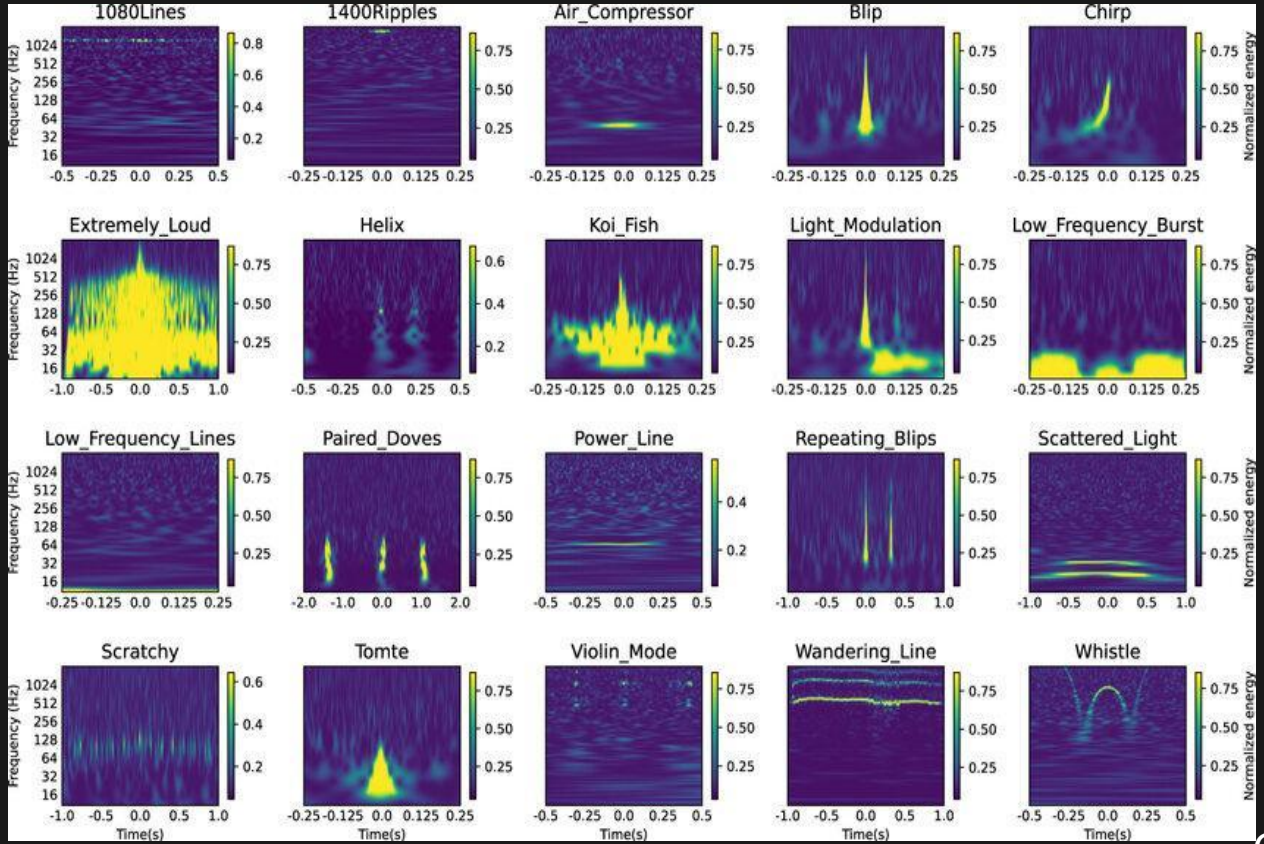
Improve Data Quality in LVK IFOs

Léo Zimmermann : l.zimmermann@ip2i.in2p3.fr
Viola Sordini : v.sordini@ip2i.in2p3.fr
Alexis Boudon : a.boudon@ip2i.in2p3.fr

Table of contents

I- Glitches

Glitches zoology

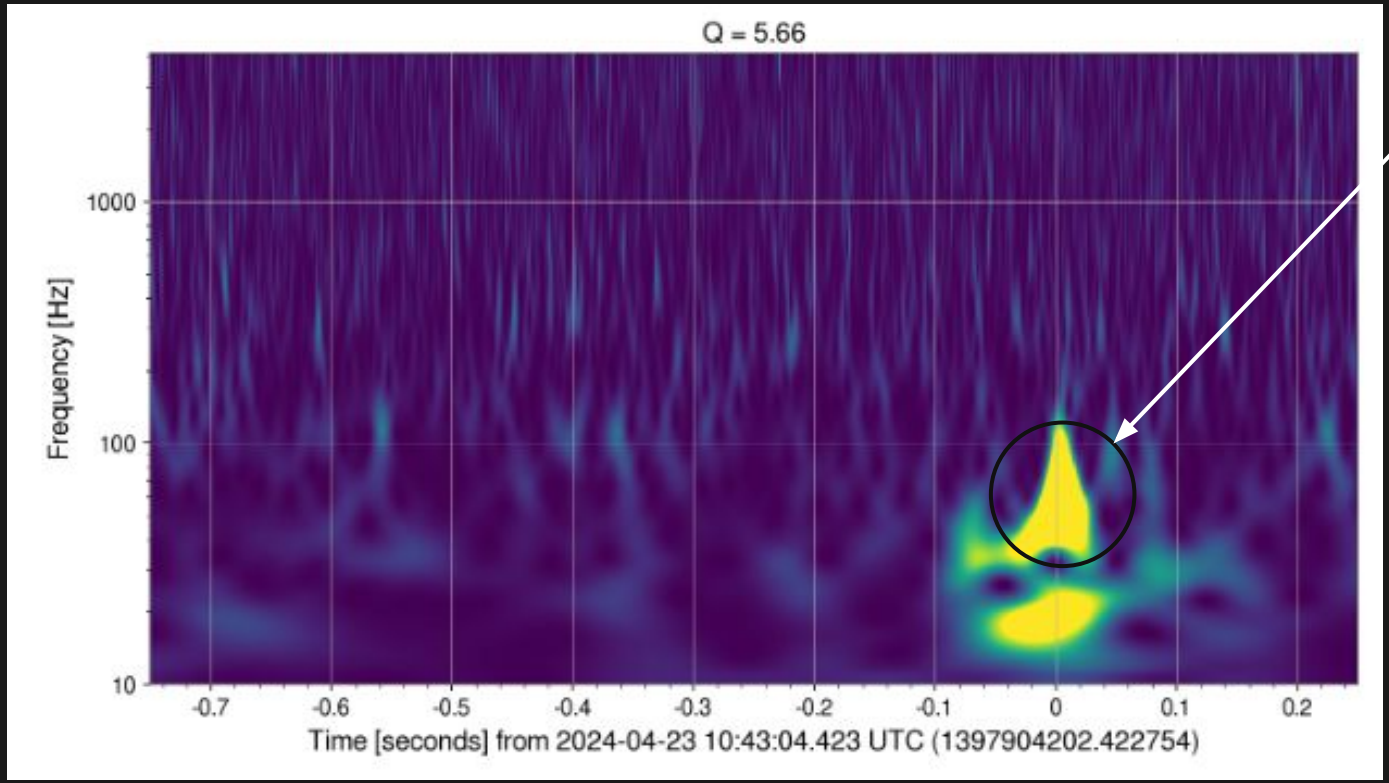


What's a Glitch ?

- -Excess of power in the detector. Not GW
- Many sources : Thunderstorm, Helicopters, light reflection… .
- Glitches classed in 23 classes.

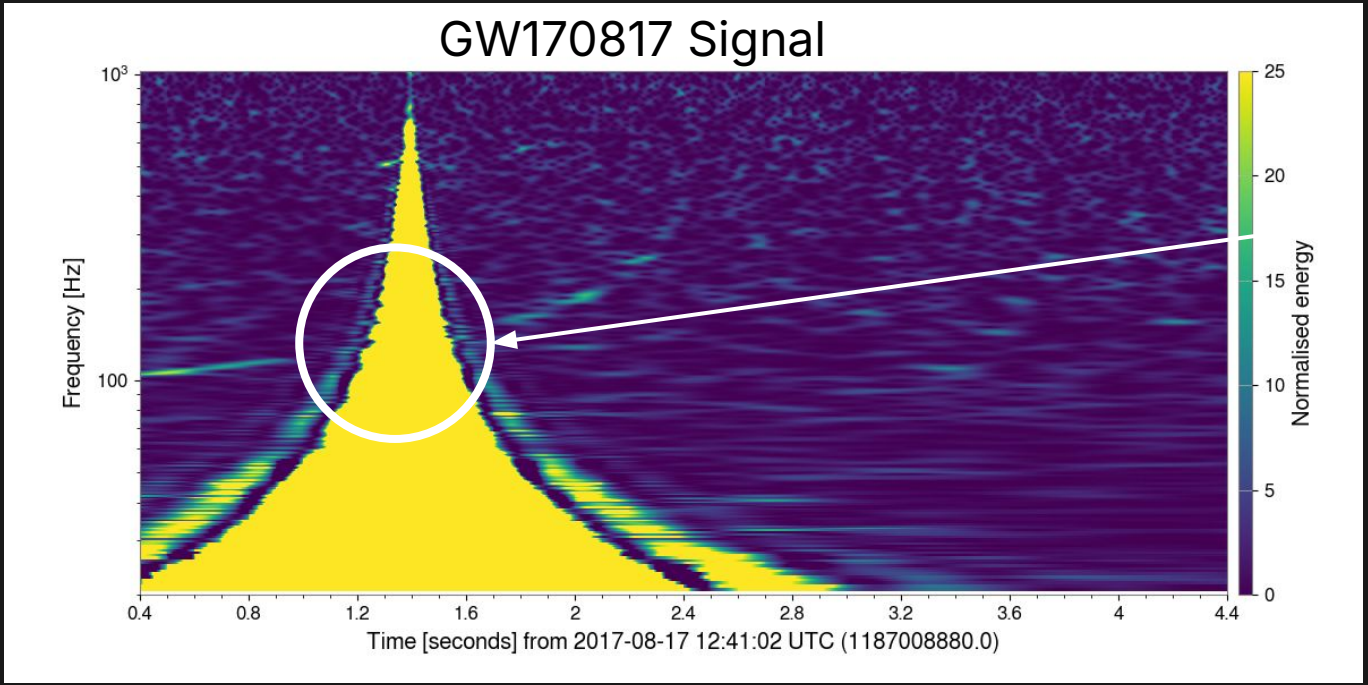
Credit : Fernandes et al 2018

Problem : Mimic Astrophysical sources of GW !



