

ISRO End-Term Evaluation Report (Team_73)

High-Resolution Elemental Mapping of Lunar Surface

Team_73

Project Description:

The aim is to develop a revolutionary approach to map lunar surface elemental composition at unprecedented kilometer-scale resolution. By ingeniously analyzing X-ray fluorescence line intensity ratios, we've transformed raw spectral data into a dynamic, interactive visualization that reveals the Moon's intricate geochemical landscape. Our solution not only advances lunar exploration but opens new frontiers in understanding the Moon's geological evolution and potential resource utilization.

Key Objectives:

- To map the ratios lunar base map
- Visualization of data on lunar map by using best ratios
- To make a subpixel resolution map

CONTENTS			IV-3	Energy Masking	9
I	INTRODUCTION	5	IV-4	Background file .	9
	I-A Background and Motivation	5	IV-5	Solar Scatter Continuum . . .	9
	I-A.1 Advancing Lunar Surface Mapping Through Dynamic Visualization . .	5	IV-6	Peak and Element Recognition	9
			IV-7	Intensity Flux . .	10
			IV-8	Catalogue	10
	I-A.2 Revolutionizing Lunar Mapping with Integrated Approaches . . .	6	IV-9	Elemental Mapping	11
			IV-10	Compositional Groups	12
	V SOLUTION OVERVIEW	14			
II	PROBLEM UNDERSTANDING	6	V-A	Road map	14
	II-A Mapping Ratios on a Lunar Base Map	6	VI DATA PROCESSING		16
	II-B Best Ratios Identification .	7	VI-A	Day-Night Classification . .	16
	II-C Sub-pixel Resolution Maps	7	VI-B	Instrumental Calibration .	16
III	LITERATURE REVIEW	7	VI-C	Background Determination	16
	III-A Selection of high- quality spectral data:	7	VI-D	Three Sigma Criterion for Confidence-Based Signal Detection	16
	III-B Comparative analysis of elemental ratios in different regions:	7	VI-E	Solar Flare Classification .	17
	III-C XSM Data Analysis:	8	VI-F	Solar Scatter Continuum and Power Law Application	18
	III-D Database selection:	8	VI-F.1	Solar Scatter Continuum Calculation . . .	18
IV	OVERVIEW OF PRIOR WORK	8	VI-F.2	Power Law Continuum . . .	19
	IV-.1 Classification of FITS files based on Time: Day and Night Region	8	VI-G	Element Correction Counts and Identification .	19
	IV-.2 ARF and RMF Calibration to obtain Corrected Counts vs Energy Graph .	8	VI-H	Intensity Flux Calculation .	19
			VI-H.1	Calculating Intensity Ratios for Silicon (Si) and a General Element (E) . . .	19

VII FINAL APPROACH (DELIVERABLE 1)	24
VII-A Dynamic Interactive Elemental Ratio map	24
VII-B Uniform Grid Formation .	24
VII-B.1 Defining Grid Data Ratios	24
VII-C Methods for Grid Interpolation	24
VII-D Database Structure (MySQL based backend) .	27
VII-E Backend Pipeline Overview	27
VII-E.1 Input: New FITS and XSM files Added	29
VII-E.2 Classifying data into day-side and night-side .	29
VII-E.3 Updation of grid_data	30
VII-E.4 Updation of kriged_data	30
VII-E.5 Additional exploration	30
VII-F Front-End Framework for Lunar Data Insights	33
VII-F.1 Representation of Elementary Ratios	33
VII-F.2 User Interaction and Customization	34
VII-F.3 Dynamic Updates	35
VIII BEST RATIOS CALCULATION & RESULTS (DELIVERABLE 2)	40
VIII-A Data Processing for Finding Best Ratios	40
VIII-B Analysis and Results	40
VIII-B.1 Mare and Highlands Plot	40

VIII-B.2 Regional Analysis (Imbrium and Australe) . .	42
VIII-B.3 Water indicators: Low Mg/Si and Fe/Si Ratios	43
VIII-B.4 Indicators of Volcanic Activity: High Ti/Si Ratios	44
VIII-B.5 Impact Features: High Fe/Si and Cr/Si Ratios	45
VIII-B.6 ISRU Ore Deposits: High Mg/Si, Fe/Si, and Ti/Si Ratios	45

IX SUB-PIXEL RESOLUTION (DELIVERABLE 3)	46
--	----

IX-A Methodology	46
IX-A.1 Naive Averaging	46
IX-A.2 Single-Patch Assignment	47
IX-A.3 Weighted Overlap Averaging at Sub-pixel Resolution	47

X Uncertainties	48
------------------------	----

X-.1 General Human Error	48
X-.2 LPGRS Data . .	49
X-.3 GOES Data	49
X-.4 Resolution and Scale Limitations	49
X-.5 Model Assumptions and Simplifications	49
X-.6 Data Quality and Preprocessing	49

	X-.7	Temporal and Spatial Variability	49		XII-B.5	Database Updates	56
					XII-B.6	Heatmap Visualization	56
XI		Challenges and Lessons Learned	49		XII-B.7	Workflow and Automation	56
	XI-.1	Auto-encoders with Attention Mechanisms (Spatial and Self-Attention) .	49	XIII		CONCLUSION	57
	XI-.2	Deep Convolutional Neural Networks (Deep CNNs)	50			References	58
	XI-.3	Random Forest and XGBoost Regressor	51				
	XI-.4	Physics-Informed Neural Networks	52				
	XI-.5	Multiclass Regression	53				
	XI-.6	Using MySQL over CSV	53				
XII		FUTURE ASPECTS	54				
	XII-A	Applications of Power BI .	54				
	XII-A.1	Input-Based Regional Analysis .	54				
	XII-A.2	Use Case: Lunar Exploration	54				
	XII-B	Cassandra-Based Backend Pipeline for Lunar Data Processing	54				
	XII-B.1	Input: File Addition and Monitoring	54				
	XII-B.2	Classification of FITS Files	55				
	XII-B.3	Background Noise Handling (Night-Side Data)	55				
	XII-B.4	Preprocessing of FITS Files	56				

ABSTRACT

This report, by using CLASS data, aims to present a comprehensive solution of the development of a dynamic and efficient pipeline for generating high-resolution elemental ratio maps of the lunar surface by leveraging X-ray fluorescence (XRF) spectroscopy to derive elemental ratios, such as Mg/Si, Al/Si, Ca/Si, etc., which provide critical insights into the compositional heterogeneity of the Moon at unprecedented spatial scales. The pipeline efficiently processes the extensive dataset of X-ray fluorescence (XRF) spectra, calculates elemental line fluxes, and maps these ratios onto a 2D and 3D lunar albedo base map for clear visualization. Through overlapping track analysis, the pipeline achieves sub-pixel resolution, enabling the identification of fine-scale geochemical variations in the lunar crust, which leads us to resource identification and enhances our understanding of lunar geological evolution. Our pipeline seamlessly integrates all steps between data processing and data visualization, resulting in dynamic, interactive maps and laying an innovative framework for gaining critical insights for geochemical studies, resource identification, and future lunar exploration, establishing a scalable approach for planetary data analysis and visualization.

I. INTRODUCTION

A. Background and Motivation

The Moon is believed to have formed about 4.5 billion years ago after a massive impact event. It holds interesting records of the early Solar System. The Moon's surface is quite unexplored, making it the epitome of resources which makes it important to study it to study planetary formations. Its composition, **characterized by crustal materials**, impact-generated ejecta, and volcanic residues, provides critical insights into its thermal and magmatic history and a broader framework for understanding planetary differentiation in the Solar System.

X-ray fluorescence (XRF) spectroscopy is one of the most important techniques used for studying planetary surfaces via elemental composition and calculating fluorescence data. Using these XRF lines that are detected, surface materials can be determined during solar flares allowing us to identify and quantify the elements thereby understanding the soil. CLASS is the Chandrayaan-2 Large Area Soft X-ray Spectrometer, an advanced XRF instrument that has been orbiting the Moon since 2019, returning high-quality data. CLASS will be looking for the key rock-forming elements of **magnesium, aluminum, and silicon**, whose presence is vital to understanding the Moon's geological evolution and resource distribution. With CLASS, scientists can finally generate high-resolution maps with compositional variation at **spatial scales finer than 12 kilometers**, details never before obtained.

High-resolution compositional mapping has very broad scientific and practical implications. It can shed light on processes like crustal differentiation, volcanic activity, and the solar wind weathering effects. Practically, it could guide future lunar missions, identifying areas of interest for sample-return missions or in-situ resource utilization (ISRU) projects. However, the existing elemental maps are precise and sufficiently formed but do not have spatial resolution to capture finer-scale regions. Therefore, it is impossible to understand the Moon's composition and geology in its entirety.

1) Advancing Lunar Surface Mapping Through Dynamic Visualization: The objective is to move ahead of the usual mapping procedures by looking into the future with dynamic and interactive visualizations of elemental data. While static maps are informative, they lack the essence of the lunar surface. Dynamic interactive maps allow us to zoom in, pan, and layer the data. This way, researchers may have a much more intuitive and exploratory approach when analyzing the surface of the Moon. These visualizations will add extra insight into compositional variability and help pinpoint some individual trends, anomalies, and areas of interest

in the composition of the Moon's surface.

This characterizes the terrains from the Moon's surface in an efficient way due to the ability to represent and visualize optimal combinations of ratios.

2) Revolutionizing Lunar Mapping with Integrated Approaches: The integrated approach of **high-resolution mapping, dynamic interactive visualizations, optimized elemental ratios, and sub-pixel resolution techniques** will redefine our understanding of the Moon's surface. It will scientifically reveal details regarding the geological history of the Moon, evolution of its crust, and magmatic processes. It will provide actionable data to further support future lunar missions, either for exploration, resource utilization, or scientific study. This would facilitate a giant leap for unlocking the full CLASS data that will advance science at a lunar level.

II. PROBLEM UNDERSTANDING

The current lunar missions have greatly expanded the study and knowledge of the Moon's surface composition. Among the most ambitious goals in lunar science, is the creation of high-resolution elemental maps of the Moon's diverse geological terrains. The study contributes to that effort, as data from the CLASS instrument on board Chandrayaan-2 is used. Since its launch into orbit of the Moon in 2019, CLASS has successfully detected and analyzed characteristic X-rays emitted by the surface of the Moon when excited by solar X-rays. Instrumental elemental abundance data are opening unprecedented opportunities to look deeply into the lunar surface detail, producing elemental ratio maps that reveal compositional variations at spatial resolutions unattainable in earlier lunar studies.

The current phase of our research focuses on providing more refined results, based on improvements in elemental ratios analysis originating from CLASS data. All the ratios will be portrayed by **dynamic and interactive visualizations on**

a lunar base map, where we can browse and investigate the surface of the Moon considering every minute detail. Such interactivity in the maps should lead to an intuitive understanding regarding the compositional diversity of the Moon, and thus the interpretation of data.

An important element of this phase is the **selection of elemental ratios that are more critical to the study of the Moon**. These elemental intensity ratios would help us to determine the various geological terrains on the Moon, and understanding this will lead us to gain insights into the spatial data via the compositions, thereby helping us find trends of the history, processes, and possible events.

Moreover, **sub-pixel resolution maps** are made to achieve proximal and more precise spatial data. The high-resolution maps will allow us to study the Moon's surface at a much smaller scale, resolving composition issues that were not resolvable earlier. At this precision, the data gives a better understanding of the Moon's compositional and geological characteristics which opens discussions to understand its origin, evolution, and potential for a huge resource body.

On a bigger scale, this study leads to a vast region for the exploration of lunar data, adding state-of-the-art processes, visualization, and analysis techniques to understand the Moon's surface better.

A. Mapping Ratios on a Lunar Base Map

The lunar data helps us to determine the elemental ratios, and the ratios of Mg/Si and Al/Si plotted against a lunar base map in a dynamic and interactive view, are some of the most useful ratios calculated among the ratios present.

By combining these dynamic maps, we will spot patterns, unusual features, and differences in the landscape of the Moon. A deeper understanding of geology on the Moon makes the interpretation of data a bit more meaningful and easier in unraveling the mystery behind its past.

B. Best Ratios Identification

This deliverable would establish the most important elemental ratios that would describe the diversity of the Moon's geology and composition. Mg/Si and Al/Si are chosen for this because they can trace key magmatic and crustal processes. These ratios would differentiate lunar terrains and give spatial patterns that might point to volcanic activity, crust evolution, or history due to impacts.

C. Sub-pixel Resolution Maps

This deliverable deals with creating high-spatial-resolution sub-pixel maps that could capture finer details of the Moon's surface. These maps provide a high degree of precision which was previously unattainable, uncovering subtle variations in composition.

With these high-resolution maps, researchers can better identify and understand local variations in composition, including micro-structures, localized events, and small-scale geological features. This is very important for the compositional exploration of the Moon, its origin, its evolution, and its possibilities as a resource-rich celestial body.

III. LITERATURE REVIEW

The Chandrayaan-2 Large Area Soft X-ray Spectrometer, or CLASS, represents a giant leap forward in the examination of lunar XRF. Since it was placed in orbit in September of 2019, CLASS has been gathering lunar X-ray spectra. Previous missions like Apollo and Lunar Prospector have demonstrated the value of such approaches in creating elemental maps, but CLASS extends this capability with higher spatial resolution and broader coverage, and we aim to analyze the relative elemental ratios Mg/Si and Al/Si, etc., and aim to enhance the resolution of lunar maps and derive the understanding of lunar crustal composition, geological evolution and identifying potential sites for in situ resource utilization. This report integrates relevant information obtained from various past missions and observations for arriving at the best possible solutions for the end-term deliverables.

A. Selection of high- quality spectral data:

To determine the threshold of the intensity of the peak to be set for selecting high-quality spectral data for elemental analysis during space observations, we discard intervals where there is significant distortion of the spectrum due to the interaction of particles. To achieve this, we identify the spectral lines with intensities above 3σ from the nominal average background and free from particle contamination as good intervals which can be considered under our observation for further analysis. This idea was inspired by the approach described in the research paper "Lunar elemental abundances as derived from Chandrayaan-2". This method is based on the concept that the spectral background is steady when the orbiter is on the night side, and a monthly average is taken as a reference background. Moreover, during geotail passes, there is a distortion in the background due to particle interaction, introducing spectral lines such as of Al-K α and Cu-K α spectral lines with fluctuating intensity ratios. So, the data intervals with significant spectral distortions due to particle interactions are not considered for spectral analysis and only the intensities exceeding three times the nominal background (3σ) and free from contamination are deemed suitable for spectral modeling. The hypothesis behind using this method is explained in detail under the section of Data Processing.

B. Comparative analysis of elemental ratios in different regions:

We can derive this analysis by using the elemental maps of Mg, Al, Si, and Fe that are made using a total of 2458 CLASS footprints at a spatial resolution of 150X12.5 km. In this analysis, Mg and Al line intensities are highlighted indicators for marking the compositional differences between mare and highland regions. As observed from the research paper "Lunar elemental abundances as derived from Chandrayaan-2", mare regions are rich in Magnesium and Iron, whereas highlands are aluminium-rich. The Silicon map does not show a clear boundary between mare and highland regions.

Moreover, diving deeper we observe that sub-regions such as Oceanus Procellarum and Mare Imbrium display variability in Mg abundance, indicating diverse magmatic sources, whereas Mare Frigoris and Mare Serenitatis have localized areas with higher Mg relative to Al, suggesting the presence of less-evolved, primitive basalts. Regions such as farside highlands show narrow ranges of Mg abundances (2–4 wt.%), contrasting sharply with the broader range in mare areas. South Pole-Aitken Basin, despite being a basin feature, exhibits Mg abundance comparable to Oceanus Procellarum, likely due to the excavation of mantle materials.

C. XSM Data Analysis:

The Solar X-ray Monitor generates different types of data files, classified into raw (level-1) and calibrated (level-2) products. The raw files include .fits, .hk, .sa files and the calibrated files include .gti, .pha, .lc files. The .lc files refer to the ‘Light Curve’ files which store time-series data representing photon count rates over time. The series of information described above is derived from the paper *“XSM Data Analysis Guide”*. These files are used to analyze variations in solar X-ray emission over time. Moreover, they cover the full energy range of 1–15 keV and have a temporal resolution of 1 second. The advantages of using .lc files include time-resolved analysis, already pre-processed in terms of variations of effective area, compatibility with visualization tools, and flexibility. The record of the XSM file would become crucial for correcting solar flux during XRF measurements because it contains all relevant information about solar radiation falling on the lunar surface. Therefore, the spectrum of solar X-rays can be significantly enhanced with a calculation involving XSM data which takes into account fluctuations of solar flux, thus refining the accuracy of an elemental composition analysis in XRF observations.

D. Database selection:

MySQL, which is a relational database management system and is widely known for being used in managing structured data in scientific

and exploratory pipelines, is integrated with our backend pipeline as our database. Its integration with Python libraries like mysql-connector-python simplifies data ingestion, transformation, and export to external analysis tools such as Pandas or NumPy. In geospatial analysis, as in our case, relational databases allow us seamless integration of spatial data with compositional metrics, enabling fast querying and updating results, as required in our workflow pipeline. Usage of MySQL as our database makes our pipeline dynamic and robust for data extraction, updation, and storage, along with implementing other methodologies.

IV. OVERVIEW OF PRIOR WORK

1) Classification of FITS files based on Time: Day and Night Region:

The first step in pre-processing involved the sorting of spectral data by the solar angle at the time of observation. We then extracted the solar angle information from the FITS header attached to each observation. As a result, the data was sorted into **day side** and **night side** observations. This distinction was important as illumination by the sun hugely impacts the spectrum measurements.

- Information about the surface elements came from the data of **day side** (where the angle of the solar angle is **less than 90°**),
- whereas the **angles above 90°** correspond mostly to background noise of **night-side data**.

This has helped us in eliminating meaningless data.

2) ARF and RMF Calibration to obtain Corrected Counts vs Energy Graph:

Although we had basic **raw counts vs channel data** as extracted from FITS files, we wanted to correct it and transform it to **counts vs Energy bins** that would be without any instrumental influence. We performed this using ARF and RMF files.

- The ARF file performed corrections for detector sensitivity variations.
- the RMF file redistributed counts to make

clear spectral peaks

Together, these corrections provided us with corrected counts vs. energy data and helped us to arrive at a better representation of the lunar surface flux.

3) *Energy Masking:*

Energy masking was applied to emphasize specific X-ray emission ranges for key elements such as **magnesium, aluminum, silicon, and calcium** (Now, with the use of XSM data, elemental analysis of Cr, Mn, Fe k, Fe l, Ti has also been done) to analyze the elemental composition of the lunar surface.

- Filtering the spectrum from **0.5 to 10 keV** to select characteristic emission lines of these elements helped remove irrelevant signals and noise that might have affected the accuracy of the calculations of elemental flux and thereby gave a better understanding of the Moon's surface composition.

4) *Background file:*

To counter the problem of background noise, a more accurate daily averaging method was adopted for the variability of background noises on a day-to-day basis. The basic steps for this method was:

- Night side data was drawn and averaged daily in order to **address the short-term variations** due to factors such as solar radiation, instrumental drift, or observational geometry.
- **Daily Background Correction:** A mean was computed for each day and was then subtracted from every observation corresponding to that day. This made the **noise removal day-specific**.

This process thus offered a more accurate and specific model of the background noise.

5) *Solar Scatter Continuum:*

In order to account for background noise due to

the scattered solar X-rays in the measured XRF spectra, the **power-law continuum model** had been used. Earlier, we had excluded the use of any external XSM data, but now we have worked upon and **successfully integrated XSM data (from Solar X-ray Monitor on Chandrayaan 2) into our data processing pipeline** (discussed under *Solar Flare Classification* in Data Processing section). This addition helps in the efficient categorization of high/low solar flare regions and helps in doing the elemental analysis of elements like Cr, Mn, Fe_k, Fe_l, Ti.

a) *Key points:*

- **Background Impact:** Scattered solar X-rays from the lunar surface, which are not meant to be added in our calculation since they do not contribute to the excitation, but yet get captured by the SCD distort the characteristic peaks like that of Mg, Al, Si, Ca, which degrades the accuracy of elemental abundances.
- **Power-Law Continuum:** The mathematics of the background model is:

$$I_{\text{scattered}}(E) = AE^{\gamma}$$

where A is a normalization constant, and γ controls the slope of the continuum.

- **Correction:** The corrected spectrum was obtained by subtracting the fitted continuum:

$$I_{\text{corrected}}(E) = I_{\text{observed}}(E) - AE^{\gamma}$$

This approach perfectly corresponds to the interaction of the X-rays with the sun, significantly reducing systematic errors so that the true XRF signals were isolated for better analysis.

