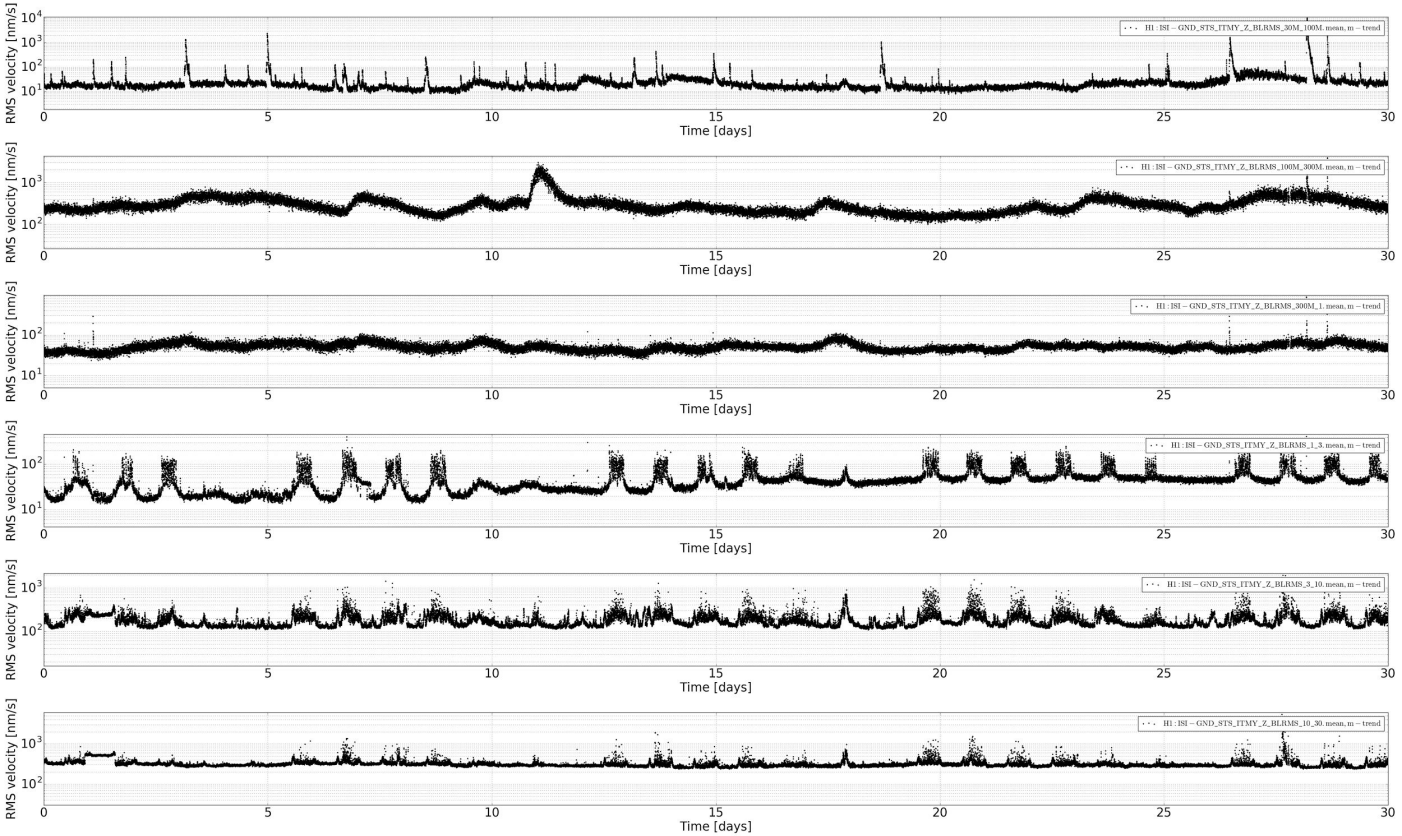
Online Detector Characterization using Neural Networks

