Clustering Supported Classification of ChemCam Data From Gale Crater, Mars

Kristin Rammelkamp, Olivier Gasnault, Olivier Forni, Candice Bedford, Erwin Dehouck, Agnès Cousin, Jeremie Lasue, Gaël David, Travis Gabriel, Sylvestre Maurice, et al.

Earth and Space Science

RESEARCH ARTICLE
10.1029/2021EA001903

Clustering Supported Classification of ChemCam Data From Gale Crater, Mars

Kristin Rammelkamp[1], Olivier Gasnault[1], Olivier Forni[1], Candice C. Bedford[2,3], Erwin Dehouck[4], Agnès Cousin[1], Jeremie Lasue[1], Gaël David[1], Travis S. J. Gabriel[5], Sylvestre Maurice[1], and Roger C. Wiens[6]

[1]Institut de Recherche en Astrophysique et Planétologie, Toulouse, France, [2]Lunar and Planetary Institute, Universitites Space Research Association, Houston, TX, USA, [3]Astromaterials Research and Exploration Science, NASA Johnson Space Center, Houston, TX, USA, [4]Univ Lyon, UCBL, ENSL, UJM, CNRS, LGL-TPE, Villeurbanne, France, [5]U.S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ, USA, [6]Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract The Chemistry and Camera (ChemCam) instrument on board the Mars Science Laboratory (MSL) rover Curiosity has collected a very large and unique data set of in-situ spectra and images of Mars since landing in August 2012. More than 800,000 single shot laser-induced breakdown spectroscopy (LIBS) spectra measured on more than 2,500 individual targets were returned so far by ChemCam. Such a data set is ideally suited for the application of statistical methods for the recognition of patterns that are difficult to observe by humans. In this work, we develop an approach relying on the feature extraction method Non-Negative Matrix Factorization (NMF) and the repetition of k-means clustering to classify ChemCam spectra. A strong consistency of the clustering results among the repetitions were found, which allowed us to identify six clusters representing the dominant compositions measured by ChemCam in Gale crater so far. By tracking clusters across the rover traverse from landing to sol 2756, we are able to provide a chemostratigraphic overview of the Gale crater from the ChemCam perspective. Transitions between major geologic groups (such as the Bradbury and the Mt. Sharp groups) are identifiable demonstrating that they are compositionally distinct, consistent with previous work. Compositional differences between their members also appear in the results. Furthermore, a first approach in which a random forest classifier was trained and validated with the obtained cluster assignments, reveals promising results for predicting cluster memberships of new ChemCam LIBS data acquired after sol 2756.

Key Points:

- With a combination of feature extraction and repeated k-means clustering applied to ChemCam LIBS data, we identify six clusters representing dominant compositions in Gale crater
- The cluster detections are in agreement with the chemostratigraphy explored by Curiosity
- We successfully trained a random forest classifier with the obtained cluster labels and used it to classify recent ChemCam targets

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Rammelkamp,
kristin.rammelkamp@irap.omp.eu

Citation:

Rammelkamp, K., Gasnault, O., Forni, O., Bedford, C. C., Dehouck, E., Cousin, A., et al. (2021). Clustering supported classification of ChemCam data from Gale crater, Mars. *Earth and Space Science*, *8*, e2021EA001903. https://doi.org/10.1029/2021EA001903

Received 13 JUL 2021
Accepted 9 OCT 2021

Author Contributions:

Conceptualization: Kristin Rammelkamp, Olivier Gasnault, Olivier Forni
Formal analysis: Kristin Rammelkamp
Investigation: Kristin Rammelkamp
Methodology: Kristin Rammelkamp, Olivier Gasnault, Olivier Forni, Candice

1. Introduction

The Mars Science Laboratory (MSL) rover Curiosity landed successfully in Gale crater on Mars on August 6, 2012. One of the primary mission goals of MSL is to determine whether Mars was ever habitable, for which the Gale crater and in particular its central mountain Aeolis Mons (informally known as Mt. Sharp) was chosen as the science target (Grotzinger et al., 2012). From orbit, it was recognized that the lower layered sedimentary strata of Mt. Sharp contains a record of environmental changes including diverse aqueous conditions over time (Milliken et al., 2010). Curiosity landed at the Bradbury site, northwest of Mt. Sharp, near the base of the alluvial fan of Peace Vallis, and explored the rocks of Aeolis Palus during the first Mars year of the mission. On sol 753, Curiosity reached the geologic transition to the Mt. Sharp group and has continued to climb the mound since then. The ancient habitability could be confirmed at Yellowknife Bay, a location in Aeolis Palus, due to several detections of early diagenesis products, which together suggest the existence of an aqueous environment, implying a higher potential for habitable conditions. (Grotzinger et al., 2014, 2015).

Several scientific instruments belong to Curiosity's payload for the characterization of Gale crater's geological evolution and one of them is the "Chemistry and Camera" (ChemCam) instrument (Maurice et al., 2012; Wiens et al., 2012). ChemCam has the capability to perform rapid multi-elemental analyses of rocks and soils from a distance with the laser-induced breakdown spectroscopy (LIBS) technique. Due to its regular use—almost every sol—ChemCam has collected a very large and unique LIBS data set containing more than 800,000 single shot LIBS spectra. Such a data set is well suited for the application of machine learning algorithms in order to identify similar targets and predict classes of new measured targets. From

C. Bedford, Erwin Dehouck, Agnès Cousin, Jeremie Lasue, Gaël David, Travis S. J. Gabriel
Project Administration: Olivier Gasnault, Sylvestre Maurice, Roger C. Wiens
Software: Kristin Rammelkamp
Supervision: Olivier Gasnault, Olivier Forni
Visualization: Kristin Rammelkamp
Writing – original draft: Kristin Rammelkamp, Olivier Gasnault, Olivier Forni, Candice C. Bedford, Erwin Dehouck, Agnès Cousin, Jeremie Lasue, Gaël David, Travis S. J. Gabriel, Sylvestre Maurice
Writing – review & editing: Kristin Rammelkamp, Olivier Gasnault, Olivier Forni, Candice C. Bedford, Erwin Dehouck, Agnès Cousin, Jeremie Lasue, Gaël David, Travis S. J. Gabriel, Sylvestre Maurice

the field of artificial intelligence research, many machine learning algorithms arose, which are used in various fields of Solar System exploration as well as in LIBS data analysis including LIBS measurements on geological samples (Anderson et al., 2011; Boucher et al., 2015; Sirven et al., 2006). Regarding ChemCam data, an artificial neural network (ANN) was successfully implemented for the identification of similar samples (Forni et al., 2019) but also unsupervised methods such as clustering algorithms were frequently used to support the ChemCam data analysis (Anderson et al., 2013; Bedford et al., 2020; Cousin et al., 2015; Gasnault et al., 2013, 2014, 2015, 2019; Lasue et al., 2011). In order to reduce dimensionality and assess variability, feature extraction techniques in combination with classification methods are often used. However, the use of supervised machine learning algorithms for ChemCam data is still challenging due to the lack of training data from Mars. The recent ChemCam quantification model based on multivariate methods (Clegg et al., 2017) uses data from more than 400 standards measured with a ChemCam replica on Earth. A study using the data from 376 of these standards to train a convolutional neural network (CNN) with different activation functions reveals promising results for the prediction of major oxide compositions (Cao et al., 2020). Machine learning algorithms are not only promising for regression and classification purposes, they can also support the preprocessing of ChemCam data (Castorena et al., 2021; Ewusi-Annan et al., 2020) and enable autonomous ChemCam target selection with the Autonomous Exploration for Gathering Increased Science (AEGIS) targeting system (Francis et al., 2017).

This study focuses on the classification of ChemCam observations from Mars through an approach relying on the repeated application of k-means clustering in combination with the feature extraction method Non-Negative Matrix Factorization (NMF). The derived clusters are compared with the stratigraphy explored by the Curiosity rover during the first 2756 sols of the mission in order to derive conclusions about the dominant compositions in each cluster and to verify the obtained classes. The results of this study are relevant in two ways: First, they summarize the observations made by ChemCam and help to identify dominant compositions detected in the Gale crater. Second, the resulting class labels can be used as training data for machine learning-based classification models, which can support a rapid classification of new targets in the future. To address the second point, we also discuss in this study a random forest classifier trained with the clustering results and the predicted cluster memberships of ChemCam observations made between sols 2756 and 3126.

2. Techniques and Methods

2.1. ChemCam Instrument

We use elemental spectroscopic LIBS data collected with the ChemCam instrument suite (Maurice et al., 2012; Wiens et al., 2012). For LIBS, a laser is focused onto the target's surface and creates a luminous micro-plasma containing excited atoms, ions, and molecules that emit characteristic photons. The spectral analysis of this light provides spectra with emission lines of species that belong to the sample. ChemCam can measure targets at remote distances up to 7 m, but most targets are chosen below 3.5 m (Maurice et al., 2016). The mast unit with the laser and the telescope system (Maurice et al., 2012) is located on top of the mast of the rover while the body unit consisting of a demultiplexer and three spectrometers is placed inside the rover's body (Wiens et al., 2012). The spectrometers cover together a spectral range of 240–850 nm with gaps at 340–385 and 469–492 nm and the three spectral ranges are referred to as UV, violet (VIO), and visible and near-infrared (VNIR). The laser spot size is in the submillimeter range (350–550 μm), which provides the ability to track elemental variations at a small scale. A LIBS measurement on a ChemCam target is typically a raster, horizontal, vertical, or as a grid, of 5–25 observation points from which each corresponds to usually 30 laser shots (i.e., 30 spectra) (Maurice et al., 2016; Wiens & Maurice, 2015). This allows to track chemical variability also with depth as each laser shot samples material slightly deeper than the previous one. In this study, we work with mean spectra from single observation points excluding the spectra from the first five shots in order to avoid signal interferences from superficial dust. Besides the LIBS technique, ChemCam has also the capability to take high-resolution images with a remote micro imager (RMI) in order to capture the context of the investigated targets (Le Mouélic et al., 2015) and to do passive spectroscopy in the 400–840 nm wavelength range (Johnson et al., 2015).

2.2. Data Set

The data set for this study consists of mean LIBS spectra from 18,719 observation points corresponding to around 2,300 individual targets acquired from landing to sol 2756 of the mission. We excluded targets that were further than 3.5 m from the instrument as these are suspected to be more influenced by experimental factors such as a lower irradiance on the sample surface (Melikechi et al., 2014). Furthermore, samples that are expected to be nearly pure iron oxides were not considered in the data set. These samples are out of the compositional range for the usual quantification routine and need to be treated separately (David et al., 2020). Even though the procedure does not use quantified major element abundances, they are used for the cluster interpretations.

2.3. Methods

In this section, a description of the data preprocessing and brief introductions to the methods used in this work are given. The whole procedure until the final cluster assignments, which employs all these methods is then described in Section 2.4. For several steps in the procedure, we use functions provided by the Python Programming Language's *scikitlearn* library (Pedregosa et al., 2011). Below, we give a description of the particular processing and scikit-learn algorithms we employed:

2.3.1. Data PreProcessing

The raw data were processed by the standard routines for ChemCam data treatment, including dark and noise removal, continuum subtraction, wavelength calibration, and the transformation from counts to photons as described in Wiens et al. (2013). The particular data preprocessing for this study starts with the normalization of the spectra where each pixel of a spectrum was divided by the sum of all counts measured in the three spectrometers, which is referred to as norm 1 (Clegg et al., 2017). Next, the spectra were standardized, thus mean centered and scaled to unity. In order to prepare the data for matrix decomposition with NMF (more details in Section 2.3.2), it has to be ensured that the spectra do not contain negative values. Hence, from each spectrum its minimum value, which is always negative due to the mean centering process, was subtracted in order to shift the whole spectrum to only nonnegative values.

2.3.2. Non-Negative Matrix Factorization (NMF)

NMF is a method to decompose a data matrix with the constraint that all involved matrices have only nonnegative entries (Pauca et al., 2006). The general principle is similar than for other source separation techniques like Singular Value Decomposition (SVD) and Independent Component Analysis (ICA), but with the assumption of nonnegativity in the data set. Considering a data matrix M with dimensions $n \times m$ and an integer r, the factorization can be written as $M = WH$ with W and H being matrices of dimensions $m \times r$ and $r \times n$, respectively. Thus, r determines the number of nonnegative ranks or sources that are stored in H while W gives for each data vector in M the scores, meaning how much of each source is needed to construct the data vector. According to the nonnegativity constraint, only additive and no subtractive combinations of the sources, also called factors, are possible, which makes it in particular useful for data with a physical interpretation. In this study, we used the NMF decomposition provided by python's *scikitlearn* library, which relies on minimizing the Frobenius norm of the residual matrix $M - WH$ by alternatingly updating W and H (Cichocki & Phan, 2009; Févotte & Idier, 2011).

2.3.3. K-Means Clustering

K-means clustering is an iterative method in which a fixed number of clusters are found in the data (Jain, 2010). The iteration is usually initiated with a random selection of cluster centroids. During the iterations, the cluster centroids are moved in order to minimize the sum of squared distances of cluster members to their cluster centroid. Among the iterations, the cluster assignments can change as they are made in such a way that the variance is minimized within each cluster. Based on this principle, several variations exist; for this study, we used the k-means clustering algorithm from python's *scikitlearn* library with random initialization and Euclidean distances.

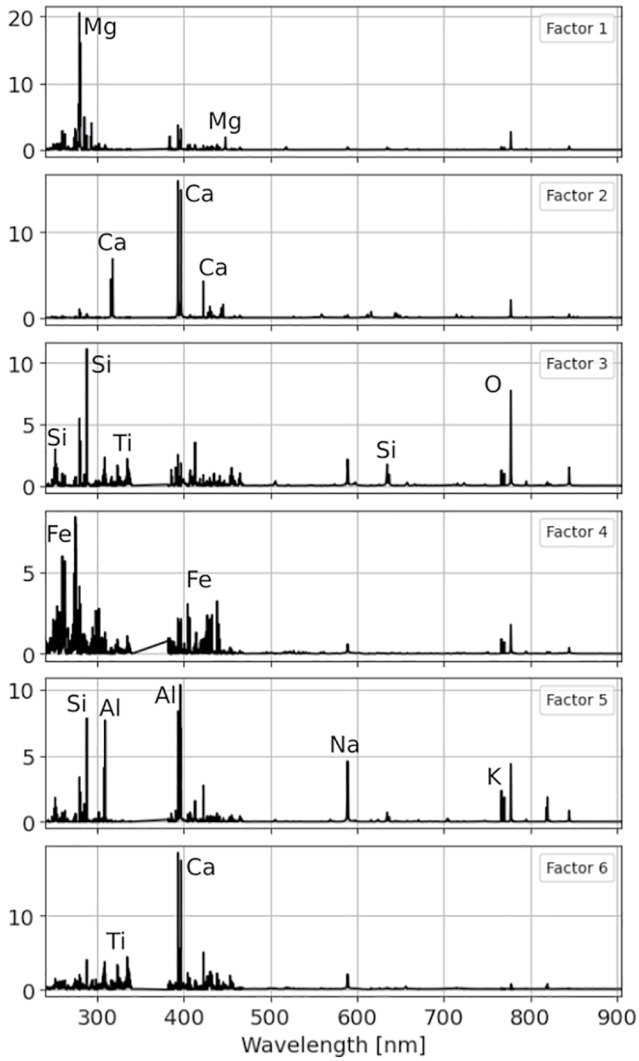

Figure 1. NMF factors of the six-dimensional fixed model, which is applied to the randomly selected sub-data sets. For each factor, dominant emission features are annotated with the corresponding element.

2.3.4. Silhouette Score

We employ the silhouette score to evaluate the cluster quality as it provides for each data point an estimation of how well the point fits in its own cluster or if it might belong to the nearest neighboring cluster. The mathematical formulation is based on the intra-cluster and nearest-neighbor Euclidean distance of each point. The silhouette score can have values in the interval (−1,1) where silhouette scores close to 1 indicate a good match to the own cluster while values close to −1 mean that the point is closer to the nearest cluster. Silhouette scores close to 0 can be interpreted as an overlap between different clusters. A detailed and illustrative description of the silhouette score can be found in Rousseeuw (1987). For this study, the silhouette scores are derived by the corresponding function of python's *scikitlearn* library based on Euclidean distances.

2.4. Procedure

The whole procedure to obtain clusters of similar characteristic from the large ChemCam data set consists of multiple steps that we chose to repeat 100 times. We discuss below (Section 2.4.2) how we chose several parameters in this study, including the repetition value of 100 and the six NMF dimensions. Below, we provide a brief overview of these steps and we provide more detail in later sections. One run of the clustering procedure is as follows:

1. A sub-data set is randomly selected using the *train_test_split* function of python's *scikitlearn* library. The sub-data set contains 4,866 observations, which is 26% of the whole data set in this study.
2. A fixed six-dimensional NMF model is applied to the sub-data set resulting in six NMF scores for each spectrum.
3. K-means clustering with the number of clusters fixed at six is applied to the NMF scores of the sub-data set.
4. The obtained clusters are ordered based on their mean NMF scores in order to get always the same cluster 1, cluster 2, etc.

