The correlation analysis based on the information-theoretic measurement methods

Special Session on Detector Characterization, KIW8 July 9, 2021

Piljong Jung[1], Sang Hoon Oh[1], Young-Min Kim[2], Edwin J. Son[1], John J. Oh[1]

[1] National Institute for Mathematical Sciences
[2] Department of Physics, Ulsan National Institute of Science and Technology

CONTENTS

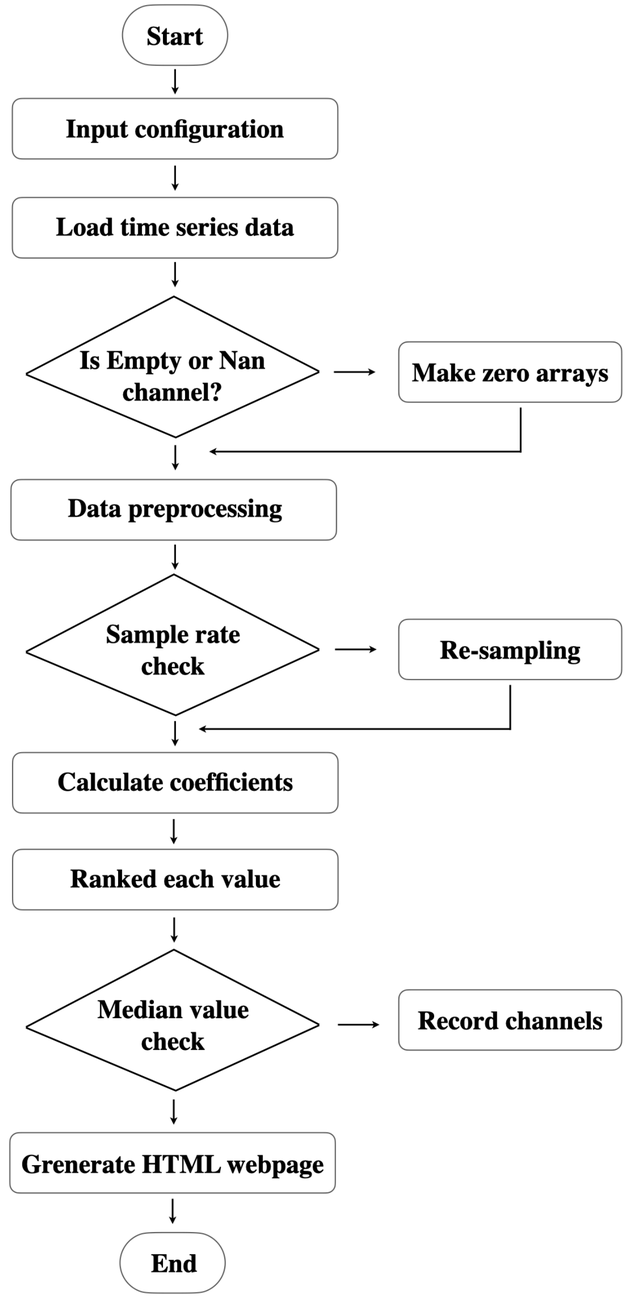
01 About CAGMon
- Overall
- Background knowledge
- Algorithm

02 Exemplary results
- Lightning strike event
- Air compressor effects

03 Summary

ABOUT CAGMON

- CAGMon
  : is the Correlation Analysis based on Glitch Monitoring
- This model evaluates the independence of channels through three coefficients,
  - The estimator for the population of maximal information coefficient (MICe)
  - Pearson Correlation Coefficient (PCC)
  - Kendall's tau
- Powered by
  Python with GWpy, scipy, numpy, and MINEpy



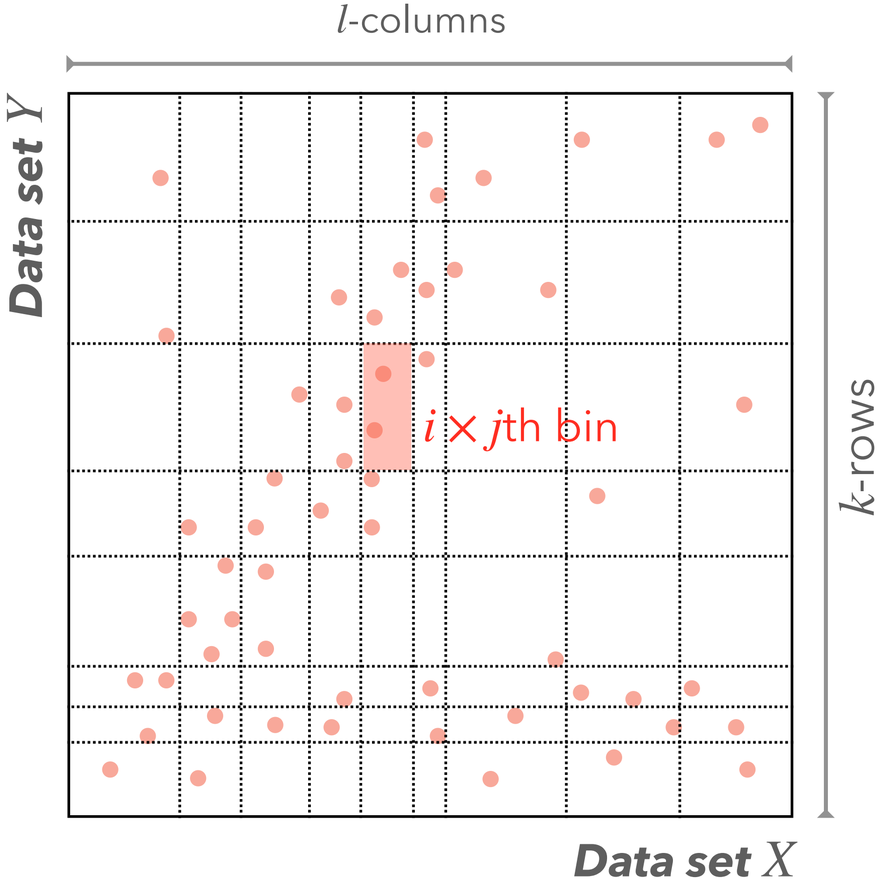
MIC*e*

Definition

The estimator for the population of maximal information coefficient (MICe) of a set D of two-variable data with sample size n and grid less than $B(n)$ is given by

$$MICe(D) = \max_{kl<B(n)} \left\{ \frac{I^{[*]}(D,k,l)}{\log_2 \min\{k,l\}} \right\}$$

, where $\omega(1) < B(n) \leq O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$.



