1. **Methods**

***2.1 Mastcam Instrument Description***

The Mastcam instrument is a pair of CCD cameras with fixed focal lengths (34-mm in the left camera and 100-mm in the right camera) mounted roughly 2 m above the surface on the rover's mast [*Malin et al.,* 2017]. Each camera obtains images through a Bayer pattern of red, green and blue (RGB) filters and telecentric microlenses bonded onto the charge-coupled device (CCD) and an eight-position filter wheel that provides the ability to obtain spectra in 12 unique wavelengths between 445 and 1013 nm, shown in Table 1 [*Bell et al.*, 2017]. Each filter wheel includes a “clear” (broadband) filter with a near-infrared cutoff for Bayer RGB images, six narrowband geology filters (three of which are redundant in the left and right cameras for stereo imaging), and a solar filter for atmospheric monitoring (e.g., tau measurements) and astronomical observations (e.g., solar transits of Mars’ moons Phobos and Deimos). The Mastcam multispectral capability derives heritage from the color imagers on previous missions, including Mars Pathfinder [*Smith et al.*, 1997], the Mars Exploration Rovers (MERs) [*Bell et al.*, 2003] and Phoenix [*Smith et al.*, 2009].

This spectral region covered by Mastcam’s geology filters is particularly sensitive to iron-bearing primary basaltic minerals [*e.g.,* *Adams*, 1974], secondary iron oxide and oxyhydroxide minerals [*e.g.*, *Singer*, 1982; *Morris et al.*, 1985; *Bell et al.,* 1990]; ferric sulfates, ferric carbonates, and iron-bearing clays [*e.g.*, *Sherman et al.,* 1982]. Mastcam’s longest wavelength filters are also expected to have some sensitivity to hydrated and/or hydroxylated minerals. Specifically, the 1013 nm near-IR filters (filters L6 and R6) can detect an absorption due to the 2ν1 + ν3 H2O combination band and/or the 3νOH overtone when this band minimum occurs longward of roughly 980-nm (e.g., as in water ice, some carbonates, chlorides and hydrated sulfates).

**Table 1.** Mastcam RGB Bayer and geology filters effective center wavelengths (λeff) and half-widths at half-maximum (HWHM) [*Bell et al.*, 2017].

|  |  |  |  |
| --- | --- | --- | --- |
| Mastcam Left (M34) | | Mastcam Right (M100) | |
| **Filter Position** | **λeff ± HWHM (nm)** | **Filter Position** | **λeff ± HWHM (nm)** |
| L0 (Red Bayer) | 640 ± 44 | R0 (Red Bayer) | 640 ± 44 |
| L0 (Green Bayer) | 554 ± 38 | R0 (Green Bayer) | 554 ± 38 |
| L0 (Blue Bayer) | 495 ± 37 | R0 (Blue Bayer) | 495 ± 37 |
| L1 | 527 ± 10 | R1 | 527 ± 10 |
| L2 | 445 ± 10 | R2 | 447 ± 10 |
| L3 | 751 ± 10 | R3 | 805 ± 10 |
| L4 | 676 ± 10 | R5 | 937 ± 10 |
| L5 | 867 ± 10 | R4 | 908 ± 10 |
| L6 | 1012 ± 21 | R6 | 1013 ± 21 |

***2.2 Mastcam Image Calibration***

Mastcam observations were calibrated to radiance using pre-flight calibration coefficients; these radiance products are available via the NASA Planetary Data System (PDS) [*Bell et al.*, 2017]. Radiance was then converted to radiance factor (*I/F*, I is equal to the measured scene radiance and π*F* is the solar irradiance at the top of the Martian atmosphere at the time of the observation) using associated observations of the Mastcam calibration target. The effects of airfall dust on the calibration target were corrected for using two-stream radiative transfer models [*Kinch et al.,* 2015]. Radiance factor was then converted to relative reflectance (*R\**) [*Reid et al.*, 1999], also known as the “reflectance factor” [*Hapke et al.*, 1993], by dividing *I/F* by the cosine of the solar incidence angle (which can be derived from values in PDS image headers). A similar procedure was used to calibrate MER Pancam images [*Bell et al.,* 2006]. Absolute calibration accuracy for Mastcam’s filter set is estimated to be 10-15% or better [*Bell et al.,* 2017].

