Comprehensive End Term Report

High Prep 4

High Resolution Elemental Mapping of the Lunar Surface

By

Team 41

Date of Submission: December 7, 2024

Contents

1 Introduction

The Moon, Earth's natural satellite, is believed to have formed as a result of a massive impact event on Earth. Its surface composition provides insights into the lunar crust, excavated materials, impact-generated formations, and the effects of space weathering due to interactions with solar wind.

X-ray mapping is one of the most effective methods to study the Moon's composition. By analyzing the spectra obtained, we can determine the elemental abundances on its surface. This report presents our methodology for analyzing, processing, and visualizing the extensive dataset collected by Chandrayaan-2. Key aspects covered include data acquisition, filtering, processing, and visualization. Additionally, sources of error and distinct compositional groups have been identified.

2 Problem Understanding

The problem involves analyzing data acquired by Chandrayaan-2 using two of its instruments: the X-ray Solar Monitor (XSM) and the Chandrayaan-2 Large Area Soft X-ray Spectrometer (CLASS). The XSM continuously records the solar spectrum at regular intervals, while the CLASS spectrometer, which faces the lunar surface, records the reflected solar spectrum whenever a solar flare is detected.

The reflected spectrum contains characteristic peaks corresponding to the primary X-ray fluorescence of elements present on the Moon's surface. By analyzing these peaks, the elemental abundances can be calculated using spectral fitting algorithms such as `x2abundance`.

In this problem, a slightly different approach is adopted: instead of calculating elemental abundances directly, a quantity called the *line intensity ratio* is determined. The line intensity ratio is defined as the ratio of the intensity of a specific element's spectral line to the intensity of a reference element's spectral line, expressed as:

$$R_{\text{line}} = \frac{I_{\text{element}}}{I_{\text{reference}}}$$

where R_{line} represents the line intensity ratio, I_{element} is the intensity of the spectral line of the element of interest, and $I_{\text{reference}}$ is the intensity of the spectral line of the reference element.

These line intensity ratios can provide valuable insights into the relative composition of elements on the lunar surface.

3 Review of Midterm Report

3.1 Question 1

Sensitivity of data smoothing to the number of FITS files(100) needs to be explored.This is comparable to flare time scale and hence there is possibility that fluctuation due to flare may be washed ?

In the previously specified dual-precision approach, there was ambiguity in the use of the number (100) for filtering spectrum points in the FITS file. In the updated pipeline for filtering, this issue has been addressed, and the iterative approach used previously has been verified for both precision and accuracy. The enhanced pipeline incorporates spectrum data from GOES (Geostationary Operational Environmental Satellite), cross-referenced with the time data of solar flares. This ensures a more robust and time-aligned analysis. Additionally, the solar flux threshold is calculated as part of this process, enabling the generation of an updated, filtered FITS file with improved accuracy and reliability.

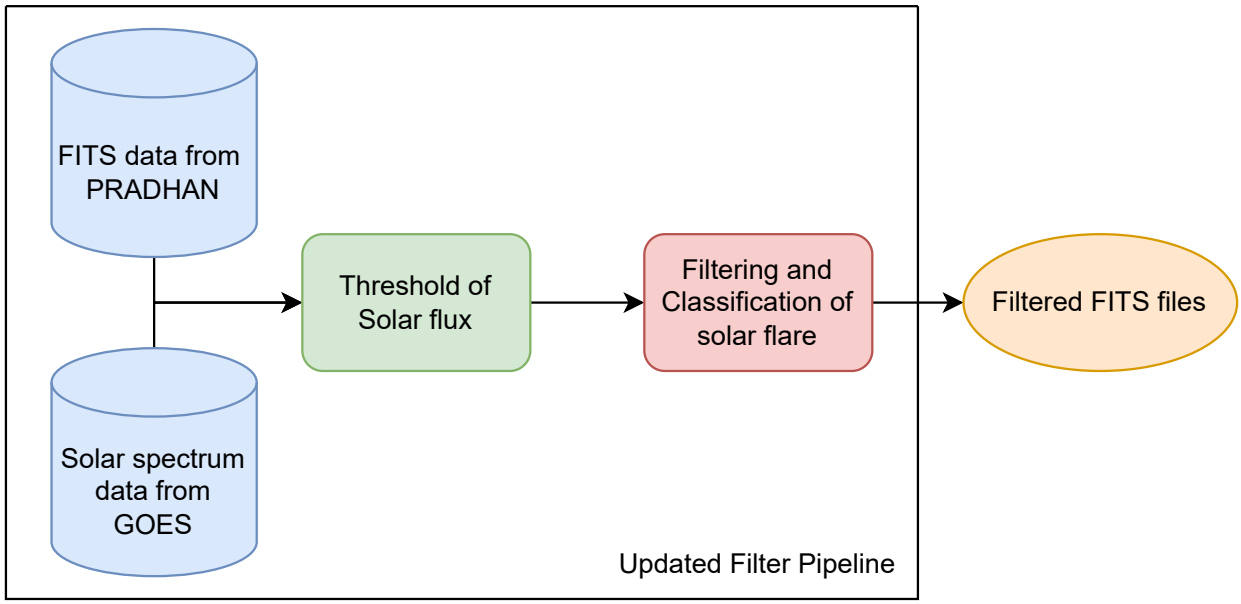

Figure 1: Filtering Process

3.2 Question 2

Applying x2abundance algorithm gives the elemental abundance directly thereby removing the need for the comparing the line-intensity ratios. Objective here is to see if the line-intensity ratios follow the trends that is expected in the elemental abundance ?

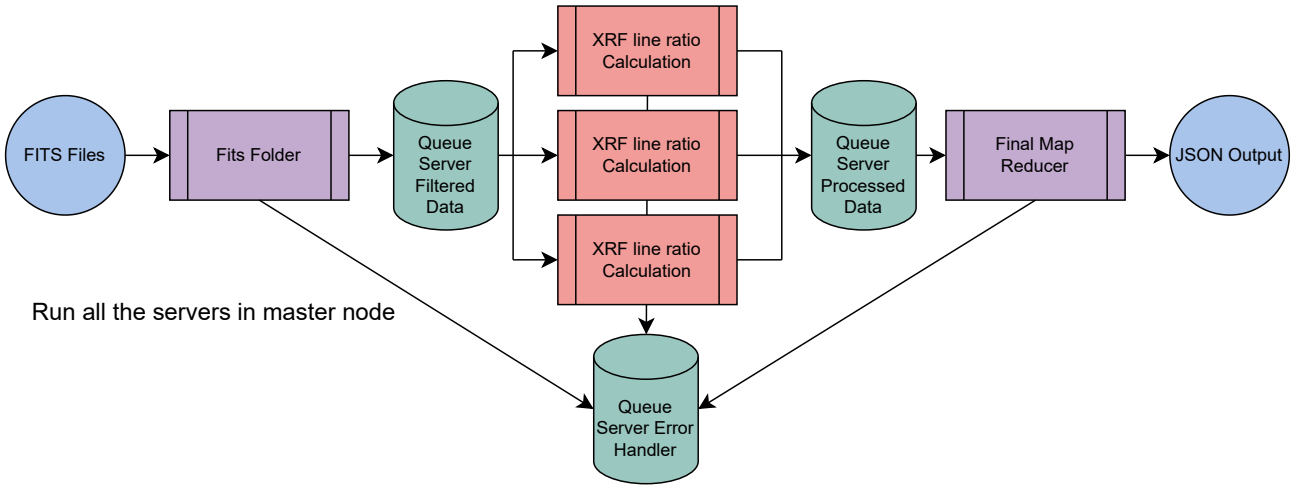

Figure 2: Line ratio architecture

Using XSPEC software and the X2abund model, we successfully analyzed the data collection process through FITS files, which also facilitated the verification of our results against XRF line ratio data and their interrelations. Furthermore, the previously developed high-performance computing architecture was utilized to calculate elemental line ratios , significantly enhancing the efficiency and reliability of the analysis.

Further this was retested with another algorithm further elaborated in the upcoming sections.

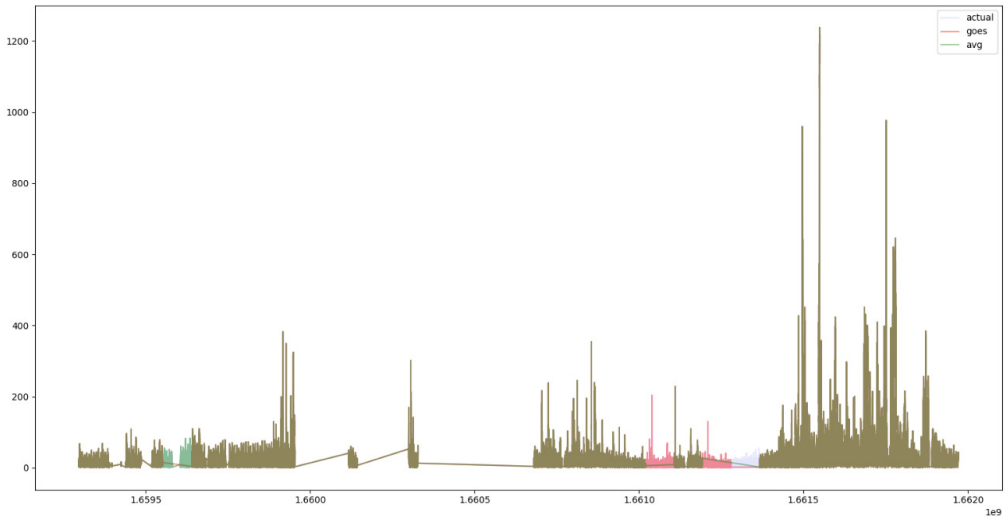

Figure 3: Comparison of the filtering method with GOES data.

