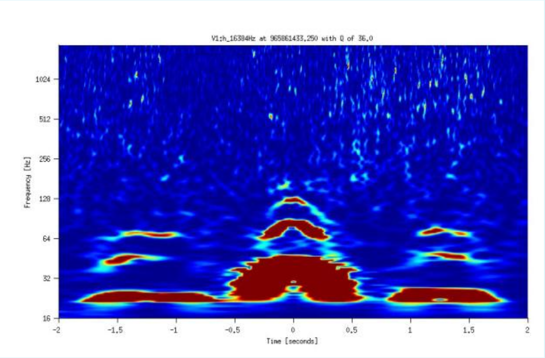
Hunting Glitches

- How do we find patterns/anomalies in glitch behavior?
  - We look at the characteristics of glitches (i.e. common frequency, signal to noise ratio (SNR), etc.)
  - We look at glitches' coincidence with auxiliary channels
    - Detector characterization groups within the LVC use tools that statistically rank the coupling between the gravitational wave channel and auxiliary channels to look for the **source** of the unusual glitches

See J. Smith et al. "A hierarchical method for vetoing noise transients in gravitational-wave detectors"

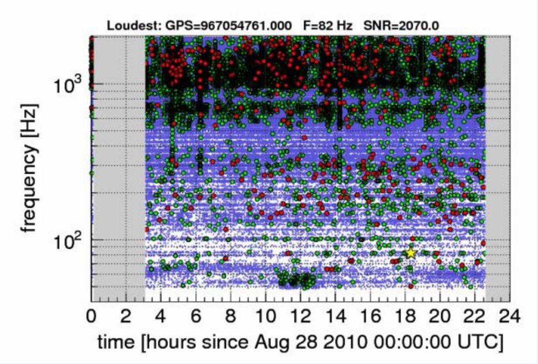
Hunting Glitches

- Useful tools
  - Frequency line finder
  - Weekly statistical results
  - Low-latency burst algorithm (Omega)



Example of Omegascan

- Other Strategies
  - Looking at the SNR of glitches in environmental channels coincident with the gravitational wave data channel
  - Other glitch characteristic studies
    - Example: how do higher SNR glitches behave vs lower SNR



Example of Glitchgram

Data Quality Flags

- **Data quality (DQ) flags** mark periods of time when a detector is suffering from a known problem or a period of suspect data quality

- Flags are assigned categories that dictate how the flag is to be used in analysis
  - Category 1 – Serious problem with the detector occurred – remove this data before analyzing
  - Category 2 – Known problem with the detector occurred – remove this data before searching for signals
  - Category 3 – Indicates an incompletely understood statistical correlation– don't use these times for "clean" data

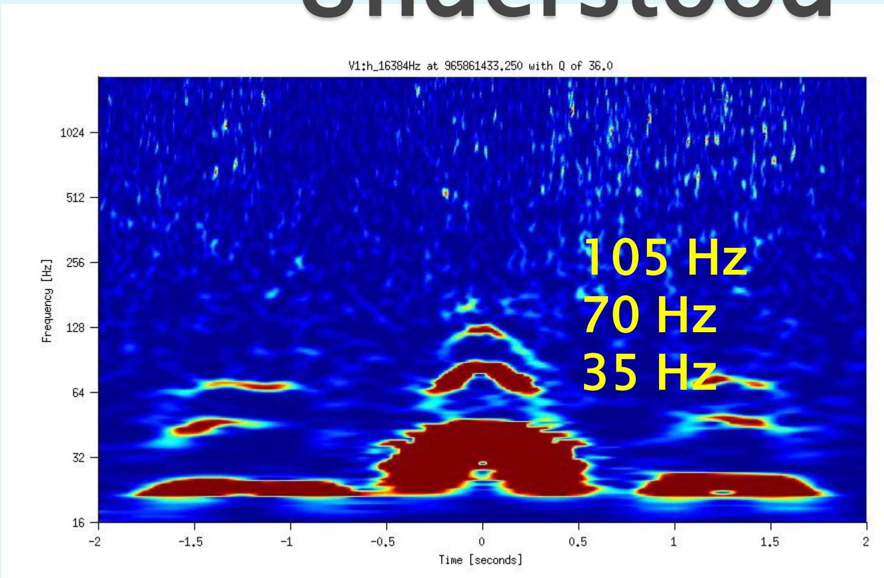
Evolution of Data Quality Flags

- Some automatically generated DQ flags are a constant presence
  - Example:
    - High wind (wind speed over a certain threshold)
- Some DQ flags are tailor-made to individual detectors or certain glitch classes
  - Examples:
    - Glitch-rate (rate of glitches over a certain SNR over a certain frequency threshold)
    - "SeisVeto" (tuned search algorithm to flag low-frequency seismic noise)
- Tailor-made flags tend to be more effective, until the glitch population shifts again

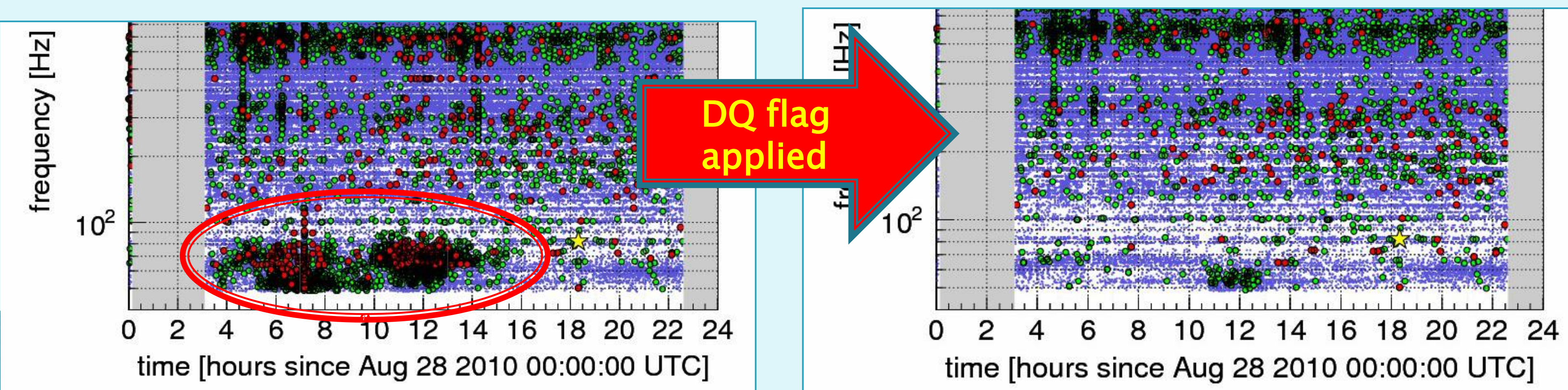
Evolution of Data Quality Flags

- Online Veto Production
  - Virgo uses the same tools and architecture developed for data acquisition for online veto production.
  - LIGO data acquisition and flag generation were separate during eLIGO
- Day or week latency vetoes:
  - Produced by statistical ranking algorithms and based on triggers produced by the a burst search algorithm. Output segments are stored offline in a database.

