Defining Surface Types of Mars Using Global CRISM Summary Product Maps

O. M. Kamps1, R. D. Hewson1, F. J. A. van Ruitenbeek1, and F. D. van der Meer1

1Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, Netherlands

Abstract For many regions on Mars, the surface composition and its geological history have been debated in the literature. Because of the limited surface coverage of in situ measurements, either new data or new processing methodologies are required to get a better understanding of the Martian geology. This paper presents the results of a multivariate, unsupervised, analysis on underutilized CRISM (Compact Reconnaissance Imaging Spectrometer) multispectral mapping mode data set for surface type analysis. The devised summary products of Pelkey et al. (2007) and Viviano-Beck et al. (2014) are averaged for the CRISM ~5° × 5° mosaic grids and used to analyze the variability in this data set. The averaged summary product values are studied using correlation coefficients calculated between summary products, as well as the correlation with dust coverage and elevation. The degree of correlation is used to interpret the summary products for the global distribution of mafic and secondary minerals and the effect of external factors such as ice, atmosphere, and dust coverage. With unsupervised clustering, all grid pixels are classified based on the spectral variability. These clusters are plotted as global maps and interpreted for geological variations in the CRISM data. Some clusters spatially correspond with previously recognized compositional distinct regions Northern Lowlands, Southern Highlands, Meridiani Planum, Syrtis Major, and Nili Fossae. Several other clusters are described here for the first time such as Solis Plenum, Ophir Planum, and Hellas Basin. These regions cover known geological units, but the interpretation of the spectral variability is uncertain whether it relates to geological or external factors.

Plain Language Summary Of all planets in the Solar System, Mars has a geological history most similar to Earth. The multiple missions to Mars carrying different instruments has taught us a lot about the geology of Mars. But for many regions, there is no published consensus on what geological process relates to the differences in surface composition. This study uses an instrument, which has rarely been used for global geology studies. The images are made mainly in the infrared light, in which, among others, specific volcanic and alteration minerals can be detected. Based on these images, we have defined several surface types, which have unique values calculated from the infrared data set (spectral parameters). Several of these surface types are similar to those defined with other orbital instruments and are known for their unique geological composition. Others are described here for the first time and are discussed whether these can be described for a unique local geology or the influence of external factors, such as the atmosphere and dust, on the spectrum.

1. Introduction

To make the most accurate reconstruction of the global surface composition and geological history of Mars, information is required from all available orbital instruments. So far, the global mapping data of CRISM (Compact Reconnaissance Imaging Spectrometer) has rarely been used for global surface analysis. Shallow absorption features in reflectance spectra and differences between orbital observations complicate the use of this data set. Nevertheless, it is thought that this data set can provide important information to understand Mars's global geology.

Most global surface composition studies that use infrared spectroscopy are performed with the TES (Thermal Emission Spectrometer) (Bandfield et al., 2000; Rogers et al., 2007) and OMEGA (Observatoire pour la Minéralogie, l’Eau, les Glaces et l’Activité) instruments (Bibring et al., 2006; Ody et al., 2012; Riu et al., 2019). Based on the analysis with both instruments, the general composition of Mars is estimated to be of basaltic to andesitic composition, with pyroxene, olivine, and feldspar as primary minerals (Ehlmann & Edwards, 2014; McSween et al., 2009). Additional geological processes contributed to the
presence of secondary minerals such as clays, sulfates, ferric oxides, and carbonates (Carter et al., 2013; Ehlmann & Edwards, 2014). The mineralogical composition is directly related to the climatic and geologic conditions during their formation and is therefore of importance to reconstruct the geological history.

The CRISM spectra are analyzed by using so-called summary products, which help to infer the presence of differing minerals by identifying important spectral features in the wavelength range between 0.4 and 4 \( \mu m \). These products have been developed for the CRISM data by Pelkey et al. (2007) and later revised by Viviano-Beck et al. (2014). Spectral features include band depths, spectral peaks, spectral band indices and ratios, spectral slope, and reflection values related to the compositions’ absorption spectral features (see Table S1.1 in the supporting information). Each of the products can be interpreted for specific mineralogy or surface property, although some have caveats for atmospheric conditions. Earlier studies showed that these products could highlight the presence of both the primary and secondary minerals (Pelkey et al., 2007; Viviano-Beck et al., 2014).

The Pelkey products have been previously calculated and analyzed from the global mapping CRISM data (Pelkey et al., 2007). The later Viviano-Beck products, however, are often used for the targeted CRISM observations. Because of the earlier described complications, the applicability of these later products for the multispectral reduced data record (MRDR) data set is published here for the first time. A visual and statistical comparison is done between the products of Viviano-Beck et al. (2014) and those of Pelkey et al. (2007) and other global mineral maps based on the data of the OMEGA and TES instruments.

The radiometric differences between orbital observations are addressed here by averaging the used summary products for a grid of circa \( 5^\circ \times 5^\circ \), where each grid tile is the size of the mosaic tile of the MRDR data set. Another advantage of using the resampled grid data is that the resolution provides a good balance between spatial coverage and spatial resolution on a global scale.

Figure 1. Flowchart of the methods presented in this study. Numbers above boxes indicate the section where the methodology is described in detail. Parallelograms indicate data sets, rectangle processes, and circles results.
An unsupervised clustering analysis will give an overview of the variability in the data set. The surface composition of these clusters and the potential geologic processes that formed them will be interpreted. The summary products are studied for correlations and anticorrelations between each other and between external data sets such as the digital elevation of the Mars Orbiter Laser Altimeter (MOLA) and the dust coverage index of the TES instrument. By studying the individual summary product maps in combination with the correlation coefficients, the global distribution of the summary product values is interpreted to understand their potential influence from mineralogy and also from possible other factors such as dust coverage and atmospheric thickness. The combination of the detectable minerals and the almost complete global coverage of the data set (Seelos & Murchie, 2018) makes the CRISM MRDR data set valuable for global surface classification. Using the global data set as input for this clustering analysis approach ensures that most diverse range of possible geological land covers are included.

2. Method and Data

The methodology of this study includes several successive analyses: testing for correlations between the summary products and surface type classification and finding the relationships between the surface types and summary products. These analyses are summarized in the flowchart in Figure 1. In the following section, the CRISM multispectral data set is introduced (section 2.1), followed by what data preparation has been done prior to the statistical analysis (section 2.2), with the statistical analysis described in the last section (section 2.3).

