Fractal analysis of LIGO data
a.k.a.
How to characterize interferometric noise in low latency

Marco Cavaglià

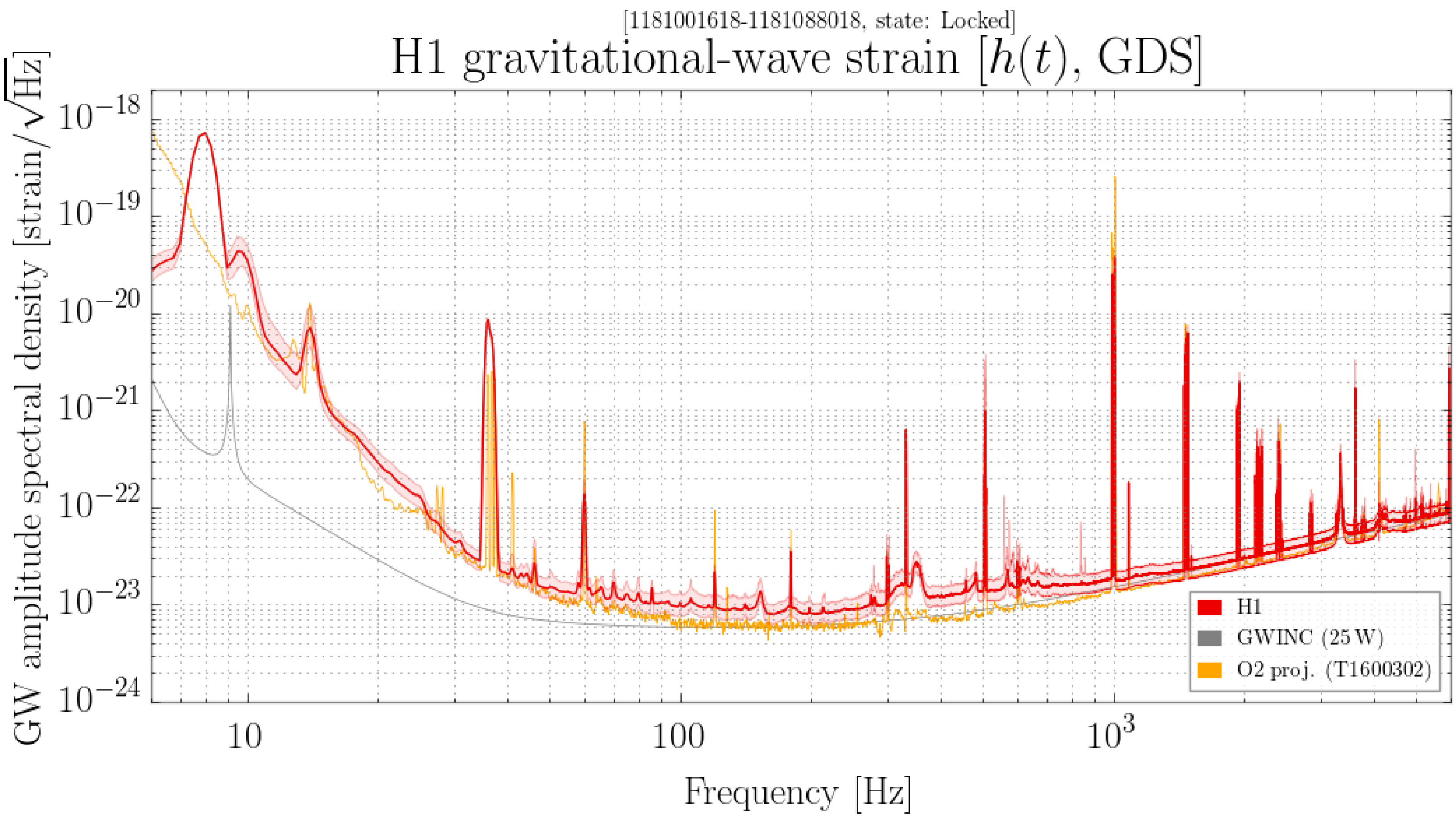
MISSOURI S&T

INSTITUTE FOR MULTI-MESSENGER ASTROPHYSICS AND COSMOLOGY

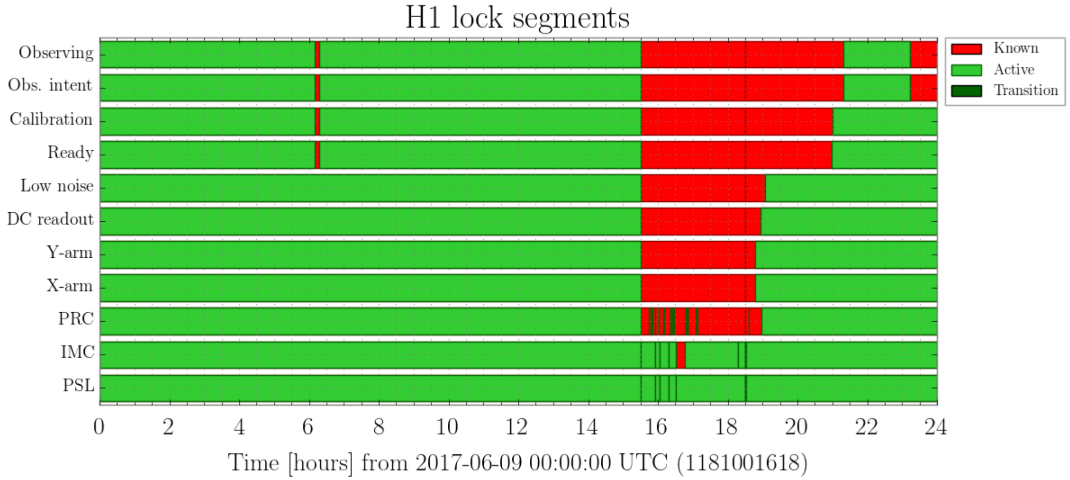
Today is Juneteenth!

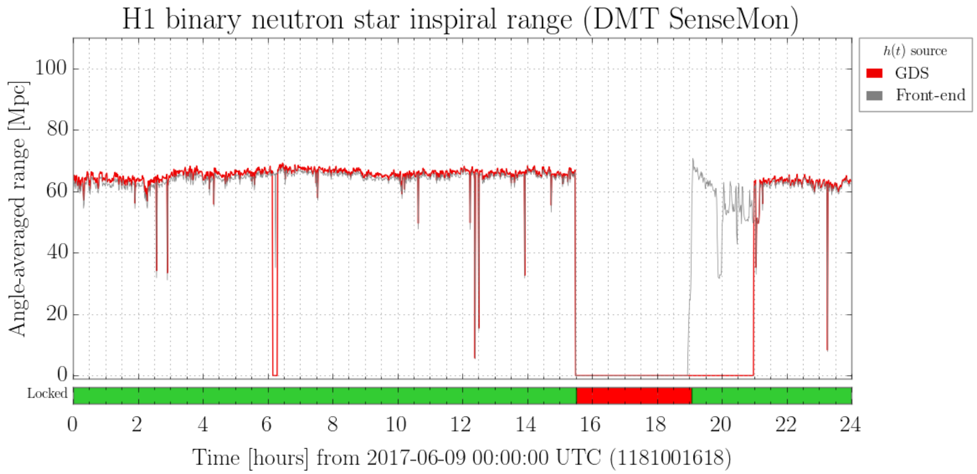
Juneteenth is the oldest nationally celebrated commemoration of the ending of slavery in the United States. Dating back to 1865, it was on June 19th that the Union soldiers, led by Major General Gordon Granger, landed at Galveston, Texas with news that the war had ended and that the enslaved were now free. It is now a federal holiday in the U.S.

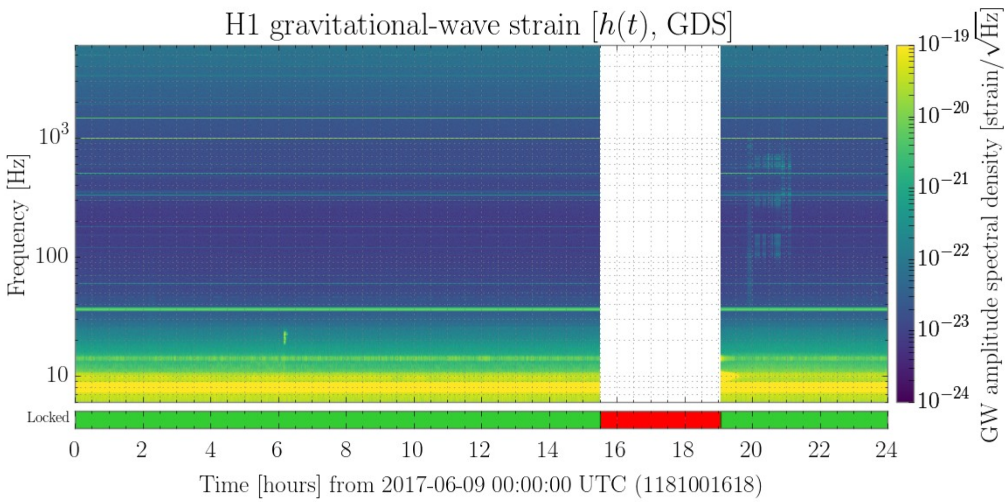
A typical day in LIGO

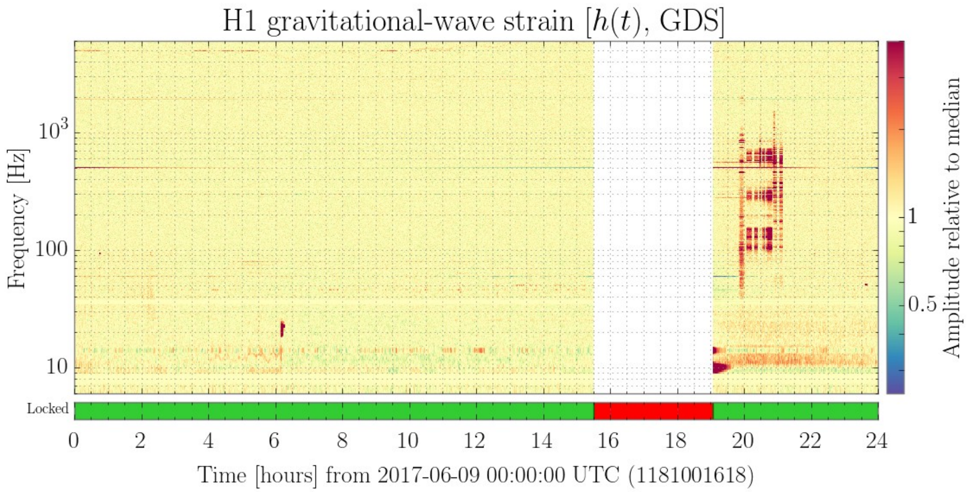


A typical day in LIGO

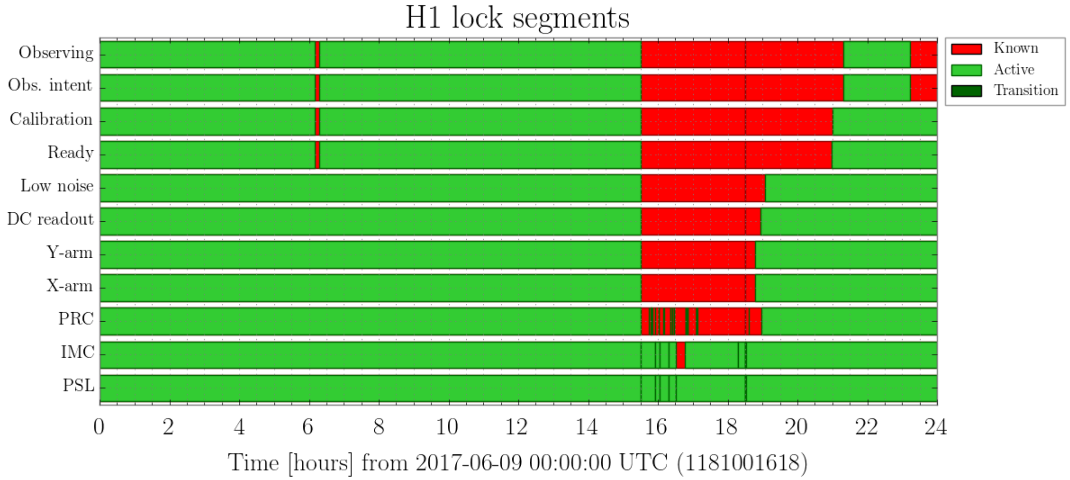


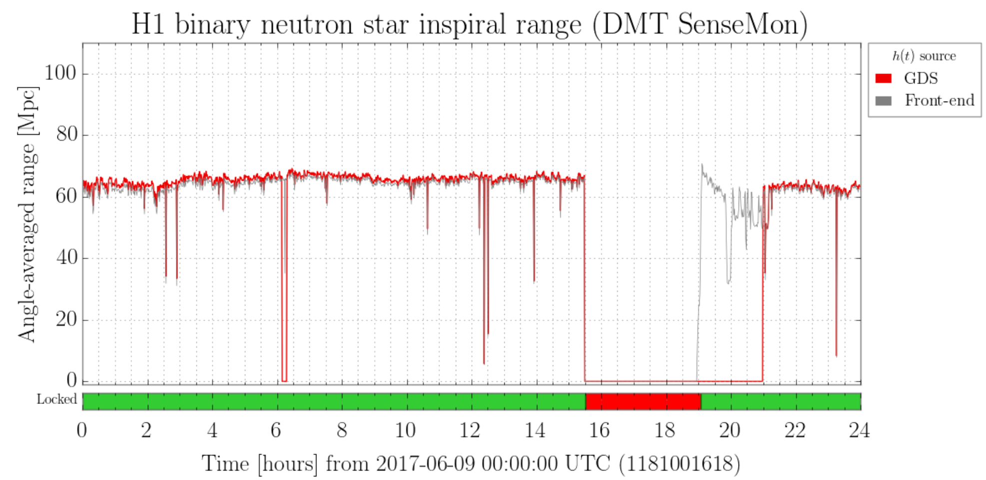


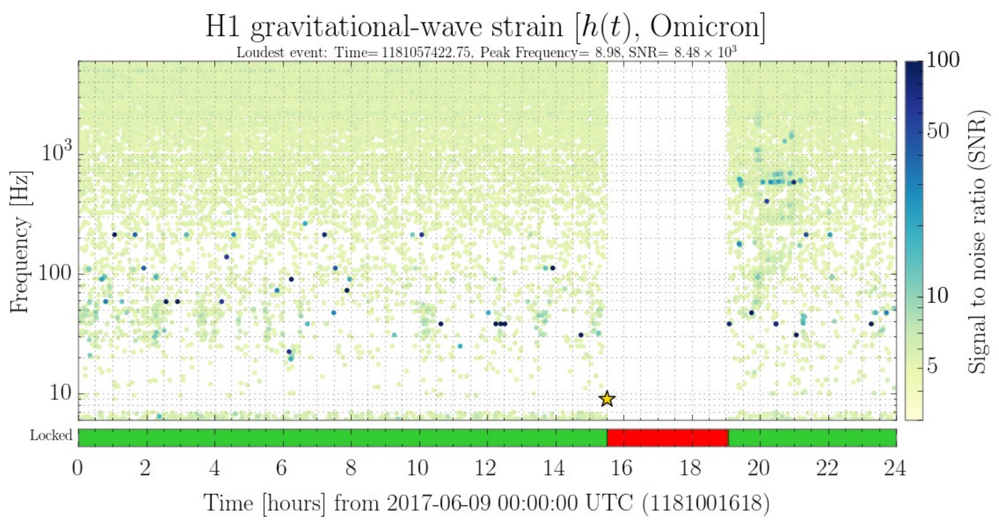


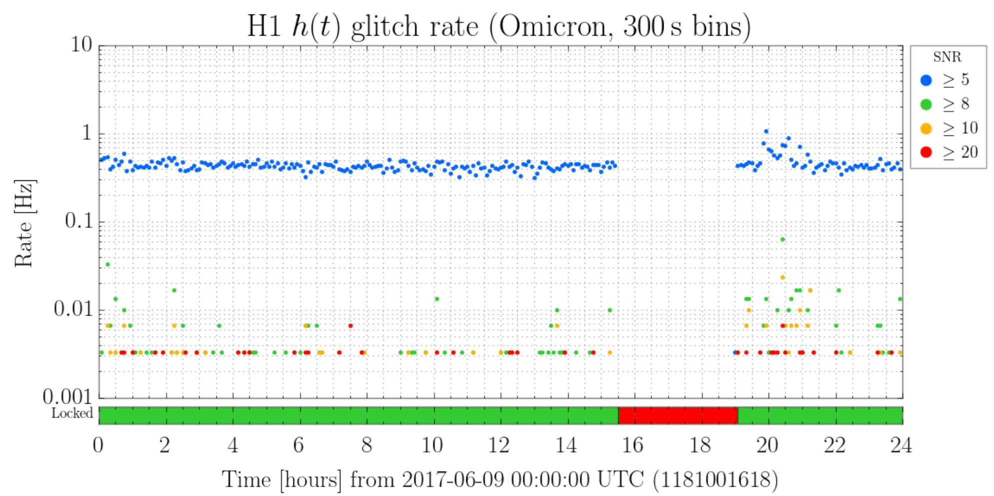


A typical day in LIGO

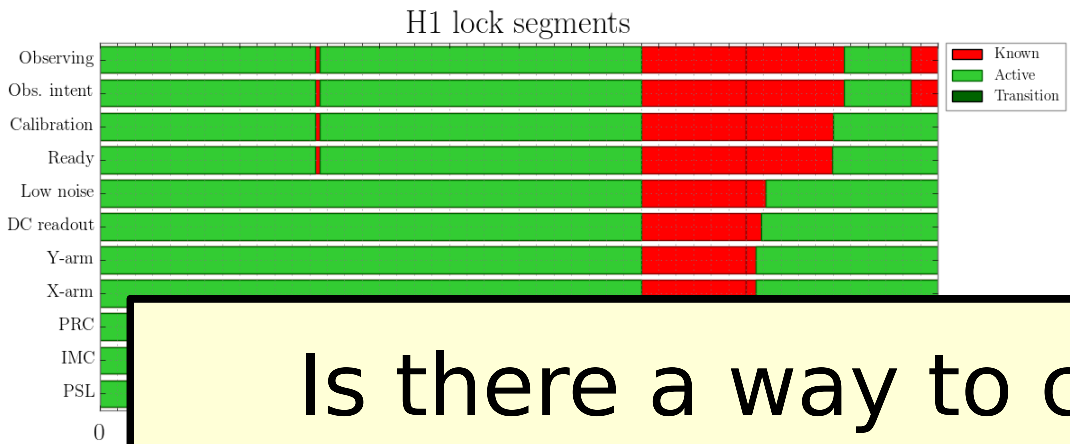


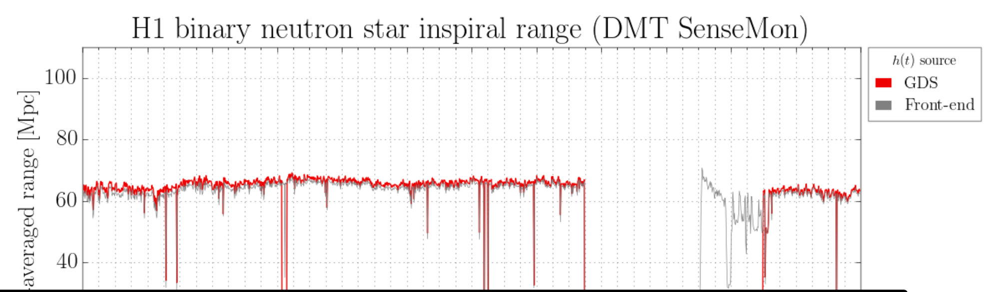




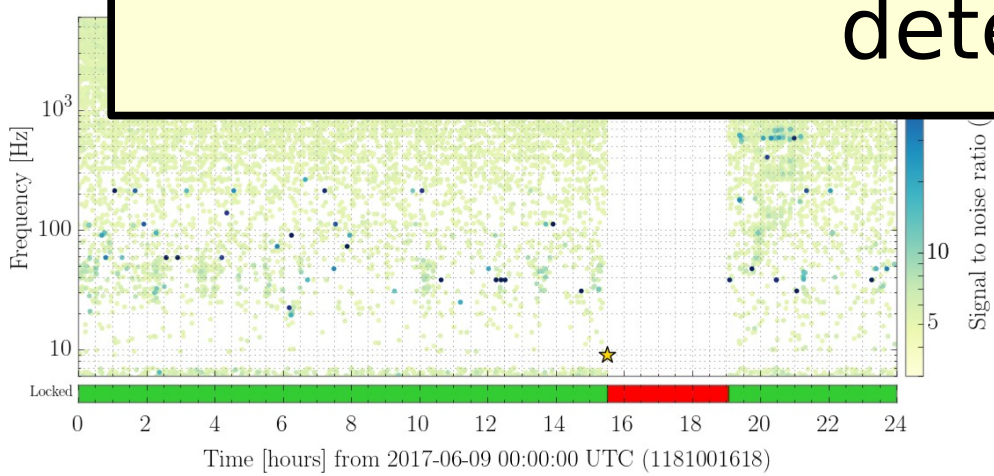


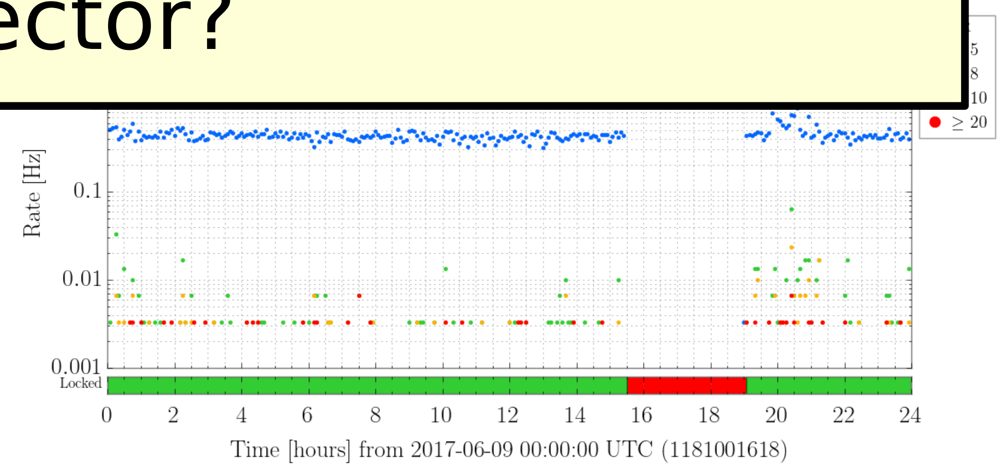
A typical day in LIGO





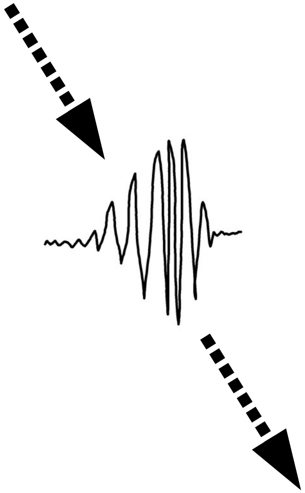
Is there a way to characterize noise variations as they happen in the detector?



