# ENHANCEMENT OF INTEGER-SECOND NOISE ANALYSIS TOOLS FOR LIGO GRAVITATIONAL DATA

Sherman Thompson
Advisor: Keith Riles

April 19th, 2022

**Abstract**

We detail the enhancement of tools used to identify noise repeated at integer-second intervals in LIGO strain data. We then detail a new tool developed to assign correlation factors between said periodic noise in the strain channels to periodic noise in auxiliary channels with the intent of identifying potential causes of the periodic noise seen in the strain channel. Finally, we demonstrate use-cases for these tools and show how they may be used to explain artifacts seen in the strain channel.

## 1 INTRODUCTION

LIGO refers to two gravitational wave observatories: one in Livingston, Louisiana which will be referred to as "L1", and one at the DOE Hanford site in Washington state, referred to as "H1". These observatories basically consist of two very large Michelson laser interferometers with arms about 4km in length. When a gravitational wave passes through one of the detectors, it changes the length of one arm relative to the other, thus causing the two originally in-phase beams of light to step slightly out of phase and interfere. The difference in length between the two channels is of the order $10^{-18}$m, which means the detector must be extremely sensitive. As such, initial LIGO achieved sensitivity of changes in strain on the order of one part in $10^{21}$. Advanced LIGO, the current phase of operation, is estimated to be six to seven times as sensitive.

Gravitational waves are an important prediction of General Relativity, and their observation provides a strong confirmation of the theory. In addition, the data from these detectors can be used to make astrophysical observations such as the calculation of the Hubble Constant. These detectors have been able to detect several astrophysical events that produce gravitational waves. First among these was GW150914, a binary black hole merger. As two black holes of masses around $36M_\odot$ and $29M_\odot$ inspiraled toward each other and eventually merged, they radiated a combined $3M_\odot$ in gravitational waves. These gravitational waves traveled from a luminosity distance of around 410 Mpc before their detection by LIGO. Since this first binary black hole merger, several more black hole mergers have been detected. However, binary black hole mergers are not the only astrophysical sources of gravitational waves powerful enough to be detected. In 2017, Advanced LIGO and VIRGO observed a binary neutron star inspiral. These two neutron stars has masses between $.86M_\odot$ and $2.26M_\odot$. In addition to simply observing the merger, LIGO and VIRGO were able to localize the event within a sky region of $28\,\mathrm{deg}^2$. This detection coincided with the detection of a $\gamma$-ray burst that Fermi-GRB detected about 1.7 seconds after coalescence, demonstrating a powerful new technique

in which electromagnetic and gravitational wave detectors worked in concert to provide a more complete description of an astrophysical event.

To achieve such sensitivity of the detectors, extreme stability of the main laser's power and wavelength, in addition to the electronics used to detect the changes in intensity is required. At this sensitivity, LIGO detectors will hear earthquakes from across the world, a car driving by, or thunderstorms. The noise of such extraneous factors can be minimized by correlating strain data from L1 and H1 together to parse out real signals from environmental noise. Another important source of noise limiting the resolution of the detector is electronic noise. Changes in voltage through the many electronics bays affect the laser's power, the detection equipment, the electronics affecting the suspension, among other things. Seemingly innocuous electronics in the same circuit as the detector can be heard in the final strain signal. Such examples range from the draw of an AC unit's compressor and fans to the regular blip of an LED in one of the electronics bays. In addition, wire loops may create back emfs which can leak into the main signal. Such back emfs also create magnetic fields in the electronics bays which can be detected. The central focus of this thesis will be the development of tools to better identify this electronic noise by repeatedly averaging 8-second bins of data from a specific channel, thereby limiting nonperiodic noise and isolating periodic noise more likely to come from electronics.

The process will proceed as follows (omitting some details): for every data channel, the data is chopped into 8-second intervals called "bins." The bins over each day are then averaged with all other bins from the same day to create a "summary file" for a specific day. This process is called "folding" and is carried out by a program called Folding Analysis. These summary files for specific days can again be treated as bins, and averaged together again to create summary files for individual months, and then over months to create summary files from a whole observing run. This process, along with the creation of graphs for the folded data, is carried out by a program called "Foldfold." Finally, we may create a function which is essentially a dot product between two sets of bins (channels). This can then be used to correlate two different data channels over the course of an observing run. The correlation application is called "Yearcompare." We will detail the implementation of each application, in addition to summarizing their results.