Here in the above Figure one can see the two different filtering methods overlayed on one another. We can notice that both the methods produce the similar filtering results.

3.3 Question 3

Weighting criteria should depend on the quality of spectral fits and not on how recent the data is ?

The weighting criteria is not based on the recency of the data. Instead, it considers the possibility that two data points may not overlap with each other but overlap with the cell individually. To account for this, the accumulated Line Intensity Ratio (LIR) in each cell is calculated using the following formula:

$$\text{New Value} = (1 - \alpha) \cdot \text{Old Value} + \alpha \cdot \text{Current Value}$$

Here:

- α is the weighting factor (where $0 \leq \alpha \leq 1$).
- Old Value represents the previously accumulated value in the cell.
- Current Value is the new data point being incorporated into the cell.

This formula ensures that the accumulated LIR in each cell reflects a balance between historical data and newly added data, weighted by α.

4 Literature Review

4.1 CLASS Instrument Overview

The Chandrayaan-2 Large Area Soft X-ray Spectrometer (CLASS) uses 16 Swept Charge Devices (SCDs) to detect X-ray fluorescence photons emitted from the lunar surface due to solar X-ray interaction. CLASS covers a 12.5 km x 12.5 km ground footprint per observation.

Operating Seasons and Duration

- CLASS operates on the sunlit side of the lunar orbit.
- **Noon-Midnight Season:** CLASS functions for 59 minutes per orbit when the Sun's vector is parallel to the orbit plane.
- **Dawn-Dusk Season:** CLASS operates for the entire 118-minute orbit when the Sun's vector is perpendicular to the orbit plane.
- **Detection Limitation:** Lunar X-ray fluorescence (XRF) is only detected during solar flares.

4.2 Numerical Estimation of Lunar X-ray Emission for SELENE X-ray Spectrometer

Findings

[7] **Spatial Resolutions:**

- **Normal Solar Conditions:** Mg, Al, and Si spatial resolution is ¡90 km.
- **Active Solar Conditions:** Mg, Al, Si resolution is 20 km; Ca, Ti, and Fe resolution is 20 km (only during solar flares).

Element Detection:

- Mg, Al, and Si are detectable under normal and active solar conditions.
- Ca, Ti, and Fe are detectable only during solar flares.

Integration Times:

- Mg, Al, and Si require shorter integration times for high-resolution mapping.
- Ca, Ti, and Fe require longer integration times, especially under normal solar conditions.

Solar Activity Impact:

- Active solar conditions enhance detection and mapping with higher spatial resolution.
- Intense solar activity efficiently excites heavier elements like Ca, Ti, and Fe.

Modeling and Analysis

- Numerical simulations estimate X-ray spectra but exclude particle size and topography effects.
- Mass ratios and fluorescence photon count ratios normalized to Si exhibit nonlinear relationships, necessitating theoretical models for accurate composition analysis.

Rock Identification:

- XRS distinguishes six rock types (e.g., highland, ferroan anorthosite) with 10% accuracy in elemental intensity ratios.
- Mg/Si and Al/Si ratios distinguish highland rocks, while Ca, Ti, and Fe ratios identify basaltic rocks.

Research on Chang'E-2 X-ray Spectrometry Data

Excitation of X-Ray Fluorescence (XRF)

[4]

- XRF is proportional to incident solar X-ray intensity and lunar surface elemental abundance.

Conditions for Data Usability:

- **Solar Flare Intensity:** Solar flare must be above class C.
- **Saturation Threshold:** XRF and XSM count rates should not exceed 3300 counts per second.
- **Incident Angle Requirements:** Solar X-ray incident angle at the observed point must be $< 90°$, and XSM solar X-ray incident angle must be $< 60°$.

Temperature and Background Spectra:

- Background spectra show minimal temperature dependence and are similar for sunlit and dark sides during quiet solar conditions.

Validation of Elemental Abundances from Chandrayaan-1 Observations

x2abundance Model

[3] **Purpose:** Converts XRF line flux measurements into elemental abundances, incorporating matrix effects up to tertiary XRF.

Steps of Execution:

1. **Flux-Fraction Definition:**
$$f_i = \frac{I_i}{\sum_{i=1}^{n} I_i}$$
where I_i is the XRF line flux of the i^{th} element.
2. Generate matrices of elemental weight percentages and compute theoretical flux fractions.
3. Use the Weighted Least Squares method to derive abundances:
$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i^{\text{obs}} - f_i^{\text{comp}})^2}{\sigma_i^2}$$
4. Estimate errors in abundances by propagating flux fraction errors.

Lunar Surface Analysis Using x2abundance:

- Corrections for heterogeneity, particle size, and coherent scatter applied for lunar regolith analysis.
- Iterative refinement ensures consistency between observed and computed line fluxes.
- Abundance results validated against expected elemental compositions of the lunar regolith.

GOES Read Me

GOES XRS Functionality and Data Corrections

The GOES X-ray Sensor (XRS) measures solar X-ray flux in two wavelength bands (0.5–4 Å and 1–8 Å), crucial for solar flare detection. The data undergoes corrections for scaling, timing, and bandpass consistency across satellites to ensure accurate and reliable measurements.

Challenges and Enhancements

- Challenges like saturation during extreme flares and contamination by energetic particles are addressed with solutions like bandpass and scaling corrections.
- These enhancements enable precise solar flare analysis.

Access to Data

Both operational and science-quality data are made accessible, supporting accurate solar flare monitoring, research, and consistent comparisons across GOES satellites.

5 Methodology

5.1 XRF Line Detection

The accurate estimation of elemental composition requires a reliable database for X-ray interactions with matter. For this purpose, we utilize a freely available software library called xraylib, which provides spectral line data for all elements. This library is instrumental in identifying X-ray spectral lines.

The data recorded by CLASS is stored in the `*.fits` format and contains the spectrum (1 keV to 15 keV) with counts for each energy level. This data includes the flare spectrum reflected from the Moon, along with background noise. To correctly identify XRF lines, the background noise is first removed, and then count values are extracted at specific energy levels corresponding to the desired elements.

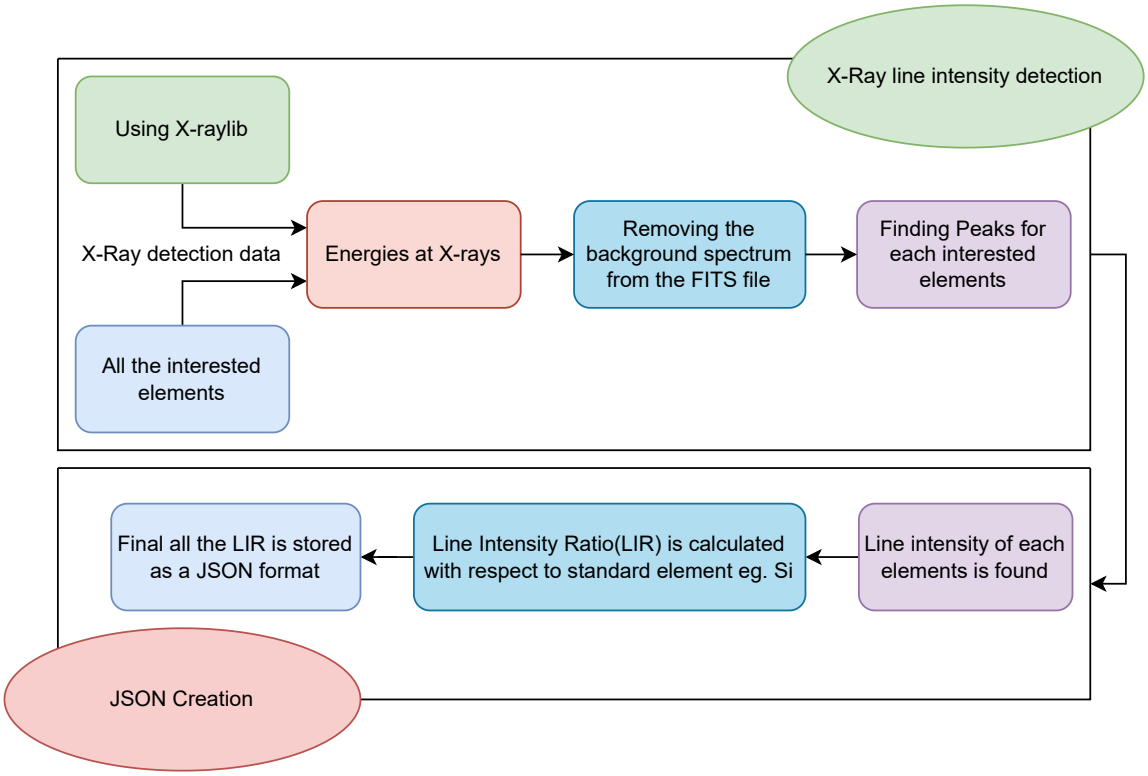

Figure 4: Line Ratio Calculation

5.2 XRF Line Ratio Estimation

To estimate line intensity ratios (LIR), the count values for specific energy levels, obtained from the `*.fits` files, are normalized against the reference element's counts. The line intensity ratio is defined as:

$$LIR = \frac{I(E_i)}{I(E_{\text{ref}})} = \frac{C(E_i)}{C(E_{\text{ref}})}$$

where $C(\cdot)$ represents the counts corresponding to a specific element, this is how the intensity is measured.