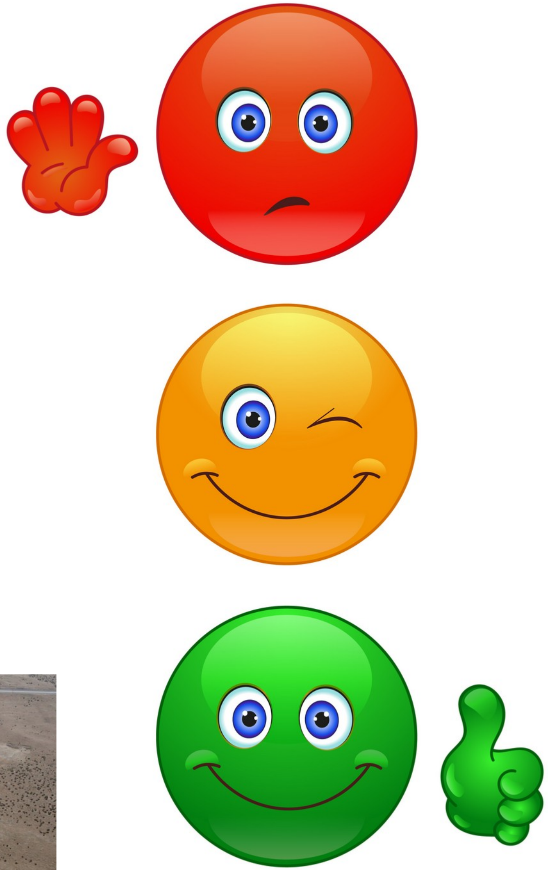


Noise monitor concept

Credit:SSU – Aurore Simonnet



Credit: LIGO Lab oratory



Bad data

Please double check...

Data OK!

Low-latency data quality assessment

Typically a lot of works goes in Data Quality assessment

EM advocate says event is okay.
Added by: Geoffrey Mo
Added: March 11, 2020, 12:15 p.m.

Superevent Information	
Superevent ID	S200311bg
Category	Production
Labels	DQOK EM_READY ADVOK EM_Selected EMBRIGHT_READY PASTRO_READY SKYMAP_READY GCN_PRELIM_SENT PE_READY
t_{start}	1267963150.37
t_0	1267963151.40
t_{end}	1267963152.44
Submitted	2020-03-11 11:59:09 UTC
Links	Data

Rapid Response Team

Analyst Comments | Data Quality | Sky Localization | External Coincidence | EM Followup

Analyst Comments

Log Comment

RRT minutes as PDF, updated to include chat text. (S200311bg RRT Minutes.pdf)

— Submitted by Jenne Driggers on March 11, 2020 17:04:24 UTC

Log Comment

RRT minutes: https://docs.google.com/document/d/1TFNX3n7ZfaXmZzTuundu5rYGhmw5pFRhDtM6jcaPuYo/edit

(S200311bg_RRT.mp3)

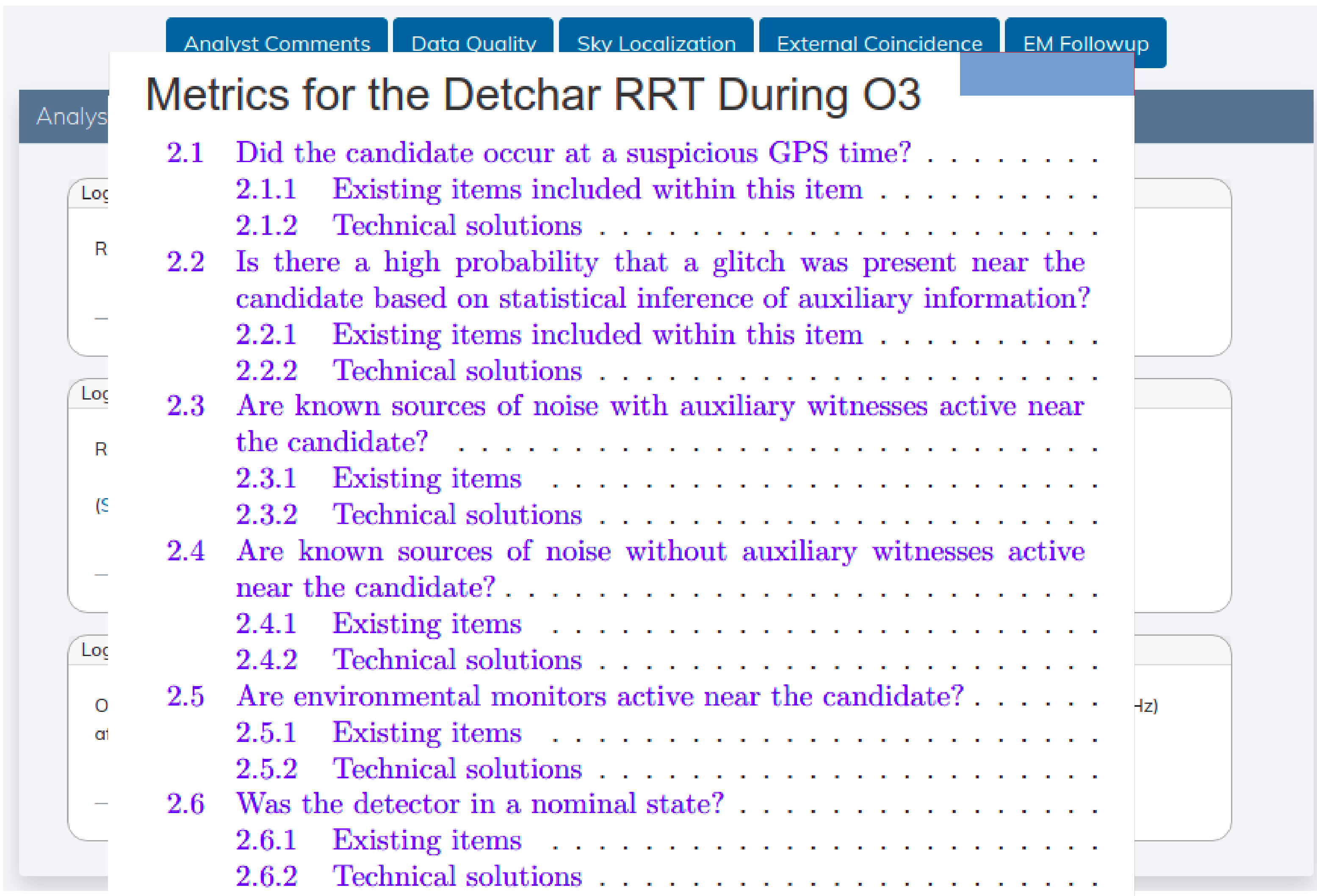
— Submitted by Geoffrey Mo on March 11, 2020 12:34:14 UTC

Log Comment

Online DQ report - H1 and L1 both look clean around the time of the event. There is some slight noise at 1.2s (below 100Hz) after the event at H1 in case anyone uses data after the event time.

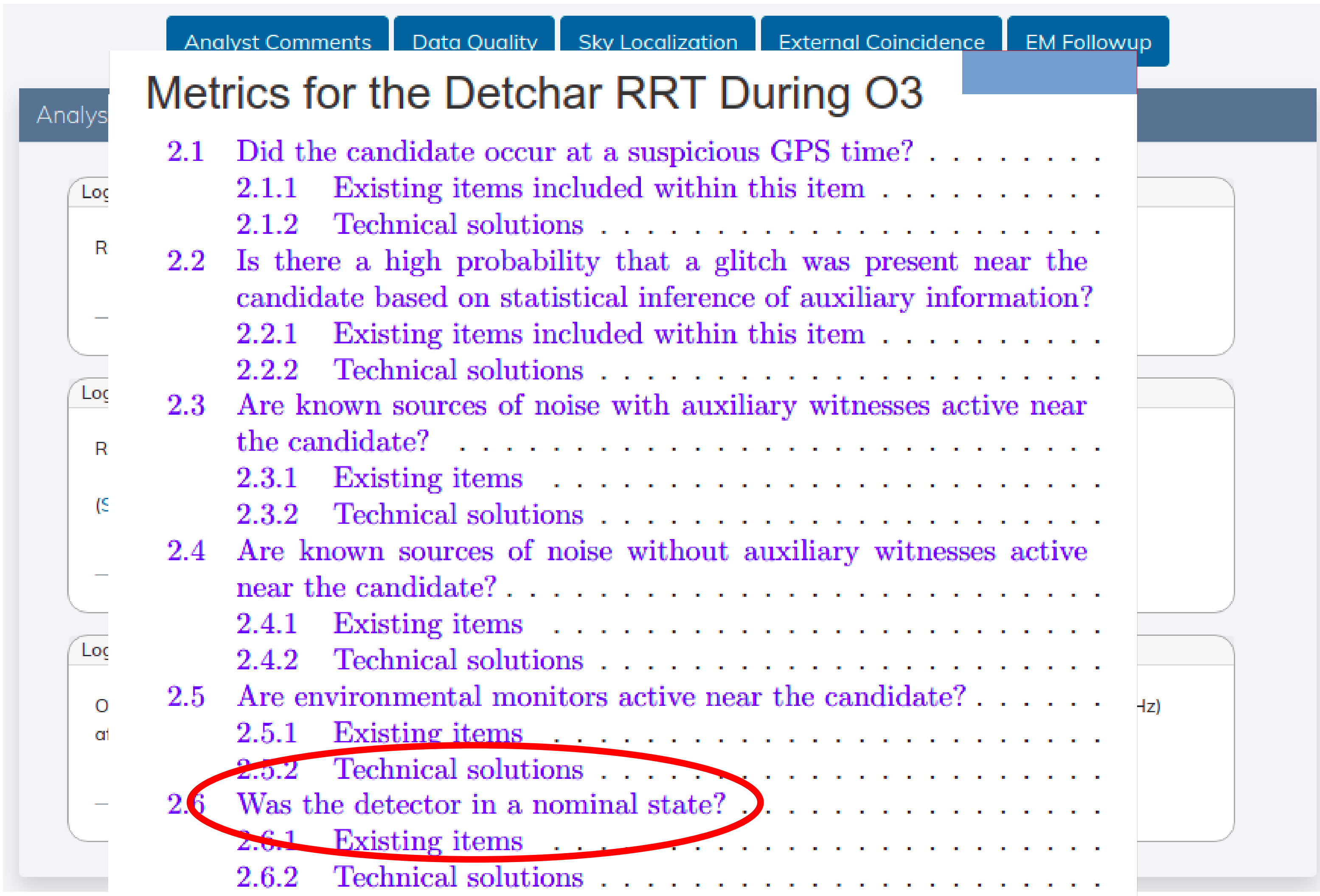
— Submitted by Laura Nuttall on March 11, 2020 12:16:24 UTC

Rapid Response Team



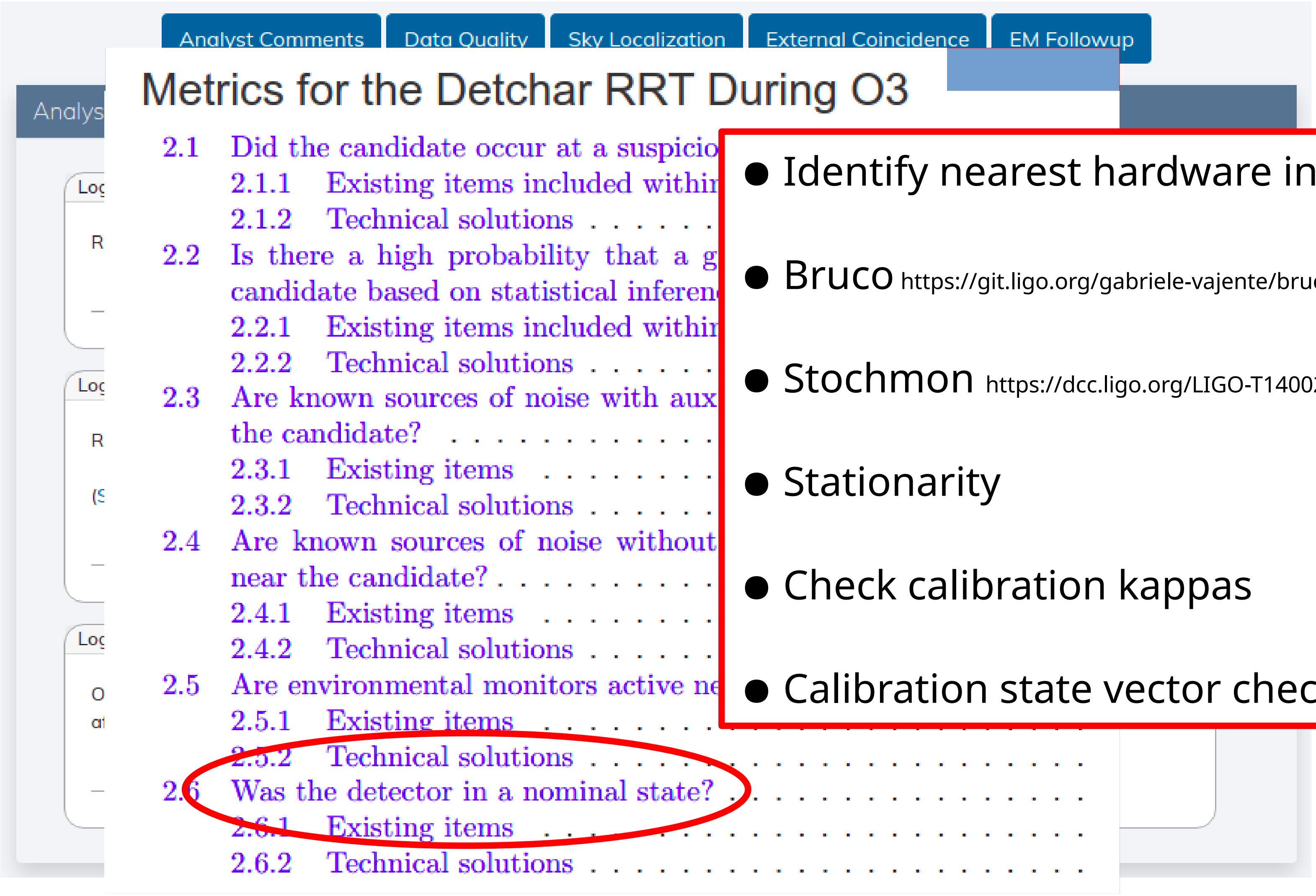
Metrics for the Detchar RRT During O3

Detector state



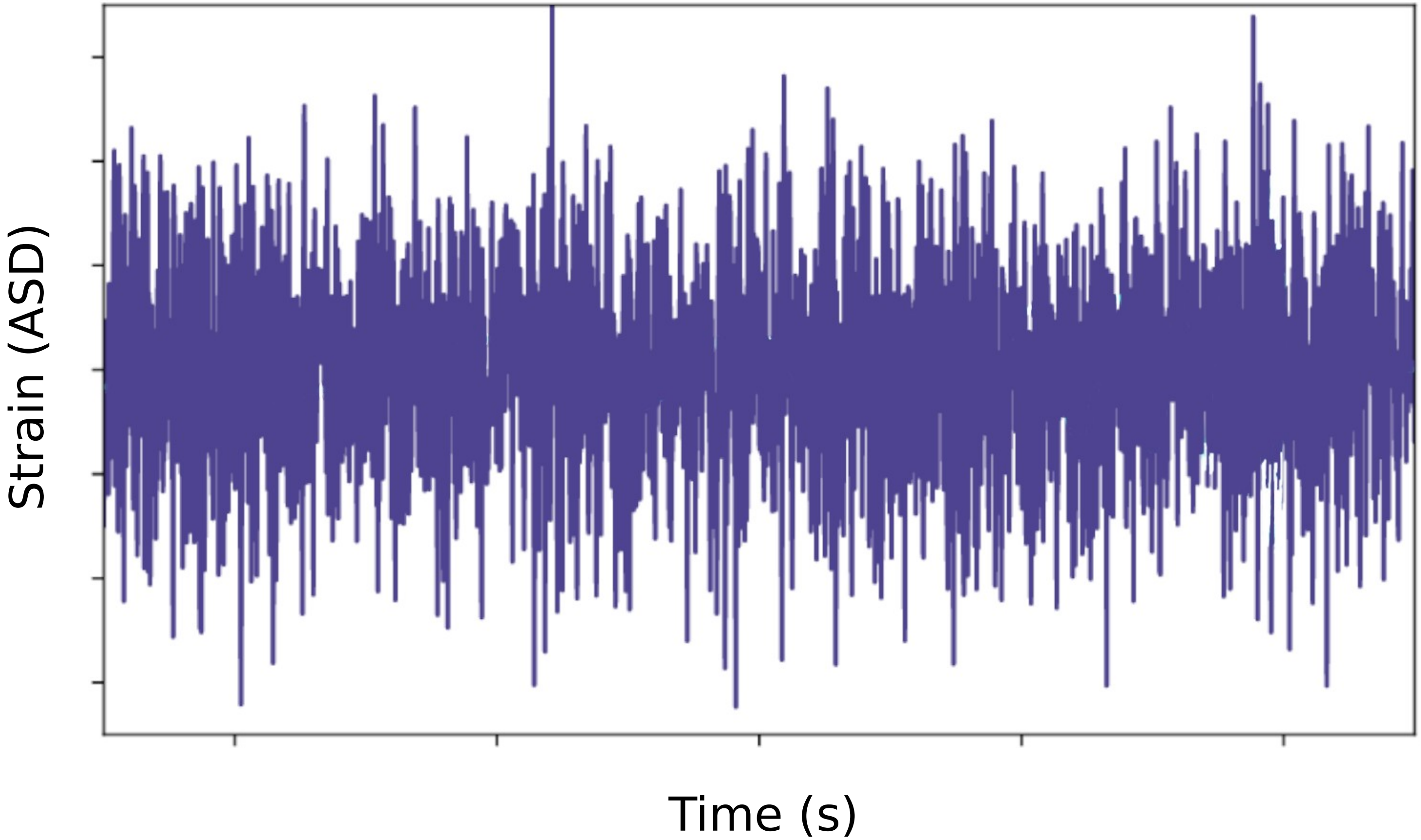
Metrics for the Detchar RRT During O3

Detector state



- Identify nearest hardware injection
- Bruco https://git.ligo.org/gabriele-vajente/bruco
- Stochmon https://dcc.ligo.org/LIGO-T1400205/public
- Stationarity
- Check calibration kappas
- Calibration state vector check

Building a nominal state metric from strain data



Noise and fractals

“In 1961 IBM was involved in transmitting computer data over phone lines, but a kind of white noise kept disturbing the flow of information—breaking the signal—and IBM looked to Mandelbrot to provide a new perspective on the problem.”

https://www.ibm.com/ibm/history/ibm100/us/en/icons/fractal/

Fractal analysis

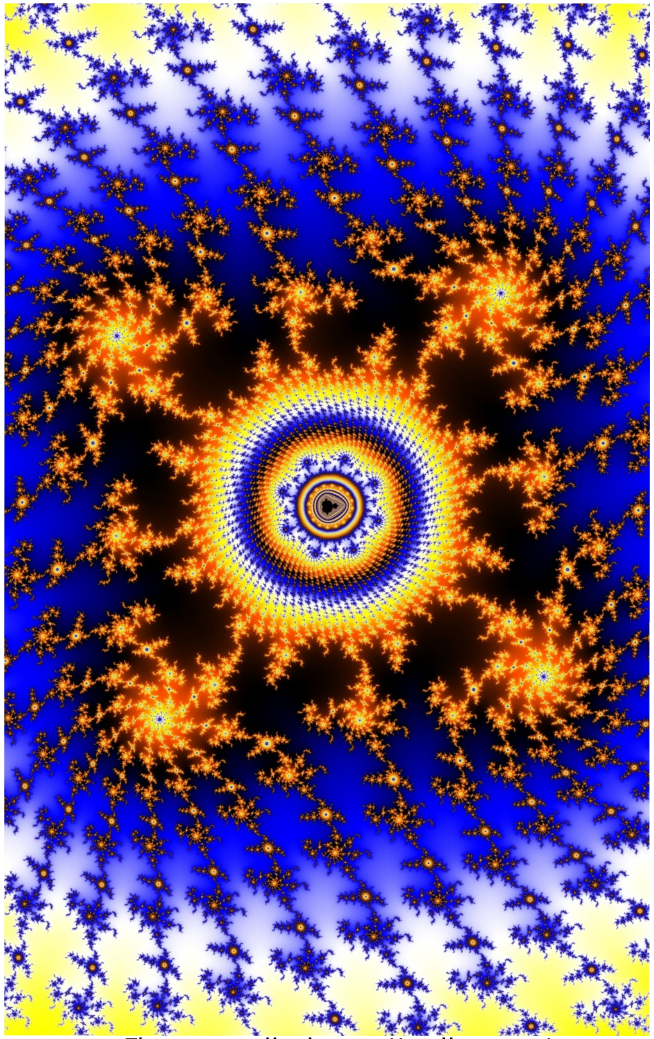


Figure credit: https://wallup.net/

- Fractal analysis can be used to characterize the degree of complexity of a set.
- The concept has been applied to different physical phenomena met in various fields from materials science to chemistry, biology, etc.

Topological dimension of a set

A topological space is an ordered pair (X, τ), where X is a set and τ is a collection of subsets of X, satisfying the following axioms:

- The empty set and X itself belong to τ.
- Any arbitrary (finite or infinite) union of members of τ belongs to τ.
- The intersection of any finite number of members of τ belongs to τ.

The elements of τ are called open sets and the collection τ is called a topology on X.