An Example of Well Understood Virgo Glitches

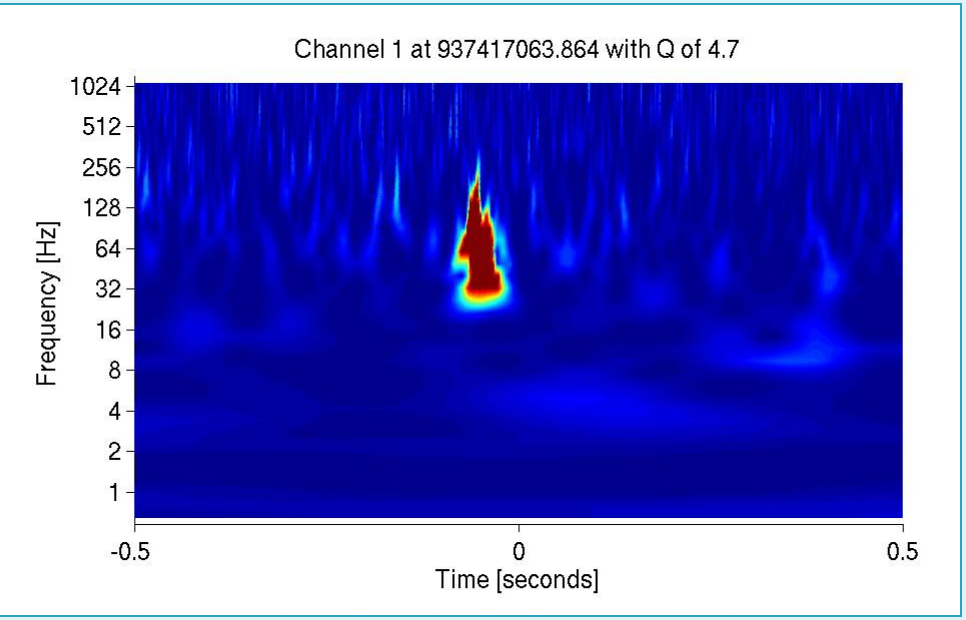


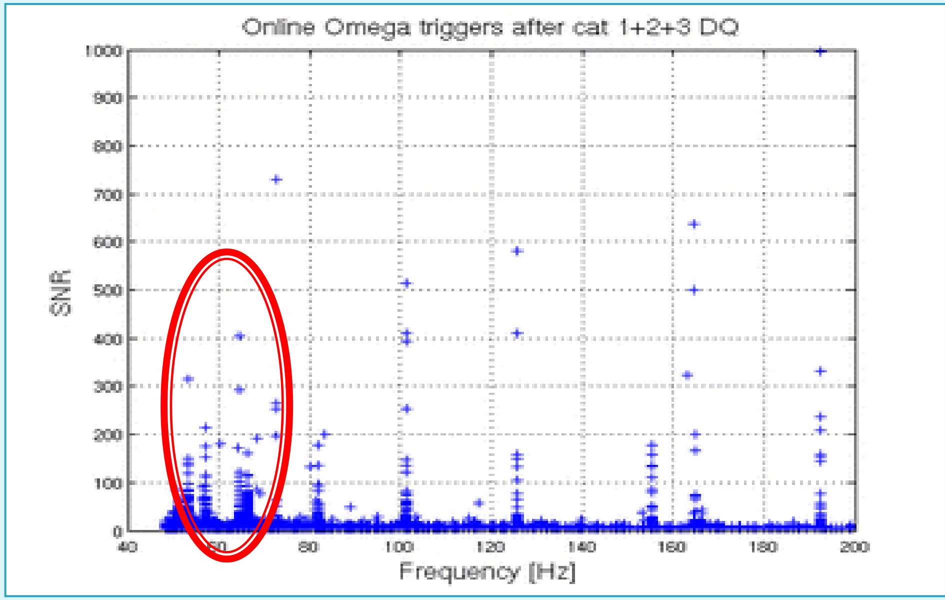
Micro-seismic <-----> scattered light
The stronger the "shaking", the more arches are seen.
Virgo has a very good seismic data quality flag
but with a large dead-time



An Example of Virgo Glitches of Unknown Origin

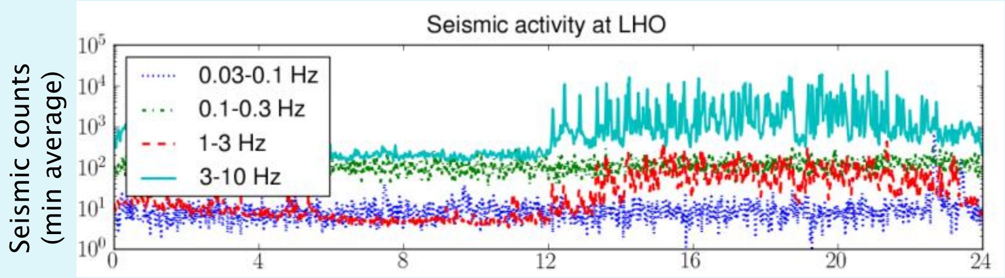
They cover the full VSR2 run with a peak in September 2009:
- They are usually loud for Virgo (SNR>15)
- No auxiliary channel found containing glitches in coincidence useful to produce vetoes against those glitches



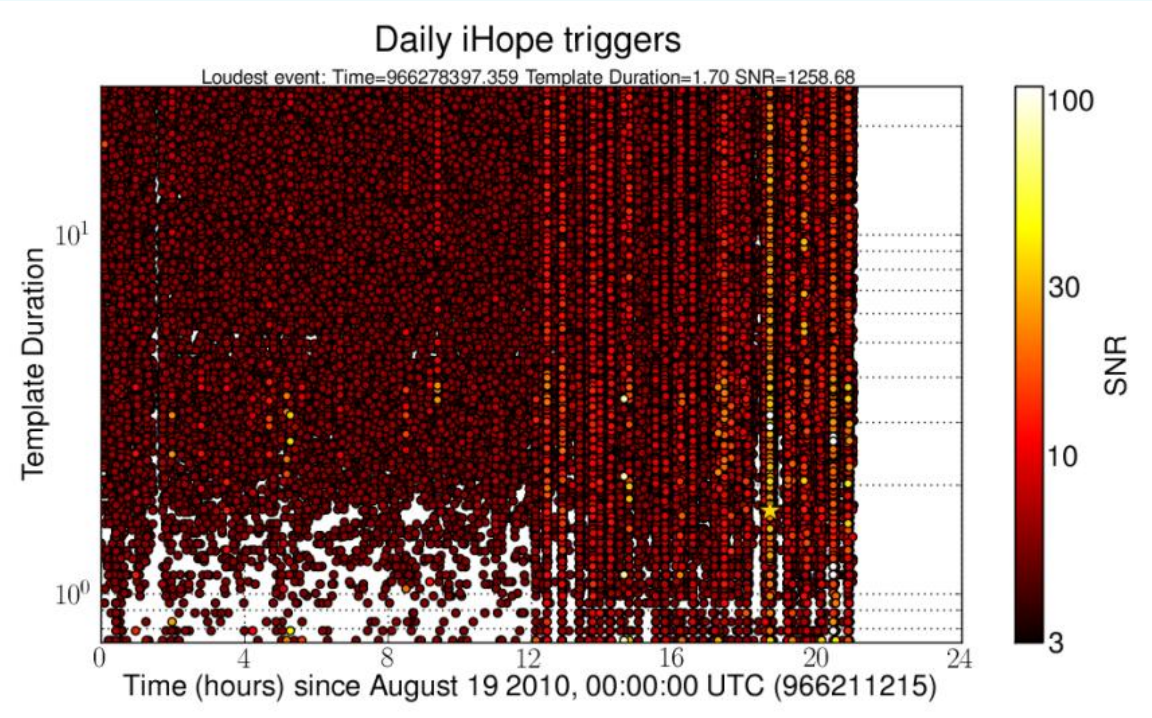


An Example of Well Understood eLIGO Glitch: Seisveto

Time series showing seismic noise during a weekday at the Hanford detector



Time series showing low-frequency glitches created by this seismic noise affecting the entire spectrum of inspiral templates in the low mass compact binary coalescent search template bank.

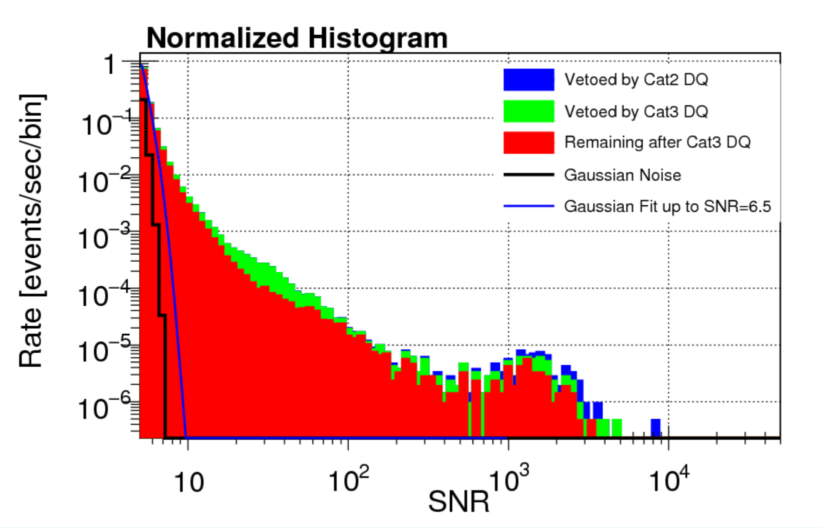


Macleod et al. 'Removing the effect of seismic noise from LIGO data by targeted veto generation'. In preparation.

