The 2024 NASA Space Apps Challenge

Challenge : Seismic Detection Across the Solar System.

Project: **Noiseless Netting**

Present:

Team **AxoloTech**

Members:

- Angel Eduardo Perez Garcia
- Jishar Adonay Tiburcio Zamudio
- Jonatan Luna Franco
- Mizar Janai Chiquito Espino
- Uriel García García
- Tanil Izquierdo Cordova

Local Event:

Xalapa, Veracruz, México

Date: October 6th, 2024

Index

Index 2
1. Introduction 3
2. Problem statement 4
3. Justification 4
4. General objective 5
5. Specific objectives 5
6. Theoretical framework 5
6.1. Seismology 5
6.1.1. Seismic waves 5
6.1.1.1. Earthquakes 6
6.1.1.2. Moonquakes 6
6.1.1.3. Marsquakes 6
6.1.1.4. Types of Seismic waves 7
6.1.1.4.1. P-waves 7
6.1.1.4.2. S-waves 7
6.1.1.5. Detection of seismic movements 7
6.1.1.5.1. STA (Short-Term Average) and LTA (Long-Term Average) 8
6.2. Datasets 9
6.2.1. Seismic data 9
6.2.1.1. MiniSeeds 9
6.3. Data processing 10
6.3.1. Denoising waves 10
6.3.2. Machine learning 10
6.3.2.1. Clustering 10
6.3.2.1.1. DBSCAN 10
6.4. Used Technologies 11
6.4.1. Python 11
7. Methodology 11
8. Results 12
10. Conclusions 14
11. References 14

1. Introduction

In space exploration, the study of planetary seismic activity is fundamental for understanding the internal structure of celestial objects such as the Moon and Mars. For instance, the Apollo and Mars InSight missions have provided a large amount of seismic data that can offer valuable information for their study. However, this data often contains significant noise, which makes it harder to process and to analyze seismic events. Transmitting this data back to Earth consumes a lot of energy, and sometimes it even makes transmission impossible. Additionally, it makes identifying truly important events more difficult.

Generally, noise is defined as any unwanted signal. In this context, noise can arise from various sources, such as environmental interference, measurement errors, and mechanical or electronic noise (Proakis & Manolakis, 1996).

Signal distortion is one of the main issues caused by noise, making it harder to identify significant events in the data. This is particularly important in seismology, where the clarity of the seismic signal is crucial for analysis.

This project aims to develop a system based on an algorithm for efficiently filtering noise in seismic data from the Apollo and Mars InSight missions.

This work will not only improve the accuracy of identifying seismic events, but it could also contribute to future research in planetary seismology.

2. Problem statement

In all their missions, NASA tried to collect all the information they could from Mars and the moon using different tools, one of them is the seismograph. It's a very useful tool because with the information it collects we can know things about celestial bodies, such as element composition, density and mass. As you know, “on average, Mars is 140 million miles (225 million km) away from

Earth. The distance can be greater or less than that depending on where the planets are in their orbits around the sun.” (Distance From Earth to Mars | Facts & Measurement | Study.com, n.d.) and the average distance between the moon and the earth is 384 400 km (238 855 miles). so, sending the collected data back to earth is a hard task to do. When the seismograph collects the data, usually all this data has noise, useless data that NASA isn’t looking for, all this noise makes seismic files have a bigger size and consume more energy than it should be.

3. Justification

The main reason for creating this project stems from our desire to make a contribution to the field of planetary seismology. This desire is deeply tied to the importance of studying celestial bodies to better understand our place in the universe. The fact that we can contribute through software development to a discipline that helps decipher the geological secrets of planets and moons is a source of inspiration for our team. We believe that, through technological solutions, it is possible to improve the analysis and interpretation of seismic data, making space exploration more efficient and accurate.

Our interest is not only scientific but also personal. As developers, we are passionate about using our skills to solve complex problems and contribute to causes we consider essential for human progress. The fact that we can combine our passion for programming with a goal as relevant as planetary seismology makes this project particularly meaningful. We want to contribute in the way we know best: by creating technological solutions that can have a real impact on space missions and the way data collected on other planets is analyzed.

The desire to help advance knowledge about seismic activity on celestial bodies has become a personal mission for us. Knowing that our work could facilitate future scientific discoveries and improve the efficiency of data collection and transmission deeply motivates us. This project is our way of collaborating with the global scientific effort, offering our skills as programmers to enhance the tools that help unveil the secrets of our universe.