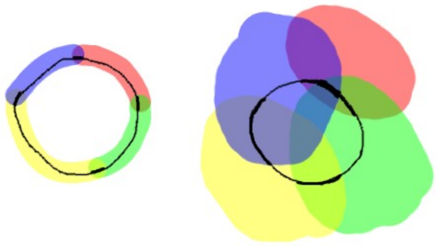
A topological space can be covered by open sets. The topological (Lebesgue covering) dimension is the smallest number n such that for every cover, there is a refinement in which every point in X lies in the intersection of no more than $n+1$ covering sets. The topological dimension is an integer number that does not change as the space is continuously deformed under an homeomorphism.

M. Yamaguti et al. “Mathematics of Fractal” Mathematical monographs, vol.167 AMS 1997s

Topological dimension of a set

Examples:

A circle requires a set of two (or more) open arcs to be covered. The topological dimension of the circle is one.



A disk requires at least three open sets to be covered. The topological dimension of the disk is two.

In general an *n*-dimensional Euclidean space has topological dimension n.

Fractal dimension

A fractal is a subset of an Euclidean space with a fractal dimension that strictly exceeds its topological dimension.

The fractal dimension *D* is defined with the number of covering areas necessary to cover the fractal structure *F*.

Box-counting

$$D = \frac{\log\left[N(\epsilon)\right]}{\log\left[1/\epsilon\right]}$$

ϵ = radius of the covering balls

Hausdorff dimension

$$D = \sup\left\{d \in \mathbb{R}^+ : \; m_d(F) = \infty\right\}$$

m_d = *s*-dimensional Hausdorff-outer measure

D is a measure of roughness.

For an *n*-dimensional Euclidean space, the Hausdorff dimension coincides with the topological dimension: Single point = 0, line segment = 1, square = 2, etc.

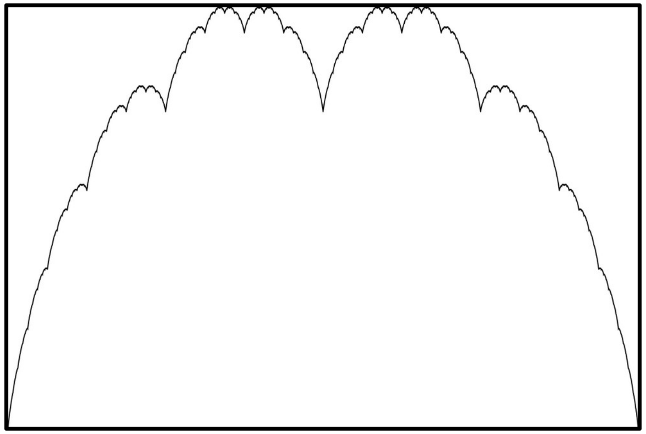
M. Yamaguti et al. “Mathematics of Fractal” Mathematical monographs, vol.167 AMS 1997s

Examples

Takagi–Landsberg curve

$$T(x) = \sum_{n=0}^{\infty} w^n s(2^n x) \qquad s(x) = \min_{n \in \mathbb{Z}} |x - n|$$

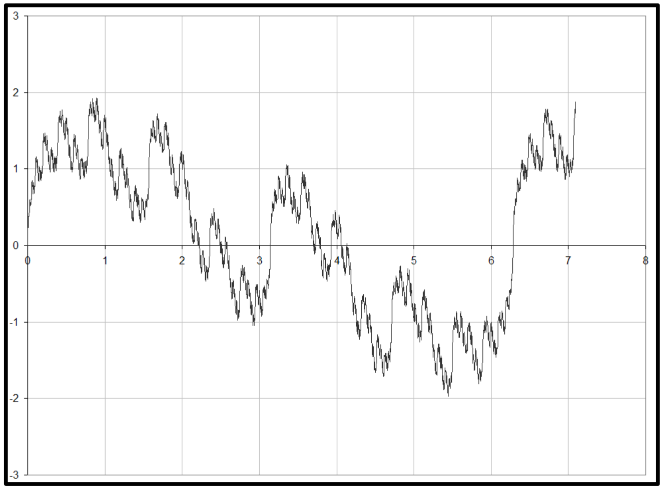
$$D = 2 + \log_2(w)$$



Weierstrass function

$$f(x) = \sum_{k=1}^{\infty} a^{-k} \sin(b^k x)\,,\ \ 1 < a < 2\,,\ b > 1$$

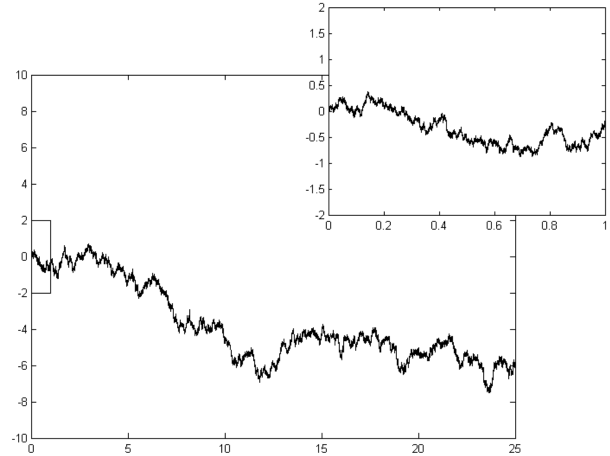
$$D = 2 - \log_b(a)$$



Examples

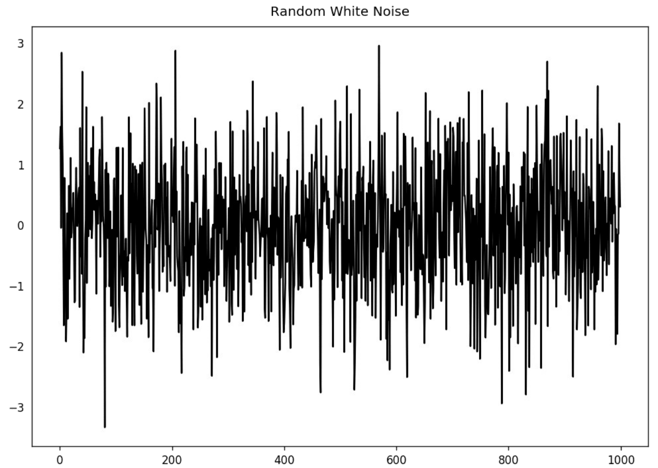
Brownian motion (Wiener process)

$$D = 3/2$$



White noise

$$D = 2$$



Calculation of fractal dimension

ANAM method

$$M_\tau^\alpha(f,x) = \left[\frac{1}{\tau^2} \int_{t_1=0}^{\tau} \int_{t_2=0}^{\tau} |f(x+t_1) - f(x-t_2)|^\alpha \, \mathrm{d}t_1 \, \mathrm{d}t_2 \right]^{1/\alpha}$$

$$K_\tau^\alpha(f,a,b) = \frac{1}{b-a} \int_{x=a}^{x=b} \left[\frac{1}{\tau^2} \int_{t_1=0}^{\tau} \int_{t_2=0}^{\tau} |f(x+t_1) - f(x-t_2)|^\alpha \, \mathrm{d}t_1 \, \mathrm{d}t_2 \right]^{1/\alpha} \mathrm{d}x$$

$$D = \lim_{\tau \to 0} \left(2 - \frac{\log K_\tau^\alpha(f,a,b)}{\log \tau} \right)$$

Bigerelle M, Iost A., C.R. Acad. Sci. Paris, t. 323, Serie II b, 1996:669±74

Calculation of fractal dimension

Variation (VAR) method

$$f : [a, b] \to \mathbb{R},$$

$$\mathrm{OSC}_\tau(f, x) = \left| \max_{|x-t|<\tau} (f(t)) - \min_{|x-t|<\tau} (f(t)) \right|$$

$$\mathrm{VAR}_\tau(f, a, b) = \frac{1}{b-a} \int_a^b \mathrm{OSC}_\tau(f, x)\,\mathrm{d}x$$

$$D = \lim_{\tau \to 0} \left(2 - \frac{\log \mathbf{VAR}_\tau(f, a, b)}{\log \tau} \right)$$

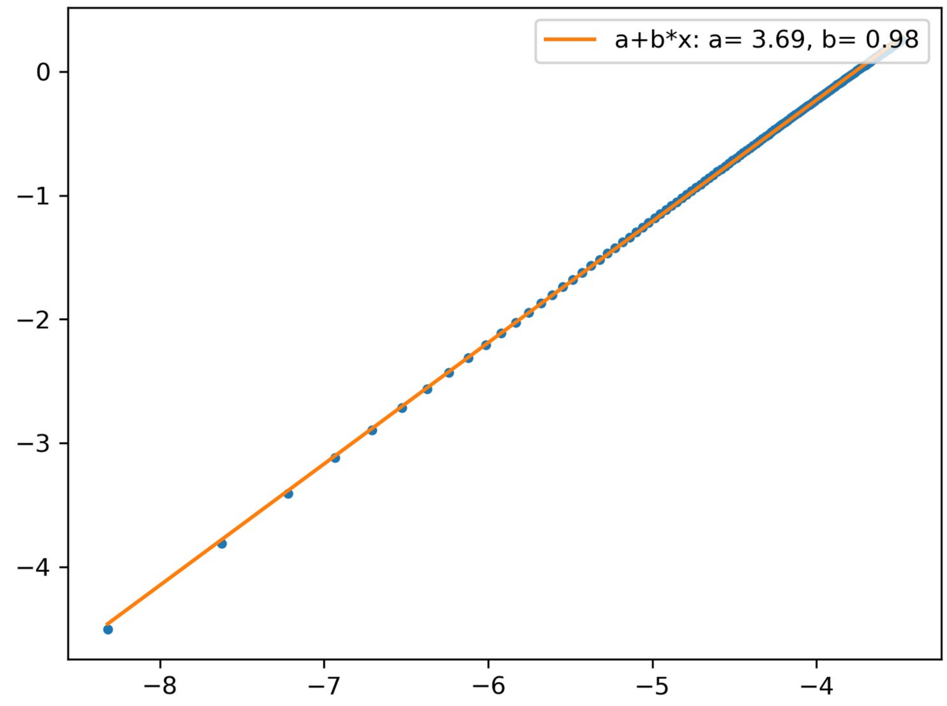
Dubuc B, Dubuc S., SIAM J Numer Anal 1996;33:602±26.

Algorithm

- Discretize either the $\mathbf{VAR}_\tau(f,a,b)$ or the $K_\tau^\alpha(f,a,b)$ estimator

- The fractal dimension of the time series is the slope of 2 – log(estimator) / log(τ)

- Theoretically this a straight line. In practice, do a fit of the log(estimator) as function of log(τ) .



Simple tests

Data length: 1 s, sampling rate: 4096 Hz. Decimate factor: 1/128 (VAR/ANAM).

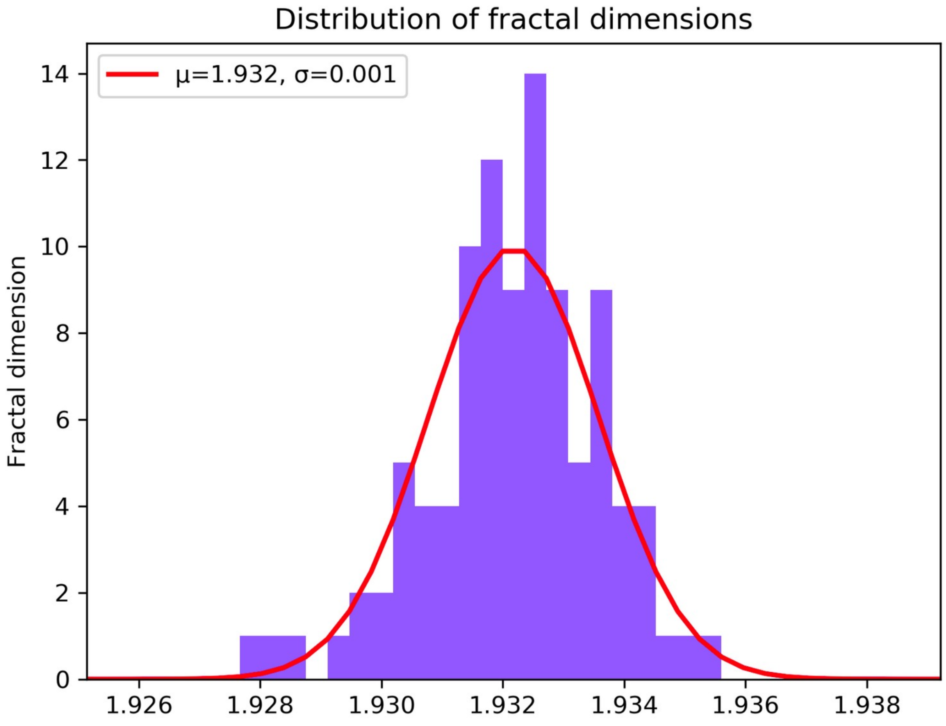
	Theory	VAR	ANAM
Simple analytic function f(x)	1.000	1.020 (+2%)	1.002 (+0.2%)
Takagi-Landberg curve (w=0.7)	1.485	1.448 (-2.5%)	1.503 (+1.2%)
Takagi-Landberg curve (w=0.8)	1.678	1.748 (+4.2%)	1.668 (-0.6%)
Weierstrass curve (a=0.7, b=9)	1.838	1.747 (-5.0%)	1.804 (-1.9%)
Weierstrass curve (a=0.5, b=12)	1.721	1.638 (-4.8%)	1.662 (-3.4%)

VAR	ANAM
Faster	Slower (x100 VAR)
Less accurate	More accurate

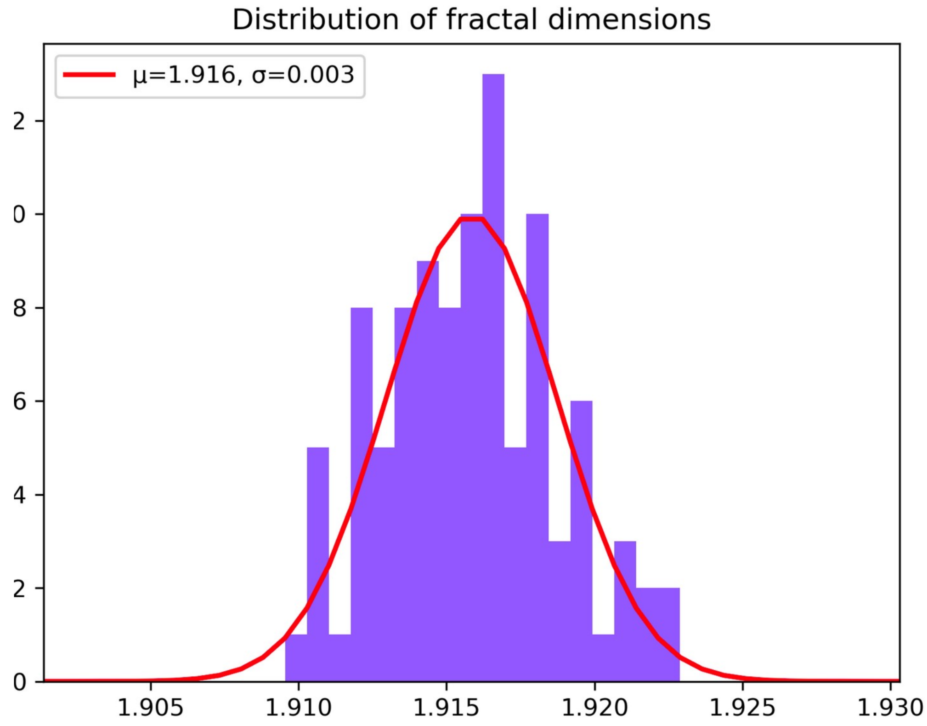
Tests on fake noise: White noise

Data length: 1 s, sampling rate: 4096 Hz. Decimate factor: 16/128 (VAR/ANAM)
Runs: 100. Theoretical fractal dimension: 2.000

VAR



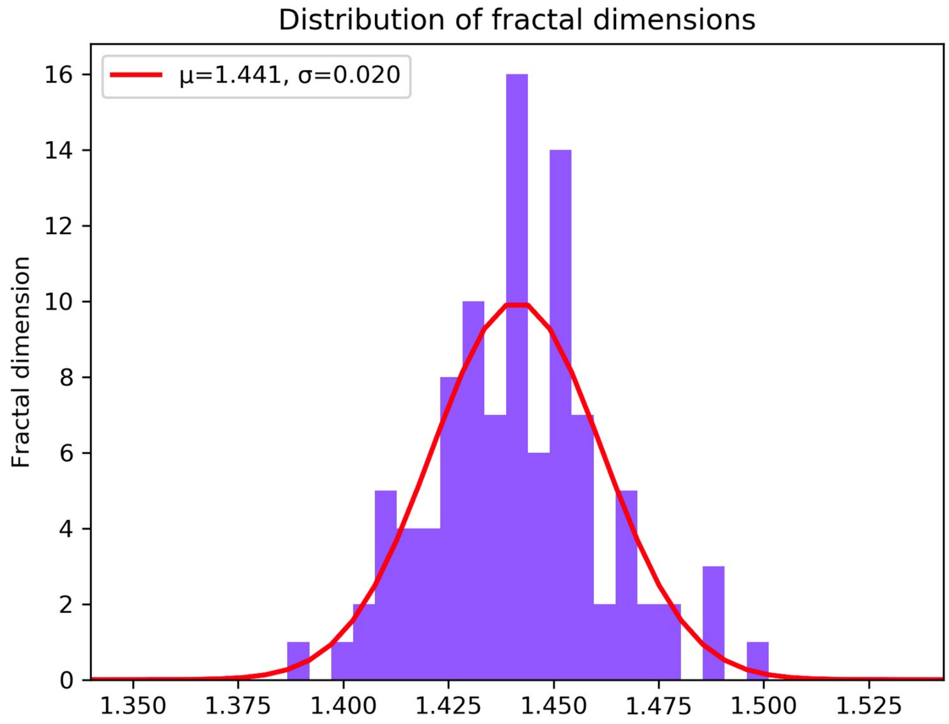
ANAM



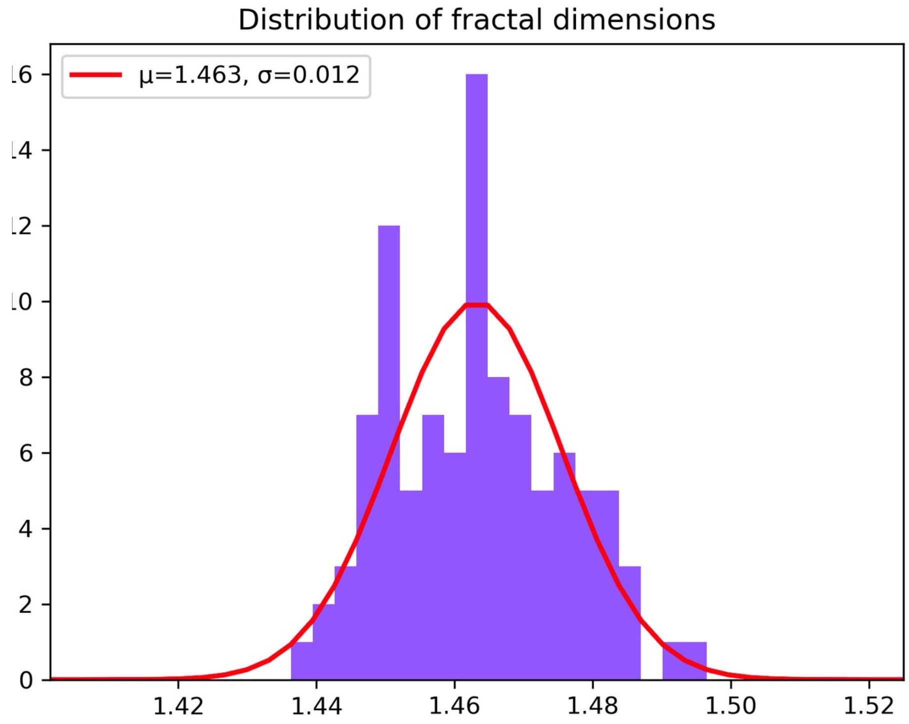
Tests on fake noise: Brownian noise

Data length: 1 s, sampling rate: 4096 Hz. Decimate factor: 16/128 (VAR/ANAM)
Brownian motion speed: 2. Runs: 100. Theoretical fractal dimension: 1.500