Properties

- $MICe$ range from 0 to +1
- $MICe$ is less affected by noises
- $MICe$ can not only show the linear relationship also Non-linear relationship
- $MICe$ is hard to calculate

D. N. Reshef , et al. , Science, vol. 334, no. 6062, pp. 1518–1524, 2011

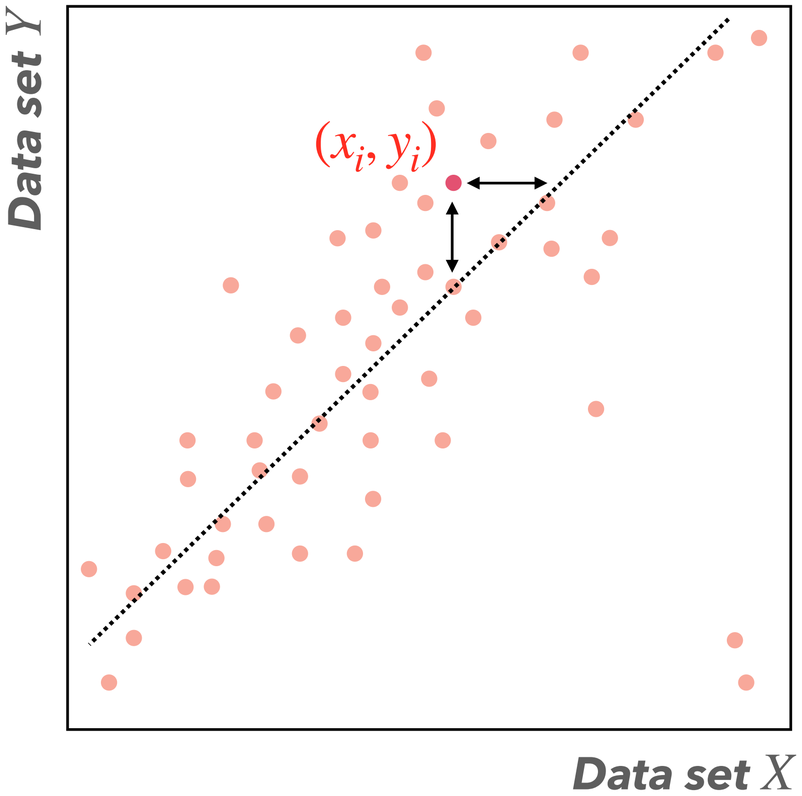
PCC

Definition

Pearson Correlation Coefficient(PCC) is a statistic that explains the amount of variance accounted for in the relationship between two (or more) variables by

$$p(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}}$$

, where $\bar{x}$ and $\bar{y}$ are the mean of x and y, respectively.



Properties

- PCC range from -1 to +1
- Absolute value larger than 0.5 is usually considered a significant linear relationship
- PCC shows Linear correlation, not causation

wikipedia, Pearson correlation coefficient

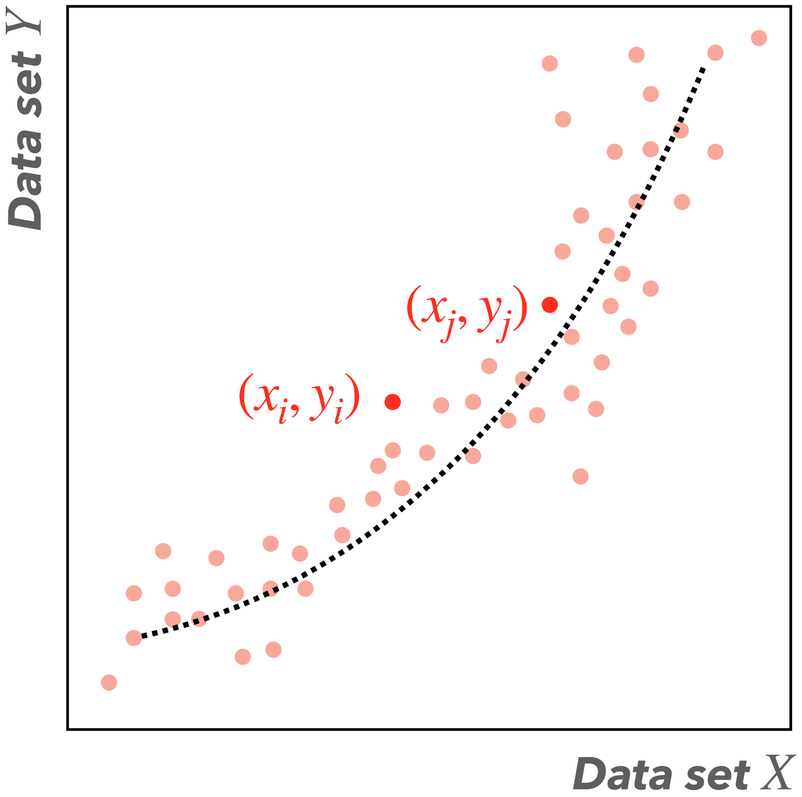
KENDALL'S TAU

Definition

Kendall's tau with a random samples n of observations from two variables measures the strength of the relationship between two ordinal level variables by

$$\tau = \frac{c - d}{\binom{n}{2}}$$

, where c is the number of concordant pairs, and d is the number of discordant pairs.