Roxana Popescu
Rana Adhikari, TJ Massinger, Jess McIver

Introduction

- Data from LIGO contains noise from many sources, that need to be characterized
- Machine learning algorithms can be used to look for patterns within the data and to cluster or classify the data into different categories
- Would help determine if changes in detector sensitivity are related to changes in environment
- Looked at seismic noise for project
- Other Environmental channels: wind, acoustic

Seismic BLRMS Data



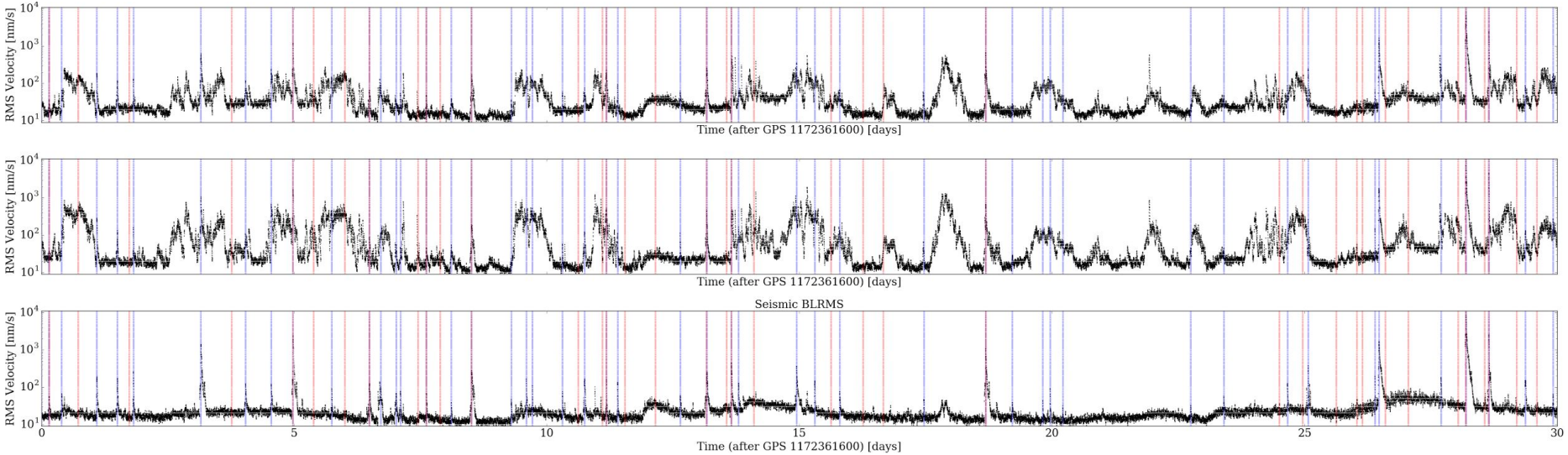
Machine Learning

- Machine learning is the field of study of programming computers so that they can learn from inputted data and improve their performance as they are given more data
- Supervised Learning vs. Unsupervised Learning
- Classification vs. Clustering