VAR

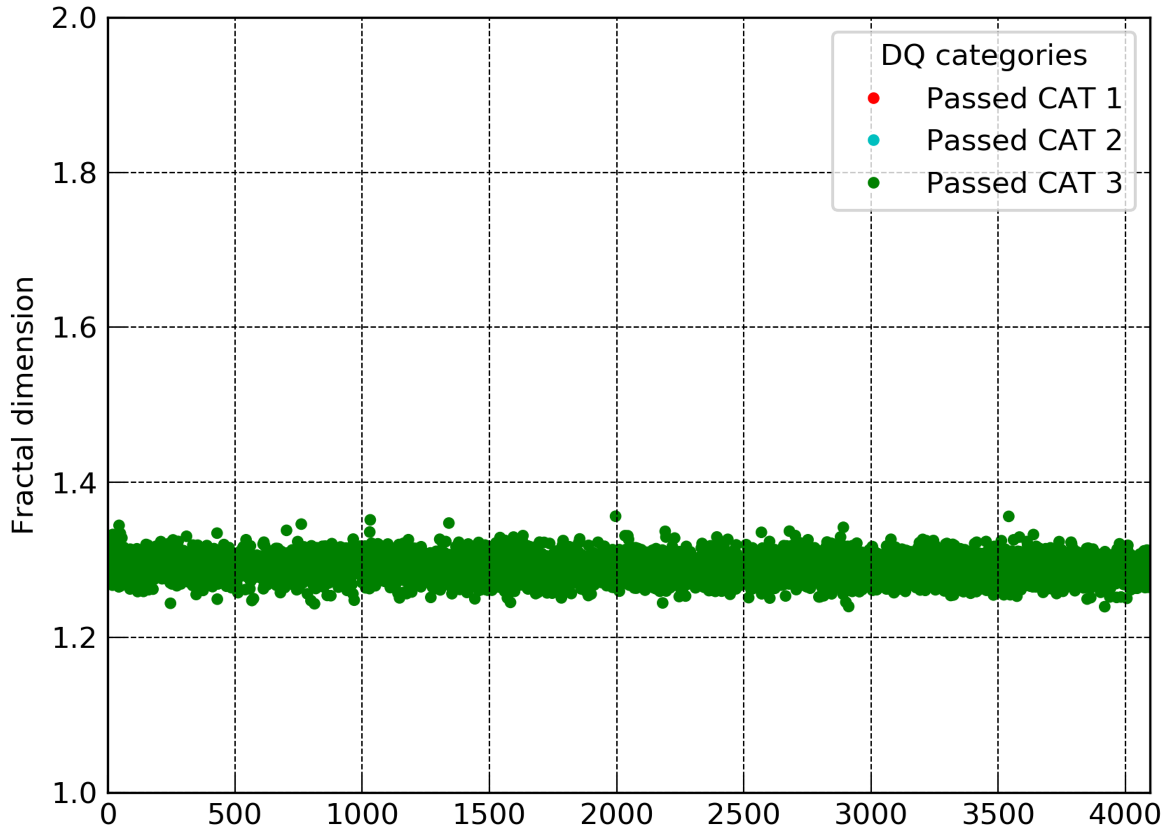


ANAM



Tests on O2 LIGO (open) data

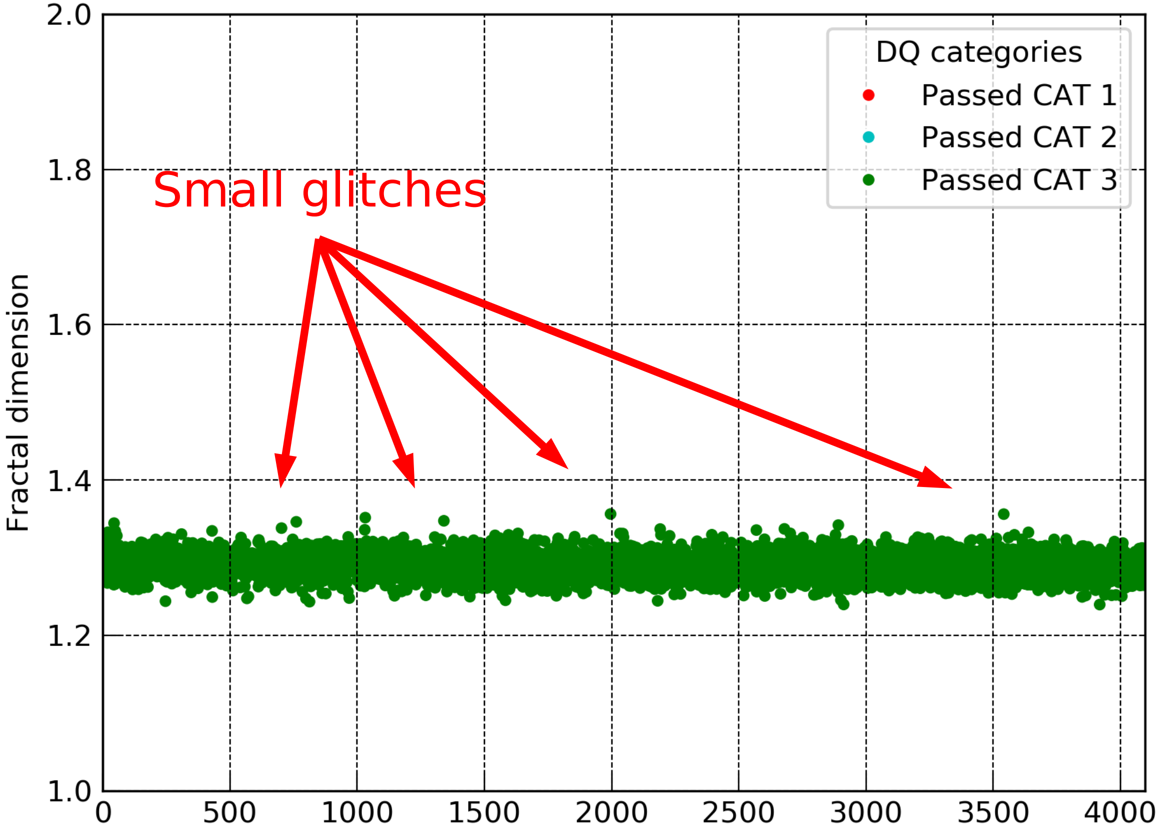
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Quiet time in H1

Tests on O2 LIGO (open) data

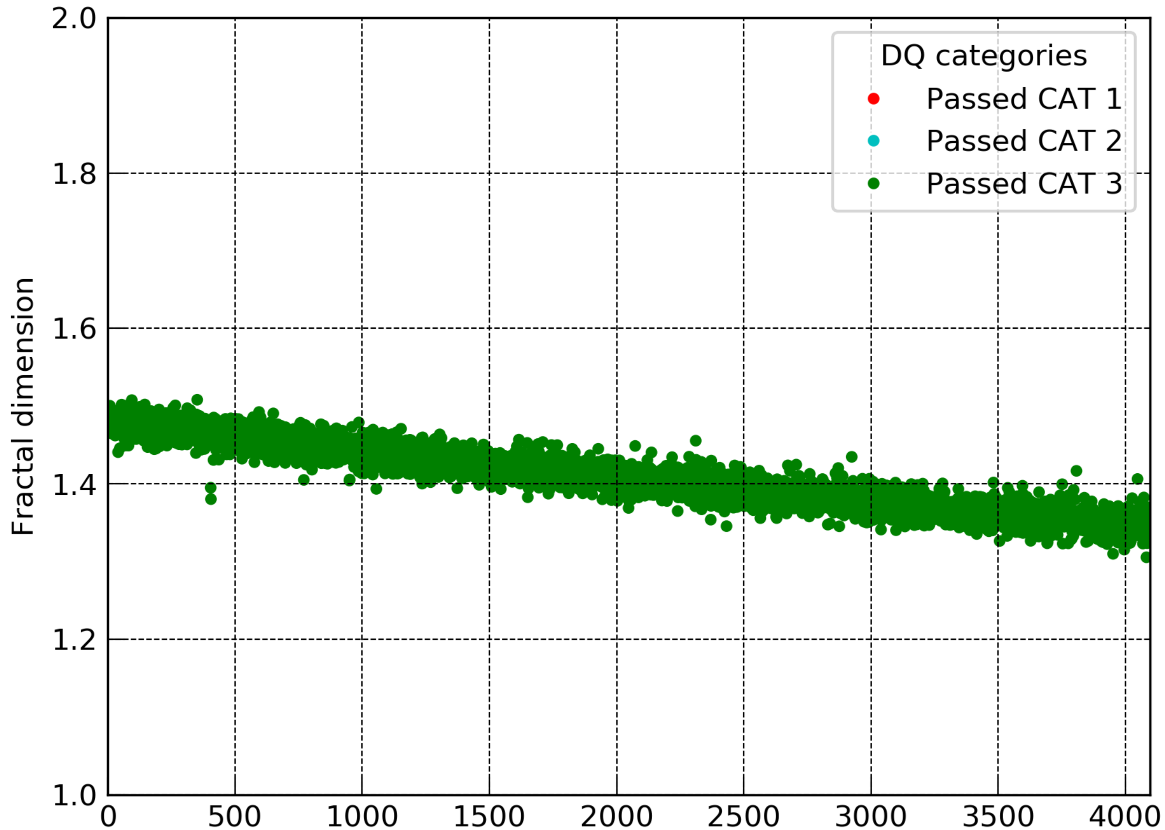
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Quiet time in H1

Tests on O2 LIGO (open) data

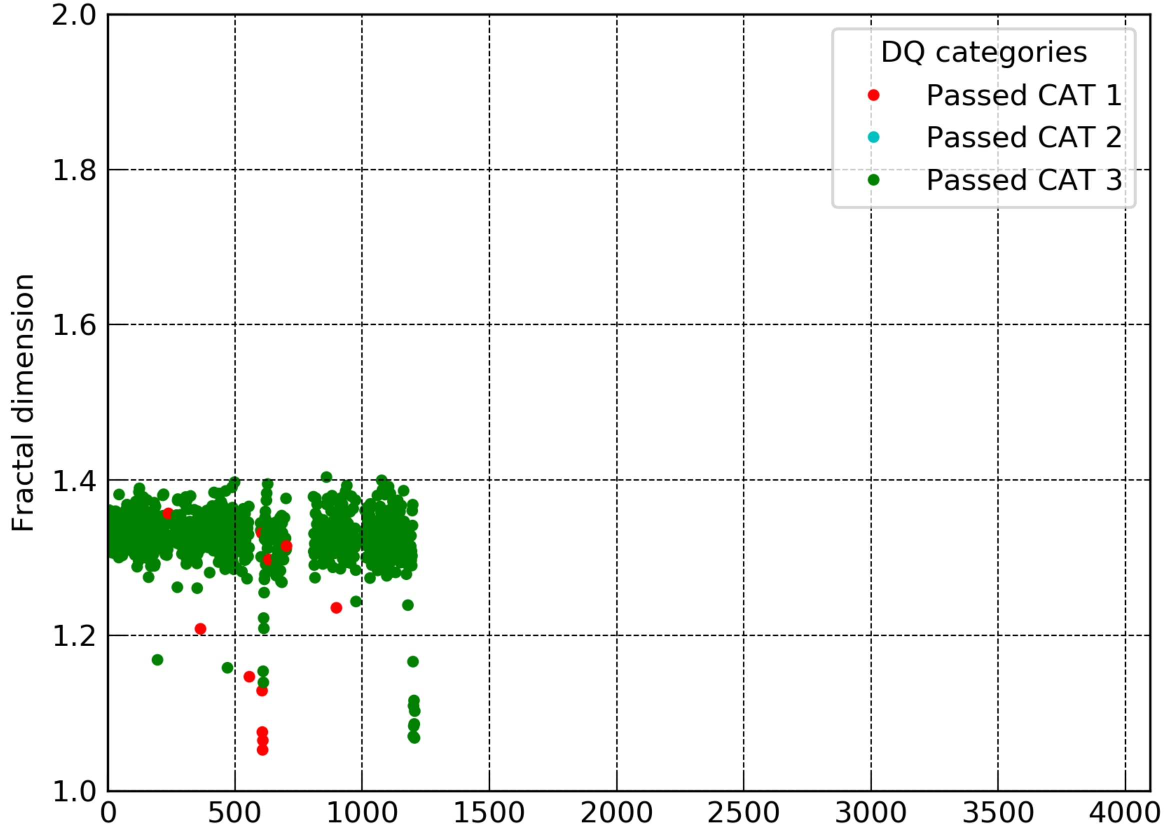
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Changing noise on long time-scale

Tests on O2 LIGO (open) data

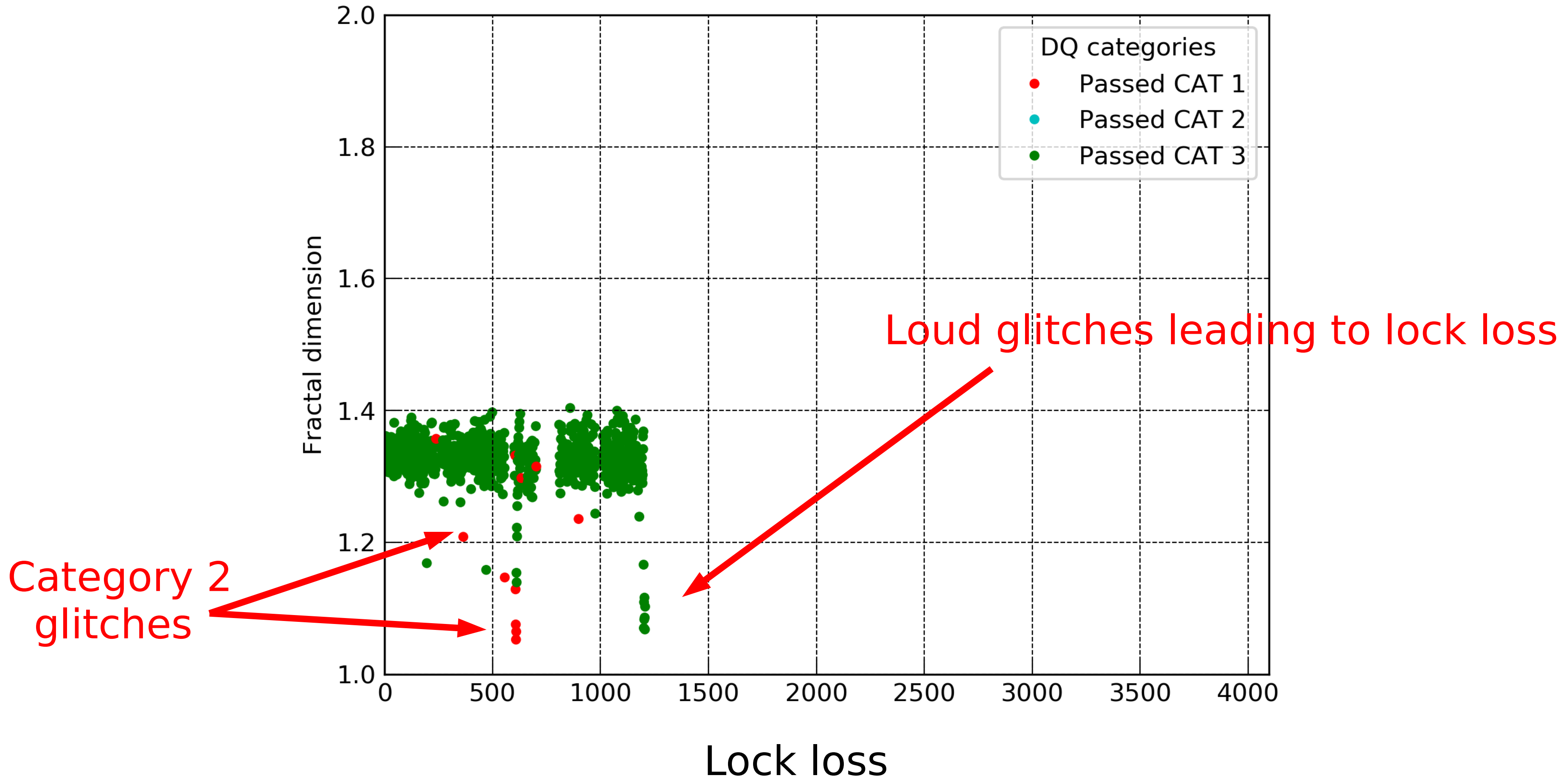
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Lock loss

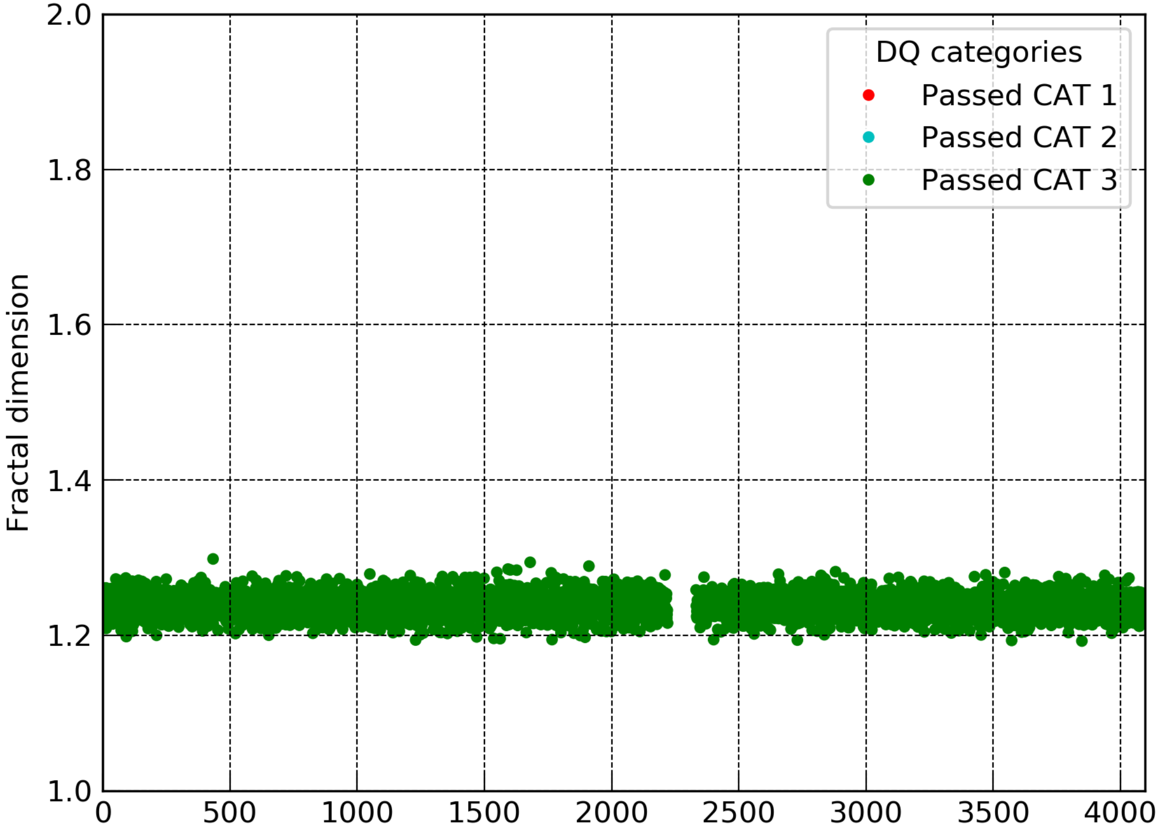
Tests on O2 LIGO (open) data

LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Tests on O2 LIGO (open) data

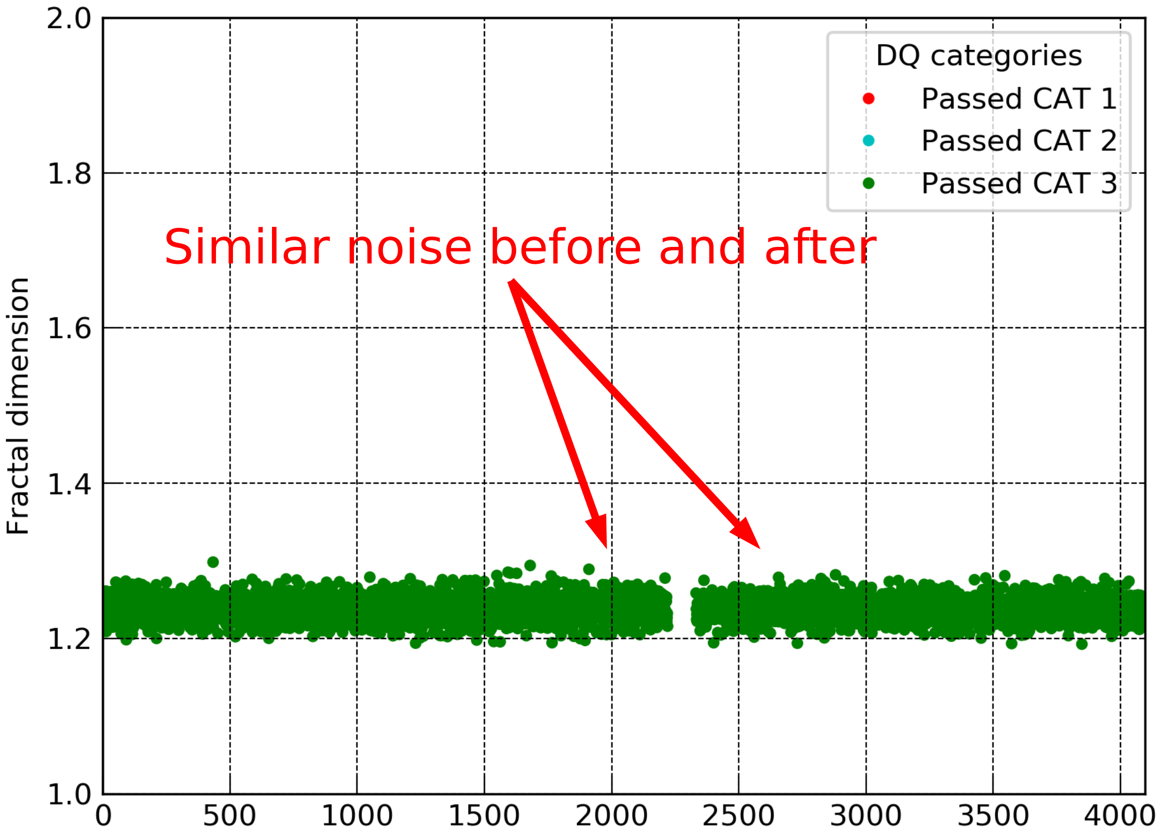
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Category 1 data (interferometer in low noise, but locked)

Tests on O2 LIGO (open) data

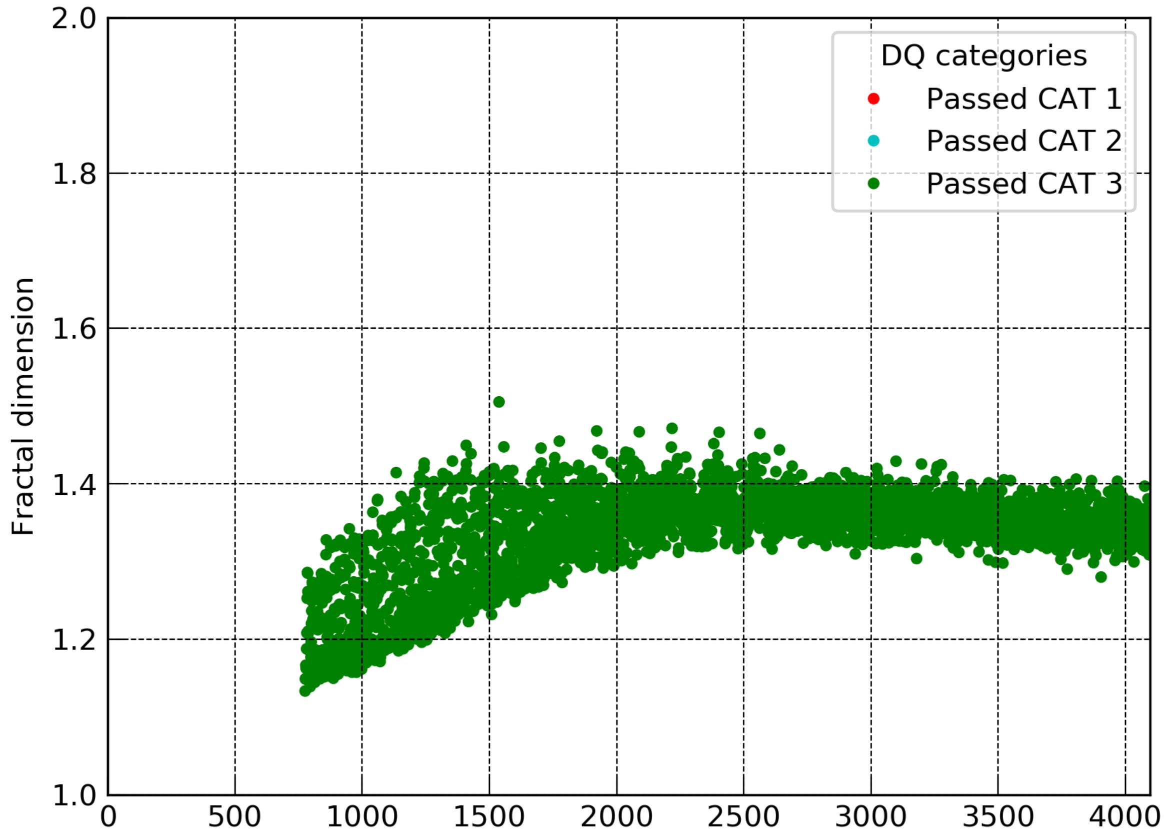
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