2.1. CRISM Multispectral Data Set

The CRISM instrument is an imaging spectrometer on the Mars Reconnaissance Orbiter, which operates in the visible/near-infrared wavelength range from 0.36 to 3.93 μm. The instrument has three types of observation modes: a targeted mode, a multispectral mapping mode, and an atmospheric mode (Murchie et al., 2007). The mapping mode, which is used for this study, corresponds to 73 wavelength channels with a spatial resolution of ~200 m/pix. Out of these 73 channels, 55 were measured by the infrared detector and 18 measured by the visible/near-infrared detector, which both work in parallel. The data set with these strips is also referred to as the Multispectral Reduced Data Record (i.e., MRDR). The latest estimation of global coverage is ~87% and approximately 80% coverage around the equator, with ~45% coverage having repeated sampling (Seelos & Murchie, 2018). MRDR includes image data of I/F (radiance/irradiance), reflectance, Lambertian albedo, summary products, and derived data records information, such as incidence, emission, phase angle, and surface temperature (Murchie et al., 2007). The MRDR data were made available on the planetary data system (PDS) of NASA, as individual strips and as mosaic tiles of 5° × 5°. The analyses in this study are all performed on the mosaic tiles. Both the Pelkey summary product mosaic tiles and albedo mosaic tiles used for our analyses are PDS version 3 from 2009.

Although the MRDR mosaic tiles are available for all latitudes, only the CRISM tiles between 67.5° north and south latitudes were used in this study. It is expected that this coverage is less affected by seasonal changes of ices in the polar regions and therefore we try to minimize the ice contribution to the CRISM spectra (Smith et al., 2001).

Spectral features in the CRISM wavelength range were described by spectral parameters, which are also called summary products (Table S1.1 in the supporting information). These products are developed by Pelkey et al. (2007) and Viviano-Beck et al. (2014) and are both used in this study, hereafter often referred as the Pelkey and Viviano-Beck data set, respectively. The summary products of Pelkey et al. (2007) are downloaded as mosaic tiles from the PDS. For the products of Viviano-Beck et al. (2014), the products are calculated using the Interactive Data Language (IDL) programming script from the CRISM processing toolbox, CAT ENVI. These summary products are calculated from the CRISM MRDR mosaic tile data set, measured in the mapping mode of the CRISM instrument. The supplied albedo data is already corrected for the atmospheric and photometric effect and shared as similar mosaic tiles for the Pelkey summary products data set (Murchie et al., 2007). The Pelkey data set has a total of 44 summary products, of which 34 products are calculated from the Lambert albedo data set, which are designed to relate with a mineralogy or surface composition (Pelkey et al., 2007). Viviano-Beck et al. (2014) revised some of these products and included new products, totally up to 60 products, of which 49 are developed for mineralogy or surface composition purposes. The parameters of Pelkey et al. (2007) were developed and tested in particular for the MRDR data.
set. The products of Viviano-Beck et al. (2014), however, are commonly used for the targeted measuring mode of CRISM but are designed such that they are also suitable to calculate for the MRDR data. However, the global maps for all of these products are assessed here for the first time.

The spectral features described with these products are one of the following four types: (1) reflectance at a specific wavelength, (2) spectral slope which is a linear slope defined by the reflectance different between two bands divided by the wavelength difference between two bands, (3) band depth, defined as the reflectance corresponding to the position of the minimum of the band divided by the reflectance calculated as a linear continuum fit between two reflectance wavelengths on each side of the band, and (4) indices derived using the ratio of reflectance values from different wavelengths. The summary products are designed so that a higher value represents a more prominent appearance of the spectral feature. For details and formulas of the used summary products, the reader is referred to Table S1.1, attached as supporting information or the papers of Pelkey et al. (2007) and Viviano-Beck et al. (2014).

2.2. Data Preparation

The Pelkey and Viviano-Beck summary products come with some challenges that are addressed below. One of the problems is that for some summary products, the MRDR data products show inconsistencies between overlapping or adjacent orbital strips. These radiometric residuals between the strips in the mosaic are a result of the atmospheric and photometric corrections done for the MRDR data strips before the mosaic is made (Seelos & Murchie, 2018). A second problem is that some pixels contain extreme values, which are more likely to be artifacts in the data rather than spectral features.

To overcome the problems with unrealistic values, lower and upper thresholds are defined based on the global mosaic data. Values that are lower or higher than these thresholds are masked and discarded from the analysis. These thresholds are defined based on quartile distances, also known as Tukey’s fences (Tukey, 1977). Quartiles or percentiles, of the data defines the percentage of lower data values. Quartile 1 (Q1) refers to the 25th percentile, which means that 25% of the data has lower values than Q1. The median refers to Q2 or 50th percentile and Q3 to the 75th percentile. The difference between Q3 and Q1 is defined as the interquartile distance (IQD). Tukey’s fences indicate that lower outliers can be defined by 1.5 times IQD minus the Q1. The upper threshold is defined by values 1.5 times the IQD plus Q3. For the band depth defined products, the minimum threshold value is set to zero, meaning that all negative band depths are excluded. A negative band depth refers to an absent spectral absorption feature and therefore indicates mineral’s absence. The resulting global summary product maps can be found as supplementary material (Data Set S1) and can be found in unannotated format in (Kamps, 2020), the maps for HCPINDEX and BD2250 are shown as two examples in Figure 2. The global maps of both summary products are presented at the original MRDR resolution (top figures) and the averaged 5° × 5° pixel size resolution (bottom figures). Underneath each global map also the upper and lower thresholds are indicated, and all pixels that are not considered for the averaging are masked out in the original resolution maps.

After masking all pixels outside the defined thresholds, we averaged the values for each individual summary product within a grid cell. The grid cells coincide for the size of a CRISM mosaic tile (~5° × 5°/pix).

