Data quality in gravitational-wave detectors

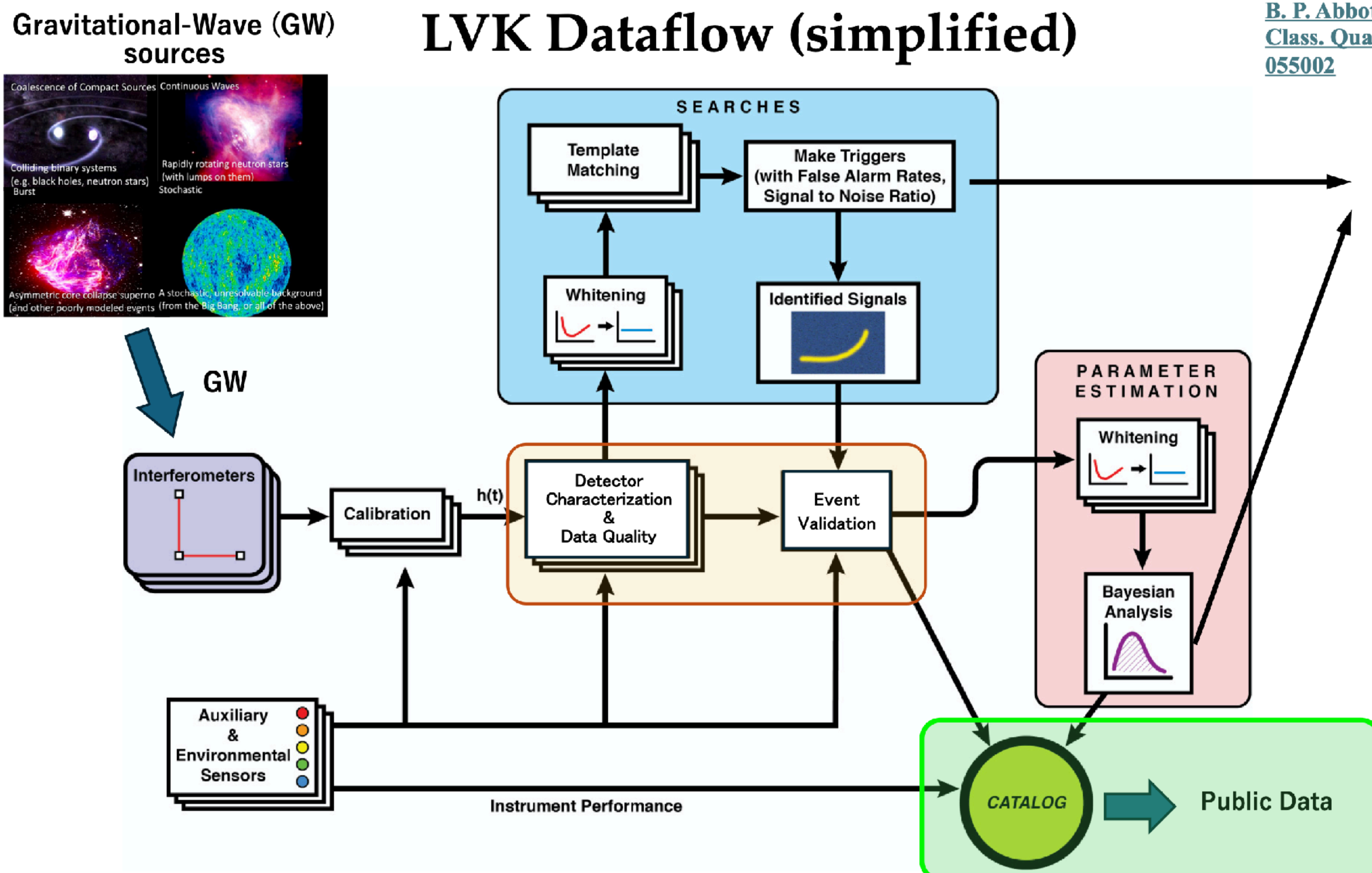
Hirotaka Yuzurihara
Gravitational Wave Open Data Workshop #7 (2024)

Acknowledgements

- Many slides are adapted from data quality presentations made by Ronaldas Macas, Laura Nuttall, Marissa Walker, and Jess McIver.
- For previous workshop slides, see https://gwosc.org/odw/

LVK Dataflow (simplified)

[B. P. Abbott et al., 2020 Class. Quantum Grav. 37 055002]



4

Slide by T.Sawada [LIGO-G2400946]

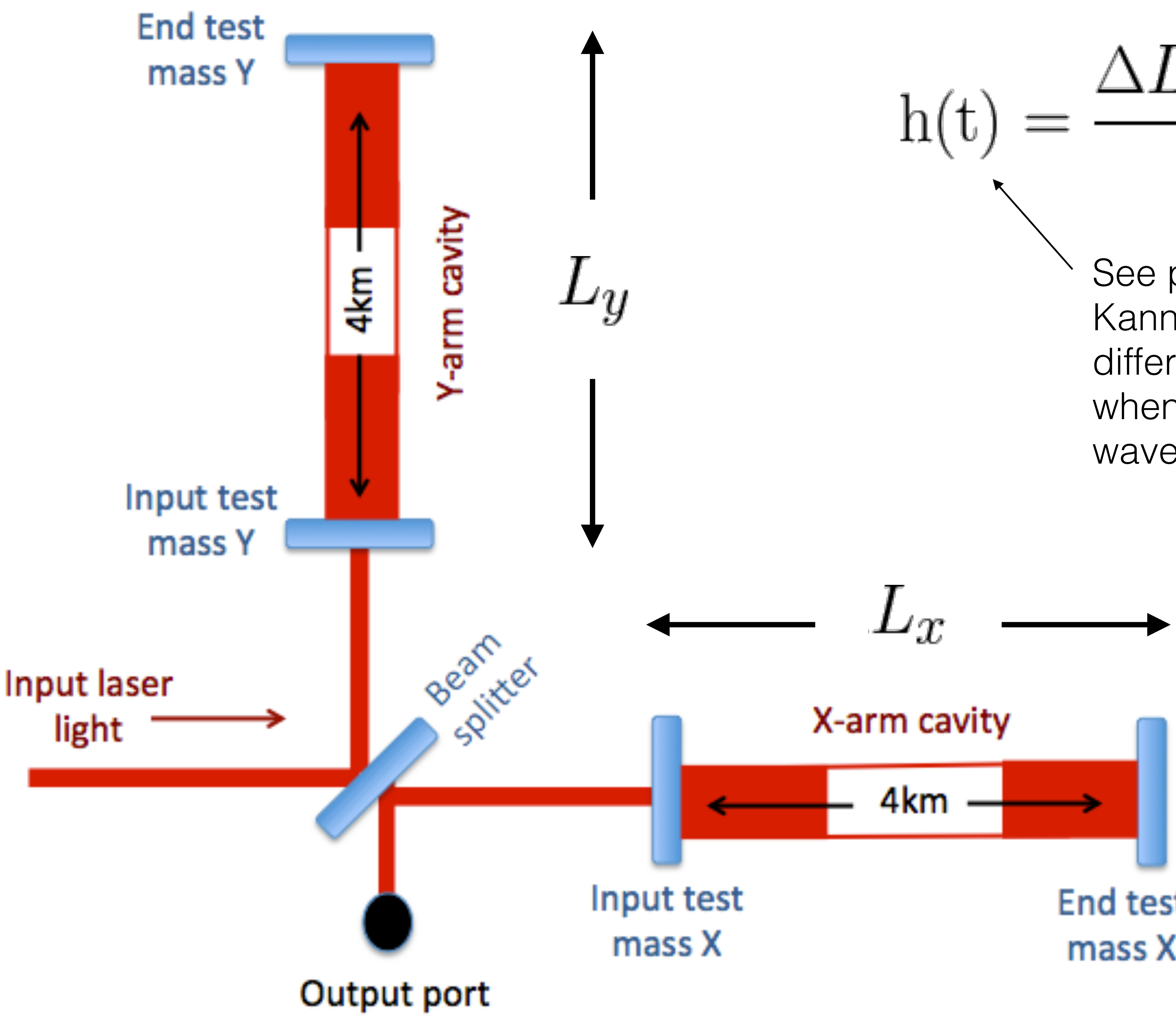
Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

What is strain data, *h(t)* ?

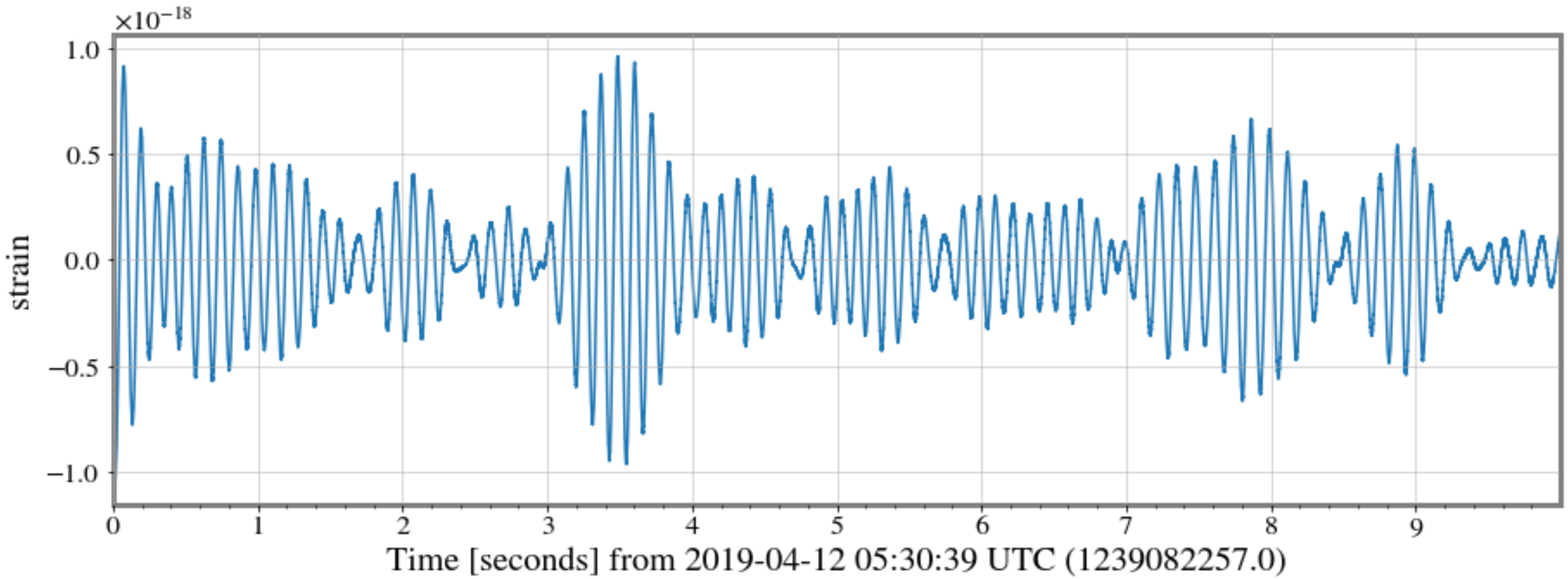


$$\mathrm{h(t)} = \frac{\Delta L_y - \Delta L_x}{L}$$

See previous talk by Jonah Kanner to find out how the differential arm behave when the gravitational wave passed.

What does strain data look like?

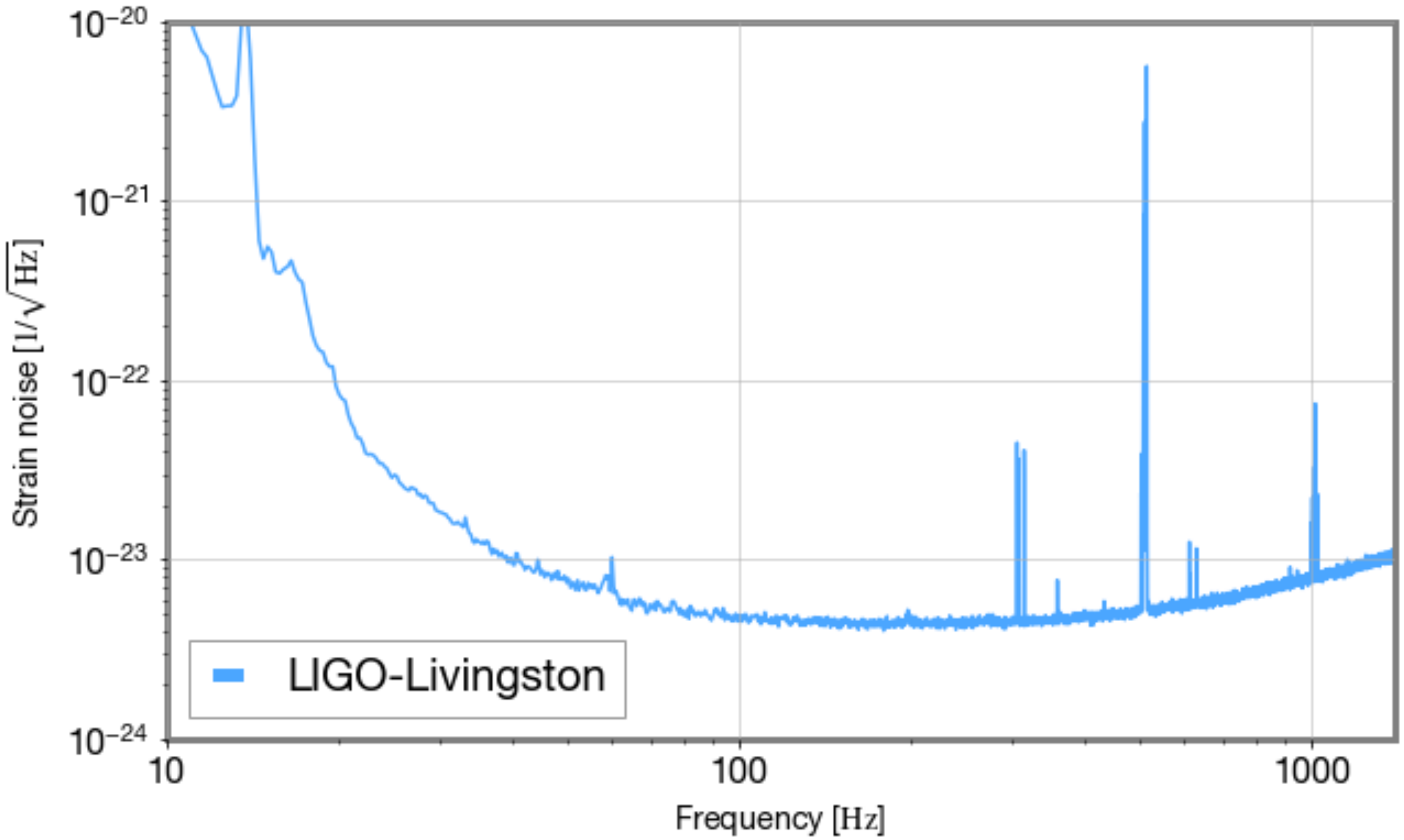
LIGO-Livingston h(t)



- h(t) sampling rate for open data: 16384 or 4096 Hz
- It’s difficult to obtain the information from the raw data.

https://github.com/gw-odw/odw-2024/blob/main/Tutorials/Day_1/Tuto_1.2_Open_Data_access_with_GWpy.ipynb