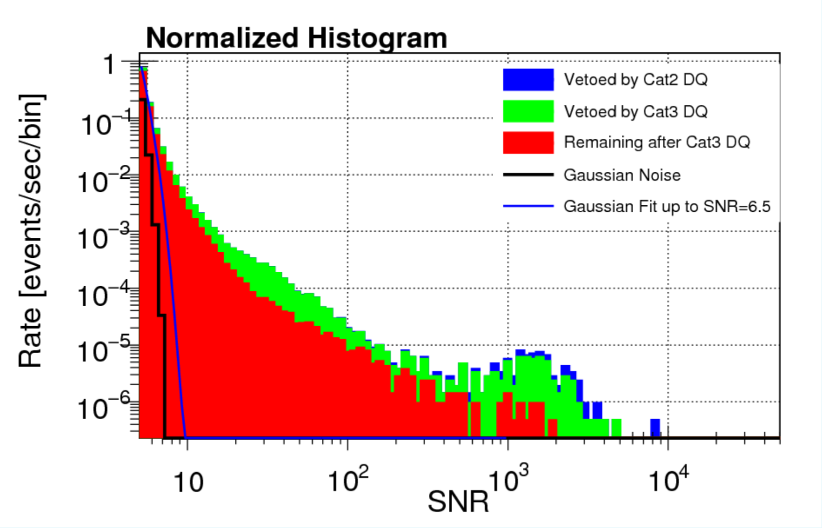
Example of eLIGO glitch well understood

- The "Seisveto" data quality flag targeted these seismic glitches

A rate histogram of Hanford detector data during S6D without Seisveto applied

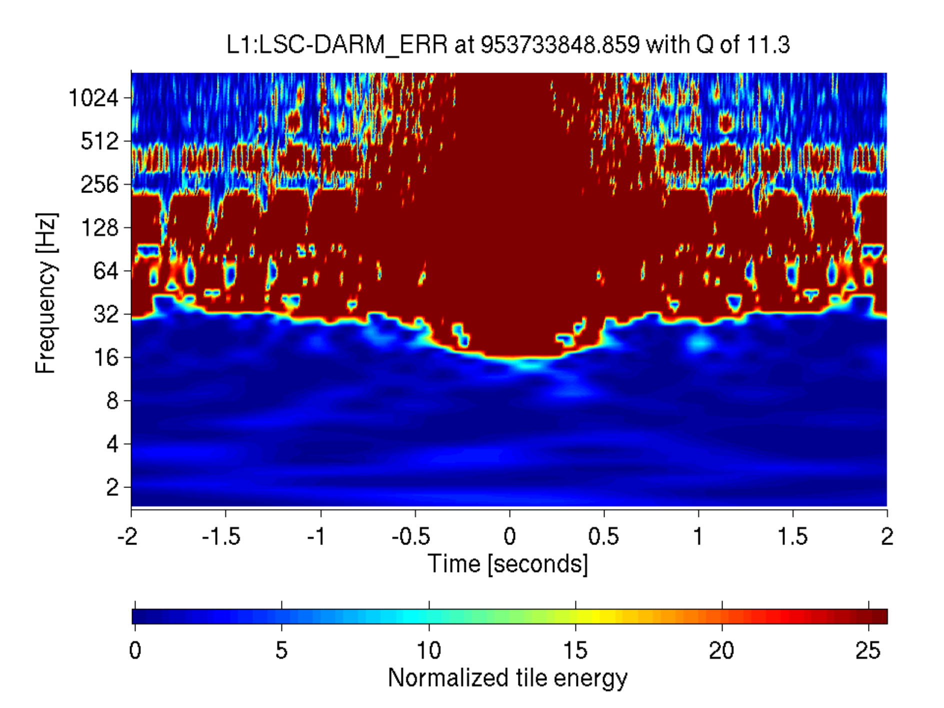


Same data as left plot with Seisveto applied as a cat3 DQ veto



Example of eLIGO glitch of unknown origin: spike glitches

- Biggest issue in S6 for Livingston
- Common and loud – problematic for low mass compact binary coalescent search



Virgo DQ Flag Performances

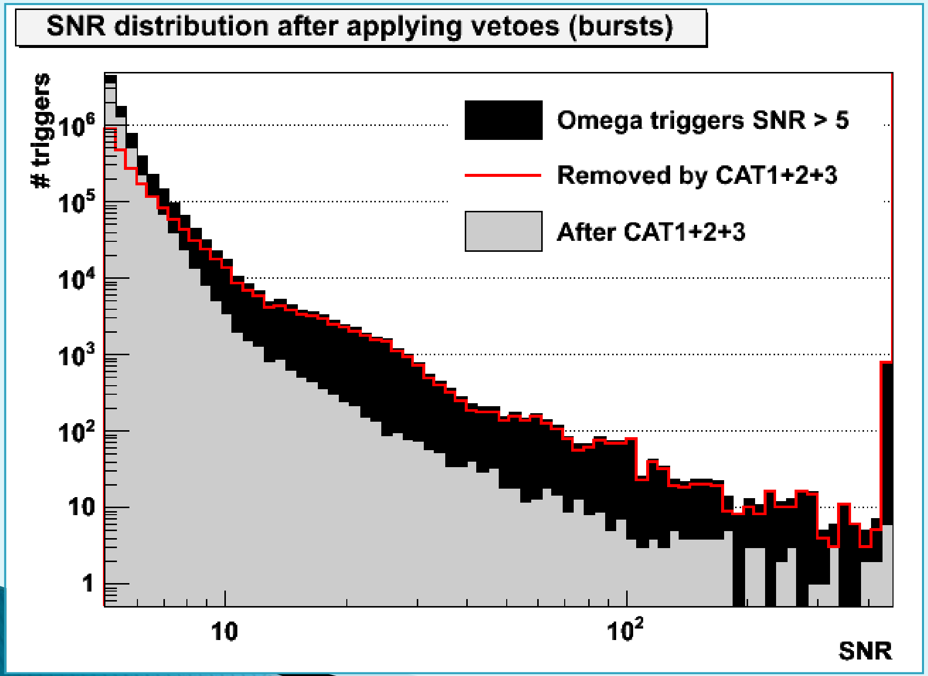
Overall performances of Burst DQ vs Omega triggers with SNR>8