Images calibrated to *R\** are partially “atmospherically corrected” because observations of the Pancam calibration target also include near-simultaneous measurements of the Mastcam sky illumination component of the scene radiance [*Kinch et al.*, 2015; *Bell et al.*, 2017]. However, the calibration of Mastcam images to *R\** involves several assumptions: (1) the Sun is a point source; (2) the solar incidence angle remains constant within an image; (3) the scene is perfectly flat (and parallel to the calibration target); and (4) the scene elements being imaged are Lambertian scatterers. To minimize uncertainties that arise from these assumptions, Mastcam multispectral observations are typically acquired as close to local noon as possible (between 10:30 and 13:30 LTST, local noon ± 1.5 hours).

***2.3 Compilation of a Mastcam Spectral Database***

A major objective of this work was to compile a comprehensive database of spectra that sample the full spectral diversity observed across Curiosity’s traverse. For each Mastcam multispectral observation (Table S1), the spectral variability in the scene was characterized by manually identifying “color end members” through a visual inspection of the approximate true color (ATC) images, false color images and decorrelation stretch (DCS) products [*Gillespie et al*., 1986] (e.g., Figure 1). False color and DCS composites were made from combinations of Mastcam filter images that which produced the largest color contrasts for each observation. While the specific filter combination that produces the most variability in false color and DCS images is not necessarily the same for each scene, we found that RXXX and LYYY produced the best color contrast in a majority of observations. We identified end members as groupings of pixels that exhibit distinct colors in the false color and DCS products and also represent geologically-distinct surfaces (as identified in the ATC images). Care was taken to distinguish spectral end members corresponding to different compositions from those representing small differences in local viewing geometry (e.g., the multiple facets of a homogenous rock). In instances of variable dust cover on an otherwise homogenous surface, end members were selected on the most- and least-dusty regions.

We extracted a representative spectrum of each end member by manually selecting pixels from common regions of interest (ROIs) in the right and left camera images. In selecting ROIs, we adhered to a system of “best practices” to ensure the extraction of geologically-meaningful spectra with minimal noise. The minimum ROI size includes 30 unsaturated pixels in the M34 (Mastcam left) images, with rare exceptions for very small features (e.g, narrow veins). Where possible, spectra were extracted from near-horizontal surfaces near the center of the image, to best match the assumptions made in the Mastcam calibration pipeline [*Bell et al.*, 2017]. We avoided surfaces exhibiting specular reflections, shadowed regions, and/or rover hardware. Each observation was inspected by multiple people to verify end member identification and ensure consistency in ROI selection following “best practices” above.

Pixels with DN values greater than XXXX were flagged as “saturated” and were excluded from ROI averages. Spectra from the two cameras were joined by normalizing to the average *R\** value of the L6 and R6 bands at 1012/1013 nm. The other stereo positions (L1/R1, L2/R2, L0B/R0B, L0G/R0G, and R0R/R0R) were then averaged. We represented error in *R\** as the variance among the selected ROI pixels, rather than from the formal instrumental noise (which is generally much lower) [*Bell et al.*, 2017].

* 1. ***Compilation of Relevant Metadata***

To enable a comprehensive analysis of the entire Mastcam multispectral dataset, each endmember spectrum was compiled with relevant metadata. Observation-specific metadata are included in Table S1, and spectrum-specific metadata are included in our archived Mastcam spectral database (REF). A number of the metadata fields specific to each observation were taken directly from the images’ Planetary Data System version 3 (PDS3) headers, including: the Mastcam sequence identifier (seq ID); target name; day of the mission (sol); time of day measured as the local true solar time (LTST) at the start of the observation; camera focal distance; site index and rover drive number (which resets after each site index increment, so that site index and drive number together give a unique rover position).