5.2.1 Estimation of Errors

The primary sources of error are:

- Residual background noise (evaluate the average variation in the noise spectrum over time).
- Variations in the spectrum due to changes in solar angle.
- Variations in the spectrum due to fluctuations in solar incidence strength.

The main source of noise in the system arises from background interference that may not be entirely removed during preprocessing. The line intensity ratio in the presence of noise is expressed as:

$$LIR = \frac{C(E_i)}{C(E_{\text{ref}})} = \frac{C_{\text{true}}(E_i) + N_i}{C_{\text{true}}(E_{\text{ref}}) + N_{\text{ref}}}$$

where $C_{\text{true}}(\cdot)$ refers to the true counts, and N_i, N_{ref} represent noise contributions.

Choosing a stronger solar flare can reduce the impact of noise because the relative contribution of noise to the overall signal becomes negligible. As the solar flare intensity increases, the line intensity ratio approaches the noiseless value:

$$\lim_{\text{flare}\to\text{stronger}} LIR(I_{\text{flare}}) = \frac{C_{\text{true}}(E_i)}{C_{\text{true}}(E_{\text{ref}})} \cdot \frac{1 + \frac{N_i}{C_{\text{true}}(E_i)}}{1 + \frac{N_{\text{ref}}}{C_{\text{true}}(E_{\text{ref}})}} \to LIR_{\text{noiseless}}$$

as $C(E) \to$ large with increasing solar flare intensity.

5.3 Determination of Appropriate Ratios

In order to determine the appropriate ratios, the following filtering steps have been applied:

- GOES data : The GOES data is used to calculate the time stamps of the solar flare, and from the `*.fits` files are filtered based on these time stamps.
- Strength of solar flare: The solar flare is characterised by the total solar flux in the duration of the event. The solar flare is classified in accordance to a logarithmic scale. Solar Flare

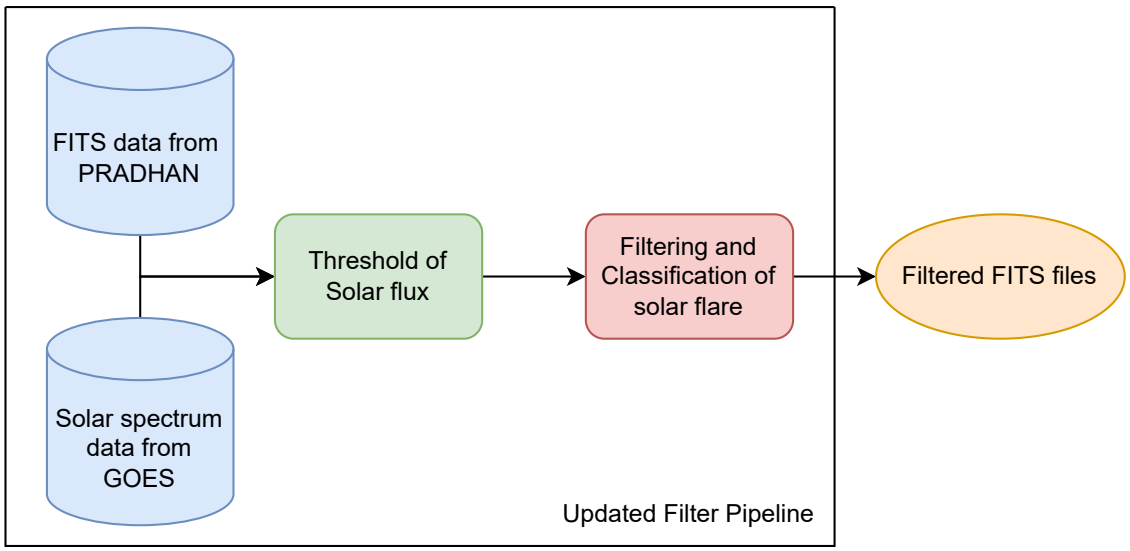

Figure 5: Filtering Process

Alternative approach for filtering was described in the mid-term report and was titled the "Dual Precision Filtering Method"

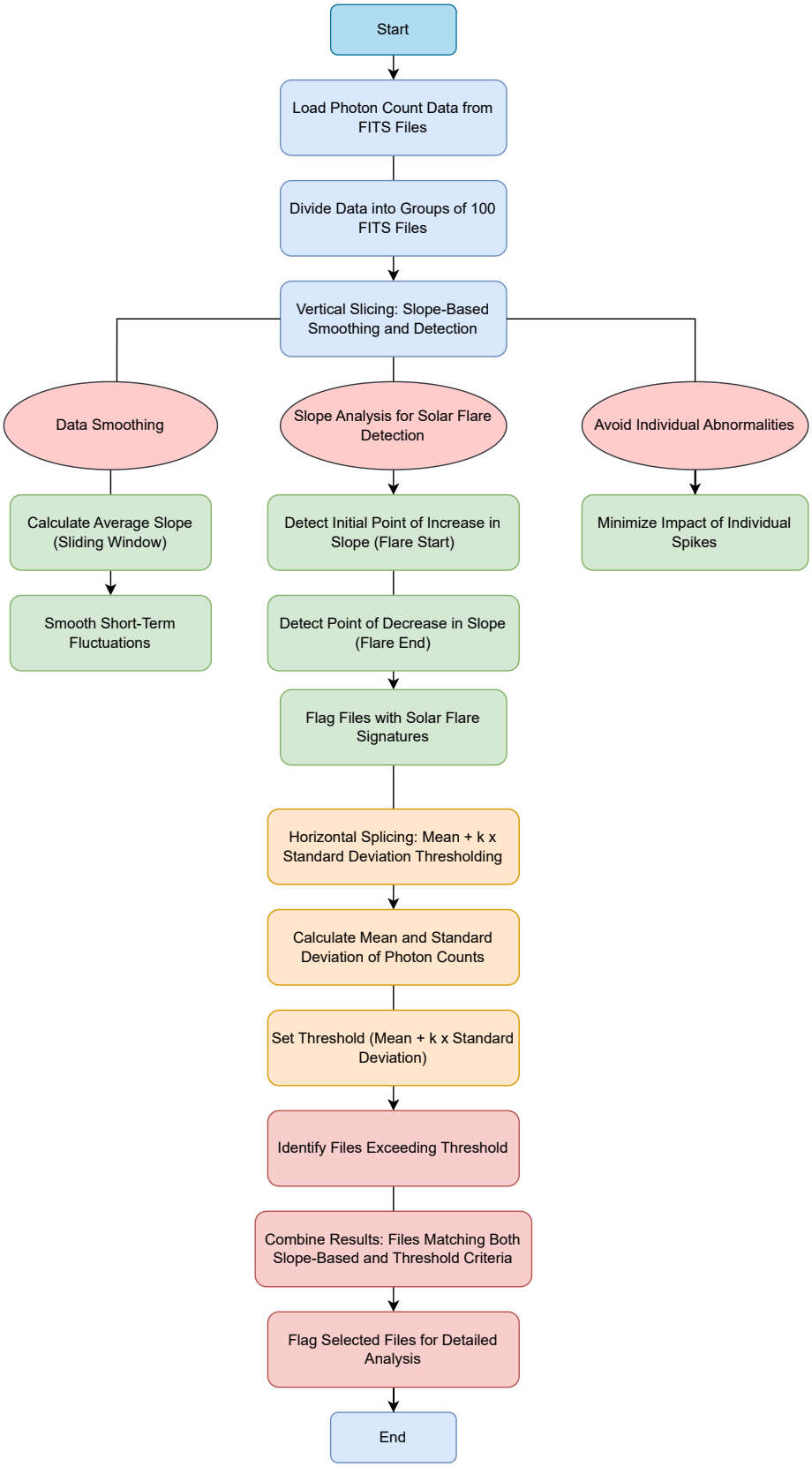

Figure 6: Algorithm for the Dual Precision Filtering Approach

5.4 Filtering the Required Datasets: Dual Precision Filtering Approach

The dataset filtering process focuses on extracting relevant photon count data from `*.fits` files, which represent X-ray fluorescence (XRF) measurements of the lunar surface. This data records metal emissions as photon counts over time, with the primary goal of identifying solar flare events that significantly influence these readings.

5.4.1 Vertical Slicing: Slope-Based Smoothing and Detection

The first stage of filtering, referred to as vertical slicing, involves analyzing photon counts to identify solar flare events through slope-based smoothing techniques. This approach operates as follows:

- **Data Smoothing**: The dataset is divided into groups of 100 `*.fits` files, and the average slope of photon counts within each group is calculated. This "sliding-window" method smooths short-term fluctuations in the data, revealing stable underlying patterns.
- **Slope Analysis for Solar Flare Detection**: Solar flare events are detected as sustained, significant changes in the slope of photon counts over time. Specifically:
  - The *start of a flare* is marked by a strict and noticeable increase in slope.
  - The *end of a flare* is identified by a subsequent strict decrease in slope.
- **Minimizing Noise from Abnormalities**: Averaging slopes over 100 files reduces the impact of isolated photon count anomalies, such as random spikes, and instead highlights broader, sustained patterns indicative of solar flares.

This vertical slicing step narrows the dataset by identifying `*.fits` files that likely contain solar flare events, enabling more focused analysis of metal composition under flare conditions.