6) *Peak and Element Recognition:*

We search for peaks in the spectrum to determine and measure the amounts of individual elements contained within spectroscopic region based on their unique X-ray emission peaks.

a) **Peak Detection:**

- **Definition of Energy Windows**

Every element has a characteristic spectral line in the spectrum, i.e., there is a pre-defined energy range for each element, and its X-rays are detected from this area in this energy range, which are called the elemental peaks. Some elemental peaks are listed below:

- The **Al** elemental peak ranges from **1.43 to 1.53 keV**.
- The elemental peak for **Si** lies between **1.68 and 1.78 keV**.
- **Magnesium** (Mg) peaks at **1.20 to 1.30 keV**.
- There is a peak for characteristic **Ca** around **3.64 to 3.74 keV**.

We referred to these energy ranges for filtering out the useful spectral information in the data.

- **Data Filtering**

After the energy ranges were determined, we filtered the data to contain only those counts that occur within each given peak range. For instance, in the case of aluminium, data between 1.43 and 1.53 keV was selected. This filter brought attention to the selected element.

- **Peak Determination:**

- **Definition:** A peak is considered to be the maximum count observed within the selected range of energy and constitutes the most prominent emission of the element.
- **Found the Peak Count:** We had implemented an algorithm that scanned the filtered spectrum to determine the energy at which the peak count occurs, and that energy had been assigned the label **peak energy**.
- **Recorded the Peak Intensity:** The count at that peak was called the **peak count** and it had given us the highest count for that particular element at that energy. This

information had formed the basis of our analysis.

7) **Intensity Flux:**

Intensity flux is the rate at which an element emits photons within a certain energy range. After we had subtracted the background noise by using the Power-Law Continuum, we **sum up all the photon counts within its emission range and get the total flux**. This measurement plays a vital role in determining the abundance of elements on the Moon.

The intensity flux F is mathematically expressed as:

$$F = \sum_{i=E_1}^{E_2} C_i$$

Where:

- C_i is the corrected photon count in the i -th energy channel.
- E_1 and E_2 define the energy range corresponding to the emission lines of the specific element.

These corrected flux values are quite imperative for calculating elemental ratios, which assist in comprehending the geological features of the lunar surface.

8) **Catalogue:**

The dataset offered us important information taken from X-ray fluorescence (XRF) observations of the lunar surface which comprised of several categories that had given us opportunities for performing further detailed analysis.

Below is a summarization of important columns and their significance;

a) **Geographical Coordinates (Latitude and Longitude):**

- `u_l_lat`, `u_r_lat`, `b_l_lat`, `b_r_lat`: These columns define the latitudes

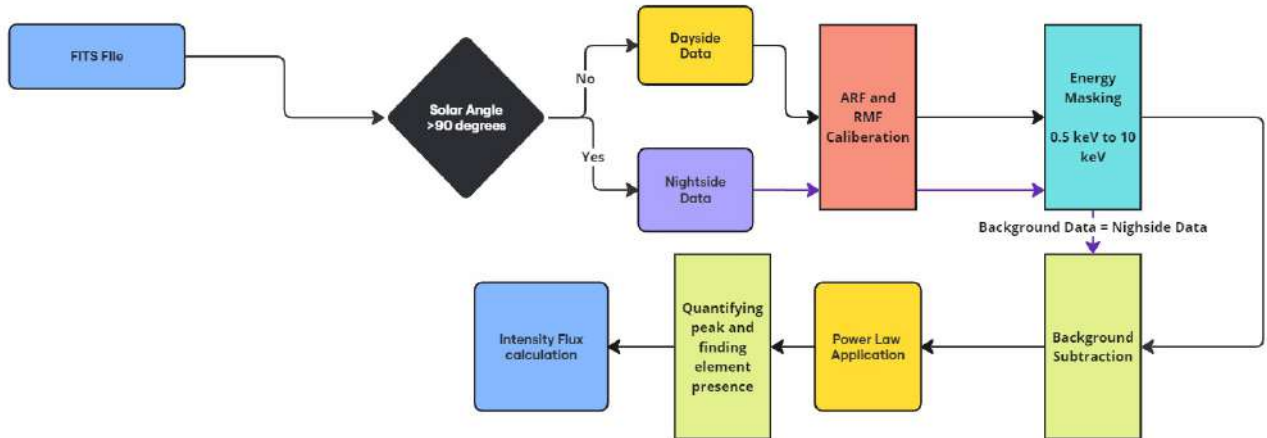


Fig. 1: Workflow of how FITS files are converted to generate corrected Element and their peaks

of the four corners of the observation area.

- `u_l_lon`, `u_r_lon`, `b_l_lon`, `b_r_lon`: These columns define the longitudes of the four corners of the observation area.

b) Time and Observation Conditions:

- `start_date`, `end_date`: The start and end dates of the observation in YYYY-MM-DD format.
- `angle`: The solar angle of the observation.
- `side`: Binary Category of whether the data is Dayside or Nightside.

c) Peak Characteristics and Quantification:

- `al_max_counts`, `si_max_counts`, `mg_max_counts`, `ca_max_counts`: Peak counts for each element representing the maximum intensity.
- `al_area`, `si_area`, `mg_area`, `ca_area`: The intensity flux or the area calculated for each element.

d) Element Abundance and Ratios:

- Relative Abundances of Elements
 - `al_abundance_relative`: Relative abundance of Aluminium
 - `si_abundance_relative`: Relative abundance of Silicon
 - `mg_abundance_relative`: Relative abundance of Magnesium
 - `ca_abundance_relative`: Relative abundance of Calcium
- `al/si`, `mg/si`, `ca/si`: Relative ratios of Aluminium, Magnesium, and Calcium to Silicon.
- `al_present`, `si_present`, `mg_present`, `ca_present`: Binary presence indicators that show whether the element is detected or not based on its abundance.

9) Elemental Mapping:

The objective here was to create high-resolution elemental maps of the Moon's surface using

spatial and geochemical data.

a) Elemental Mapping of the Lunar Surface:

The catalogue had given us considerable insights about the lunar composition, yet it had many drawbacks. For several regions, overlapping areas has to be taken into consideration. The method that we used to overcome this was based on the concept that smaller patches have their elemental ratios proportional to their region of overlap with the larger patches. *In other words:*

- Spatial data analysis was combined with geochemistry to deduce the surface composition of the Moon.
- Large datasets were used to generate high resolution maps by interpolating ratios for finer areas.
- The most challenging part was to refine the elementary data while considering the details about overlaps and distributions.

b) Uniform Grid Approach:

To systematically organize and analyze the lunar surface, we had implemented the uniform grid approach. A uniform grid divides the whole surface into equal-sized cells of $0.4^\circ \times 0.4^\circ$ ($\sim 12 \text{ km} \times \sim 12 \text{ km}$).

Steps for grid division and ratio calculation:

- 1) **Grid Formation:** We had divided the lunar surface into a uniform grid of cells, each measuring $0.4^\circ \times 0.4^\circ$.
- 2) **Larger region to grid mapping:** We had calculated the centroid for each region, and then had checked in which grid the centroid lied in. That was the grid assigned for that particular region.
- 3) **Ratio Calculation:** We had taken the weighted average of the ratios (of a particular element) of all the regions mapped with that grid cell and had assigned that ratio value to

that grid cell. We had performed this for all the grid cells, for all elements.

c) Interpolation with Kriging:

Kriging is a **statistical interpolation** method for finding or determining missing data from spatial interdependence among data points. It smoothens data and produces values reliable to us for regions with insufficient information. Important elements of Kriging:

Important elements of Kriging:

- **Semivariogram:** A mathematical method that defines the relationship between data points on the basis of their distance, i.e. as distance increases, similarity decreases.
- **Models:** Statistics such as spherical, exponential, and Gaussian tell us how data depends on distance.
- **Weights:** Data points are weighted based on spatial data with respect to them to estimate values for undefined regions. Weights are calculated to minimize estimation errors.
- **Variance:** This measures the confidence in the interpolated values. Lower variance indicates higher confidence, while higher variance highlights uncertainty in sparse regions.

d) Geo-visualization with GeoTIFF:

The estimated as well as assigned ratios were saved in a GeoTIFF format which is compatible to work on almost all GIS software. Using this, we were able to overlay the plot of the elemental ratios onto a lunar map, which allowed us to visualize the elemental distribution on the lunar surface, on which we further allowed spacial and pattern analysis and use its conclusions.

10) Compositional Groups:

The lunar regions which were identified and classified with similar compositions, such as

the **Feldspathic Highlands, Mare Basaltic Plains, Procellarum KREEP Terrane, and South Pole-Aitken Basin**. Elemental ratios are clustering parameters. Through a clustering approach that was driven by machine learning and underpinned by geological expertise, areas with similar geochemical signatures were grouped together.

a) Challenges and Solutions:

Traditional unsupervised clustering methods resulted in too much density of outliers dominating patterns and thus preventing the formation of meaningful clusters. Noise also contributed by interrupting the picking of clusters. A geology-aware clustering method was therefore used to address this. Spatial coordinates and elemental composition criteria for each region were included to successfully label and group the data points, thereby overcoming the limitations of traditional methods. .

b) Methodology Overview: The methodology consisted of several steps.

- First, key variables such as latitude, longitude, and elemental ratios (e.g., Al/Si, Mg/Si, Ca/Si) were selected for analysis. To minimize the effect of outliers, **elemental ratios were clipped to a range of 0–2**, and datasets were cleaned to handle missing values and anomalies.
- Next, a **dictionary of mare regions** was created, including central coordinates (latitude and longitude), geographic boundaries (radius), and **compositional criteria** (elemental abundance ranges specific to each region). For example, Mare Australe was defined by coordinates (-36°, 93°E), while Mare Tranquillitatis was characterized by coordinates (8.5°, 31.4°E).

Labeling and filtering were performed by adding a `label` column initialized with "other" and ap-

plying spatial and compositional filters.

Spatial filtering involved calculating the **Haversine distance** to determine proximity to a region's center, while compositional filtering ensured that data points matched the regional elemental criteria. Points meeting both criteria were labeled accordingly.

Descriptive statistics, such as mean and variance, were calculated for elemental abundances and ratios, and comparative analysis was conducted to identify unique compositional signatures for each region. Visualization techniques, including **box plots, spatial distribution maps, and 3D scatter plots**, were employed to enhance interpretability and reveal compositional trends and clusters.

c) Results and Insights:

The analysis revealed distinct elemental abundance patterns across mare regions.

- Mare Australe exhibited the highest aluminum concentrations (30–35 units), indicating unique geological processes.
- Mare Insularum and Oceanus Procellarum had lower aluminum content (10–15 units), typical of basaltic regions.
- Magnesium levels were highest in Mare Serenitatis (20–25 units), suggesting enrichment with magnesium-bearing minerals such as olivine.
- Silicon content showed high variability in Mare Insularum, potentially due to differences in silicate mineral crystallization.

Elemental ratio analysis provided further insights.

- The Al/Si ratio was higher in Mare Frigoris, reflecting aluminum enrichment, while most regions displayed balanced ratios with a median around.
- Mg/Si ratios were consistent across regions, except for outliers in Mare Serenitatis, indicating localized geological anomalies.

- Ca/Si ratios were elevated in Mare Imbrium and Serenitatis, possibly due to calcium-rich anorthosite crustal material.

d) *Key Observations:*

Each mare region exhibited unique elemental profiles, reflecting variations in geological history and processes. The integration of geological knowledge improved clustering and highlighted patterns that traditional methods overlooked. Visualization techniques, such as box plots and 3-D scatter plots, proved crucial for uncovering hidden patterns and enhancing data interpretability.

V. SOLUTION OVERVIEW

The goal is to create detailed and interactive maps of elemental ratios across the Moon's surface to reveal the geologic and compositional heterogeneity of the Moon. This effort will make data not only scientifically correct but also readable and visually engaging. There are 3 deliverables to solve.

First, there's the task of **mapping the elemental ratios** onto a base map of the Moon. This step includes **aligning spatial data with exact lunar coordinates** so that data points represent their actual locations on the lunar surface.

We intend to generate **dynamic interactive maps** with much more value as researchers can dive into data at greater levels of detail. The features that include panning and overlaying information will greatly help observe patterns and anomalies in elemental compositions in the lunar regions, as well as updating the maps as and when more and more data becomes available.

The second one is about the **identification of the most meaningful elemental ratios and visualization techniques**. Al/Si, Mg/Si, and Ca/Si are of importance in identifying geological and compositional differences between different regions of the Moon but should be well communicated using intuitive visualizations. Based on the trends observed while plotting the data, we have gathered the results which are shown in *Deliverable 2*.

Lastly, the final task is to produce **sub-pixel resolution maps in which fine-scale details** of the lunar surface are recorded. High-resolution maps may even be able to detect subtle variations in elemental ratios that might indicate some unique geological processes or perhaps compositional anomalies. These maps **have high resolutions so that small-scale features** are captured that would otherwise go unnoticed at lower resolutions.

This integration of elements would aim toward reducing the gap between chemical raw data and outcomes that further open topics of research and advancements. The outcome would bring about not only an advance understanding of lunar geology but, more importantly, a robust tool for future lunar exploration and resource identification.

A. *Road map*

The solution mapping and analyzing lunar elemental composition is made around a carefully made approach which not only gives good results but also makes it easy to bring new data into the system. There are several important steps which help us to obtain our desired results.

Since our data is raw and not sustainable enough to perform operations so we start with data processing, which is important to make sure the data is clean. Changes to the data are made to account for new elements, including adjustments for **energy masking and background noise removal**. To improve the data accuracy, **power law continuum** are considered to account for the **solar flare influence**. This is particularly important for the new elements since their **response under high solar flare conditions is critical** and deserves special consideration. XSM solar flare data is utilized to identify intervals of high and low solar flares, which allows us to detect elements which required more energy for excitation of photons.

After raw data pre-processing, the most critical step is **data classification**, which categorizes data by **solar angle**. The process allows for sorting data into dayside and nightside data, according

to how sunlight affects the lunar surface, hence only relevant data being used in further analysis. After this step, the value calibration becomes more reliable, making the later stages of the analysis accurate in comparison. In fact, **comparisons of count values to energy values** at this point become very effective.

The process of **peak identification** involves **defining energy ranges, filtering the data** to isolate key features, and then identifying the maximum peaks associated with each element. These **peaks are quantified** to assess elemental presence, supported by hypothesis testing based on Gaussian distribution. This rigorous approach allows us to identify and confirm the elements present on the lunar surface with a high degree of certainty.

Next comes **Intensity calculation** and **area analysis**, where the area under the spectral graph between pre-defined regions is calculated to estimate the intensity of different elements.

The catalogue lays a basis that helps in processing data on a uniform grid and ensures spatial consistency by adopting the centroid method across the data points. Further, the catalogue is also significant as it defines how the construction of the grid happens as well as how the actual data is mapped to the lunar surface.

We continue with the following critical step, known as **Kriging interpolation**, transforming the data to make it possible to draw out gridded values fit for visualization. Filling gaps in the data ensures a uniform distribution of information with results that will be free of seams and accurate when talking about the elemental composition of the Moon.

For *Deliverable 1*, we aim to map ratios of elemental compositions onto the comprehensive lunar albedo base map. The maps are made to be dynamic as well as interactive.

- The back-end of the system is made by a **MySQL database**. This improves data management because it reduces query time by

a big margin, and it is more scalable than traditional methods that work with CSV files.

- New files which are added to system, update the ratios in the database automatically.
- For the visualizations, **Plotly** is utilized in the front-end for better view of the Kriged result in an interactive way such that the front-end remains connected to the back-end all the time, ensuring an undisturbed experience to display the dataset.

In *Deliverable 2*, the focus is on identifying the **Best Ratios** for distinguishing between different lunar regions.

- A custom algorithm will be developed to optimize these calculated ratios, which will be used to help differentiate areas with distinct elemental compositions.
- This algorithm will be used to visualize the various trends on a map, allowing users to easily identify compositional groupings in different lunar regions.

For *Deliverable 3*, the challenge is to develop **sub pixel resolution maps** that can represent the data at a much higher level of detail.

- This is where **Kriging interpolation** comes into play, enabling the creation of high-precision maps. The feasibility of this approach will be explored in a later section of the report.

This will allow for the visualization of elemental compositions at a much finer spatial scale, providing deeper insights into the Moon's geological features.

Through this structured, step-by-step approach, the solution ensures that the mapping and analysis of lunar elemental compositions is not only accurate but also flexible, capable of adapting to new data and evolving as research progresses.

VI. DATA PROCESSING

A. Day-Night Classification

The data is classified into dayside and nightside categories based on the solar angle, θ . Data with $\theta \leq 90^\circ$ fall in the category of **dayside**, while those with $\theta > 90^\circ$ fall in the category of **nightside**. This classification ensures that the data are segregated for later analysis based on solar exposure.

B. Instrumental Calibration

- ARF (Auxiliary Response Function) and RMF (Response Matrix File) are used for **correcting errors in instruments**. These instrumental corrections transform raw spectrum counts by adjusting the energy bins of the effective area of an instrument.
- The corrected counts, C_{corr} , are given by:

$$C_{\text{corr}} = \frac{C_{\text{raw}}}{\text{EffArea}(E)}$$

where;

- C_{raw} is the raw count
- $\text{EffArea}(E)$ is the effective area as a function of energy, E .

This step normalizes the data for consistent spectral analysis.

C. Background Determination

Background data analysis is derived from the nightside data. It represents the aggregate of all the nightside data present in the current day file, ensuring that it is neither too specific nor too general, but sufficiently accurate for later analysis based on solar exposure.

- **Standard Deviation:**
In addition to the mean of these counts, the standard deviation is used to quantify the variability in the counts. It highlights the consistencies in the background signals and indicates whether fluctuations exist in the data for each energy range.
- **Structured Table:**
Once the calculations are done, the results are

saved into a structured table. The rows of the table represent a single file, while the columns contain the mean and standard deviation for each element's background contribution. This clear and interpretable format facilitates easy analysis.

- **Usage:**

This background data is later used to calculate elemental presence criteria and also for the correction of counts in subsequent analysis.

D. Three Sigma Criterion for Confidence-Based Signal Detection

The Three Sigma Criterion, also known as the $\mu + 3\sigma$ criterion, is a statistical model used to distinguish between true signals and background noise in spectral analysis. By employing measures such as the mean (μ) and standard deviation (σ), this approach identifies significant spectral peaks with high confidence and eliminates random noise.

- **Three-Sigma Criterion and Statistical Analysis:**

We calculate the background mean (μ_{bg}) and standard deviation (σ_{bg}) for pre-defined individual elemental ranges. Peaks with intensity I_{peak} greater than $\mu_{bg} + 3\sigma_{bg}$ are considered statistically significant, ruling out fluctuations as the cause.

- **Key Formulae for Noise Parameters:**

- **Mean of the Background Intensity (μ_{bg}):**

$$\mu_{bg} = \frac{1}{N} \sum_{i=1}^N I_{bg,i}$$

where N is the total number of background intensity measurements and $I_{bg,i}$ is the intensity of the background in the i -th measurement.

- **Background Intensity Standard Deviation (σ_{bg}):**

$$\sigma_{bg} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_{bg,i} - \mu_{bg})^2}$$

- **Detection Threshold for Significant**

Peaks:

$$I_{peak} > \mu_{bg} + 3\sigma_{bg}$$

- **Why 3σ Is the Desired Threshold:**

- **Avoids False Positives:** The probability that noise exceeds the 3σ threshold is only 0.3
- **Balances Sensitivity and Specificity:** Lower thresholds, such as 1σ , increase sensitivity but may lead to false positives, while higher thresholds, like 5σ , reduce false positives but might miss weaker true peaks.

- **Usage in Spectral Analysis:**

- **Calculating Background Noise Parameters:** Noise parameters (μ_{bg} and σ_{bg}) are derived from noisier parts of the spectrum, often obtained during the night or in areas with weak signals.
- **Identifying Significant Peaks:** Peaks with amplitudes larger than $\mu_{bg} + 3\sigma_{bg}$ are considered statistically valid. Signals from elements like Fe, Mg, or Al above this threshold and repeating over time are considered real.
- **Noisy Data Elimination:** The Three-Sigma Criterion eliminates false positives due to instrumental artifacts (e.g., Al-K or Cu-K lines) or random fluctuations.

- **Three-Sigma Criterion Hypothesis Testing Framework:** The Three-Sigma Criterion can be integrated into hypothesis testing to classify a signal.

- **Null Hypothesis (H_0):** The observed peak is noise:

$$I_{peak} \leq \mu_{bg} + 3\sigma_{bg}$$

- **Alternative Hypothesis (H_1):** The ob-

served peak represents a real signal:

$$I_{peak} > \mu_{bg} + 3\sigma_{bg}$$

- **Test Statistic (Z):**

$$Z = \frac{I_{peak} - \mu_{bg}}{\sigma_{bg}}$$

If $Z > 3$, reject H_0 and the signal is meaningful. If $Z \leq 3$, fail to reject H_0 and the signal may be noise.

