Are Mountain-forming Rocks of Thaumasia Highlands, Mars, other than Basalt/Basaltic Andesite?: Machine Learning-based Evaluation of TES Data and Implications on Early Evolution of Mars

J.M. Dohm[a,b], L. Scharenbroich[d], R. Wang[c], J.B. Dalton[e],

R. Castano[d], T.M. Hare[f], V.R. Baker[a,b]

[a]*Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ, 85721*
[b]*Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ*
[c]*Harvey Mudd College*
[d]*Jet Propulsion Laboratory, Pasadena, CA, 91101*
[e]*NASA Ames Research Center, Moffett Field, CA, 94035*
[f]*U.S. Geological Survey, Flagstaff, AZ, 86001*

E-mail: jmd@hwr.arizona.edu

Manuscript pages: 62

Figures: 17

1

Proposed running title: Evaluation of Thaumasia Highlands, Mars.

Editorial correspondence to:

James Dohm
Department of Hydrology and Water Resources,
University of Arizona,
Tucson, AZ, 85721

Phone: 520-626-8454
Fax: +34 91 717 68 45

E-mail: jmd@hwr.arizona.edu

ABSTRACT

Do the mountain building rocks of the ancient Thaumasia highlands mountain range differ from the basalt-basaltic andesite compositions inferred for much of Mars? Distinct characteristics of the mountain range include magnetic signatures, complex tectonic structures, cuestas, hog backs, and valley networks (characteristics similar to the mountain ranges of Earth), revealed through Viking, Mars Odyssey (MO), and Mars Global Surveyor (MGS) information. We have applied Machine Learning and Geographic Information Systems (GIS) techniques to published geologic and Thermal Emission Spectrometer (TES) hyperspectral image cube information. When compared to the ancient mountain-forming materials, the younger plains-forming materials record a different TES emissivity signature. This finding is consistent with Viking-era, geological mapping-based interpretations that the mountain-forming materials could be comprised of a diversity of rock types. These include a suite of mostly crystalline igneous and/or metamorphic rocks that generally underlies the sedimentary rock sequence, known as basement complex. In contrast, the plains-forming materials are mostly volcanic. Here, we describe one of the ancient mountain ranges of Mars, Thaumasia highlands, the importance of determining its composition, a useful Machine Learning approach that may help improve our understanding of Mars' geology, the significance of the mountain ranges to unraveling the early evolutionary phases of Mars, and a rationale for continued investigation of such ancient features of Mars through MGS, MO, Mars Reconnaissance Orbiter (MRO), and future science-driven reconnaissance missions.

KEYWORDS: Mars, mountains; tectonism; machine learning; basement complex.

1. Introduction

Detailed geologic mapping indicates that ancient mountain ranges form the southern and southeastern margins of the Thaumasia igneous plateau in the southeast part of Tharsis, Thaumasia highlands and Coprates rise, respectively (**Figures 1 and 2**) (Scott and Tanaka, 1986; Dohm et al., 2001a,b). Stratigraphic and cross-cutting relations, impact crater statistics, an order of magnitude greater density of tectonic structures in the mountains compared to the younger lava plains of the shield complex of Syria Planum, and distinct magnetic signatures indicate that the mountain ranges began to form prior to the shut down of the magnetosphere, prior to the evolution of Tharis (Dohm and Tanaka, 1999; Acuña et al., 1999, 2001; Connerney et al., 1999, 2005; Dohm et al., 2001a,b; Anderson et al., 2001, 2004; Arkani-Hamed, 2003). As such, large tectonic structures such as the Thaumasia highlands mountain range are important markers of an ancient Mars. In order to improve our understanding of the early geological and geochemical evolution of Mars, it is critical to determine the mineralogic compositions (e.g., rock types) of the ancient structures (e.g., for assessing bulk composition, etc.). For example, is the Thaumasia highlands mountain range (**Figures 1 and 2**) comprised of materials other than basalt and basaltic andesite? Is there a distinct ancient rock record at existing orbiter resolutions such as in the Thaumasia highlands that is not obscured by Tharsis-era geologic activity, activity which dates back to at least the Middle Noachian epoch (Dohm et al., 2001a) or more than 3.7Ga to present (age chronology based on Hartmann an Neukum (2001))? More specifically, are there rocks other than basalt/basaltic-andesite such as basement complex rocks in the ancient mountain range (a suite of mostly crystalline igneous and/or metamorphic rocks that generally underlies the sedimentary rock sequence) that are poking up