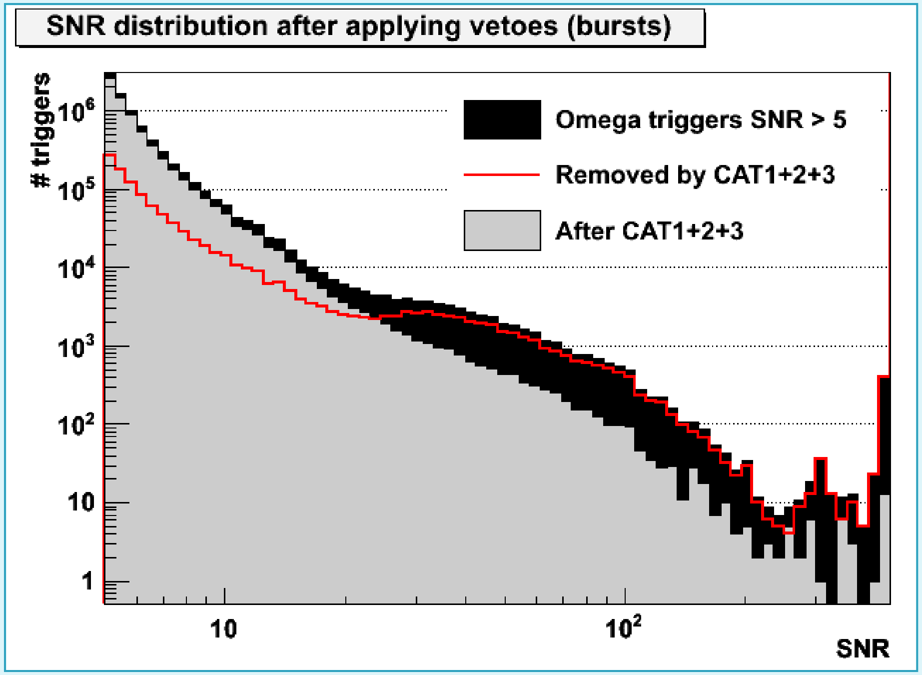
VSR2

	cat1+2	cat1+2+3
Efficiencies:	61.7%	74.6%
Deadtimes:	5.0%	8.1%

VSR3

	cat1+2	cat1+2+3
Efficiencies:	19.3%	27.1%
Deadtimes:	4.2%	8.8%

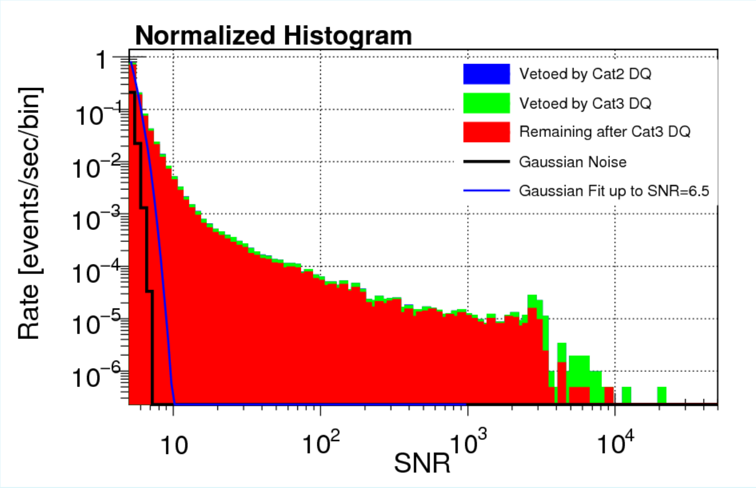


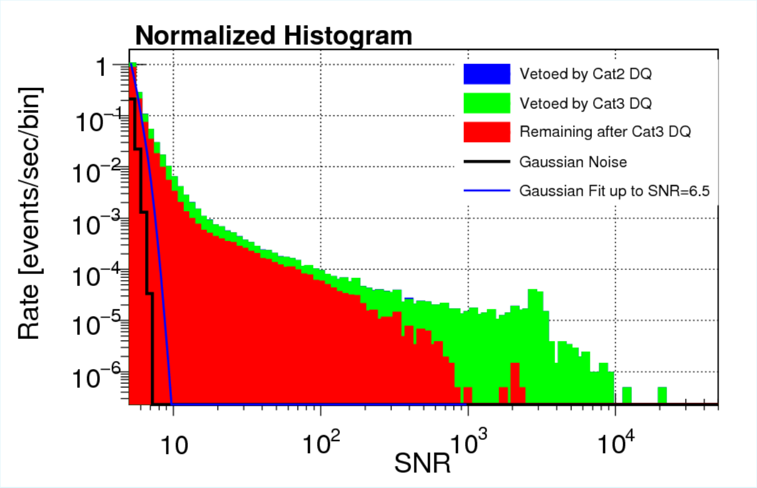


S6 DQ Veto eLIGO Summary

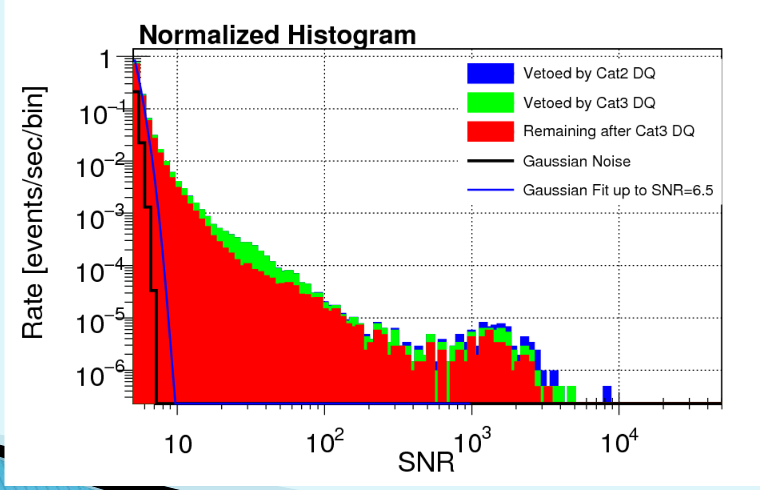
With only pre-S6D (online) DQ vetoes applied | With DQ vetoes tailor-made for S6D applied

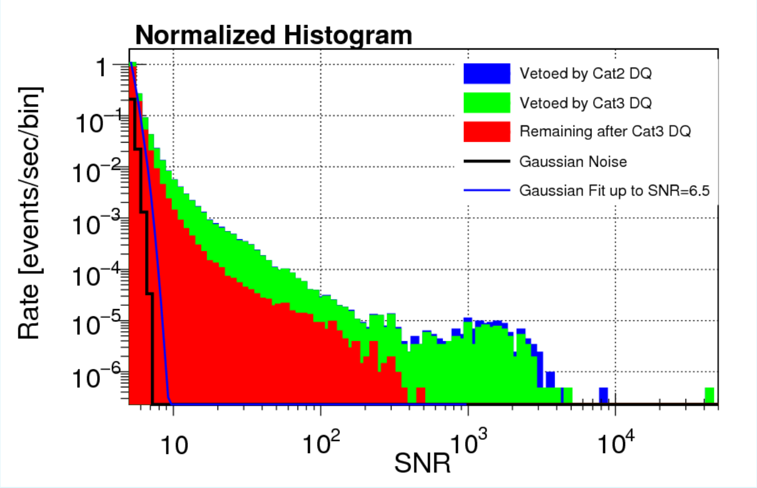
Livingston rate histograms during S6D





Hanford rate histograms during S6D





S6 DQ Veto eLIGO Summary

S6D final statistics

Hanford

Cat2

DeadTime:	7677/	6365144	= 0.121%
SNR >5.0e+00	7301/	3257728	= 0.224%
SNR >1.0e+01	1033/	42833	= 2.412%
SNR >2.0e+01	282/	8753	= 3.222%
SNR >5.0e+01	80/	1752	= 4.566%
SNR >1.0e+02	76/	727	= 10.454%
SNR >1.0e+03	43/	180	= 23.889%

Cat3

DeadTime:	678951/	6365144	= 10.667%
SNR >5.0e+00	802821/	3257728	= 24.644%
SNR >1.0e+01	34435/	42833	= 80.394%
SNR >2.0e+01	7403/	8753	= 84.577%
SNR >5.0e+01	1409/	1752	= 80.422%
SNR >1.0e+02	616/	727	= 84.732%
SNR >1.0e+03	180/	180	= 100.000%

Livingston

Cat2

DeadTime:	8619/	6397996	= 0.135%
SNR >5.0e+00	3722/	3316769	= 0.112%
SNR >1.0e+01	227/	45491	= 0.499%
SNR >2.0e+01	61/	12655	= 0.482%
SNR >5.0e+01	20/	4739	= 0.422%
SNR >1.0e+02	8/	2562	= 0.312%
SNR >1.0e+03	0/	563	= 0.000%

Cat3

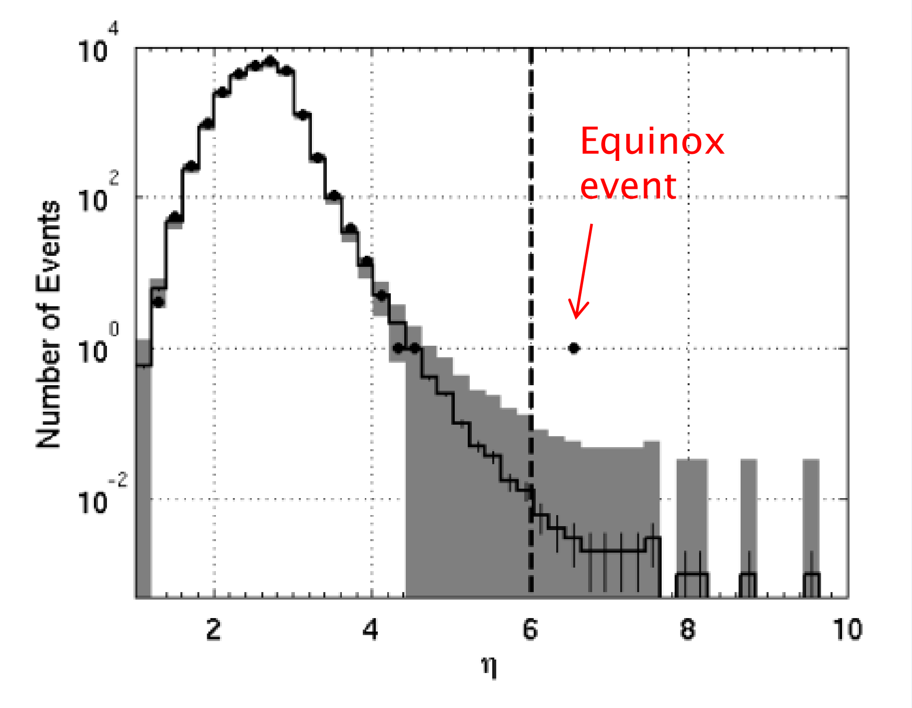
DeadTime:	782001/	6397996	= 12.223%
SNR >5.0e+00	731936/	3316769	= 22.068%
SNR >1.0e+01	21485/	45491	= 47.229%
SNR >2.0e+01	5269/	12655	= 41.636%
SNR >5.0e+01	2479/	4739	= 52.311%
SNR >1.0e+02	1741/	2562	= 67.955%
SNR >1.0e+03	557/	563	= 98.934%

Milestones in enhanced detector era data quality

The equinox event

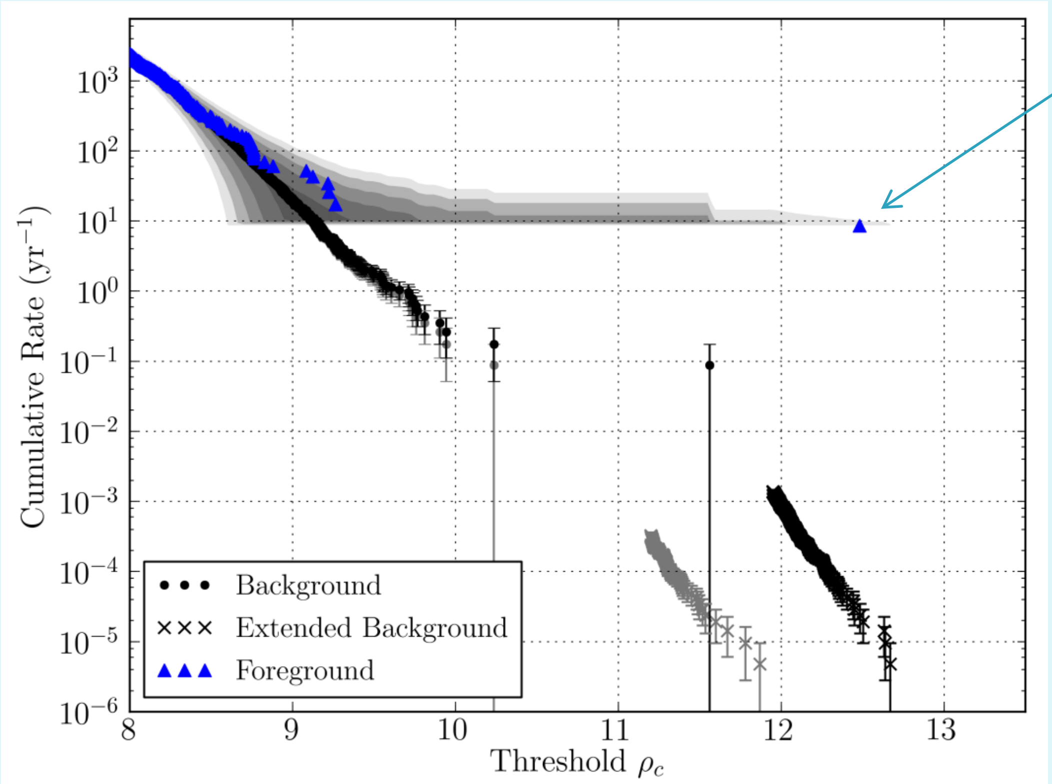
- The “equinox” event was the only event in the coherent wave burst search above threshold in S5
  - Later turned out to be a blind injection