- **Illustrative Computation:**

For magnesium (Mg) in a spectral dataset:

- $\mu_{bg} = 2.88$, $\sigma_{bg} = 0.05$
Observed peak intensity: $I_{peak} = 7.37$
- The test statistic is:

$$Z = \frac{7.37 - 2.88}{0.05} = 89.8$$

Since $Z > 3$, the peak is statistically significant, corresponding to a genuine signal.

- **Importance of the Three-Sigma Criterion:**

- **Consistency Across Elements:** The criterion delivers accurate results for various elements such as Fe, Mg, and Al, ensuring that actual signals are detected in diverse datasets.
- **Sensitivity to Complex Spectral Environments:** If the background intensity is not normally distributed, advanced statistical methods can further refine μ_{bg} and σ_{bg} to improve robustness.
- **High Confidence:** The Three-Sigma Criterion minimizes false positives while maintaining sensitivity, ensuring high-quality, scientifically valid datasets.

E. Solar Flare Classification

The classification of solar flare intensity is crucial for identifying regions of interest in spectral analysis. This work adopts a statistically robust approach by utilizing photon count rates from XSM light curve (.lc) files. It involves thresholding with the monthly mean (μ) and standard deviation (σ) of

photon count data, along with temporal constraints for robust event detection.

- 1) **Threshold Definition for High and Low Solar Flares:** A solar flare interval is classified as high intensity if the measured photon count rate (I_{rate}) satisfies the inequality:

$$I_{\text{rate}} > \mu + \sigma,$$

where:

- μ : Average photon count rate during the month.
- σ : Standard deviation of photon count rates during the month.

For intervals to be classified as high solar flare regions, they must also meet the duration criterion:

$$\Delta T = \text{MET}_{\text{end}} - \text{MET}_{\text{start}} > 8 \text{ s},$$

where ΔT is the event time difference, ensuring that only temporally consistent events are considered. The MET (Mission Elapsed Time) values are then converted to UTC (Universal Time Coordinated) for readable start and end times.

Intervals where the photon count rate does not exceed $\mu + \sigma$ are regarded as low-intensity regions:

$$I_{\text{rate}} \leq \mu + \sigma.$$

- 2) **Data Processing Workflow:** For each lightcurve file:

- **Monthly Statistics:** The mean (μ) and standard deviation (σ) are computed from the photon count rates:

$$\mu = \frac{1}{N} \sum_{i=1}^N I_{\text{rate},i},$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_{\text{rate},i} - \mu)^2},$$

where N is the total number of photon count measurements.

- **Setting threshold:** Intervals exceeding the

threshold $\mu + \sigma$ are identified. Adjacent data points that meet the criteria are grouped as contiguous intervals.

- **Duration Analysis:** For each grouped interval, the duration (ΔT) is calculated. Intervals with $\Delta T > 8 \text{ s}$ are classified as high solar flare intervals.
- **CSV Accumulation:** The high-intensity regions are written to a CSV file containing:
 - Filename
 - Date
 - Start Time
 - End Time
 - Duration (seconds)

3) Element Analysis by Flare Intensity:

- High solar flare regions are used to study elements with strong emission lines during high-intensity conditions, such as {Si, Mg, Al, Ca, Fe-K, Fe-L, Cr, Mn, Ti}.
- Low solar flare regions are used to study elements detectable during low-intensity conditions, such as {Mg, Si, Al, Ca}.
- This two-fold classification allows for the targeting of spectral features relevant to different intensity levels of flares.

F. Solar Scatter Continuum and Power Law Application

Background subtraction is crucial for isolating true signals in X-ray spectra.

This power-law model helps in quantifying the energy distribution and separating solar scatter from elemental contributions.

- 1) **Solar Scatter Continuum Calculation:** The solar scatter continuum accounts for scattered solar X-rays, which are adjusted based on the intensity of solar activity. The elements involved are categorized as follows:

- **High Solar Flare Requirement Elements:** Elements including Ca, Mg, Al, Si, Mn, Cr, Ti, Fe-K, and Fe-L, require data during strong solar flares to be measured accurately.
- **Low Solar Flare Requirement Elements:** Elements like Ca, Mg, Al, and Si can be measured during low solar flare activity.

2) **Power Law Continuum:** The corrected counts are obtained after applying background noise correction and fitting a power-law continuum to the spectrum. These corrections are essential to isolate the true signals from each element. The power-law model is given by:

$$F(E) = A \cdot E^{-B},$$

where;

- $F(E)$ is the flux at energy E
- A is the normalization constant
- B is the photon index

which characterizes the spectral slope of the continuum.

This distinction ensures that the background for all elements is accurately modeled, considering its dependence on solar activity.

G. Element Correction Counts and Identification

Element identification is crucial for detecting specific elements such as Al, Si, Mg, and Ca, based on their characteristic X-ray emission lines. To isolate these elements, **background noise correction** is applied to the photon count data, followed by peak detection within the **energy ranges** corresponding to the emission lines of the target elements. This process ensures that only true signals from the elements of interest are considered, allowing for more accurate quantification of their abundances.

For key elements like Silicon (Si), Magnesium (Mg), Aluminum (Al), and Calcium (Ca), their corresponding photon counts are corrected by applying appropriate calibration factors such as the

Ancillary Response Function (ARF) and the Redistribution Matrix Function (RMF). After this correction, the counts are integrated over predefined energy bins that correspond to the energy ranges where these elements emit X-rays.

H. Intensity Flux Calculation

The flux for each element is calculated by summing the corrected photon counts over the energy bins specific to the element's emission range. The general formula for flux calculation is:

$$\text{Flux} = \sum_{\text{Energy Bins}} \text{Corrected Counts},$$

where the corrected counts represent the counts adjusted for background noise and other calibration factors.

1) Calculating Intensity Ratios for Silicon (Si) and a General Element (E):

To further interpret the elemental composition, intensity ratios are calculated by comparing the flux of Silicon (Si) to that of another element, E , such as Calcium or Magnesium. The process can be broken down into the following steps;

- 1) **Flux Calculation for Silicon (Si):** The flux for Silicon is determined by summing the corrected photon counts across its energy range (for example, 1.68–1.78 keV):

$$F_{\text{Si}} = \sum_{i=E_1^{\text{Si}}}^{E_2^{\text{Si}}} C_{\text{Si},i}.$$

- 2) **Flux Calculation for the General Element (E):** Similarly, the flux for any other element E (for example, Calcium or Magnesium) is calculated by summing the corrected photon counts within its specific energy range:

$$F_E = \sum_{i=E_1^E}^{E_2^E} C_{E,i}.$$

Variable	Description	Usage
file name	Name of the data file being analyzed	Identifies the specific file corresponding to the data for each element and observation
u_l_lat u_r_lat b_r_lat b_l_lat	Latitude coordinates for the upper-left, upper-right, bottom-right, and bottom-left corners of the region of interest	Defines the spatial region of interest for the data analysis. These values are used to map the area
u_l_lon u_r_lon b_r_lon b_l_lon	Longitude coordinates for the upper-left, upper-right, bottom-right, and bottom-left corners of the region of interest	Similar to latitudes, these define the spatial extent of the region under study
start_date end_date	Start and end dates of the observation	Used to filter or organize the data by observation period
start_time end_time	Start and end times of the observation	Helps identify the exact time frame of data acquisition
angle	Angle of observation	Used to adjust data for viewing angles, affecting how the observations are interpreted
side	Side of the observation (e.g., left or right)	Refines the region of interest by specifying the observation side (if applicable)
flare	Solar flare status (indicating if a flare occurred during the observation period)	Indicates the influence of solar activity on the data, used to adjust for solar flare effects
geotail	Indicator of whether the data was taken from the Earth's geotail	Assists in understanding the geomagnetic influence on the data and its location
al_max_counts	Maximum counts detected for Aluminum (Al)	Represents the highest photon count for Aluminum, indicating the intensity of detection
si_max_counts	Maximum counts detected for Silicon (Si)	Represents the highest photon count for Silicon, indicating the intensity of detection
mg_max_counts	Maximum counts detected for Magnesium (Mg)	Represents the highest photon count for Magnesium, indicating the intensity of detection
ca_max_counts	Maximum counts detected for Calcium (Ca)	Represents the highest photon count for Calcium, indicating the intensity of detection

Variable	Description	Usage
mn_max_counts	Maximum counts detected for Manganese (Mn)	Represents the highest photon count for Manganese, indicating the intensity of detection
cr_max_counts	Maximum counts detected for Manganese (Mn)	Represents the highest photon count for Manganese, indicating the intensity of detection
ti_max_counts	Maximum counts detected for Titanium (Ti)	Represents the highest photon count for Titanium, indicating the intensity of detection
fe_k_max_counts	Maximum counts detected for Iron (Fe-K)	Represents the highest photon count for Iron (Fe-K), indicating the intensity of detection.
fe_l_max_counts	Maximum counts detected for Iron (Fe-L)	Represents the highest photon count for Iron (Fe-L), indicating the intensity of detection.
al_energy	Energy corresponding to the maximum counts detected for Aluminum (Al)	Represents the highest photon count for Iron (Fe-L), indicating the intensity of detection.
si_energy	Energy corresponding to the maximum counts detected for Silicon (Si)	Represents the energy at which maximum counts were observed for Silicon
mg_energy	Energy corresponding to the maximum counts detected for Magnesium (Mg)	Represents the energy at which maximum counts were observed for Magnesium
ca_energy	Energy corresponding to the maximum counts detected for Calcium (Ca)	Represents the energy at which maximum counts were observed for Calcium
mn_energy	Energy corresponding to the maximum counts detected for Manganese (Mn)	Represents the energy at which maximum counts were observed for Manganese
cr_energy	Energy corresponding to the maximum counts detected for Chromium (Cr)	Represents the energy at which maximum counts were observed for Chromium
ti_energy	Energy corresponding to the maximum counts detected for Titanium (Ti)	Represents the energy at which maximum counts were observed for Titanium
fe_k_energy	Energy corresponding to the maximum counts detected for Iron (Fe-K)	Represents the energy at which maximum counts were observed for Iron (Fe-K)
fe_l_energy	Energy corresponding to the maximum counts detected for Iron (Fe-L)	Represents the energy at which maximum counts were observed for Iron (Fe-L)

Variable	Description	Usage
al_area	Area corresponding to the detection of Aluminum (Al)	Defines the physical area covered by the detection for Aluminum
si_area	Area corresponding to the detection of Silicon (Si)	Defines the physical area covered by the detection for Silicon
mg_area	Area corresponding to the detection of Magnesium (Mg)	Defines the physical area covered by the detection for Magnesium
ca_area	Area corresponding to the detection of Calcium (Ca)	Defines the physical area covered by the detection for Calcium
mn_area	Area corresponding to the detection of Manganese (Mn)	Defines the physical area covered by the detection for Manganese
cr_area	Area corresponding to the detection of Chromium (Cr)	Defines the physical area covered by the detection for Chromium
ti_area	Area corresponding to the detection of Titanium (Ti)	Defines the physical area covered by the detection for Titanium
fe_k_area	Area corresponding to the detection of Iron (Fe-K)	Defines the physical area covered by the detection for Iron (Fe-K)
fe_l_area	Area corresponding to the detection of Iron (Fe-L)	Defines the physical area covered by the detection for Iron (Fe-L)
al_abundance_relative	Relative abundance of Aluminum (Al) based on the counts and flux	Helps in determining the relative abundance of Aluminum in the region
si_abundance_relative	Relative abundance of Silicon (Si) based on the counts and flux	Helps in determining the relative abundance of Silicon in the region
mg_abundance_relative	Relative abundance of Magnesium (Mg) based on the counts and flux	Helps in determining the relative abundance of Magnesium in the region
ca_abundance_relative	Relative abundance of Calcium (Ca) based on the counts and flux	Helps in determining the relative abundance of Calcium in the region
mn_abundance_relative	Relative abundance of Manganese (Mn) based on the counts and flux	Helps in determining the relative abundance of Manganese in the region

Variable	Description	Usage
cr_abundance_relative	Relative abundance of Chromium (Cr) based on the counts and flux	Helps in determining the relative abundance of Chromium in the region
ti_abundance_relative	Relative abundance of Titanium (Ti) based on the counts and flux	Helps in determining the relative abundance of Titanium in the region
fe_k_abundance_relative	Relative abundance of Iron (Fe-K) based on the counts and flux	Helps in determining the relative abundance of Iron (Fe-K) in the region
fe_l_abundance_relative	Relative abundance of Iron (Fe-L) based on the counts and flux	Helps in determining the relative abundance of Iron (Fe-L) in the region
al/si	Ratio of Aluminum (Al) to Silicon (Si)	Provides comparative insight into the relative abundances of Aluminum and Silicon
mg/si	Ratio of Magnesium (Mg) to Silicon (Si)	Provides comparative insight into the relative abundances of Magnesium and Silicon
ca/si	Ratio of Calcium (Ca) to Silicon (Si)	Provides comparative insight into the relative abundances of Calcium and Silicon
mn/si	Ratio of Manganese (Mn) to Silicon (Si)	Provides comparative insight into the relative abundances of Manganese and Silicon
cr/si	Ratio of Chromium (Cr) to Silicon (Si)	Provides comparative insight into the relative abundances of Chromium and Silicon
ti/si	Ratio of Titanium (Ti) to Silicon (Si)	Provides comparative insight into the relative abundances of Titanium and Silicon
fe_k/si	Ratio of Iron (Fe-K) to Silicon (Si)	Provides comparative insight into the relative abundances of Iron (Fe-K) and Silicon
fe_l/si	Ratio of Iron (Fe-L) to Silicon (Si)	Provides comparative insight into the relative abundances of Iron (Fe-L) and Silicon

- 3) **Intensity Ratio Calculation:** The intensity ratio between Silicon and the general element E is computed as:

$$\text{Intensity Ratio (E/Si)} = \frac{F_E}{F_{\text{Si}}}.$$

Where:

- $C_{\text{Si},i}$ and $C_{E,i}$ are the corrected photon counts for Silicon and element E in the i -th energy channel.
- E_1^{Si} and E_2^{Si} represent the energy range for Silicon's emission lines.
- E_1^{Si} and E_2^{Si} represent the energy range for Silicon's emission lines.
- E_1^E and E_2^E represent the energy range for the emission lines of element E .

The intensity ratio tells us how abundant Silicon is compared to any other element E in the sample. This ratio is crucial for understanding the relative composition of materials.

VII. FINAL APPROACH (DELIVERABLE 1)

A. Dynamic Interactive Elemental Ratio map

Dynamic Interactive Lunar Map, helps us to visualize the elemental ratios across the lunar surface which is both **real time and interactive**. It helps to display the elemental ratios which are calculated using **Kriging Method** on a uniform grid that has resolution of $0.4^\circ \times 0.4^\circ$ for the ratios Mg/Si, Ca/Si, Al/Si onto a lunar base map. It not only adds features like zooming, panning, and hover capabilities but also ensures that the map **updates the new data when available**, making sure that the user interacts with up-to-date representation of Moon's elemental composition.

B. Uniform Grid Formation

A consistent grid system divides the lunar surface with each grid cell measuring $0.4^\circ \times 0.4^\circ$ (about 12 km x 12 km on the actual surface). This grid system divides the surface into pieces of uniform size to examine spatial data. Each cell can be

given a representative ratio, such as an elemental abundance ratio, based on the data points that fall within its boundaries.

However, some of the cells do not have overlaying data points and therefore do not have a computed ratio.

1) Defining Grid Data Ratios:

- **Grid Division:** The surface of the moon has been divided into a grid of $0.4^\circ \times 0.4^\circ$ cells systematically. Every cell is therefore an area of about 12 km x 12 km by which a systematic approach to explore the surface of the moon can be taken.
- **Centroid Overlap:** Determine for each cell, which data points have centroids that fall inside the cell.
- **Ratio Retrieval:** Once these relevant data points have been determined, retrieve the ratios (for example, elemental abundance) of those points.
- **Mean Value Calculation:** Calculate the average of ratios. The average ratio of a cell gives a typical ratio of that cell. This yields the region's representative value.

When there are no overlapping data points within a cell, then a cell loses its ratio leading to data gaps. In that case, the system assists in the proper organization as well as analysis of large spatial data but has to handle sparse data points scenarios.

C. Methods for Grid Interpolation

Data points for which information was not available, we applied several techniques both Machine Learning and Statistical models to predict the values in these regions.

A lot of effort was put into improving how calculations were done for the lunar surface grid using advanced machine learning methods. Methods based on **probabilities or data-driven approaches** were chosen because they were expected to give more accurate predictions. These methods are often more

effective when data is incomplete compared to traditional techniques.

The challenges and reasons for rejecting the ML model approach are specified in the Challenges and Lessons Learned section of the report.

Kriging Interpolation: Kriging is a geostatistical spatial interpolation method where the variable to be estimated at unspecified locations is the value of a spatially correlated variable determined by known values at nearby locations. The method assumes that spatially close points are likely to take similar values, and this spatial dependency is quantified using **covariance function or variogram**.

a) Covariance Function (Semivariogram): In Kriging, the spatial correlation between two data points x_1 and x_2 is described by the covariance function $C(x_1, x_2)$, which measures how values at two locations are related. The general form of the covariance function is:

$$C(x_1, x_2) = \text{Cov}(z(x_1), z(x_2))$$

$$C(x_1, x_2) = \text{Var}(z(x)) - \gamma(x_1, x_2)$$

where $\gamma(x_1, x_2)$ is the semivariogram, which is defined as:

$$\gamma(x_1, x_2) = \frac{1}{2} E[(z(x_1) - z(x_2))^2]$$

The semivariogram γ describes the spatial structure of the data by measuring the dissimilarity between values at different locations. It is assumed that the semivariogram depends only on the distance between the points, i.e.,

$$\gamma(x_1, x_2) = \gamma(\|x_1 - x_2\|)$$

b) Variogram Models:

The most common models for the semivariogram are the spherical model, the exponential model, and the Gaussian model. The variogram model of the spherical function is given by the following expression:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left(1 - \frac{3h^2}{2a} + \frac{h^3}{2a^3}\right), & \text{if } h \leq a \\ C_0 + C_1, & \text{if } h > a \end{cases}$$

where:

- $h = \|x_1 - x_2\|$ is the lag distance between the two points.
- C_0 is the nugget effect, representing small-scale spatial variability.
- C_1 is the sill, representing the total variance of the field.
- a is the range, where the correlation between data points is negligible.

The choice of the variogram model is critical, and it is estimated from the observed data using a process known as *variogram fitting*.

c) Kriging Estimation:

Given observed values $z(x_1), z(x_2), \dots, z(x_n)$ at known locations x_1, x_2, \dots, x_n , Kriging predicts the value $\hat{z}(x_0)$ at a new location x_0 . The Kriging estimator $\hat{z}(x_0)$ is a weighted average of the observed values:

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i)$$

where λ_i are the Kriging weights, which are computed by solving a system of linear equations derived from the covariance matrix.

d) Kriging System of Equations:

The Kriging weights λ_i are determined by minimizing the prediction variance subject to the constraint that the estimator is unbiased. The system of equations used to compute the Kriging weights is as follows:

$$\begin{bmatrix} C(x_1, x_1) & C(x_1, x_2) & \dots & C(x_1, x_n) & 1 \\ C(x_2, x_1) & C(x_2, x_2) & \dots & C(x_2, x_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C(x_n, x_1) & C(x_n, x_2) & \dots & C(x_n, x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C(x_0, x_1) \\ C(x_0, x_2) \\ \vdots \\ C(x_0, x_n) \\ 1 \end{bmatrix}$$

where:

- $C(x_i, x_j)$ represents the covariance between observed locations x_i and x_j .
- μ is the Lagrange multiplier that ensures the estimator is unbiased (i.e., the sum of the weights λ_i should be 1).

The last equation, $\sum_{i=1}^n \lambda_i = 1$, ensures that the Kriging estimator is unbiased.

e) *Variance of the Kriging Estimator:*

The variance of the Kriging estimator $\hat{z}(x_0)$ provides us an indication of the uncertainty in the prediction at location x_0 . The variance is defined as:

$$\text{Var}(\hat{z}(x_0)) = C(x_0, x_0) - \sum_{i=1}^n \lambda_i C(x_0, x_i)$$

where;

- $C(x_0, x_0)$ is the self-covariance at the predicted location
- $C(x_0, x_i)$ is the covariance between the predicted location and each of the known data points

The **lower the variance, the more reliable the prediction**. Kriging provides not only the predicted value but also this uncertainty estimate, which is a major advantage over probabilistic machine learning models that typically do not give an explicit measure of prediction uncertainty.

f) *Advantages of Kriging:*

- **Unbiased Prediction:** Kriging prevents the predictions from being biased by forcing that the weighted sum of observed data is equal to the predicted value.
- **Spatial Autocorrelation:** Kriging incorpo-

rates spatial dependence in the model through the covariance function, which makes it ideal for problems where data are spatially correlated, like the lunar surface grid.