through an aeolian mantle chiefly composed of basaltic/basaltic-andesite materials, free of thin-skinned secondary weathering products (e.g., see Squyres et al., 2004a,b for information on MER-based investigations that reveal secondary weathering rinds), or contributing fragmented materials with fresh surfaces to the development of fan materials on and along the southern flanks of the Thaumasia highlands mountain range (Dohm et al., 2001b), all at sufficient areal extent to be detected from orbital platforms ranging from Thermal Emission Spectrometer (TES) to Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on the Mars Reconnaissance Orbiter (MRO)?

Ground-truthing of IKONOS satellite imagery at 1 m/pixel resolution and ASTER imagery at 90 m/pixel (whereas TES has a 3 km/pixel resolution) revealed a diversity of rock types not recognized from IKONOS or ASTER of a structurally controlled basin in the Atacama Desert (Life in the Atacama Ground-truthing Workshop, 2006). The Atacama Desert is considered to provide a unique testing ground for rover exploration on Mars, especially since much of this temperate desert, which covers the northern one-third of Chile (18°S to 28°S), is hyperarid, and has well established magmatic, tectonic, aqueous, and aeolian histories, all of which are not necessarily mutually exclusive (e.g., Chong et al., 1999). For example, while some rock materials such as clay and volcanics were accurately identified from the satellite data, other rock types such as granites and granodiorites, which were poking up through alluvial fan materials and partly contributing locally to the alluvial fan materials, were not identified from the satellite data. These materials represent a significant part of the ancient geologic record of the region that would have otherwise been disregarded, if not for field-based examination. In addition, addressing whether the mountain range is comprised of rocks other than basalt is further prompted by satellite-based characteristics that are similar in many respects to the mountain ranges of Earth of

diverse rock compositions (Dohm and Tanaka, 1999; Dohm et al., 2001a,b 2002a, 2005), which includes magnetic signatures (e.g., Acuña et al., 1999, 2001; Connerney et al., 1999, 2005; Arkani-Hamed, 2003), Gamma Ray Spectrometer (GRS)-based elevated silicon (when compared to younger volcanics of Tharsis; see **Figure 3**), complex tectonic structures, cuestas, hog backs, and valley networks, and the potential significance to the early evolution of Mars.

Mars is commonly viewed as a one-plate planet since its incipient development based largely on geophysical and geochemical arguments mainly from analysis of SNC meteorites (e.g., Halliday et al., 2001; Nimmo and Tanaka, 2005). Similar to the potential problem of obscuration from satellite-based perspective of an ancient rock record through the mantling and secondary weathering processes, is it possible that the SNC—associated sampling is only revealing a part of the geological and geochemical histories of Mars? Whether the Thaumasia highlands mountain range is other than basalt/basaltic andesite is a significant query to address since the range records an ancient part of Mars' evolution, which includes hypothesized Earth-like evolutional phases (Baker et al., 2002), including plate tectonism (Sleep, 1994; Márquez et al., 2004), particularly during its embryonic stages of evolution (Dohm et al., 2002; Fairén et al., 2002; Fairén and Dohm, 2004; Connerney et al., 2005). These are invoked to account for many enigmatic features in regard to the geological history of Mars. Included in these features are the ancient mountain ranges, Thaumasia highlands and Coprates rise, extremely large ancient tectonic structures such as Tempe Mareotis Fossae and Phlegra Montes, and ancient geologic provinces such as Terra Cimmeria and Arabia Terra (**Figures 1 and 2**) that record pronounced magnetic anomalies (see Dohm et al., 2005 and Connerney et al., 2005).

Below we detail the approach that was used to evaluate the TES data, which includes the variables that may influence its signature, present the results, and discuss why such an investigation may improve our understanding of the early geological evolution of Mars. We also explain why it is necessary for further investigation of ancient features and geologic provinces of Mars (see Dohm et al., 2005) through MGS, MO, Mars Reconnaissance Orbiter (MRO), and future science-driven reconnaissance missions (e.g., Fink et al., 2005; Schulze-Makuch, 2005).