4. General objective

“Write a computer program to analyze real data from the Apollo missions and the Mars Interior Exploration using Seismic Investigations, Geodesy and Heat Transport (InSight) Lander to identify seismic events.”(*Space Apps Challenge*, 2024)

5. Specific objectives

- Reduce the amount of noise in each seismic file to reduce its size.
- Avoid losing relevant data while reducing the amount of noise.
- Make seismic events detection easier by reducing the noise.

6. Theoretical framework

6.1. Seismology

Seismology is the scientific study of earthquakes and the propagation of elastic waves through the Earth or other planetary bodies. It involves analyzing the causes, effects, and mechanics of earthquakes, as well as the internal structure of the Earth.

6.1.1. Seismic waves

Seismic waves are caused by the sudden movement of materials within the Earth, such as slip along a fault during an earthquake. Volcanic eruptions, explosions, landslides, avalanches, and even rushing rivers can also cause seismic waves. Seismic waves travel through and around the Earth and can be recorded with seismometers.

6.1.1.1. Earthquakes

Earthquakes are any sudden shaking of the ground caused by the passage of seismic waves through Earth’s rocks. Seismic waves are produced when some form of energy stored in Earth’s crust is suddenly released, usually when masses of rock straining against one another suddenly fracture and "slip."

6.1.1.2. Moonquakes

Moonquakes are seismic events occurring on the Moon, analogous to earthquakes on Earth but with significant differences in origin, frequency, and intensity. They are produced as a result of meteoroids hitting the surface or by the gravitational pull of the Earth squeezing and stretching the moon's interior. (*Moonquakes and Marsquakes: How We Peer Inside Other Worlds*, 2020)

6.1.1.3. Marsquakes

Marsquakes are seismic events occurring on Mars, they provide crucial insights into the interior structure and geological activity of Mars.

"Mars doesn't seem to have plate tectonics, but other things can make the ground shake too, like:

- Cracking caused by contraction from cooling.
- Magma moving and creating pressure deep underground.
- Or even meteorite impacts." ("Student Video: Mars in a Minute: Are There Quakes on Mars? | NASA/JPL Edu," 2020)

6.1.1.4. Types of Seismic waves

"There are three basic types of seismic waves – P-waves, S-waves and surface waves. P-waves and S-waves are sometimes collectively called body waves."

6.1.1.4.1. P-waves

"P-waves, also known as primary waves or pressure waves, travel at the greatest velocity through the Earth." (*Seismic Waves*, n.d.)

6.1.1.4.2. S-waves

"S-waves, also known as secondary waves, shear waves or shaking waves, are transverse waves that travel slower than P-waves. In this case, particle motion is perpendicular to the direction of wave propagation."(*Seismic Waves*, n.d.)

6.1.1.5. Detection of seismic movements

Seismometers enable us to detect and measure earthquakes by transforming vibrations from seismic waves into electrical signals, which can then be visualized as seismograms on a computer screen. Seismologists use this data to analyze earthquakes, determining their location and magnitude.

To accurately capture ground motion in three dimensions, seismologists use three separate sensors within the same instrument. Each sensor records vibrations in a specific direction:

- The Z component captures vertical (up/down) motion.
- The E component records east-west motion.
- The N component measures north-south motion.

(British Geological Survey, 2021)

6.1.1.5.1. STA (Short-Term Average) and LTA (Long-Term Average)

The STA is an algorithm used in the detection of seismic events that averages the amplitude of a seismic signal in a short window of time, typically 1 to 5 seconds. This calculation is performed continuously as the window slides over the signal.

- Purpose: The purpose of the STA is to quickly capture changes in the amplitude of the seismic signal that indicate the occurrence of a seismic event, such as an earthquake. An increase in the value of the STA suggests increased seismic activity in the near term.

- Calculation: It is calculated by adding the amplitudes of the signal in a short period of time and dividing the result by the duration of the window. As the window scrolls, a new average is obtained.

Definition: LTA is a complementary algorithm to STA that averages the amplitude of the seismic signal over a longer time window, usually 10 to 30 seconds or more. This calculation is performed continuously, similar to the STA.

- Purpose: The LTA provides background context about seismic activity and helps identify significant changes compared to background noise. An increase in the LTA value indicates an overall level of seismic noise.