DQ’s role in lack of confidence in a detection:
The glitch rate with data quality vetoes applied was still too high to make a confident detection



"Big Dog" Blind Injection

The good news: the low mass CBC search was able to set a very low false alarm rate on this blind injection



Virgo: Summary and Next Steps

- Fairly good results in VSR2 and VSR3

- but several unexplained and/or unvetoed loud glitches remain

- VSR4 run has started since June 3rd 2011, in coincidence with GEO detector
  - Glitchiness back to VSR2 level ~ 0.7 Hz
  - Detchar activities concentrate on low frequency (<50 Hz) where noisy lines may prevent pulsar searches and high frequency (>800 Hz) where glitches may pollute GEO-Virgo coincident burst searches
  - VSR4 will be also useful to test new ideas for vetoes in Adv. Virgo (glitch families, vetoes categorization, information about non-stationary frequency lines much more exploited…)

- More details about Virgo Data Quality and glitch investigations in http://wwwcascina.virgo.infn.it/DataAnalysis/VDBdoc

- More details about noise monitoring and spectral lines investigations in specific Virgo talks at this conference.

The advanced detector era

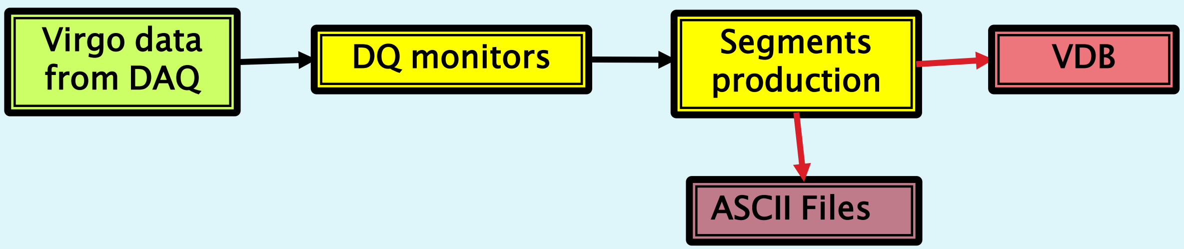
- Low-latency detector characterization enabled us to rapidly point telescopes during the latest science run
- In the advanced detector era, the detector characterization groups at LIGO and Virgo are preparing for whatever new populations of glitches we may see, and are improving DQ tools for lower latency DQ
- See Duncan Macleod’s poster: “Detector Characterisation for the Advanced LIGO detectors”

Questions?

Additional Information

Virgo DQ Flags
Online production and access

- **Virgo data are sent via Ethernet to DQ monitors**
- **DQ monitors produce DQ flags sent online to the DQ segments production**
- **DQ flags segments are stored in a database (VDB) and in ASCII files**



- **VDB provides also**
  - VDBtk_segments : a line command to access segments lists in VDB
  - VDB web interface : a web interface to access segments lists in VDB (https://vdb.virgo.infn.it/VDB/main.php)

KW-based vetoes

KW-based vetoes (UPV and hveto) have low deadtime and allows to get rid of few loud glitches not vetoed by the DQ flags

UPV on all mbta triggers

		eff	use perc.	eff/dt
SNR> 5.0	42033 / 1534162 =	2.7 %	45.9 %	2.4
SNR> 8.0	6785 / 27903 =	24.3 %	13.4 %	21.5
SNR> 15	1659 / 5342 =	31.1 %	3.6 %	27.4
SNR> 30	145 / 1257 =	11.5 %	0.3 %	10.2
deadtime:	146230 / 12903048 =	1.133 %		

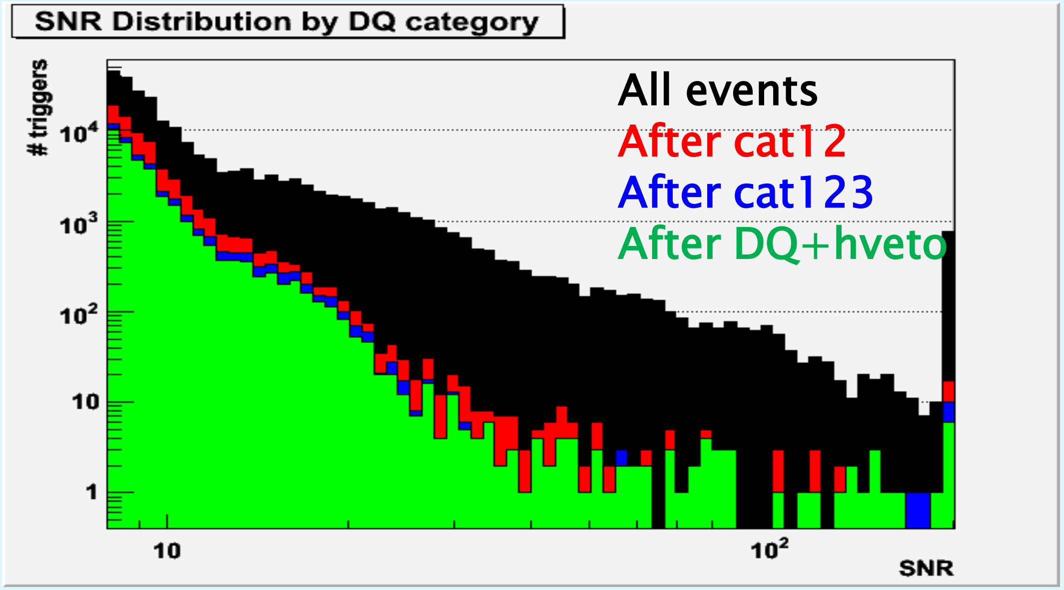
UPV on mbta triggers after cat123

		eff	use perc.	eff/dt
SNR> 5.0	6123 / 963256 =	0.6 %	9.6 %	0.56
SNR> 8.0	75 / 4152 =	1.8 %	0.2 %	1.59
SNR> 15	13 / 627 =	2.1 %	0.0 %	1.83
SNR> 30	9 / 148 =	6.1 %	0.0 %	5.37
deadtime:	146230 / 12903048 =	1.133 %		

hveto :