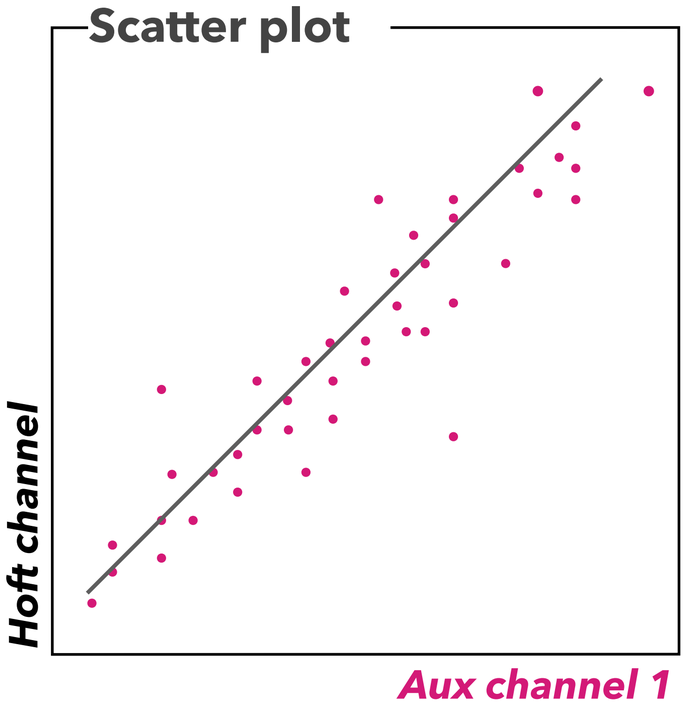
Properties

- Kendall's tau coefficients range from -1 to +1
- Kendall's *Tau* measures the strength of the relationship between two ordinal level variables