5.4.2 Horizontal Splicing: Mean + $k \times$ Standard Deviation Thresholding

The second stage, horizontal splicing, refines the selection further using statistical thresholding:

- **Threshold Calculation**: The mean photon count and standard deviation are calculated across the entire dataset. A threshold is then established as:

  $$\text{Threshold} = \text{Mean} + k \times \text{Standard Deviation}$$

  Files with photon counts significantly exceeding this threshold are flagged as containing solar flare events.
- **Refinement of Selection**: This statistical thresholding is combined with the slope-based detection from vertical slicing. Only `*.fits` files meeting both conditions—sharp slope changes and photon counts exceeding the threshold—are selected for further analysis.

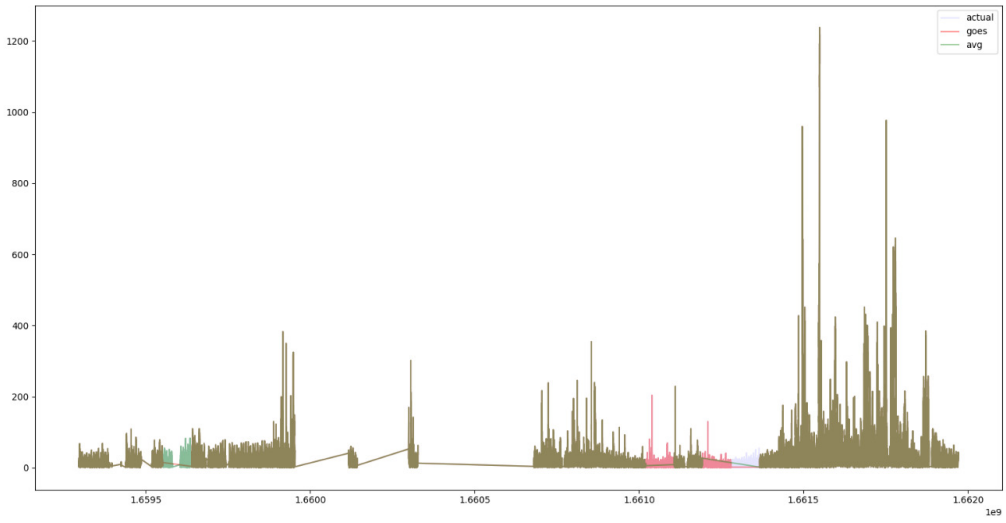

Figure 7: Comparison of the two filtering methods

5.4.3 Combined Filtering Method

The dual precision filtering approach, integrating vertical slicing and horizontal splicing, enhances the accuracy of solar flare detection. By cross-verifying `*.fits` files against both metrics, this method minimizes false positives and ensures that only the most relevant data is used for advanced analysis of the lunar surface's metal composition during solar flare events.

5.4.4 Comparing the two filtering methods

In the figure above, the two different filtering methods are overlaid, revealing that both approaches yield similar filtering outcomes.

The prediction of solar flares is effectively achieved using the dual precision algorithm. Initially, this method was based on a simple heuristic, which was later refined into a more accurate approach.

5.5 Moon Mapping Algorithm

This algorithm processes raw sensor data and visualizes it on a 2D map of the lunar surface. The primary goal is to quantify the contribution of each data point within specified grid cells, creating a systematic representation of data distribution across the Moon's surface.

The process begins by plotting the sensor data points on the 2D lunar map, where each point is overlaid based on its latitude and longitude coordinates. A grid is then drawn over the map, dividing the lunar surface into an array of cells. Each cell represents a specific area of the surface, for which the data contributions will be calculated.

For each cell, the algorithm calculates the contribution of sensor data points located within its boundaries. In cases where multiple data points fall within a cell, an exponential moving average (EMA) is applied to determine the overall contribution for that cell. The EMA provides a weighted average of the data points, giving more weight to recent or nearby data points. This allows the algorithm to better capture temporal and spatial relevance.

The EMA calculation is expressed as:

$$\text{EMA}_{\text{cell}} = (\text{Previous Contribution} \cdot a) + (\text{Current Contribution} \cdot (1 - a))$$

where a is the smoothing factor, which adjusts the influence of recent data points, depending on the desired level of emphasis on them.

This method ensures that data contributions within each cell are accurately aggregated, taking into account both the spatial density and intensity of the sensor readings. The resulting grid-based map offers a detailed, visual, and quantitative representation of data contributions across the lunar surface, supporting further analysis of spatial patterns and trends in sensor readings (Johnson et al., 2022).

5.5.1 Algorithm

When an element is selected on the frontend, the backend processes the request in Figure 8.

Input Data Parsing

1. Extract `quad` and `data`.
2. Compute minimum and maximum latitudes and longitudes for each quadrilateral:

$$\text{min\_lat} = \min_i(\text{quad}[i][\text{"lat"}]), \quad \text{max\_lat} = \max_i(\text{quad}[i][\text{"lat"}])$$

$$\text{min\_lon} = \min_i(\text{quad}[i][\text{"lon"}]), \quad \text{max\_lon} = \max_i(\text{quad}[i][\text{"lon"}])$$

3. Collect bounds:

$$\text{grid\_min\_lat} = \min(\text{min\_lat}), \quad \text{grid\_max\_lat} = \max(\text{max\_lat})$$

$$\text{grid\_min\_lon} = \min(\text{min\_lon}), \quad \text{grid\_max\_lon} = \max(\text{max\_lon})$$

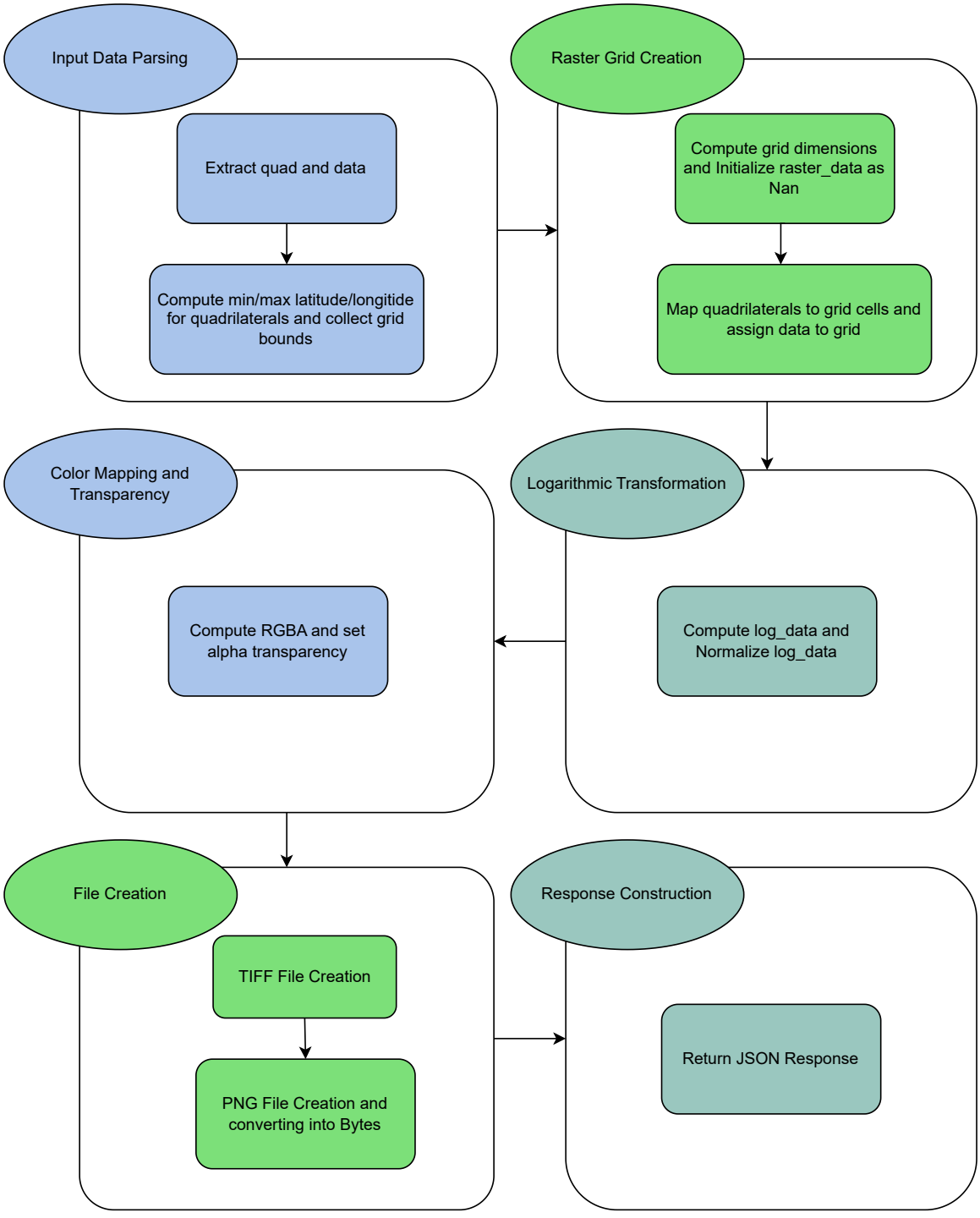

Figure 8: Moon Mapping

Raster Grid Creation

1. Compute grid dimensions:

$$\text{height} = \left\lceil \frac{\text{grid\_max\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rceil, \quad \text{width} = \left\lceil \frac{\text{grid\_max\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rceil$$

2. Initialize `raster_data` as `NaN`.

3. Map quadrilaterals to grid cells:

$$\text{min\_lat\_idx} = \left\lfloor \frac{\text{min\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rfloor, \quad \text{max\_lat\_idx} = \left\lceil \frac{\text{max\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rceil$$

$$\text{min\_lon\_idx} = \left\lfloor \frac{\text{min\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rfloor, \quad \text{max\_lon\_idx} = \left\lceil \frac{\text{max\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rceil$$

4. Assign data to grid.

Logarithmic Transformation

$$\text{log\_data}_{i,j} = \log(1 + r_{i,j})$$

$$\text{norm\_data}_{i,j} = \frac{\text{log\_data}_{i,j} - \min(\text{log\_data})}{\max(\text{log\_data}) - \min(\text{log\_data})}$$

Color Mapping and Transparency

$$\text{RGBA}_{i,j} = C_e(\text{norm\_data}_{i,j})$$

$$\text{alpha}_{i,j} = \begin{cases} 0 & \text{if raster\_data}_{i,j} \text{ is NaN} \\ 255 & \text{otherwise} \end{cases}$$

TIFF File Creation

$$T = \begin{bmatrix} \text{grid\_resolution} & 0 & \text{grid\_min\_lon} \\ 0 & -\text{grid\_resolution} & \text{grid\_max\_lat} \end{bmatrix}$$

PNG File Creation

$$\text{base64\_image} = \text{Base64Encode}(\text{PNG\_bytes})$$

Response Construction

The server returns a JSON response containing:

- Elemental metadata (name, color, gradient limits).
- Base64-encoded PNG images for visualization.