	eff	use percentage
SNR> 8.0	3.2 %	20.2 %
SNR> 15	1.7 %	2.0 %
SNR> 30	1.3 %	0.3 %

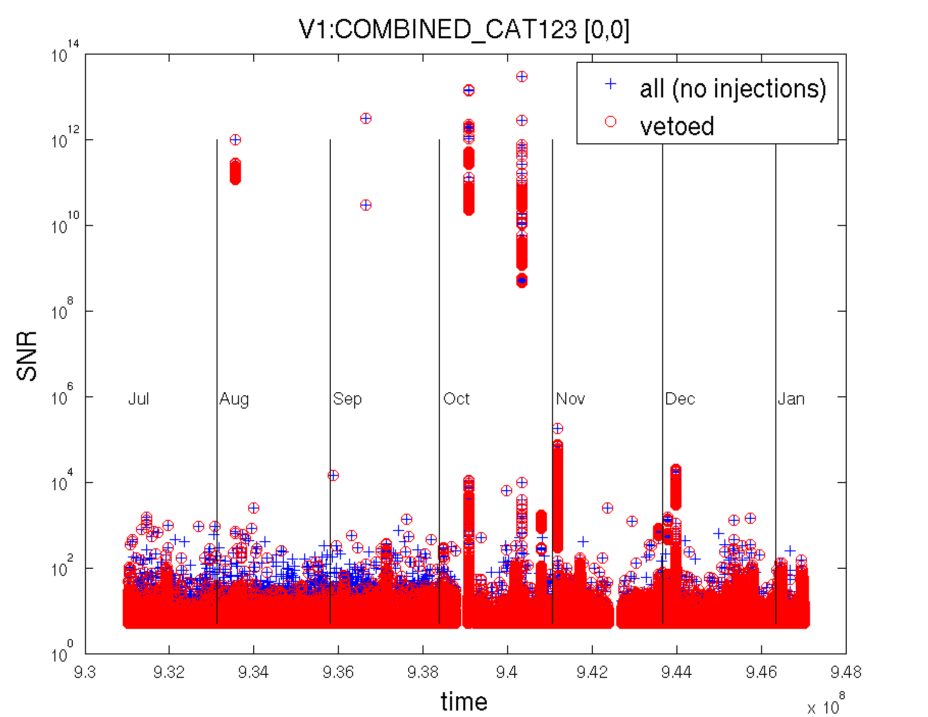
Deadtime=0.06%

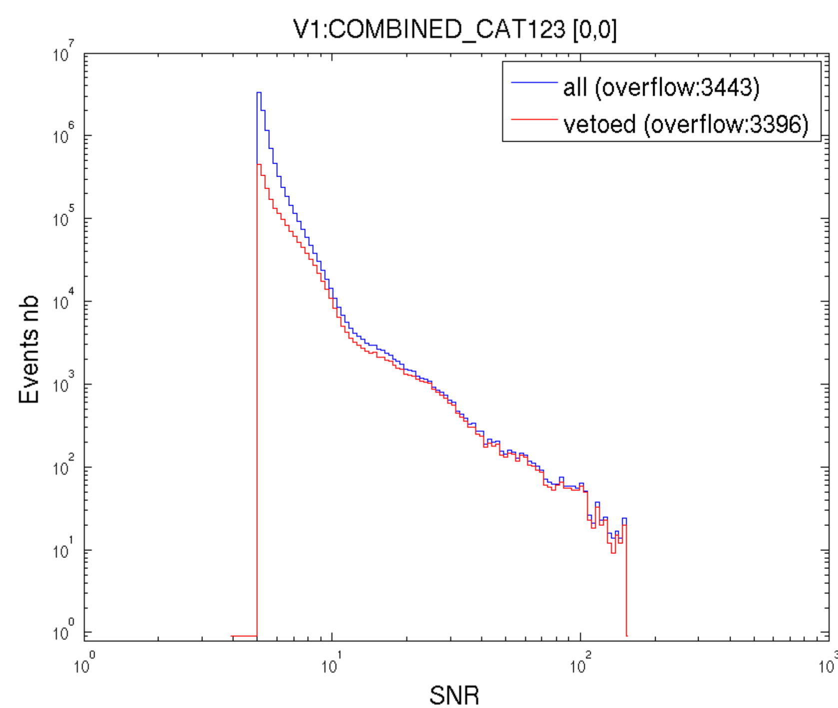


DQ performances on Omega triggers

Cat1+Cat2+Cat3 **deadtime=8.1%**

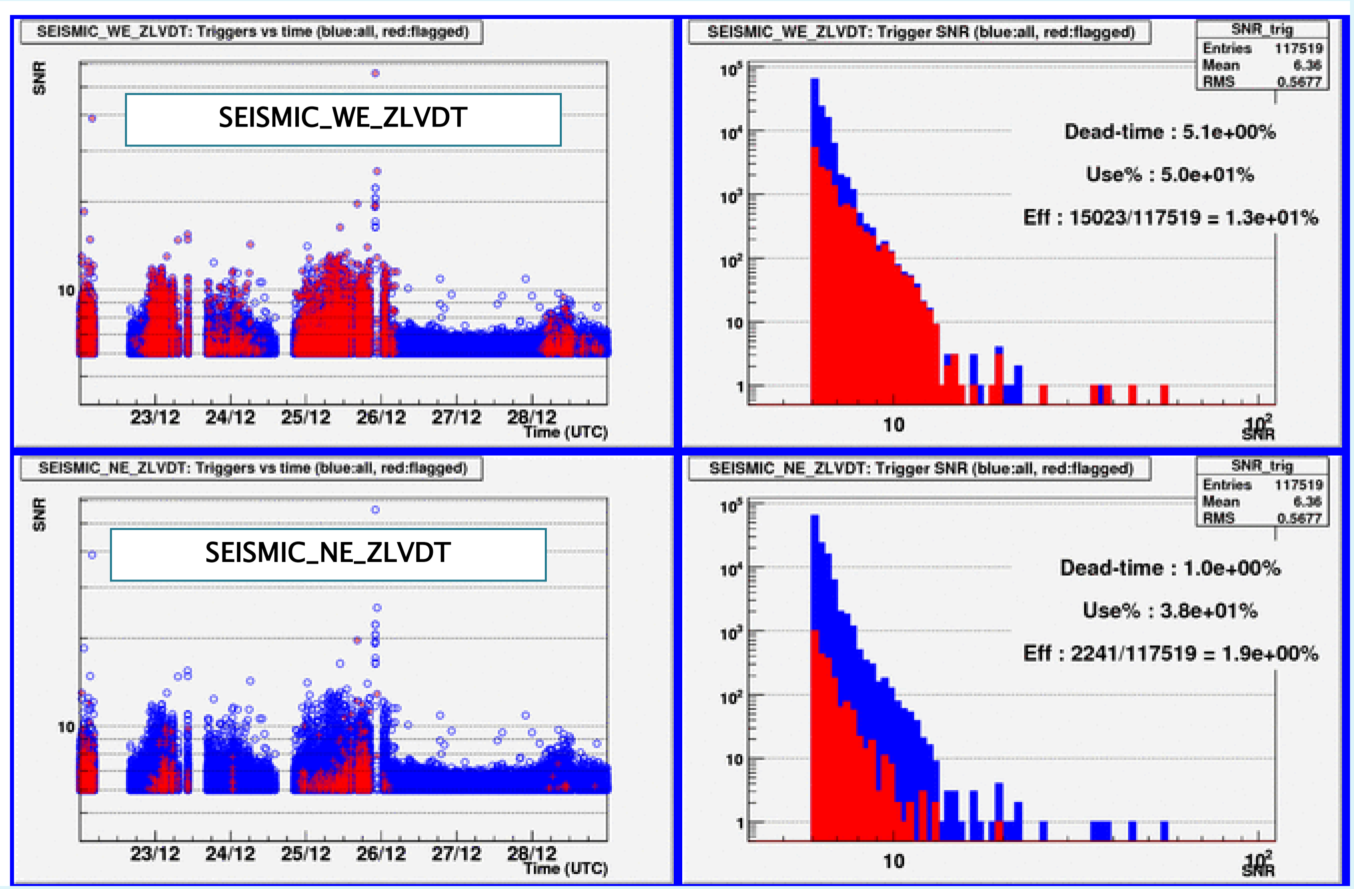
	efficiency	use percentage	efficiency/deadtime
SNR> 5	22.635 %	78.33 %	2.80
SNR> 8	74.644 %	22.26 %	9.22
SNR> 15	88.635 %	4.81 %	10.95
SNR> 30	94.018 %	0.60 %	11.62



