Observing the Correlation Between Scattering and Anthropogenic Noise in LIGO Detectors

Nii-Boi Quartey

Faculty Adviser: Dr. Marissa Walker

April 28, 2021

Abstract

The Laser Interferometer Gravitational-wave Observatory (LIGO) is a large scale observatory capable of detecting gravitational waves. However, transient, systematic noise events are present in the detector's data. We call these events "glitches," and they can produce false gravitational-wave triggers in our data analysis pipelines. To best identify true gravitational-wave events, these glitches must be well understood and removed. After using a custom pipeline called Omicron to find glitches in the data, the Gravity Spy project helps LIGO scientists better understand these glitches by using machine learning to classify them by their shape. By calculating the rate at which glitches occur for each category, we can see how different types of glitches increase or decrease in frequency. I will present plans for a study exploring how we can use the average rate of glitches over time to characterize different types of noise in the LIGO detectors.

Introduction

LIGO is composed of two laser interferometers capable of detecting gravitational-waves. These gravitational waves are caused by ripples in spacetime caused by colliding black holes and neutron stars. Gravitational-wave detectors were developed in order to measure these waves, but sources of noise known as glitches interfere with the LIGO time series of data. These glitches can be removed by being vetoed, but when they are vetoed the amount of time data that can be analyzed becomes significantly smaller. One of the ways this issue was through the Gravity Spy program, which allows LIGO scientists and machines to classify glitches by their shape. Glitches have been classified by their characteristics such as their shape, duration, and signal-to-noise ratio (SNR).

One of these classified glitches was scattered light, which occurs when a small amount of light strays out of the main detector beam. As a result, noise is introduced into the gravitational-wave data [Figure 1]. Scattered light typically has a glitch rate between 10 to 120 Hertz. The scattered light glitch also contributes to anthropogenic noise, which is defined as high intensity seismic vibrations that can create high levels of noise and interfere with the detectors.

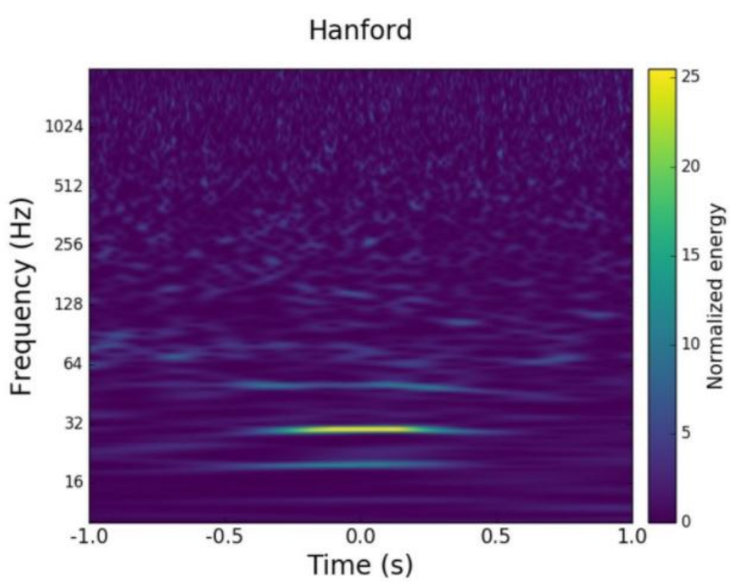

Figure 1: Scattered Light Glitch Detected by Gravity Spy

The goal of this project is to verify that there is a correlation between scattering and anthropogenic noise by analyzing one day when there was a large amount of scattered light glitches detected in the Livingston observatory. The day that will be observed is December 9, 2019 [Figure 2].