Strain data in the frequency domain



Fourier transform of data, called the Frequency spectrum

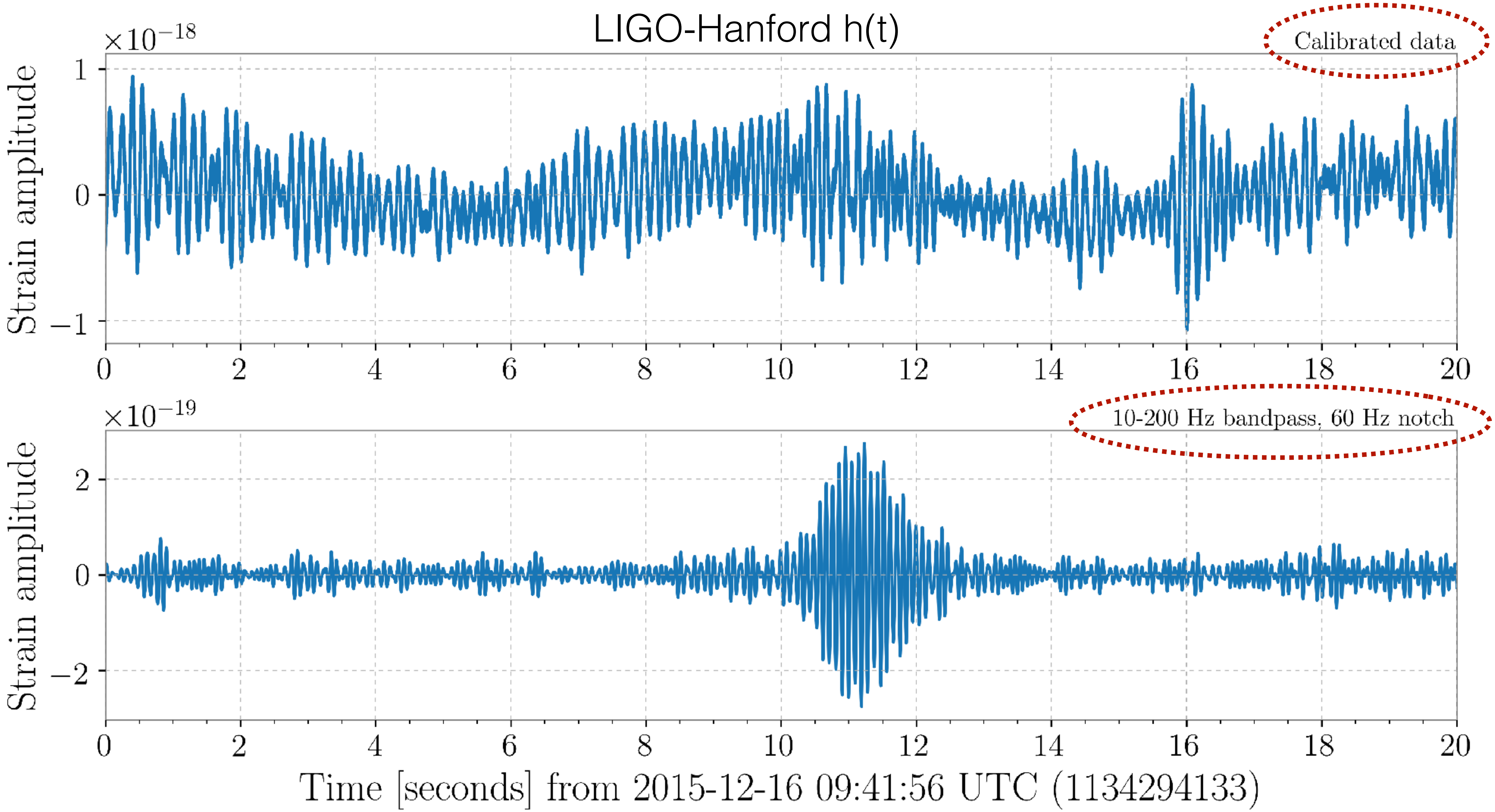
- Fourier transform: x(t) —> X(f)
- Power spectrum density (PSD)
- Amplitude spectrum density (ASD)
  - Median, mean
  - ASD is often used to represent the sensitivity.

$$\mathrm{PSD}(f) = \frac{2\left\langle |X(f)|^2 \right\rangle}{T}$$

$$\mathrm{ASD}(f) = \sqrt{\mathrm{PSD(f)}}$$

https://github.com/gw-odw/odw-2024/blob/main/Tutorials/Day_1/Tuto_1.2_Open_Data_access_with_GWpy.ipynb

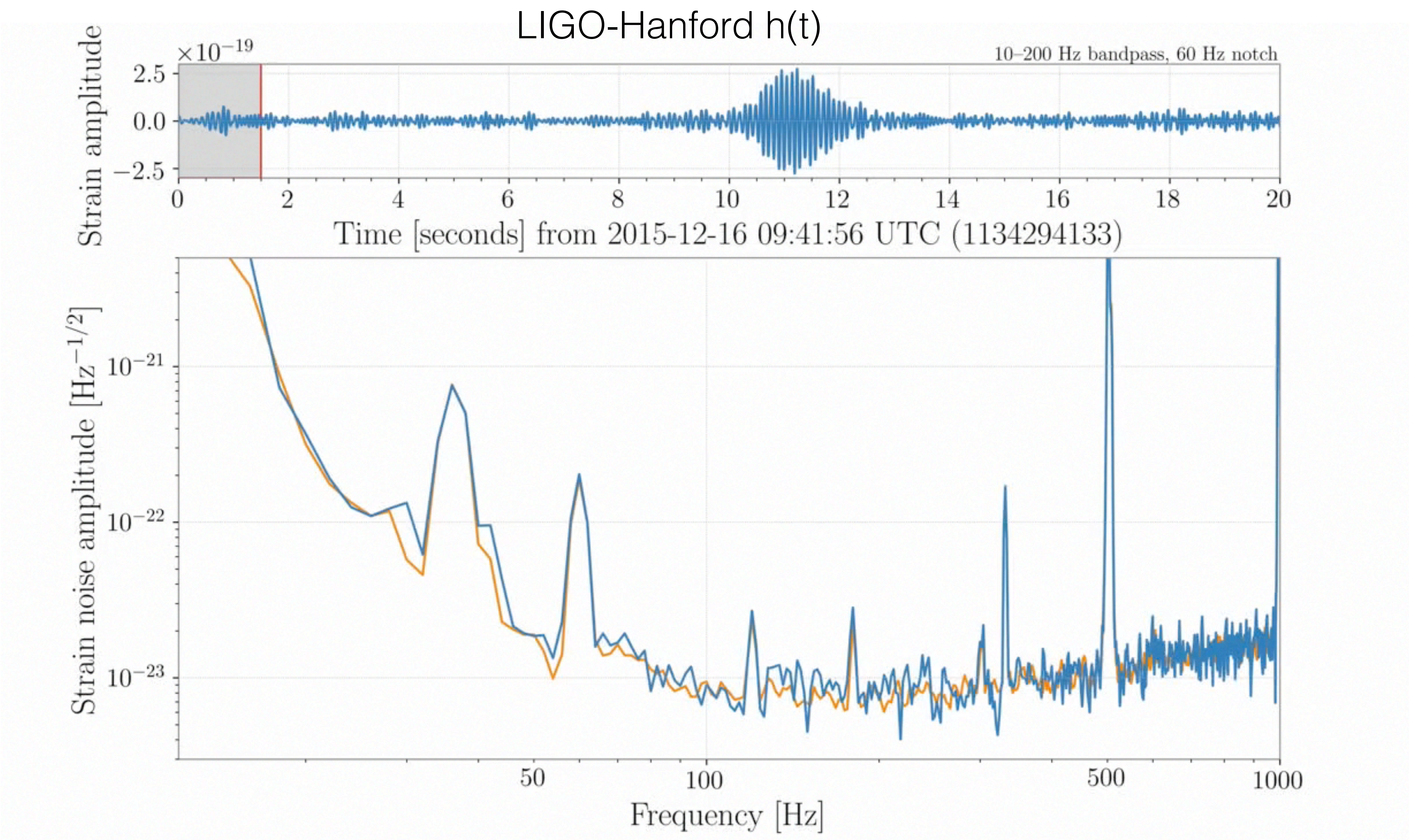
What does strain data look like?



- Signal processing by using time domain filter.

Made with GWpy by Duncan Macleod

Strain data in the frequency domain

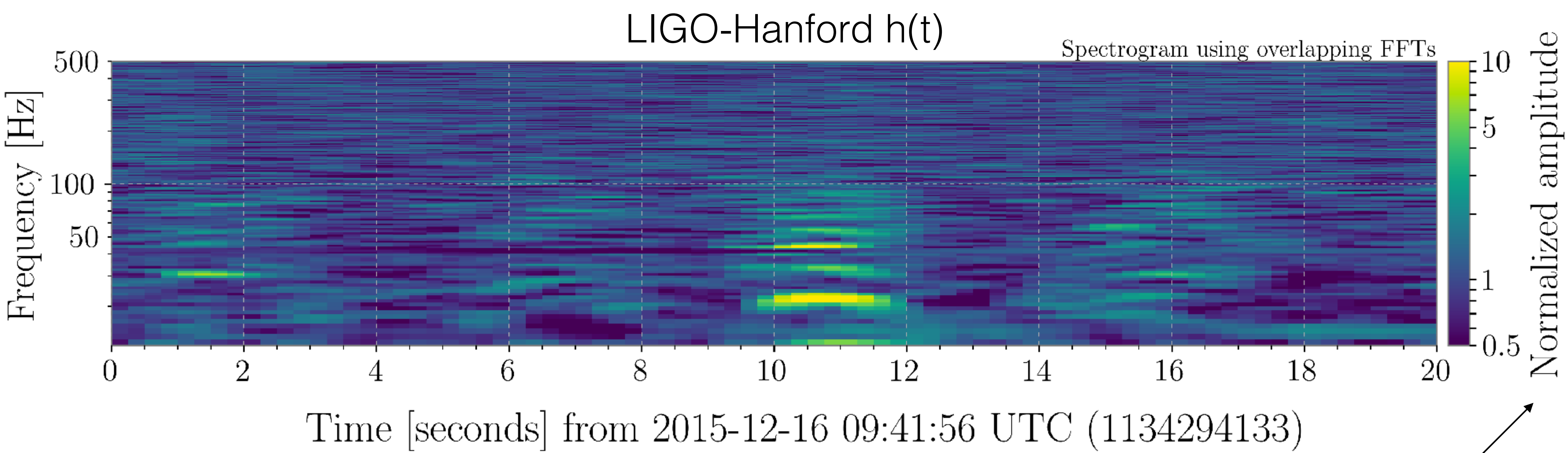


- h(t) data is not always stationary —> non-stationary

Made with GWpy by Duncan Macleod. Code: https://git.io/gwpy-ligo-scattering-animation

0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

Time-frequency map (normalized spectrogram in gwpy)



Roughly speaking, normalized amplitude
= **(blue ASD)** / **(orange ASD)**
in the previous page

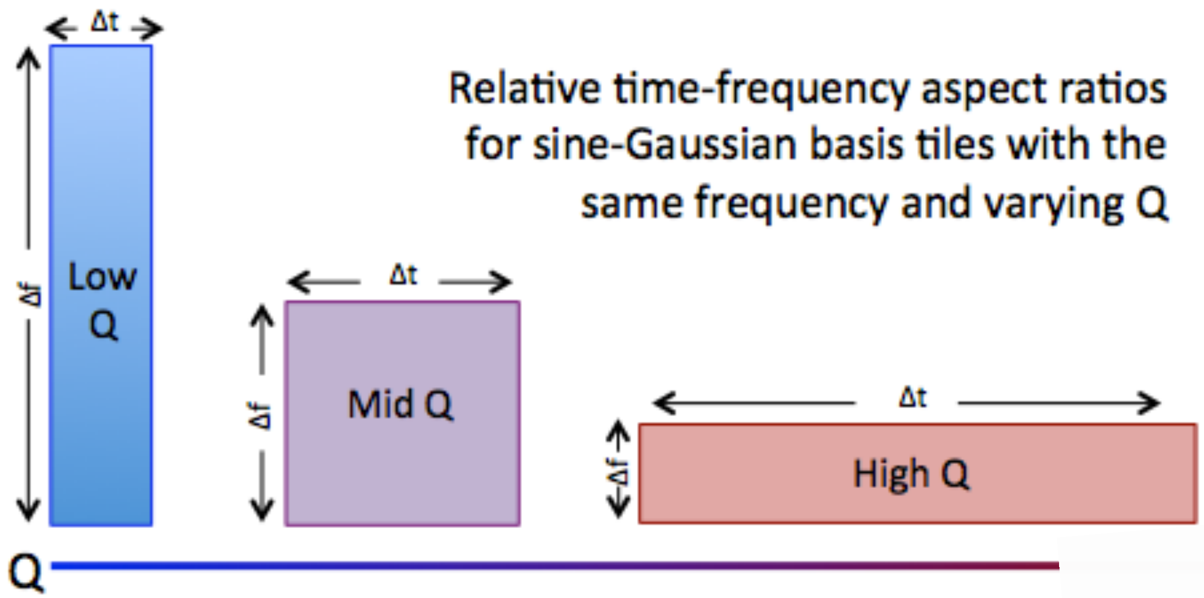
- Normalized spectrum is convenient to find the excess.

Made with GWpy by Duncan Macleod

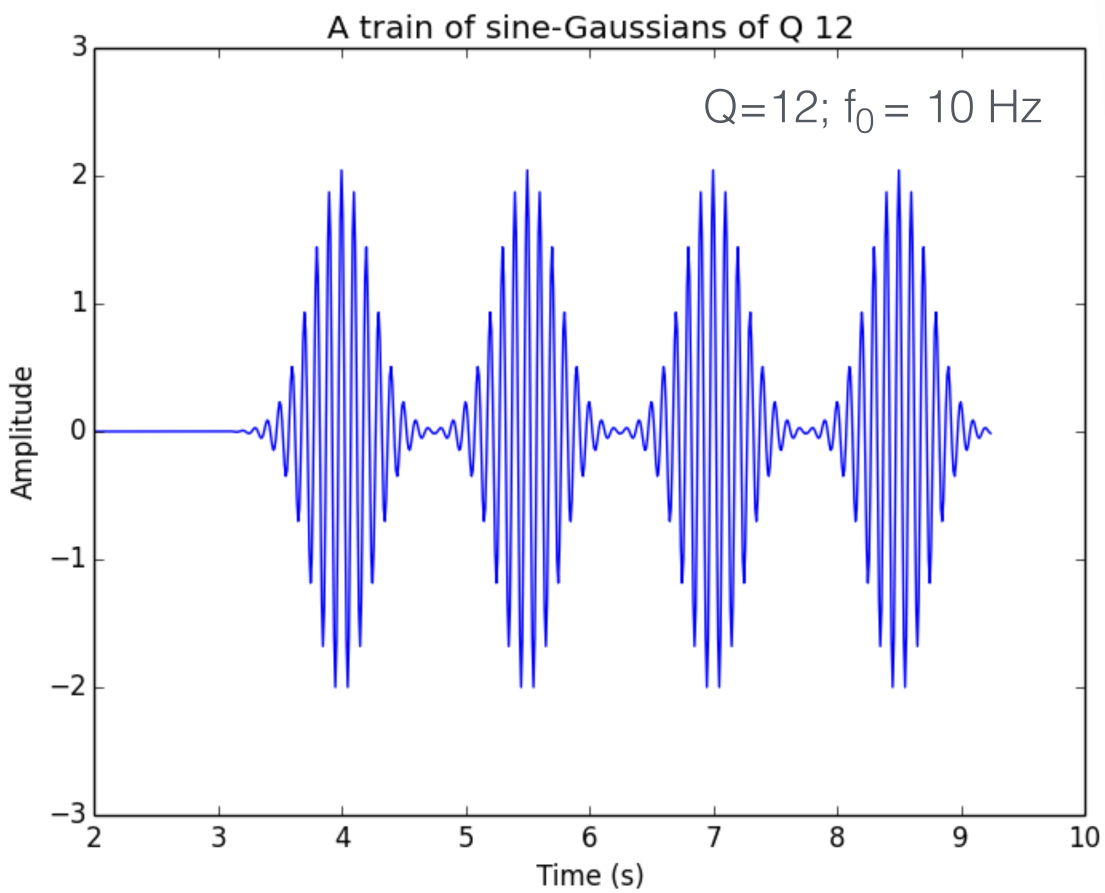
See also https://gwpy.github.io/docs/stable/examples/spectrogram/ratio/