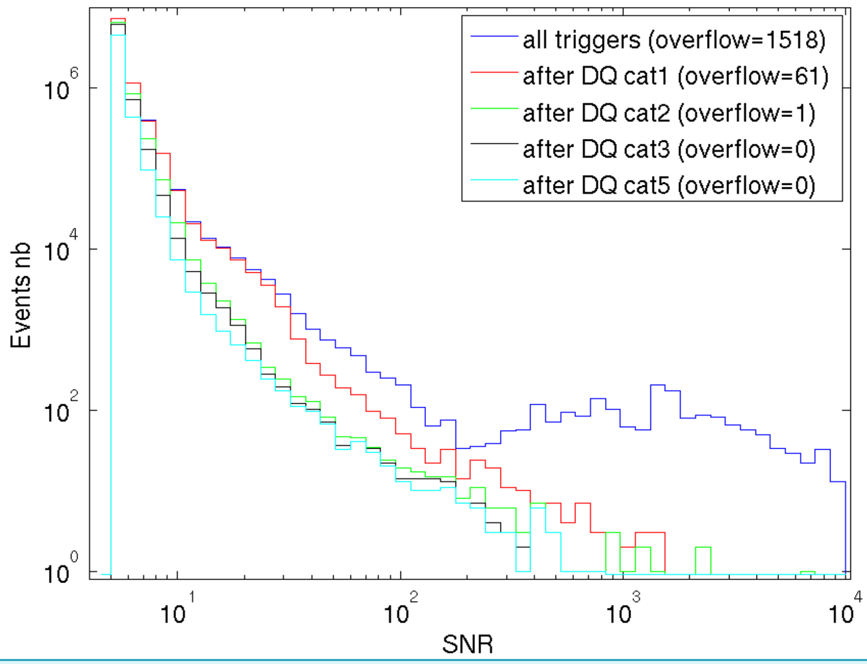


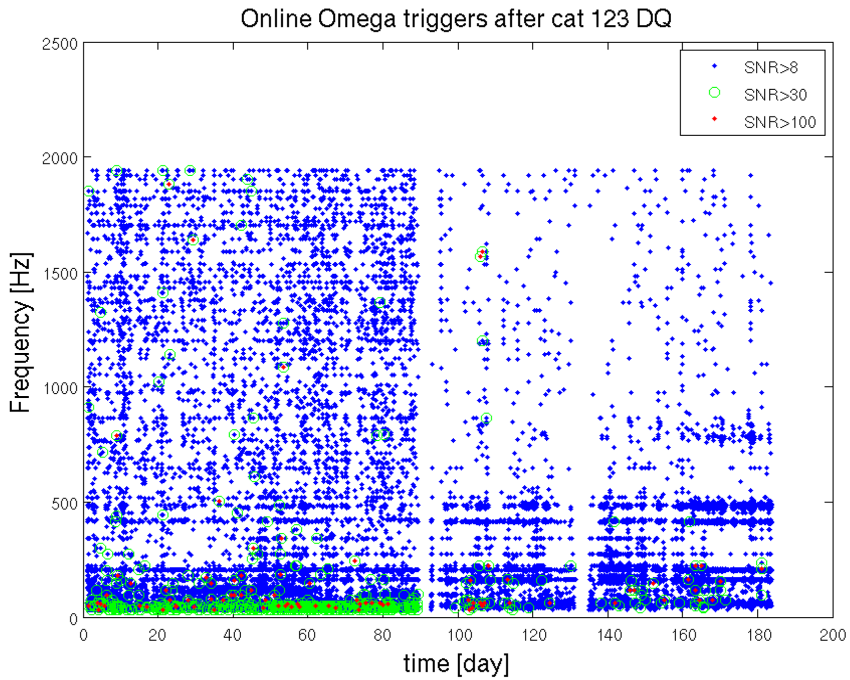
Example of good DQ flags

The SEISMIC flags



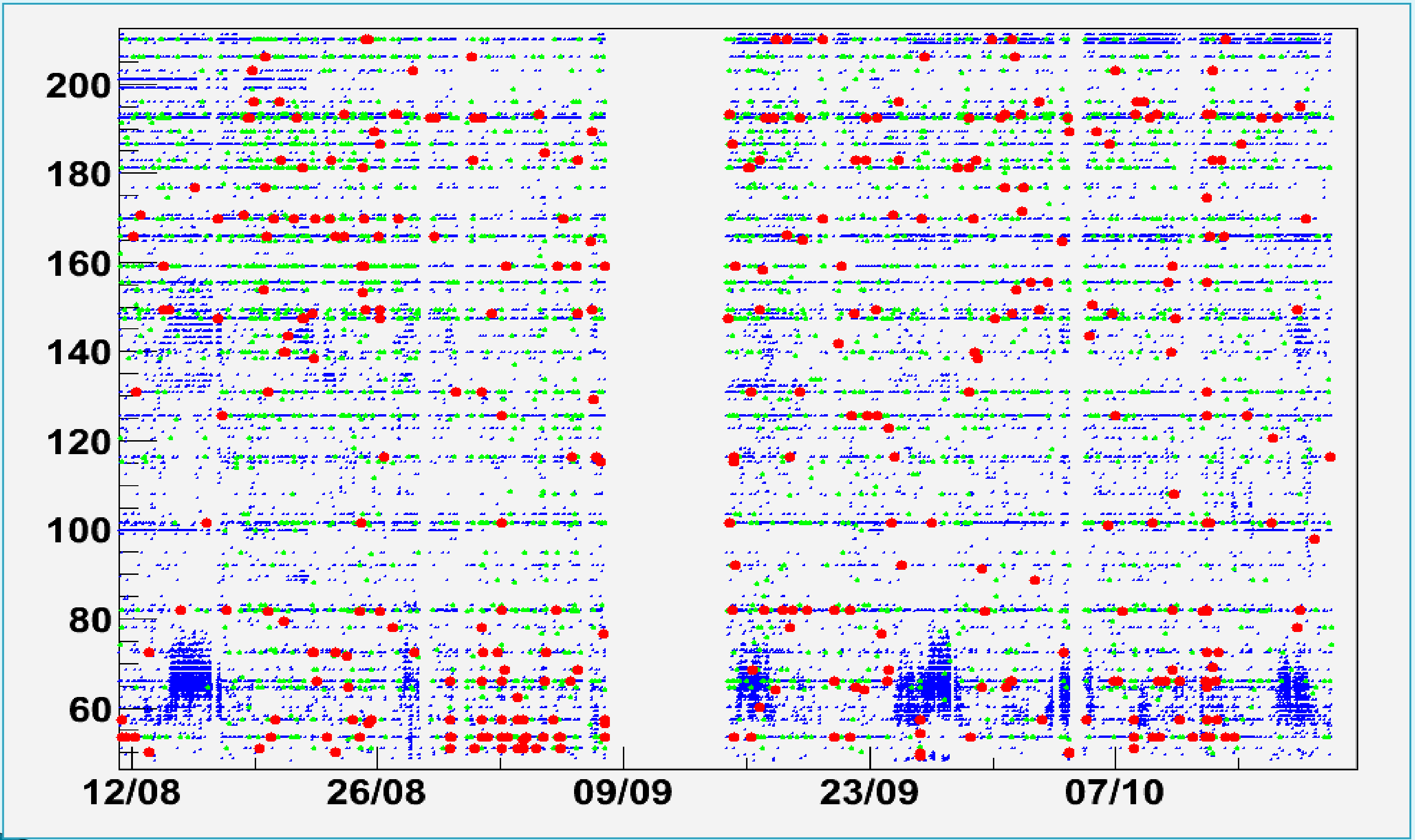
VSR2 remaining glitches





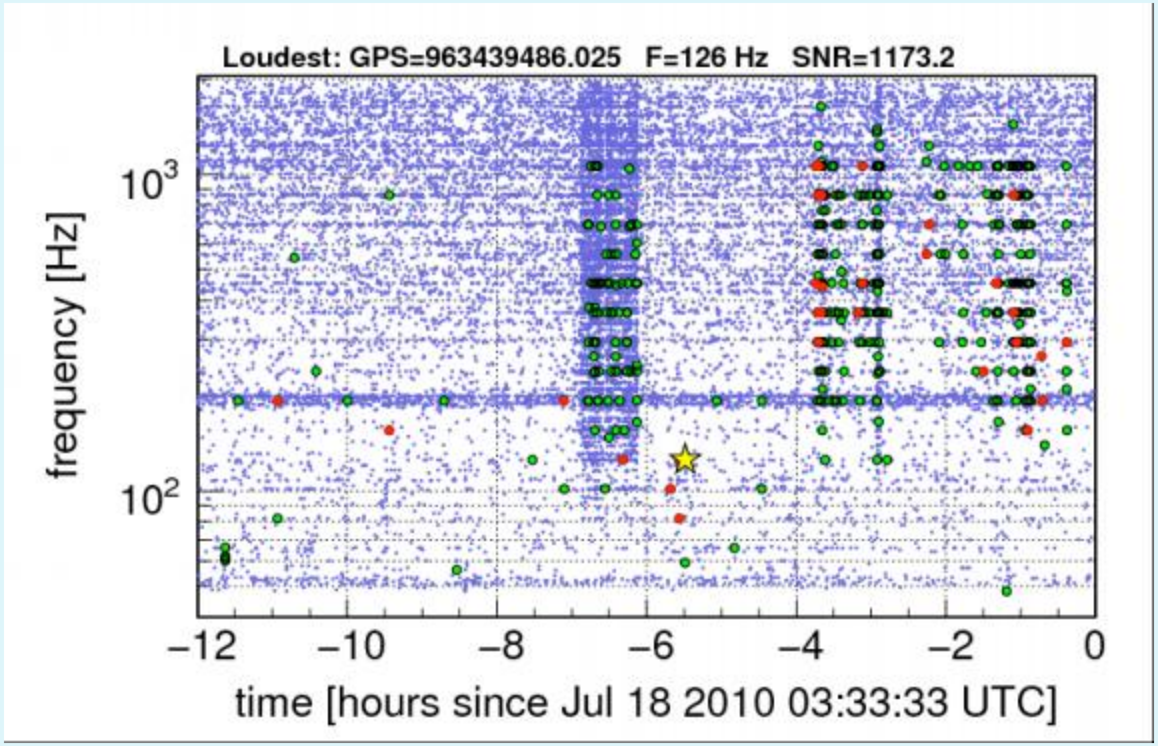
VSR3 Remaining glitches

Low frequency Omega triggers after cat123 (green:SNR>20, red:SNR>50)



A case study: “grid” glitches

- Very frequent glitches plaguing the Hanford detector for about two months late in eLIGO