This work continues off that of previous students. While this thesis includes the production of data from Folding Analysis and Foldfold, Folding Analysis and Foldfold were built and employed by previous students. However, Yearcompare is a new tool built as a part of this thesis using the data produced with the infrastructure constructed by previous students.

# 2 FOLDING ANALYSIS

All data from the detector–strain, magnetometer, voltage monitor, or otherwise–comes in the same form of counts vs time. These are accessed through LIGO frame files. All data is acquired from and all applications execute on the LIGO data grid. First, Folding Analysis accesses the frame files stored in directories `/frames/O3/raw/L1/L-L1_R-*` and `/frames/O3/raw/H1/H-H1_R-*`. These frame files contain both the strain data along with magnetometer and other auxiliary channels. These raw data streams are then queried using the *frgetvect* interface to get the count vs time data for a specific day and ensure that only data taken during observation mode (so called "science segments") is used.

Once we have the time series data for a specific channel on a specific day, Folding Analysis then begins the folding process. First, the raw time series data arrives from *frgetvect* in science segments of varying length in time called $a_i$. One may think of these $a_i$ as simply $2 \times n$ arrays which have the counts and time listed out for $n$ values, where $n = LS$ where $S$ is the sample rate in samples

per second of the channel analyzed and $L$ is the time length of the science segment. Because science segments are not continuous over a whole day we need to ensure that we cut the $a_i$ into 8-second bins called $b_i$ so that periodic noise remains after averaging. To do this, we let the first data point in $a_1$ have time $t_0$ in seconds. We then require that the first data point of every bin $b_i$ to have time $t_0 + 8k$ and final time $t_0 + 8(k+1)$ for $k \in \mathbb{Z}$, rejecting all other data. This ensures that noise with integer-second periodicity remains after repeated averaging. We now have cut the science segments for a day $a_i$ into 8-second bins $b_i$, which are now thought of as $2 \times m$ arrays where $m = 8S$. The first of these entries is the counts–or data. The second of these is now a time modulo 8 seconds beginning at 0 and ending at 8 seconds across the $m$ entries.

Since science segments often do not span entire days, we may have jump-discontinuities across $b_i$ and $b_{i+1}$. These jump discontinuities will contaminate the spectrum of each $b_i$ once Fourier transformed. Therefore, we must apply a process called windowing to the set of $b_i$ to limit jump discontinuities. This process begins with querying each $b_i$. If a $b_i$ does not contain a boundary of a science segment, do nothing. Should a $b_i$ be a bin on the boundary of a science segment–meaning that its first or last entry is the beginning or end of a science segment–we apply a quadratic envelope. This envelope multiplies the boundary point by zero and the data point .5 seconds inward from the boundary by one. This quadratic window suppresses spectral contamination while also allowing all consecutive $b_i$ to be continuous across boundaries, in addition to the first derivatives of the consecutive $b_i$ to be continuous across science segment boundaries. Let such windowed $b_i$ be called $c_i$.

The last step before averaging is the rejection of any major glitches throughout a day as these are not relevant to the background noise we wish to isolate, but can be loud enough to distort the daily average for affected bins. Let $\mathcal{D}' = \{c_i\}$ for a specific day. We calculate the mean counts $\overline{x}$ across all $c_i \in \mathcal{D}'$ and the standard deviation $\sigma$ of counts for the $c_i \in \mathcal{D}'$. If a $c_i$ contains a data point $x$ with value $x > \overline{x} + 5\sigma$ or $x < \overline{x} - 5\sigma$, we remove that $c_i$ from $\mathcal{D}'$, thereby excluding it from future analysis. Let $\mathcal{D} = \{d_i \in \mathcal{D}' : d_i \text{ does not contain a } 5\sigma \text{ outlier}\}$ be the set of $c_i$ remaining after this rejection process.