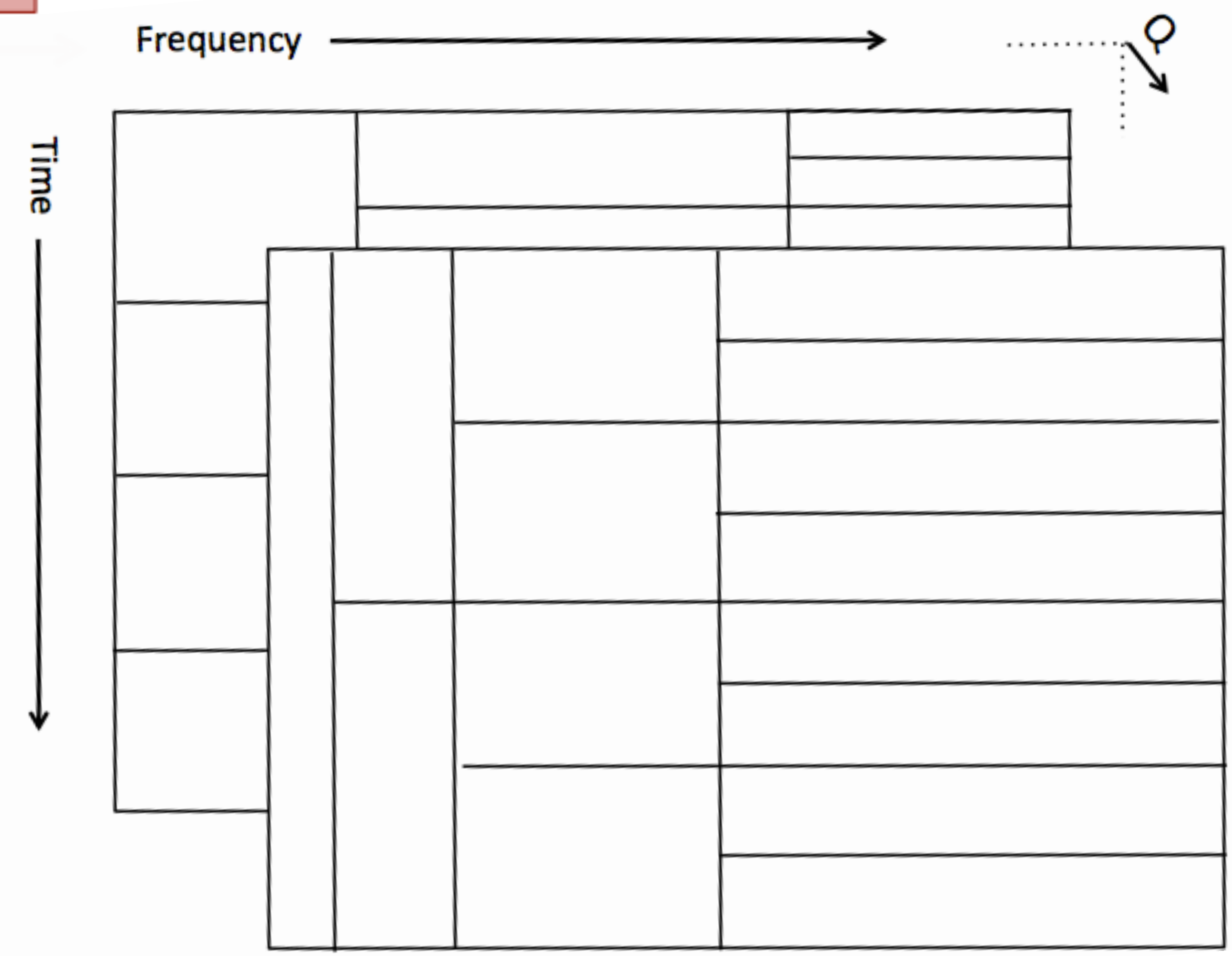
Q transform

S. Chatterji et al. CQG (2010)
Images: McIver

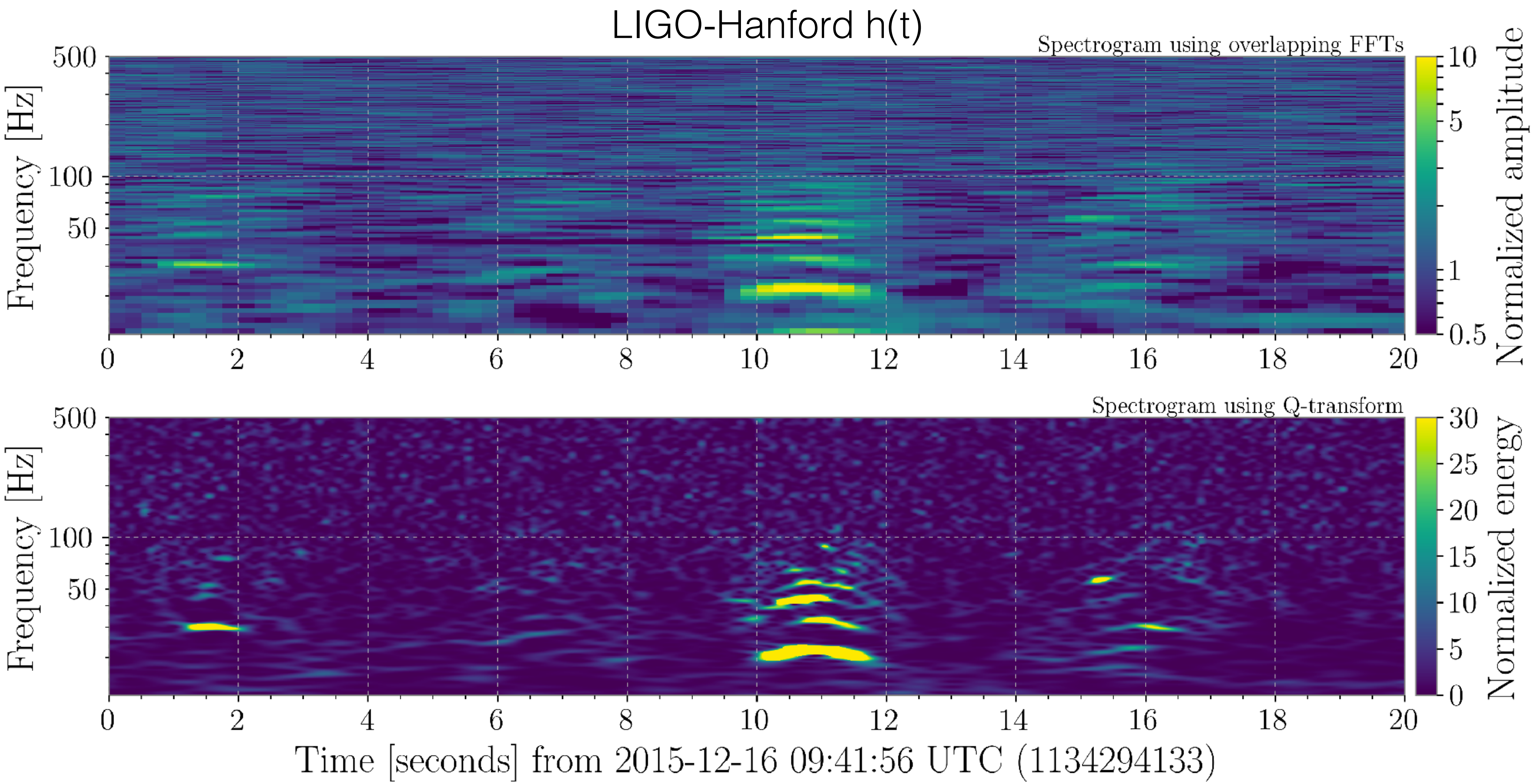


$$Q = \frac{f_0}{\Delta f}$$: Quality factor
how many times it oscillate before dumping





Time-frequency spectrograms



Made with GWpy by Duncan Macleod

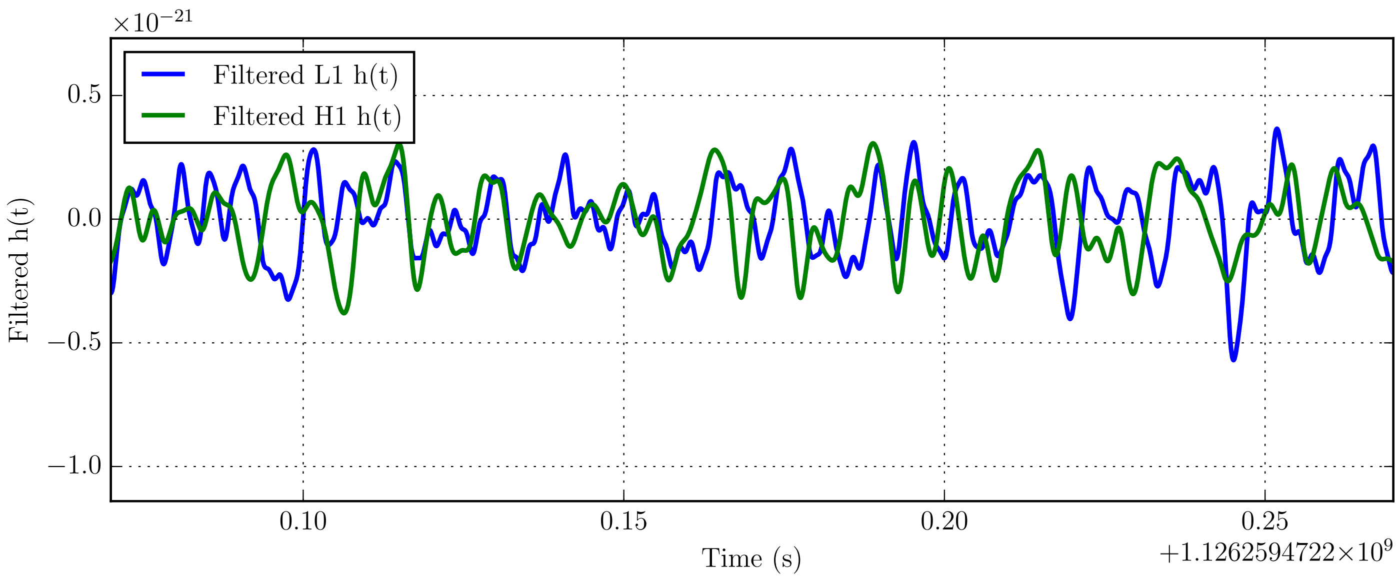
See also https://gwpy.github.io/docs/stable/examples/timeseries/qscan/

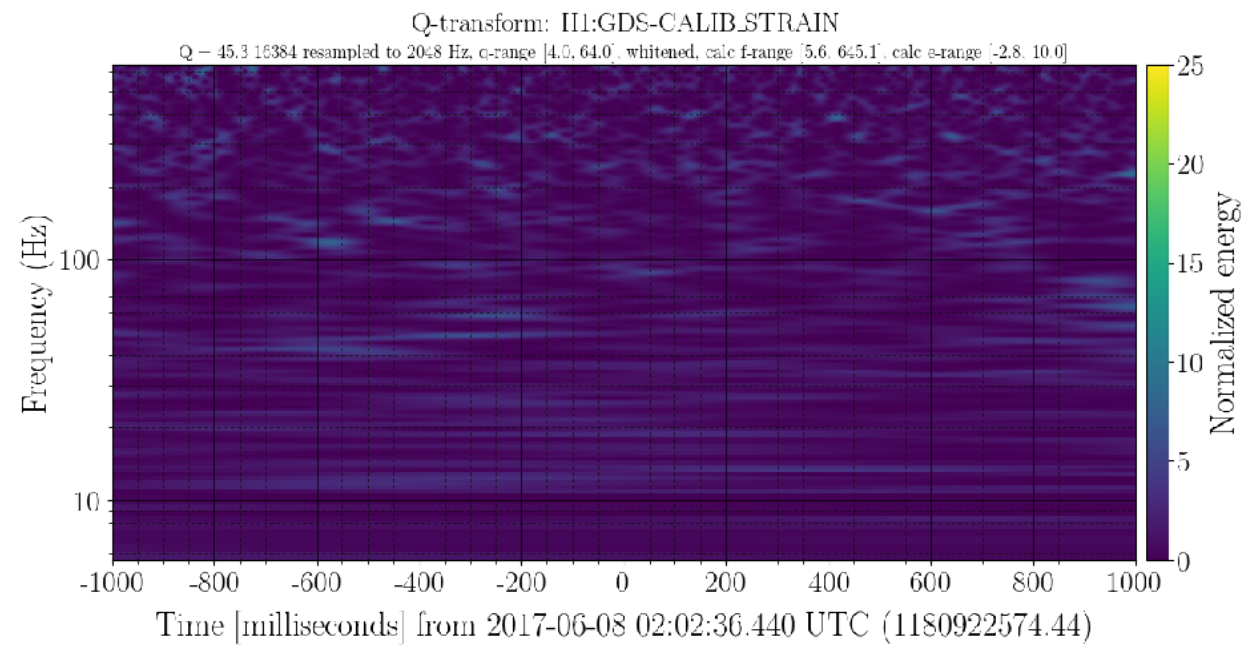
Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

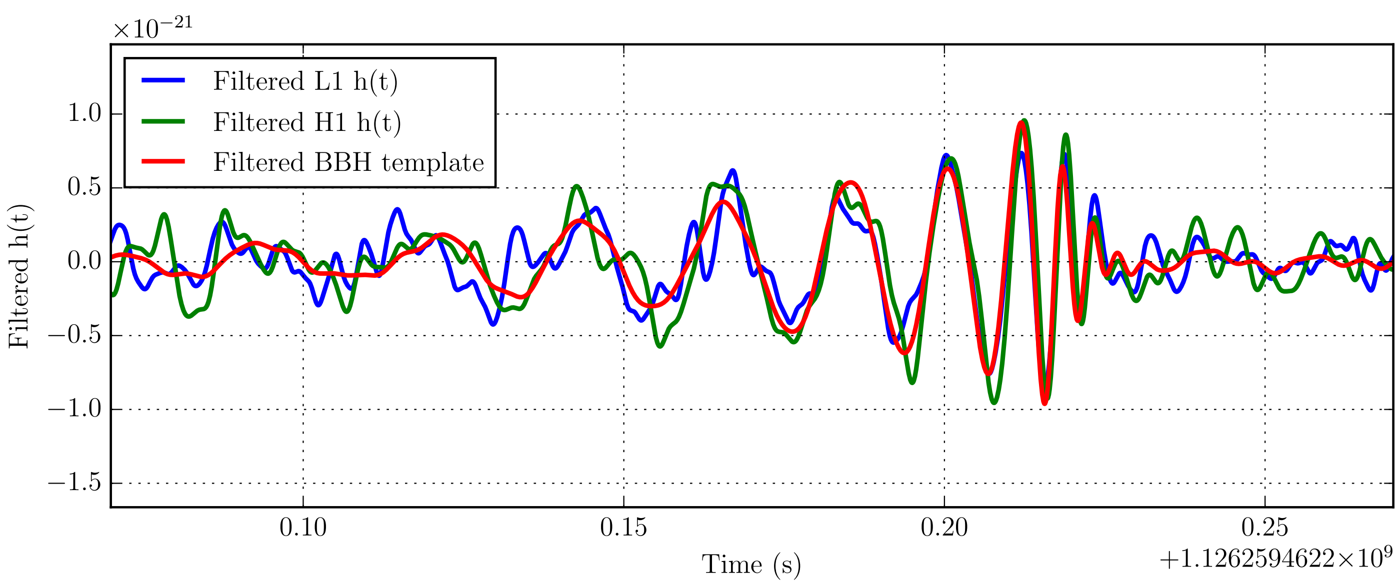
GW data in a perfect world…

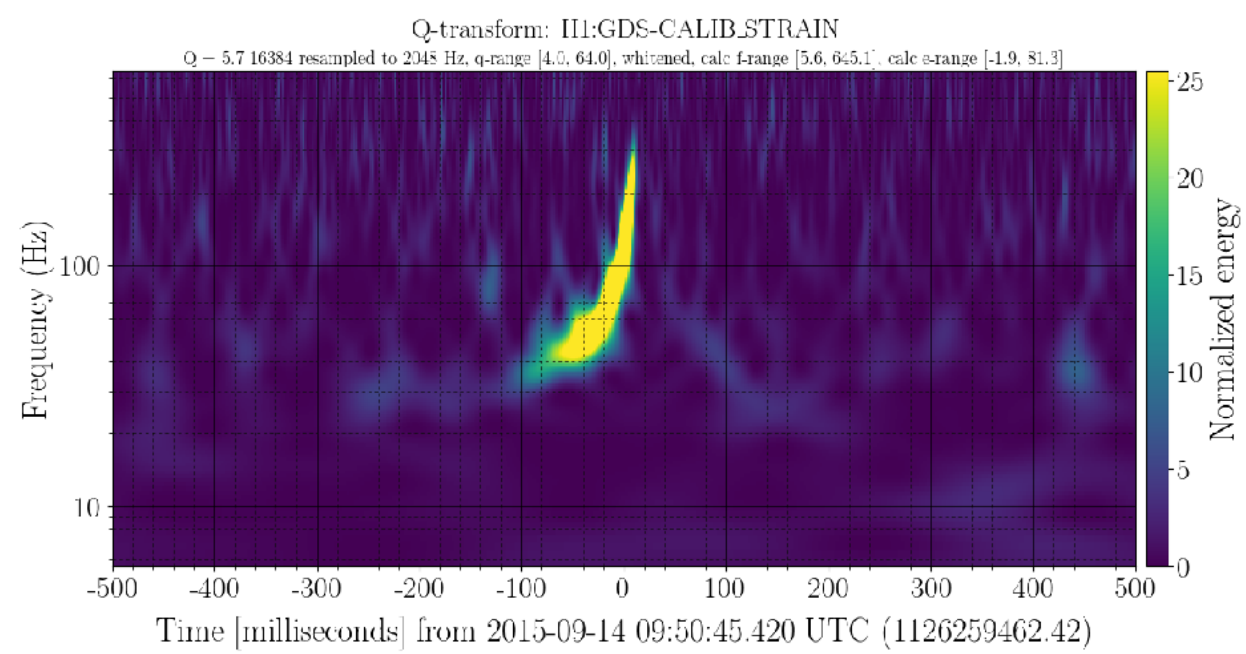
No signal



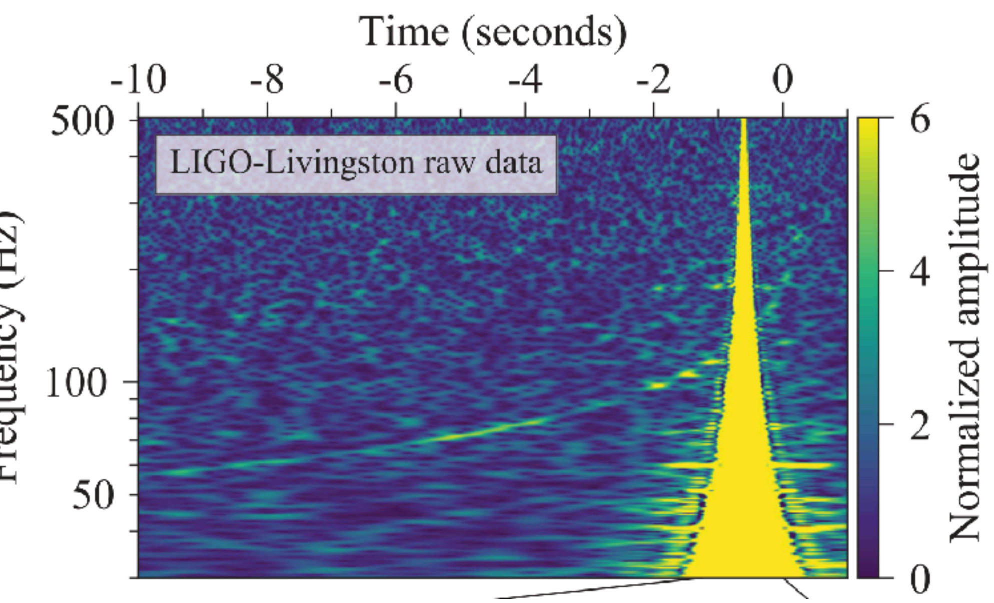


Signal (GW150914)