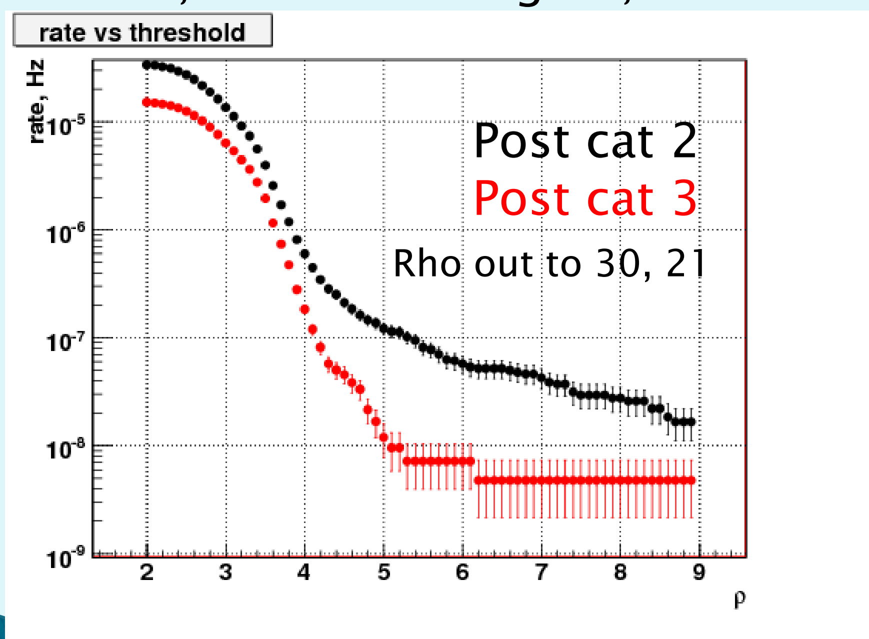


A case study: "grid" glitches

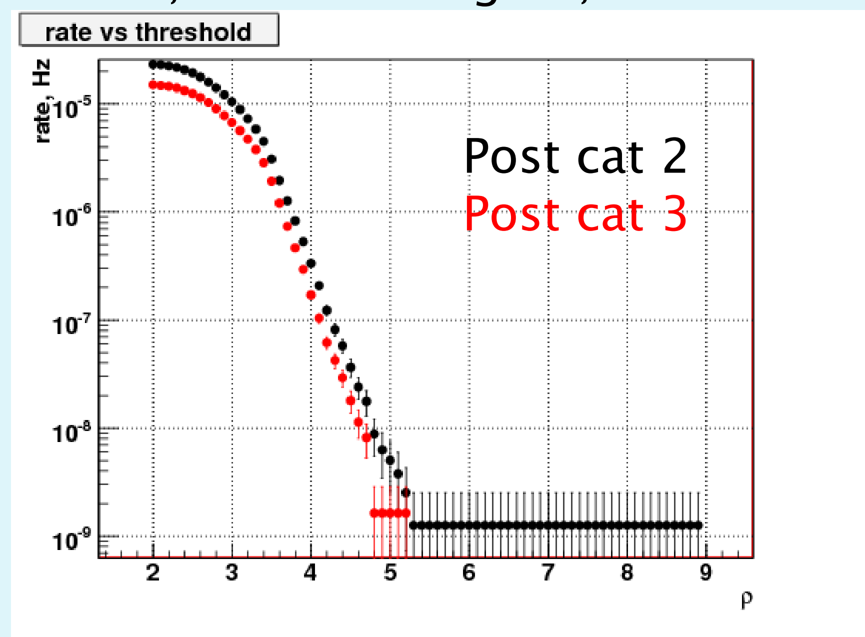
- How grid glitches limited the searches

Impact on the LIGO network coherent wave burst search search

L1, H1 **Before** Aug 17, 2010



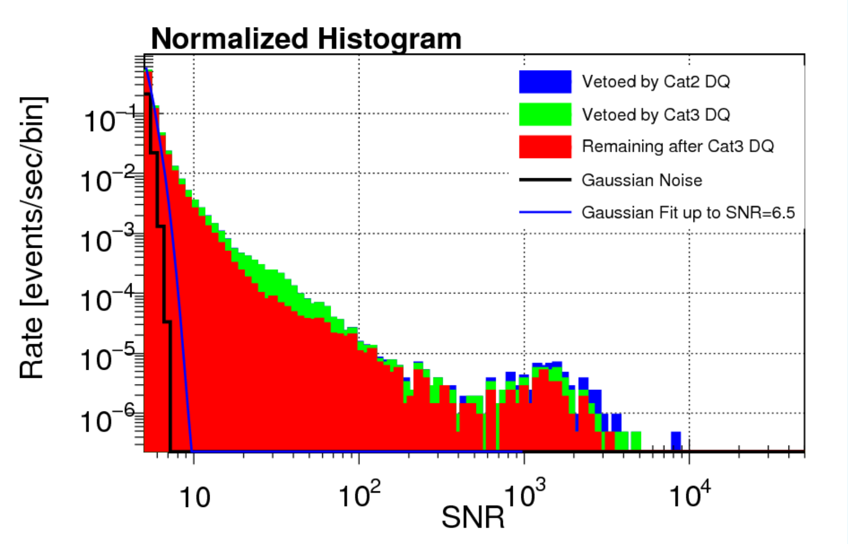
L1, H1 **After** Aug 17, 2010



A case study: "grid" glitches

- The glitch rate flags
  - A series of DQ flags designed to combat grid glitches
    - Search groups adopted flags vetoing times containing glitches over a certain SNR occurring more often than a certain frequency
    - Resolved by re-soldering of Piezo heater driver chassis

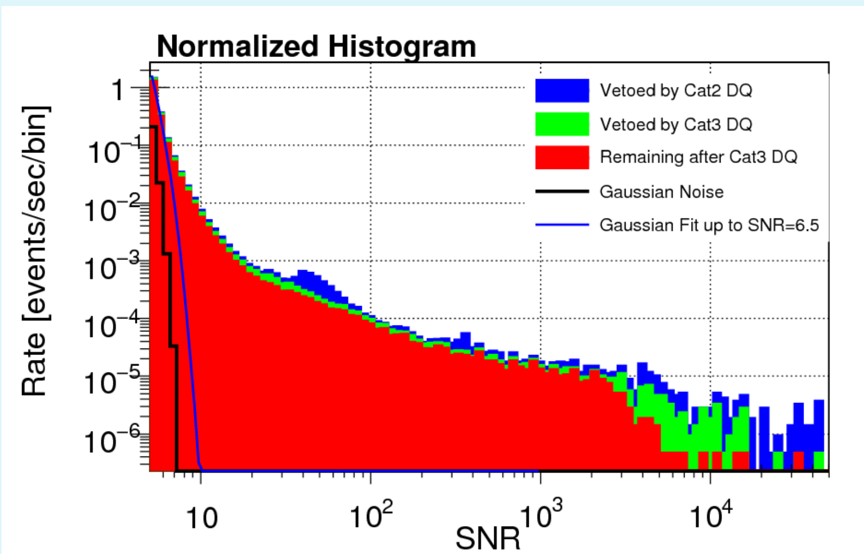
Hanford detector during grid glitches (June 26 – Aug 21)
Histogram of glitches found by a single detector burst search after the burst group adopted a conservative glitch rate flag as a cat 3 veto



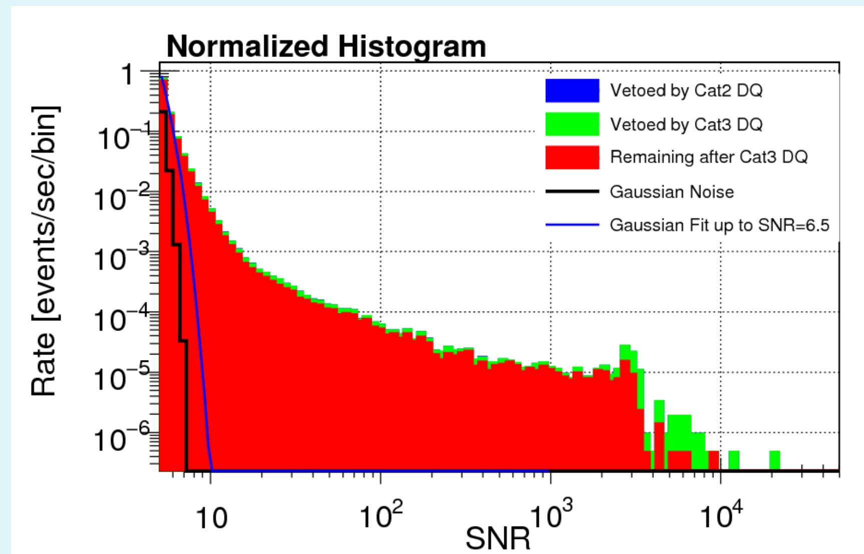
Data Quality Vetoes

- The interferometers are constantly changing
  - We see a major shift in how effective individual DQ flags are after a commissioning period

A rate histogram of the Livingston detector for “S6C” (Jan–May 2010)



A rate histogram of the Livingston detector for “S6D” (June–October 2010)



Seisveto stats

H1:DCH-SEISVETO_CBC:1-cat1cat4-961545615_971654415

DeadTime:		287949/ 6365144	= 4.524%
SNR	>5.0e+00	270131/ 3257728	= 8.292%
SNR	>1.0e+01	8677/ 42833	= 20.258%
SNR	>2.0e+01	2404/ 8753	= 27.465%
SNR	>5.0e+01	792/ 1752	= 45.205%
SNR	>1.0e+02	368/ 727	= 50.619%
SNR	>1.0e+03	139/ 180	= 77.222%