- **Prediction Uncertainty:** Kriging does not only provide predictions but also the variance, giving a measure of confidence in the predictions.
- **Optimal Linear Estimator:** Kriging is considered an optimal linear estimator because it minimizes the prediction variance, resulting in the best possible predictions under the given spatial structure.

g) *Deterministic v/s Probabilistic Method:*

Advantages of Kriging over probabilistic machine learning models for spatially correlated data are **unbiased predictions** with guaranteed accuracy. However, Kriging's predicted results are unbiased since the weight values in its weighted average summation are set to 1 so, inherently accurate.

In contrast, many probabilistic machine learning models lack this feature. While these models rely on data to make predictions, their **results can sometimes be biased** due to incorrect assumptions or poor data quality, especially in spatially correlated scenarios, which can make their estimates less reliable.

• **Including Spatial Autocorrelation**

- Kriging uses covariance function to determine how data points are related based on their distance. Hence, it works best when **data points are close to each other**. Kriging has proved to be quite helpful in fields like, geology and environmental modeling.

Unlike probabilistic machine learning models that often demand more complicated modifications before taking spatial correlation into consideration, for example, Gaussian Processes; Kriging naturally allows such dependency and is requires lesser

computation for such tasks in predicting uncertainty.

- Kriging not only predicts values but also directly measures uncertainty through the variance of its estimates. This makes it easy to understand how reliable the predictions are. While machine learning models might provide confidence intervals, Kriging's clear way of showing uncertainty is more intuitive and useful, especially for making decisions in uncertain or high-stakes scenarios.

- **Optimal Linear Estimator:**

Kriging is treated as an optimal linear estimator because it has the smallest possible variance while predicting; thus, leading to more accuracy with less uncertainty. Machine learning models that implement say random Forest and/or Neural Networks predict the data correctly but **minimize the variance**. In fact, their predictions could have a lot of variability, especially in spatial data, for which Kriging provides precision.

- **Deterministic Nature:**

Kriging is deterministic, meaning once the data and covariance function are given, it creates consistent and repeatable results. Its predictions possess **no randomness**, which makes it ideally suited for places where precision and consistency are required.

In addition, probabilistic models often include randomness or assumed distributions, which introduce variability and make decision-making much harder in applications that require consistent outcomes.

Summarizing by saying that Kriging is the least biased, has explicit uncertainty quantification, can take care of spatial correlations, and has a deterministic approach. It would be best used in those tasks that deal with spatially dependent data.

D. Database Structure (MySQL based backend)

Our main MySQL database, which consists of all tables and is integrated with our dynamic pipeline is named 'isro_team3'. This database consists of the following tables: 'background_mean_sigma_counts', 'subpixel_resolution_data', 'grid_data', 'kriged_data', 'high_solar_flare_intervals', 'geotail', 'update_data' and 'catalogue'. For better visualization of the structure of the database, a flowchart is given below.

E. Backend Pipeline Overview

Using MySQL database, the pipeline is designed for dynamic, efficient, and seamless processing and updating of our database of the scientific data acquired from observations of the Chandrayaan 2 orbiter. New FITS files are automatically fetched, processed, and put into a MySQL database which is integrated with the visualization system through our pipeline. The whole process, starting from FITS file detection to heatmap update, is automated into a single unified pipeline, and does not require manual intervention of any kind.

After performing the query analysis, we arrived at the following query- time analysis which shows a comparative analysis between using CSV versus using MySQL database. The below table is formulated by considering 1 FITS file of size 26 kb.

Through pipeline integration with MySQL and avoiding the repeatative use CSV files, we achieved the following:

- **LATENCY:** Direct database interaction instead of file input-output operation overhead.
- **ACCURACY:** Integrates MySQL database and eliminates CSV file to avoid any risk of data loss or corruption due to intermediate CSV files.
- **REAL-TIME PROCESSING:** Automates the end-to-end workflow, thereby increasing the speed of updates and visualization.

isro_team_73

Field	Type
date	varchar(255)
al_mean	float
mg_mean	float
si_mean	float
ca_mean	float
mn_mean	float
cr_mean	float
ti_mean	float
fe_k_mean	float
fe_l_mean	float
al_sigma	float
mg_sigma	float
si_sigma	float
ca_sigma	float
mn_sigma	float
cr_sigma	float
ti_sigma	float
fe_k_sigma	float
fe_l_sigma	float
nightside_...	float

background_mean_
sigma_counts

Field	Type
u_l_lat	float
u_l_lon	float
u_r_lat	float
u_r_lon	float
b_r_lat	float
b_r_lon	float
b_l_lat	float
b_l_lon	float
al/si	float
mg/si	float
ca/si	float
mn/si	float
cr/si	float
ti/si	float
fe_k/si	float
fe_l/si	float
counts	float

subpixel_resolution_data

Field	Type
u_l_lat	float
u_l_lon	float
u_r_lat	float
u_r_lon	float
b_r_lat	float
b_r_lon	float
b_l_lat	float
b_l_lon	float
al/si	float
mg/si	float
ca/si	float
mn/si	float
cr/si	float
ti/si	float
fe_k/si	float
fe_l/si	float
centroi...	float

grid_data

Field	Type
u_l_lat	float
u_l_lon	float
u_r_lat	float
u_r_lon	float
b_r_lat	float
b_r_lon	float
b_l_lat	float
b_l_lon	float
cent_lat	float
cent_lon	float
al/si_kri...	float
mg/si_kr...	float
ca/si_kr...	float
mn/si_kr...	float
cr/si_kri...	float
ti/si_kri...	float
fe_k/si_...	float

kringed_data

Field	Type
file_name	varchar(255)
u_l_lat	float
u_r_lat	float
b_r_lat	float
b_l_lat	float
u_l_lon	float
u_r_lon	float
b_r_lon	float
b_l_lon	float
start_date	varchar(255)
end_date	varchar(255)
start_time	varchar(255)
end_time	varchar(255)
angle	float
side	varchar(255)
flare	varchar(255)
geotail	varchar(255)
al_max_c...	float
si_max_c...	float
mg_max_c...	float
ca_max_c...	float
mn_max_c...	float
cr_max_c...	float
ti_max_c...	float
fe_k_max...	float
fe_l_max...	float
al_energy	float
si_energy	float
mg_energy	float
ca_energy	float
mn_energy	float
cr_energy	float
ti_energy	float
fe_k_ene...	float
fe_l_ene...	float
al_area	float
si_area	float
mg_area	float
ca_area	float
mn_area	float
cr_area	float
ti_area	float
fe_k_area	float
fe_l_area	float
al_abund...	float
si_abund...	float
mg_abun...	float
ca_abund...	float
mn_abun...	float
cr_abund...	float
ti_abunda...	float
fe_k_abu...	float
fe_l_abun...	float
al/si	float
mg/si	float
ca/si	float
mn/si	float
cr/si	float
ti/si	float

Catalogue

Field	Type
File_Name	varchar(255)
Date	varchar(255)
Start_Time	varchar(255)
End_Time	varchar(255)
Duration	float

high_solar_flare_
intervals

Field	Type
Full_Moon_Date	varchar(255)
Start_Timestamp	varchar(255)
End_Timestamp	varchar(255)

geotail

Field	Type
u_l_lat	float
u_l_lon	float
u_r_lat	float
u_r_lon	float
b_r_lat	float
b_r_lon	float
b_l_lat	float
b_l_lon	float
cent_lat	float
cent_lon	float
al/si_kri...	float
mg/si_kr...	float
ca/si_kr...	float
mn/si_kr...	float
cr/si_kri...	float
ti/si_kri...	float
fe_k/si_...	float
fe_l/si_...	float

update_data

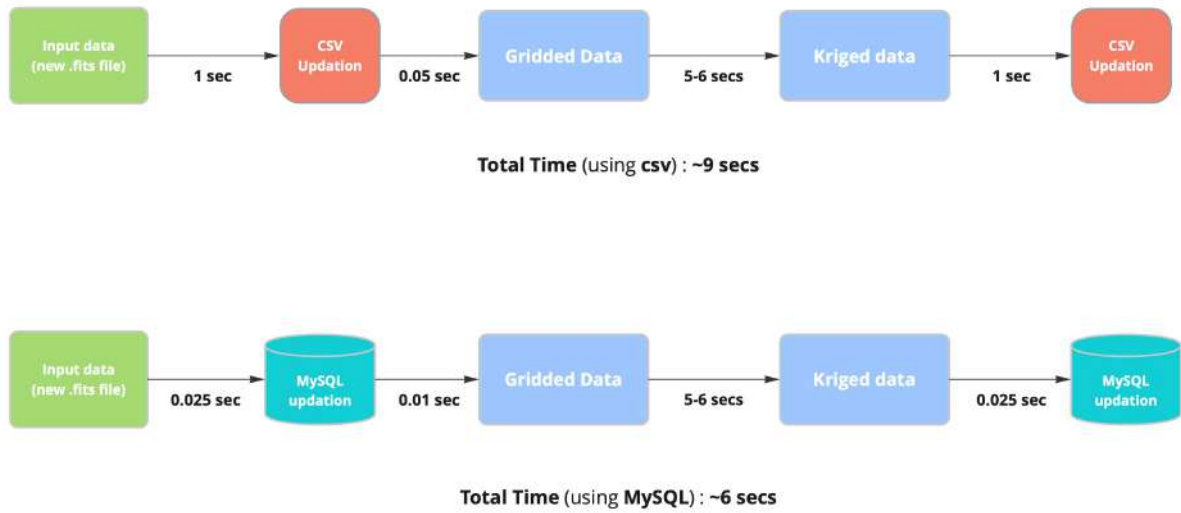


Fig. 2: MySQL vs CSV: Time comparison

- **SCALABILITY:** Pipeline supports big data and parallel processing of many FITS files. MySQL ensures robust data handling and querying.
- **MAINTAINABILITY:** Centralized storage and processing of data ensure traceability and makes it easier to maintain the system.

Here's a technical description of the pipeline:

1) *Input: New FITS and XSM files Added :*

The pipeline begins with the addition of observational data in the FITS file format and XSM file from the Chandrayaan 2 orbiter in the "New_Data" directory. Under the directory of New_Data, we have two directories, "fits_files" and "xsm_files". The process of addition of files is flexible, robust, and scalable, that is, the FITS files may be added in the form of individual FITS files of 8 seconds, or the form of a directory under the fits.file directory.

Similarly, the XSM file may be added as a single file or in the directory format with the convention of YYYY\MM\DD. The addition of files is allowed in any format, as our script restructures the files by converting it into directory format. Thus, the directory of New_Data here acts as a repository of incoming raw data and accepts all files added to it.

2) *Classifying data into day-side and night-side:*

After a new file is added to the directory of New_Data, the FITS files are passed into a Python script, where the FITS files are classified into day-side and night-side on the basis of solar angle. If the solar angle greater than 90° , the file is classified as night-side, otherwise it is classified as day-side.

• **Handling day-side files**

If the newly added file is day-side, we check whether the background file corresponding to that file exists or not from the "background.mean.sigma.counts" table in our database. If the background

file exists, the “catalogue” table in the MySQL database is updated accordingly, and further the “grid_data” and “kriged_data” tables are also updated. Finally, that FITS file is moved to the directory of `processed_data/dayside/YYYYMMDD`. If the background file corresponding to the newly added file does not exist, then the file is left unprocessed as it is.

- **Handling night-side files**

If the newly added file is night-side, we access the date of that file and check if the background file corresponding to that already exists or not in our database. If it exists, we update the background file data and the row corresponding to that file in catalogue, grid_data and kriged_data, background_mean_sigma_counts table, and this file is moved to “processed_data/nightside/YYYYMMDD” directory. If the background file for that night-side does not exist, a new background is generated and a new row for that file is added in the background_mean_sigma_counts table. Further, it also checks if that newly generated background file can be used to process any unprocessed FITS file present in our directory.

3) *Updation of grid_data:*

To update the grid data when a new FITS file is added, we use the “**centroid- method**”. Here, we calculate the centroid latitude and centroid longitude from the meta-data of the newly added FITS file, and check in which grid region those centroid coordinates lie. After finding the grid corresponding to that file, we will update the relative elemental ratios of that grid by taking the weighted average of the old ratio and the new ratio obtained from the newly added file.

This method ensures that we include the new observations inferred from the newly added file

without discarding the observations derived from historical data.

$$R_{\text{new}} = \frac{(W_{\text{old}} \cdot R_{\text{old}}) + (W_{\text{new}} \cdot R_{\text{new}})}{W_{\text{old}} + W_{\text{new}}}$$

where:

- R_{new} : Updated ratio value.
- R_{old} : Existing ratio value in the database.
- W_{old} : Previous weight (number of observations for the region).
- W_{new} : Weight assigned to the new observation (usually 1 for a single FITS file).

The field of weights is incremented according to the number of new files used to update the ratios of a particular grid. For e.g. If one new file is used to update the ratios, then the weight is incremented by 1.

4) *Updation of kriged_data:*

When a new FITS file is added to our directory, we identify to which kriged point it belongs and update the relative elemental ratio by the one achieved in the grid_data update. Further, we see the 10 nearest interpolated points (which currently have null values) and apply kriging again to those points.

5) *Additional exploration:*

After completing the full-fledged pipeline with MySQL database, we built another pipeline using Neo4j as our database, which is a graph-type database generally used for handling and analyzing complex relationships within our dataset. Here, our data is stored in the form of nodes and edges, and we used ‘sandbox’ to build this database for easier

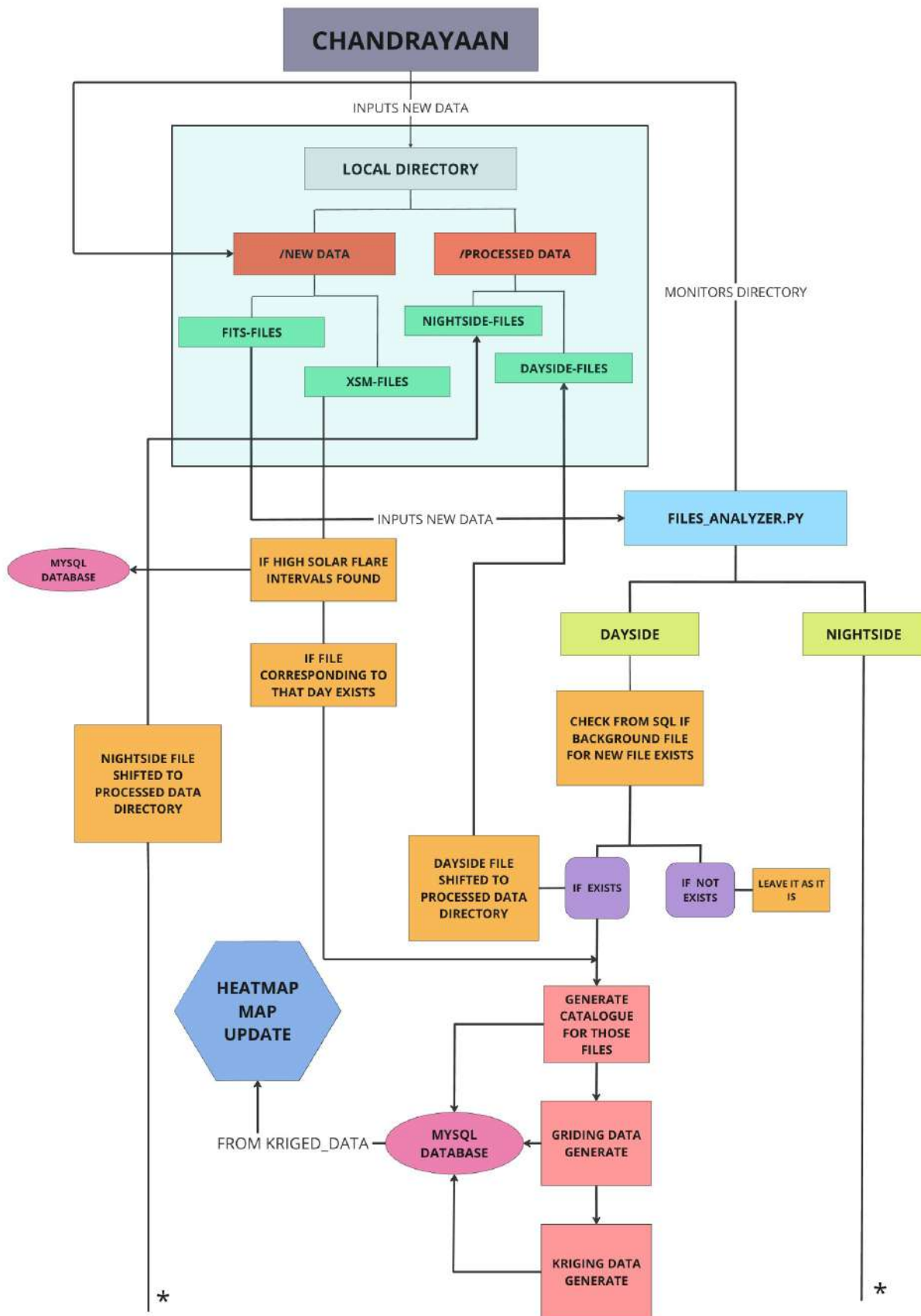


Fig. 3: MySQL Backend Pipeline

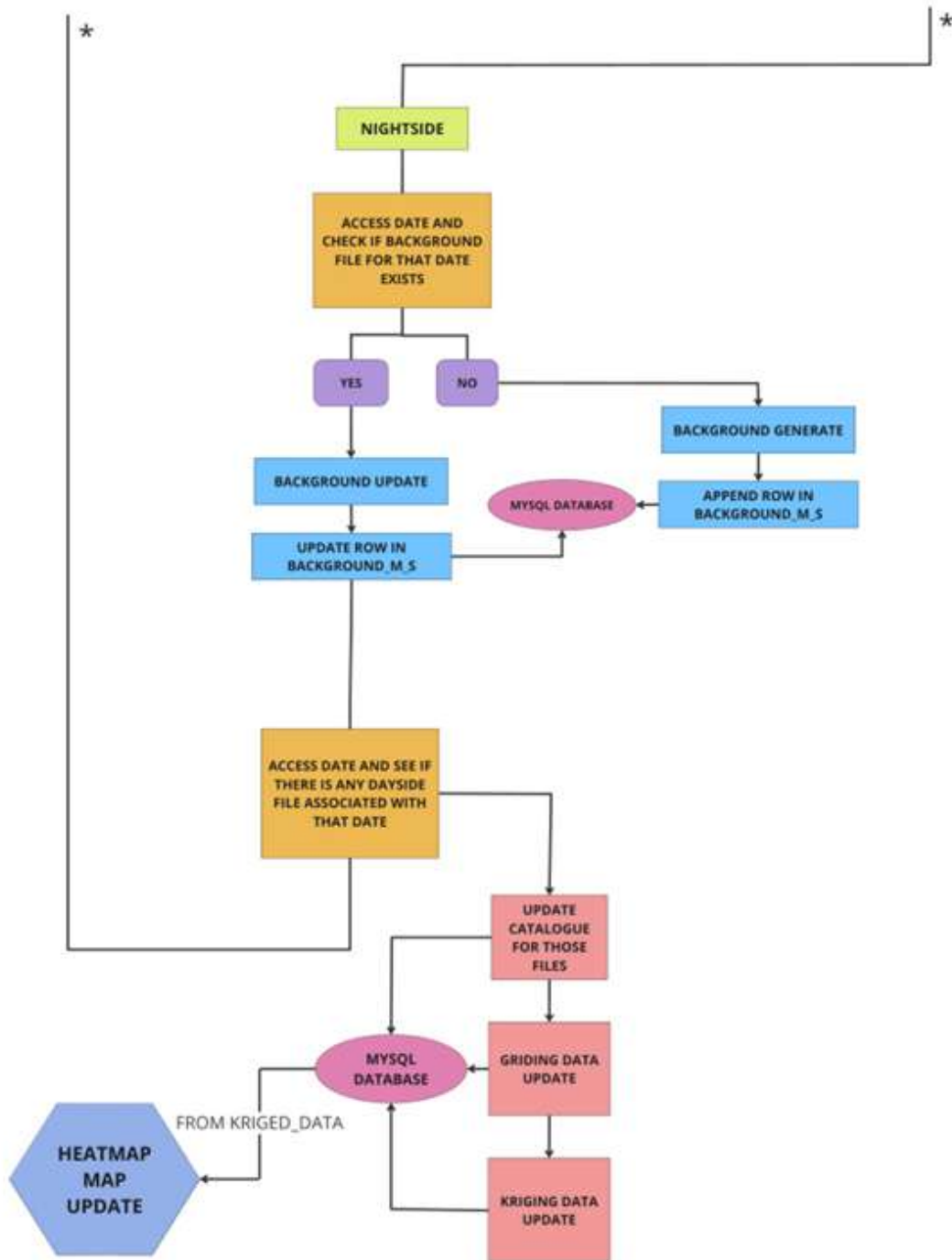


Fig. 4: MySQL Pipeline

extraction from CSV files. When a new file is added in our directory, its spatial and compositional features are transformed into a graph structure, after preprocessing of the FITs file. The logic for the updatation of `grid_data` and `kriged_data` remains the same as in the case of the pipeline with MySQL. The in-built visualization tools of Neo4j allow us to visualize the graph structure, facilitating intuitive exploration of the data. Moreover, Neo4j offers advantages such as enhanced interpretability and scalability. The visualization of data in the database can be seen figure 3.

