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

Technical Note **LIGO-T2300147–v** 2023/06/05

LIGO Seismic State Characterization using Machine Learning Techniques

Isaac Kelly

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/

1 Introduction

The Laser Interferometer Gravitational-wave Observatory (LIGO) uses laser interferometers to detect gravitational waves. These distortions in spacetime appear as changes in the relative lengths of the interferometer arms, which causes a phase shift in the light reflected by the test masses. The detector signal indicates the current strain on spacetime. Analysis of the strain over time allows extraction of signals from individual events, and these events provide insight on astrophysical and relativistic phenomenae. Current technological limitations constrain the observable emissions to those from inspiraling compact binary objects. Other sources are predicted by models, but no detection has been made yet.

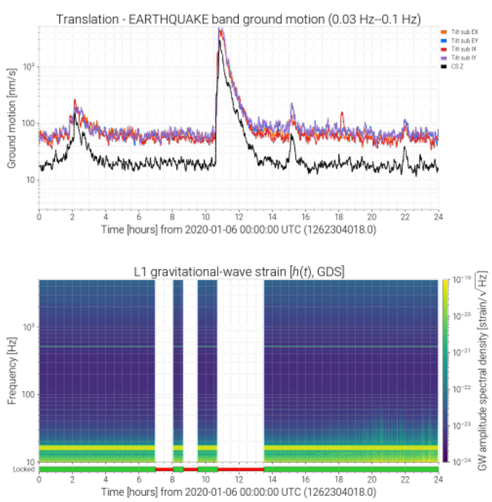

Figure 1: Earthquake event and corresponding data loss

Besides gravitational radiation, terrestrial detectors are subject to a range of other strains, all of which can interfere with the correct operation of the detectors. A significant part of this interference is ground motion; the 4-km interferometer arms are susceptible to distortion caused by movements in the earth beneath them. Even the strongest gravitational waves require a detector sensitivity of approximately $1 * 10^{-21}/\sqrt{Hz}$ to be evaluated with scientific significance.[1] One common type of ground motion, called the 'secondary microseism', is over 10 orders of magnitude stronger than the real signal at 10 Hz. [1] Earthquakes and human-caused (or anthropogenic) noise also cause distortions or loss of the strain signal. Ground motion is a significant contributor to noise and glitches in the detector (see figure 1) with both active and passive isolation utilized in the system to reduce its effects. [1] However, this isolation cannot completely nullify its effects, and so it is necessary to determine when the detector is being affected by this interference in order to find noise sources, test new isolation methods, and avoid misinterpretation of such noise as a gravitational wave event.

In order to monitor external noise sources, LIGO maintains many physical environmental monitors, or PEMs. These include seismometers, accelerometers, and microphones, all of which produce streams of time-series data on disturbances in the LIGO system. We will use traditional clustering methods to evaluate the seismic state of the detector via analysis of PEM sensor data. The time-series data from PEM channels will be divided into time segments. Two methods will be used to create clusters from these segments. In one method, a feature extraction process will create a reduced dataset to which the clustering algorithms will be applied. In the other approach, the algorithms will be applied directly to the raw time-series segments.

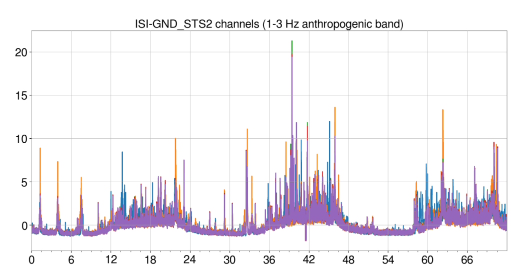

Figure 2: Periodic elevation in nosie floor in the anthropogenic band during daytime hours, followed by reduction during nighttime hours.

These complementary methods will be used to implement a pipeline which can in real time determine the seismic state of the detector.

2 Objectives

Given environmental data from PEM sensors, we plan to determine the seismic state of the LIGO detector using clustering algorithms. The specific objectives are summarized below.

- **Objective 1: Dataset creation.**

  From the time-series PEM data, fixed-length segments of time will be extracted, and a feature extraction process will transform these segments into a scalar dataset. Alternatively, a raw time-series dataset may be created.