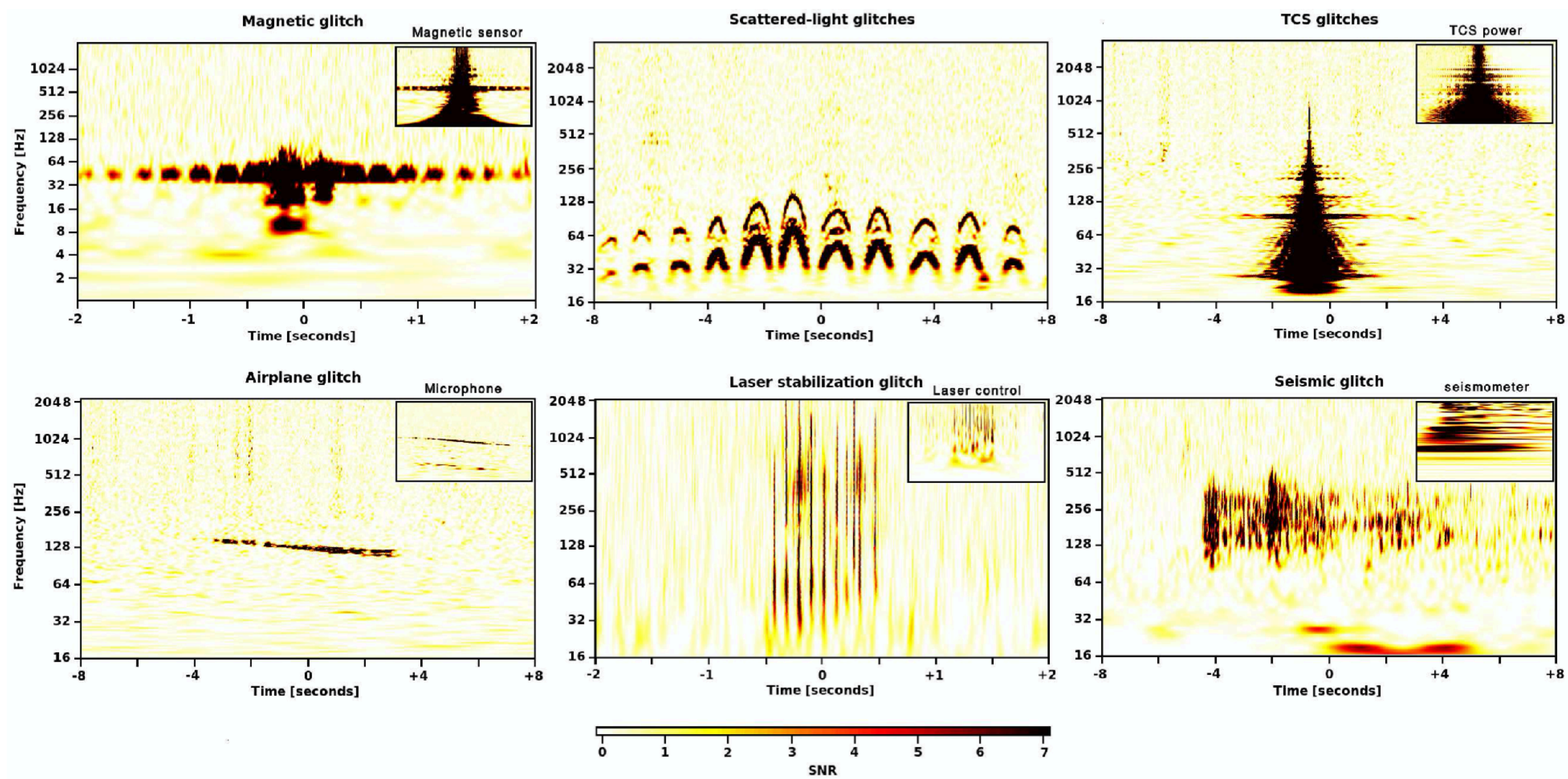




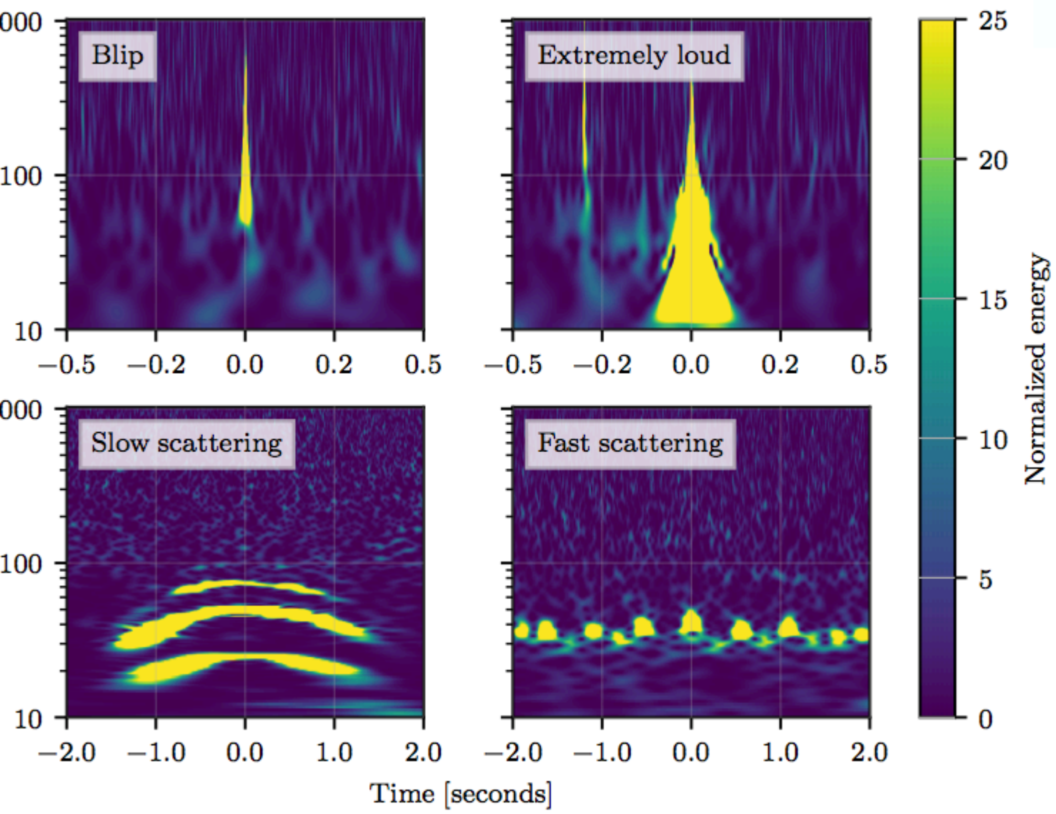
Strain data is non-stationary!



GW170817, arXiv: 1710.05832 (2017)



J. Aasi et al.,*The characterization of Virgo data and its impact on gravitational-wave searches*, arXiv: 1203.5613 (2012)

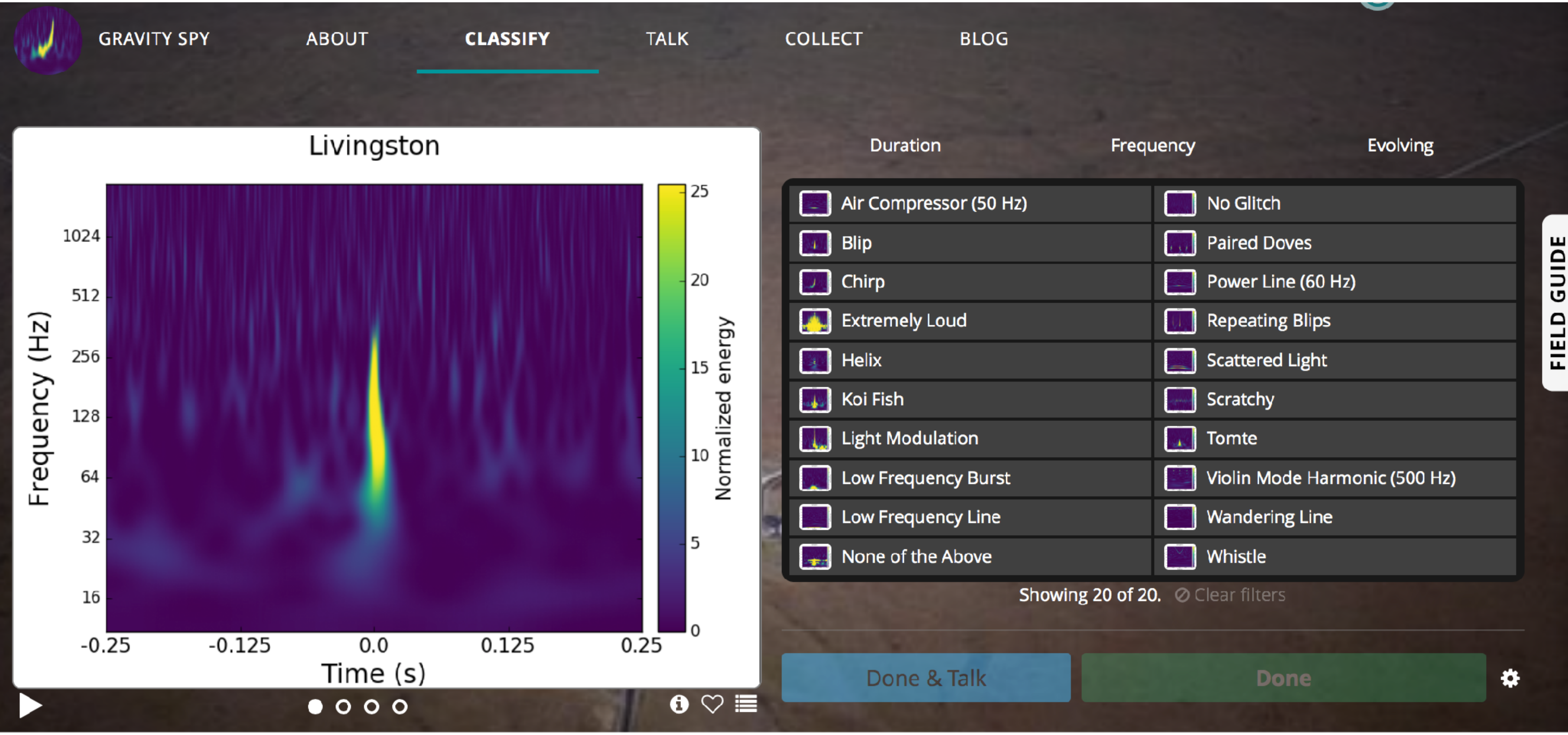


D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

- Multiple type of non-stationary noise
  - Duration
  - Central frequency
  - Characteristic shape
- Glitch
  - Short duration disturbance

Gravity spy: citizen science for glitch classification

gravityspy.org



Instrumental lines

Instrumental lines
= narrow band, persistent artifacts

Sources:

- 60Hz power line harmonics,
- Violin modes of suspension
- Environmental disturbances, etc…

Instrumental lines can degrade the CW search sensitivity because they share the similar features with CGWs leading the increase of the false-alarm rate. Also, lines much affect on the PSD estimation.