F. Front-End Framework for Lunar Data Insights

We used **Plotly** to create interactive and dynamic visualizations of the data. With Plotly, users can zoom in and pan across data while hovering over points to get a more detailed piece of information, thus exploring and understanding the data.

We set up real-time updating visualizations using **Plotly and the Dash framework**. This automatically changes data within the visualization as the latest data is collected and processed, so users don't need to reload the page to see the latest change performed. This was one of the key reasons that made this feature so important to our project since changes in data on the lunar surface occur often, and for this reason, we enable the users to access the latest updates without having to refresh their view.

1) Representation of Elementary Ratios: We have represented the different ratios of elements such as magnesium to silicon (mg/si), calcium to silicon (ca/si), and aluminum to silicon (al/si) as **heatmaps**, which reflect their variations across the lunar surface.

`plotly.graph_objects.Heatmap` has been used to **generate 2D heat maps**, which will represent elemental ratios over a grid. This point in the grid will correspond to a given location over the lunar surface, and the color intensity at such a point will represent the value of the ratio at that point. In

each case, the color scale will automatically adjust to help the users quickly understand the magnitude of each ratio.

Using `plotly.graph_objects.Surface` which is a class in the plotly library, we have built our **3D heatmap**, which provides control over how the surface is being rendered. By adjusting features such as color scales, surface properties, and data formatting, we integrate this map.



Fig. 5: Visualization of 10 data-points from Neo4j database

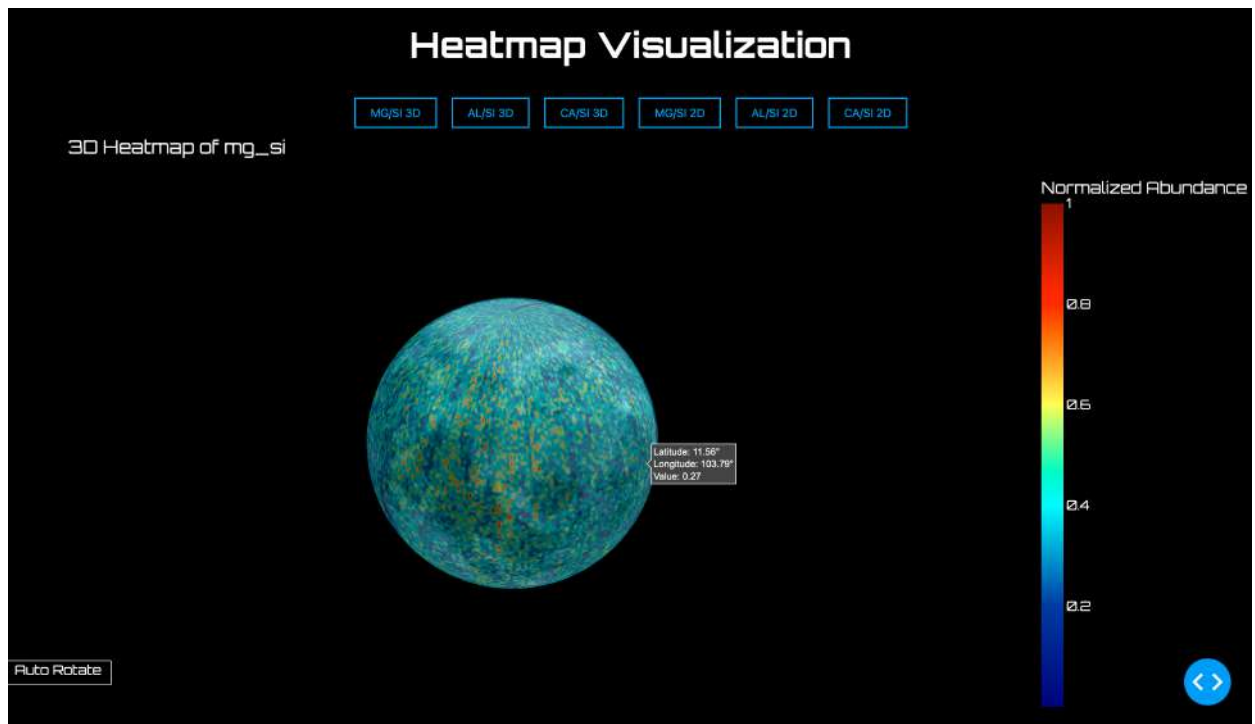


Fig. 6: Mg/Si 3D Heatmap

The heatmap utilizes a latitude-longitude grid and surface color properties to represent elemental ratios, providing users with insights into their variations.

2) User Interaction and Customization: Plotly offers interactive functionality. We included a number of features that make it easy for users to view and analyze the data.

a) Zooming and Panning:

- The lunar map can be zoomed into details in the heatmap or the 3D surface for better clarity.
- Additionally, we allow users to pan across the lunar surface to see various regions and look at useful region.

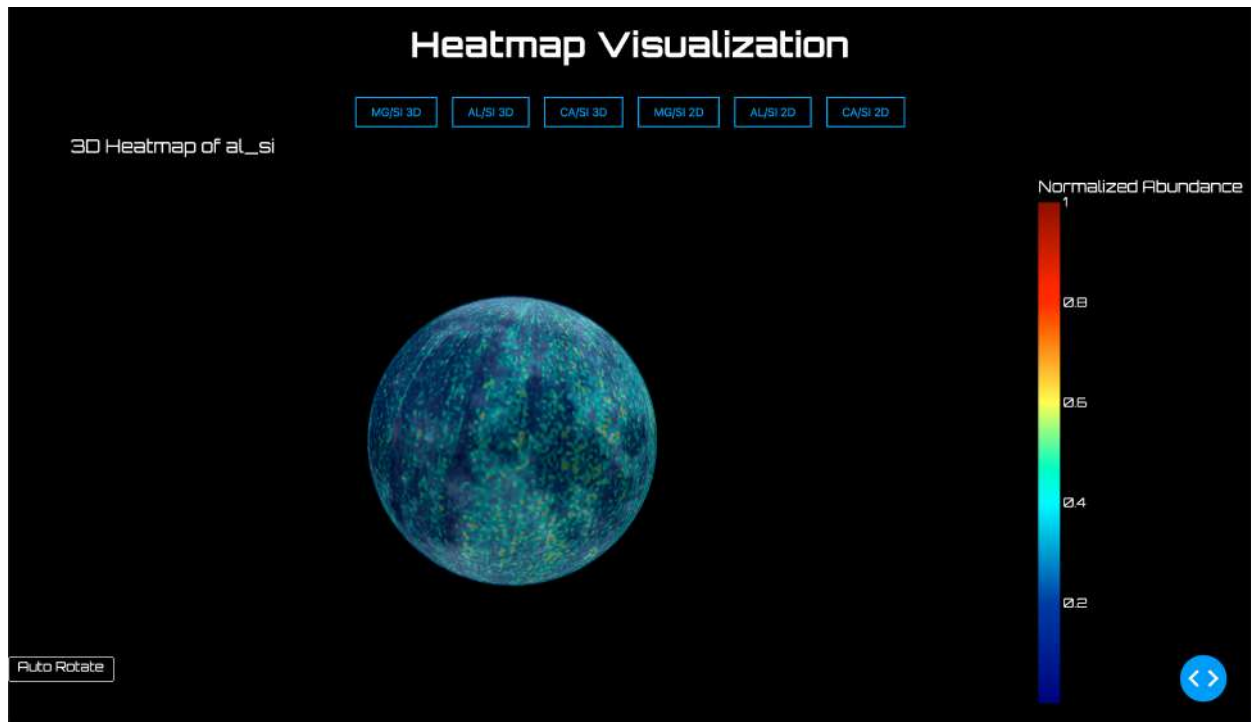


Fig. 7: Al/Si 3D Heatmap

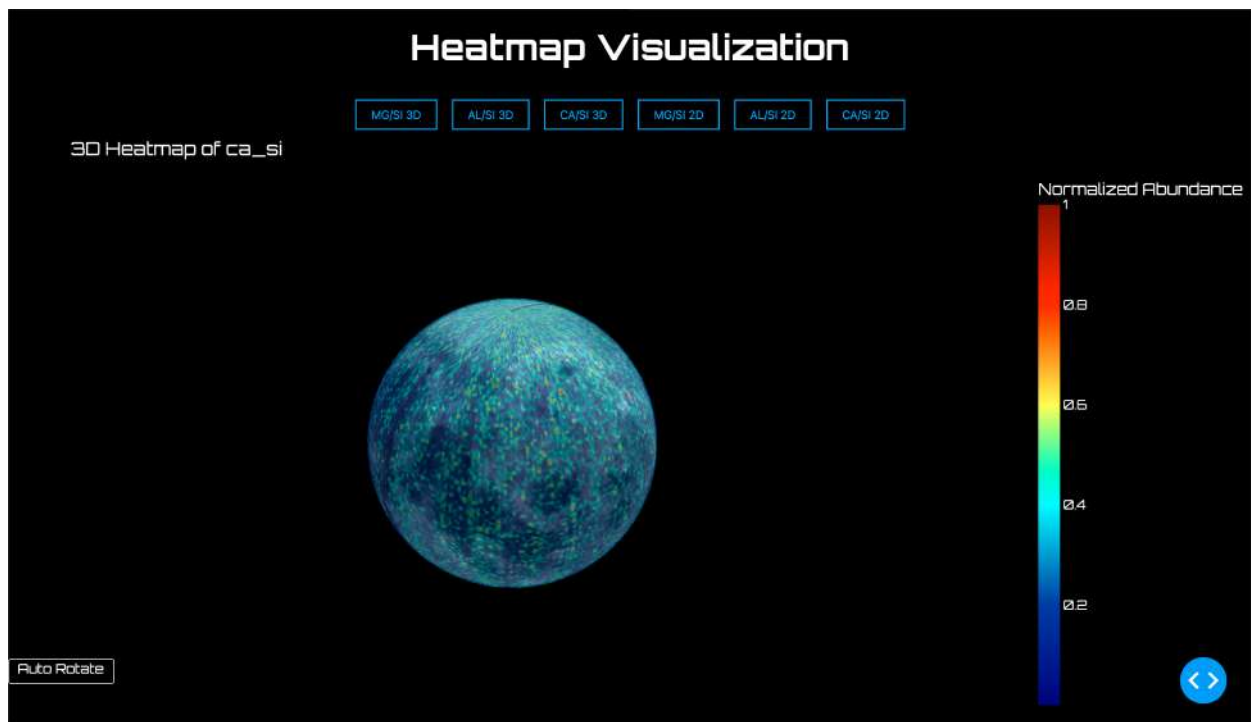


Fig. 8: Ca/Si 3D Heatmap

b) Hover Interactions: As we look upon our map, we have enabled display of information corresponding to the location being hovered over so that we know the exact trend of the corresponding elemental ratio at any surface.

3) Dynamic Updates: The weighted average method is automatically updating the grid value in the page without refreshing it so that we have an unhindered user experience. **The data we work with constantly changes**, and we are enabling the users to view the most current values when

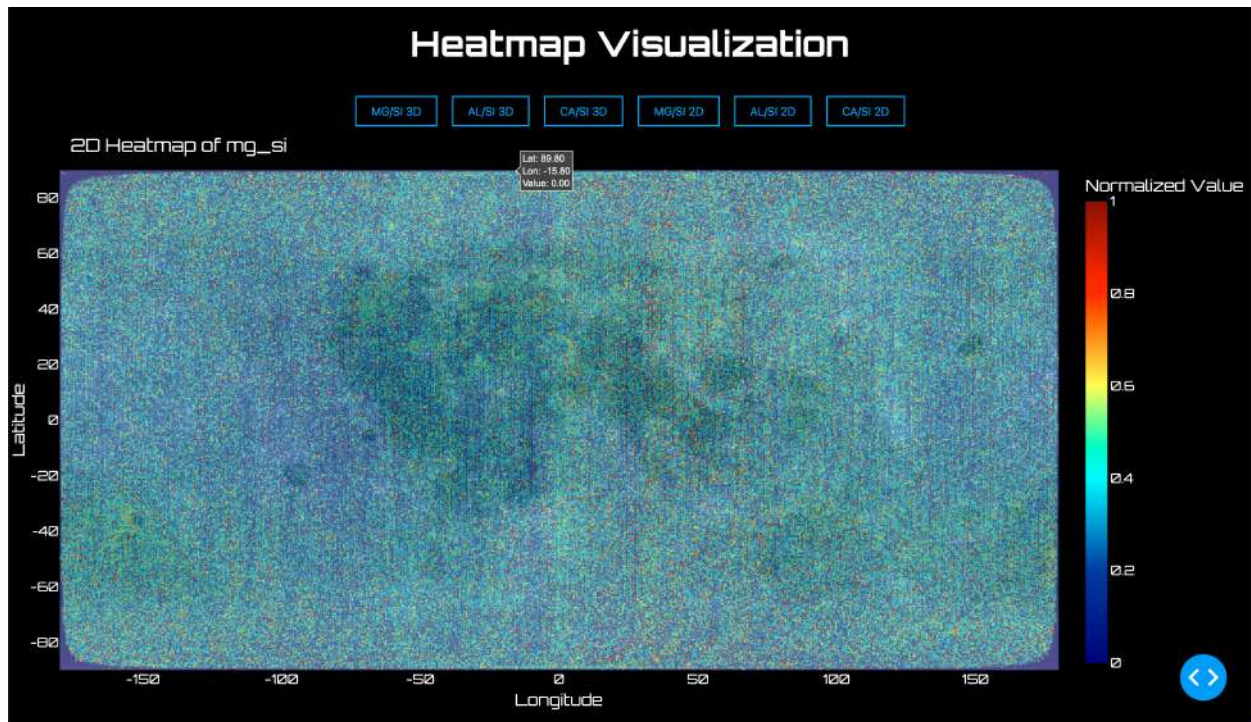


Fig. 9: Mg/Si 2D Heatmap

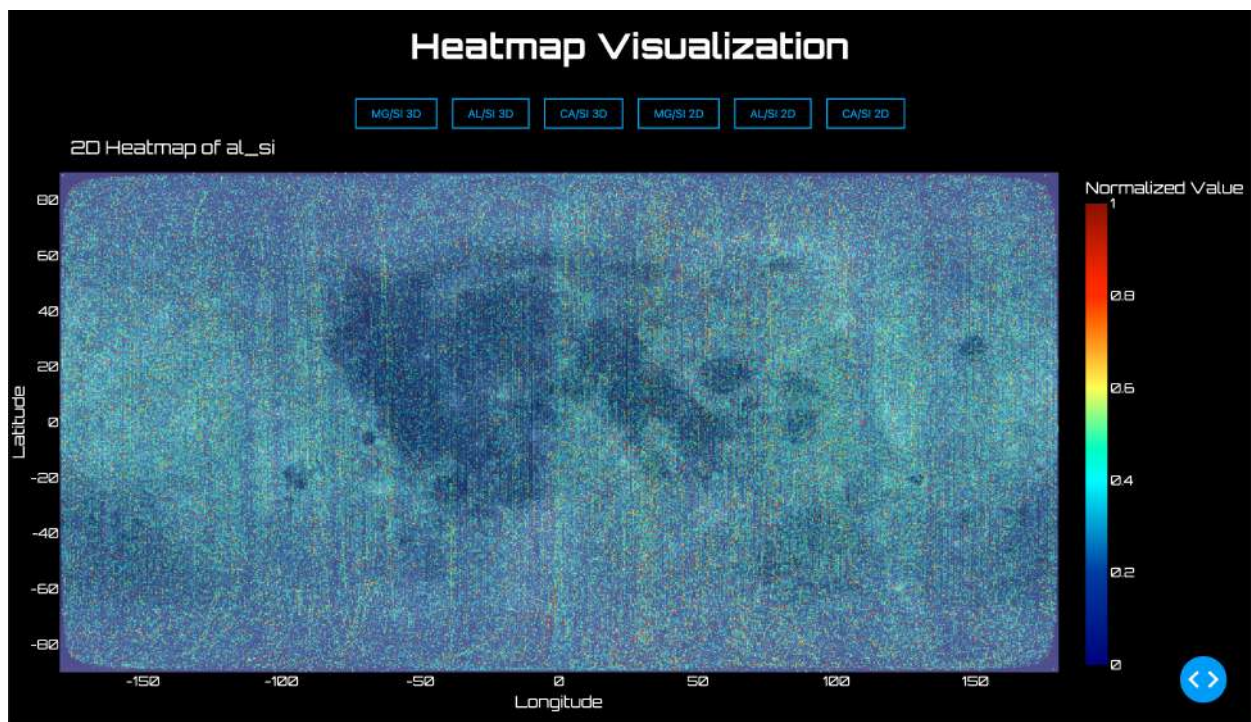


Fig. 10: Al/Si 2D Heatmap

interacting with the visualizations.

We achieved this by using the `dcc.Interval` of Dash, which calls an **update every 10 seconds** to fetch the values like elemental ratios `mg_si`, `al_si`, `ca_si` etc. This way, when a request is

made, it fetches the newest available data. It does so by overlaying updates over the old grid value, including both 2D and 3D heatmaps without interrupting user interactions.

a) *Efficient data handling:*

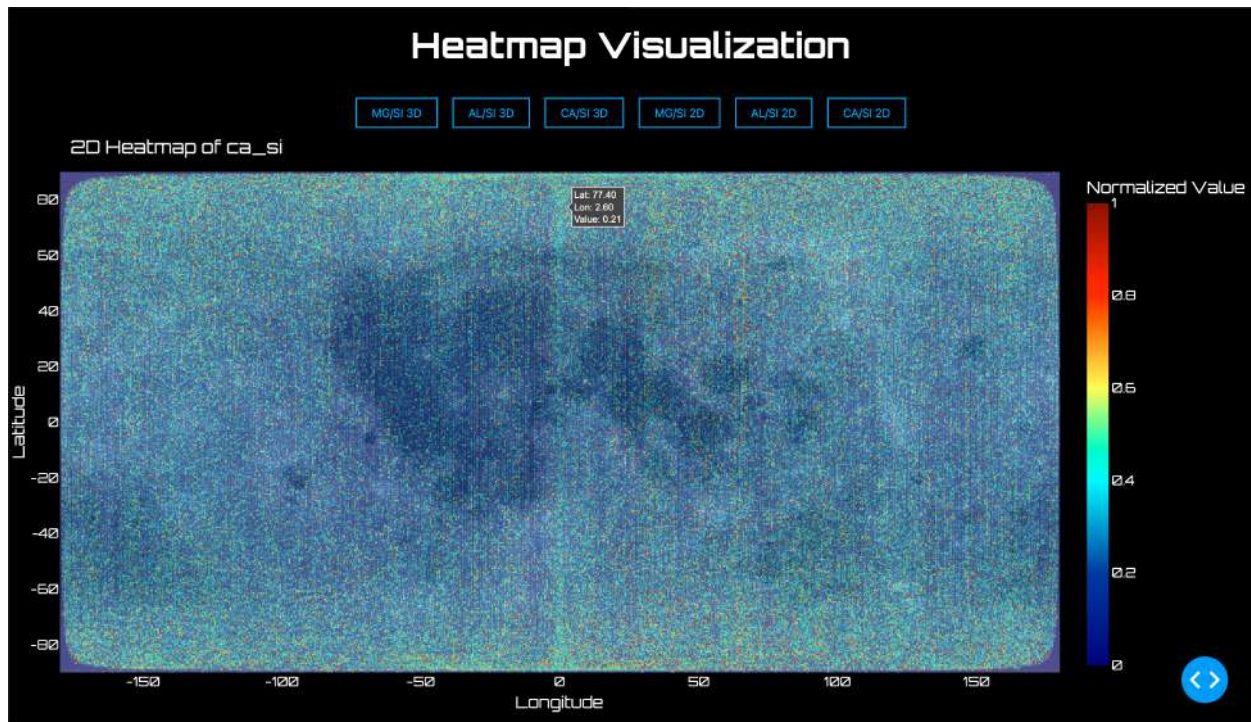


Fig. 11: Ca/Si 2D Heatmap

The process updates the new values of the grid in the existing graphs rather than recreating the complete graph as a whole, which otherwise would be of no use and make the system slow down.

b) *Smooth User Experience:*

The data updates in the background, enabling users to interact with the heatmaps— zoom, rotate, and pan while the visualization updates in real time without interruption.

c) *Instant Reflectivity of the Latest Data:*

The system continuously scans and checks for new data every 10 seconds, automatically fetching and integrating any new files into the visualizations. Therefore, moon's surface and elemental ratios are always up to date.