5.5.2 Explanation

Input Data Parsing

The input is a JSON object containing a set of quadrilaterals (`quad`) and associated elemental data values (`data`). Each quadrilateral represents a region defined by four corners in latitude and longitude coordinates.

- Extract `quad` (coordinates of the four corners) and `data` (elemental values) for each entry.
- Compute the minimum and maximum latitude and longitude for each quadrilateral:

$$\text{min\_lat} = \min_i(\text{quad}[i][\text{"lat"}]), \quad \text{max\_lat} = \max_i(\text{quad}[i][\text{"lat"}])$$
$$\text{min\_lon} = \min_i(\text{quad}[i][\text{"lon"}]), \quad \text{max\_lon} = \max_i(\text{quad}[i][\text{"lon"}])$$

- Collect the bounds across all quadrilaterals to determine the overall spatial extent:

$$\text{grid\_min\_lat} = \min(\text{min\_lat}), \quad \text{grid\_max\_lat} = \max(\text{max\_lat})$$
$$\text{grid\_min\_lon} = \min(\text{min\_lon}), \quad \text{grid\_max\_lon} = \max(\text{max\_lon})$$

This provides the bounding box for the entire dataset, which is essential for creating a consistent raster grid.

Raster Grid Creation

A raster grid is a 2D array that represents the spatial data over a defined region. Each cell corresponds to a specific geographic area (determined when the grid resolution has the value 0.1).

1. Compute the dimensions of the raster grid:

$$\text{height} = \left\lceil \frac{\text{grid\_max\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rceil, \quad \text{width} = \left\lceil \frac{\text{grid\_max\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rceil$$

2. Initialize a 2D array (`raster_data`) for each element, filled with `NaN` to indicate no data.
3. Map each quadrilateral's data values to the corresponding grid cells:
   - For a quadrilateral with bounds $(\text{min\_lat}, \text{min\_lon})$ to $(\text{max\_lat}, \text{max\_lon})$, compute the grid cell indices:

$$\text{min\_lat\_idx} = \left\lfloor \frac{\text{min\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rfloor, \quad \text{max\_lat\_idx} = \left\lceil \frac{\text{max\_lat} - \text{grid\_min\_lat}}{\text{grid\_resolution}} \right\rceil$$
$$\text{min\_lon\_idx} = \left\lfloor \frac{\text{min\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rfloor, \quad \text{max\_lon\_idx} = \left\lceil \frac{\text{max\_lon} - \text{grid\_min\_lon}}{\text{grid\_resolution}} \right\rceil$$

   - Fill the corresponding grid cells with the elemental value.

This step converts irregular spatial data into a uniform grid format suitable for visualization and analysis.

Logarithmic Transformation

To emphasize variations in elemental data, logarithmic scaling is applied. The log transformation is especially useful when data spans several orders of magnitude, as it compresses high values and expands lower values.

Given the raster data $r_{i,j}$, the transformation is:

$$\text{log\_data}_{i,j} = \log(1 + r_{i,j})$$

Adding 1 ensures that zero values are mapped correctly ($\log(1) = 0$).

The logarithmic data is then normalized to the range $[0, 1]$:

$$\text{norm\_data}_{i,j} = \frac{\text{log\_data}_{i,j} - \min(\text{log\_data})}{\max(\text{log\_data}) - \min(\text{log\_data})}$$

Color Mapping and Transparency

Each element is visualized using a predefined colormap. The normalized data values are mapped to RGBA (Red, Green, Blue, Alpha) colors.

1. Use the colormap function C_e for the given element e:

$$\text{RGBA}_{i,j} = C_e(\text{norm\_data}_{i,j})$$

2. Handle transparency by setting the alpha channel to zero for grid cells with `NaN`:

$$\text{alpha}_{i,j} = \begin{cases} 0 & \text{if raster\_data}_{i,j} \text{ is NaN} \\ 255 & \text{otherwise} \end{cases}$$

This step ensures that invalid or missing data is visually excluded.

File Generation

TIFF File Creation

Each grid is saved as a TIFF file with georeferencing metadata:

- The transformation matrix T defines the relationship between pixel indices and geographic coordinates:

$$T = \begin{bmatrix} \text{grid\_resolution} & 0 & \text{grid\_min\_lon} \\ 0 & -\text{grid\_resolution} & \text{grid\_max\_lat} \end{bmatrix}$$

PNG File Creation

The RGBA image is saved as a PNG file using the Python `Pillow` library.

Base64 Encoding

PNG files are converted to a byte array and encoded in Base64 format for transmission to the front end:

$$\text{base64\_image} = \text{Base64Encode}(\text{PNG\_bytes})$$

Response Construction

The server returns a JSON response containing:

- Elemental metadata (name, color, gradient limits).
- Base64-encoded PNG images for visualization.

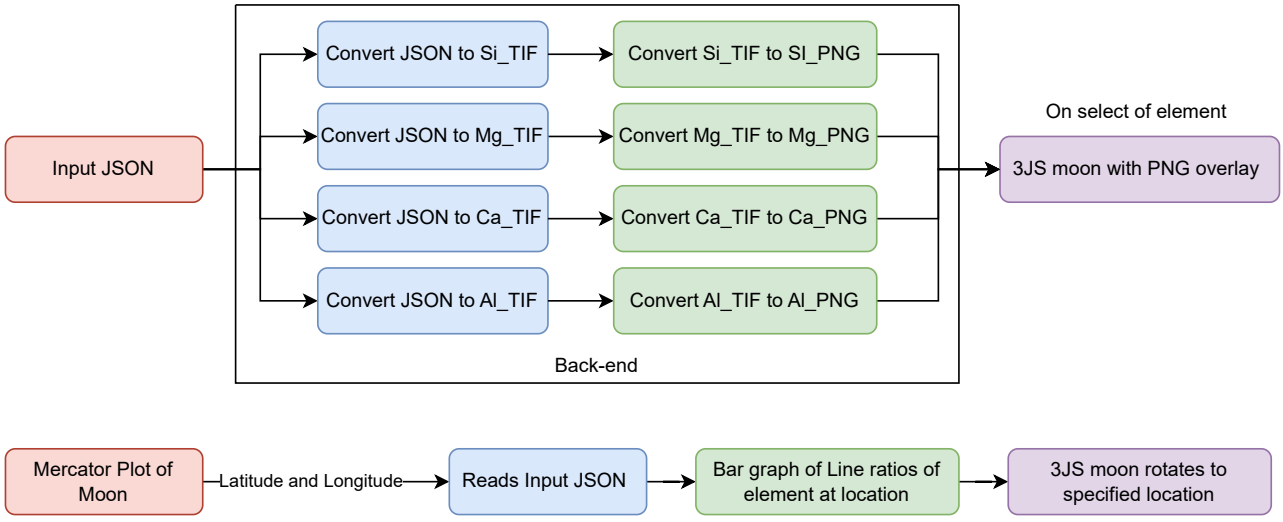

Figure 9: Website Backend

5.6 Moon Mapping Website

1. The complete front end is developed using Svelte. The website accepts a JSON file as input. The JSON is processed in the backend, where the following operations are performed:
    - A TIFF file is created using the discussed methodology.
    - The TIFF file is converted to a PNG image, which is overlaid on a 3D moon model created using Three.JS (3JS).
2. A mini-map is displayed at the bottom-right corner of the website, using a Mercator projection of the moon. Key features of the mini-map include:
    - A tracker that allows users to point to a specific location.

- The latitude and longitude from the tracker are used to fetch elemental line ratio data from the input JSON at the specified location.
- These line ratio values for each element are displayed as a bar graph in the top-right corner of the website for the selected location.
- The 3D moon model rotates to highlight the current location.

3. Additional features of the website include:
   - A zoom slider to adjust the zoom level on the 3D moon model.
   - A gradient scale for each TIFF file, indicating the minimum and maximum values of the line ratios of elements across the moon's surface.

