

# Analyzing Sequenced Spectra and Major Oxide Composition Data for NASA's Curiosity Rover ChemCam Instrument

Matthew King<sup>1</sup> and Patrick Gasda<sup>1,2</sup>

<sup>1</sup>The Institute for Computing in Research

<sup>2</sup>Los Alamos National Laboratory

## Abstract

*The Chemistry and Camera (ChemCam) instrument onboard NASA's Curiosity rover has collected a large data set of in-situ spectra over the course of its almost 11 years on the surface of Mars. This large data set, specifically the almost 27,000 Clean Calibrated Spectra (CCS) and predicted Major Oxide Composition (MOC) derived from the CCS, is prime to be analyzed for difficult to identify patterns. In this research, a novel algorithm, named the Sequencer, designed to automatically reveal the main trend of a dataset, was applied to identify patterns and features of the CCS and MOC datasets.*

*There have been different attempts to analyze these datasets in the past; our goal was to determine if this new method yields anything new or different. The Sequencer did effectively identify trends in both datasets, and some unexpected patterns were found. This could suggest there is something of interest previously missed in these datasets; however, it is also plausible that these differences can be explained by limitations in the Sequencer's methods. These results are promising for future applications of the Sequencer in more nuanced ways to these, or similar, datasets.*

## Introduction

### ChemCam Data

NASA's ChemCam instrument consists of two remote sensing instruments: the Laser-Induced Breakdown Spectrometer (LIBS) and the Remote Micro-Imager (RMI). These components work together to identify the elemental composition and the geomorphologic context of rocks and soil on Mars. ChemCam's LIBS instrument uses a laser to vaporize a small spot on samples, which causes light emitting plasma to be produced [1]. This light is collected and carried by fiber optic cables to three spectrometers which record the spectra over a range of 240–850 nm with gaps at 340–385 nm and 469–492 nm [1] [2]. The spectral data consists of emission lines of the elements present in the sample. This is a method by which the elemental composition of Martian rocks and soil can be determined. For each sample recorded, a clean calibrated spectra (CCS) file is created containing a processed version of the spectral data recorded that has been corrected for instrument background, denoised, and calibrated for wavelength and intensity. The first dataset analyzed in this research is the CCS collected to sol 3644. The RMI instrument is used as a context imager for the LIBS; this data is not used in this research [1].

Using the spectral information recorded by the LIBS instrument, the predicted major oxide compositions (MOC) of the sample can be calculated [2]. Predicted MOC calculated include: SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>,

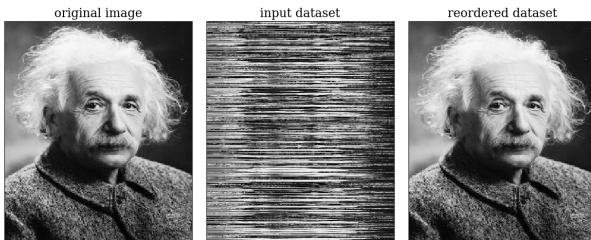
FeO<sub>T</sub>, MgO, CaO, Na<sub>2</sub>O, and K<sub>2</sub>O [3]. Elemental composition is recorded as oxides because that is preferred by geologists; however, in this research, oxides effectively map to their non-oxide elemental counterpart and will be referred to interchangeably throughout this paper. The quantification model used to predict the MOC makes use of multivariate methods, inducing Partial Least Squares. This model was trained using a ChemCam replica on Earth, specifically: a laboratory LIBS instrument, Mars-like atmospheric conditions, and a large set of standard rock and soil samples [2] [3]. While this model for calculating MOC is better than previous models, it is important to remember that this data is the predicted MOC, not necessarily the actual MOC [3]. For each sample of rock or soil taken on Mars, the predicted MOC is calculated; the second, and last, dataset analyzed in this research is the predicted MOC to sol 3644.

### The Sequencer

In certain datasets, information is not already ordered in ways that reveal interesting patterns. However, if the data does follow a trend, it must be possible to rearrange the dataset to reveal that underlying pattern. The Sequencer is an algorithm designed to automatically reveal the main, one-dimensional trend in an arbitrary dataset, assuming one exists [4]. One difficulty in finding these one-dimensional trends, often called sequences, is knowing the correct point of view to "look" at the data [5] [4]. To overcome this,

the Sequencer can use up to four different metrics each with unique properties to analyze the data: Euclidean Distance (L2), Kullback-Leibler Divergence (KL), Monge-Wasserstein / Earth Mover Distance (EMD), and Energy Distance (ED or energy) [4]. The Sequencer can also automatically consider different scales of the data by dividing the dataset into parts and considering the similarities between these parts [4] [5]. This is important because observable signatures of an underlying trend can exist on smaller or larger scales [4].