For several reasons, such as the lack of pixel values within the defined thresholds or the atmospheric effect on some of the summary products, several products were discarded. In particular, from the Pelkey data set, the products VAR, BD1750, and BD2100 were excluded from the analysis because the data include too little number of pixels with band depths larger than zero. In the averaged data set, this resulted in a lack of spatially coherent patterns, which were found unlikely to represent a mineralogy. For the same reason, summary product BD860_2 was excluded from the Viviano-Beck data set. The summary products BD2500_2 and MIN2295_2480 from the Viviano-Beck data set were excluded because all values were negative. The discarded products are categorized separately in Table S1.1 in the supporting information and are not used in either of the analysis.

After averaging, several products appear to have no-data values for some pixels. These are the products related to reflectance values >3 μm in the vicinity of Hellas Basin (BD3000, BD3100, BD3200, BD3400, and CINDEX). In the Pelkey data set, also some pixels of the product BD860 have no data values for several pixels in the dust-covered regions. Because all products with no-data values are band depth products, the pixel values are set to zero which can be considered as no spectral feature.
2.3. Data Analysis

A combination of multivariate data analysis techniques is used to define the surface types and find the products that contributed significantly to defining these surface types. Prior to the statistical analysis, the values for each product are normalized and standardized, which means that for each summary product, the average is subtracted from each observation and divided by the standard deviation.

In the statistical analysis, both the variables (summary products) and samples (pixels) are studied. The summary products are studied to understand their relationship with the mineralogical composition. This is done by testing how the summary products are correlated to each other and to other variables such as elevation and the dust cover index from the TES instrument (resolution 16 pixels per degree) (Ruff & Christensen, 2002). The spatial patterns in the global summary product maps are compared with mineral maps based on studies with the CRISM, TES, and OMEGA instruments (Bandfield et al., 2003; Bibring et al., 2006; Christensen et al., 2001; Koeppen & Hamilton, 2008; Ody et al., 2012, 2013; Riu et al., 2019; Rogers et al., 2007; Ruff & Christensen, 2002).

2.3.1. Correlations Between Summary Product

The relationships between the different summary products are studied by calculating the Pearson’s correlation coefficients ($r$), described in section 3.1 (Härdle & Simar, 2007). The dust cover index (Ruff & Christensen, 2002) and MOLA digital elevation (Zuber et al., 1992) are also averaged for the size of a CRISM mosaic tile and included in the correlation analysis. In particular, the Pearson’s correlation coefficient calculated between summary products and digital elevation enables the effect of dust and atmosphere on the summary product values to be studied (Zuber et al., 1992).

Figure 2. Global maps of the summary products HCPINDEX and BD2250, at original MRDR resolution (~200 m/pix) and averaged resolution (5°/pix). At the original resolution the white pixels are those considered as outliers (section 2.2).
The correlation coefficient is an indicator of the influence of one variable on the other and is often used as an effect size. A Pearson’s correlation coefficient ranges between −1 and 1, where −1 means an absolute negative correlation, 1 a positive correlation and 0 no correlation at all (Cohen, 1988, 1992). Categorizing the effect size in low, moderate, and high effect is arbitrary. Here it is classified into the following categories: −0.6 < r or r < 0.6 indicating low correlation, 0.6 < r < 0.8 or −0.6 > r > −0.8 for moderate correlation, and −0.8 > r or r > 0.8 for high correlation. These thresholds are used in Figure 3 to indicate any effects between summary products, between summary products and dust coverage, and between summary products and elevation/atmosphere. As described in section 2.2, the conservative lower correlation coefficient of 0.6 is used to allow for excluding summary products with possible atmosphere effects.

**2.3.2. Defining Surface Types by Clustering Analysis**

To classify the data, we use hierarchical clustering analysis, which is an unsupervised clustering method (Härdle & Simar, 2007). Here hierarchical clustering analysis is favored above other cluster analysis strategies such as k-means, because hierarchical clustering does not require a prior assumption about the number of clusters. Instead, by using a tree diagram, also known as a dendrogram, the relationship between clusters can be studied. The surface types are studied with a divisive (top down) approach. With this approach, the clustering analysis is used to find clusters in the data being the most dissimilar, so those with the most unique surface composition (Hastie et al., 2009). Pixels are clustered by calculating the unweighted averages of the Euclidean distances (section 3.2).

Although hierarchical clustering analysis does not require a prior assumption about the number of clusters, it is of interest to this study to know how many clusters describe the variability in the data set best. At some point, defining more clusters would not indicate major variability in the data but smaller changes within clusters instead. The decision on the number of clusters is based on the knowledge of the data and the geology. As described by Hardy (1994), the validity of the number of clusters was tested with the elbow method of a graph plotting the mean Euclidean distance against the number of clusters (section 3.2). The elbow method assumes that significant clusters have a high Euclidean distance. At some point, adding new clusters would cause a decrease in slope because these new clusters are explaining minor spectral differences within a cluster instead of significant new clusters (Hardy, 1994).

**2.3.3. Relationships Between Surface Types and Summary Products by PLS-DA**

Summary products can be used to draw conclusions about the mineralogy and related surface types. Because the clustering was performed with multiple summary products, a multivariate analysis is preferred above comparing each individual map, to define the importance of each summary product on the definition of each surface type.

A common method to reduce the number of axes in the data set is principal component analysis (PCA), which defines new axes in the data set that describe the most variance (Härdle & Simar, 2007). Since we are interested in the variance between each cluster and the rest of the data and not in the variance within the complete data set, we used partial least-squares discriminant analysis (PLS-DA) (see section 3.2). This is a method that originates from the field of chemometrics (Brereton & Lloyd, 2014). Just like PCA, it creates new axes in the data set, where the first axes in PLS-DA describe most of the variance between groups. In our study, the PLS-DA is done for each cluster defined by the hierarchical clustering analysis. All pixels of the cluster we study for that specific PLS-DA are considered as one group and all other pixels as another group. The two groups are used as input for the PLS-DA to create a new axis that describes the most variance between these groups. The outcomes of the PLS-DA are components with weight values for all the variables and score values, which are pixel values projected on the new component axis. These can be analyzed as a bivariate plot, which is a scatter plot presenting both of these results in one figure. Based on this figure, it can be observed which variables, that is, summary products, relate to which surface type (section 3.2).