6 Analysis of Processed Results

6.1 Compositional Groups

The lunar surface can be categorized into distinct compositional groups based on the elemental abundances of various rock types. These groups are classified as follows:

Rock Type	**Key Elemental Abundances (Approximate)**
Basaltic Rocks	Iron (Fe): 10-20%, Magnesium (Mg): 5-15%, Calcium (Ca): 10-12%, Aluminum (Al): 5-8%, Silicon (Si): 20-25%
Anorthositic Rocks	Aluminum (Al): 20-25%, Calcium (Ca): 10-15%, Silicon (Si): 25-30%, Iron (Fe): 1-3%, Magnesium (Mg): 1-3%
Breccias	Varies: Iron (Fe): 5-15%, Magnesium (Mg): 5-10%, Aluminum (Al): 10-20%, Silicon (Si): 20-30%, with traces of other elements depending on components
Regolith	Silicon (Si): 20-25%, Aluminum (Al): 10-15%, Iron (Fe): 5-15%, Magnesium (Mg): 5-10%, Titanium (Ti): 1-5%

Table 1: Elemental Abundances in Lunar Rock Types

The compositional groups are divided into four main categories to simplify the classification process. These four groups, outlined in the table above, represent the primary types of rocks found on the Moon's surface.

To classify these rocks, a decision tree algorithm can be utilized. The line intensity ratios (LIR) are approximately proportional to the ratio of the elemental densities and the X-ray interaction coefficients. The relationship can be expressed as:

$$LIR = \frac{N_1 A_1}{N_2 A_2} \Rightarrow \log(LIR) = \log\left(\frac{W_1}{W_2}\right) + \log(\text{scaling constants})$$

where W_1 and W_2 represent the weights of the elements being analyzed.
A decision tree algorithm can then be constructed based on these ratios.

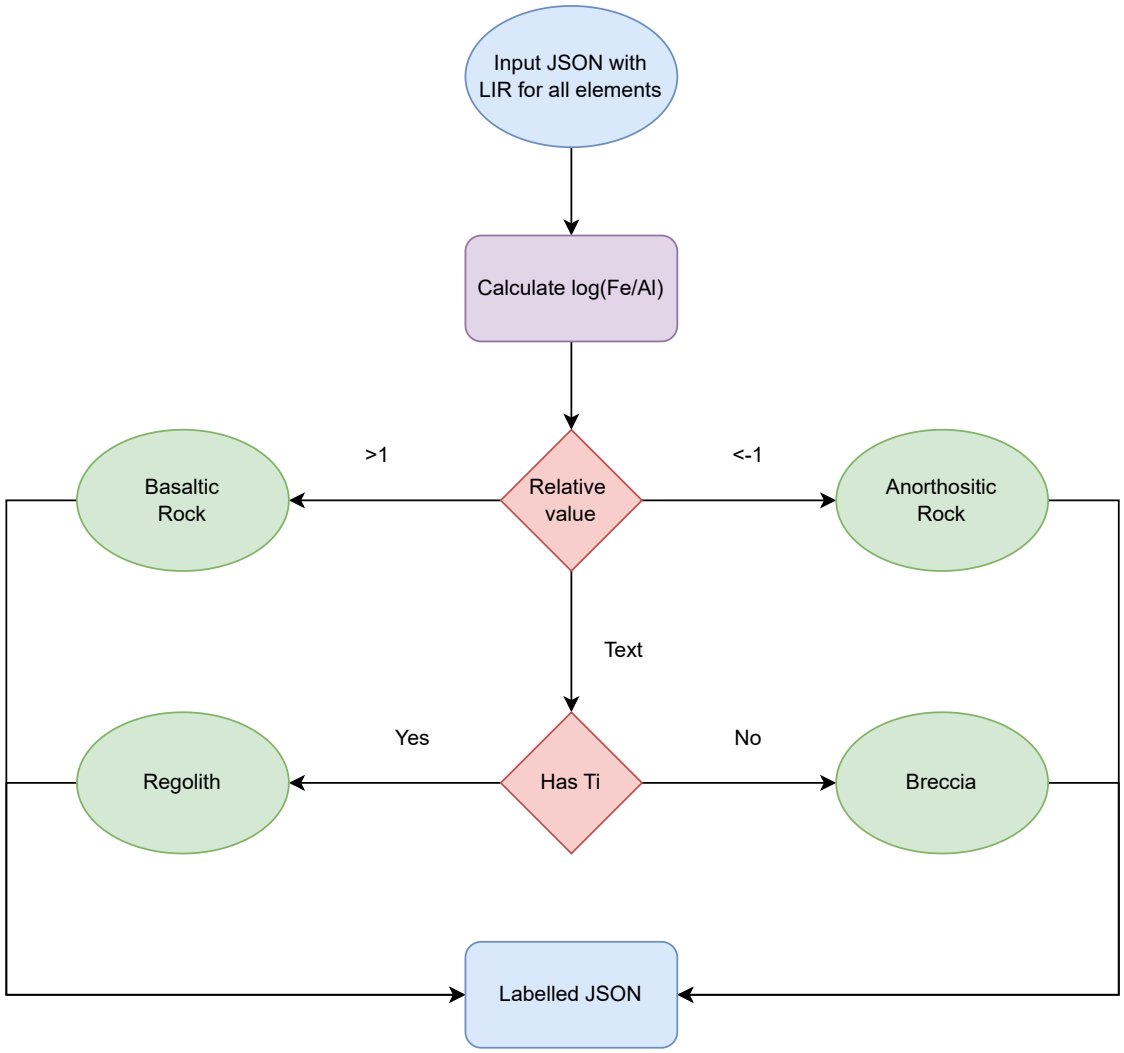

Figure 10: Decision Tree

The first step involves comparing the ratios of iron and aluminum. A low iron-to-aluminum ratio indicates an anorthositic rock, as it contains a relatively small amount of iron. On the other hand, a high ratio suggests the rock is basaltic.

When the iron-to-aluminum ratio is comparable, the decision tree further distinguishes between basaltic and breccia rocks by checking for the presence of titanium. If titanium is present, the data point is classified as either a breccia or regolith based on the titanium content.

In summary, the classification process involves:

- Comparing the iron and aluminum ratios to distinguish between basaltic and anorthositic rocks.
- Evaluating the presence of titanium to differentiate breccia and regolith.

This systematic approach allows for accurate classification of lunar rock types based on elemental abundances and line intensity ratios.

7 Best ratios to use and visualization of data on a lunar map

The optimal ratios to use for the lunar base map will be determined based on the entire data pipeline. Initially, the data undergoes filtering based on solar flare timing and intensity. In the final interactive map, the tiling process aggregates the data by using the average value, as described in the moon mapping algorithm.

Alternatively, an approach based on **spatial locality** can be applied. This concept assumes that the Line Intensity Ratios (LIRs) from overlapping data points should be consistent, as the content in the same area is being sampled twice under different conditions.

Two methods can be employed to determine the best ratios for the lunar base map:

- Selecting the data with the **higher solar flare flux**.
- Calculating the mean and standard deviation of the data and then plotting the mean.

8 Lunar Base Maps

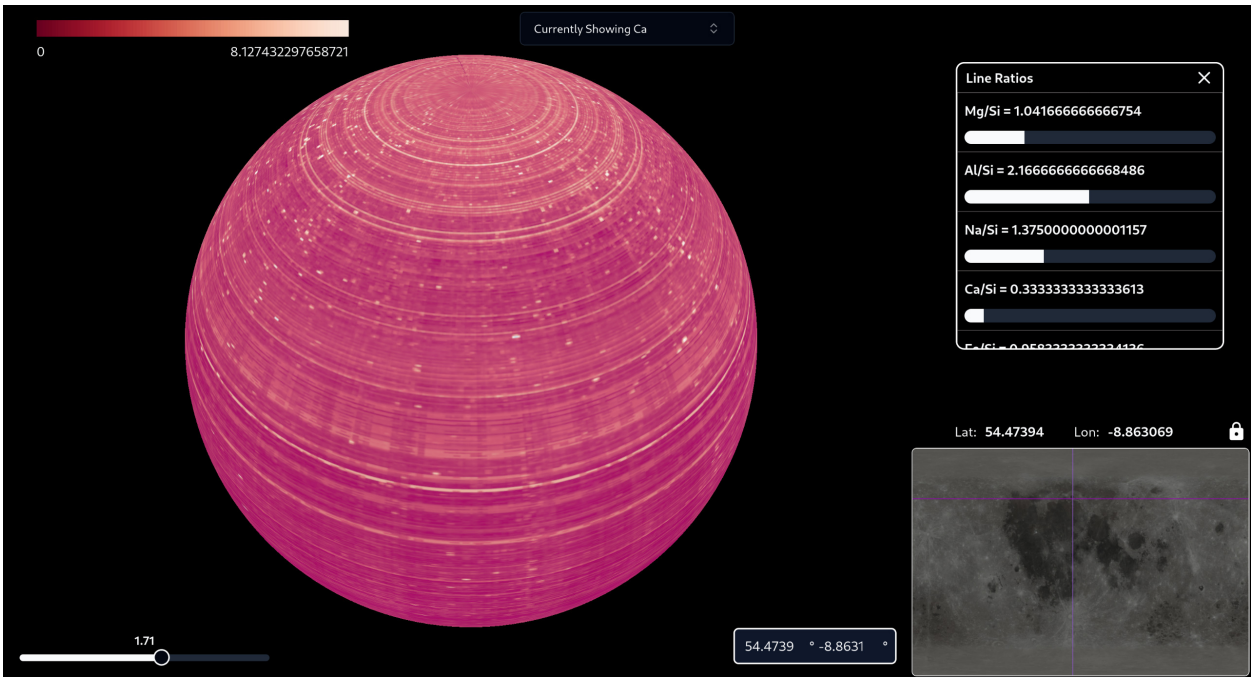

Figure 11: UI

Initial Data Upload

The website interface begins with an option to upload a JSON file. The JSON structure adheres to the following format:

- **Root Structure:** A dictionary with a key `data`, which contains a list of dictionaries.
- **Each dictionary contains:**
  - `quad`: The latitude and longitude of the quadrilateral generated by the Moon Mapper algorithm.
  - `data`: A list of line ratios corresponding to various elements being plotted.