- Calculation: It is calculated by adding the amplitudes of the signal over a longer period of time and dividing the result by the duration of that window. The LTA is used to establish a baseline level against which the STA can be compared.

STA/LTA Ratio: The STA/LTA ratio is calculated by dividing the STA value by the LTA value at any given time. This index provides a measure of seismic activity in relation to the level of background noise.

- Use: It is used as a criterion for detecting seismic events. When the STA/LTA ratio exceeds a specific threshold, it can be inferred that a significant seismic event has occurred.

- Detection Threshold: Establishing a threshold for the STA/LTA ratio is critical to avoid false alarms. This threshold can be adjusted based on the specific conditions of the seismic environment.

6.2. Datasets

A dataset is a structured collection of data stored together for analysis or processing. The data within a dataset is typically related in some way, in this case, the dataset was provided by NASA.

6.2.1. Seismic data

Data obtained from various levels of the celestial bodies' subsurface using seismic techniques.(Oliver et al., 2021)

6.2.1.1. MiniSeeds

The International Federation of Digital Seismograph Networks (FDSN) defines miniSEED as "a format for digital data and related information. The primary intended uses are data collection, archiving and exchange of seismological data".(*Overview — FDSN MiniSEED 3*, n.d.)

6.3. Data processing

Data processing refers to essential operations executed on raw data to transform the information into a useful format or structure that provides valuable insights to a user or organization. The outcomes of data processing operations flow into various data outputs as designed by a data scientist, including data analytics, business intelligence, machine learning and artificial intelligence. (Kerner, 2024)

6.3.1. Denoising waves

In the context of wavelets, "denoising" means reducing the noise as much as possible without distorting the signal. Denoising makes use of the time-frequency-amplitude matrix created by the wavelet transform. (Intro. To Signal Processing:Wavelets And Wavelet Denoising, s. f.)

6.3.2. Machine learning

"Machine Learning, often abbreviated as ML, is a subset of artificial intelligence (AI) that focuses on the development of computer algorithms that improve automatically through experience and by the use of data. In simpler terms, machine learning enables computers to learn from data and make decisions or predictions without being explicitly programmed to do so." (Crabtree, 2023)

6.3.2.1. Clustering

Clustering is an unsupervised machine learning algorithm that organizes and classifies different objects, data points, or observations into groups or clusters based on similarities or patterns.(Ibm, 2024)

6.3.2.1.1. DBSCAN

Clustering analysis or simply Clustering is basically an Unsupervised learning method that divides the data points into a number of specific batches or groups, such that the data points in the same groups have similar properties and data points in different groups have different properties in some sense. It comprises many different methods based on differential evolution.(GeeksforGeeks, 2023)

6.4. Used Technologies

6.4.1. Python

Python is a high-level, interpreted programming language used across various fields like web development, data science, machine learning, automation, and scientific computing.

6.4.1.1. ArgParser

ArgParser is a Python module that makes it easy to write user-friendly command-line interfaces. It is the channel through which we connect the UI with the logic. It accepts arguments to pass to the main denoising function.

6.4.1.2. Scikit-learn

With the goal of offering easy-to-use and effective tools for data mining, data analysis, and machine learning tasks, Scikit-learn is a popular open-source machine learning library for Python. Building upon core libraries like NumPy, SciPy, and matplotlib, it is renowned for its user-friendliness, adaptability, and integration with the Python community.

7. Methodology

For the development of this project, a documentary methodology was implemented that allowed us to thoroughly investigate the principles and techniques involved in noise filtering in seismic data, as well as the functioning of the DBSCan algorithm. This approach was based on the collection and analysis of existing information through various sources, including educational videos, technical documentation of Python and multiple online searches.

The information search process began with watching videos that addressed the logical functioning of the DBSCan algorithm. These visual presentations facilitated the understanding of fundamental concepts and allowed us to observe practical examples of how the algorithm is applied in data processing. Additionally, we reviewed official documentation for Python and familiarized ourselves with the libraries and tools that would be used in the algorithm's implementation.

As we progressed in our research, we conducted online searches to delve deeper into the concept of noise in seismic waves, including primary (P) waves and secondary (S) waves. This phase was crucial, as many of these concepts were unknown to the team, making it necessary to establish a solid understanding of how noise affects the quality of seismic data. Through the review of articles, technical reports, and educational resources, we managed to build a theoretical framework that underpins our approach to noise filtering.