2.4.1. NMF Model

In step 2 of the described procedure, a fixed NMF model is applied to the spectra of the sub-data sets. We find that the derived NMF components are generally the same for each split of the data, therefore, their averages were used to build the fixed and final NMF model whose six components are shown in Figure 1 with annotations for dominant features on each factor. Using a fixed model has several advantages: (a) it ensures the comparability of the outcomes from all repetitions, (b) it can be used to make predictions of new targets and to compare them with the derived clusters, and (c) it reduces the computational effort. Even though the NMF factors contain no quantitative information, the score values of a transformed spectrum can indicate the major elemental contributions. Factor 1 is characterized by strong Mg emission lines, which are mainly in the UV while factor 2 consists almost only of Ca emission lines in all three spectral ranges. On factor 3, Si emission lines have a major contribution but also strong features from Ti and O are present. Factor 4 is mainly characterized by multiple strong Fe emission lines in the UV and VIO spectral range. Emission lines from elements typical of felsic minerals like Si, Al, and the alkalis Na and K are dominant on factor 5. Factor 6 shows mainly multiple Ti emission lines in the UV spectral range, but also Ca lines in the VIO.

2.4.2. Consistency Among the Repetitions

Before arriving at the previously described procedure, several parameters were varied and investigated to determine the appropriate steps. We assessed different numbers of fixed clusters, NMF dimensions, and the

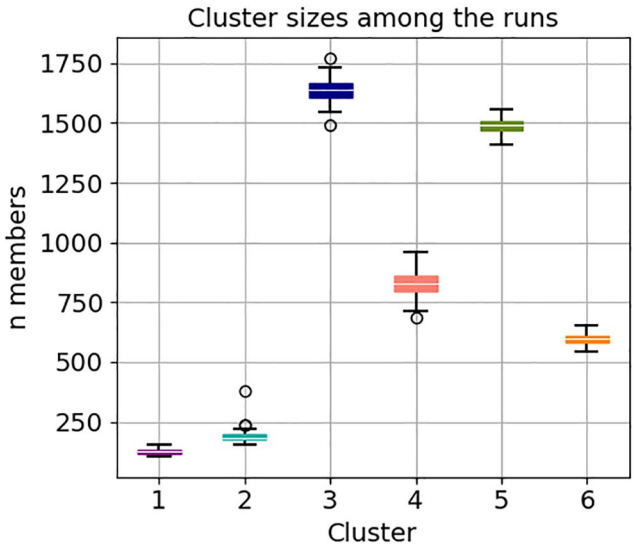

Figure 2. Boxplot of the number *n* of members in each cluster among the 100 runs. Besides one outlier run, which becomes apparent in particular for cluster 2, the number of members *n* is consistent among the runs indicating the robustness of the approach.

selected observations. Additionally, we tested different types of clustering algorithms like agglomerative hierarchical clustering with two types of linkage (ward and complete), spectral clustering and k-means clustering. It was found that k-means clustering with six fixed clusters in a six-dimensional NMF space is the most robust, which was assessed by consistency among the 100 runs. In Figure 2, the distribution of members in each cluster for the 100 repetitions is shown. Besides one outlier, which is best observable for cluster 2, the cluster sizes are consistent. In each run, two comparatively small clusters (cluster 1 and cluster 2), two mid-size clusters (4 and 6), as well as two rather large clusters (3 and 5) are obtained. Besides the cluster sizes among the repetitions, the cluster centroids in the NMF space are also consistent. Figure 3 shows the 100 centroids for each cluster on all six dimensions and they appear mostly at similar positions during the runs. Again, one outlier run for cluster 2 is observable, which is the same as seen in Figure 2. Smaller clusters such as cluster 1 and cluster 2 show a larger scatter of their centroid positions than the large clusters 3 and 5, which is not surprising as small clusters are more sensitive to positions of individual points.

2.5. Final Cluster Assignment

During the 100 repetitions, each observation was chosen multiple times to be in the random sub-data sets on which the clustering was applied. On average, each observation was selected 26 times; however, how often an observation was selected varies between 14 and 36 times. From the 18,719 spectra in this study, 17,293 were always assigned to the same cluster when they were selected, which is ≈92%. There are different reasons why an observation is not always assigned to the same cluster, for example, the composition can be very distinct and different from the dominant compositions in the data set. Another possibility is that an observation is located at the border between two clusters and depending on the distributions in the sub-data sets, the assignment alternates between the two clusters among the runs, in which the particular observation was selected. However, the ambiguous 8%, which was not always assigned to the same

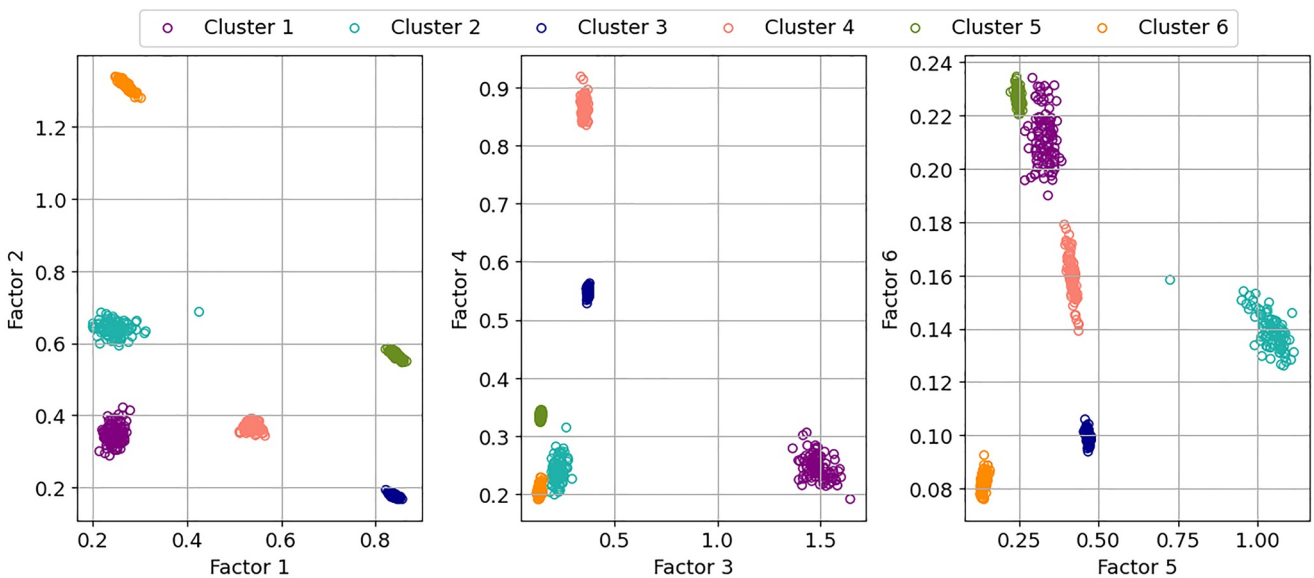

Figure 3. Centroids of the 100 runs found by the k-means algorithm shown for all NMF factors. Except for one outlier of cluster 2, which corresponds to the outlier run seen in Figure 2, the positions of the centroids are stable in each repetition. The centroids of the smaller clusters 1 and 2 have a larger scatter than those of the biggest clusters 3 and 5.

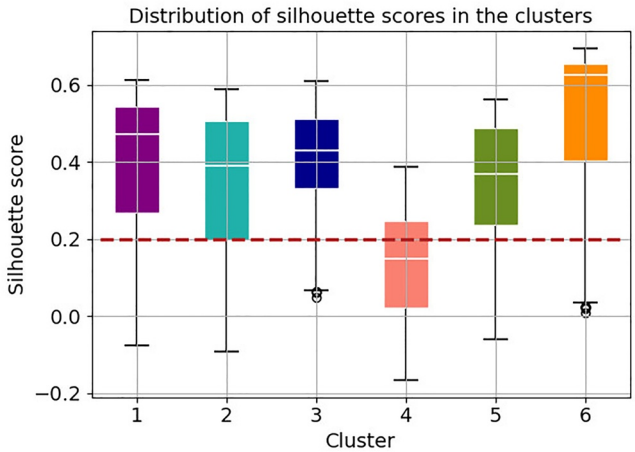

Figure 4. Boxplot of the silhouette scores of those samples that were always assigned to the same cluster. A negative silhouette score indicates that the point is closer to another cluster than its actual assigned cluster. The red dashed line shows the threshold of 0.2, which was applied for all clusters except for cluster 4.

cluster, was not further considered in this analysis. Additionally, in order to strengthen the distinction of the clusters from one another, further observations were rejected based on their cluster quality. The decision to discard observations even when they were always assigned to the same cluster among the repetitions is driven by the aim to use the obtained labels for the training of machine learning classifiers as the quality of training data has a strong impact on the performance of such models. We computed the silhouette score for the consistent observations whose results can be seen as a boxplot in Figure 4. Silhouette scores smaller than zero indicate that the observation belongs more likely to another cluster, while a score close to one indicates an observation very close to the other members of its cluster compared to the average distance to members of the nearest cluster. We decided for a threshold of 0.2 for all clusters except for cluster 4. Observations with a silhouette score smaller than this value are clearly not considered to belong to their clusters and were rejected. However, more than half of the cluster 4 observations fall below this limit, which can be explained by a large spread within the cluster and a small distance to its nearest neighbors at the same time, which can already be seen for the centroids positions in Figure 3. Therefore, the threshold for cluster 4 members was set to the first quantile, which is 0.023. Thus 25% of unambiguous cluster 4 observations were excluded. Following the described procedure, the cluster assignments were finalized and their distribution is shown in Figure 5. The final distribution is in agreement with the repetitions in terms of cluster size. Cluster 1 (n = 364) and cluster 2 (n = 485) are the smallest clusters, while cluster 3 (n = 5,512) and cluster 5 (n = 4,340) contain more than half of the observations.

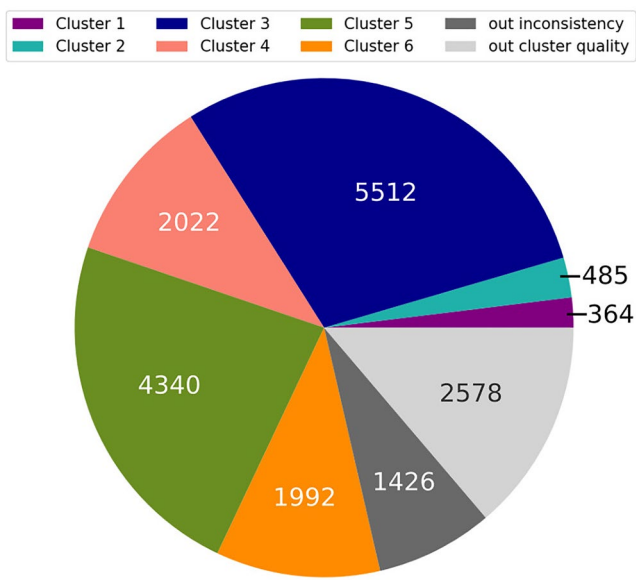

Figure 5. The final distribution of all observations among the clusters. The gray parts correspond to observations that were rejected because of either inconsistency among the repetitions or the silhouette score. As observed for the cluster sizes in each run, see Figure 2, the clusters vary in size.

3. Results

3.1. General Observations

For the interpretation of the derived clusters, we considered different properties such as the score values on the NMF factors, the Major Oxide Compositions (MOC), Pearson correlation coefficients of both NMF scores and MOC values in each cluster (see Tables S1–S6), and textural characteristics observed in the RMI images. The MOC values are derived using the standard procedure of quantification of ChemCam LIBS data described in (Clegg et al., 2017) for which a balanced combination of partial least squares regression and ICA is used. In Table 1, the mean, median, and standard deviation of each cluster for the concentrations of SiO_2, TiO_2, Al_2O_3, FeO_T, MgO, CaO, Na_2O, and K_2O as well as the totals (sum of oxides) are summarized. Usually, the MOCs are not normalized to 100 wt% so that the totals can indicate over- or underestimations of elemental abundances or which is probably more often the case, the presence of an element, which is not routinely quantified such as sulfur and hydrogen (Nachon et al., 2014; Rapin et al., 2017; Schröder et al., 2015). Even though each cluster has a distinct composition in terms of mean and median concentrations, the standard deviations suggest that their concentrations overlap for some elements. Therefore, a more visual comparison is given in Figure 6, which shows density plots of the abundances of each major oxide in the six clusters. Similar plots can be seen in Figure 7 but with the NMF score values instead of the MOCs. Overall, the observations from MOCs and NMF scores are in agreement as expected. For example, cluster 2 observations have the highest CaO concentrations and the highest score values on factor 2, which has almost exclusively Ca emission characteristics. Without detailing the particular compositions,

Table 1
Mean MOCs and Sum of Oxides (Totals) for Each Cluster Together With the Median and the Standard Deviation, all Values are in wt%

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
SiO_2	mean	75.77	56.16	54.70	48.92	44.42	28.32
	median	76.0	55.30	54.90	49.90	44.10	32.80
	stddev	5.48	4.27	2.63	6.14	3.39	16.65
TiO_2	mean	2.47	0.74	1.00	1.23	0.92	0.59
	median	2.32	0.75	0.97	1.03	0.90	0.65
	stddev	1.11	0.21	0.15	0.84	0.17	0.34
Al_2O_3	mean	5.54	19.30	11.95	10.94	9.42	5.32
	median	5.20	19.30	11.80	10.90	9.10	5.90
	stddev	3.33	2.59	1.45	2.50	2.03	3.01
FeO_T	mean	5.43	7.44	18.78	21.15	18.96	10.74
	median	4.90	7.40	18.90	20.10	19.10	11.80
	stddev	3.61	4.91	1.59	4.40	1.88	5.74
MgO	mean	2.43	1.50	5.79	3.43	7.91	2.69
	median	2.30	1.40	5.50	3.50	7.60	2.70
	stddev	0.85	0.77	1.32	0.92	2.30	1.11
CaO	mean	1.60	6.44	1.76	3.32	6.68	23.08
	median	1.10	6.80	1.70	2.80	6.70	21.50
	stddev	1.59	2.66	0.56	1.87	1.61	8.55
Na_2O	mean	1.69	5.70	2.72	2.91	2.28	1.01
	median	1.50	5.60	2.65	2.80	2.15	0.97
	stddev	0.86	1.26	0.63	0.77	0.66	0.67
K_2O	mean	0.66	1.54	1.43	1.16	0.62	0.30
	median	0.54	1.14	1.39	1.10	0.43	0.26
	stddev	0.67	1.40	0.48	0.58	0.63	0.28
Totals	mean	95.59	98.82	98.13	93.07	91.19	72.05
	median	95.45	98.79	98.18	93.01	91.13	77.88
	stddev	2.72	2.40	2.75	4.85	4.31	17.27

two different types of clusters can be generally observed: The first type are clusters that are characterized by high score and/or high MOC values for one particular factor/oxide. For example, cluster 1 has the highest SiO_2 concentrations and the highest scores on factor 3 of all clusters. Besides cluster 1, clusters 2, 4, and 6 also belong to this category. The second type are clusters which have high score and/or MOC values (but not the highest) for multiple factors/oxides. This applies to the largest clusters, 3 and 5, which have, for example, overlapping and high score values on factor 1. In Figure 8, five exemplifying RMI images of ChemCam targets are shown together with annotations indicating the cluster membership for each observation point. Some of these targets illustrate that the submillimeter ChemCam points represent diverse compositions within one target, represented by different clusters. The RMI images in Figure 8 also show different types of matrices, which can differ in their physical characteristics. Those differences can have a strong impact on the LIBS plasma and therefore on the data, which are known as matrix effects (Takahashi & Thornton, 2017). However, in this study we did not identify clusters based solely on a certain type of matrix or rock texture. Detailed descriptions and discussions of the individual clusters also relating to Figure 8 are given in the following.