13

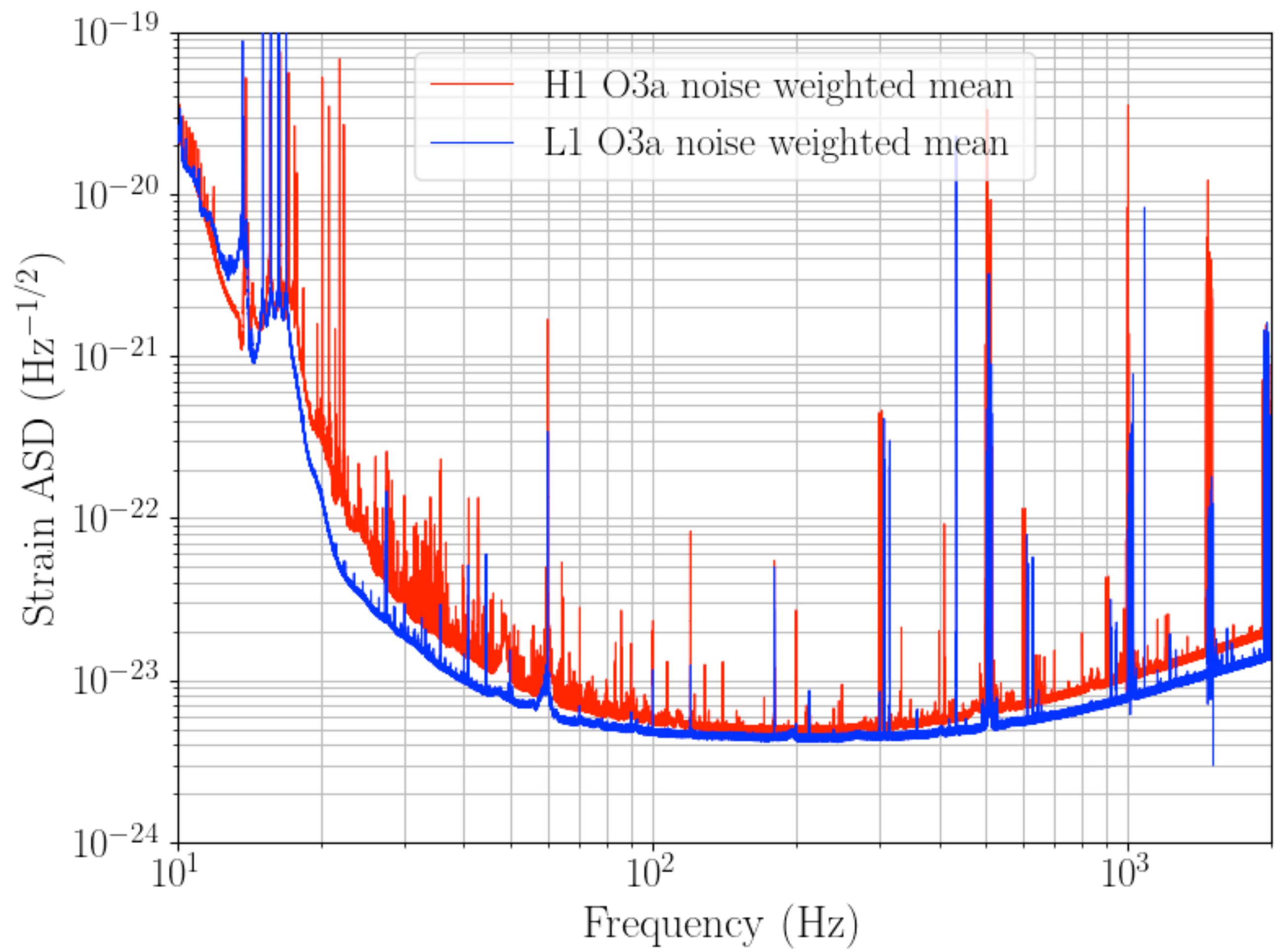


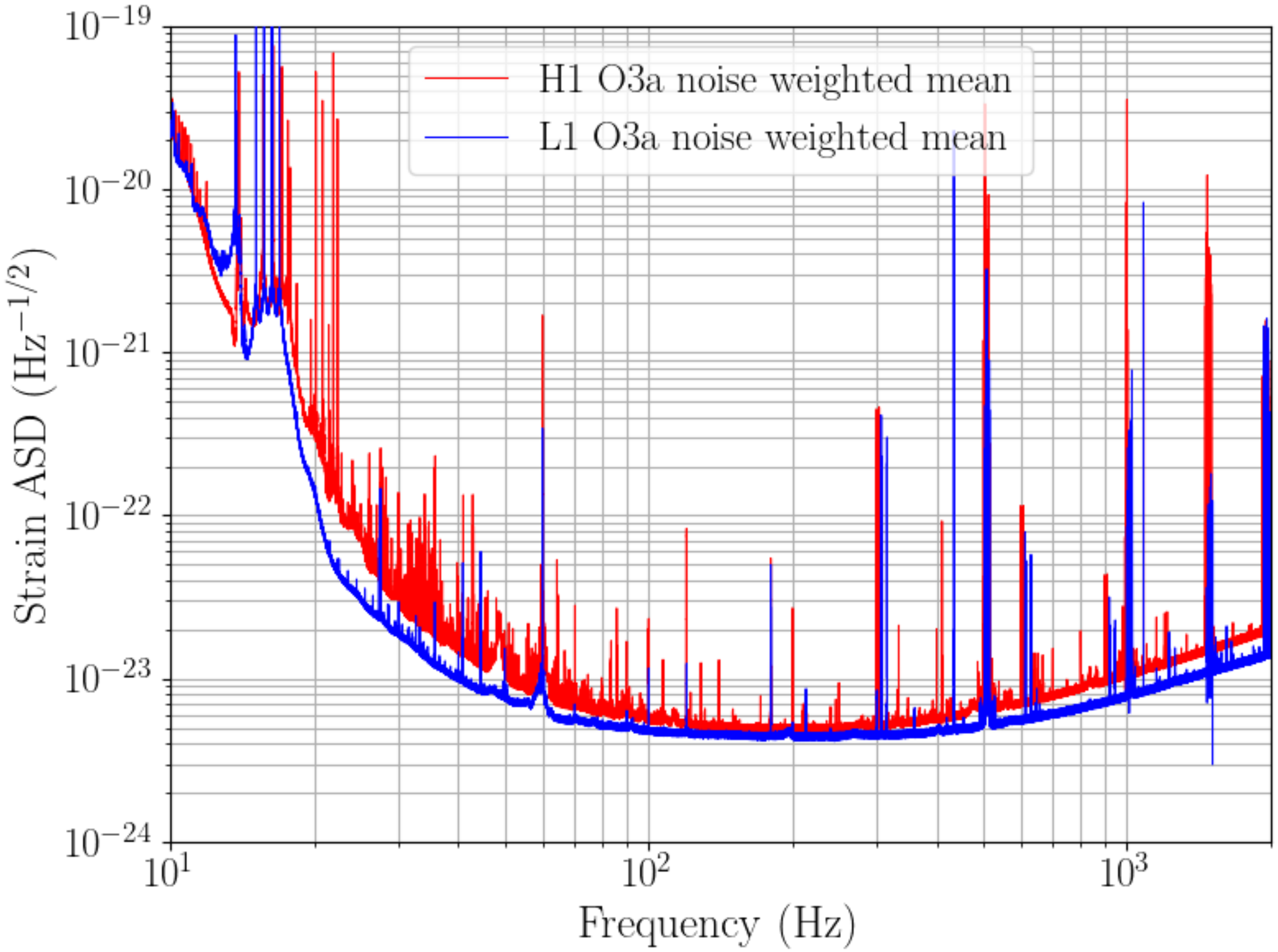
Fig: https://gwosc.org/O3/o3speclines/

Slide by T.S.Yamamoto [LIGO-G2400946]

In reality…GW data also contains instrumental and environmental artifacts

In the frequency domain, it is clear to see many combs of lines in the data.

More information at: https://www.gw-openscience.org/O3/o3aspeclines/



Noise Subtraction for O3 dataset

After data collection we remove several independently measured terrestrial contributions to the detector noise:

- LIGO - remove calibration lines and 60Hz AC power mains harmonics. We also remove some additional noise due to non stationary couplings
- Virgo - remove broadband noise, including frequency noise from the laser, noise introduced when controlling the displacement of the beam splitter and amplitude noise of the 56 MHz modulation frequency.

For details, see https://gwosc.org/O3/o3_details/

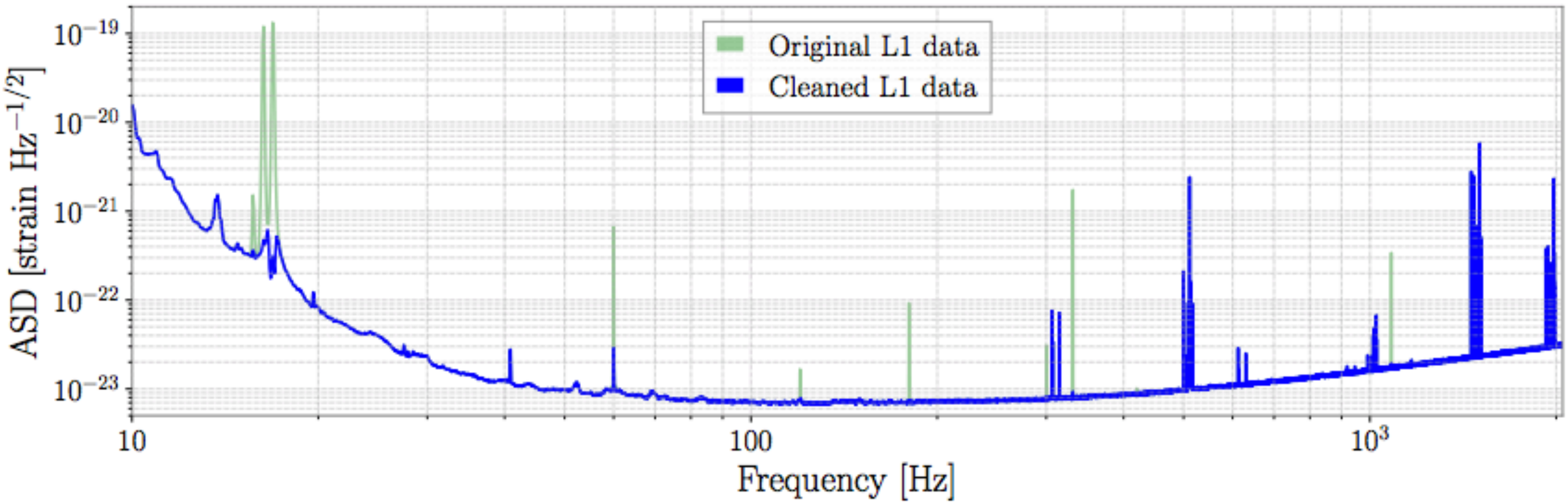


Image: D. Davis et al., *Improving the sensitivity of Advanced LIGO using noise subtraction*, CQG **36** 055011, arXiv: 1809.05348 (2019)

Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

Data Quality Information available in GWOSC

- The non-stationary noise remains in the data of observation run, even though many attempts to identify the origin.
- To reduce the effect of the non-stationary noise to the analysis, we provide the data quality information with several categories (classes)

Bit	Short Name	Description
Data Quality Bits		
0	DATA	data present
1	CBC_CAT1	passes the cbc CAT1 test
2	CBC_CAT2	passes cbc CAT2 test
3	CBC_CAT3	passes cbc CAT3 test
4	BURST_CAT1	passes burst CAT1 test
5	BURST_CAT2	passes burst CAT2 test
6	BURST_CAT3	passes burst CAT3 test
Injection Bits		
0	NO_CBC_HW_INJ	no cbc injection
1	NO_BURST_HW_INJ	no burst injections
2	NO_DETCHAR_HW_INJ	no detchar injections
3	NO_CW_HW_INJ	no continuous wave injections
4	NO_STOCH_HW_INJ	no stoch injections

See https://gwosc.org/archive/dataset/O3a_16KHZ_R1/

Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

CAT1 (Category 1): Failing a data quality check at this category indicates **a critical issue with a key detector component not operating in its nominal configuration**.

- These times are identical for each data analysis group.
- *Times that fail CAT1 flags are not available as LIGO open data.*

Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

CAT1 (Category 1): Failing a data quality check at this category indicates **a critical issue with a key detector component not operating in its nominal configuration**.

- These times are identical for each data analysis group.
- *Times that fail CAT1 flags are not available as LIGO open data.*

CAT2 (Category 2): Failing a data quality check at this category indicates times when there is a **known, understood physical coupling to the gravitational wave channel**. For example, high seismic activity.

Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

CAT1 (Category 1): Failing a data quality check at this category indicates **a critical issue with a key detector component not operating in its nominal configuration**.

- These times are identical for each data analysis group.
- *Times that fail CAT1 flags are not available as LIGO open data.*

CAT2 (Category 2): Failing a data quality check at this category indicates times when there is a **known, understood physical coupling to the gravitational wave channel**. For example, high seismic activity.

CAT3 (Category 3): Failing a data quality check at this category indicates times when there is **statistical coupling to the gravitational wave channel** which is not fully understood.

Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

CAT1 (Category 1): Failing a data quality check at this category indicates **a critical issue with a key detector component not operating in its nominal configuration**.

- These times are identical for each data analysis group.
- *Times that fail CAT1 flags are not available as LIGO open data.*

CAT2 (Category 2): Failing a data quality check at this category indicates times when there is a **known, understood physical coupling to the gravitational wave channel**. For example, high seismic activity.

CAT3 (Category 3): Failing a data quality check at this category indicates times when there is **statistical coupling to the gravitational wave channel** which is not fully understood.

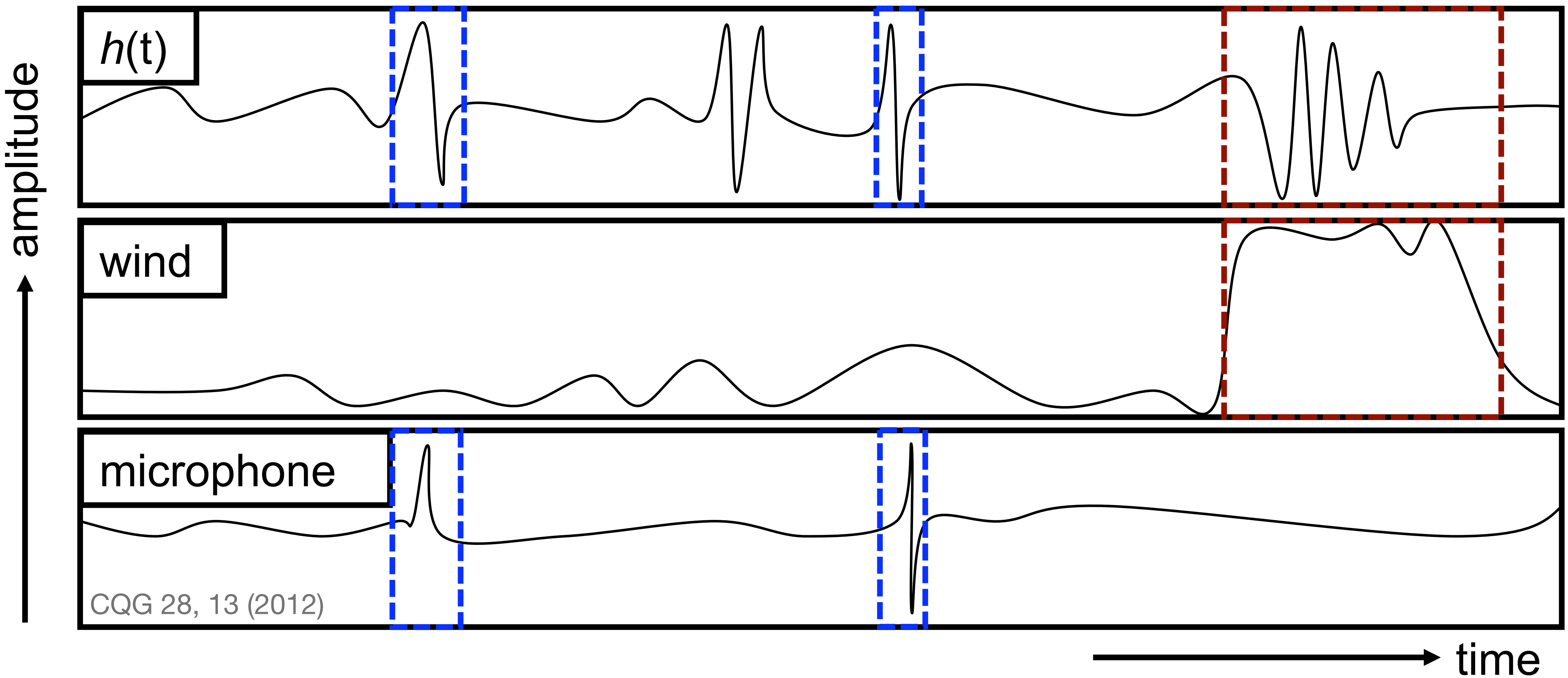
Data quality levels are defined in a cumulative way: a time which fails a given category automatically fails all higher categories.

Data quality categories are defined independently for different analysis groups: if something fails at CAT2_BURST, it could pass CAT2_CBC.

Auxiliary channels

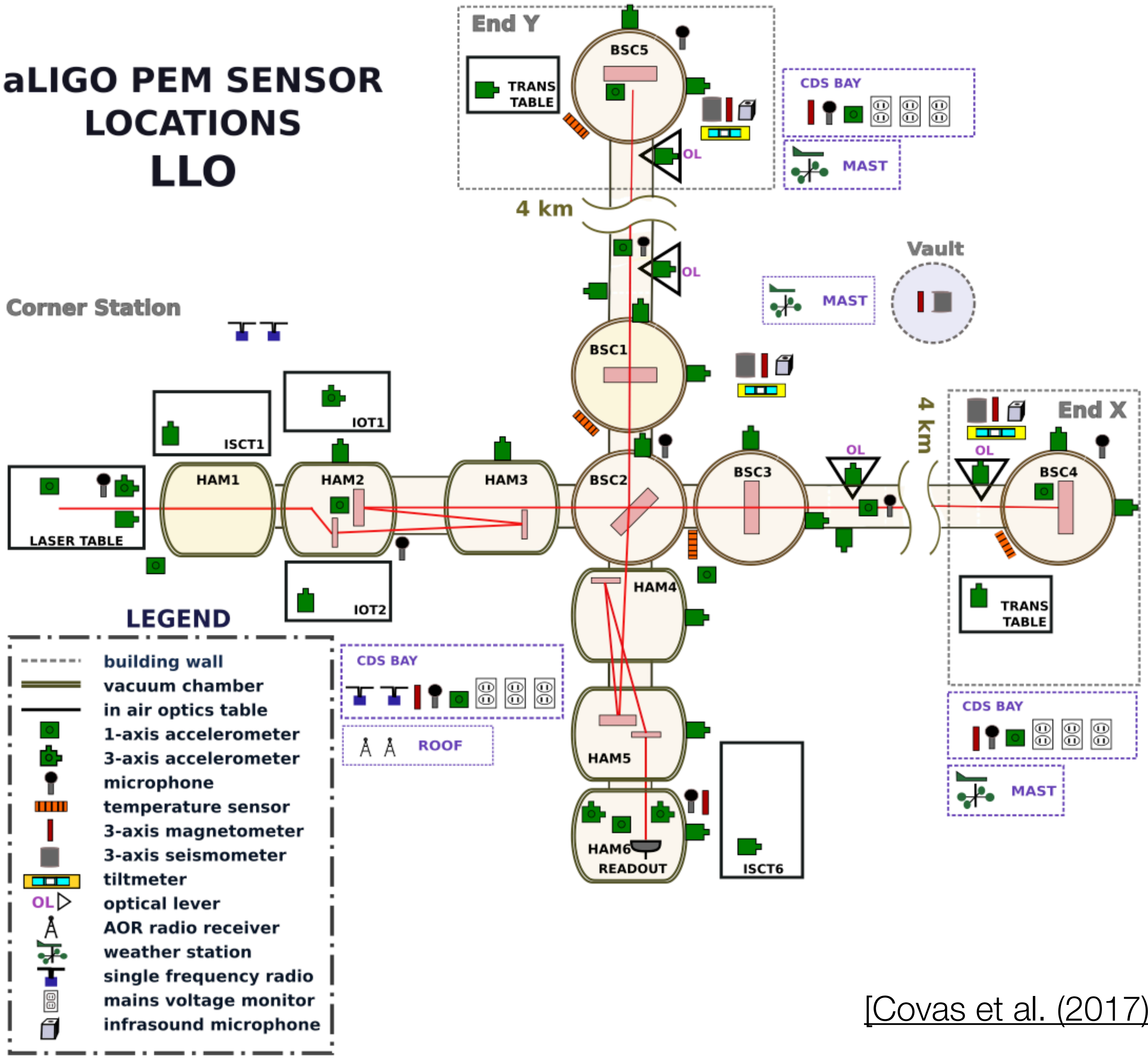
We record **over 200,000 channels per detector** that monitor the environment and detector behavior.

We can use these to help trace the instrumental causes of glitches that pollute the search backgrounds.