As part of the geological study in the region of interest on Mars, the thermal emission spectroscopy (TES) data taken by the Mars Global Surveyor (MGS) spacecraft are analyzed by machine learning methodologies using both unsupervised (clustering) and supervised (classification) techniques. Canonical spectral endmembers representing atmospheric contributions as well as surface types and pure mineral endmembers from the Arizona State University (ASU) Spectral Library (e.g., Christensen et al., 2000; Bandfield et al., 2000a; Smith et al., 2000; Bandfield and Smith, 2003; Wyatt and McSween, 2002; Ruff, 2003) were used for the Machine-Learning clustering investigation and the published Viking-era geologic information of Dohm et al. (2001b) was used to select prime mountain- and plains-forming TES spectral types for the Machine-Learning supervised investigation.

2. Methodology

To determine whether there are distinctions in spectral signatures among the ancient mountain-forming materials of Thaumasia Highlands and the younger lava plains of the complex shield volcano, Syria Planum (e.g., Syria, Sinai and Solis Planae; for detailed mountain range and shield complex information, see Dohm et al., 2001a and Anderson et al., 2004), our approach (described in detail below) requires the

following steps; (1) determine distinguishing TES information using unsupervised and supervised Machine Learning techniques, which included TES-based Type 1 vs. Type 2 spectral information (based on Bandfield, 2000) and selected prime mountain-forming pixels vs. prime plains-forming pixels (pixels chosen in prime geologic-based locales based on Dohm et al., 2001b), (2) select "quality" tracks of TES data that have the least contributions from clouds, atmospheric and surface dust, or noise for comprehensive analysis, (3) apply unsupervised and supervised techniques to the "quality" tracks of TES data to visualize whether there is a spectral distinction between the ancient mountain-building materials and plains-forming materials, (4) separate out ancient mountain-forming materials from relatively young plains-forming materials based on published geologic information using Geographic Information Systems (GIS) for coupling with the Machine Learning-based results, and (5) quantify the comparison among the Machine Learning-based results and the published geologic information, which includes determining the density of untrained Type 1 and Type 2 pixels and trained mountain-forming and plains-forming pixels (e.g., total amount of pixels of a specific type divided by the total area of either mountain-forming or plains-forming units).

2.1. Definition of the Two Distinct TES Mineralogical Units, Types 1 and 2.

In addition to the often-referred-to "global dust component" of Mars and hematite, two distinct mineralogical units have been identified, characterized, and interpreted based on mid-infrared spectra from the Thermal Emission Spectrometer (TES) on the Mars Global Surveyor spacecraft (Bandfield et al., 2000). Material comprising "surface type 1" (ST1) occurs primarily in low-albedo regions in the southern hemisphere. Due to its spectral signature, which compares to terrestrial volcanic rocks, and its domination

of the Syrtis Planitia region, it has been referred to as "Syrtis type" spectra (Wyatt and McSween, 2002; Ruff, 2003). Syrtis is a shield volcano that began to from during the Late Hesperian (Greeley and Guest, 1987).

On the other hand, spectral comparisons with terrestrial mineral assemblages have not yielded a unique rock composition signature for the materials denoted as "surface type 2" that occur in the northern low-albedo regions (e.g., Bandfield et al., 2002), and often referred to as "Acidalia type". Oxidized Shergotty-Nakhla-Chassigny (SNC) type basalts, palagonitized basalt, basaltic andesite, silica-coated basalts, and aqueous weathering processes of basalts appear consistent with the mineralogy of ST2 (e.g., Ruff, 2003). This is in part based on Pathfinder Alpha Proton X-Ray Spectrometer (APXS) data, consistent with an andesitic composition (Wänke et al., 2001). Data acquired from the Mars Odyssey Thermal Emission Imaging System (THEMIS) also indicates that evolved silica-enriched lava could be an important component of ST2 materials (Christensen et al., 2005). Silica-enriched rocks on Mars would have significant geophysical implications, as it may strengthen the case for hypothesized past crustal recycling processes (Sleep,, 1994; Baker et al., 2002; Dohm et al., 2002; Fairén et al., 2002; Fairén and Dohm, 2004).