The information gathered was analyzed and organized into relevant topics, allowing us to integrate the knowledge acquired into the development of our solution. However, it is important to highlight that the documentary methodology presents limitations, as reliance on existing information can restrict innovation and may not always address the practical complexities that may arise during the algorithm's implementation in real-world scenarios.

8. Results

9. Thanks to our algorithm, it was possible to process the seismic data before sending them back to earth, reducing the noise of these, as shown in the following comparison:

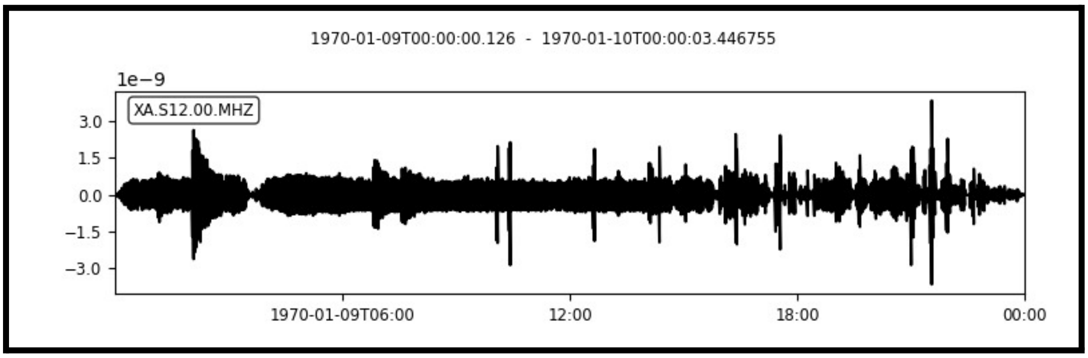

Figure 1: Data before denoising.

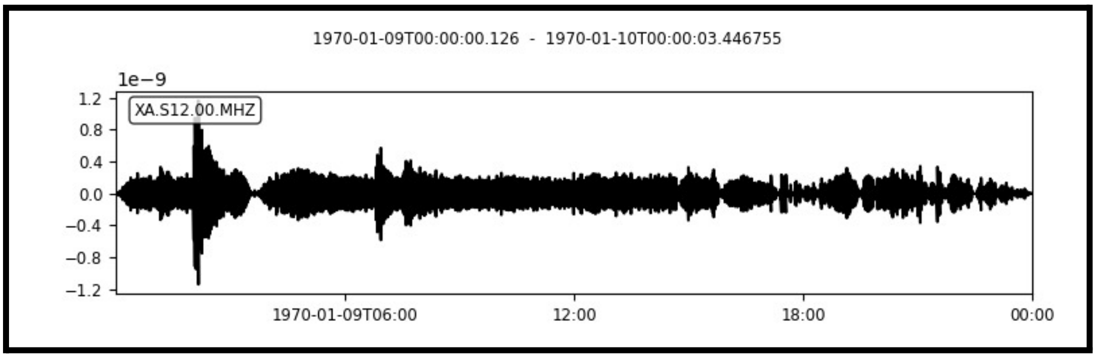

Figure 2: Data after processing and denoising.