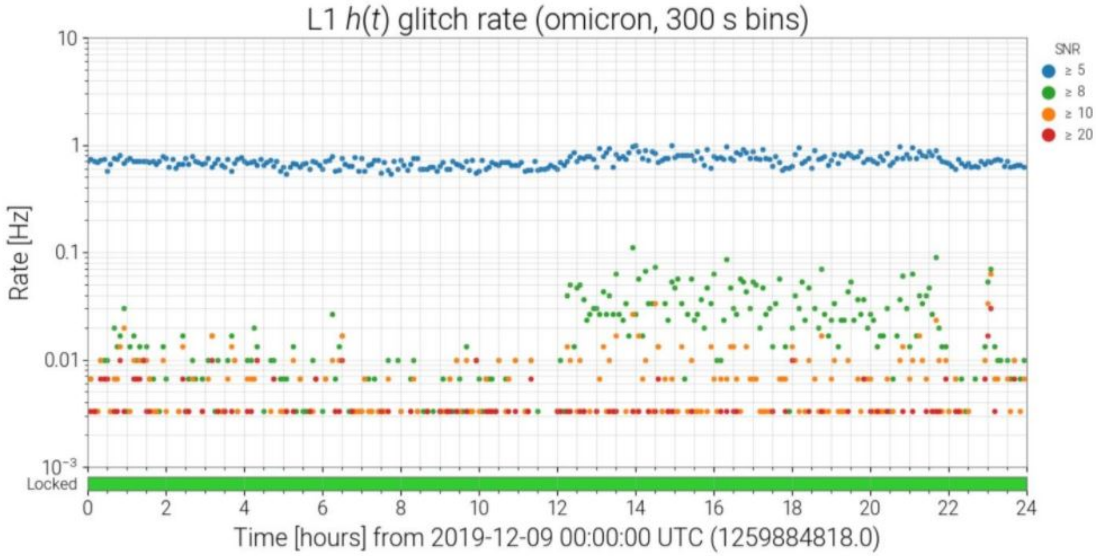

Figure 2: Glitch Rate on December 9, 2019

The figure above shows the entire glitch rate of the Livingston observatory on December 9, 2019. The y-axis indicates the number of glitches per second and the x-axis is the time in hours. The color of the glitches indicate their signal-to-noise ratio (SNR), which tells us how loud the glitches are. Glitches in red are very loud glitches whereas glitches in blue are glitches are not as loud.

Theory

Reports from the Advanced LIGO Livingston Observatory Logbook on December 9 state that scattering is correlated with anthropogenic noise that occurs within distinct periods of time. To explore the relationship between scattering and seismic noise, the simple moving average [Equation 1] of the scattered light glitch rate will be compared with the seismic channel being used.

$$SMA_i = \frac{\sum_{k=i-n}^{i} x_k}{n}$$

Equation 1: Simple Moving Average

This formula will be used by taking the scattered light glitch rate of the first 30 minutes and average those values together. The value returned from the simple moving average will then be appended to an array. The data will move over by one minute, and the next 30 minutes of the simple moving average will be calculated and appended to the array. This process will be repeated until the moving average has all the data from December 9.

Once the scattered light glitch rate is in the form of a moving average, it can be compared with a seismic channel. The data will be converted into a scatter plot and the Pearson correlation coefficient [Equation 2] will be used to indicate the linear relationship between the two variables.

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2 \sum(y-\overline{y})^2}}$$

Equation 2: Linear Regression R-value

The correlation coefficient takes the difference of the x variable subtracted by its mean and multiplies it by the same difference for the y variable. The sum of this product is then divided over the square root of the sum of squared deviations for both variables. The value that is returned from this calculation is the R-value, which measures how close the points in the scatter plot are to the linear regression line plotted with it. This value gives us an idea of how strongly correlated the scattering and anthropogenic noise are.

Methods

Retrieving the Scattered Light Glitch Data

Constructed inside the Caltech JupyterHub, the first steps required the retrieval of all the scattered light glitches on December 9. LIGO specific packages can be used inside of the JupyterHub, so one of these packages was used to retrieve the required data. Known as the Gravitational Wave Python (GWpy) package, the command used allowed me to specify where I wanted to retrieve the data from. In this project, I am looking for glitches in Gravity Spy, so that is what I specified in my code. Once I had specified that I was looking for Gravity Spy glitches, I was able to specify the data even further by putting in my code that I was looking for scattered light glitches. With this information specified in the command parameters, I was able to retrieve just the scattered light glitch rate in the form of a step plot. In the plot parameters, I only had to specify the the stride of the step plot, which was set to 60. This means that the plot will return the glitch rate every minute for the entirety of December 9. The only other parameters involved setting the range of the step plot to the start and end time of December 9. The plot containing scattered light glitch rate is shown below [Figure 3].