Category 1 data (interferometer in low noise, but locked)

Tests on O2 LIGO (open) data

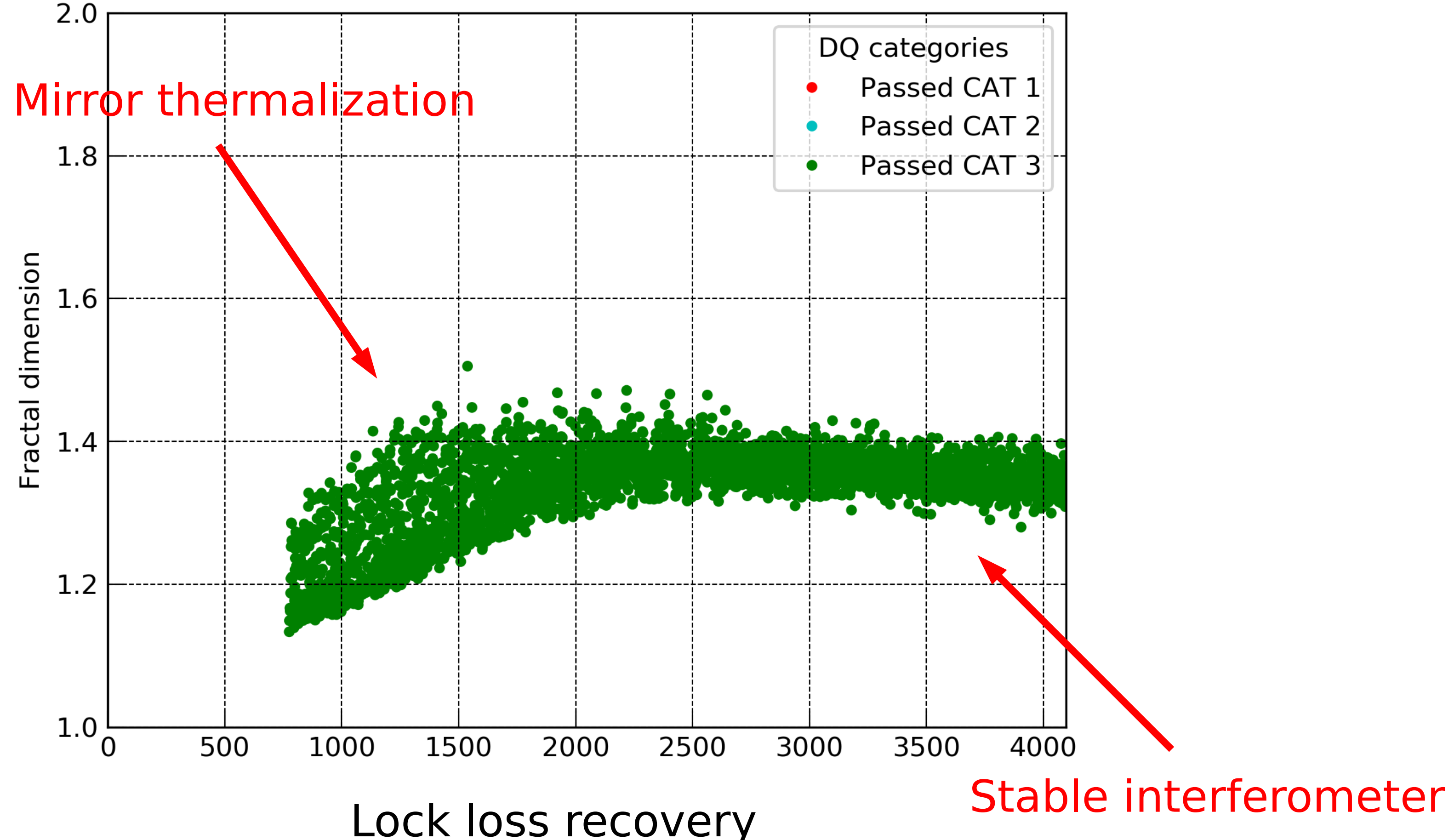
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



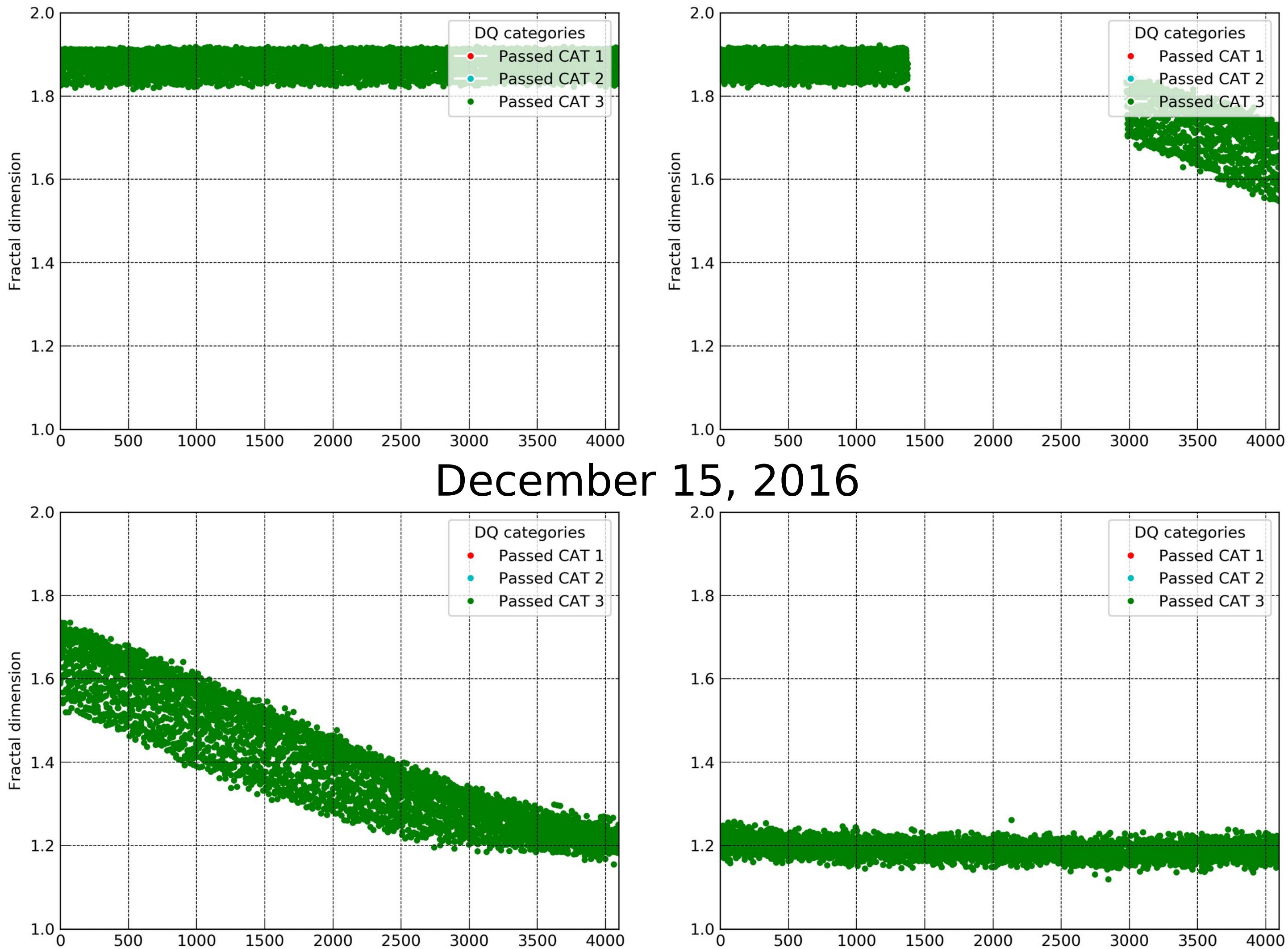
Lock loss recovery

Tests on O2 LIGO (open) data

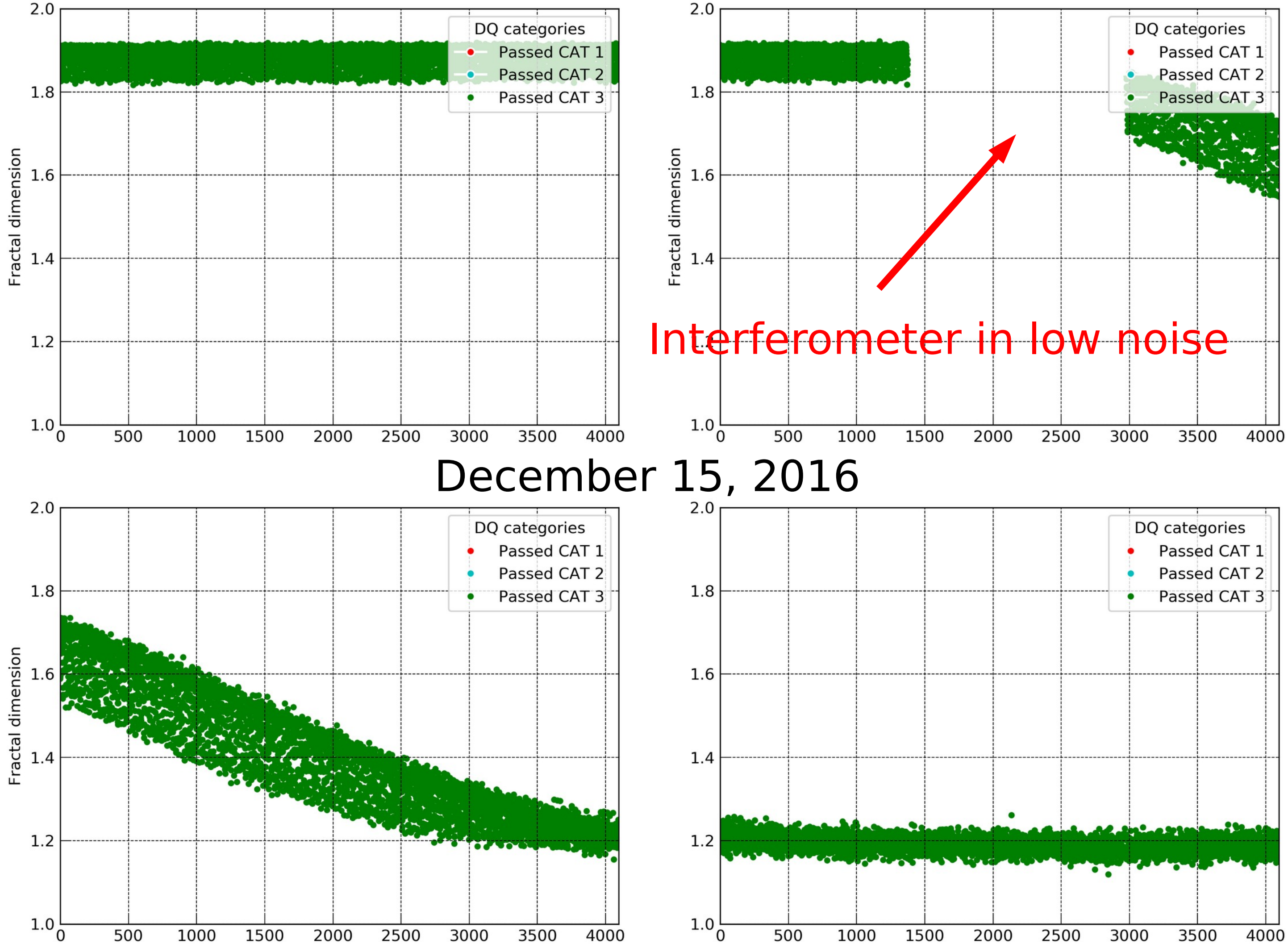
LIGO Hanford O2 data. Sampling rate: 4096 Hz. Fractal dimension over 1 sec.
Method: VAR. Decimate factor: 16



