LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note **LIGO-T1900418-v1-** July 2019

Extending the Reach of Gravitational-wave Detectors with Machine Learning Interim Report 1

Morgan Nanez

Distribution of this document:
LIGO Scientific Collaboration

Draft

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW22-295
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

`http://www.ligo.caltech.edu/`

LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note	**LIGO-T1900418-v1-**	July 2019

Extending the Reach of Gravitational-wave Detectors with Machine Learning Interim Report 1

Morgan Nanez

Distribution of this document:

LIGO Scientific Collaboration

Draft

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW22-295
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

`http://www.ligo.caltech.edu/`

Abstract

This report presents the idea of using current machine learning techniques and algorithms to reduce the overall noise floor of the LIGO detectors. There will be a hard emphasis on techniques that analyze time series data, such as utilizing long short-term memory and non-linear regression algorithms. While other sources of noises in the detectors are outlined in the proposal, there will be a focus on using machine learning algorithms to hone in on noise sources coming from the physical attributes of the instrument itself. The goal is to increase the sensitivity of the detectors by subtracting linear and non-linear noise coupling mechanisms.

1 Introduction

In recent years LIGO has made strides in the discovery of gravitational waves from stellar mass black holes and neutron star mergers. However, there are still many more waves viable for detection below the current surface of noise. With the application of machine learning algorithms to the gravitational-wave detector data and auxiliary channels on-site, there is a possibility to reduce the noise in the time-series due to instrumental artifacts. By reducing the current noise floor there will be greater sensitivity in the instrument, leading to a greater rate of detection.

2 Current Noise Sources

Glitches in the data can sometimes mimic astrophysical events, causing confusion in detecting gravitational waves. The causes of glitches can be identified through analyzing the witness channels during the time of the glitch and comparing multiple similar glitches within the same channel. Past glitches can be attributed to phone rings, airplanes, and trucks passing by. These glitches are examples of removable sources that machine learning techniques can be trained to learn and remove from the outputted data.[1] When picking which witness channels to subtraction from the data in order to get rid of certain glitches, they must be determined incapable of subtracting potential gravitational-wave signals. This can be tested by inserting simulated signals into a test data set and then processing them accordingly.

3 Physical Intuition

Rather than solely focusing on the environmental factors as causes for noise and glitches, looking into the instrument itself can lead to other sources for noise. The LIGO detector is very sensitive and must have passive vibration isolation, which is achieved through mechanisms such as a system of pendulums. It must also have active vibration isolation which is achieved through LIGO's active damping systems, consisting of various suspensions. By subtracting noise from the instrumental noise, LIGO detectors will have an increased sensitivity

3.1 Beam size and Angle

An instrumental cause of small fluctuations in the data is the jittering of the pre-stabilized laser. [2] The jittering was introduced when upgrades were added to the subsystem. A high powered oscillator was added to the system in an effort to increase laser power. However, the high powered oscillator required continuous heat dissipation via water cooling. Vibrations originated from the water flow, introducing jitter into the beam angle and size, resulting in noise. Other efforts, such as adding sensors that measured radial beam distortions and thermal compensation systems were used to mitigate the noise from the jittering. However, the jitter remained causing noise throughout the second LIGO observation.

4 Mock Data

In order to train and develop our neural network, it must be given mock data as an input. Previously, many filters were created to help simulate the real data being outputted by the interferometer. However, the past filters were not able to recreate the spikes as seen in the real data. To generate mock data, you first determine which model you want to use to generate the data. Depending on the desired output, we decide on the length of time series and the sampling frequency. Various parameters are inputted into the specific models, and outputted are an array of times, array of mock DARM background with no additional noise added, which will be used to evaluate the regression efficacy, the background with nonlinear noise added, which will be used as the subtraction target, a number of witness signals to be used as input to the nonlinear regression methods, and an optional dictionary containing auxiliary data that varies on which model is being used. DARM is the term used to describe the degree of freedom describing the differential length of the arms in the instrument. Using the outputs, we also generate a plot to clearly see what is being generated. The purpose of creating this mock data is so that we can realistically simulate outputs from the interferometer to accurately train our neural network.