**3. Results**

**3.1. Correlation Between Summary Products**

Figure 3 shows a correlation coefficient matrix of all summary products and the two additional variables, elevation, and dust coverage. The brighter the color the higher the positive (green) or negative (red) correlation coefficient. The boxes highlight the correlation coefficient larger than 0.8 and the circles those larger than 0.6. The summary products are grouped into six categories based on our interpretation and the
Figure 3. Correlation coefficient matrices of summary products of the Pelkey (a) and Viviano-Beck (b) data sets. The brightness of the color shows the positive (green) and negative (red) Pearson’s correlation values. The circle highlights the correlation values with moderate correlation (−0.6 > r or r > 0.6) and squares + circles those with high correlation (−0.8 > r or r > 0.8); see the color bar in the upper right of the figure.
descriptions of Pelkey et al. (2007) and Viviano-Beck et al. (2014). Those products describing a mineralogy are categorized in mafic minerals, that is, olivine and pyroxene, ferric iron, and secondary minerals. We used the term secondary minerals to summarize the mineral groups of carbonates, sulfates, phyllosilicates, and hydrous silicates. These mineral groups have overlapping spectral features in the wavelength range between 2 and 2.5 μm, which makes it difficult to distinguish them from each other based on an individual spectral parameter. Besides mineralogical summary products, some are interpreted to be related to the dust coverage, the atmosphere conditions, and the ices.

The atmospheric effect on the summary products is tested by determining their Pearson’s correlation coefficients ($r$) with the digital elevation (see section 2.3.1). For the products with a moderate correlation coefficient ($r < -0.6$ or $r > 0.6$) with the digital elevation, the spectral features are considered to be significantly affected by atmospheric absorption and thereby biasing the values of the summary products, assuming that the atmospheric effects are linearly related with elevation. From the Pelkey data set, these are the products ICER2 and BDCARB and from the Viviano-Beck data set BD1400, BD1435, BD1900R2, BD2200, BD2355, ICER2, and BD3000. The reason of this relation between these products and the elevation can be that the products are sensitive for the spectral features of atmospheric CO$_2$ near 1.4, 1.9, and 2 μm (Brown et al., 2010; McGuire et al., 2009).

Because of the atmospheric effect on these summary products, these are excluded for the clustering and PLS-DA analysis. Therefore, for the following analysis, both the products categorized as discarded and atmospheric in Table S1.1 in the supporting information are not considered.

### 3.2. Classification Into Surface Types

The results of the clustering analysis are presented as global maps and are shown in Figure 4 with the corresponding dendrograms in Figure 5. The clustering analysis based on the Pelkey and Viviano-Beck data sets shows many similarities. In both analyses, the main branches in the dendrograms relate to the following surface types: northern lowlands, southern highlands, Hellas Basin, dust covered regions, and Syrtis Major and Meridiani. A total of 18 clusters have been defined by the summary products of Pelkey and a total of 17 clusters by those of Viviano-Beck. The names of the clusters in Figure 4 will be used the remainder of this paper. The results of the elbow method are attached in the supporting information. In the elbow plot (Figure S2.1), it shows that the number of clusters chosen are around the tipping point (elbow) where the change in Euclidean distance is constant. As mentioned in the section 2.3.2, this is the point where more cluster describe internal variance of cluster instead of significant new clusters.

### 3.3. Relationships of Summary Products and Surface Types

For each surface type, a PLS-DA is performed to test the contribution of each summary product to the definition of that surface type. Because this involves a total of 35 individual analyses (18 surface types derived from Pelkey et al., 2007, parameters and 17 surface types derived from Viviano-Beck et al., 2014, parameters), four geologically interesting clusters are shown here as an example (Figure 6). The examples include Syrtis Major + Sinus Meridiani, Nili Fossae + Meridiani Planum, northern lowlands, and the transition zone, all performed with the Viviano-Beck products. These figures present the score and weight values of the first two components of the PLS-DA. As described in the section 2.3.2, the first components explain most of the variance between the groups. The figures essentially display the same as a bivariate plot from a principal component analysis. The score values are the pixel values on a projected axis (PC). Weights are indicators of how much the summary product contributed to the axis, so the higher the weight, the higher the contribution.

Table 1 summarizes all PLS-DA results and can be found in plot form in the supporting information (Figures S4.1–S4.35). This table indicates for each surface type of both the Pelkey and Viviano-Beck data sets, the summary products that contributed to their classification, based on the weight values of the PLS-DA (see Figure 6). The variables that plot close to the specific surface type (encircled in the bivariate plot in Figure 6) have a positive contribution and those that plot opposite have a negative contribution in defining the surface type. Figure 6 shows that it is not always clear which summary products contributed most in defining the surface types. Therefore, the global maps of the distribution of each summary products are used to evaluate the interpreted importance of the summary product for a specific surface type.

The distance between the pixels of one surface type to all other pixels in the bivariate plots indicates how distinct the surface type is compared to all other pixels. For example, the pixels classified as transition
zone (Figure 6c) plot between the dust covered pixels and either the northern lowlands pixels or southern highland pixels. Therefore, no specific summary products were listed for the transition zone in Table 1.

4. Discussion

The results show that the CRISM multispectral mapping mode data are useful to assess the global surface geology. The novel approach with the use of summary products in combination with unsupervised data-analysis techniques has proved to be a transparent method to test for the variability in the CRISM data and evaluate for the local geology. The PLS-DA allows us to study the variance of each surface type in multi-dimensions as shown in Figure 6 and summarized in Table 1. Some are defined based on distinct geological phenomena, and others are related to nongeological processes or to artifacts in the data sets. The method shows to be consistent in that it exhibits similar surface types for the Pelkey and Viviano-Beck data sets and correspond to surface type classification studies based on TES (Bandfield et al., 2000; Rogers et al., 2007; Rogers & Hamilton, 2015), OMEGA (Riu et al., 2019), and GRS (Gasnault et al., 2010).
The spectral differences between the southern highlands and northern lowlands (Figure 4) are the most consistent in all global surface types studies. Just as it was observed by the TES and OMEGA instruments, the northern lowlands have limited spectral features related to the mafic minerals olivine and pyroxene in comparison with the southern highlands (Figure 6). The CRISM data shows that besides the mafic mineral difference, many secondary mineral summary products have high values for the northern lowland region (Figure 6). This could suggest a chemical weathering process in aqueous conditions. However, because of the low values of the band depth products (e.g., third decimal place numbers for BD2250 in Figure 2) and the lack of a spatial coherent pattern at the original MRDR resolution, this study is unable to be conclusive regarding the presence of secondary minerals in the northern lowlands. Furthermore, previous studies with the OMEGA instrument concluded that secondary mineral absorption features are rare to absent in the northern lowlands (Carter et al., 2013). The compositional change between the major surface types in the southern highlands, northern lowlands, and dust-covered regions seems to be gradual and classified as a separate surface type, called the transition zone here (Figure 4: cluster 15 Viviano-Beck, cluster 5 Pelkey).