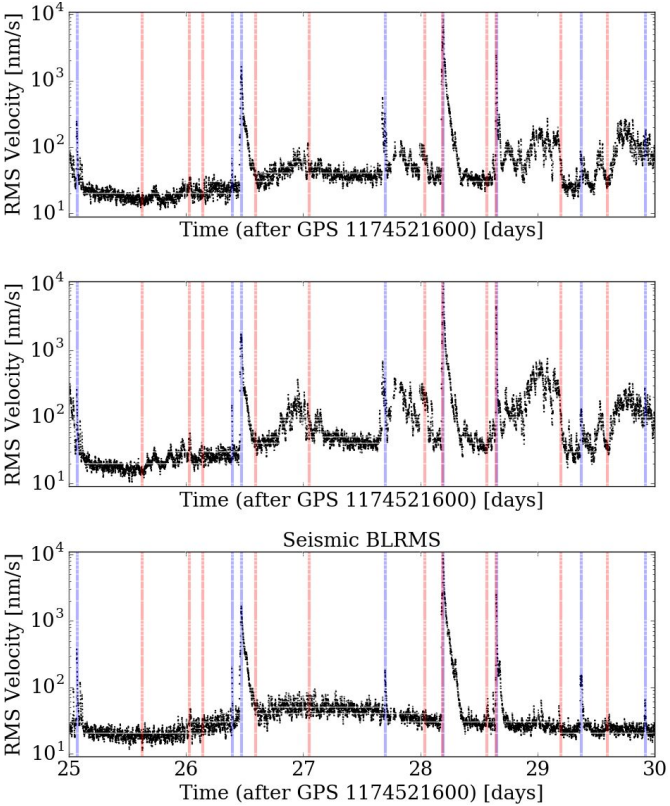
Evaluating How Well Clustering Works

- Calinsky Harabaz-Score
  - Ratio of between-clusters dispersion mean to within-cluster dispersion mean
- Comparison to recorded earthquake times
  - Add up cluster labels that occur 10 minutes before/after an earthquake
  - Add total number of cluster labels
  - For each cluster determine score , E(k), by dividing cluster labels near earthquake, N_e, by total cluster labels, N_t
  - $E(k) = N_e/N_t$