wikipedia, Kendall's tau coefficient

COMPARISON

Linear correlation

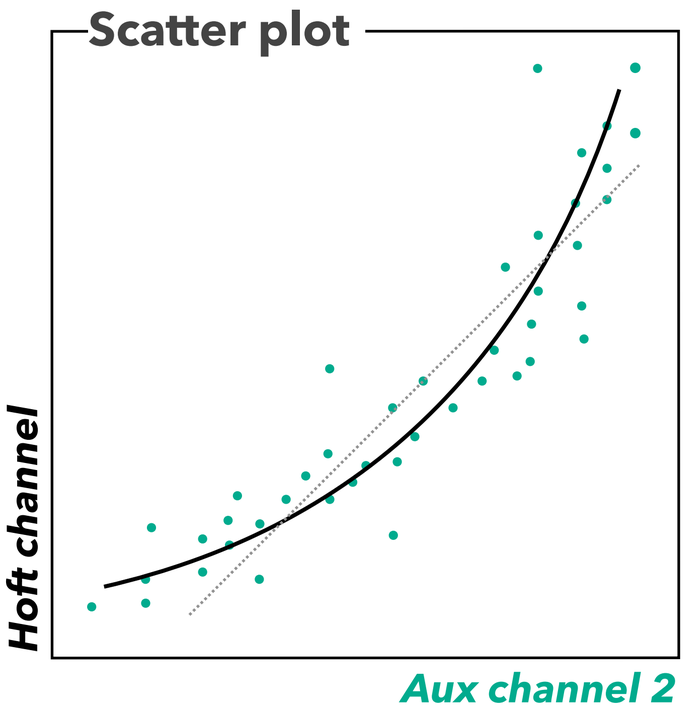


MICe: 0.9

PCC: 0.9

Kendall: 0.9

Ordinal correlation

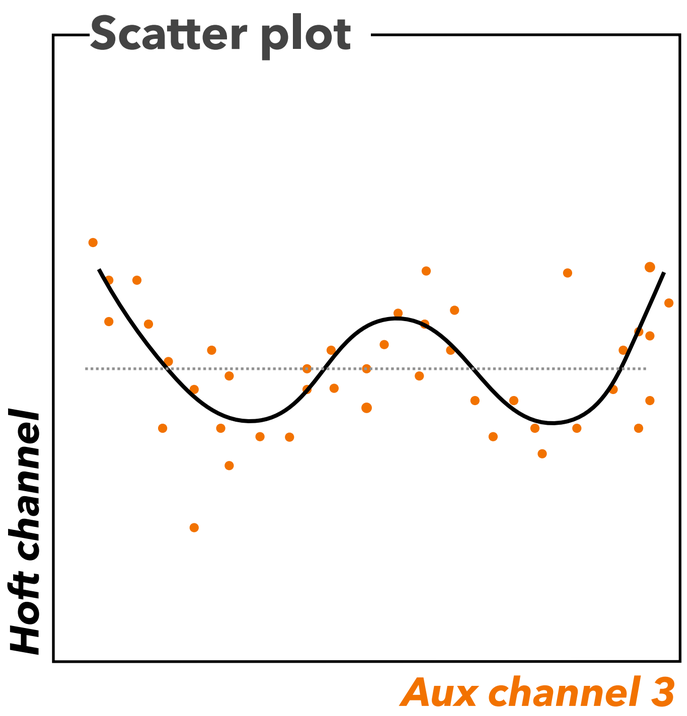


MICe: 0.9

PCC: 0.6

Kendall: 0.8

Non-linear correlation

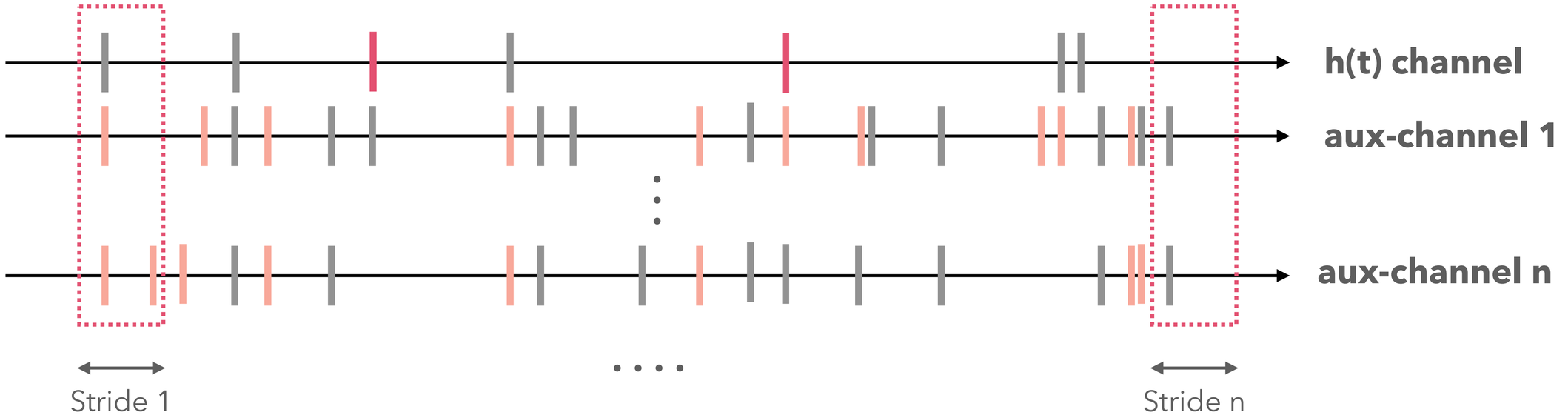


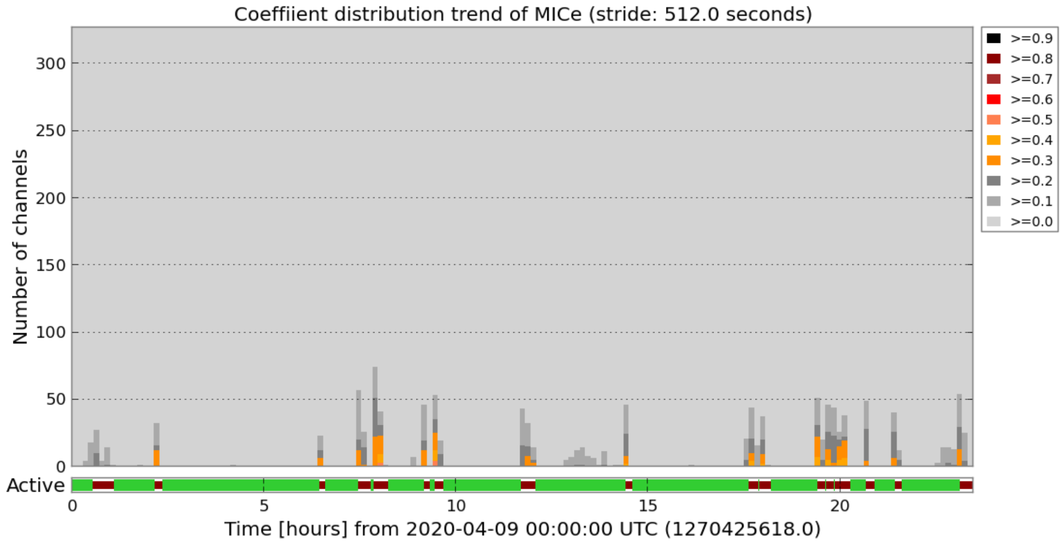
MICe: 0.9

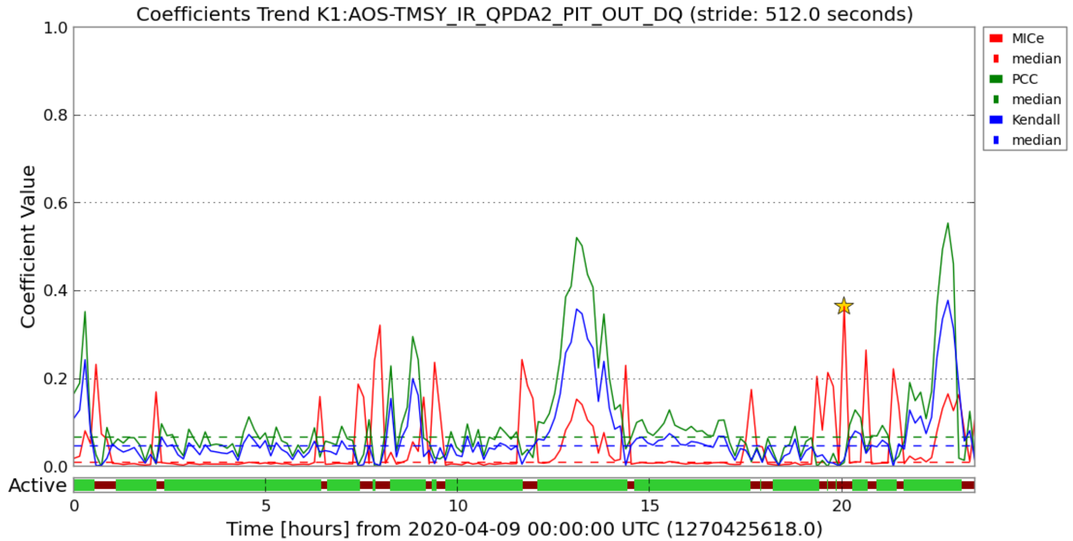
PCC: 0.3

Kendall: 0.2

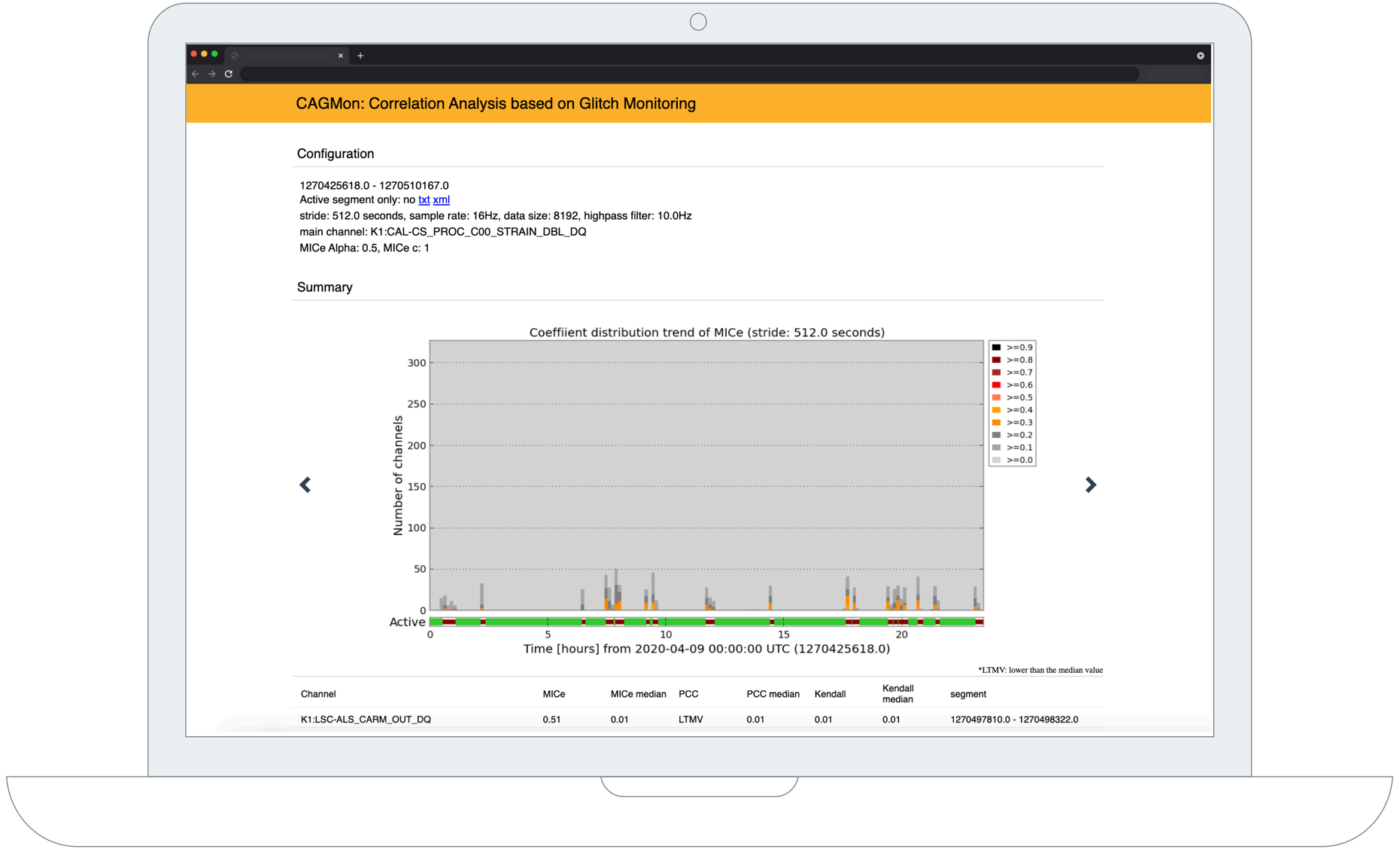
ALGORITHM


h(t) channel
aux-channel 1
aux-channel n
Stride 1
Stride n

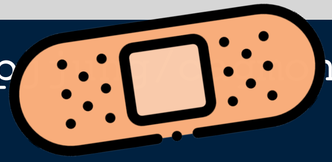

Coeffiient distribution trend of MICe (stride: 512.0 seconds)
Number of channels
Active
Time [hours] from 2020-04-09 00:00:00 UTC (1270425618.0)
>=0.9
>=0.8
>=0.7
>=0.6
>=0.5
>=0.4
>=0.3
>=0.2
>=0.1
>=0.0


Coefficients Trend K1:AOS-TMSY_IR_QPDA2_PIT_OUT_DQ (stride: 512.0 seconds)
Coefficient Value
Active
Time [hours] from 2020-04-09 00:00:00 UTC (1270425618.0)
MICe
median
PCC
median
Kendall
median

ALGORITHM


CAGMon: Correlation Analysis based on Glitch Monitoring
Configuration
1270425618.0 - 1270510167.0
Active segment only: no txt xml
stride: 512.0 seconds, sample rate: 16Hz, data size: 8192, highpass filter: 10.0Hz
main channel: K1:CAL-CS_PROC_C00_STRAIN_DBL_DQ
MICe Alpha: 0.5, MICe c: 1
Summary
Coeffiient distribution trend of MICe (stride: 512.0 seconds)
Number of channels
300
250
200
150
100
50
0
Active
0
5
10
15
20
Time [hours] from 2020-04-09 00:00:00 UTC (1270425618.0)
>=0.9
>=0.8
>=0.7
>=0.6
>=0.5
>=0.4
>=0.3
>=0.2
>=0.1
>=0.0
*LTMV: lower than the median value
Channel
MICe
MICe median
PCC
PCC median
Kendall
Kendall median
segment
K1:LSC-ALS_CARM_OUT_DQ
0.51
0.01
LTMV
0.01
0.01
0.01
1270497810.0 - 1270498322.0

ALGORITHM



```
[KAGRA@USER ~]git clone https://github.com/p                    n.git