4.1 Previous Models

The three previous models are known as resonance, biliniear, and scatter. The resonance model uses the previous inputs as well as a resonant frequency and a quality factor to produce the subtraction target and witness signals. The bilinear model attempts to represent data similar to the beam-spot motion and mirror angulary control signal. The beam-spot motion is caused by microseism, or very small earthquakes. Lastly, the scatter model attempts to mimic one witness moving slowly and one acoustically active witness coupling together. While all these model have there own strengths, they all fail to accurately mimic the real data being outputted by the interfometer.

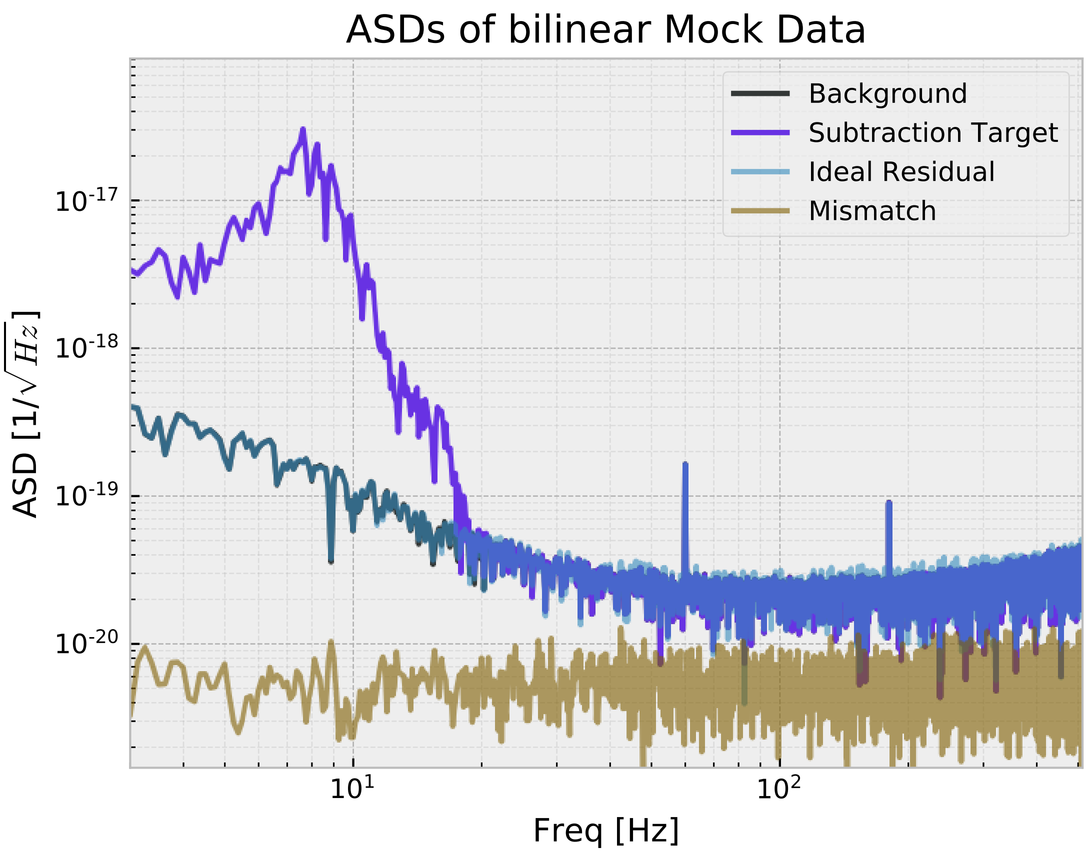

Figure 1: Spectral density of the data being generated by the bilinear filter

4.2 Current Model

While we have not mimicked the real data very accurately, one of the harder features to remove from the data are the irregular spikes and cusps that are present. To mimic these features we are using a filter called ellip. It is quite simple at the moment but it is just a preliminary filter to test and develop the neural network on. Right now, it is a simple bandstop filter but also contains the types of peaks and troughs we are aiming at removing. The idea is that if we test the neural network on this model, it will be easier to build up the model and adjust the network accordingly. This is simply a first step in the right direction and the ellip model does not contain all the factors needed to make an accurate data set. The ellip model is appealing because it time dependent. In order to create a good neural network capable of mimicing this filter, it should in theory need to look at previous inputs in order to correctly predict the output of the current input.