These interactive and dynamic features make it easier for users to engage with the data, whether they are zooming in for a closer look, using hover interactions to get more details, or relying on real-time updates to ensure they are always

working with the latest information.

Power BI Data Visualization:

a) ***Introduction:*** We used Power BI to produce data visualizations. Our catalog table resides in our MySQL database, and data is drawn from this to select the last five FITS files for visualization. The aim is to visualize elemental abundances, regional features, and energy spectra of solar phenomena.

b) ***Data Sources:***

The FITS files contain the following data:

- Elemental abundances for Al, Mg, Ca, Cr, Mn, Fe_K, Fe_L, and Si.
- Region characteristics such as geotail classification, solar angle, and solar flare activity (high/low).
- Metadata including year, date, start and end times, upper-left latitude, upper-right latitude, bottom left latitude, bottom-right latitude, upper-left longitude, upper-right longitude, bottom left longitude, bottom-right longitude

Then using the Power BI's feature of data transform, we have created two extra columns namely centroid_lat and centroid_lon by averaging the above mentioned latitude and longitude columns respectively.

Data is dynamically fetched from the MySQL catalog table, ensuring that the most recent files are always analyzed.

c) *Visualizations:*

The visualization provides three different table options, and each of them can be accessed using interactive buttons:

- **Table Options**

- **Ratios Table:**

- * The ratio of the elements (Al, Mg, Ca, Cr, Mn, Fe_K, Fe_L) to Si.
 - * One can compare the elemental composition of various regions.

- **Characteristics Table:**

- * Classifications of regions, such as:
 - Geotail region or not.
 - Dayside or Nightside.
 - Solar angle range.
 - High or Low Solar Flare intensity.
 - * Includes the FITS file name for traceability.

- **Abundances Table:**

- * Shows relative abundances of elements Al, Mg, Ca, Cr, Mn, Fe_K, Fe_L, Si.
 - * Includes additional metadata such as year, date, start and end times, and centroid latitude/longitude.

- **Pie Chart**

- Displays the relative element abundances of a selected region.
 - Helps in developing an understanding of which elements are more dominant and by how much.

- **X-ray Spectrum**

- Displays the X-ray spectrum data derived from the FITS files.
 - **Key features:**
 - * A power-law continuum showing the overall trend.
 - * Maximum peak highlighted, representing the dominant energy level.
 - * Purpose: Analyze energy distributions in solar flare or geotail regions.

d) *Case Studies:*

While we were inputting our data, we took snapshots twice of the visualizer for the most recent files. The first one was taken when we encountered consecutive files for low solar flare regions and then once we encountered consecutive files for high solar flare regions. Specifically these two have been chosen to see diverse variation in visualizations.

- **Low Solar Flare Region**

- **Description:**

A region with low solar flare intensity. Key observations:

- * Weak detection of elements Cr, Mn, Ti, Fe_K, Fe_L
 - * Region classification: Geotail, Dayside.
 - * Solar angle less than 75°

- **Visualization Snapshots:**

- * Ratios Table: Emphasizing low flare characteristics.
 - * Pie Chart: Elemental distributions.
 - * X-ray Spectrum: Highlighting the spectrum for low flare intensity.

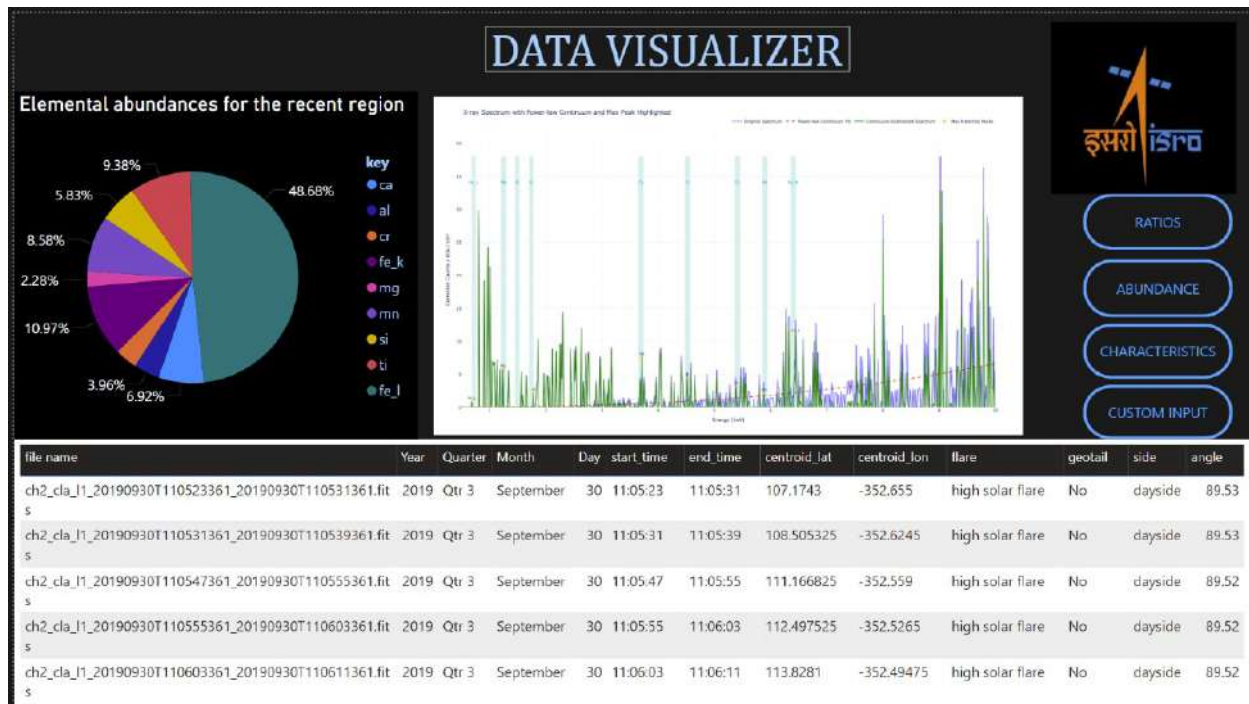


Fig. 12: Power BI Data Visualizer: High Solar Flare, depicting Characteristics

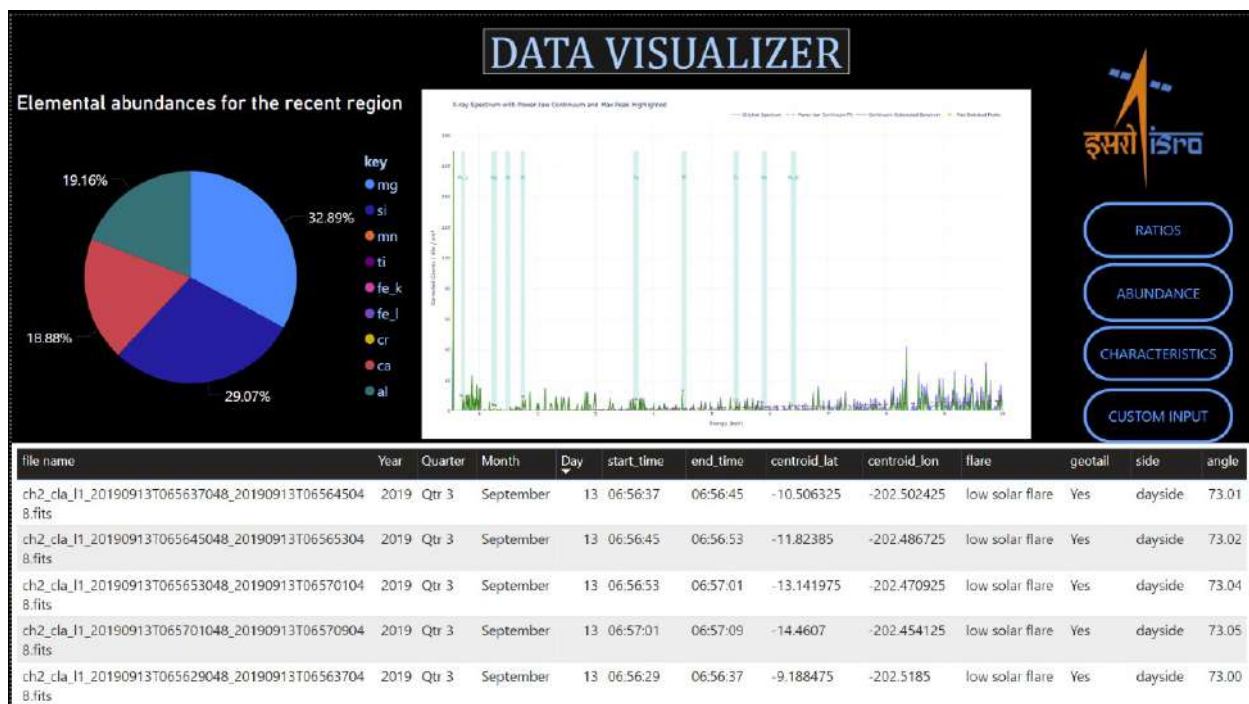


Fig. 13: Power BI Data Visualizer: Low Solar Flare, depicting Characteristics

• High Solar Flare Region

– Description:

A region with high solar flare intensity. Key observations:

- * Increased Cr, Mn, Ti, Fe_K, Fe_L ratios relative to Si.

- * Region classification: Non-Geotail, Nightside.

- * Solar angle greater than 75°.

– Visualization Snapshots:

- * Characteristics Table: Highlighting the

high solar flare region.

- * Pie Chart: High Fe concentrations.
- * X-ray Spectrum: Enhanced maximum peak. Also better peaks for elements like Cr, Mn, Ti, Fe_K, Fe_L.

e) Insights and Observations:

- Elemental ratios vary significantly between high and low solar flare regions, with Fe_K and Fe_L more dominant in high flare regions.
- Geotail regions exhibit different elemental characteristics compared to non-geotail regions.
- Solar angle influences the energy distribution observed in the X-ray spectrum

f) Conclusion:

From our current visualizations, we can get a great analysis of elemental properties like ratios, abundances and characteristics of regions like if it is a geotail region or not, etc. Expansion of this visualizer from five fits files to more is possible, we have currently just restricted it to five for the ease of display. With our future aim of coordinate-based neighbouring region analysis (discussed in detail in *Future Aspects* section), this tool will definitely prove to be quite helpful.

VIII. BEST RATIOS CALCULATION & RESULTS (DELIVERABLE 2)

In the second deliverable, we aim to extract the best ratios i.e. the most relevant from the MySQL database that gives us desirable insights about the moon's surface composition. We identify regions present on the lunar surface based on its trend of elemental ratio in its surface.

To achieve this, we query in the database and filter out elemental ratios based on their relevance to lunar geology. From this, we get a curated set of ratios which represents the compositional features of different regions which further help us with visual analysis.

A. Data Processing for Finding Best Ratios

Data processing is an important step in the lunar surface analysis which makes sure that the dataset used is consistent. We achieve this by **calculating the deviations** with respect to the Lunar Prospector Gamma Ray Spectrometer (LPGRS) as reference data.

After comparing the datasets;

- **Raw Data:** Observations that are not processed and prone to noise and inconsistencies.
- **Grid Data:** Data that is spatially structured but most likely to fluctuate on the basis of processing methods.
- **Kriged Data:** A statistically interpolated dataset designed to deal with the inconsistencies of spatial trends.

The **Kriged Data** has shown to be least deviated from the LPGRS reference showing best reliability. Hence Kriging Data was used as the basis of further analysis and its reduced deviation makes sure a robust foundation for making meaningful conclusions.

B. Analysis and Results

An analysis of six distinct regions has been carried out after identification of minimum deviation data points from the kriged data. Each of this analysis gives us unique viewpoints of the moon's geological features, history, and resource potential.

1) Mare and Highlands Plot: This analysis aims to differentiate the lunar surface into two primary types: Mare and Highland regions, on the basis of special characteristics and compositions.

This analysis calculates how variations which take place in Al/Si and Mg/Si ratios influence the classification boundaries providing flexibility and robustness to identify these distinct terrains.

a) Dynamic Threshold-Based Classification:

- **Threshold Percentiles:**

- Two ranges of percentiles were defined for Al/Si and Mg/Si ratios:

- * **High thresholds (e.g., 20th, 40th, 60th, and 80th percentiles):** To recognize areas with higher values indicating richer composition.
- * **Low thresholds (e.g., 10th, 30th, 50th, and 70th percentiles):** To identify regions with lower values.

- These ranges in percentiles make the analysis dynamic in changing the classification rules.

• **Combinatorial Exploration**

- We explored all the possible combinations of high and low thresholds for Al/Si and Mg/Si. This allows us to see how changing those thresholds would affect the classification of the different regions.

b) Classification Logic:

• **Dynamic Region Assignment:**

- A custom classification function used these rules for each data point.
 - * **Highland:** If Al/Si greater than Al_high and Mg/Si less than Mg_low, indicating anorthositic terrains typical of stable highlands.
 - * **Mare:** If Mg/Si greater than Mg_high and Al/Si less than Al_low, which is a characteristic of basaltic volcanic plains.
 - * **None:** This case has been skipped from the classification function.
- This flexible approach adjusts the classification based on the thresholds set for each combination.

• **Spatial Labeling:**

- The classified data points were plotted on a map using the longitude (cent_lon)

and latitude (cent_lat) coordinates of the centroid of the observed area.

c) Visualizing the Impact of Thresholds:

• **Scatter Plots:**

- We plot all the combinations of threshold on a scatter plot by coloring classified points differently:
 - * **Red:** Mare regions.
 - * **Blue:** Highland regions.
- These plots provide an ease to understanding of how the selected thresholds influence the classification.

• **Iterative Plotting:**

- With multiple threshold combinations, upto 16 plots, distributing the results across several pages for clarity.
- Each plot's title includes the thresholds used, such as "Al High: 40th, Low: 30th; Mg High: 60th, Low: 50th," to help with interpretation.

d) Fixed Regional Definitions:

In addition to the dynamic threshold, fixed geographical definitions have been used to establish mare regions to a sufficient degree of precision to align with known lunar geography. The above-defined mare regions are:

- **Mare Tranquillitatis:** Having its centre at 5°N, 25°E, with a radius of 250 km.
- **Oceanus Procellarum:** Having its centre at 25°N, -60°E, with a radius of 650 km.
- **Mare Imbrium:** Having its centre at 30°N, -20°E, with a radius of 600 km.
- Other mare regions, including **Mare Serenitatis** and **Mare Crisium**, were similarly specified.

For each point, we computed the **Haversine**

distance to the predefined centers of these mare regions to find out how close the point was. We classified points within given radii as belonging to the corresponding mare region, and others as part of the highlands.

e) Outcomes of the Analysis:

- **Dynamic Threshold Results:**

The visualizations showed us how varying these threshold combinations can be when it comes to classification of Mare and Highland regions, which gave us insightful outcomes about how flexible and robust the classification criterion is.

- **Spacial Region Mapping**

The regions already defined as mare regions closely matched the dynamically classified Mare areas, which in a way validated our classification by confirming its accuracy. The Highlands extended over much larger areas beyond the mare boundaries, which highlights their stability and ancient nature.

- **Scientific Insights:**

The analysis was consistent with already known mare regions such as **Mare Tranquillitatis** and **Oceanus Procellarum** which are characterized by high Mg/Si ratios. The Highlands, with their high Al/Si ratios, were **matching their results** with the **anorthositic compositions**.

Analyzing how different threshold combinations influence classifications has deepened our understanding of the Moon's compositional regions.

2) Regional Analysis (Imbrium and Australe):

The regional analysis concentrates on the defined regions of interest on the lunar surface, including Imbrium and Australe. Such analysis utilizes spatial data as well as elemental composition for an in-depth characterization of the regions. This will determine the geological history,

compositional variability, and possible availability of resources in each region.

a) Regional Analysis Process:

- **Definition of Regions:** Specific regions of interest are defined based on their central coordinates (latitude and longitude) and approximate radius, which encompasses the geographical extent of these regions. They include:

- **Mare Imbrium:** The large mare region with a rich history of volcanic activity in the past.

- **Australe:** A highland region near the lunar south pole and is stable.

- Other regions include prominent mare areas such as **Tranquillitatis**, **Serenitatis**, and **Oceanus Procellarum**, and smaller well-known features such as **Mare Crisium** and **Mare Nubium**.

For each region, spatial boundaries were defined using:

- **Center Coordinates:** Representing the approximate center of the region.

- **Radius** (in kilometers): Determining the area around the center to include within the region.

Once these were labeled, the analysis would calculate the mean and the variance of elemental ratios per region.

The elemental ratios considered are as follows:

Al/Si: aluminum abundance, an indicator of the **anorthositic composition highland**
Mg/Si: it will directly relate to the basaltic material, a characteristic signature of s.
Fe/Si, Ti/Si, Cr/Si: indicators of the **metallic** and **volcanic components** in the composition. Ca/Si and Mn/Si: additional ratios that reflect various geological processes.

These will help to identify dominant composition in each region. The compositional variability within a region can thus indicate geological diversity or even past activity.

Result Compilation The outcome of this analysis was compiled into a CSV file, where:

- ➔ Rows are referring to different regions (such as Imbrium, Australe, Tranquillitatis).
- Columns include the **mean** and **variance** for each elemental ratio; thus, it captures not only the average composition but also the variability in every region.

This structured output, therefore, allows direct comparison across regions and will lay the foundation for further and more advanced analyses such as resource potential or historical reconstruction.

b) Justifications for Regional Analysis:

- **Imbrium:**
 - **Significance:** Mare Imbrium is one of the biggest moon maria and is also considered to be an important research location regarding volcanic activities. High Mg/Si and Ti/Si ratios show its basaltic nature and significant lava flow extent.
 - **Findings:** A great variance in elemental ratios that suggests complex volcanic activities occurred and a rich source exists.
- **Australe:**
 - **Significance:** This location is in the near area of the lunar south pole and Australe is related to understanding the ancient and intact terrains, along with preserved water-ice potentials.
 - **Findings:** Low variance in elemental ratios indicates a stable composition, providing evidence for the hypothesis that water-ice deposits exist in shadowed areas.
- **Other Regions:**

- **Tranquillitatis and Serenitatis:** These are sites of intense volcanic history and characterized by the elevated Mg/Si and Ti/Si ratios.
- **Oceanus Procellarum:** It is the largest mare region, kind of a hotspot for studying basaltic volcanic flows and ISRU potential due to its rich titanium content.
- **Highlands:** Examples include Mare Crisium and Mare Nubium. These regions are small and serve as interfaces between mare and highland terrains.

c) Outcome of the Regional Analysis:

Comprehensive Compositional Profiles: The mean and variance statistics clearly show the predominant materials and compositional variability of each region.

- **Geological Implications:** Regions such as Imbrium and Procellarum were confirmed to be volcanic hotspots with high Mg/Si and Ti/Si ratios.
- Regions such as **Australe showed stability** and low compositional variability, in accordance with their ancient terrain characteristics.
- **Resource Potential:** Regions with high Mg/Si and Ti/Si ratios were marked as possible ISRU sites, while stable highland regions were noted as probable sites for the preservation of water-ice.

The application of regional analysis bridges the gap from raw data to meaningful geological insight by providing rich and better insight into the moon's surface. Combining spatial classification with elemental composition opens the diversity of the moon, leading pathways toward future exploration.

3) Water indicators: Low Mg/Si and Fe/Si Ratios: Regions likely to have water in them on the moon should depict low elemental variability

and not very significant volcanic influence. For such conditions, some substitutions are possible using Mg/Si and Fe/Si ratios.

a) *Process:*

- **Threshold Selection:**

- Mg/Si Low (1st percentile): Focuses on areas with low magnesium content, excluding areas rich in volcanic basalts.
- Fe/Si Low (1st percentile): Regions of lowest Fe abundance typically correspond to quiet, non-fragmented terrains.

Selection of 1st percentile of the Mg/Si and Fe/Si ratio sets boundaries to emphasize those terrains with the lowest perturbed regions, further **eliminating probable volcanic or basaltic terrains**, in which the H-I may not have a preserved record. These low ratios are indicative of geological regions with minimum disturbances, like highlands or shadowed polar regions which are optimal conditions for water-ice retention.