For each metric, scale combination, a distance matrix and its minimum spanning tree is computed. This information is combined by taking an elongation-weighted average of the minimum spanning trees to create a global distance matrix [4]. The minimum spanning tree of the global distance matrix is then computed; this is the critical piece of information that allows us to determine the best sequence for the dataset. A Breadth First Search (BFS) is applied to the combined minimum spanning tree starting at the least connected node. If this minimum spanning tree has a notable elongation, the BFS will identify the main branch of the tree as the main trend of the dataset. The order in which we visit the nodes of the minimum spanning tree with BFS is the order in which our dataset should be re-ordered to clearly show its one-dimensional trend [4].



**Figure 1:** Application of the Sequencer to a natural image [4]. The middle panel is a row-shuffled version of the original image (left panel). The Sequencer intakes the middle panel and determines how the data set should be reordered, outputting the right panel. This demonstrates how the Sequencer can identify complex patterns in datasets.

## The Focus

This paper will focus on applying the Sequencer to both the CCS collected and MOC data computed from the Curiosity rover's landing in August of 2012 to sol 3644 of the mission. Comparing and contrasting the sequenced versions of these two datasets to find interesting features is our primary goal. Specifically, we will analyze certain wavelengths of light that correspond to the major oxides. Depending on the features identified, and current scientific knowledge of the geology of Mars, we will speculate on

---

**Algorithm 1:** Sequencer pseudo-code

---

```

set list of metrics;
set list of scales;
for each metric k do
    for each scale l do
        set list of parts;
        split objects X into m parts Xm;
        for each part m do
            Dklm = distance matrix(Xm);
            MSTklm = Minimum Spanning Tree(Dklm);
            ηklm = elongation(MSTklm);
        end
        Dkl = η-weighted average of individual Dklm;
        MSTkl = Minimum Spanning Tree(Dkl);
        ηkl = elongation(MSTkl);
    end
P = combined proximity matrix populated by
    η-weighted edges of MSTkl;
D = 1/P;
MST = Minimum Spanning Tree(D);
Sequence = Breadth First Search path(MST);

```

---

**Figure 2:** Pseudo-code outline of the Sequencers algorithm, as described in "Introduction, The Sequencer" [4].

what our sequenced data could mean and why the Sequencer could have ordered these two datasets in the way it did.

## Sequencing Spectral Data

### The Data Set

As previously mentioned, for each rock or soil sample analyzed by the LIBS, the spectral data is recorded and processed to produce the clean calibrated spectra (CCS) data file. When a sample is being analyzed, a series of shots, or laser pulses, is taken to increase precision by having a larger sample size. For each of these shots, a radiance value is recorded in each of its corresponding channels. After all shots are taken, an average radiance value is calculated for each of the 6144 wavelength channels. It is important to note that the wavelength channels for which radiance values were recorded span 240 - 850 nm and remained constant across all samples [1]. In this paper specifically, the average radiance data from CCS files publicly available on NASA's Planetary Data System (PDS) in the ChemCam database will be passed through the Sequencer to identify trends.

To gather this data, a program was created that can intake a local directory of CCS files and output two parts of the CCS data set as .csv files: average radiance for each wavelength and the wavelengths of light for which radiance was recorded. The average radiance file will have N number of rows where N is the number of samples taken. Each row will be a

list of the average radiance values recorded for that sample. The wavelengths .csv file is simply a list of the wavelengths of light for which radiance values were recorded. While the wavelengths required no modification to be compatible with the Sequencer, this was not the case for the average radiance values. In certain cases, a negative radiance value can be recorded. Some of the metrics the Sequencer uses are not compatible with negative or zero values, so this program also “processes” the average radiance data so that the minimum value of the dataset is greater than or equal to one. It is important to note that the data we are sequencing has been modified.