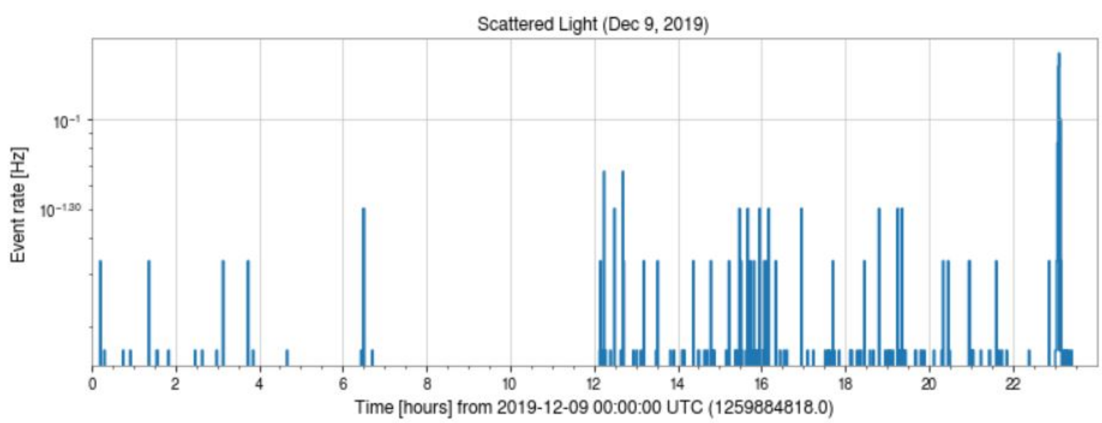

Figure 3: Scattered Light Glitch Rate

As the discrete data from this plot makes it difficult to determine the overall trends, there is not much that can be done with it in this form. However, the data can be converted to a more usable form when changed to a simple moving average.

Simple Moving Average

The simple moving average was created using a while loop so that the data retrieved would not exceed December 9, 2019. Inside the loop, the moving average was calculated by summing the glitch rate for the first 30 minutes and then averaging that sum. The while loop allowed the data to move every minute, so the data was averaged for the entire day. Compared to the scattered light glitch rate in the form of a step plot [Figure 4], the simple moving average proves to be a smoother way to analyze the data.

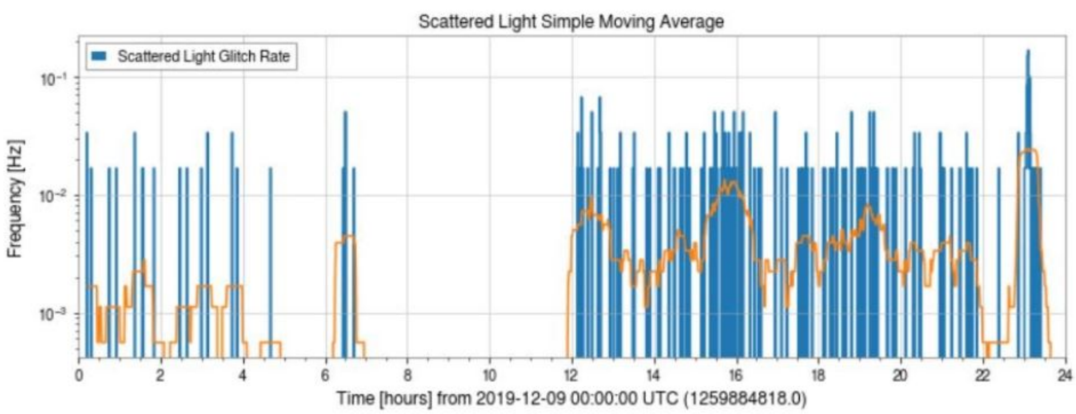

Figure 4: Scattered Light Event Rate Compared with Simple Moving Average

Retrieval of Seismic Channels

This anthropogenic noise channel was retrieved using GWpy's "get_data" command. Similar to the fetch command, get_data retrieves the data from the channel of choice and and range of time that the data will be retrieved. Three seismic channels were plotted together on one plot (ETMX in blue, ETMY in green, ITMY in red) and each channel would later be compared with the scattered light glitch rate in the form of a scatter plot to observe the correlation between the glitch rate and seismic noise. Each channel was plotted so that it would fit the time range of December 9, 2019 [Figure 5].

