11-1-2019

# Exploring Planetary Surfaces with Remote Sensing

Donald R. Hood

# EXPLORING PLANETARY SURFACES WITH REMOTE SENSING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geology and Geophysics

by
Donald Ramsey Hood
B.S. Carnegie Mellon University, 2014
December 2019

*To my mother, who never squashed my curiosity.*

*To my wife, the picture of perseverance and dedication.*

*To God, who made the beautiful worlds I am lucky to study.*

## Acknowledgements

I thank the following collaborators for their roles in the work presented below: Taylor Judice, Deanne Rogers, James Dohm, and Lorraine Carnes for their contributions to Chapter 1, Olivier Gasnault, Amy Williams, Luju Ojha, Shannon Kobs, Kyeong Kim, Jen Heldmann, and Cayla Fralick for their contributions to Chapter 2, and Caleb Fassett, Steven Sholes, and Patrick Brothers for their contributions to Chapter 3.

I thank the LSU Department of Geology and Geophysics for supporting me as a TA during my graduate work. I also thank the LSU Foundation, the New Orleans Geological Society, and the American Association of Petroleum Geologists for supporting through a variety of scholarships, fellowships, and grants. Further grants from NASA's Mars Data Analysis Program (NNH17ZDA001N-MDAP) and the Louisiana Space Consortiums Research Enhancement Award, Research Award Program, and Graduate Student Research Assistantship made much of this work and my graduate career possible.

I thank my committee members Juan Lorenzo and Barb Dutrow. Their grounded advice and focus on fundamental scientific principles kept me rooted in the often-lofty world of planetary science.

I thank David Susko, who joined me for the better part of my time at LSU. I've never met anyone who believed so truly in mankind's destiny in space, and his commitment is an inspiration.

Finally, I thank my advisor Suniti. In my years at LSU he has challenged me and asked much of me, but he always believed in me. So, I thank him for that faith, the opportunities and training he has given me, and for helping me push my limits and become a better scientist.

## Table of Contents

## Abstract

Aside from Earth, Mars is the most well-explored planetary body in our solar system. Much has been learned about its geologic past including a history of volcanism, ice ages, and potentially long-lasting liquid water. Much of this knowledge of Mars comes from remote observations of the surface and atmosphere from Mars-orbiting satellites. Such remote observations remain the only way to examine the entirety of Mars' surface as rovers and landers can only examine small areas. The chapters of this work cover three projects that explore the surface of Mars through a variety of remote sensing methods.

In chapter 1, Gamma-ray spectroscopy and infrared spectroscopy are used to assess the changes in what may have been an evolving volcanic system in Thaumasia Planum, Mars. The geochemical changes within this region are consistent with models for martian mantle evolution rather than other formation hypotheses. In Chapter 2, global trends in soil geochemistry and surface dust cover are examined, looking for distinct geochemical trends on either side of the martian topographic dichotomy. The volatile chemistry of soils on either side of this dichotomy are found to deviate from one another and may indicate much greater aqueous interaction in the southern highlands. In Chapter 3, a novel algorithm is presented that provides an automated tool for location and measuring boulders on the martian surface, called the Martian Boulder Automatic Recognition System or MBARS. This tool enables rapid analysis of boulder populations, assisting widespread investigation of surface processes.

These chapters display the utility of a wide range of remote sensing methods as well as the challenges and rewards that come from careful combination of multiple remote sensing methods. Remote observations remain the most productive avenue to study large-scale processes on Mars due to the relative ease of collection and large data volume. Even beyond Mars, remote observations are an important part of planetary exploration where in-situ access can be prohibitively expensive or even dangerous. The continued analysis of remote data, application of new techniques, and development of new tools for remote analysis are crucially important for the advancement of planetary science and exploration.

## Introduction: Eyes Over the Red Planet

Exploring Mars is one of the most exciting and challenging things we as a species have ever pursued. We are faced with the challenge of unraveling the history of a planet that is simultaneously similar to our own Earth and yet completely exotic. On Earth, many of the rocks that record ancient events are buried deeply or exposed as outcrops in select areas. However, on Mars, large areas of the surface are over three billion years old, and even some of the youngest rocks would rank among the oldest rocks seen on Earth's surface.

Mars surface ages are broken into three defined time periods: The Noachian (~4.65-3.7 Ga), Hesperian (~3.7-3.0 Ga) and Amazonian (<3 Ga) (e.g. J. W. Head, 2007). Mars's time periods are more difficult to distinguish than those on Earth, as surface ages are generally derived from crater abundance (e.g. Hartmann et al., 2010). Based on radiometric ages from lunar samples, models were developed for cratering flux through time, allowing a derivation of age based on crater population. Extrapolation of cratering rates and uncertainties to Mars' surface conditions through time make these crater-based ages less reliable, but they remain the most effective way to date most of martian surface as in-situ dating is currently limited to results from the Curiosity Rover(Farley et al., 2014). The three martian time periods can be summarized based on the dominant geologic processes. The Noachian, Mars' most geologically active time period, had widespread volcanism, intense cratering, and fluvial activity. Much of the southern highlands, ~2/3 of the planet's surface, is interpreted as Noachian volcanic material, although much of the surface is modified such that distinct volcanic forms are not readily visible. Many of the fluvial features most consistent with flowing water (e.g. valley networks, deltaic deposits, etc.) are Noachian in age, suggesting stability of liquid (most likely water) on the surface. While the abundance of fluvial features demands widespread liquid water, it is currently unclear if a long, stable period of water or ice activity is required to generate the observed features (e.g. Fastook & Head, 2015). In the Hesperian, widespread fluvial activity is generally reduced, but the massive, catastrophic outflow channels leading from Valles Marineris into Chryse Planitia formed in this time. Volcanic activity persisted in this time and most of the major volcanic regions on Mars (Elysium, Tharsis, Syrtis Major, Hesperia) show some degree of Hesperian-aged terrain. It is suggested, in the Hesperian, that a major transition occurred where explosive volcanism on Mars ceased, leading to the effusive-dominated volcanic constructs formed in later time periods. During this time, the northern lowlands were covered in volcanic material forming the modern Vastitas Borealis formation (Tanaka et al., 2014) and creating the young surface of the northern lowlands.

The Amazonian is the longest and most quiescent of these periods. While the Tharsis and Elysium volcanic constructs continued to grow, most other geological processes slowed considerably. Fluvial activity in the Amazonian is restricted to local occurrences, the origin of which is currently unclear (e.g. Kite et al., 2017). Polar and glacial processes are a dominant force in landscape change, in part due to large changes in obliquity. Mars' chaotic obliquity cycles makes it difficult to precisely predict obliquity past tens of millions of years. However obliquity is known to have shifted wildly, leading to rotation axis migration from 23° down to 48° (Touma & Wisdom, 1993). This shifting pole is thought to lead to the polar layer deposits (J.

Head et al., 2003) and provides a mechanism to deposit modern mid-latitudinal ice bodies (Stuurman et al., 2016). Modern Mars is much like the rest of the Amazonian: cold, dry, and tectonically quiet (Pike et al., 2019), although it is in a unique period of stable and Earth-like obliquity. While some volcanic activity has been suggested in the past few million years (Hauber et al., 2011) eruptions have never been recorded on Mars. Fluvial activity seems completely absent, although some briny flows are possible (Edwards & Piqueux, 2016), and the polar regions show the greatest degree of change and activity.

The above history has been determined primarily from surface or near-surface observations, and rare martian meteorites. However, even with this limited information, several important questions can be pursued: How long did water persist on and near the surface? Did life ever exist there? What caused the martian environment to reach its barren state? In pursuit of these questions a veritable armada of clippers and orbiters, in addition to several landed missions, have been sent to explore Mars, providing a wealth of orbital images and observations without disturbing the pristine surface conditions. This ability to gather information about planetary objects without disturbing them is crucial and is an important stage of exploration for planetary bodies, especially those that may harbor life. However, remote sensing methods also present several challenges: multiple datasets are required to provide a more complete picture of the target and differing spatial resolutions and sampling depths among datasets can make them hard to compare. Accommodating and incorporating these fundamental challenges is essential to any remote sensing investigation.

This dissertation focuses on three works, each depending uniquely on the view provided by orbital remote sensing techniques. These projects leverage the regional and global perspective of remotely-sensed data to approach regional and global-scale questions while carefully and rigorously handling the inherent challenges of such datasets.

In **Chapter 1**, two distinct remote sensing techniques: Gamma-ray spectroscopy and Infrared spectroscopy are combined with mapped geology to assess the changes in what may have been an evolving volcanic system in Thaumasia Planum, Mars. The region surrounding Thaumasia Planum contains volcanic units spanning much of Mars' history and acts as a useful recorder of the general evolution of volcanism on Mars. The shifts in geochemistry within this region are consistent with models for martian mantle evolution, although evidence is found that suggests some deviation from the overall mantle.

In **Chapter 2**, global trends in soil geochemistry and surface dust cover are examined, looking for distinct geochemical trends on either side of the martian topographic dichotomy. The volatile chemistry of soils on either side of this dichotomy are found to deviate from one another and may indicate much greater soil alteration occurring in the southern highlands.

In **Chapter 3**, a novel algorithm is presented that provides an automated tool for location and measuring boulders on the martian surface, called the Martian Boulder Automatic Recognition System or MBARS. Estimates of boulder population morphometry have been used to examine surface weathering (de Haas et al., 2013), periglacial processes (T. C. Orloff et al.,

2011), and cratering dynamics (Krishna & Kumar, 2016). This tool enables such investigations to examine larger areas and more boulders, leading to more refined results.

These three chapters use remote sensing methods including gamma-ray spectroscopy, remote imaging, and thermal-infrared spectroscopy, each of which provides a unique perspective of the martian surface. Collectively, they demonstrate many facets of working with remote sensing data, its utility in modern geology, and indispensability in planetary exploration to explore global and regional-scale questions.

# Chapter 1. Assessing the Geologic Evolution of Greater Thaumasia, Mars

## 1.1 Introduction

Mars Odyssey Gamma Ray Spectrometer (GRS) mapping indicates that the chemical variability of the martian subsurface is subdued at the approximately 450 km spatial scales observed by GRS (Boynton et al., 2007) compared to rover, lander, and meteorite observations. Nevertheless, recent investigations with Odyssey GRS data have revealed statistically distinct chemical provinces. A trilogy of papers (Gasnault et al., 2010; Suniti Karunatillake et al., 2009; G. Jeffrey Taylor et al., 2010) discuss a preliminary view of these provinces. Geologic mapping investigations also reveal provinces based on stratigraphy, paleotectonic history, impact crater statistics, and geophysical map information (e.g. magnetic data) (J. M. Dohm et al., 2005; J M Dohm et al., 2013). The geologic provinces based on these datasets have been shown to have bulk chemical distinctions consistent with their respective lines of evidence (J. M. Dohm et al., 2005; J M Dohm et al., 2013; James M. Dohm et al., 2007, 2008; James M. Dohm, Anderson, et al., 2009). For example, the extremely ancient geologic provinces, which contain large expanses of Noachian terrain, show relatively high crater retention ages and remnant magnetic anomalies (Acuña et al., 1999; Connerney et al., 2001). When compared to terrain with lower crater-retention ages and no magnetic anomalies, such as major volcanic provinces (e.g. Tharsis, Syrtis Major, Elysium), these ancient geologic provinces show elevated Th, K, and Fe (J. M. Dohm et al., 2005; J M Dohm et al., 2013).

In this work, we focus on the Thaumasia region because it is a significant record of the geological evolution of Mars, which includes Early Noachian (possibly > 4.0 Ga) mountain building such as the formation of the Thaumasia Highlands (James M. Dohm & Tanaka, 1999), the Middle Noachian (~ 3.93 Ga based on (Robbins et al., 2013)) Argyre impact event, and the Noachian to Amazonian development of Tharsis (J. M. Dohm et al., 2005; J M Dohm et al., 2013; James M. Dohm, Ferris, et al., 2001; James M. Dohm, Tanaka, et al., 2001; Scott & Tanaka, 1986; Tanaka et al., 2014). The region, which we refer to here as the Greater Thaumasia region, therefore likely contains diverse rock materials (i.e. poorly consolidated rocks and regolith), ranging from crustal basement (including mountain-forming and impact-exposed) to younger Tharsis-associated volcanics. Two of the three previously mentioned papers that sought to define provinces based on geochemical data noted unique signatures within the Greater Thaumasia region. Within the Greater Thaumasia region, these two works observe a transition between province types (G. Jeffrey Taylor et al., 2010) and three spatially contiguous chemical provinces (Suniti Karunatillake et al., 2009).

Karunatillake et al. (2009) defined “chemical provinces” on Mars using the GRS-derived 5°×5° chemical maps of the martian mid-latitudes. Each chemical province is a broad area delineated by contiguous bins for which individual compositions deviate by more than 1 standard deviation from the global average for two or more elements. We refer to these signatures of 1 standard deviation above or below as “enriched” or “depleted” respectively, without specific

Previously published by AGU as D. R. Hood *et al.*, “Assessing the geologic evolution of Greater Thaumasia, Mars,” *J. Geophys. Res. Planets*, pp. 1–17, 2016. Reproduced here in compliance with AGU’s Usage Permissions.

inference on the process that produced the anomalous abundance. This simultaneous deviation in two elements, in addition to filtering the results to maintain robust statistics despite spatial autocorrelation and smoothing, reduces the possibility of chemical provinces being erroneous or insignificant. Some of these defined chemical provinces overlap spatially with anomalous areas of other datasets (J. M. Dohm et al., 2005; J M Dohm et al., 2013; James M. Dohm et al., 2007, 2008; James M. Dohm, Anderson, et al., 2009; James M. Dohm, Williams, et al., 2009), making them compelling for further investigation.

In particular, we investigate three chemical provinces of the Thaumasia region identified by Karunatillake et al. (2009): provinces 70, 75, and 90 (Figure 1.1). While the chemical signatures of 75 and 70 are unique to the Thaumasia region, three major regions (Thaumasia, Elysium, Aonia Terra) share the same chemical signature classified as province 90. In this work, we only consider the province type 90 located in the vicinity of Thaumasia. Province 70 shows a depletion in H along with an enrichment in Si, and province 90 shows a depletion in both K and Th while province 75 shows all four of these anomalies(Suniti Karunatillake et al., 2009). The signature that constitutes province 70 covers 45 GRS pixels (5° x 5° each), roughly 60° x 15° (~4 million $km^2$), while the signature for province 90 covers 17 pixels (~1.5 million $km^2$). Province 75 overlaps provinces 70 and 90, covering a 10-pixel area (~0.9 million $km^2$). Based on the geologic mapping investigations (James M. Dohm, Tanaka, et al., 2001; James M. Dohm & Tanaka, 1999; Tanaka et al., 2014), province 90 includes Late Hesperian lava flow materials near the summit of Syria Planum in two areas, one of which occurs to the southeast of the summit and one to the east near the southern margin of Valles Marineris. Province 75 includes a large part of the southeastern flanks of Syria Planum, composed largely of Hesperian lava-flow materials. Province 70 includes the Coprates Rise mountain range, the eastern portion of the Thaumasia Highlands mountain range, and the transition zone between Tharsis and Argyre basins. The transition zone includes fluvial, alluvial, colluvial, and possibly glacial materials originating from the mountain ranges, mixing with Argyre-impact modified terrain (J. M. Dohm, Hare, et al., 2015).

Geologic mapping investigations and regional geomorphologic analyses have motivated several hypotheses concerning the geological evolution of Greater Thaumasia, such as the existence of complex tectonic features that may relate to early plate tectonics (Anguita et al., 2001; V R Baker et al., 2002; Victor R Baker et al., 2007; Connerney et al., 2005; J. M. Dohm, Spagnuolo, et al., 2015; J M Dohm et al., 2002, 2013; J M Dohm, Maruyama, et al., 2001; A;lberto G. Fairén et al., 2002; Alberto G. Fairén & Dohm, 2004; Maruyama et al., 2001), and a mega-landslide (Montgomery et al., 2009). Likewise, the Claritas Rise and surrounding terrains, which overlap the Greater Thaumasia region, may impart a magmatic signature that pre-dates Tharsis (James M. Dohm, Anderson, et al., 2009) and provide preliminary clues to a lithospherically active period of Mars (J. M. Dohm, Spagnuolo, et al., 2015). It is also the site of anomalously low crustal thermal energy production in the vicinity of Solis Planum, which is inferred from regional geochemical observations (Hahn et al., 2011). Summarily, Greater Thaumasia may contain important clues of regional-scale processes that modify the crust and surface of the planet both before and after the shutdown of the dynamo. Pre-shutdown mountain building is focused in the Thaumasia Highlands, which includes the Claritas rise (James M.

Dohm, Anderson, et al., 2009). Post-shutdown activity in Tharsis includes the development of Thaumasia Plateau and the magmatic-driven opening of Valles Marines. In addition, a major magmatic-tectonic center, Syria Planum, has been mapped and interpreted to have emplaced voluminous lava flows post-shutdown.(R. C. Anderson et al., 2001; James M. Dohm, Tanaka, et al., 2001; James M. Dohm & Tanaka, 1999; Tanaka, Robbins, et al., 2014 and references therein).

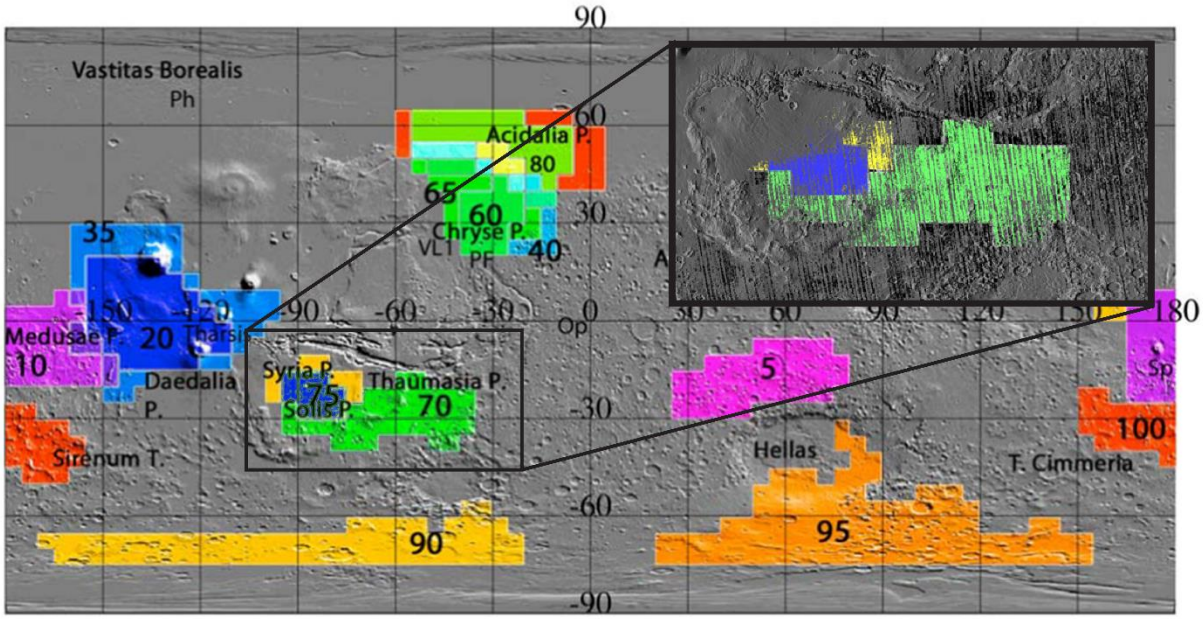



Figure 1.1. Global map of Mars, which highlights various chemically anomalous provinces. These regions show elemental abundances of two or more elements that are 1 standard deviation above the global average. The data are derived from GRS observations of gamma spectra. Each color represents a different chemical signature. In yellow, near Syria Planum, is chemical province 90 (K, Th depleted), in blue, surrounding Solis Planum, is chemical province 75 (K, Th, and H depleted; Si enriched), the green spanning Thaumasia Planum is chemical province 70 (H depleted; Si enriched). These regions constitute our study area. Insert in the upper right shows the coverage of TES data in the study area, produced by Deanne Rogers. The lack of data for the western provinces is mainly due to dust cover in those regions. Modified from Karunatillake et al. (2009)

We characterize Greater Thaumasia in terms of chemistry, mineralogy, and geologic age. Given the regional importance of Thaumasia, this work uses chemical maps derived from GRS, and mineralogical characterization from thermal infrared spectroscopy (i.e., Mars Global Surveyor Thermal Emission Spectrometer (TES)), allowing regional scale and decimeters depth sensitivity, while aiming to retain the local outcrop-driven mineralogic insight from TES. We seek to assess how significantly and in what manner the processes of the martian shallow subsurface are coupled with bulk crustal processes. We also employ the latest global geologic map information compiled by Tanaka et al. (2014) for preliminary geological insight. Based on the collective observations, driven primarily by composition, we assess the implications to

existing hypotheses concerning the geological evolution of this region of Mars informally referred to here as Greater Thaumasia.

## 1.2 Data and Methods

Compositional comparison of Greater Thaumasia with average martian crust relies upon the suite of chemical maps derived from GRS (Boynton et al., 2007), including Al, Ca, Cl, Fe, H, S, Si, K and Th. These maps extend roughly from 45°N to 55°S with some latitudinal variance dependent on longitude. High H abundance in the polar regions effectively moderates neutron energies, which invalidates the assumptions used to model elemental abundances (Boynton et al., 2007). This specifically prevents the calculation of Si abundance, which in turn prevents the normalization of all other elements (excepting K, Th, and H). The resulting maps do not constrain elemental abundances in the polar regions and we do not examine areas of the map that have incomplete data. We refer to this data coverage constrained to the middle and lower latitudes as "mid-latitudinal" throughout the work. For comparisons of chemical subdivisions within Thaumasia, four of these elements – H, K, Si, and Th – distinguish the chemical provinces, and are used to estimate proportions of enrichment and depletion that may in turn provide insight into the nature of processes (including geological and hydrological). Though sulfur content obtained from GRS (S Karunatillake et al., 2014) is typically expressed in 10°×10° chemical maps, we re-binned the data to 5°×5° pixels, assigning the same composition to each as reported in the original 10°×10° bin. We avoid using interpolation methods to upsample the sulfur maps as these may alter the values we primarily use in analysis (median, $25^{th}$ and $75^{th}$ percentile), which are unaltered under our simpler method.

Specifically, we use the same temporal compilation of gamma coverage as the pilot study(Suniti Karunatillake et al., 2009) of the radar "stealth" region west of Arsia and Pavonis Montes characterized by low radar reflectance. This temporal compilation consists of the two mapping periods of 8 June 2002 to 2 April 2005 and 30 April 2005 to 22 March 2006. Furthermore, for consistency with motivating work (S Karunatillake et al., 2014; Suniti Karunatillake et al., 2009; G. Jeffrey Taylor et al., 2010), and for allowing the inclusion of recently released S maps, we use the same chemical map version as used by Karunatillake et al. (2014) dated to 2010/04/29.

### 1.2.1 Modified box-and-whiskers comparisons for compositional analysis

Among statistical tools, we use modified box-and-whisker comparisons (c.f. fig 3, Suniti Karunatillake et al., 2011) to characterize the relative distribution of chemical compositions from one chemical province to another. We use global comparisons between the three chemical provinces of Greater Thaumasia – 70, 75, and 90 (Figure 1.1) - and the elemental chemistry of the martian crust to characterize inter-provincial chemical trends. Here, we use an average of mid-latitudinal GRS pixels, excluding the study region, as a proxy for the "average martian crust". This measure of average is similar to previous estimations (G. Jeffrey Taylor et al., 2010) in the abundances of Fe, Ca, and Th, however our measure of average martian crust is lower in K and Si abundances at the four and five sigma level respectively. The disparity in the chemical character of "average martian crust" Si abundance is possibly related to our exclusion of Greater

Thaumasia, a Si-enriched region, from the calculation. The lower abundances in K likely arise from the methodological differences in calculating the chemistry of the "average martian crust", however this only serves to make our observations of regionally depleted K more robust.

The modified box-and-whiskers comparisons utilized here show three key values: the ratio of the 75th percentile of the sample region over the 25th percentile of the reference region, the 25th percentile of the sample over the 75th percentile of the reference, and the 50th percentile over the 50th percentile (i.e., ratio of the medians, as a more informative parameter than the averages). We consider a chemical province enriched or depleted in a specified element with respect to another chemical province if the ratio of the median of two provinces is greater than or less than 1, respectively. To calculate error bars, we use the Median Absolute Deviation (MAD) (Bevington & Robinson, 2003), which provides the uncertainty of each median. Specifically, we calculate the median of the absolute deviations from the distribution median, which provides a measure of the dispersion of the dataset, rather than the uncertainty of the median. Because we are calculating dispersion, the presented error bars are conservative. These errors are used to assess the significance of any observed enrichment or depletion. Our graphical distributional analyses capture not only the comparison of the centers of the distributions, but also the high and low extents of the distribution. As discussed by Karunatillake et al. (2011), this enables a graphically summarized comparison of two distributions with immediate insight into the maximum and minimum compositional differences.

The ratios of some elements such as Fe/Si (major rock-forming elements, lithophiles), K/Th (large ion lithophiles), and $Cl/H_2O$ (key volatiles) are also included in the modified box-and-whiskers analysis. These help to determine the comparative likelihood of igneous, chemical alteration, or mixing processes.

### 1.2.2 Calculation of Mg and Na Content

In our analysis of GRS and TES data, we use TAS diagrams (Section 1.3.2) and several ternary diagrams (Section 1.3.4) to compare our provinces to GRS, TES, and in situ observations of the martian surface. These diagrams, however, call for abundances of Na and Mg, which are not constrained by GRS observations. The work by Baratoux et al. in (2014) used oxide ratios from analyses of basaltic shergottite meteorites (Table 1.1) to estimate the Na content of the martian surface as observed by GRS. The methods presented there also calculated Mn, P, and Ti content based on shergottite meteorite oxide ratios. After these four elements are calculated, Mg is calculated as the deficit of observed elements. The $Na_2O/K_2O$ ratio used in this method (6.8 ± 2.73) (David Baratoux et al., 2014) near the ratios observed at Gusev (6.8±1.28) (Blake et al., 2013) and within uncertainty of the ratios observed at Rocknest soil (~5.5±.23) and Meridiani (~4.6±.61) (Blake et al., 2013). We applied the method presented in Baratoux's (2014) work to our chemical datasets to produce Mg maps that are consistent with those presented by Baratoux. Ultimately, this method results in an uncertainty for $Na_2O$ content (~2% w($Na_2O$)) that is large relative to the average $Na_2O$ content (~2.5% w($Na_2O$)). However, Baratoux's (2014) work shows that this method, when used on martian meteorites, produced CIPW-normalized mineralogy that matched well with observed mineralogy. We observe convincing overlap between Provinces 70,

75, and 90 and the Greater Thaumasia TES observations (Section 1.2.4), which suggests that this method has produced accurate results for the Greater Thaumasia region.

Table 1.1. Basaltic Shergottite-derived ratios used in calculation of Mg and Na content from GRS observations. Ratios and errors from Baratoux et al. (2014)

| Ratio | $Na_2O/TiO_2$ | $TiO_2/P_2O_5$ | MnO/FeO | $P_2O_5/K_2O$ |
|---|---|---|---|---|
| SNC | 1.374 | 1.151 | .025 | 4.297 |
| Error (1σ) | .337 | .265 | .001 | .939 |

Previous efforts to estimate Mg abundance based on GRS observations have utilized Mg/Si ratios derived from meteorite observations (H. Y. J. McSween et al., 2009). We did not use this method for several reasons. First, the Mg/Si ratio can be altered by igneous processes. The region we are examining covers a spatially and temporally widespread igneous history, and as such, we cannot expect that this ratio would remain constant within the Greater Thaumasia region. In fact, when calculated with the method used by Baratoux et al. in (2014), Mg and Si abundance show an inverse relationship. As Si content increases to the east, Mg content decreases. Second, this region is distinguished by its unusually high Si content (Suniti Karunatillake et al., 2009) making it unique from both average martian crust and other volcanic regions of Mars. This unique abundance of Si could lead to a distinct Mg/Si ratio, causing our use of previously observed Mg/Si ratios to be inaccurate.

### 1.2.3 Geologically mapped units

To supplement our chemical and mineralogical interpretations, we also establish the dominant geological properties with mapped geologic units updated recently by Tanaka et al. (2014) (via http://webgis.wr.usgs.gov/pigwad/down/mars_geology.htm#Mars%20Global). To assess the variation of age and rock type between the three chemical provinces, we used an algorithm coded in *Python* (Rossum, 1995) to calculate the distribution of geologic units within the three provinces. Using maps generated by Tanaka et al. (2014), we separated each of the chemical provinces into 5°x5° pixels and calculated areal fractions of geologic units within each pixel. These pixels were then recombined and total province statistics were calculated. Since the study area is near the equator and does not extend greatly toward the poles, we were able to work in a cylindrical projection without significantly distorting the true area of each pixel. Tanaka et al. (2014) utilizes an extensive and detailed categorization system, which we condensed into 6 simpler categories. For each age (Noachian, Hesperian, Amazonian) we refer units with a volcanic designation as "(Age) Volcanic." Units that are primarily non-volcanic in origin are referred to by their age alone.

### 1.2.4 Assessing mineralogy

A large part of the Greater Thaumasia region lacks widespread, unconsolidated layers of globally-derived dust mantles(N. Putzig et al., 2005, fig 5), enabling insight into locally-derived soils and bedrocks from infrared spectral data. Mineralogical characterization of the surface from

infrared spectra complements the bulk chemical characterization, particularly given the vastly different sampling depths of the instruments. Advancing earlier TES-derived mineral areal fraction maps (http://mars.asu.edu/data/), work by Rogers et al. (2007), and by Rogers and Christensen (2007), we analyzed TES-derived mineralogy within each of the three defined provinces 70, 75, and 90 of Greater Thaumasia. Even though such mineralogy may not unambiguously constrain elemental composition (A. Deanne Rogers & Hamilton, 2014; A .D. Rogers & Aharonson, 2008), it has been reported that it can establish whether the surficial mineralogy is representative of the chemical composition at depth (c.f. A. Deanne Rogers & Hamilton, 2014; Wyatt & McSween, H. Y., 2007).

While visible to near infrared (VNIR) observations, such as from CRISM and OMEGA, are well-suited for the detection of a wide variety of minerals, we exclude them for the following reasons. First, VNIR spectral features exist primarily for Fe-bearing minerals, carbonates, and hydrated phases (A Deanne Rogers & Hamilton, 2015). Second, due to non-linear mixing in these wavelengths, extraction of relative abundance from these observations is generally impractical, though minerals with low abundance are easily detected. Both of these factors make it difficult to address the nature of the chemical anomalies within Greater Thaumasia using VNIR spectra.

We seek collocated chemical signatures in TES-derived mineralogical and bulk chemical data that may provide further support for the chemical anomalies in this broad region. For this possibility, we created a mineralogical map that bounded the three chemical province types (i.e., 70, 75, and 90) of Greater Thaumasia, with dust-covered areas excluded as shown in Figure 1.1. The area generally resembles the mineralogic province eight defined by Rogers and Hamilton (2014).

## 1.3 Results

### 1.3.1 Compositional comparisons

Using the modified box-and-whiskers comparison method discussed in section 1.2.1, the three chemical provinces of Greater Thaumasia are compared to the average martian crust. These results are also expressed concisely in Table 1.2, with a graphical representation in Figure 1.2. Si enrichment is observed throughout the three regions, whereas $H_2O$ is depleted in all three regions, though the degree of enrichment or depletion varies. Al content is high in all regions, though only chemical province 75 is enriched beyond our conservative error bars. Both regions 75 and 90 are depleted in K and Th, however K shows less depletion than Th. This yields a high K/Th ratio relative to the average martian crust, which is rarely seen elsewhere on Mars (G. Jeffrey Taylor et al., 2006). Province 70 does not show the K and Th depletions seen in 75 and 90, but it does show a unique Ca value that is lower than crustal average. In addition, Cl is below average in provinces 70 and 75, but approximates crustal values in province 90. All other elements not mentioned are near the average martian crust values and show no sign of significant depletion or enrichment.

Table 1.2. Summarized results from modified box and whiskers analysis. Boxed marked with an E show enrichment, where D show depletion. Empty boxes denote results that are consistent with average crust, showing no significant depletion or enrichment. These results are discussed in section 1.3.1.

| *Province* | *Al* | *Ca* | *Cl* | *Fe* | *$H_2O$* | *K* | *S* | *Si* | *Th* |
|---|---|---|---|---|---|---|---|---|---|
| *Province 70* | | | D | | D | D | | E | |
| *Province 75* | E | | D | | D | D | | E | D |
| *Province 90* | | | | | D | D | | E | D |

The three ratios we examined (Fe/Si, K/Th, and Cl/$H_2O$) also vary between provinces and provide context to our elemental comparisons. Fe/Si is relatively constant between provinces, staying below but consistent with unity. This is expected due to enriched Si content and lower than average Fe content observed throughout Greater Thaumasia. Cl/$H_2O$ is generally high throughout Greater Thaumasia, though it is only enriched in Province 70. This may suggest processes that deplete $H_2O$ without affecting the other volatile elements in Province 70. K/Th is consistent with unity in Province 70, and is enriched in Provinces 75 and 90. This suggests a distinct geological process to create this enrichment in provinces 75 and 90 that is lacking in Province 70.

We assessed the inter-provincial compositional variations relying on the same modified box-and-whiskers comparison method described in Section 1.2.1. This compared each chemical province to the other two provinces within Greater Thaumasia (e.g., 90 vs 70+75) (Figure 1.3). This reveals provinces 75 and 90 to have higher abundances of Al, Ca, and $H_2O$ compared to the rest of Greater Thaumasia, while lower abundances of Al, Ca, and $H_2O$ are observed in province 70. K and Th show the lowest values in provinces 90 and 75. While chemical differences exist between provinces 90 and 70 given the inter-provincial comparisons, we also note the disproportionality in areal coverage between the two provinces. Province 90 is much smaller than province 70, containing only seven 5°x5° pixels, while province 70 contains thirty-five pixels. While this makes the distributional comparisons less compelling, we note that the MADs (errors, see Section 1.2.1) for each province are typically comparable to the size of their distribution (Figures 1.2, 1.3). This suggests that, despite the disparity in size, the distributions in 90 are comparable to those in 70 in terms of their variance and how representative the median is of the dataset. We therefore expect the qualitative implications to remain valid.

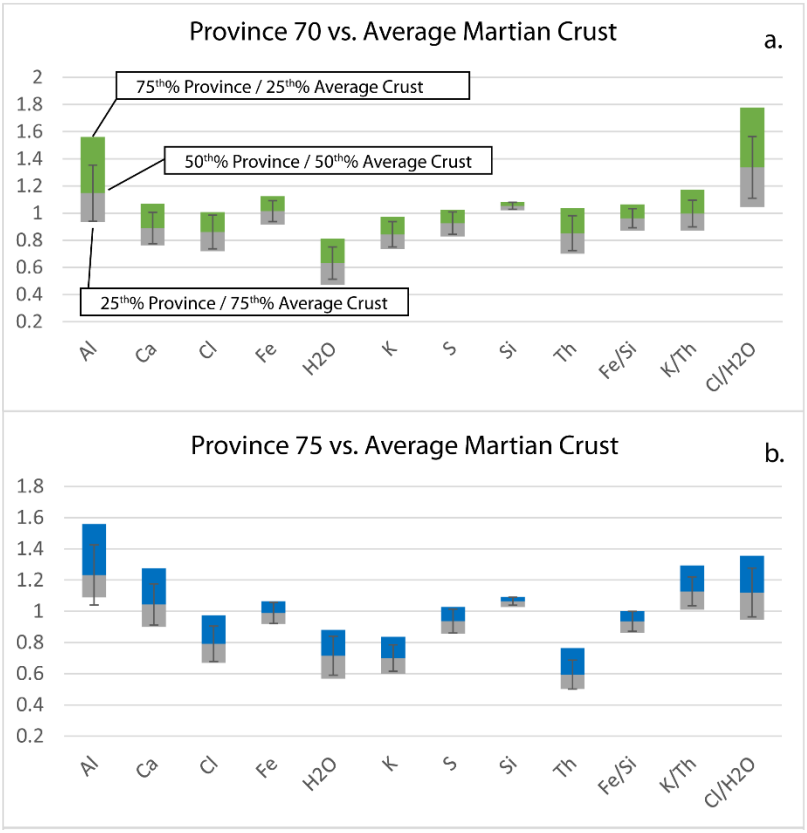
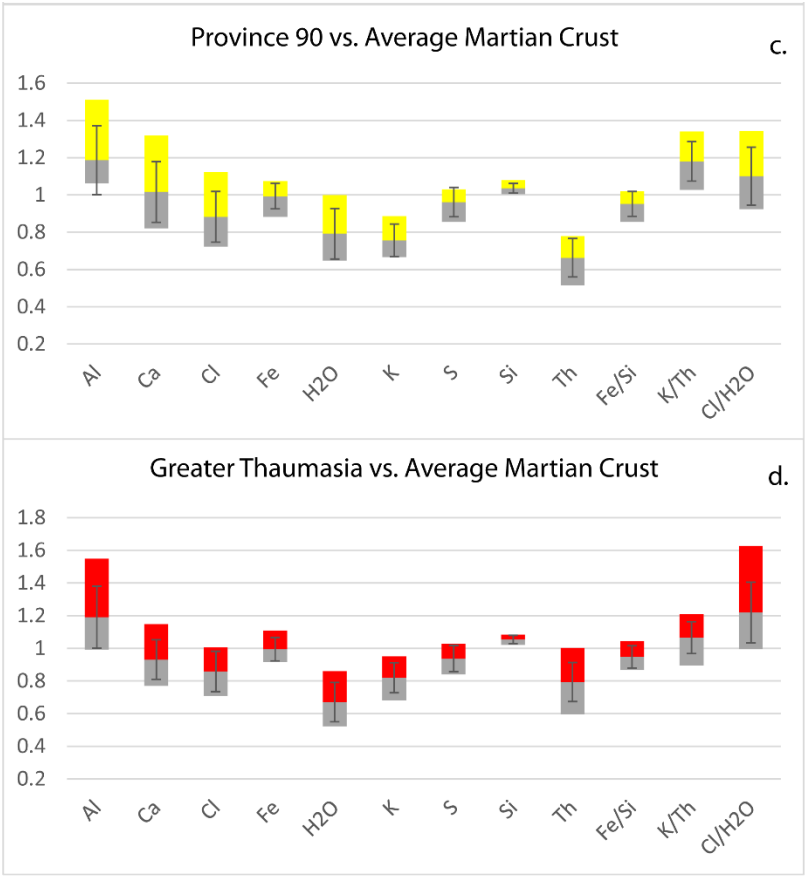



Figure 1.2. Modified box-and-whiskers diagram to compare the distribution of elements in (a) province 70, (b) province 75, (c) province 90, and (d) all of the Greater Thaumasia provinces, to the global geochemistry of the average martian crust. For the purpose of this work, the martian crustal composition is quantified as an average of mid-latitudinal GRS pixels outside our study area: chemical provinces 70, 75, and 90 over Greater Thaumasia. We designate this the "average martian crust". The 75$^{th}$ percentile/25$^{th}$ percentile ratio is represented by the top of the upper box, the 25$^{th}$/75$^{th}$ percentile ratio by the bottom of the lower box, and the ratio of the medians by the boundary between them. The whiskers here represent the uncertainty on the ratio of medians using the Median Absolute Deviation (MAD), see section 1.2.1 for further detail. When a regional median ratio falls above or below unity on the y-axis by at least 1 sigma, the province is enriched or depleted in that element with respect to the martian crust; a graph lacking any overlap with unity will indicate a decisive distinctness from the martian crust, even for small relative differences.

(figure cont'd.)




For the obtained mineralogy derived from infrared spectra, we converted the mineral abundance of the chemical provinces into mass fractions of elemental oxides and elemental abundances using the method of Hamilton and Christensen (2000). This facilitated comparing mineralogical and chemical data of provinces 70, 75, and 90 of Greater Thaumasia. Mineral abundances are shown in Figure 1.4 and are found to generally agree with global averages within each province. Presence of amorphous silica may suggest some degree of aqueous alteration, though 5-6% abundance of olivine suggests otherwise.

From our modified box-and-whisker comparisons (Figures 1.2, 1.3) we observe that mineralogical observations, when compared to chemical observations, show high variance in elemental and oxide distributions within the chemical provinces. While some of this variance may be attributed to systematic uncertainties related to processing techniques (e.g. sensitivity to choice of library spectra), much of this variance is likely related to the greater compositional heterogeneity at local (~3 km) scales relative to the compositions observed at regional (hundreds of km) scales. The resulting contrast with bulk chemical data (Figure 1.5) may reflect the greater compositional variability at local, shallow scales on Mars, despite regionally subdued chemical contrast. Table 1.3 provides a quantitative comparison of chemical variance across the two data sets, which makes the distinction clearer. For example, variance relative to the average computed by mineralogical observations, such as of FeO or $K_2O$, is more than three times greater than the

variance ratios computed for chemically derived mass fractions. In addition, the TES-derived modified box-and-whisker comparisons (Figure 1.6) do not show the same inter-provincial trends observed in GRS-derived analyses. While it may be that the variance of the data is obscuring the trends beyond recognition, it is also possible that the regional trends observed by GRS are not present in the TES observations. This strongly points to the influence of regionally distributed soils or weathering products obscuring changes in the underlying bedrock.

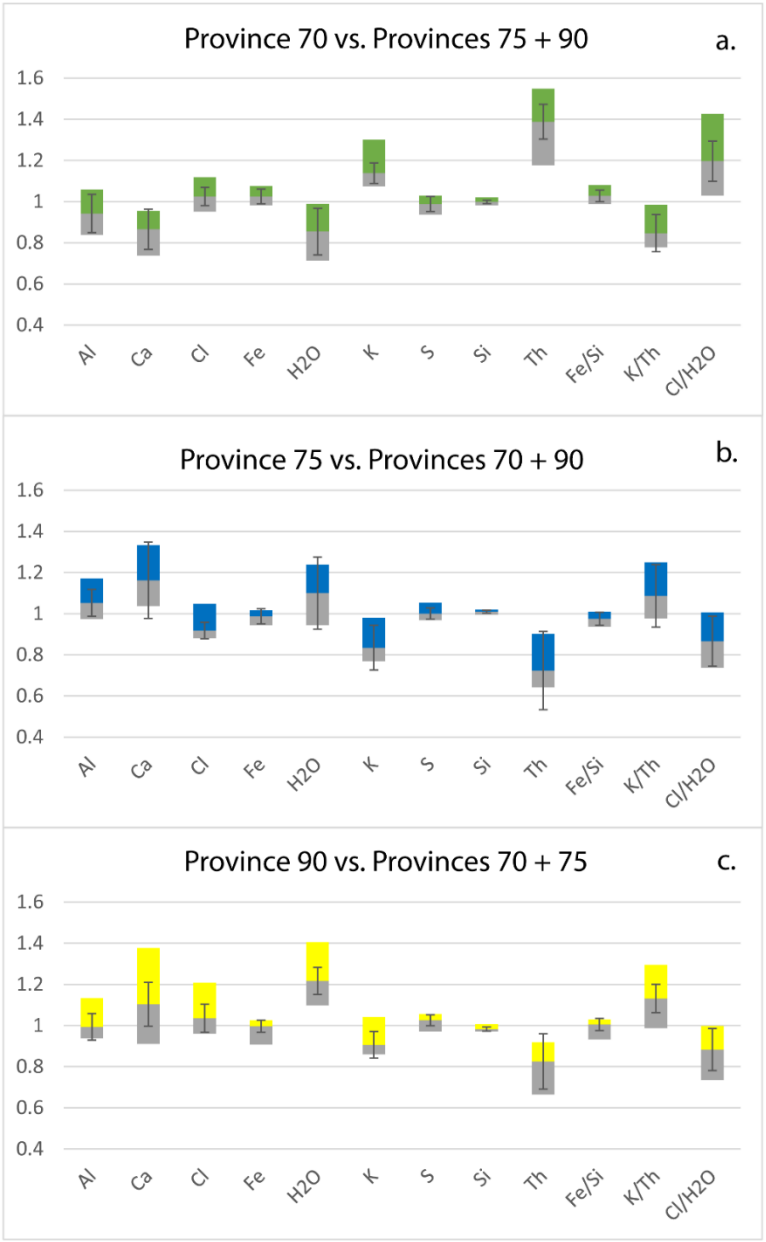



Figure 1.3. Modified box-and-whiskers diagrams that compare the distribution of elements on an inter-provincial basis. This divides the elemental distribution for one of the three chemical provinces of Greater Thaumasia (e.g. province 90) by the combined distribution of elements within the remaining two chemical provinces of Greater Thaumasia (e.g. province 75 and 70). See Figure 1.2 and section 1.2.1 for discussion of our modified box-and-whiskers comparison method.

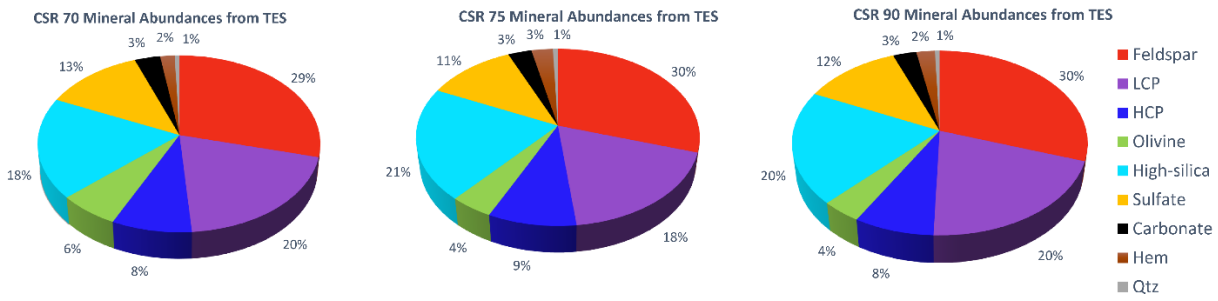



Figure 1.4. TES-derived mineral abundances for Greater Thaumasia. The mineral assemblages of Greater Thaumasia are comparable to global mineral averages (Table 2, A. Deanne Rogers & Hamilton, 2014). The presence of olivine supports evidence shown from ternary diagrams (Figs. 9-11) that suggest minimally altered basalt. Low Ca pyroxenes (LCP) include minerals such as pigeonite, enstatite, and bronzite. High Ca pyroxenes (HCP) represent diopside and augite."High-Silica" refers to zeolites or to poorly-crystalline to amorphous silicate or aluminosillicate phases with a Si/O ratio >~.35, such as volcanic glass, secondary silica, opal, or palagonite (Michalski et al., 2005; A. Deanne Rogers & Hamilton, 2014). These phases are spectrally similar to one another in the thermal infrared range.

Table 1.3. Computed regional mass variance ratios (as standard deviation in mass fractions divided by mean mass fraction) converted from mineral abundances through CIPW normalization are computed for GRS-derived chemistry. The variance ratios indicate that at regional scales the variance of TES data is much greater than the variance shown in GRS data.

| **TES of Provinces 70,75,90** | **SiO2** | **Al2O3** | **FeO** | **CaO** | **K2O** | **SO3** | **H2O** |
|---|---|---|---|---|---|---|---|
| Average (wt%) | 51.04 | 11.54 | 15.74 | 10.08 | 0.70 | 6.50 | 3.81 |
| Standard deviation (wt%) | 3.17 | 1.69 | 3.69 | 1.86 | 0.32 | 2.29 | 0.98 |
| TES Variance Ratio | 0.06 | 0.15 | 0.23 | 0.19 | 0.46 | 0.35 | 0.26 |
| | | | | | | | |
| **GRS of Provinces 70,75,90** | **SiO2** | **Al2O3** | **FeO** | **CaO** | **K2O** | **SO3** | **H2O** |
| Average (wt%) | 46.09 | 8.65 | 18.79 | 7.83 | 0.31 | 5.16 | 2.59 |
| Standard deviation (wt%) | 0.44 | 0.83 | 0.80 | 0.88 | 0.03 | 0.35 | 0.36 |
| GRS Variance Ratio | 0.01 | 0.10 | 0.04 | 0.11 | 0.09 | 0.07 | 0.14 |

### 1.3.2 Trends of Total Alkali Silica diagram

As an important step of compositional analysis, we generate a Total Alkali Silica (TAS) diagram (LeMaitre et al., 2002) to reveal trends from igneous rock compositions that the TES-derived mineralogy and GRS-derived chemistry of Greater Thaumasia represent. While TAS diagrams are primarily reserved for comparison of igneous samples, previous work has shown that useful and informative comparisons can be made between remotely sensed data and in situ data using a TAS diagram (Agee et al., 2013; H. Y. J. McSween et al., 2009). Both the mineralogical and chemical data for the TAS diagram are plotted on a volatile-free basis by removing Cl, H, and S and renormalizing to unity. While limited to only nine elements using

GRS chemical data, converting elemental abundance to mass fractions of oxides allows us to assess regional trends across Greater Thaumasia in relation to the bulk martian chemistry. For the TES-derived analysis, mineral abundance is converted to oxide mass fractions within the assumptions of CIPW normalization. Plotting mineralogical and chemically derived mass fractions on a TAS diagram allows us to compare the data to that compiled through *Spirit*- and *Opportunity*-based Alpha Particle X-Ray Spectrometer (APXS) observations, which were obtained from the MER Analyst Notebook (http://an.rsl.wustl.edu/mer/), as well as martian meteorite data reported from the martian meteorite compendium (http://curator.jsc.nasa.gov/antmet/mmc/) by Meyer (2003).

As we note earlier, the greater bulk-composition sensitivity of the GRS instrument may also help to highlight regional compositional trends distinct from local heterogeneity. Chemical provinces 70, 75, and 90 in Greater Thaumasia plot as outliers of the martian crust in the TAS diagram, showing enrichment of $SiO_2$ and depletion of alkalis overall (Figure 1.5). The majority of MER-obtained APXS data of the traversed landscapes of Meridiani and Gusev plot as outliers to the chemically derived average martian crustal composition indicating greater amounts of alkalis and less $SiO_2$ compared to the Greater Thaumasia chemical provinces. In general, martian meteorites do not represent the bulk martian crust (H. Y. J. McSween et al., 2009) as evidenced by figure 1.5, where most martian meteorites do not overlap with the GRS data.. However, the composition of some martian shergottites, such as NWA 5298 and Dhofar 378, plot within the chemical provinces on the TAS diagram. While the recently characterized NWA 7034 (Agee et al., 2013) falls within the bounds of the GRS data, it does not fall near our provinces, which are more enriched in Si and lower in alkali content. Despite these outliers, the general contrast of martian meteorites with remote and in situ crustal analyses, which are broadly consistent with one another, led us to adopt comparison to in situ measurements over meteorite-derived compositions.