Determining Earthquake Times



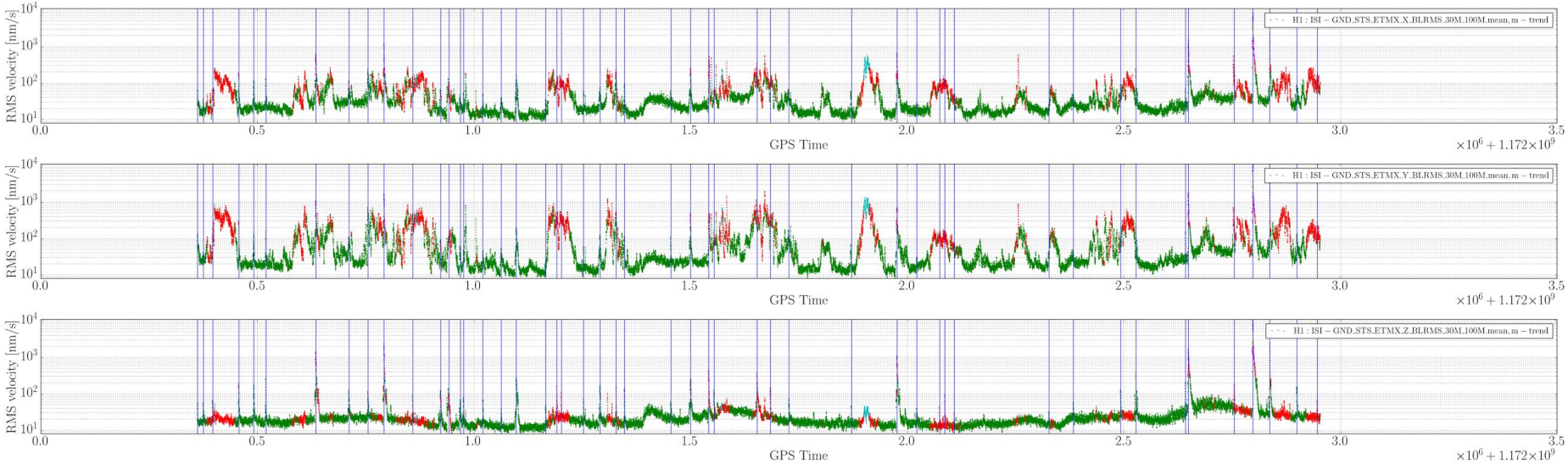
Determining Earthquake Times



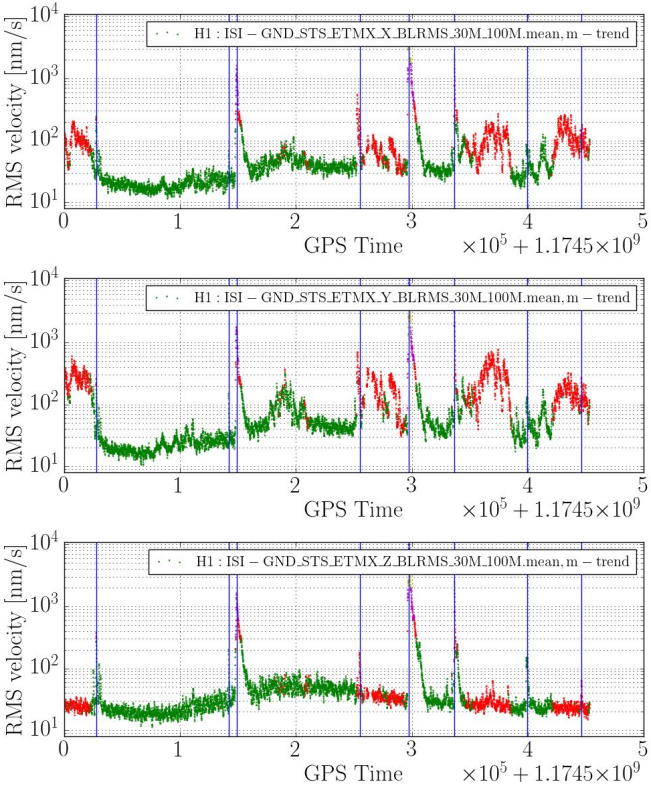
Clustering Algorithms

- Kmeans
  - Splits data into k number of clusters by minimizing distances between points and average point in cluster
- DBSCAN
  - Splits data into clusters to create clusters out of high density areas
- Agglomerative Clustering
  - A type of hierarchical clustering that builds clusters by merging data points into clusters
- Birch
  - Makes a tree data structure

Kmeans


RMS velocity [nm/s]
GPS Time
$\times 10^6 + 1.172 \times 10^9$
H1 : ISI – GND_STS.ETMX_X_BLRMS_30M_100M.mean, m – trend
H1 : ISI – GND_STS.ETMX_Y_BLRMS_30M_100M.mean, m – trend
H1 : ISI – GND_STS.ETMX_Z_BLRMS_30M_100M.mean, m – trend

Kmeans


H1 : ISI − GND_STS_ETMX_X_BLRMS_30M_100M.mean, m − trend
RMS velocity [nm/s]
GPS Time
×10^5 + 1.1745×10^9
H1 : ISI − GND_STS_ETMX_Y_BLRMS_30M_100M.mean, m − trend
RMS velocity [nm/s]
GPS Time
×10^5 + 1.1745×10^9
H1 : ISI − GND_STS_ETMX_Z_BLRMS_30M_100M.mean, m − trend
RMS velocity [nm/s]
GPS Time
×10^5 + 1.1745×10^9

Kmeans

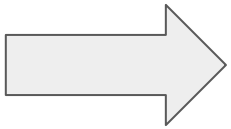
Number of Clusters	Calinsky-Harabaz Score	Cluster of Max Earthquake Score	Maximum Earthquake Score
2	40172.1	1	0.03
3	37282.1	1	0.04
4	43960	1	0.07
5	44224.7	4	0.08
6	45616.4	3	0.08
7	46338.4	3	0.08
8	46348.9	7	0.11
9	46095.1	1	0.11
10	46746.5	6	0.13
Average	44087.1	N/A	0.08

DBSCAN

Epsilon Value	Minimum Samples	Number of Clusters	Calinsky-Harabaz Score	Cluster of Maximum Earthquake Score	Maximum Earthquake Score
1	15	1	14.2	-1	0.0125
2	10	15	5.1	-1	0.0126
2	15	5	6.3	-1	0.0125
2	20	1	14.2	-1	0.0125
2	25	1	14.2	-1	0.0125
2	30	1	14.2	-1	0.0125
3	15	6	123.1	-1	0.0141
4	15	6	194.1	-1	0.0159
5	15	8	372.5	-1	0.0176