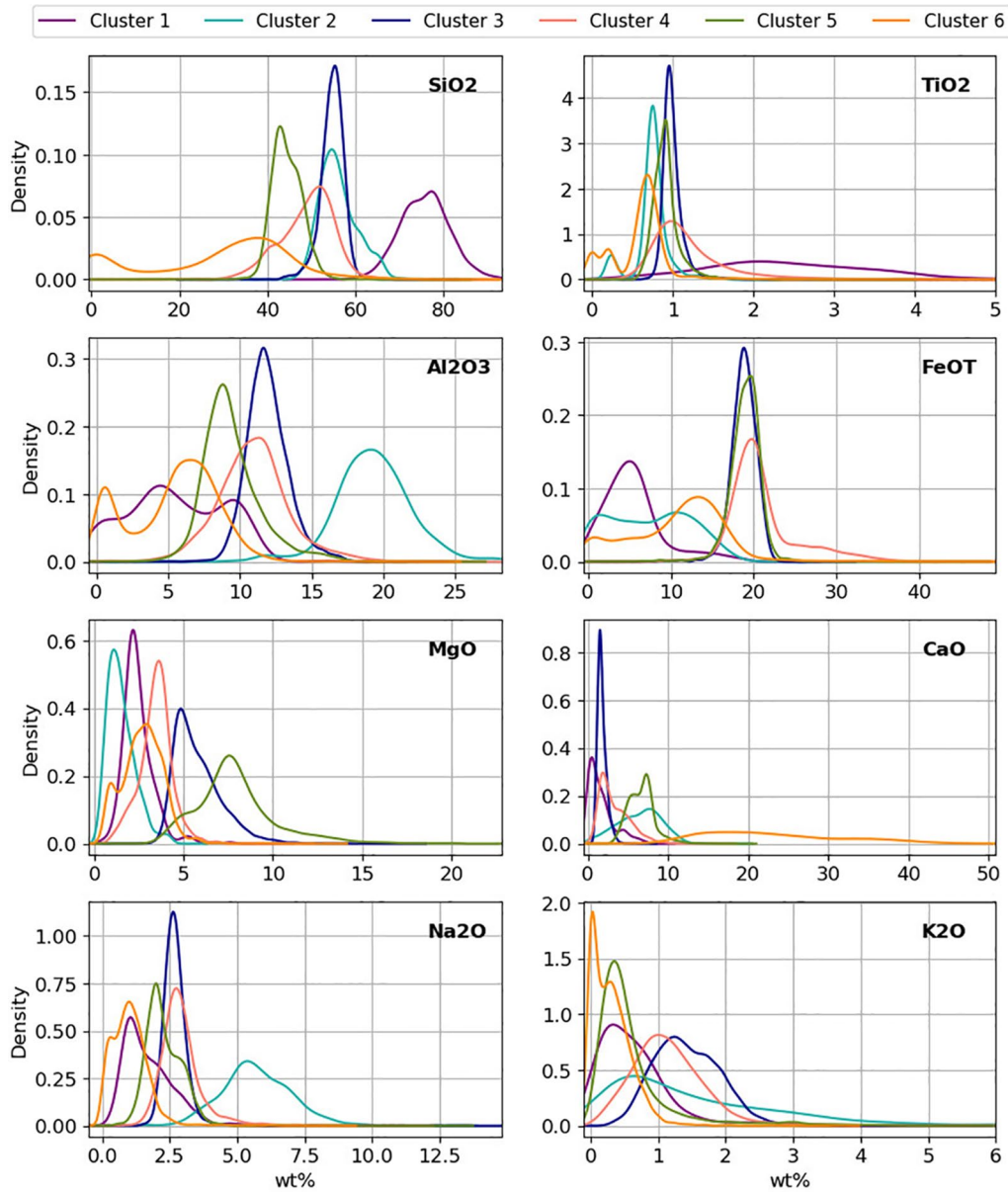

Figure 6. Another support for the cluster interpretation are the MOCs, which are shown here as density plots for each geological major element with its distribution within the clusters. The density curves were generated with Gaussian kernels and the bandwidths were selected following Scott's rule. The area below each curve is normalized to one. For TiO_2 and K_2O, the figures were limited to 5 wt% and 6 wt%, respectively, as there were only single observations with higher abundances of these oxides.

3.2. Detailed Cluster Descriptions

3.2.1. Cluster 1 — High SiO_2

Cluster 1 is characterized by elevated SiO_2 concentrations whose average is 75.8 wt%, see Table 1. Furthermore, members of cluster 1 have the highest score values on factor 3 (Figure 7), which is characterized by strong Si emission lines (Figure 1). Regarding Pearson correlations of NMF score values and MOCs among the members in cluster 1, an anti-correlation of the SiO_2 concentration with concentrations of typical elements from felsic compositions like Al_2O_3 and the alkalis is found (see Table S1). Moreover, the SiO_2 concentration has only a positive correlation with NMF scores on factor 3 and is not correlated with those on factor 5, which shows typical spectral signature of felsic compositions. This indicates that cluster 1 members

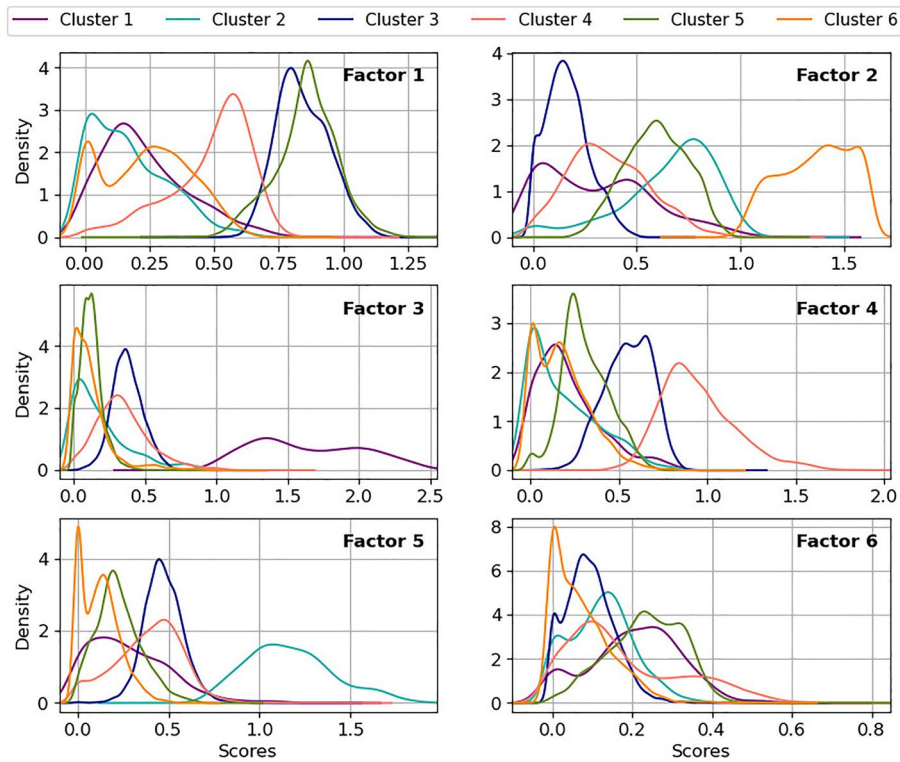

Figure 7. Another group of density plots but here showing the score values on the NMF factors. The density contours were generated in the same way as in Figure 6.

are not composed of felsic minerals even though SiO_2 is particularly elevated. The example of a cluster 1 observation in Figure 8e shows light-toned material forming a "halo" around a fracture, which was found to be typical for cluster 1 members. However, this type of material was not seen for all cluster 1 observations.

3.2.2. Cluster 2 — Felsic Compositions

The members of cluster 2 have their highest score values on factor 5 (Figure 1), which is represented by strong emission lines from typical elements of felsic minerals (Al, Na, and K). Also, the composition indicates a felsic origin with elevated Al_2O_3 (on average 19.3 wt%) compared to the other clusters, see Table 1. Pearson correlations between the cluster 2 members reveal an anti-correlation between the sum of Al_2O_3, Na_2O and K_2O, and the sum of FeO_T and MgO (see Table S2). Anticorrelations between these variables in the ChemCam data for Gale crater's sedimentary units have been shown to relate to the varying abundances of mafic and felsic minerals (Bedford et al., 2020; Meslin et al., 2013). As such, it is likely that this cluster relates to a high proportion of felsic minerals. Regarding correlations between Al_2O_3, Na_2O, and K_2O, no distinct positive correlations can be observed. This suggests that even though cluster 2 is dominated by these elements, different types of felsic minerals are assigned to it. The interpretation of cluster 2 is further supported when comparing its members to igneous rocks classified by Cousin, Sautter, et al. (2017) using images and ChemCam data. Most of the discussed rocks fall in cluster 2 except for those assigned to group 3 in Cousin, Sautter, et al. (2017), which is the group with the lowest Al_2O_3 concentration. Furthermore, this group 3 was found to be enriched in K-feldspars, whereas the other groups contain more plagioclase-type of feldspars (Cousin, Sautter, et al., 2017). Regarding cluster 2, this comparison could indicate that cluster 2 contains rather plagioclase feldspars than K-feldspars. The example shown in Figure 8d exhibits a typical texture of igneous rocks with light-toned phenocrysts. However, cluster 2 observations do not only cover igneous rocks, but also small pebbles like those in Figure 8a belong to cluster 2. Even though ChemCam has a small footprint, mixtures are often analyzed and only in a few cases pure mineral crystals are sampled (Payré et al., 2020). This is reflected for cluster 2 observations in Figure 9 where molar ratios of Al/Si are plotted versus those of (Fe + Mg)/Si. Despite the absence of a distinct mixing trend between felsic and mafic minerals, the diagram shows that cluster 2 contains at least some pure feldspar observations. Nevertheless,

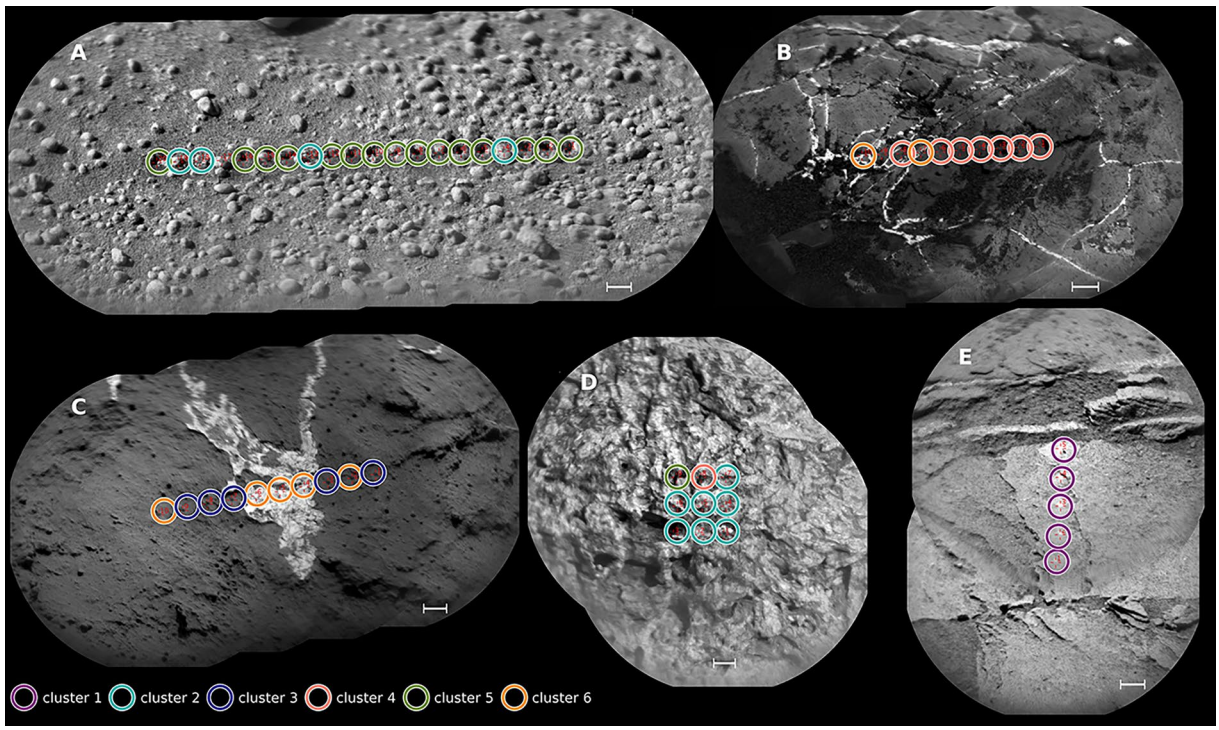

Figure 8. Remote micro imager images of exemplary Chemistry and Camera (ChemCam) targets representing typical members of the clusters. The scale bars in each image correspond to 5 mm. (A) *CC_BT_0607a*, sol 607, and 20 × 1 raster: Points in the soil are assigned to cluster 5, while those on small pebbles belong to cluster 2. (B) *Forres*, sol 2246, and 10 × 1 raster: Points targeting the dark bedrock are cluster 4 members, #7 and #10 on the light-toned veins are cluster 6 members. (C) *Chibanda*, sol 1470, and 10 × 1 raster: Points on the bright material as well as #2 and #10 belong to cluster 6, while the remaining points are assigned to cluster 3. (D) *Albee*, sol 654, and 3 × 3 raster: Large light-toned phenocrysts are visible and points on them belong to cluster 2, while points #8 and #9 on darker material are assigned to cluster 4 and cluster 5, respectively. (E) *Etunda*, sol 1317, and 5 × 1 raster: All points targeted a brighter-toned, fracture-associated halo and were assigned to cluster 1.

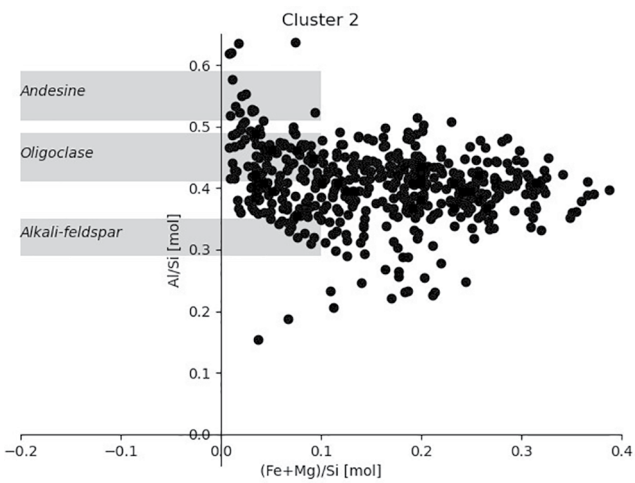

Figure 9. Element ratio diagram of Al/Si versus (Fe + Mg)/Si for all cluster 2 observations. The gray patches indicate ranges in which most likely a pure feldspar grain was sampled, while observations with (Fe + Mg)/Si ratios larger than 0.1 are considered to be mixtures but with a dominant feldspar contribution.

there are many mixtures of mafic and felsic minerals in cluster 2 with the felsic contribution being dominant. Moreover, the Al/Si ratio close to 0.5 supports the interpretation that the feldspar component of cluster 2 members is of a plagioclase type rather than a K-feldspar type (Al/Si ratio of 0.33). Usually these campaigns last longer than 10 sols and the stars mark approximately the temporal center of the corresponding campaign.

3.2.3. Cluster 3 — Low CaO

Cluster 3 shows no distinct high abundance of a particular geologic major element and for most of them cluster 3 has abundances in between those of other clusters, which can be best seen in the density plots in Figure 6. An exception is CaO whose abundances in cluster 3 are low compared to the other clusters. Except for cluster 1, all other clusters are on average higher in CaO. Therefore, we will refer to cluster 3 as the low-CaO cluster. Furthermore, cluster 3 is the only one in which CaO concentrations are not correlated with other oxides and there is no correlation between elements from felsic and mafic minerals (see Table S3). The average composition of cluster 3 overlaps with the silica-rich basalt endmember described in Bedford et al. (2019) when taking the standard deviations into account. Slight compositional deviations can occur due to the different approaches as targets with sum of oxides outside the 95–105 wt% range and diagenetic features were excluded in Bedford

et al. (2019). This silica-rich basalt endmember was mainly found in the Mt. Sharp group with increasing contributions toward higher elevation.

3.2.4. Cluster 4 — High FeO_T

Members of cluster 4 have their highest score values on NMF factor 4, which shows strong Fe emission characteristics (Figure 1). The average composition (Table 1) also indicates elevated FeO_T abundances compared to the other clusters. However, the density plot in Figure 6 shows for FeO_T an overlap with clusters 3 and 5, but cluster 4 has a characteristic extended wing toward high FeO_T concentrations. Despite this overlap, cluster 4 is interpreted to contain high FeO_T targets as suggested by the high factor 4 scores in Figure 7. The limitations of high FeO_T concentration predictions from ChemCam data are known and a dedicated calibration model for high FeO_T targets was studied and proposed in (David et al., 2020). The average sum of oxides in cluster 4 of ≈93 wt% (Table 1) is a further indication for FeO_T abundances being underestimated for high FeO_T targets (David et al., 2020). No relevant Pearson correlations for MOCs or NMF scores are obtained within cluster 4, except for the already discussed slight deviation between quantified FeO_T concentrations and the scores on factor 4 as they correlate only weakly (0.47, see Table S4). A positive correlation between the sum of oxides and the sum of Al_2O_3, Na_2O, and K_2O can also be observed, but not between the sum of oxides and the sum of FeO_T and MgO. In addition, the absence of other distinct correlations could point out that the elevated FeO_T abundances of cluster 4 members cannot be attributed to a single Fe-rich mineral.

3.2.5. Cluster 5 — Low SiO_2

Cluster 5 is neither characterized by high abundances of a particular element nor by high NMF scores on a particular NMF factor, similar to the other large cluster, cluster 3. It is noticeable that members of cluster 5 are on average somewhat low in SiO_2 and have a low sum of oxides, see Table 1. In comparison to cluster 3, the abundance of CaO is distinctly higher for cluster 5, which is well visible in Figure 6. In particular for cluster 5, the RMI images of the corresponding points were important for interpretation as many of them show loose material like fine and coarse grained soils (a discussion about the terminology of soils can be found in Certini et al., 2020). This agrees with the low sum of oxides that is typical for soils due to volatiles (like S, P, C, and Cl) that are not routinely quantified and which are usually elevated in soils (Cousin, Dehouck, et al., 2017; Cousin et al., 2015; Meslin et al., 2013). Regarding Pearson correlations in cluster 5, an observable weak negative correlation (−0.45, see Table S5) between the sum of $Al_2O_3 + Na_2O + K_2O$ and the sum of FeO_T + MgO agrees with the typical mineral mixing of mafic and felsic minerals in soils. However, consolidated rocks with sandstone as well as mudstone textures are also frequently observed within cluster 5 so that it cannot be interpreted to be solely a soil cluster.