TES-derived mineralogical observations of Greater Thaumasia differ from the work by McSween et al (2009, fig. 1), which presents global TES data. Our analysis indicates that the chemical provinces contain a wider range of $SiO_2$ than predicted by GRS observations, varying from picrobasalt to basaltic andesite (Figure 1.5). In general, TAS values derived from mineralogical data across the pixels reinforce our observation from box-and-whiskers comparisons that these data show greater variance than chemical data and do not reflect the same inter-provincial trends. However, the center of the TES-derived data overlaps with the GRS-derived values for Greater Thaumasia, showing that these datasets are not entirely unrelated. We can attribute the lower variance ratio of chemical data (Table 1.3) to the presumably subdued regional-scale variations in elements relative to local variability as on Earth. The lack of coherent distributional shifts from one chemical province of Thaumasia to another in TES-derived mineralogy may reflect the greater local variance in mineralogy, such as that seen in Gusev crater (H. Y. McSween, Ruff, et al., 2006; H. Y. McSween, Wyatt, et al., 2006). Dramatic sampling depth differences between GRS (10s of cm) and TES (several microns) could also contribute to their apparent divergence. The shallower TES-sampling depths, high compositional variability due to weathering, alteration, and possible regional distribution of soils may conceal provincial trends that occur in the bulk regolith observed by GRS.

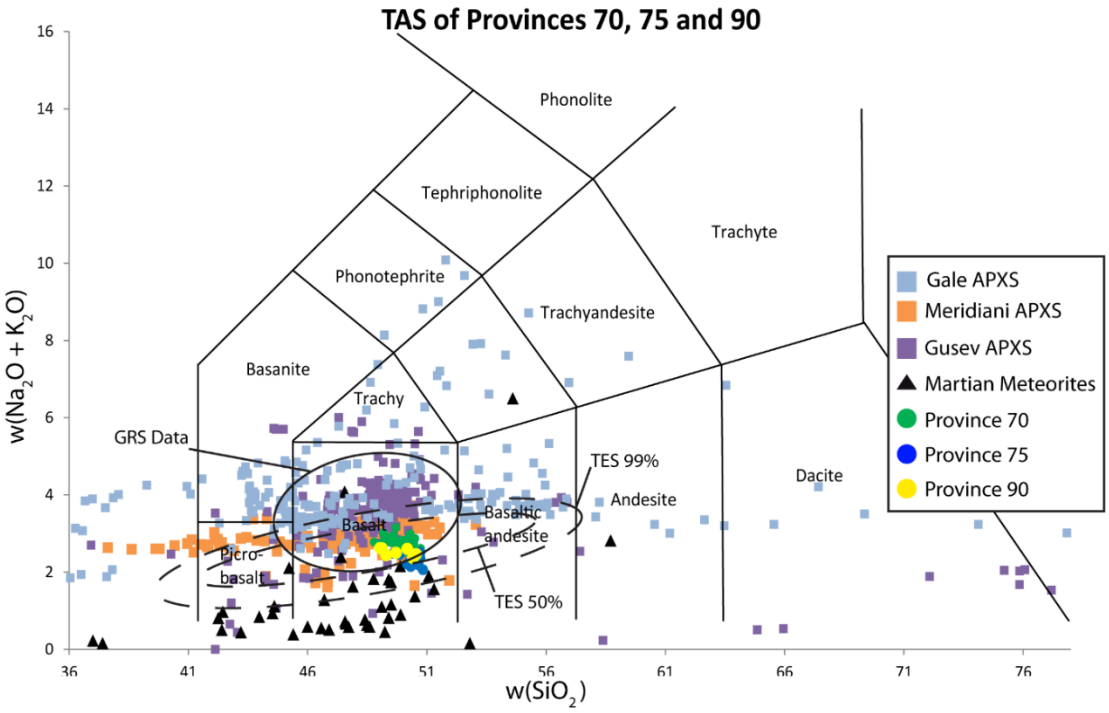



Figure 1.5. Total Alkali Silica (TAS) diagram for igneous rocks (LeMaitre et al., 2002) for rock type classification purposes. Plotted are the APXS observations from Gale, Gusev, and Meridiani (Curiosity, Spirit, and Opportunity rovers), and martian meteorite compositions compiled from the Martian Meteorite Compendium (see acknowledgments for link). GRS-derived elemental chemistry is represented by the solid ellipse marked "GRS Data". GRS-derived chemistry has been converted to oxide mass fraction percentages (e.g., w(SiO2)), with $Na_2O$ values derived using a Na/K ratio (See section 1.2.2 for detailed discussion).With respect to the global gamma-derived chemistry of Mars, the chemical provinces of Greater Thaumasia plot as a chemical outlier in that they are $SiO_2$ enriched and among the most alkali-poor regions observed by GRS. TES-derived oxide concentrations for the Greater Thaumasia region are plotted as the dashed ellipse, 50% of the data lies inside the inner ellipse, while 99% is contained by the outer ellipse. Despite the spread in the TES data, the provinces overlap convincingly with the center of the TES ellipse.

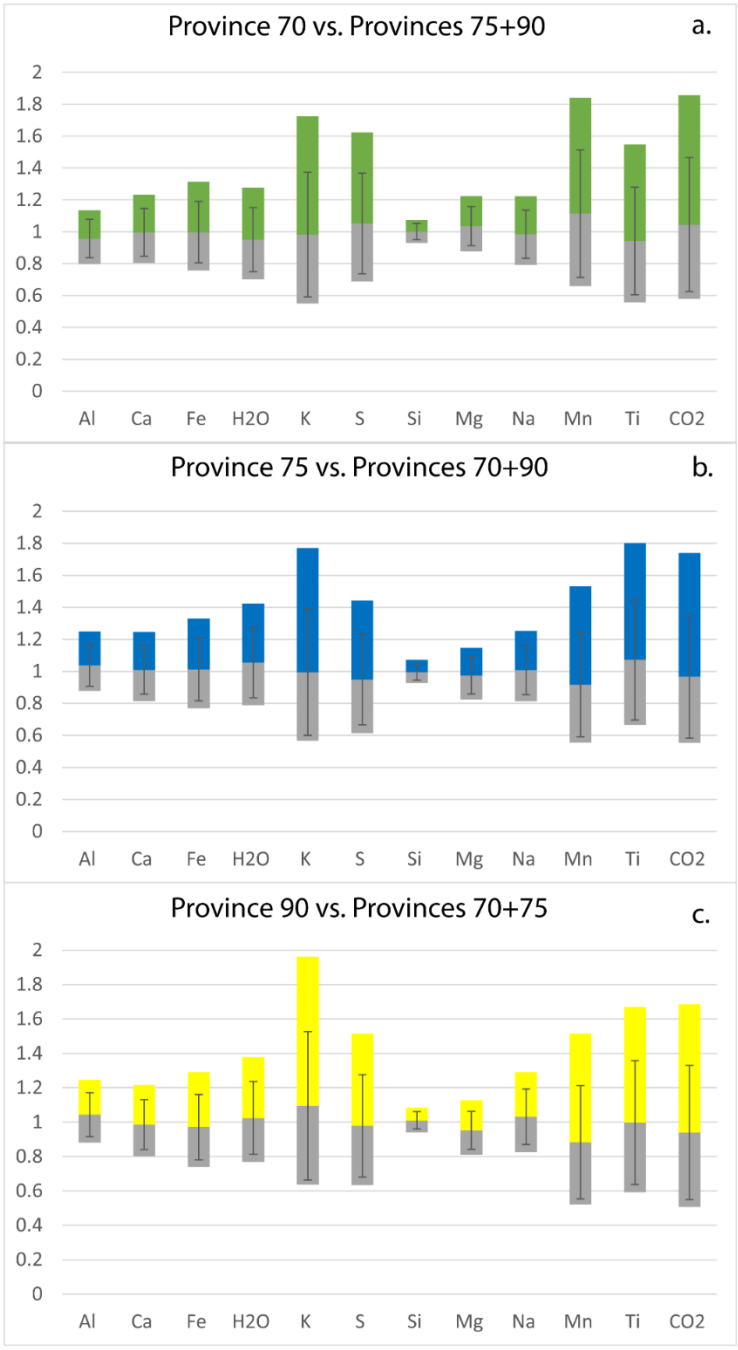



Figure 1.6. Modified box-and-whisker diagram of elemental mass fractions equivalent to mineralogical abundances derived from infrared spectra as observed by TES. Compared to the previous modified box-and-whiskers diagrams for gamma-derived comparisons, the infrared-derived comparison shows a greater variation of elemental distributions at regional scales. This makes inter-regional trends difficult to discern. See Figure 1.2 and section 1.2.1 for discussion of our modified box-and-whiskers comparison method.

### 1.3.3 Distribution of geologic units

We use the mapped geology by Tanaka (2014) to characterize the age and geologic units across the three chemical provinces. These maps are highly descriptive, providing extensive information on each unit. However, we primarily focus on relative age and provenance. As shown in Figure 1.7, the three chemical provinces show a clear age progression of younger (mainly Hesperian) terrain in provinces 75 and 90, to older (primarily Noachian) terrain in province 70. As noted above, chemical provinces 75 and 90 are dominated by Hesperian-aged volcanics related to the development of Syria Planum, which may include fissure-fed and centralized volcanism (James M. Dohm, Tanaka, et al., 2001; Tanaka et al., 2014). However province 90 contains many Noachian units, as the province includes portions of Thaumasia Planum. The geology of Thaumasia Planum is considered related to the development of the shield complex of Syria Planum (i.e., the southeastern flank of Syria Planum (James M. Dohm, Tanaka, et al., 2001)), although older bedrock shed from the Thaumasia Highlands and Coprates Rise mountain ranges likely also contribute to the spectroscopic-based signatures (e.g., possible mixing among the lava flows and mountain outcrops from chemical and mechanical weathering). Chemical province 70 is largely Noachian in age, with a geologic record that includes Noachian development of Syria Planum (north-central portion) as well as mountain building comprised of crustal materials mapped as middle Noachian highland (mNh) and Noachian Highland undivided (Nhu) (Thaumasia Highlands and Coprates Rise). Noachian terrains are also found to the east, southeast, and south of the mountain ranges, which have been highly modified by the Argyre impact event, estimated to be ~ 3.93 Ga (Robbins et al., 2013), with subsequent surface modification by eolian, aqueous, and gravity-driven processes (J. M. Dohm, Spagnuolo, et al., 2015). This includes degradation of the mountain ranges and related sediment transfer into the transition zone within chemical province 70 (area between Thaumasia plateau and Argyre, including the Thaumasia Highlands and Coprates rise along its southern and eastern margins, respectively (James M. Dohm, Tanaka, et al., 2001)). However, we note that three Hesperian geologic units as defined by Scott and Tanaka (1986) (Hr, Hsu, and Hsl) constitute more than 90% of the K and Th depleted chemical provinces 75 and 90. Consistent with the K and Th depletion signature, the work by Hahn et al. (2011), which also utilizes crustal thickness models, shows that Thaumasia contains anomalously low crustal thermal energy production, perhaps indicating a regionally low crustal thermal flux. Figure 1.7 shows the results of our analysis, classifying each chemical province into distinct mapped geologic groups.

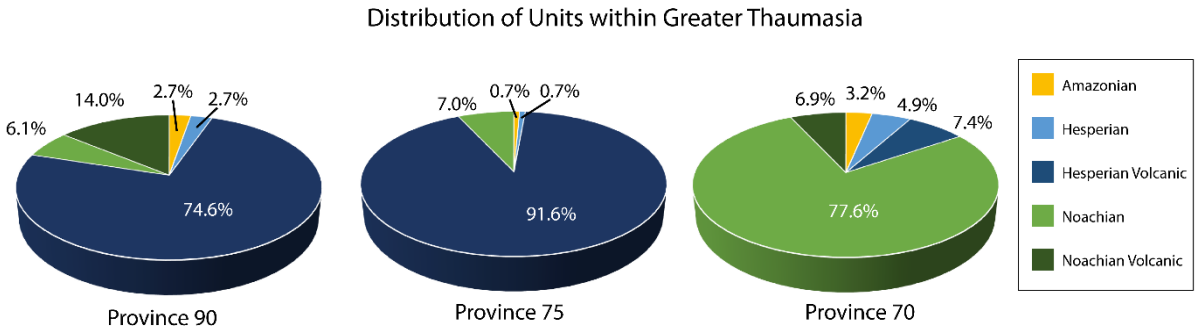



Figure 1.7. The pie charts above display the distribution of geologic age units by means of percentage for each chemical province in Greater Thaumasia based on geologic maps from Tanaka et al.(2014). The specific geologic classifications from Tanaka et al. (2014)have been simplified by grouping all non-volcanic type units of the same age together (e.g. Noachian, Hesperian, Amazonian) while also grouping all volcanic type units of the same age together as well (e.g. Noachian Volcanic, Hesperian Volcanic). The geologic ages derived from crater counting indicate Hesperian aged volcanic rock to dominate the western provinces of Great Thaumasia (i.e. province 90 and 75), while Noachian aged non-volcanic rocks appear to dominate the eastern province of Greater Thaumasia.

### 1.3.4 Ternaries

To gain further insight into the chemical composition of Greater Thaumasia and regional chemical processes, we use three molar-fraction based ternary diagrams. Through creating ternary diagrams, we assess evidence for regional-chemical processes as well as aqueous-alteration behavior of martian igneous rocks. Using Haneberg's (2004) computational geoscience package within Wolfram Mathematica 10.0, we produce geochemical ternary diagrams and plot GRS-derived mass fractions, martian meteorites, and in situ chemical composition information to provide qualitative insight complementing the compositional comparisons (Sections 1.3.1 and 2.3.4). These analyses allow us to assess the chemical consistency of this area with geomorphic (Montgomery et al., 2009) and chemical (Balta & McSween, 2013; D. Baratoux et al., 2013) models. This region contains a compositional mix consisting of the ancient, highly weathered (e.g., weathering rinds), mantled (e.g., aeolian, alluvial, fluvial, colluvial deposits), and degraded rocks that compose the Thaumasia Highlands, Claritas Rise and Coprates Rise mountain ranges, and other extremely ancient geologic provinces of Mars. These include Early Noachian units up to 3.96 Ga old, as displayed in Table 1 by Tanaka et al. (2014).

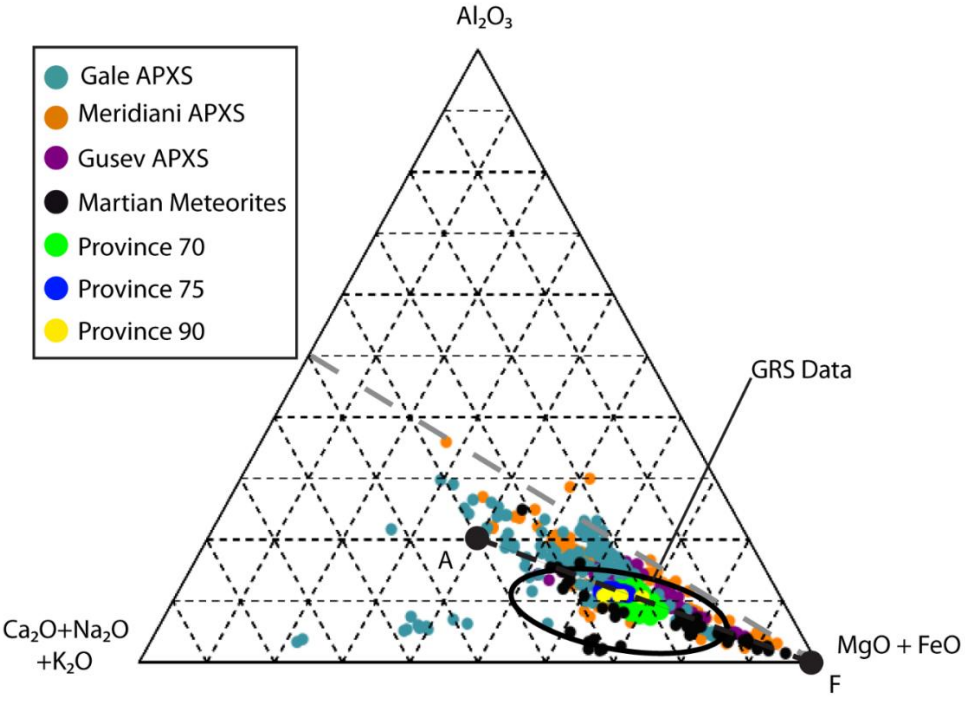



Figure 1.8. ($CaO+Na_2O+K_2O$) - ($Al_2O_3$) - (MgO+FeO) (CNK-A-MF) Ternary Diagram. Adapted from Hurowitz and McLennan (2007), the dashed gray line represents the olivine-feldspar join. Compositions that plot below the line are considered to be igneous rocks that are not weathered under moderate pH conditions. The black dashed line represents subaqueous alteration trends at low-pH, similar to those observed on Mars (Joel A. Hurowitz & McLennan, 2007) .Basalts weathered under these conditions trend toward the CNK-A join (A) and alteration fluids towards the MF apex (F). Gamma data across the three provinces do not suggest aqueous alteration trends commonly observed on Earth. The GRS data is represented by the black ellipse, see Section 1.2.2 for further detail on calculating $Na_2O$ and MgO values from GRS data.

In order to understand the aqueous alteration of olivine-bearing basalts, such as the ones that are generally expected on Mars, we plot the first ternary diagram (Figure 1.8) with apices of ($CaO+Na_2O+K_2O$) – ($Al_2O_3$) – (MgO+FeO), denoted CNK-A-MF. Similar to work by Nesbitt and Wilson (1992), which characterizes low pH aqueous alteration conditions of igneous rocks in terrestrial settings. We illustrate the feldspar-olivine join as a line that extends from the MF apex to the midpoint between the A and CNK apices. Igneous rocks altered under low-pH aqueous conditions plot below the feldspar-olivine join and trend away from the MF apex to a rock composition containing more $Al_2O_3$ from its initial unaltered state (Joel A. Hurowitz & McLennan, 2007; Nesbit & Wilson, 1992). As shown in figure 1.8, both the average martian crust and our study area plot beneath the Ol-Fsp join, which is typical for martian basalts (Joel A. Hurowitz & McLennan, 2007). In addition, they fall along the experimental low pH path examined by Hurowitz and McLennan (2007) and do not appear substantially altered along this path.

The second ternary (Figure 1.9) uses apices of MgO, $SO_3$, and FeO, (M-S-F) to describe the distribution of evaporitic or sulfate minerals (i.e., gypsum and anhydrite (McLennan, 2000)) produced by subaqueous sedimentary processes, compared to igneous minerals (i.e. FeO and MgO). An igneous rock composition containing greater amounts of $SO_3$ might be expected to have undergone some degree of subaqueous alteration (Suniti Karunatillake et al., 2010; McLennan, 2000). In Figure 1.9, we see that while the study region does have some presence of $SO_3$, it falls within, and even on the unaltered end of, the GRS data. This suggests that subaqueous alteration and evaporitic deposition have played a minimal role in the history of Greater Thaumasia.

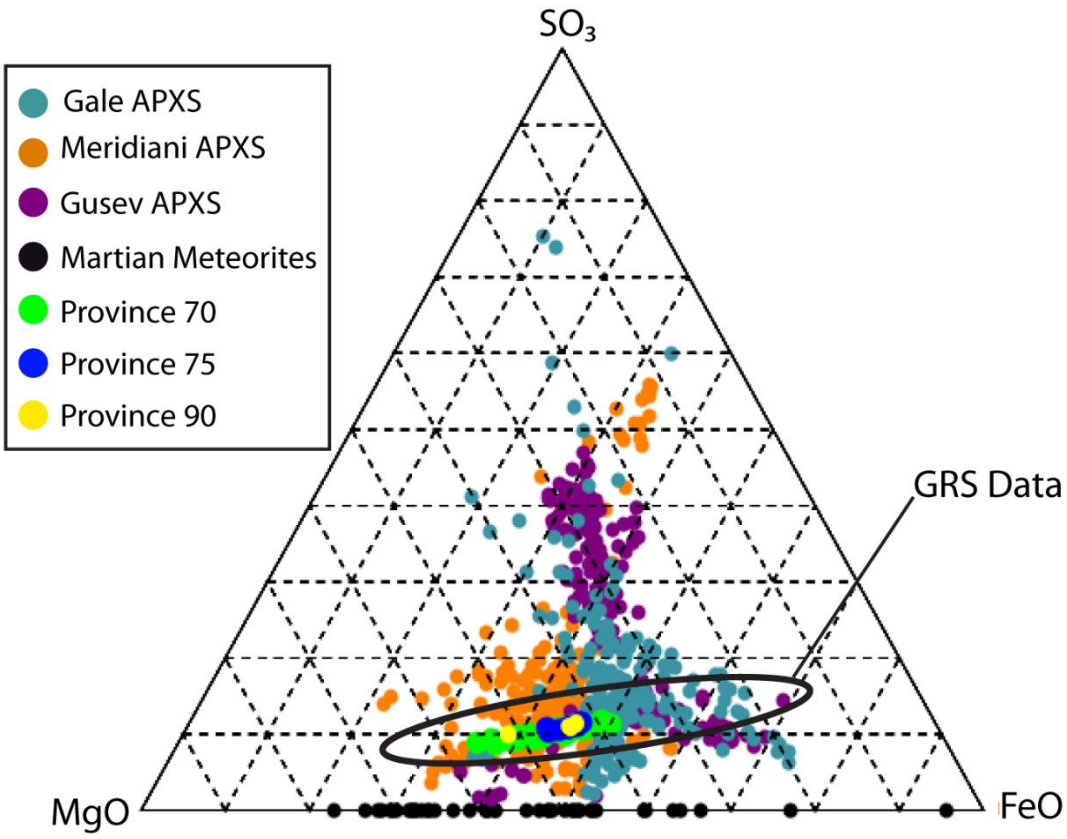



Figure 1.9. (MgO) – ($SO_3$) – (FeO) Ternary Diagram. Adopted after McLennan (2000), the composition that plots closer to the FeO-MgO join represents igneous compositions that are less altered by subaqueous conditions and contain less evaporitic minerals. The lower $SO_3$ in the chemical provinces of Greater Thaumasia (green, blue and yellow) compared to several MER in situ sampling and the global martian surface chemistry could indicate a minimal presence of evaporitic or alteration salts on the surface in Greater Thaumasia at regional scales. GRS data is represented by the black ellipse, see section 1.2.2 for discussion of calculating MgO values for GRS data.

The third ternary (Figure 1.10) has apices ($Al_2O_3$+$SiO_2$) – (CaO+FeO+MgO) – ($SO_3$), denoted ASi-CFM-S. This accounts for the divalent cations that are mobile under possible aqueous sulfuric alteration conditions on Mars and shows the degree to which the igneous composition has been altered (T. M. McCollom & Hynek, 2007; Thomas M. McCollom & Hynek, 2005). Highly altered samples are expected to plot closer to the $SO_3$ apex and away from CFM apex, whereas pristine igneous material will plot along the ASi-CFM join. Our study area falls along a join between unaltered basalts (martian meteorites) and pure $SO_3$, which is consistent with the rest of the GRS data (Figure 1.10). Samples falling along this line have been interpreted to represent mixing between pure endmembers of martian basalts and sulfate-rich material, rather than aqueous alteration of basalts to sulfates (Thomas M. McCollom & Hynek, 2005). Considering the trends in the ASi-CFM-S and M-S-F ternaries, these analyses do not suggest the presence of significant amounts of evaporitic material. In addition, what sulfates may exist regionally may derive from a process that mixed sulfate rich material with local regolith, such as explosive volcanism (T. M. McCollom & Hynek, 2007) or aeolian mixing, rather than subaqueous alteration.

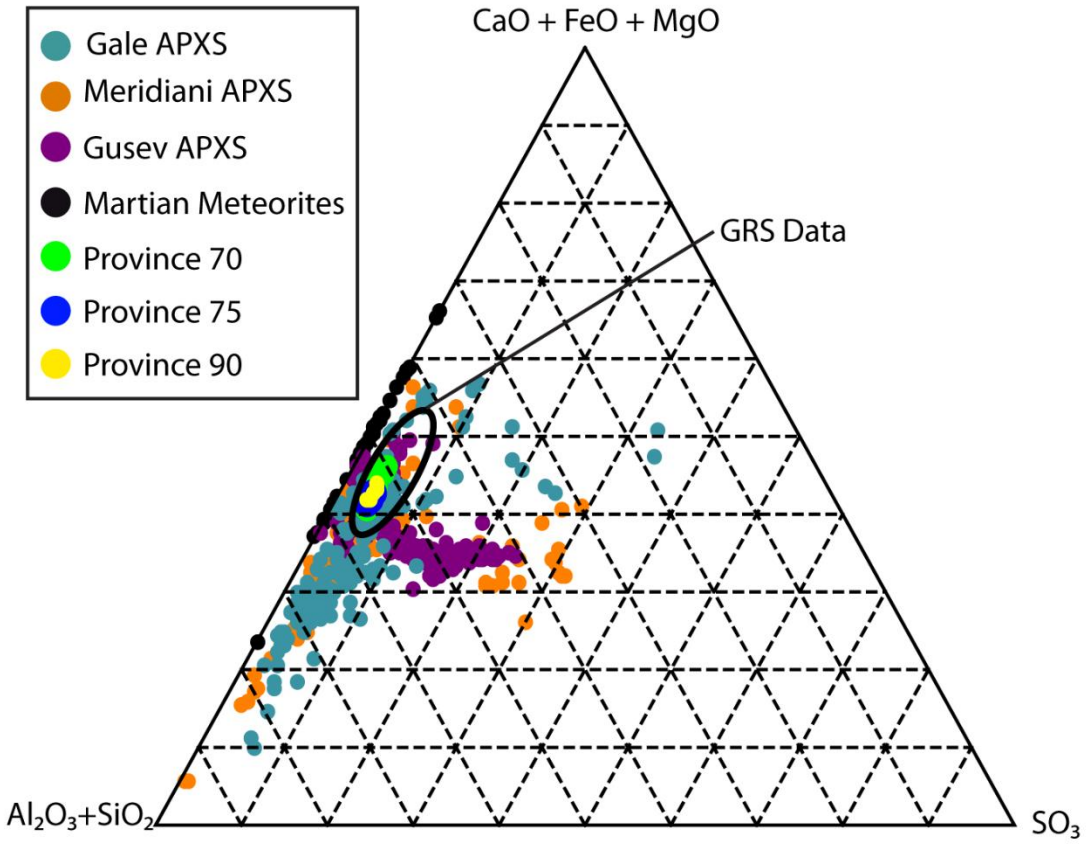



Figure 1.10. ($Al_2O_3$+$SiO_2$) – (CaO+FeO+MgO) – ($SO_3$) (ASi-CFM-S) Ternary Diagram. Adopted after McCollom and Hynek (2005) to account for major igneous elements, divalent cations that are mobile under aqueous conditions, and for the potential of sulfates in martian aqueous conditions. GRS data is represented by the black ellipse, see section 1.2.2 for discussion of calculating MgO values for GRS data.

## 1.4 Discussion

Grounded in chemistry, we defined Greater Thaumasia as the total extent of three chemical provinces (i.e. 70, 75, and 90) as first observed in the work by Karunatillake in (2009). These chemical provinces overlap Sinai, Solis, and Thaumasia Planae, the Coprates Rise mountain range, and the eastern part of the Thaumasia Highlands mountain range. This includes the transition zone linking the Argyre basin with the Thaumasia Highlands and Coprates Rise mountain ranges. Greater Thaumasia varies in a compositionally compelling manner trending to the southeast outwards away from the complex shield volcanoes through to the Argyre basin. Four chemical observations highlight the results of our investigation across provinces 70, 75, and 90 within Greater Thaumasia. First, Si is enriched and Al is above average throughout Greater Thaumasia relative to the martian bulk crust. Second, Ca approximates global crustal values in provinces 75 and 90 within Greater Thaumasia, with below average abundances in province 70. Lastly, the $H_2O$ depletion in 90 increases across 75 to 70. Inter-provincial comparisons across 70, 75, and 90 provide additional insight; 75 and 90 have the highest Al content compared to province 70. While 90 and 75 have higher values in Ca, 70 shows depletion. The trend reverses for Th: 90 and 75 are depleted while 70 is enriched. Our modified box-and-whiskers diagrams (Figures 1.2, 1.4 and Table 1.2) and related discussion in Section 1.3.1 present the underlying details.

Of the ratios examined in the modified box-and-whisker diagrams, K/Th showed significant variation between provinces and has direct bearing on possible regional aqueous processes. Work by Taylor et al. in (2006) thoroughly explored variations in K/Th across Mars, and noted the Greater Thaumasia region (Taylor terms this “Valles Marineris region”) as a region of high K/Th. In our observations, this high K/Th is restricted to provinces 75 and 90, which are dominated by Hesperian lava flows, likely sourced from Syria Planum. Since this shift in K/Th occurs in the transition into Syria-sourced lavas, this may suggest that igneous fractionation is responsible for the enriched ratio rather than aqueous processes. It remains possible, however, that aqueous processes could still play a role, though it is unclear to what degree these processes affect the regional K/Th ratio.

Several hypotheses have been proposed to explain the geologic history of the Thaumasia and surrounding regions, including Greater Thaumasia. We summarize comparisons of our results with these models in Table 1.4. It was suggested (Montgomery et al., 2009) that a continent scale landslide occurred, which formed the modern Thaumasia Planum and Thaumasia Highlands, and was related to the formation of Valles Marineris and its outflow channels. This model was proposed to fit the observed deformation patterns within and around the Thaumasia Plateau, as well as provide a unified explanation for Valles Marineris and surrounding geomorphology. This model called on buried ice or salt to lubricate the landslide, which otherwise could not occur due to low regional slopes, and to provide a water source for the Valles Marineris outflow channels.

Table 1.4. Summarized expectations from and comparisons to models that may inform the origin of the Greater Thaumasia region. This is discussed in more detail in the discussion section (section 1.2.4)

| Evolution model | Chemical expectations from presented models | Comparisons with this work |
|---|---|---|
| Montgomery et al. 2009, *Continental salt tectonics on Mars and the origin of Valles Marineris and associated outflow channels* | Presence of subsurface salt may lead to diapir or namakir-like features, especially where faults reach the surface<br><br>Regional aqueous alteration likely, either from groundwater or ancient salt-depositing water body | Evidence for exposed (TES) or buried (GRS) salt bodies Is not found. In addition, evidence for regional aqueous alteration (Figures 8, 9) is lacking. Though both these signatures could be buried beyond instrument sensitivity depths<br><br>The older age of the Thaumasia Highlands compared to Syria Planum is difficult to reconcile with Syria-sourced slide |
| Baratoux et al 2013, *The petrological expression of early Mars volcanism* | Sharp shift from low Ca to high Ca pyroxenes at Noachian-Hesperian boundary | Noachian dominated chemical province 70 shows lower Ca abundance than Hesperian dominated provinces 75 and 90, consistent with model expectations |
| Balta et al. 2013, *Water and the composition of Martian magmas* | High $SiO_2$, high $H_2O$, low Th magmas in the Noachian period, progressing into low $SiO_2$, low $H_2O$, high Th magmas in the Amazonian | The shift from province 70 (Noachian dominated) to 75 (Hesperian dominated) is marked by a substantial increase in $H_2O$ and a decrease in Th, contrary to predictions<br><br>Changes in magma Th and $H_2O$ content through time supported by our analysis |

If salt-lubricated landslide(s) formed the Thaumasia Plateau (Montgomery et al., 2009), salt diapirism could occur over geological time, potentially producing an abundance of namakir-type deposits (Jackson et al., 1994; J. Li et al., 2010). Salt diapirs on Earth are observed to rise at speeds near 2 km/Ma (Jackson et al., 1994) and occur in both extensional and contractional settings. Given well over one billion years to rise, in addition to abundant pathways to the surface via landslide-associated faulting, the presence of namakirs or salt diapirs on the surface or in the shallow subsurface of Thaumasia Planum is expected. However, we do not find compositional evidence for such formations. GRS investigations show that all three regions have Cl and S abundances that are below crustal averages, which does not suggest abundant surface or shallow-buried salt bodies. TES-based analysis did detect sulfates, however at a level that it typical for the martian crust, and does not suggest an abundance of exposed salts. In addition,

analysis of ternary diagrams (section 1.3.4) suggest that the soils observed by GRS were not altered under aqueous conditions, and that whatever sulfates are regionally present were not likely created as evaporites. It remains possible that salt deposits and evidence for aqueous alteration could be obscured by volcanic-, wind-, or water-driven mantles, as well as by the purported landslide to depths that block both TES-based and GRS-based investigations. However, other geomorphologic observations and interpretations also question the landslide hypothesis. The Early Noachian terrain, including the Thaumasia Highlands, predate the formation of Syria Planum, the putative source region of the landslide. This challenges the idea of a Syria Planum-sourced slide resulting in the Thaumasia Highlands mountain range. However, salt tectonism at the base of the Tharsis lavas may have contributed to the deformational history of the Thaumasia region (Montgomery et al., 2009; Wezel & Baioni, 2010). This may have included the interaction among Tharsis magmatism, a hypothesized regional ancient (i.e. pre to early Tharsis development) Tharsis basin and related aquifer system, and possible evaporite deposits beneath the Tharsis lavas (James M. Dohm, Williams, et al., 2009). With the Thaumasia Highlands as its southern margin and older parts of the Tempe Terra igneous plateau as its northern margin, the proposed basin by Dohm et al. (2001) was a catchment not only for Tharsis lavas, exemplified by the kilometers-thick, layered lava materials exposed in the walls of Valles Marineris, but also pre-lava evaporite deposits.

In contrast to a landslide, the study area serves as a partial regional analog to what Balta and McSween (2013) and Baratoux et al. (2013) modeled globally for the temporal evolution of martian magmatism. Distinct shifts in $H_2O$, K, and Th appear across chemical provinces (Figures 1.2, 1.3), especially between the Noachian-dominated province 70 and the Hesperian-dominated province 75. In general, we observe increasing $H_2O$ and decreasing K, Th abundances moving forward in time, which differs from the Balta and McSween model which predicts decreasing H2O and increasing Th. These inconsistencies may be caused by igneous processes unique to the Greater Thaumasia region, though remains possible that the chemical trends seen in K, Th, and $H_2O$ may be related to sustained aqueous alteration in the region despite the lack of observed signatures of aqueous weathering. Despite inconsistencies with the specifics of compositional trends, our observations suggest the regional-scale existence of global magmatic evolution, reinforcing earlier work (Balta & McSween, 2013; David Baratoux et al., 2011; Hahn et al., 2011). Furthermore, the higher Ca content in the Hesperian chemical provinces of Greater Thaumasia relative to their Noachian counterparts supports the possibility of low Ca, pyroxene-rich rocks in more ancient volcanics as suggested by Baratoux's work in (2013). This is also consistent with IR observations globally, which show low Ca pyroxene in Noachian terrains and high Ca pyroxenes in Hesperian terrains (cf. Mustard et al., 2005; A. D. Rogers & Hamilton, 2014; A. Deanne Rogers & Christensen, 2007; A. Deanne Rogers & Hamilton, 2014).

## 1.5 Conclusion

The Thaumasia region records significant geological activity on Mars, which includes Early/Middle Noachian mountain building related to the formation of the Thaumasia Highlands and Coprates Rise mountain ranges, Noachian/Amazonian Tharsis development, and the Middle Noachian Argyre impact. The region, which we refer to here as the Greater Thaumasia region, therefore likely contains diverse rock materials, ranging from crustal basement (including

mountain-forming and impact-exposed) to younger Tharsis-associated volcanics. Three chemically distinct provinces, defined by Karunatillake et al. (2009), overlap spatially with anomalous areas of other datasets, making them compelling for further investigation. We therefore performed an investigation of the Greater Thaumasia region which included an assessment of the geology and geochemistry (i.e., MGS TES mineralogical and MO GRS elemental). In addition to the rich geologic history, the hypothesized megaslide by (Montgomery et al., 2009) provided the motivation for our investigation.

Our collective observations are inconsistent with a proposed mega-landslide origin, and yet show GRS- and TES-based chemical distinctions when comparing among the chemical provinces, geological features, and stratigraphic ages. In particular, our observations support distinct magmatic processes that may impart the Si-enriched signature of the eastern portion of Greater Thaumasia, which includes the Coprates Rise, part of the Thaumasia Highlands, and the transition zone between the Argyre basin and the Thaumasia Highlands, which overlaps with K and Th depletion of the southeastern flank of Syria Planum. Geological activity, including varying magmatic conditions through time, likely contributed to the distinctions. In addition to magmatism, aqueous activity, which includes weathering and leaching of elements, may have also contributed to the elemental signatures of the Greater Thaumasia region. More geological investigation is necessary to further explore the extent to which the Greater Thaumasia region compares chemically and mineralogically to the other volcanic provinces, such as the Tharsis complex, as well as the more ancient martian basement complex.

# Chapter 2. Contrasting Regional Soil Alteration across the Topographic Dichotomy of Mars

## 2.1 Introduction

The martian topographic dichotomy is one of Mars' oldest features (e.g. Andrews-Hanna et al., 2008), and separates the northern lowlands from the southern highlands. While the dichotomy is likely early Noachian in age (e.g. Frey et al., 2002), the surface ages of these two areas are distinctly different due to a large amount of Hesperian-age material in the northern lowlands (Tanaka et al., 2014). Global assessments of geochemistry also reveal distinctions in the distribution of $H_2O$ and S in the shallow subsurface soils that may associate with the dichotomy boundary (S Karunatillake et al., 2014). In situ results show that correlations and trends among volatile elements (e.g. S, $H_2O$, Cl) can elucidate the amount of alteration (e.g. Haskin et al., 2005), the mineralogy of altered phases (Dehouck et al., 2014), and the former presence and migration of fluids (Haskin et al., 2005). Such trends, identified in situ, are considered in soils using GRS and applied to regional-to-global scale soils to examine potential signals of widespread aqueous interaction and alteration in martian soil. This work explores the hypothesis that regionally distinct aqueous histories may cause the geochemical distinction in volatile chemistry across the dichotomy. Older, more altered soils may occur in the south formed by acidic processes such as hydrothermal leaching , aerosols, and low-pH brine films (J. A. Hurowitz & Fischer, 2014; Thomas M. McCollom & Hynek, 2005) or by more neutral-to-alkaline processes (J. A. Hurowitz et al., 2017). In contrast in the northern lowlands, younger, less altered soils may have been derived primarily from physical reworking of underlying material with limited fluid interaction.

Aqueous interaction can be supported by a variety of indicative minerals (e.g. clays), but remote Infrared observations that identify directly these minerals are limited to the optical surface (<10 micrometers depth) and typically only examine consolidated or lithic material in dust-free areas. These constraints make Gamma-Ray Spectrometer (GRS) chemical data key to examining regional and global-scale trends in soil chemistry. Because GRS generally measures soil chemistry up to ~0.5 meters with continuous coverage, it is possible to identify trends in soils at the global-scale and evaluate the same soil volatile trends seen at rover and lander sites. Prior work examining the latitudinal variability of the $H_2O$:S ratio in martian bulk soil suggested sulfate minerals are the primary hydrated phase in the southern highlands, but that this is not the case in the northern lowlands (Suniti Karunatillake et al., 2016). However, this work focused solely on $H_2O$ and S; the influence of Cl and non-volatile elements were not considered. In addition, recent work (Ojha et al., 2018) has highlighted dust as a reservoir of S and Cl on the martian surface and specific areas of the surface are suggested to have mantles of sufficient thickness to influence GRS observations (Viviano et al., 2019). The inclusion of Cl, non-volatile soil chemistry, and a metric for surface dust adds additional dimensions to build upon and advance, these previous findings.

Here, Principal Component Analysis (PCA) is applied to the GRS chemical maps, and to Thermal Emission Spectrometer (TES)-derived Dust Cover Index (DCI) data to identify chemical correlations across the dichotomy that may indicate fluid interaction at hemispheric to global scales. This work is the first global multivariate analysis of the newest set of GRS chemical maps, and the first such analysis to incorporate Ca, Al, and S maps. This work incorporates nine chemical maps and DCI and examines chemical changes across the geologically meaningful division: the topographic dichotomy.

## 2.2 Data and Methods

### 2.2.1. Datasets

The primary dataset used here is the GRS-derived surface chemistry (referred to as "chemical maps") covering approximately the area from 45° N to 50° S projected at 5°×5° resolution (Figure 2.1). These maps provide chemical abundances for nine elements: The key rock-forming elements Si, Ca, Al, and Fe, the incompatible elements K and Th, and the volatile elements H (presented as stoichiometrically equivalent $H_2O$), S, and Cl. An important caveat of these maps is that Mg and Na , important rock-forming elements used in several metrics of soil maturity and alteration (Thomas M. McCollom & Hynek, 2005), are not included in chemical maps due to technical limitations (Boynton et al., 2007). Some chemical maps are available at higher resolutions and with improved areal coverage (e.g. 2° K and Th maps), but the multivariate analysis is ultimately limited by the coarsest resolution map and by the areas where all nine maps are available (an extent referred to as 'midlatitudinal' in this work). Further details on the chemical maps and their derivation can be found in the Supplement section A 1.1.

The Dust Cover Index (DCI) (Ruff & Christensen, 2002) is a measure of surface dust abundance derived from Thermal Emission Spectrometer (TES) data, specifically the emissivity between 1350-1400 $cm^{-1}$ related to the abundance of fine (<<100 micron) silicate material. DCI is incorporated into analysis to evaluate the correlation between areal coverage of the dust and regional soil chemistry. DCI is available at high lateral resolution (16 pixels per degree) and represents a much shallower depth (10s of microns) compared to the decimeter-scale depth represented by chemical maps. The sampling depth disparity suggests little correlation between DCI and the chemical maps. Nevertheless, DCI was selected over other measures of surface dust (e.g. Thermal Inertia (Putzig et al., 2005)) because it most directly measures the abundance of <100 micron silicate material.

### 2.2.2. Methods

Principal Component Analysis (PCA) applies multicomponent chemistry to derive chemical correlations and is applied to the combined chemical maps and DCI datasets at both global and regional scales following previous applications in the planetary context (e.g. Gasnault et al., 2010). In addition to analyzing the entire spatial dataset (i.e. midlatitudinal extent), twelve regions (i.e. spatial subsets of the data) were analyzed with the same methods (Figure B 1) to capture correlations that are significant at more regional scales. Six of these regions were selected for either high or low abundance of one of the volatile elements ($H_2O$, S, Cl), and two for high and low dust cover. These eight regions are intended to isolate the trends that may be prominent only in statistically distinct regions. Two more regions isolating volcanic terrains and non-volcanic terrains were analyzed to examine trends potentially driven by lithology. Two regions, the northern lowlands and southern highlands (Tanaka et al., 2014) are isolated to examine trends unique to each region and examine changes across the dichotomy boundary that may indicate distinct soil histories.

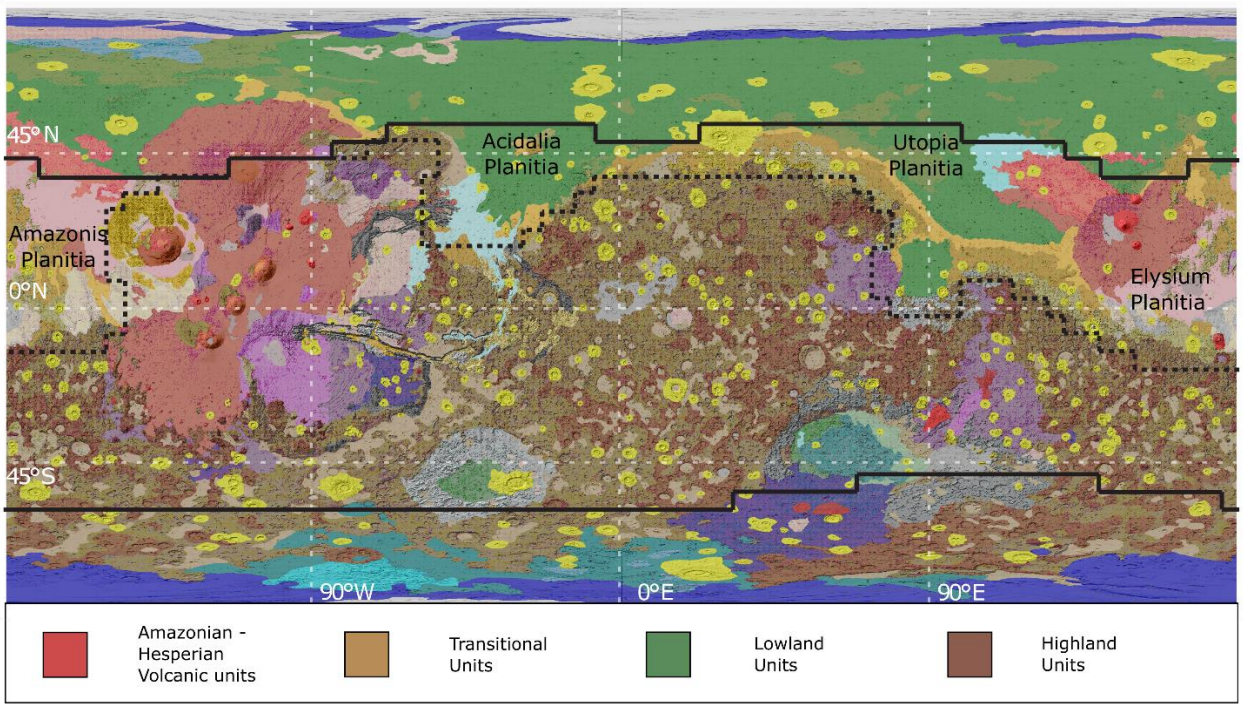



Figure 2.1. Map of the martian surface showing the extent of the 'midlatitudinal' region (solid black lines) for which chemical map data are available overlain on the geologic map of Mars (Tanaka et al., 2014) with several locations labeled. The black dotted line shows the boundary between the northern lowlands and southern highlands regions used in analysis. The simplified legend shows the colors for four of the relevant geologic units: Amazonian and Hesperian volcanism in shades of red, Transitional units in shades of gold, Lowland units in shades of green and Highland units in shades of brown. Details on the spatial extent of GRS data and the derivation of these regions can be found in the supplement.

Results from the northern lowlands region and southern highlands region are discussed in detail as these are the areas of greatest interest (Section 2.3.1). The remaining ten regions provide the broader results in Section 2.3.2 and give context to the specific findings in Section 2.3.1. Further details on region selection and coverage can be found in the supplement, section A.1.4.

Most applications to date fail to incorporate uncertainty into PCA, which is addressed here by normalizing GRS and DCI data using both the standard deviation of the dataset and the uncertainty of individual values. This method reduces the influence of imprecise data and prevents such data from disproportionately contributing to the variance. The results of this normalization are referred to as "t-values" throughout this work, with details in Supplement S 1.3. T-values also simplify the process of selecting the statistically distinct regions discussed earlier. Each region is defined with a mask, selecting ≥2-pixel contiguous areas that are above +1 or below -1 in t-values. This threshold is comparable to one standard deviation above or below average at each pixel, with spatial contiguity conferring a statistical confidence in excess of 95% (Suniti Karunatillake et al., 2009).

The primary results of each analysis are the loadings of the principal components, which describe the variations in chemistry and DCI within each region in order of significance. In order to examine inter-regional patterns of correlation, the cosine similarity of vectors in PC-space is used as a quantitative measure of correlation in each analysis (Figure 2.2). The cosine similarity uses the angle between vectors in the projected space as a measure of their similarity or correlation and is commonly used in computer science and information retrieval (Salton & McGill, 1983). To emphasize the most significant correlations, only the first two Principal Components (i.e. the directions of primary and secondary variance in the region, typically capturing 50-60% of the total variance) are used to calculate the cosine similarity. Two related values are used throughout this paper: The cosine similarity, which ranges from 0 to 1, as the proxy for the correlation coefficient, and the cosine similarity angle (CSA), which ranges from 0 to 180 degrees, to visualize the PCA space (Figure 2.2). The distribution of CSAs to a reference variable are compiled from the regional analyses in Cumulative Distribution Functions (CDFs) (e.g. Figure 2.3) and examined with two-tailed Kolmogorov-Smirnov (K-S) tests (Massey, 1951). The CDFs show the range of observed CSAs, placing the results from each individual region in context and showing the larger trends in CSAs. The K-S tests show if the distribution of CSAs for each variable are distinct from the distributions for other variables. These tests provide confidence intervals to identify significant mutual correlations that persist at both regional and global scales.

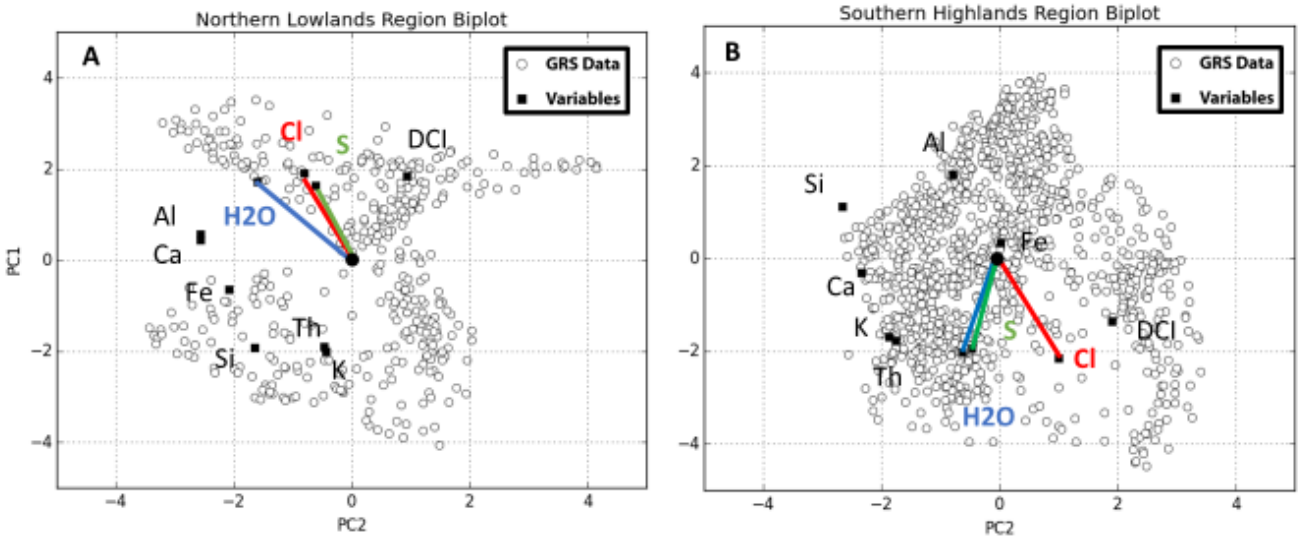



Figure 2.2. An example of the PC1-PC2 space examined using cosine similarity in both the northern lowlands (a) and southern highlands (b) (See supplement sections A 1.2, A 1.5 for details). The white circles show the GRS dataset projected into PC1-PC2 space, while the black squares represent the vectors of the original axes (i.e. the loadings of PC1 and PC2 for each variable scaled by a common factor for clarity). The vectors for $H_2O$ (blue), S (green), and Cl (red) are marked for clarity. These two biplots highlight the distinction between S-Cl-$H_2O$ trends in the two regions, with S and Cl most closely correlated in the northern lowlands, while S and $H_2O$ are more closely correlated in the southern highlands.