Frontend Features

Once the JSON data is uploaded, it is processed by a Flask backend and loaded into a Three.js-based frontend. The UI includes the following features:

Customizable Globe

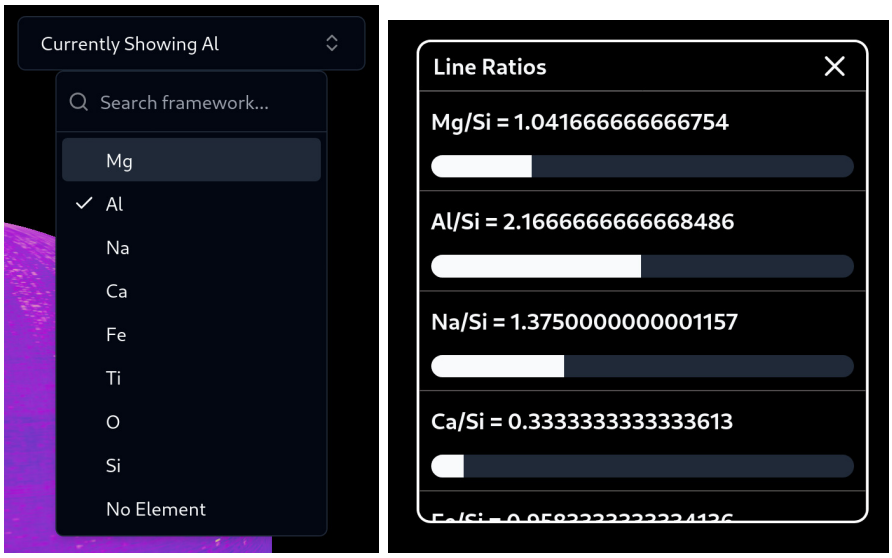

Figure 12: Drop Down Menu(Left) and Current line ratios of element(Right)

- A 3D globe is displayed, customizable with options to select various elements from a dropdown menu.
- A gradient section is present to visualize the element line ratios using color gradients.

Mini Map

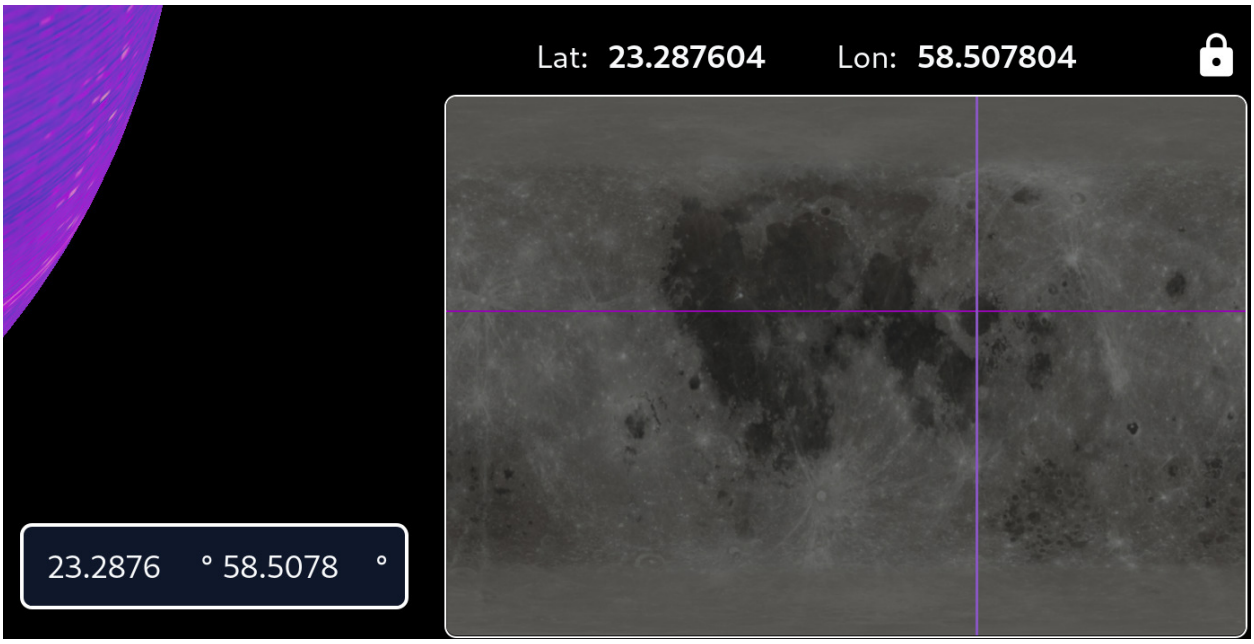

Figure 13: Mini map

- A dynamic interactive mini map is located at the bottom right corner of the screen.
- Clicking on the mini map:
  - Stops the globe's autorotation.
  - Spins the globe to the clicked latitude and longitude.
  - Opens a dialog box displaying the elemental ratios at the selected point on the Moon.
- Closing the dialog box restarts the globe's autorotation.

Zoom Control

Figure 14: Zoom

- A zoom slider is available at the bottom left of the screen, allowing users to zoom in and out of the globe for better observation.

Manual Latitude and Longitude Input

- Located beside the mini map, this input field allows users to manually enter latitude and longitude values, which are linked to the globe for navigation.

9 Complete Pipeline

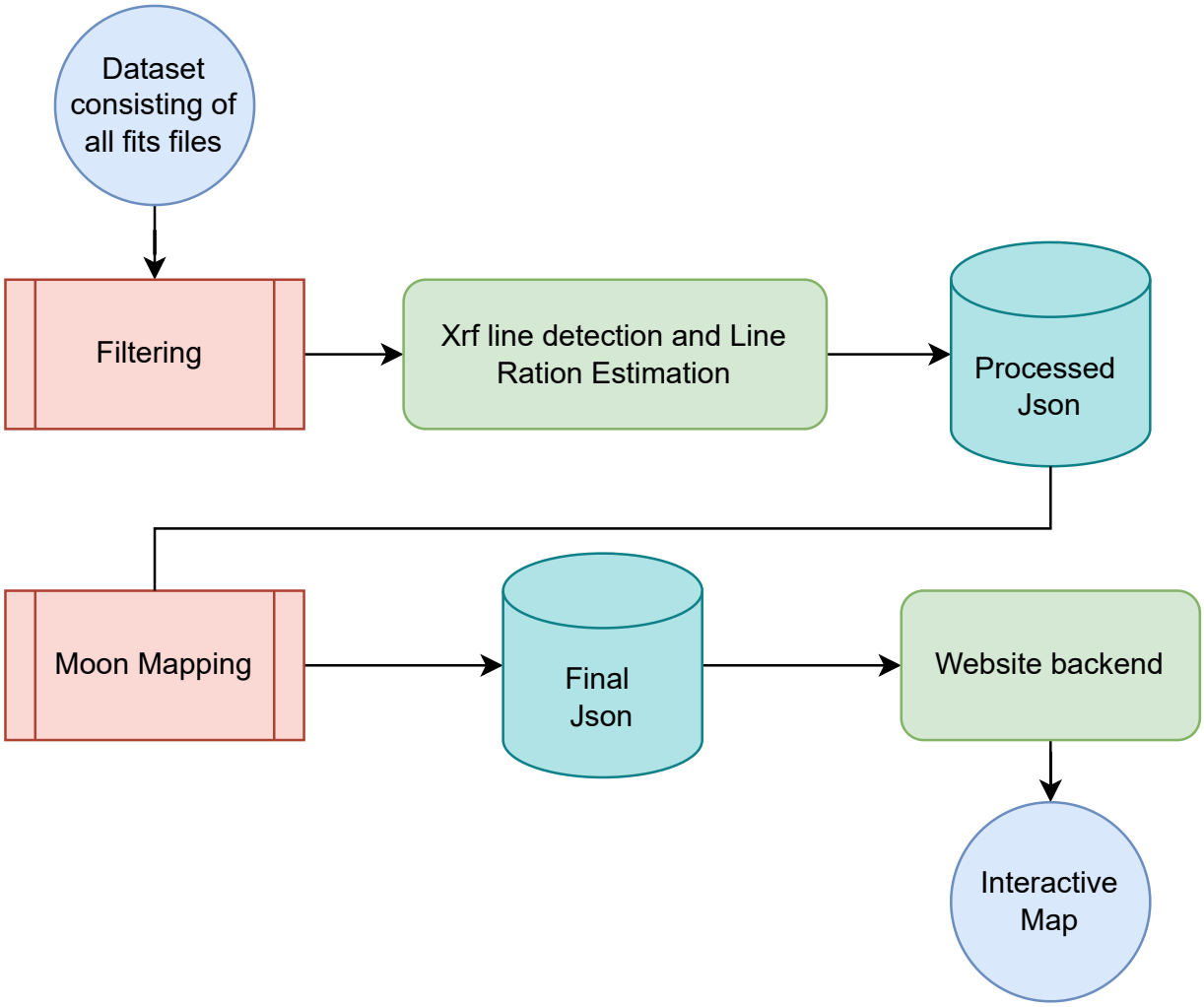

Figure 15: Final Data Processing Pipeline

10 What We Have Achieved

10.1 Subpixel Resolution

The original resolution of the *Chandrayaan-2 Large Area Soft X-ray Spectrometer (CLASS)* was 12.5 km $\times$ 12.5 km, equating to an area of 156.25 km^2 per pixel. Through our advancements, we achieved subpixel resolution of approximately 37 km^2 per pixel. This significant improvement was made possible by dividing the lunar surface into a grid of equal-area cells and projecting each FITS file onto the grid. By utilizing the overlap of the readings from multiple files, we were able to calculate individual cell values with higher precision. The code for this process was written in *Rust* and distributed among a cluster of nodes, enabling us to process data spanning from 2021-04-01 to 2024-07-31 in under 8 hours.