[KAGRA@USER ~]cd cagmon
[KAGRA@USER cagmon]ls
  LICENSE          README.md          cagmon          setup.py
[KAGRA@USER cagmon]python setup.py install
running install
running bdist_egg
running egg_info
.
.
Finished processing dependencies for CAGMon==0.8.0
[KAGRA@USER cagmon]cagmon ——help
usage: cagmon [-h] [-v] [-c CONFIG]

optional arguments:
  -h, --help            show this help message and exit
  -v, --version         Show version of CAGMon
  -c CONFIG, --config CONFIG
                        the path of CAGMon configuration file
```

ALGORITHM

```
[KAGRA@USER ~]cagmon --config test_config.ini
```

```
[Configuration Information]
 Start GPS time: 1271289618.0
 End GPS time: 1271376018.0
 Main channels: K1:DAS-RANGE_BNS
 Sample rate: 16Hz
 Whitening option: no
 Active segment only option: no
 Show additional plots option : no
 Defined segment condition: K1:GRD-LSC_LOCK_STATE_N == 1000
 Coefficient trend stride: 512.0 seconds
[Computing Resources]
 Given CPUs: 72 cores
 Given memory: 377.636459351 GB
[Configuration Validation Check]
 [OK] Cache
 [OK] Main channel
 [OK] Aux-channels
 [OK] Segment
 [OK] Stride
[Process Begins]
```

EXEMPLARY RESULTS

- **Development status and needs**
  - It is still in the development stage
  - It is required to diagnosis its reliability and limitations
- **Exemplary result 1**
  - The lightning stroke event (Washimi et al. arXiv:2103.06516, KIW8 poster P5)
  - If this method could not provide us with the same result of the lightning stroke event, we cannot believe any analysis result from them.
- **Exemplary result 2**
  - An air compressor effect
  - It is a general test to verify the reliability of association analysis if the BNS range channel is applied to a primary channel.