The spectral differences between the southern highlands and northern lowlands (Figure 4) are the most consistent in all global surface types studies. Just as it was observed by the TES and OMEGA instruments, the northern lowlands have limited spectral features related to the mafic minerals olivine and pyroxene in comparison with the southern highlands (Figure 6). The CRISM data shows that besides the mafic mineral difference, many secondary mineral summary products have high values for the northern lowland region (Figure 6). This could suggest a chemical weathering process in aqueous conditions. However, because of the low values of the band depth products (e.g., third decimal place numbers for BD2250 in Figure 2) and the lack of a spatial coherent pattern at the original MRDR resolution, this study is unable to be conclusive regarding the presence of secondary minerals in the northern lowlands. Furthermore, previous studies with the OMEGA instrument concluded that secondary mineral absorption features are rare to absent in the northern lowlands (Carter et al., 2013). The compositional change between the major surface types in the southern highlands, northern lowlands, and dust-covered regions seems to be gradual and classified as a separate surface type, called the transition zone here (Figure 4: cluster 15 Viviano-Beck, cluster 5 Pelkey).

Much of the variability in the data is related to the dust coverage on Mars. This can be observed in the number of pixels that classify in the group dust-covered region (~38% of the pixels classify as cluster 9 in Figure 4) and the number of summary products with a moderate or high correlation with the dust cover index of Ruff and Christensen (2002) (Figure 3). The products related to this surface type are interpreted as the result of the high albedo of the dust and the dusts’ ferric component. The ferric component of the dust is often referred as nanophase ferric oxide (Ehlmann & Edwards, 2014). Here the products BD530_2 and RPEAK1 are

Figure 5. Dendrograms of the hierarchical clustering analysis. The main branches are named after the geographic regions in Figure 4 that are covered by the surface types.
interpreted to be related to the ferric component of the dust. The products OLINDEX and CINDEX are revised by Viviano-Beck et al. (2014) because of their sensitivity to high albedo, giving false positives for the interpretation of olivine and carbonates.

In contrast with previous studies, our study has highlighted a few different regional groupings (Syrtis Major together with Sinus Meridiani and Meridiani Planum with Nili Fossae), as well as a few new regions of spectral distinctions (e.g., Ophir Planum, Solis Planum, and Hellas Basin). As will be described, the previously unrecognized regional groupings show spectral similarities which can be interpreted for the local geology. However, also the averaging and thresholding process used to create the input maps had significant influence on the regional grouping (section 4.1). The new spectral distinctions (section 4.2) are discussed.

Figure 6. Bivariate plots presenting the score values (colored dots: pixels) and weights (black points and labels: summary product variables) of the summary product of the principal component, resulting from the PLS discriminant analysis. These are the surface types and summary products of (Viviano-Beck et al., 2014). The plots show the results for the surface types (a) Syrtis Major + Meridiani, (b) Nili Fossae + Meridiani Planum, (c) transition zone, and (d) northern lowlands. Dots correspond to the colors used for the global maps (Figure 4) of the main branches shown in Figure 5. The circles highlight the pixels that belong to the surface type labeled with the name above each subplot.
whether these might represent previously unrecognized regional differences in composition or surface properties (section 4.2.1) or are the result of influences of external factors such as dust and ice. The uncertainties of our work and additional explanation on how to interpret our results are discussed in section 4.3.

4.1. New Regional Groupings
4.1.1. Nili Fossae and Meridiani Planum

The regions Nili Fossae and Meridiani Planum classify in this study as similar surface types because of the interpreted high olivine and secondary mineral content. The high olivine content is indicated by the positive impact of the OLINDEX3 product and secondary minerals content indicated based on the products BD1900 and D2300 (Figure 5 and Table 1).

The interpretation of the summary products of secondary minerals is found difficult because at the original resolution the products lack a clear spatial coherent pattern (BD2250 as an example of a secondary mineral product in Figure 2). Also, the summary product values of the secondary mineral products are small (third decimal values for BD2250), which introduces some uncertainty because the features cannot easily be

Table 1
Summary of the Most Important Products of Pelkey and Viviano-Beck for Each Surface Unit (ST)