We are finally ready for the averaging process. We take all the processed bins $d_i \in \mathcal{D}$ and average them together. Let $k$ is the number of $d_i \in \mathcal{D}$. The averaging algorithm is then:

$$f' = \frac{1}{k} \sum_{i=1}^{k} d_i \tag{1}$$

We have found that it is helpful to transform the time series data $f'$ to frequency space, and filter out all Fourier coefficients outside the 10 to 50 Hz band and then transform back into counts vs time space. Frequencies below 10Hz tend to dominate the strain channel, thus removing them greatly reduces the dynamic range of our data. As we go above 50Hz we run into less and less contributions from internal electronics. With this process completed, we have created a final summary file for a day called $f$. Note that $f$ also contains the prefiltered data for comparison. Below is the full process for Folding Analysis:

$$\begin{aligned} \text{Frame Files} \xrightarrow{frgetvect} \{a_i\} \xrightarrow{\text{align and cut}} \{b_i\} \xrightarrow{\text{window}} \{c_i\} \\ \{c_i\} \xrightarrow{\text{reject outliers}} \{d_i\} \xrightarrow{\text{averaging}} f' \xrightarrow{\text{bandpass filter}} f \end{aligned} \tag{2}$$

It is important to note that this is the process for one day. We run Folding Analysis over the entire O3 observation to get summary files for every day in the O3 run. We employ the Condor infrastructure to run each channel in parallel. Condor is a cluster management software which allows

users to queue jobs on multiple computing arrays so that several applications may run in parallel. We first compile Folding Analysis and Foldfold into executables. Then, we create a properly-formatted job list using a python script. This job list contains all of the jobs set to run in parallel. For Folding Analysis, only one month and one channel runs per instance. Therefore, our job list needs to contain execution commands for each channel and each month. After proper formatting, we submit our job list in `.dag` format to the computing cluster using the Condor infrastructure. Condor dramatically reduces the time required to produce data, as executing every single Folding Analysis command as a user for every channel and every month takes vasts amounts of time compared to creation and submission of the above job lists.

# 3 FOLDFOLD

Once we complete Folding Analysis, we now have summary files in `.mat` format for every day in the O3 run. We need to create monthly summary files, then O3 summary files for each channel, and then create graphs to display the data. This is the job of the next application, Foldfold. Foldfold is much simpler than Folding Analysis. Foldfold runs once for a given channel. The user designates the directory containing all `.mat` summary files (called $f_i$) for each day in an observation run, and Foldfold produces graphs for the spectra and final folded data for each day–both pre and post bandpass filter. Next, Foldfold averages each $f_i$ over a specific month to create a summary file for every month. Finally, Foldfold averages every summary file for each month into a grand summary file over an entire observation run for a specific channel, and the user is left with the `.mat` summary files for each month and a grand summary file for the whole observation period in addition to graphs displaying the results.

# 4 RESULTS FROM FOLDING ANALYSIS AND FOLDFOLD

LIGO data channels are named using a system of acronyms. Below is a glossary of such terms:

| Data Channel Naming | | | | | |
|---|---|---|---|---|---|
| 1st Term | Definition | 2nd Term | Definition | 3d/4th/5th Terms | Definition |
| CAL-DELTAL_EXTERNAL | Calibrated Strain | CS | Corner Station Near Mirrors | ADC_0_12 | Analog to Digital Converter Voltage Monitor |
| PEM | Physical Environment | ETMY | End Test Mass - Y Arm | DQ | Data Quality |
| SUS | Position Senors on Supended Mirror | EX | End X Arm | EBAY | Electronics Bay |
| | | EY | End Y Arm | LVEA | Laser and Vaccum Equipment Area |
| | | | | MAG | Magnetometer Channel |
| | | | | SUSRACK | Near Suspension Rack |
| | | | | VERTEX | Near Main Beam Splitter Chamber |
| | | | | X,Y,Z | Axes of a Detector |