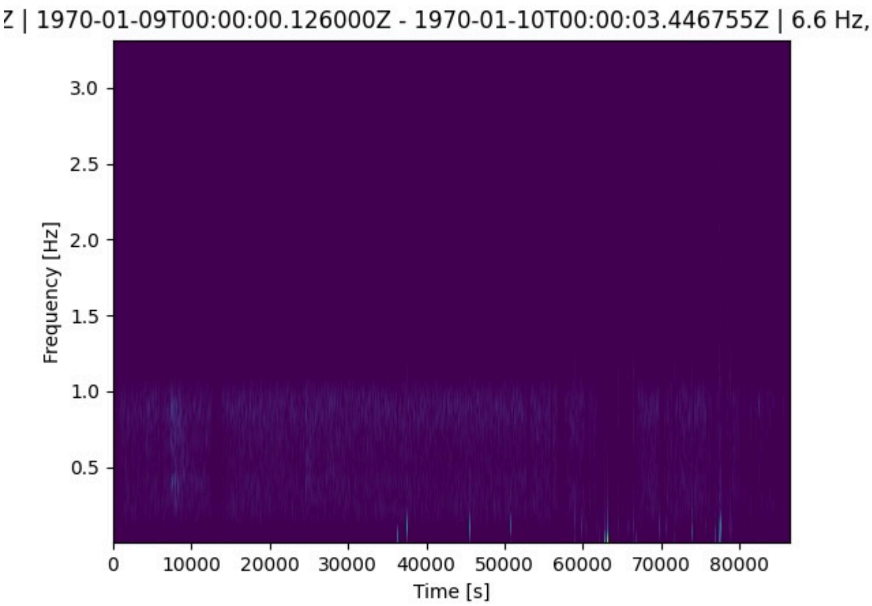

Figure 3: Frequency spectrogram of the denoised data.

We were also able to represent the filtered data as a spectrogram:

It should be noted that the new graphs are generated from a new file ". mseed" which is the result that would be sent to earth in case of a formal implementation, and whether it is profitable to send the data or not.

10. Conclusions

During the project, we learned new concepts that changed our perspective on seismic activities and their characteristics.

Key takeaways include understanding the importance of designing an algorithm that dynamically adapts to different sampling rates and noise types, as seismic conditions vary widely.

We also identified opportunities to improve the algorithm for more efficient real-time processing, valuable for continuous seismic monitoring.

This has given us a solid foundation to advance seismic analysis and create more robust solutions for diverse environments.

11. References

- Michigan Technological University. (2021, 4 octubre). Seismology. https://www.mtu.edu/geo/community/seismology/learn/seismology-study/#:~:text=Seismology%20is%20the%20study%20of,studies%20earthquakes%20and%20seismic%20waves
- Bolt, B. A. (2024, 12 septiembre). Earthquake | Definition, Causes, Effects, & facts. Encyclopedia Britannica. https://www.britannica.com/science/earthquake-geology/Tectonics
- Proakis, J. G., & Manolakis, D. G. (1996). Digital Signal Processing: Principles, Algorithms, and Applications (3rd ed.). Prentice-Hall International, Inc.
- Kerner, S. M. (2024, 19 abril). data processing. Data Backup. https://www.techtarget.com/searchdatabackup/definition/data-processing
- Intro. to Signal Processing:Wavelets and wavelet denoising. (s. f.). https://terpconnect.umd.edu/~toh/spectrum/wavelets.html#:~:text=In%20the%20context%20of%20wavelets,created%20by%20the%20wavelet%20transform.
- Crabtree, M. (2023, July 19). What is Machine Learning? Definition, Types, Tools & More. https://www.datacamp.com/blog/what-is-machine-learning
- Ibm. (2024, 13 agosto). Clustering. Clustering. https://www.ibm.com/topics/clustering
- GeeksforGeeks. (2023, May 23). DBSCAN Clustering in ML | Density based clustering. GeeksforGeeks. https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/
- *Seismometers, seismographs, seismograms - what's the difference? How do they work? | U.S. Geological Survey*. (2024, June 6). https://www.usgs.gov/faqs/seismometers-seismographs-seismograms-whats-difference-how-do-they-work

- *How far away is the Moon?* (s. f.). Royal Museums Greenwich. https://www.rmg.co.uk/stories/topics/how-far-away-moon#:~:text=How%20far%20away%20is%20the%20Moon%20from%20the%20Earth%3F,km%20(238%20855%20miles).
- DataScientest. (2023, October 30). *Machine Learning & Clustering: el algoritmo DBSCAN*. Formación En Ciencia De Datos | DataScientest.com. https://datascientest.com/es/machine-learning-clustering-dbscan
- *DBSCAN*. (n.d.). Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html
- Student video: Mars in a minute: Are there quakes on Mars? | NASA/JPL Edu. (2020, December 31). *NASA/JPL Edu*. https://www.jpl.nasa.gov/edu/learn/video/mars-in-a-minute-are-there-quakes-on-mars/
- *Moonquakes and marsquakes: How we peer inside other worlds*. (2020, August 10). Horizon Magazine. https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/moonquakes-and-marsquakes-how-we-peer-inside-other-worlds
- *Seismic waves*. (n.d.). Science Learning Hub. https://www.sciencelearn.org.nz/resources/340-seismic-waves
- Databricks. (2024, June 6). *What is a Dataset? | Databricks*. https://www.databricks.com/glossary/what-is-dataset