| Surface type                  | ST | Pelkey et al. (2007) | ST | Viviano-Beck et al. (2014) |
|-------------------------------|----|----------------------|----|---------------------------|
| Southern highlands            | 4  | BD11000VIS HCPINDEX  | 5  | HCPINDEX2 LCPINDEX2 BD920_2 SINDEX2 |
| Syrtis Major                  | 15 | R770 RRR RPEAK1 BD1000VIS | 12 | OLINDEX3 R1330 BD1300 VAR ICER1_2 BD12000 BD2100_2 BD2165 BD2190 BD2230 BD2250 MIN2250 CINDEX2 BD3400_2 R770 RPEAK1 RBR BD530_2 VAR |
| Nili Fossae + Meridiani Planum| 14 | D2300 BD11000IR      | 14 | OLINDEX3 D2300 LCPINDEX2 BD1000IR SINDEX2 |
| Sinus Meridiani               | 12 | BD12000 BD300 SH600  | 11 | BD1900_2 MIN2200 ICER1_2 BD3400_2 |
| Southern Latitude Zone        | 1  | SH700 R1330 BD11000IR R770 BD2210_2 |
| Southern Latitude Zone II     | 4  | SH700 BD2265         |
| Promethei Terra               | 13 | CINDEX               | 13 | CINDEX2 BD3400_2 |
| Northern Lowlands             | 17 | LCINDEX HCPINDEX     | 16 | OLINDEX3 LCPINDEX2 HCPINDEX2 ICER1_2 ISLOPE1 BD2200 BD2220 BD2240 BD2300 BD1435 BD2290 |
| Transition Zone               | 5  | Composition between dust covered region and adjacent surface type |
| Northern Latitude Zone        | 18 | D2300 BD11000IR ISLOPE1 | 17 | MIN2200 BD1500_2 HCPINDEX2 LCPINDEX2 SH600_2 |
| Equatorial Region             | 9  | R770 RRR BD530 RPEAK1 IRAC CINDEX OLINDEX BD11000R BD1000VIS | 9 | R1330 R770 BD530_2 RPEAK1 RBR SH770 BD1000IR VAR |
| Volcanoes                     | 10 | ICER1 BD1500 SINDEX BD3100 BD3200 | 10 | MIN2200 CINDEX2 BD3400_2 BD3200 BD1500_2 BD2210_2 ICER1_2 |
| Solis Planum                  | 8  | ISLOPE 1 BD2290 BD12000 | 8  | ISLOPE1 BD1900_2 BD2210_2 OLINDEX3 |
| Medusae Fossae                | 16 | BD1900               | 16 | |
| Hellas Basin                  | 3  | BD3100 CINDEX BD2200 BD3400 | 3  | BD3100 CINDEX2 BD3200 BD3400 VAR SINDEX2 SH600_2 OLINDEX3 D2300 RRR SH700 MIN2250 |
| Hellas North                  | 6  | SINDEX ICER1 BD2210 SH600 BD1435 | 6  | VAR SINDEX2 SH600_2 OLINDEX3 D2300 RRR SH700 MIN2250 |
| Hellas Middle                 | 7  | BD3200 SH600 BD3400 | 7  | VAR BD3100 |
| Hellas External               | 1  | BD3000 SINDEX BD3400 BD640 BD1500 BD1900 | 2  | SH600_2 RRR SH700 |
| Hellas South West             | 2  | CINDEX               |    |                           |

Note. Products with negative contribution are underlined. Numbers in the column Surface type correspond to those in Figures 4 and 5.
recognized in the spectra. The summary products D2300 is displayed at original resolution in Figure 7 for the regions Meridiani Planum and Nili Fossae to demonstrate that the pixel values are spatially coherent and can be related to the local geology (Figure 7). The local geology is indicated in black, outlining the boundaries of Meridiani Planum and Syrtis Major from the geological map of Tanaka et al. (2014). Gray squares indicate the outline of the clusters in this study. At the original resolution, it can be noticed that the high values of D2300 are occurring within the outline of Meridiani Planum. On this plain, both in situ as orbital observations have detected sulfate evaporates and Fe/Mg phyllosilicates silicates formed by aqueous alteration (Squyres & Knoll, 2005). Both D2300 and BD1900 describe spectral features related to Fe/Mg phyllosilicates (Viviano-Beck et al., 2014). The low values of the sulfate index (SINDEX Table 1) and lack of a BD1750 feature region suggest that the presence of sulfates could not be detected with our analysis.

For Nili Fossae, the product D2300 shows locally high values (Figure 7). The region Nili Fossae is often described in literature in relation with the presence of secondary minerals such as carbonates and Fe/Mg phyllosilicates (Brown et al., 2010; Ehlmann & Edwards, 2014). In these studies, hydrothermal alteration is often named as the circumstance where these minerals have formed. Both carbonates and Fe-Mg
phylosilicates have spectral features around 2.3 and 1.9 μm, and both minerals can form under hydrothermal circumstances.

The OLINDEX3 map is most similar to the mean fractional contribution map for olivine by Koeppen and Hamilton (2008) based on the data of the TES instrument. In their study, the mean abundance of olivine is relatively high in the regions that overlap with the surface types Nili Fossae and Meridiani Planum in this study. Other surface type studies such as Rogers and Hamilton (2015), Hamilton and Christensen (2005), and Hoefen et al. (2003) have also found Nili Fossae to be distinctly mapped because of the high olivine content. Although the values for OLINDEX3 are relatively high for Nili Fossae compared to the rest of Mars, in our maps, these are not significantly different compared to the rest of Meridiani and Syrtis. This is different from the olivine maps produced with the TES (Bandfield et al., 2000; Koeppen & Hamilton, 2008) and OMEGA (Ody et al., 2012) data. This difference can be explained by the effect of averaging and thresholding of the pixel values that causes masking of the extreme values. A close look at OLINDEX3 product in Nili Fossae shows that some high values in this region are masked out because of the defined threshold.

Besides the presence of secondary minerals, other orbital studies have found that Meridiani is unique because of the presence of iron-oxide minerals, especially hematite (Christensen et al., 2001). Product SH600 is able to capture the presence of hematite Viviano-Beck et al. (2014). The values for this product are high in Meridiani Planum but not significantly higher than the adjacent region Margaritifer Terra (which is not a separately classified surface type in this study).

4.1.2. Syrtis Major and Sinus Meridiani

One of the aspects that distinguishes the surface types Syrtis Major and Sinus Meridiani class from the other regions in the southern highlands is the low dust coverage, with the lowest values for all the dust coverage products (see negative weights for dust products in Figure 6). Beside the dust, the regions are characterized by high positive values for several secondary mineral products and the OLINDEX3.

In contrast to the previous section on Nili Fossae and Meridiani Planum, there are some doubts whether the OLINDEX3 indicates a high olivine content in Syrtis Major and Sinus Meridiani. These doubts are based on the fact that none of the other olivine maps by the TES or OMEGA instruments indicated high olivine content in this region. This difference might be explained by the fact that the OLINDEX3 is also sensitive to the 1 μm feature of high-calcium pyroxene (Viviano-Beck et al., 2014). The high-calcium pyroxene content is often related to what makes Syrtis unique in other surface type studies (Riu et al., 2019; Rogers & Hamilton, 2015). The elevated high-calcium content in this region would result in high values of the HCPINDEX, which is not the case in this region. Similar to the olivine index values in Nili Fossae, this could be related to the fact that the high values of the HCPINDEX parameter are masked out by the thresholding process and averaging of pixel values (see white pixels in Figure 2).