We run Folding Analysis and Foldfold for the following LIGO data and auxiliary channels over the entire O3 run for both L1 and H1:

```
CAL-DELTAL_EXTERNAL_DQ
PEM-EY_MAG_EBAY_SUSRACK_Z_DQ
PEM-EY_MAG_EBAY_SUSRACK_Y_DQ
PEM-CS_MAG_LVEA_VERTEX_Y_DQ
PEM-CS_MAG_LVEA_VERTEX_Y_DQ
PEM-CS_MAG_EBAY_SUSRACK_Z_DQ
PEM-CS_MAG_LVEA_VERTEX_X_DQ
SUS-ETMY_L3_MASTER_OUT_LL_DQ
PEM-EX_MAG_EBAY_SUSRACK_Y_DQ
PEM-EX_MAG_EBAY_SUSRACK_X_DQ
PEM-CS_MAG_EBAY_SUSRACK_X_DQ
PEM-CS_MAG_LVEA_VERTEX_Z_DQ
PEM-EX_MAG_EBAY_SUSRACK_Z_DQ
PEM-CS_MAG_EBAY_SUSRACK_Y_DQ
PEM-EY_MAG_EBAY_SUSRACK_X_DQ
PEM-EX_ADC_0_12_OUT_DQ
PEM-EY_ADC_0_12_OUT_DQ
PEM-EY_MAG_EBAY_SUSRACK_Z_DQ
PEM-EY_MAG_EBAY_SUSRACK_Y_DQ
PEM-CS_MAG_LVEA_VERTEX_Y_DQ
```

The results of Folding Analysis alone is not the focus of this thesis, however, it is useful to inspect three example outputs from the Folding Analysis–Foldfold process:

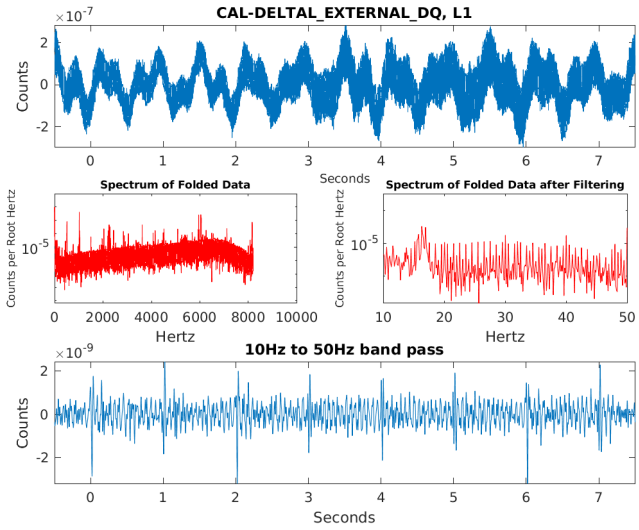


Figure 1: The results of Folding Analysis and Foldfold are plotted for the L1 strain channel over the entire O3 run.

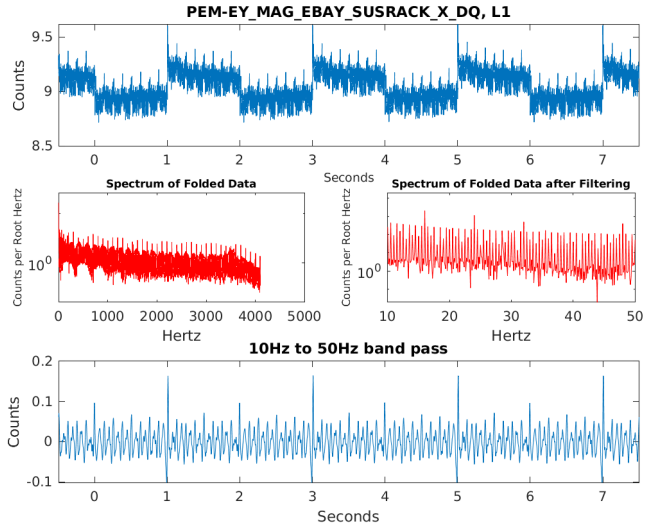


Figure 2: The results of Folding Analysis and Foldfold are plotted for an L1 magnetometer channel monitoring an electronics bay at the end of the Y arm over the entire O3 run. Note that are periodic artifacts with periods 1 and 2 seconds.