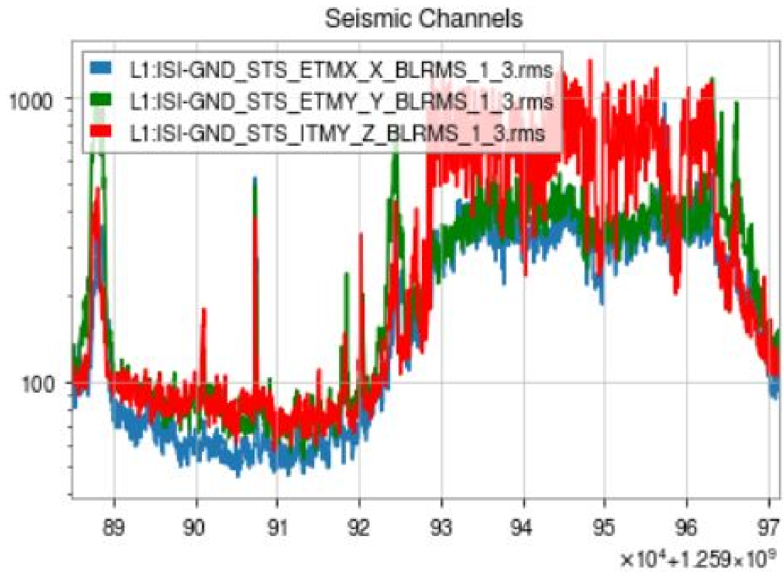

Figure 5: Seismic Channels in Raw Form

Pearson Correlation Coefficient

Before the scatter plot observing the correlation between each channel and the scattered light glitch rate can be made, all the data must be normalized. If the data is not normalized, the correlation between the glitch rate and seismic channel becomes significantly smaller (for reference, refer to Figure 10). This is because the scattered light glitch rate is measured in Hertz and all the seismic channels are measured in nanometers per second.

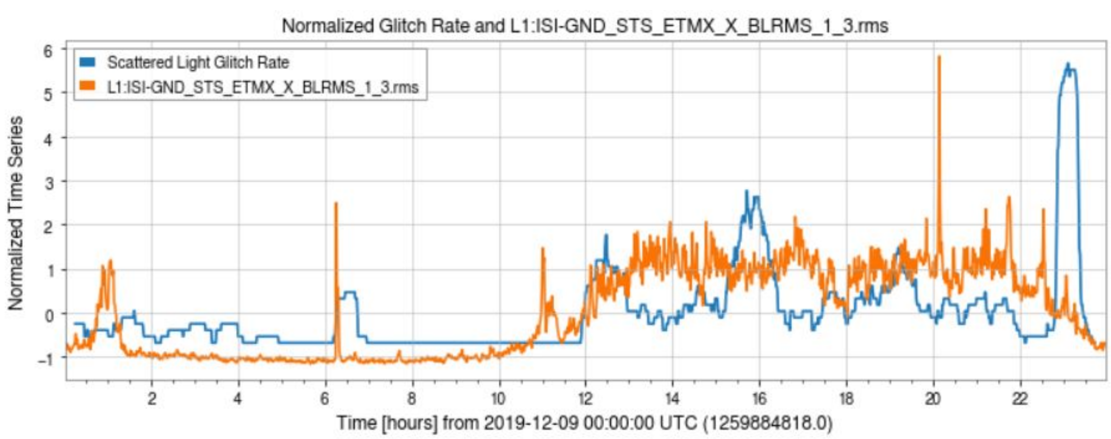

Figure 6: Normalized Glitch Rate and Seismic Channel (ETMX)

With normalized data, the scatter plot showing the correlation between the glitch rate and each seismic channel can be plotted. The scatter plot shows the glitch rate on the y-axis and the seismic channel being used on the x-axis. Once again, each seismic channel is differentiated by color. The linear regression showing the correlation between the two variables is shown using sklearn's LinearRegression command and the numerical value of the regression (otherwise known as the R-value) is calculated using numpy's corrcoef command. Each comparison between the glitch rate and respective seismic channel is shown below [Figure 7].