In addition, ST2 may mark diverse rock assemblages sourcing from ancient crustal rock materials (both shallow and deep seated) deposited in the northern plains by Tharsis-driven flooding (Baker et al., 1991; Dohm et al., 2001a, 2001c; Clifford and Parker, 2001), and to a lesser extent, spring-fed activity along parts of the highland-lowland boundary (Tanaka et al., 2003, 2005) and Elysium-triggered flooding. The flood events are hypothesized to have resulted in water bodies ranging from oceans to lakes (e.g., Fairén et al., 2003). The hypothesis that acidic aqueous conditions for the most extensive and oldest putative water body with a shoreline located to the south and

above the Opportunity Landing may have inhibited carbonate formation at the martian surface (Fairén et al, 2004) was later supported by MER-based investigations, which includes the identification of jarosite (Moore, 2004; Catling, 2004). Evaporites may have also formed in such environments from flood and spring-fed brines, ponding of flood waters, and eventual depletion of the standing water. The existence of a diverse rock assemblage in the reported Type 2 regions would not be inconsistent with findings based on the Mars Odyssey, Mars Express, and the Mars Exploration Rovers Missions, which are increasingly showing a greater diversity of rock types for Mars such as hematite, andesite, sulfates, layered sedimentary deposits, and even quartz-bearing granitoids, etc. (e.g., Malin and Edgett, 2000; Bandfield et al., 2000; Christensen et al., 2001a,b, 2004, 2005; Rieder et al., 2004; Gendrin et al., 2005). The source regions for the flood deposits in the northern plains materials (e.g., Vastitas Borealis Formation; e.g. Scott and Tanaka, 1986; Tanaka et al., 2005) may have been comprised of such materials, consistent with published Viking-era geologic mapping information (e.g., Scott and Tanaka, 1986; Dohm et al., 2001b)

2.2. Selection of High-Quality TES data tracks for Evaluation

The spectral datasets are composed of band emissivity values obtained from the TES archives. TES emissivity spectra acquired on separate orbits were compiled and registered into a hyperspectral image cube. All TES observations exhibited influences of CO_2, atmospheric dust, surface dust, and/or water ice clouds. Spectra were categorized based upon strengths of atmospheric dust features centered near 1079 cm^{-1} and water ice features at 500-900 cm^{-1}. A dust cover index based on dust emissivity at 1350-1400 cm^{-1} (Ruff and Christensen, 2002) was also used to distinguish orbits dominated by dust

emissivity. Observations exhibiting high dust or water ice content reveal little about the surface composition.

To ensure the highest data quality for our analysis, we applied a Minimum Noise Fraction (MNF) transformation (Green et al., 1998; Boardman and Kruse, 1994) to the assembled TES image cube data. The MNF transform is essentially a decorrelation stretch, similar to that used in Principal Components Analysis (PCA) with the distinction that an estimated noise covariance matrix is first applied to the observations prior to the decorrelation. This allows us to highlight orbits dominated by atmospheric effects (Stockstill et al., 2005; Dalton et al., 2004) and determine the inherent dimensionality of the data while segregating noise from higher-level information content. In the MNF-transformed data of **Figure 4**, color variations represent the information content of each pixel, projected into the PCA coordinate space. The variations are more important than the colors themselves: where the variations of adjacent orbits correlate with surface features such as craters, the emissivity values are more representative of the surface; however orbital tracks whose color variations "stand out" from their neighbors and tend to be uncorrelated, typically are dominated by atmospheric effects. As in the methods outlined by Stockstill et al. (2005), orbits exhibiting high levels of atmospheric interference and dust loading were removed and the process repeated until color variations correlated primarily with surface units. Orbits with low surface temperatures, high atmospheric dust or water ice content, low signal levels, or instrumental artifacts were also removed, as were side-looking limb profiles.