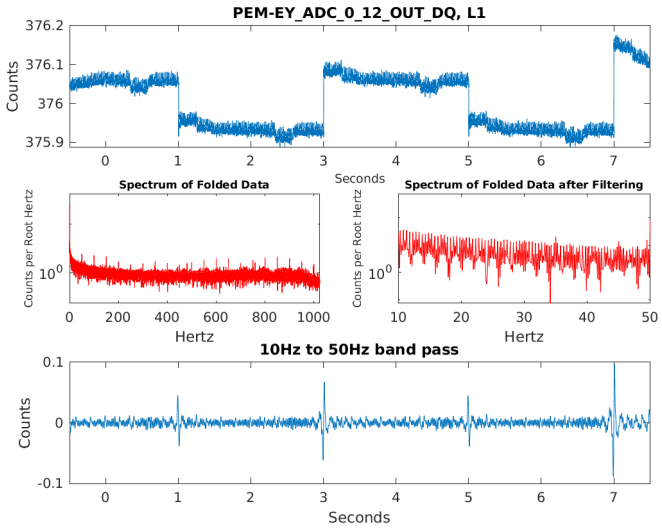


Figure 3: The results of Folding Analysis and Foldfold are plotted for an L1 voltage monitor at the end of the Y Arm. Note that there is a periodic artifact with period 2 seconds.

Of relevance to this thesis is the correlation between these three data channels in the bandpass graph. We can clearly see that strain channel features small periodic noise at sub-integer intervals with large peaks at integer second intervals. This is similar to the magnetometer channel seen in

Figure 2. Figure 3 has an artifact with period 2 seconds, but less prevalent is regular noise with period equal to or shorter than one second like those seen in Figures 1 and 2. While this qualitative observation may be helpful in identifying the causes of potential artifacts in data, Folding Analysis and Foldfold give us a much more powerful framework that we can use to correlate noise between data channels. All folded data is accessible with the following link:
https://ldas-jobs.ligo.caltech.edu/~sherman.thompson/DotProducts/clickme/

# 5 YEARCOMPARE

We may treat the filtered data channel of each 8-second folded summary file as function $f : [0, 8] \to \mathbb{R}$. One may think of the space of these summary files as a "Hilbert Space" of sorts over the interval $[0, 8]$ seconds with the following "dot product:" If $f$ and $g$ are in this Hilbert Space and have the same sample rate, then the inner product $\langle f, g \rangle$ is:

$$\langle f, g \rangle = \frac{1}{\sigma_f \sigma_g \sqrt{S}} \sum_{t \in D} (f(t) - \overline{f})(g(t) - \overline{g}) \tag{3}$$

where:

$\sigma_f$ - The standard deviation of $f$ over the domain.
$\sigma_g$ - The standard deviation of $g$ over the domain.
$S$ - The sample rate of $f$ and $g$ in samples/second.
$D$ - The domain. This is technically discrete and consists of times ranging from 0 to 8 seconds.
$\overline{f}$ - The mean value of $f$ over the domain.
$\overline{g}$ - The mean value of $g$ over the domain.

If $f$ and $g$ do not have the same sample rate, then the finer sampled data set is downsampled by averaging consecutive entries. Since LIGO data has sample rates in powers of 2, we can always apply this trick to ensure that $f$ and $g$ have the same sample rate. This dot product is coded into an application called Matcompare, which calculates the dot product between two `.mat` summary files' bandpass-filtered data. This dot product can be thought of as assigning a correlation factor between two data channels. To see what values of the dot product are significant, we first take $f$ and $g$ to be random arrays generated from a Gaussian distribution with sample rate $16,384$ samples/second– typical of LIGO data. For this sanity check, we tell the random number generator to have $\sigma_f = 10$ and $\sigma_g = 5$ with $\overline{f} = \overline{g} = 0$, means we expect from our data. We repeat this process $10,000$ times to produce the following histogram.