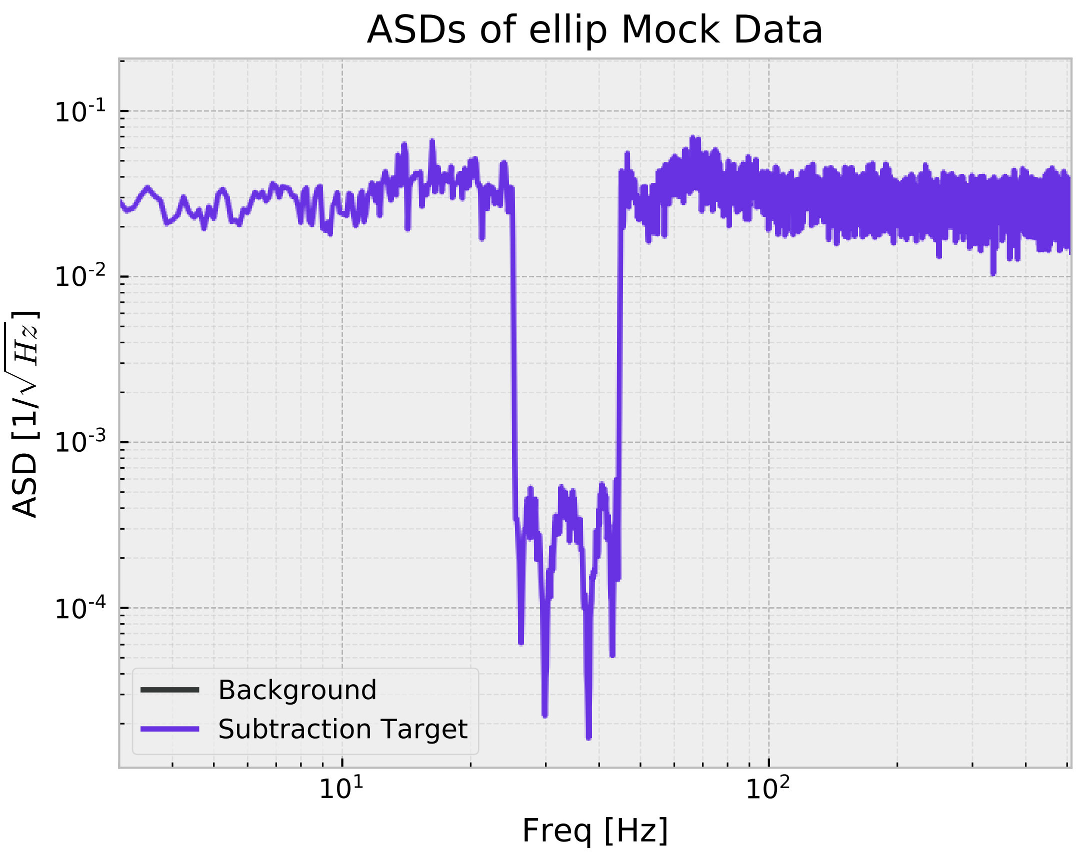

Figure 2: Spectral density of the data being generated by the ellip filter

Throughout the rest of the summer we hope to improve it and combine the previously mentioned models into one model that will accurately represent the real data from the instrument.

5 Machine Learning Techniques

There are two types of noise which contribute to the overall sensitivity of the detector. Non-removable noise sources, such as thermal noise, determine the underlying sensitivity of the detector. These noise sources can only be removed by improving the design of the detector itself. Other noise sources, known as removable noises, such as seismic noise, can be removed by monitoring the witness channels. Witness channels are the channels that monitor the physical environment around the detector. Machine learning algorithms coupled with data from the witness channels can be used on removable noise sources while keeping the signals we want to detect intact. By running the algorithms on past data, noise regression algorithms will be able to predict future noise sources and subtract them from the incoming data. It is also important to define neural networks as a set of algorithms that are designed to recognize patterns. These networks are made up of multiple layers that do different computations on the passing inputs, reducing the outputted data. These layers are identified as the input layer, the hidden layer where computations are preformed, and the output layer. Because the data we are looking at is sequential in nature, we consider the outputted data to be time series data.