Based on the MNF transformation results, dust cover index, and evaluation of known spectral features, orbital tracks were categorized as excellent, good, acceptable, dusty, very dusty, dusty and cloudy, cloudy, partially cloudy, and noisy. Almost none of the data were considered to be of "good" or "excellent" quality as all tracks contained

some atmospheric dust. We have attempted to limit our data to only "acceptable" tracks. All "cloudy", "dusty" and "dusty and cloudy" tracks were excluded from the analysis. However, in order to obtain better coverage of the region under investigation, some "noisy" and "partially cloudy" tracks are included. Of the original 238 orbital tracks assembled into the hypercube, only 11 were judged of sufficient quality to be retained for the analysis. In the MNF-transformed image cube shown in **Figure 5**, these orbits can be seen to correspond primarily to either plains (blue and green) or mountain (red to yellow) units. Some atmospheric dust (orange) and water ice cloud (white) influence remains, and must be taken into account in the final discussion.

2.3. Application of Machine Learning to the TES Information.

Prior to analysis, all the spectral datasets were normalized to common scales by aligning spectral profiles of the minerals in the spectral library to the wavelengths of the observed TES data. After preprocessing, two varieties of machine learning techniques are applied to the TES observations; unsupervised clustering (k-means, mixtures of Gaussians and mixtures of *t*-distributions), (MacQueen, 1967; Bishop, 1995; Peel and McLachlan, 2000) and supervised classification (e.g., Bayesian). These algorithms are applied directly to the TES spectral data and to the mineral abundance estimates obtained from a linear spectral unmixing algorithm (Vélez-Reyes and Rosario, 2004) using endmembers from the ASU Spectral Library and the set of eight canonical endmembers (Bandfield, 2000; Smith, 2000; Bandfield, 2003).

Each algorithm was applied to the data using Euclidean and correlation distance metrics. The Euclidean metric groups data according to a L_2 metric while the correlation metric (e.g., spectral angular mapping, or SAM) groups spectra according to their

angular separation in a vector space representation. Effectively, the Euclidean metric groups spectral data by magnitude and the correlation metric groups data by shape.

2.3.1 Spectral Alignment

Since the TES spectral wavelengths differ from end member wavelength, the spectra must be aligned. For this purpose, Gaussian Process (GP) regression (Quinonero-Candela, 2004; MacKay, 1998) is applied to estimate the value of the endmember spectrum at the wavelengths of the TES instrument. The spectral library emissivities are interpolated at the observed TES wavelengths rather than vice-versa for three reasons: (1) it is more computationally efficient, as there are only a few tens of endmembers but many thousand TES observations, (2) the spectral library data is of higher quality, and (3) to interpolate the observations would result in loss of observational information. The decision to use GP regression instead of a simpler regression or interpolation technique, such as linear or spline regression, is motivated by the fact that GP regression is a formal probabilistic technique which produces a measure of the uncertainty in addition to a value estimate. This uncertainty is propagated though the spectral unmixing process to more accurately bound the margin of error in the abundance estimates.

A Gaussian Process is uniquely defined by a covariance function which imposes a prior on the functional form of the regressor. For interpolating the spectral library endmember emissivities, we employ a Gaussian covariance function of the form,

$$C(x,y) = \frac{\theta_1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-y}{\sigma}\right)^2\right\} + \theta_2$$

The most sensitive parameter is the scale parameter, sigma, which controls the influence of nearby data points on the regression estimate. Increasing sigma forces a smooth regression function at the expense of deviating from the given spectral values. **Figure 6** shows the alignment of a regression curve to a canonical endmember. The second image zooms in on the region near 1um to show the uncertainty in the estimate.