This threshold ensures that the analysis extracts out the promising regions for potential water resource.

- **Data Extraction:**

Points satisfying both conditions were located and graphed.

- **Contextual Insights:**

These regions were cross-referenced with lunar topographic and temperature data to refine the potential for water-ice preservation.

- **Outcome:**

From the above analysis we can say that highland regions and the regions near the lunar poles can act as potential water-ice reservoirs. These can be helpful in future research where we can infer about the presence of water on lunar surface which can be a huge step towards water resource for life support.

4) *Indicators of Volcanic Activity: High Ti/Si Ratios:*

Volcanic activity on the Moon is highly correlated with high titanium-bearing regions, mainly in the form of ilmenite. The Ti/Si ratio is an effective indicator for identifying volcanic regions, particularly within the mare of the Moon.

a) *Process:*

- **Threshold Selection:**

We chose a threshold based on the 95th percentile, which we called `ti_si_high`, to focus on the most extreme Ti/Si values.

This will bring to light the areas with the highest enrichment of titanium, the ones that are most representative of volcanic activity. This ensures that only the **most extreme volcanic regions** are considered after removing the more moderate variations.

The 95th percentile of the Ti/Si ratio was chosen since it tends to highlight the exceptional titanium enrichment that characterizes volcanic activity at the lunar surface. Such titanium-rich ilmenite basalts in mare regions are thus defined by this **high concentration**.

Targeting the upper 5% of Ti/Si values ensures that areas with the highest probability of volcanic origin are selected and captures major volcanic plains such as **Oceanus Procellarum** and **Mare Tranquillitatis**.

This approach eliminates all kinds of noise associated with moderately varying Ti/Si ratios, which will considerably strengthen the credibility of identified **notable volcanic features** and present the complete view of thermal and geologic evolution of the moon.

- **Explanation about the Ratio Ti/Si:**

Titanium is abundantly found in the rocks of **volcanic basalts** along with ilmenite-rich resources. Silicon, although has somewhat uniform abundance throughout the Lunar Sur-

face, serves as the steady denominator for this particular ratio.

- **Data Extraction:**

Extract all data points in the data record for which Ti/Si exceeds the 95th percentile and plot them.

b) Results:

Regions which most probably contain major volcanic plains like parts of Oceanus Procellarum or Mare Tranquillitatis contain good records of the thermal history through time of the Moon itself as the titanium-rich basalts are a remainder from older lava flows.

5) Impact Features: High Fe/Si and Cr/Si Ratios

: It does impart a noticeable compositional change when impact events occur. Fe and Cr are then possibly introduced or redistributed within the Earth's crust. Such elemental ratios as Fe/Si and Cr/Si that are elevated can become quite useful in the determination of possible impact features.

a) Process:

- **Cutoff Choice:**

- Fe/Si High (75th percentile): Marks areas where the iron content is anomalously high, indicative of **meteoritic deposition**, or of the exhumation of iron-rich subsurface material from beneath the lunar crust.
- Cr/Si High: 95th percentile Chromium enrichment is less frequently observed but makes a good fingerprint for **high energy impacts**

- **Data Extraction:**

Data points for which both of the two conditions (Fe/Si greater than 75th percentile and Cr/Si greater than 95th percentile) are met are extracted here, thus emphasizing locations where elemental **anomalies caused by impact are maximal**.

b) Results:

On a cross check we concluded that we were able to point out both previously established impact sites and potentially new ones as well.

6) ISRU Ore Deposits: High Mg/Si, Fe/Si, and Ti/Si Ratios:

The location of ore deposits will be essential for upcoming future lunar missions. Regions with high magnesium, iron, and titanium will be very useful as they are mainly used in construction and fueling processes.

a) Methodology:

- Identify important Areas

- High Mg/Si Ratio (Top 25%): Regions that are high in magnesium usually contain basaltic material.
- High Fe/Si Ratio (Top 25%): This marks areas of high potential for metallic ore, especially iron-bearing deposits.
- High Ti/Si Ratio (Top 5%): These regions have abundant ilmenite that will be needed for the extraction of oxygen and other ISRU processes.

Targeting areas at the 75th percentile for Fe/Si and at the 95th percentile for Cr/Si may highlight areas of contrasting compositional patterns that have a tendency to be associated with **impacts of meteorites**.

High Fe/Si values suggest iron-rich environments that can be produced either by **deposition of meteorites** or the **uplift of subsurface iron-rich** layers resulting from impact events. This will give better identification of rare chromium anomalies with a 95th percentile threshold and is indicative of high-energy impacts. With this combination, analysis be done to identify regions that are pronounced in their elemental anomalies due to impacts which reduces the number of false positives.

On top of this, we not only identify crater impacts like **Tycho** and **Copernicus** but also reveal new potential sites of impact helping us to make a throwback to lunar bombardment history.

- **Data Extraction:** Regions that met all three requirements were chosen so that regions identified by the thresholds would be resource-rich.
- **Strategic Mapping:** Overlap of thresholds was mapped to show regions as resource-rich.
- **Outcome:** Regions in mare basalts in Mare Imbrium and Oceanus Procellarum were identified as **hotspots for ISRU** potential. Findings give a roadmap for the extraction of lunar resources toward sustainable exploration.

By conducting these six detailed analyses, we were able to take out a very large amount of insights into the moon's geography, resources and history of various activities. We have performed analysis from classifying mare and highland regions to identifying volcanic activity, water presence, and ISRU opportunities, where each step makes it even clear for an in-depth understanding of the lunar surface.

IX. SUB-PIXEL RESOLUTION (DELIVERABLE 3)

High-resolution surface mapping of the lunar body is essential in order to bring out fine-scale details and subtle variations in elemental composition, which may indicate characteristic geological processes or compositional anomalies. Our algorithm enhances intensity ratios up to a **sub pixel resolution** and transitions from the current $0.4^\circ \times 0.4^\circ$ resolution to a finer $0.1^\circ \times 0.1^\circ$. The grid of latitude and longitude forms our basic framework. We refine the ratios by breaking **larger spatial blocks** into **smaller pieces**, which increases the resolution of the data significantly, allowing for a more detailed analysis of the lunar surface. With this higher resolution, we hope to reveal features that are not yet visible at coarser resolutions.

A. Methodology

Simplified methods like **Naive Averaging** and **Single Patch Assignment** were tested to increase the resolution of the lunar base map for high resolution grids. While these methods were working with good efficiency in terms of computational operations, they were not able to provide us with accurate and quality data whereas **Weighted Overlap Averaging** is a more generalized approach which also gives better results.

The following is a detailed analysis of why we chose Weighted Overlap Averaging as our algorithm for Sub-pixel Resolution.

1) *Naive Averaging:*

a) *Mechanism:*

All overlapping patches receive the same weight when their values contribute towards a grid cell. It computes the average, in particular, irrespective of the number of overlaps.

b) *Strengths and Weaknesses:*

• **Strength:**

The process of averaging is quite simple, so it is computationally efficient. The overlapping patches are treated uniformly so that no patch goes out of favor.

• **Weakness:**

When the values in overlapping data patches are quite different, uniform averaging mindlessly smoothens over differences. Resolution sharpening is lost; equalizing all overlaps resets finer distinctions of the grid, so it can no longer capture fine sub-pixel variation.

c) *Limitations:*

Naive averaging captures the best trade off between simplicity and accuracy but does not really preserve much of what the overlapping patches have in common. Because there's **no**

proportional weighting, what happens is that the grid may not reflect where data should actually be with the distribution in regions like the regions with densely overlapping data.

2) *Single-Patch Assignment*:

a) *Mechanism*:

The grid cell assumes the value of the dominant or primary patch that it overlaps. Subsequent overlapping patches are completely ignored.

b) *Strengths and Weaknesses*:

- **Strengths:**

- Extremely simple and computationally very fast.
- No averaging needed.
- Facilitates ease of implementation.

- **Weakness:**

- **Dominant Patch Bias:** The grid reflects the first or largest patch dominantly over other useful information overlapping with patches.
- **Low Resolution in the Gaps:** The spaces where patches do not overlap are quite large, thus generating non-overlapping abrupt transitions to create a grid that lacks continuity and precision.
- **Low Data Fidelity:** The removal of overlapping patches leads to a significant loss of data richness.

c) *Limitations*:

The Single Patch Assignment method oversimplifies overlap handling. It does not account for supplementary data contributions, and the underlying distribution is poorly represented in areas with multiple patches or sparse coverage in the grid.

3) *Weighted Overlap Averaging at Sub-pixel Resolution*:

When interpreting astronomical data, for example, elemental abundance maps created from twelve FITS files, one would expect the data to be presented over a regular grid. However, the data patches usually overlap with many patches contributing to a particular grid cell. Thus, Weighted Overlap Averaging ensures that in the end, every grid cell is indeed a faithful representation of how all the overlapping patches collectively function, leading to a highly resolved and smooth grid at $0.1^\circ \times 0.1^\circ$ sub-pixel resolution.

a) *Sequential Process*:

- **Setting up grid data:**

A uniform grid of cells of desired dimensions, sized $0.1^\circ \times 0.1^\circ$ in our case, is created and is made to cover the entire Lunar surface.

- **Overlap Resolution:**

The patches that overlap into the same cell of the grid, rather than replacing or omitting the pre-existing values, get added up to calculate weighted averages of values for the cells of a grid. Here is how it works:

- **First Installment:**

For the first patch affecting a grid cell, the grid cell takes directly the values of the patch.

- **Further contributions:**

Should the grid cell include contributions from n patches, the updated value is:

$$\text{Update Value} = \frac{n \cdot \text{Previous Value} + \text{New Patch Value}}{n+1}$$

This results in a proportionate influence for every patch, ensuring no individual patch dominates.

- **Updating Counts**

It increments the count of contributions for each patch that contributes to a grid cell. These counts are then used to compute the

weighted average and track how much data has affected each grid cell.

This will capture the high-resolution details of a grid by incorporating information from every patch that overlaps the patch of interest. The higher the number of contributions to the value by more patches, the higher the resolution of the $0.1^\circ \times 0.1^\circ$ map.

b) Method Reliability:

- **Accuracy:**

In including all data from each patch, nothing is omitted. Weighted averaging balances out all influences of overlapping patches.

- **Scalability:**

Resolution map can be generated for any value and not just $0.1^\circ \times 0.1^\circ$.

- **High Definition:**

It generates a total sub-pixel representation through the refinement of each subsequent patch.

- **Overlap Management:**

Overlaps instead of conflicts makes the grid more accurate by adding more data.

c) Illustration:

If we imagine trying to make a map of the lunar elemental distribution, each FITS file will provide a portion of the data, and together they will cumulatively cover the entire lunar surface. Some regions are covered by multiple patches, resulting from overlaps. This methodology ensures that grid cells in such areas portray all contributing patches accurately by calculating weighted averages for every elemental value, thereby providing a reliable high-resolution map to the research community of the lunar surface.

The Weighted Overlap Averaging for Sub-Pixel Resolution is a robust technique for mapping data from overlapping patches onto a uniform grid. It ensures that the resulting grid not only

contains all the available data but also represents it accurately and at a higher resolution. This technique is especially useful in planetary science, where high-resolution maps are necessary for proper analysis and exploration.

d) Key Advantages of Weighted Overlap Averaging:

- **Proportional Influence:**

Unlike Naive Overlap Averaging, Weighted Overlap Averaging accounts for the number of contributions to a grid cell. Patches of overlap affect the grid in proportion to their cumulative contribution, thereby retaining finer discrimination.

- **Improved Resolution:**

Weighted handling ensures that every patch contributes something to the grid while incorporating resolution to fill out holes in Single-Patch Assignment.

Although more computationally intensive than the other two approaches, Weighted Overlap Averaging achieves very high data fidelity without any excess of complexity.

It catches overlaps dynamically rather than equitably treating all patches or discarding them; therefore, it will always provide an appropriate representation of elemental distribution.

X. UNCERTAINTIES

The project has some uncertainties which may affect the accuracy and reliability of results to some little extent. The sources of these uncertainties include human error, methodological constraints, data limitations, etc.

1) General Human Error: Human error is always an element of uncertainty in any process where human effort is involved. Throughout the project, data entry, interpretation, and analysis are examples of such tasks where errors may have crept in. These may be errors due to coding, misinterpretation of data, or inconsistencies while handling data at

various stages. Inaccuracies could be introduced in the results in case subjective judgments are made during the processing or visualization of the data.

2) **LPGRS Data:** The LPGRS data was prepared manually by color mapping from a heatmap presented in a research paper. This method may have several errors. The interpretation of color gradients is subjective, which may lead to inconsistencies between persons extracting data. Another factor with variations in the visual representation of a heatmap and difficulties in definable boundaries for the local peaks and gradients does further go to increase the chances of inaccuracies. Thus, the LPGRS data could not be able to exactly represent phenomena underlying the heatmap analysis, and hence there exists an element of uncertainty.

3) **GOES Data:** The absence of GOES data introduces a lot of uncertainty. GOES data, which contains real-time atmospheric and environmental information, wasn't included in the project. This means that some environmental factors, such as atmospheric pressure, humidity, and temperature, that could have been easily captured by GOES satellites were not considered in the research. This data being absent limits what the model can do since it might not be able to give highly accurate or complete results.

4) **Resolution and Scale Limitations:** Resolution-cut input data determines how accurate the analysis should be. Low spatial or temporal resolution data may not pick up details that may have gone unreported and, therefore, resulted in generalized conclusions that miss important variations. Scale of analysis may also determine outcome since local patterns may not be clearly observable in larger analysis.

5) **Model Assumptions and Simplifications:** Several assumptions and simplifications were made during the process of modeling to help analysis and computations. For instance, in some cases leaving out some variables in the

environment that may influence results. Though it can be practical, often what has been simplified may not conform to reality. Then, the entire model as built would not be representative if it were to be applied in real life and would therefore possess an uncertain predictive capability.

6) **Data Quality and Preprocessing:** The quality of the input data is of utmost importance and plays an important role in determining the accuracy of the final results. If the new data point have noise, or error of any kind then that inconsistency could propagate through the entire analysis and distort our final result. Also, the errors and biases during the preprocessing procedure can compromise with the integrity of the output.

7) **Temporal and Spatial Variability:** The results of the project are subject to temporal and spatial variability. Environmental conditions can change over time and across different locations, which may not be fully captured in the analysis. This can lead to discrepancies between the model's outputs and real-world observations, especially if the analysis is based on static data or short-term observations. Temporal shifts in environmental variables and regional differences could contribute to variations in the outcomes, highlighting the need for broader, longitudinal data for more accurate conclusions.

XI. CHALLENGES AND LESSONS LEARNED

One of the major challenges we faced was during Grid interpolation. For this, we applied several probabilistic methods which included various Machine Learning models, whose information is stated below,

1) **Auto-encoders with Attention Mechanisms (Spatial and Self-Attention):**

a) **A Basic Idea:**

Auto-encoders are a type of neural network designed to provide an efficient representation of data by compressing the input into a smaller, meaningful size and reconstructing it into its original form.

For our task, the auto-encoder aimed to identify

patterns in the given data and reconstruct the missing values in the lunar grid.

b) Improving Auto-encoders with Attention Mechanisms:

To enable the auto-encoder to handle sparse and uneven data more effectively, we incorporated attention mechanisms. These mechanisms allow the model to focus on the most important parts of the input data.

Spatial Attention:

Spatial attention assigns importance weights to different regions of the input grid, enabling the model to focus on areas with more reliable data. This mechanism emphasized well-sampled regions of the grid, helping the model to fill in gaps in nearby areas.

Self-Attention:

Self-attention enables the model to learn relationships across all parts of the grid by comparing every part of the input with every other part. Self-attention exploited correlations between distant regions, capturing global spatial patterns in the lunar grid.

c) Key Advantages:

By combining spatial and self-attention mechanisms, the auto encoder was designed to tackle both local and global relationships in the data:

- **Spatial Attention** enhanced focus on dense regions with reliable data.
- **Self-Attention** captured global dependencies, allowing distant but relevant regions to inform each other.

This approach seemed particularly suitable for the uneven data distribution in the lunar grid.

d) Challenges and Limitations:

Despite its promise, the approach suffered from significant issues:

- **High Variance in Sparse Regions:** In areas with little to no data, the probabilistic nature of the neural network introduced significant uncertainty, resulting in unreliable predictions.
- **Bias Toward Dense Regions:** Spatial attention overly emphasized well-sampled regions, leading to overfitting and poor performance in sparse regions where accurate predictions were most needed.
- **Limited Global Modeling:** While self-attention aimed to identify global patterns, many correlations proved insignificant due to the spatial complexity of the lunar grid and the dominance of local trends.
- Compared to more straightforward techniques like Kriging, this method dramatically increased computational cost and complexity, making it harder to justify.

2) Deep Convolutional Neural Networks (Deep CNNs):

Convolutional Neural Networks (CNNs) are a class of neural networks specifically designed to detect spatial patterns and hierarchies in data. CNNs excel at analyzing images or spatial grids where features are detected at various levels of granularity.

The output is a feature map, highlighting the patterns detected by the filter.

a) How Deep CNNs Work: In deep CNNs, this process is repeated across multiple layers, each with its unique set of filters.

- **Lower Layers:** Detect simple features like edges or textures.
- **Deeper Layers:** Identify more complex patterns by combining simpler elements detected earlier.
- **Fully Connected Layers:** Pass the extracted features through fully connected layers to generate predictions.

In our case, these predictions aimed to fill in missing ratios in the lunar grid.

b) Key Advantages:

For the lunar grid, deep CNNs were used to predict missing values based on the spatial structure of the data. They contributed in the following ways:

- **Sparsity in Patterns:** CNNs were adept at identifying local spatial patterns, such as clusters or trends within the data.
- **Hierarchical Representation:** The multi-layered architecture allowed CNNs to combine small-scale patterns (e.g., local trends in ratios) into larger structures, effectively capturing regional characteristics.
- **Data-Driven Prediction:** CNNs leveraged the available data to learn the relationships between nearby values, enabling predictions for regions with missing data.

c) Challenges and Limitations:

Despite their potential, CNNs encountered several challenges in this task:

- **Local Focus:** CNNs captured local spatial patterns effectively but struggled with long-range dependencies, making it difficult to model global relationships between distant regions.
- **Overfitting to Dense Regions:** Similar to other machine learning methods, CNNs overfit well-sampled regions, leading to poor performance in sparse regions where predictions are most needed.
- **High Variance:** In sparse locations with limited data, predictions were highly ambiguous, as the model lacked sufficient context to form reliable estimates.

3) Random Forest and XGBoost Regressor:

a) Random Forest:

Another very popular machine learning technique that uses a set of decision trees to generate

predictions is called Random Forest. Each tree is trained over a collection of data and features, and the output is computed by summing all the predictions generated by the trees. This type of ensemble approach reduces overfitting and improves the generative capability of the model.

How Random Forest Would Have Assisted:

Random Forests provided an intuitive yet powerful means of dealing with missing values for the lunar grid:

- **Robustness to Noise:** Because Random Forests make multiple trees average their predictions, noisy data or inconsistent values could smooth out errors in the final predictions.
- **Dealing with Nonlinear Relationships:** Decision trees are effective for detecting complex, nonlinear relationships common within spatial data.
- **Feature Importance:** Random Forest automatically generates feature importance from the input variables, which helps determine the spatial qualities important in identifying missing ratios.

b) XGBoost:

Extreme Gradient Boosting, or XGBoost, is an advanced form of machine learning algorithm where decision trees are added iteratively to minimize residuals created by the previous set of trees. Unlike Random Forest, where decision trees are created without dependencies on each other, XGBoost optimizes performance through a combination of loss and regularization terms.

c) Key Advantages:

XGBoost was very attractive for the lunar grid because:

- **Iterative Refinement:** By focusing on residual errors, it iteratively improved predictions, making it capable of refining estimates for

missing values.

- **Handling Sparsity:** XGBoost has built-in optimizations for sparse data, which align well with the unevenly sampled lunar grid.
- **Regularization:** The regularization terms helped control model complexity, reducing overfitting to dense regions.

d) *Challenges and Limitations:*

- **Overfitting Sparse Regions:** Both Random Forest and XGBoost failed in sparse regions where there was little or no data. Their overdependence on the patterns in existing data resulted in poor generalization in sparse areas, making the predictions unreliable.
- **High Variance:** For Random Forest, averaging the predictions reduced some variance, but not enough to result in consistent results in sparse regions. XGBoost, being iterative, often amplifies uncertainty in sparse regions with limited data points.
- **Local Dependency:** Both models were highly dependent on the local characteristics of the data. This led to a failure to capture long-range spatial dependencies, which are critical for accurate interpolation in spatial tasks.