Include Shifted Data in Clustering

A	B	C	D	E	F
0	1	2	3	4	5



A	B	C	D	E	F
0	1	2	3	4	5
1	2	3	4	5	
2	3	4	5		

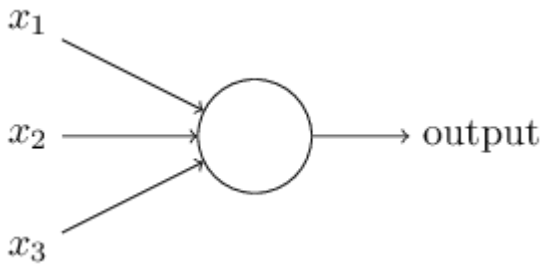
Shifting Data by Two Indices

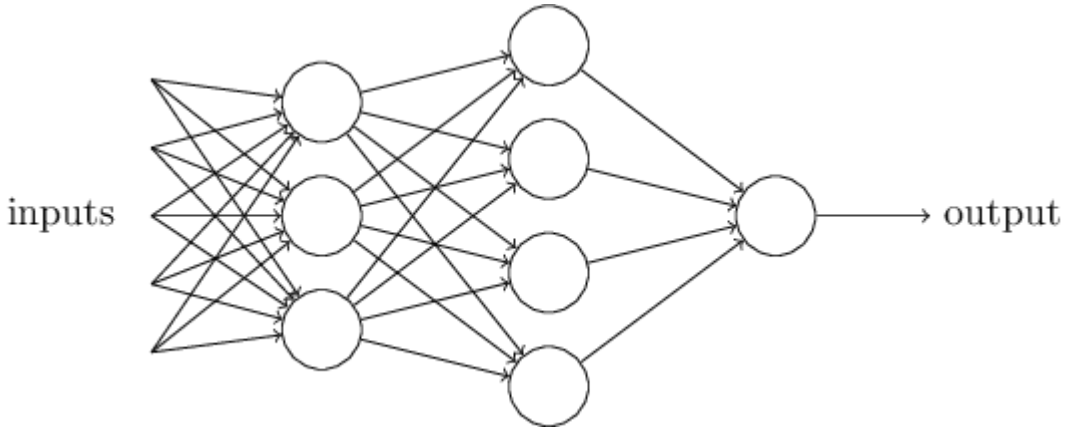
Include Shifted Data in Clustering

Timeshift (minutes)	Calinsky-Harabaz Average	Maximum Earthquake Score Average
0	44087.1	0.08
10	49251.1	0.08
30	44081.2	0.09
60	44066.1	0.08

Neural Networks

- Neural networks can be used to find relationships in data by using hidden layers of connections within the data