2.3.2 Linear Spectral Unmixing

In order to estimate the abundances of minerals present in a spectral observation, a linear unmixing algorithm is used (Velez-Reyes, 2004) where the problem is recast as a least distance, least squares problem. There are two major steps in this procedure. First, the QR decomposition of the endmember matrix A is taken and the problem rewritten as:

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{R}\mathbf{x} - \mathbf{c}_1\|$$
$$s.t. \quad \mathbf{x} \geq 0 \text{ and } \sum_i x_i \leq 1$$

where $\mathbf{c}_1$ is the appropriate upper portion of the vector $\mathbf{Q}^T\mathbf{b}$ and the sum-to-one constraint is relaxed in order to incorporate a dark endmember. After performing the following transformation,

$$\mathbf{z} = \mathbf{R}\mathbf{x} - \mathbf{c}_1$$

and solving the optimization,

$$\hat{\mathbf{z}} = \arg\min \|\mathbf{z}\|$$
$$s.t. \quad \mathbf{G}\mathbf{z} \geq \mathbf{g}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{1}^T \end{bmatrix} \mathbf{R}^{-1}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{1} \end{bmatrix} - \begin{bmatrix} \mathbf{I} \\ \mathbf{1}^T \end{bmatrix} \mathbf{R}^{-1}\mathbf{c}_1$$

and then defining the optimization problem,

$$\min \|\mathbf{Eu} - \mathbf{f}\|$$
$$s.t. \quad \mathbf{u} \geq 0$$

where

$$\mathbf{E} = \begin{bmatrix} \mathbf{G}^{\mathrm{T}} \\ \mathbf{g}^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$$

we are left with a non-negative least squares (NNLS) problem, which can be solved using conventional tools.

This method is motivated by the property that it allows the total endmember abundances to sum to less than one. The effect of this relaxation over other unmixing algorithms is that a portion of the spectrum that is not well modeled by the given endmembers may be ignored. This added flexibility can give rise to more robust estimates of endmember abundances.

We only consider wavelengths in the range 232 - 507 cm^{-1} and 825 - 1301 cm^{-1} for the following reasons:

- The data at low wavenumber is generally of low quality.

- A strong CO_2 band dominates in the mid-range.

- Most of the signal is dust and water at high wavenumbers.

From a practical point of view, performing spectral unmixing may be viewed as a type of *dimensionality reduction* where the dimensionality of the data is reduced from n to m dimensions. In this case, n is the 143 TES bands and m is the size of our spectra library. Each end member spectra may be considered a *basis* in the projected linear space. Even though we have no guarantee that the set of spectra basis vectors in the endmember library are orthogonal, they do retain the advantage of interpretability since each basis coefficient, i.e. abundance estimate, is directly linked to the relative amount of endmember expression in a given TES observation.

The unmixing was conducted as two separate investigations: in the first, the eight canonical endmembers were used as inputs to the unmixing algorithm. In the second investigation, these were supplemented with 13 minerals from the ASU Spectral Library. The first investigation revealed the levels of atmospheric and surface dust present at the time of the observations, as well as water ice clouds. Also the first investigation indicated where the emissivity spectrum was well-matched using the Syrtis Type 1 and Acidalia Type 2 endmember spectra; it can be inferred that these areas would be most similar to the generic Mars surface compositions described in earlier works. Those areas not well-matched by the simple canonical unmixing analysis (those with the highest residuals, or the greatest amount of "other" endmember category) can be reasonably expected to exhibit the highest concentrations of minerals which depart from the simple characterizations of basaltic, basaltic/andesitic, type 1, and type 2 terrains.

The second unmixing investigations attempts to use the canonical endmembers along with a suite of minerals in order to better describe the surface mineralogy. We chose a suite of rock-forming minerals which are prevalent among basaltic, andesitic, and granitic rocks. This included albite, anorthite, anorthoclase, augite, biotite, bytownite, fayalite, forsterite, labradorite, microcline, muscovite, oligoclase, and quartz. While some of these are found in both basaltic and granitic rocks, their relative proportions can be used to constrain the rock types.

2.3.3 Stability Analysis

Before proceeding with the analysis of the TES data, the stability of the spectral unmixing algorithm is assessed in the presence of noise and under the assumption that the endmembers that generate the data are present in the library. To test the stability, we

create synthetic spectra based on a set of k end members and then corrupt the spectra with Gaussian noise. We run multiple tests with different noise levels as well as creating spectra with different numbers of end members. Out results are shown in **Figures 7 and 8**. In both plots, the vertical axis is the average reconstruction error and the horizontal axis is the number of end members used to construct the synthetic spectra.

Intuitively, we may expect the amount of error in the spectral unmixing to increase as the amount of noise is introduced. Ideally the expected error will be independent of the number of constitutive components and only depend on the amount of noise. This is the result we observe in **Figure 7**.