3.2.6. Cluster 6 — High CaO

This cluster has a clear signature of high CaO abundances: It shows the highest score values on NMF factor 2 (Figure 7) characterized by almost only Ca emission lines (Figure 1) and its quantified CaO abundances are high and discriminant (Figure 6). Pearson correlation coefficients reveal negative correlations of CaO abundance with all other oxides included in the MOC (see Table S6). Similarly, the NMF scores of factor 2 scores have a negative correlation with the scores on all other factors. Furthermore, cluster 6 has the highest median silhouette score value, see Figure 4, which implies that its members are mostly well separated from the other clusters. Such high CaO observations are often associated with Ca-sulfate occurrences either as veins (Nachon et al., 2014), as nodules (Bedford et al., 2019; Nachon et al., 2017), or as cements in the bedrock (Rapin et al., 2019). The images in Figures 8b and 8c with cluster 6 observation points are, for example, characterized by bright material as thin veins and a slightly broader vein features, respectively. Even points, which do not target the bright material in Figure 8c are assigned to cluster 6 indicating that Ca-sulfate is present as a cement within the bedrock or exists below the surface. However, not all points in cluster 6 are Ca-sulfate, there are other minerals with elevated CaO, for example, targets with strong CaF molecular bands, interpreted to be fluorite or fluorapatites (Forni et al., 2015, 2020), belonging to cluster 6. What is also noteworthy about this cluster is that it is the only cluster that contains observations with score values of 0 on all other factors except for factors 2 and 6. In total, 106 observation points have this characteristic and all of them sampled bright material as observed in the RMIs, which suggest the presence

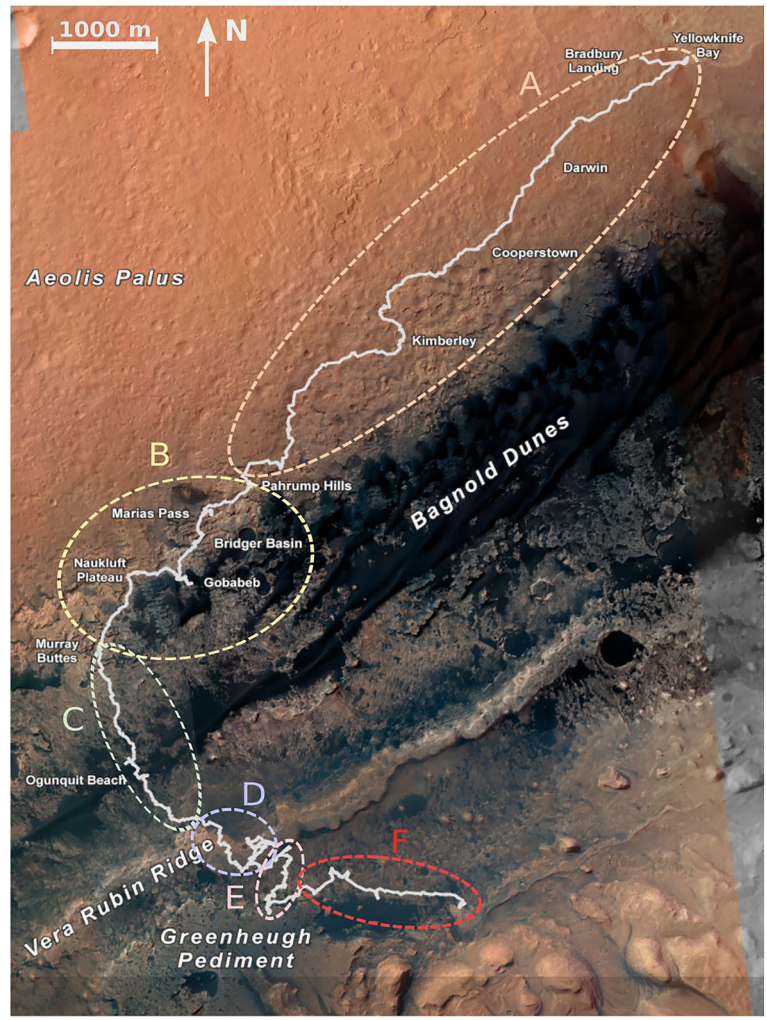

Figure 10. Context map showing Curiosity's 25 km long traverse (in white) from Bradbury landing until sol 3088 of the mission. Background image from High Resolution Imaging Experiment (HiRISE) (NASA/JPL-Caltech/). This study exploits Chemistry and Camera laser-induced breakdown spectroscopy data measured from the beginning until sol 3007. The highlighted sections of the traverse correspond to the discussion of cluster detections in the context of the traverse and the various figures showing cluster detection fractions as bar plots. (A) Sol 0–750, Bradbury, Figure 11. (B) Sol 750–1380, Ascending Mt. Sharp, Part I, Figure 12. (C) Sol 1350–1800, Ascending Mt. Sharp, Part II, Figure 14. (D) Sol 1800–2300, Vera Rubin Ridge (VRR), Figure 15. (E) Sol 2300–2756, Beginning Glen Torridon, Figure 16. (F) Sol 2756–3007, Glen Torridon continued, Figure 17 showing predictions of cluster memberships.

of almost pure Ca-sulfate in these points. Overall, cluster 6 is distinct from the other clusters due to its high CaO composition but also because its main component Ca-sulfate exists in nearly pure form.

4. Cluster Detections Along the Traverse

In this section, the cluster detections are put in context of the traverse of Curiosity and the observations reported in multiple studies. This comparison aims to demonstrate that the clustering results support and contribute to the overall observations made during the mission by the ChemCam instrument. The cluster detections along the traverse are presented in intervals based on geological units visited by the rover. A HiRISE (High Resolution Imaging Experiment) orbital image from Gale crater with Curiosity's traverse in Figure 10 provides the context and the different sections, which will be discussed in the following. More detailed views of the path can be found in multiple studies, which are also referred to in the text. Barplots of

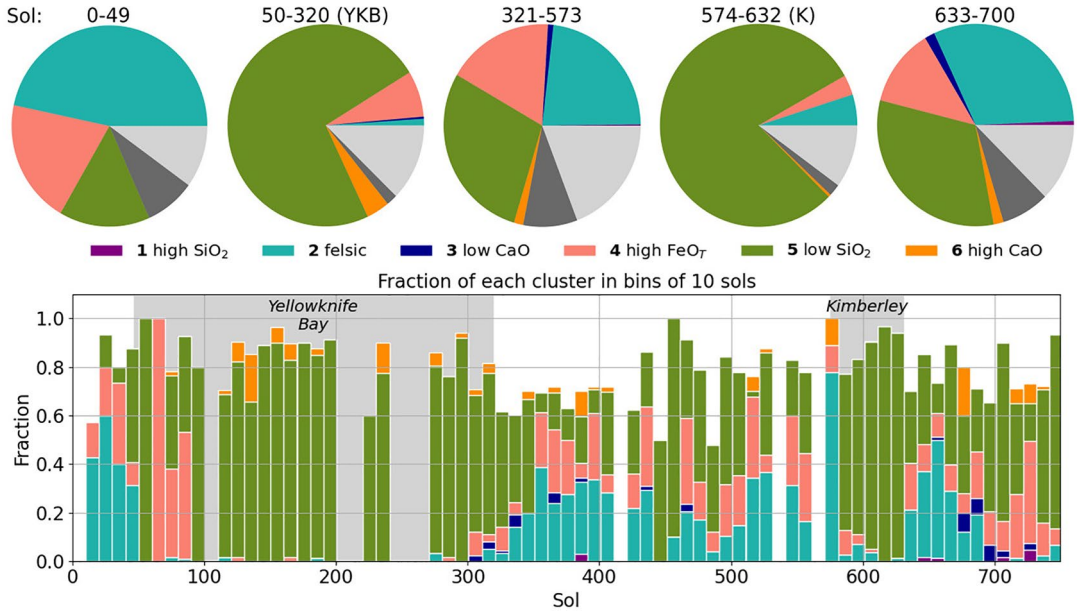

Figure 11. Bradbury with YKB: Yellowknife Bay and K: Kimberley. Proportion of each cluster observations in a constant bin of 10 sols. Rejected targets were counted, too, which explains the summed fractions smaller than 1. No bar was plotted when there were no ChemCam observations in the particular time period. It has to be noted that the number of total observations in one bin can vary for each bin. The gray fractions in the pie diagrams represent rejected observations. The dark gray corresponds to rejections due to inconsistencies while the light gray were rejected according to the silhouette score criterion.

cluster observation proportions in constant time bins of 10 sols are shown to visualize the outcomes as well as pie diagrams for particular time periods (Figures 11, 12 and 14–16). In both cases, it was also accounted for rejected observations.

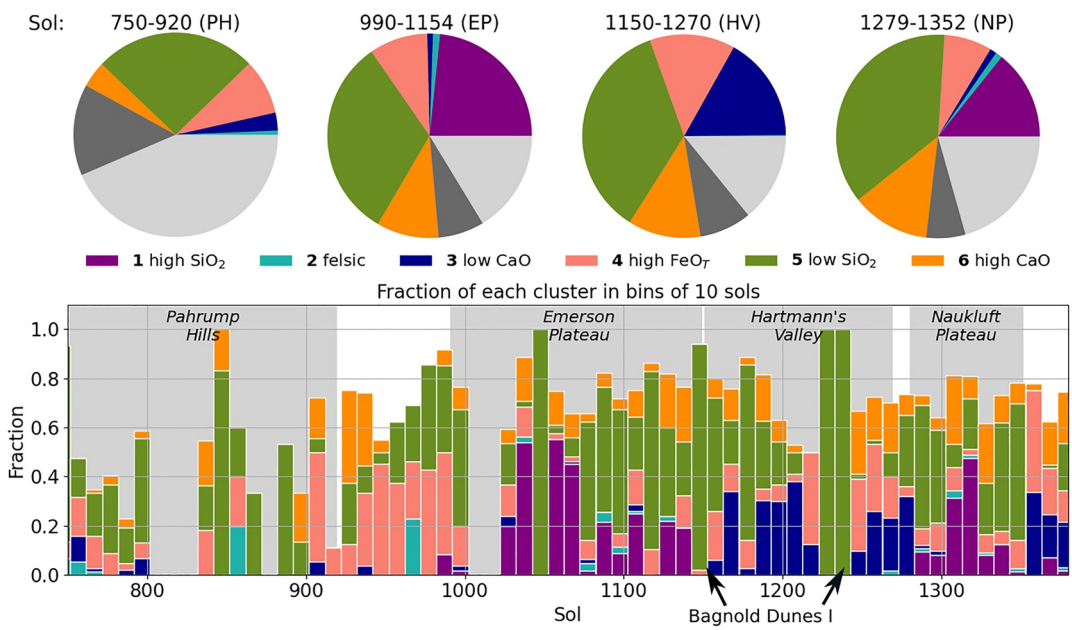

Figure 12. Same type of plot as in Figure 11, here: Ascending Mt. Sharp with PH: Pahrump Hills, EP: Emerson Plateau, HV: Hartmann's Valley, and NP: Naukluft Plateau. The gray fractions in the pie diagrams represent rejected observations. The dark gray corresponds to rejections due to inconsistencies while the light gray were rejected according to the silhouette score criterion.

4.1. Bradbury

The Bradbury group was studied from the Bradbury Landing until the Pahrump Hills area (Figure 10, section A) where the transition to the Mt. Sharp group takes place (Grotzinger et al., 2015). The corresponding visualizations are shown in Figure 11. The Bradbury Landing is characterized by dominant cluster 2 detections as well as observations assigned to clusters 4 and 5. As the landing site is close to the alluvial fan of Peace Vallis, a fluvial channel cutting through the northern rim of the Gale crater (Anderson & Bell, 2010; Palucis et al., 2014), the frequent cluster 2 observations, and targets with dominant felsic compositions, are interpreted as sediments derived from either a plagioclase-phyric basalt (Siebach et al., 2017) or a trachybasalt (Bedford et al., 2019; Edwards et al., 2017),which have been transported from the rim to their recent location by fluvial activity (Cousin, Sautter, et al., 2017; Sautter et al., 2014, 2015). The rover made a slight detour southeast from the direct path up to Mt. Sharp and visited Yellowknife Bay located at a contact between three different geologic units (hummocky plains of landing site, bedded fractured unit and cratered surface unit) (Grotzinger et al., 2014). The arrival at this location becomes apparent in the cluster analysis due to an abrupt decrease of cluster 2 detections after sol 50. This change in composition becomes apparent when comparing the pie diagrams in Figure 11 of the time periods before the rover arrived at Yellowknife Bay (sol 0–49) and the time spent there (sol 50–320). The latter is dominated by cluster 5 detections and has only a small proportion of cluster 2 observations, which dominated the previous pie diagram. One of the first outcrops investigated at the Glenelg member of Yellowknife Bay is the Rocknest outcrop, which is characterized by elevated FeO_T abundances (Blaney et al., 2014) and which can be identified in the cluster analysis based on primarily cluster 4 detections around sol 70. ChemCam observations within the other members, Gillespie Lake and Sheepbed, and their outcrops at Yellowknife Bay are mostly assigned to cluster 5. This agrees with ChemCam findings at Yellowknife Bay reported in Mangold et al. (2015) showing that the upper strata of Yellowknife Bay, mainly the Rocknest outcrop, which is closer to the hummocky plains, are different than lower lying strata. Overall, the geochemical analyses with ChemCam and APXS demonstrate a composition close to the average Martian crust (McLennan et al., 2014) at Yellowknife Bay suggesting isochemical alteration, which explains why soils and these rocks at Yellowknife Bay both fall in cluster 5. Furthermore, all three members of Yellowknife Bay contain Ca-sulfate observations mostly as light-toned veins (Nachon et al., 2014), which appear in the cluster analysis as several cluster 6 detections for the first time.

When the rover left Yellowknife Bay and continued its traverse toward Mt. Sharp, frequent cluster 2 as well as cluster 4 detections reappear, while the cluster 5 proportion becomes quite variable and sometimes rare. The resurgence of felsic cluster 2 analyses correlates to the appearance of conglomerate and float rock targets that had abundant feldspar mineral grains (Mangold et al., 2015). ChemCam analyses of these felsic conglomerate and float targets suggest that they are likely derived from a different sediment source compared to sediments from Yellowknife Bay with a chemistry similar to trachybasalt (Bedford et al., 2019; Cousin, Sautter, et al., 2017; Edwards et al., 2017).

No distinct changes in composition based on the cluster analysis can be observed until the rover reaches a region informally named Kimberley on sol 576, whereas for Yellowknife Bay, cluster 2 detections significantly decrease in abundance and almost all targets are assigned to cluster 5. Furthermore, the pie diagram of the Kimberley period (sol 274–632) is notably similar to that of Yellowknife Bay (sol 50–320). The main finding at Kimberley is an enrichment in K reported by ChemCam (Le Deit et al., 2016) as well as by APXS (Thompson et al., 2016). Furthermore, the CheMin instrument detected 21 wt% of sanidine, an alkali feldspar, in the *Windjana* drill sample at Kimberley (Treiman et al., 2016). Because cluster 2 contains mainly plagioclase feldspars and not alkali feldspars, the Kimberley observations are mostly assigned to cluster 5 and not to cluster 2. It can also be that sanidine mineral grains are too small to be targeted individually with ChemCam and therefore no pure or dominant sanidine observation was made with ChemCam (Payré et al., 2020), except in some single shot LIBS spectra (Cousin et al., 2021). This does not become apparent in this study as average spectra were used for the cluster analysis. Furthermore, the abundances of Al_2O_3 of ChemCam observations at Kimberley (Le Deit et al., 2016) are overall lower than those of cluster 2 observations (Figure 6). When the rover left the Kimberley region around sol 632, clusters 2 and 4 detections similarly reappear as observed after Yellowknife Bay. After sol 700, the number of cluster 2 observations decreases again, but cluster 4 observation points remain present. At this time, Curiosity arrived at Hidden

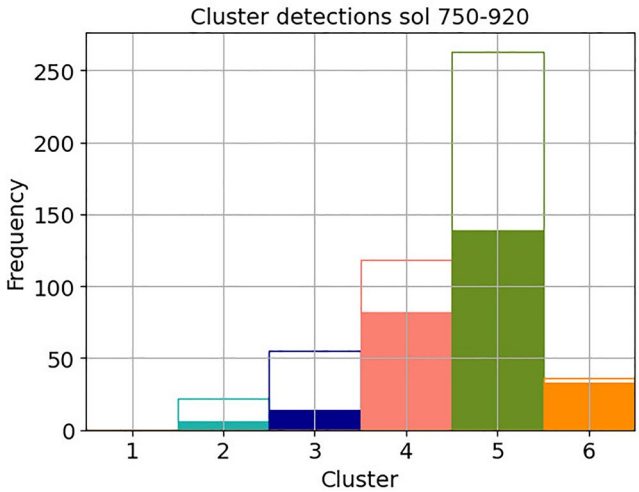

Figure 13. Detection frequency of the six clusters during the time period the rover explored the Pahrump Hills member (sol 750–920). The filled bars show the detections after the silhouette score criterion was applied while the outlined bars indicate the cluster detection frequencies beforehand. Relatively strong reductions are present for clusters 2, 3, and 5, but proportions are preserved.

Valley and then took a traverse westwards arriving at the transition to Mt. Sharp strata on sol 750 of the mission.