Time series data is defined to be taken sequentially. Time series forecasting is the process of predicting future events based on previous data. The most important components of time series are the trend, cycle, seasonality and error. This type of forecasting will be used on the time series data outputted by the detector to analyze the witness channels and the environmental factors contributing to removable noise sources.

We also want to be able to see how the data evolves and reacts to the data around it. For this reason, we want to deal with recurrent neural networks, specifically looking at long short-term memory networks. These are a type of recurrent neural network that process sequential inputs, taking into consideration past inputs to analyze the current input. This process is appealing because it can capture long-term dependencies in the data. Some of the removable noise sources, like seismic waves, will take several seconds to get to the witness channels, so having the ability to take in longer inputs is needed.

Because the data we are dealing with it highly complicated, I plan on trying nonlinear regression algorithms. Nonlinear regression is used to find nonlinear relations between sets of data. It is ideal for the task at hand because it can estimate models with arbitrary relationships between independent and dependent variables.

5.1 Simple Model

Using the mock data generated by both the bilinear filter and the ellip filter, we designed a very simple neural network designed to learn the behaviors of these models and produce similar outputs.

Before we build the model we have to prepare the data that is being inputted into the model. We take the inputs that were inputted into the bilinear and ellip filters, along with their respected filtered outputs and normalize everything. The goal of normalizing the data is to change the values within the data set to a common scale, without distorting the differences in the ranges of values. Without normalized data, different features will have more or less influence on the model compared to others, which is not what we want at the moment. Once the data is normalized, we split the data into training, testing and validation data.

For the simple model, we utilized Keras, an open-source neural-network library, which allowed for fast experimentation with the deeper networks we are interested in. We chose to use the Sequential model. This model linearly stacks layers allowing us to simply create a layer by layer neural network. For this first network, we used four dense layers. The dense layer is a linear operation in which each input is connected to each output by a weight. Our first three layers consist of 128, 64, 32 neurons respectfully and condensing to one neuron in the last layer.

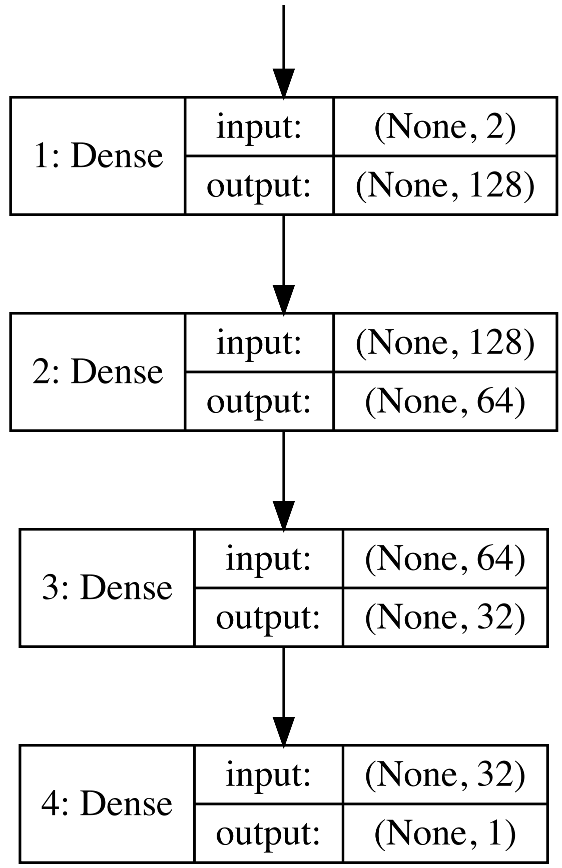

Figure 3: Our simple neural network, consisting of 4 dense layers.

We plan to implement a dropout layer, which will help with over fitting. The dropout layer sets the weights of a fraction of the data to 0, essentially dropping them from the network.

Our current model uses an Nesterov Adam optimizer, incorporating Nesterov momentum. We used scaled exponential linear unit (SELU) as activation for the first three layers, and then used sigmoid activation for the final layer. We currently are using mean square error to calculate the loss.