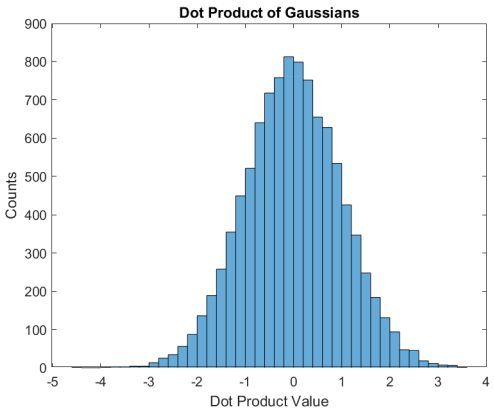


Figure 4: This shows the histogram of many dot products between Gaussian random number arrays. Note that this distribution is consistent with a Gaussian of zero mean and unit variance. Since we normalized our dot product by the standard deviation and sample rate, we note that changing the values of $\sigma_f$, $\sigma_g$, $\overline{f}$, $\overline{g}$, or $S$ in this example does not affect the above distribution.

The purpose of this sanity check is to ensure that the normalization factor $\frac{1}{\sigma_f \sigma_g \sqrt{S}}$ works such that, no matter the relative means, sample sizes, or standard deviations, two Gaussian random number functions $f$ and $g$ will always have a dot product with zero mean and unit variance. These all appear to be the case, as repeated trials with different values of these produced practically-identical histograms. Therefore, we expect that any dot product with absolute value larger than 4 ought to be classified as a significant correlation as data outside $4\sigma$ has a 1 in $15,787$ change of occurrence.

With Matcompare now able to produce reliable correlation factors between individual `.mat` files, we can now produce correlation data from the grand summary files of channels. Even better than this, we can take the grand summary file of the strain channel produced from Foldfold, and plot the dot product with a specific channel over time. This is the function of Yearcompare. Yearcompare takes the grand summary file of one channel and dots it with the daily summary files of another channel to produce a correlation vs time graph that can be used to pinpoint not only which channels have correlated periodic noise, but also on which dates the correlation began, peaked, or subsided.

# 6 RESULTS FROM YEARCOMPARE

For both L1 and H1, we have Yearcompare dot the grand summary file of the strain channel with all of the folded data produced before. While there are several interesting results, this thesis will highlight two use-cases for how Yearcompare might be effectively deployed. After investigating the dot products of multiple channels, the following Yearcompare result gives a very clear trend:

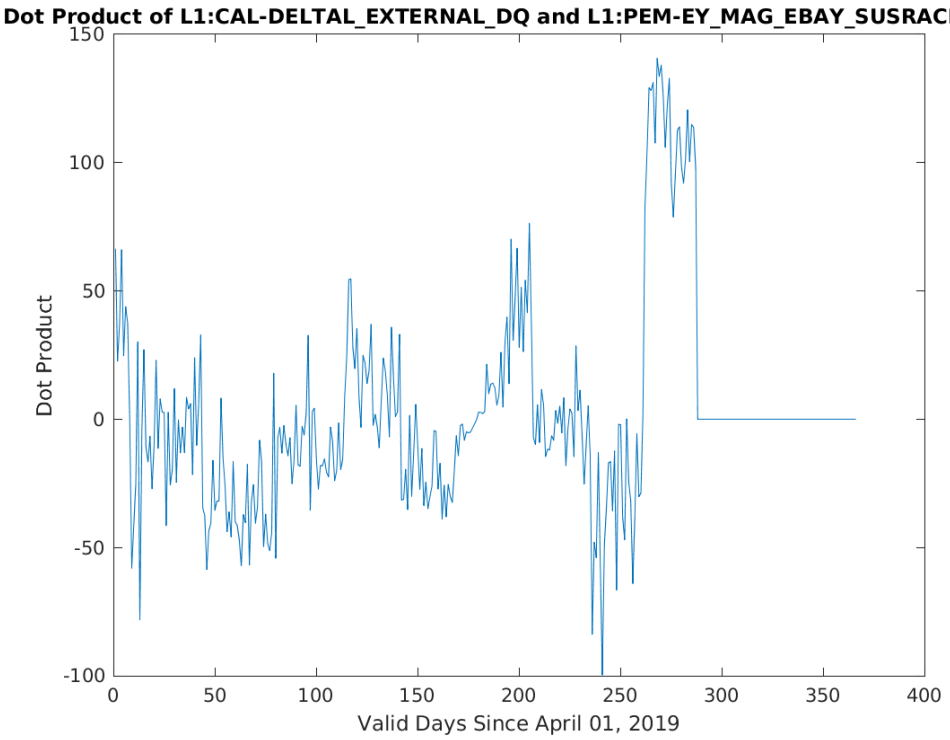


Figure 5: The dot product between the L1 strain channel and a magnetometer channel at the end of the Y arm plotted as valid days since the beginning of O3. Valid day is used as there may be some gaps, as not every day of the O3 timeframe was used for observation. Recall that any value with absolute value larger than 4 is a significant correlation.