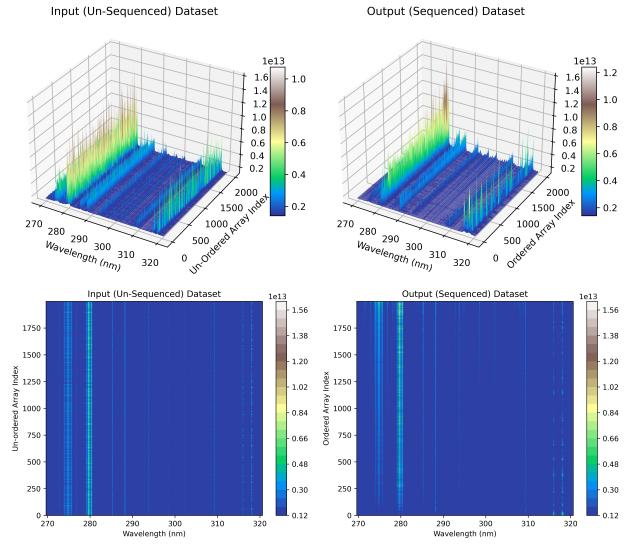
## A Random Sample of 2000

First we will sequence and analyze a random sample of 2000 rows of the average radiance with two metrics each at four different scales: EMD at scales of 1, 2, 128, and 256; ED at scales of 1, 2, 4, and 8. These specific parameters were chosen from previous knowledge about which metrics perform the best: smaller samples of 500 rows were run through the Sequencer using all metrics and numerous scales. The elongations calculated for each metric, scale are as follows:

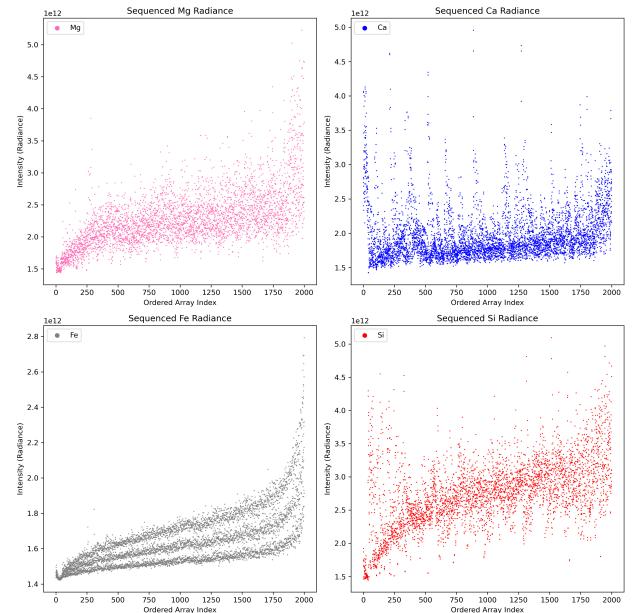
- EMD 1, elongation: 52.29
- EMD 2, elongation: 23.05
- EMD 128, elongation: 57.15
- EMD 256, elongation: 64.97
- Energy 1, elongation: 45.97
- Energy 2, elongation: 25.23
- Energy 4, elongation: 15.02
- Energy 8, elongation: 19.45

As can be seen, certain metric, scale combinations performed better than others; so only the best three were used in creating our final dataset for graphing. The resulting elongation from the best three elongations was 64.80. Elongations of around one suggest no clear pattern in the dataset, and elongations of around N where N is the number of objects in the sample mean a substantial trend was identified [6]. While N in this sample would be 2000, and 64.80 is not close to 2000, this does not mean the dataset has no patterns.

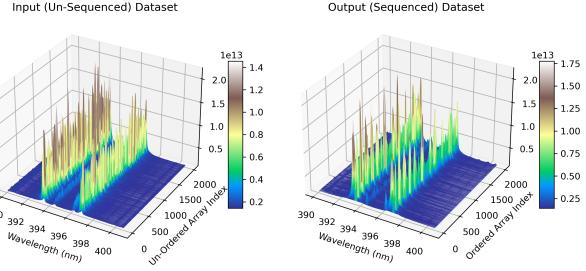
For context, wavelengths of 275, 280, 288, and 316 nm represent Fe, Mg, Si, and Ca respectively [7]. When viewing Figure 3 and Figure 4, a few notable features can be discerned. Radiance values at wavelengths related to Fe, Mg, and Si were all sequenced so that higher values are at higher index. However, all increase at different rates as if their base functions are different. Of the three, Fe has the clearest trend where each of the three bands are their own unique wavelength (as recorded by ChemCam) and



**Figure 3:** Two and Three dimensional graphs comparing “Input (Un-Sequenced) Dataset” (left) and “Output (Sequenced) Dataset” (right). All graphs show the same subset of data, specifically only wavelengths between 270 nm and 320 nm. Each row represents the average radiance, across the specified wavelength channels, for one rock or soil sample. Color and vertical elevation are used to denote different radiance values.



**Figure 4:** “Side Profile” scatter plot views of select sequenced wavelengths: 285.18, and 285.23 nm (Mg); 422.68, 422.72, and 422.76 nm (Ca); 404.50, 404.54, and 404.59 nm (Fe); 288.15, and 288.20 nm (Si). Note: Mg, Ca, and Si panels have slightly different scales; however, Fe has a noticeably smaller scale. The patterns shown in this Figure match those of Figure 3.