4.2. Ascending Mt. Sharp

On sol 750 of the mission, the Curiosity rover reached the transition from the Bradbury group to the Mt. Sharp group. In Figures 12 and 14, the bar- and pie plots of cluster detections in particular time intervals until the rover reached the Vera Rubin Ridge (VRR) are shown. In both figures, we highlight the time periods, in which the rover explored the five lithostratigraphic members of the Murray formation up to VRR (Pahrump Hills, Hartmann's Valley, Karasburg, Sutton Island and Blunts Point) and two plateaus (the Emerson and the Naukluft) of the Stimson formation. Sections B and C in Figure 10 correspond to these stages discussed in the following. The members are not discussed in the chronological order in which they were visited by the rover, but based on the Murray and Stimson formation as well as the Bagnold Dunes.

4.2.1. Pahrump Hills

The Pahrump Hills member is distinct in the cluster analysis because of many rejected observations rather than particular cluster detections, which results in a pie diagram (sol 750–920) with rejections of more than half of the ChemCam observations in this period. To understand these aggregated rejections, a histogram of cluster detections before and after the sorting out based on the silhouette score (see Section 2.5) is shown in Figure 13 for the time period the rover spent at Pahrump Hills. The outlines correspond to the detection frequency before the sorting out, while the filled bars are the final cluster detections. Even though cluster 3 detections are already relatively rare before the rejection process, there are even less afterward as 75% were discarded during the routine. Cluster 5 is also affected by the quality control: almost half of its members were rejected. The reason for the enhanced rejection rate can be that CaO concentrations measured by ChemCam at Pahrump Hills have an average composition of 4.8 wt% (Mangold et al., 2019), which is lower than the cluster 5 and higher than the cluster 3 average, see Table 1. Cluster 4 has the closest CaO average concentration to the Pahrump Hills average, but the FeO_T abundances are too low at Pahrump Hills to be assigned to cluster 4. Thus, based on the CaO abundances, the rejected observations are most likely not well bounded to their clusters and fall in between cluster 3 and cluster 5. Without the cluster quality check, the average CaO concentrations of cluster 3 and cluster 5 at Pahrump Hills are 3.6 wt% and 6.0 wt%, respectively. After the rejection step, a clear shift is observable: The remaining 13 cluster 3 members have on average 2.9 wt% CaO while cluster 5 has on average 7.1 wt% CaO. In addition, the mineralogical analysis with the CheMin instrument of two Pahrump Hills targets, *Confidence Hills* and *Mojave2*, indicate a distinct mineralogy compared to Bradbury targets but also to Murray formation targets at higher elevations (Rampe, Blake, et al., 2020; Rampe et al., 2017). In particular, the Pahrump Hills drill targets have a higher magnetite to hematite ratio and higher abundances of igneous mafic minerals such as pyroxene (Rampe et al., 2017).

4.2.2. Other Murray Members

After the exploration of Pahrump Hills, the rover arrived at the Emerson plateau, which will be discussed in Section 4.2.4 together with the Naukluft plateau. In between these plateaus, the rover traversed Hartmann's Valley in the Murray formation. During this time (sol 1150–1270), cluster 3 detections substantially contribute to the corresponding pie diagram for the first time. Besides the low detection frequency of cluster 3 observations at Pahrump Hills, this is in agreement with the distinct compositional difference between Bradbury and Mt. Sharp mudstones reported in Bedford et al. (2019). The latter were hypothesized to be derived from a different sedimentary provenance compared to the Bradbury group, which most probably could be a more silica-rich basalt that likely underwent some chemical weathering (Bedford et al., 2019). The silica-rich basalt endmember of Bedford et al. (2019) is similar to the average cluster 3 composition as described in Section 3.2.3. However, the dominant contribution in the pie diagram is still coming from

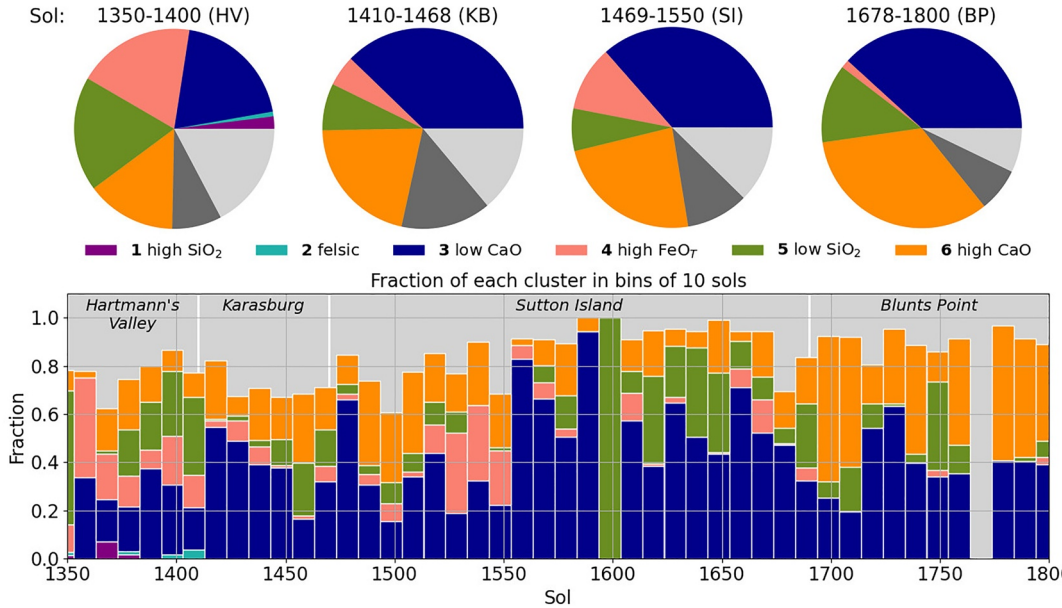

Figure 14. Same type of plot as in Figure 11, here: Ascending Mt. Sharp with HV: Hartmann's Valley, KB: Karasburg, SI: Sutton Island, and BP: Blunts Point. The gray fractions in the pie diagrams represent rejected observations. The dark gray corresponds to rejections due to inconsistencies while the light gray were rejected according to the silhouette score criterion.

cluster 5, which is mainly due to the investigation of the Bagnold Dunes described in more detail in Section 4.2.3. Hartmann's Valley continues after the Naukluft plateau, see Figure 14, and the pie diagram (sol 1350–1410) shows a similar distribution as the previous one except for a decreased contribution from cluster 5. With the transition to the Karasburg member at higher elevation, the frequency of cluster 3 detections further increases. This trend proceeds also for the following members at higher elevations seen in the pie diagrams from the Karasburg (sol 1410–1470), the Sutton Island (sol 1470–1690), and the Blunts Point (sol 1690–1800) member, which all have dominant cluster 3 contributions. Even though the number of cluster 6 (high CaO) detections increases, at the same time, the transition from pie diagram distributions with major cluster 5 contributions to dominant cluster 3 contributions can be interpreted as a drop of CaO in the bedrock, as cluster 6 members can be mainly related to Ca-sulfate. This drop was observed with ChemCam (Frydenvang et al., 2020; Mangold et al., 2019) and is in agreement with the mineralogical analyses of drill holes from members above the Pahrump Hills member (Bristow et al., 2018). Even though the drill material from Bradbury and Murray (above Pahrump Hills and until VRR) targets both contain clay minerals, there are general differences regarding pyroxene and Ca-sulfate abundances as well as the type of Fe-oxides. The Bradbury targets contain more pyroxene, which was suggested to be in particular pigeonite with higher Ca fractions, than the Murray drill targets (Morrison et al., 2018). This can be a further explanation for the dominant cluster 5 detections at Bradbury in comparison to more cluster 3 (low CaO) detections at Murray. Mangold et al. (2019) investigated the chemical index of alteration (CIA), which serves as a proxy of open system weathering as it refers to the leaching of Ca, Na, and K relative to Al, for the Murray formation until VRR. They observed an increase in the CIA with elevation and a negative correlation with CaO abundances. This suggests that cluster 3 contains observations from more altered material than cluster 5, which was detected in areas with lower CIA such as Bradbury. Furthermore, the increasing contribution of cluster 6 detections, which is particularly strong in the pie diagram of the Blunts Point member, is in agreement with Ca-sulfate detected as cements in bedrock and not only in fracture fills (Rapin et al., 2019). Ca-sulfate as a bulk component in these Murray members is also observed by the CheMin instrument (Bristow et al., 2018).

4.2.3. Bagnold Dunes

During the first exploration of Hartmann's Valley (sol 1510–1270, Figure 12), the rover stopped at two locations for the investigation of the Bagnold Dunes. This is the first active extraterrestrial dune field explored by a rover (Bridges & Ehlmann, 2018). The two stops of the first Bagnold Dunes campaign can be seen in

Table 2
Cluster Assignments Related to Drill Campaigns Within Altered Halos and Their Parent Rocks at Two Locations in the Stimson Formation

	Altered		Parent	
	Greenhorn	*Lubango*	*Big Sky*	*Okoruso*
Before drill	1,6	–	5	5
Drill hole	4,6	1,6	5	5
Tailings	4,6	6	5	5

Note. Clusters are listed when more than one observation of the corresponding ChemCam rasters were assigned to this cluster.

Figure 12 in terms of bars with almost only cluster 5 detections around sol 1180 and sol 1230. Even though the dunes were found to be coarser grained than previously analyzed soils and exhibit a lower H signal in the ChemCam spectra (Cousin, Dehouck, et al., 2017), the similarities with soils are sufficient for making all observation points of Bagnold Dunes fall into cluster 5. As discussed earlier, the corresponding pie diagram of Hartmann's Valley (sol 1150–1270) is strongly influenced by this campaign and might therefore not be representative.

The Bagnold Dunes were further investigated in a second phase starting on sol 1602 at Mapleton and continued until sol 1658 including four stops (Lapotre & Rampe, 2018). The first stop clearly appears in the cluster analysis due to a bar showing only cluster 5 detections, see Figure 14. Nevertheless, besides observations on the dunes, rock targets were also analyzed with ChemCam during the campaign in between the four stops. This one distinct cluster 5 appearance around sol 1600 is more an effect of slightly varying observation numbers in one bin of 10 sols, which contains here only the 5 observations of the ChemCam target *Mapleton*. However, all Bagnold Dunes observations of Phase 2 were assigned to cluster 5 as already observed for phase 1. Possibly misleading visual impressions due to the varying numbers of observations in one bin can appear but they are rare. Representations with constant numbers of observations in each bin were also checked, which gave overall the same results. Similar as for the pie diagram of Hartmann's Valley containing the first Bagnold Dunes campaign, a slight increase of cluster 5 detections in the pie diagram of the Sutton Island member (sol 1470–1690) is observable due to the second Bagnold Dunes campaign.

4.2.4. Emerson and Naukluft Plateaus

The Emerson and Naukluft plateaus belong to the Stimson formation, which in turn is part of the Siccar Point group unconformably overlying the Murray formation (Fraeman et al., 2016). The Stimson formation is younger than the Murray formation and is interpreted to be an ancient dune field consisting therefore mostly of sandstones (Banham et al., 2018). ChemCam data from the Stimson formation of the two plateaus give an average bulk composition (Bedford et al., 2020) similar to the average composition of cluster 5 (Table 1) with only slight deviations. Both pie charts of the plateaus (sol 990–1154 and sol 1279–1352) indeed show a dominant contribution from cluster 5 observations. Similar to the active Bagnold Dunes (Cousin, Dehouck, et al., 2017), mineral sorting of mafic and felsic minerals was found for the Stimson formation (Bedford et al., 2020); hence, indicating a further relation to cluster 5.

One major difference of the Emerson and Naukluft pie diagrams to those of all other regions is the presence of the high SiO_2 cluster 1. These local detections are in agreement with high silica enrichments at Marias Pass and Bridger Basin reported in Frydenvang et al. (2017). The observed light-toned materials are enriched in SiO_2 up to >80 wt% and were mainly observed in fracture-associated halos whose origin is contested Frydenvang et al. (2017); Yen et al. (2017, 2021). Since those features were also found in the underlying Murray formation at these areas, the diagenetic event has to postdate the deposition and lithification of the Stimson formation and indicates therefore late stage groundwater activity (Frydenvang et al., 2017). The CheMin instrument analyzed the material from drill holes in the parent bedrock outside the halo (*Big Sky* and *Okoruso*) and in the altered halos (*Greenhorn* and *Lubango*) at two positions in the Stimson formation (Yen et al., 2017). They found that the crystalline components of parent and altered targets are mainly similar except for additional Ca-sulfate in the altered material. On the other hand, the alteration halos have a higher fraction of amorphous material, which is enriched in SiO_2. With the cluster analysis of the parent targets, we find that ChemCam observation points on the surface, in the drill hole, and in the drill tailings that are mainly assigned to cluster 5 representing the dominant composition of the Stimson formation. On the other hand, the observations from the drill campaigns of the altered material indicate mainly cluster 6 (high CaO) observations with a few points assigned to cluster 1 (high SiO_2) and some cluster 4 (high FeO_T) points for *Greenhorn*. These observations are also summarized in Table 2 and agree with CheMin findings, which report an increase of crystalline Ca-sulfates in the altered samples compared to the parent material and somewhat higher FeO_T in *Greenhorn* (Yen et al., 2017). Furthermore, CheMin identified tridymite, which is a high silica polymorph, in one drill sample called Buckskin at the Marias Pass location belonging

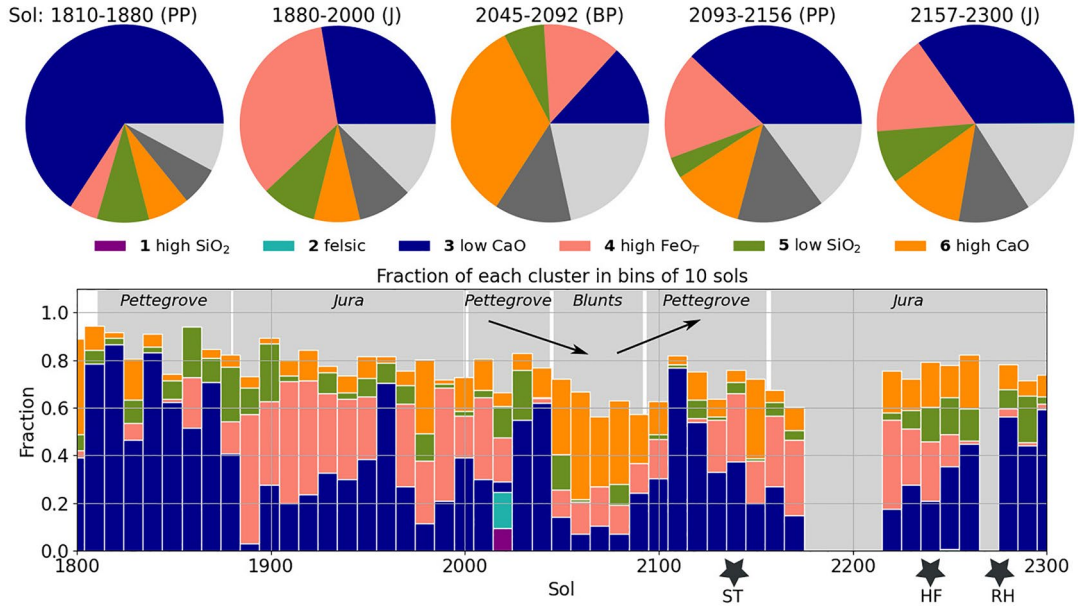

Figure 15. Same type of plot as in Figure 11, here: Vera Rubin ridge with PP: Pettegrove Point and J: Jura and BP: Blunts Point. The arrows indicate where the rover was going down section back to the Blunts Point member and up section returning onto the ridge. The stars indicate the time periods of successful drill campaigns on VRR with: ST: *Stoer*, HF: *Highfield,* and RH: *Rock Hall*. Usually, the drill activities last longer than 10 sols and the stars mark approximately the temporal center of the corresponding campaign. The gray fractions in the pie diagrams represent rejected observations. The dark gray corresponds to rejections due to inconsistencies, while the light gray were rejected according to the silhouette score criterion.

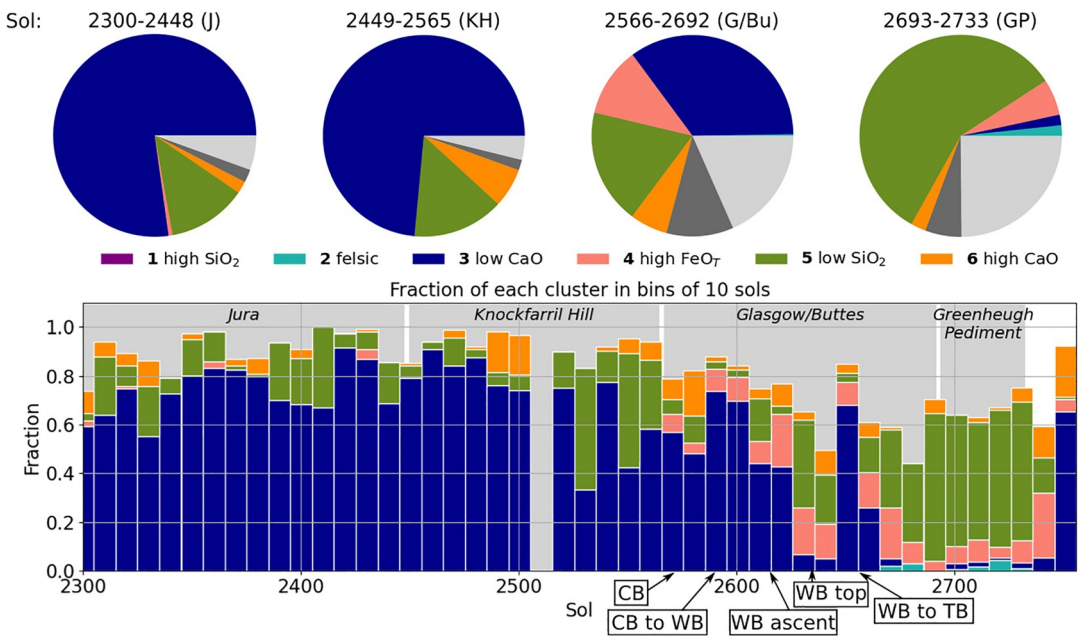

Figure 16. Same type of plot as in Figure 11, here: Beginning of Glen Torridon. Because transitions between the Glasgow member, the buttes but also still the Knockfarril Hill member (KH) occurred in a short time scale during the sols 2570–2690 period, this entire time was designated as Glasgow/Buttes (G/Bu). Nevertheless, it has to be noted that the parts of central butte, which could be sampled with ChemCam, were classified as Knockfarril Hill member. The buttes and the traverses in between them are annotated where CB: Central butte, WB: Western butte, and TB: Tower butte. The gray fractions in the pie diagrams represent rejected observations. The dark gray corresponds to rejections due to inconsistencies while the light gray were rejected according to the silhouette score criterion.

to the Murray formation (Morris et al., 2016), which are suggested to be related to the high-silica Stimson halos (Yen et al., 2021). Almost all points of Buckskin on the surface prior to drilling but also in the hole and the tailings are assigned to cluster 1, emphasizing the highly elevated SiO_2 at this location.