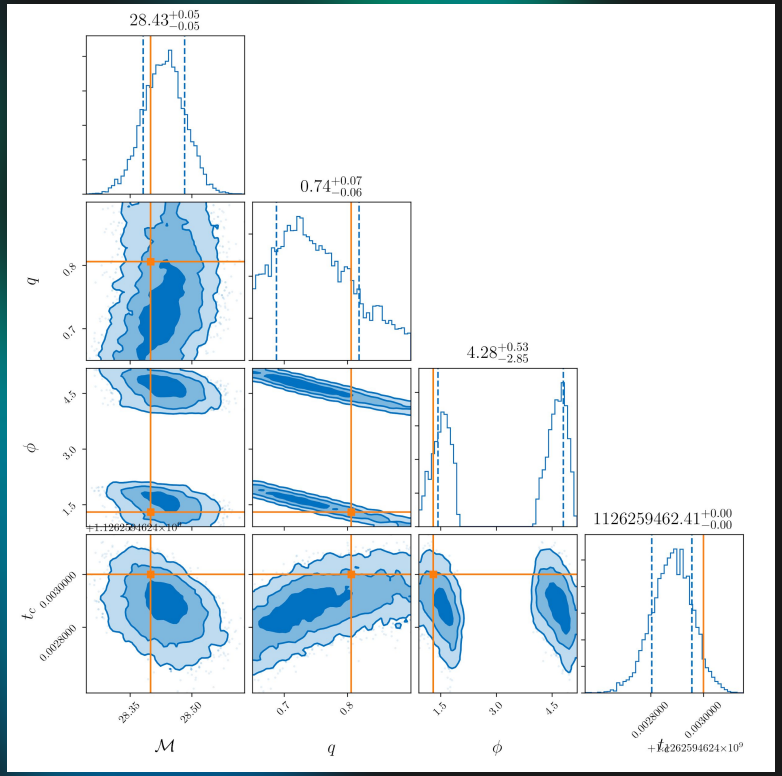
This Glitch has a chirp form which is the typical form of Gws signals

Problem : They can overlap True Signal !



The glitch overlap the GW signal. A good part of the signal is hidden by this glitch.

In O3 ~20% (16/70) of the detected GW where near a Glitch

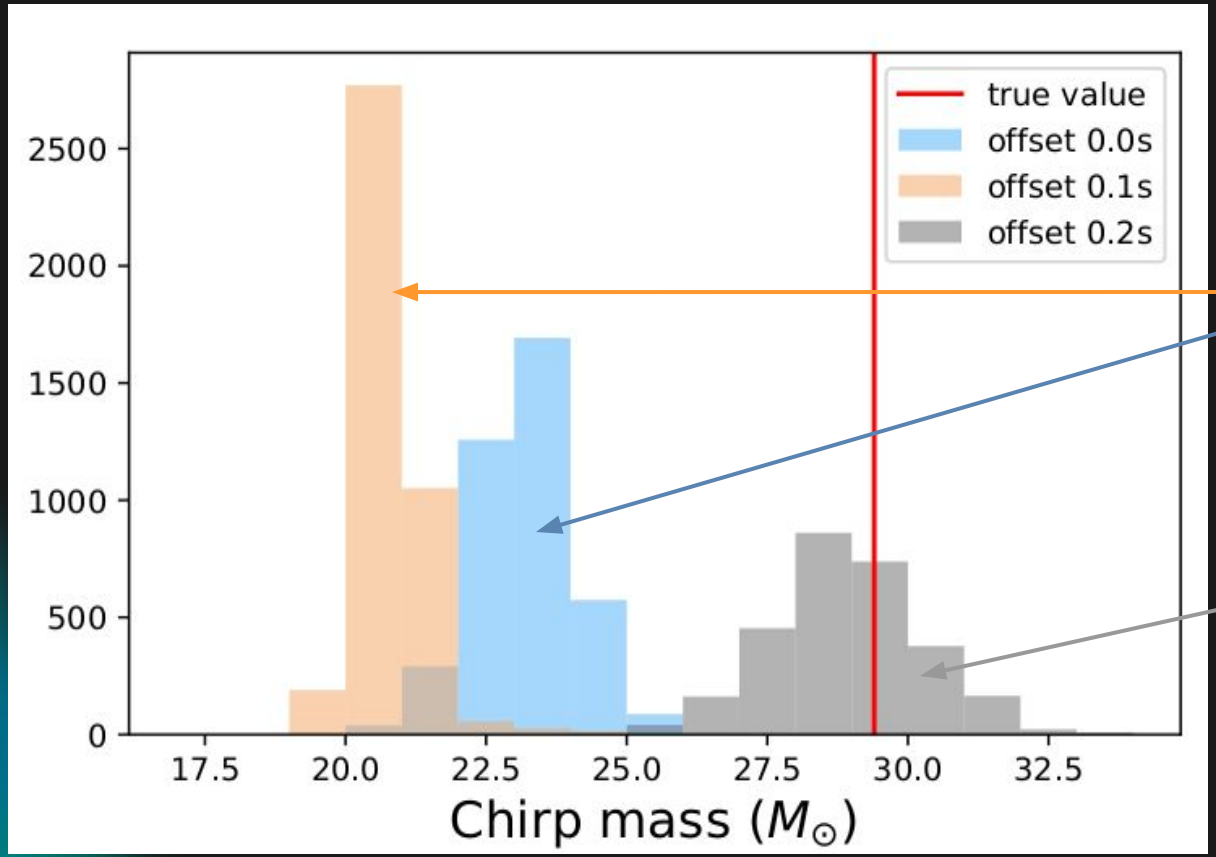


Ⅱ - Impact of noise on observations

a) Impact on $\mathcal{M}$.

a) Impact on Luminosity distance and H0.

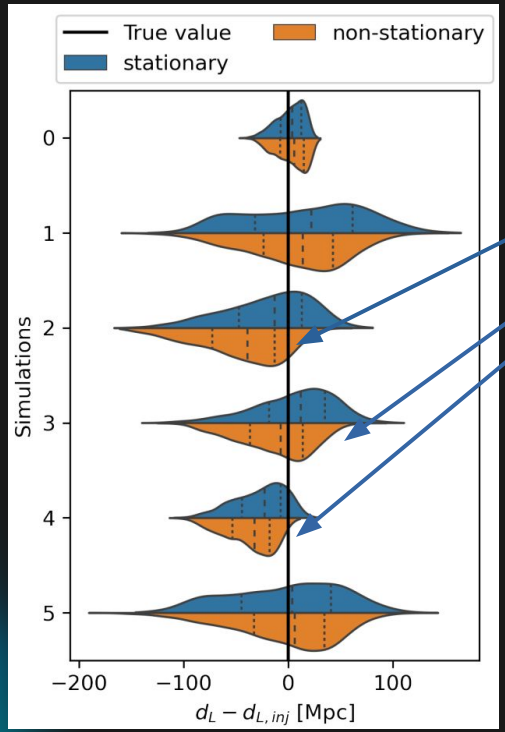
Impact of Glitches on parameters estimation (1/2)



- For a Glitch overlapping the signal, we bias the signal.
- Fortunately, when not overlapping, there is no bias

Credit :Powell 2018

Impact of Glitches on parameters estimation (2/2)



There is a shift towards smaller luminosity distance values.

Systematic under-estimation of the measured luminosity distance by up to 6.8%.

$$H_0 \approx \frac{cz}{d_L}, \; z \text{ small}$$

This lead to a miss estimation of H0 of 4.42σ.

Credit : Mozzon et al 2022

Ⅲ- IDQ : A solution to identify glitches.

a) Basic ideas.

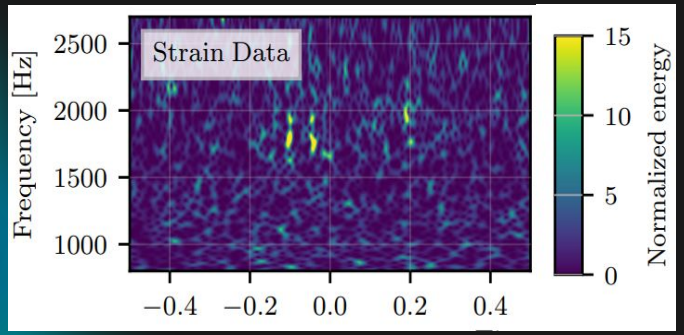
a) First test on O4b 1st week Virgo data.

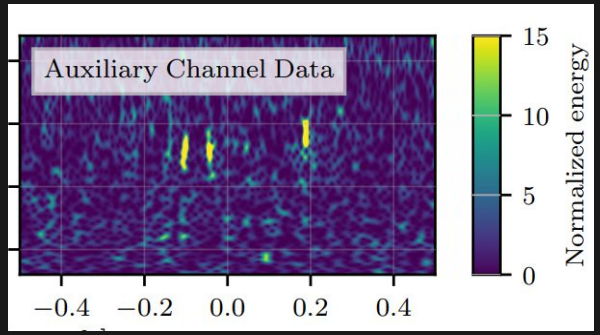
a) First look at 25min glitches.