LIGHTNING STROKE EVENT

arXiv:2103.06516

Effects of lightning strokes for underground gravitational wave observatories

T. Washimi[a,1] T. Yokozawa,[b] M. Nakano,[b] T. Tanaka,[b] K. Kaihotsu,[c] Y. Mori,[c] and T. Narita[d]

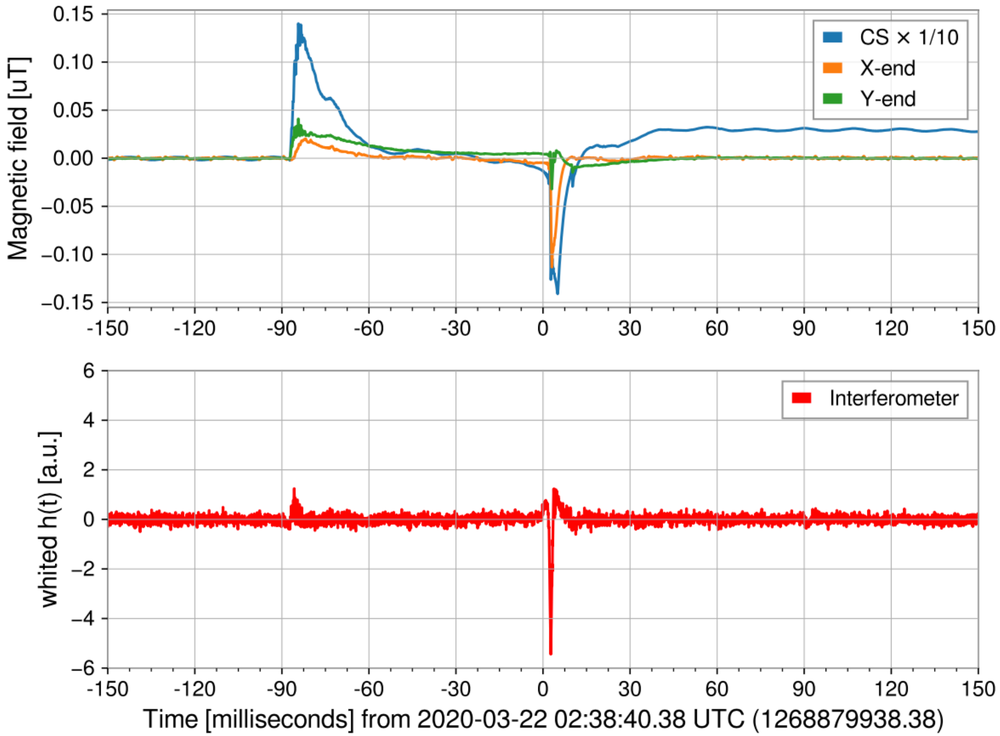
[a] *Gravitational Wave Science Project (GWSP), Kamioka branch, National Astronomical Observatory of Japan (NAOJ), Kamioka-cho, Hida City, Gifu 506-1205, Japan*

[b] *Institute for Cosmic Ray Research (ICRR), KAGRA Observatory, The University of Tokyo, Kamioka-cho, Hida City, Gifu 506-1205, Japan*

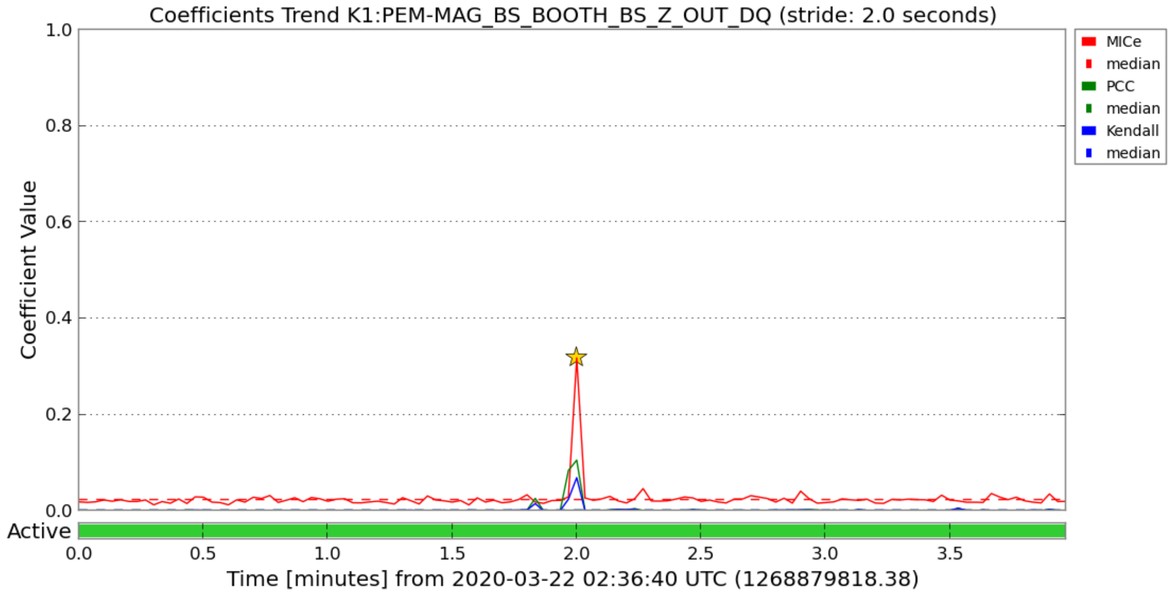
[c] *Department of Physics, the University of Toyama, Gofuku, Toyma City, Toyama 930-8555, Japan*