4.3. Vera Rubin Ridge

On sol 1809 of the mission, the rover reached the VRR, a prominent ridge for which a particular strong hematite spectral signature was observed from orbit (Milliken et al., 2010). The cluster detections during the time of the VRR exploration (section D in Figure 10) are shown in Figure 15. Stratigraphically, the ridge belongs to the Murray formation and is divided into two members, the Pettegrove Point and the Jura member (Fraeman et al., 2020). The rocks of the latter appear in gray and red colors resulting in a differentiation between gray and red Jura. With the dedicated FeO_T ChemCam calibration for high FeO_T targets (David et al., 2020), it was found that gray Jura targets have on average higher FeO_T abundances than Pettegrove Point and red Jura targets. Furthermore, nearly pure iron oxides were identified with ChemCam in dark-toned diagenetic features (David et al., 2020; L'Haridon et al., 2020). However, these spectra were not included in the cluster analysis due to the anomalously high FeO_T abundances. The rover traverse on the VRR did not proceed directly over the ridge, but rather the member boundaries were crossed several times to address science goals, including a short return to the Blunts Point member (Fraeman et al., 2020).

Based on the cluster analysis, see Figure 15, the transitions between the different members of the VRR can be identified, demonstrating that they are also chemically distinct: The first transition from the Pettegrove Point member (sol 1810–1880) to the Jura member (sol 1880–2000) can be seen by means of higher proportions of cluster 4 detections in the Jura member. The transition back from Jura to Pettegrove Point (not shown as a pie chart) becomes apparent in the clustering somewhat later than the stratigraphy suggests as the detection of cluster 4 only declines after sol 2020. The next transition from Pettegrove Point to Blunts Point is clearly observable due to an increased proportion of cluster 6 observations, which make up the major part in the pie diagram (sol 2045–2092). Those clearly decrease again when the rover returned to the Pettegrove Point after sol 2100. Finally, the second transition from Pettegrove Point to Jura is not as distinct as the first one, but can still be observed based on higher cluster 4 proportions starting already somewhat earlier than sol 2120 when Curiosity crossed the stratigraphic boundary. The rover left the VRR on sol 2300 and started to explore the clay-bearing unit Glen Torridon. The clustering analysis in Figure 15 on the other hand suggests a change in chemistry already around sol 2250. From the pie diagrams, a difference between the Pettegrove Point and the Jura member is only noticeable at the beginning of VRR exploration where the Jura member is characterized by a larger cluster 4 contribution than the Pettegrove Point member. In contrast, the latter pie charts of Pettegrove Point (sol 2093–2156) and Jura (sol 2157–2300) have rather similar distributions. Even though the average bedrock FeO_T concentration of both VRR members is similar to that of Mt. Sharp members at lower elevations, the higher detection frequency of cluster 4 (high FeO_T) at VRR is in agreement with a higher variability of FeO_T abundances observed with ChemCam, especially in the Jura member but also in the upper part of the Pettegrove Point member (David et al., 2020). The latter can explain the higher cluster 4 detections in some Pettegrove Point locations, especially close to the boundaries to the Jura member.

CheMin results from three VRR drill holes, one in the Pettegrove Point member (*Stoer*), one in gray (*Highfield*), and one in red Jura (*Rock Hall*), show that hematite is more abundant in the Pettegrove Point (14.7 wt% in the bulk) and gray Jura (8.5 wt% in the bulk) than in the red Jura drill target (2.9 wt% in the bulk) (Rampe, Bristow, et al., 2020). However, the *Rock Hall* drill sample contains the Fe-bearing minerals akaganeite and jarosite in higher abundances (Rampe, Bristow, et al., 2020). Thus, minerals containing Fe are present in all three areas. The clustering analysis suggests that points with high FeO_T abundances occur more frequently in the Jura member, see Figure 15. In contrast, CheMin measured the highest abundances of Fe-bearing minerals in the crystalline phase of the *Stoer* target from the Pettegrove Point member (Rampe, Bristow, et al., 2020). But since the drill campaign of *Stoer* (sol 2132–2155) was done at an elevation close to the boundary to the Jura member (Fraeman et al., 2020), thus in an area with a larger variability of FeO_T abundances (David et al., 2020), the clustering results are still in agreement with CheMin observations. Indeed, the bars corresponding to the *Stoer* drill campaign in Figure 15 show contributions from cluster 4

observations. The two drill campaigns for the Jura targets *Highfield* and *Rock Hall* took place in the second Jura period during sols 2223–2249 and 2258–2296, respectively. The bars of the *Highfield* campaign have, similar to the Pettegrove Point target, cluster 4 contributions, whereas the bars of the *Rock Hall* campaign show only rare cluster 4 detections. The latter is the drill target among the VRR drills with the lowest hematite abundance, which could explain the few cluster 4 detections even though other Fe-bearing minerals were measured with CheMin (Rampe, Bristow, et al., 2020).

4.4. Beginning Glen Torridon

On sol 2302, Curiosity left the VRR and started the exploration of the Glen Torridon region, which was formerly known as the clay-bearing unit due to a strong spectral signal of clay minerals observed from orbit (Milliken et al., 2010). As VRR, Glen Torridon is part of the Murray formation and the Jura member from the VRR further continues into the Glen Torridon region. The cluster detections within Glen Torridon are shown in Figure 16 until sol 2756 of the mission corresponding to section E in Figure 10. During this time, three different members of the Murray formation were visited by the rover: the Jura, the Knockfarril Hill, and the Glasgow member (also previously known as fractured intermediate unit) (Fox et al., 2021). Within the latter, three buttes, the Central, Western, and Tower butte were on the rover's route. Furthermore, Curiosity climbed onto the Greenheugh pediment, which is capped by rocks belonging to the Stimson formation previously investigated at the Emerson and Naukluft plateaus.

The Jura member belonging to Glen Torridon is different from the Jura observed at VRR in terms of absent cluster 4 detections. However, as discussed for the VRR, the cluster 4 detections are mostly related to the gray Jura. Thus, the red VRR Jura and the Glen Torridon Jura are similar in terms of cluster detections, which are dominated by the low CaO cluster 3. However, the prevalence of cluster 3 is more distinct in the Glen Torridon Jura and also Knockfarril Hill member as well. Furthermore, there are less rejections in both regions resulting in pie diagrams with around 75% cluster 3 proportions (sol 2300–2448 and sol 2449–2565). This agrees with a regional comparison of CheMin data from VRR and Glen Torridon drill targets, which indicate lower hematite and higher phyllosilicate abundances in Glen Torridon, the latter being the highest observed so far (Thorpe et al., 2021). Around sol 2400, slightly higher fractions of cluster 5 observation points appear, which correspond to targets measured in the Rigg ripple field and close to it in the area of the *Kilmarie* and *Aberlady* drill holes. This relation to soil and loose material measurements suggest that cluster 5 detections in this area correspond primarily to these types of targets rather than to consolidated targets as seen for the Bradbury and Stimson formation. No distinct difference in cluster detections shows up at the boundary between the Jura and the Knockfarril Hill member suggesting that, seen from the cluster analysis, Glen Torridon is rather homogeneous in composition. However, there is a difference in chemistry between two types of facies, the rubbly and the coherent bedrock, which is not observable in the clustering results. The coherent bedrock has higher MgO (6–10 wt%) and lower K_2O abundances, while the rubbly bedrock is enriched in K_2O (>1.5 wt%) and SiO_2 (Cousin et al., 2020, 2021; Dehouck et al., 2020). A possible explanation why both types of rocks fall mostly into cluster 3 is that rubbly and coherent bedrock in Glen Torridon both have a similar low CaO concentration. Even though cluster 5 is on average higher in MgO than cluster 3, there are also cluster 3 observations with even higher MgO in the range of 8–12 wt%. Thus, the CaO content is more discriminant than MgO and K_2O concentrations.

A change in composition becomes apparent when the rover arrived at the buttes within the Glasgow member. A detailed distinction in pie diagrams between Glasgow member and buttes material is challenging and we decided to summarize this region as Glasgow/Buttes (sol 2566–2692) even though parts of Central butte are classified as Knockfarril Hill member. The summarized pie chart shows an increase of cluster 4, cluster 5, and ambiguous points as well as less cluster 3 contributions compared to the previous Glen Torridon members. Nevertheless, some differences between Knockfarril Hill member, Glasgow member, and butte points become apparent in the bar plots representation where bars corresponding to the buttes or the transitions between them have additional annotations (Figure 16). The first butte, Central Butte, has the smallest height of the three buttes and its top part is stratigraphically assigned to the Glasgow member. However, the lower part, which was reached by the rover and investigated by ChemCam is classified as the Knockfarril Hill member. A slight decrease in cluster 3 detections and an increase in rejections suggest a difference in composition. During the drive to Western Butte (sol 2590–2610), cluster 3 observations increase again

and also on the butte's base (sol 2610–2620) the fraction of cluster 3 detections is prevalent, but it clearly decreases when rocks during the ascent and the top of the butte were measured. While cluster 3 detections decrease, cluster 5 as well as cluster 4 detections increase. The compositional difference between the higher elevated Western Butte material and the underlying Glasgow member shows up again during the descent of the butte and the drive toward the Greenheugh pediment (sol 2643–2660) by means of dominant cluster 3 detections. This is in general also the reason why the Glasgow/Buttes pie diagram is different from the previous Glen Torridon pie diagrams. While the Glasgow base member is dominated by the low CaO cluster 3 similar as the Jura and the Knockfarril Hill member, the material of the buttes is characterized by the low SiO_2 cluster 5. The Tower Butte area is not distinguishable based on the clustering from the Greenheugh pediment capping unit and the area below the pediment. The rover explored the contact, the Siccar Point unconformity, between the Murray and the Stimson formation and the area below until sol 2693 when it climbed on top of the pediment. Already in the area below the unconformity, the number of rejected observations as well as cluster 5 observations increase as seen for the Western butte ascent. Both areas are also characterized by multiple fluorine measurements with the ChemCam instrument (Forni et al., 2021). On top of the pediment, cluster 5 detections clearly dominate the clustering results. Even though the Greenheugh Pediment as well as the Emerson and the Naukluft Plateau belong to the Stimson formation, small deviations regarding the composition were observed with ChemCam indicating the recycling of local material and mineral sorting across the ancient Stimson dune field (Bedford et al., 2021). Furthermore, the aforementioned F detections stop on top of the pediment, they are only present in the area below the contact (Forni et al., 2021). Such changing abundances of minor and trace elements like F with only a few emission signatures in LIBS spectra and small compositional deviations between the different Stimson areas do not show up in the cluster analysis.

4.5. Implications From Low-SiO_2 Cluster Detections Along Traverse

From the previous sections discussing the cluster detections along the traverse, it is worth noting that the low SiO_2 cluster 5 is the dominant cluster of different members and types of targets: The Bradbury sandstones (Figure 11), the rocks of the Stimson formation (Figures 12 and 16), and loose materials like soils are assigned to this cluster. This suggests a certain similarity based on chemistry and raises the question of the reason for these similarities. Previous research has shown that the composition of the Bradbury group low-silica sediments and the Stimson formation are compositionally similar (Bedford et al., 2020). The compositional similarity between the Stimson formation and low-silica rocks of the Bradbury group were hypothesized to relate to both units sharing a similar subalkaline basaltic source. Alternatively, these compositional similarities could indicate that the ancient aeolian dune field recycled and preferentially transported the sand grains of the Bradbury group to where the Stimson sandstone is now preserved relative to the fine grains of the Murray formation (Bedford et al., 2020). The modern Bagnold dunes and soils also likely have a basaltic sediment source similar to the Stimson formation and Bradbury group; however, minor geochemical differences exist suggesting that the basaltic sources are unlikely to be the same (Bedford et al., 2020). These differences are not large enough for our large-scale clustering approach to differentiate between them; as such, all sediments relating to a low-silica basaltic sediment source are placed into cluster 5. In these processes, the grain size plays an important role as coarse grains are more likely to be reworked than fine-grained material, which is easily blown away by the wind. Consequently, the Stimson formation is similar to Bradbury sandstones and distinct from Murray formation mudstones. This interpretation is in agreement with previous works on the provenance of soils and their sorting based on grain sizes (Cousin et al., 2015).

5. Predictions With Random Forest Classifier

This section explores the use of our clustering results for training machine learning classifiers and how this can be used to predict cluster memberships of new targets. The aim is to demonstrate how the clustering can be used rather than to present a detailed study on training, validating, and comparing machine learning models. Here, we focus on the model-free and nonlinear classifier random forest, which is usually built of multiple de-correlated decision tree classifiers (Breiman, 2001). Those make predictions on random sub-data sets and their averages are used to make the final predictions. Such a combination of weak learners

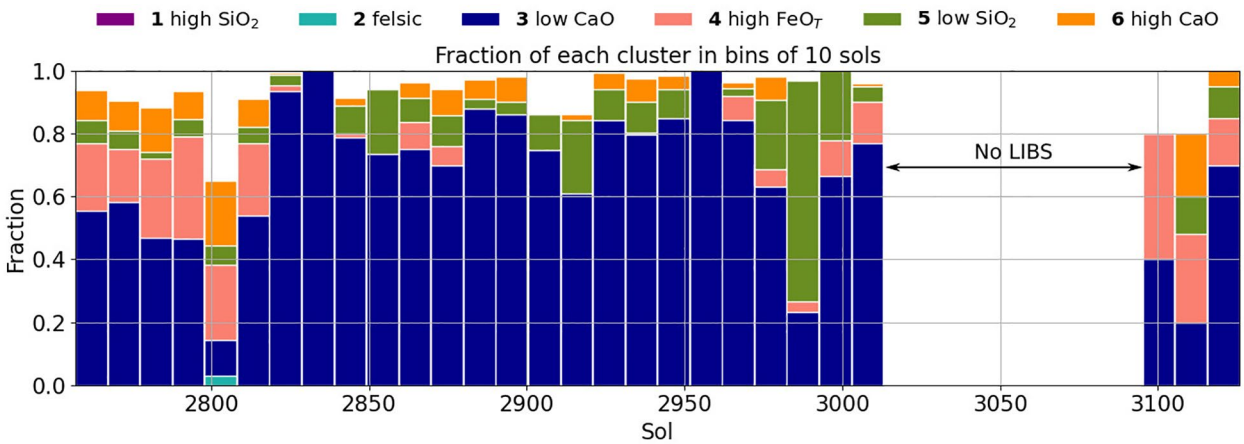

Figure 17. Same type of plot as in Figure 11 but with cluster membership predictions from a random forest classifier, which was trained and validated with clustering results. Predictions that have a probability of less than 0.7 for belonging to the predicted cluster were rejected, which is the reason why not all bars sum up to one.

following the principle of voting can improve the accuracy of predictions and is less prone to overfitting than single classifiers, a common machine learning principle.

We used the random forest classifier class provided by python's *scikitlearn* library and set up a model with a maximum depth of 10 and with 50 estimators, meaning 50 de-correlated decision trees. With the *train_test_split* function, the 14,715 observations with a final cluster assignment were split into training and test data in an 80:20 ratio. Thus, the NMF scores and the corresponding cluster assignments of 11,772 spectra were used to train the random forest. Next, predictions with the classifier were made on the NMF scores of 2,943 spectra in the test data set. This process was repeated 50 times with randomly changing the selection of training and test data but keeping their sizes constant. The lowest accuracy for the predictions on the test data was 0.998 among the 50 repetitions and in 14 runs an accuracy of 1 was obtained. Such high accuracies allow the trained random forest models to be used for predictions on new data unknown to the models.