Auxiliary Channels in LVK IFOs (1/2)

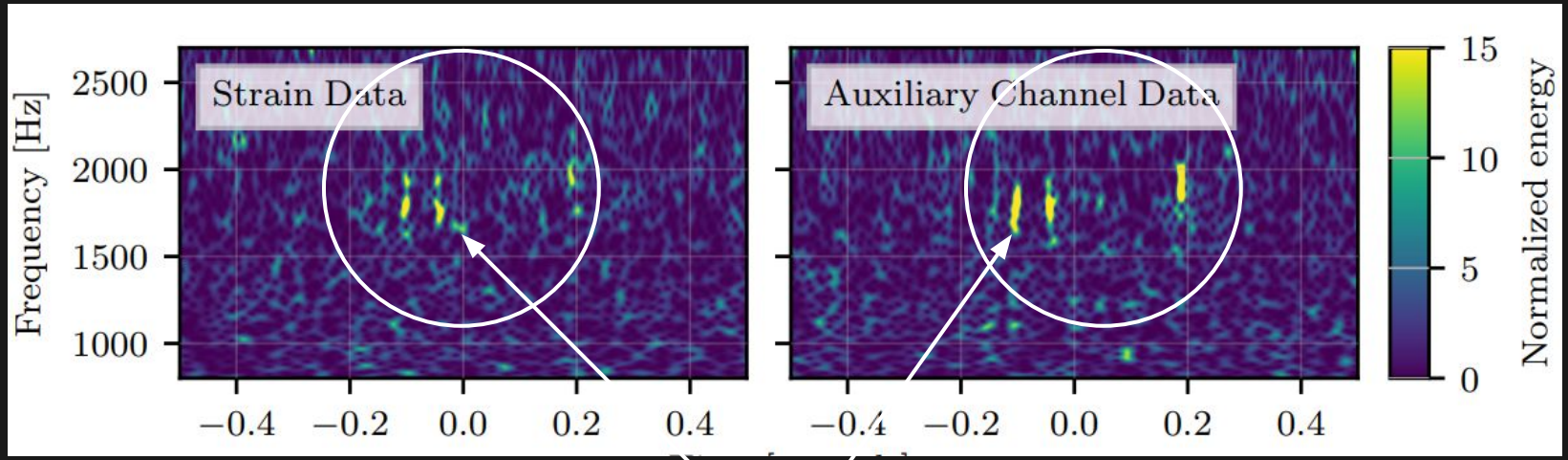
Aim to detect GWs in the primary channel

Also measure numerous auxiliary channels



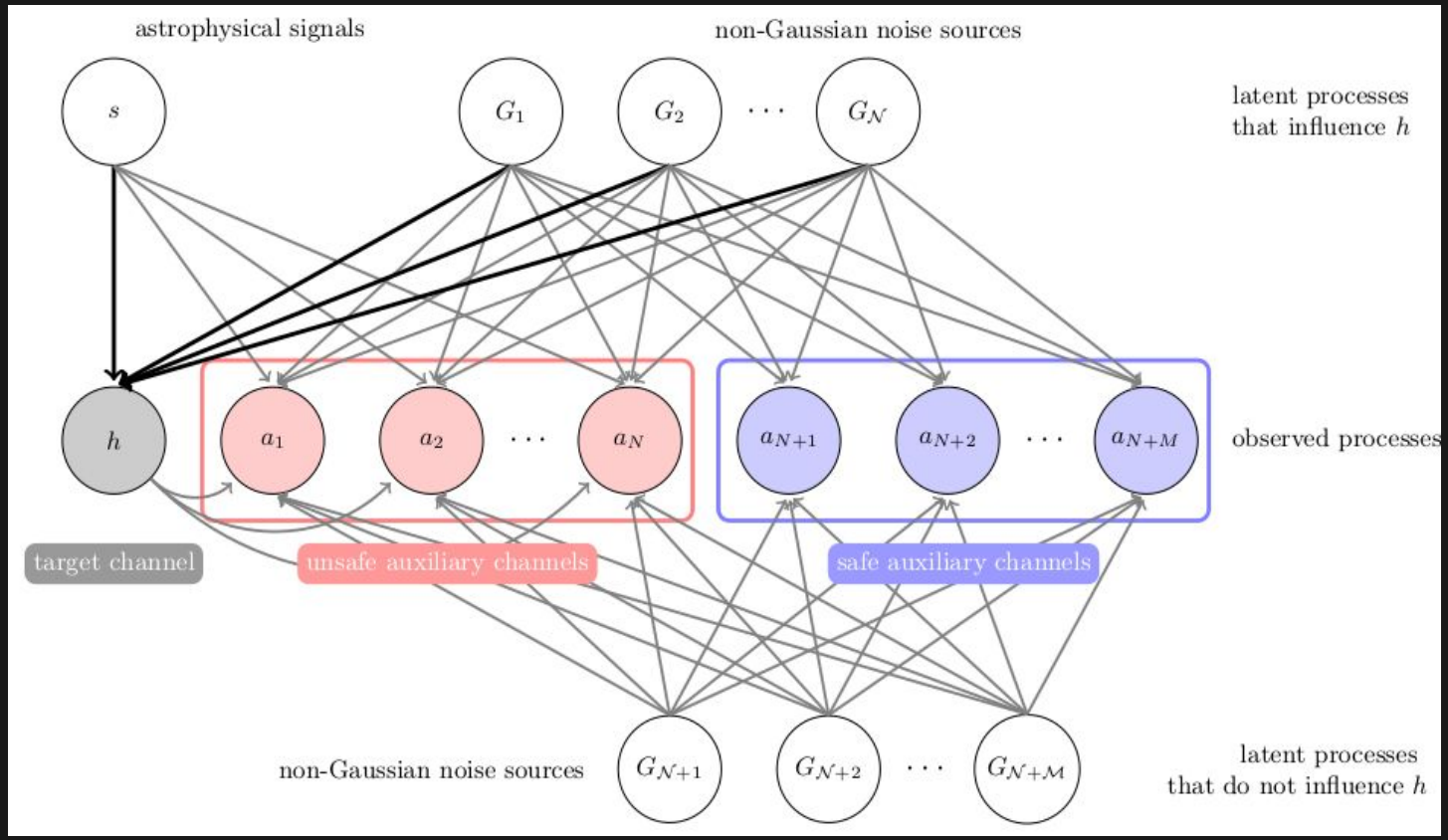


Auxiliary Channels in LVK IFOs (2/2)



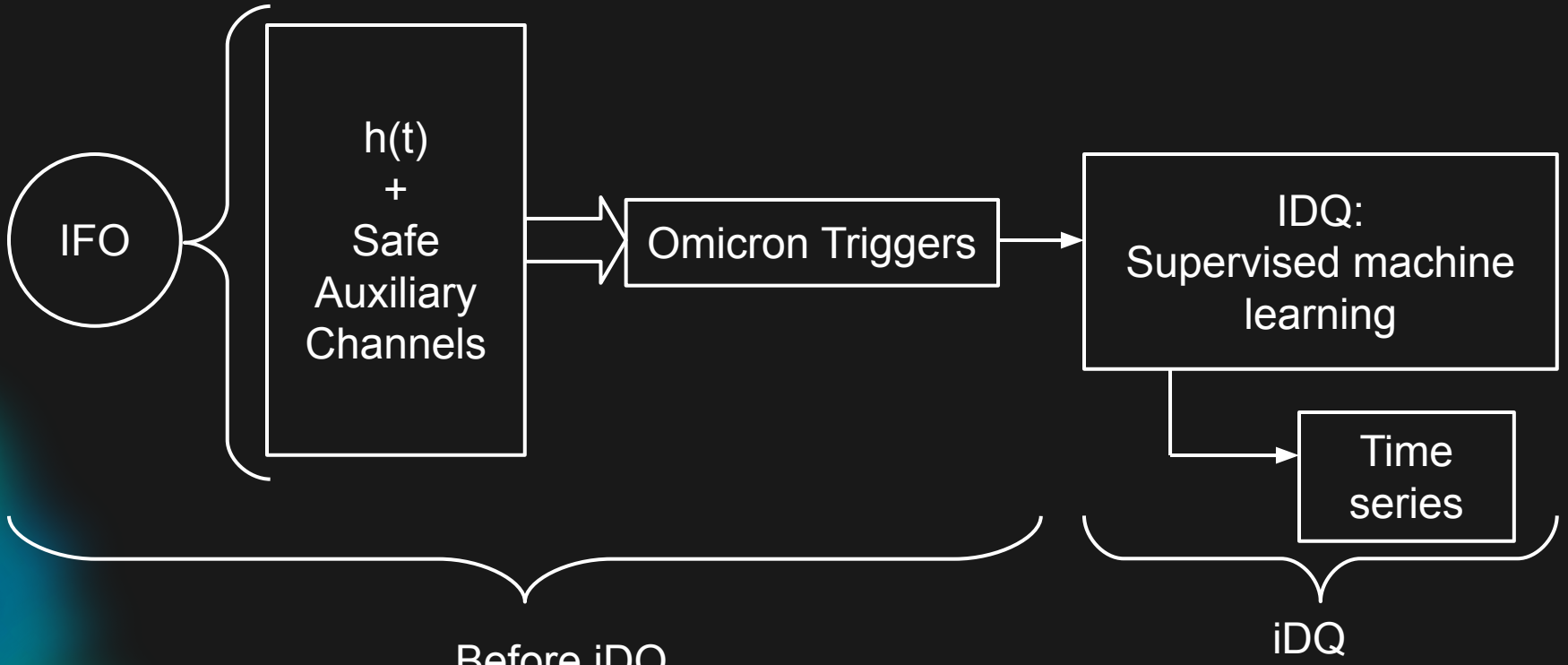
One can find correlation between Auxiliary channels and the target channel !