The situation for the abundance estimation is more complicated since it may be reasonable to expect that the reconstruction error will increase as the number of end members increases (same constraining conditions, but more variables). This is confirmed by the approximately linear relationship in **Figure 8**, showing an increase in the abundance estimate error that depends on the amount of noise and the number of end members. The combination of results can give confidence that the spectral unmixing algorithm is well behaved under the assumption that the spectral data is a linear combination of spectra end members with additive Gaussian noise.

2.3.4 Effect of Atmospheric Dust on Reconstruction Error

One may hypothesize that as the amount of dust increases, the reconstruction error of the spectra may also increase. **Figure 9** shows a typical plot of the residual error of the spectral unmixing versus abundance estimation. There is no significant correlation between the two, although a causal link cannot be ruled out since the quantities are not independent. Also, it is difficult to use the results from the synthetic

unmixing experiments to place a bound on the amount of residual error since we do not have a way of estimating the amount of noise in the TES spectra.

2.3.5 Separation of Mountain- and Plains-forming Materials Using ML

The goal of any machine learning technique is to automatically construct a model from existing data with the property that the model *generalizes* well. That is, the model should be able to extrapolate and perform well on new data not seen before, as well as the data on which it was trained. The expected generalization power of a model is dependent on the quantity and quality of the data given. A data set of few, low-noise observations should be able to produce a model of equivalent power as a data set composed of many, high-noise observations.

With that goal in mind, we have applied two broad classes of machine learning techniques, unsupervised and supervised classification, to the problem of separating regions of the Thaumasia Highlands based on the TES spectral data. In both type of analysis the goal is to obtain a *label* for each data point that identifies the class to which it belongs. *Unsupervised* clustering takes only the data as the input and produces a classification based on an internal model trained on the data. *Supervised* classification requires that a subset of the data be labeled.

Supervised methods are potentially more powerful than unsupervised because a domain expert can provide better information. However, if the data can be well approximated by an unsupervised method's internal model, then one may expect to see significant parity between the supervised and unsupervised methods.

2.3.6 Unsupervised Classification

We apply unsupervised clustering to the spectral emissivity data under the assumption that there are two dominant geologic types in the Thaumasia highlands plains regions with high basalt content and weathered mountainous regions. The clustering algorithms used (*k*-means, mixtures of Gaussians, and mixtures of *t*-distributions) require that the number of classes, denoted by the variable *k*, be specified *a priori*. Since we do not know how many components are required to accurately model the data, we try all possible numbers of components from 2 to 8.

The key difference between the three unsupervised methods lies in the underlying assumptions on the nature of the data. The *k*-means algorithm assigns each data point to the nearest class prototype. Under a Euclidean metric, this partitions the data space into a *Voronoi* diagram.

The Mixture of Gaussian (MoG) model can be viewed as a probabilistic extension to *k*-means where data points have an affinity to every class proportional to the data point's evaluation under a Gaussian probability density function. The underlying distance metric is the *Mahalanobis* distance metric, which is simply the exponential term of a Gaussian density function.

$$D_M^2(\mathbf{x},\mathbf{y}) = (\mathbf{x} - \mathbf{y})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \mathbf{y})$$

In our model, we restrict the covariance matrix to be diagonal which improves the numerical stability. This covariance matrix, which is learned from the data for each class, allows each dimension of the data space to be independently weighted.

The Mixture of *t*-distributions (MoT) model is a further extension of the MoG model and has an interpretation as a model robust to numerical outliers. The Student-*t* distribution may be interpreted as a Gaussian distribution marginalized over an unknown variance drawn from a Gamma distribution, e.g.

$$t_\nu(\mu, s) = \int_\sigma N(\mu \mid \sigma)\Gamma(\sigma)d\sigma$$

thus the estimates from a MoT model should be more robust to outliers in the data and give better estimates of the spectral prototypes.