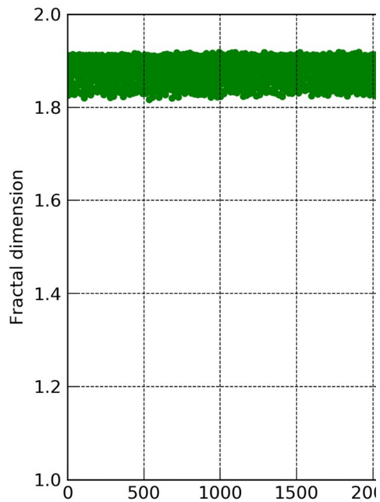
Tests on O2 LIGO (open) data

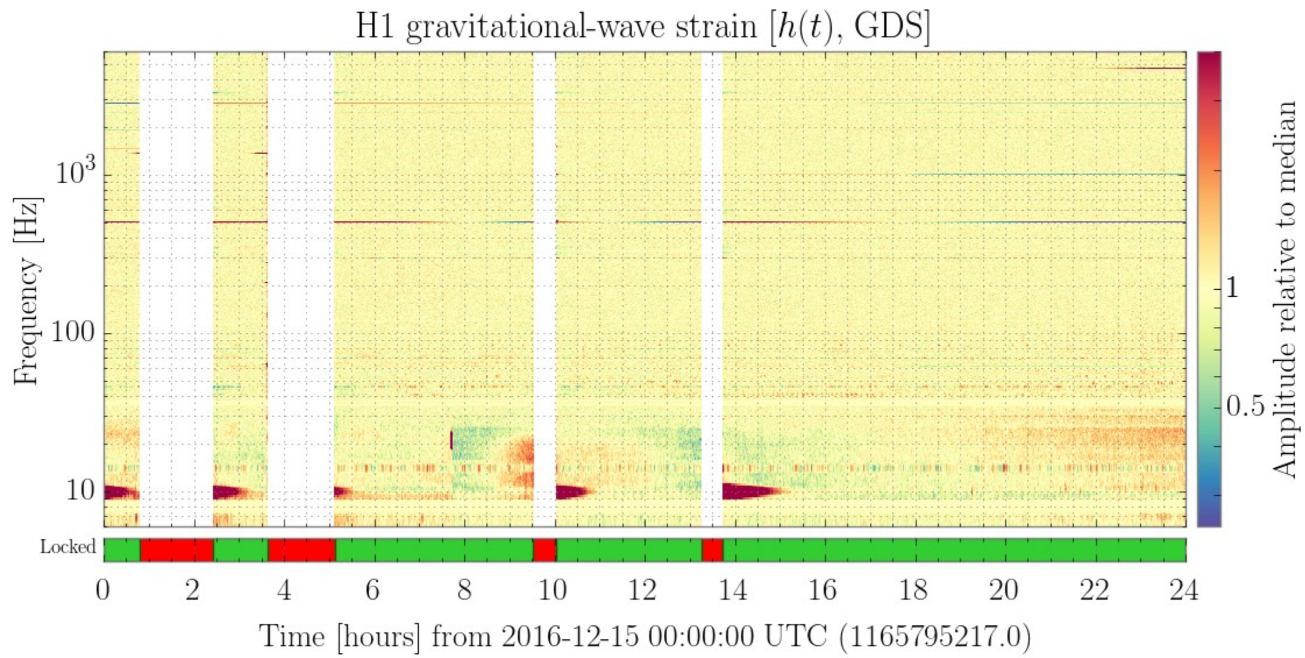


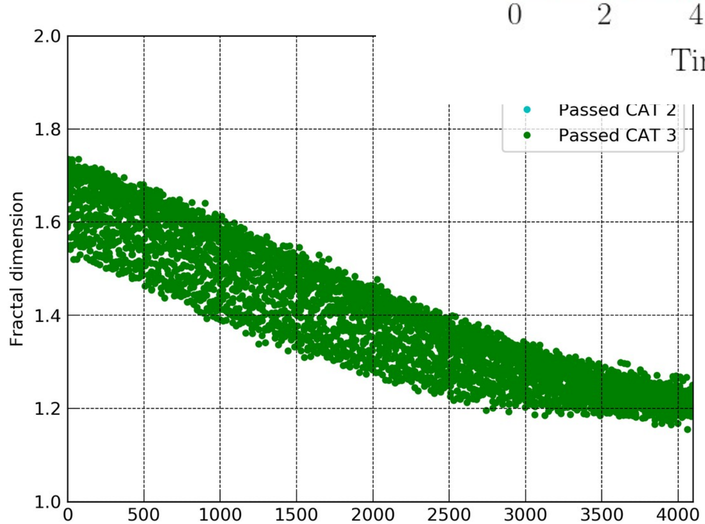
Tests on O2 LIGO (open) data

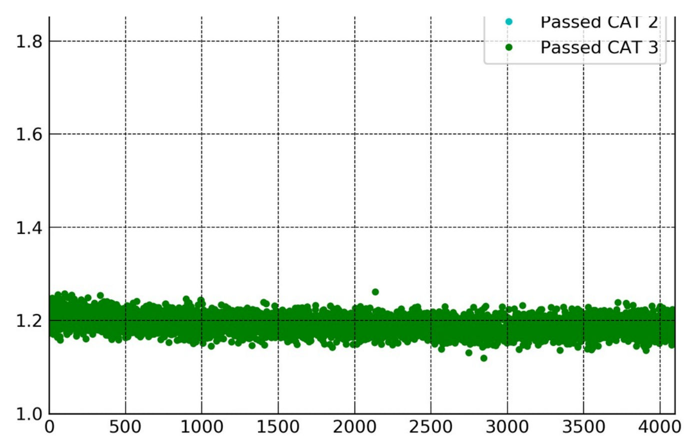


Tests on O2 LIGO (open) data

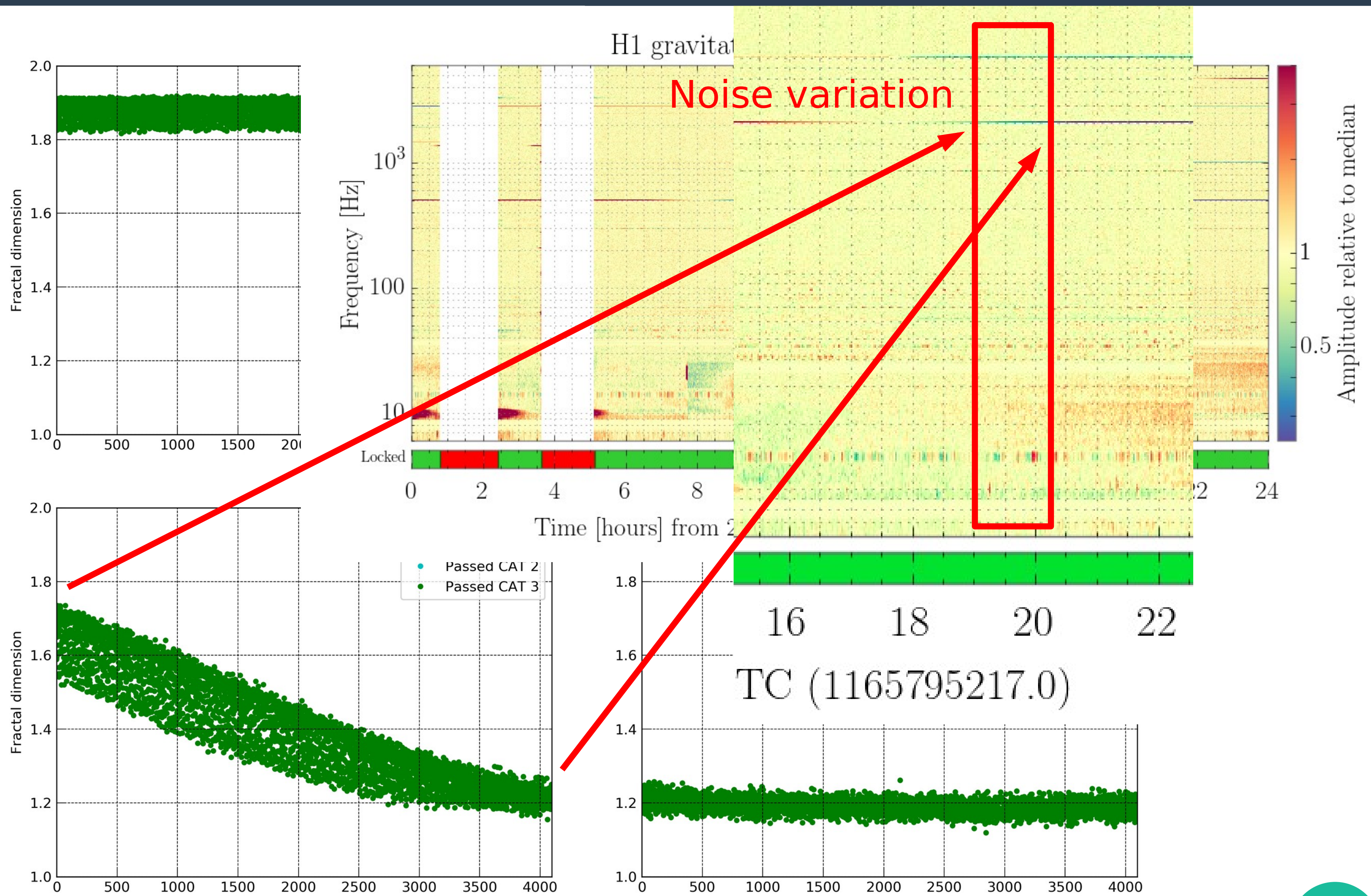






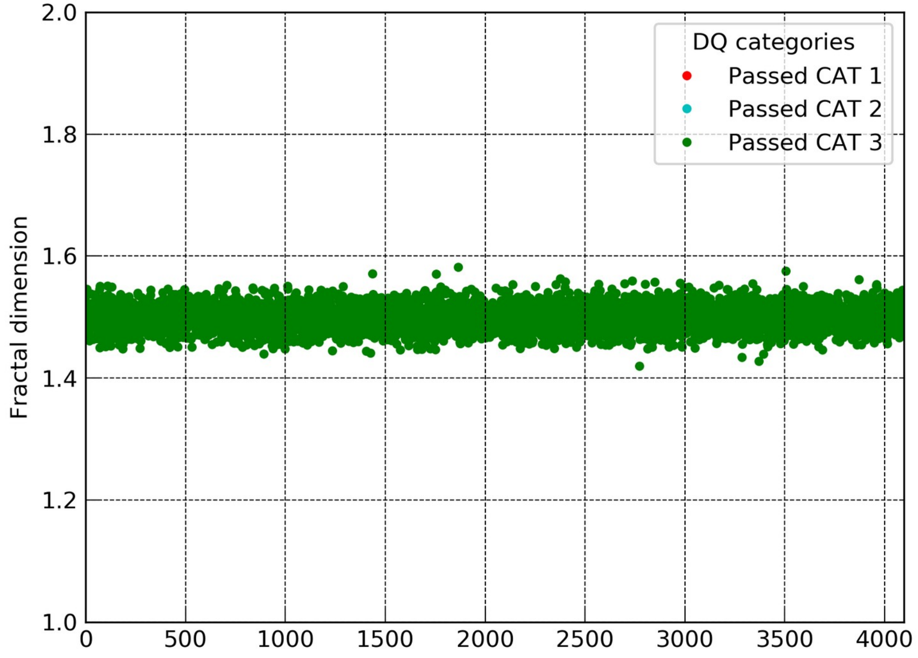


Tests on O2 LIGO (open) data

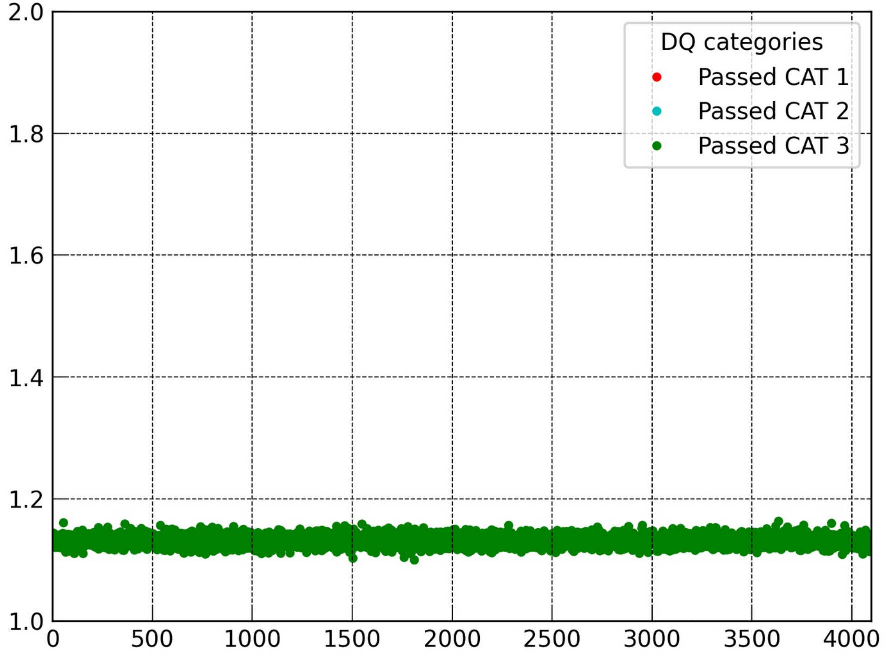


Tests on O2 LIGO (open) data

Background noise is not the same across the run!