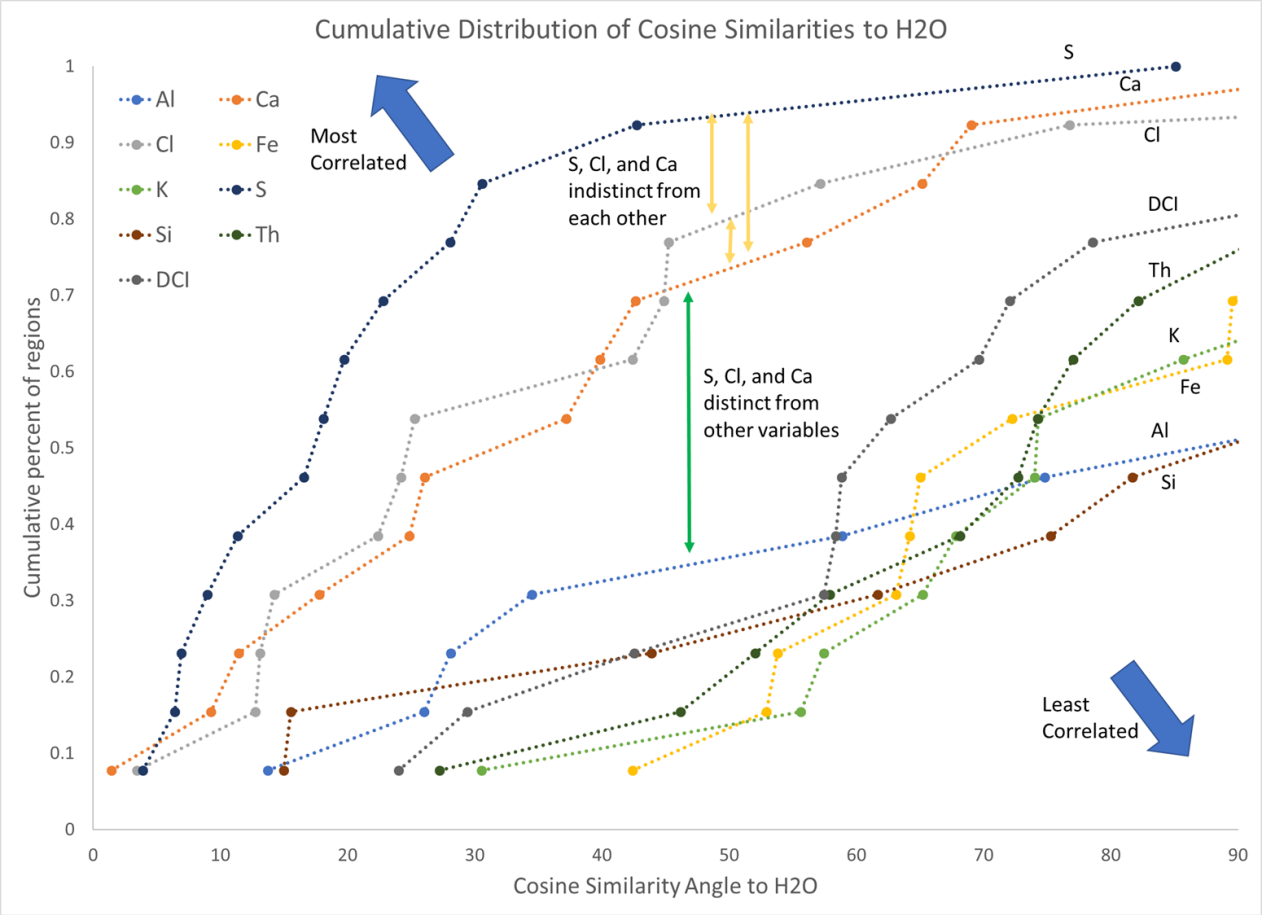



Figure 2.3. Cumulative PCA results from all regions for $H_2O$ as the reference variable. The blue arrows generally indicate where variables strongly correlated with $H_2O$ would plot relative to more weakly correlated variables. The cumulative trend of S, with a large portion of regions at low relative angles, suggests strong spatial correlation relative to the other variables. Ca and Cl show similar cumulative trends, indicating weaker correlation with $H_2O$ than S, but more than the other variables. Important K-S results are noted, with yellow lines indicating that the S, Cl, and Ca distributions are statistically indistinct from one another and the green line indicating that S, Cl, and Ca distributions are statistically distinct from the remaining variables. K-S results are enumerated in Table B 3

## 2.3. Results

The analytical methods provide two sets of results: PC loadings that describe the dominant variances within each region and CDFs that describe the global to regional distributions of CSAs for each variable. PC loadings concisely describe the most important variations within each region, providing insight into intra-regional trends in chemistry and DCI. Comparatively, the CDFs that compile regional correlation data describe correlations and trends that occur in multiple regions, capturing correlations that exist at both regional and global scales. Among the 13 PCA results and nine CDFs, the results most relevant to the conclusions are discussed in detail, with more discussion provided in the supplementary text (Section A 2) that supports and reinforces the conclusions.

### 2.3.1. Comparison between northern lowlands and Southern Highlands

The PCA results for the northern lowlands and southern highlands regions (Figure 2.2, Table 2.1) shed light on the geochemical distinction between the two regions, primarily for $H_2O$, S, and Cl (Figure 2.2). In the southern highlands, $H_2O$ and S are strongly correlated (Cosine Similarity = 0.998, CSA = 3.9°) but S and Cl are weakly correlated (0.782, 38.5°, Table 2.1). Comparatively, the northern lowlands show a weaker correlation between $H_2O$ and S (0.941, 19.7°) and a stronger correlation between S and Cl (0.999, 2.6°, Table 2.1). Cumulative results for S (Figure 2.4) show where these CSAs fall within the distribution of CSAs for all analyses. The $H_2O$-S CSA in the southern highlands is the smallest (i.e. strongest correlation) among the 13 regions, where more than half of the regions have $H_2O$-S CSAs smaller than that seen in the northern lowlands. Similarly, the S-Cl CSA in the northern lowlands is the smallest observed among the 13 regions, and the S-Cl CSA in the highlands is larger than more than half of observed regions. The size of these changes in correlation support these distinctions as meaningful changes in the correlation of $H_2O$, S, and Cl.

Table 2.1. The loading values for PC1 and PC2 (i.e. the components of each PC in the original data space) in key analysis including the midlatitudinal, northern lowlands and southern highlands regions (See section 2.2.2 and Figure 2.1 For details on regions). ±.316 was used as the cutoff for "significant" associations with a PC (See supplement section A 2 for more detail) values outside this range are bolded and highlighted in blue for positive loadings, blue for negative. The variance captured by PC1 and PC2 is below each region name. The cosine similarity angle to $H_2O$ and S is calculated for each variable in each region, with values below 25° highlighted in green.

PCA results for Selected Analyses

| Region | | Al | Ca | Cl | Fe | H2O | K | S | Si | Th | DCI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Midlatitudinal | PC1 | **0.318** | -0.087 | **-0.478** | -0.128 | **-0.448** | -0.199 | **-0.443** | 0.156 | -0.238 | **-0.360** |
| Var: 51% | PC2 | 0.117 | -0.187 | 0.158 | -0.231 | 0.031 | **-0.546** | 0.121 | **-0.512** | **-0.491** | 0.237 |
| Cosine Similarity Angle to H2O | | 155.8 | 69.0 | 14.2 | 65.0 | - | 74.0 | 11.3 | 111.0 | 68.1 | 29.4 |
| Cosine Similarity Angle to S | | 144.5 | 80.4 | 2.9 | 76.4 | 11.3 | 85.3 | - | 122.3 | 79.5 | 18.0 |
| Northern Lowlands | PC1 | 0.116 | 0.091 | **0.329** | -0.127 | **0.345** | **-0.402** | **0.381** | **-0.384** | **-0.379** | **0.371** |
| Var: 63% | PC2 | **-0.512** | **-0.513** | -0.122 | **-0.417** | **-0.320** | -0.086 | -0.162 | **-0.327** | -0.093 | 0.188 |
| Cosine Similarity Angle to H2O | | 34.5 | 37.2 | 22.4 | 64.2 | - | 125.2 | 19.7 | 96.7 | 123.5 | 69.6 |
| Cosine Similarity Angle to S | | 54.2 | 56.9 | 2.6 | 83.9 | 19.7 | 144.9 | - | 116.5 | 143.2 | 49.9 |
| Southern Highlands | PC1 | **0.361** | -0.062 | **-0.430** | 0.066 | **-0.403** | **-0.338** | **-0.387** | 0.224 | **-0.356** | -0.269 |
| Var: 53% | PC2 | -0.159 | **-0.467** | 0.202 | 0.004 | -0.126 | **-0.373** | -0.092 | **-0.534** | **-0.350** | **0.381** |
| Cosine Similarity Angle to H2O | | 138.9 | 65.2 | 42.4 | 166.3 | - | 30.5 | 3.9 | 95.5 | 27.2 | 72.1 |
| Cosine Similarity Angle to S | | 142.8 | 69.0 | 38.5 | 170.2 | 3.9 | 34.4 | - | 99.3 | 31.1 | 68.2 |

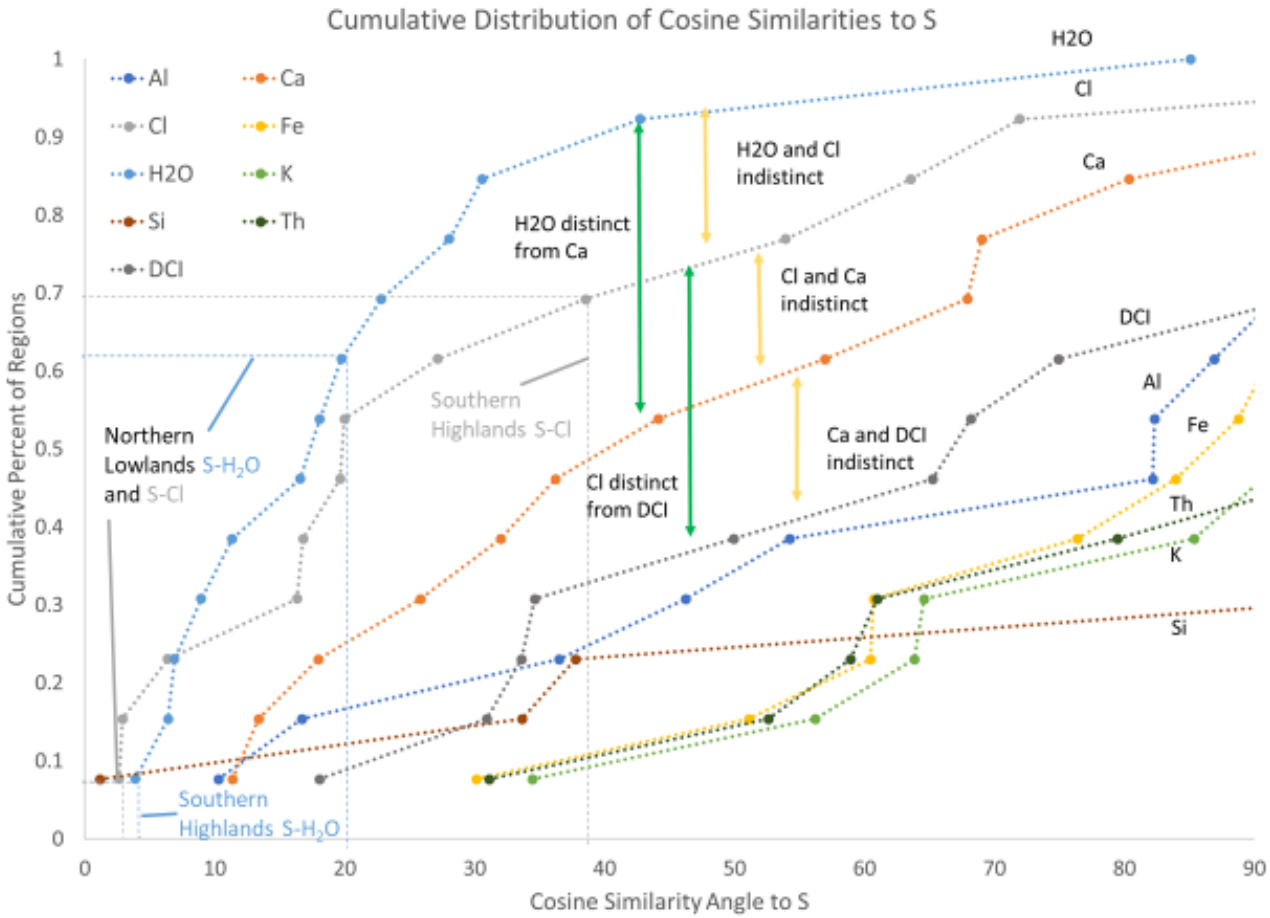



Figure 2.4. Cumulative PCA results for cosine similarity to S with specific results from the northern lowlands and southern highlands noted. Important K-S results are marked with green lines indicating distinction and yellow lines indicating indistinction. The distribution of $H_2O$ is distinct from all variables except Cl and Cl is indistinct from both $H_2O$ and Ca. Ca is distinct only from $H_2O$ and the two variables least correlated with S: K and Si. Because $H_2O$ is distinct from Ca and Cl is not distinct from Ca, $H_2O$ may be more strongly correlated with S compared to Cl, although only slightly.

### 2.3.2. Cumulative PCA results

The cumulative results from all regions for $H_2O$ (Figure 2.3) and S (Figure 2.4) give context to the result for the northern lowlands and southern highlands and highlight the general significance of correlations among the volatile elements $H_2O$, S, and Cl. The detailed discussion of the cumulative results for DCI is given in the supplementary material (Section A 2.4). In summary, DCI does not show notable correlations with any of the chemical maps, although there is tentative evidence for a correlation with Cl. Cumulative results for the remaining elements (Si, Al, Fe, Ca, K, Th) (Section A 2.4) show generally weak correlations with the other variables. Further discussion of cumulative results is in the supplement (Section A 2.4).

The cumulative PCA results and K-S tests for $H_2O$ (Figure 2.3, Table B 3) highlight the strong correlations among the volatile elements. CSA distributions of S, Ca, and Cl are distinct from the other six variables (Si, Al, Fe, K, Th, DCI) and not distinct from one another. This

distinction shows that the elements S, Cl, and Ca are more significant predictors of $H_2O$ variability than the other six variables, although relative significance among them cannot be confidently established. The cumulative PCA results for S (Figure 2.4) and Cl (Figure B 7) reinforce these results: S, Cl and $H_2O$ generally show mutual correlations that are distinct from the remaining variables, and Ca shows correlation with S and $H_2O$ beyond that of the other non-volatile elements.

The cumulative PCA results and K-S tests for S (Figure 2.4, Table B 3) generally reinforce the trends seen in the cumulative PCA results for $H_2O$ (Figure 2.3). Distributions of CSAs for $H_2O$ and Cl are distinct from most of the other variables, notably excepting Ca. However, in this case, a rough hierarchy can be established among $H_2O$, Cl, and Ca. The CSA distribution for $H_2O$ is distinct from all other CSA distributions except Cl; the Cl distribution is indistinct from the Ca and $H_2O$ distributions; and Ca is only distinct from $H_2O$, K, and Si. In the context of the other CSA distributions, $H_2O$ appears to correlate more strongly with S compared to Cl, and the remaining variables (K, Th, Ca, Al, Si, Fe) are generally poorly correlated.

## 2.4. Discussion

### 2.4.1 Aqueous Interaction with Regional soil across the topographic dichotomy

These observations refine prior bivariate insight into volatile spatial trends (Suniti Karunatillake et al., 2016) in three ways: (1) the S-$H_2O$ trend is preserved in multivariate space, (2) the change in the S-$H_2O$ correlation, previously associated with the hemispheres, persists when comparing each side of the topographic dichotomy, and (3) there is a newly discovered change in the S-Cl correlation associated with the topographic dichotomy. The first two points reinforce existing findings that S and $H_2O$ are strongly correlated in the southern highlands and that their correlation may indicate that hydrated sulfate compounds are an important reservoir of $H_2O$ in martian soil (S Karunatillake et al., 2014; Suniti Karunatillake et al., 2016). The discovery that the S-$H_2O$ correlation is paired with a shift in the S-Cl correlation leads to a more complete understanding of the contrast in soil geochemistry across the topographic dichotomy. The change in the S-$H_2O$ correlation across the topographic dichotomy (Table 2.1, Figure 2.2) (decreased S-$H_2O$ correlation in the north) may indicate a decreased abundance or decreased prevalence of hydrated sulfate compounds in northern lowlands soils. The associated change in S-Cl correlation across the dichotomy boundary (increased S-Cl correlation in the north) is similar to the contrast between excavated and surface soils seen in Gusev crater (Haskin et al., 2005). In Gusev soils, S, Cl, and Br correlated strongly in the lightly altered surface soils but aqueous $H_2O$ (possibly as brines or thin films) caused these three species to decouple in the older, more altered soils due to their differential solubility (Haskin et al., 2005). Furthermore, previous work at Gusev and Meridiani found that relative trends among S and Cl in the shallow subsurface are more consistent with aqueous processes rather than volatilization of Cl (e.g. via photolysis (Suniti Karunatillake et al., 2013)). Therefore, the regionally weaker correlation of S and Cl seen in the southern highlands, compared to the northern lowlands, would be consistent with greater (either longer-lasting or greater intensity) interaction with aqueous $H_2O$ in the southern highlands, similar to the Gusev soils at the Boroughs and Big Hole sites (Haskin et al., 2005). The relative weakness of the S-Cl correlation in the southern highlands provides evidence for aqueous interaction that may lead to alteration and leaching or migration of soluble soil components (e.g. salts containing S, Cl, and/or $H_2O$) across much of the southern highlands. However, long-range transport or concentration of these components is not noted. By contrast,

the increased correlation of S and Cl in the northern lowlands may indicate less interaction with $H_2O$ and a lack of alteration at the regional scale. This finding is similar to the unaltered basaltic material observed in Gusev and contrary to previous interpretations (A Deanne Rogers & Hamilton, 2015). This lack of alteration may be expected for the Amazonian volcanic units in Amazonis and Elysium Planitia. However, these volcanic units only constitute 25% of the northern lowlands region (Figure 2.1). The remaining 75% consists of lowlands and basin units, including mixed transitional units along the dichotomy boundary. Therefore, the absence of aqueous alteration is associated not only with young volcanic units, but also with the late Hesperian lowlands units found in Utopia, Chryse, and Acidalia Planitia (Tanaka et al., 2014).

Based on the results discussed in Section 2.3.2, it is possible that Ca plays an important role in the changing geochemistry between the northern lowlands and southern highlands. Figure 2.3 shows that the correlation of Ca and $H_2O$ is comparable with the Cl-$H_2O$ correlation at the regional to global scale. Because Ca-(per)chlorate compounds are likely less common than other compounds (e.g. Mg-perchlorates (Hecht et al., 2009)), this correlation may implicate Ca-sulfates as important reservoirs of both $H_2O$ and S in martian soil. This result is consistent with remote and in-situ observation of Ca-sulfates in various states of hydration (Langevin et al., 2005; Rapin et al., 2016). As a significant reservoir of S and $H_2O$, Ca-sulfates would play a role in the geochemical differences between the northern lowlands and southern highlands, and changes in the mineralogy or hydration states of Ca-sulfates could implicate specific aqueous processes in martian soils. However, the trends among Ca, $H_2O$, and S between the northern lowlands and southern highlands (Figure 2.2, Table 2.1) do not imply a clear pattern of presence, absence, or change in hydration state among Ca-sulfate species. While evidence is present that Ca-sulfates are an important component of the changes occurring between the northern lowlands and southern highlands, these results to do not clearly indicate any changes in Ca-sulfate mineralogy or abundance between these two regions that contribute to the proposed history of aqueous interaction.

Another consideration is the potential influence of surface dust abundance on soil chemistry and, therefore, the trends among the volatile elements. The moderate correlations between volatile elements and DCI in the midlatitudinal analysis (Section A 2.1) exceeds initial expectations and may suggest weak correlation between DCI and soil chemistry. However, the generally weak cumulative correlation between DCI and soil chemistry (excepting Cl, Section A 2.4) and the lack of strong correlations with DCI in either of the southern highlands or northern lowlands (Figure 2.2, Table 2.1) do not support a significant influence of surface dust in the noted shift on correlation among S, Cl, and $H_2O$. Relatedly, the atmospheric deposition of salts (Catling et al., 2010) may be an important process in martian soil that could influence the S and Cl correlation and contribute a substantial portion of soil S and Cl content (Ewing et al., 2006). However, more research is needed on predicted patterns of atmospheric salt deposition in order to substantiate differential deposition of salts on either side of the dichotomy such that it could cause the distinction between the northern lowlands and southern highlands.

Soils in the southern highlands may have interacted with $H_2O$ via leaching (e.g., hydrothermal vents and fumaroles), acid fog (i.e., aerosols and low-pH brine films reacting with the soil over geologic time) (J. A. Hurowitz & Fischer, 2014; Thomas M. McCollom & Hynek, 2005) or by more neutral to alkaline processes (J. A. Hurowitz et al., 2017). Minerals produced by these processes (e.g. perchlorates and Mg-sulfates) are typically soluble or deliquescent and

the preservation of such phases over geologic time in the near surface necessitates low water-to-rock ratios (e.g. Kounaves et al., 2010). The lack of such evidence of aqueous interaction in the north is not likely due to the lack of aqueous processes occurring in the northern lowlands. Instead, aqueous interaction likely preceded the emplacement of the Hesperian-aged units in the northern lowlands. Relative quiescence of widespread aqueous process and limited soil redistribution since the Hesperian would allow the poorly altered soil at the surface of the northern lowlands to arise from physical weathering of the underlying material with limited chemical weathering. This history is consistent with recent evidence from mineralogical observations (Pan et al., 2017) that suggest much of the upper (<2 km) volcanic or volcaniclastic rock in the northern lowlands lacks signatures of significant alteration.

## 2.5. Conclusions

Based on PCA analysis of GRS chemical maps and DCI data, the interpretation of distinct soil chemistries across the topographic dichotomy is supported. Spatial correlations among in S, Cl, and $H_2O$ abundances are consistent with widespread interaction with aqueous $H_2O$ in the southern highlands, and relatively pristine soils in the northern lowlands. These findings suggest laterally extensive aqueous interaction with bulk soil in southern highlands soils followed by geologically hyper-arid, low-temperature conditions that preserve metastable hydrated minerals over geologic time. The absence of such signatures in the northern lowlands supports a hypothesis of regionally pervasive mantling by minimally weathered soil that postdates widespread aqueous processes in the highlands. The preserved distinctness of soils suggests curtailed eolian, fluvial, mass wasting, or other kinds of mixing between the lowlands and highlands, at least since the emplacement of the Hesperian units in the north. In addition, the distinct correlation of Ca and $H_2O$ relative to other rock-forming elements supports the significance of hydrated Ca-compounds as a reservoir of $H_2O$ in martian soil.

The possible existence of hydrated minerals buried under several kilometers of unaltered material in the northern lowlands (Pan et al., 2017) is consistent with the proposed period of widespread aqueous interaction and alteration that formed both the present southern highlands soil and the now-buried paleosol in the northern lowlands. In addition, widespread aqueous processes and soil alteration can create biologically significant redox-sensitive minerals. With the suggestion that organic carbon is found on Mars in low but ubiquitous concentrations (Eigenbrode et al., 2018; Freissinet et al., 2015; Navarro-González et al., 2010), the evidence presented here for aqueous interaction throughout highlands soils may be indicative of a widespread habitable environment in martian soils.

# Chapter 3. The Martian Boulder Automatic Recognition System: MBARS

## 3.1 Introduction

Images taken by the High Resolution Imaging Science Experiment (HiRISE) aboard the Mars Reconnaissance Obiter (MRO) show that meter-scale boulders and blocks are common on the entire surface of Mars (Golombek et al., 2008, 2012). While it is not known how long these boulders have resided at the surface, they may serve as important geologic records for surface processes. Quantitative estimates of these boulder populations, including their size and location, have been used to address outstanding questions regarding surface weathering and impact processes (de Haas et al., 2013; Krishna & Kumar, 2016).

Boulder populations have also been used to examine potential boulder migration processes occurring atop patterned ground in the northern lowlands of Mars. Patterned ground is characterized by a polygonal network of surface fractures and occurs nearly ubiquitously at latitudes above ~60°N. Polygonal fractures, like their Earth analogs, form from the seasonal contraction of the ice-rich soil. The fracture patterns are visible in images, but no direct activity from fracture contraction has been detected through repeat HiRISE observations. Boulders clustered along the margins of the polygons or fracture networks, however, may demonstrate that the fractures likely are or have been sufficiently active to migrate the boulders to the fractures; a process well recognized in periglacial processes on Earth (e.g. Schlyter, 1992)Both large polygons (called Subdued Polygons in (J. S. Levy et al., 2010)) with diameters of ~20 m and the smaller polygons (3-6m diameter) have boulders accumulated in their bounding troughs (J. Levy et al., 2009; T. C. Orloff et al., 2011). Despite the ubiquity of the patterns and associated boulders, the link between the two features has heretofore been limited, and a gap remains as to whether the boulders are at meaningfully correlated (Barrett et al., 2017). Boulder sorting atop polygonal terrain in the martian northern lowlands may represent one of the few processes that are actively altering the martian surface in the modern environment. If active, cryoturbation may be overturning the upper several meters of the regolith, which is an important factor in interpreting surface observations. Furthermore, this process would actively erase shallow craters and other small-scale landforms in areas where polygons exists. Surface boulders act as an indicator of polygon activity, and therefore an indicator of the extent to which soil overturn and landform erasure have occurred. However, to date, examinations of boulder locations relative to underlying polygonal fractures have been constrained to small study areas, or to larger areas with more qualitative insight. The time and labor-intensive nature of detailed measurement and identification of boulders is generally a limiting factor in these investigations.

Previous attempts to automate the boulder identification process generated useful and accurate results (Golombek et al., 2008, 2012, 2016) but the process has not been applied outside the team that designed it and has seen little application outside of mission-critical landing site assessment. Automated boulder identification has recently been applied on the lunar surface (Y. Li & Wu, 2018), although it is unclear how successfully such methods will translate to Mars. Here, a Python-based algorithm is developed that can automatically identify, locate, and measure boulders using remote images of the martian surface. This set of tools and programs is collected in a python library called the Martian Boulder Automatic Recognition System (MBARS). MBARS is designed to be autonomous and fully reproducible, that is, it does not rely on expert oversight or trained input. Due to these capabilities, MBARS is the first fully open-source and user-independent framework for examining boulder populations and morphologies at large

scales. An application of MBARS, using boulder locations and morphometries to constrain boulder clustering in the martian northern lowlands, is presented as a demonstration of the utility of MBARS.

## 3.2. Methods

### 3.2.1. The HiRISE Dataset

The primary data for identifying boulders on Mars are images from the HiRISE instrument, a high-resolution camera that provides ~25-30 cm/pixel images of the martian surface. Typical HiRISE RED filter images are ~3km × 6km, and coverage is global but not contiguous. For the purposes of automatic boulder detection, certain attributes of the HiRISE dataset must be considered. The Point-Spread Function (PSF), or expected distribution of a point light source, of HiRISE is ~1.5 pixels (Kirk et al., 2008; McEwen et al., 2007). Thus, objects ~1 meter across are resolvable at typical resolutions. The PSF sets an important lower bound on measurements made in HiRISE images, as accuracy below ±1.5 pixels is not achievable without deconvolution or other image-reconstruction methods. Because the objects of interest are near the resolution limit of HiRISE, foreshortening (the distortion of objects in the foreground of an image relative to those further away) could have strong effects on interpreted boulder size. However, due to instrument design, foreshortening effects are compensated for (McEwen et al., 2007), so no considerations or corrections are required to correct or compensate for this issue.

Because the detection and measuring method is based on boulder shadows, the incidence angle of the Sun is an important factor. Lower incidence angles (sun closer to zenith) will cause shadows to shorten, making shadow detection more challenging, while higher incidence angles (sun closer to horizon) will stretch shadows, possibly to the point of extensive overlap. A shadow three pixels long and three pixels wide is considered a detection. Taking that size as a lower limit, the minimum detectable height in pixels can be calculated as $h_{min} = 2/\tan(\theta)$, where θ is the solar incidence angle measured from zenith. On the Moon, it was found that boulder height over diameter (h/D) approximated 0.54, roughly consistent with experimental studies of impact fragmentation of bedrock (Demidov & Basilevsky, 2014). Compared to the moon, it is more likely that rocks on the martian surface undergo a wider range of erosional processes that may influence boulder shape. However, given the abundance of impacts and impact-derived boulders on Mars, this is at least a likely ratio for fresh boulders or those deposited in the Amazonian. Using this ratio as an estimate for Mars, the minimum detectable diameter can be predicted as $D_{min} = 3.7/\tan(\theta)$. This equation can be used to derive an important transition: the incidence angle at which a "detectable" boulder (i.e. 3 pixels wide) is no longer detectable due to the shortness of the shadow. This transition occurs at ~51°, although boulders six pixels (1.5-1.8m) across remain theoretically detectable down to ~31.5°, close to the lowest incidence angles among HiRISE images. These values can then be used as lower limits for suitability for use in MBARS, although in general lower incidence angles are less favored since smaller shadows do not fit the boulder model (Figure 3.1) as well as larger shadows. High incidence angles may cause shadows long enough to overshadow other boulders and increase the chance that boulder shadows intersect panel boundaries. An upper incidence angle limit is difficult to determine, as it depends sensitively on boulder size distribution, spatial distribution, and local topography. In the

set of test images run at this point (five entire images with six image subsets), with sun incidence angles as high as 58°, extensive overlap from long shadows has yet to cause issues that are not overcome by shadow-splitting processes within the algorithm.

### 3.2.2. Detection by Shadow Segmentation

MBARS is motivated by the established shadow-detection approach published by Golombek (Golombek et al., 2008). In this approach, shadows are isolated from HiRISE images and measured in order to determine the morphometry of the shadow-casting boulder. Here the method is automated to improve repeatability and decrease reliance on user expertise.

The assumed boulder model is that of a half-buried spheroid, with a circular shape in map view and an elliptical cross section when viewed from ground level (Figure 3.1). Because the boulder is partially obscured by its own shadow to an unknown degree, it is difficult to estimate the width of the boulder along the direction of illumination. For this reason, boulders are assumed to be circular in this dimension. Throughout this work, directions and axes in the images are referenced in relation to the Sun's compass direction: Sunward, anti-Sunward, and Sun-perpendicular. Because the model is fully dependent on Sun orientation, typical geographic directions, as well as ellipse terminology (minor and major axes) are less useful. The cast shadow of a spheroidal object oriented vertically is an ellipse with its Sun-perpendicular axis equal to the boulder width, and the Sunward axis length determined by the boulder height, boulder width, and incidence angle (See Figure 3.1). The shadow length is more specifically related to the point on the boulder surface where the shadow is cast, labeled in the diagram as the "casting point". In the model, this is the point on the boulder where the surface slope is equal to the incidence angle of the Sun. The measured height will consistently be an underestimate of the actual boulder height. Equations 1 and 2 in Figure 3.1 quantify the relationship between the actual height and the height at the casting point for a spheroidal boulder. Because boulder shape cannot be known a priori it is difficult to correct this systematic underestimate. Therefore, the predicted boulder heights must be taken as minima, although boulder widths remain accurate within instrumental uncertainty.

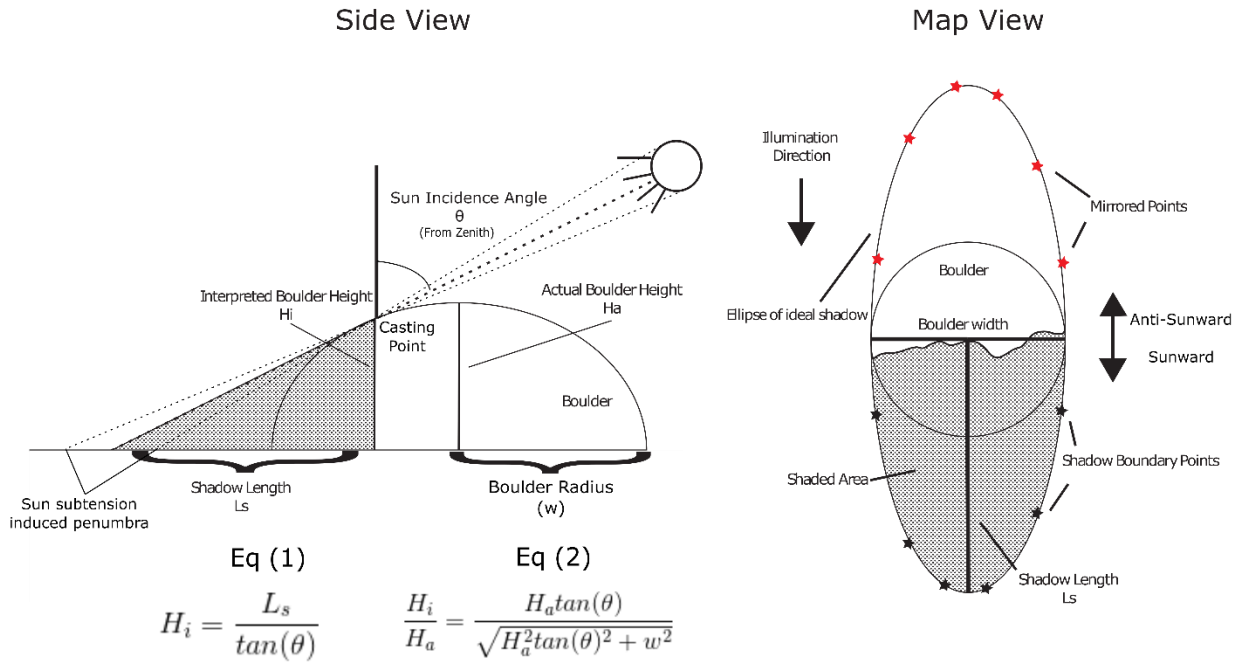



Figure 3.1. A geometric model of the map-view and side view of an observed boulder in a HiRISE image. The boulder is modeled as a spheroid oriented vertically (i.e. directly upwards) with horizontal radius *w*, and height $H_a$. Such a boulder casts a shadow expressed as an ellipse with a horizontal axis equivalent to the boulder width and a vertical axis equivalent to twice the length of the shadow. The Sunward-most extent of the boulder shadow is used as the vertical center of the boulder, and the shadow boundary is mirrored about this axis to create the mirrored points (red stars). Anti-Sunward side of the ellipse defined by the black and red points defines the ideal shadow cast by a hemispheric boulder. The side view of the boulder and shadow shows the relationships between the boulder height $H_a$, the interpreted height $H_i$, the shadow length $L_s$, and the sun angle $\theta$. Equation 1 shows how $H_i$ is determined from $L_s$ and $\theta$, which is used in the boulder measuring process. Equation 2 shows the relationship between $H_i$ and $H_a$, which assumes the spheroidal shape of the boulder. The casting point and $H_i$ can be determined by deriving the point on the boulder surface where the slope is equal to the sun incidence angle. As equation 2 shows, this is intricately related to $H_a$, *w*, and $\theta$.

Shadow penumbrae, the transition from shadowed areas to fully lit areas, can create uncertainty in these calculations. Atmospheric scattering is a major cause of penumbrae on Earth, but the thin martian atmosphere (~6 mbar) and scattering cross section of $CO_2$ for HiRISE RED filter wavelengths (~$1.214\times10^{-38}$ $m^2$ at 694nm) lead to expected scattering roughly ten orders of magnitude smaller than green light on Earth. Furthermore, the size of the Sun in the sky (.33° (Golombek et al., 2008) causes a penumbra in the shadow (Figure 3.1), but this effect creates penumbrae that are on the order of 3% of the shadow length. For boulders between 1.5 and 5m wide, this penumbra is on the order of 10 cm, small compared to the PSF of HiRISE, ~1.5 pixels or ±37.5 cm for a 25 cm/px image (Kirk et al., 2008; McEwen et al., 2007).

### 3.2.3. Algorithm Description

MBARS is designed as a library (MBARS) and two scripts (MBARS_RUN and MBARS_ANALYSIS) to access and run the algorithm (Appendix C). The workflow described below is also described succinctly in flowchart in Figure 3.2.

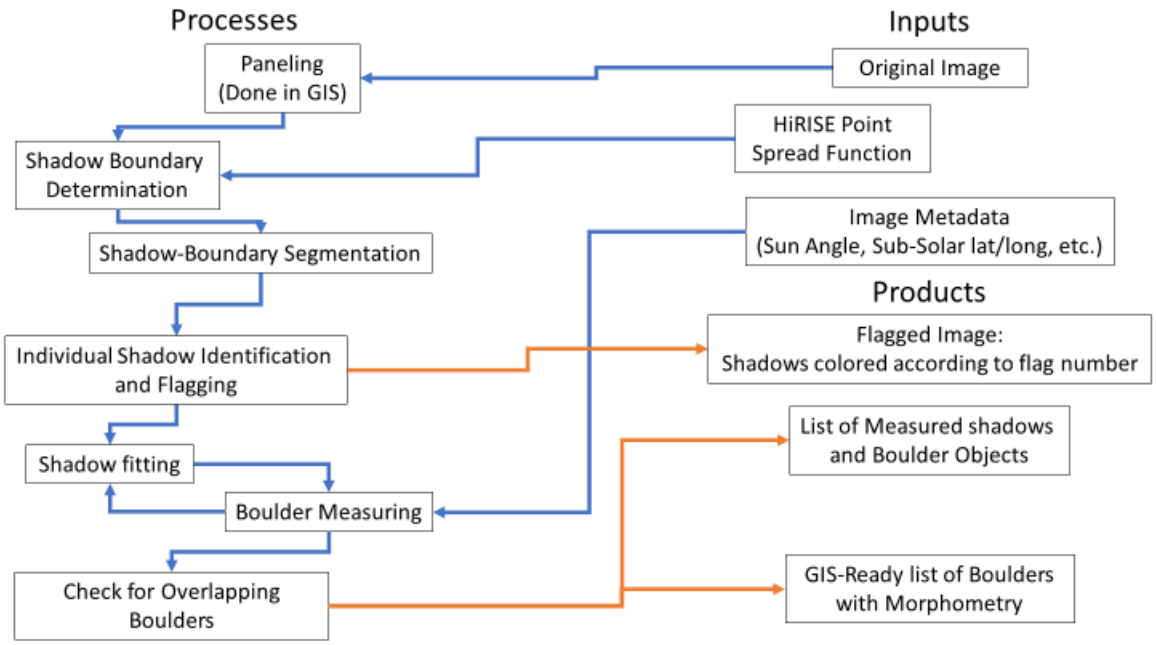



Figure 3.2. A flowchart of the MBARS algorithm. The three inputs are the HiRISE image (binned Red Reduced Data Record in JP2 format), the HiRISE PSF, and the Image Metadata, which is automatically fetched from the RDRCUMINDEX.TAB file. The original image must be paneled by an external GIS program, which also handles the conversion to PNG from JP2 files. Performing the following steps in loop on each image panel produces the listed products independently for each panel. Inbuilt into MBARS are tools to make the auxiliary files required for importing to ArcGIS for the produced images. Once segmented, the shadows, i.e. clusters of shadow pixels, are identified and flagged. With shadows identified, the algorithm fits ellipses to the shadows and begins the boulder measuring process.

#### *3.2.3.1. Image Preparation*

Prior to MBARS processing, several steps are taken to prepare the image for analysis. First, the HiRISE image is separated into panels using a GIS program. The GIS program provides crucial map orientation information to MBARS and splits the HiRISE image into manageable sizes for processing. The size of each panel can be controlled by the user, although 500-pixel square image panels (~ 125m) are used for test purposes. To segment shadows from non-shadowed areas, the individual image statistics and HiRISE Point-spread function (PSF) are used to predict a “shadow boundary” for each image (Figure 3.3). To make this prediction, three key pieces of information are required: the darkness of shadows, the brightness of non-shadow pixels, and the PSF of the HiRISE instrument. The interior of shadows larger than ~5 pixels are

generally dark enough to register a one in I/F values for HiRISE images, so the "model" shadow has a dark (I/F = 1) interior. The brightness of non-shadow pixels varies between images due to changing surface properties (e.g. albedo) and lighting conditions. Instead of predicting these changes, the modeled non-shadow pixels are randomly sampled from the entire image population. Finally, the HiRISE PSF is well-quantified from in-flight imaging of on-board targets and stars (McEwen et al., 2007), although other factors (spacecraft jitter, atmospheric conditions, etc.) are more difficult to constrain. Following previous HiRISE work (Kirk et al., 2008), a Lorentzian model (gamma=0.77) for the HiRISE PSF will be used here, and other factors that may influence the PSF will be assumed to be accounted for within the PSF. The modeling process is shown in Figure 3.3 and is repeated 100 times with the values within the "shadow" recorded. From the 100 results, the average of the shadow maxima is taken as the shadow boundary, i.e. the brightest pixel intensity expected to be found within a shadow.

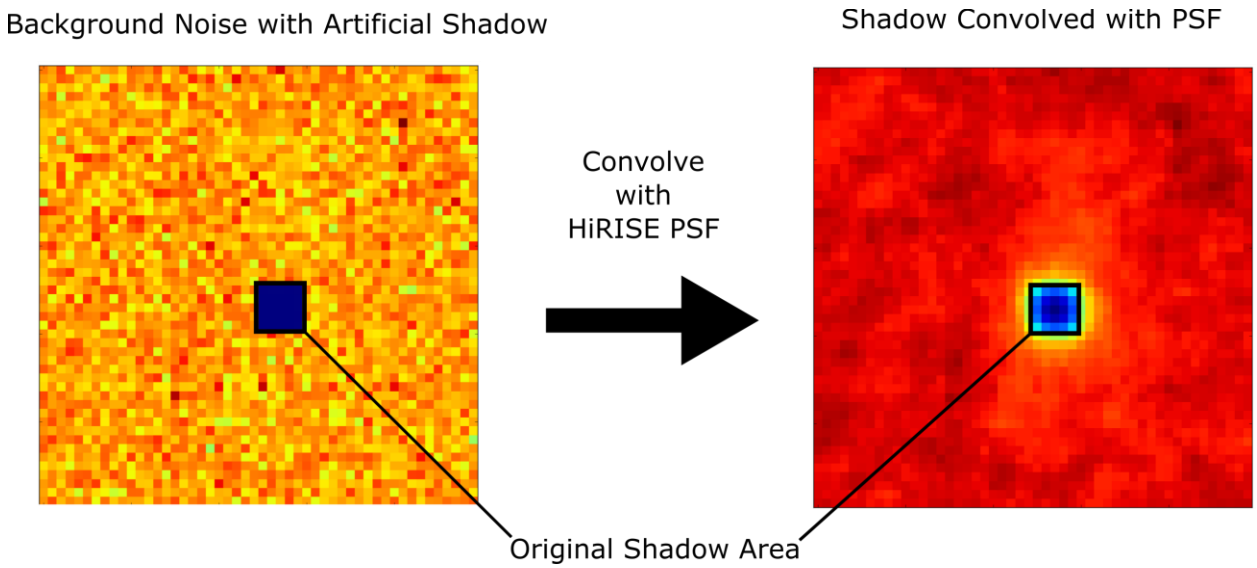



Figure 3.3. An example of the shadow convolution method applied to artificial data. The shadow area is fixed at a value of one and the rest of the image is randomly sampled from the image population. Convolution with the HiRISE Point Spread Function (PSF) yields the theoretical observation by HiRISE. After convolution, the original shadow area, marked in black, is examined to understand the post-convolution brightness within the shadow.

After the shadow boundary is defined, the program segments the sub-images based on the provided pixel intensity, below which the pixels are considered members of a shadowed area. During this step, MBARS also retrieves relevant metadata (sub-solar lat/long, resolution, incidence angle, etc.) from the RDRCUMINDEX file provided by the HiRISE team. The image is also rotated according to the Sun direction calculated from the sub-solar and sub-spacecraft coordinates provided in the HiRISE metadata. Rotating the image such that the Sun is coming from the top of the image speeds up later calculations and simplifies the overall process. This first collection of functions results in one primary product: the original image rotated and filtered such that pixel intensities above the shadow boundary are set to a fixed value. The "segmented" image is passed onto the next major function, boulder segmentation.

*3.2.3.2. Boulder Segmentation*

The second major step is to select and define individual shadows in the "segmented" image created in the first step, image preparation. A watershed algorithm (Meyer, 1994) is used to identify individual shadows, filling the shadow "basins" and separating shadows that touch one another (Figure 3.4). Local minima in the pixel intensity are used as seeds (i.e. starting points) for the watershed algorithm, readily identifying individual shadows and divides multiple shadows that touch and form a merged shadow region. The watershed process can create some false splitting in shadows with multiple minima that is corrected for later in the process. The watershed process also creates unique IDs, called flags, for each shadow. After shadows are isolated, shadows that are too small (<4 pixels) to reliably identify, or too large to be boulders (>30m) are disregarded. With each shadow identified and flagged, MBARS begins the process of determining boulder morphometry based on the shadow dimensions. MBARS utilizes Python's object-oriented functionality and builds shadow objects that store the factors required to fully describe both the shadow and the boulder. Lists of all boulder objects in an image panel are stored as ".shad" files that serve as the master record of boulders identified in the image.

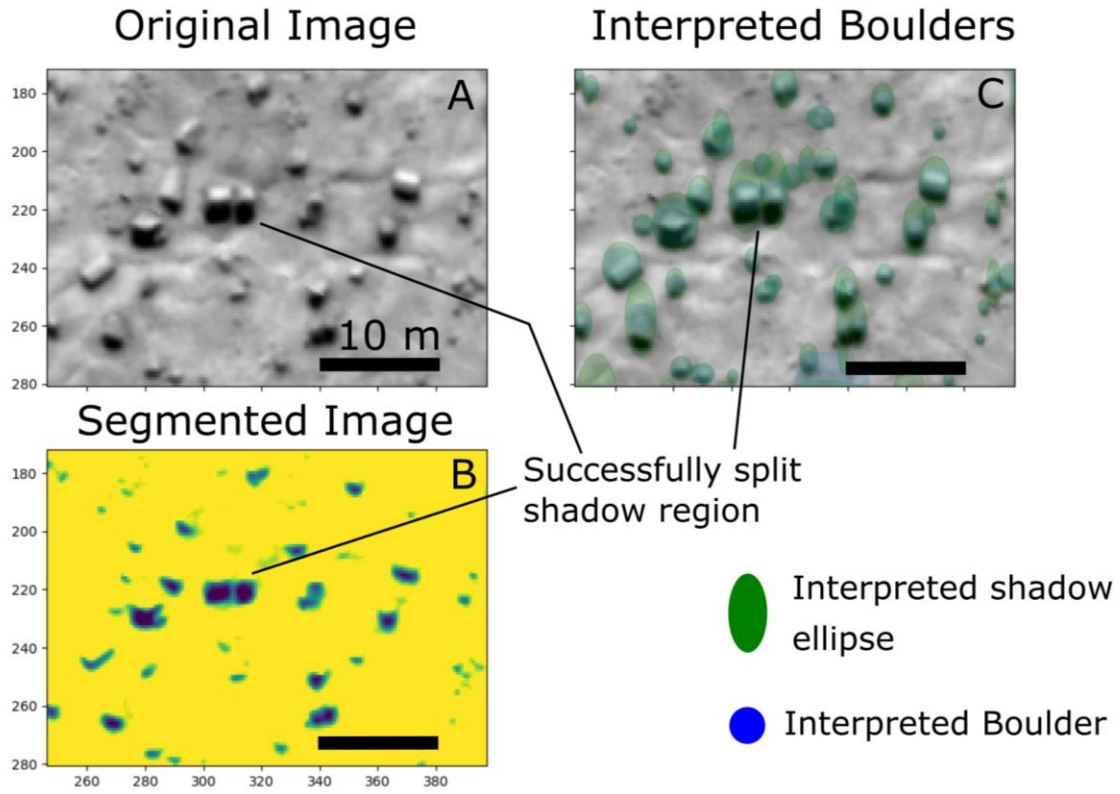



Figure 3.4. A subset of HiRISE image TRA_000828_2495 (Credit NASA/JPL) shown unaltered (A), segmented (B), and with interpreted shadows and boulders (C). In B, the background is segmented along the determined shadow boundary, with values above the boundary set to a fixed value and value below unchanged to preserve intra-shadow information. In C, shadows are shown with best-fit ellipses (green) and boulder models (blue circles). Ellipses are fit to the boundary of the shadows using Orthogonal Distance Regression. Shadows that lack ellipses were either removed due to small size or did not have good-fit ellipses. Two boulders with shadows that touch, forming a merged shadow region are successfully split by the algorithm.

*3.2.3.3. Measuring Boulder Dimensions*

Once a shadow is identified and defined as an individual shadow, MBARS constructs a boulder object and begins morphometric analysis. Previous automated methods (Golombek et al., 2008) used an area-preserving fit, i.e. the resultant shadow ellipse had the same area as the initial shadow. In practice, preserving shadow area did not improve fits and often led to less convergent solutions. Using a new approach, the anti-sunward (pointing away from the Sun) shadow boundary is reflected across the sunward (towards the Sun) end of the shadow and an ellipse is fit to both sets of points, retrieving the ideal ellipse (Figure 3.1). Effectively, the boulder shadow is fit to a half-ellipse, which is the expected shape of the shadow (Figure 3.1) and has proven to be effective and computationally fast. The algorithm delineates the border of each shadow (Figure 3.4) and fits an ellipse to it using Orthogonal Distance Regression (ODR) based on the Levenburg-Marquardt procedure (Boggs et al., 1992). The fit is not fixed to align with the Sunward and Sun-perpendicular axes, as these constraints caused poor fits. However, without constraining the orientation, ellipses are generally closely aligned with these axes. Fits that return diameters or heights that are too large to be a boulder (i.e. >10m) are not discarded but are marked as bad results. Generally, these bad results come from shadow casting topography (e.g. craters, knobs) or very small detections that produce shadows that are not well described by an ellipse. With the shadow ellipse described, the boulder diameter and boulder height ($2w$ and $H_i$, Figure 3.1) are retrieved from the sun-perpendicular and sun-parallel axes of the shadow ellipse (Figure 3.4). When measuring the boulders, the ellipse axes are first projected onto the image axes for consistency, but most ellipses are very closely aligned to the image axes despite not constraining the orientation. At this stage, a list of shadow objects, each containing information about the shadow, shadow ellipse, and boulder, is recorded for each image.