We ran this model on the bilinear model with 100 epochs. Once trained, the simple neural network we constructed was able to correctly predict the output we wanted and correctly mimic the bilinear filter. We believe that this network works on the bilinear model because its output is not dependent on the previous input and is there time independent. Over each epoch the loss from the training data got smaller. Once applied to the validation data, the loss was still relatively small.

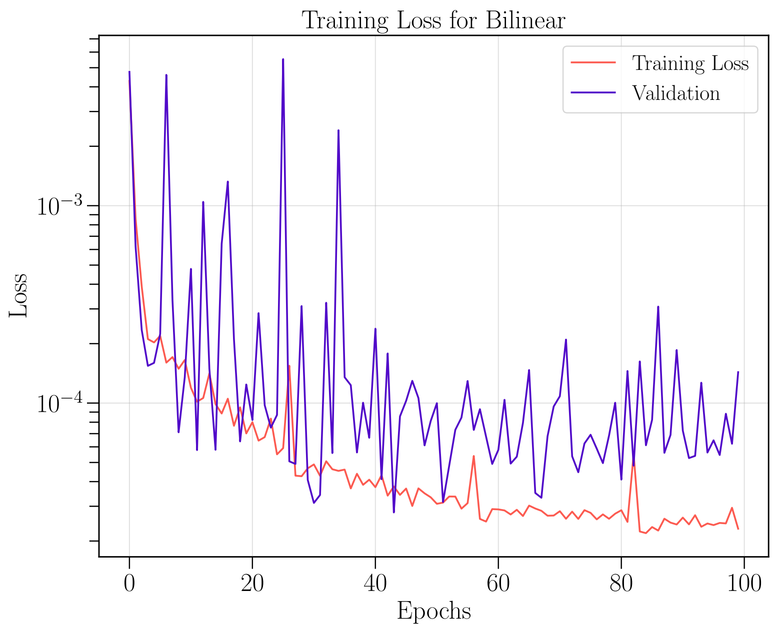

Figure 4: Loss for each epoch using bilinear data

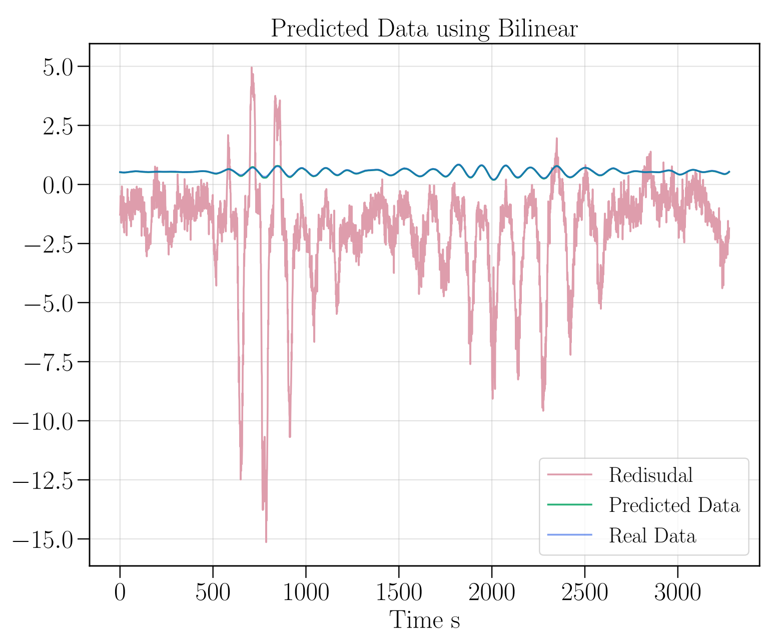

Figure 5: Data predicted by the model after being trained compared to the real data

We ran this model on the ellip model with 100 epochs. It is immediately clear that this neural network does not produce the the desired patterns. We expected this because our network is quite simple but we are dealing more complicated data. The percentage of residual in this model was substanially higher than in the bilinear test. This difference between outputs of the bilinear filter and ellip filter can be attributed to the fact that the ellip filter is dependent on the previous inputs, whereas bilinearly filtered data is not.