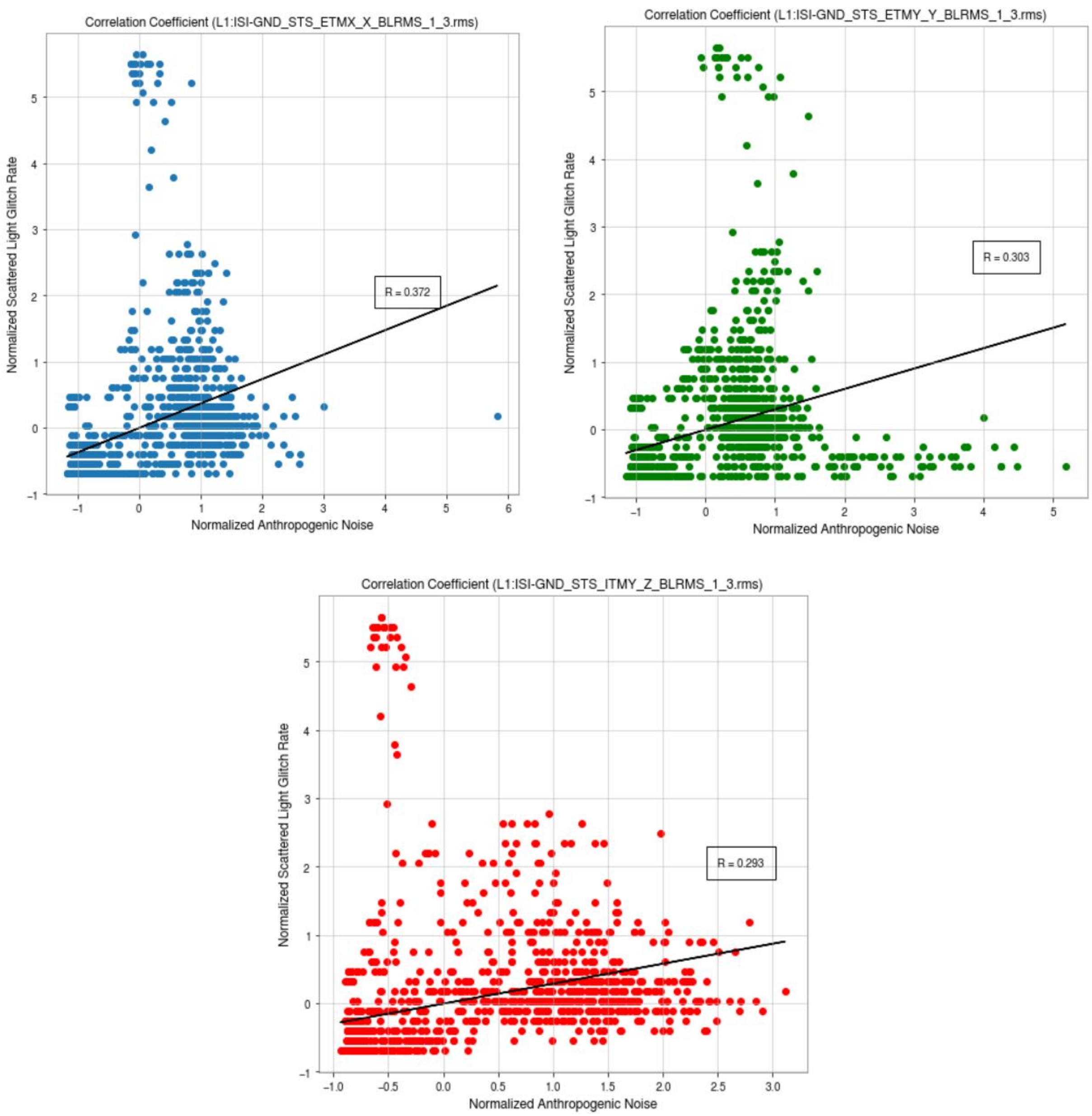

Figure 7: Scatter Plot (Glitch Rate vs ETMX in blue, vs ETMY in green, vs ITMY in red)

The R-values of each correlation is as follows: 0.372 for the ETMX channel, 0.302 for ETMY, and 0.293 for ITMY. The ITMY (red) channel is the channel that measures seismic motion in the part of the interferometer that was mentioned in the logbook report, so this is the channel that will be used moving forward.