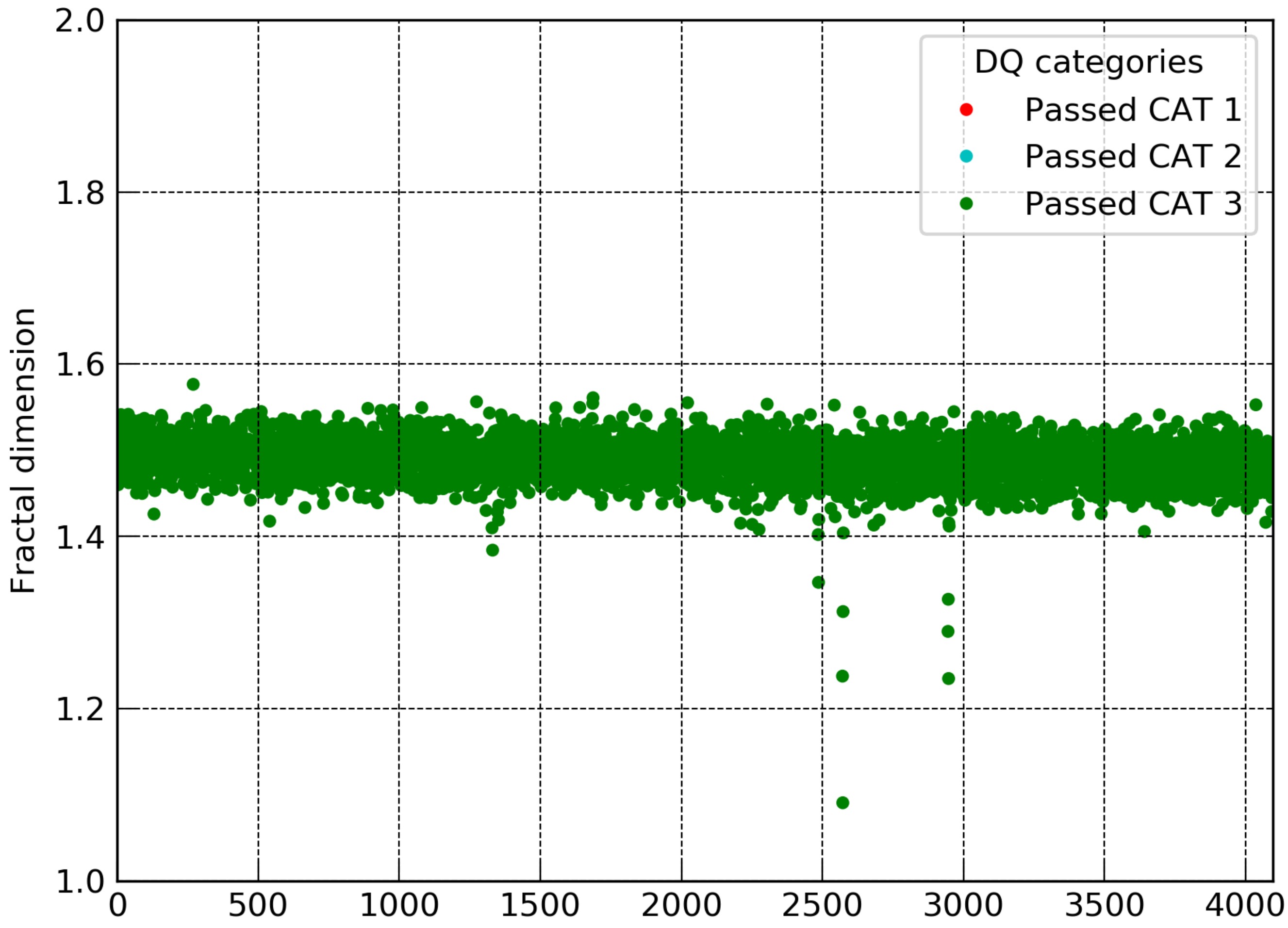


December 8, 2016



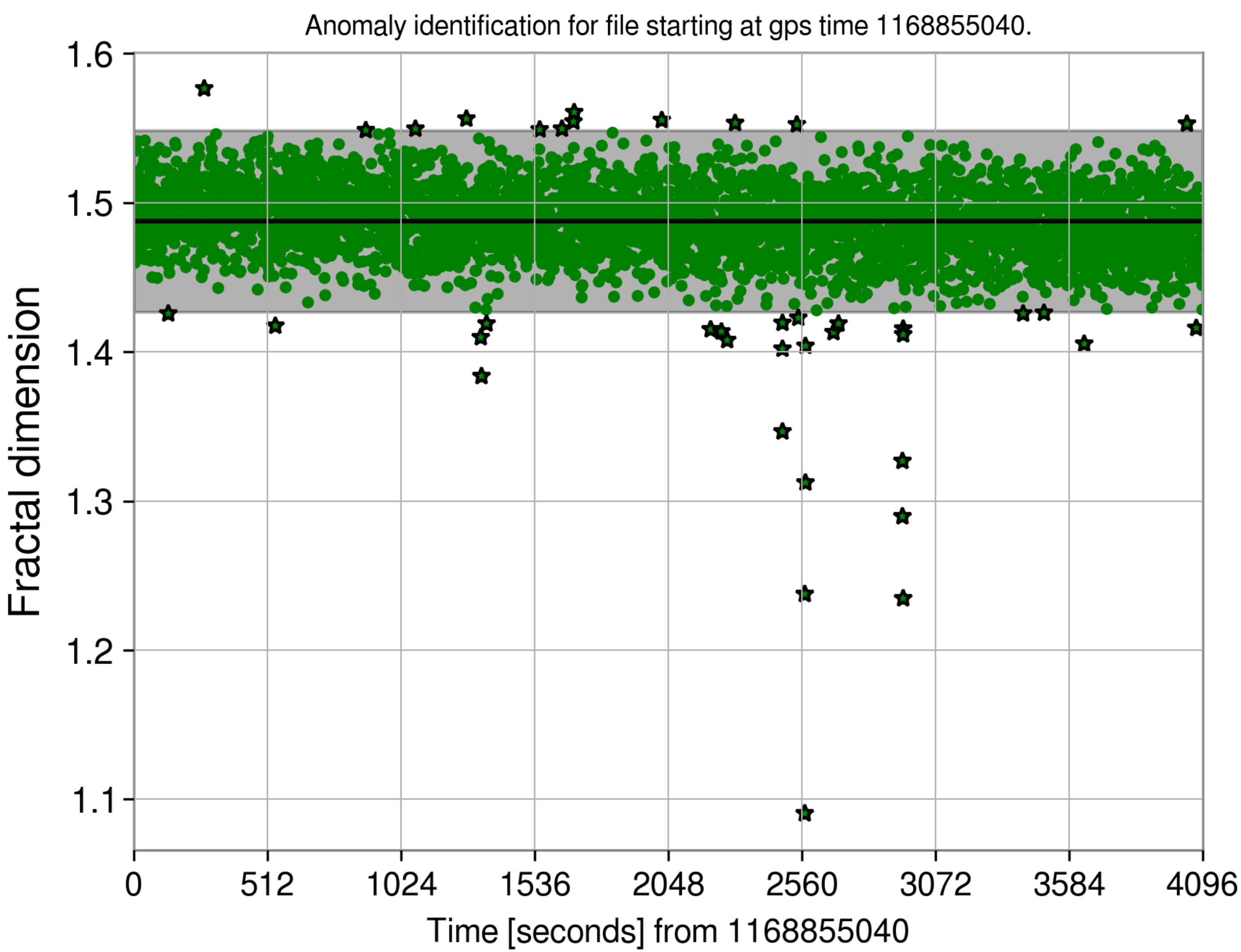
February 24, 2017

Glitch identification



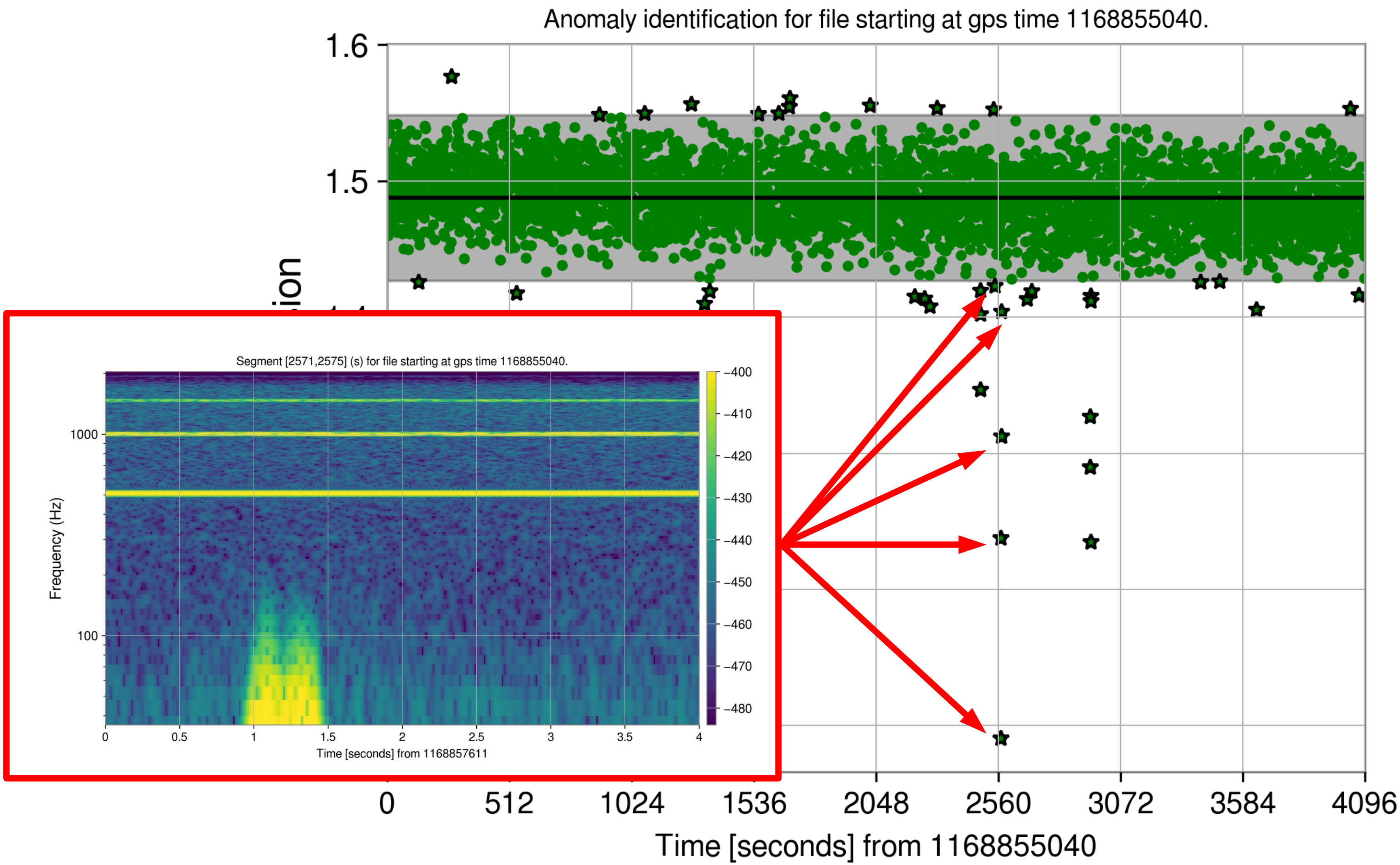
Glitch identification

Deviations > 2.5 σ



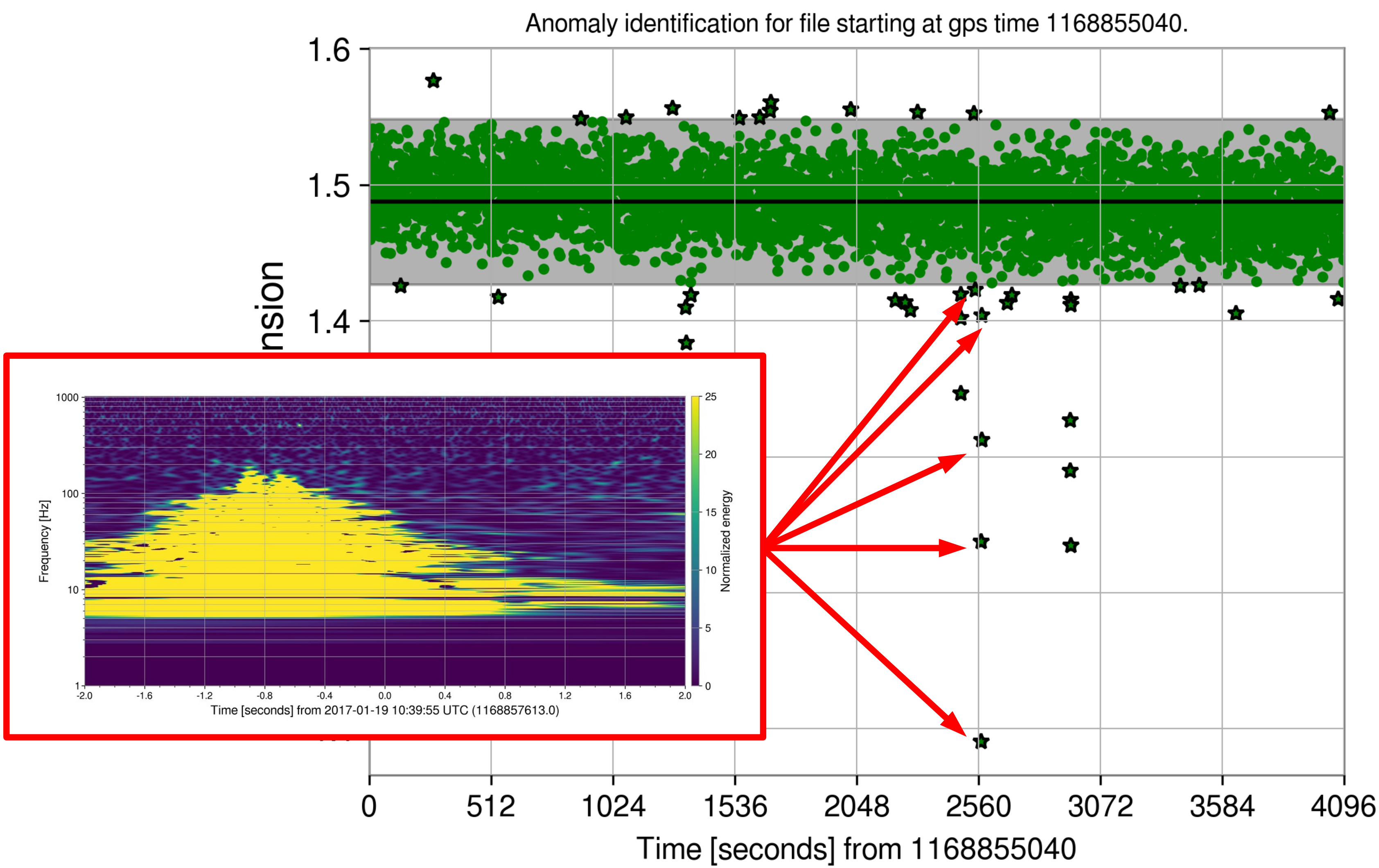
Glitch identification

Deviations > 2.5 σ



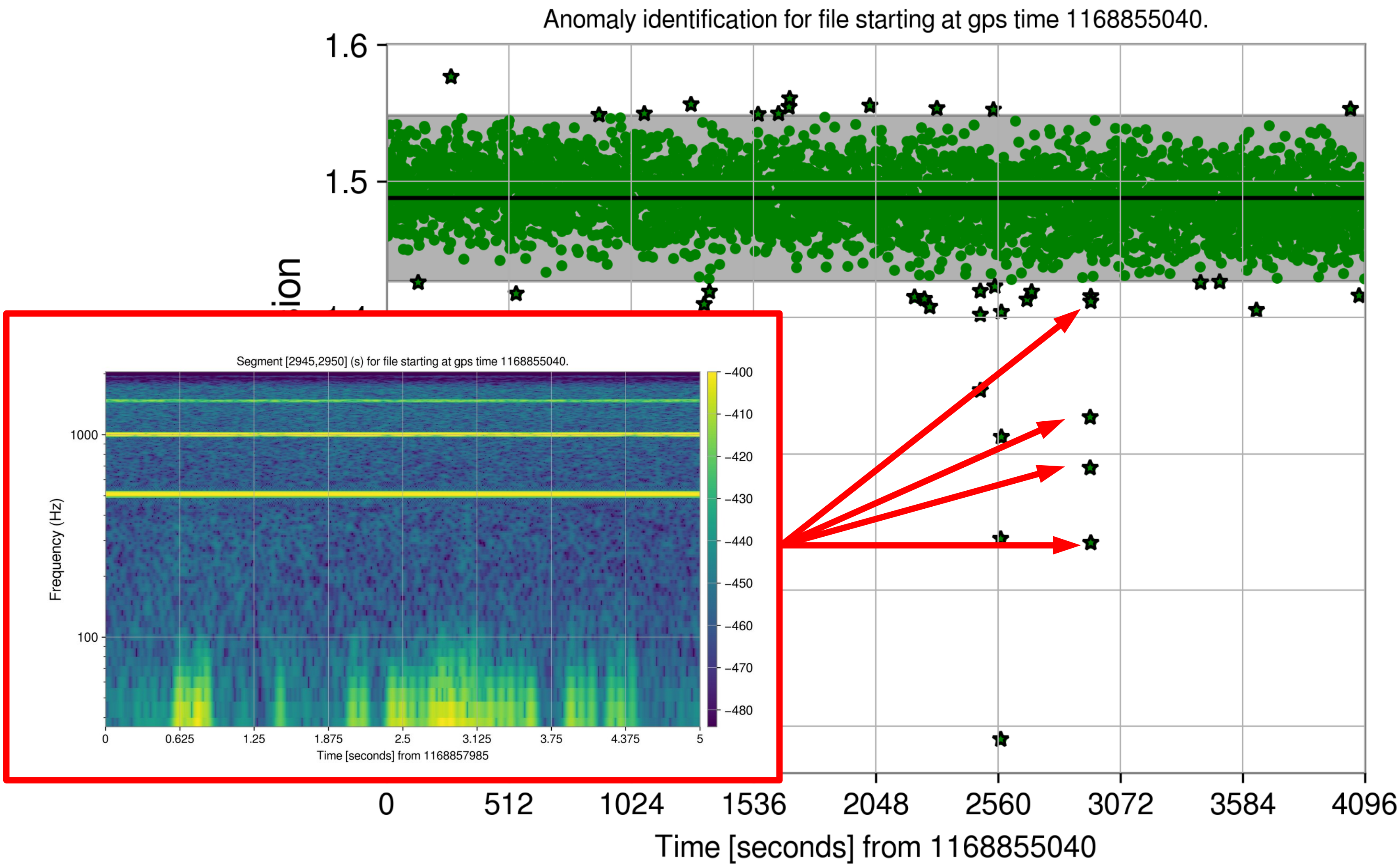
Glitch identification

Deviations > 2.5 σ



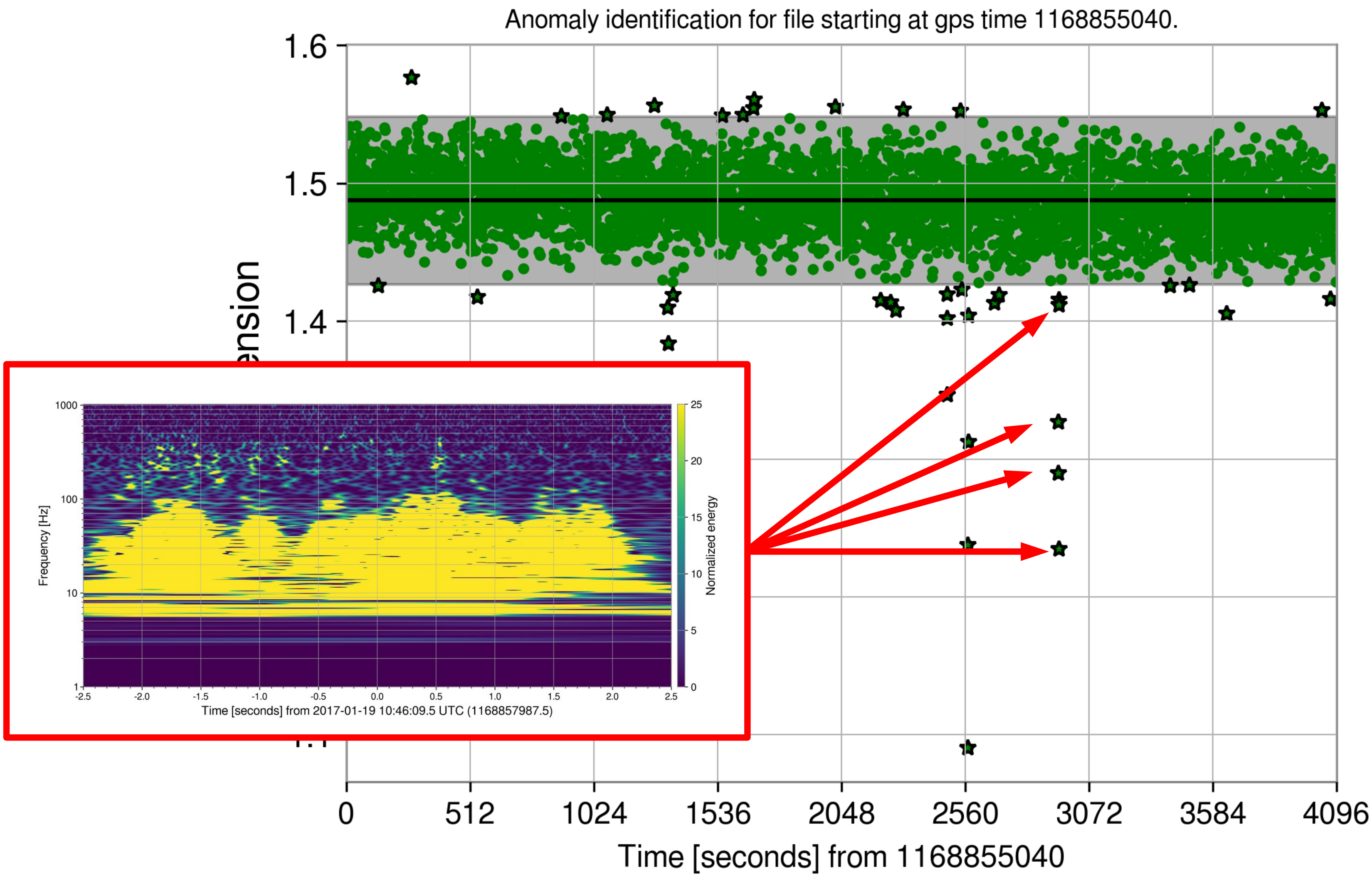
Glitch identification

Deviations > 2.5 σ

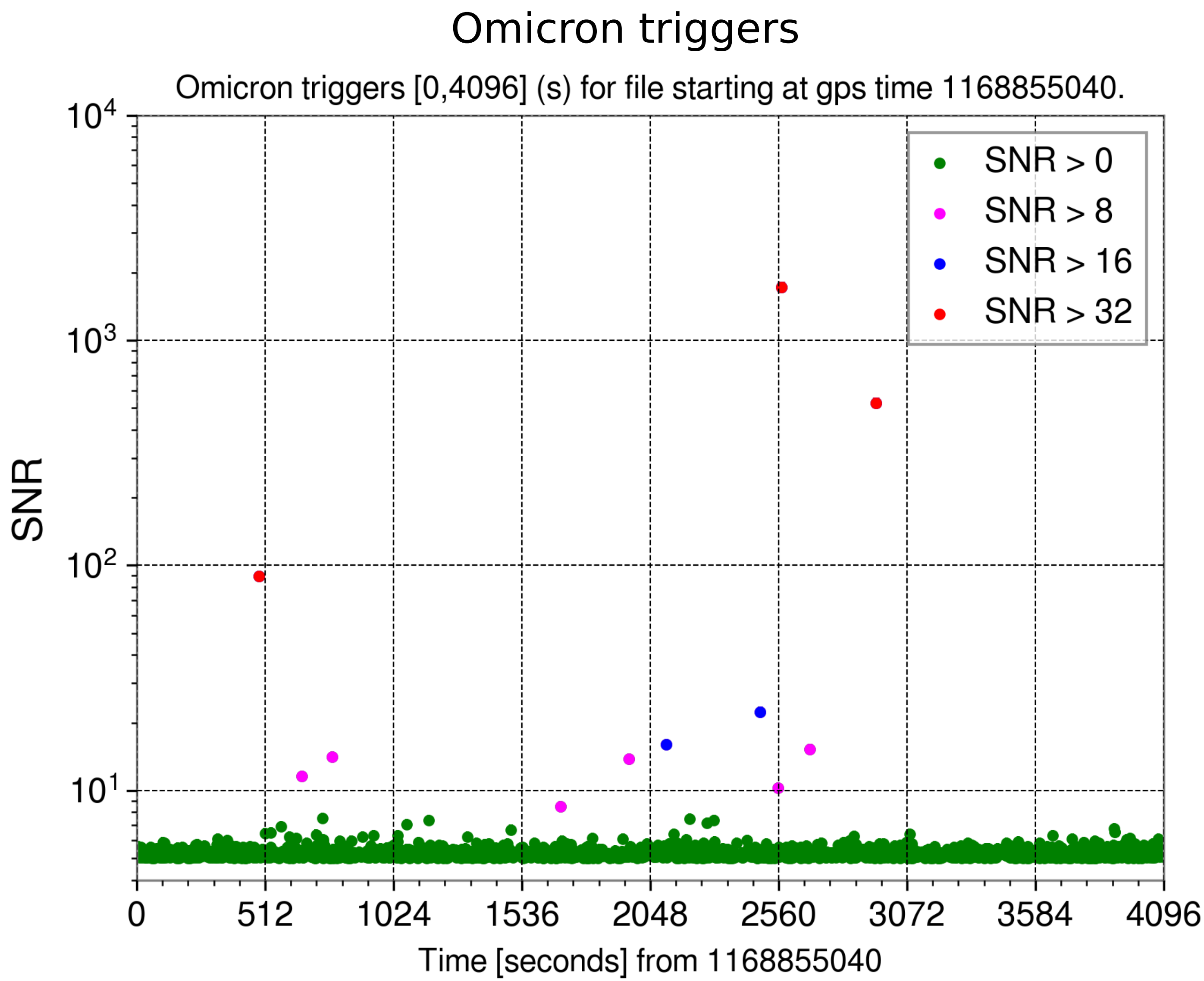


Glitch identification

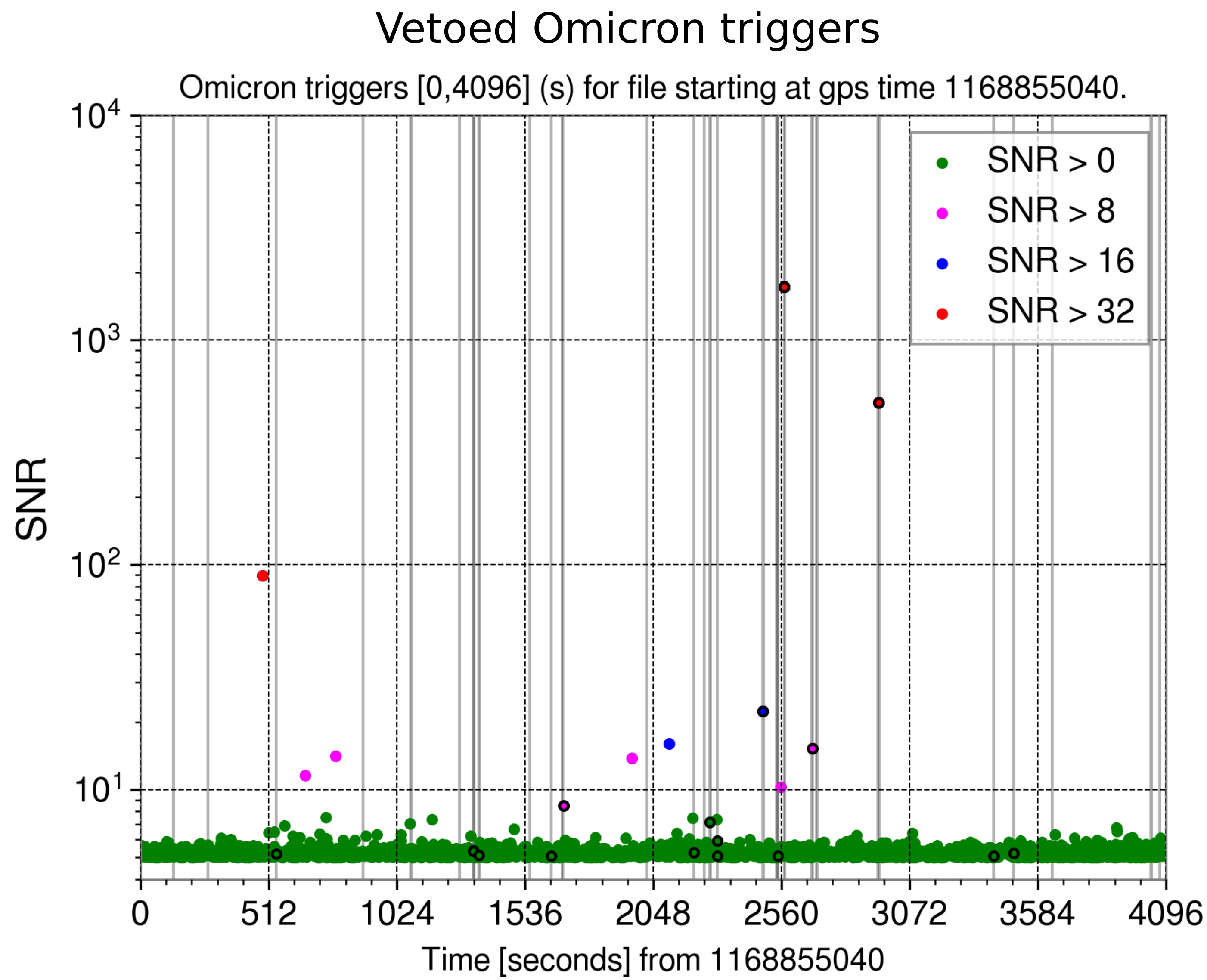
Deviations > 2.5 σ



Glitch identification

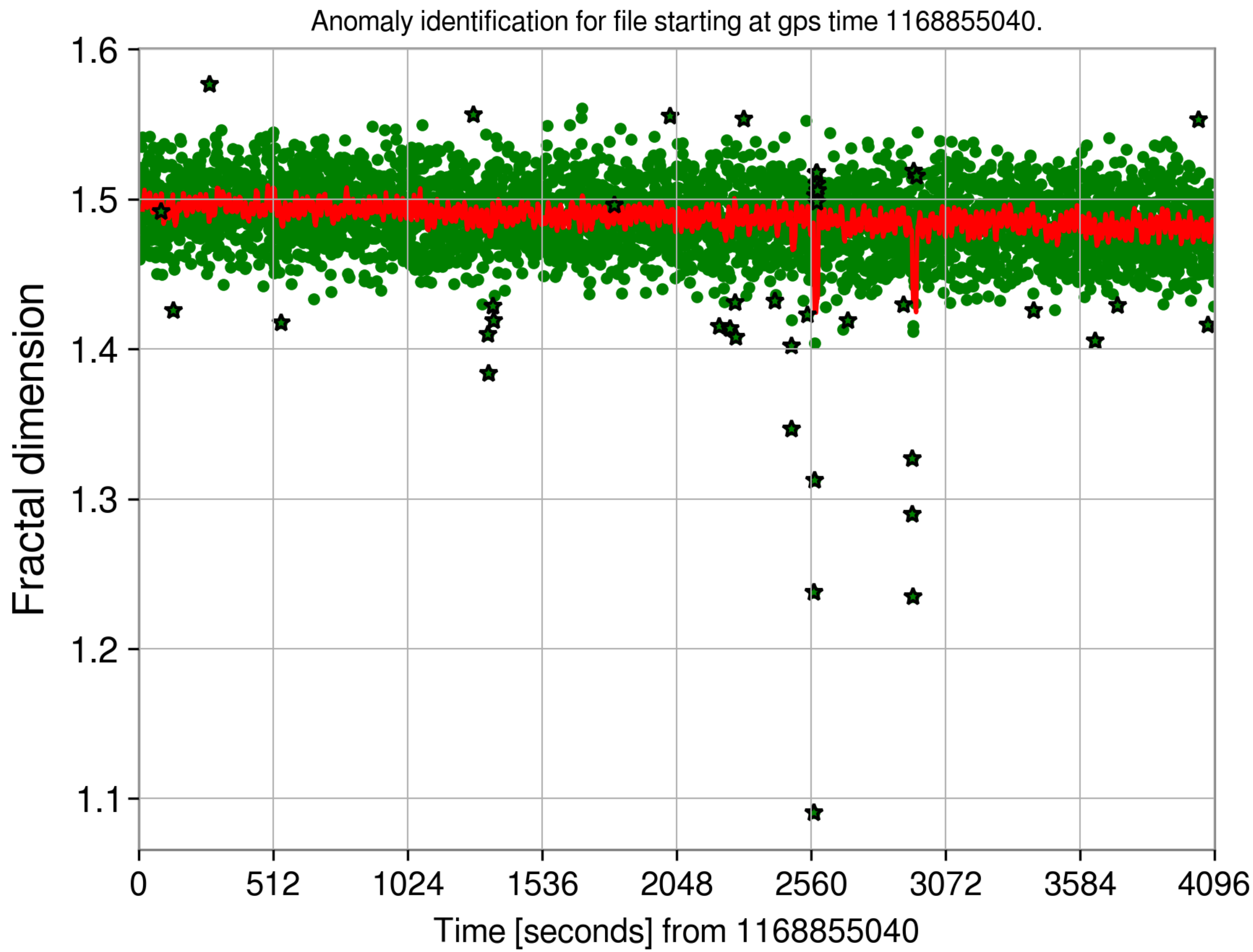


Glitch identification

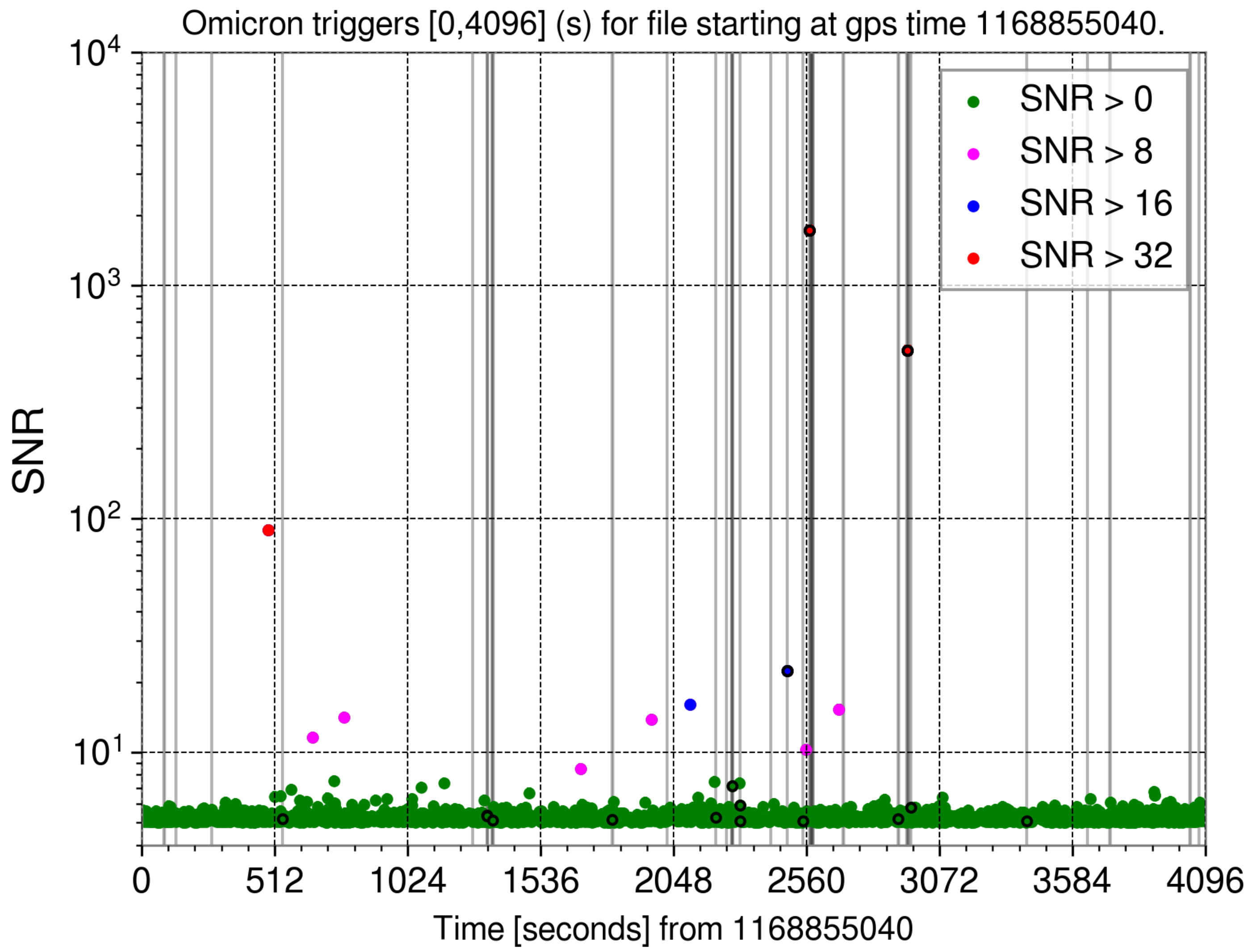


Glitch identification

Rolling mean + anomaly detection with Local Outlier Factor (LOF)