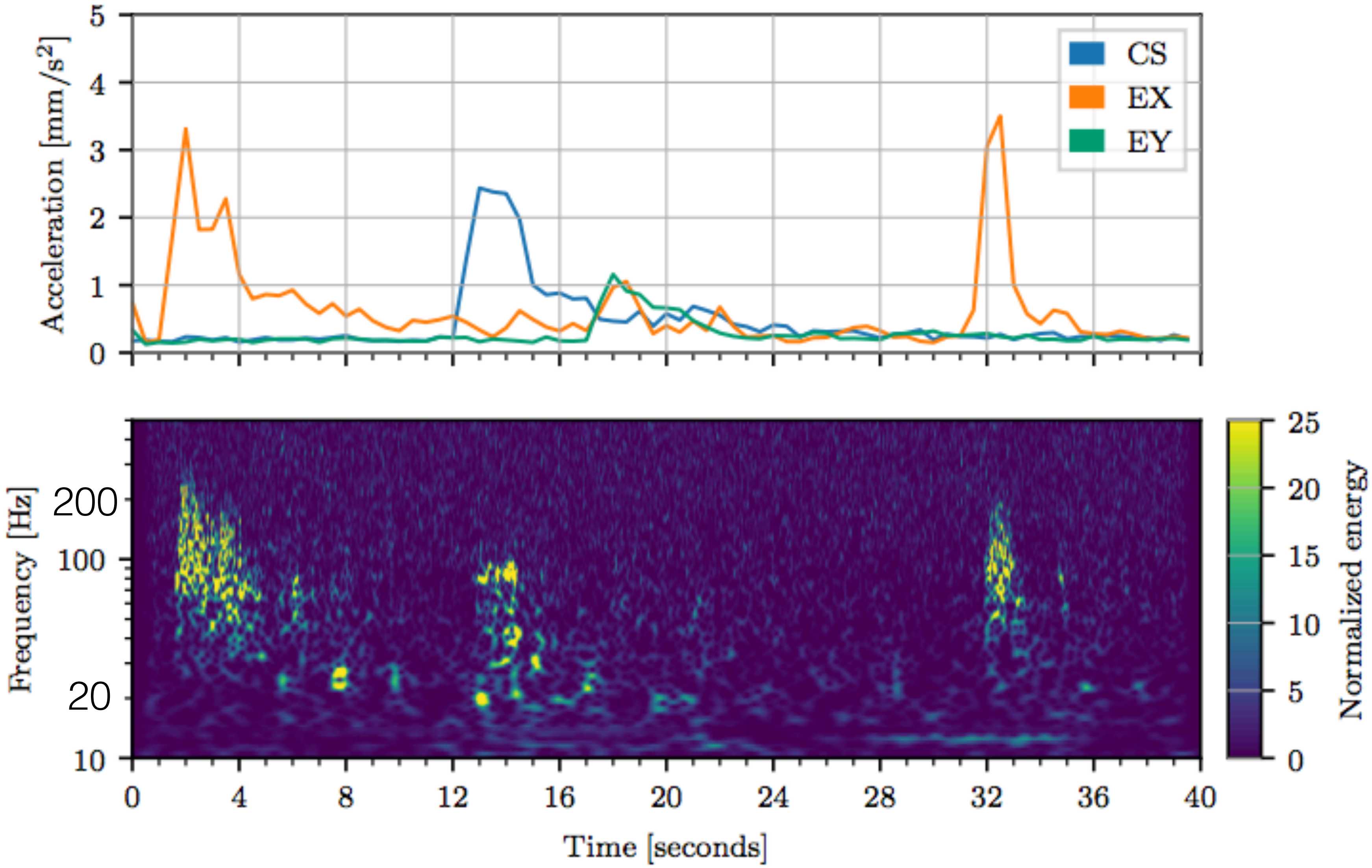
Subset of LIGO's auxiliary channels for O3 are publicly available.
See https://gwosc.org/O3/auxiliary/

Physical environment channels



[Covas et al. (2017), arXiv 1801.07204]

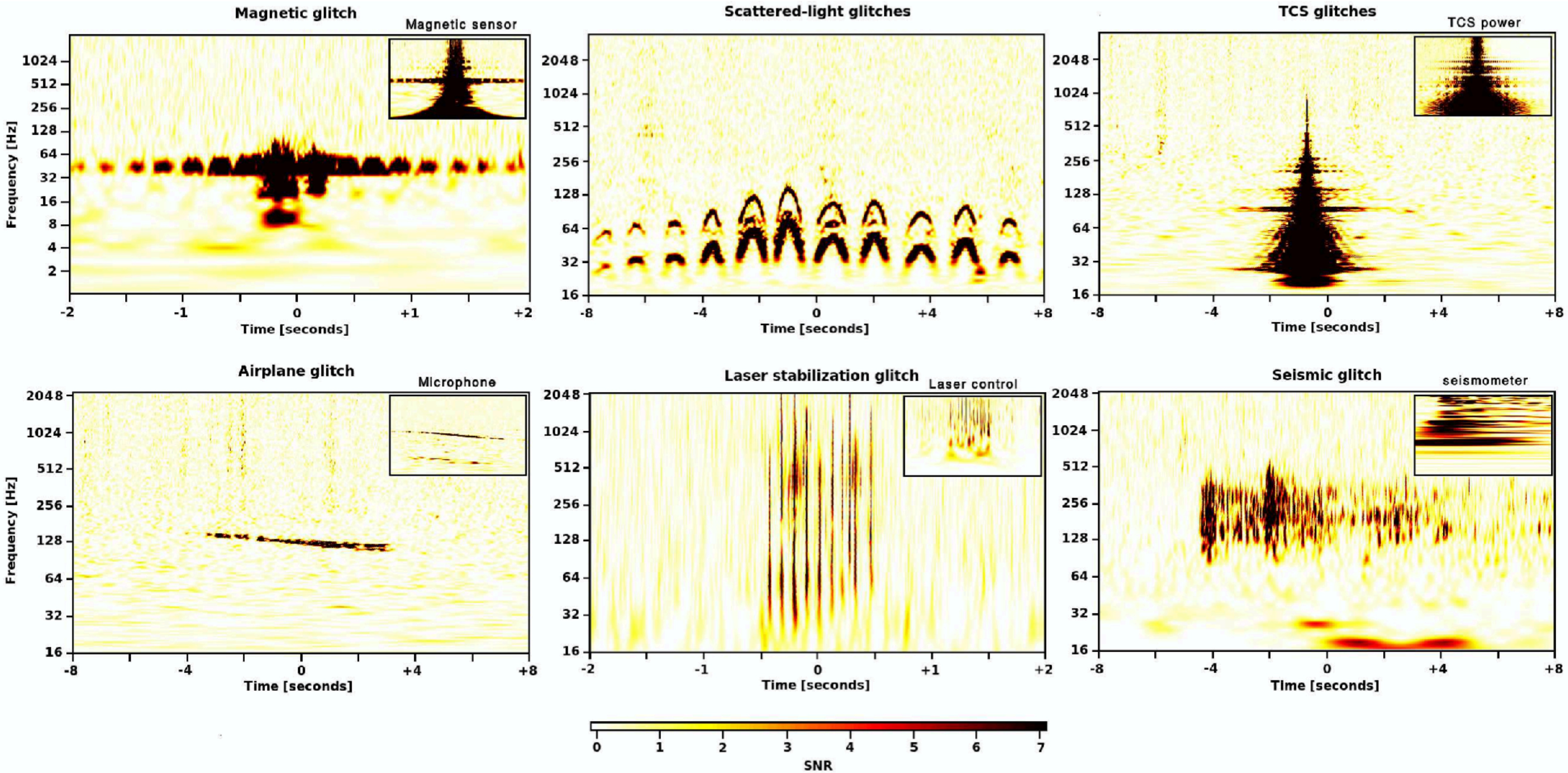
Example of noise coupling: thunderstorms



- Top: Data between 10-100 Hz from accelerometers located in the corner station (CS), End X station (EX) and End Y station (EY)
- Bottom: Spectrogram of the GW strain channel at the same time. Excess noise in the frequency range of 20 Hz to 200 Hz coincides with the thunderclaps, with intensity depending on the thunder's location.

D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

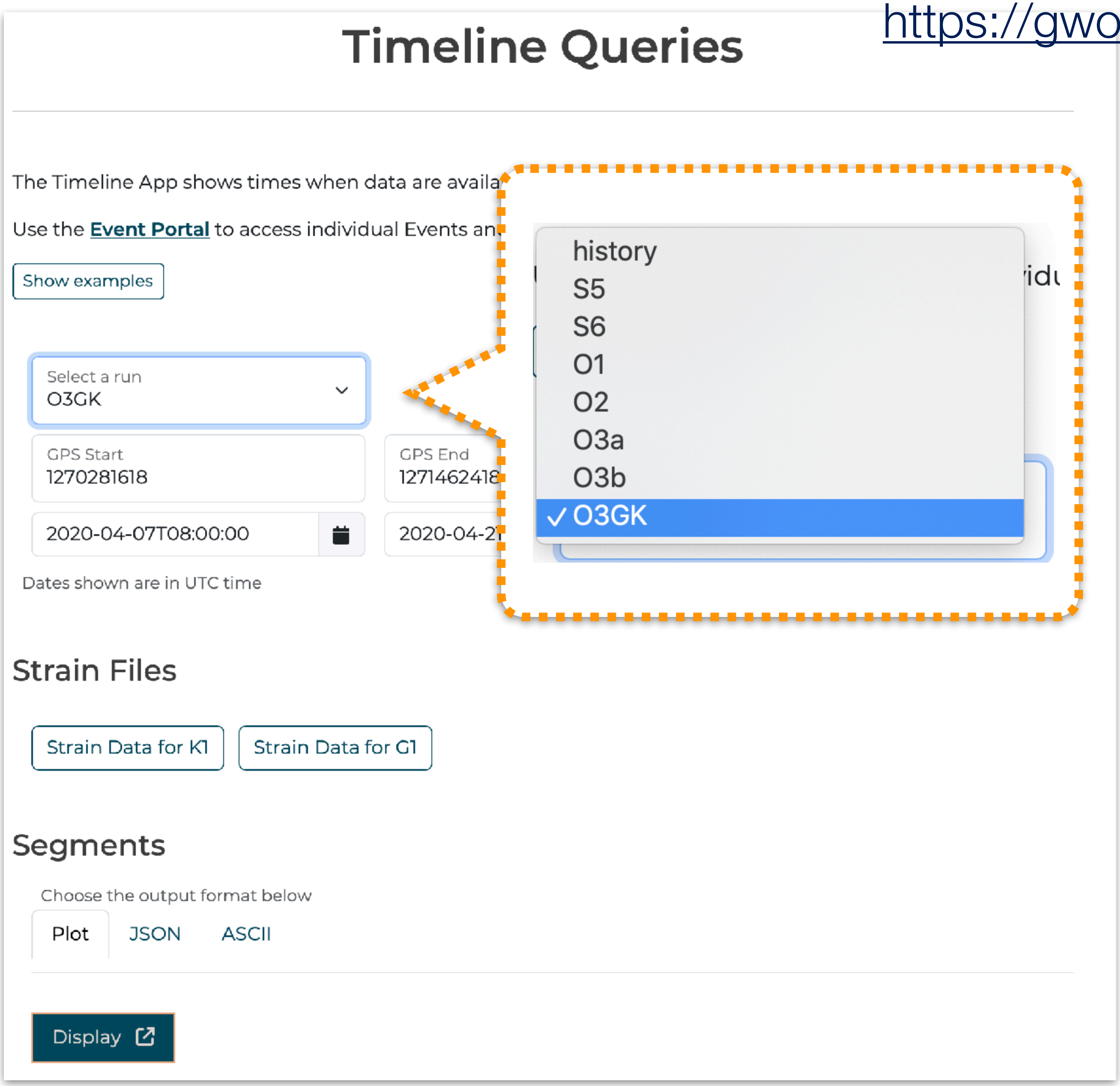
Example of noise coupling: Virgo data



J. Aasi et al.,*The characterization of Virgo data and its impact on gravitational-wave searches*, arXiv: 1203.5613 (2012)

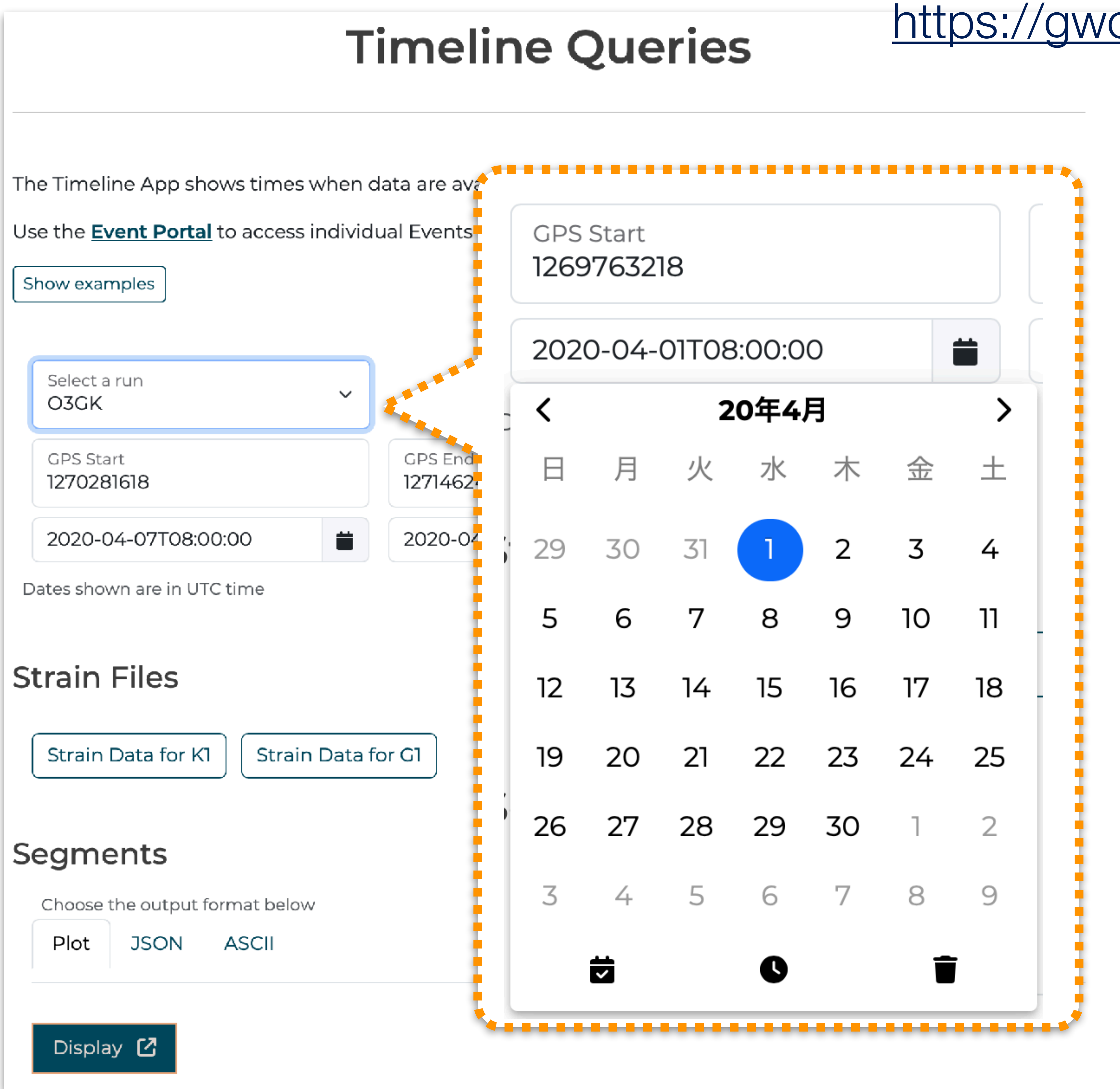
How to get Data Quality Segments: (1) GWOSC Timeline Query

https://gwosc.org/timeline/query/Run/



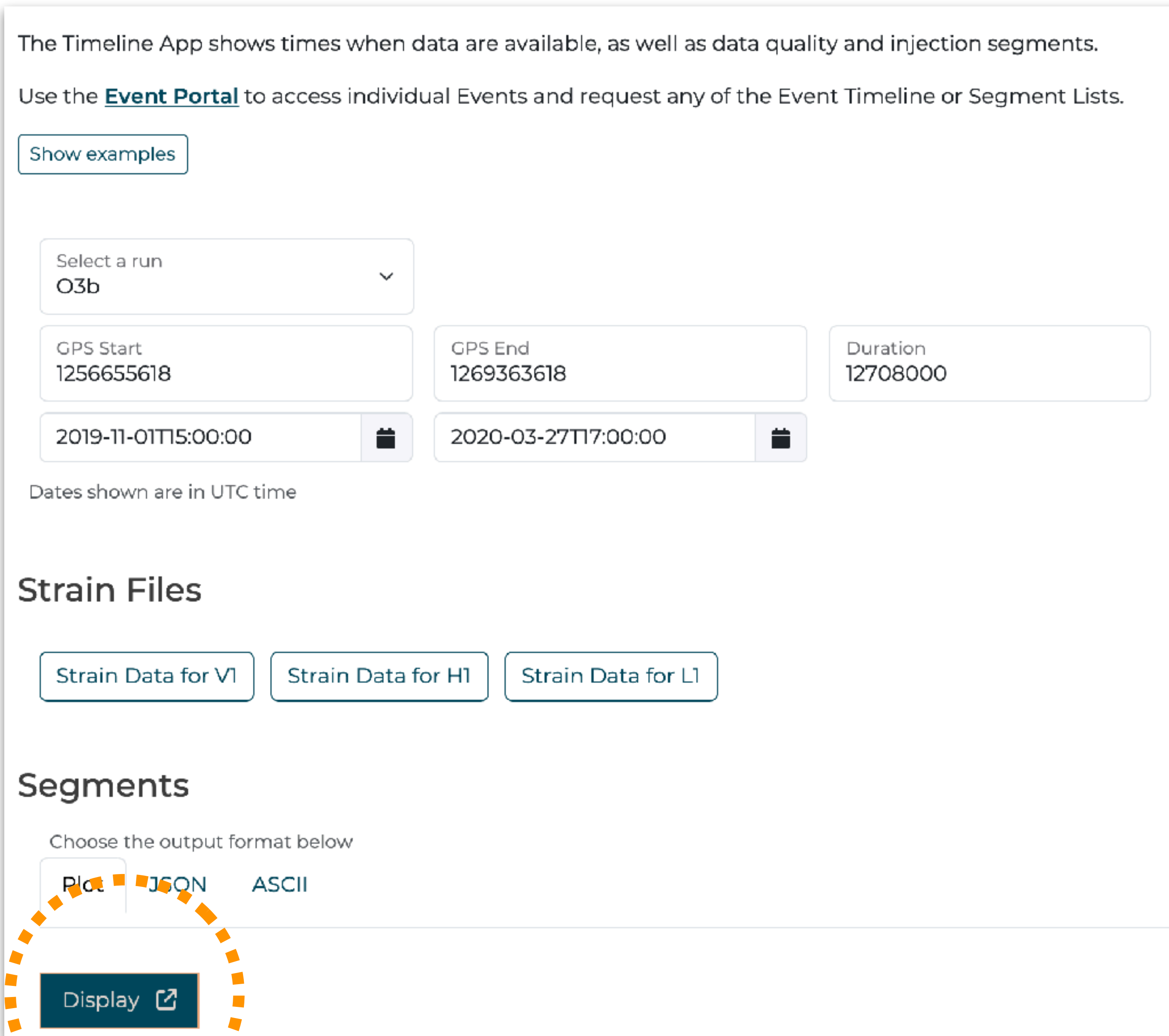
How to get Data Quality Segments: (1) GWOSC Timeline Query