One thing to note in each of the channels is that there are several points at the beginning that are nowhere near the linear regression. It is believed that these points were the the result of an earthquake that occurred towards the end of the day (refer to Figure 3, specifically around 22 hours). These points potentially interfere with the R-value, as the value could potentially be lower in each instance. Because of this, the scattered light glitch rate is increased so that it would average the glitches every five minutes rather than every minute [Figure 8].

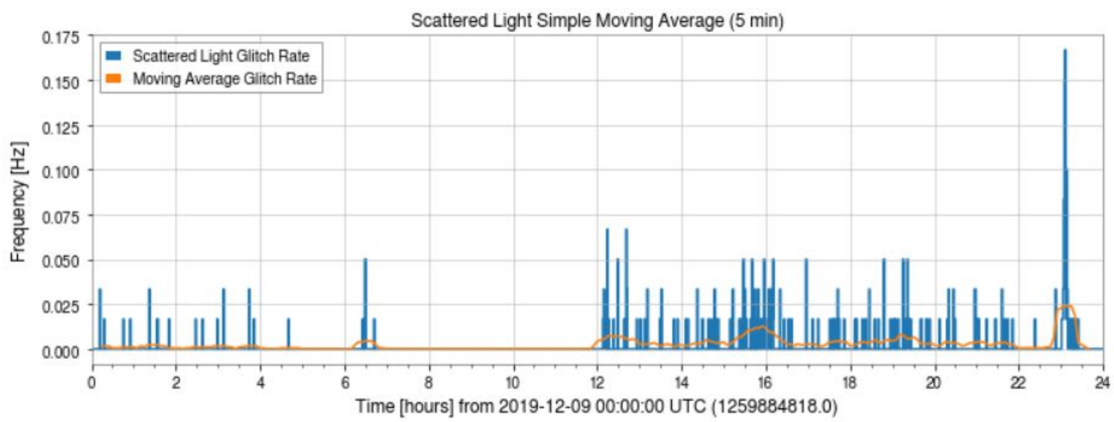

Figure 8: Scattered Light Glitch Rate (5 min)

Once normalized, this new glitch rate reduces the total number of points that appear on the scatter plot. With less points appearing on the scatter plot [Figure 9], it can be seen that there are much less points that appear on the top that interfere with the data. There was also a substantial increase in the R-value.

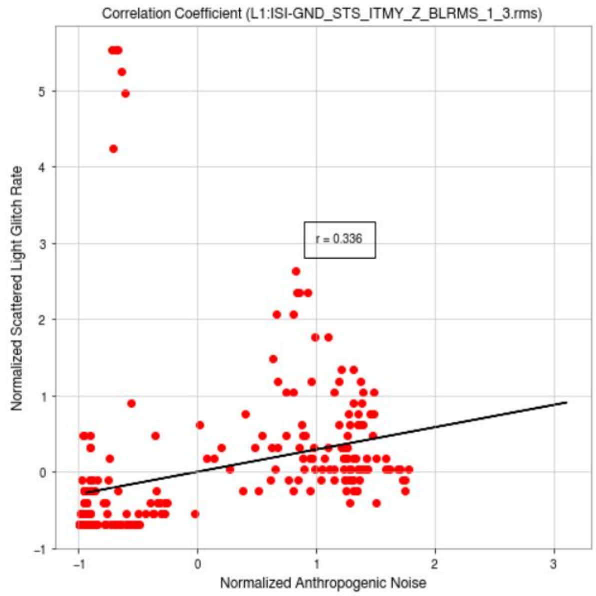

Figure 9: Scatter Plot (Glitch Rate vs ITMY, 5 min)

It was mentioned earlier that the unnormalized correlation has a significantly smaller R-value than the normalized correlation. However, the unnormalized data also takes the raw data between the glitch rate and seismic channel. This means that it plots the scattered light glitch rate without its moving average with the seismic channel. The result of this plot is shown below [Figure 10].