IDQ : machine learning algorithm to class noise (1/2)



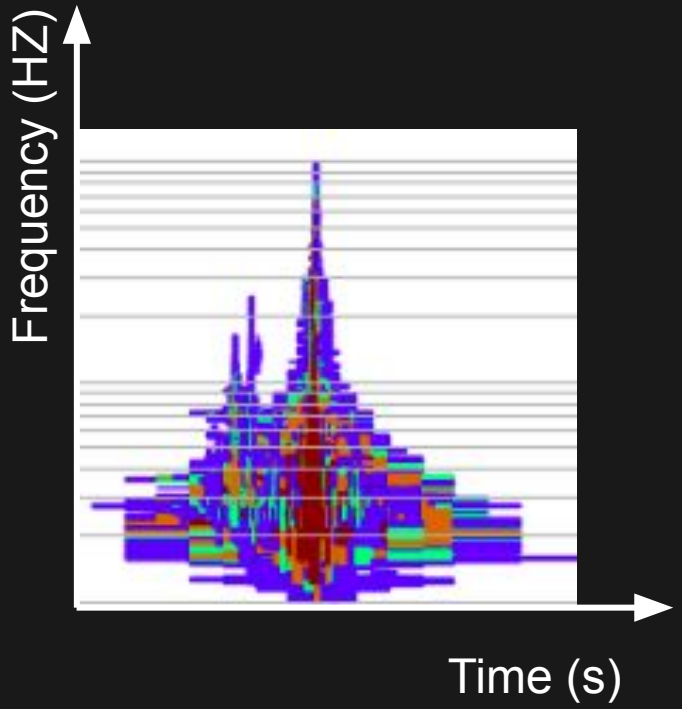
Credit: Essick et al 2020

IDQ : machine learning algorithm to class noise (2/2)

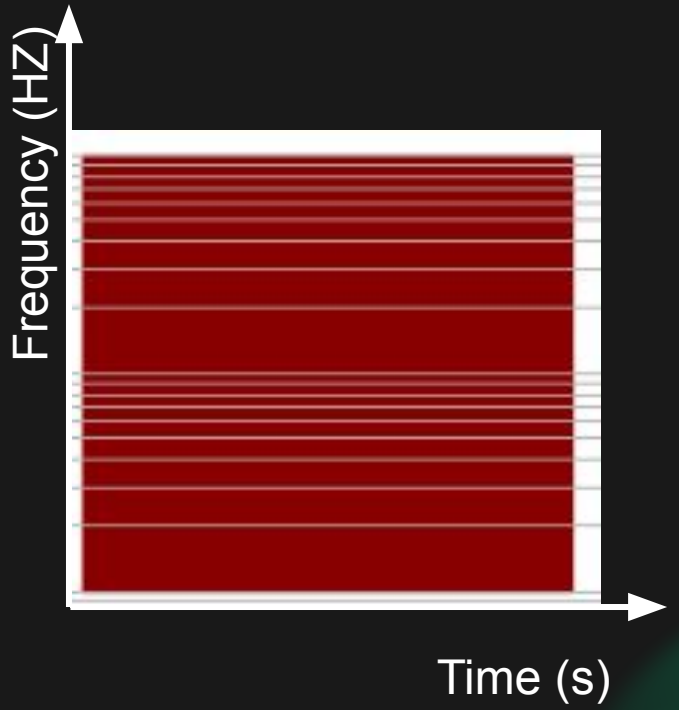


Credit: Essick et al 2020

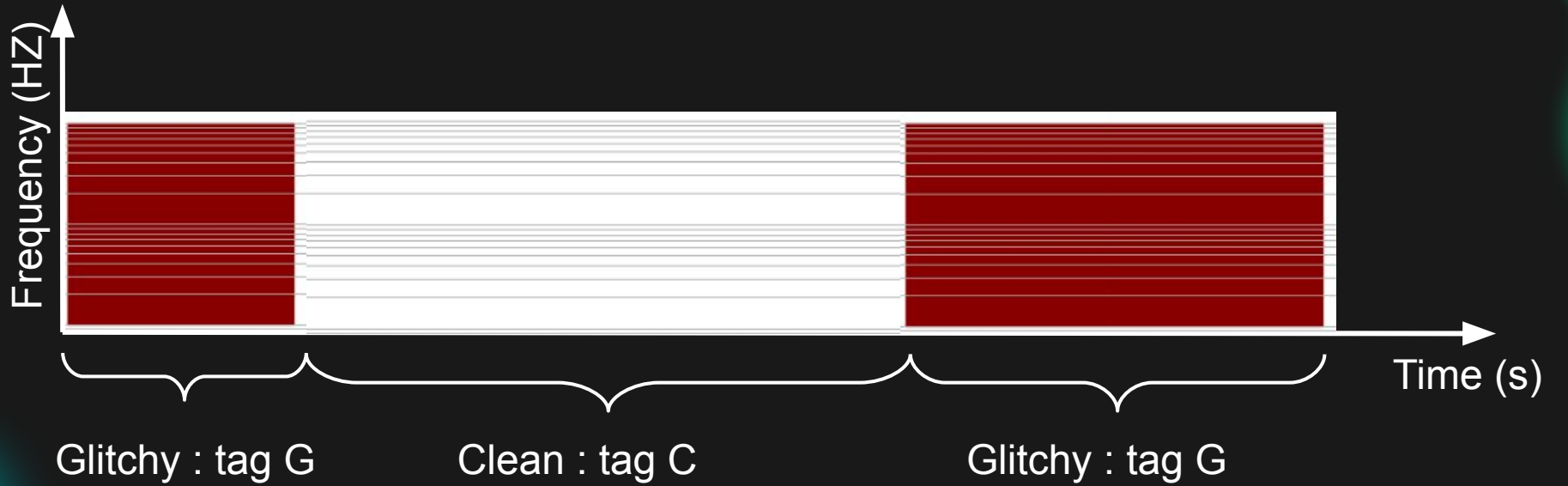
Omicron Triggers



Simplification of the Glitch



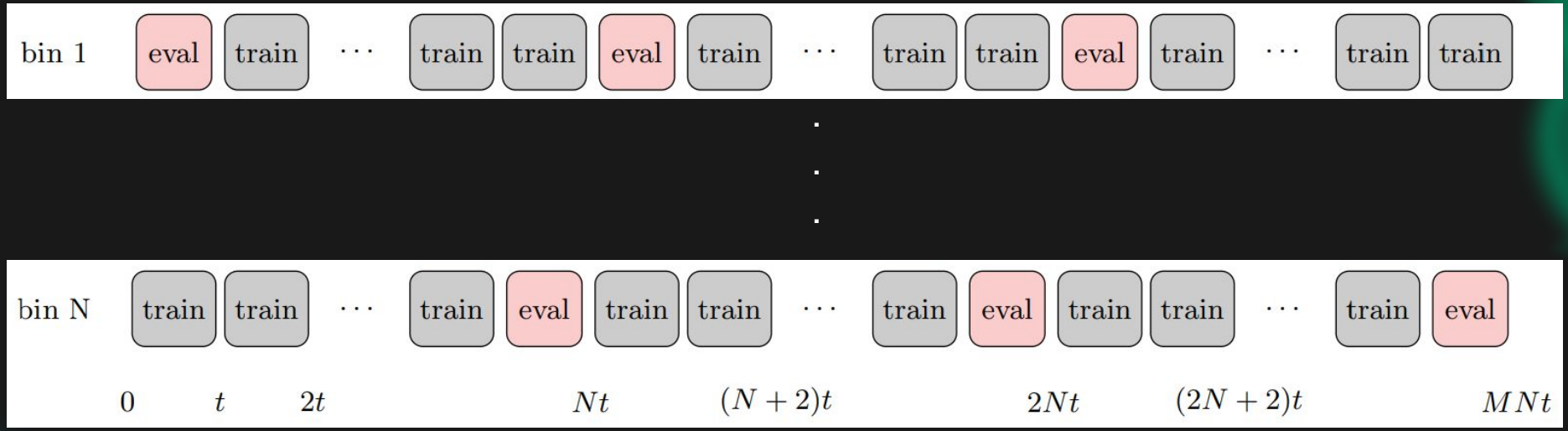
IDQ : How does it work in practice ? (1/2)



1st Step : iDQ looks at h(t) and tag segments.

2nd Step : Extract transient from Auxiliary channel and tag them according to step 1

IDQ : How does it work in practice ? (2/2)



3rd Step : Divide the samples into M segments used to build N bins.
Train to separate C from G from N-1 bins and evaluate on the last bin.
Confirm evaluation with h(t). Then rotate to do it on each bin.

Credit: Essick et al 2020

Ⅲ- IDQ : A solution to identify data.

a) Basic ideas.

a) First test on O4b 1st week Virgo data

a) First look at 25min glitches.

First test

April, 10 2024
(3PM UTC)
GPS time : 1396797018

April, 16 2024
(3PM UTC)
GPS time : 1396311218

It corresponds to a live time of : “4 days, 15h18 and 35 seconds”

Technical aspects

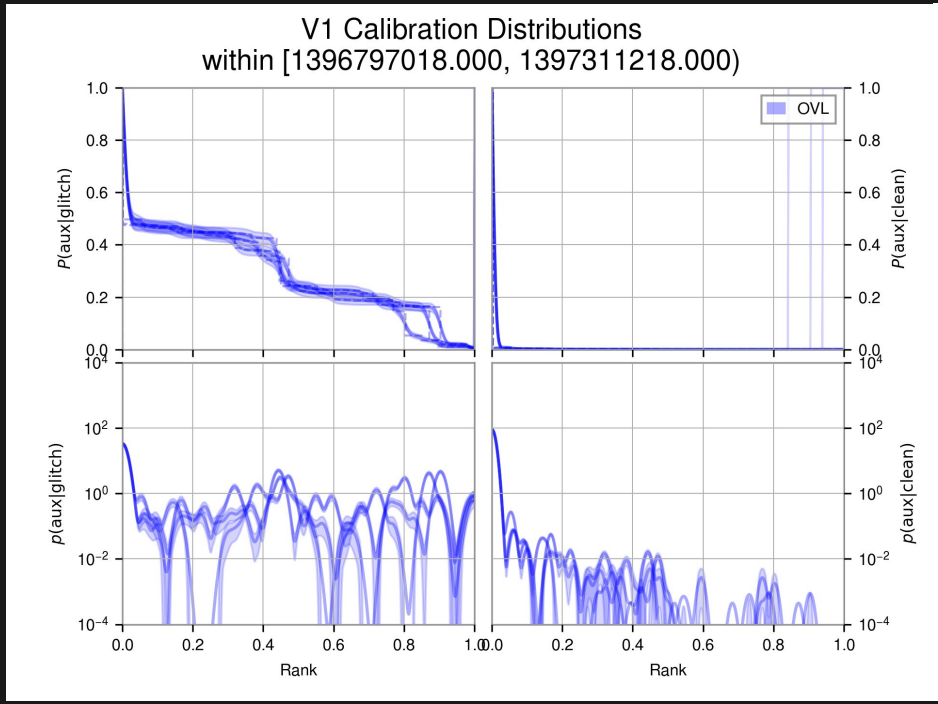
The git repo to run iDQ is here : git to run iDQ.