**Figure 5:** Three dimensional graph showing “Output (Sequenced) Dataset” along a wavelength range indicative of Ca: 390 nm to 400 nm. The pattern shown in this data is dissimilar to the pattern shown at wavelengths of 275 nm (Fe) and 280 nm (Mg) in Figure 3. However, the grouping is very similar to the grouping of data shown at wavelength 316 nm in Figure 3; this is expected because both represent the same element, Ca.

each resembles a cubic function. At 316 nm in Figure 3, and in the top right panel of Figure 4, groupings of relatively higher radiance values, independent of the more pronounced sequencing of the other oxides, can be noticed. Using these graphs, it can be concluded that the concentration of Fe, Mg and Si are all positively correlated in rock and soil samples. Notably, Ca does not share the same trend as the other oxides: there appears to be high and low groupings with no clear trend. Even when looking at the wavelengths of light with the highest radiance values that correspond to Ca, specifically 393 nm and 397 nm, the same, possibly more nuanced pattern of erratic high and low groupings, is demonstrated: see Figure 5.

## Sequencing MOC Data

### The Data Set

As explained in the introduction, after spectral information about a sample is recorded by the LIBS instrument, the predicted major oxide composition (MOC) of the sample is computed using a multivariate approach [2]. For each of the almost 27,000 average radiance samples, a corresponding MOC entry is made in an online NASA database. In this research, the publicly available MOC files will be passed through the Sequencer so that trends may be identified.

A program similar to the one used for collecting and compiling the average radiance dataset was used to congregate the MOC dataset. This program also intakes a directory of relevant files and outputs two .csv files: a 2D array of MOC data and a list of numbers each corresponding to an oxide. The MOC dataset will have N number of rows where N is the number of rock or soil samples analyzed. Each row will contain nine values, each a weighted percent

corresponding to the predicted concentration of an oxide. The list of numbers mapping to oxides did not require modification to be compatible with the Sequencer; however, because the MOC dataset could contain values of zero, all data was shifted up so that the minimum value in the dataset was greater than or equal to one. It is important to note that the data we are sequencing has been modified.

### A Random Sample of 2000

First, a random sample of 2000 rows of predicted MOC was sequenced. The metrics EMD, ED, KL, and L2 were used. The scales for each metric were not specified, so the Sequencer was left to autonomously determine appropriate scales: scales of one were chosen for each metric. A scale of one means the dataset was looked at as a whole; a scale of two would mean the dataset was divided into two parts perpendicular to the oxides axis. Because there are only nine oxides, the Sequencer did not deem it relevant to divide the dataset into any smaller parts. The resulting elongations for each metric, scale are as follows:

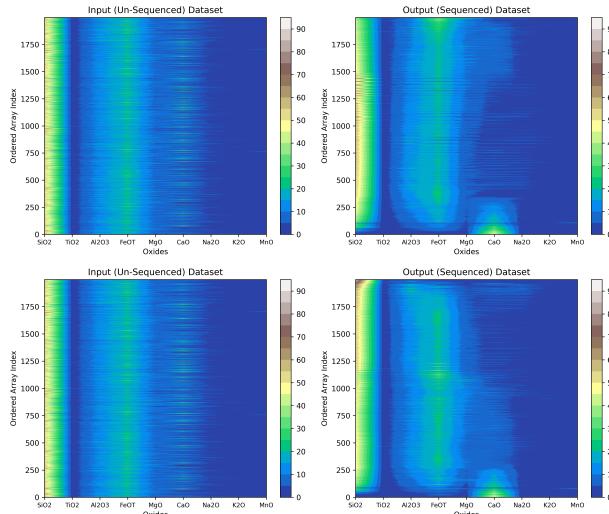
- EMD 1, elongation: 27.42
- ED 1, elongation: 27.58
- KL 1, elongation: 19.05
- L2 1, elongation: 20.17

Because both KL and L2 had relatively weaker elongations, a graph was also produced using just EMD and ED at scales of 1. To see this comparison, refer to Figure 6. Using only the two best elongations results in a subjectively better re-ordering of the input dataset. This graph will be used for interpretation and drawing conclusions going forward, and it had a combined elongation of 27.58.