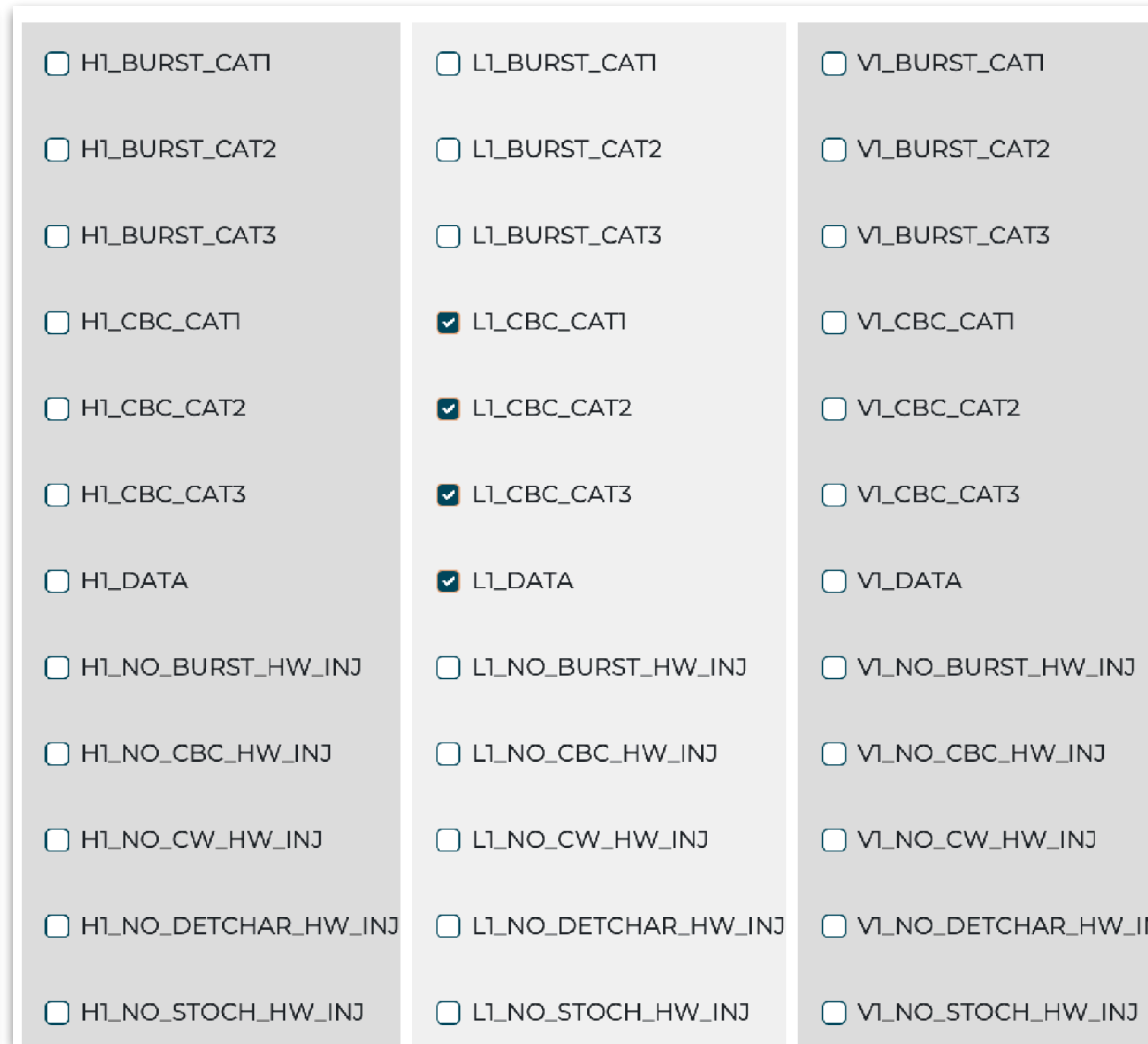
https://gwosc.org/timeline/query/Run/



How to get Data Quality Segments: (1) GWOSC Timeline Query

https://gwosc.org/timeline/query/Run/





Here I selected the O3b data flags for L1.

How to get Data Quality Segments: (1) GWOSC Timeline

https://gwosc.org/timeline/query/Run/

	Time Active	Duty Cycle	Segments
L1_CBC_CAT1	9810816 s	77.20%	352
L1_DATA	9810816 s	77.20%	352
L1_CBC_CAT2	9782946 s	76.98%	9495
L1_CBC_CAT3	9782946 s	76.98%	9495

1256655666	1256657739	2073
1256663958	1256673192	9234
1256673691	1256756709	83018
1256766682	1256772748	6066
1256775368	1256798948	23580
1256809201	1256821151	11950
1256821199	1256821222	23
1256827739	1256835798	8059
1256838093	1256838252	159
1256838369	1256838835	466
1256838869	1256838894	25
1256874558	1256890940	16382
1256891012	1256903452	12440
1256903719	1256910776	7057
1256915116	1256916066	950
1256918388	1256920060	1672
1256920132	1256920380	248
1256920452	1256923518	3066
1256958170	1256959868	1698
1256965469	1256967668	2199
1256971138	1256973055	1917
1256975290	1256977603	2313
1256983637	1257001290	17653

Segments can be plotted (with interactive zooming) or downloaded

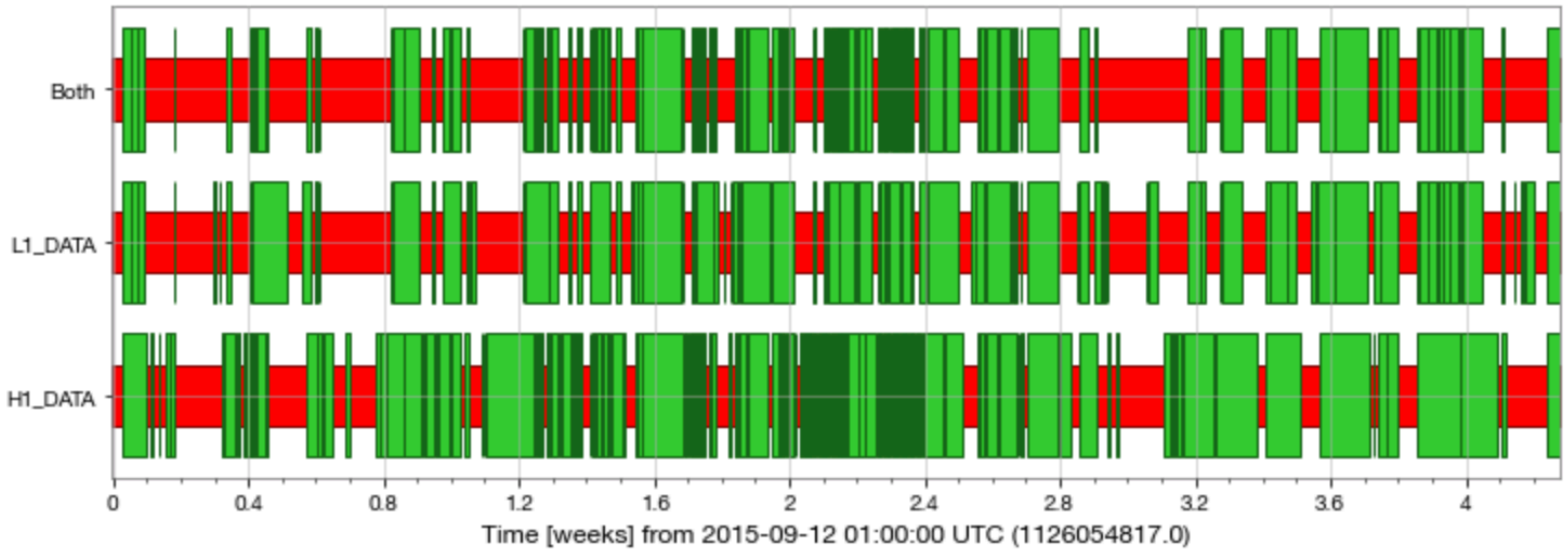
How to get Data Quality Segments: (2) using gwpy

Example of how to find and plot data quality segments:
https://gwpy.github.io/docs/stable/examples/segments/open-data/

[11]:
```
from gwpy.segments import DataQualityFlag
%matplotlib inline

h1month1 = DataQualityFlag.fetch_open_data('H1_DATA', 'Sep 12 2015',
                                           'Oct 12 2015')
l1month1 = DataQualityFlag.fetch_open_data('L1_DATA', 'Sep 12 2015',
                                           'Oct 12 2015')
bothon = h1month1 & l1month1
plot = h1month1.plot()
ax = plot.gca()
ax.plot(l1month1)
ax.plot(bothon, label='Both')
```

[11]: [<matplotlib.collections.PatchCollection at 0x15b4bfe80>]



How to get Data Quality Segments: (2) using gwpy

Example of how to find and plot data quality segments:
https://gwpy.github.io/docs/stable/examples/segments/open-data/

```
[12]: bothon

[12]: <DataQualityFlag('H1:DATA',
                      known=[[1126051217 ... 1128643217)]
                      active=[[1126073529 ... 1126086553)
                              [1126088511 ... 1126098729)
                              [1126100142 ... 1126111138)
                            ...
                              [1128466792 ... 1128504361)
                              [1128538016 ... 1128539723)
                              [1128618578 ... 1128641735)]
                      description=None)>
```

Period which you referred

Period which `H1_DATA` and `L1_DATA` flags are active.

```
[13]: bothon.active

[13]: <SegmentList([Segment(1126073529, 1126086553)
                    Segment(1126088511, 1126098729)
                    Segment(1126100142, 1126111138)
                    Segment(1126164689, 1126165915)
                    Segment(1126256829, 1126264691)
                    Segment(1126301250, 1126301382)
                    Segment(1126301717, 1126301897)
                    Segment(1126301957, 1126302050)
                    Segment(1126302274, 1126302437)
                    Segment(1126302497, 1126302557)
                    Segment(1126302617, 1126304365)
                    Segment(1126304777, 1126304897)
                    Segment(1126305077, 1126305197)
                    Segment(1126305377, 1126305437)
                    Segment(1126305557, 1126305617)
```

Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

Public daily detector status (O2, O3, O4a)

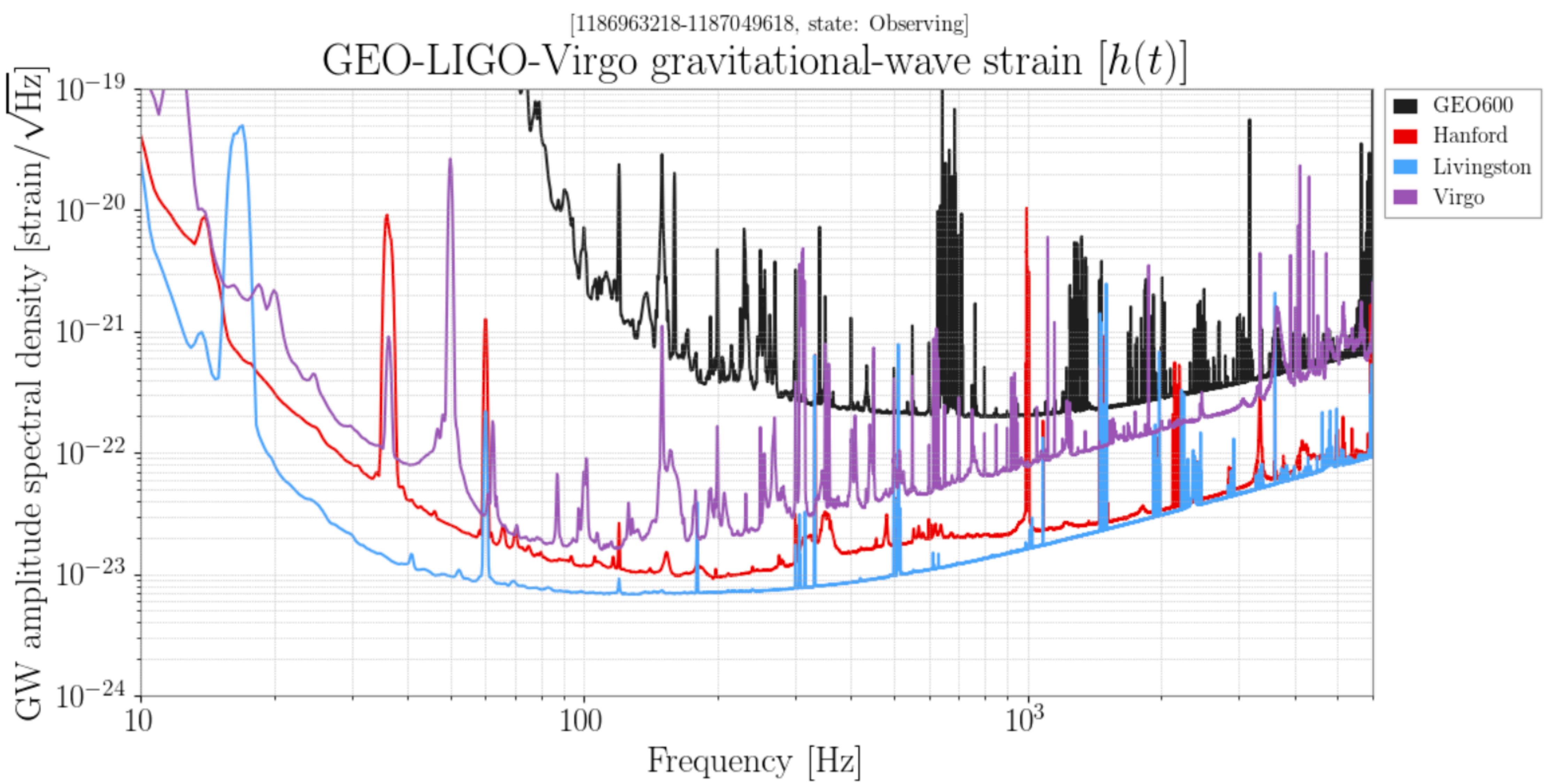
https://www.gw-openscience.org/detector_status/day/20170817/