Geographic information was taken from localization data compiled for each rover position by the MSL team [*e.g.*, *Vasavada et al., 2015*], including: latitude, longitude, total traverse distance (odometry), and rover elevation. We also include metadata for the season, measured as solar longitude (Ls), which is the Mars-Sun angle measured from the Northern Hemisphere spring equinox. Atmospheric optical depth is also included for each observation, given as τ (“tau”) [*Lemmon et al.*, 2015]; for sols where a direct measurement of tau is not available, we include an estimated tau value via linear interpolation. Target elevation was calculated from the camera focal distance (as an approximation of the distance from Mastcam to the surface in the center of the image), instrument elevation, and rover elevation, using an instrument height of 1.9 m above the ground and assuming a flat surface. For observations of targets in the mid- to far-field, where the camera focal distance was set to infinity, we calculated the target elevation using stereo range data for the distance to an ROI in the center of the image.

Observation geometries were calculated using instrument data in the PDS3 headers: the incidence angle (*i*) is taken as the solar elevation minus 90 degrees, and the emission angle (*e*) is taken as the instrument elevation plus 90 degrees. The phase angle (*g*) is defined as the angle between the incidence and emission vectors, which is determined from the following relationship:

(1)

where is the angle between the projection of the incidence vector and emission vector on the surface, given as the difference between the two absolute azimuths, or [*e.g., Shepherd et al.*, 2008].

Each spectrum was assigned one of the geologic “feature types” listed in Table 2. All of the rocks were also designated as either “float” (not attached to outcrop) or “in-place” (outcrop), and assigned specific lithology information (group, formation, and member) [*Latest Strat Column,* 2019]. Float rocks were

**Table 2.** Feature types assigned to each Mastcam spectrum.

|  |  |  |
| --- | --- | --- |
|  | Feature Type | Description |
| Soils | Undisturbed Soil | Sand and/or soil that has not been disturbed by the rover |
| Disturbed Soil | Sand and/or soil that has clearly been disrupted by the rover’s wheels, scoop, and or/or drilling activities |
| Drill Fines | Drill Tailings | The annulus of fine-grained material surrounding a the hole after drilling |
| Dump Piles | Drill core material that was crushed and sieved and dumped on the ground following a drill campaign |
| Rocks | DRT Target | Rock surfaces that have been brushed by the Dust Removal Tool (and are relatively dust-free) |
| Broken Rock | Freshly-exposed interior surfaces of rocks that have been broken open by Curiosity’s wheels and/or drill activities |
| Nodule-rich Rock | Undisturbed rock surfaces that include a high density (≲ 50% by area) of concretions and/or nodules |
| Rock | Undisturbed rock surfaces that do not include a high density of concretions and/or nodules |
| Other | Vein | Light-toned fracture-fill material |
| Other | Features that do not fit into any of the above categories |

***2.3 Spectral Classification***

Spectra were visually inspected for known VNIR features associated with iron-bearing mineralogy [e.g., *Burns*, 1970, 1993; *Gaffey* et al., 1993; *Morris et al.,* 1985] and/or hydration [*Rice et al.*, 2010].We quantified a number of parameters previously used to characterize Mastcam spectra [*Wellington et al.,*2017] and adapted others that had been used to characterize Pancam spectra [*Farrand et al.,* 2008; 2014; *Rice et al.,* 2011], summarized in Table 3. These include combinations of two or three Mastcam filters as reflectance ratios, slope values (the difference in reflectance values divided by the difference in wavelength) and band depth values. The band depth at a given Mastcam wavelength position was quantified depth beneath a continuum line, using the definition of *Clark & Roush* [1984]:

(2)

where *Rb­* is the reflectance at the band center *λb* and *Rc* is the reflectance of the continuum at the same wavelength as *Rb*. The continuum is a straight line defined by two “shoulder” positions on either side of the absorption feature. For left and right shoulder reflectance values *RL* and *RR* at wavelength positions *λL* and *λR*, the reflectance of the continuum is:

(3)

where

(4)

and

(5)

When possible, we use spectral parameters defined by only right- or left-eye filters (to avoid errors introduced by combining different ROI spectra from the two Mastcams).