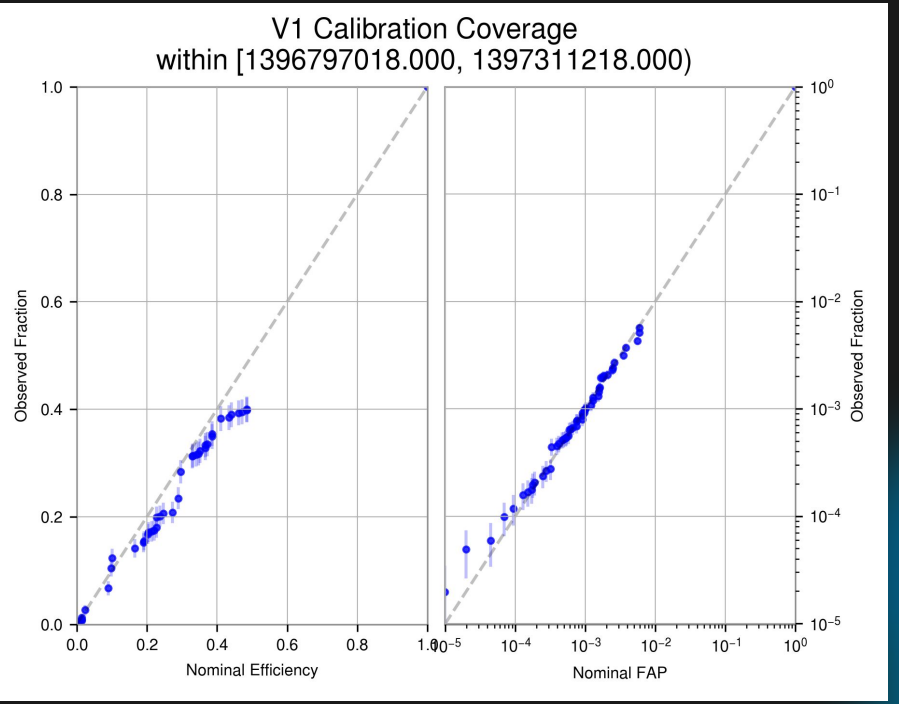
We launch as 1 single SLURM job (Condor is hidden behind SLURM) at CC IN2P3 with data already clustered.

ldq-batch -v –num-bins=3 –num-segs-per-bin=56 config.toml start_time end_time

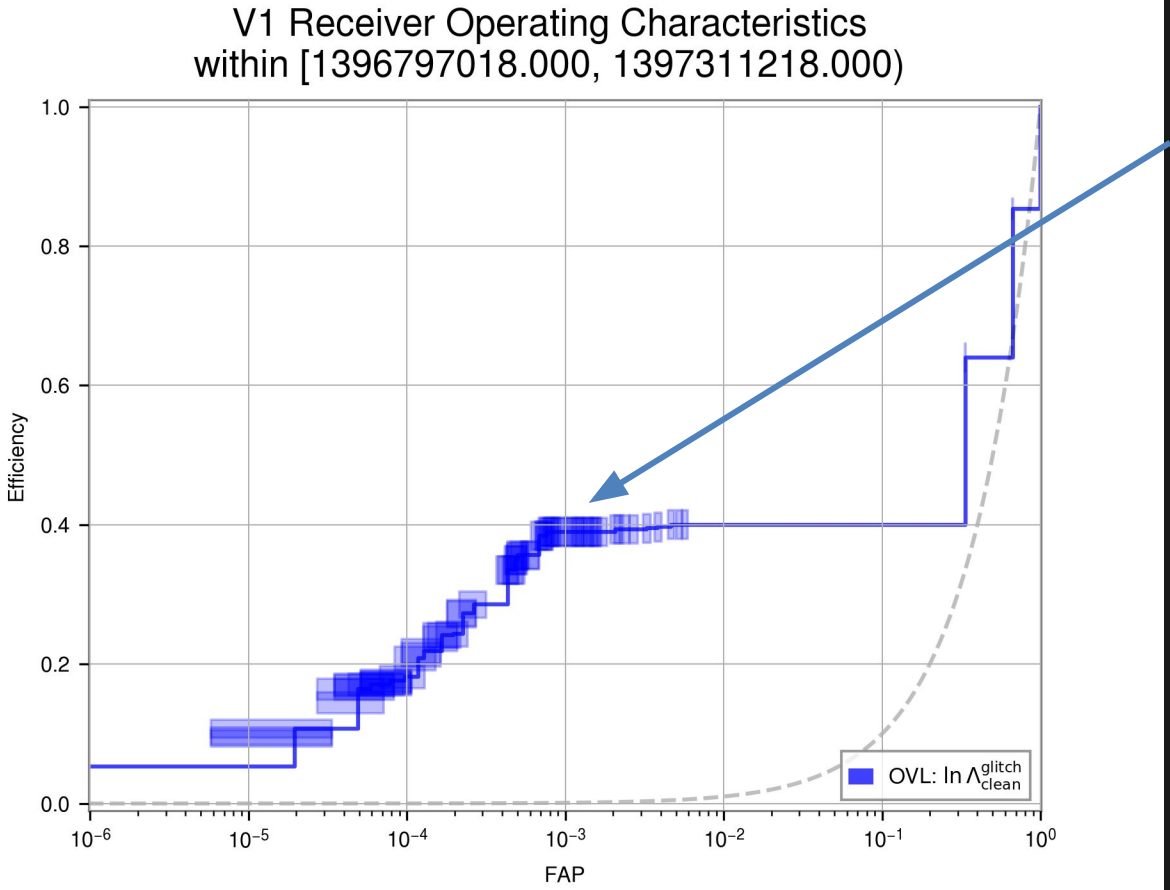
Execution time for ~ 2100 channels :
20H59 minutes and 9s

IDQ calibration (summary)





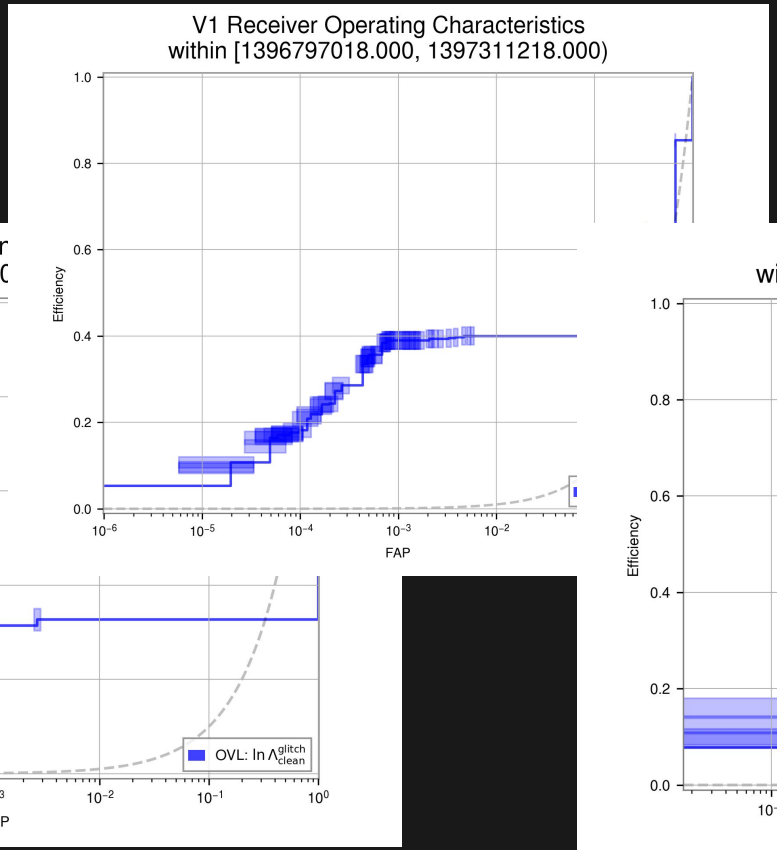
Efficiency of iDQ for 500 000 seconds



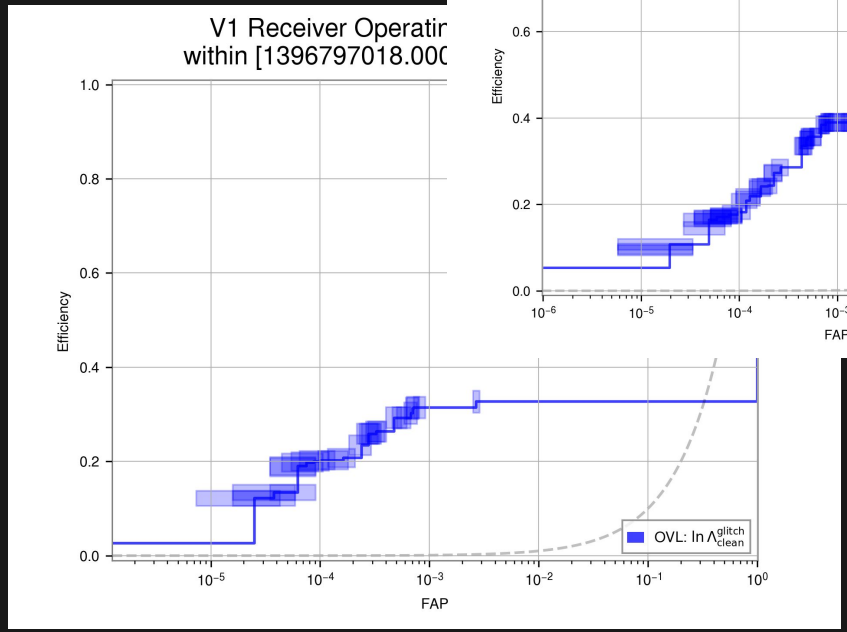
This point at (10^-3, 0.4) means that there is a rank at which you identify 40% of glitches and only veto 0.001 of clean time