In comparing the bottom two panels in Figure 6, some notable features can be seen allowing for certain conclusions to be drawn.  $\text{SiO}_2$ ,  $\text{FeO}_T$ ,  $\text{Al}_2\text{O}_3$ , and  $\text{MgO}$  were the most common oxides, appearing in  $\sim 98\%$  of the rock and soil samples. When these elements appeared in samples, they made up the following amounts of the sample by mass:

- $\text{SiO}_2$ :  $\sim 50\%$  (with a peak to  $\sim 90\%$ )
- $\text{FeO}_T$   $\sim 20\%$
- $\text{Al}_2\text{O}_3$ :  $\sim 12\%$
- $\text{MgO}$ :  $\sim 8\%$

$\text{CaO}$  also appeared relatively frequently; however, its concentration has a very strong inverse correlation to the oxides mentioned above. When  $\text{CaO}$  is at its highest concentration of approximately 40%,  $\text{SiO}_2$ ,  $\text{FeO}_T$ ,  $\text{Al}_2\text{O}_3$ , and  $\text{MgO}$  all have concentrations of almost 0%. Another notable feature appears around index 1100. At this index, the concentration of  $\text{SiO}_2$  appears to drop off while the concentration of  $\text{MgO}$



**Figure 6:** Two 2D graphs showing “Input (Un-Sequenced) Dataset” (left) and “Output (Sequenced) Dataset” (right). Top right panel shows the effects of using all four metrics, bottom right panel shows the effect of using only the two best metrics. Each row represents the predicted MOC for one rock or soil sample. Color is used to represent percent composition of the samples.

and  $\text{FeOT}$  rise. Other than this peak of  $\text{MgO}$  and  $\text{FeOT}$  at 1100, and decreases in concentration at the start and end of the indices, both oxides have relatively constant concentrations over the range of indices. This can be more easily seen in a “side profile” scatter plot view of  $\text{SiO}_2$ ,  $\text{FeOT}$ ,  $\text{Al}_2\text{O}_3$ , and  $\text{MgO}$ , see Figure 7.  $\text{TiO}_2$ ,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ ,  $\text{Al}_2\text{O}_3$ , and  $\text{MnO}$  all appeared in-frequently with little to no pattern to be discerned. However,  $\text{Na}_2\text{O}$  and  $\text{Al}_2\text{O}_3$  appear to be at their highest concentration of  $\sim 5\%$  and  $\sim 20\%$  respectively when  $\text{FeOT}$  is at its lowest and  $\text{SiO}_2$  is nearing its highest concentration.

## Comparing Sequenced Spectral and MOC data

### The Same Conclusion?

One of the goals of this research was to determine how the Sequencer would sort spectral and MOC data, and to identify any similarities and or differences. Since predicted MOC data is derived from the spectral data set analyzed, it is expected similar features and trends will be identified by the Sequencer in both datasets.

First, we will compare the ordering of Si between the sequenced spectral and MOC data. In the MOC data, Si is clearly sequenced so that smaller Si concentrations trend towards higher Si concentrations from index 0 to 2000. However, this pattern is less defined in sequenced spectral data: see the bottom

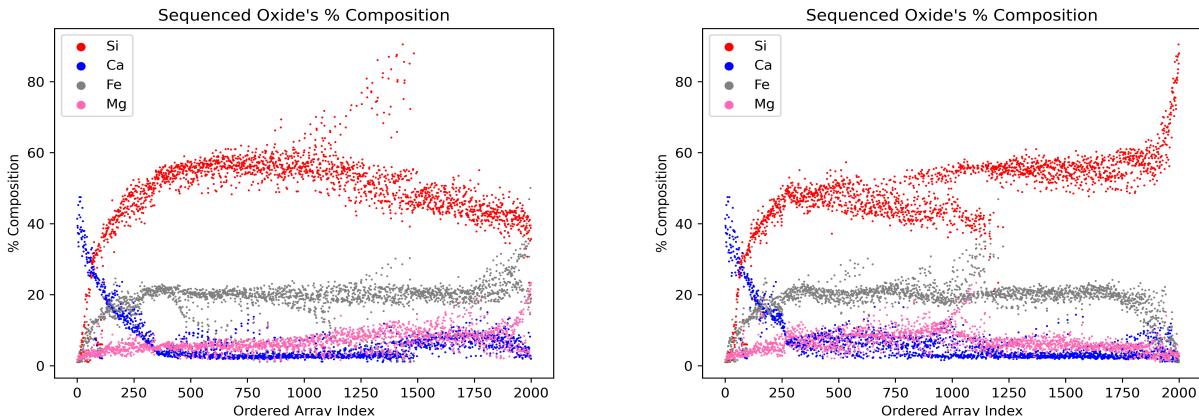
right panel of Figure 4. The radiance values have a very noisy trend upward with a slight trend of increasing radiance near the start of the range of indices. The pattern of increasing Si concentrations (in the sequenced MOC dataset) do not really translate to the sequenced radiance values for Si.