- **Objective 2: Clustering and evaluation.** While individual sensors can provide clear information about seismic states in some situations, analysis of the entire corpus of sensor data should improve the reliability of state determination. Given N time segments, K discrete clusters will be identified, classifying time periods according to statistical similarities.

- **Objective 3: State identification.** Manual labeling will permit correlation of clusters with known detector states; this may allow discovery of new states, and provides room for future exploration. The final goal for this project is the automatic labeling of the detector's ground-motion state as a veto provider and data quality metric.

3 Approach

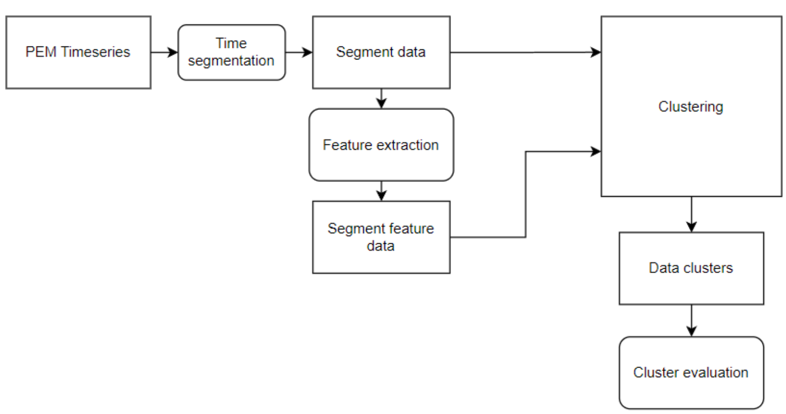

Figure 3: Programming approach

Time-series data will be acquired from various ground motion sensors as well as other instruments which show related noise. This data will be separated into fixed-length segments, with the length as an experimental parameter. Representations of these segments will be evaluated in two forms. First, the **tsfresh** Python library [2] will be used to extract meaningful statistical features. This significantly reduces the dimensionality of the dataset. Second, in a shape-based approach, the raw data itself will be treated as a representation of the segment.

In either approach, the next step is the application of clustering algorithms, found in the **scikit-learn** [3] Python library. Initial algorithms in use include k-means [4] and DBSCAN. [5] Various clustering parameters will be explored. The most interesting parameter is the specified number of clusters, k. Because there may be undefined and unknown ground-motion states, the 'true value' of k is unknown, so comparing clustering results with various values for k will provide interesting data. Time segment length can also be adjusted; this may affect the types of noise which are detectable, as noise events may vary in length.

Cluster quality will be evaluated using metrics available in **scikit**. These may include the variance ratio criterion and the Davies-Bouldin index. Because the 'true state' of each time segment, known as the ground truth, is not necessarily known, many metrics cannot be used for cluster evaluation. However, because some ground motion states are well defined, clusters may be compared to known states in some situations. And clusters with unknown states may give clues to the existence of entirely new states. A diverse evaluation approach will be utilized, and should offer a broad field for analysis and experimentation.

4 Project Schedule

We present an outline of the projected project schedule.

- Before arrival: Familiarization with coding toolkits by reading documentation and experimentation with
- Week 1: Initial orientation to gain familiarity with the toolkits and manipulating LIGO datasets.
- Week 2: Creation of feature-extraction pipeline using tsfresh
- Week 3-4: Exploration of clustering algorithms, their results, and the corresponding quality metrics
- Week 5-6: Interim report. Further statistical analysis of clusters
- Week 7: Exploration of correlation between cluster properties and novel seismic states
- Week 8: Begin final report. Continued exploration of states and their relationship to clusters.
- Week 9-10: Completion of final report

References

[1] Rana X. Adhikari. Gravitational radiation detection with laser interferometry. *Reviews of Modern Physics*, 86(1):121–151, feb 2014.

[2] tsfresh documentation. https://github.com/blue-yonder/tsfresh. Accessed: 2023-05-14.

[3] scikit-learn. https://scikit-learn.org/stable/. Accessed: 2023-05-17.

[4] J MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297. University of California Los Angeles LA USA, 1967.

[5] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.