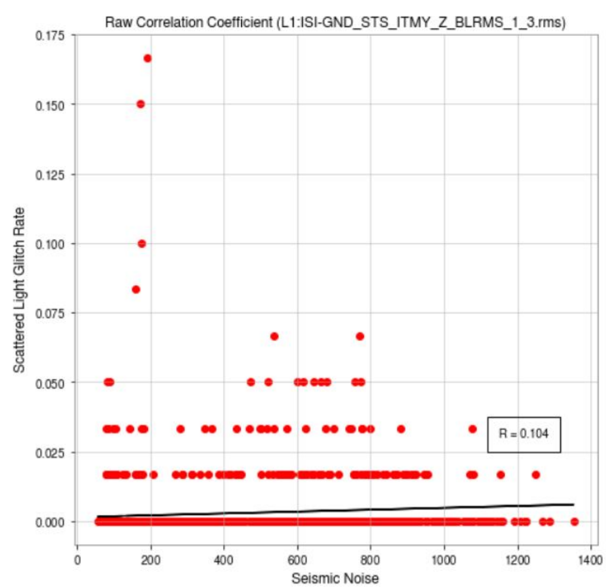

Figure 1: Scatter Plot (Raw, Unnormalized Glitch Rate vs ITMY)

This plot essentially shows why the moving average of the scattered light glitch rate must be used over the raw data. Without the smoothness of the moving average, the scatter plot takes a very odd shape and the R-value becomes much smaller than it should be.

Final Results

Using the five minute trend of data, the final results show that the R-value between scattering and anthropogenic noise is 0.336. This R-value does show that a correlation exists between the two components, but this value is not as high as we expected. This is most likely because there are more types of seismic noise affecting the data. Anthropogenic noise is just one type of seismic noise. The earthquake that occurred late in the scattered light glitch rate plot was one of the many other factors that affected the scattering. If this project looked at all the other factors that affect scattering, it is likely that the R-value would be greater than the resulting value.

Conclusion and Future Steps

Though the correlation between scattering and the glitch rate is complete, there is more work that can be done beyond the scope of this project. For instance, this project only observes the correlation between scattering and anthropogenic noise on December 9, 2019. There are other days besides this one where the scattering was high that can be observed. It was also mentioned that this project looked only compared anthropogenic noise to the scattering. Other types of seismic noise can be compared to the scattering to see how the correlation between the two variables change. This project also observed the scattered light glitch rate for one day. However, the glitch rate can be looked at for longer periods of time such as two days, one week, or one month.

References

[1] Torres-Forné, Alejandro et al. "Application of dictionary learning to denoise LIGO's blip noise transients". *Physical Review D*, vol. 102, no. 2, 8 July 2020

[2] Johnston, FR et al. "Some properties of a simple moving average when applied to forecasting a time series". *Journal of the Operational Research Society*, vol. 50, no. 12, Dec 1999

[3] Abbott, B.P. and others. "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs". *Phys. Rev. X*, vol. 9, no. 3, pp. 31-40, 2019

[4] Nuttall, LK. "Characterizing transient noise in the LIGO detectors". *Philosophical Transactions A*, 10 January 2018

[5] Ottaway, David J et al. "Impact of upconverted scattered light on advanced interferometric gravitational wave detectors". *OSA*, vol. 20, pp. 8329-8336, Apr 2012

[6] Matichard, F. and others. "Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance". *Class. Quant. Grav.*, vol. 32, no. 18, pp. 185003, 2015

[7] Abbott, B.P. and others. "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914". *Class. Quant. Grav.*, vol. 33, no. 13, pp. 134001, 2016

[8] Aasi, J. and others. "Advanced LIGO". *Class. Quant. Grav.*, vol. 32, pp. 074001, 2015