The most obvious trend identified in the MOC data is that Ca enriched samples are depleted in the other major oxides measured. However, a somewhat different pattern for Ca appears in the sequenced radiance values. There are erratic Ca groupings spread across the entire index range, many of which directly overlap with high radiance values of Mg and Fe (see the upper right panel of Figure 4 and Figure 5). This is not expected, as almost all Ca is grouped as one in the sampled MOC dataset, (see Figure 6) and these samples have very low concentrations of Mg and Fe. These groupings of Ca seem to indicate that the Sequencer believes there are distinct high or low groups of Ca values. This could indicate that the presence of Ca is discrete; either there is or isn’t Ca.

The last trend of note is the extremely strong one-dimensional trend the Sequencer identified for Fe: specifically at wavelengths of 404.50, 404.54, and 404.59 nm (see bottom left panel of Figure 4). Unlike the trends identified for Mg, Ca, and Si, this trend has very little noise and looks to be a very clean distribution of data. Whether this is due to the Sequencer favoring Fe above the other oxides, or from some inherent property of Fe, it is unclear. Possibly, ChemCam’s measurements at these wavelengths were more accurate, or Fe produces a more regular and consistent radiance emission when vaporized by ChemCam’s laser. It is relevant to note that this extremely strong trend for Fe is unique to the sequenced spectral data and is not present in the sequenced MOC data: see Figure 7. The main similarity between the two sequenced data sets, as it pertains to Fe, is the general shape of the graph. As can be seen in the left panel of Figure 7 and the the bottom left panel of Figure 4, the shape of the sequenced Fe data resembles a cubic function.

### What Happened?

What mechanism of the Sequencer’s algorithm is responsible for the scattered grouping of high Ca radiance values we see? In theory, the Sequencer should try to order rows to maximize similarities (using its four metrics); so, in theory, it would make sense if the Sequencer would prioritize “fixing” data with high magnitudes of variation that have enough intermediate values between those large variations. However, this is only partly seen in practice. The wavelengths of light shown in Figure 5 contain some of the largest radiance values; however, Ca is still



**Figure 7:** A “side profile” scatterplot version of the sequenced MOC using two metrics, EMD and ED, both at scales of one (right) and all metrics at scales of one (left). While these two plots may look very different at first glance, the same general patterns exist in both. One can think of the Si peak around index 1250 in the left panel being moved to index 2000 to create the plot of the right panel. The same is true of the downward trend of Si from index 400-1000 in the right panel: it has been transposed to the right to create the plot of the left panel. The increasing concentration of Si over the index, the drop off of Si and increase in Mg and Fe, and Ca’s existence in the absence of the other oxides can all clearly be seen. This dataset is the same as the right panels in Figure 3.

grouped erratically (as mentioned previously, this could simply indicate that Ca’s presence is discrete). The Fe radiance values graphed in the bottom left panel of Figure 4 are  $\sim 3/25$  the magnitude of the Ca radiance values, but the sequenced Fe values exhibit an extremely clean trend. A deeper understanding of the Sequencer and more research into these irregularities is needed to understand why the Sequencer did what it did to the spectral dataset. In contrast, the trends exhibited in the sequenced MOC data make more sense intuitively. The main differences between these two datasets is the fact that there are only nine oxides but there are 6144 wavelength channels. So, possibly the Sequencer is being overwhelmed by such a large dataset and having to make many compromises.

## Considering Domain Knowledge

Through the Curiosity rover missions, ChemCam has picked up on many regions of high Ca concentration that were also extremely low in the other major oxides [7]. These regions of high Ca concentration are sulfur calcium peaks, and appear as independent veins that are distributed irregularly across the surface analyzed by the Curiosity rover. This information matches with part of the findings from our sequenced spectral data: the existence of either high or low Ca radiance values. This discrete behavior implies that there either is or isn’t Ca. However, the fact that there are high Ca radiance values when there are also high Fe, Si, and Mg radiance values does not match with this knowledge or the MOC data. While high radiance values don’t alone mean higher concentrations, intuitively, this would make sense; so,

more research into this pattern is needed.

## Future Work: Other Methods for Grouping Data

This section of the paper will include other methods by which data has been or could be sequenced and or grouped. While many of these methods have not been tested yet, they could reveal more interesting trends by overcoming limitations and add to our understanding of the already identified differences between the way spectral and MOC data was sequenced. These are some of the next steps that could be taken in future research to better understand these sequenced dataset.