After an image panel is complete, the measured boulders are examined, looking for overlapping shadows. First, all boulder objects with shadows that touch (i.e. have adjacent pixels) are considered as possible candidates for merging and are compiled into “webs”, sets of boulder objects with touching shadows. In practice, these are typically chains of adjacent boulders or boulders that have been incorrectly split by the watershed algorithm. For each web, the summed fit error of each boulder object (from the ODR ellipse fit) is calculated, and boulder objects that lack a fit or exceed the maximum expected diameter are assigned an error of 1000, where typical fit errors are less than 20. The sum of fit errors is used, rather than the average, to select for simplicity, preferring fewer boulders to describe the observed shadow. Initially, the shadow pixels of all boulder objects in the web are combined and is interpreted as a single boulder. If the new boulder is an improvement on the original set of boulders (i.e. lower summed error), the new boulder object is kept, and the original set of boulder objects is removed. Otherwise, the algorithm uses k-means clustering to split all the pixels in the web into two or more shadows, performs fits on these new shadow divisions, and calculates the summed fit error. The original number of boulder objects (i.e. the number of objects in the web) is the highest number of means considered in the k-means split. Ultimately, among the original solution and the k-means derived splits, the solution with the lowest summed fit error is considered final.

A second program, MBARS_ANALYSIS, has several inbuilt tools to analyze the boulder morphometry results and can export the data to a GIS-ready data table enabling further analysis. Most significantly, the GIS-exporting function uses the .PGw files produced as a part of the GIS-

based image paneling to convert the boulder data from image-specific coordinates into GIS projected coordinates.

### 3.2.4. Spatial Statistical Analysis

To demonstrate the utility of MBARS, a HiRISE image is analyzed and examined with a variety of spatial statistical techniques that provide preliminary exploration of boulder clustering in the martian northern lowlands. A variety of well-developed metrics exist for quantifying clustering, as it is a common property of interest in spatial statistics. Three spatial statistics are calculated on the rock locations: Moran's *I*, Getis & Ord's General *G* statistic, and Ripley's *k*, all of which are already implemented in the ArcGIS Spatial Analyst Toolbox (Bailey & Gatrell, 1995; Getis & Ord, 2010). All these measures test the general hypothesis that features (e.g. boulders) or values (e.g. boulder sizes) are distributed in a spatially random way. Ripley's *k* is applied through the Multi Distance-Spatial Clustering Analysis and quantifies the clustering of boulders by their location and provides insight into the spatial scale at which the clustering occurs. Specifically, it distinguishes between clustering at the polygon scale (<10 m) compared to clustering at larger scales (~1km). A predominance of boulder clustering at smaller scales would suggest that a clustering mechanism is at work. Both Moran's I and Getis and Ord's G statistic measure the clustering tendency of values in different ways. Moran's I provides a measure of value clustering, specifically investigating whether boulders cluster according to size (i.e. larger or smaller boulders tending to cluster together). The G statistic will instead measure whether clusters of large boulders or clusters of small boulders are dominant. Both these methods give an image-wide assessment of the clustering of boulder morphometry, but they may be influenced by long-wavelength clustering, such as that from impact craters. A previous study (T. Orloff et al., 2013) measured the characteristic length scales of polygons using a 2-D Fourier Transform of subsets of HiRISE images. Similarly, a 2-D Fourier transform is performed on boulder location data using a similar methodology, power-weighted average of spatial frequencies, to determine a characteristic length for boulder clusters. Comparing the resultant characteristic length of boulder locations within an image to the characteristic length of polygons in the image informs if boulders are clustering along polygon edges, as has been suggested (T. Orloff et al., 2013).

## 3.3. Results

Three different methods are used to check the accuracy of MBARS: Measuring objects of known size, comparing results to manual analysis, and comparing results to previous algorithms. Throughout the results, error bars are not shown as they are typically smaller than the symbols on the plot. The principal source of uncertainty is the PSF of HiRISE, which is taken to be ±1pixel (typically 25cm) for all observations (e.g. boulder shadow width). While the uncertainty is on the order of 5-15% for each measurement, the large number of observed boulders yields confident averages and cumulative statistics.

In several cases, the Cumulative Fractional Area (CFA) of boulders is used to compare MBARS results to reference data. The CFA is calculated as the sum of all boulder areas above a given diameter divided by the analysis area. As a metric of boulder populations, CFA tests both

the detection rate of boulders (i.e. more detected boulders will lead to a higher CFA) as well as the measurement accuracy of boulders (systematic over or underestimates of boulders will be apparent). However, because CFA comparison conflates these two measurements other methods are used to independently test boulder measurement accuracy, and boulder detection rate accuracy.

### 3.3.1. Measuring Known Objects

Robotic landers represent one of the few objects on the martian surface with fully known shape and size, making them the ideal target to measure to "ground truth" MBARS. Six images of the Viking 1 and 2 landers (Table 3.1) were used to test the ability of MBARS to accurately measure objects. In these runs, the algorithm was only applied to the area around the lander, and care was taken to make sure that the landers were not split by the algorithm or combined with nearby boulders. These results test the maximum possible accuracy and precision of the boulder model and test the variability of results between multiple images of the same object. Generally, MBARS accurately measures the diameter of the landers within 1pixel error. The estimated heights are less accurate, although the inaccuracy may be due to the central antenna which is a significant deviation from the model boulder shadow shape (Figure 3.5). Compared to previous works, MBARS is comparably accurate in measuring diameter and height, although in general MBARS has a tendency to overestimate size, where previous work tends to underestimate the lander dimensions. These results demonstrate that, under the proper conditions, MBARS can measure objects on the martian surface accurately, verifying the boulder model and boulder measuring portions of the algorithm.

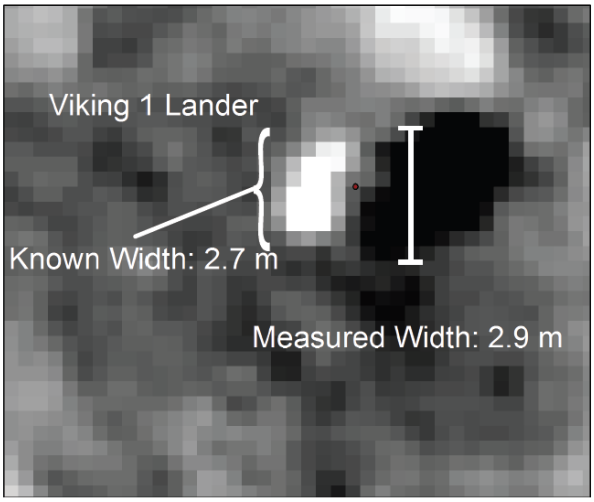



Figure 3.5. MBARS results for one of the Viking landers, an object of known size on the martian surface. Table 3.1 gives results for five other images of landers. The long end of the shadow appears to show the central antenna on the lander.

### 3.3.2. Comparison to Manual Results

Further comparison can be made between MBARS results and manually-measured images where a human has outlined and measured the boulders. Some discrepancies are expected between these two methods, because the human user was directly outlining the boulders, rather

than using shadows to measure the boulders. The comparison is shown in Figure 3.6, using three areas within HiRISE Image PSP_007718_2350. These three areas (10, 20, and 30) are each a one hectare section of the image, have no large topographic features or craters, and are dominated by their boulder populations (see Figure 3.7). As shown in Figure 3.6, results are generally comparable between manual measurements and MBARS results. In images 10 and 30, MBARS underpredicts boulder populations relative to the manual measurements, but this inaccuracy is likely due to a difference in analysis method. In these two images, bright spots that lack shadows are abundant, and these spots may be counted as boulders by the human observer, whereas MBARS ignores them due to lack of shadow. The solar incidence angle in the image is 42° above the horizon, which is in the middle range for HiRISE images, so this result is not due to low solar incidence. These bright spots may be boulders less than ~0.5m (see Figure 3.1) tall, although that would suggest unusually large aspect ratios given their apparent diameters. More likely, these are buried boulders or exposed underlying rock that are not tall enough to cast shadows. Image 20, where the two methods are most comparable, seems to show a larger degree of shadow-casting boulders, allowing MBARS to properly measure more of the surface boulders.

Table 3.1. Example of objects of known size measured with MBARS. Most images provide results that are within 1-pixel uncertainty of the known diameter. The measured heights are less reliable. Images marked N/A were either not measured in previous works, or not available at the time of publication. A small hollow near the Viking 1 lander is the apparent cause of some of the mismatch between observations and known dimensions. This hollow can extend the shadowed area if the lighting conditions are correct.

| Lander | HRISE image | Image Res | Sun Incidence Angle | Lander Dimensions | | Golombek observations | | MBARS Dimensions | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | D | H | D | H | D | H |
| VL1 | PSP_001521_2025 | 25 cm/px | 48 | 2.7 | 2 | 2.01 | 1.67 | 2.42 | 2.25 |
| VL1 | PSP_001719_2025 | 25 cm/px | 50 | 2.7 | 2 | 2.18 | 2.22 | 2.83 | 2.22 |
| VL1 | ESP_046170_2025 | 25 cm/px | 50 | 2.7 | 2 | N/A | N/A | 2.14 | 1.88 |
| VL2 | PSP_001976_2280 | 25 cm/px | 58 | 2.7 | 2 | 2.54 | 1.41 | 2.94 | 2.01 |
| VL2 | PSP_002055_2280 | 25 cm/px | 57 | 2.7 | 2 | 2.46 | 1.73 | 3.15 | 2.5 |
| VL2 | PSP_001501_2280 | 25 cm/px | 51 | 2.7 | 2 | 2.61 | 1.64 | 2.74 | 2.57 |

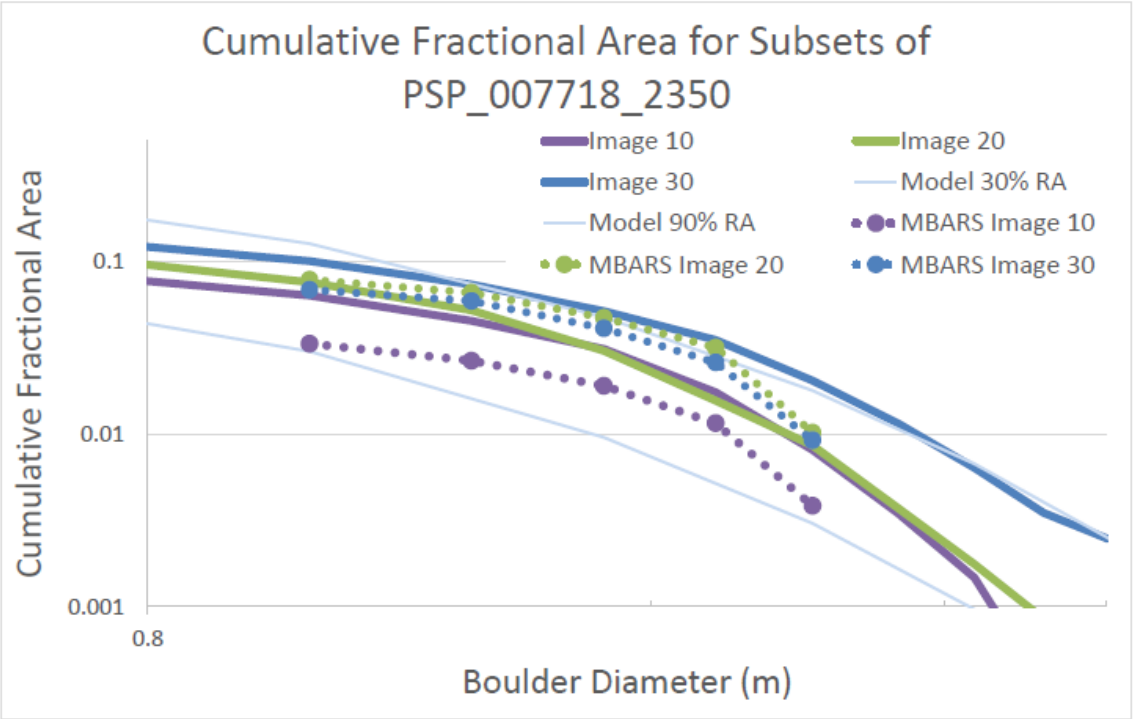



Figure 3.6. Comparison of MBARS results (dotted lines) to manual measurements (solid lines) in subsets of PSP_007718_2350. Three images (10, 20, and 30) were isolated from the original image and counted manually.

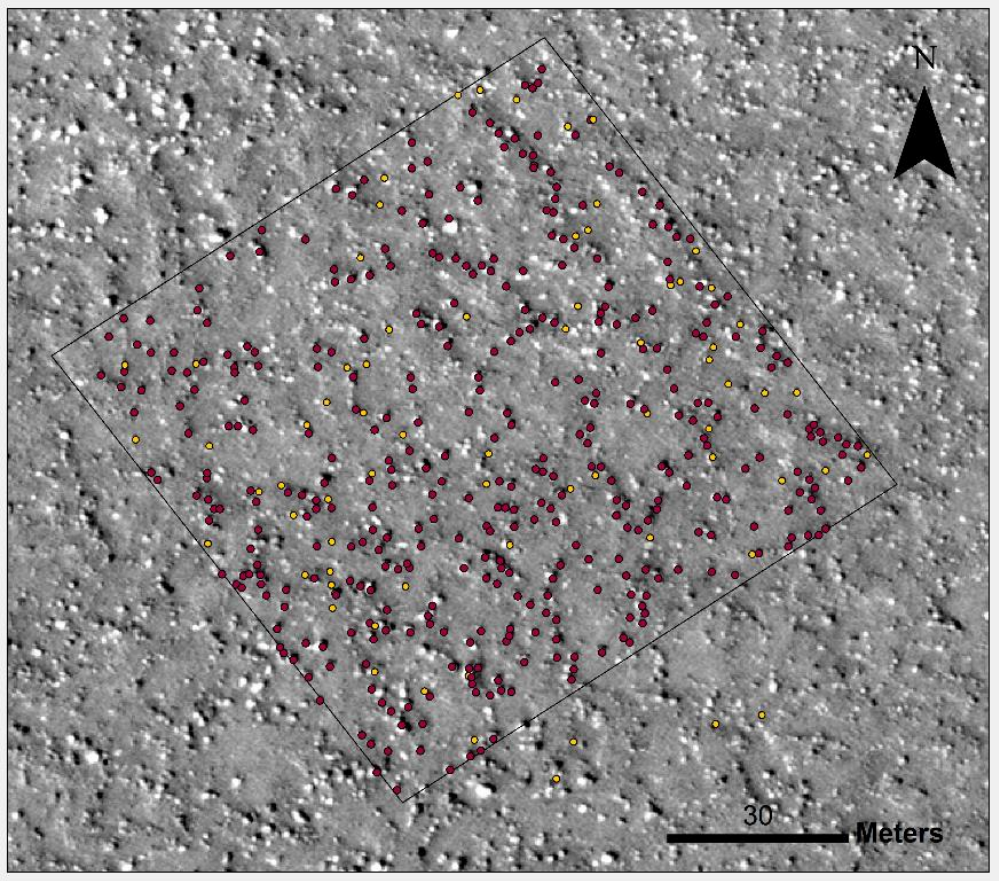



Figure 3.7. Subset of PSP_007718_2350, with the black box indicating “Image 20” in Figure 3.6. Red dots indicate measured boulders with good fits, yellow dots indicate detected boulders with poor fits. This image illustrates a fundamental challenge in comparing MBARS results with results from manual analysis: MBARS will miss boulders that lack shadows; humans will not. Several such boulders are seen in this image, although at lower abundances than in the other image subsets.

### 3.3.3. Comparison to Previous Algorithms

Comparison can be made between MBARS and previous semi-automatic methods of boulder detection presented in previous work (Golombek et al., 2008). Because the algorithm utilizes the same fundamental boulder detection methodology as the previous work, the results are expected to be nearly identical.

HiRISE image TRA_000828_2495 was used extensively in previous work (Fig 17, M. P. Golombek et al., 2008) and is used as the primary comparison to previous work. The CFA of boulders is compared in three areas (Figure 3.8), selected to exemplify areas of high, moderate, and low rock abundance. Figure 3.9 shows results from several MBARS tests compared to results from previous work. Of the two tests, the “MBARS_Low” results fit most closely, a near perfect match to previous results. These results were generated under more conservative running parameters (i.e. a darker shadow threshold) compared to the “MBARS_High” results and yield excellent results for this image. The MBARS_High results clearly overestimate either the abundance of boulders or their sizes, leading to an overall larger CFA at all boulder diameters.

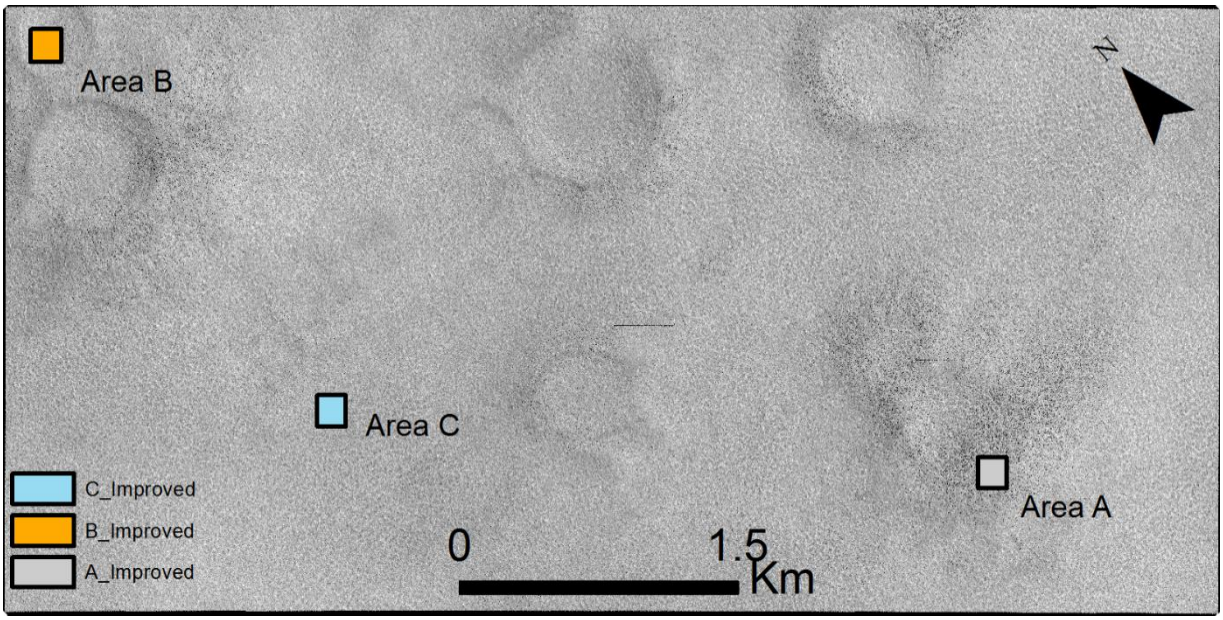



Figure 3.8. HiRISE image TRA_000828_2495 annotated with analysis areas A, B, and C as measured in Figure 3.9. This image is an example of typical terrain in the northern lowlands: low relief, moderate craters, and polygonal terrain covering most of the surface.

A second image, PSP_001391_2465, is also used as for comparison. In this case, the derived rock abundance (rocks/hectare) is compared between the two results (Figure 3.10). Compared to the use of CFA in Figure 3.8, comparison of rock abundance only tests the detection of boulders, not their interpreted size. Figure 3.10 clearly shows that the MBARS_High results again overestimate the boulder population, specifically the abundance of boulders, not necessarily individual boulder size. The MBARS_Low map shows much better agreement with previous results, and differences could be attributed to map processing differences (e.g. smoothing radius). However, closer inspection of the MBARS_Low results (Figure 3.11) shows that the conservative running parameters have caused a measuring issue. Nearly half of the boulders in image shown have poor fits on their boulder measurements, likely due to small shadows being reduced close to the detection limit. The varying quality of results shows that the "Low" settings are not a perfect fit to all images. Compared to TRA_00828_2495 (Figure 3.9), PSP_001391_2465 is generally darker, either due to lower surface albedo or dimmer lighting conditions. It is suspected that the difference in brightness between these two images is the primary cause of the differing results. Fine-tuning the boundary selection method in order to produce quality results independent of lighting and surface conditions is a topic of current development.

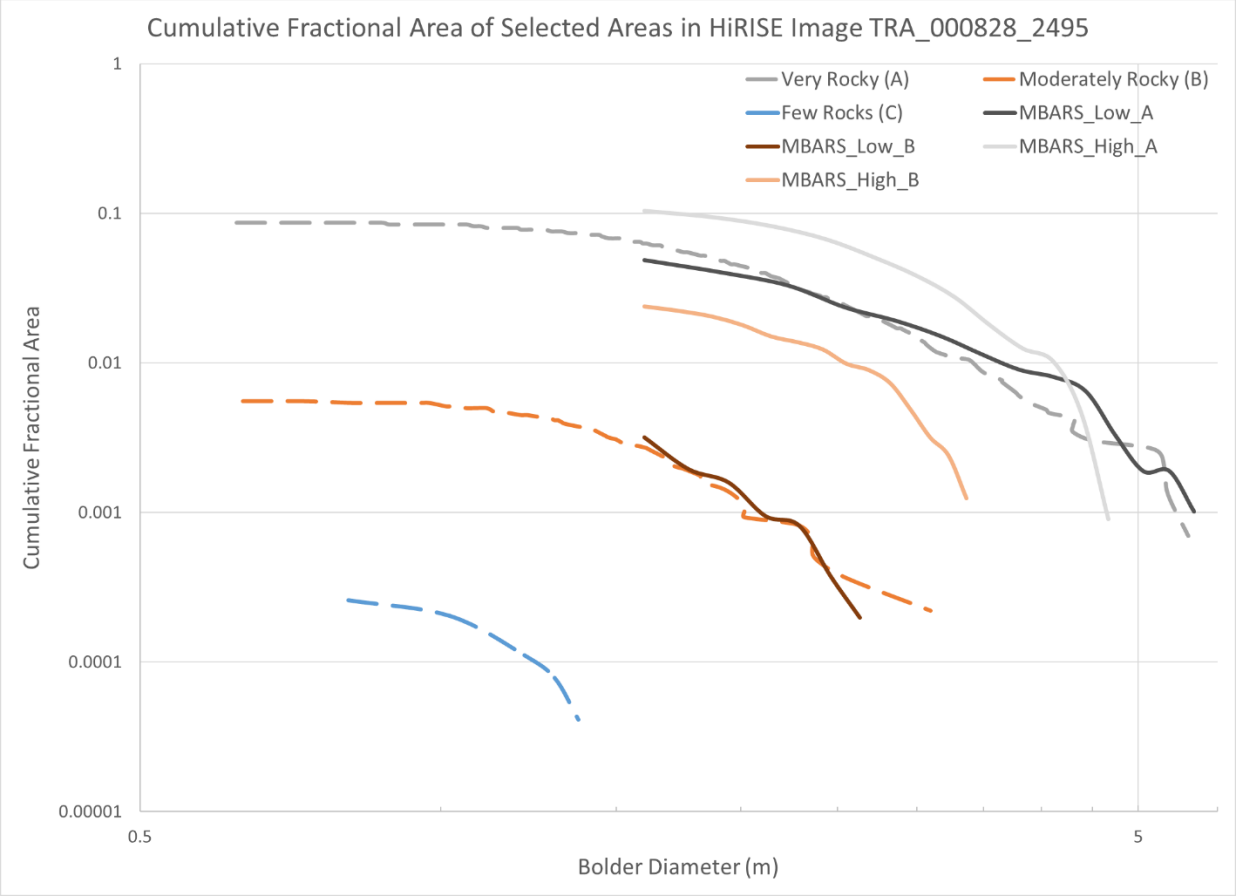



Figure 3.9. Comparison of MBARS results with Golombek's results (Golombek et al., 2008) under varying running conditions. The dashed lines show Golombek's results for areas with Few Rocks (C), Moderately Rocky (B), Very Rocky (A) (See Figure 3.8). Results for the same areas are in the same color with "Low" (conservative) results in dark shades and "High" (generous) results in lighter shades. The "Low" results are most consistent with previous results, while the "High" results are overly generous, resulting in high boulder counts.

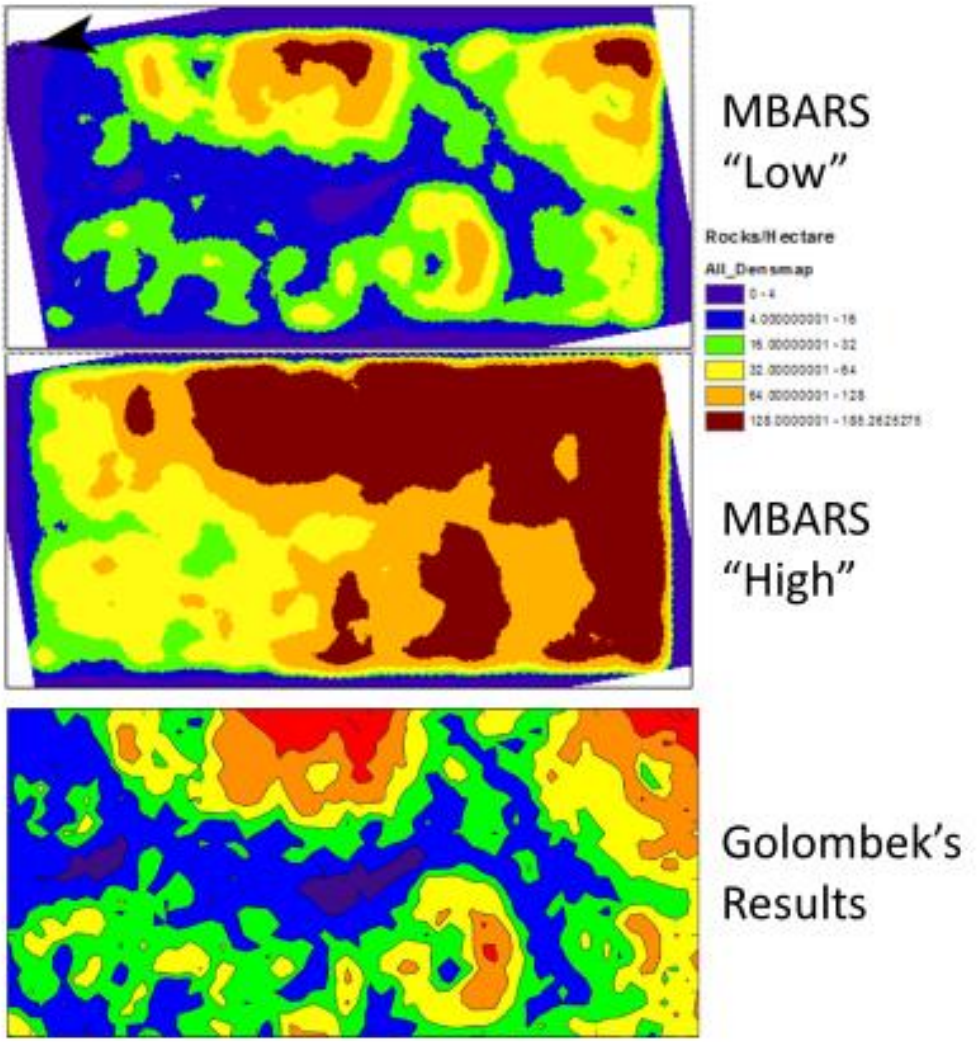



Figure 3.10. Comparison of Rock density derived from two different MBARS running parameters and results from previous work (Fig. 19, M. P. Golombek et al., 2008) on HiRISE image PSP_001391_2465. All images use the same color scale, shown on the right. The "Low" results closely match boulder abundance found in previous works, whereas the "High" results clearly overestimate the boulder population. This comparison shows that the "Low" parameters have very few false positive boulder detections but does not assess the accuracy of size assessment.

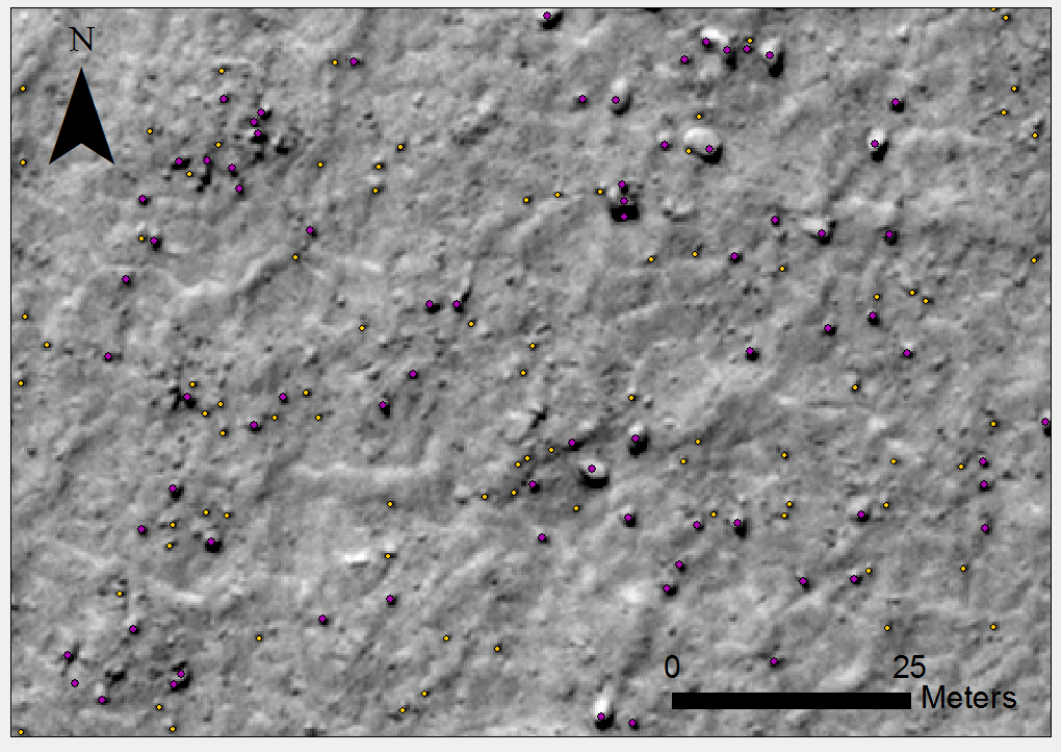



Figure 3.11. Closeup of Boulders in PSP_001391_2465 with results from the MBARS "Low" shown. Purple dots denote detected boulders, and yellow dots denote detected but unmeasured boulders. Large boulders are well described, but smaller boulders are systematically unmeasured, leading to poor results when detected boulder morphometry is considered.

## 3.4. Application

The image selected for analysis, PSP_001415_2470 (Figure 3.12), is in Acidalia Planitia at ~66°N latitude and is representative of high-latitude northern lowlands terrain containing both thermal contraction polygons and boulders that may have been sorted along polygon margins. The polygons within the image have been examined and measured in previous work (J. Levy et al., 2009; T. Orloff et al., 2013), but the boulders in the image have not been measured or otherwise analyzed. The existing analysis of polygons is essential to connect the MBARS results, which only examine boulders, with the underlying polygonal terrain.

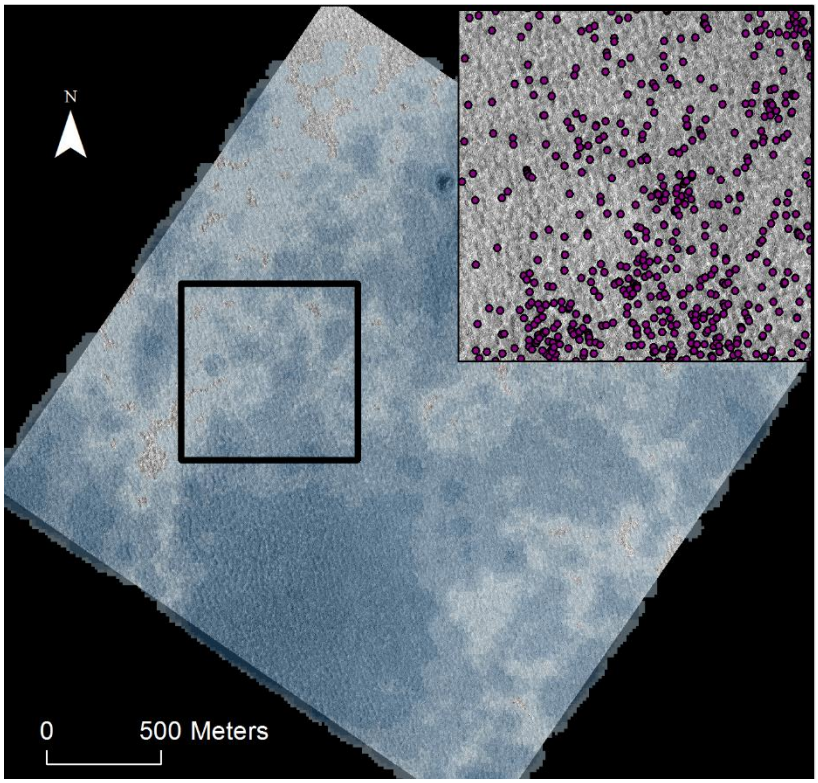



Figure 3.12. Boulder populations in HiRISE image PSP_001415_2470. The boulder density in boulders/hectare is shown in the main image, and individual objects are marked as purple dots in the inset. The entire image contains ~16,000 boulders. MBARS completed this analysis in ~1 hour owing to the relatively low boulder population and small image size.

Application of the spatial statistical methods to the selected image is promising, providing several indicators of clustering at many scales. In the test image, MD-SCA shows clustering at all spatial scales (Figure 3.13) and higher clustering at lower scales is consistent with polygon-scale clustering. The image-wide tests, Moran's I, Getis & Ord's G, and 2-D Fourier analysis are consistent with the above results, showing boulders clustering according to size (particularly large boulders) and showing small-scale characteristic lengths (Table 3.2). The polygon-derived $\lambda$ for this image is 4.9m (T. Orloff et al., 2013), and while the MBARS result is not identical, it is in the range of expected values for polygon-aligned clustering (~3-10m). It is also worth noting that even in this small image, 16,000 boulders were found. At a rate of one boulder measurement per 10 seconds, it would take a human user ~44 hours to measure all the boulders in this image. This image was completed in ~1 hour by MBARS, a 44× improvement over manual methods.

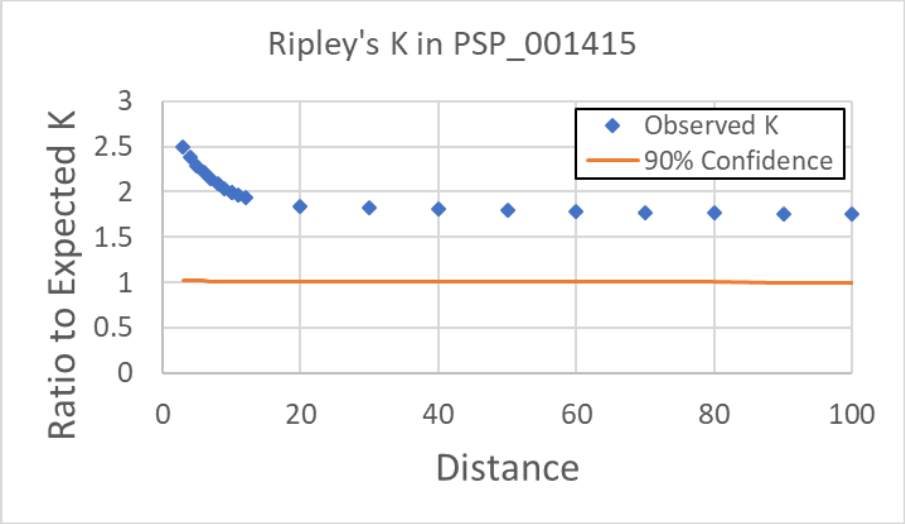



Figure 3.13. Results of Multi Distance Spatial Clustering Analysis in PSP_001415. The observed k (blue dots) is at all scales higher than the expected k, as well as the 90% confidence interval (orange line). Using the ratio to expected k highlights the increased clustering at small scales that may indicate polygon-scale clustering.

Table 3.2. Initial Results for G, I, and Orloff's characteristic length method on PSP_001415. Both G and I are consistent with clustering of large boulders. While MBARS size measurements are less reliable than the locations, relative size may be informative. Moran's I is notable significantly above the threshold to suggest clustering (z = 2.58). Orloff's method returned a λ of 3.9, this result is near the 4.9m estimated by Orloff (T. Orloff et al., 2013), and the general range of λ in the original work.

| Preliminary Results for PSP_001415 | | |
|---|---|---|
| Test | Result | Interpretation |
| Moran's I | z = 12.8 | Boulders are clustered according to size |
| Getis & Ord General G | z = 3.42 | Large boulders are more clustered than small boulders |
| Orloff's Characterstic Length (λ) | λ = 3.9m | Boulders cluster at a characteristic length of ~4m |

These initial findings and preliminary analyses show the utility of MBARS as a rapid and powerful tool to study boulder populations. Application to PSP_001415_2470 has yielded results that are self-consistent and may provide initial support to several hypotheses regarding boulder clustering patterns. As MBARS becomes more refined, rapid, and reliable, these analyses can be expanded to providing a hemispheric-scale assessment of boulder clustering patterns and testing competing hypotheses on boulder migration in the northern lowlands of Mars

## 3.5. Conclusions

MBARS is an accessible, open source tool for planetary scientists to locate and measure boulders in HiRISE images accurately and rapidly. The results establish the accuracy of MBARS and the consistency with previous methods of automated boulder measurement. MBARS is capable of reproducing previous results without oversight and can accurately determine boulder locations and diameters under proper conditions. The application in Section 4 shows the utility of MBARS in assessing boulder populations in a single image. In addition to outpacing manual measurements by a factor of more than 40, these results provide preliminary support for existing theories on boulder clustering in the northern lowlands of Mars.

The boulder identification, location, and measurement across multiple HiRISE images will expedite investigations regarding boulder populations, including ejecta processes (Krishna & Kumar, 2016), erosional processes, and fluvial processes (de Haas et al., 2013). Furthermore, the ability to measure boulders at such a large spatial scales in reasonable times enables large-scale statistical analysis of phenomena that have predominantly been studied at more local scales (e.g. Barrett et al., 2017; T. C. Orloff et al., 2011). These capabilties makes MBARS a new and necessary tool for the study of martian surface processes that will advance the understanding of a wide variety of processes on Mars. Furthermore, the open source and low cost design of MBARS will enable rapid expansion of the algorithm, combining it with other complementary image analysis methods, as well as machine-learning approaches.

## Conclusions: The Mission Continues

These three chapters display essential components of working in planetary remote sensing. In the first chapter, established methods are applied (e.g. Box and Whisker plots, TAS diagrams) in an area that had not yet been regionally assessed. This displays the importance of identifying areas where more can be learned from a dataset using established methods. In the second chapter, established methods (Principal Component Analysis, Kolmogorov-Smirnov tests, cosine similarity) are combined in new ways to advance on previous global mapping efforts (e.g. Gasnault et al., 2010) and examinations of volatile chemistry (Suniti Karunatillake et al., 2016). This shows how established techniques can be combined in new ways to re-examine data and improve existing analyses. In the third chapter, new methods are developed that enable entirely new analysis from the HiRISE image dataset. Novel methods of analysis, can lead to new applications of existing data, expanding the return from individual missions and enabling insight outside expectations.

Planetary science is a growing field, with many governments and private entities working to send robots and people to the Moon, Mars, and other bodies in our solar system. Even with this great effort, planetary science data are still mission-limited: ultimately the missions, their payloads, and goals determine what kind of data will be retrieved and when. It is this constraint that drives analysis of mission data beyond its original scope and leads to innovative applications of exiting data (e.g. determining gravity measurements from navigational accelerometers (Lewis et al., 2019)). The works presented above reflect this necessity and demonstrate the need and capability to extract more information from existing data. Within our lifetimes, we will not likely witness the time when gathering geologic information on Mars is as straightforward as accessing even the most remote places on Earth. However, the prospect and planning of human missions to Mars and an increasing number of rovers show that we are on the cusp of many incredible discoveries about Mars. Our view of Mars has changed vastly in the past 50 years since some of the first crafts touched down on its surface, and we can only begin to imagine how our view will continue to evolve and grow in the next 50 years. I hope that this work, and the work I will do in the future, will contribute to that evolution and expand our understanding of Mars.

## References


Acuña, M. H., Connerney, J. E. P., Ness, N. F., Lin, R. P., Mitchell, D., Carlson, C. W., et al. (1999). Global Distribution of Crustal Magnetization Discovered by the Mars Global Surveyor MAG/ER Experiment. *Science*, *284*(5415), 790–793. https://doi.org/10.1126/science.284.5415.790

Agee, C. B., Wilson, N. V., McCubbin, F. M., Ziegler, K., Polyak, V. J., Sharp, Z. D., et al. (2013). Unique Meteorite from Early Amazonian Mars: Water-Rich Basaltic Breccia Northwest Africa 7034. *Science*, *339*(6121), 780–785. https://doi.org/10.1126/science.1228858

Anderson, R., Bridges, J. C., Williams, A., Edgar, L., Ollila, A., Williams, J., et al. (2015). ChemCam results from the Shaler outcrop in Gale crater, Mars. *Icarus*, *249*, 2–21. https://doi.org/10.1016/j.icarus.2014.07.025

Anderson, R. C., Dohm, J. M., Golombek, M. P., Haldemann, A. F. C., Franklin, B. J., Tanaka, K. L., et al. (2001). Primary centers and secondary concentrations of tectonic activity through time in the western hemisphere of Mars. *Journal of Geophysical Research*, *106*(E9), 20563. https://doi.org/10.1029/2000JE001278

Andrews-Hanna, J. C., Zuber, M. T., & Banerdt, W. B. (2008). The Borealis basin and the origin of the martian crustal dichotomy. *Nature*, *453*(7199), 1212–1215. https://doi.org/10.1038/nature07011

Anguita, F., Farelo, A.-F., Lopez, V., Mas, C., Munoz-Espadas, M.-J., Marquez, A., & Ruiz, J. (2001). Tharsis dome, Mars: New evidence for Noachian-Hesperian thick-skin and Amazonian thin-skin tectonics. *Journal of Geophysical Research*, *106*(2000), 7577–7589.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis* (Vol. 413). Longman Scientific & Technical Essex.

Baker, V R, Maruyama, S., & Dohm, J. M. (2002). A theory of early plate tectonics and subsequent long-term superplume activity on Mars. *Electronic Geosciences*, *7*.

Baker, Victor R, Maruyama, S., & Dohm, J. M. (2007). Tharsis superplume and the geological evolution of early Mars. In *Superplumes: Beyond plate tectonics* (pp. 507–522). Springer.

Balta, J. B., & McSween, H. Y. (2013). Water and the composition of Martian magmas. *Geology*, *41*(10), 1115–1118. https://doi.org/10.1130/G34714.1

Baratoux, D., Toplis, M. J., Monnereau, M., & Sautter, V. (2013). The petrological expression of early Mars volcanism. *Journal of Geophysical Research E: Planets*, *118*(1), 59–64. https://doi.org/10.1029/2012JE004234

Baratoux, David, Toplis, M. J., Monnereau, M., & Gasnault, O. (2011). Thermal history of Mars inferred from orbital geochemistry of volcanic provinces. *Nature*, *472*(7343), 338–341. https://doi.org/10.1038/nature09903

Baratoux, David, Samuel, H., Michaut, C., Toplis, M. J., Monnereau, M., Wieczorek, M., et al. (2014). Petrological constraints on the density of the Martian crust. *Journal of Geophysical Research: Planets*, *119*(7), 1707–1727. https://doi.org/10.1002/2014JE004642

Barrett, A. M., Balme, M. R., Patel, M. R., & Hagermann, A. (2017). Clastic patterned ground in

Lomonosov crater, Mars: examining fracture controlled formation mechanisms. *Icarus*, *295*, 125–139. https://doi.org/10.1016/j.icarus.2017.06.008

Bevington, P. R., & Robinson, D. K. (2003). Data reduction and error analysis. *McGraw-Hill*.

Blake, D. F., Morris, R. V., Kocurek, G., Morrison, S. M., Downs, R. T., Bish, D., et al. (2013). Curiosity at Gale Crater, Mars: Characterization and Analysis of the Rocknest Sand Shadow. *Science*, *341*(6153), 1239505-1-1239505–7. https://doi.org/10.1126/science.1239505

Boggs, P. T., Byrd, R. H., Rogers, J. E., & Schnabel, R. B. (1992). User's Reference Guide for ODRPACK Version 2 . 01 Software for Weighted Orthogonal Distance Regression. *U.S. Department of Commerce*, (June), 99.

Boynton, W. V., Taylor, G. J., Evans, L. G., Reedy, R. C., Starr, R., Janes, D. M., et al. (2007). Concentration of H, Si, Cl, K, Fe, and Th in the low- and mid-latitude regions of Mars. *Journal of Geophysical Research*, *112*, E12S99. https://doi.org/10.1029/2007JE002887

Catling, D. C., Claire, M. W., Zahnle, K. J., Quinn, R. C., Clark, B. C., Hecht, M. H., & Kounaves, S. (2010). Atmospheric origins of perchlorate on mars and in the atacama. *Journal of Geophysical Research E: Planets*, *115*(1), 1–15. https://doi.org/10.1029/2009JE003425

Certini, G., & Ugolini, F. C. (2013). An updated, expanded, universal definition of soil. *Geoderma*, *192*, 378–379. https://doi.org/10.1016/j.geoderma.2012.07.008

Connerney, J. E. P., Acuña, M. H., Wasilewski, P. J., Kletetschka, G., Ness, N. F., Rème, H., et al. (2001). The Global Magnetic Field of Mars and Implications for Crustal Evolution. *Geophysical Research Letters*, *28*(21), 4015–4018. https://doi.org/10.1029/2001GL013619

Connerney, J. E. P., Acuña, M. H., Ness, N. F., Kletetschka, G., Mitchell, D. L., Lin, R. P., & Reme, H. (2005). Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 14970–14975. https://doi.org/10.1073/pnas.0507469102

Dehouck, E., Mclennan, S. M., Meslin, P., & Cousin, A. (2014). Constriants on abundance, composition, and nature of X-ray amorphous components of soils and rocks at Gale Crater, Mars. *Journal of Geophysical Research: Planets*, *119*, 2640–2657. https://doi.org/10.1002/2014JE004716

Demidov, N. E., & Basilevsky, A. T. (2014). Height-to-diameter ratios of moon rocks from analysis of Lunokhod-1 and -2 and Apollo 11–17 panoramas and LROC NAC images. *Solar System Research*, *48*(5), 324–329. https://doi.org/10.1134/s0038094614050013

Dohm, J. M., Anderson, R. C., Baker, V. R., Ferris, J. C., Hare, T. M., Strom, R. G., et al. (2000). System of gigantic valleys northwest of Tharsis, Mars: Latent catastrophic flooding, northwest watershed, and implications for northern plains ocean. *Geophysical Research Letters*, *27*(21), 3559–3562. https://doi.org/10.1029/2000GL011728

Dohm, J. M., Kerry, K., Keller, J., Baker, B. R., Boynton, W., Maruyama, S., & Anderson, R. C. (2005). Mars Geological Province Designations for the Interpretation of GRS data. In *36th Annual Lunar and Planetary Science Conference* (pp. 3–4). League City. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2005/pdf/1567.pdf

Dohm, J. M., Hare, T. M., Robbins, S. J., Williams, J. P., Soare, R. J., El-Maarry, M. R., et al. (2015). Geological and hydrological histories of the Argyre province, Mars. *Icarus*, *253*, 66–98. https://doi.org/10.1016/j.icarus.2015.02.017

Dohm, J. M., Spagnuolo, M. G., Williams, J.-P., Viviano-Beck, C. E., Karunatillake, S., Álvarez, O., et al. (2015). THE MARS PLATE-TECTONIC-BASEMENT HYPOTHESIS. In *Lunar and Planetary Science Conference* (p. 1741). The Woodlands. Retrieved from http://www.hou.usra.edu/meetings/lpsc2015/pdf/1741

Dohm, J M, Maruyama, S., & Baker, V. R. (2001). Mars plate tectonics: Surface geology and analyses of topographic and geophysical data. In *AGU Fall Meeting Abstracts* (p. abstract #P32C-0561).

Dohm, J M, Maruyama, S., Baker, V. R., Anderson, R. C., Ferris, J. C., & Hare, T. M. (2002). Plate tectonism on early Mars: diverse geological and geophysical evidence. In *Lunar and Planetary Science Conference* (Vol. 33, p. 1639).

Dohm, J M, Miyamoto, H., Maruyama, S., Baker, V., Anderson, R. C., Hynek, B. M., et al. (2013). Mars evolution. In *Nova Science Publishers, Inc.* (pp. 1–33).

Dohm, James M., & Tanaka, K. L. (1999). Geology of the Thaumasia region, Mars: plateau development, valley origins, and magmatic evolution. *Planetary and Space Science*, *47*(3–4), 411–431. https://doi.org/10.1016/S0032-0633(98)00141-X

Dohm, James M., Ferris, J. C., Baker, V. R., Anderson, R. C., Hare, T. M., Strom, R. G., et al. (2001). Ancient drainage basin of the Tharsis region, Mars: Potential source for outflow channel systems and putative oceans or paleolakes. *Journal of Geophysical Research*, *106*(E12), 32943. https://doi.org/10.1029/2000JE001468

Dohm, James M., Tanaka, K. L., & Hare, T. M. (2001). Geologic map of the Thaumasia region, Mars: U.S. Geological Survey Geologic Investigations Series I-2650, 3 sheets., 1–3.