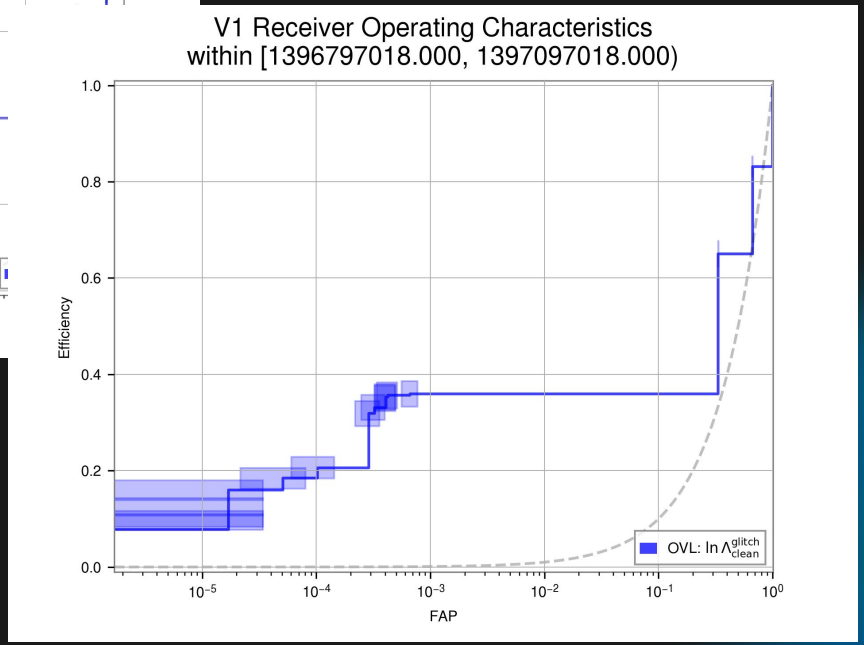
1 000 000 seconds of data.

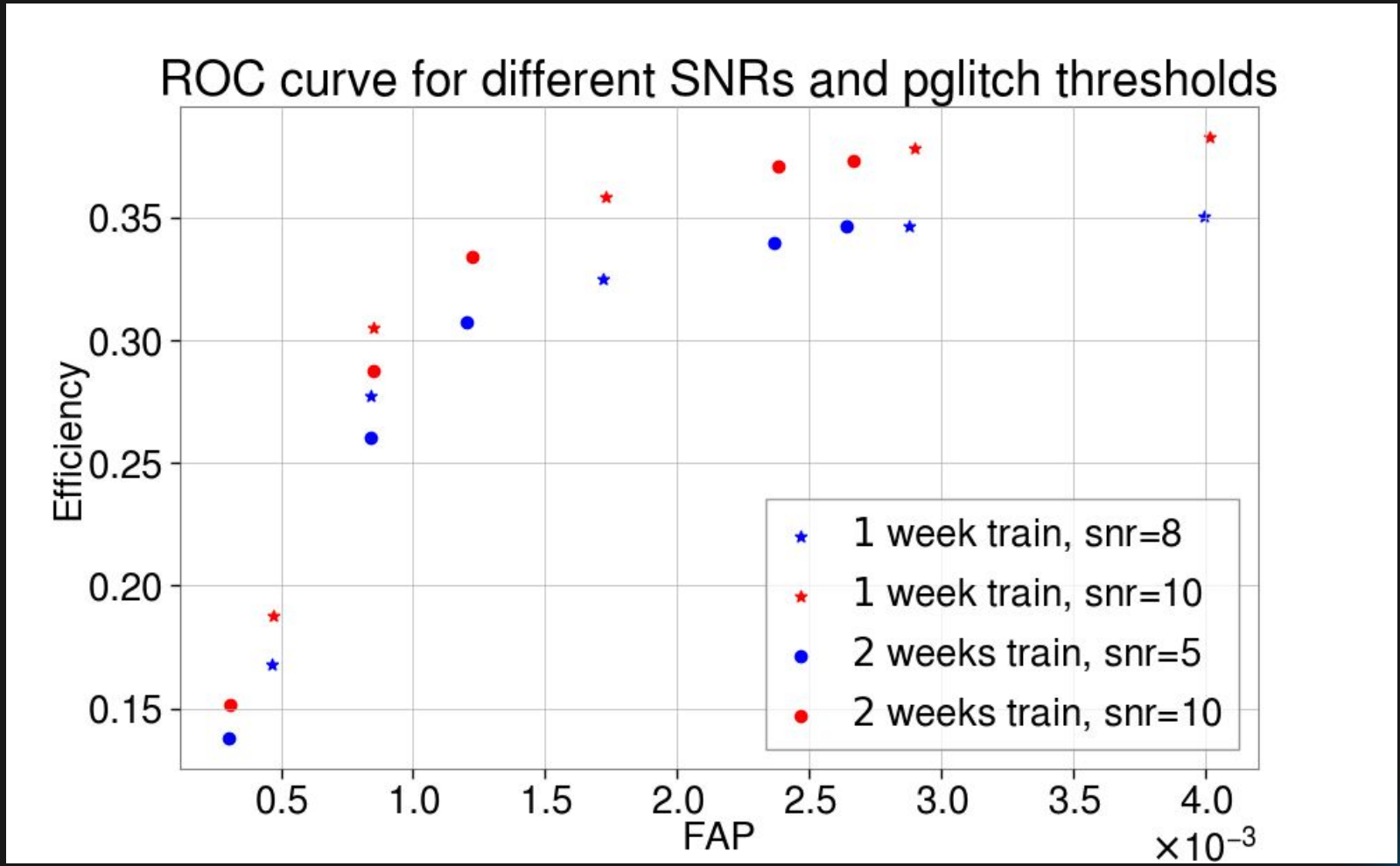
500 000 seconds



300 000 seconds of data.






ROC curve for different SNRs and pglitch thresholds
Efficiency
0.35
0.30
0.25
0.20
0.15
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FAP
×10⁻³
1 week train, snr=8
1 week train, snr=10
2 weeks train, snr=5
2 weeks train, snr=10

Ⅲ- IDQ : A solution to identify glitches.

a) Basic ideas.

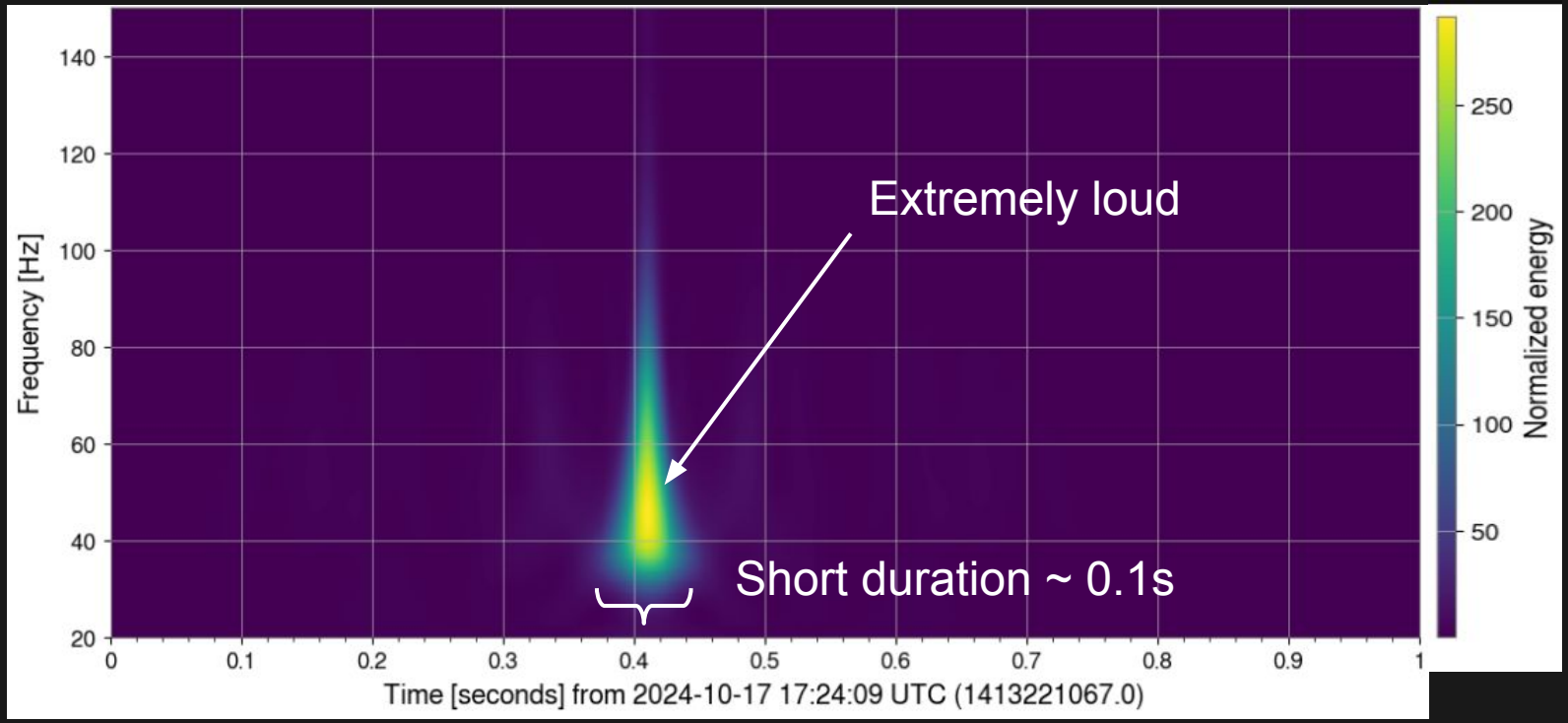
a) First test on Virgo 1st week.

a) First look at 25min glitches.