We apply one of the random forest classifiers with an accuracy of 1 to predict cluster memberships of 2,846 ChemCam spectra measured within sols 2757–3126 of the mission, that is the most recent period that is not included in the cluster analysis described so far in this paper. Prior to the model application, the spectra were transformed with the fixed NMF model, see Section 2.4.1. In the trained random forest, each tree returns a decision about the cluster membership of an observation and the cluster with the most votes is then the final prediction. In this way, the random forest gives the opportunity to return probabilities for belonging to each of the possible classes. Such a probability is the number of decision trees voted for the cluster divided by the total number of trees, which is 50 in our random forest. Regarding just the probabilities for the actual assigned cluster of each observation, we find mean and median probabilities of 0.94 and 0.999, respectively. These values indicate an agreement of the single trees for most of the observations, however, there are also predictions with probabilities of belonging to its assigned cluster smaller than 0.5. We decided therefore to reject predictions with probabilities smaller than 0.7 and with this limit, 93% of the predictions were kept. Interestingly, for some predictions, linear trends between the probability of belonging to the assigned cluster and the probability of belonging to one other cluster could be observed. In other words, for some observations the decision trees in the random forest disagreed between two clusters, which could indicate the presence of mixtures between endmembers from two clusters. The outcomes of the non-rejected predictions are shown in Figure 17 in the same type of bar plot as the clustering results. The period corresponds to section F in Figure 10 and is also discussed in Dehouck et al. (2021a, 2021b) in more detail. Due to a technical issue with the ChemCam laser, no LIBS measurements were possible in the time period of sols 3008–3106, which appears as a gap in Figure 10. The traverse includes a return from the Glasgow member to the Knockfarril Hill member where the *Mary Anning* drill site was investigated for about 90 sols (sol 2830–2920). The transition back to Knockfarril Hill member appears in the clustering results around sol 2820 due to less cluster 4 and low probability observations. This change in composition

is in agreement with ChemCam results reported in Dehouck et al. (2021a) supporting the predictions from the random forest model. Also the rover's stop at the dune field Sands of Forvie around sol 2990 can be identified in the predicted cluster memberships as the fraction of cluster 5 assignments is clearly increased. On sol 3107, LIBS analyses resumed on top of a cliff called Mont Mercou, which is located ≈600 m to the east and ≈19 m higher in comparison to the rover's location when the technical issue occurred on sol 3007. The bedrock composition at Mont Mercou is surprisingly similar to that measured before the pause in LIBS measurements (Dehouck et al., 2021b). Since the present study also considers ChemCam observations of diagenetic features and veins besides the bedrock, the results cannot be compared directly. However, the apparent increase of cluster 4 (high FeO_T) is in accordance with high FeO_T observations but can also be explained by the sampling of diagenetic features. Cluster 3 is, moreover, still the dominant cluster indicating that the chemistry at Mont Mercou is similar to that of the Glasgow member.

6. Conclusion

This study highlights the broad compositional variations seen by ChemCam between the stratigraphic units traversed by the Curiosity rover as summarized in the pie diagrams. This work can therefore support the effort to map the chemostratigraphy of the Gale crater. But it is also striking to see the variability when the data are displayed by groups of 10 sols in the bar plots, which might be an expression of the various degrees of mixing of the sedimentary sources. Even though the six clusters are able to display distinct changes in composition, the approach is not sensitive to varying abundances of minor and trace elements, which can be crucial for understanding environmental changes. In addition, we cannot discount the potential impact of bias in target selection since most ChemCam targets were selected according to scientific and operational priorities of the time.

Regarding the approach, six clusters with distinct compositions were identified based only on LIBS data as no empirical model for quantification of elemental abundances were employed. The MOCs were solely used for interpretations and comparison with the obtained NMF factors. However, with this approach, the distinction between the clusters can be more driven by elements like Ca, which have several strong emission lines in ChemCam LIBS spectra, while elements like K with only two emission lines have a smaller impact. In particular, this becomes apparent for the Kimberley region where high K_2O abundances were measured with ChemCam (Le Deit et al., 2016), but most of these points were assigned to cluster 5, which has on average low K_2O concentration compared to the other clusters (see Table 1). Such effects should always be kept in mind when making interpretations and this possible bias could be investigated in future studies.

Our cluster results largely highlight the geochemical diversity of the sediment source rocks and main alteration features. Clusters 1 and 6 relate to the silica-rich halo and calcium-sulfate alteration features, which are geochemically distinct to average bedrock compositions. Meanwhile, cluster 2 correlates to the presence of feldspar and the geological units previously interpreted to be derived from a trachybasaltic source (Bedford et al., 2019; Edwards et al., 2017). Clusters 3 and 5 both relate to basaltic sediment sources, with the iron-rich cluster 4 relating to the abundant iron-oxide diagenetic cement, which likely formed through the alteration of basaltic materials. Cluster 3 correlates to the sediments of the Murray formation hypothesized to be from a distinct silica-rich basaltic source that likely underwent some chemical weathering, and cluster 5 relates to the isochemically altered or relatively unaltered basaltic sediments of the Bradbury group, Stimson formation, and soils (Bedford et al., 2019, 2020).

The random forest classifier trained with ChemCam data measured until sol 2756 of the mission gives promising results for predicting cluster memberships of new data. The validation process returned overall high accuracies, which might be due to the exclusion of ambiguous points identified by inconsistency among the 100 clustering repetitions or bad silhouette scores. For the predictions on the data, which was not used in the clustering and the training of the random forest, the probabilities for belonging to the assigned cluster served as a useful criterion for removing uncertain predictions. For some predictions with low probabilities, we observed that the decision trees in the forest disagreed between only two of the six clusters. This indicates a certain degree of mixing between end-members of the particular clusters. All these characteristics make the classifier a useful support for the rapid interpretation of new ChemCam LIBS spectra through the identification of cluster memberships and thus their association with specific geochemistry. Even if the

chemistry changes in such a way that the targets are not well described by any of the clusters anymore, this will be noted based on the probabilities given for each observation. This might become important at higher elevations when the transition to the sulfate-bearing unit takes place.

For this study, averaged LIBS spectra from multiple shots were used, which means that variability or trends with depth at one point are not considered. Tracing such trends or variabilities can reveal correlations between elements indicating particular mineral phases or support the identification of soils. The latter are characterized by a larger variability among the shots than consolidated rock targets (Cousin et al., 2015), which could give, for example, the possibility to distinguish between rock and soil targets in the well-populated low SiO_2 cluster 5. Including shot to shot data in this type of analysis could also support the identification of smaller subclusters, especially in the major clusters we identified. Furthermore, the influence of the preprocessing on the results could be studied by using raw data. The combination with machine learning models for preprocessing (Castorena et al., 2021; Ewusi-Annan et al., 2020) is also conceivable. Another aspect, which was not discussed in detail here is the investigation of the rejected observation points. As one objective of this study was to use the clustering results for the training of classifiers such as the random forest classifier, we wanted the clusters to be distinct from each other, which motivated the rejections due to the silhouette score criterion. However, as seen for example for the Pahrump Hills regions, rejections can indicate compositions that are not well described by the six clusters, but nevertheless these points still hold important geochemical information. Therefore, a dedicated study on the rejections could be part of a future follow-up work of this classification study.

The developed method may further be relevant for other LIBS instruments on robotic missions. The results show that regular LIBS measurements make it possible to document geochemical transitions on a large scale. Recently, two rovers safely arrived on the surface of Mars, which are equipped with LIBS instruments. These are the Zhurong rover from the Chinese Tianwen-1 mission (Mallapaty, 2021) and the Perseverance rover from NASA's Mars 2020 mission whose payload includes the SuperCam instrument (Maurice et al., 2021; Wiens et al., 2021). The latter is the successor of ChemCam and already returned LIBS data from the Jezero crater (Anderson et al., 2021) at the time of writing this manuscript.

Data Availability Statement

All the ChemCam data used in this paper are released and can be found on the Planetary Data System (https://pds-geosciences.wustl.edu/missions/msl/chemcam.htm).

Acknowledgments
Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This work was enabled by a post-doc grant by the Centre National d'Etudes Spatiales (CNES), which also supports the French operations of the ChemCam instrument. The authors are grateful to the MSL science and engineering teams who operate the rover.

References

Anderson, R., & Bell, J. F., III. (2010). Geologic mapping and characterization of Gale Crater and implications for its potential as a Mars Science Laboratory landing site. *The Mars Journal*, *5*, 76–128. https://doi.org/10.1555/mars.2010.0004

Anderson, R., Lasue, J., Wiens, R. C., Clegg, S. M., Lanza, N. L., Ehlmann, B. L., & Team, M. S. (2013). Spectral classification and variability in ChemCam data from Bradbury landing to Rocknest. In *44th lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2013/pdf/2750.pdf

Anderson, R. B., Forni, O., Clegg, S. M., Cousin, A., Frydenvang, J., Pilleri, P., Legett, C., et al. (2021). SuperCam laser induced breakdown spectroscopy (LIBS) data processing, calibration, and first results. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/1606.pdf

Anderson, R. B., Morris, R. V., Clegg, S. M., Bell, J. F., Wiens, R. C., Humphries, S. D., & McInroy, R. (2011). The influence of multivariate analysis methods and target grain size on the accuracy of remote quantitative chemical analysis of rocks using laser induced breakdown spectroscopy. *Icarus*, *215*(2), 608–627. https://doi.org/10.1016/j.icarus.2011.07.034

Banham, S. G., Gupta, S., Rubin, D. M., Watkins, J. A., Sumner, D. Y., Edgett, K. S., & Vasavada, A. R. (2018). Ancient Martian aeolian processes and palaeomorphology reconstructed from the Stimson formation on the lower slope of Aeolis Mons, Gale crater, Mars. *Sedimentology*, *65*(4), 993–1042. https://doi.org/10.1111/sed.12469

Bedford, C., Bridges, J., Schwenzer, S., Wiens, R., Rampe, E., Frydenvang, J., & Gasda, P. (2019). Alteration trends and geochemical source region characteristics preserved in the fluviolacustrine sedimentary record of Gale crater, Mars. *Geochimica et Cosmochimica Acta*, *246*, 234–266. https://doi.org/10.1016/j.gca.2018.11.031

Bedford, C., Schwenzer, S., Bridges, J., Banham, S., Wiens, R., Gasnault, O., & Gasda, P. (2020). Geochemical variation in the Stimson formation of Gale crater: Provenance, mineral sorting, and a comparison with modern Martian dunes. *Icarus*, *341*, 113622. https://doi.org/10.1016/j.icarus.2020.113622

Bedford, C. C., Banham, S., Bowden, D., Bridges, J. C., Smith, R., Forni, O., Cousin, A., et al. (2021). Identifying ancient dune processes in the Stimson formation of Gale crater using geochemical data from ChemCam: New insights from the Greenheugh capping unit. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/1569.pdf

Blaney, D. L., Wiens, R. C., Maurice, S., Clegg, S. M., Anderson, R. B., Kah, L. C., & Madsen, M. B. (2014). Chemistry and texture of the rocks at Rocknest, Gale Crater: Evidence for sedimentary origin and diagenetic alteration. *Journal of Geophysical Research: Planets*, *119*(9), 2109–2131. https://doi.org/10.1002/2013JE004590

Boucher, T. F., Ozanne, M. V., Carmosino, M. L., Dyar, M. D., Mahadevan, S., Breves, E. A., & Clegg, S. M. (2015). A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *107*, 1–10. https://doi.org/10.1016/j.sab.2015.02.003

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Bridges, N. T., & Ehlmann, B. L. (2018). The Mars Science Laboratory (MSL) Bagnold Dunes campaign, phase I: Overview and introduction to the special issue. *Journal of Geophysical Research: Planets*, *123*(1), 3–19. https://doi.org/10.1002/2017JE005401

Bristow, T. F., Rampe, E. B., Achilles, C. N., Blake, D. F., Chipera, S. J., Craig, P., et al. (2018). Clay mineral diversity and abundance in sedimentary rocks of Gale crater, Mars. *Science Advances*, *4*(6). https://doi.org/10.1126/sciadv.aar3330

Cao, X., Zhang, L., Wu, Z., Ling, Z., Li, J., & Guo, K. (2020). Quantitative analysis modeling for the ChemCam spectral data based on laser-induced breakdown spectroscopy using convolutional neural network. *Plasma Science and Technology*, *22*(11), 115502. https://doi.org/10.1088/2058-6272/aba5f6

Castorena, J., Oyen, D., Ollila, A., Legget, C., & Lanza, N. (2021). Deep spectral CNN for laser induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *178*, 106125. https://doi.org/10.1016/j.sab.2021.106125

Certini, G., Karunatillake, S., Zhao, Y.-Y. S., Meslin, P.-Y., Cousin, A., Hood, D. R., & Scalenghe, R. (2020). Disambiguating the soils of Mars. *Planetary and Space Science*, *186*, 104922. https://doi.org/10.1016/j.pss.2020.104922

Cichocki, A., & Phan, A.-H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor Factorizations. In *IEICE transactions on fundamentals of electronics, communications and computer sciences*. *E92-A*(3), 708–721. https://doi.org/10.1587/transfun.E92.A.708

Clegg, S. M., Wiens, R. C., Anderson, R., Forni, O., Frydenvang, J., Lasue, J., et al. (2017). Recalibration of the Mars Science Laboratory ChemCam instrument with an expanded geochemical database. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *129*, 64–85. https://doi.org/10.1016/j.sab.2016.12.003

Cousin, A., Dehouck, E., Forni, O., David, G., Berger, G., Meslin, P. Y., et al. (2020). Potassium enrichments and minor element abundances measured by ChemCam at Glen Torridon, Gale crater, Mars. In *51st lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2020/pdf/2148.pdf

Cousin, A., Dehouck, E., Meslin, P.-Y., Forni, O., Williams, A. J., Stein, N., et al. (2017). Geochemistry of the Bagnold dune field as observed by ChemCam and comparison with other aeolian deposits at Gale crater. *Journal of Geophysical Research: Planets*, *122*(10), 2144–2162. https://doi.org/10.1002/2017JE005261

Cousin, A., Desjardins, M., Dehouck, E., Forni, O., David, G., Berger, G., et al. (2021). K-rich rubbly bedrock at Glen Torridon, Gale crater, Mars: Investigating the possible presence of illite. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/2127.pdf

Cousin, A., Meslin, P., Wiens, R., Rapin, W., Mangold, N., Fabre, C., et al. (2015). Compositions of coarse and fine particles in martian soils at Gale: A window into the production of soils. *Icarus*, *249*, 22–42. https://doi.org/10.1016/j.icarus.2014.04.052

Cousin, A., Sautter, V., Payré, V., Forni, O., Mangold, N., Gasnault, O., et al. (2017). Classification of igneous rocks analyzed by ChemCam at Gale crater, Mars. *Icarus*, *288*, 265–283. https://doi.org/10.1016/j.icarus.2017.01.014

David, G., Cousin, A., Forni, O., Meslin, P., Dehouck, E., Mangold, N., et al. (2020). Analyses of high-iron sedimentary bedrock and diagenetic features observed with ChemCam at Vera Rubin ridge, Gale crater, Mars: Calibration and characterization. *Journal of Geophysical Research: Planets*, *125*(10), 1–26. https://doi.org/10.1029/2019JE006314

Dehouck, E., Cousin, A., Mangold, N., Frydenvang, J., Gasnault, O., David, G., et al. (2020). Is the clay-bearing unit distinct in Gale crater? Geochemical diversity and open-system alteration revealed by MSL and ChemCam at Glen Torridon, Mars. In *51st lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2020/pdf/2770.pdf

Dehouck, E., Cousin, A., Mangold, N., Frydenvang, J., Gasnault, O., Rammelkamp, K., et al. (2021a). Bedrock geochemistry measured by ChemCam along a 2-km eastward traverse in the Glen Torridon region, Gale crater, Mars. In *Europlanet science congress*.

Dehouck, E., Cousin, A., Mangold, N., Frydenvang, J., Gasnault, O., Rapin, W., et al. (2021b). Leaving Glen Torridon: Bedrock geochemistry measured by ChemCam en route to the sulfate unit of Gale crater. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/1858.pdf

Edwards, P. H., Bridges, J. C., Wiens, R., Anderson, R., Dyar, D., Fisk, M., et al. (2017). Basalt-trachybasalt samples in Gale crater, Mars. *Meteoritics & Planetary Sciences*, *52*(11), 2391–2410. https://doi.org/10.1111/maps.12953

Ewusi-Annan, E., Delapp, D. M., Wiens, R. C., & Melikechi, N. (2020). Automatic preprocessing of laser-induced breakdown spectra using partial least squares regression and feed-forward artificial neural network: Applications to Earth and Mars data. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *171*, 105930. https://doi.org/10.1016/j.sab.2020.105930

Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Computation*, *23*(9), 2421–2456.