« August 17 2017 ▾ » Summary Home Environment ▾ Instrument performance ▾

Summary

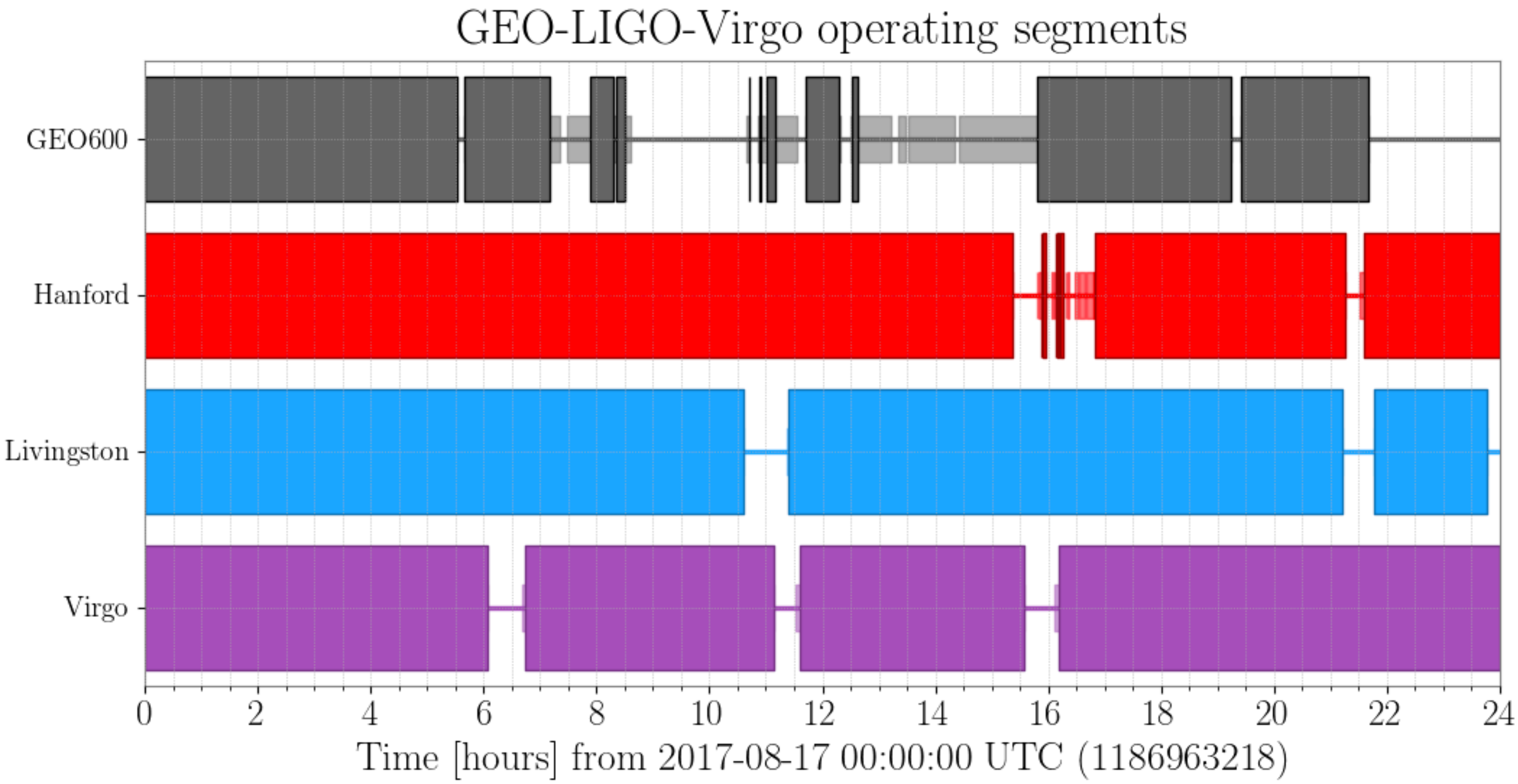
Date selection

The plots shown below characterize the sensitivity and status of each of the LIGO interferometers as well as the Virgo detector in Cascina, Italy and the GEO600 detector in Hanover, Germany.
For more information about the plots listed below, click on an image to read the caption. Use the tabs in the navigation bar at the top of the screen for more detailed information about the LIGO, Virgo, and GEO interferometers.



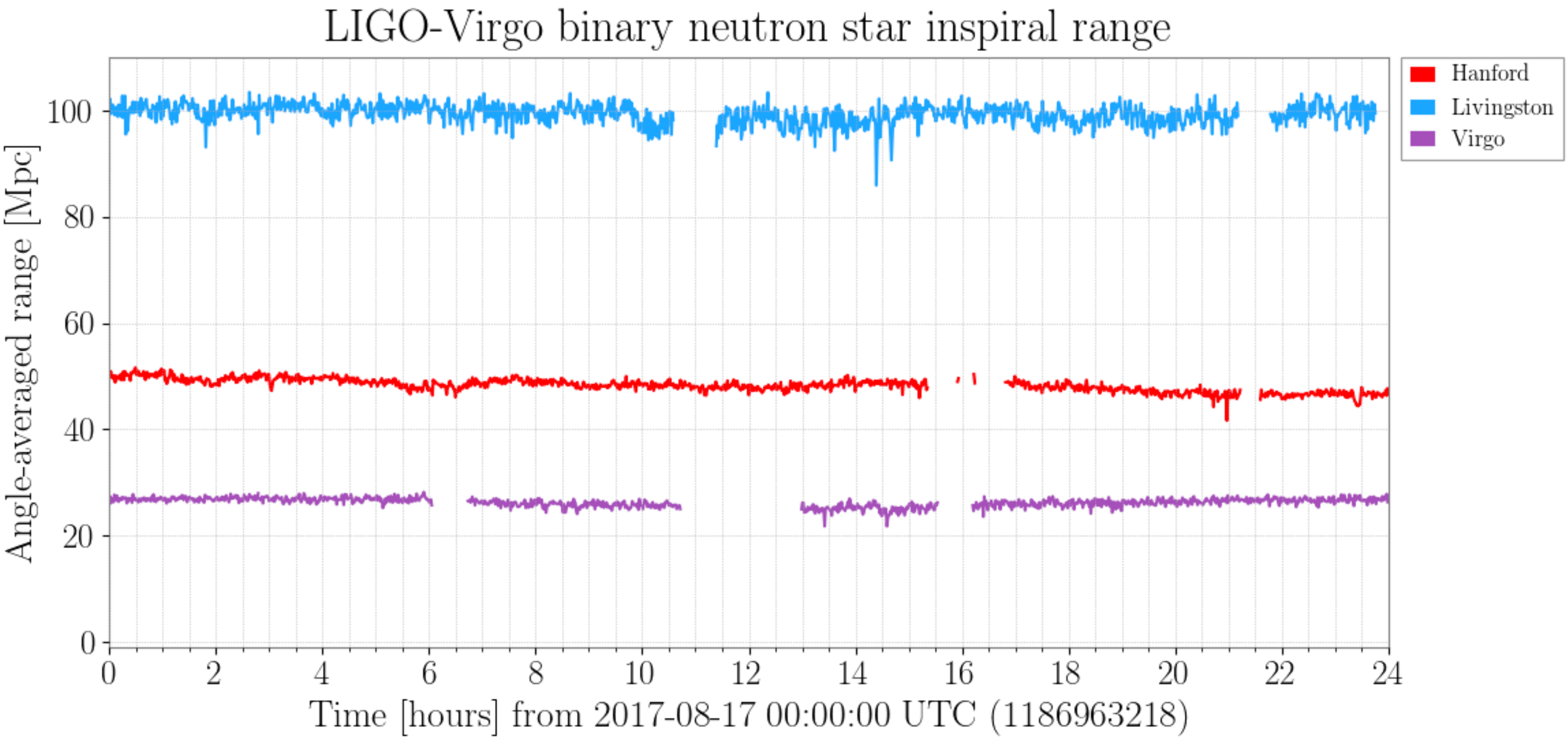
Public daily detector status (O2, O3, O4a)

https://www.gw-openscience.org/detector_status/day/20170817/



Public daily detector status (O2, O3, O4a)

https://www.gw-openscience.org/detector_status/day/20170817/

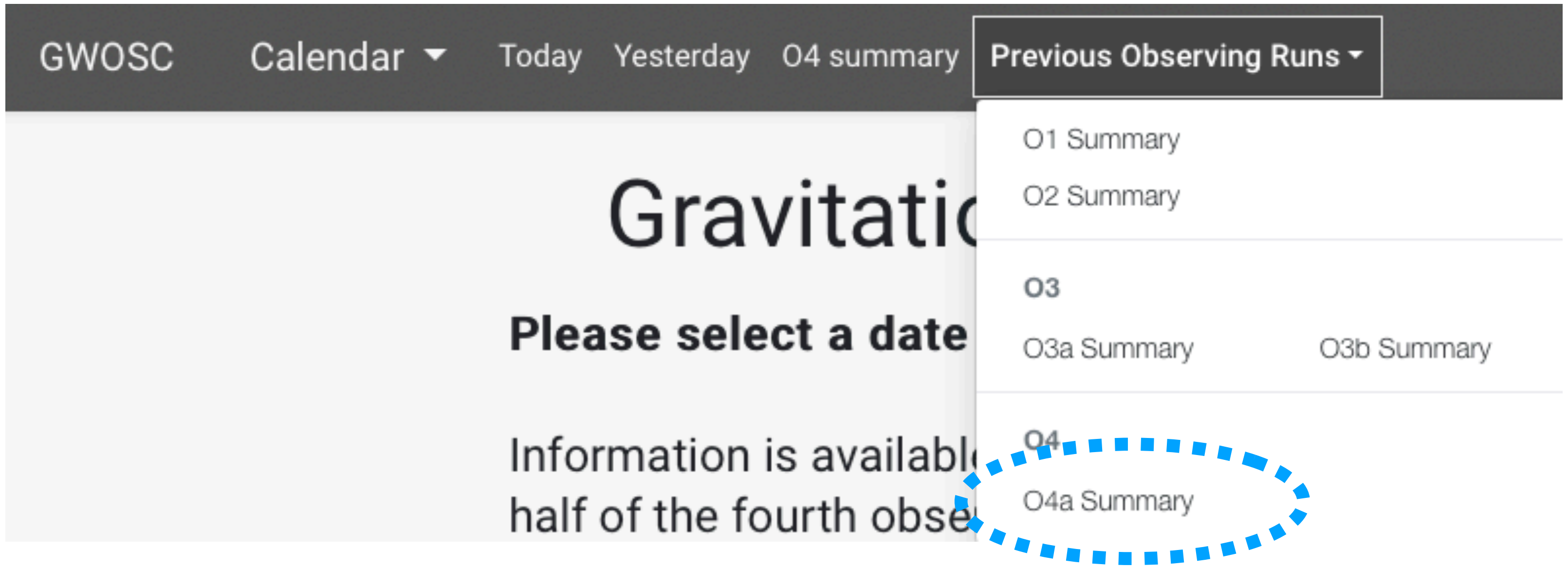


Inspiral range: indicator of sensitivity
how far we can detect the signal from BNS, on average

For inspiral range, see https://gwpy.github.io/docs/stable/examples/miscellaneous/range-timeseries/

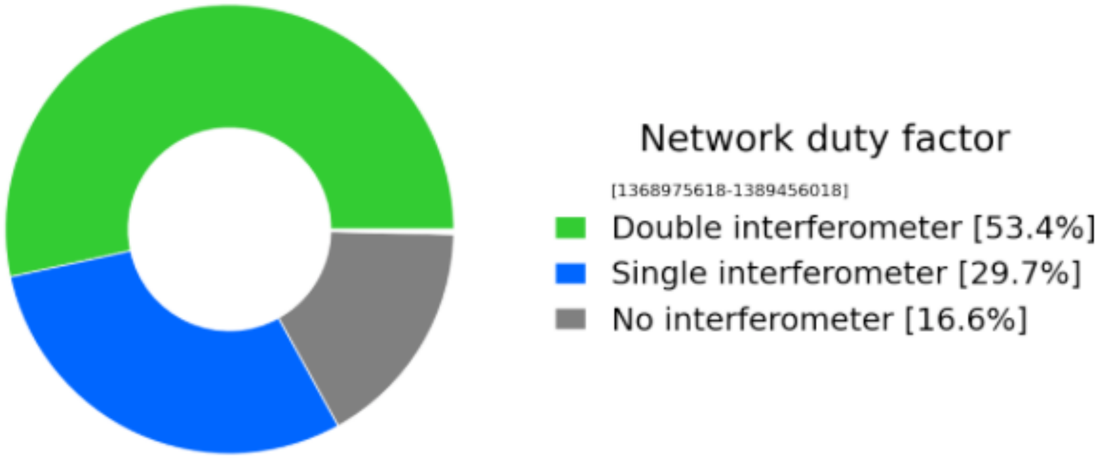
Public daily detector status (O2, O3, O4a)

https://gwosc.org/detector_status/



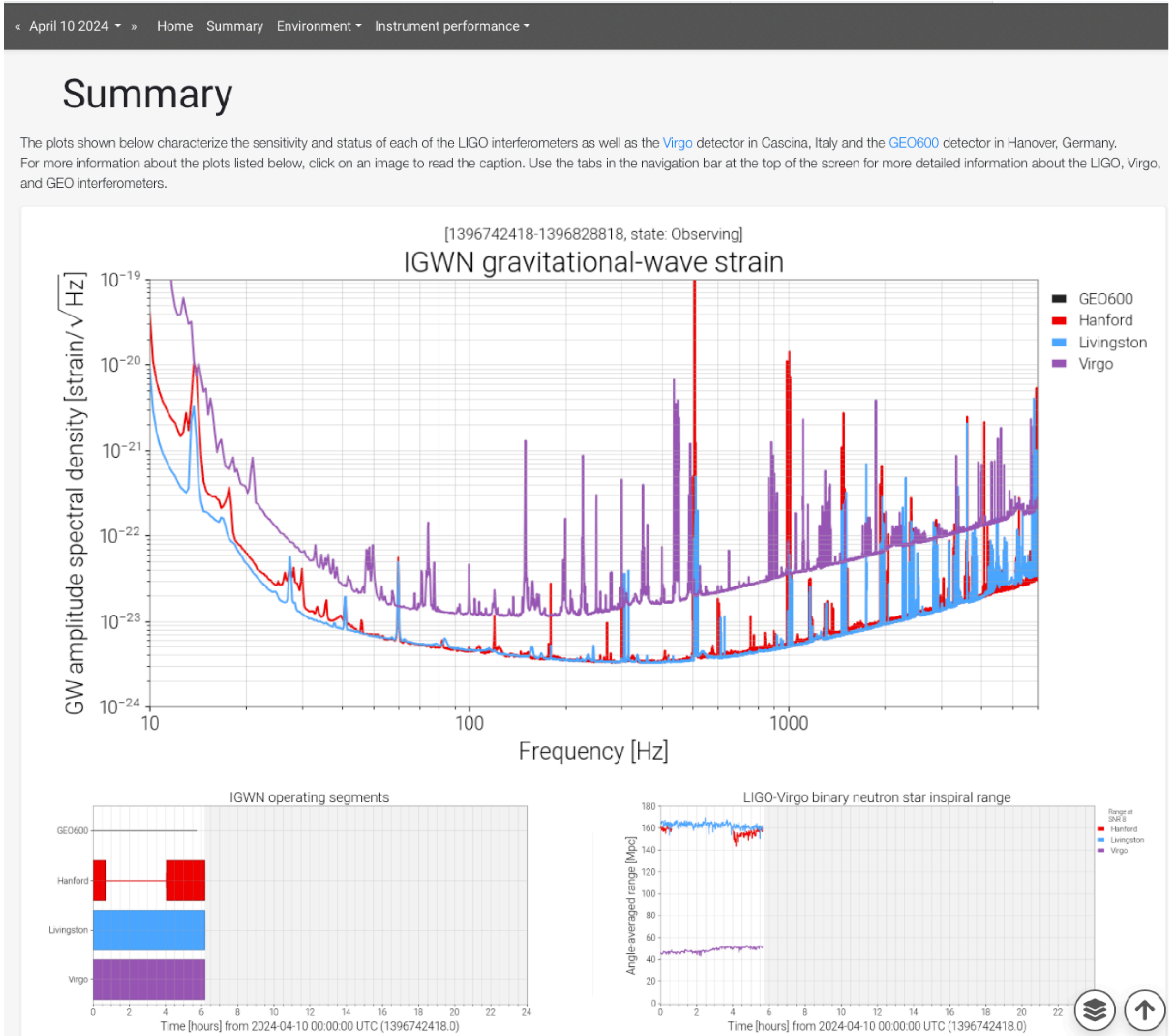
https://gwosc.org/detector_status/O4a/

Includes summary plots of LIGO segments and sensitivity over the run



Observing run summaries (O4b)

https://gwosc.org/detector_status/day/20240410/



Outline

- What does GW data look like?
  - Time domain, Frequency domain, Time-frequency map
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines
- Mitigating noise artifacts
  - Data quality information
  - Noise coupling, Physical environment channels
- Monitor tool, Public summary page
- Reference

Useful references for data quality

For glitches:
GW150914 Detector Characterization paper: arXiv 1602.03844
O2/O3 LIGO Detector Characterization paper: arXiv: 2101.11673
O3 Virgo Detector Characterization paper: arXiv: 2205.01555
Gravity Spy: gravityspy.org

For lines:
O1/O2 lines paper: arXiv 1801.07204
O2 lines catalog on the GWOSC: https://www.gw-openscience.org/o2speclines/
O3 lines calico on GWOSC: https://www.gw-openscience.org/O3/o3aspeclines/

Data Quality around events:
GWTC-2 paper: arXiv: 2010.14527
GWTC-3 paper: arXiv: 2111.03606

Data quality segments:
Data quality timelines: https://www.gw-openscience.org/timeline/

O3a Data Set technical Details: https://www.gw-openscience.org/O3/o3a_details/

Public interferometer status monitoring: https://www.gw-openscience.org/detector_status/

O4a public alerts: https://gracedb.ligo.org/superevents/public/O4/

GWpy documentation: https://gwpy.github.io/