We can see that there is a significant jump in the above dot product at around 275 valid days since April 01. Even though there are significant values beforehand, earlier values of the dot product vary much more smoothly compared to the significant jump seen around day 275. While this data does not contain explicit dates for space reasons, a directory containing `.fig` files for every dot product featuring precise date information is available at the following link:
https://ldas-jobs.ligo.caltech.edu/~sherman.thompson/Yearcompare/

After accessing the above directory and downloading and opening the appropriate `.fig` file, the sudden spike can be identified to have began on 18-02-2020 and ended 22-02-2020. After identifying the dates of the anomaly, the next step is to identify potential changes in the logbook which may explain the data. If one searches through the appropriate dates of the logbook available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?content=1, one finds the following entry:

michael.laxen@LIGO.ORG - 11:38, Tuesday 18 February 2020 (51702)
XEND Complete, PEM voltage monitor is now monitoring the ESD +24VDC/MV (moved from Beckhoff 24V power strip)
YEND Complete, PEM voltage monitor is now monitoring the ESD +24VDC/MV (moved from Beckhoff 24V power strip)

This is a strong culprit for the cause of our sudden discontinuity. ESD is the Electrostatic drive, a set of electrodes that inductively control the mirrors across a thin vacuum gap. These are meant to keep the mirror in in place, and any slight variation in their voltage directly affects the position of the mirror, thus directly affecting the signal. Therefore, any adjustment to the ESD almost directly affects the primary signal of the detector. Therefore, with any adjustments to the ESD circuit as seen above, we expect a sudden change in correlation between ambient magnetic fields and the main signal.

These sudden glitches in dot product are not unique to the above example. Here is another example:

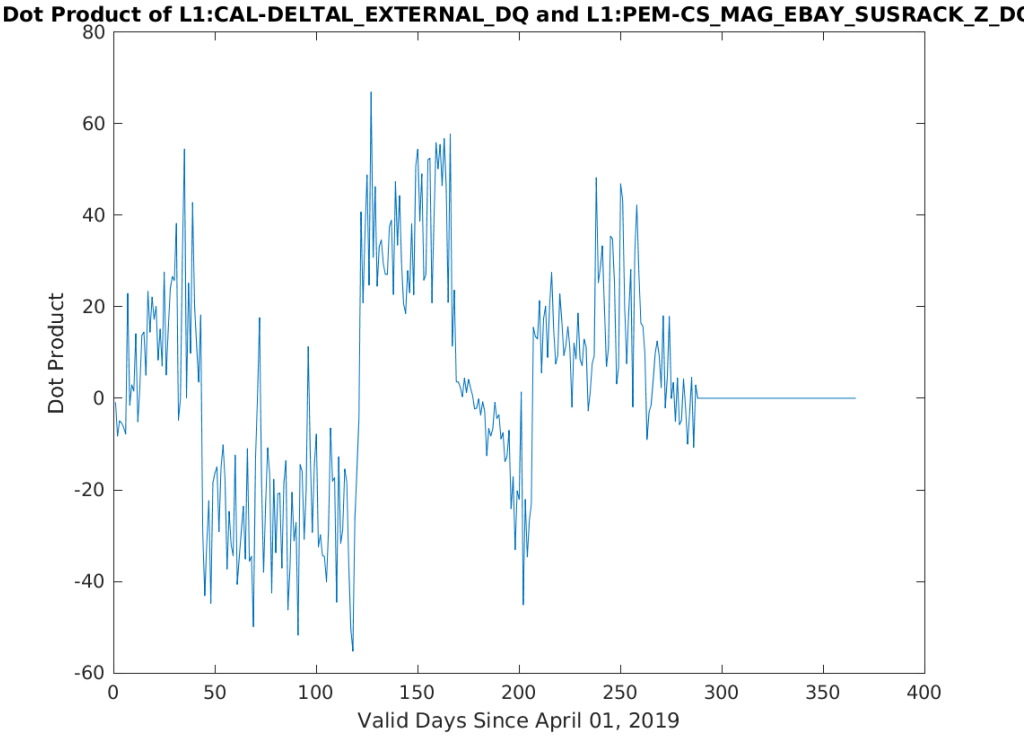


Figure 6: The dot product between the L1 strain channel and a magnetometer channel at the Corner Station plotted as valid days since the beginning of O3.