Dohm, James M., Barlow, N. G., Anderson, R. C., Williams, J.-P., Miyamoto, H., Ferris, J. C., et al. (2007). Possible ancient giant basin and related water enrichment in the Arabia Terra province, Mars. *Icarus*, *190*(1), 74–92. https://doi.org/10.1016/j.icarus.2007.03.006

Dohm, James M., Anderson, R. C., Barlow, N. G., Miyamoto, H., Davies, A. G., Jeffrey Taylor, G., et al. (2008). Recent geological and hydrological activity on Mars: The Tharsis/Elysium corridor. *Planetary and Space Science*, *56*(7), 985–1013. https://doi.org/10.1016/j.pss.2008.01.001

Dohm, James M., Anderson, R. C., Williams, J.-P., Ruiz, J., McGuire, P. C., Buczkowski, D. L., et al. (2009). Claritas rise, Mars: Pre-Tharsis magmatism? *Journal of Volcanology and Geothermal Research*, *185*(1–2), 139–156. https://doi.org/10.1016/j.jvolgeores.2009.03.012

Dohm, James M., Williams, J. P., Anderson, R. C., Ruiz, J., McGuire, P. C., Komatsu, G., et al. (2009). New evidence for a magmatic influence on the origin of Valles Marineris, Mars. *Journal of Volcanology and Geothermal Research*, *185*(1–2), 12–27. https://doi.org/10.1016/j.jvolgeores.2008.11.029

Edwards, C. S., & Piqueux, S. (2016). The Water Content of Recurring Slope Linea on Mars. *Geophysical Research Letters*, *43*, 8912–8919. https://doi.org/10.1002/2016GL070179

Eigenbrode, J. L., Summons, R. E., Steele, A., Freissinet, C., Millan, M., Navarro-González, R., et al. (2018). Organic matter preserved in 3-billion-year-old mudstones at Gale crater, Mars, (June), 1–6. https://doi.org/10.1126/science.aas9185

Ewing, S. a., Sutter, B., Owen, J., Nishiizumi, K., Sharp, W., Cliff, S. S., et al. (2006). A threshold in soil formation at Earth's arid-hyperarid transition. *Geochimica et Cosmochimica Acta*, *70*, 5293–5322. https://doi.org/10.1016/j.gca.2006.08.020

Fairén, A;lberto G., Ruiz, J., & Anguita, F. (2002). An Origin for the Linear Magnetic Anomalies on Mars through Accretion of Terranes: Implications for Dynamo Timing. *Icarus*, *160*, 220–223. https://doi.org/10.1006/icar.2002.6942

Fairén, Alberto G., & Dohm, J. M. (2004). Age and origin of the lowlands of Mars. *Icarus*, *168*(2), 277–284. https://doi.org/10.1016/j.icarus.2003.11.025

Farley, K. A., Malespin, C., Mahaffy, P., Grotzinger, J. P., Vasconcelos, P. M., Milliken, R. E., et al. (2014). In Situ Radiometric and Exposure Age Dating of the Martian Surface. *Science*, *343*(January), 1247166.

Fastook, J. L., & Head, J. W. (2015). Glaciation in the Late Noachian Icy Highlands: Ice accumulation, distribution, flow rates, basal melting, and top-down melting rates and patterns. *Planetary and Space Science*, *106*, 82–98. https://doi.org/10.1016/j.pss.2014.11.028

Feldman, W. C., Prettyman, T. H., Maurice, S., Plaut, J. J., Bish, D. L., Vaniman, D. T., et al. (2004). Global distribution of near-surface hydrogen on Mars. *Journal of Geophysical Research E: Planets*, *109*(9), 1–13. https://doi.org/10.1029/2003JE002160

Freissinet, C., Glavin, D. P., Mahaffy, P. R., Miller, K. E., Eigenbrode, J. L., Summons, R. E., et al. (2015). Organic molecules in the Sheepbed Mudstone, Gale Crater, Mars. *Journal of Geophysical Research : Planets*, *120*, 495–514. https://doi.org/10.1002/2014JE004737

Frey, H. V., Roark, J. H., Shockey, K. M., Frey, E. L., & Sakimoto, S. E. H. (2002). Ancient lowlands on Mars. *Geophysical Research Letters*, *29*(10), 22-1-22–4. https://doi.org/10.1029/2001gl013832

Gasnault, O., Jeffrey Taylor, G., Karunatillake, S., Dohm, J., Newsom, H., Forni, O., et al. (2010). Quantitative geochemical mapping of martian elemental provinces. *Icarus*, *207*(1), 226–247. https://doi.org/10.1016/j.icarus.2009.11.010

Getis, A., & Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis* (pp. 127–145). Springer.

Golombek, M., Huertas, A., Marlow, J., McGrane, B., Klein, C., Martinez, M., et al. (2008). Size-frequency distributions of rocks on the northern plains of Mars with special reference to Phoenix landing surfaces. *Journal of Geophysical Research*, *113*, E00A09. https://doi.org/10.1029/2007JE003065

Golombek, M., Huertas, A., Kipp, D., & Calef, F. (2012). Detection and Characterization of Rocks and Rock Size-Frequency Distributions at the Final Four Mars Science Laboratory Landing Sites. *Mars*, *7*, 1–22. https://doi.org/10.1555/mars.2012.0001

Golombek, M., Kipp, D., Warner, N., Daubar, I. J., Fergason, R., Kirk, R. L., et al. (2016). Selection of

the InSight Landing Site. *Space Sci Rev*, 5–95. https://doi.org/10.1007/s11214-016-0321-9

de Haas, T., Hauber, E., & Kleinhans, M. G. (2013). Local late Amazonian boulder breakdown and denudation rate on Mars. *Geophysical Research Letters*, *40*(14), 3527–3531. https://doi.org/10.1002/grl.50726

Hahn, B. C., McLennan, S. M., & Klein, E. C. (2011). Martian surface heat production and crustal heat flow from Mars Odyssey Gamma-Ray spectrometry. *Geophysical Research Letters*, *38*(14), 1–5. https://doi.org/10.1029/2011GL047435

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Hamilton, V. E., & Christensen, P. R. (2000). Determining the modal mineralogy of mafic and ultramafic igneous rocks using thermal emission spectroscopy. *Journal of Geophysical Research: Planets*, *105*(E4), 9717–9733.

Haneberg, W. (2004). *Computational Geosciences with Mathematica* (2004 Editi). Springer.

Hartmann, W. K., Quantin, C., Werner, S. C., & Popova, O. (2010). Do young martian ray craters have ages consistent with the crater count system? *Icarus*, *208*(2), 621–635. https://doi.org/10.1016/j.icarus.2010.03.030

Haskin, L. a, Wang, A., Jolliff, B. L., McSween, H. Y., Clark, B. C., Des Marais, D. J., et al. (2005). Water alteration of rocks and soils on Mars at the Spirit rover site in Gusev crater. *Nature*, *436*(7047), 66–69. https://doi.org/10.1038/nature03640

Hauber, E., Broz, P., Jagert, F., Jodowski, P., & Platz, T. (2011). Very recent and wide-spread basaltic volcanism on Mars. *Geophysical Research Letters*, *38*(10), 1–5. https://doi.org/10.1029/2011GL047310

Head, J., Mustard, J., Kreslavsky, M., Milliken, R., & Marchant, D. (2003). Recent ice ages on Mars. *Nature*, *426*(6968), 797–802. https://doi.org/10.1038/nature02114

Head, J. W. (2007). The Geology of Mars: new insights and outstanding questions. In *The Geology of Mars* (pp. 1–46).

Hecht, M. H., Kounaves, S. P., Quinn, R. C., West, S. J., Young, S. M. M., Ming, D. W., et al. (2009). Detection of Perchlorate and the Soluble Chemistry of Martian Soil at the Phoenix Lander Site. *Science*, (July).

Hood, D. R., Judice, T., Karunatillake, S., Rogers, D., Dohm, J. M., Susko, D., & Carnes, L. K. (2016). Assessing the geologic evolution of Greater Thaumasia, Mars. *Journal of Geophysical Research: Planets*, 1–17. https://doi.org/10.1002/2016JE005046

Hunter, J. D. (2007). Matplotlib: A 2D graphic environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hurowitz, J. A., & Fischer, W. W. (2014). Contrasting styles of water-rock interaction at the Mars Exploration Rover landing sites. *Geochimica et Cosmochimica Acta*, *127*, 25–38. https://doi.org/10.1016/j.gca.2013.11.021

Hurowitz, J. A., Grotzinger, J. P., Fischer, W. W., McLennan, S. M., Milliken, R. E., Stein, N., et al. (2017). Redox stratification of an ancient lake in Gale crater, Mars. *Science*, *356*(6341). https://doi.org/10.1126/science.aah6849

Hurowitz, Joel A., & McLennan, S. M. (2007). A ~3.5 Ga record of water-limited, acidic weathering conditions on Mars. *Earth and Planetary Science Letters*, *260*(3–4), 432–443. https://doi.org/10.1016/j.epsl.2007.05.043

Hyvärinen, a., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4–5), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Jackson, M. P. A., Vendeville, B. C., & Schultz-Ela, D. D. (1994). Structural Dynamics of Salt Systems. *Annual Review of Earth and Planetary Sciences*, *22*, 93–117. https://doi.org/10.1146/annurev.ea.22.050194.000521

Jones, E. (2015). Identifying an index of subsurface volatiles on Mars through an anlysis of impact crater morphometry using principal component analysis. *Journal of Geophysical Research: Planets*, 1023–1043. https://doi.org/10.1002/2014JE004735.Received

Karunatillake, S, Wray, J. J., Gasnault, O., McLennan, S. M., Rogers, A. D., Squyres, S. W., et al. (2014). Sulfates hydrating bulk soil in the Martian low and middle latitudes. *Geophysical Research Letters*, *41*(22), 7987–7996. https://doi.org/10.1002/2014GL061136

Karunatillake, Suniti, Wray, J. J., Squyres, S. W., Taylor, G. J., Gasnault, O., McLennan, S. M., et al. (2009). Chemically striking regions on Mars and Stealth revisited. *Journal of Geophysical Research*, *114*(E12), E12001. https://doi.org/10.1029/2008JE003303

Karunatillake, Suniti, McLennan, S. M., & Herkenhoff, K. E. (2010). Regional and grain size influences on the geochemistry of soil at Gusev crater, Mars. *Journal of Geophysical Research*, *115*(E7), 1–17. https://doi.org/10.1029/2010JE003637

Karunatillake, Suniti, Squyres, S. W., Gasnault, O., Keller, J. M., Janes, D. M., Boynton, W. V., & Finch, M. J. (2011). Recipes for Spatial Statistics with Global Datasets: A Martian Case Study. *Journal of Scientific Computing*, *46*(3), 439–451. https://doi.org/10.1007/s10915-010-9412-z

Karunatillake, Suniti, Gasnault, O., Squyres, S. W., Keller, J. M., Janes, D. M., Boynton, W. V., & Newsom, H. E. (2012). Martian Case Study of Multivariate Correlation and Regression with Planetary Datasets. *Earth, Moon, and Planets*, *108*, 253–273. https://doi.org/10.1007/s11038-012-9395-x

Karunatillake, Suniti, Zhao, Y.-Y. S., McLennan, S. M., Skok, J. R., & Button, N. E. (2013). Does martian soil release reactive halogens to the atmosphere? *Icarus*, *226*(2), 1438–1446. https://doi.org/10.1016/j.icarus.2013.07.018

Karunatillake, Suniti, Wray, J. J., Gasnault, O., Mclennan, S. M., Rogers, a. D., Squyres, S. W., et al. (2016). The association of hydrogen with sulfur on Mars across latitudes, longitudes, and compositional extremes. *Journal of Geophysical Research*, *121*, 1–29. https://doi.org/10.1002/2016JE005016

Kirk, R. L., Rosiek, M. R., Anderson, J. A., Archinal, B. A., Becker, K. J., Cook, D. A., et al. (2008). Ultrahigh resolution topographic mapping of Mars with MRO HiRISE stereo images : Meter-scale

slopes of candidate Phoenix landing sites, *113*, 1–31. https://doi.org/10.1029/2007JE003000

Kite, E. S., Gao, P., Goldblatt, C., Mischna, M. A., Mayer, D. P., & Yung, Y. L. (2017). Methane bursts as a trigger for intermittent lake-forming climates on post-Noachian Mars. *Nature Geoscience*, *10*(10), 737–740. https://doi.org/10.1038/NGEO3033

Kounaves, S. P., Hecht, M. H., Kapit, J., Gospodinova, K., Deflores, L., Quinn, R. C., et al. (2010). Wet Chemistry experiments on the 2007 Phoenix Mars Scout Lander mission: Data analysis and results. *Journal of Geophysical Research E: Planets*, *115*(1), 1–16. https://doi.org/10.1029/2009JE003424

Krishna, N., & Kumar, P. S. (2016). Impact spallation processes on the Moon: A case study from the size and shape analysis of ejecta boulders and secondary craters of Censorinus crater. *Icarus*, *264*, 274–299. https://doi.org/10.1016/j.icarus.2015.09.033

Langevin, Y., Poulet, F., Bibring, J.-P., & Gondet, B. (2005). Sulfates in the North Polar Region of Mars Detected by OMEGA/Mars Express. *Science*, *307*(5715), 1584–1586. https://doi.org/10.1126/science.1109091

LeMaitre, R. W., Streckeisen, A., Zanettin, B., LeBas, M. J., Bonin, B. ., Bateman, P., et al. (2002). Igneous Rocks: A classification and glossary of terms. In R. W. Le Maitre (Ed.), *International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks*. New York.

Levy, J., Head, J., & Marchant, D. (2009). Thermal contraction crack polygons on Mars: Classification, distribution, and climate implications from HiRISE observations. *Journal of Geophysical Research E: Planets*, *114*(1), 1–19. https://doi.org/10.1029/2008JE003273

Levy, J. S., Marchant, D. R., & Head, J. W. (2010). Thermal contraction crack polygons on Mars: A synthesis from HiRISE, Phoenix, and terrestrial analog studies. *Icarus*, *206*(1), 229–252. https://doi.org/10.1016/j.icarus.2009.09.005

Lewis, K. W., Peters, S., Gonter, K., Morrison, S., Schmerr, N., Vasavada, A. R., & Gabriel, T. (2019). A surface gravity traverse on Mars indicates low bedrock density at Gale crater. *Science*, *363*(6426), 535–537. https://doi.org/10.1126/science.aat0738

Li, J., Cheng, H., Wang, K., & Deng, G. (2010). Salt Tectonics in the Kuqa Foreland fold and thrust belt, Tarim basin, northwest China.

Li, Y., & Wu, B. (2018). Analysis of Rock Abundance on Lunar Surface From Orbital and Descent Images Using Automatic Rock Detection. *Journal of Geophsical Research: Planets*. https://doi.org/10.1029/2017JE005496

Maruyama, S., Dohm, J., & Baker, V. (2001). Mars plate tectonics (1): An Earth prospective. In *AGU Fall Meeting Abstracts* (Vol. 1, p. 565).

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78.

McCollom, T. M., & Hynek, B. M. (2007). A geochemical model for formation of Meridiani Planum layered sulfate deposits by volcanic processes. In *Seventh Interntional Conference on Mars* (p. 3257). Pasadena. Retrieved from http://www.lpi.usra.edu/meetings/7thmars2007/pdf/3257.pdf

McCollom, Thomas M., & Hynek, B. M. (2005). A volcanic environment for bedrock diagenesis at Meridiani Planum on Mars. *Nature*, *438*(7071), 1129–1131. https://doi.org/10.1038/nature04390

McEwen, A. S., Eliason, E. M., Bergstrom, J. W., Bridges, N. T., Delamere, W. A., Grant, J. A., et al. (2007). MRO's High Resolution Imaging Science Experiment (HiRISE). *Journal of Geophysical Research*.

McLennan, S. M. (2000). Chemical composition of Martian soil and rocks: Complex mixing and sedimentary transport. *Geophysical Research Letters*, *27*(9), 1335–1338. https://doi.org/10.1029/1999GL008432

McSween, H. Y., Ruff, S. W., Morris, R. V., Bell, J. F., Herkenhoff, K., Gellert, R., et al. (2006). Alkaline volcanic rocks from the Columbia Hills, Gusev crater, Mars. *Journal of Geophysical Research E: Planets*, *111*(9), 1–15. https://doi.org/10.1029/2006JE002698

McSween, H. Y., Wyatt, M. B., Gellert, R., Bell, I. F., Morris, R. V., Herkenhoff, K. E., et al. (2006). Characterization and petrologic interpretation of olivine-rich basalts at Gusev Crater, Mars. *Journal of Geophysical Research E: Planets*, *111*(2), 1–17. https://doi.org/10.1029/2005JE002477

McSween, H. Y. J., Taylor, G. J., & Wyatt, M. B. (2009). Elemental composition of the Martian crust. *Science*, *324*(5928), 736–739. https://doi.org/10.1126/science.1165871

Meyer, F. (1994). Topographic distance and watershed lines. *Signal Processing*, *38*(1), 113–125. https://doi.org/10.1016/0165-1684(94)90060-4

Michalski, J. R., Kraft, M. D., Sharp, T. G., & Christensen, P. R. (2005). Palagonite-like alteration products on the earth and Mars I: spectroscopy (0.4-25 microns) of weathered basalts and silicate alteration products. In *36th Annual Lunar and Planetary Science Conference* (Vol. 36, p. 1188).

Montgomery, D. R., Som, S. M., Jackson, M. P. A., Schreiber, B. C., Gillespie, A. R., & Adams, J. B. (2009). Continental-scale salt tectonics on Mars and the origin of Valles Marineris and associated outflow channels. *Geological Society of America Bulletin*, *121*(1–2), 117–133. https://doi.org/10.1130/B26307.1

Mustard, J. F., Poulet, F., Gendrin, A., Bibring, J.-P., Langevin, Y., Gondet, B., et al. (2005). Olivine and pyroxene diversity in the crust of Mars. *Science (New York, N.Y.)*, *307*(5715), 1594–7. https://doi.org/10.1126/science.1109098

Navarro-González, R., Vargas, E., De La Rosa, J., Raga, A. C., & McKay, C. P. (2010). Reanalysis of the Viking results suggests perchlorate and organics at midlatitudes on Mars. *Journal of Geophysical Research E: Planets*, *115*(12), 1–11. https://doi.org/10.1029/2010JE003599

Nesbit, H. W., & Wilson, R. E. (1992). Recent chemical weathering of basalts. *American Journal of Science*, *292*(10), 740–777. https://doi.org/10.2475/ajs.292.10.740

Nowicki, S. a., & Christensen, P. R. (2007). Rock abundance on Mars from the Thermal Emission Spectrometer. *Journal of Geophysical Research*, *112*(E5), E05007. https://doi.org/10.1029/2006JE002798

Ojha, L., Lewis, K., Karunatillake, S., & Schmidt, M. (2018). The Medusae Fossae Formation as the single largest source of dust on Mars. *Nature Communications*, *9*(1), 1–7.

https://doi.org/10.1038/s41467-018-05291-5

Orloff, T., Kreslavsky, M., & Asphaug, E. (2013). Distribution of polygon characteristic scale in Martian patterned ground terrain in the northern hemisphere using the Fourier transform. *Journal of Geophysical Research: Planets*, *118*, 1558–1566. https://doi.org/10.1002/jgre.20111

Orloff, T. C., Kreslavsky, M., Asphaug, E., & Korteniemi, J. (2011). Boulder movement at high northern latitudes of Mars. *Journal of Geophysical Research*, *116*(E11), E11006. https://doi.org/10.1029/2011JE003811

Pan, L., Ehlmann, B. L., Carter, J., & Ernst, C. M. (2017). The stratigraphy and history of Mars' northern lowlands through mineralogy of impact craters: A comprehensive survey. *Journal of Geophysical Research: Planets*, *122*(9), 1824–1854. https://doi.org/10.1002/2017JE005276

Pike, W. T., Lognonné, P., Banerdt, W. B., Calcutt, S. B., Standley, I. M., Giardini, D., et al. (2019). Results from the Short-Period (SP) Seismometers on the Mars Insight Mission: From Launch to Sol 40. *50th Lunar and Planetary Science Conference*, *50*(2132), 2109. Retrieved from http://adsabs.harvard.edu/abs/2019LPI....50.2109P

Putzig, N., Mellon, M., Kretke, K., & Arvidson, R. (2005). Global thermal inertia and surface properties of Mars from the MGS mapping mission. *Icarus*, *173*(2), 325–341. https://doi.org/10.1016/j.icarus.2004.08.017

Rapin, W., Meslin, P. Y., Maurice, S., Vaniman, D., Nachon, M., Mangold, N., et al. (2016). Hydration state of calcium sulfates in Gale crater, Mars: Identification of bassanite veins. *Earth and Planetary Science Letters*, *452*, 197–205. https://doi.org/10.1016/j.epsl.2016.07.045

Robbins, S. J., Hynek, B. M., Lillis, R. J., & Bottke, W. F. (2013). Large impact crater histories of Mars: The effect of different model crater age techniques. *Icarus*, *225*(1), 173–184. https://doi.org/10.1016/j.icarus.2013.03.019

Rogers, A. D., & Hamilton, V. E. (2014). COMPOSITIONAL PROVINCES OF MARS FROM STATISTICAL ANALYSES OF TES, GRS AND CRISM DATA. In *45th Lunar and Planetary Science Conference* (p. 1585). The Woodlands. Retrieved from http://www.hou.usra.edu/meetings/lpsc2014/pdf/1585

Rogers, A. Deanne, & Christensen, P. R. (2007). Surface mineralogy of Martian low-albedo regions from MGS-TES data: Implications for upper crustal evolution and surface alteration. *Journal of Geophysical Research*, *112*(E1), 1–18. https://doi.org/10.1029/2006JE002727

Rogers, A. Deanne, & Hamilton, V. E. (2014). Compositional provinces of Mars from statistical analyses of TES, GRS, OMEGA and CRISM data. *Journal of Geophysical Research : Planets*, *120*(1), 62–91. https://doi.org/10.1002/2014JE004690

Rogers, A. Deanne, Bandfield, J. L., & Christensen, P. R. (2007). Global spectral classification of Martian low-albedo regions with Mars Global Surveyor Thermal Emission Spectrometer (MGS-TES) data. *Journal of Geophysical Research*, *112*(E2), 1–29. https://doi.org/10.1029/2006JE002726

Rogers, A .D., & Aharonson, O. (2008). Mineralogical composition of sands in Meridiani Planum determined from Mars Exploration Rover data and comparison to orbital measurements. *Journal of Geophysical Research*, *113*(E6), 1–19. https://doi.org/10.1029/2007JE002995

Rogers, A Deanne, & Hamilton, V. E. (2015). Compositional provinces of Mars from statistical analyses of TES, GRS, OMEGA and CRISM data. *Journal of Geophysical Research : Planets*, *120*(1), 62–91. https://doi.org/10.1002/2014JE004690.Received

Rossum, G. van. (1995). *Python Tutorial, Technical Report CS-R9526*. Amsterdam.

Ruff, S. W., & Christensen, P. R. (2002). Bright and dark regions on Mars: Particle size and mineralogical characteristics based on Thermal Emission Spectrometer data. *Journal of Geophysical Research*, *107*(E12), 5127. https://doi.org/10.1029/2001JE001580

Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. New York, NY: McGraw-Hill Book Company.

Saunders, R. S., Arvidson, R. E., Badhwar, G. D., Boynton, W. V, Christensen, P. R., Cucinotta, F. A., et al. (2004). 2001 MARS ODYSSEY MISSION SUMMARY. *Space Science Reviews*.

Schlyter, P. (1992). Large sorted stone polygons, and ventifact distribution, in the Syrkadal area, Scania, S. Sweden. *Geografiska Annaler: Series A, Physical Geography*, *74*(2–3), 219–226.

Scott, D. H., & Tanaka, K. L. (1986). Geologic map of the western equatorial region of Mars. *USGS*.

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 1–9.

Stuurman, C. M., Osinski, G. R., Holt, J. W., Brothers, T. C., Levy, J. S., & Kerrigan, M. C. (2016). SHARAD detection and characterization of subsurface water ice deposits in Utopia Planitia, Mars. *Geophysical Research Letters*, 9484–9491. https://doi.org/10.1002/2016GL070138

Świąder, A. (2014). Identification and visualisation of possible ancient ocean shoreline on Mars using submeter-resolution Digital Terrain Models. *Geologos*, *4*, 289–301. https://doi.org/10.2478/logos-2014-0017

Tanaka, K. L., Robbins, S. J., Fortezzo, C. M., Skinner, J. A., & Hare, T. M. (2014). The digital global geologic map of Mars: Chronostratigraphic ages, topographic and crater morphologic characteristics, and updated resurfacing history. *Planetary and Space Science*, *95*, 11–24. https://doi.org/10.1016/j.pss.2013.03.006

Taylor, G. J., Martel, L. M. V., Karunatillake, S., Gasnault, O., & Boynton, W. V. (2010). Mapping Mars geochemically. *Geology*, *38*(2), 183–186. https://doi.org/10.1130/G30470.1

Taylor, G. Jeffrey, Stopar, J. D., Boynton, W. V., Karunatillake, S., Keller, J. M., Brückner, J., et al. (2006). Variations in K/Th on Mars. *Journal of Geophysical Research*, *112*(E3), 1–20. https://doi.org/10.1029/2006JE002676

Taylor, G. Jeffrey, Martel, L. M. V., Karunatillake, S., Gasnault, O., & Boynton, W. V. (2010). Mapping Mars geochemically. *Geology*, *38*(2), 183–186. https://doi.org/10.1130/G30470.1

Touma, J., & Wisdom, J. (1993). The Chaotic Obliquity of Mars. *Science*, *259*(5099), 1294–1297. https://doi.org/10.1126/science.259.5099.1294

Viviano, C., 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*, #pagerange#. https://doi.org/10.1016/j.icarus.2019.03.001

Wezel, F. C., & Baioni, D. (2010). Evidence for subaqueously resedimented sulphate evaporites on Mars. *Planetary and Space Science*, *58*(11), 1500–1505. https://doi.org/10.1016/j.pss.2010.07.003

Wyatt, M. B., & McSween, H. Y., J. (2007). TES and GRS Compositions of the Martian Surface: Evidence for Igneous and Secondary Chemical Fractionation Processes. In *Lunar and Planetary Science* (p. 2258). League City. Retrieved from http://www.lpi.usra.edu/meetings/lpsc2007/pdf/2258.pdf

# Appendix A. Supplementary Text for Chapter 2

## A.1. Materials and Methods

### A.1.1. Dataset

The GRS aboard Mars Odyssey (Saunders et al., 2004) samples the surface of Mars to decimeter depths and has extensive spatial coverage, which makes it an ideal choice for this work. Currently, GRS-derived chemical maps are available for nine elements, H (presented as stoichiometrically equivalent $H_2O$ ), S, Cl, Si, Al, Fe, Ca, K, and Th (Boynton et al., 2007; Hood et al., 2016; S Karunatillake et al., 2014) . These maps provide elemental mass fraction percentages in 5°×5° pixels, covering approximately from 45° North to 50° South, this coverage is referred to as "midlatitudinal" throughout this work. Due to low rock abundance at large scales (Nowicki & Christensen, 2007), GRS maps are taken as representative of "soil" in the planetary context: unconsolidated, chemically and mechanically weathered sediment on a planetary surface without implications for life, organic material, or habitability (Certini & Ugolini, 2013). The abundant subsurface ice in polar regions moderates neutron energies, impeding estimation of elements aside from H, K, and Th (Boynton et al., 2007; Feldman et al., 2004), as such these regions are excluded from analysis. The details of the GRS processing are discussed at length in the work by Boynton et. al. in (2007).

The Dust Cover index (DCI) is used as a proxy for surface dust abundance. This index is calculated from the average emissivity between 1350 and 1400 $cm^{-1}$ in TES spectra associated with fine-grained (~100μm diameter or smaller) silicate material (Ruff & Christensen, 2002). Other related properties, such as albedo or thermal inertia, could be used as well, however DCI more directly assesses the coverage of silicate dust where the other measures are less focused. Including more than one of these optical properties would only serve to inject strong correlations into the multivariate analysis, diluting the signal from the elemental variations. In practice, these three parameters have broadly similar spatial distributions (Ruff & Christensen, 2002) and are likely spatially very similar once coarsened to GRS-pixel scales. The data for DCI (Ruff & Christensen, 2002) were originally provided at a higher resolution (16ppd) than GRS maps, and were coarsened to 5°×5° resolution. In addition, values were inverted from the original dataset (in which low DCI values represent high dust coverage) during conversion to t-values (Section A 1.3). This modification was done so that positive correlations with DCI represented a positive correlation with dust abundance, more consistent with the rest of the data. Other variables, such as lithology, were considered as additions to the dataset, but PCA is not well-suited to examine categorical data, so these data are excluded from quantitative analysis.

### A.1.2. Principal Component Analysis implementation

To analyze the 10-variable dataset, multivariate analysis techniques are applied to reduce the dimensionality and focus on the prominent, bulk variations in soil chemistry. Two of the most common multivariate analysis techniques available are Independent Component Analysis (ICA) (Hyvärinen & Oja, 2000) and PCA (Gasnault et al., 2010; Jones, 2015), of these methods PCA is better suited for this work. ICA selects the components by maximizing *statistical independence* (Hyvärinen & Oja, 2000). Therefore, ICA is ideal for drawing out non-gaussian components(A Deanne Rogers & Hamilton, 2015) that additively combine to form the original data distribution, maximizing uniqueness between components. PCA, on the other hand, will

reduce the dimensionality of the data, emphasizing linear combinations (Principal Components, PCs) of the input data that describe as much of the variance as possible. The first two PCs (PC1, PC2) describe the primary and secondary bulk variation in the soil chemistry, collectively describing up to 50-60% of the total variance in the GRS data. Examining correlations in these two components captures bulk variance that should be associated with the primary mechanisms of soil chemistry variation and reduce the influence of less significant mechanisms.

PCA was implemented in the *Python* (Rossum, 1995) language using *matplotlib* (Hunter, 2007). Starting with chemical maps, the data were filtered to restrict analysis to regions that have measured abundance for all nine elements, so analysis is restricted to the midlatitudinal region. The PCA process imports the data and expresses it as standard deviations from arithmetic means. This processing removes any bias towards elements that have unusually high (Si) or low values (K, Th) as well as normalizing the variance. The primary result of the PCA program are the 10 orthonormal PC axes, as well as chemical maps projected into PC-space. These data serve as input into further analysis described in the following sections.

### A.1.3. Error Incorporation

One of the challenges associated with applying PCA is the lack of robust error propagation method. Previous applications of PCA have used various approaches to assess which PCs were significant (R. Anderson et al., 2015; Gasnault et al., 2010) , however few have attempted to increase the robustness of the resultant PCs. For this work, a modified Student's-t test parameter, t (also called t-values throughout), is used as shown in Karunatillake 2009 (Suniti Karunatillake et al., 2009) to incorporate uncertainty into the input to PCA.

$$t_i = \frac{x_i - \bar{x}}{\sqrt{\sigma_i^2 + s^2}}$$

Where $x_i$ is the mass fraction of an element in the *ith* bin, $\bar{x}$ is the average value of that element, $\sigma_i$ is the error on $x_i$, and $s$ is the standard deviation of the element distribution in the midlatitudinal region. Processing the data before applying PCA allows the incorporation of error without modifying the standard PCA procedure. Using t-values in place of mass fraction data produced results that were consistent with previously published findings (Gasnault et al., 2010) while increasing the robustness of the results.

### A.1.4. Regional Analyses

To examine covariation at smaller scales than the global scales observed in the broader midlatitudinal analysis, PCA is applied to twelve spatial subsets of the data (Figure B1). In general, these regions provide a diverse set of analyses to support an understanding of global to regional correlations, although some will be individually examined. Two of the regional analyses focused on the surface on either side of the topographic dichotomy that distinguishes the martian northern lowlands from the southern highlands. To form these two regions, the midlatitudinal dataset is divided along the topographic dichotomy (Figure 2.1), derived from the USGS maps (Tanaka et al., 2014). The 5°×5° grid of GRS pixels was superimposed on this mapped boundary and pixels south of the boundary were included in the southern highlands region, while pixels

intersecting and north of the boundary were included in the northern lowlands region. Despite debate on the origin of this dichotomy, there is an apparent shift in geologic age and provenance at this boundary, as well as the possibility of an ancient northern ocean (J. M. Dohm et al., 2000; Świąder, 2014) which motivates separate analyses. Additional regions are defined based on anomalously high (enriched) or low (depleted) abundances of the variables S, $H_2O$, Cl, and DCI. Here, the terms "enriched" and "depleted" are used to indicate a statistically significant deviation (±1 t-value) of the abundance of a given variable with no inference into processes that may have caused this deviation (Table B 1). By selecting regions enriched or depleted in certain variables, the variance of the selected variable is necessarily depressed (e.g. the variance of DCI is lower in DCI-enriched or DCI-depleted regions compared to the total DCI variance). In these regions, the lower variance of the defining variable (the one that is enriched or depleted) would likely preclude it from being an important component of PC1 or PC2. Therefore, these enriched and depleted regions allow insight into chemical correlations that may exclude the defining variable. Finally, two regions were defined based on the areal percentage of volcanic terrain (Tanaka et al., 2014) with each 5°×5° area classified as either volcanic (>50% volcanic terrain) or non-volcanic (<50% volcanic terrain). These regions provide more spatial subdivisions not defined by chemistry and may highlight trends within each terrain type.

### A.1.5. Correlation Analysis

The PCA process projects the original data into a new, orthonormal space defined in relation to the original data space by the PC loadings. In PC-space, the original variables (e.g. chemical abundances) can be projected as vectors using these PC-loadings. Correlation of these vectors can be measured using the cosine similarity, a commonly used measure in computer science (Salton & McGill, 1983; Singhal, 2001). In this work, the cosine similarity of two vectors is taken as the dot product of the two vectors and is analogous to a correlation coefficient. The Cosine Similarity Angle (CSA) (i.e. the angle between the two vectors in PC-space) is primarily used in this analysis and is more intuitively connected to the biplots (e.g. Figure 2.2, Figure B 2). Similar comparisons were made in previous work (Gasnault et al., 2010), although only in a qualitative fashion. Correlation analysis using the cosine similarity can easily be compared across regions, leading to insight within each analysis and across all analyses. This inter-regional comparison is necessary, because covariance between two variables in a single PC, or even a single region is not sufficient to establish a mineralogical or genetic correlation between variables. This analysis could expand into n-dimensions, but the information-compressive nature of PCA is used to focus the analysis to only the most significant components. In most analyses there was a notable drop in variance from PC2 (≥20%) to PC3 (~15%), and a further drop from PC4 (~13%) to PC5 (<10%). Because the isotropic distribution of variance in the 10-dimensional space would be ~10% variance in each dimension, PC1 and PC2 capture double the variance expected from any of the original axes, PC3 and PC4 are only slightly more significant, and PC5 and above are insignificant. To emphasize bulk trends, analysis is restricted to the 2-D space formed by PC1 and PC2 (Figure 2.2), because these two dimensions efficiently capture ≥50% of the variance. It is important to note that the use of only PC1 and PC2 in calculating the cosine similarity means that the cosine similarity is not a correlation coefficient. Instead, it is a quantified measure of correlation, limited to the PC1-PC2 space, that describes only the correlations that exist among the variables within these primary directions of variance.

A.1.6. Analysis of Cumulative PCA results

The above methods assess the correlation of the 10 variables across midlatitudinal analyses and the 12 smaller regions. The results for each region can be interpreted independently, but bulk properties that are present across all analyses are more robust. One of the challenges of this method is that the primary results, principle components, are not immediately comparable from one analysis to the next. Between each region the composition of PC1 and PC2 will change, but the PC1-PC2 space consistently represents the space of dominant variation for each region. The cosine similarity described above provides a quantitative measure of correlation that is comparable from one region to the next despite changes in the components of PC1 and PC2. Distributions of cosine similarities, plotted as cumulative distribution functions (CDFs), compile these results to reveal correlative trends among regions. For these plots, the area of each region is not factored in, considering each analysis with equal weight. The purpose of comparing all regions equally is to emphasize variation in bulk chemistry without emphasizing correlation only at large spatial scales. If an areal weighting were included, the results from the midlatitudinal analysis, as well as the pairwise complementary datasets (i.e. northern lowlands and southern highlands) would dominate the results, de-emphasizing correlations that persist at small spatial scales.

For each set of cumulative PCA results, a variable is chosen as the response variable for comparison (e.g. $H_2O$, Figure 2.3). All angles discussed are measured relative to the PC1-PC2 space vector of the response variable. The cosine similarity angle between the response variable vector and all other variables (independent variables) is calculated in PC1-PC2 space in each region as discussed in section 2.3.4. The plot shows the percentage of regions (y-axis) where the angle between the response and independent variable vectors is at or below a given angle (x-axis). For each independent variable, this plot captures the frequency and strength of correlations between the independent variable and the response variable, with greater correlation in many regions suggesting a strong correlation in bulk soil. Comparison between different independent variables informs the relative strength of correlation of the independent variables with the response variable. As discussed in Section 2.2.2, the results rely exclusively on error incorporated, t-value data for PCA. Use of t-values addresses the general impracticalities of carrying uncertainty through PCA to yield precision bounds for derived results (Gasnault et al., 2010) without the burden of Monte-Carlo methods (Haining, 2003; Suniti Karunatillake et al., 2012) to simulate the effect of uncertainty. Consequently, PCA derived results using t-values as input data, including cumulative results, effectively weigh for precision without the use of error bars.

Cumulative distributions can be quantitatively distinguished utilizing a two-tailed Kolmogorov-Smirnov (K-S) (Massey, 1951) test to estimate the likelihood that two samples are derived from the same underlying population. This test calculates a K-S statistic, D, the maximum distance between the cumulative distribution of the two samples. The D value, along with the sample size of the two distributions, determines the probability (p) that the null hypothesis, that two samples are drawn from the same population, is correct. These tests are sensitive discriminators of distributions, but poor identifiers of similarity. K-S tests on two samples that are drawn from the same population can have p values ranging from .25-1. As such, the p-value is not used to establish a continuum of distinctness, only a binary assessment of distinction. For the purposes of this work, if variables A and B have distinct distributions ($p<.05$)

of cosine similarity to the chosen response variable C, then A is more correlated to C than B is to C. If not, a distinction cannot be made between them and neither A nor B is more or less correlated to C than the other.

## A.2. Results

Here the midlatitudinal analysis that serves as a guide for further analyses is discussed in detail, and the analyses presented in the manuscript that focus on the northern lowlands, southern highlands are expanded upon. The Dust Extrema analysis, focused on only dust enriched and dust depleted regions, is also discussed. The remaining analyzed regions did not warrant individual discussion and provide data for the cumulative analysis in Section 2.3.2 and A 2.4.

For each analysis, the components of PC1 and PC2 are discussed in terms of the 10 variables used in the study (Section 2.2.1). These PCs are linear combinations of the original variable axes expressed as unit vectors, and the components of this vector are typically referred to as "weights" or "loadings". Nominally, an equal contribution from each variable would lead to a PC with individual loadings of $\pm\frac{1}{\sqrt[2]{10}}$ (~.316) in each variable (i.e. PC1=.316 $H_2O$+.316 Si+…etc.). For the purpose of discussion, only the variables that exceed this threshold are explicitly discussed, as they are increased in loading relative to an assumption of isotropy. For clarity, both the cosine similarity and the angle between variables are presented as pairs (e.g. (.42, 65°). The primary data in support of these sections is found in Table B 2.

### A.2.1. Midlatitudinal analyses

The term "midlatitudinal" is used here to represent the complete dataset after filtering out areas that had incomplete data (Figure 2.1). Key loadings in PC1 are (+Al, -Cl, -$H_2O$, -S), regions high in PC1 are marked by high Al abundances and lower volatile abundances (Table B 2). Spatially, PC1 is low in two major regions; Arabia Terra and regions to the NW and SE, and a large lobe extending from Olympus Mons westward to Isidis Basin, excluding Elysium Mons (Figure B2). Key loadings in PC2 are (-K, -Si -Th) regions with low PC2 are enriched in these three elements (Table B 2). PC2 shows consistently high values in many of the major volcanic provinces (Syrtis, Elysium, Olympus, Tharsis), while consistently negative values highlight the northern lowlands (Figure B2). This spatial distribution is consistent with "Chemical Province 3" in Jeff Taylor's work (G. J. Taylor et al., 2010), and mirrors the low K, Th, and Si that characterizes the province. In this analysis (Figure B3), $H_2O$ shows correlates with S (.98, 11°) and Cl (.97, 14°), and shows weaker correlation with DCI (.87, 29°). S shows similar patterns, correlating with $H_2O$ (.98, 11°) and Cl (1.0, 2.9°) more strongly than DCI (.95, 18°).

### A.2.2. Southern Highlands & Northern lowlands

This section expands upon the analysis of these two regions presented in the manuscript (Section 2.3.1, Figure 2.2) and provides greater details on the results from each individual region. Analysis in the southern highlands yields results that, on the surface, appear similar to those of the midlatitudinal analysis. Key loadings in PC1 are (+Al, -Cl, -$H_2O$, -S, -K, -Th), similar to the midlatitudinal analysis (Table B 2). Spatially, this similarity continues, highlighting Syrtis Major, Argyre Basin, Hellas Basin, and Sinai Planum as regions of high PC1 and Arabia Terra, Tharsis Montes, and the highlands south of Elysium Mons as low in PC1 (Figure B4). Key loadings in PC2 are (-Ca, -K, -Si, -Th, +DCI) again similar to the

midlatitudinal analysis (Table B 2). The spatial distribution of PC2 is similar to that of PC2 in midlatitudinal analyses, both highlighting volcanic provinces as high in PC2 (Figure B4). Despite these similarities, the southern highlands and midlatitudinal analyses differ greatly in the mutual correlations of S, Cl and $H_2O$ (Table 2.1).

In the northern lowlands, key loadings in PC1 are (+Cl, +$H_2O$, -K, +S, -Si, -Th) (Table B 2). This PC seems to distinguish between the volcanic dominated areas between Elysium and Olympus (high PC1) and the rest of the northern lowlands (low PC1) (Figure B5). This PC highlights volcanic regions of the surface, much like PC2 in the midlatitudinal analysis. Key loadings for PC2 are (-Al, -Ca, -Fe, -$H_2O$, -Si), and the less significant loadings are mostly negative as well (Table B 2). This trend is likely driven by the enrichment in most GRS-detected elements in Chryse Planitia and may serve as a proxy for the elements not represented in the dataset, such as Mg.

A.2.3. Dust Extrema

To directly investigate the correlations between DCI and the chemistry of martian bulk soil, chemical data from regions with DCI extrema (>1σ above or below average, see supplement section A 1.4) were analyzed with PCA (Figure B3, Figure B6). Using only these regions and excluding intermediate values of DCI, maximizes the variance of DCI and makes it likely to substantially contribute to PC1. In application, this choice biases the analysis to highlight any geochemical trends that correlate with variable dust abundance and is useful to explore any geochemical contrast between dust-rich and dust-poor regions. Due to this bias, the DCI-extrema region is excluded from the cumulative results presented elsewhere (e.g. Section 2.3.2).

Data from both the DCI-Enriched and DCI-Depleted regions are used to create a region that contains only these areas. Key loadings of PC1 are (+Al, -Cl, -$H_2O$, -S, -DCI), with DCI included in PC1 as predicted (Table B 2). The spatial distribution of PC1 in this analysis resembles the distribution of midlatitudinal PC1 (although inverted in parity), with depletions in Tempe and Tyrrhena Terrum as well as Terra Sirenum, and enrichments in Chryse Planitia, Arabia Terra, Olympus Mons and south of Elysium Planitia (Figure B6). Key loadings in PC2 are (-K, -Th), notably lacking DCI as a key factor (Table B 2). PC2 in this region highlights Olympus Mons and the surrounding areas, Elysium, and Hellas as depleted in PC2, while Chryse, Arabia, Terra Sabea, and Terra Cimmeria are enriched in PC2. Although flipped in parity, this PC closely follows the spatial distribution of the midlatitudinal PC2.

In PC1-PC2 space, $H_2O$ shows high cosine similarity with S (0.993, 7.0°) and Cl (0.999, 2.4°) while DCI correlates more weakly with $H_2O$ (0.911, 24.4°), S (.954, 17.4°), and Cl (0.927, 22.0°). The correlations among DCI and the volatile elements in the DCI-extrema region are not enhanced compared to the correlations present in the entire martian midlatitudes (Table B 2, Figure B3), supporting the initial evaluation that little meaningful correlation would be present. Alternatively, strong spatial correlation between DCI and either PC1 or PC2 could support regionally significant and chemically homogenous dust layers and reinforce any chemical correlations apparent in the results. In this analysis, the spatial distribution of PC1 does not follow enrichments and depletions in DCI and regions of enriched DCI are not uniformly enriched or depleted in PC1 or PC2 (Figure B6). Collectively, these results show that the mutual correlation among $H_2O$, S, and Cl is stronger than any correlation between these three elements and DCI, even with the choice of DCI-extrema regions. In addition, the similarity of the DCI-

extrema results with results from the midlatitudes suggests that any chemical shift between regions with high DCI to those with low DCI is either not present or too weak to influence either PC1 or PC2.

### A.2.4. Cumulative PCA Results

Here, more details of the cumulative cosine similarities which used Cl, DCI, and Fe as their response variables are discussed, as well as some direction in interpreting the cumulative results.

When assessing the correlations among the presented variables in the cumulative results, it is important to note the systematic dependence of GRS-derived chemistry due to mass balance. Chemical maps are not available for all elements (namely Mg (Boynton et al., 2007)), so the data do not sum to 100% and there is not a strict mass-balance effect. However, these mass-balance effects are still present and will reduce the true degrees of freedom, suppressing positive correlation among elements. Data that is not limited by this mass balance (i.e. DCI) is not impacted by this suppression. Therefore, any correlations between DCI and chemistry must be taken as upper limits due to the possible suppression of correlations among elements.

The cumulative PCA results for Cl (Figure B7) show overlapping CSA distributions for $H_2O$, S, and DCI. K-S tests (Table B 3) shows that indeed these three distributions are indistinct from one another, although the DCI CSA distribution is also indistinct from the Ca distribution, suggesting a weaker correlation between DCI and Cl. It is worth noting that DCI, which was expected to have generally weak correlations, shows a correlation with Cl that is comparable to the correlations among Cl, S, and $H_2O$.

As stated in section 2.2.1, the disparity in sampling depth between DCI and the chemical maps suggests that there would be little correlation between these two datasets and any correlation that exists must be examined with caution. With this caution in mind, the cumulative PCA results for DCI across all regions (Figure B8) show an apparent Cl-DCI correlation, while the remaining variables plot along a continuum. However, K-S tests show that correlation of DCI and Cl is not distinct from the other volatile elements, and is only distinct from Ca and Al, which are very poorly correlated. This lack of distinct correlation initially supports the expected conclusion, that little correlation exists between DCI and the GRS-derived chemistry. However, the cumulative results for Cl (Figure B7) show that the Cl-DCI correlation is distinct from many of the poorly correlated variables (Si, Fe, K, Th, Al) and is indistinct from the Cl-$H_2O$ and Cl-S correlations. These data suggest that each of these three variables (S, $H_2O$, and DCI) are equal predictors of Cl abundance, although as mentioned above, this represents an upper limit of DCI correlation and should be taken with caution. Contrary to expectations, there may be a causative connection between DCI and Cl abundance, however a more thorough investigation is warranted before conclusions can be made.

The cumulative results for Ca (Figure B9) are similar to the results for the remaining variables (Si, Fe, Al, K, and Th). K-S tests show that there is very little distinction among the distribution of cosine similarities of the nine variables, preventing the identification of unique correlations with Ca. Analysis of the remaining CDFs yielded similar results, and no distinct correlations are found. The exception to this pattern is that the strong correlation between K and Th is apparent in their respective CDFs, although no further correlations are noted.

## Appendix B. Supplementary Figures and Tables for Chapter 2

B.1. Figures

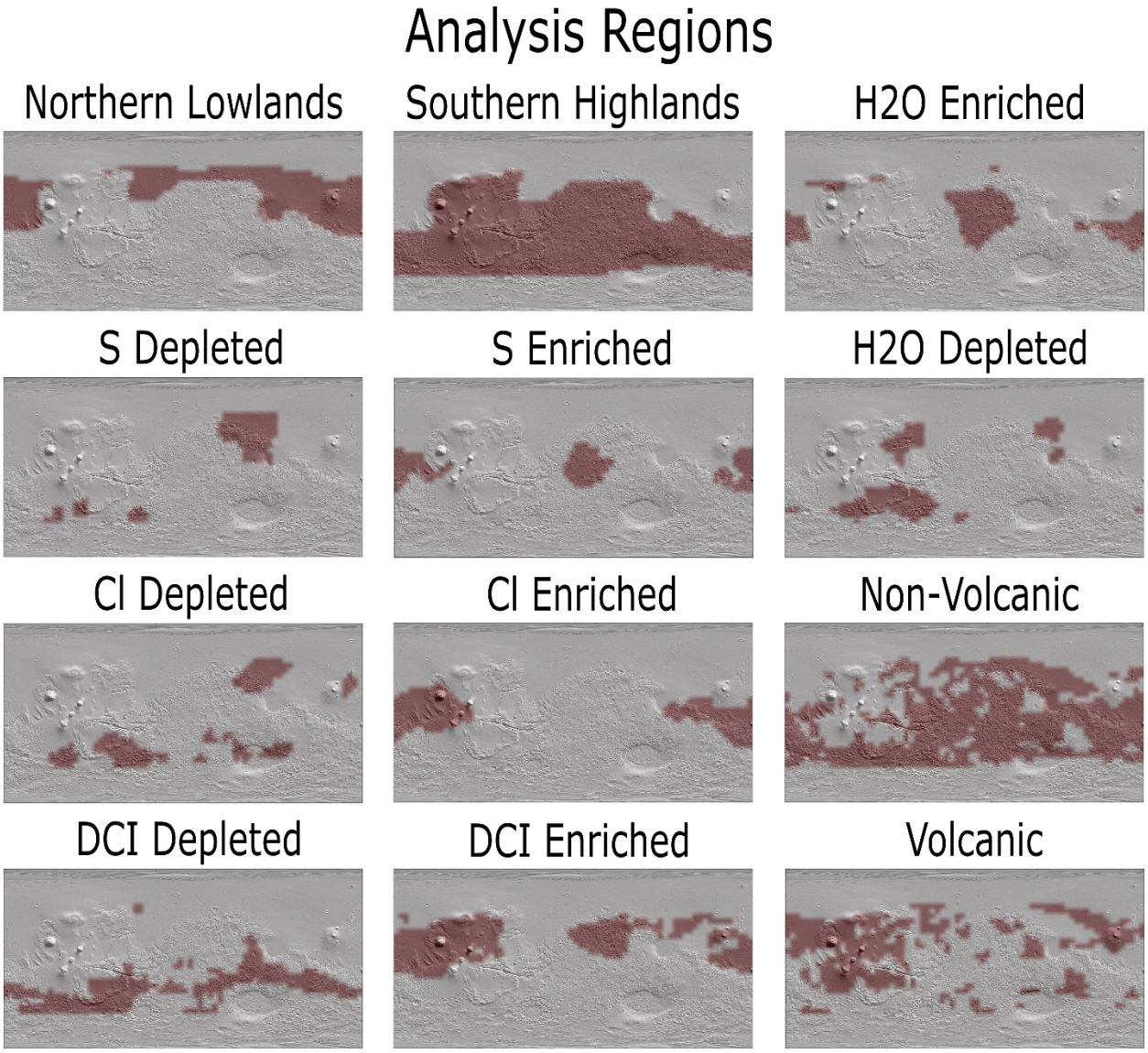



Figure B 1: Maps of 12 analysis regions overlain on MOLA hillshade. Regions enriched or depleted in a selected variable (e.g. S Enriched) are selected by separating areas that have ±1 in t-values (Sections 2.2.2, A 1.3, A 1.4), while other divisions (northern lowlands, southern highlands, volcanic, non-volcanic) are based on geologic maps of Mars (Tanaka et al., 2014). Boundaries for the midlatitudinal analysis (Section A 2.1) and DCI-extrema regions (Section A 2.3) are shown in Figure 1.1 and B 6 respectively.