**Table 3.** Summary of spectral parameters used to characterize Mastcam spectra.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Formula** | **Possible Mineralogic Indicators** |
| 527 nm band depth | 1 – (*R\*527* / (0.23*R\*447*) + (0.77*R\*551*)]) | Larger value can indicate higher degree of Fe oxidation [*e.g.*, *Farrand et al.*, 2008] |
| 554 – 751 nm slope | (*R\*751 – R\*554*) / (751 – 554) | Larger values are consistent with higher degrees of oxidation [*e.g.*, *Farrand et al.*, 2008] and indicate “redder” spectra |
| 527 – 805 nm peak | λ where *R\**λ = max(*R\*527…R\*805*) | Positions < 600 nm are consistent with reduced Fe; Longer wavelength peak positions are consistent with Fe oxidation |
| 805 – 937 nm slope | (*R\*937 – R\*805*) / (937 – 805) | Negative slopes may indicate broad Fe absorptions in the NIR; “Flat” slopes are consistent with phases that are spectrally-neutral in the NIR (e.g., pure sulfates); Positive slopes are consistent with hematite. For vein spectra, this slope defines the continuum beneath which we measure the 1013 nm band depth |
| 908 nm band depth | 1 – (*R\*527* / (0.23*R\*447*) + (0.77*R\*551*)]) | Broad absorption in NIR consistent with pyroxene, Fe-oxides and various Fe-sulfates |
| 1013 nm band depth | 1 – (*R\*527* / (0.23*R\*447*) + (0.77*R\*551*)]) | Narrow absorption consistent with hydration when NIR spectrum is otherwise featureless [*e.g., Rice et al.*, 2010] |

To first order, the rock spectra along Curiosity’s traverse are similar and highly correlated. In order to maximize the variance within the Mastcam spectral dataset, therefore, we also used a principal components approach. Principal Component Analysis (PCA) is a dimensional reduction technique in which a linear orthogonal transformation transforms a dataset into a new coordinate space [*e.g.*, *Davis*, 1973]. In the new coordinate space, data can be represented as a linear sum of orthogonal principal components, which are chosen in the PCA process to be in order of decreasing variance (*i.e*., the most variance in the dataset lies along the first principal component (PC1), the second most variance lies along the second component (PC2), etc.). When using PCA on Mastcam spectral data, the principal components themselves can be represented as Mastcam spectra. This approach has previously been applied to identify spectral classes from Pancam datasets along the Spirit and Opportunity rover traverses [*e.g.*, *Farrand et al.*, 2006; 2008; 2013].

We used PCA to examine the two subsets of spectra in our database that exhibit the least amount of obscuration by the ubiquitous Mars surface dust: a “dust-cleared rock” group consisting the DRT target and Broken Rock feature types; and a “drill fines” group consisting the Drill Tailings and Dump Pile feature types (as summarized in Table 2). In the drill fines group, the data averaged from the L2 R2 and L0 R0 (Blue Bayer) filter pairs was omitted to avoid “blue artifacts”, a known calibration error that appears in many drill fines spectra [*Wellington et al.*, 2017]. After these corrections, we examined the two sets of 12-point (dust-cleared rock group) and 10-point (drill fine group) Mastcam spectra. These spectra were represented as a point in a higher dimensional space, where each axis corresponds to the reflectance at a specific wavelength. PCA was performed on the dataset using the sklearn Python package.  Contributions of the various principal components to each spectrum were plotted against one another to represent Mastcam spectra in component space, where unique spectral endmembers can be easily identified by their separation from the rest of the data cloud.