When we access the appropriate `.fig` file, we note that there is a sudden downtrend in dot product from 18-05-2019 to 18-05-2019 and a sudden uptrend from 10-08-2019 to 12-08-2019. Looking through the logbook, the sudden downtrend coincides with the following entry:

stuart.aston@LIGO.ORG - 00:09, Sunday 19 May 2019 (46057)CDS, SUSLink [Michael L, Stuart A] After arriving on-site we checked in the DC power room and confirmed all SUS PSU's were not tripped, next we inspected the l1sush45 Coil Drivers in the CER and all was operational. However, we discovered that the AA chassis +15V rail had failed in the SUS C1 Rack, so it was replaced with a spare chassis, which has resolved the fault.

All SUS/ISI platforms were recovered to there nominal states ALIGNED/ISOLATED by 0:00 [CST] 05:00 [UTC], alignments were restarted for the suspension models that had been restarted, and the IMC re-locked without issues.
Given the earthquake motion has now subsided the Operator is running an alignment.

This may be a culprit for the sudden change in dot product as there was a significant adjustment in power delivery. The detector is extremely sensitive to adjustments in power circuits, so this adjustment likely affected the correlation between the ambient magnetic field in the electronics bay which may result from wire loops and the strain signal.

# CONCLUSION

These examples illustrate how Yearcompare and Matcompare could be used to identify (and eventually allow mitigation of) noise sources in LIGO data channels. While this thesis was not focused on the results of Folding Analysis and Foldfold since the primary development of those tools is detailed elsewhere, those programs even without the extra new dot-product infrastructure detailed here provide unique observations on the causes of noise in LIGO detectors.

The extra functionality that Yearcompare and Matcompare provide enable users to pinpoint the causes of electronic noise in detectors, not only weighing different locations as causing more or less noise than other locations, but also leveraging the correlation vs time data to highlight specific, seemingly innocuous electronics changes as the root cause of noise. Even more than this, should Yearcompare or Matcompare identify a particular time or set of circumstances where a usually-noisy channel becomes uncharacteristically quiet, a user can search the logbook and discover a potentially-new method to reduce noise in the main detector.

# References


[1] J. Aasi et al. *Advanced LIGO*, Classical and Quantum Gravity 32 (7): 074001. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.161101

[2] B. P. Abbott et al. *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Phys. Rev. Lett. 119, 161101. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.161101

[3] B. P. Abbott et al. *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. 116, 061102. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.116.061102

[4] B. P. Abbott et al. *Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA*, Living Reviews in Relativity (2020) 23:3. https://dcc.ligo.org/DocDB/0152/T1800208/001/description-folding-analysis.pdf

[5] R. Abbot et al. *Constraints on the Cosmic Expansion History from GWTC-3.* https://arxiv.org/abs/2111.03604

[6] P.B. Covas et al. *Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of Advanced LIGO*, Phys. Rev. D 97, 082002. https://dcc.ligo.org/DocDB/0152/T1800208/001/description-folding-analysis.pdf

[7] P. Nguyen et al. *Environmental Noise in Advanced LIGO Detectors*, Classical and Quantum Gravity 38 (14): 145001. https://doi.org/10.1088/1361-6382/ac011a

[8] K. Rao, W. Liu, and K. Riles. *Description of a Folding Analysis for Detector Characterization.* https://dcc.ligo.org/DocDB/0152/T1800208/001/description-folding-analysis.pdf

[9] D. Thain, T. Tannenbaum, and M. Livny. *Distributed Computing in Practice: The Condor Experience.* https://research.cs.wisc.edu/htcondor/doc/condor-practice.pdf