## Sequencing Select Wavelengths and Normalizing Radiance Values

As theorized in “Comparing Sequenced Spectral and MOC Data, The Same Conclusion?” larger radiance values could be prioritized over smaller values (despite this pattern not being present in actual sequenced data). Because the Sequencer prioritizes “fixing” larger variations in data, and therefore larger radiance values, this limitation could be overcome by normalizing the radiance values at each desired wavelength (so that the maximum value at the wavelengths are the same) and only sequencing that data. This method would bypass the potential issues of simultaneous sequencing of large and small radiance values. If a more favorable sequence of spectral data is identified, this sequence could be applied to the en-

tire dataset. This method makes use of the fact that the Sequencer does not automatically reorder the dataset it's given, but produces a list of the way the original dataset's index values should be rearranged.

## Applying the Spectral Sequence to MOC

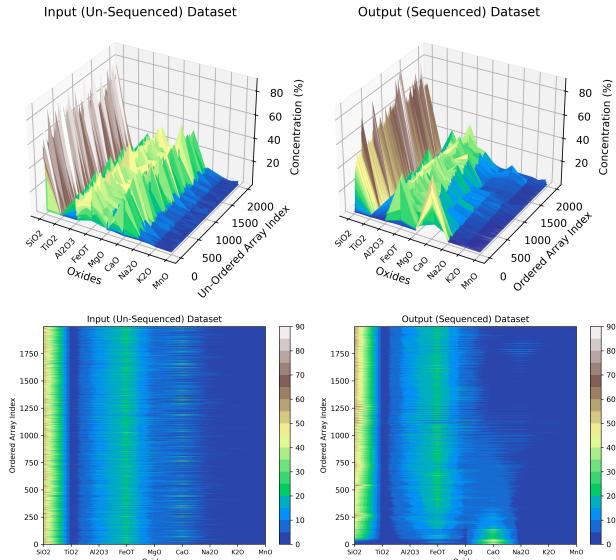
An entirely different way of grouping data would be to apply the spectral sequence to the MOC dataset and vice versa. By applying the list created when sequencing spectral data (which describes how index values should be rearranged) to the MOC dataset we could learn more about the differences between the sequencing of the MOC and the spectral data. This would allow for us to see which patterns of the sequenced spectral data map to certain features of the MOC dataset. To do this, it would be important to select chunks of corresponding spectral and MOC data. Taking a random sample of the entire dataset, as outlined above, would be ineffective because the key part of this method is the fact that the spectral data maps to the MOC data.

## Sequencing MOC at Difference Scales

In "Sequencing MOC Data, Random Sample 2000" it was mentioned that the Sequencer autonomously determined that scales of one were appropriate for each metric. However, by manually setting the scale list, each of the four metrics were run on scales of 1, 2, 4, and 8; their resulting elongations are as follows:

- EMD 1, elongation: 18.87
- EMD 2, elongation: 32.99
- EMD 4, elongation: 20.38
- Energy 1, elongation: 22.28
- Energy 2, elongation: 35.69
- Energy 4, elongation: 20.03
- KL 1, elongation: 15.35
- KL 2, elongation: 17.22
- KL 4, elongation: 18.83
- L2 1, elongation: 19.69
- L2 2, elongation: 30.73
- L2 4, elongation: 17.68

By taking the five best elongations, a total elongation of 30.74 was found for the sequence. This is slightly larger than the resulting elongation of 27.58 found when only evaluating each metric at a scale of one and taking the two best elongations. The resulting graphs have some noticeable differences compared to original sequencing of MOC data [See Figure 8]. First, there is no longer a clean and gradual increase of SiO<sub>2</sub> concentration over the index. There are now jagged peaks and the greatest concentration is found around index 1250 as opposed to index 2000



**Figure 8:** 2D and 3D graphs showing "Input (Un-Sequenced) Dataset" (left) and "Output (Sequenced) Dataset" (right). Each row represents the predicted MOC for one rock or soil sample. Color and vertical elevation is used to represent percent composition of the samples. The trends of increasing Si and decreasing Fe concentrations over the index values are still present however to a lesser extent. More noticeably, the Ca grouping does not have a band of low values through its center.

in the original sequence. Secondly, while the grouping of CaO is all in the same general position, this new sequence does not share the gap in the higher CaO concentration noticed around index 100 in the original sequence [See bottom right panel of Figure 6]. These differences are caused by the Sequencer now mostly using metrics at scales of two and four. These smaller scales mean the data was split into more parts along the Oxides axis. By comparing more parts, the Sequencer favors the variations in concentration of the less concentrated oxides more in this dataset.