10.2 Data Processing Computer Architecture

We developed a distributed Data Processing Computer Architecture to handle FITS files with high efficiency. This system, utilizing multiple laptops connected locally, was designed for parallel processing, real-time error handling, and robust output generation. It provided minimal downtime and high accuracy even for extensive datasets.

10.3 FITS File Filtering

- Implemented slope-based analysis across 100 files to detect stable patterns while avoiding transient noise and spikes.
- Applied threshold filtering using the formula:

$$\text{Threshold} = \text{Mean} + k \cdot \text{Standard Deviation},$$

  where k determined the filter's sensitivity to anomalies.
- Cross-verified anomalies with time-stamped solar flare data from the *Geostationary Operational Environmental Satellite (GOES)* to enhance detection accuracy.
- Resolved ambiguities in dual-precision approaches by verifying calculations against real-time solar flux data.

10.4 Parallelized Moon Mapping and Line Ratio Calculations

- Developed a multi-threaded program in *Rust* for parallel computation, leveraging its concurrency model for high performance.
- Processed 2,500 filtered data chunks, each containing 1,000 data points, via a *Queue Server* implemented using SQLite.
- Calculated elemental line ratios for key elements (e.g., Si, Mg, Al, Ca, O, Ti, Fe, Na) with structured results stored in a database exceeding 10GB.

- Future file processing is handled by the the moon mapper automatically and its flow diagram is given in Figure 15.

10.5 Final Map Reduction

The processed data represented maps of the lunar surface for each job chunk. To generate a comprehensive final map:

- Averaged values from all overlapping grid cells across job chunks.
- Completed the final reduction in under 55 seconds, resulting in a single, high-resolution JSON file of over 500MB.

Summary: These achievements mark a significant leap in lunar surface analysis under solar flare conditions, providing high-resolution data and efficient processing capabilities. This pipeline ensures robust and reliable outputs suitable for advanced scientific research.

11 Challenges and Lessons Learned

11.1 Challenges Faced

Data Quality and Noise:

- Handling noisy XRF spectra due to instrument limitations or environmental interference.
- Difficulty in distinguishing between genuine signals and background noise in lunar surface data.

High-Resolution Mapping:

- Managing the immense computational demand for processing large datasets at high resolutions.
- Overcoming the limitations of interpolation techniques for sub-pixel accuracy.

Spectroscopic Analysis:

- Identifying and analyzing subtle variations in elemental abundance from XRF data.
- Ensuring calibration accuracy across different instruments or datasets.

Integration and Visualization:

- Challenges in merging data from multiple sources while maintaining consistency.
- Designing clear, interpretable visualizations to represent complex datasets effectively, like rendering the moon data on the website, due to large amounts of data.

Domain Knowledge:

- Bridging the gap between astronomical concepts and technical implementation.

11.2 Lessons Learned

Importance of Preprocessing:

- Rigorous data cleaning and normalization are critical for reliable results in spectroscopy and mapping.

Effective Planning:

- Breaking down complex goals (e.g., sub-pixel mapping) into manageable tasks ensures steady progress.

Iterative Testing:

- Incremental testing and debugging prevent compounded errors and save time in the long run.

Team Collaboration and Feedback:

- Engaging with peers or mentors provides diverse perspectives that enhance problem-solving approaches.

Resource Optimization:

- Utilizing optimized algorithms (e.g., multithreading between 10-15 laptops) and hardware accelerations for handling large datasets.

Documentation and Organization:

- Keeping detailed notes and organized codebases improves project scalability and reproducibility.

Adaptability:

- Being flexible with methodologies, adapting to new tools, or revisiting hypotheses strengthens outcomes.

12 Future Scope and Recommendations

12.1 Future Scope

Geotechnical Applications on Earth: The line ratio-abundance method can be effectively utilized to analyze elemental abundances and soil composition across various terrains on Earth. This approach provides valuable insights into the structural and compositional characteristics of the soil, aiding in the study of geotechnical properties and resource exploration.

Search for Water on the Moon: Given that water is primarily composed of hydrogen and oxygen, the line ratio methodology, when coupled with gamma-ray spectroscopy instead of X-ray spectroscopy, can precisely determine the abundance of hydrogen- and oxygen-rich areas on the lunar surface. This information would significantly contribute to identifying potential water reservoirs and advancing future lunar exploration missions.

12.2 Recommendations

Filtering: The current approach employs a dual-precision system for filtering data, which is computationally intensive as it is compared with GOES (Geostationary Operational Environmental Satellite) data. Simplifying this process with a more robust filtering mechanism strictly aligned with the available GOES data would enhance the reliability and accuracy of the filtering process.

Enhanced Plotting Techniques: The existing system plots line ratios for individual elements. Expanding this methodology to include the overlay of multiple TIF files for different elements would enable a more comprehensive visualization of relative elemental abundances. Such an approach could provide deeper insights into elemental distributions across studied regions.

User Interface Enhancements: The user interface can be enriched with additional features, such as the integration of multiple JSON data files as input for a combined analysis. Furthermore, advanced visualization techniques, including 3D projections of line ratios mapped onto a 3D lunar surface, could replace simpler bar graph visualizations. These enhancements would provide a more intuitive understanding of the spatial distribution of elements.

Error Reduction Approaches: The current method averages line ratio values for specific locations (latitude and longitude) when multiple datasets from continuous yearly readings are available. To improve accuracy and precision, alternative approximation techniques can be explored. Methods such as weighted averaging, statistical smoothing, or machine learning-based prediction models could be employed to mitigate data variability and enhance the reliability of the results.

13 Conclusion

In conclusion, the development of high-resolution elemental maps using CLASS data represents a breakthrough in lunar surface mapping, enabling more refined compositional analysis. The methodology of deriving intensity ratios and mapping these onto a lunar albedo base map provides a novel approach to understanding surface heterogeneity. This report outlines the steps taken to process CLASS spectra, detect elemental signatures, and create composition maps. The resulting maps, along with the sub-pixel resolution projections, offer valuable insights for geochemical studies and potential lunar resource exploration. This work not only

enhances the scientific knowledge of the Moon's surface but also contributes significantly to ISRO's goals in planetary exploration and resource mapping.

References

[1] L Alha, Juhani Huovelin, K Nygård, H Andersson, Eero Esko, CJ Howe, BJ Kellett, S Narendranath, BJ Maddison, IA Crawford, et al. Ground calibration of the chandrayaan-1 x-ray solar monitor (xsm). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 607(3):544–553, 2009.

[2] KA Arnaud. Xspec: the first ten years. In *Astronomical Data Analysis Software and Systems V, ASP Conference Series, Vol. 101, 1996, George H. Jacoby and Jeannette Barnes, eds., p. 17.*, volume 101, page 17, 1996.

[3] PS Athiray, S Narendranath, P Sreekumar, SK Dash, and BRS Babu. Validation of methodology to derive elemental abundances from x-ray observations on chandrayaan-1. *Planetary and Space Science*, 75:188–194, 2013.

[4] Chao Ban, Yongchun Zheng, Yongchao Zhu, Feng Zhang, Lin Xu, and Yongliao Zou. Research on the inversion of elemental abundances from chang'e-2 x-ray spectrometry data. *Chinese Journal of Geochemistry*, 33:289–299, 2014.

[5] John W Criss and LS Birks. Calculation methods for fluorescent x-ray spectrometry. empirical coefficients versus fundamental parameters. *Analytical Chemistry*, 40(7):1080–1086, 1968.

[6] S Narendranath, Netra S Pillai, M Bhatt, K Vadodariya, Radhakrishna Vatedka, Srikar P Tadepalli, A Sarwade, A Tyagi, and V Sharan. Lunar elemental abundances as derived from chandrayaan-2. *Icarus*, 410:115898, 2024.

[7] Kazunori Ogawa, Tatsuaki Okada, Kei Shirai, and Manabu Kato. Numerical estimation of lunar x-ray emission for x-ray spectrometer onboard selene. *Earth, planets and space*, 60:283–292, 2008.

[8] Netra S Pillai, S Narendranath, K Vadodariya, Srikar P Tadepalli, V Radhakrishna, Anurag Tyagi, Reena Yadav, Brajpal Singh, Vaishali Sharan, PS Athiray, et al. Chandrayaan-2 large area soft x-ray spectrometer (class): calibration, in-flight performance and first results. *Icarus*, 363:114436, 2021.

[9] V Radhakrishna, A Tyagi, S Narendranath, Koushal Vadodariya, Reena Yadav, Brajpal Singh, G Balaji, Neeraj Satya, Akash Shetty, HN Suresha Kumar, et al. Chandrayaan-2 large area soft x-ray spectrometer. *Current Science*, 118(2):219–225, 2020.