Several secondary mineral summary products have high values in Syrtis Major. One of those products, BD2100_2, is displayed in Figure 7 for both Syrtis Major and Meridiani Planum. BD2100_2 appears to be noisier and does not show a clear distribution compared to D2300. This lack of a clear spatial distribution was found for all the secondary mineral summary products related to Syrtis and Sinus Meridiani. Also, several secondary mineral summary products appear to be correlated with the dust coverage (Figure 3). These summary products have high values within Syrtis Major but not significantly different from other low dust-covered regions such as northern lowlands and locally in the southern highlands. Therefore, it remains difficult to conclude on the presence of secondary minerals in this region.

4.2. New Spectral Distinctions

As described in section 2, the clustering analysis is performed with a top-down approach so that the clusters that are most distinct are identified first. Most of the most distinct clusters have been recognized as spectrally distinctive in previous studies that utilized infrared spectroscopic data sets. But several clusters of this studies appear new, in comparison with other surface type studies with the OMEGA and TES data sets such as: Hellas, Volcanoes, Promethei Terra, Solis Planum, Ophir Planum, and Medusae Fossae. However, for some of these regions, the summary products with high impact or defining these regions are interpreted to be related to nongeological conditions such as dust coverage and local atmosphere conditions. These regions are for example the dust-covered regions and the volcanic regions, and Hellas (section 4.2.2 clusters of nongeological origin). For the clusters Ophir Planum, Solis Planum, Promethei Terra, and the higher northern
and southern latitude zones, it is not well understood whether the spectral uniqueness is related to geological processes or external factors. Since these clusters are new and overlap with established geological units, the uniqueness of these clusters in the data is discussed in section 4.2.1 clusters of potential geological origin.

### 4.2.1. Clusters of Potential Geological Origin

Cluster Solis Planum covers in both data sets the Hesperian volcanic region on the geological map of Tanaka et al. (2014). Although the cluster is named the same in both the Viviano-Beck and Pelkey data set, their spatial coverage is different in that the cluster in the Viviano-Beck data extends more to the North toward the Tharsis region. In both data sets, the spectral difference with other units is minor, which can be observed from the low Euclidean distance (Figure 5). The regions are characterized by high spectral slope (ISLOPE) values and low values for secondary mineral and olivine summary products. Previous laboratory spectral studies have shown that a continuous spectral slope between 1.8 and 2.5 μm can arise from either the result of acidic alteration of basaltic glass (Horgan et al., 2017), rock exposure (Harloff & Aarnold, 2000), or ferric coating on dark rocks (Fischer & Pieters, 1993). The lower values for hydrated mineral summary products could be explained by more recent volcanism, of which the rocks might have remained unaltered. The low values for OLINDEX3 remains unclear, especially because the detailed studies of Viviano et al. (2019) concluded that the composition of this region is similar to other Hesperian regions on Mars.

Ophir Planum is a cluster which consists in both data sets of several of pixels north of Valles Marineris. The high values in this region for the summary products SINDEX and BD1900 could indicate the presence of sulfate minerals in the region. Cluster Ophir Planum overlaps with some of the chaos terrains in Margaritifer Terra that contain earlier identified light toned deposits (Glotch & Rogers, 2007). The identified sulfate deposits in these interior layered deposits could be a reason for the higher values for the products SINDEX and BD1900 (Gendrin et al., 2005). However, both products are also known to be sensitive for the spectral features of ice. This in combination with the lower values of the ICER product in both data sets could also imply the presence of atmospheric water ice. The presence of atmospheric ice in this region is also mentioned for its occurrence during the aerocentric solar longitudes Ls 0–180 (Benson et al., 2006).

Several clusters, in both data sets, cover the larger basins on Mars, Hellas, and Agyle (Figure 4). Clusters 3-Hellas Basin and 13-Promethei Terra are in both data sets classified first and were distinguished with high Euclidean distances (Figure 5). The spectral uniqueness of this region is related to the missing values of the summary products using wavelengths >3 μm as described in section 2.2. Also, the clusters Hellas North and Hellas middle appear spectrally unique with high Euclidean distance with the other clusters. The secondary mineral summary products, for example, have low values for this region, while the products OLINDEX3, SH600_2, and SININDEX have high values. However, studying the individual global summary product maps, it seems that all products have unique values in this region. The known circumstances of high opacity due to dust in the atmosphere (Ogohara & Satomura, 2008) and the presence of ice clouds (Kahre et al., 2020) complicate the interpretation whether mineralogy contributed to the spectral uniqueness or these external effects.

### 4.2.2. Clusters of Nongeological Origin

In addition to the surface types of potentially geological origin, some classes in the clustering analysis such as the volcanoes and the higher latitude zones result from unique summary product values which are interpreted to be of nongeological origin. In section 2.2, it was already mentioned that several summary products were not used for the analysis because of their high correlation with the elevation (Pelkey: ICER2, BDCARB; Viviano-Beck: BD1400, BD1435, BD1900R2, BD2200, BD2355, ICER2, and BD3000; Figure 3). These summary products are interpreted here to be related to the atmosphere thickness.

Not just globally but also local atmospheric conditions seem to affect the spectra and therefore the summary products. In particular, the pixels covering the elevated volcanoes on Mars are classified as separate surface types because of high values of BD1500, SININDEX, CINDEX, BD3100, BD3200, and BD3400 (Table 1). These products can be associated with spectral features of sulfate and carbonate, but here we think that they are affected by the local atmospheric conditions around the volcano. It is known that the atmosphere around the volcanoes hosts ice clouds (Benson et al., 2006) and from the products BD1500 and SINDEX it is known that these are sensitive for H2O and CO2 ice, respectively (Viviano-Beck et al., 2014). Also, the other products capture spectral features that overlap with any of the H2O and CO2 ice spectra.

Our two classification maps (Figure 4) have both have regions in the higher latitudes that cluster separately. It seems that although most of the polar pixels are excluded, the regions up to 60° latitude are affected by
seasonal changes, resulting in spectral differences between orbital strips within the mosaics. Such abrupt spectral changes around 60° latitude were also observed by Pelkey et al. (2007) in the OMEGA and CRISM data set and related to season changes.

4.3. Uncertainties

In this study, choices were made that impacted on the uncertainties/accuracy of the results. First of all, the summary products were normalized and standardized for the region between the 67.5° latitudes. Therefore, some products that were developed to study ice around the poles are used to study surface composition where ice might not be expected. We assumed that when the poles are excluded, these summary products describe spectral features that can be related to mineralogy. Also, the normalization and standardization, together with the defined thresholds, could have resulted in enhancement of noise within areas of low spectral variation.