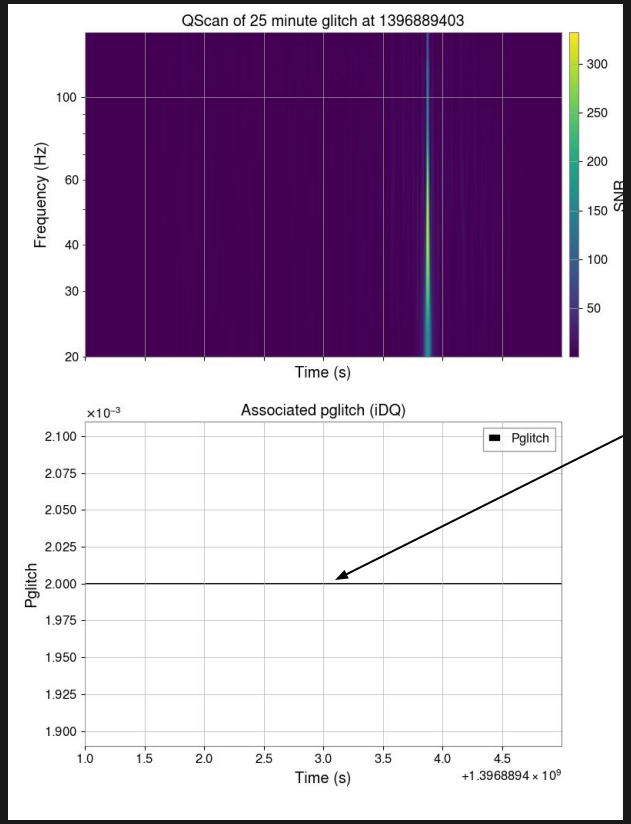
25 minutes Glitches (2/2)

- Unknown origin
- Huge SNR of about $O(10^2)$
- At the beginning of O4b, every 25 minutes. Not so true anymore.
- Can IDQ give any hint on their origin?

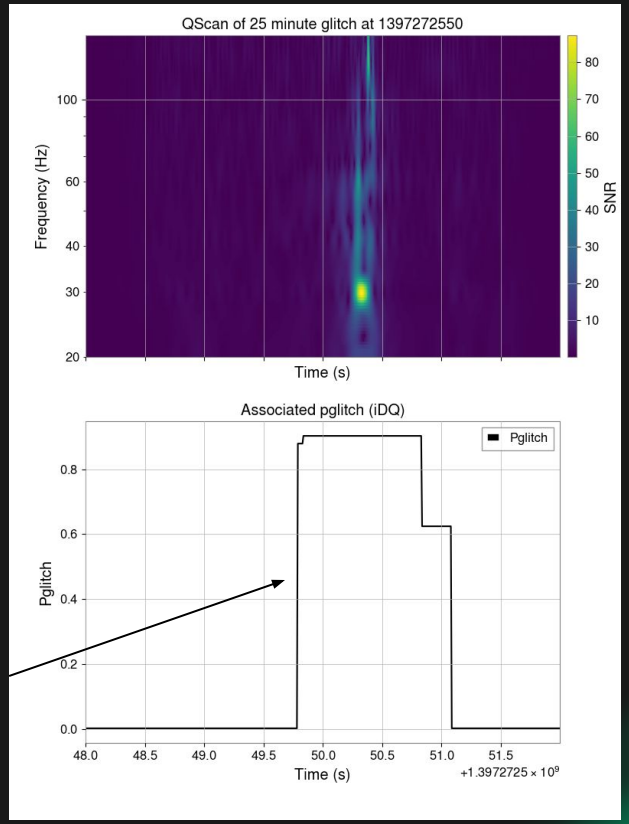
25 minutes Glitches (1/2)



IDQ for 25 minute glitches

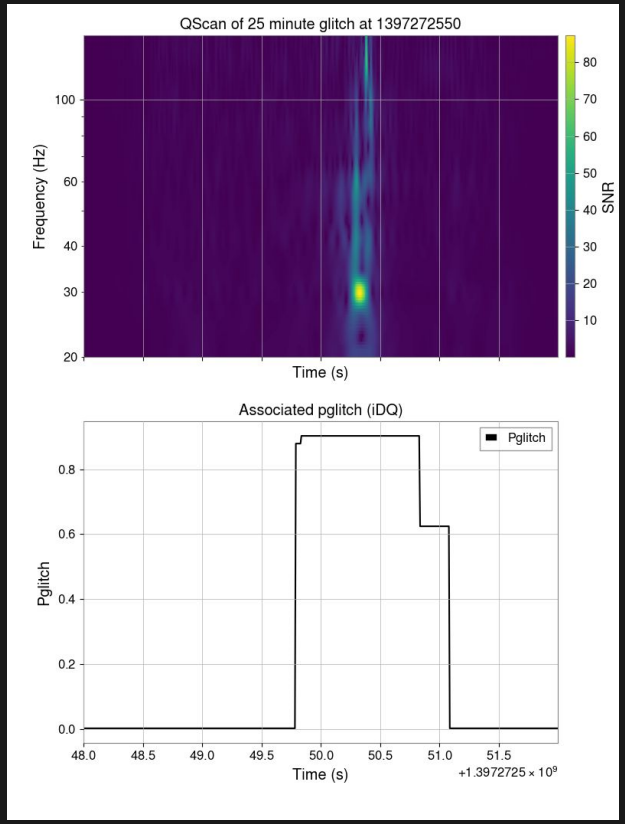


Not detected



Detected

Detection with pglitch

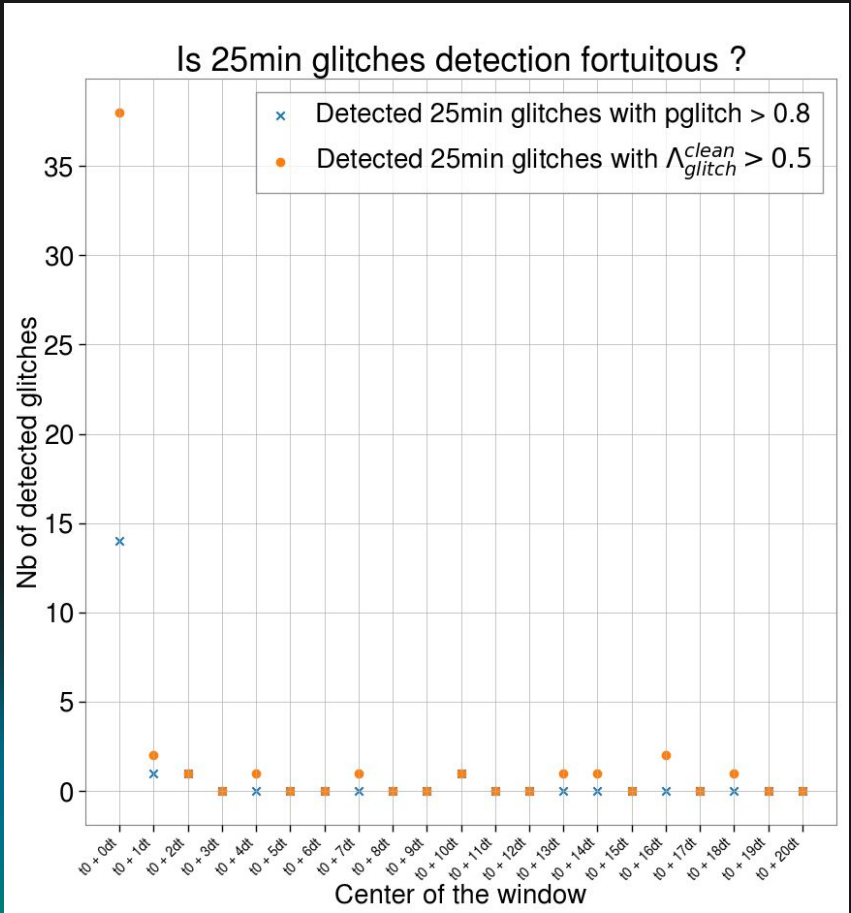


With pglitch > 0.5 : 38 out of 203 (18.7%)

With 0.01 < pglitch < 0.5 : 12 out of 203 (5.9%)

With pglitch < 0.01 : 153 out of 203 (75,4%)

Sanity check



How did we do ?

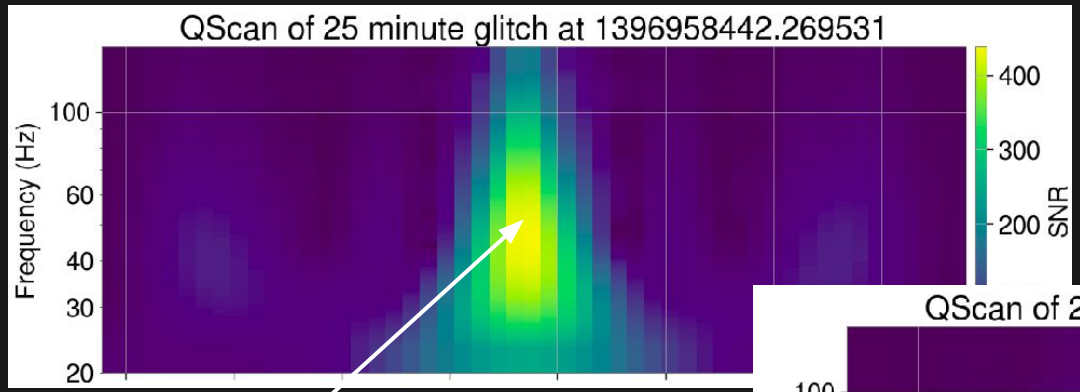
We counted the number of time we had pglitch > 0.8 in 0.5 second windows around a known 25min glitch

For windows from 0.1s to 0.5s, we still have the 13 detections of glitches.