Forni, O., Dehouck, E., Cousin, A., Bedford, C. C., David, G., Schwenzer, S. P., et al. (2021). Elevated fluorine abundances below the Siccar Point unconformity: Implications for fluid circulation in Gale crater. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/1503.pdf

Forni, O., Gaft, M., Toplis, M. J., Clegg, S. M., Maurice, S., Wiens, R. C., et al. (2015). First detection of fluorine on Mars: Implications for Gale crater's geochemistry. *Geophysical Research Letters*, *42*(4), 1020–1028. https://doi.org/10.1002/2014GL062742

Forni, O., Gasnault, O., Cousin, A., Anderson, R. B., Dehouck, E., David, G., et al. (2019). Machine learning applied to MSL/ChemCam data. In *50th lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2019/pdf/1404.pdf

Forni, O., Meslin, P.-Y., Drouet, C., Cousin, A., David, G., Mangold, N., et al. (2020). Apatites in Gale crater. In *51st lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2020/pdf/2192.pdf

Fox, V. K., Bennett, K. A., Bryk, A. B., Arvidson, R. E., Fedo, C., & Dehouck, E. (2021). Contextualizing CRISM observations of the clay-bearing Glen Torridon region with the Mars Science Laboratory Curiosity rover. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/2765.pdf

Fraeman, A. A., Edgar, L. A., Rampe, E. B., Thompson, L. M., Frydenvang, J., Fedo, C. M., et al. (2020). Evidence for a diagenetic origin of Vera Rubin ridge, Gale crater, Mars: Summary and synthesis of Curiosity's exploration campaign. *Journal of Geophysical Research: Planets*, *125*(12), 1–34. https://doi.org/10.1029/2020JE006527

Fraeman, A. A., Ehlmann, B. L., Arvidson, R. E., Edwards, C. S., Grotzinger, J. P., Milliken, R. E., et al. (2016). The stratigraphy and evolution of lower Mount Sharp from spectral, morphological, and thermophysical orbital data sets. *Journal of Geophysical Research: Planets*, *121*(9), 1713–1736. https://doi.org/10.1002/2016JE005095

Francis, R., Estlin, T., Doran, G., Johnstone, S., Gaines, D., Verma, V., et al. (2017). AEGIS autonomous targeting for ChemCam on Mars Science Laboratory: Deployment and results of initial science team use. *Science Robotics*, *2*(7), eaan4582. https://doi.org/10.1126/scirobotics.aan4582

Frydenvang, J., Gasda, P. J., Hurowitz, J. A., Grotzinger, J. P., Wiens, R. C., Newsom, H. E., et al. (2017). Diagenetic silica enrichment and late-stage groundwater activity in Gale crater, Mars. *Geophysical Research Letters*, *44*(10), 4716–4724. https://doi.org/10.1002/2017GL073323

Frydenvang, J., Mangold, N., Wiens, R. C., Fraeman, A. A., Edgar, L. A., Fedo, C. M., et al. (2020). The chemostratigraphy of the Murray formation and role of diagenesis at Vera Rubin ridge in Gale crater, Mars, as observed by the ChemCam instrument. *Journal of Geophysical Research: Planets*, *125*(9), 1–21. https://doi.org/10.1029/2019JE006320

Gasnault, O., Anderson, R. B., Forni, O., Maurice, S., Wiens, R. C., & Pinet, P. (2014). Classification of ChemCam targets along the traverse of Curiosity at Gale. In *Eighth international conference on mars*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/8thmars2014/pdf/1269.pdf

Gasnault, O., Forni, O., Mangold, N., Wiens, R., Meslin, P.-Y., Lasue, J., et al. (2015). Updated perspective on ChemCam data through clustering. In *46th lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2015/pdf/2789.pdf

Gasnault, O., Forni, O., Meslin, P.-Y., Maurice, S., Wiens, R. C., Anderson, R. B., et al. (2013). ChemCam target classification: Who's who from Curiosity's first ninety sols. In *44th lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2013/pdf/1994.pdf

Gasnault, O., Pinet, P., Wiens, R., Dehouck, E., Gasda, P., Forni, O., et al. (2019). Targeting and classifying drill holes on Mars with ChemCam. In *Ninth international conference on mars*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/ninthmars2019/pdf/6199.pdf

Grotzinger, J. P., Crisp, J., Vasavada, A. R., Anderson, R. C., Baker, C. J., Barry, R., et al. (2012). Mars Science Laboratory mission and science investigation. *Space Science Reviews*, *170*(1–4), 5–56. https://doi.org/10.1007/s11214-012-9892-2

Grotzinger, J. P., Gupta, S., Malin, M. C., Rubin, D. M., Schieber, J., Siebach, K., et al. (2015). Deposition, exhumation, and paleoclimate of an ancient lake deposit, Gale crater, Mars. *Science*, *350*(6257), aac7575. https://doi.org/10.1126/science.aac7575

Grotzinger, J. P., Sumner, D. Y., Kah, L. C., Stack, K., Gupta, S., Edgar, L., et al. (2014). A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale crater, Mars. *Science*, *343*(6169), 1242777. https://doi.org/10.1126/science.1242777

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Johnson, J. R., Bell, J., Bender, S., Blaney, D., Cloutis, E., DeFlores, L., et al. (2015). ChemCam passive reflectance spectroscopy of surface materials at the Curiosity landing site, Mars. *Icarus*, *249*, 74–92. https://doi.org/10.1016/j.icarus.2014.02.028

Lapotre, M. G. A., & Rampe, E. B. (2018). Curiosity's investigation of the Bagnold Dunes, Gale crater: Overview of the two-phase scientific campaign and introduction to the special collection. *Geophysical Research Letters*, *45*(19), 10200–10210. https://doi.org/10.1029/2018GL079032

Lasue, J., Wiens, R. C., Stepinski, T. F., Forni, O., Clegg, S. M., & Maurice, S. (2011). Nonlinear mapping technique for data visualization and clustering assessment of LIBS data: Application to ChemCam data. *Analytical and Bioanalytical Chemistry*, *400*(10), 3247–3260. https://doi.org/10.1007/s00216-011-4747-3

Le Deit, L., Mangold, N., Forni, O., Cousin, A., Lasue, J., Schröder, S., et al. (2016). The potassic sedimentary rocks in Gale crater, Mars, as seen by ChemCam on board Curiosity. *Journal of Geophysical Research: Planets*, *121*(5), 784–804. https://doi.org/10.1002/2015JE004987

Le Mouélic, S., Gasnault, O., Herkenhoff, K., Bridges, N., Langevin, Y., Mangold, N., et al. (2015). The ChemCam remote micro-imager at Gale crater: Review of the first year of operations on Mars. *Icarus*, *249*, 93–107. https://doi.org/10.1016/j.icarus.2014.05.030

L'Haridon, J., Mangold, N., Fraeman, A. A., Johnson, J. R., Cousin, A., Rapin, W., et al. (2020). Iron mobility during diagenesis at Vera Rubin ridge, Gale crater, Mars. *Journal of Geophysical Research: Planets*, *125*(11), 1–24. https://doi.org/10.1029/2019JE006299

Mallapaty, S. (2021). What China's Mars rover will do next. *Nature*, *593*.

Mangold, N., Dehouck, E., Fedo, C., Forni, O., Achilles, C., Bristow, T., et al. (2019). Chemical alteration of fine-grained sedimentary rocks at Gale crater. *Icarus*, *321*, 619–631. https://doi.org/10.1016/j.icarus.2018.11.004

Mangold, N., Forni, O., Dromart, G., Stack, K., Wiens, R. C., Gasnault, O., et al. (2015). Chemical variations in Yellowknife Bay formation sedimentary rocks analyzed by ChemCam on board the Curiosity rover on Mars. *Journal of Geophysical Research: Planets*, *120*(3), 452–482. https://doi.org/10.1002/2014JE004681

Maurice, S., Clegg, S. M., Wiens, R. C., Gasnault, O., Rapin, W., Forni, O., et al. (2016). ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars. *Journal of Analytical Atomic Spectrometry*, *31*(4), 863–889. https://doi.org/10.1039/C5JA00417A

Maurice, S., Wiens, R. C., Bernardi, P., Caïs, P., Robinson, S., Nelson, T., et al. (2021). The SuperCam instrument suite on the Mars 2020 rover: Science objectives and mast-unit description. *Space Science Reviews*, *217*(3), 47. https://doi.org/10.1007/s11214-021-00807-w

Maurice, S., Wiens, R. C., Saccoccio, M., Barraclough, B., Gasnault, O., Forni, O., et al. (2012). The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: Science objectives and mast unit description. *Space Science Reviews*, *170*(1–4), 95–166. https://doi.org/10.1007/s11214-012-9912-2

McLennan, S. M., Anderson, R. B., Bell, J. F., Bridges, J. C., Calef, F., Campbell, J. L., et al. (2014). Elemental geochemistry of sedimentary rocks at Yellowknife Bay, Gale crater, Mars. *Science*, *343*(6169), 1244734. https://doi.org/10.1126/science.1244734

Melikechi, N., Mezzacappa, A., Cousin, A., Lanza, N., Lasue, J., Clegg, S., et al. (2014). Correcting for variable laser-target distances of laser-induced breakdown spectroscopy measurements with ChemCam using emission lines of martian dust spectra. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *96*, 51–60. https://doi.org/10.1016/j.sab.2014.04.004

Meslin, P. Y., Gasnault, O., Forni, O., Schroder, S., Cousin, A., Berger, G., et al. (2013). Soil diversity and hydration as observed by ChemCam at Gale crater, Mars. *Science*, *341*(6153), 1238670. https://doi.org/10.1126/science.1238670

Milliken, R. E., Grotzinger, J. P., & Thomson, B. J. (2010). Paleoclimate of Mars as captured by the stratigraphic record in Gale crater. *Geophysical Research Letters*, *37*(4), 1–6. https://doi.org/10.1029/2009GL041870

Morris, R. V., Vaniman, D. T., Blake, D. F., Gellert, R., Chipera, S. J., Rampe, E. B., et al. (2016). Silicic volcanism on Mars evidenced by tridymite in high-SiO_2 sedimentary rock at Gale crater. *Proceedings of the National Academy of Sciences*, *113*(26), 7071–7076. https://doi.org/10.1073/pnas.1607098113

Morrison, S. M., Downs, R. T., Blake, D. F., Vaniman, D. T., Ming, D. W., Hazen, R. M., et al. (2018). Crystal chemistry of martian minerals from Bradbury landing through Naukluft plateau, Gale crater, Mars. *American Mineralogist*, *103*(6), 857–871. https://doi.org/10.2138/am-2018-6124

Nachon, M., Clegg, S. M., Mangold, N., Schröder, S., Kah, L. C., Dromart, G., et al. (2014). Calcium sulfate veins characterized by ChemCam/Curiosity at Gale crater, Mars. *Journal of Geophysical Research: Planets*, *119*(9), 1991–2016. https://doi.org/10.1002/2013JE004588

Nachon, M., Mangold, N., Forni, O., Kah, L., Cousin, A., Wiens, R., et al. (2017). Chemistry of diagenetic features analyzed by ChemCam at Pahrump Hills, Gale crater, Mars. *Icarus*, *281*, 121–136. https://doi.org/10.1016/j.icarus.2016.08.026

Palucis, M. C., Dietrich, W. E., Hayes, A. G., Williams, R. M., Gupta, S., Mangold, N., et al. (2014). The origin and evolution of the Peace Vallis fan system that drains to the Curiosity landing area, Gale crater, Mars. *Journal of Geophysical Research: Planets*, *119*(4), 705–728. https://doi.org/10.1002/2013JE004583

Pauca, V. P., Piper, J., & Plemmons, R. J. (2006). Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, *416*(1), 29–47. https://doi.org/10.1016/j.laa.2005.06.025

Payré, V., Siebach, K. L., Dasgupta, R., Udry, A., Rampe, E. B., & Morrison, S. M. (2020). Constraining ancient magmatic evolution on Mars using crystal chemistry of detrital igneous minerals in the sedimentary Bradbury group, Gale crater, Mars. *Journal of Geophysical Research: Planets*, *125*(8), e2020JE006467. https://doi.org/10.1029/2020JE006467

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rampe, E. B., Blake, D., Bristow, T., Ming, D., Vaniman, D., Morris, R., et al. (2020). Mineralogy and geochemistry of sedimentary rocks and eolian sediments in Gale crater, Mars: A review after six earth years of exploration with Curiosity. *Geochemistry*, *80*(2), 125605. https://doi.org/10.1016/j.chemer.2020.125605

Rampe, E. B., Bristow, T. F., Morris, R. V., Morrison, S. M., Achilles, C. N., Ming, D. W., et al. (2020). Mineralogy of Vera Rubin ridge from the Mars Science Laboratory CheMin instrument. *Journal of Geophysical Research: Planets*, *125*(9), 1–31. https://doi.org/10.1029/2019JE006306

Rampe, E. B., Ming, D., Blake, D., Bristow, T., Chipera, S., Grotzinger, J., et al. (2017). Mineralogy of an ancient lacustrine mudstone succession from the Murray formation, Gale crater, Mars. *Earth and Planetary Science Letters*, *471*, 172–185. https://doi.org/10.1016/j.epsl.2017.04.021

Rapin, W., Ehlmann, B. L., Dromart, G., Schieber, J., Thomas, N. H., Fischer, W. W., et al. (2019). An interval of high salinity in ancient Gale crater lake on Mars. *Nature Geoscience*, *12*(11), 889–895. https://doi.org/10.1038/s41561-019-0458-8

Rapin, W., Meslin, P.-Y., Maurice, S., Wiens, R., Laporte, D., Chauviré, B., et al. (2017). Quantification of water content by laser induced breakdown spectroscopy on Mars. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *130*, 82–100. https://doi.org/10.1016/j.sab.2017.02.007

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sautter, V., Fabre, C., Forni, O., Toplis, M. J., Cousin, A., Ollila, A. M., et al. (2014). Igneous mineralogy at Bradbury rise: The first ChemCam campaign at Gale crater. *Journal of Geophysical Research: Planets*, *119*(1), 30–46. https://doi.org/10.1002/2013JE004472

Sautter, V., Toplis, M. J., Wiens, R. C., Cousin, A., Fabre, C., Gasnault, O., et al. (2015). In situ evidence for continental crust on early Mars. *Nature Geoscience*, *8*(8), 605–609. https://doi.org/10.1038/ngeo2474

Schröder, S., Meslin, P.-Y., Gasnault, O., Maurice, S., Cousin, A., Wiens, R., et al. (2015). Hydrogen detection with ChemCam at Gale crater. *Icarus*, *249*, 43–61. https://doi.org/10.1016/j.icarus.2014.08.029

Siebach, K. L., Baker, M. B., Grotzinger, J. P., McLennan, S. M., Gellert, R., Thompson, L. M., & Hurowitz, J. A. (2017). Sorting out compositional trends in sedimentary rocks of the Bradbury group (Aeolis Palus), Gale crater, Mars. *Journal of Geophysical Research: Planets*, *122*(2), 295–328. https://doi.org/10.1002/2016JE005195

Sirven, J.-B., Bousquet, B., Canioni, L., & Sarger, L. (2006). Laser-induced breakdown spectroscopy of composite samples: Comparison of advanced chemometrics methods. *Analytical Chemistry*, *78*(5), 1462–1469. https://doi.org/10.1021/ac051721p

Takahashi, T., & Thornton, B. (2017). Quantitative methods for compensation of matrix effects and self-absorption in laser induced breakdown spectroscopy signals of solids. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *138*, 31–42. https://doi.org/10.1016/j.sab.2017.09.010

Thompson, L. M., Schmidt, M. E., Spray, J. G., Berger, J. A., Fairén, A. G., Campbell, J. L., et al. (2016). Potassium-rich sandstones within the Gale impact crater, Mars: The APXS perspective. *Journal of Geophysical Research: Planets*, *121*(10), 1981–2003. https://doi.org/10.1002/2016JE005055

Thorpe, M. T., Bristow, T. F., Rampe, E. B., Grotzinger, J. P., Fox, V. K., Bennett, K. A., et al. (2021). The mineralogy and sedimentary history of the Glen Torridon region, Gale crater, Mars. In *52nd lunar and planetary science conference*. Houston: Lunar and Planetary Institute. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2021/pdf/1519.pdf

Treiman, A. H., Bish, D. L., Vaniman, D. T., Chipera, S. J., Blake, D. F., Ming, D. W., et al. (2016). Mineralogy, provenance, and diagenesis of a potassic basaltic sandstone on Mars: CheMin X-ray diffraction of the Windjana sample (Kimberley area, Gale crater). *Journal of Geophysical Research: Planets*, *121*(1), 75–106. https://doi.org/10.1002/2015JE004932

Wiens, R., Maurice, S., Lasue, J., Forni, O., Anderson, R., Clegg, S., et al. (2013). Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *82*, 1–27. https://doi.org/10.1016/j.sab.2013.02.003

Wiens, R. C., & Maurice, S. (2015). ChemCam: Chemostratigraphy by the first Mars microprobe. *Elements*, *11*(1), 33–38. https://doi.org/10.2113/gselements.11.1.33

Wiens, R. C., Maurice, S., Barraclough, B., Saccoccio, M., Barkley, W. C., Bell, J. F., et al. (2012). The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: Body unit and combined system tests. *Space Science Reviews*, *170*(1–4), 167–227. https://doi.org/10.1007/s11214-012-9902-4

Wiens, R. C., Maurice, S., Robinson, S. H., Nelson, A. E., Cais, P., Bernardi, P., et al. (2021). The SuperCam instrument suite on the NASA Mars 2020 rover: Body unit and combined system tests. *Space Science Reviews*, *217*(1), 4. https://doi.org/10.1007/s11214-020-00777-5

Yen, A. S., Ming, D., Vaniman, D., Gellert, R., Blake, D., Morris, R., et al. (2017). Multiple stages of aqueous alteration along fractures in mudstone and sandstone strata in Gale crater, Mars. *Earth and Planetary Science Letters*, *471*, 186–198. https://doi.org/10.1016/j.epsl.2017.04.033

Yen, A. S., Morris, R. V., Ming, D. W., Schwenzer, S. P., Sutter, B., Vaniman, D. T., et al. (2021). Formation of tridymite and evidence for a hydrothermal history at Gale crater, Mars. *Journal of Geophysical Research: Planets*, *126*(3), 1–16. https://doi.org/10.1029/2020JE006569