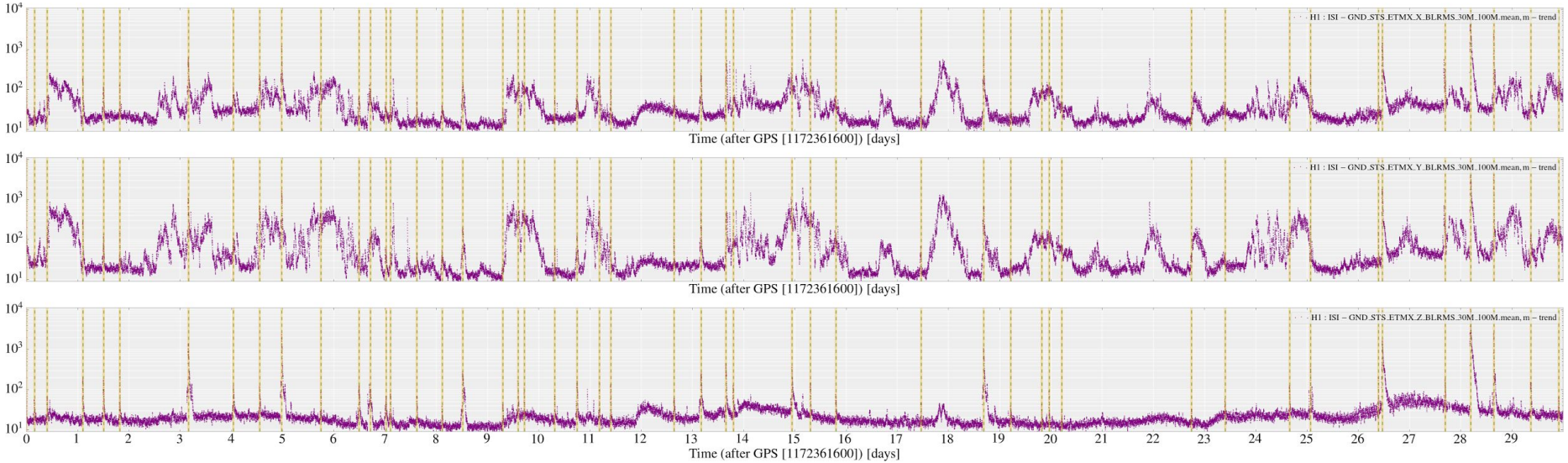


Figures from: http://neuralnetworksanddeeplearning.com/chap1.html

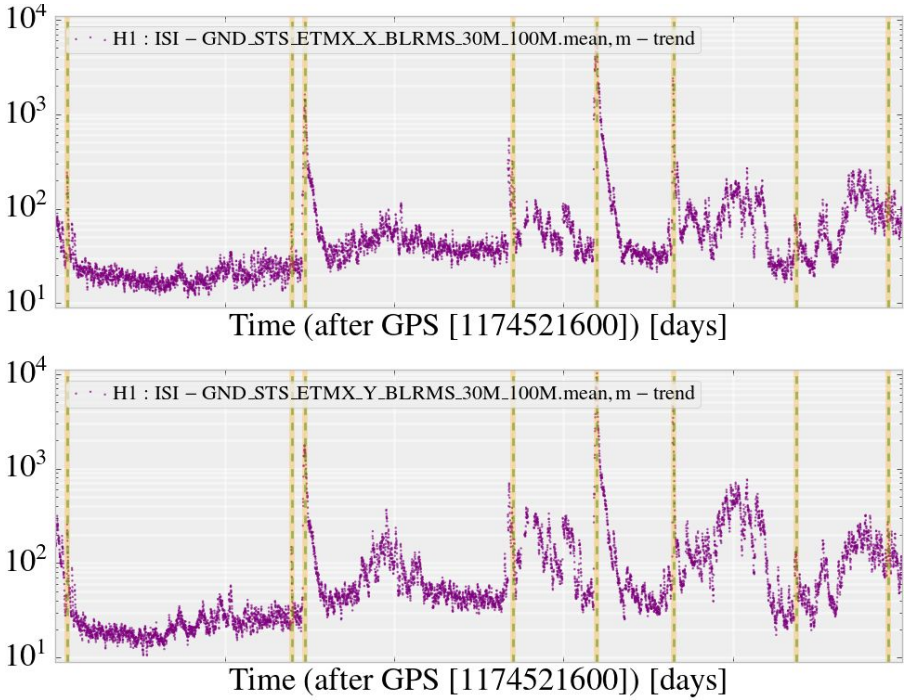
Neural Networks

- We used keras with tensorflow backend
- Timeshift the data by 30 min
- Read in whether an earthquake occurs at a given time
- Use Sequential model to add four layers
- Use sigmoid activation
- Accuracy: 0.998

Neural Networks



Neural Networks



Future Work

- Obtain six months of data to use for training the neural network
- Improve the neural network
- Compare neural network results to results from clustering
- Cluster and classify DARM channel BLRMS