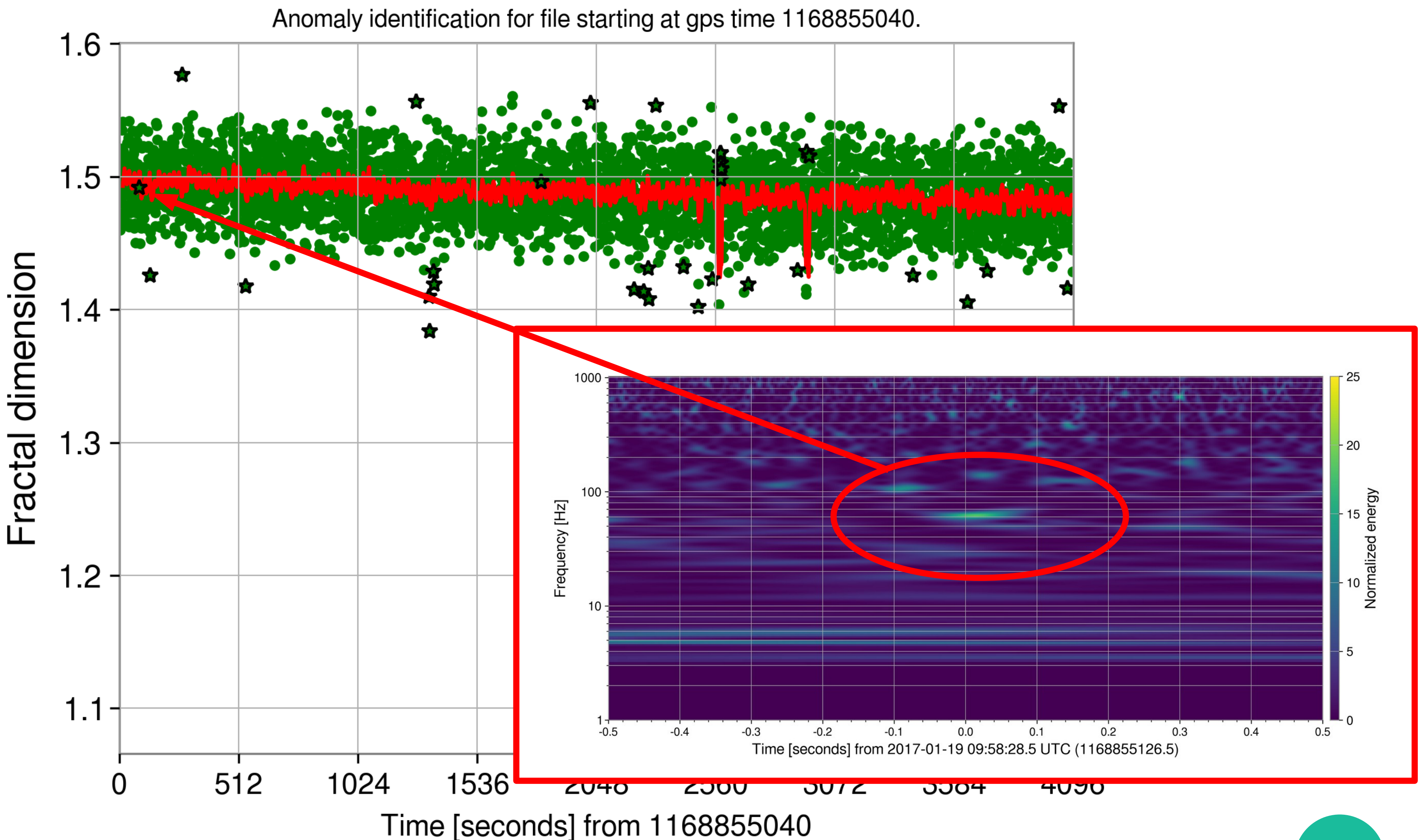


Glitch identification



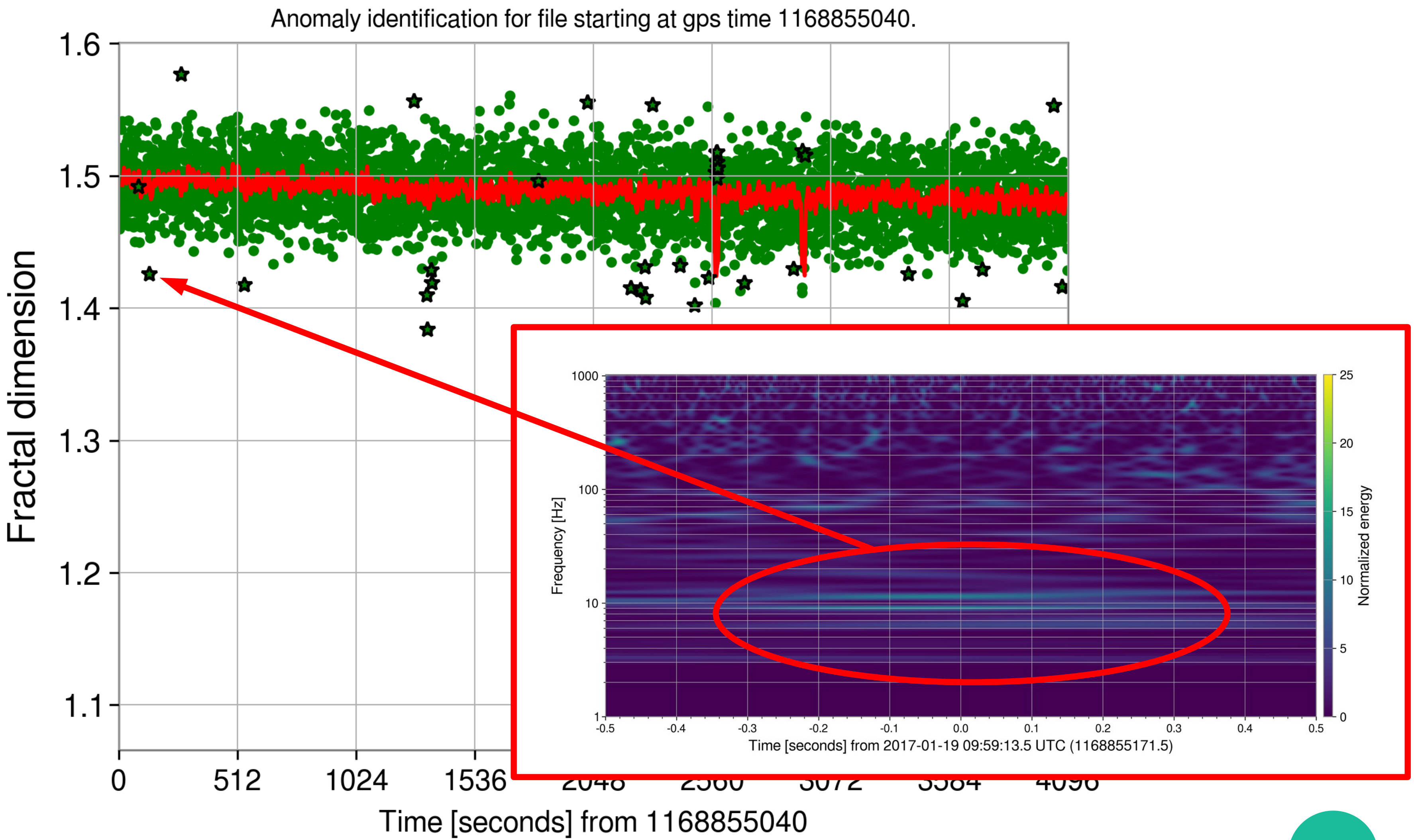
Glitch identification

Rolling mean + anomaly detection with Local Outlier Factor (LOF)

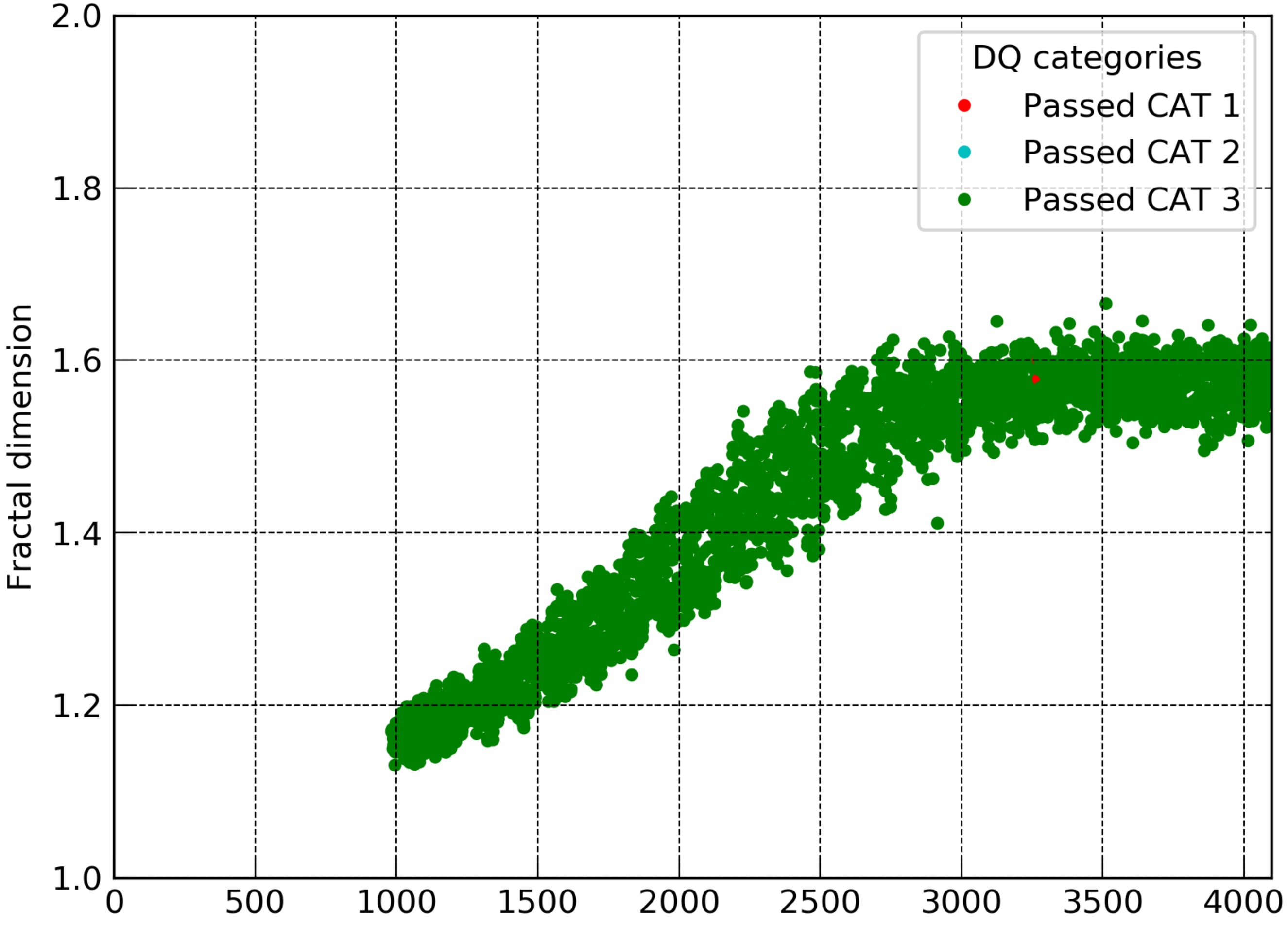


Glitch identification

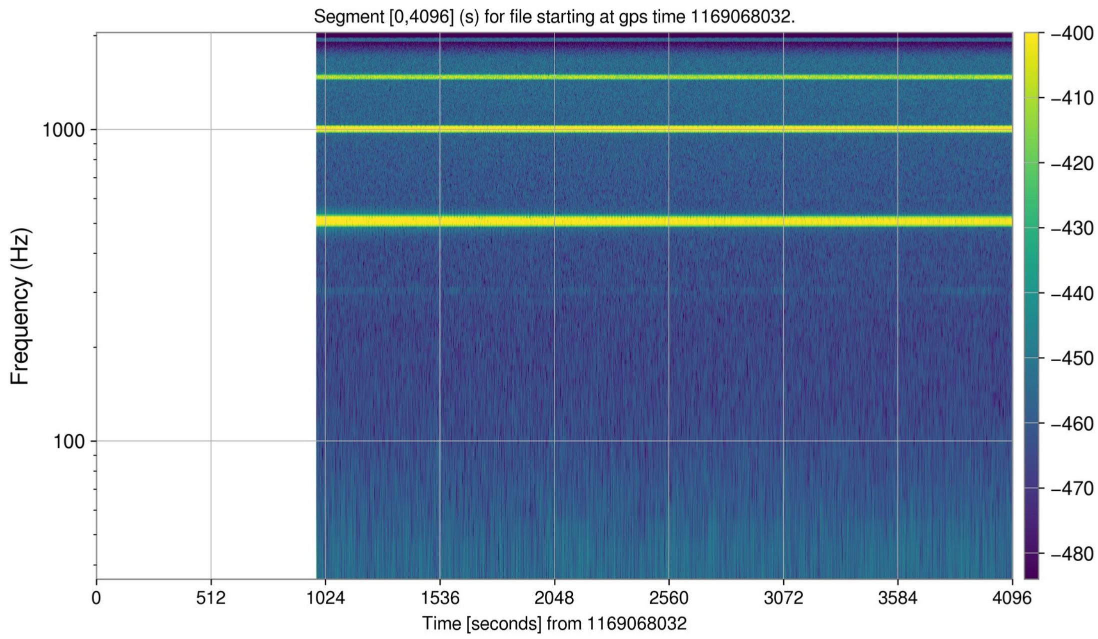
Rolling mean + anomaly detection with Local Outlier Factor (LOF)

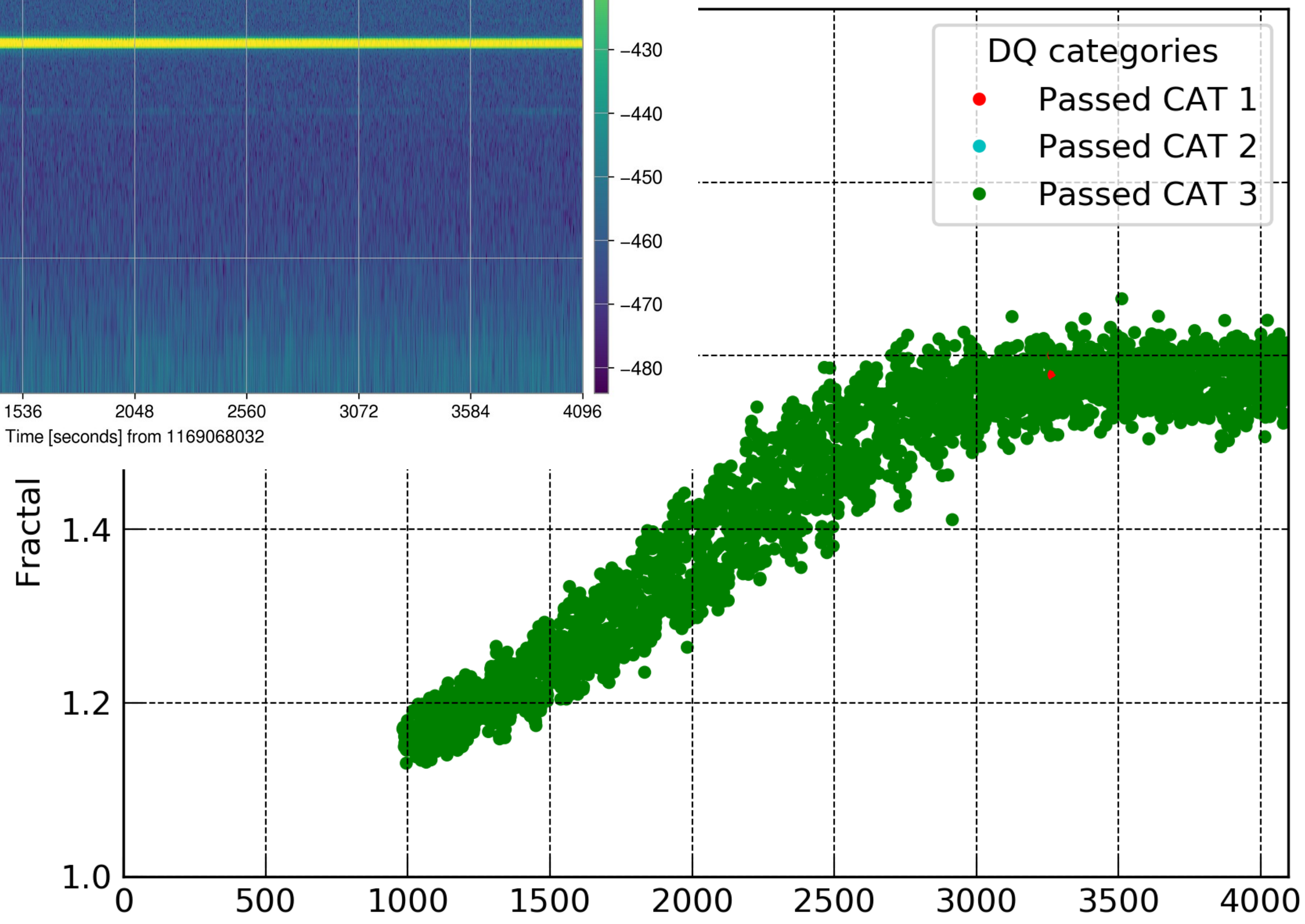


Long-term noise variation

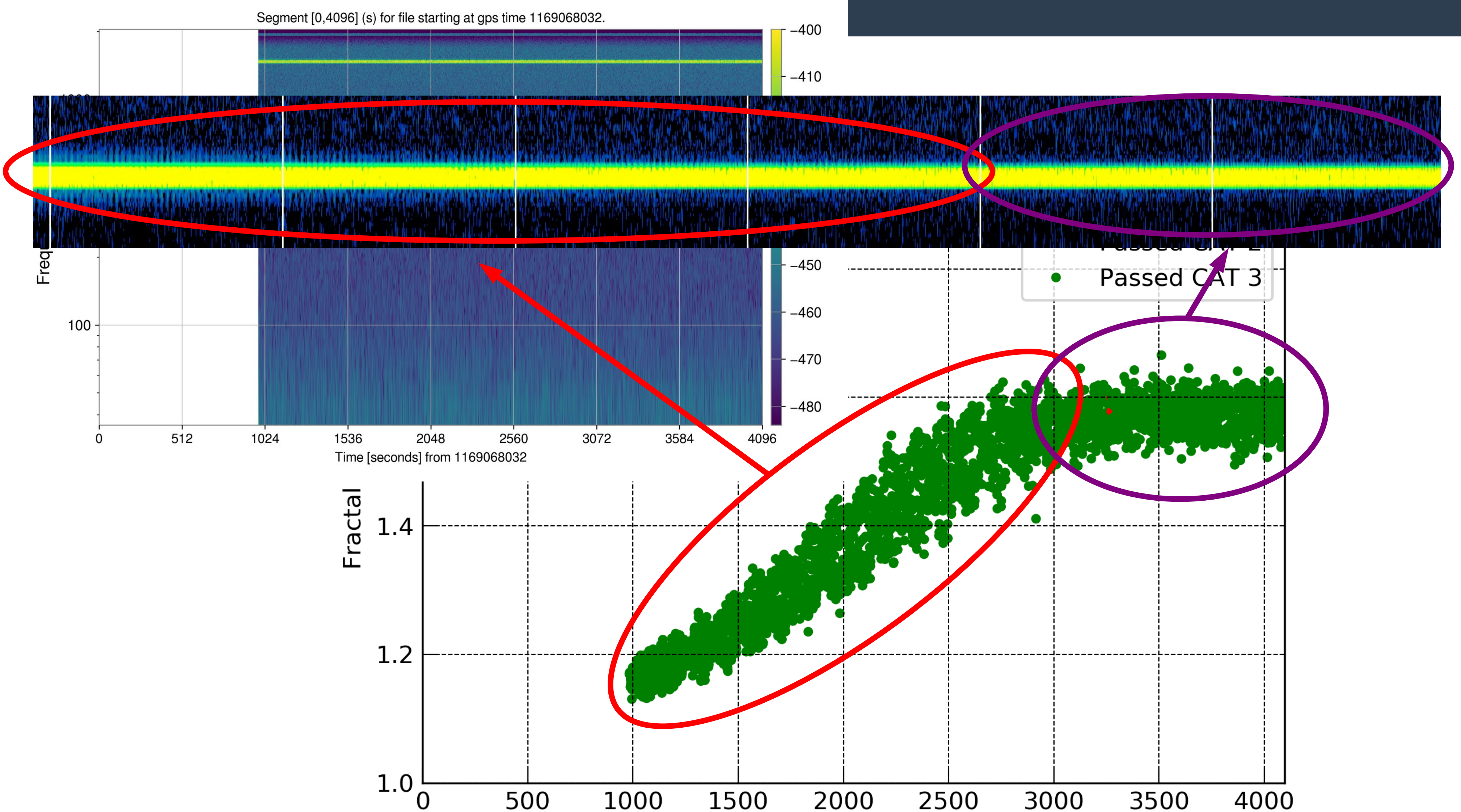


Long-term noise variation





Long-term noise variation



Conclusions

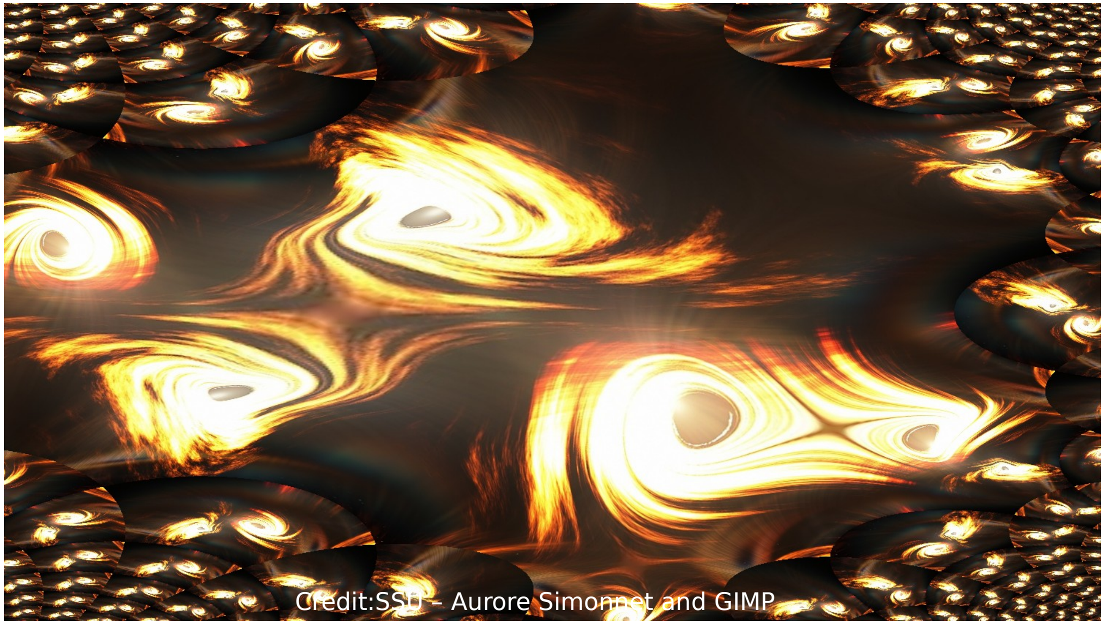
- Simple algorithms that can run on GPUs with numba → speed x 100
- One second of O2 open data at 4kHz decimated by a factor 16 is processed with the VAR algorithm on Caltech's GPU's pcdev11 in less than 0.6 s (including I/O)
- ~ 1/3 of O2 open data processed in a few days:
  https://ldas-jobs.ligo.caltech.edu/~marco.cavaglia/Fractals/
- Fractal dimension can characterize the noise and can be processed in real time!

Useful references:

- M. Bigerelle, I. Alain, *Fractal dimension and classification of music*, Chaos Solitons & Fractals 11(14):2179-2192 (November 2000), DOI:10.1016/S0960-0779(99)00137-X
- P. Maragos, A. Potamianos, *Fractal dimensions of speech sounds: Computation and application to automatic speech recognition*, The Journal of the Acoustical Society of America 105(3):1925-32 (April 1999) DOI:10.1121/1.426738.
- M. Yamaguti et al. "Mathematics of Fractal" Mathematical monographs, vol.167 AMS 1997.

Thank you! Questions(†) ?

(†) “If you ask me a question I do not know, I’m not going to answer it”
– Yogi Berra



Credit:SSU - Aurore Simonnet and GIMP

The author thankfully acknowledges the human and material resources of the LIGO Scientific Collaboration and the Virgo Collaboration that have made possible the results presented in this talk, and the National Science Foundation for its continuous support of LIGO science and basic and applied research in the United States. **This work has been partially supported by NSF grant PHY-2011334.**

LIGO VIRGO NSF LSC

MISSOURI S&T

INSTITUTE FOR MULTI-MESSENGER ASTROPHYSICS AND COSMOLOGY