Our results seems to indicate that the 25 minute glitches spotted by iDQ are not fortuitous :)

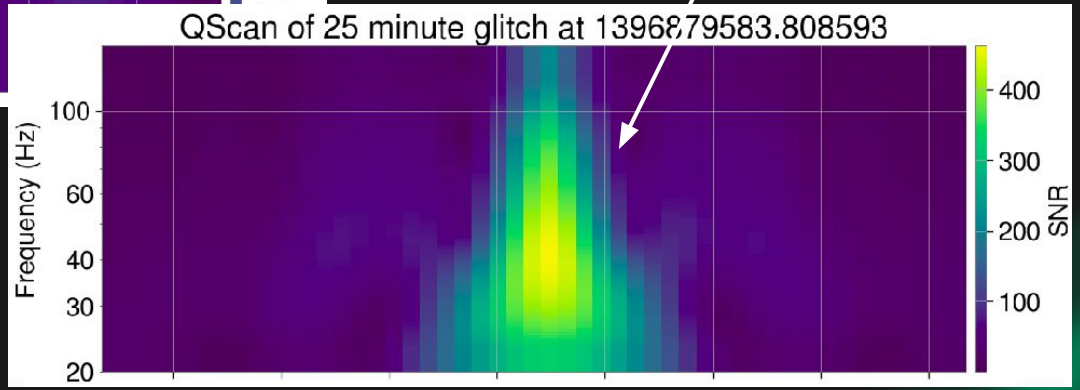
Glitches morphology

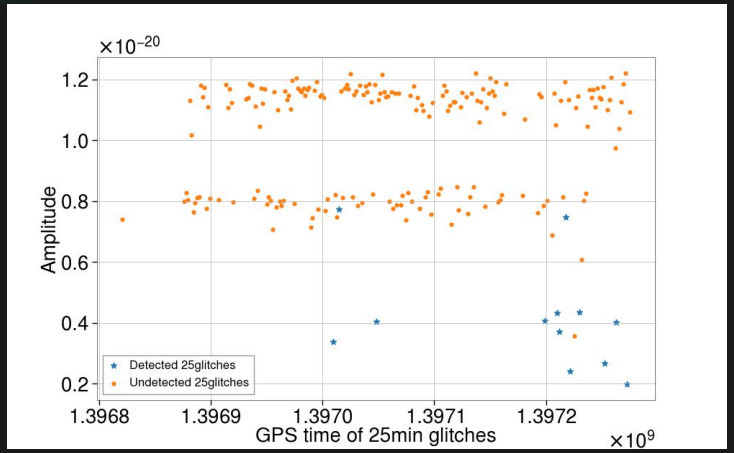
Detected and undetected 25min glitches looks similar

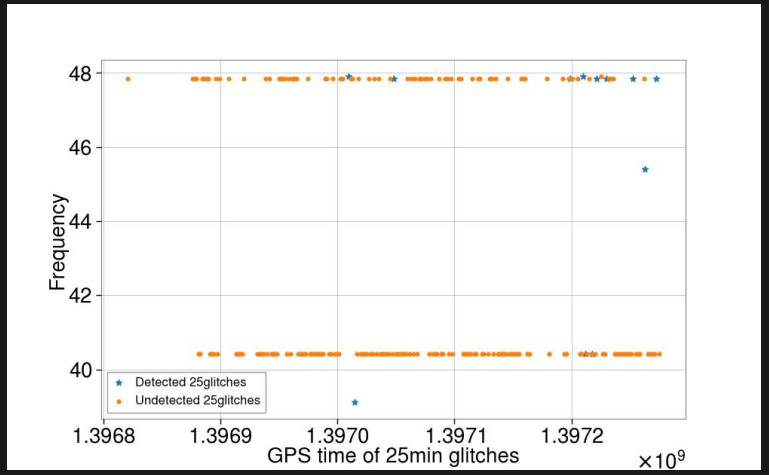


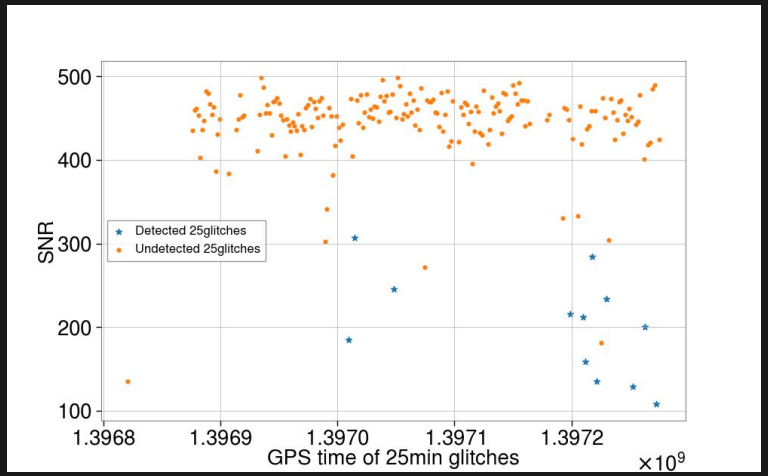
Detected one

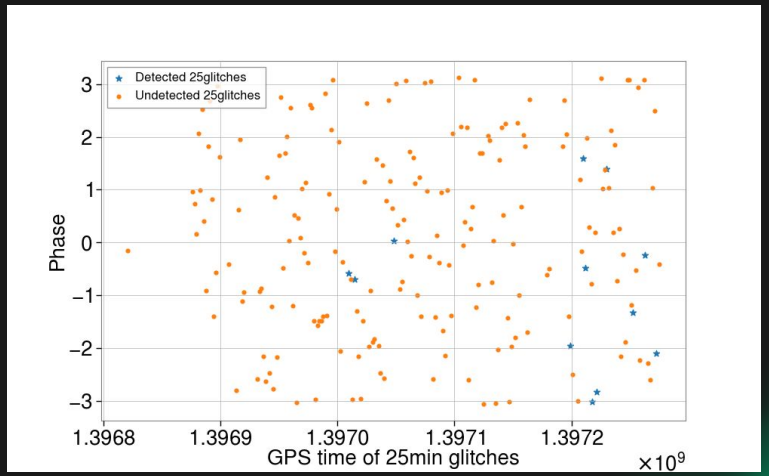
Undetected one




×10−20
1.2
1.0
0.8
0.6
0.4
0.2
Amplitude
Detected 25glitches
Undetected 25glitches
1.3968
1.3969
1.3970
1.3971
1.3972
GPS time of 25min glitches
×10^9


48
46
44
42
40
Frequency
Detected 25glitches
Undetected 25glitches
1.3968
1.3969
1.3970
1.3971
1.3972
GPS time of 25min glitches
×10^9


500
400
300
200
100
SNR
Detected 25glitches
Undetected 25glitches
1.3968
1.3969
1.3970
1.3971
1.3972
GPS time of 25min glitches
×10^9


3
2
1
0
−1
−2
−3
Phase
Detected 25glitches
Undetected 25glitches
1.3968
1.3969
1.3970
1.3971
1.3972
GPS time of 25min glitches
×10^9

Is there auxiliary channels correlated ?

See more details in this txt file.

```
detected 25 minute glitch GPS time : 1397198805.316406
                    V1:Sc_PR_MIR_VOUT_UL_7.0_0.25:1 0.3
                    V1:Sc_PR_MIR_VOUT_UL_10.0_1.0:1 0.34
                    V1:Sc_PR_MIR_VOUT_UL_10.0_0.25:1 0.34
                    V1:Sc_PR_MIR_VOUT_UL_7.0_0.25:1 0.36
detected 25 minute glitch GPS time : 1397209894.570312
                    V1:TCS_NI_CO2_PWRLAS_9.0_0.125:1 0.27
                    V1:TCS_NI_CO2_PWRLAS_9.0_0.0625:1 0.29
                    V1:Sc_PR_MIR_VOUT_UL_7.0_0.25:1 0.3
                    V1:TCS_NI_CO2_PWRIN_50.0_1.0:1 0.33
                    V1:Sc_PR_MIR_VOUT_UL_10.0_0.25:1 0.34
                    V1:Sc_PR_MIR_VOUT_UL_10.0_1.0:1 0.34
                    V1:Sc_PR_MIR_VOUT_UL_7.0_0.25:1 0.36
                    V1:ENV_CEB_UPS_CURR_S_2000Hz_20.0_1.0:1 0.36
                    V1:TCS_NI_CO2_POWER_CH_PICKOFF_100.0_1.0:1 0.38
                    V1:TCS_NI_CO2_PWRLAS_10.0_1.0:1 0.42
                    V1:ENV_CEB_UPS_CURR_S_2000Hz_20.0_1.0:1 0.49
                    V1:ENV_CEB_UPS_CURR_S_2000Hz_25.0_1.0:1 0.5
                    V1:ENV_CEB_UPS_CURR_S_2000Hz_25.0_1.0:1 0.51
                    V1:TCS_NI_CO2_PWRLAS_15.0_1.0:1 0.55
                    V1:TCS_NI_CO2_PWRLAS_10.0_1.0:1 0.56
                    V1:TCS_NI_CO2_PWROUT_50.0_1.0:1 0.59
                    V1:TCS_NI_CO2_PWRLAS_10.0_1.0:1 0.63
                    V1:TCS_NI_CO2_PWRLAS_12.0_1.0:1 0.71
                    V1:TCS_NI_CO2_PWRLAS_12.0_1.0:1 0.78
                    V1:ENV_CEB_UPS_CURR_S_2000Hz_25.0_1.0:1 0.8
                    V1:TCS_NI_CO2_PWRLAS_25.0_1.0:1 0.87
                    V1:TCS_NI_CO2_PWRLAS_250.0_0.5:1 0.89
                    V1:TCS_NI_CO2_PWRLAS_25.0_1.0:1 0.9
```

Some correlated channels !

TCS_NI_CO2_PWLAS

TCS_NI_CO2_POWER_CH_PICKOFF

TCS_NI_CO2_PWROUT

ENV_CEB_UPS_CURR_S

SDB2_B1_PD2_6MHz

Ⅳ- Conclusion

Conclusion / Perspectives

- Thanks to Alexis, iDQ run for Virgo !

- 1 days to run on 1 week of data, 1.5 days to run on 2 weeks of data

- In contact with iDQ developers to finalise configuration (study was presented at an iDQ meeting)
- will start soon production, iDQ data products will be available for pipelines (There is some interest from pycbc)
- iDQ seems to see a correlation of some aux channels with a fraction of 25 mins glitches - will follow up with detchar and detector experts

- We are investigating the safe channel list with DetChar experts.

Thanks!

Do you have any questions?

l.zimmermann@ip2i.in2p3.fr