[d] *Department of Electrical and Electric Engineering, Shonan Institute of Technology, Tsujido, Fujisawa City, Kanagawa 251-8511, Japan*

E-mail: tatsuki.washimi@nao.ac.jp



LIGHTNING STROKE EVENT



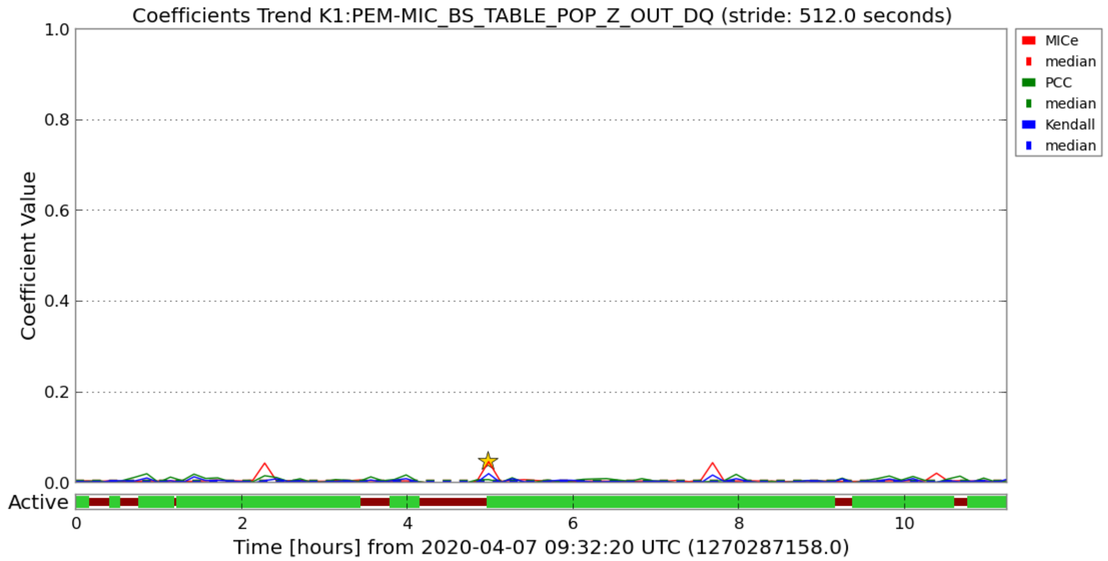
Configuration

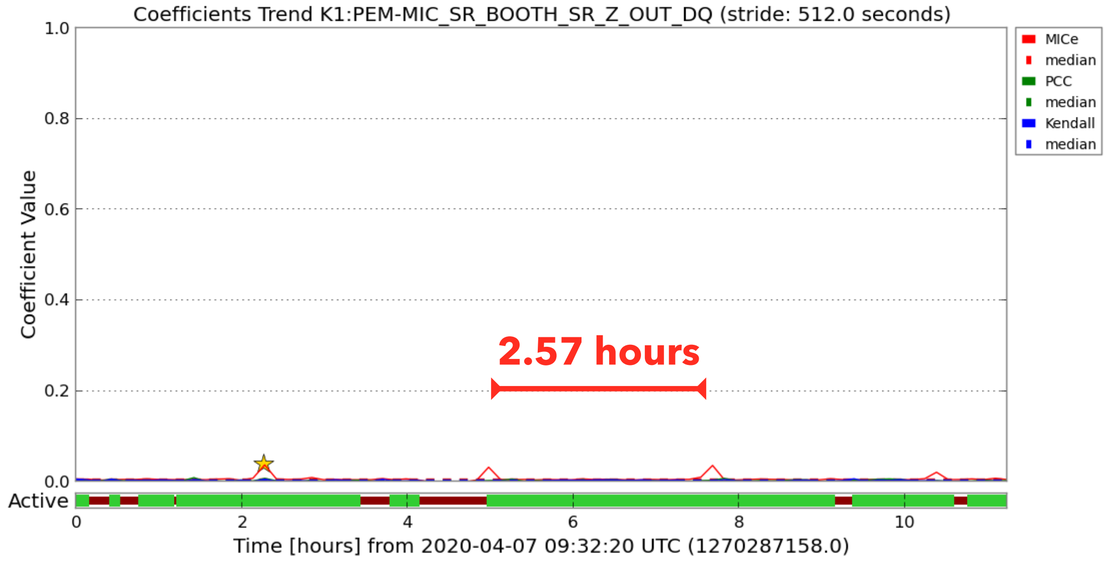
- GPS times: 1268879818.38±120s
- Stride: 2.0 seconds
- MICe alpha/c: 0.5/1
- Main channel: K1:CAL-CS_PROC_C00_STRAIN_DBL_DQ

Result

- Coefficient trends were changed at the same point of the lightning event.
- CAGMon sensed the signal what appeared in both channels within a sub-segment simultaneously.
- It achieved the minimum requirement to estimate the channel correlation.

AIR COMPRESSOR EFFECTS





Configuration

- Applied data: O3GK period
- Stride: 512.0 seconds
- MICe alpha/c: 0.5/1
- Main channel: K1:DAS-RANGE_BNS