## Other Methods

Some other methods that could be tested in future research include: sequencing more of the data, analyzing minimum spanning trees, and applying the Sequencer to similar datasets. In this research, samples of only 2000 were taken; sequencing of the entire dataset could reveal trends that were not identified in these samples. The Sequencer has the functionality of showing the final minimum spanning tree (MST) of the datasets it sequences. Analyzing the MSTs could help confirm the strength of the patterns found and reveal more information about the datasets. Lastly, ChemCam data that is currently not publicly available could also be run through the Sequencer to find trends. In this research only publicly

available data was considered.

## Conclusion

### Sequencer Performance

The Sequencer was able to identify trends in both the spectral and MOC dataset; however, there were key differences in how this data was sequenced. Spectral sequencing seemed to strongly favor wavelengths of light associated with Fe: a very strong trend was identified. Sequencing of spectral data also resulted in a clear trend of increasing Mg and Si related radiance values over the sampled index. There were scattered groupings of Ca related radiance values across many of Ca's wavelengths. It was theorized that since most Ca radiance values were either high or low, the Sequencer prioritized the stronger one-dimensional trend of the other oxides mentioned in this paper. Sequencing of predicted MOC produced a clear trend of gradually increasing Si concentration over the sampled index. Graphs produced of sequenced MOC showed a clear separation between rock and soil samples containing Ca and those containing the other oxides measured. Unlike in the sequenced spectral data, Ca was largely grouped together. The difference in the grouping of Ca in spectral and MOC data is the largest difference identified in this research.

### Key Findings

1. The Sequencer is effective at identifying one-dimensional trends in both datasets. Some unexpected patterns were found, primarily related to Fe, and it is unclear why the Sequencer identified them.
2. Sequenced spectra and MOC data supports that Ca is either in very high or very low quantities. This matches with other research [2].
3. However, the sequenced spectra data showed high Ca radiance values when there were high Fe, Si, and Mg radiance values. This pattern does not match with sequenced MOC data or other research in this field [2].

### Difficulties

The primary obstacle in this research was operation of the Sequencer and managing its extreme run times on larger datasets, specifically CCS. This required reducing the MOC and CCS datasets from the almost 27,000 rocks or soil samples collected down to samples of 2000 for analysis. Hardware limitations also required the sample of 2000 CCS to be sequenced on another much more powerful device, and this

computation took almost 15 hours. The Sequencer is intended to be run on extremely powerful devices, including super computers, so this issue is expected when sequencing data on less powerful devices. The issues with run time were a limiting factor in how much data could be sequenced, and how many conclusions could be drawn.

## Acknowledgements

I would like to thank my mentor, Dr. Patrick Gasda, of the Los Alamos National Laboratory, for guiding my research and answering my countless questions. Thank you, Alex Huynh, for sequencing the spectral data so that I could load and analyze the results for this paper. Lastly, I would like to thank the Institute for Computing in Research for this opportunity, and Mark Emry for introducing me to the program.

## References

- [1] NASA Science Mars Exploration. *ChemCam for Scientists*. Apr. 2021. URL: <https://mars.nasa.gov/msl/spacecraft/instruments/chemcam-for-scientists/>.
- [2] Kristin Rammelkamp et al. "Clustering Supported Classification of ChemCam Data From Gale Crater, Mars". In: *Earth and Space Science* 8.12 (Dec. 2021). doi: 10.1029/2021EA001903.
- [3] Samuel M. Clegg et al. "Recalibration of the Mars Science Laboratory ChemCam instrument with an expanded geochemical database". In: *Spectrochimica Acta Part B: Atomic Spectroscopy* 129 (2017), pp. 64–85. doi: <https://doi.org/10.1016/j.sab.2016.12.003>.
- [4] Dalya Baron and Brice Menard. "Extracting the main trend in a dataset". In: (Apr. 2020). URL: [http://sequencer.org/static/docs/Sequencer\\_paper.pdf](http://sequencer.org/static/docs/Sequencer_paper.pdf).
- [5] Dalya Baron and Brice Menard. *The Sequencer Algorithm*. 2023. URL: <http://sequencer.org/documentation>.
- [6] Dalya Baron and Brice Menard. *The Sequencer README*. 2020. URL: <https://github.com/dalya/Sequencer/blob/master/README.md>.
- [7] S. Maurice et al. "ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars". In: *J. Anal. At. Spectrom.* 31 (4 2016), pp. 863–889. doi: 10.1039/C5JA00417A.