4) *Physics-Informed Neural Networks:*

Physics-Informed Neural Networks (PINNs) are the amalgamation of traditional neural networks with physical laws defined using partial differential equations (PDEs). The aim of PINNs is to predict outputs by satisfying the underlying physical laws that rule the system.

a) *Loss Function of PINNs:*

The loss function of PINNs consists of two components:

- **Data Loss:** The error between the prediction and observed data.

- **Physics Loss:** Ensures that the predictions obey the constraints of the governing PDEs.

The first term ensures that the model predictions meet the PDE, while the second term minimizes the error between predictions and actual observations.

b) *Key Advantages:*

PINNs would have been a potential tool in reconstructing the lunar grid's missing values because:

- **Physics:** The lunar grid likely satisfies specific laws derived from physical knowledge concerning some lunar terrain or spatial dependence. PINNs will directly enforce these laws into the models, thus guiding predictions within regions with little to no observations.
- **Improved Predictions at Areas of Sparsity:** By applying physics-based constraints, PINNs are better suited for making predictions in sparse areas.
- **Generalization Across Data Gaps:** In contrast to purely data-driven approaches, PINNs enforce physical consistency, enabling generalization of predictions across regions with poor sampling.
- **Flexible Integration of Observations:** The data loss term allows PINNs to leverage observed data, combining empirical information with theoretical principles.

c) *Challenges and Limitations:*

Though PINNs are well-grounded theoretically, they have their limitations:

- **Complexity of Governing Equations:** The physical laws governing the lunar grid might be too complex or poorly understood, making it difficult to define the PDEs accurately.
- **Balance Between Physics and Data Loss:** Selecting the correct value for λ is important. If the balance is off, the model may either overfit the data or fail to respect the physics.

- **Computational Intensity:** The optimization process includes solving PDEs, which may be computationally intensive, especially for large and high-dimensional datasets like the lunar grid.
- **Dependence on Initial Assumptions:** The performance of PINNs largely depends on the accuracy of the physical models used. If these assumptions are flawed, the predictions are not reliable.
- **Handling Sparse Data:** The softmax function used in multiclass regression can be helpful in sparse data scenarios because it allows the model to consider the relative probability of each class.
- **Generalizing to Unobserved Regions:** The multiclass model generalizes well to other regions or parts of the lunar grid, where perhaps data are sparse or missing, in that it predicts class probabilities for all possible classes.

5) *Multiclass Regression:*

Multiclass regression refers to a machine learning model where the model predicts a categorical label or class from many classes.

The multiclass regression differs from binary classification as it only has two potential outcomes, but the multiclass has more than two classes.

Its working basis is the usage of softmax algorithms to find out the most probable class for the given data point and then translate the outcome into a probability across all categories. The forecast is therefore the class having the highest likelihood.

The cross-entropy loss is the objective function. Multi-class regression aims to fit the model so that its cross-entropy loss—the difference between its predicted and the true class probabilities—is minimal.

a) *Key Advantages:*

A multiclass regression would have proven useful in handling the grid data, especially if the task were spatial areas to be categorized into different classes or types rather than prediction. Multiclass regression could have helped in the following ways:

- **Classification of Regions Based on Observed Features:** If the data from the lunar grid can be categorized, say by regions that are similar in terms of their terrain, lunar properties, or environment, then a multiclass regression can be used to classify each point, or region, to its correct class.

- **Improving Predictions Across Different Classifications:** If the grid could be divided into several broad categories, such as high-altitude regions, low-altitude regions, cratered terrain, etc., then multiclass regression would be able to categorize missing values based on the features of surrounding data.

b) *Challenges and Limitations:* While multiclass regression has several advantages, it may have failed with the lunar grid data for the following reasons:

- **Complexity of Spatial Relationships:** If the data contained complex, nonlinear spatial relationships, multiclass regression might have struggled to capture the nuances of these dependencies.
- **Imbalanced Class Distribution:** If certain classes were underrepresented in the lunar grid data, the model could have been biased toward the more frequent classes, leading to poor predictions in sparsely populated regions.
- **Challenges With Continuous Variables:** Since multiclass regression is designed for categorical variables, it was not aptly applied in tasks requiring the forecasting of continuous spatial data, say height of terrain or mineral content concentration.

6) *Using MySQL over CSV:* One of the primary challenges we faced was ensuring the continuous

processing of FITS files without causing bottlenecks or delays. At the beginning, the pipeline that we had chosen based on CSV took nearly 9 seconds to calculate the output for one file. This proved to be a challenge as, to make the process real-time, the FITS files would be added at a rate of approximately one file every 8 seconds, which means possible traffic buildup.

We then migrated to a MySQL-based pipeline. With query handling optimized and leveraging database efficiencies, the MySQL pipeline cut down computation time to 6 seconds per file. This reduced traffic build-up and allowed seamless, real-time processing of FITS files. This upgrade enables continuous addition and processing of FITS files with no break in data flow, thereby maintaining the integrity of the real-time analysis pipeline. After performing the query analysis, we could arrive at the following query-time analysis which shows a comparative analysis between using CSV versus using MySQL database. Figure 11 shows a comparison between query-time of operations in CSV files and MySQL table. This table is formulated by considering 1 FITS file of size 26 kb.

Operation	Using CSV (in s)	Using MySQL (in s)	Remarks
Data Reading	0.05	0.01	MySQL parsing overhead dominates here, even for small files.
Preprocessing	0.15	0.15	Includes ,fits calibration and adjustments where time remains constant.
Data Updation	1	0.025	CSV requires rewriting the file, whereas MySQL performs direct updates efficiently.
Refreshing Heatmap	0.25	0.04	CSV involves re-parsing the updated file, whereas MySQL uses indexed queries.
End-to-End Processing	0.65	0.23	Significant reduction due to MySQL's optimized operations.

Fig. 14: Comparative query-time analysis

XII. FUTURE ASPECTS

A. Applications of Power BI

1) **Input-Based Regional Analysis:** In the future, we will extend the use of Power BI to create advanced visualization functionalities for input-based regional analysis. It will provide the

user with the possibility of inputting a specific latitude-longitude coordinate, through which the visualization tool can display elemental and spectral data for a predefined range—like 600 x 600 km rectangle region, keeping the inputted coordinates at the centre. This feature will give an in-depth view of the composition and characteristics of the chosen lunar region.

2) Use Case: Lunar Exploration :

- **Lunar Exploration:** This feature is likely to have important uses in missions for lunar exploration. For instance, by positioning a rover like Chandrayaan at a point on the Moon, this ability will make it possible to visualize information about elements for the 600-km surrounding area.
- **Selective Investigation:** This capability will be instrumental in performing selective investigations of regions exhibiting unusual elemental abundances or interesting spectral properties, enhancing the precision and focus of lunar studies.

B. Cassandra-Based Backend Pipeline for Lunar Data Processing

To scale out for the future, a NoSQL database like Cassandra will be able to manage growing datasets while keeping the real-time processing demands in check.

- Node Distribution in Cassandra does not increase Query Times as the Datasets Grow, thus helps to make the entire flow real-time.
- Data replication with automatic fail-over ensures that there will always be availability for uninterrupted analytical work and data integrity.

Below is the proposed pipeline using Cassandra:

1) Input: File Addition and Monitoring:

a) Data Sources:

New observational data in FITS and XSM formats

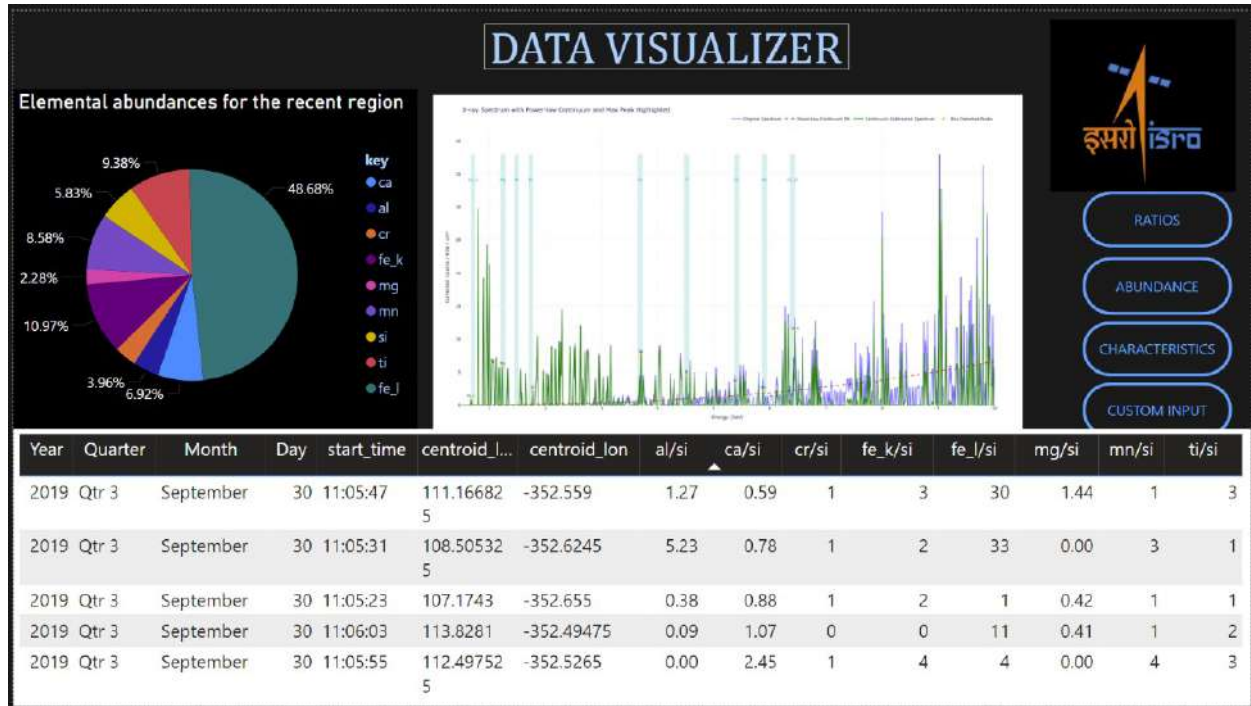


Fig. 15: Power BI Data Visualizer: High Solar Flare, dynamic output for a custom input

are added to a dedicated New_Data remote directory on the server. The server directory serves as the repository for raw data. Subdirectories include:

- **fits_files/**: Accepts individual FITS files (8 seconds) or entire directories.
- **xsm_files/**: Accepts XSM files with the structure YYYY/MM/DD.

b) **Monitoring:**

A Python watchdog service continuously monitors these directories for new files and triggers preprocessing when changes are detected.

2) **Classification of FITS Files:**

a) **Classification Script:**

A Python-based classification script determines whether a FITS file corresponds to the day-side or night-side:

- Day-Side: Solar angle less than or equal to 90° .
- Night-Side: Solar angle greater than 90° .

b) **Cassandra Table for Metadata:**

• Table: file_classification

- file_id (UUID): Unique ID for each FITS file.
- file_path (TEXT): Path to the file.
- solar_angle (FLOAT): Solar angle of the file.
- classification (TEXT): Day-side or Night-side.
- processed (BOOLEAN): Tracks processing status.

After classification, the pipeline flow will remain more or less similar to the current Mysql integrated one.

3) **Background Noise Handling (Night-Side Data):**

a) **Daily Background Computation:**

- Night-side data is averaged daily to calculate background noise levels.
- Statistical parameters (mean μ and standard deviation σ) for energy ranges are derived.

- Element intensities are flagged if they exceed $\mu + 3\sigma$.

b) Cassandra Table for Background Noise:

- Table: background_mean_sigma_counts
 - date (DATE): Observation date.
 - element (TEXT): Element name (e.g., Mg, Al).
 - mean_counts (FLOAT): Daily mean counts for the element.
 - sigma_counts (FLOAT): Standard deviation of counts.

4) Preprocessing of FITS Files :

Preprocessing occurs on the server to ensure centralized data consistency:

- ARF and RMF Calibration
- Power-Law Continuum Adjustment:
 - Background noise due to scattered solar X-rays is removed.
 - XSM data is analyzed to classify files into high or low solar flare regions.
- Cassandra Table for Preprocessed Data:
 - Table: preprocessed_data
 - * file_id (UUID): Unique ID of the FITS file.
 - * calibrated_counts (MAP<INT, FLOAT>): Energy vs. corrected photon counts.
 - * solar_flare_class (TEXT): High or Low flare classification

5) Database Updates:

a) Grid-Level Updates:

- FITS files are mapped to lunar grids using centroid latitude/longitude.
- Elemental ratios are updated via weighted averaging.
- Table: grid_data

- grid_id (UUID): Unique ID for the grid region.
- latitude_range (TUPLE (FLOAT, FLOAT)): Latitude bounds.
- longitude_range (TUPLE (FLOAT, FLOAT)): Longitude bounds.
- element_ratios (MAP (TEXT, FLOAT)): Elemental ratios such as Al/Si, Mg/Si .
- weights (INT): Number of files contributing to the ratio.

b) Kriged Data Updates:

- Interpolates missing values for subpixel grids using kriging.
- Table: kriged_data
 - kriged_id (UUID): Unique ID for the kriged point.
 - latitude (FLOAT): Latitude of the point.
 - longitude (FLOAT): Longitude of the point.
 - interpolated_ratios (MAP (TEXT, FLOAT)): Interpolated elemental ratios.

6) Heatmap Visualization:

a) Real-Time Updates:

- Elemental ratio maps are dynamically updated when grid_data or kriged_data changes.
- Heatmaps are generated on the server and served to the frontend via an API.

7) Workflow and Automation:

a) Monitoring Service:

Uses a Python thread to detect file uploads. Automates triggering of classification, preprocessing, and database updates.

b) Queue-Based Integration:

- Preprocessed results are queued for database updates.
- A FIFO mechanism ensures sequential and consistent updates.

c) Cassandra's Scalable Design:

- Distributed architecture ensures high availability and fault tolerance.
- Efficient query handling for large-scale data via partitioning on `grid_id`, `file_id`, and `date`.

This implementation plays to the strengths of scalability and fast read/write performance in Cassandra, making lunar data processing pipeline robust, efficient and capable of handling high volumes of observational data.

XIII. CONCLUSION

In this project, we have completed the task of mapping the heterogeneous elemental composition of the lunar surface using the CLASS instrument onboard the Chandrayaan-2. The major results include detailed and high-resolution dynamic lunar base maps based on elemental ratios such as Mg/Si, Al/Si, Ca/Si, Mn/Si, Cr/Si, Fe_k/Si, etc, as well as a map made at a subpixel level. The maps projected onto a lunar albedo base reveal insight into the compositional diversity of the lunar surface on kilometer-level resolution, aiding proper lunar geological understanding.

Additional outputs include a catalogue of all line detections for XRF and the corresponding elemental identifications, along with algorithms and source codes for spectral modeling, and visualizations of compositional groupings. It sets the precedent of demonstrating the capability of achieving sub-pixel resolution maps through the integration of multiple overlapping tracks, an attempt that would change the course of planetary mapping techniques. An additional layer of function in making the results usable for research and exploration purposes is the **dynamic and interactive map** format.

Hence, this project places a benchmark of providing strong methodologies with reproducible results and interactive tools in the pursuit of lunar science. Such outcomes, therefore, add to the understanding of the geological history of the Moon but at the same time, give a direction toward the detection of areas that might potentially be resource-rich. Moreover, these results act as a source for the planning of planetary missions, and in addition, they set the pace in geochemical research and space exploration technologies.

REFERENCES

- [1] S. Narendranath, N. S. Pillai, M. Bhatt, K. Vadodariya, R. Vatedka, S. P. Tadepalli, A. Sarwade, A. Tyagi, V. Sharan, *Lunar elemental abundances as derived from Chandrayaan-2*, *Icarus*, vol. 410, pp. 115898, 2024. doi:10.1016/j.icarus.2023.115898.
- [2] S. Narendranath, P. S. Athiray, P. Sreekumar, R. Vatedka, A. Tyagi, B. J. Kellett, and the CLASS team, *Mapping lunar surface chemistry: New prospects with the Chandrayaan-2 Large Area Soft X-ray Spectrometer (CLASS)*, *Advances in Space Research*, vol. 54, no. 10, pp. 1993-1999, 2014. doi:10.1016/j.asr.2013.04.008.
- [3] P. S. Athiray, S. Narendranath, P. Sreekumar, S. K. Dash, B. R. S. Babu, *Validation of methodology to derive elemental abundances from X-ray observations on Chandrayaan-1*, *Planetary and Space Science*, vol. 75, pp. 188-194, 2013. doi:10.1016/j.pss.2012.10.003.
- [4] N. S. Pillai, S. Narendranath, K. Vadodariya, S. P. Tadepalli, et al., *Chandrayaan-2 Large Area Soft X-ray Spectrometer (CLASS): Calibration, In-flight performance and first results*, *Icarus*, vol. 363, pp. 114436, 2021. doi:10.1016/j.icarus.2021.114436.
- [5] R. K. Sinha, V. Sivaprahasam, M. Bhatt, et al., *Geological characterization of Chandrayaan-2 landing site in the southern high latitudes of the Moon*, *Icarus*, vol. 337, pp. 113449, 2020. doi:10.1016/j.icarus.2019.113449.
- [6] A. J. Gloudemans, E. Kuulkers, R. Campana, et al., *Re-evaluation of Lunar X-ray observations by Apollo 15 and 16*, *Astronomy & Astrophysics*, vol. 649, A174, 2021. doi:10.1051/0004-6361/202140321.
- [7] M. Shanmugam, S. V. Vadawale, A. R. Patel, et al., *Alpha Particle X-Ray Spectrometer (APXS) On-board Chandrayaan-2 Rover – Pragyan*, arXiv, 2019. arXiv:1910.09232.
- [8] F. Zaman, L. W. Townsend, W. C. de Wet, et al., *Composition variations of major lunar elements: Possible impacts on lunar albedo spectra*, *Icarus*, vol. 369, pp. 114629, 2021. doi:10.1016/j.icarus.2021.114629.
- [9] R. C. Elphic, D. J. Lawrence, W. C. Feldman, et al., *Lunar rare earth element distribution and ramifications for FeO and TiO₂*, *Journal of Geophysical Research: Planets*, vol. 105, no. E8, pp. 20333-20345, 2000. doi:10.1029/1999JE001176.
- [10] O. Gasnault, C. d'Uston, W. C. Feldman, et al., *Lunar fast neutron leakage flux calculation and its elemental abundance dependence*, *Journal of Geophysical Research: Planets*, vol. 105, no. E2, pp. 4263-4271, 2000. doi:10.1029/1999JE001124.
- [11] N. Hasebe, N. Yamashita, O. Okudaira, et al., *The high precision gamma-ray spectrometer for lunar polar orbiter SELENE*, *Advances in Space Research*, vol. 42, no. 2, pp. 323-330, 2008. doi:10.1016/j.asr.2007.05.046.
- [12] A. L. Turkevich, *Average chemical composition of the lunar surface*, *The Moon*, vol. 8, no. 3, pp. 365-367, 1973.
- [13] J. K. Wilson, H. E. Spence, N. A. Schwadron, et al., *Precise detections of solar particle events and a new view of the moon*, *Geophysical Research Letters*, vol. 47, no. 1, pp. e2019GL085522, 2020.
- [14] F. A. Zaman, L. W. Townsend, W. C. de Wet, et al., *Modeling the Lunar Radiation Environment*, *Space Weather*, vol. 20, no. 8, pp. e2021SW002895, 2021. doi:10.1029/2021SW002895.
- [15] K. Yumoto, Y. Cho, J. A. Ogura, et al., *Elemental analyses of lunar soils using laser-induced breakdown spectroscopy*, *Spectrochimica Acta Part B*, vol. 221, pp. 107049, 2024. doi:10.1016/j.sab.2024.107049.
- [16] H. Nagaoka, M. Ohtake, Y. Karouji, et al., *Studies of lunar crust using Kaguya observations*, *Icarus*, vol. 392, pp. 115370, 2023.
- [17] A. S. Laxmiprasad, R. V. L. N. Sridhar, A. Goswami, et al., *Laser Induced Breakdown Spectroscopy on Chandrayaan-2 Rover*, *Current Science*, vol. 118, no. 4, pp. 573-581, 2020.
- [18] L. Zhang, Y. Yang, Z. M. Chen, et al., *Elemental compositions of lunar plagioclase: Chang'e 5 insights*, *Icarus*, vol. 413, pp. 116002, 2024.
- [19] S. Narendranath, P. S. Athiray, P. Sreekumar, et al., *Lunar X-ray fluorescence observations by Chandrayaan-1*, *Icarus*, vol. 214, no. 1, pp. 53-66, 2011.