Correlated aux-channels

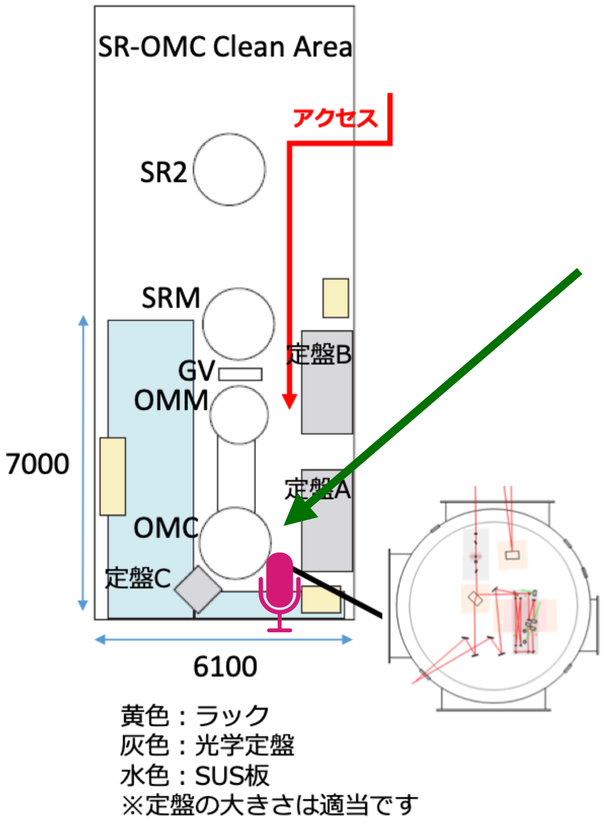
- K1:PEM-MIC_SR_BOOTH_SR_Z_OUT_DQ
- K1:PEM-MIC_MCF_TABLE_REFL_Z_OUT_DQ
- K1:PEM-MIC_BS_TABLE_POP_Z_OUT_DQ
- K1:PEM-MIC_BS_TABLE_POS_Z_OUT_DQ

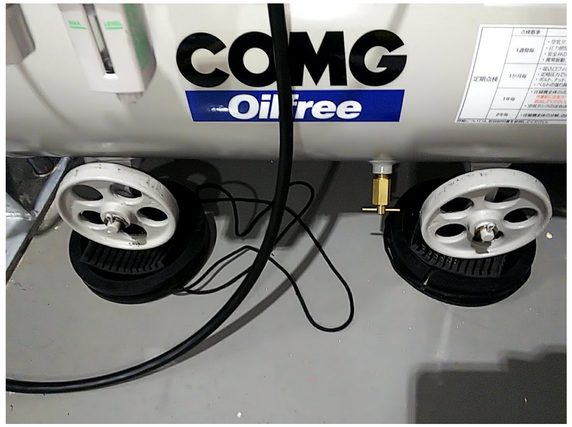
Properties of correlation

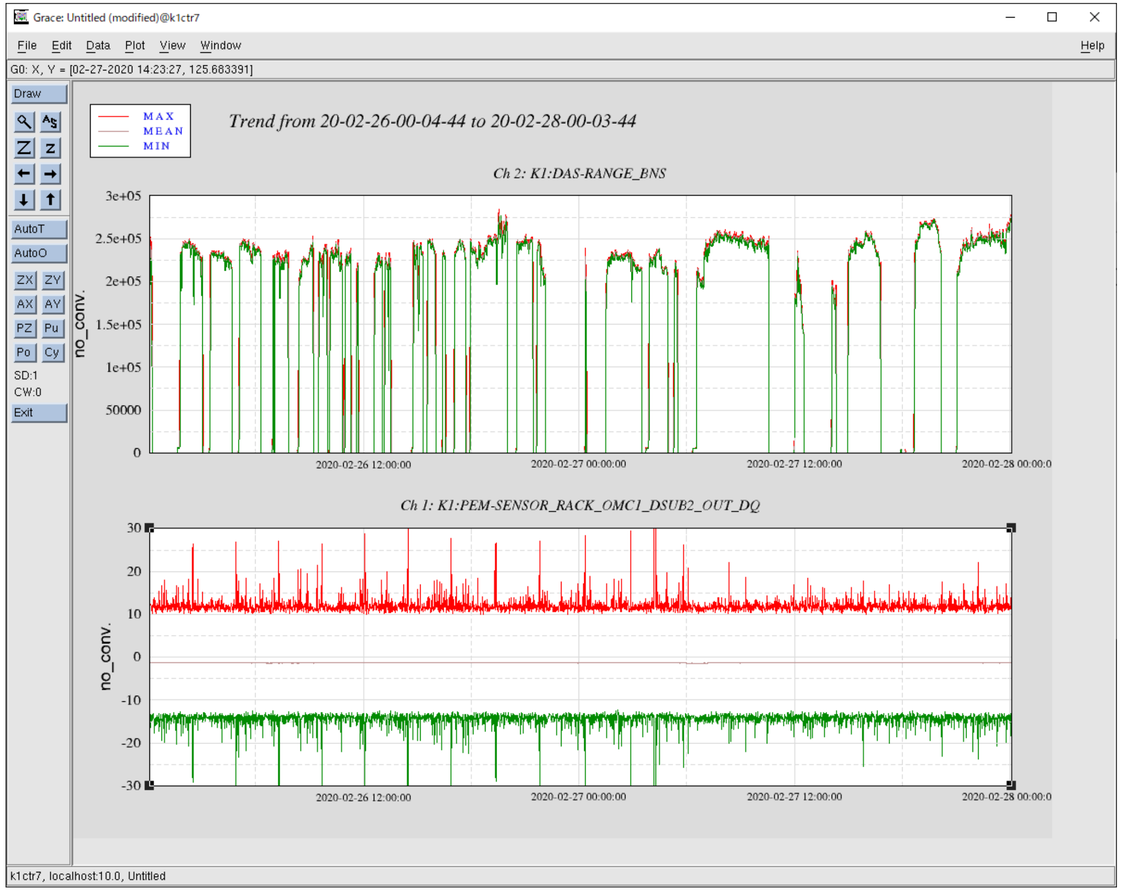
- It was a periodic phenomenon for 2.57 hours during whole O3GK period
- Probably, it affected to the BNS range drop
- It was placed around BS area
- It created 58Hz of frequency

AIR COMPRESSOR EFFECTS

klog:13206 klog:13181 klog:13197







SUMMARY

- **Development of CAGMon**
  - The goal of this project is to find a systematic way of identifying the abnormal glitches/ non-linear correlations in the gravitational-wave data using the information-theoretic measurement methods

- **Analysis results by CAGMon**
  - It sensed the association between the stain and magnetometer channels when the lightning strikings
  - It also detected the relationship between the BNS range and OMC chamber vibration due to an air compressor

- **Future plan**
  - Keep testing the reliability, performance, and limitation