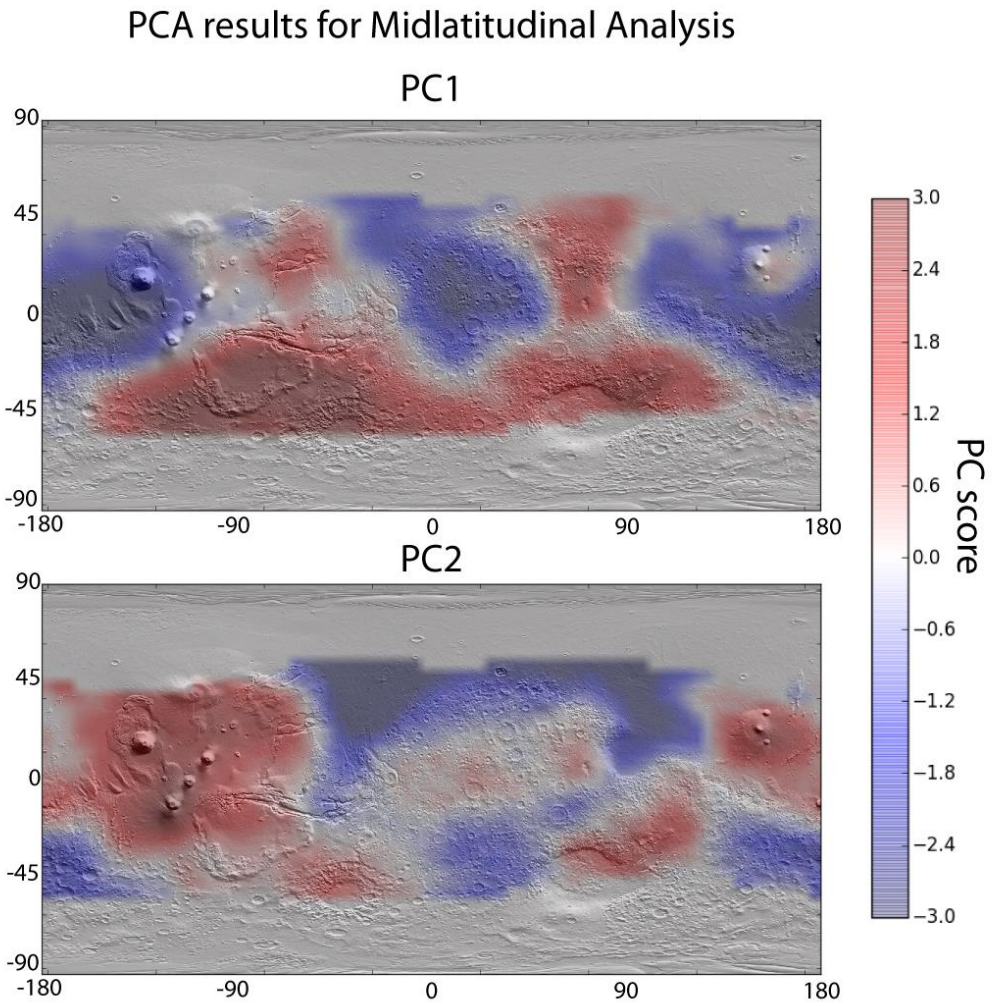



Figure B 2: Map of PC1 and PC2 in the midlatitudinal analysis overlain on MOLA hillshade (Credit: NASA/JPL). Key loadings in PC1 are (+Al, -Cl, -$H_2O$, -S), regions high in PC1 are marked by high Al abundances and lower volatile abundances. PC1 is low in two major regions; Arabia Terra and regions to the NW and SE, and a large lobe extending from Olympus Mons westward to Isidis Basin, excluding Elysium Mons. Key loadings in PC2 are (-K, -Si -Th) regions with low PC2 are enriched in these three elements. PC2 shows consistently high values in many of the major volcanic provinces (Syrtis, Elysium, Olympus, Tharsis), while consistently negative values highlight the northern lowlands.

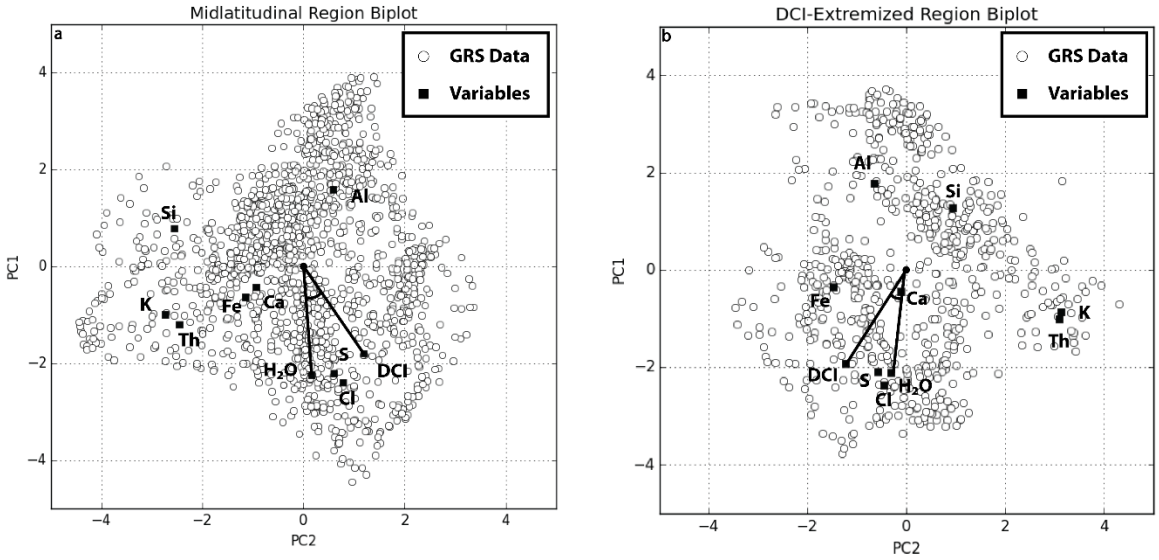



Figure B 3. An example of the PC1-PC2 space examined using cosine similarity in both the midlatitudinal (a) and DCI-extrema (b). The white circles show the GRS dataset projected into PC1-PC2 space, while the black squares represent the vectors of the original axes (i.e. the loadings of PC1 and PC2 for each variable scaled by a common factor for clarity). The relative angles of $H_2O$ and DCI are marked for clarity in each plot.

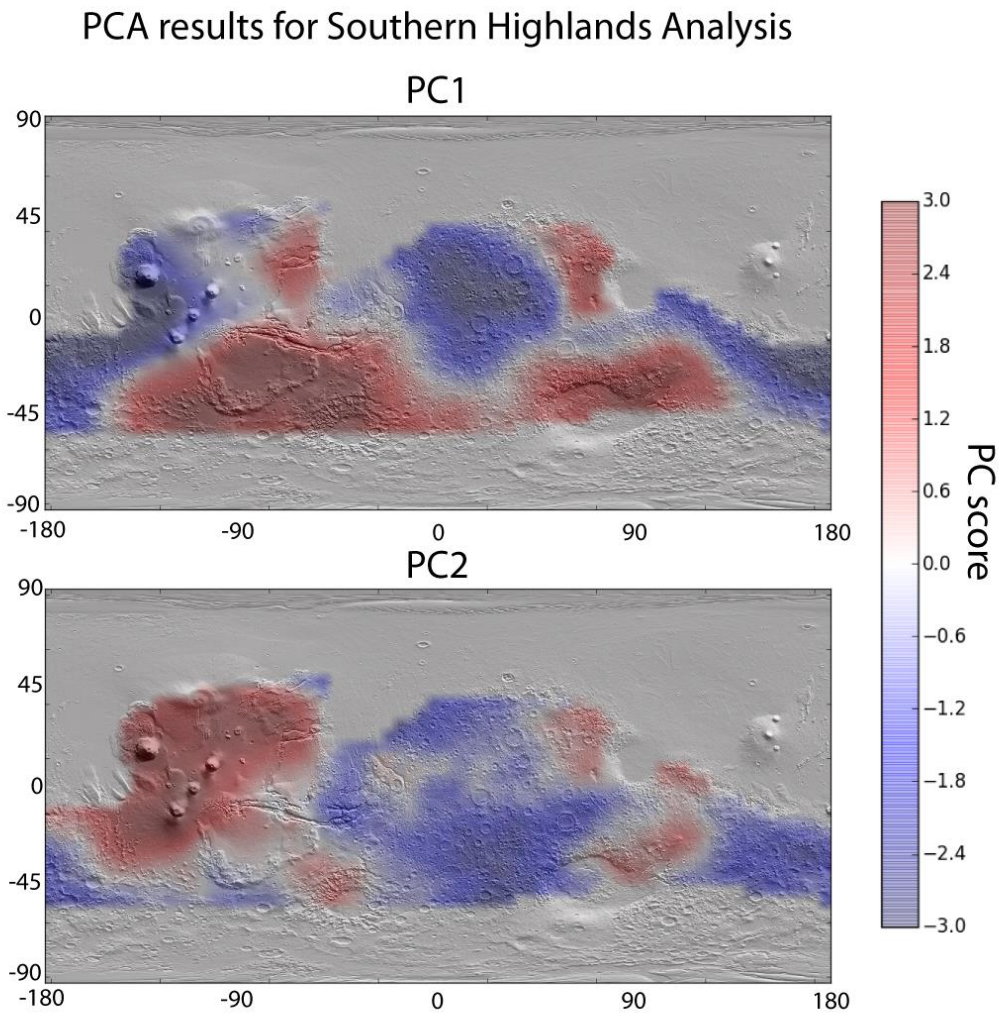



Figure B 4: Map of PC1 and PC2 in the southern highlands analysis overlain on MOLA hillshade (Credit: NASA/JPL). Key loadings in PC1 are (+Al, -Cl, -$H_2O$, -S, -K, -Th), similar to the midlatitudinal analysis. Spatially, this similarity continues, highlighting Syrtis Major, Argyre Basin, Hellas Basin, and Sinai Planum as regions of high PC1 and Arabia Terra, Tharsis Montes, and the highlands south of Elysium Mons as low in PC1. Key loadings in PC2 are (-Ca, -K, -Si, -Th, +DCI) again similar to the midlatitudinal analysis. Despite these differences in chemical trends, the spatial distribution of PC2 is similar to that of PC2 in midlatitudinal analyses, both highlighting volcanic provinces as high in PC2.

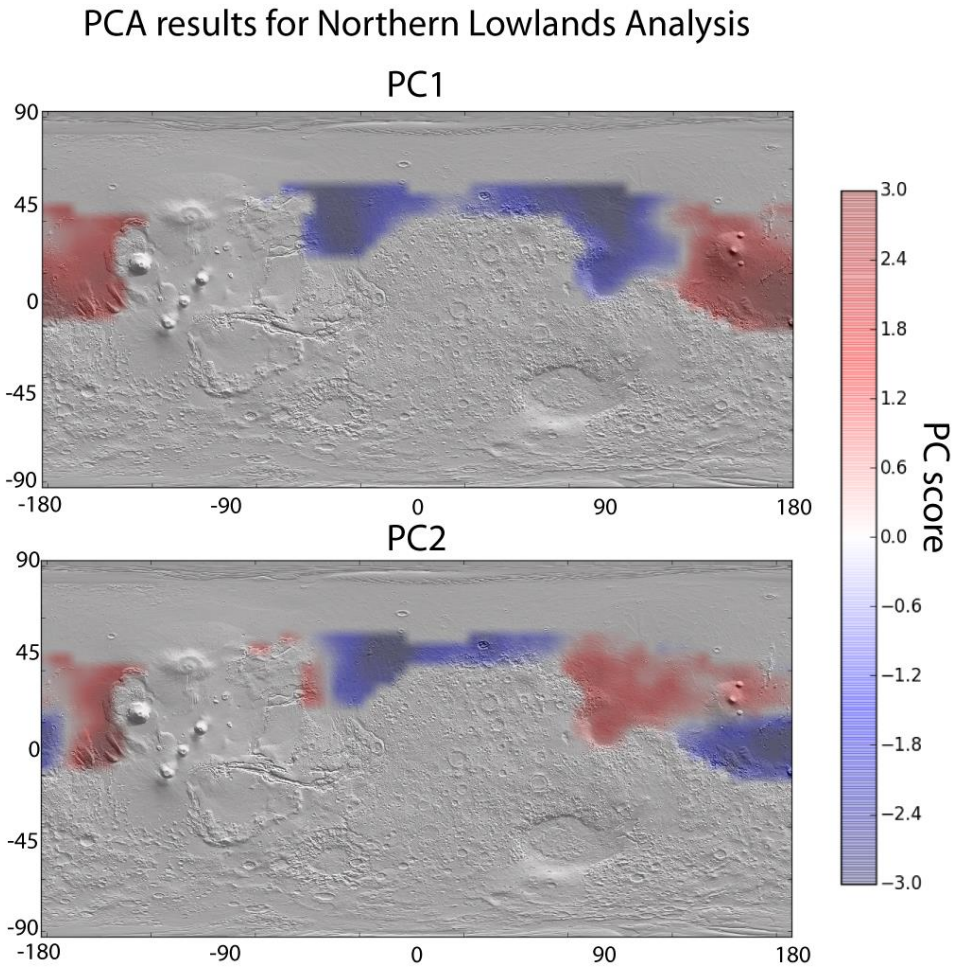



Figure B 5: Map of PC1 and PC2 in the northern lowlands analysis overlain on MOLA hillshade (Credit: NASA/JPL). Key loadings in PC1 are (+Cl, +$H_2O$, -K, +S, -Si, -Th). This PC seems to distinguish between the volcanic dominated areas between Elysium and Olympus (high PC1) and the rest of the northern lowlands (low PC1). This highlights volcanic regions of the surface, much like PC2 in the midlatitudinal analysis. Key loadings for PC2 are (-Al, -Ca, -Fe, -$H_2O$, -Si), and the less loadings are mostly negative as well. This is possibly driven by the enrichment in most GRS-detected elements in Chryse Planitia and may serve as a proxy for the elements not represented in this dataset, such as Mg.

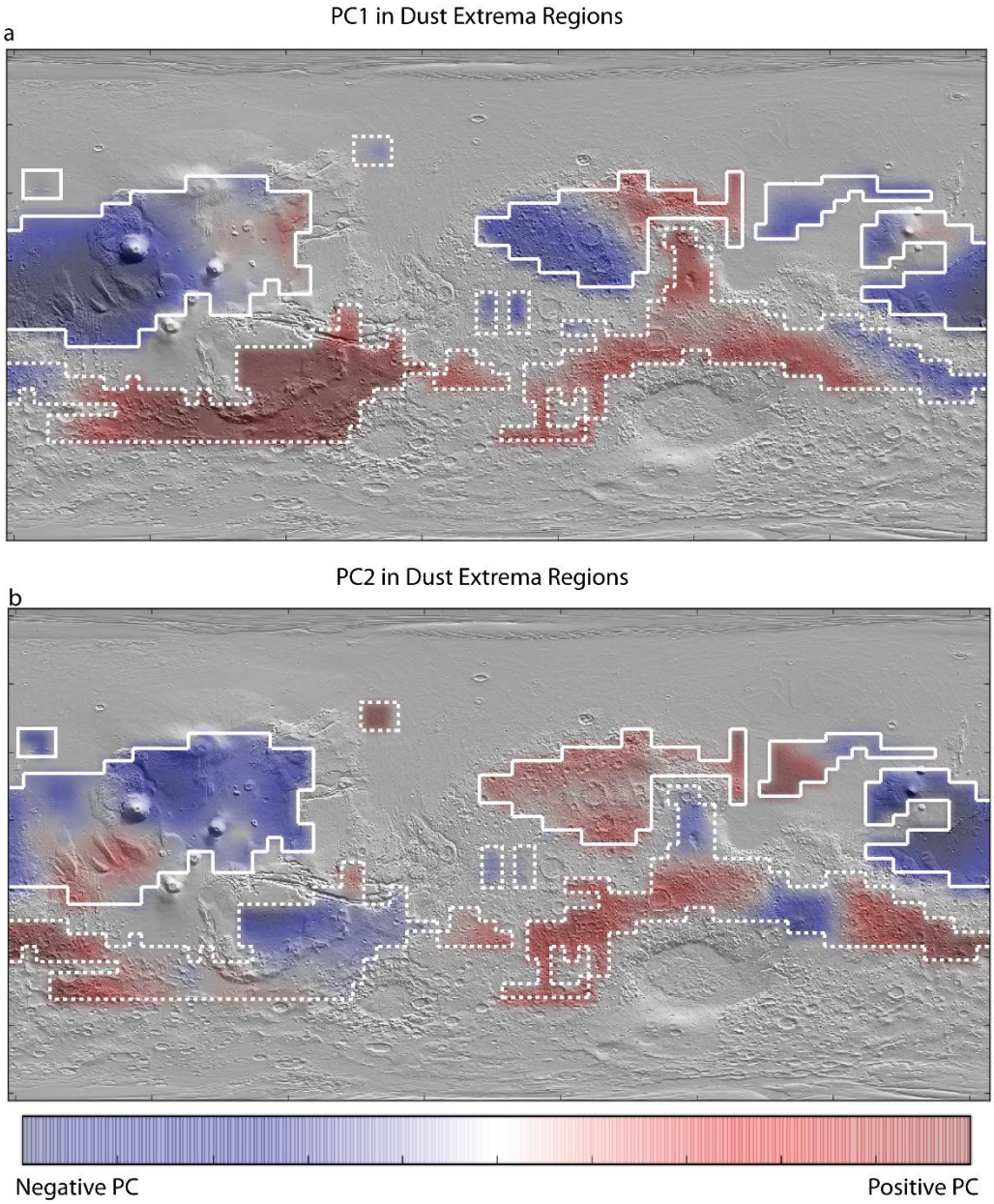



Figure B 6. A map of Mars showing the spatial distribution of PC1 (a) and PC2 (b) in the DCI-Extrema analysis. The outlines of dust enriched (white boxes) and dust depleted (dotted white boxes) are superimposed here. Note that this pattern of positive or negative values in either PC1 or PC2 does not follow high and low values of DCI and large areas of dust enrichment over Tharsis and Arabia Terra show both high and low values of PC1 and PC2.

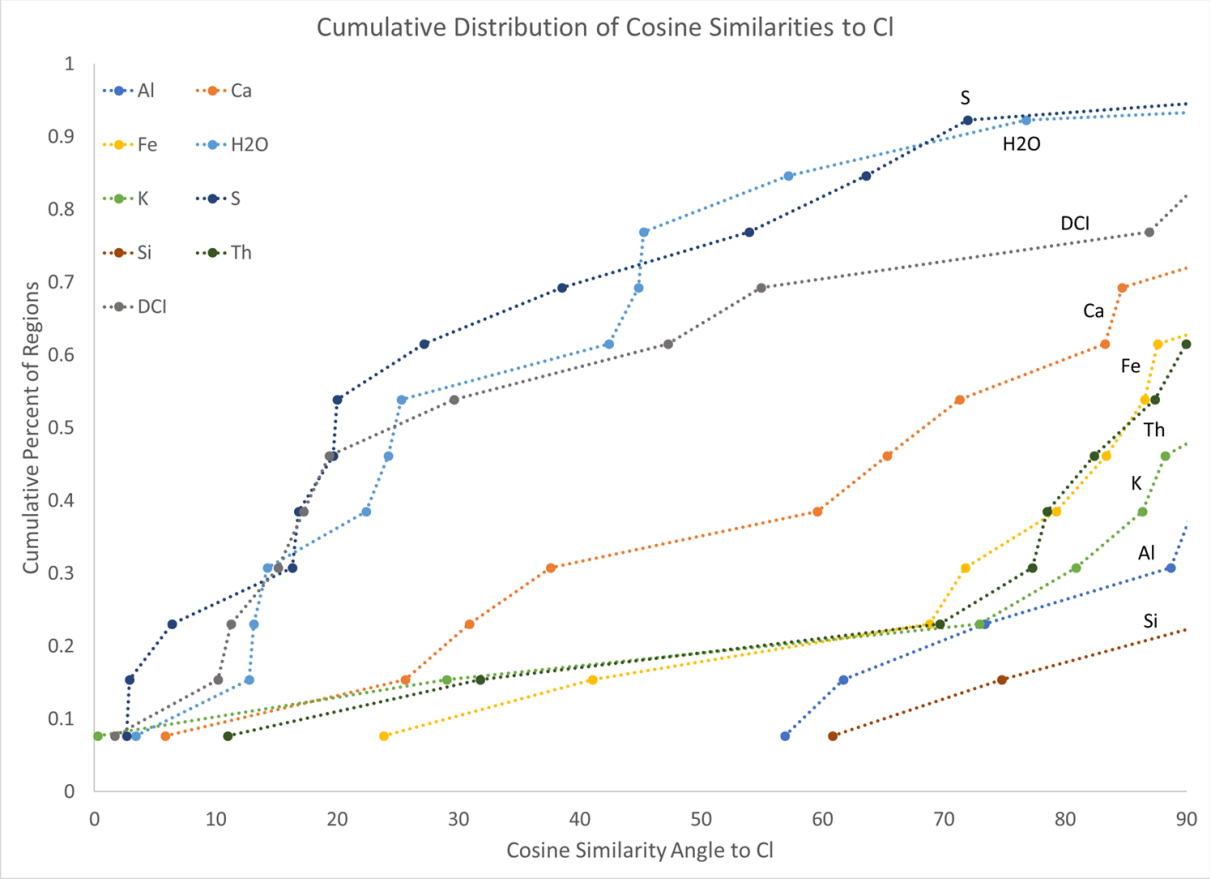



Figure B 7. Cumulative PCA results for Cl as the response variable. As with the other volatile elements, correlations with $H_2O$ and S are prominent, however the correlation with DCI is comparable here. Correlations with other variables are much weaker, although Ca deviates from this trend somewhat.

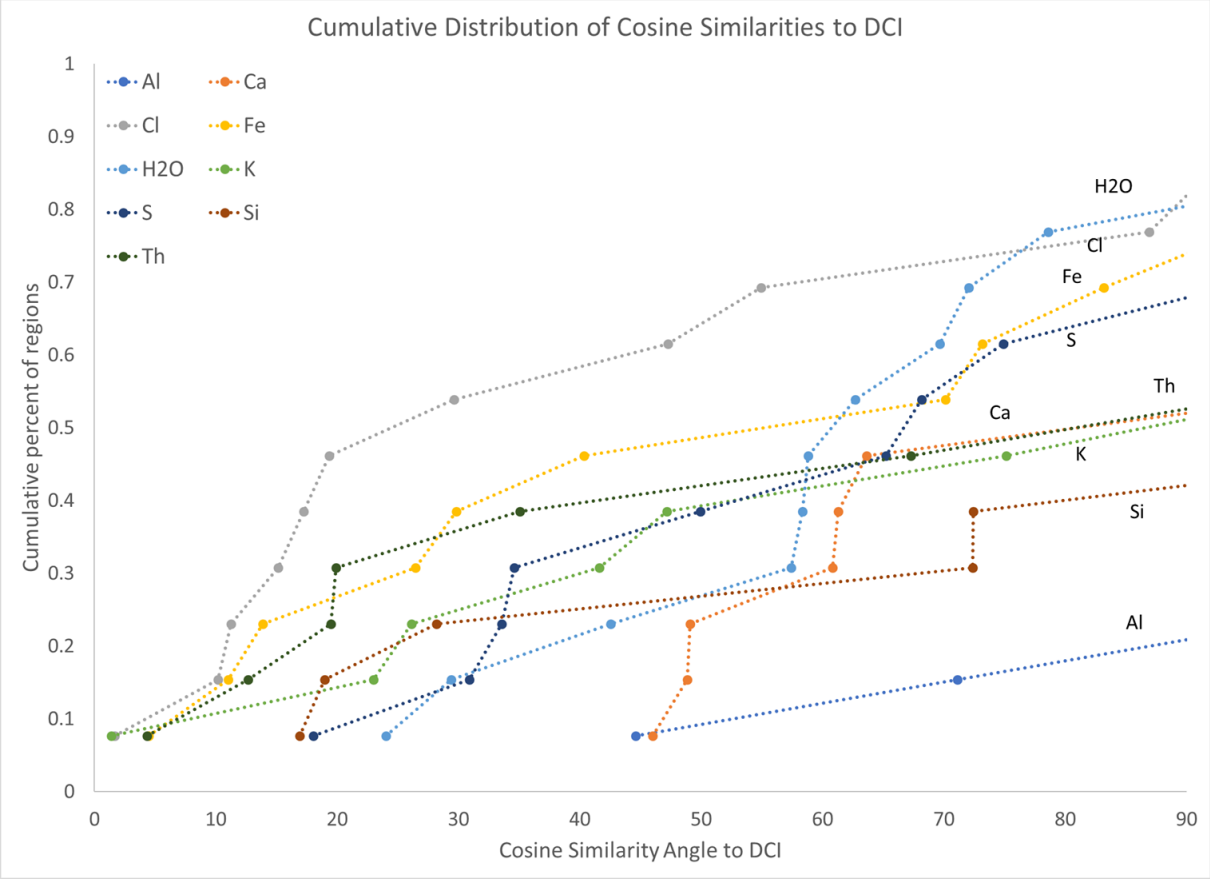



Figure B 8. Cumulative PCA results of variables vs. DCI. Of the considered variables, Cl shows the strongest correlation with DCI, although correlations with $H_2O$ and S are also present. Unlike the other cumulative results plots, there are no real outliers present in the plot, instead there is more of a continuum of cumulative correlation from Cl to Al. The K-S test results (table 2.2) aid in interpreting this plot, showing that Cl-DCI correlation is only distinct from two variables, Ca and Al.

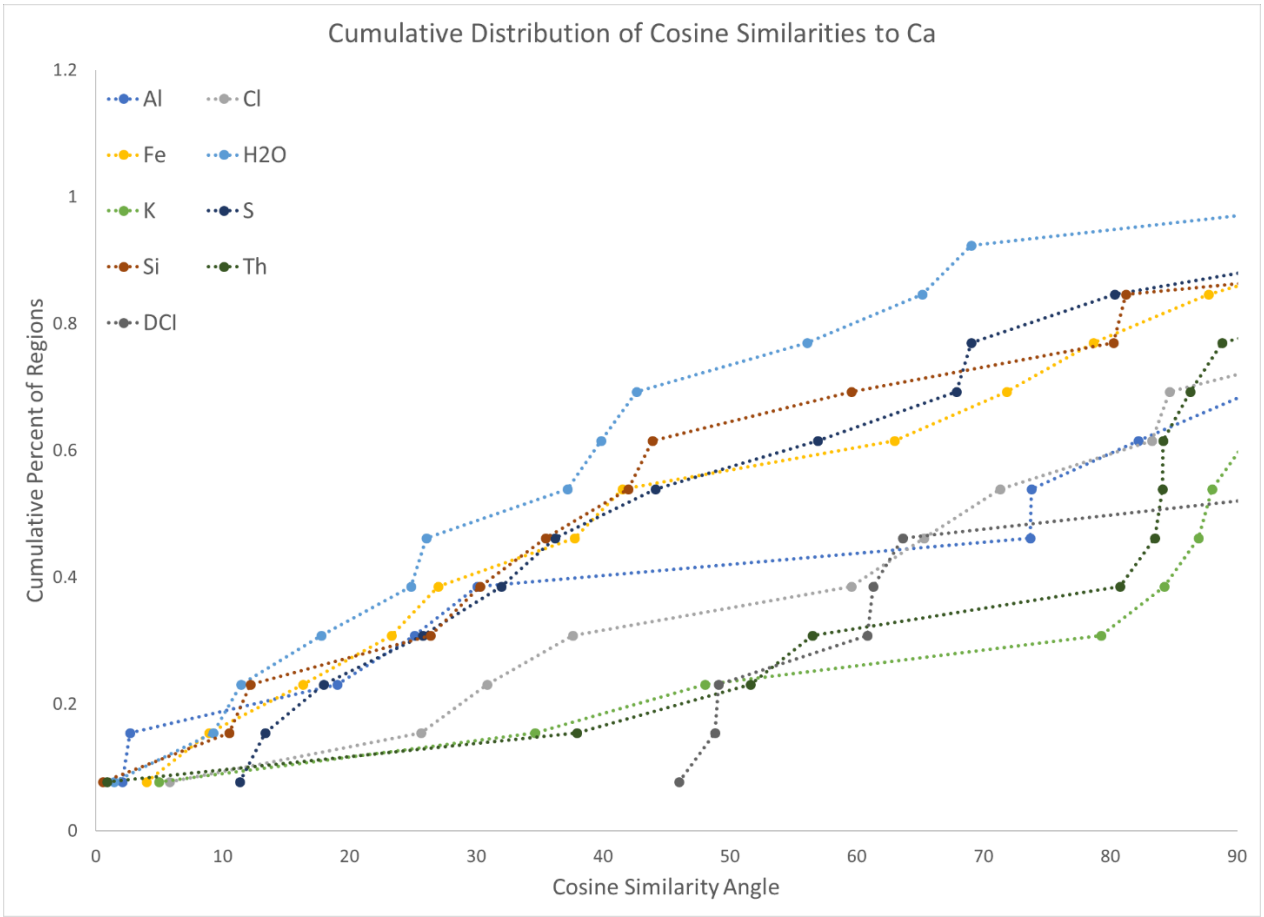



Figure B 9. Cumulative PCA results of Ca as the response variable. Despite the distinction of the Ca-$H_2O$ correlation shown in Figure 3, the same distinctness is not present here, and most distinctions among distributions are due to the very low correlation of DCI, K, and Th. CDFs for the remaining variables (Si, Al, Fe, K, and Th) show similar results lacking any distinct correlations with other variables.

## B.2. Tables

Table B 1. Table of Averages, Standard Deviations, and approximate boundaries for enriched and depleted regions. Only approximate boundaries can be given for the enriched and depleted regions because the t-value calculation (Section A 1.3) depends on the uncertainty of the individual pixel. Only regions enriched and depleted in the bolded variables were used in analysis.

| **Dataset Averages, Standard Deviations, and Boundaries for Enriched and Depleted Regions** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Element | Al (wt%) | Ca (wt%) | **Cl (wt%)** | Fe (wt%) | **$H_2O$ (wt%)** | K (wt%) | **S (wt%)** | Si (wt%) | Th (ppm) | **DCI** |
| Average | 4.48 | 5.00 | 0.47 | 12.65 | 3.90 | 0.35 | 2.18 | 19.87 | 0.61 | 0.95 |
| Standard Deviation | 1.14 | 0.78 | 0.08 | 1.36 | 1.02 | 0.06 | 0.26 | 0.88 | 0.12 | 0.03 |
| Enriched | 5.63 | 5.78 | 0.55 | 14.01 | 4.92 | 0.41 | 2.44 | 20.75 | 0.74 | 0.98 |
| Depleted | 3.34 | 4.22 | 0.40 | 11.29 | 2.89 | 0.30 | 1.92 | 18.98 | 0.49 | 0.93 |

Table B 2. Numerical results for PCA on the 13 analyzed regions. For each region, the loadings of PC1 and PC2 are given, with those exceeding an absolute value of .316 (see section A 2) highlighted in blue and red depending on the parity. Data for the first three regions, Midlatitudinal, Northern Lowlands and Southern Highlands are also found in Table 2.1. Two more regions focus only on volcanic and non-volcanic terrains. The remaining regions are either enriched in a variable (e.g. $H_2O$-Enriched regions) or depleted in a variable (H2O-Depleted). The variance captured by the first two PCs is reported below the region name, typically close to 50%, and sometimes exceeding 60%. The last listed region, DCI-Extrema, is not included the Cumulative Distribution Functions (e.g. Figure S 7).

| PCA Results for All Analyzed Regions | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Region | | Al | Ca | Cl | Fe | H2O | K | S | Si | Th | DCI |
| Midlatitudinal | PC1 | **0.318** | -0.087 | **-0.478** | -0.128 | **-0.448** | -0.199 | **-0.443** | 0.156 | -0.238 | **-0.360** |
| Var: 51% | PC2 | 0.117 | -0.187 | 0.158 | -0.231 | 0.031 | **-0.546** | 0.121 | **-0.512** | **-0.491** | 0.237 |
| | Average | 0.01 | -0.03 | -0.01 | -0.03 | -0.01 | 0.00 | -0.01 | -0.04 | 0.00 | 0.00 |
| | Std. Deviation | 0.71 | 0.64 | 0.92 | 0.87 | 0.91 | 0.98 | 0.74 | 0.61 | 0.93 | 1.00 |
| Northern Lowlands | PC1 | 0.116 | 0.091 | **0.329** | -0.127 | **0.345** | **-0.402** | **0.381** | **-0.384** | **-0.379** | **0.371** |
| Var: 63% | PC2 | **-0.512** | **-0.513** | -0.122 | **-0.417** | **-0.320** | -0.086 | -0.162 | **-0.327** | -0.093 | 0.188 |
| | Average | -0.59 | 0.07 | 0.25 | 0.76 | 0.16 | 0.46 | 0.18 | 0.17 | 0.27 | 0.53 |
| | Std. Deviation | 0.60 | 0.60 | 1.07 | 0.73 | 0.84 | 1.30 | 0.85 | 0.65 | 1.00 | 0.82 |
| Southern Highlands | PC1 | **0.361** | -0.062 | **-0.430** | 0.066 | **-0.403** | **-0.338** | **-0.387** | 0.224 | **-0.356** | -0.269 |
| Var: 53% | PC2 | -0.159 | **-0.467** | 0.202 | 0.004 | -0.126 | **-0.373** | -0.092 | **-0.534** | **-0.350** | **0.381** |
| | | 0.23 | -0.07 | -0.11 | -0.33 | -0.08 | -0.18 | -0.08 | -0.13 | -0.11 | -0.20 |
| | | 0.61 | 0.65 | 0.83 | 0.72 | 0.93 | 0.75 | 0.69 | 0.57 | 0.88 | 0.99 |
| NonVolcanic | PC1 | 0.298 | -0.139 | **-0.457** | -0.198 | **-0.445** | -0.227 | **-0.419** | 0.139 | -0.278 | **-0.347** |
| Var: 52% | PC2 | **0.366** | 0.203 | 0.148 | -0.085 | 0.264 | **-0.574** | **0.323** | -0.218 | **-0.493** | 0.041 |
| | Average | 0.09 | 0.06 | -0.08 | -0.06 | 0.02 | 0.13 | -0.01 | 0.08 | 0.14 | -0.19 |
| | Std. Deviation | 0.69 | 0.66 | 0.94 | 0.88 | 0.99 | 0.90 | 0.80 | 0.46 | 0.86 | 0.95 |
| Volcanic | PC1 | 0.200 | 0.214 | **-0.460** | 0.306 | -0.261 | 0.119 | -0.316 | **0.469** | 0.084 | **-0.447** |
| Var: 55% | PC2 | **-0.378** | 0.119 | 0.186 | 0.240 | 0.282 | **0.522** | 0.249 | 0.268 | **0.514** | 0.037 |
| | Average | -0.16 | -0.20 | 0.12 | 0.05 | -0.09 | -0.28 | -0.01 | -0.28 | -0.28 | 0.35 |
| | Std. Deviation | 0.71 | 0.56 | 0.86 | 0.84 | 0.72 | 1.07 | 0.62 | 0.76 | 0.99 | 1.00 |
| Cl Depleted | PC1 | **0.481** | -0.043 | 0.162 | **-0.378** | -0.030 | **-0.468** | 0.237 | -0.239 | **-0.422** | -0.293 |
| Var: 60% | PC2 | -0.171 | **0.529** | **0.423** | -0.016 | **0.531** | 0.009 | **0.459** | 0.021 | 0.034 | 0.131 |
| | Average | 0.31 | -0.26 | -1.22 | -0.09 | -0.91 | -0.26 | -0.73 | 0.21 | -0.43 | -0.45 |
| | Std. Deviation | 0.81 | 0.62 | 0.20 | 0.79 | 0.62 | 1.03 | 0.59 | 0.43 | 0.93 | 0.88 |
| Cl Enriched | PC1 | **0.391** | **0.424** | 0.099 | **0.453** | **0.403** | -0.103 | 0.244 | **0.447** | -0.131 | -0.053 |
| Var: 57% | PC2 | -0.018 | -0.005 | 0.281 | -0.138 | 0.192 | **0.603** | 0.232 | 0.078 | **0.595** | -0.295 |
| | Average | -0.47 | -0.12 | 1.59 | 0.00 | 0.68 | -0.01 | 0.80 | -0.58 | 0.06 | 0.86 |
| | Std. Deviation | 0.61 | 0.68 | 0.42 | 1.23 | 0.81 | 0.68 | 0.54 | 0.67 | 0.61 | 0.73 |
| DCI Depleted | PC1 | **0.384** | -0.258 | **-0.370** | 0.137 | **-0.383** | **-0.375** | -0.313 | 0.211 | **-0.413** | -0.172 |
| Var: 55% | PC2 | -0.221 | **-0.634** | 0.231 | **0.562** | -0.181 | 0.236 | -0.193 | -0.126 | 0.158 | 0.114 |
| | Average | 0.41 | -0.10 | -0.44 | -0.55 | -0.49 | -0.13 | -0.33 | 0.10 | -0.07 | -1.16 |
| | Std. Deviation | 0.64 | 0.71 | 0.69 | 0.59 | 0.69 | 0.95 | 0.54 | 0.34 | 0.88 | 0.10 |
| DCI Enriched | PC1 | **0.356** | **0.498** | -0.094 | **0.493** | 0.292 | -0.016 | 0.079 | **0.496** | 0.038 | 0.184 |
| Var: 49% | PC2 | -0.027 | 0.131 | **0.570** | -0.074 | **0.410** | -0.147 | **0.610** | -0.188 | -0.098 | -0.211 |
| | Average | -0.29 | -0.13 | 0.61 | 0.04 | 0.48 | -0.19 | 0.29 | -0.46 | -0.08 | 1.33 |
| | Std. Deviation | 0.70 | 0.60 | 0.99 | 0.96 | 0.91 | 0.75 | 0.85 | 0.67 | 0.80 | 0.22 |
| H2O Depleted | PC1 | **0.479** | -0.134 | -0.218 | **-0.388** | -0.134 | **-0.460** | -0.009 | -0.137 | **-0.442** | **-0.332** |
| Var: 54% | PC2 | -0.074 | -0.265 | **-0.571** | -0.002 | **-0.426** | 0.239 | **-0.525** | 0.221 | 0.169 | -0.084 |
| | Average | 0.38 | -0.50 | -0.88 | 0.05 | -1.34 | -0.32 | -0.65 | 0.17 | -0.39 | -0.56 |
| | Std. Deviation | 0.93 | 0.38 | 0.55 | 0.79 | 0.27 | 1.06 | 0.45 | 0.37 | 0.97 | 0.81 |
| H2O Enriched | PC1 | -0.087 | -0.276 | **-0.421** | **-0.565** | -0.089 | **0.364** | -0.253 | -0.068 | **0.440** | -0.130 |
| Var 48% | PC2 | **-0.492** | **-0.391** | 0.078 | -0.134 | -0.203 | **-0.376** | -0.240 | **-0.440** | **-0.383** | -0.020 |
| | Average | 0.13 | 0.26 | 0.77 | 0.13 | 1.57 | 0.08 | 0.92 | -0.09 | 0.35 | 0.69 |
| | Std. Deviation | 0.33 | 0.52 | 0.91 | 0.82 | 0.43 | 0.53 | 0.68 | 0.37 | 0.63 | 0.87 |
| S Depleted | PC1 | **0.359** | -0.283 | -0.049 | **-0.412** | -0.183 | **-0.373** | 0.280 | **-0.356** | **-0.411** | -0.270 |
| Var: 63% | PC2 | 0.232 | **0.397** | **0.606** | -0.128 | **0.574** | -0.237 | 0.117 | 0.067 | -0.014 | -0.030 |
| | Average | -0.01 | -0.33 | -1.02 | 0.36 | -0.72 | 0.21 | -1.30 | 0.34 | 0.22 | 0.03 |
| | Std. Deviation | 0.75 | 0.49 | 0.46 | 0.72 | 0.62 | 1.04 | 0.20 | 0.42 | 0.74 | 1.09 |
| S Enriched | PC1 | **-0.423** | **-0.413** | 0.155 | **-0.401** | -0.264 | 0.274 | -0.290 | **-0.400** | 0.247 | 0.128 |
| Var 61% | PC2 | 0.257 | 0.008 | -0.269 | -0.298 | **0.411** | **0.431** | 0.075 | 0.095 | **0.527** | **-0.356** |
| | Average | -0.09 | 0.20 | 1.11 | 0.18 | 1.38 | -0.01 | 1.37 | -0.30 | 0.17 | 0.75 |
| | Std. Deviation | 0.59 | 0.62 | 0.96 | 1.04 | 0.75 | 0.60 | 0.27 | 0.48 | 0.54 | 0.85 |
| DCI-Extrema* | PC1 | **0.354** | -0.089 | **-0.473** | -0.069 | **-0.422** | -0.172 | **-0.420** | 0.255 | -0.204 | **-0.384** |
| Var: 53% | PC2 | -0.132 | -0.021 | -0.089 | -0.296 | -0.061 | **0.626** | -0.114 | 0.187 | **0.618** | -0.246 |
| | Average | 0.03 | -0.12 | 0.13 | -0.22 | 0.04 | -0.16 | 0.01 | -0.21 | -0.07 | 0.21 |
| | Std. Deviation | 0.76 | 0.65 | 1.01 | 0.86 | 0.95 | 0.85 | 0.79 | 0.61 | 0.84 | 1.25 |

Table B 3. KS-Test results (p-values) for $H_2O$, S (A), Cl, and DCI (B) with tables divided along the diagonal. P-values that show confident distinction ($p<.05$) between two distributions are highlighted in green. In the H2O results (Table A, blue highlight): S, Cl, and Ca are all indistinct from one another and distinct from all other variables, separating them as uniquely well correlated with $H_2O$. The remaining variables show little distinction, except Al which is distinctly uncorrelated. S results (Table A, yellow highlight) show that $H_2O$ and Cl are indistinct from one another, but Cl is also indistinct from Ca. In the table for Cl (Table B, gray highlight) the correlation of $H_2O$, S, and DCI with Cl are indistinct from one another. The correlation between DCI and Cl is also indistinct from the Ca-Cl correlation, suggesting possibly lower significance of DCI compares to $H_2O$ and S. Results for DCI (Table B, orange highlight) show very few distinctions among the variables, and the Cl distribution is only distinct from the distributions of Al and Ca which are the least correlated with DCI.

| K-S test Results for H2O and S | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | Al | Ca | Cl | Fe | H2O | K | S | Si | Th | DCI |
| Al | - | 0.0870 | **0.0280** | 0.8281 | **0.0017** | 0.4889 | - | 0.0870 | 0.4889 | 0.8281 |
| Ca | 0.0280 | - | 0.4889 | 0.0870 | **0.0280** | **0.0280** | - | **0.0076** | 0.0870 | 0.4889 |
| Cl | 0.0280 | 0.9950 | - | **0.0076** | 0.8281 | **0.0017** | - | **0.0017** | **0.0076** | **0.0280** |
| Fe | 0.2264 | 0.0076 | 0.0017 | - | **0.0001** | 0.8281 | - | 0.2264 | 0.8281 | 0.4889 |
| H2O | - | - | - | - | - | **0.0001** | - | **0.0003** | **0.0001** | **0.0003** |
| K | 0.8281 | 0.0076 | 0.0017 | 0.8281 | - | - | - | 0.8281 | 0.9950 | 0.4889 |
| S | 0.0076 | 0.2264 | 0.4889 | 0.0001 | - | 0.0001 | - | - | - | - |
| Si | 0.9950 | 0.0076 | 0.0076 | 0.2264 | - | 0.8281 | 0.0003 | - | 0.4889 | 0.4889 |
| Th | 0.2264 | 0.0076 | 0.0017 | 0.9950 | - | 0.8281 | 0.0001 | 0.2264 | - | 0.4889 |
| DCI | 0.4889 | 0.0280 | 0.0076 | 0.4889 | - | 0.4889 | 0.0017 | 0.2264 | 0.4889 | - |
| | | | | | | | | | | |
| K-S Test Results for Cl and DCI | | | | | | | | | | |
| **B** | Al | Ca | Cl | Fe | H2O | K | S | Si | Th | DCI |
| Al | - | **0.0280** | **0.0017** | **0.0076** | **0.0076** | 0.0870 | **0.0280** | 0.8281 | 0.0870 | - |
| Ca | 0.0870 | - | **0.0280** | 0.0870 | 0.4889 | 0.4889 | 0.4889 | 0.2264 | 0.2264 | - |
| Cl | - | - | - | 0.8281 | 0.0870 | 0.2264 | 0.0870 | 0.0870 | 0.2264 | - |
| Fe | 0.2264 | 0.4889 | - | - | 0.4889 | 0.4889 | 0.4889 | 0.2264 | 0.4889 | - |
| H2O | 0.0003 | 0.0280 | - | 0.0017 | - | 0.4889 | 0.9950 | 0.0870 | 0.4889 | - |
| K | 0.8281 | 0.2264 | - | 0.8281 | 0.0017 | - | 0.8281 | 0.4889 | 0.8281 | - |
| S | 0.0003 | 0.0870 | - | 0.0017 | 0.8281 | 0.0001 | - | 0.4889 | 0.8281 | - |
| Si | 0.4889 | 0.0280 | - | 0.0280 | 0.0001 | 0.0017 | 0.0003 | - | 0.4889 | - |
| Th | 0.4889 | 0.4889 | - | 0.8281 | 0.0017 | 0.8281 | 0.0001 | 0.0870 | - | - |
| DCI | 0.0017 | 0.2264 | - | 0.0280 | 0.8281 | 0.0280 | 0.0017 | 0.0017 | 0.0280 | - |

# Appendix C. MBARS Code and Scripts

## C.1. MBARS_RUN

```
import MBARS
import matplotlib.pyplot as plt
import scipy.misc as spm
import numpy as np
import time
import cPickle as pickle
import threading
'''
set number of images, this is expecting a series of images with the same root name
and numerals at the end, i.e. "root0.PNG" and "root23.PNG"
'''
filenames = []

#filename += 'ESP_011357_2285_RED300PX'


#filenameA = 'ESP_036437_2290_RED500PX'
#filenames  = [filenameB,filenameC,filenameD]

#GOlombek Comparison Images
filenames += ['TRA_000828_2495_RED500PX']
#filenames +=['TRA_000828_2495_RED_16bit']
#filenames += ['PSP_001391_2465_RED500PX']
#filenames += ['PSP_001391_2465_RED16bit500PX']

#viking 1 lander setup:

#filenames += ['PSP_001521_2025_RED16bit100PNL44']
#filenames +=['PSP_001719_2025_RED16bit100PNL52_500PX']
#filenames +=['ESP_046170_2025_RED16bit100PNL52_500PX']



#viking 2 lander images
#filenames+=['PSP_001976_2280_RED16bit100PNL52_500PX']
#filenames+=['PSP_002055_2280_RED16bit100PNL57_500PX']
#filenames+=['PSP_001501_2280_RED16bit100PNL47']


#PSP_007718 subset images
#filenames = ['PSP_007718_2350_']
#filenames += ['PSP_007718_2350_RED16bit500PX']
#filenames += ['PSP_007718_2350_RED300px']

#JoesImages
#filenames += ['PSP_007693_2300_RED500PX']

#Proposal Test Images
#filenames += ['PSP_001415_2470_RED500PX']
#filenames+= ['PSP_001415_2470_RED16bit500PX']


######SOME CONTROLS##################
#produce intermediate plots, False unless you are debugging something
plot = False

#for continuing broken runs, use Startat to specify which panel to begin on for the first run
#Keep in  mind that threaded runs do not complete the files in order, use with caution
startat = 0

```

1

```
#Process is largely processer-limited, so benefit to large number of threads is minimal
#setting no limit causes memory errors.
thread_limit = 2

def core(num,gam,plot,manbound,bound,odr_keycard):
    '''core function of the run file, does gamfun and boulder detect
    '''
    #seg, good, runfile = MBARS.gamfun(num,gam, plot,manbound)
    seg,good,runfile = MBARS.autobound(num,bound)
    #print 'step 1 done'
    if good:
        if any(seg.compressed()):
            bads = MBARS.boulderdetect_threadsafe(num,seg,runfile,odr_keycard)
            #MBARS.overlapcheck_threadsafe_DBSCAN(num,runfile, odr_keycard, overlap=.001)
            MBARS.overlapcheck_shadbased(num,runfile,odr_keycard)
    if num%200 == 0:
        print 'Done with image %s'%(num)
    return runfile

def thread_run(filename,plot,startat):

    MBARS.FNM, MBARS.ID, MBARS.NOMAP,panels = MBARS.RunParams(filename)

    MBARS.PATH = 'C://Users//dhood7//Desktop//MBARS//Images//%s//'%(MBARS.FNM)
    MBARS.INANGLE, MBARS.SUNANGLE, MBARS.RESOLUTION, MBARS.NAZ, MBARS.SAZ, MBARS.ROTANG = MBARS.s
    #mangam,manbound = MBARS.FindIdealParams(filename,True)
    bound = MBARS.getimagebound(panels)
    mangam = 0
    manbound = 0
    t1 = time.clock()

    threads = []
    krange = range(startat,panels)
    print '%s images to run'%(panels)
    odr_keycard = threading.Lock()
    threads = [threading.Thread(target = core, args=(a,mangam,plot,manbound,bound,odr_keycard),na
    count=0
    for i in range(len(threads)):
        runfile = threads[i].start()
        while threading.active_count() > thread_limit:
            #print 'Waiting at %s on thread room'%(i.name)
            time.sleep(5)
        # make sure the dragging end doesnt get too far behind
        # This way, when it encounters a big image it wont move on too far
        if i>thread_limit:
            threads[i-thread_limit].join()
            if (i-thread_limit)%200 == 0:
                print 'completed thread %s'%(threads[i-thread_limit].name)

    #make sure it does not execute any more code until all threads done
    for i in threads:
        i.join()

    startat = 0
    t2 = time.clock()
    ttime = (t2-t1)/3600.
    print ('total time: '+str(ttime)+'hours')
    #this is to note the last running conditions:
    string = "This image was last run with the parameters mangam = %s, manbound=%s. \nIt took %s
    record = open('%s%s%s_runinfo.txt'%(MBARS.PATH,runfile,MBARS.FNM),'w')
```

```
    for item in string:
        record.write(item)
    record.close()
    return

#setup all the parameters before running
#for i in filenames:
    #mangam, manbound = MBARS.FindIdealParams(i)

for i in filenames:
    thread_run(i,plot,startat)
```