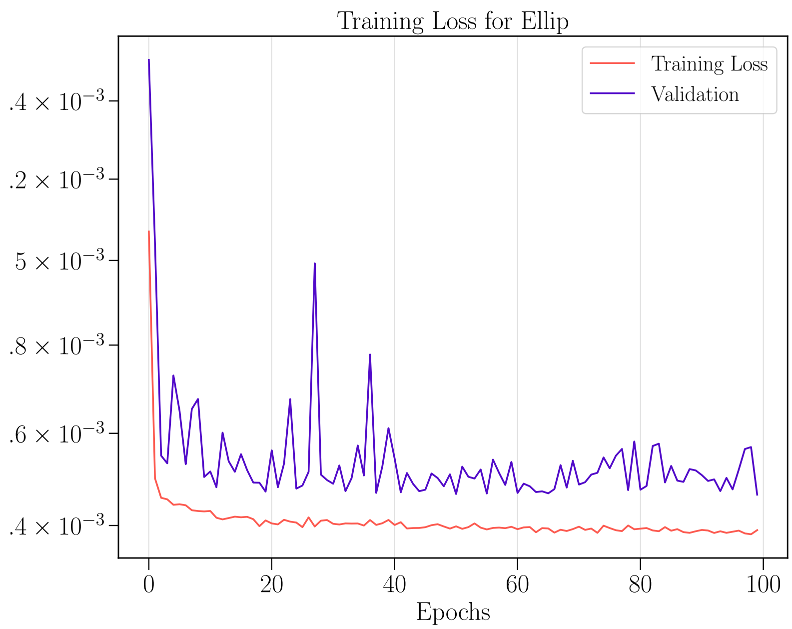

Figure 6: Loss for each epoch using ellip data

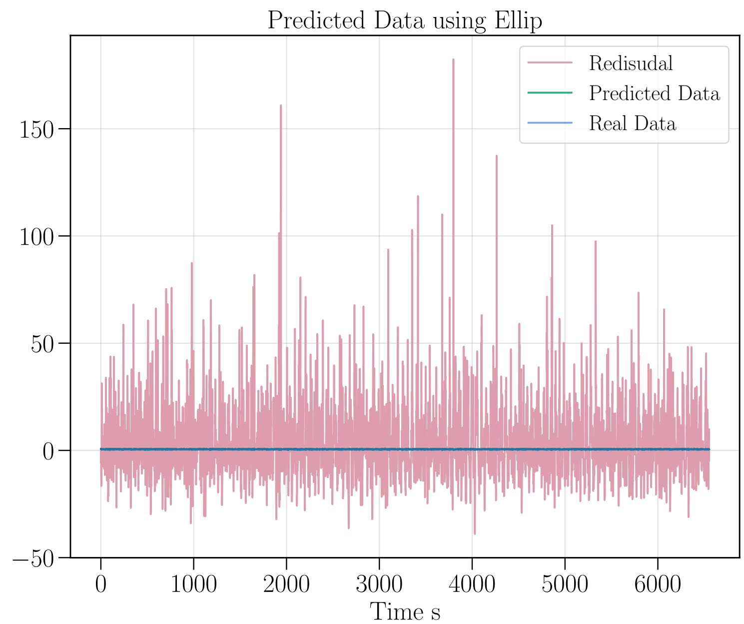

Figure 7: Data predicted by the model after being trained compared to the real data

6 Future

While testing, I will check to make sure that when removing noise sources, it does not remove any potential gravitational wave sources. I will then create and test the neural network. This will take the most time due to all the testing that is required to get the desired technique and algorithms. Once the network works adequately with the mock data, I will test on real LIGO data. During the process of working with the real LIGO data, I plan on analyzing where the noise sources that are being subtracting are originating from. In doing so, we can get a better intuition of what physical properties of the instrument are causing noise.

References

[1] Beverly K. Berger. Identification and mitigation of advanced ligo noise sources. *Journal of Physics: Conference Series*, 957:511–653, 2018.

[2] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall. Improving the sensitivity of Advanced LIGO using noise subtraction. *Classical and Quantum Gravity*, 36(5):055011, March 2019.