The use of thresholds based on quartile distances has proven to be a good method to exclude unrealistic data values. But for some products such as OLINDEX3 in Nili Fossae, it might have masked relevant information. The averaging over such a large grid helped to overcome some of the problems with radiometric differences between strips and smoothened the data.

It was also mentioned in the papers of both Pelkey et al. (2007) and Viviano-Beck et al. (2014) that the use of summary products contains caveats and interpretations have to be done with care. Therefore, as a reminder, the interpretations resulting from this study are based on averaged summary products over 5° × 5° pixels and for a global perspective. Some interpretations might not apply to the original resolution of the MRDR summary product maps since higher resolution data provides more detailed information.

5. Conclusions

The global multispectral CRISM data has shown to be a useful data set for global surface type classification. For the first time, both the spectral parameters of Pelkey et al. (2007) and Viviano-Beck et al. (2014) are used to capture the small spectral features of the CRISM mapping mode data and interpreted for the local mineralogy. However, several assumptions had to be made such as averaging for a 5° × 5° grid and defining thresholds to mask unlikely values. These assumptions had to be made to deal with radiometric differences between strips in the CRISM MRDR data. With a novel and transparent approach, including a combination of multivariate data-analysis techniques, we have classified Syrtis Major, Nili Fossae, Meridiani Planum, Sinus Meridiani, northern lowlands and southern highlands, and the global dust deposits as separate surface types, consistent with previous global spectral studies (Bandfield et al., 2000; Riu et al., 2019; Rogers et al., 2007; Rogers & Hamilton, 2015). The effect of external factors such as the dust coverage and the elevation on the summary products is studied using correlation coefficients. The Viviano-Beck products BD1400, BD1435, BD1900R2, D2200, BD2355, ICER2, and BD3000 and the Pelkey products ICER2 and BDCARB were found to be correlated with elevation, which is considered to be inversely related with atmosphere thickness. The clusters covering the higher latitude zones, Hellas, and the volcanoes are interpreted to be of nongeological origin and found to be related to atmosphere conditions, dust-coverage, atmosphere conditions, or missing values in the data. Surface types Solis Planum and Ophir Planum are new in comparison with surface type studies done with other global data sets. For these, the geological interpretation of the spectral variability is more uncertain whether it is can be related to geology or possible atmospheric ice.

References

Bandfield, J. L., Glotch, T. D., & Christensen, P. R. (2003). Spectroscopic identification of carbonate minerals in the Martian dust. Science, 301(5636), 1084–1087. https://doi.org/10.1126/science.1088054

Bandfield, J. L., Hamilton, V. E., & Christensen, P. R. (2000). A global view of Martian surface compositions from MGS-TES. Science, 287(5458), 1626–1630. https://doi.org/10.1126/science.287.5458.1626

Benson, J. L., James, P. B., Cantor, B. A., & Remigio, R. (2006). Interannual variability of water ice clouds over major Martian volcanoes observed by MOC. Icarus, 184(2), 365–371. https://doi.org/10.1016/j.icarus.2006.03.014

Bibring, J. P., Langevin, Y., Mustard, J. F., Poulet, F., Arvidson, R., Gendrin, A., et al. (2006). Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. Science, 312(5772), 400–404. https://doi.org/10.1126/science.1122659

Breteron, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. Journal of Chemometrics, 28(4), 213–225. https://doi.org/10.1002/cem.2609

Acknowledgments

We thank Christina Viviano for her feedback on the manuscript, which significantly improved the quality of the paper and our understanding of summary products. The original CRISM data used for this study was obtained using the PDS Geosciences Node Mars Orbital Data Explorer. The newer versions of the albedo data are downloaded from the MROCR_3101 (Northern Hemisphere) and MROCR_3102 (Southern Hemisphere) MRDR folders. The summary product end products of the Pelkey data set was downloaded from the MROCR_3001 and MROCR_3002 MRDR folders. The data developed for the analysis presented here are made available in a DANS repository (https://doi.org/10.17026/dans-25g-tt32), and the global mosaics can be requested via DANS's email address (info@dans.knaw.nl).
Seelos, F. P., & Murchie, S. L. (2018). CRISM mapping data empirical radiometric reconciliation for the next-generation Mars global multispectral map. Paper presented at the Lunar and Planetary Science Conference.

Smith, D. E., Zuber, M. T., & Neumann, G. A. (2001). Seasonal variations of snow depth on Mars. Science, 294(5549), 2141–2146. https://doi.org/10.1126/science.1066556

Squyres, S. W., & Knoll, A. H. (2005). Sedimentary rocks at Meridiani Planum: Origin, diagenesis, and implications for life on Mars. Earth and Planetary Science Letters, 240(1), 1–10. https://doi.org/10.1016/j.epsl.2005.09.038

Tanaka, K. L., Skinner, J. A., Dohm, J. M., Irwin, R. P. III, Kolb, E. J., Fortezzo, C. M., et al. (2014). Geologic map of Mars (3292). Reston, VA. Retrieved from http://pubs.er.usgs.gov/publication/sim3292

Tukey, J. W. (1977). Exploratory data analysis. (Vol. 2). Reading, MA.

Viviano, C. E., Murchie, S. L., Daubar, I. J., Morgan, M. F., Seelos, F. P., & Plescia, J. B. (2019). Composition of Amazonian volcanic materials in Tharsis and Elysium, Mars, from MRO/CRISM reflectance spectra. Icarus, 328, 274–286. https://doi.org/10.1016/j.icarus.2019.03.001

Viviano-Beck, C. E., Seelos, F. P., Murchie, S. L., Kahn, E. G., Seelos, K. D., Taylor, H. W., et al. (2014). Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars. Journal of Geophysical Research: Planets, 119, 1403–1431. https://doi.org/10.1002/2014JE004627

Zuber, M. T., Smith, D. E., Solomon, S. C., Muhleman, D. O., Head, J. W., Garvin, J. B., et al. (1992). The Mars observer laser altimeter investigation. Journal of Geophysical Research, 97(E5), 7781–7797. https://doi.org/10.1029/92JE00341