3

## C.2. MBARS_ANALYSIS

```
import MBARS
import matplotlib.pyplot as plt
import scipy.misc as spm
import numpy as np
import time
import cPickle as pickle
import shutil
import matplotlib.patches as patches
import os
from numpy import ma as npma
import skimage.transform as sktrans
import skimage.util as skutil
import imageio
'''
Catch all code to look at results
'''


#Sholes Images
#filename = 'PSP_007718_2350_'
#filename = 'PSP_007718_2350_RED300px'
#filename = 'PSP_007718_2350_RED16bit500px'
#runfile='autobound//'
#runfile='gam600_manbound102//'

#Golombek Comparison Images
#filename = 'TRA_000828_2495_RED500PX'
#filename = 'TRA_000828_2495_RED_16bit'
runfile = 'autobound//'
#runfile = 'gam600_manbound159//'
#filename = 'PSP_001391_2465_RED500PX'
#filename = 'PSP_001391_2465_RED16bit500PX'
#runfile = 'gam600_manbound135//'

#viking 1 lander images:
#filename = 'PSP_001521_2025_RED100PNL47_500PX'
filename = 'PSP_001521_2025_RED16bit100PNL44'
#filename = 'PSP_001719_2025_RED100PNL52_500PX'
#filename = 'ESP_046170_2025_RED_100PNL52_500PX'

#filename = 'PSP_001521_2025_RED16bit100PNL44'
#filename ='PSP_001719_2025_RED16bit100PNL52_500PX'
#filename ='ESP_046170_2025_RED16bit100PNL52_500PX'

#Viking 2 lander:
#filename = 'PSP_001501_2280_RED100PNL47_500PX'
#filename = 'PSP_001976_2280_RED100PNL52_500PX'
#filename = 'PSP_002055_2280_RED100PNL57_500PX'

#filename='PSP_001976_2280_RED16bit100PNL52_500PX'
#filename='PSP_002055_2280_RED16bit100PNL57_500PX'
#filename='PSP_001501_2280_RED16bit100PNL47'

#Joes Images
#filename = 'PSP_007693_2300_RED500PX'
#runfile = 'gam600_manbound173//'

#Proposal Test Images
#filename = 'PSP_001415_2470_RED500PX'
#runfile = 'gam600_manbound85//'
#filename = 'PSP_001415_2470_RED16bit500PX'
#runfile='autobound//'

```

```
#specify which running file to examine, include '//' on the end of the name

#runfile = 'gam600_manbound147//'

MBARS.PATH = 'C://Users//dhood7//Desktop//MBARS//Images//%s//'%(filename)
MBARS.FNM = filename
root, MBARS.ID, MBARS.NOMAP, panels = MBARS.RunParams(filename)
MBARS.INANGLE, MBARS.SUNANGLE, MBARS.RESOLUTION, MBARS.NAZ, MBARS.SAZ, MBARS.ROTANG = MBARS.start


######################PICK THE ANALYSES#################
ManualMerge = False
OutToGIS = True
MakeCFAs = False
bigs = False
imageanalysis = True

#Can manually reduce the panels to look at results subsets
#panels = 3000

#manual Merge Controls:
#uses trgetfile as the runfile
#A and B flags mark the boulders to merge
MMnum = 28
MMflags = [243,244]

#controls for making CFAs
#determines diameter past which the program does not count boulders
maxd = 10

#controls for examining images & bigs
#diameter past which boulders are considered "big" and plotted
maxdiam = 2
#when running exmaine images, do you want to see images with no boulders?
showblanks = False


#check for the proper path:
if not os.path.exists('%s%s'%(MBARS.PATH,runfile)):
    print 'runfile does not exist for this image, check runfile'
    exit()
#where the actual work happens
if ManualMerge:
    query = 'Did you mean to do a manual Merge?\n y/n?\n'
    answer = raw_input(query)
    if answer != 'y':
        print 'OK I wont do it'
    else:
        MBARS.ManualMerge(runfile,MMnum,MMflags)
if OutToGIS:
    MBARS.OutToGIS(runfile,panels)

if MakeCFAs:
    record = file('%s%s_record.csv'%(MBARS.PATH,MBARS.FNM),mode='wb')
    record.write('Filename ,Best Fit Rock Abundance,R^2,Upper limit RA, lower limt RA, maxd=%s\n'
    plt.figure(runfile)
    fit_k,upfit_k,downfit_k, fit_r2 = MBARS.bulkCFA(runfile,panels+1,maxd,2.25,root)

    print'Best fit rock abundance for file %s is %s percent, up to %s, or down to %s with an R^2
    #MBARS.plotCFArefs()
    plt.show()
    record.write('%s,%s,%s,%s,%s\n'%(runfile,fit_k,fit_r2,upfit_k,downfit_k))

    record.close()
```

```
133 if bigs:
134     MBARS.FindExcluded(runfile,panels,maxdiam)
135     for i in range(panels+1):
136         bigs = MBARS.FindBigs(runfile,i, maxdiam)
137         plt.show()
138 
139 if imageanalysis:
140     while True:
141         query = "Which image do you want to analyze?\n"
142         num = raw_input(query)
143         try:
144             MBARS.ExamineImage(runfile,num, showblanks)
145         except:
146             print "Failed to examine, image does not exist"
        query = 'Look at another? y/n\n'
        answer = raw_input(query)
        if answer != 'y':
            break
```

3

## C.3. MBARS

```
import sys
sys.path.append('C:\\Python27\\Lib\\site-packages')
import numpy as np
import scipy as sp
import matplotlib as mpl
import matplotlib.pyplot as plt
import scipy.misc as spm
from scipy import odr
import winsound
from scipy.optimize import curve_fit
import time
import cPickle as pickle
import os
import shutil
import matplotlib.patches as patches
#import pysal
import numpy.ma as npma
import skimage.transform as sktrans
from math import *
import skimage.feature as skfeat
import skimage.morphology as skmorph
import skimage.filters.rank as skrank
import skimage.restoration as skrestore
import skimage.util as skutil
import sklearn.cluster as skcluster
from scipy.ndimage import filters
import scipy.stats as sps
import scipy.signal as spsig
import imageio
import threading

#This is the MBARS library, it contains all the functions needed to run MBARS

#Global Variables, adjust as needed:
#REFPATH is where important reference files are stored, the key one is
# the HiRISE info file (RDRCUMINDEX.TAB) needs to be in the REFPATH folder
REFPATH = 'C://Users//dhood7//Desktop//MBARS//RefData//'
#BASEPATH is where MBARS will look for the provided filename
BASEPATH = 'C://Users//dhood7//Desktop//MBARS//Images//'
PATH = None
FNM = None

#enter product ID
ID = None
NOMAP = None

WIDFACT = 2.
LENFACT = 2.
#assigned in start as i, s, and r
INANGLE = None
NAZ = None
SAZ = None
SUNANGLE = None
RESOLUTION = None
ROTANG = None
#MAX EXPECTED BOULDER SIZE, expressed in pixels
#max diameter (MD) and max height (MH), done after measurements
MD = 30
MH = 30
#maximum shadow area, taken before measrements, based soleley on the shadow area
MA = 300
#minimum accepted shadow size expressed in pixels
minA = 4
#minimum distance between maxima in the  watershed splitting
```

```
mindist = 3


'''this is called at the end to initialize the program'''
def start():
    i,s,r,n, saz,rotang = getangles(ID)
    current()
    return i,s,r,n,saz, rotang

#Inputs: image = Image to be analyzed (a 2 or 3-d np.array), gam = value for Gamma,
# png = boolean dictating if the image is a 3-D array (True if 3D), plot = True to plot things,
#bound = where to cutoff the shadows, expressed as a fraction of gaussian max (.01 will select t
#outputs: imagelin and imagemodlin, 1-D arrays of the image data before and after gamma modifica
# imageseg= np array with shadows marked as 0.


def gamfun(num,gam=.6, plot=False, manualbound = 130,quickrun = False):
    #print(num)
    #this section causes some errors to pop up between threads, two can try and make it at the s
    if quickrun:
        runfile = 'temp//'
    elif manualbound:
        runfile = 'gam%s_manbound%s//'%(int(gam*1000),int(manualbound))
    else:
        runfile = 'gam%s_bound%s//'%(int(gam*1000),int(boundfrac*1000))
    if not os.path.exists('%s%s'%(PATH,runfile)):
        try:
            os.makedirs('%s%s'%(PATH,runfile))
        except(WindowsError):
            #this is in case two threads try and make something at the same time
            #is it dirty? yes, does it work? also yes.
            pass
    try:
       image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
    except(ValueError, SyntaxError):
        return None, False, runfile
    if isinstance(image[0][0], list):
         #Black and White PNGs store three (identical) values per pixel, this flattens pngs into
        imagetemp = np.zeros((len(image),len(image[0])),dtype=np.uint16)
        for i in range(len(imagetemp)):
            for j in range(len(imagetemp[0])):
                imagetemp[i][j]=image[i][j][0]
        image=imagetemp
    #sharpening was attempted, it destroyed any boulder shapes, left for now
    #image = skrank.enhance_contrast(image, skmorph.disk(3))
##    if NOMAP:
##        rotang = 90+SAZ
##            else:
##        #rotang = NAZ-SAZ
##        rotang = -sunangle

    image = sktrans.rotate(image,ROTANG, resize=True, preserve_range=True)
    image = npma.masked_equal(image, 0)

    ''' rotation seems to cause some stray data to appear at the edge, this is often
        categorized as shadows because it is very dark but not zero, this code will
        essentially erode the edges a bit, shifting the mask up, down, left and right
        and taking the logical OR of the two masks, masking things that are adjacent to
        the mask in any direction
        '''
    shift = np.roll(image.mask,1,1)
    image.mask = np.logical_or(image.mask, shift)
    shift = np.roll(image.mask,-1,1)
    image.mask = np.logical_or(image.mask, shift)
```

```
    shift = np.roll(image.mask,1,0)
    image.mask = np.logical_or(image.mask, shift)
    shift = np.roll(image.mask,-1,0)
    image.mask = np.logical_or(image.mask, shift)

    #image.dump('%s%s%s%s_rot_masked.npy'%(PATH,runfile,FNM,num))
    #rotation incurs some kind of offset to the pixel value, meaning much of the area isnt '0' a
    #recasting as int may solve this problem

    pixels = len(image)*len(image[0])

    #if the image is just blank, side effect of projected images
    if not np.any(image.compressed()):
        return None, False, runfile

    #do the actual gamma correction,

    top = 0.
    image = image**float(gam)
    top = np.max(image)
    scale = 255./top
    image = image*scale
    image = image.astype(int)

    #making sure the boundary is a float
    bound = float(manualbound)

    #print ("Shadow value boundary at:%s"%(bound))
    #doing the indexing in the fastest way always passes the mask, so we are unmasking
    #filling in the values with an impossible number, then putting it all back in
    imageseg = npma.copy(image)
    imageseg = imageseg.astype(float)
    imageseg = imageseg.filled(-1)

    imageseg[imageseg>bound+1] = bound+1
    imageseg = npma.masked_equal(imageseg, -1)
    imageseg = imageseg.astype(int)
    imageseg.fill_value = 0

    imageseg.dump("%s%s%s%s_SEG.npy"%(PATH,runfile,FNM,num))
    #this is the full figure suite, which struggles when you have a big image
    if plot:
        fig,ax = plt.subplots(1,2,sharex = True, sharey = True)
        ax[0].imshow(image, cmap='binary_r', interpolation='none')
        ax[1].imshow(imageseg,vmax = bound, vmin = bound-1, interpolation='none')
##        plt.figure(1)
##        plt.title('Gamma stretched image')
##        plt.imshow(imagemod, cmap='binary_r')
##        plt.figure(2)
##        plt.title('shadows isolated image')
##        plt.imshow(imageseg)
        #plt.figure (3)
        #plt.imshow(sktrans.rotate(image,-SUNANGLE,resize=True, preserve_range=True), cmap='bina

        plt.show()
    #This makes sure that images with data but no shadows do not go through the segmentation pro
    if np.min(imageseg)>= bound:
        good = False
    else:
        good=True
    return(imageseg, good ,runfile)


#this method uses a gaussian fit to the data to find the shadow boundary
#it takes in the linearized image that you want to use and returns the shadow boundary
```

```
def gauss(x,sig,mu):
    y = (1./(sig*np.sqrt(2.*np.pi)))*np.exp((-1./2.)*(((x-mu)/sig)**2.))
    return y
def gauss_unnorm(x,sig,mu):
    y = np.exp((-1/2)*((x-mu)/sig)**2)
    return y

def autobound(num,bound):
    ''' An automatic boundary-finder for HiRISE images, relies on input statistics
'''

    runfile = 'autobound//'
    if not os.path.exists('%s%s'%(PATH,runfile)):
        try:
            os.makedirs('%s%s'%(PATH,runfile))
        except(WindowsError):
            #this is in case two threads try and make something at the same time
            #is it dirty? yes, does it work? also yes.
            pass
    try:
       image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
    except(ValueError, SyntaxError):
        return None, False, runfile

    if not np.any(image.flatten()):
        return None, False, runfile
    image = sktrans.rotate(image,ROTANG, resize=True, preserve_range=True)
    image = npma.masked_equal(image, 0)

    ''' rotation seems to cause some stray data to appear at the edge, this is often
        categorized as shadows because it is very dark but not zero, this code will
        essentially erode the edges a bit, shifting the mask up, down, left and right
        and taking the logical OR of the two masks, masking things that are adjacent to
        the mask in any direction
        '''
    shift = np.roll(image.mask,1,1)
    image.mask = np.logical_or(image.mask, shift)
    shift = np.roll(image.mask,-1,1)
    image.mask = np.logical_or(image.mask, shift)
    shift = np.roll(image.mask,1,0)
    image.mask = np.logical_or(image.mask, shift)
    shift = np.roll(image.mask,-1,0)
    image.mask = np.logical_or(image.mask, shift)

    imageseg = npma.copy(image)
    imageseg = imageseg.astype(float)
    imageseg = imageseg.filled(-1)

    imageseg[imageseg>bound+1] = bound+1
    imageseg = npma.masked_equal(imageseg, -1)
    imageseg = imageseg.astype(int)
    imageseg.fill_value = 0

    imageseg.dump("%s%s%s%s_SEG.npy"%(PATH,runfile,FNM,num))

    #guard against images with no shadows
    if np.min(imageseg)>= bound:
        good = False
    else:
        good=True

    return(imageseg, good, runfile)

def getimagebound(panels):
    '''TO retrieve the overall image stats and calculate the absolute shadow boundary
```

```
265
266     assumes a 2-pixel wide (4 total pixels) shadow to the minimum shadow size'''
267
268     bins = np.linspace(0,1023,1024)
269     bins = map(int,bins)
270     hist = None
271     cum_hist = None
272     #panels = 400
273     for i in range(0,panels):
274         #print i
275         im = imageio.imread('%s%s%s.PNG'%(PATH,FNM,i))
276         n,bins = np.histogram(im,bins)
277         #cum_n,bins,c=plt.hist(im.flatten(),bins,cumulative=True)
278         n[0] = 0
279         if i==0:
280             hist = n
281             #cum_hist = cum_n
282         else:
283             hist+=n
284         #print hist
285             #cum_hist+=cum_n
286     #we have the histogram for the whole image now.
287     #kill the zeros
288     #hist[0] = 0
289     mode = np.argmax(hist)
290     #print mode
291     #make and normalize the cumulative histogram
292     cum_hist = np.cumsum(hist)
293     ncum_hist = cum_hist/float(max(cum_hist))
294     #this is now a map to the actual image distribution
295     runs = 100
296     #this is important, what boundary should be chosen? I am going with the
297     #average of the 100th percentile
298     stats = []
299     for i in range(runs):
300         img = ImageMaker(ncum_hist,bins)
301         #plt.imshow(img)
302         #plt.show()
303         #.77 for the lorentzian taken from
304         #Kirk et al 2008, 10.1029/2007JE003000
305         img_con = convolve_lorentzPSF(img,mode,.77)
306         #plt.imshow(img_con)
307         #plt.show()
308         #shadow size comes in HERE, must be same as below
309         shad_con = img_con[25:30,25:30]
310         stat = np.percentile(shad_con,100)
311         stats+=[stat]
312
313
314     bound = np.average(stats)
315     print ("Selected boundary at %s"%(bound))
316
317     #return bound,ncum_hist,hist
318     return bound
319
320 def ImageMaker(mapping_hist,mapping_bins,dimx=50,dimy=50):
321     ''' returns a value from the HiRISE image population, meant to replicate what I see in MBARS
322
323     img = np.random.rand(dimy,dimx)
324     for i in range(len(img)):
325         for j in range(len(img[0])):
326             for k in range(len(mapping_hist)):
327                 if mapping_hist[k]>=img[i][j]:
328                     img[i][j] = float(mapping_bins[k])
329
330
```

```
    #shadow size comes in here, must be same as above
    img[25:30,25:30] = 1
    return img

def convolve_lorentzPSF(image,avg,gam=.77):
    '''convolve an array with a lorentzian HiRISE PSF'''
    kern = lorentz_kern(31,gam)
    #print avg
    newimage=spsig.convolve2d(image,kern,mode='same',boundary='fill',fillvalue=avg)
    return newimage

def lorentz(x,xo,gam):
    #this is a lorentzian, a more accurate PSF for HiRISE
    l = (1/(np.pi*gam))*((gam**2)/(((x-xo)**2)+gam**2))
    return l

def lorentz_kern(dim=21,gam=.77):
    #method to make a square lorentzian kernal for deconvolution
    #VERY IMPORTANT, MUST BE ODD numbered
    if dim%2==0:
        print("Must be odd, adding 1 to dim")
        dim+=1
    kern = np.empty((dim,dim))
    mid = dim/2
    for i in range(len(kern)):
        for j in range(len(kern[0])):
            y= float(i-mid)
            x = float(j-mid)
            r = np.sqrt(y**2+x**2)
            kern[i][j] = lorentz(r,0,gam)
    total = sum(kern.flatten())
    kern = kern/total
    #plt.imshow(kern)
    #plt.show()
    return(kern)

#########This is the measuring side of the code#####################################

def boulderdetect(num,image,runfile):
    #deprecated, uses threadsafe version now
    #flag must be dtype long, otherwise it will wrap at high numbers and reset the flag to 1
    flag = long(1)
    coords = [0,0]
    save = open("%s%s%s%s_shadows.shad"%(PATH,runfile, FNM,num), 'wb')
    im_area = len(image)*(len(image[0]))
    #flagged image has the boulders marked
    #print 'Running Watershed %s\n'%(num)
    fimage = watershedmethod(image)
    #have to explicitly pass the mask on from the input image
    fimage = npma.masked_array(fimage)
    fimage.dump('%s%s%s%s_flagged.npy'%(PATH, runfile,FNM,num))
    fimage.mask = image.mask

    #close the seg image to open up memory
    image = None
    fmax = np.max(fimage)
    #shadows = []
    #print 'finding shadows %s \n'%(num)
    shade=None
    for i in range(fmax+1):
        #print 'flag %s image %s\n'%(i,num)
        pixels = np.argwhere(fimage == i)
        #must be converted to list for the neighbor-checking function to work
        pixels = pixels.tolist()
```

```
397         if len(pixels) >=minA and len(pixels)<MA:
398
399             shade = shadow(i, pixels, im_area)
400
401             shade.run()
402
403             pickle.dump(shade,save)
404     save.close()
405     return
406
407 def boulderdetect_threadsafe(num,image,runfile,odr_keycard):
408     #flag must be dtype long, otherwise it will wrap at high numbers and reset the flag to 1
409     flag = long(1)
410     coords = [0,0]
411     save = open("%s%s%s%s_shadows.shad"%(PATH,runfile, FNM,num), 'wb')
412     im_area = len(image)*(len(image[0]))
413     #flagged image has the boulders marked
414     #print 'Running Watershed %s\n'%(num)
415     fimage = watershedmethod(image)
416     #have to explicitly pass the mask on from the input image
417     fimage = npma.masked_array(fimage)
418     fimage.dump('%s%s%s%s_flagged.npy'%(PATH, runfile,FNM,num))
419     fimage.mask = image.mask
420     fmax = np.max(fimage)
421     #clear the seg image to save memory
422     image = None
423     shade=None
424     for i in range(fmax+1):
425
426         pixels = np.argwhere(fimage == i)
427         #must be converted to list for the neighbor-checking function to work
428         pixels = pixels.tolist()
429
430         if len(pixels)<MA and len(pixels)>0:
431
432             shade = shadow(i, pixels, im_area)
433
434             #broken into 3 steps to narrow the thread-unsafe part into one function
435             shade.run_prep()
436             with odr_keycard:
437                 shade.run_fit()
438             shade.run_post()
439
440             pickle.dump(shade,save)
441     save.close()
442
443     return
444
445 def watershedmethod(image):
446     #this is the new way of finding the shadows in an image
447     #first find the "plateau" value, we will need this for masking
448     plat = sps.mode(image.compressed())
449     #fill the image with a known value in order to preserve the mask
450     temp = image.filled(np.max(image)+1)
451     #invert the image so the shadows are peaks not lows
452     temp = temp*(-1)+np.max(image)+1
453     #find the peaks in the image, return the points as a nx2 array
454     #min_distance is a super important argument,changes the minimum distance allowed betwen maxi
455     points = skfeat.peak_local_max(temp,min_distance=mindist, indices=True)
456
457     #put in a guard against images with no shadows
458     threshold = len(image.compressed())/2
459     if len(points)>threshold:
460         return np.ones_like(image)
461     #prepare to convert the points matrix to an image-like array
462
```

```
    #this could perhaps be done with DBSCAN

    cores,labels = skcluster.dbscan(points,2,2)

    view = np.zeros_like(image)
    flag = 2
    excl = []
    for i in range(len(points)):
        if labels[i] == -1:
            view[points[i][0]][points[i][1]] = flag
            flag+=1
        elif labels[i] not in excl:
            view[points[i][0]][points[i][1]] = flag
            excl+=[labels[i]]
            flag+=1

    #this will make the mask on view mask out the originally masked points (outside data)
    # as well as the plateau pixels
    temp = npma.masked_equal(image, plat[0][0])
    view.mask = np.logical_or(temp.mask, view.mask)

    '''details on the arguments being handed to watershed:
        in general ,the algorithm doesnt like masked arrays, hence everything is filled
        the masks are added back in the makeshadows code above
        the mask at the ennd prevents it from trying to segment all the no-data areas
        however it takes the mask in the opposite sense of the masked arrays
        in ndarrays, the mask is True where you want it masked, watershed wants 0
    '''

    boulds = skmorph.watershed(image.filled(np.max(image)+1), view.filled(0),mask=~view.mask)

    return boulds
def overlapcheck_threadsafe_DBSCAN(num,runfile,odr_keycard,overlap=.1):
    '''
    Code to get rid of double-counts in returned boulders, inputs:
    num - num of target image
    runfile - current runfile for shad file finding purposes
    overlap - allowed overlap between boulders, expressed in fraction of boulder area
    '''
    shadow_file = getshads(runfile, num,mode='r')
    if shadow_file == None:
        return
    parameters = []
    while True:
        try:
            data = pickle.load(shadow_file)
        except(EOFError):
            break
        #pull in the flag, the y coordiante, the x coordinate, and the width
        parameters+=[[data.flag,data.bouldcent[0],data.bouldcent[1],data.bouldwid, data.fitgood]
    #get rid of boudlers with unreliable measurements
    parameters = [a for a in parameters if a[4]]
    if len(parameters) == 0:
        shadow_file.close()
        return

    #compute modified distance matrix
    #make meshgrids of every number in the calculation
    #includes some precomputing to minimize the memory footprint
    ya,yb = np.meshgrid([a[1] for a in parameters],[a[1] for a in parameters])
    ydist = (ya-yb)**2
    ya = None
    yb = None
    xa,xb = np.meshgrid([a[2] for a in parameters],[a[2] for a in parameters])
```

```
    xdist = (xa-xb)**2
    xa = None
    xb = None
    tdist = np.sqrt(ydist+xdist)
    ydist = None
    xdist = None
    da,db = np.meshgrid([a[3] for a in parameters],[a[3] for a in parameters])
    dsum = da+db
    da = None
    db = None
    #calculate the distance ratio
    #D = euclidean distance between points/sum of radii, this way boulders that touch have a D o
    D = 2*tdist/(dsum)
    dsum = None
    tdist = None
    #this is now the Distance matrix to pass to DBSCAN
    #arguments that go to DBSCAN: D, distance matrix, 1=distance to be considered neighbors, 2 m
    #DBSCAN is a little generous to help catch more things
    cores,labels = skcluster.dbscan(D,1.2,2,metric='precomputed')
    #cores is a list of 'core' points, less important than labels, which groups the points into
    #add these to the parameters list
    for i in range(len(parameters)):
        parameters[i]+=[labels[i]]
    #make our 'web' of flags, a list where every sublist is flags that are interconnected
    webs = []
    for i in range(max(labels)+1):
        subweb=[]
        for j in parameters:
            if j[5] == i:
                subweb+=[j[0]]
        webs+=[subweb]

    #now lets re-read in the shadow data
    shadow_file.seek(0)
    og_data = []
    while True:
        try:
            og_data+=[pickle.load(shadow_file)]
        except(EOFError):
            break
    #close and re-open for re-writing
    shadow_file.close()
    shadow_file = getshads(runfile,num,mode='w')


    problem_flags = [a[0] for a in parameters if a[5]!=-1]
    problem_boulders = []
    for i in og_data:
        if i.flag in problem_flags:
            problem_boulders+=[i]
        else:
            pickle.dump(i,shadow_file)
    og_data = None
    #print(webs)
    #OK with just the problem_boulders list, we can do the work
    for subweb in webs:
        #print subweb
        current_web = []
        #take the web list (which is just flags) and make a list of shadow objects
        for bould in problem_boulders:
            if bould.flag in subweb:
                current_web+=[bould]


        #merge = True
```

```
595         while True:
596             #print 'Current_web length: %s'%(len(current_web))
597             #print(subweb)
598             #reset to no merge
599             merge = False
600             #current_web = newboulds
601             newboulds = []
602             for a in current_web:
603                 #print'looking for merges'
604                 if merge:
605                     continue
606                 #A_flag = j.flag
607                 for b in current_web:
608                     if merge:
609                         continue
610                     #B_flag = k.flag
611                     if a.flag != b.flag:
612                         #look for overlapping pairs in the web
613 
614                         area = checkpos(a,b)
615                         #print(area)
616 
617                         if area > overlap:
618                             #print 'Merging Boulders'
619                             #if they overlap too much, make a new shadow that combines their pix
620                             newbould=shadow(a.flag, a.pixels+b.pixels, a.im_area)
621                             newbould.run_prep()
622                             with odr_keycard:
623                                 newbould.run_fit()
624                             newbould.run_post()
625 
626                             #compare the fiterr
627                             #the ones too small to fir are assigned stupid high fiterr values
628                             if a.fiterr:
629                                 fita = a.fiterr
630                             else:
631                                 fita = 100
632                             if b.fiterr:
633                                 fitb = b.fiterr
634                             else:
635                                 fitb=100
636                             fitavg = (fita+fitb)/2.
637                             #trying to not filter until the end to avoid killing intermediate bo
638                             #boulders that are smaller than the min and are improvements on the
639                             if newbould.bouldwid<MD and newbould.fiterr<fitavg:
640                             #if True
641                                 #you found two that needed to merge and it made a decent boulder
642                                 A_flag = a.flag
643                                 B_flag = b.flag
644                                 #print 'Merged %s and %s'%(A_flag,B_flag)
645                                 merge = True
646                                 #if the new boulder does not look good, Merge is not set to True
647 
648 
649             if merge:
650                 #since you found two that merged, dont add them to the next web and put the new
651                 #print 'I found a merge!'
652                 for bould in current_web:
653                     if bould.flag != A_flag and bould.flag != B_flag:
654                         newboulds+=[bould]
655                         #print bould.flag
656                 newboulds+=[newbould]
657                 current_web = newboulds
658             else:
659 
660 
```

```
                break


            #if no merges are found in the inner loops, the code will continue to the next web
        #now save all these new ones you just made
        for boulder in current_web:
            #print'Dumping shadows'
            #print 'Saving Boulders'
            #print(boulder.flag)
            #if boulder.fitgood:
            pickle.dump(boulder,shadow_file)
    shadow_file.close()
    return

def overlapcheck_shadbased(num,runfile,odr_keycard):
    '''
    Another take on the overlap merging function. This one is NOT based on the interpreted bould
    if shadows touch each other, the function considers whether combining it with other shadows
    '''
    shadow_file = getshads(runfile, num,mode='r')
    if shadow_file == None:
        return
    parameters = []
    while True:
        try:
            data = pickle.load(shadow_file)
        except(EOFError):
            break
        #pull in the flag, y center of the shadow, the x center of the shadow, the border
        parameters+=[[data.flag,data.center[0],data.center[1],data.border,data.pixels]]

    if len(parameters) == 0:
        shadow_file.close()
        return

    #each element in parameter is a boulder, first things first is to get the "adjacency" list
    #each item in "adjacency" has three items, flag a, flag b, adjacency score (number of adjace
    adjacency = []
    for i in range(len(parameters)):
        aflag = parameters[i][0]
        ay = parameters[i][1]
        ax = parameters[i][2]
        abord = parameters[i][3]
        apix = parameters[i][4]
        if len(abord)==0:
            abord = apix
        #dont double-count pairs, this should speed things up
        for j in range(i+1,len(parameters)):
            #quicker check if they are super far apart
            bflag = parameters[j][0]
            by = parameters[j][1]
            bx = parameters[j][2]
            bbord = parameters[j][3]
            bpix = parameters[j][4]
            if len(bbord) ==0:
                bbord=bpix

            dist = np.sqrt((ay-by)**2+(ax-bx)**2)
            if dist > sum([len(apix),len(bpix)]):
                adjacency+=[[aflag,bflag,0]]
                continue
            #they are not super far away, so we can check adjacency
            D = 0
            for k in abord:
```

```
                for l in bbord:
                    dist = np.sqrt((k[0]-l[0])**2+(k[1]-l[1])**2)
                    #print(k,l,dist)
                    if dist<=1:
                        D+=1
            adjacency+=[[aflag,bflag,D]]
    #print [a for a in adjacency if a[2]>0]
    #OK, we have the adjacency, now we need to determine the clusters.
    #these are organized by the A flag, so lets make some webs
    clusters = []
    for adj in adjacency:
        #the two flagged shadows dont touch
        if adj[2] == 0:
            continue
        #they do touch
        else:
            #see if webs exist that contain these flags
            if len(clusters) == 0:
                clusters+=[[adj[0],adj[1]]]
                continue
            #hits will mark the clusters in which the flags are found
            hits = []
            for i in range(len(clusters)):
                if adj[0] in clusters[i] or adj[1] in clusters[i]:
                    hits+=[i]
            if len(hits) == 0:
                clusters+=[[adj[0],adj[1]]]
            else:
                newclust = []
                hits.sort(reverse=True)
                for i in hits:
                    newclust+= clusters.pop(i)
                #need to amke sure new flags get added, repeats will be there, but it doesnt act
                newclust+=[adj[0],adj[1]]
                clusters+=[newclust]

    #clusters is now a list of clusters, adjacency is still available to reference the adjacency
    #return to the start of the shadows and start the merging process
    #now lets re-read in the shadow data
    shadow_file.seek(0)
    og_data = []
    while True:
        try:
            og_data+=[pickle.load(shadow_file)]
        except(EOFError):
            break
    #close and re-open for re-writing
    shadow_file.close()
    shadow_file = getshads(runfile,num,mode='w')

    prob_flags = [a for clust in clusters for a in clust]
    problem_boulders = []
    for i in og_data:
        if i.flag in prob_flags:
            problem_boulders+=[i]
        else:
            #pass
            pickle.dump(i,shadow_file)
    og_data = None

    #for testing purposes
    report = []
    for clust in clusters:
        #lets collect the shadow objects for the relevant clusters
```

```
793
794         boulds = []
795         avg_fiterr = 0.
796         all_pixels = []
797         for rock in problem_boulders:
798             if rock.flag in clust:
799                 boulds+=[rock]
800                 #if the rock is one of the big ones, anything is an improvement
801                 if rock.bouldwid > MD:
802                     avg_fiterr+=1000
803
804                 elif rock.fiterr:
805                     avg_fiterr+=rock.fiterr
806                 else:
807                     avg_fiterr+=1000
808                 all_pixels+=rock.pixels
809         base_flag = boulds[0].flag
810         #identify areas that are way too big, likely shadow-casting topography
811         if len(all_pixels) > 5000:
812             #print base_flag
813             #print (len(all_pixels))
814             continue
815
816         avg_fiterr = avg_fiterr
817         #now, merge the pixels and see if it is better.
818         newbould=shadow(base_flag, all_pixels, boulds[0].im_area)
819         newbould.run_prep()
820         with odr_keycard:
821             newbould.run_fit()
822         newbould.run_post()
823         #consider which is better?
824         #new one exists, is better, and is smaller than max
825         #print 'considering cluster %s'%(clust)
826         #print 'New fiterr = %s, avg_fiterr = %s'%(newbould.fiterr,avg_fiterr)
827         if newbould.fiterr and newbould.fiterr < avg_fiterr and newbould.bouldwid < MD:
828             pickle.dump(newbould,shadow_file)
829             #report+=['New boulder was better']
830
831         #this is the method where we filter based on connectivity scores,
832 ##        else:
833 ##            #report+=['old boulders were better']
834 ##            #OK, lets attempt to calculate "connectivity", the sum of adjacency scores:
835 ##            connectivity = []
836 ##            for rock in boulds:
837 ##                connect = sum([a[2] for a in adjacency if a[0] == rock.flag or a[1] == rock.fl
838 ##                connectivity+=[connect]
839 ##            #zip connectivity into the boulds list
840 ##            boulds = [list(a) for a in zip(boulds,connectivity)]
841 ##            #print'Entering Recursive Portion'
842 ##            #print (zip([a[0].flag for a in boulds],connectivity))
843 ##            exclusive_shadowmerge(boulds,2,shadow_file,odr_keycard)
844         #Lets try a k-means based method
845         else:
846             #print newbould.bouldwid
847             kmeans_shadowmerge(boulds,shadow_file,odr_keycard,avg_fiterr)
848
849
850     return
851     #return clusters,adjacency,report
852
853
854 def exclusive_shadowmerge(boulds,mincon,shadowfile,odr_keycard):
855     ''' Calls the shadowmerge method, recursively tries to make new boulders from adjacent shado
856     '''
857
858
```

```
    all_pixels = []
    avg_fiterr = 0
    inboulds = [rock for rock in boulds if rock[1]>=mincon]
    outboulds = [rock for rock in boulds if rock[1]<mincon]
    if len(inboulds)>0:
        for rock in inboulds:
            if rock[0].bouldwid > MD:
                avg_fiterr+=100
            elif rock[0].fiterr:
                avg_fiterr+=rock[0].fiterr
            else:
                avg_fiterr+=100
            all_pixels+=rock[0].pixels
        base_flag = inboulds[0][0].flag
        avg_fiterr = avg_fiterr/(float(len(inboulds)))
        #now, merge the pixels and see if it is better.
        newbould=shadow(base_flag, all_pixels, inboulds[0][0].im_area)
        newbould.run_prep()
        with odr_keycard:
            newbould.run_fit()
        newbould.run_post()
        if newbould.fiterr and newbould.fiterr < avg_fiterr and newbould.bouldwid < MD:
            pickle.dump(newbould,shadowfile)
            if len(outboulds)>0:
                exclusive_shadowmerge(outboulds,1,shadowfile,odr_keycard)
        else:
            exclusive_shadowmerge(boulds,mincon+1,shadowfile,odr_keycard)
    else:
        for rock in boulds:
            pickle.dump(rock[0],shadowfile)

def kmeans_shadowmerge(boulds,shadowfile,odr_keycard,avg_fiterr):
    #boudlds is a list of shadow objects, shadowfile is the targeted shadow file, should be in "
    #odr-keycard is the thread lock object to prevent multiple access to ODR, maxboulds is the h
    #avg_fiterr is the average fit error on the original boulders, we have to be better
    all_pixels = [b for a in boulds for b in a.pixels]
    #this is a list of all flags, maxbolds is limited by the length of this list
    all_flags = [a.flag for a in boulds]
    maxboulds = len(all_flags)
    newerrs_list = []
    newboulds_list = []
    #print all_flags
    for i in range(2,maxboulds+1):
        #lets try manually seeding to limit splits along the sun-line
        #by seeding the k-means as an equal x-spread, this should strongly favor lateral boulder
        minx = min([a[1] for a in all_pixels])
        maxx = max([a[1] for a in all_pixels])
        avgy = np.average([a[0] for a in all_pixels])
        #make a linspace that goes from min to max, then remove the ends.
        seeds  = list(np.linspace(minx,maxx,i+2))
        seeds.pop(0)
        seeds.pop(-1)
        seeds = map(lambda x: (x,avgy),seeds)
        seeds = [list(a) for a in seeds]
        seeds = np.array(seeds)
        #kmeans = skcluster.KMeans(n_clusters=i,init = seeds).fit(all_pixels)
        kmeans = skcluster.KMeans(n_clusters=i).fit(all_pixels)
        labelpix = [list(a) for a in zip(kmeans.labels_,all_pixels)]
        #cents = kmeans.cluster_centers_
        #print(kmeans.labels_)
        #for display purposes
        #print(kmeans.cluster_centers_)
##        cols = [a[0] for a in labelpix]
##        ys = [a[1][0] for a in labelpix]
```

```
925 ##        xs = [a[1][1] for a in labelpix]
926 ##        plt.scatter(xs,ys,c=cols,cmap='rainbow')
927 ##        plt.show()
928
929         #check if this is a good fit!
930         newboulds = []
931         new_fiterr = 0
932         for j in range(i):
933             pix = []
934             for a in labelpix:
935                 if a[0] == j:
936                     pix+=[a[1]]
937             newbould = shadow(all_flags[j],pix,boulds[0].im_area)
938             newbould.run_prep()
939             with odr_keycard:
940                 newbould.run_fit()
941             newbould.run_post()
942             newboulds+=[newbould]
943             if newbould.bouldwid > MD:
944                 new_fiterr+=1000
945             elif newbould.fiterr:
946                 new_fiterr+=newbould.fiterr
947             else:
948                 new_fiterr+=1000
949         new_fiterr = new_fiterr
950         #add the results to these list
951         newerrs_list+=[new_fiterr]
952         newboulds_list+=[newboulds]
953     #pick the best solution:
954     #print avg_fiterr
955     #print min(newerrs_list)
956     if min(newerrs_list)<= avg_fiterr:
957         #we found a better solution
958         #print "%s is best fit"%(np.argmin(newerrs_list)+2)
959         final_boulders = newboulds_list[np.argmin(newerrs_list)]
960         for i in final_boulders:
961             pickle.dump(i,shadowfile)
962     else:
963         #no better solution was found
964         #print "no better fit was found"
965         for i in boulds:
966             pickle.dump(i,shadowfile)
967     return()
968
969
970
971
972 def touch(array,pos,wid,indflag,indpos,indwid,ind, plus=False,minus=False):
973     '''Corrolary to the above overlap check function '''
974     #no longer in use
975
976     rad = wid/2.
977     neighborflags = []
978     #check the neighbors
979     try:
980         if (array[ind+1][indpos]-(array[ind+1][indwid])/2.) < (pos+rad) and plus:
981             neighborflags+=[array[ind+1][indflag]]
982             try:
983                 neighborflags+=touch(array,pos,wid,indflag,indpos,indwid,ind+1,True,False)
984             except(RuntimeError):
985                 pass
986     except(IndexError):
987         pass
988     try:
989
990
```

```
        if (array[ind-1][indpos]+(array[ind-1][indwid]/2.)) > (pos-rad) and minus:
            neighborflags+=[array[ind-1][indflag]]
            try:
                neighborflags+=touch(array,pos,wid,indflag,indpos,indwid,ind-1,False,True)
            except(RuntimeError):
                pass
    except(IndexError):
        pass
    return neighborflags

def webfinder(array,web):
    '''Corrolary to the overlap check function '''
    #no longer in use
    oldweb = web
    for i in web:
        for j in array:
            if i in j:
                for k in j:
                    if k not in web:
                        web+=[k]
    if oldweb != web:
        web = webfinder(array,web)
    else:
        return web

def checkpos(shadow1,shadow2):
    '''expects two shadow objects , checks how much they overlap
    returns a fraction of the area of the smaller boulder overlapped by the larger
    '''
    bouldwid1 = shadow1.bouldwid
    bouldwid2 = shadow2.bouldwid

    pos1 = shadow1.bouldcent
    pos2 = shadow2.bouldcent
    #if one is unmeasured or has no dimensions, forget it
    if pos1[0] == None or pos2[0] == None:
        return 0.
    #distance between the two circles
    d = np.sqrt((pos1[0]-pos2[0])**2+(pos1[1]-pos2[1])**2)
    r1 = bouldwid1/2.
    r2 = bouldwid2/2.
    if r1>=r2:
        ra = r1
        rb = r2
    else:
        ra = r2
        rb = r1
    #lets throw out some simple scenarios
    #the circles exactly reach or do not overlap
    if d>= ra+rb:
        return 0.
    #if the smaller is fully in the larger:
    if d+rb <=ra:
        return 1.
    #if the center of one is within the other, but not entirely
    if ra>d and d+rb>ra:
        #returning 1 beccause I always want these to merge
        return 1.
    else:
        #note: all this work is done in pixels not meters
        x = (ra**2 - rb**2 + d**2)/(2*d)
        y = np.sqrt(ra**2-x**2)
        #note: this returns in radians
        thetaA = abs(np.arccos(x/ra))
```

```
        thetaB = abs(np.arccos((d-x)/rb))
        #relevant EQ, area of circle slice: piR^2*(2theta/2pi), triangle .5*b*h
        area = (ra**2)*thetaA - x*y + (rb**2)*thetaB - (d-x)*y
        #print('ra = %s, rb = %s, d = %s'%(ra,rb,d))
        area = area/(np.pi*rb**2)
        return area


class shadow(object):

    def __init__(self, flag, pixels, im_area):
        #marker for this boulder, unique
        self.flag = flag
        #immediate pixels in this shadow
        self.pixels = pixels
        #pixels in the mirrored shadow
        self.mpixels = None
        #area of the shadow, based on how many pixels are in it
        self.area = len(pixels)
        #area for the mirrored shadow, simply 2*area for now, may change
        self.marea = 2*self.area
        #incedence angle of the sun, fixed by target image
        self.inangle = INANGLE
        #angle of the sun, fixed as coming from top (-y), this is holdover till
        #code sentive to angle is entirely gone
        self.beta = np.radians(270)
        #image resolution, tied to each shadow image for safekeeping
        self.resolution = RESOLUTION
        #image area (pixels^2) tied to each shadow for safekeeping
        self.im_area = im_area
        #find the center of the shadow
        tempx = 0.
        tempy = 0.
        for i in range(len(pixels)):
            tempy +=pixels[i][0]
            tempx +=pixels[i][1]
        self.center = [tempy/float(len(pixels)),tempx/float(len(pixels))]
        #border of the shadow
        self.border = []
        #initially, the border of the shadow without the -y(upper) side,
        #finally the rim of the mirrored shadow
        self.mborder = []
        #the flipping axis of the shadow
        self.flipaxis = []
        #initial conditions for the non Area-preserving and AP fits,
        #this gets overwritten in the run_prep section, leaving here for now
        self.fitinit = [self.center[0], 2.0,self.center[1], 2.0, 0.]
        self.AP_fitinit = [self.center[0],2.0,self.center[1], 0.]
        #Empty variable, will be the return from the ODR fit
        self.fitbeta = None
        #Is the fit good assumed False, see ODRFit functions for conditions
        self.fitgood = False
        #Measurement of how well the fit actually fits, used to determine if merges improve the
        self.fiterr = None
        #records stoppping condition of the ODR, <4 is good, >4 is bad, 4 is OK
        self.fitinfo = None


        #has the boulder been assessed?
        self.measured = None

        #boulder morphometry
        #boulder width in pixels and meters
        self.bouldwid = None
        self.bouldwid_m = None
```

```
        #height of boulder in pixels and meters
        self.bouldheight = None
        self.bouldheight_m = None
        self.bouldheight_m_actual = None
        #center of fit ellipse
        self.bouldcent = [None, None]
        #semi-axis along the sun direction in pixels and meters
        self.shadlen = None
        self.shadlen_m = None
        #placeholder for the patch object
        self.ellipsepatch = None

    def run_prep(self):
        #main function that does most things we want it to do

        #self.findborder()
        self.findborder_cents()
        if len(self.border) != 0:
            flipval = self.mirror()
            self.fitinit = [flipval, 2.0,self.center[1], 2.0, 0.]

    def run_fit(self):
        #to change the kind of border fit used, alter this line
        if len(self.mborder)!=0:
            self.odrfit_m()

    def run_post(self):
        #turned off this guard to see if large ones are being tossed
        #if self.fitgood:
        self.shadowmeasure_m()

    def findborder(self):
        #we will look at each point and see if it has neighbors, if not, there is a border
        #this cannot be the most efficient way to do this....

        for i in self.pixels:
            top=False
            other = False
            if [i[0]-1,i[1]] not in self.pixels:
                self.border+= [[i[0]-.5, i[1]]]
                self.border+= [[i[0]-.5, i[1]+.5]]
                self.flipaxis+= [[i[0]-.5, i[1]]]
                self.flipaxis+= [[i[0]-.5, i[1]+.5]]
                top = True

            if [i[0]+1,i[1]] not in self.pixels:
                self.border+= [[i[0]+.5, i[1]]]
                self.border+= [[i[0]+.5, i[1]-.5]]
                self.mborder+= [[i[0]+.5, i[1]]]
                self.mborder+= [[i[0]+.5, i[1]-.5]]


            if [i[0],i[1]+1] not in self.pixels:
                self.border+= [[i[0], i[1]+.5]]
                self.border+= [[i[0]+.5, i[1]+.5]]
                if not top:
                    self.mborder+= [[i[0], i[1]+.5]]
                    self.mborder+= [[i[0]+.5, i[1]+.5]]


            if [i[0],i[1]-1] not in self.pixels:
                self.border+= [[i[0], i[1]-.5]]
                self.border+= [[i[0]-.5, i[1]-.5]]
                if not top:
                    self.mborder+= [[i[0], i[1]-.5]]
```

```
                    self.mborder+= [[i[0]-.5, i[1]-.5]]
    def findborder_cents(self):
        '''
        Alternate method for finding the shadow border, uses pixel cetners rather than pixel edg
        This may generally struggle with shadows that are linear
        must set self.border, self.mborder, self.flipaxis in otder to swap in for findborder
        '''

        #quick check for linear shadows and small shadows, these need to be tossed
        xs = [i[1] for i in self.pixels]
        ys = [i[0] for i in self.pixels]
        if min(xs) == max(xs) or min(ys) == max(ys):
            #this is a linear shadows, it will bomb the ODR fitting
            return
        if len(self.pixels) <=minA:
            return
        for i in self.pixels:
            top=False
            other = False
            bord = False
            mirror = False
            if [i[0]-1,i[1]] not in self.pixels:
                self.border+= [i]
                self.flipaxis+= [[i[0]-.5,i[1]]]
                top = True
                bord = True

            if [i[0]+1,i[1]] not in self.pixels:
                self.mborder+= [i]
                mirror = True
                if not bord:
                    self.border+= [i]

            if [i[0],i[1]+1] not in self.pixels:
                if not bord:
                    self.border+= [i]
                if not top and not mirror:
                    self.mborder+= [i]

            if [i[0],i[1]-1] not in self.pixels:
                if not bord:
                    self.border+= [i]
                if not top and not mirror:
                    self.mborder+= [i]
        return

    def mirror(self):
        ''' this takes the boulder shadow and border and flips it over the
            sun_perpendicular vector, which is the x axis in all cases.
            By  doing this we can fit an ellipse to what is actually an ellipse,
            theoretically giving better fits. the points flip along the MIN (most sunward)
            pixel value,
            '''
        #step 1, find the flip axis
        yvals = np.array(self.flipaxis)
        yvals = yvals[range(len(yvals)),[0]]
        #this is the flip value
        #flipval = np.average(yvals)
        flipval = np.min(yvals)
        self.mborder = [[float(k[0]),float(k[1])] for k in self.mborder]
        #make a copy of the mborder, which lacks the -y boundary
        temp = np.copy(self.mborder)
        dist = range(len(temp))
        #invert that about the flip value
```

```
1255            temp[dist,[0]]*=-1
1256            temp[dist,[0]]+=2*flipval
1257            #mborder is stored as a list, lets keep it that way
1258            temp = temp.tolist()
1259            #append it and you are done
1260            self.mborder+=temp
1261
1262            return flipval
1263
1264        def AP_odrfit_m(self):
1265            #not used, odrfit_m is used
1266            input_y = list(map(lambda f:f[0], self.mborder))
1267            input_x = list(map(lambda f:f[1], self.mborder))
1268            input_dat = [input_y, input_x]
1269            fit_data = odr.Data(input_dat, y=1)
1270            fit_model = odr.Model(self.AP_ellipse, implicit=True)
1271            fit_odr = odr.ODR(fit_data, fit_model, self.AP_fitinit)
1272            fit_out = fit_odr.run()
1273            self.fitinfo = str(fit_out.info)
1274            temp = fit_out.beta
1275            self.fitbeta = [temp[0],temp[1], temp[2], self.marea/(np.pi*temp[1]),temp[3]]
1276
1277            #cutoff for areas of boulders, throws out any fits that are too big
1278            area = self.marea
1279            if area > MA or self.fitbeta[1] > MD or self.fitbeta[3] > MD:
1280                self.fitgood = False
1281            elif self.fitinfo == "2" or self.fitinfo == "3" or self.fitinfo == "1" or self.fitinfo =
1282                self.fitgood = True
1283            return
1284
1285        def odrfit_m(self):
1286            input_y = list(map(lambda f:f[0], self.mborder))
1287            input_x = list(map(lambda f:f[1], self.mborder))
1288            input_dat = [input_y, input_x]
1289            fit_data = odr.Data(input_dat, y=1)
1290            fit_model = odr.Model(self.ellipse, implicit=True)
1291            fit_odr = odr.ODR(fit_data, fit_model, self.fitinit)
1292            #print 'doing ODR'
1293
1294            fit_out = fit_odr.run()
1295
1296            #print 'ODR done'
1297            self.fitinfo = str(fit_out.info)
1298
1299            self.fitbeta = fit_out.beta
1300            self.fiterr = fit_out.sum_square
1301
1302            area = abs(np.pi*self.fitbeta[1]*self.fitbeta[3])
1303            if area> MA or self.fitbeta[3]*2 > MD:
1304                self.fitgood = False
1305            elif self.fitinfo == "2" or self.fitinfo == "3" or self.fitinfo == "1" or self.fitinfo =
1306                self.fitgood = True
1307            return
1308            #fit_out.pprint()
1309
1310        def shadowmeasure_m(self):
1311            '''shadow measuring now that we are doubling the shadow, very straightforward'''
1312            if len(self.mborder) == 0:
1313                self.bouldwid = mindist*2
1314                self.shadlen = 0
1315                self.bouldcent = self.center
1316                self.measured = True
1317                self.fitgood = True
1318                self.fitbeta = [self.center[0],mindist,self.center[1],mindist,0]
1319                self.bouldheight = self.shadlen/np.tan(np.radians(self.inangle))
1320
```

```
            self.bouldwid_m = self.bouldwid*self.resolution
            self.shadlen_m = self.shadlen*self.resolution
            self.bouldheight_m = self.bouldheight*self.resolution
            return
        #despite not being constrained, the fit ellipses are pretty much either veritcal or hori
        # so np.cos(alpha) is essentially either 0,1, or -1, or at least close to it. With this
        #non-zero results (~1 or -1) will be negative, others will be positive

        test = .5 - abs(np.cos(self.fitbeta[4]))

        #build in exception here for small boulders

        #with a minimum of four pixels, no problems were had

        if test <= 0:
            factor = np.cos(self.fitbeta[4])
            self.bouldwid = 2*abs(factor*self.fitbeta[3])
            self.shadlen = abs(factor*self.fitbeta[1])
            self.bouldcent = [self.fitbeta[0],self.fitbeta[2]]
            self.bouldheight = self.shadlen/np.tan(np.radians(self.inangle))
            self.measured = True
        if test > 0:
            factor = np.sin(self.fitbeta[4])
            self.bouldwid = 2*abs(factor*self.fitbeta[1])
            self.shadlen = abs(factor*self.fitbeta[3])
            self.bouldcent = [self.fitbeta[0],self.fitbeta[2]]
            self.bouldheight = self.shadlen/np.tan(np.radians(self.inangle))
            self.measured = True
        self.bouldwid_m = self.bouldwid*self.resolution
        self.shadlen_m = self.shadlen*self.resolution
        self.bouldheight_m = self.bouldheight*self.resolution
        if self.bouldwid > MD or self.bouldheight > MH:
            self.measured = False
        return

    def ellipse (self, beta, coords):
        '''note alpha (beta[4]) is recorded in radians'''
        y = coords[0]
        x = coords[1]

        yc = beta[0]
        xc = beta[2]
        ay = beta[1]
        ax = beta[3]
        alpha = beta[4]
        #alpha is the clockwise angle of rotation
        #alpha = np.arctan(self.sunangle[1]/self.sunangle[0])

        val = (((y-yc)*np.cos(alpha)+(x-xc)*np.sin(alpha))/ay)**2 + (((x-xc)*np.cos(alpha)-(y-yc

        #print beta
        return val

    def AP_ellipse(self, beta, coords):
        y = coords[0]
        x = coords[1]
        yc = beta[0]
        xc = beta[2]
        ay = beta[1]
        ax = self.marea/(np.pi*ay)
        #alpha is the clockwise angle of rotation
        alpha = beta[3]
        return (((y-yc)*np.cos(alpha)+(x-xc)*np.sin(alpha))/ay)**2 + (((x-xc)*np.cos(alpha)-(y-y

    def patchplot(self,filt):
```

```
        #this will be the new ellipse plotting function that uses the matplotlib patches functio
        #need to figure out how to make sure they are plotting correctly though...
        #filter controls whether big ones will be plotted

        color = 'g'
        alpha = .2
        if not filt:
            angle = np.degrees(np.pi-(self.fitbeta[4]%(2*np.pi)))
            shadow = patches.Ellipse([self.fitbeta[2],self.fitbeta[0]], self.fitbeta[3]*2., self
            boulder = patches.Circle([self.bouldcent[1],self.bouldcent[0]], (self.bouldwid/2.),
            return [shadow, boulder]
        if self.measured:
            angle = np.degrees(np.pi-(self.fitbeta[4]%(2*np.pi)))
            shadow = patches.Ellipse([self.fitbeta[2],self.fitbeta[0]], self.fitbeta[3]*2., self
            boulder = patches.Circle([self.bouldcent[1],self.bouldcent[0]], (self.bouldwid/2.),
            return [shadow, boulder]
        else:
            return []


##############Some tools for analysis##########
def current():
    print 'Current Path is: %s'%(PATH)
    print 'Current Filename is: %s'%(FNM)
    print 'current Product ID is: %s'%(ID)
    return
def getshads(runfile, num, silenced = True, mode='r'):
    #open and prep a shadow file, returns open file object and endpoint
    #current()
    try:
        load = open('%s%s%s%s_shadows.shad'%(PATH,runfile,FNM,num), mode)
    except IOError:
        if not silenced:
            print "No shadow file exists"
        return None
    except:
        if not silenced:
            print'Likely broken file'
        return None
    return load
def bulkCFA(runfile,maxnum,maxd,fitmaxd,root):
    ''' runs the CFA protocol on a bunch of files and gives an average
'''
    #set value for maximum on plots (in meters)
    plotmax = 3
    while True:
        new = raw_input('make new CDFs? y/n\n')
        if new == 'y' or new == 'n':
            break
    allCFAs = []
    if new == 'n':
        for i in range(maxnum):
            try:
                data = open('%s%s%s%s_CFA.csv'%(PATH,runfile,FNM,i),'r')
                dat = np.loadtxt(data,delimiter=',')

            except(IOError):
                continue

            allCFAs+=[dat]
        if allCFAs == []:
            print "No CDFs Present"
            new = 'y'
    if new == 'y':
```

```
        for i in range(maxnum):
            #acuire all the CFA data, this will come in the form:
            #CFA[0] = bins
            #CFA[1] = lower error bound CFA
            #[2] = actual CFA data
            #[3] = Upper bound error CFA
            dat = CFA(runfile,i, maxd)
            #Keep an eye on this, would break quickly if CFA is changed
            if len(dat) == 3:
                #plt.plot(dat[0],dat[2],'b-',alpha=.05)
                allCFAs+=[dat]
                #print len(dat)

##    if not any(allCFAs):
##        return None, None
    #extract bins and the three CFAs (upCFAs = upper end of CFA uncertainty, downCFAs lower end
    #above method is being revised, shifting to a sigma calculation
    bins = allCFAs[0][0]
    allCFAs = np.array(allCFAs)
    #CFAs = allCFAs[:,2]
    #upCFAs = allCFAs[:,1]
    #downCFAs = allCFAs[:,3]
    CFAs = allCFAs[:,1]
    CFAsigmas = allCFAs[:,2]
    #tranpose the files to go from (image,bins) to (bins,image) format
    T_CFAs = np.transpose(CFAs)
    T_CFAsigmas = np.transpose(CFAsigmas)
    #T_upCFAs = np.transpose(upCFAs)
    #T_downCFAs = np.transpose(downCFAs)
    #average accross bins, results in 1-D array of length 'bins'
    #also fetch a 25% and 75% CFA curve
    avgCFAs = map(np.average,T_CFAs)
    topqCFAs = map(lambda x: np.percentile(x,75),T_CFAs)
    botqCFAs = map(lambda x: np.percentile(x,25),T_CFAs)
    avgCFAsigmas = []

    for i in T_CFAsigmas:
        sigma = (1/float(len(i)))*(np.sqrt(sum(map(lambda x: x**2,i))))
        avgCFAsigmas+=[sigma]
    avgupCFAs = []
    avgdownCFAs = []
    for i in range(len(avgCFAs)):
        avgupCFAs+=[avgCFAs[i]+avgCFAsigmas[i]]
        avgdownCFAs+=[avgCFAs[i]-avgCFAsigmas[i]]
    #avgupCFAs = map(np.average,T_upCFAs)
    #avgdownCFAs = map(np.average,T_downCFAs)
##    print bins
##    print CFAs
##    print upCFAs
##    print downCFAs
    #fit to each end of the spectrum
    fit_k, fit_r2 = fittoRA(bins,avgCFAs, [1.5,fitmaxd])
    #upfit_k, upfit_r2 = fittoRA(bins,avgupCFAs, [1.5,maxd])
    #downfit_k, downfit_r2 = fittoRA(bins,avgdownCFAs, [1.5,maxd])
    upfit_k, upfit_r2 = fittoRA(bins,avgupCFAs, [1.5,fitmaxd])
    downfit_k, downfit_r2 = fittoRA(bins,avgdownCFAs, [1.5,fitmaxd])
    topqfit_k, topqft_r2 = fittoRA(bins,topqCFAs,[1.5,fitmaxd])
    botqfit_k, botqft_r2 = fittoRA(bins,botqCFAs,[1.5,fitmaxd])

    #calculate RA for each side of uncertainty
    fit_bins = np.linspace(min(bins),max(bins),100)
    fitRA = GolomPSDCFA(fit_bins,fit_k)
    upfitRA = GolomPSDCFA(fit_bins,upfit_k)
    downfitRA = GolomPSDCFA(fit_bins,downfit_k)
```

```
    topqfitRA = GolomPSDCFA(fit_bins,topqfit_k)
    botqfitRA = GolomPSDCFA(fit_bins,botqfit_k)

    #plot them all!
    #errors in format of array [2,bins] all must be positive, lower errors first, see matplotlib
    errors = []
    errors+=[fitRA-downfitRA]
    errors+=[upfitRA-fitRA]
    #plots stack in reverse, so plot what you want on top first (I think...)
    plt.plot(bins,avgCFAs,label = root,zorder=3, c = 'g', marker='|')

    plt.errorbar(fit_bins,fitRA,zorder=2,label = 'RA Envelope',yerr = errors,ecolor = 'k',c = 'k
    plt.plot(fit_bins,topqfitRA,zorder=3,label = '75th Percentile RA')
    plt.plot(fit_bins,botqfitRA,zorder=3,label = '25th Percentile RA')
    plotCFArefs(plotmax)
    plt.xscale('log')
    plt.yscale('log')
    plt.xlim(xmin= 1, xmax = plotmax)
    plt.xlabel('Boulder Diameter (m)')
    plt.ylabel('Cumulative Fractional Area')
    plt.ylim(ymin=10**(-5))
    plt.legend(loc=3)
    plt.title('CFA for image %s at parameters %s'%(root,runfile))
    plt.savefig('%s%sCFAPlot.png'%(PATH,runfile))

    #save the CFA data:
    savecfa = open('%s%sCFAdata_maxd_%s_fitmaxd_%s.csv'%(PATH,runfile,maxd,fitmaxd),'w')
    savecfa.write('Bins,')
    for i in bins:
        savecfa.write('%s,'%i)
    savecfa.write('\n')
    savecfa.write('avgCFA,')
    for i in avgCFAs:
        savecfa.write('%s,'%i)
    savecfa.write('\n')
    savecfa.write('sigma,')
    for i in avgCFAsigmas:
        savecfa.write('%s,'%i)
    savecfa.write('\n')
    savecfa.write('topq,')
    for i in topqCFAs:
        savecfa.write('%s,'%i)
    savecfa.write('\n')
    savecfa.write('botq,')
    for i in botqCFAs:
        savecfa.write('%s,'%i)
    savecfa.write('\n')
    savecfa.close()
    #plt.show()
    return fit_k,topqfit_k,botqfit_k, fit_r2

def CFA(runfile,num,maxd):
    #Produces data for Cumulative Fractional Area, saves and produces plot
    #plt.show must be called after to plot all the data
    load = getshads(runfile,num)
    sizes = []
    if not load:
        return [None]
    while True:
        try:
            dat = pickle.load(load)
        except EOFError:
            break
        if dat.measured:
            sizes+=[dat.bouldwid_m]
```

```
1585
1586            res = dat.resolution
1587            #may need to re-run some boulder detection to get this into the shad file
1588            im_area = dat.im_area
1589    if not any(sizes):
1590        return [None]
1591    #im = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
1592    #im = npma.array(im)
1593    #get the image area in meters
1594    #im_area = float(len(im.compressed()))*res*res
1595    #print area
1596    #create the two ends of the uncertainy spectrum, assumes 1 pixel uncertainty
1597    err = 1
1598    sigma = err*res
1599    sizes = np.asarray(sizes)
1600
1601    sizes = [x for x in sizes if x < maxd]
1602
1603    sizes = np.asarray(sizes)
1604
1605    bins = np.linspace(0,maxd,20*maxd+1)
1606    SFD, binsconf = np.histogram(sizes, bins=bins)
1607    #SFD = np.append(SFD,0)
1608    SFD = SFD/im_area
1609    sizes.sort()
1610
1611    #Total  Area, sum of all boudlers
1612    #TA = sum(areas)
1613    #CFA will be a list of y points, with the x points in sizes
1614    CFA = np.zeros_like(bins)
1615    CFAsigma = np.zeros_like(bins)
1616    #upCFA = np.copy(CFA)
1617    #downCFA = np.copy(CFA)
1618    for i in range(len(bins)):
1619        CFA[i] = float(sum(map(lambda x: np.pi*((x/2.)**2),sizes[sizes>bins[i]])))/im_area
1620        CFAsigma[i] = sigma*np.sqrt((np.pi/im_area)*CFA[i])
1621        #upCFA[i] = float(sum(map(lambda x: np.pi*((x/2.)**2),upsizes[upsizes>bins[i]])))/im_are
1622        #downCFA[i] = float(sum(map(lambda x: np.pi*((x/2.)**2),downsizes[downsizes>bins[i]])))/
1623    #save1 = [bins,upCFA.tolist(),CFA.tolist(),downCFA.tolist()]
1624    save1 = [bins,CFA.tolist(), CFAsigma.tolist()]
1625    save2 = [binsconf[:-1].tolist(),SFD.tolist()]
1626    cfasavefile = open('%s%s%s%s_CFA.csv'%(PATH,runfile,FNM,num),'w')
1627    np.savetxt(cfasavefile, save1,delimiter=',')
1628    sfdsavefile = open('%s%s%s%s_SFD.csv'%(PATH,runfile,FNM,num),'w')
1629    np.savetxt(sfdsavefile, save2,delimiter=',')
1630
1631 ##    CFA[0] = TFA-(areas[0]/area)
1632 ##    for i in range(1,len(sizes)):
1633 ##        CFA[i] = CFA[i-1] - (areas[i]/area)
1634
1635    #plt.plot(bins, CFA, 'b-', alpha=.01)
1636    #plt.xscale('log')
1637    #plt.yscale('log')
1638    #return [bins,upCFA.tolist(),CFA.tolist(),downCFA.tolist()]
1639    return [bins, CFA.tolist(), CFAsigma.tolist()]
1640
1641
1642 def plotCFArefs(xmax):
1643    ''' plots data from golombek 2008 for comparison, user controlled which one
1644 '''
1645    query = 'Plot: 1= just reference \n 2= just TRA_000828_2495 \n 3=Sholes PSP_007718_2350 Hand
1646    option = int(raw_input(query))
1647    if option == 1 or option == 5:
1648        #dat = np.loadtxt('%sGolomRefCFACurves.csv'%(REFPATH),delimiter=',')
1649        xs = np.linspace(.1,xmax)
1650        y20 = GolomPSDCFA(xs,.2)
```

```
        y30 = GolomPSDCFA(xs,.3)
        y40 = GolomPSDCFA(xs,.4)
        y50 = GolomPSDCFA(xs,.5)
        x = np.tile(xs,4)
        y = np.concatenate((y20,y30,y40,y50))
        plt.plot(x,y,'m*',alpha=.2,label = '20,30,40,50% RA',zorder=1)
    if option == 2 or option == 5:
        print 'cannot do until data is provided'
        #dat = np.loadtxt('%sGolomCFADat_TRA_000828_2495.csv'%(REFPATH),delimiter=',')
        #plt.plot(dat[0],dat[1],'k*',alpha=.7)
    if option == 3 or option == 5:
        dat = np.loadtxt('%sPSP_007718_2350Ref.csv'%(REFPATH),delimiter=',')
        plt.plot(dat[0],dat[1],'b*',label='Manual Results')
    return

def PSD(k):
    ''' the particle size distribution for the martian surface, according to Charalambous 2014 (
'''
    #These values based on Charambolous
    t = 3
    p = .79
    return sp.special.binom(t+k-1,k)*(p**k)*(1-p)**t

def GolomPSDCFA(D,k):
    ''' The Model curves used in Golombek's 2012 work, similar to those used in Li 2018
    k is fractional area covered by rocks
    this produces a CFA
'''
    q = 1.79 + .152/k
    F = k*np.exp(-q*D)
    return F

def fittoRA(xdat,ydat,RNG = [1.5,2.25]):
    '''Function to fit CFA results to a rock abundance
    Takes in bins as xdat, CFA as ydat, lower and upper bounds as RNG (assumed 1.5-2.25 to match
'''
    fit_xdat = []
    fit_ydat = []
    for i in range(len(xdat)):
        if xdat[i] <= RNG[1] and xdat[i] >= RNG[0]:
            fit_xdat+=[xdat[i]]
            fit_ydat+=[ydat[i]]
    popt,pcov = sp.optimize.curve_fit(GolomPSDCFA,fit_xdat,fit_ydat,p0=[.1])
    #calculate the R2 on the fit maybe??
    ybar = np.average(fit_ydat)
    SStot = np.sum((fit_ydat-ybar)**2)
    predicted = GolomPSDCFA(fit_xdat,popt)
    SSres = np.sum((fit_ydat-predicted)**2)
    R2 = 1-SSres/SStot
    return popt, R2

def checkbads(runfile,num):
    #this is for when you want to get all the shadows that went awry and plot them
    load = getshads(runfile,num)
    bads = []
    while True:
        try:
            dat = pickle.load(load)
        except EOFError:
            break
        if not dat.measured:
            bads+=[dat]
    patches = []
    for j in bads:
        j.shadowmeasure()
```

```
        patches+=j.patchplot()
    fig=plt.figure(1)
    ax = fig.add_subplot(111)
    #image = np.load('%s%s%s%s_rot_masked.npy'%(PATH, runfile,FNM, num))

    plt.imshow(image, cmap='binary_r')
    for j in patches:
        ax.add_patch(j)
    plt.show()
    return bads

def exportdata(runfile,num):
    #this takes a shadow file and converts it to a csv file with relevant data
    load = open('%s%s%s%s_shadows.shad'%(PATH,runfile,FNM,num),'rb')
    attributes = [['flag','bouldwid','bouldcent_y','bouldcent_x','alpha']]
    while True:
        try:
            dat = pickle.load(load)
        except EOFError:
            break
        attributes +=[[dat.flag, dat.bouldwid, dat.bouldcent,dat.fitbeta[3]]]
    datfile = open('%s%s%s%s_data.csv'%(PATH,runfile, FNM,num),'w')
    for item in attributes:
        datfile.write('%s\n'%item)
    datfile.close()
    #np.savetxt('%s%s%s_data.csv'%(PATH,FNM,num),attributes, delimiter="'")
    return attributes

def FindIdealParams(filename, oldvals = False):
    '''Code to identify ideal parameters for running images, will assume single set of values fo
    returns gam,bound
    '''
    global ID,FNM,PATH,NOMAP,SAZ,NAZ,INANGLE,SUNANGLE,RESOLUTION,ROTANG
    root, ID, NOMAP, num = RunParams(filename)
    PATH = 'C://Users//dhood7//Desktop//MBARS//Images//%s//'%(filename)
    FNM = filename

    #first see if it has been run before and offer to use those:
    if os.path.isfile('%slastrun.txt'%(PATH)):
        #temporary while testing something else
        lastrun = open('%slastrun.txt'%(PATH))
        data = lastrun.readline()
        data = data.rstrip()
        oldgam,oldbound = data.split(',')
        oldgam = float(oldgam)
        oldbound = float(oldbound)
        if oldvals:
            print 'Using old values\n'
            return oldgam,oldbound
        prompt = 'Use old params, gam = %s, bound = %s y/n?\n'%(oldgam,oldbound)
        answer = raw_input(prompt)

        if answer == 'y':
            print 'Using old values\n'
            return oldgam,oldbound
        else:
            if answer !='n':
                print 'Lets assume you meant \'n\''
            print 'OK, lets make new values\n'

    while True:
        imnum = np.random.random_integers(0,num-1)
        image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,imnum))
        print 'Is the following sub-image representative of the entire image?\n'
```

```
        plt.imshow(image, cmap='binary_r')
        plt.show()
        answer = raw_input('y/n\n')
        if answer == 'y':
            print 'image %s selected, lets get params'%(imnum)
            break
        else:
            print 'trying new image...\n'


    userinterp = True
    while userinterp:

        INANGLE, SUNANGLE,RESOLUTION,NAZ,SAZ,ROTANG = start()
        current()
        gam = .6
        image = npma.masked_equal(image, 0)
        imagemod = np.zeros_like(image, dtype=np.uint16)
        top = 0.
        imagemod = image**float(gam)
        top = np.max(imagemod)
        scale = 255./top
        imagemod = imagemod*scale
        imagemod = imagemod.astype(int)
        bound = int(np.average(imagemod.compressed()))
        imagemod = sktrans.rotate(imagemod,ROTANG,resize=True,preserve_range=True)
        image = sktrans.rotate(image,ROTANG,resize=True,preserve_range=True)
        while True:
            print "Let set an initial boundary"
            imageseg = npma.copy(imagemod)
            imageseg = imageseg.astype(float)
            imageseg = imageseg.filled(-1)
            imageseg[imageseg>bound] = bound+1
            imageseg = npma.masked_equal(imageseg, -1)
            imageseg = imageseg.astype(int)
            imageseg.fill_value = 0
            print ('Boundary at %s\n'%(bound))

            print ('(I)ncrease or (D)ecrease the boundary? or is it (C)orrect?\n')

            fig,ax = plt.subplots(1,2,sharex = True,sharey = True)
            ax[0].imshow(image,cmap = 'binary_r')
            ax[1].imshow(image,cmap = 'binary_r',alpha = .5, zorder = 1)
            ax[1].imshow(imageseg,vmin = bound,vmax = (bound+1), zorder = 0)
            plt.show()

            prompt = 'I,D,C?\n'
            answer = raw_input(prompt)
            if answer == 'C' or answer == 'c':
                break
            prompt = 'Move by how much? (integer please)\n'
            size = raw_input(prompt)
            try:
                size = int(size)
            except(ValueError):
                print 'that was not an int'
                size = 5

            if answer == 'I' or answer == 'i':
                bound+=size
            else:
                bound-= size
        while True:
            print "lets refine the boundary"
            imageseg = npma.copy(imagemod)
```

```
            imageseg = imageseg.astype(float)
            imageseg = imageseg.filled(-1)
            imageseg[imageseg>bound] = bound+1
            imageseg = npma.masked_equal(imageseg, -1)
            imageseg = imageseg.astype(int)
            imageseg.fill_value = 0
            print ('Boundary at %s\n'%(bound))

            seg, good, runfile = gamfun(imnum,gam,False,bound,quickrun=True)
            bads = boulderdetect_threadsafe(imnum,seg,runfile,threading.Lock())
            overlapcheck_threadsafe_DBSCAN(imnum,runfile,threading.Lock(),overlap=.0001)
            print ('(I)ncrease or (D)ecrease the boundary? or is it (C)orrect?\n')
            load = getshads(runfile,imnum)
            patches = []
            while True:
                try:
                    dat = pickle.load(load)
                except EOFError:
                    break
                patches += dat.patchplot(True)
            fig,ax = plt.subplots(1,3,sharex = True,sharey = True)
            ax[0].imshow(image,cmap = 'binary_r')
            ax[1].imshow(image,cmap = 'binary_r',alpha = .5, zorder = 1)
            ax[1].imshow(imageseg,vmin = bound,vmax = (bound+1), zorder = 0)
            ax[2].imshow(image,cmap='binary_r')
            for j in patches:
                ax[2].add_patch(j)
            plt.show()
            prompt = 'I,D,C?\n'
            answer = raw_input(prompt)
            if answer == 'C' or answer == 'c':
                save = open('%slastrun.txt'%(PATH),'w')
                save.write('%s,%s'%(gam,bound))
                return gam,bound
            prompt = 'Move by how much? (integer please)\n'
            size = raw_input(prompt)
            try:
                size = int(size)
            except(ValueError):
                print 'that was not an int'
                size = 5

            if answer == 'I' or answer == 'i':
                bound+=size
            else:
                bound-= size




    return gam,bound


def ExamineImage(runfile,num, showblanks,filt = True):
    hasshads = True

    if hasshads:
        load = getshads(runfile,num)
        #need two so you dont reuse same artist, silly but necessary
        patches1 = []
        patches2 = []
        while True:
            try:
                dat = pickle.load(load)
            except EOFError:
```

```
                break
            patches1 += dat.patchplot(filt)
            patches2 +=dat.patchplot(filt)
        #image = np.load('%s%s%s%s_rot_masked.npy'%(PATH,runfile,FNM,num))
        image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
        segimage = np.load('%s%s%s%s_SEG.npy'%(PATH,runfile,FNM,num))
        image = sktrans.rotate(image,ROTANG, resize=True, preserve_range=True)
        #segimage = sktrans.rotate(segimage,ROTANG, resize=True, preserve_range=True)
        image = npma.masked_equal(image, 0)
        filtimage = np.load('%s%s%s%s_flagged.npy'%(PATH,runfile,FNM,num))

        fig,ax = plt.subplots(2,2,sharex = True, sharey = True)
        ax[0][0].imshow(image, cmap='binary_r', interpolation='none')
        ax[0][1].imshow(image,cmap='binary_r',interpolation='none')
        ax[1][0].imshow(segimage,interpolation='none')
        ax[1][1].imshow(filtimage,interpolation='none')
        for j in patches1:
            ax[0][1].add_patch(j)
        for j in patches2:
            ax[1][1].add_patch(j)
        plt.show()
##        plt.figure(1)
##        plt.imshow(image)
##        plt.figure(2)
##        plt.imshow(sktrans.rotate(image, -ROTANG))
##        plt.show()
    elif showblanks:
        #image = np.load('%s%s%s%s_rot_masked.npy'%(PATH,runfile,FNM,num))
        image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
        image = sktrans.rotate(image,ROTANG, resize=True, preserve_range=True)
        image = npma.masked_equal(image, 0)
        filtimage = np.load('%s%s%s%s_flagged.npy'%(PATH,runfile,FNM,num))
        fig,ax = plt.subplots(2,2,sharex = True, sharey = True)
        ax[0][0].imshow(image, cmap='binary_r', interpolation='none')
        ax[0][1].imshow(image,cmap='binary_r',interpolation='none')
        ax[1][0].imshow(filtimage,interpolation='none')
        ax[1][1].imshow(filtimage,interpolation='none')
        plt.show()
    return


def FindBigs(runfile,num,diam = 3):
    '''code to find and identify large boulders thay may be causing issues in the CFA'''
    shads = []
    load = getshads(runfile,num)
    if load == None:
        return None
    while True:
        try:
            dat = pickle.load(load)
            shads+=[dat]
        except(EOFError):
            break
    bigs = []
    patches = []
    total = 0
    for i in shads:
        if i.bouldwid_m>diam:
            bigs+=[i]
            patches+= i.patchplot()
        if i.measured:
            total+=1
    if np.any(bigs) == False:
        return None
    tossout = float(len(bigs)/total)
```

```
    #image = np.load('%s%s%s%s_rot_masked.npy'%(PATH,runfile,FNM,num))
    #replacing this with the original image to save on drive space.
    image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,num))
    image = sktrans.rotate(image,ROTANG, resize=True, preserve_range=True)
    image = npma.masked_equal(image, 0)
    fig,ax = plt.subplots(1,2, sharex = True, sharey = True)
    ax[0].imshow(image,cmap='binary_r',interpolation='none')
    ax[1].imshow(image,cmap='binary_r',interpolation='none')
    for j in patches:
        ax[1].add_patch(j)
    return bigs
def FindExcluded(runfile,maxnum,maxdiam):
    '''finds how many boulders were ignored due to exclusion of large boulders'''
    total = long(0)
    used = long(0)
    for i in range(maxnum):
        shads = []
        load = getshads(runfile,i)
        if load == None:
            continue
        while True:
            try:
                dat = pickle.load(load)
                shads+=[dat]
            except(EOFError):
                break
        for j in shads:
            if j.measured:
                total+=1
                if j.bouldwid_m<maxdiam:
                    used+=1
    exclpercent = 100.*(float(total-used)/float(total))
    print ('%s boulders found in image, %s percent ignored due to diameter'%(total,exclpercent))
    return

def ManualMerge(runfile,num,flags):
    '''Code to manually merge two boulders, only to be used in exception circumstances
        runfile and num specify the image, boulders listed in flags will be merged into one with
        '''
    shads = getshads(runfile,num,mode = 'r')
    boulds=[]
    mergeboulds = []
    while True:
        try:
            dat = pickle.load(shads)
        except(EOFError):
            break
        if dat.flag in flags:
            mergeboulds += [dat]
        else:
            boulds+=[dat]
    print len(mergeboulds)
    shads.close()
    shads = getshads(runfile,num,mode='w')
    for obj in boulds:
        pickle.dump(obj,shads)
    if len(mergeboulds) == len(flags):
        finalflag = mergeboulds[0].flag
        finalarea = mergeboulds[0].im_area
        finalpixels = []
        for i in mergeboulds:
            finalpixels+=i.pixels

        newbould=shadow(finalflag, finalpixels, finalarea)
        newbould.run_prep()
```

```
        newbould.run_fit()
        newbould.run_post()
        pickle.dump(newbould,shads)
        shads.close
        print 'succesfully merged input boulders with flags %s'%(flags)
        return

    else:
        print'Did not find %s boulders, aborting merge'%(len(flags))
        for obj in mergeboulds:
            pickle.dump(obj,shads)
        shads.close()
        return




def OutToGIS(runfile,maxnum,dlow = 1.0, dhigh = 5,extension='.PGw'):
    '''this code will take an entire run and export the boulder data to a GIS-interpretable form
    ''' A key part of this is interpreting the PGW files, which follow this convention:
        6 values on 6 lines:
        A
        D
        B
        E
        C
        F
        these are inputs to two equations for xmap and ymap, the coordinates of the pixel in the
        xmap = Ax + By +C
        ymap = Dx + Ey +F
    '''
    if not os.path.exists('%sGISFiles//%s'%(PATH,runfile)):
        os.makedirs('%sGISFiles//%s'%(PATH,runfile))
    datafile = open("%sGISFiles//%s%s_All_boulderdata.csv"%(PATH,runfile,FNM),'w')
    datafile2 = open("%sGISFiles//%s%s_Confident_boulderdata.csv"%(PATH,runfile,FNM),'w')
    #put in the headers, we will start small with the boulders:
    headers = 'image,flag,xloc,yloc,bouldwid_m,bouldheight_m,shadlen,measured,fitgood,fiterr\n'
    datafile.write(headers)
    datafile2.write(headers)
    #bring in the original rotation information
    rotang_r = np.radians(ROTANG)
    for i in range(maxnum+1):
        #bring in the image each time, wouldnt have to do this if they were all identical
        #but that cant be gauranteed
        try:
            seg = np.load('%s%s%s%s_flagged.npy'%(PATH,runfile,FNM,i))
        except(IOError):
            continue
        shads = getshads(runfile,i)
        if not shads:
            continue
        lycent = len(seg)/2.
        lxcent = len(seg[0])/2.
        seg = None
        o_image = imageio.imread('%s%s%s.PNG'%(PATH,FNM,i))
        o_lycent = len(o_image)/2.
        o_lxcent = len(o_image[0])/2.
        o_image = None

        worldfile = open('%s%s%s%s'%(PATH,FNM,i,extension),'r')
        constants = []
        for line in worldfile:
            val = float(line.rstrip())
            constants+=[val]
```

```
        boulds = []
        marks = []
        while True:
            try:
                dat  = pickle.load(shads)
            except:
                break
            #taking out the filter for now to see if good ones are getting tossed
            try:
                flag = dat.flag
                bouldwid_m = dat.bouldwid_m
                bouldheight_m = dat.bouldheight_m
                shadlen = dat.shadlen
                AR = float(bouldheight_m/bouldwid_m)
                try:
                    fiterr = dat.fiterr
                except:
                    fiterr = None
                #GIS doesnt like mixing data types in csv
                measured = int(dat.measured)
                fitgood = int(dat.fitgood)
                #weve got to do something a little tricky here, we need the pixel location of th
                #first we make a pointmap with flags to keep track
                #boulds+=[dat]

                xpos = dat.bouldcent[1]
                ypos = dat.bouldcent[0]
            except:
                print "failed to retrieve parameters"
                continue
                #change xpos and ypos to origin on the image center
            xpos_c = xpos-lxcent
            ypos_c = ypos-lycent

            #rotate them, must give the negative rotation
            xpos_rot_c = xpos_c*np.cos(rotang_r) - ypos_c*np.sin(rotang_r)
            ypos_rot_c = xpos_c*np.sin(rotang_r) + ypos_c*np.cos(rotang_r)

            #re-reference to the corner of the image
            xpos_rot = xpos_rot_c + o_lxcent
            ypos_rot = ypos_rot_c + o_lycent
            xmap = constants[0]*xpos_rot + constants[2]*ypos_rot + constants[4]
            ymap = constants[1]*xpos_rot + constants[3]*ypos_rot + constants[5]
            #write it all down
            info = '%s,%s,%s,%s,%s,%s,%s,%s,%s,%s\n'%(i,flag,xmap,ymap,bouldwid_m,bouldheight_m,
            datafile.write(info)
            #putting the aspect ratio filter in here, going to use numbers based on Demidov and
            #average = ~.5
            #stdev = ~.28
            #low boundary = .22
            #high boundary = .78
            if shadlen!=0 and bouldwid_m>1.5:
                datafile2.write(info)


        shads.close()
    datafile.close()
    datafile2.close()
    return




def LROCAdapter():
```

```
    '''this code is  intended as a quick fix to looking at LROC images, longer term
better infrastructure should be put in place to make this smoother'''
    #the only real problem is reaching for the metadata, we have to shortcut this
    #these are called when MBARS initializes and with MBARS.start, so we can call this instead a

    INANGLE = None
    NAZ = None
    SAZ = None
    SUNANGLE = None
    RESOLUTION = None

    return


def plotborder(array):
    #this is to speed up plotting up borders
    #array is in form [[y,x],[y,x]....]
    ydat = []
    xdat = []
    l = len(array)
    array = np.copy(array)
    ydat = array[range(l),[0]*l]
    xdat = array[range(l),[1]*l]
    plt.plot(xdat, ydat, "o")
    return
#Range of Floats List, tool to make lists of floats not ints
def ROFL (start, stop, step):
    nums = []
    i = float(start)
    while i < stop:
        nums+=[i]
        i+=step
    return nums


def getangles(ID, path = REFPATH):
    '''
    input is the product ID of the HiRISE image, returns key observation values:
    Incedence angle (sun's angle below peak
    sun direction (emission angle of the sun, clockwise from right of image)
    Resolution(m/px)
    North Azimuth (clockwise north direction from right hand side of image
    Sun Azimuth (
    '''
    cumindex = open(path+'RDRCUMINDEX.TAB','r')
    for line in cumindex:
        try:
            dat = line
        except EOFError:
            print 'No such HiRISE image, check product ID or update CUMINDEX file'
            return None, None, None
        dat = dat.split(',')
        pid = dat[5]
        pid = pid.replace(" ","")
        pid = pid.replace('"','')
        if pid == ID:
            break
    """ key factors we are looking for in the TAB file are as follows:
        Attribute                INDEX
        Product ID:              5
        SubSpacecraft lat/long: 29/30
        SubSolar lat/long:      27/28
        Incedence Angle:        20
        North Azimuth:          25
```

```
        Sun Azimuth:            26
        Resolution:             40
        Projection_type:        41
    """

    naz = float(dat[25])
    saz = float(dat[26])
    inangle = float(dat[20])
    glat = float(dat[29])
    glong = float(dat[30])
    slat = float(dat[27])
    slong = float(dat[28])
    sunangle = groundaz(glat,glong,slat,slong)

    if NOMAP:
        rotang = 90+saz
    else:
        projection = dat[41]
        projection = projection.replace("\"","")
        projection = projection.rstrip()
        if projection == 'EQUIRECTANGULAR':
            #print 'Map Projection is listed as %s, make sure to project to Equirectangualr proj
            rotang = sunangle
        elif projection == 'POLAR STEREOGRAPHIC':
            rotang = -glong+sunangle
        else:
            print 'Projection listed as %s, I am error'%(projection)
            return (None)

    ''' Inangle is returned as measured from azimuth, sunangle as clockwise from North'''

    #HiRISE_INFO has the resolution, PID is on index 11, resolution on the last one
    info = open(path+'HiRISE_INFO.txt','r')
    for line in info:
        try: dat=line
        except EOFError:
            print 'No such HiRISE image, check product ID or update CUMINDEX file'
            return None, None, None
        dat = dat.split(',')
        if ID == dat[11]:
            break
    res = dat[-1].strip()
    res = float(res)



    return inangle, sunangle, res, naz, saz, rotang

def decon_PSF(image, iterations = 10, binned = False):
    '''my kernel for the HiRISE Point Spread Function, used to deconvolve the image
        from McEwen 2007, the FWHM of the PSF is 2 if the image is unbinned, or 1 if binned at 2
        as such, we will allow for both options
        '''
    #THIS DOES NOT WORK< NOT SURE WHY
    if binned:
        sigma = 1./(2*np.sqrt(2*np.log(2)))

    else:
        sigma = 2./(2*np.sqrt(2*np.log(2)))
    #PSF = np.zeros((11,11))
    k1d = np.linspace(0,10,11)
    k1d = gauss(k1d,sigma,5)
    k2d = np.outer(k1d,k1d)
    k2d = k2d/k2d.sum()
```

```
    image = image.astype(float)
    #trying out manual iterating since the internal iterations arent working

    #image_decon = skrestore.richardson_lucy(image, k2d,1,False)
    image_decon = skrestore.wiener(image, k2d, 1.0, None, True, False)

    return image_decon

def groundaz(glat, glon, slat, slon):
    """
    Translated directly from the ISIS GroundAzimuth C function
    inputs and outputs in degrees
    first pair is ground lat/lon, second is either subsun or subspacecraft
    Originally Authored by Caleb Fassett
    """

    if (glat >= 0):
        a=radians(90-slat)
        b=radians(90-glat)

    else:
        a=radians(90+slat)
        b=radians(90+glat)

    cslon=slon
    cglon=glon
    if cslon>cglon:
        if ((cslon-cglon)>180):
            while ((cslon-cglon)>180):
                cslon=cslon-360

    if cglon>cslon:
        if ((cglon-cslon)>180):
            while ((cglon-cslon)>180):
                cglon=cglon-360

    if (slat>=glat):
        if (slon>=glon):
            quad=1

        else:
            quad=2

    else:
        if (slon>=glon):
            quad=4

        else:
            quad=3

    dlon=radians(glon-slon)
    if dlon<0:  dlon=-dlon

    c=acos(cos(a)*cos(b)+sin(a)*sin(b)*cos(dlon))
    az=0

    if (((sin(b)==0)|(sin(c)==0))|((sin(b)*sin(c))==0)):

        return az

    else:

        try:
            bigA=degrees(acos((cos(a)-cos(b)*cos(c))/(sin(b)*sin(c))))
        except:
```

```
            return az

        ### This is a kludge -- domain errors (perhaps near 180 lon)

        if (glat>=0):
            if ((quad==1)|(quad==4)):
                az=bigA

            else:
                az=360-bigA

        else:
            if ((quad==1)|(quad==4)):
                az=180-bigA

            else:
                az=180+bigA

        return az

def RunParams(filename):
    '''a long term solution to running multiple images, this will
create text files in each image file so that various parameters dont need to be
reentered, files read:
root
ID
NOMAP
Panels
'''
    parampath = '%s%s//runparams.txt'%(BASEPATH,filename)
    if not os.path.isdir('%s%s'%(BASEPATH,filename)):
        print 'no such directory\n'
        return None,None,None,None
    if os.path.isfile(parampath):
        print 'Running Parameters found\n'
        #get these parameters
        paramfile = open(parampath,'rb')
        info = paramfile.readline()
        info = info.rstrip()
        root,mbarsid,mbarsnomap,panels = info.split(',')
##        root = paramfile.readline()
##        root = root.rstrip

##        mbarsid = paramfile.readline
##        mbarsid = mbarsid.rstrip()

##        mbarsnomap = paramfile.readline()
        if 'True' in mbarsnomap:
            mbarsnomap = True
        else:
            mbarsnomap = False

##        panels = paramfile.readline()
##        panels = panels.rstrip()
        panels = int(panels)


    else:
        print "no parameters found, making new file\n"

        params = open(parampath,'wb')
        q1 = 'enter MBARS ID\n'
        mbarsid = raw_input(q1)
```

```
        q2 = 'Is the image projected (i.e. north is up?) y/n \n'
        while True:
            answer = raw_input(q2)
            if answer == 'y':
                mbarsnomap = False

                break
            elif answer == 'n':
                mbarsnomap = True
                break
            else:
                print 'y/n \n'
        #retrieve number of panels
        files = os.listdir('%s%s'%(BASEPATH,filename))
        files = [s for s in files if '.PNG' in s]
        files = [s.replace(filename,'') for s in files]
        files = [filter(lambda s: s in '0123456789',j) for j in files]
        files = [int(s) for s in files]
        panels = np.max(files) +1
        print('%s panels found'%panels)

        root = filename

        #write it all in:
        params.write('%s,'%root)
        params.write('%s,'%mbarsid)
        params.write('%s,'%mbarsnomap)
        params.write('%s'%panels)
        params.close()

    return root, mbarsid, mbarsnomap, panels


########INITIALIZATION#########
INANGLE,SUNANGLE,RESOLUTION,NAZ, SAZ, ROTANG = start()
```

# Appendix D. Copyright agreement for Chapter 1

**Journal of Geophysical Research: Planets**

**Published by Wiley on behalf of American Geophysical Union (the "Owner")**


**COPYRIGHT TRANSFER AGREEMENT**

Date: 2016-09-09

Contributor name: Don, R Hood

Contributor address: Louisiana State University E235 Howe Russel Louisiana State University Baton Rouge LA 70803 United States

Manuscript number: 2016JE005046

Re: Manuscript entitled Assessing the geologic evolution of Greater Thaumasia, Mars (the "Contribution")

for publication in Journal of Geophysical Research: Planets (the "Journal")

published by Wiley Periodicals, Inc. ("Wiley")

Dear Contributor(s):

Thank you for submitting your Contribution for publication. In order to expedite the editing and publishing process and enable the Owner to disseminate your Contribution to the fullest extent, we need to have this Copyright Transfer Agreement executed.
**Publication cannot proceed without a signed copy of this Agreement.**

---

**INFORMATION CONCERNING YOUR COPYRIGHT**

The Copyright Law enacted in 1978 requires the American Geophysical Union to obtain specific rights to articles published. This legal requirement does not alter in any way the long-standing relationship between AGU and its authors, nor does it change the philosophy behind our practice of copyrighting our journals and books.

AGU's philosophy recognizes the need to ensure that authors have a say in how their works are used and the necessity to foster broad dissemination of scientific literature while protecting the viability of the publication system. Authors still retain all proprietary rights other than copyright (such as patent rights), the right to present the material orally, the right to reproduce figures, tables, and extracts properly cited.

**Permission to Deposit an Article in an Institutional Repository**

AGU allows authors to deposit their journal articles if the version is the final published citable version of record, the AGU copyright statement is clearly visible on the posting, and the posting is made 6 months after official publication by the AGU.

**Public Domain**

Work prepared by U.S. Government employees in their official capacities is not subject to copyright in the United States. Such authors may place their papers in the public domain, which means that it may be freely copied, republished, and redistributed. In order for the work to be placed in the public domain, all authors must be official U.S. Government employees. If at least one author of the work was privately employed, copyright should be transferred to AGU by any of the privately employed authors.

**Works Prepared by Both Privately Employed Persons and Official U.S. Government Employees**

To ensure that we can continue to promote wide availability of our publications, AGU requests all non-U.S. Government employees to transfer copyright to us. This transfer permits us to continue publishing our journals and books in all their various formats, to grant permission to abstracting and indexing services to cover our publications, and to grant permission for photocopying beyond the limits defined in the law. In instances where authorship consists of both U.S. Government and privately employed individuals, we require at least one privately employed author to transfer copyright to AGU. This kind of transfer provides ultimate protection and broad dissemination of the work.

**Rights Other Than Copyright**

Privately employed authors who have written articles in their official capacities as employees should also transfer copyright to AGU. The author's employer retains the same rights as individual authors. AGU claims no right to the work other than copyright; the author's employer retains all other rights such as patent rights.

**Works Supported by U.S. Government Grants or Contracts**

Authors who are publishing works supported by a U.S. Government grant or contract are requested to transfer copyright to AGU. This kind of transfer permits the broad dissemination of the work while recognizing the U.S. Government's prior license to use the work for noncommercial purposes.

**Copyright Permission for Reprinted/Modified Figures/Tables**

If any of the figures/tables in your article are reprinted or modified from another source (this includes any that were redrawn but are basically unaltered or have only slight modifications), you must obtain and provide AGU with original letter(s) (fax or e-mail not acceptable) from the copyright holder(s) granting permission to use them in your article before your article may be published. If the original copyright holder has given blanket permission for reuse with credit, a copy of the published permission statement is sufficient. If the material is in the public domain, please provide confirmation of this in an original letter. Material originally published in AGU publications does not require copyright permission as long as proper credit is given.

**[ X ]** I agree to the COPYRIGHT TRANSFER AGREEMENT as shown above and have obtained written permission from all other contributors to execute this Agreement on their behalf.

Contributor's signature (type name here): Donald R Hood

Date: 9/9/2016

**SELECT FROM OPTIONS BELOW:**

**[ X ] Contributor-owned work**

**[ ] Certification of U.S. Government work**

*Note to U.S. Government Employees*

*Select this option and type Contributor name into the Contributor's signature section above only if ALL authors were U.S. Government employees at the time the work was prepared. (See above for explanation.) I certify that the article referenced above was prepared solely by U.S. Government employees as part of their official duties and therefore legally cannot be copyrighted. I confirm that this article has not been published previously elsewhere, nor is it under consideration by any other publisher.*

**[ ] Certification of Crown Government Copyright Reserve**

*Note to Crown government Employees*

**For Crown Copyright this form cannot be completed electronically and should be printed off, signed in the Contributor's signatures section above by the appropriately authorised individual and returned to the Journal production editor by email.** For production editor contact details please visit the Journal's online author guidelines.

This is to certify that ALL authors are or were bona fide officers of a Crown government which reserves its own copyright under national law, and the work was prepared on behalf of the _____ (Government or Agency Name) as part of their official duties. Copyright in the paper is therefore reserved and may not be transferred.

The following nonexclusive rights are hereby given to the American Geophysical Union (AGU).

1. The right to use, print, and/or publish in any language the above-mentioned work or any part thereof, provided that the name of the author(s) and the Crown government affiliation are clearly indicated.
2. The right to grant the same rights to others to print and publish the work subject to the above proviso.
3. The right to collect royalty fees.

In return for these rights, AGU hereby certifies that whenever approached by third parties for individual permission to use, reprint, or republish specified full articles (except for classroom use, library reserve, or to

reprint in a collective work of the AGU) the corresponding author's or employer's permission will also be required. AGU is not obligated to print the Crown government copyright credit line. *If your status as a government employee legally prevents you from signing this Agreement, please contact the Journal production editor.*

**[ ] Other**

Including Other Government work or Non-Governmental Organisation work

*Note to Non-U.S., Non-U.K. Government Employees or Non-Governmental Organisation Employees*

**For Other Government or Non-Governmental Organisation work this form cannot be completed electronically and should be printed off, signed in the Contributor's signatures section above by the appropriately authorised individual and returned to the Journal production editor by email.** For production editor contact details please visit the Journal's online author guidelines. *If your status as a government or non-governmental organisation employee legally prevents you from signing this Agreement, please contact the Journal production editor.*

Name of Government/Non-Governmental Organisation:

_____

**[ ] Company/institution owned work (made for hire in the course of employment)**

**For "work made for hire" this form cannot be completed electronically and should be printed off, signed and returned to the Journal production editor by email.** For production editor contact details please visit the Journal's online author guidelines.

Name of Company/Institution:

_____

Authorized Signature of Employer: _____

Date: _____

Signature of Employee: _____

Date: _____

## Vita

Donald Ramsey Hood, born in Dallas, Texas acquired his bachelor's degree in physics from Carnegie Mellon University. For his graduate education, he shifted fields, choosing to study Geology at Louisiana State University. Here, he studied planetary geology, with an emphasis in remote sensing of the martian surface. His time at LSU has seen him give talks in the U.S. at the Lunar and Planetary Institute, China at the International Symposium on Lunar and Planetary Science, and Sri Lanka at the National Institute of Fundamental Studies and the University of Sri Jayawardenepura. He has also participated in several international meetings including regularly at the Lunar and Planetary Science Conference. He has published a first-authored paper and led fieldwork as far as Sri Lanka exploring serpentinite outcrops along a tectonic suture zone and as nearby as St. Helena Parish investigating a potential impact structure. He has been awarded numerous academic scholarships from the New Orleans Geological Society, as well as several from LSU. He has successfully proposed to two Louisiana Space Consortium Graduate Assistantships and been a member of successful proposal teams to the Louisiana Space Consortium and the NASA Mars Data Analysis Program. After completing his Ph.D., he will begin his post-graduation career in College Station, Texas.