2.3.6.1 Two Clusters (k = 2)

Under the hypothesis that there may be two dominant regions (basaltic plains-forming materials and mountains comprised of a diversity of rock types including basement complex), the data is partitioned into exactly two clusters using both a Euclidean and correlation metric. The spectral endmember abundances are estimated independently for each observed pixel and the abundance estimates are used as the input data to the algorithm. The Euclidean metric is sensitive to the absolute magnitude of the abundances, which creates a bias as described in Section 2.3.5.

The results of the clustering using the Mixture of Gaussian clustering algorithm are shown in **Figure 10** and are similar to the results from applying k-means and Mixture of t-distributions.

2.3.6.2 Eight Clusters (k = 8)

While the two-class clustering results show some separability between the plains and mountainous regions, it is likely that the data is not well characterized by only two Gaussian distributions. In order to more accurately model the underlying data distribution, we increase the number of clusters to a reasonable amount based on the estimated number of different types of mineralogy in the region. We hope that this will more clearly delineate the two regions of interest.

Our results from this clustering are shown in **Figure 11** with all the clusters superimposed in the upper-left image and the remaining eight images corresponding the eight distinct classes. The first set of nine images was produced using the Euclidean distance metric and the second set, shown in **Figure 12**, used a correlation metric.

Both clustering methods are able to separate the plains and mountainous regions, but increasing the number of clusters delineates the two regions of interest with greater accuracy due to the more precise characterization of the data.

2.3.7 Supervised Classification

Some supervised classification algorithms (Kohonen, 1989; Vapnik, 1995) were also used for the purpose of distinguishing the two main surface types (mountain- and plains-forming materials) in the TES dataset. For training purposes, a small number of TES data points were labeled to represent either mountain or plain regions, based on the terrain information (obtained from, for example, previous geological mapping efforts on Mars, and the thermal emission imaging system THEMIS), as well as the preliminary results of the unmixing methods discussed above. This is shown in **Figure 13**. Based on these training samples, various feature extraction and supervised classification methods were then applied to classify the entire TES dataset.

2.3.7.1 Feature Extraction

The supervised classification could be carried out in the high dimensional feature space spanned by all n spectral bands of the TES dataset. However, for better computational efficiency, the dimensionality of the feature space could be drastically reduced by some feature extraction techniques, such as those based on the within-class

and between-class and scatter matrices S_w and S_B, which could be obtained once the training samples were available as follows:

$$S_w = \sum_{i=1}^{c} P_i S_i = \sum_{i=1}^{c} P_i \Sigma_i$$

$$S_B = \sum_{i=1}^{c} P_i (M_i - M)(M_i - M)^T$$

which add up to the total scatter matrix (covariance matrix):

$$S_T = \frac{1}{N} \sum_{X} (X - M)(X - M)^T = \frac{1}{N} \sum_{i=1}^{c} \sum_{X \in w_i} (X - M)(X - M)^T = S_W + S_B$$

Here M_i is the mean vector of the training samples for class i and M_0 is the mean vector of all training samples. The purpose here is to find a subspace of dimensionality $m < n$ with maximum between-class scatter matrix S_B, so that the classes could be best separated with greatly reduced computational cost. The m-dimensional subspace could be obtained by directly choosing m out of the original n TES spectral bands, or, alternatively, generated as a linear combination of all n bands based on the principal component analysis (PCA). Specifically, we can find the m eigenvectors corresponding to the largest m eigenvalues of S_B to form an m by n matrix that transforms the original n dimensional space into a new m dimensional space without losing any separability.

This dimensionality reduction technique is similar to using the end member spectra as a set of basis functions, but, in this case, an optimal set of basis functions are learned from the data which may lead to improved generalization of the trained model, but the interpretability of the result may suffer as the spectral eigenvector cannot be directly related to the spectra of known minerals.

2.3.7.2 Bayesian and Minimum-Distance Classifications

In the *m* dimensional space, all data points in the TES dataset would be classified into either of the two classes of mountain and plain. To do so different algorithms could be used, such as the classical Bayesian method that classifies a given data sample to a class with maximum likelihood, of which the MoG and MoT are examples. Alternatively, some distance-based methods could also be used that classify a given sample to a class with minimum distance, such as the Mahalanobis distance based on the covariance matrix of the training samples as well as their mean vectors, or simply the Euclidean distance between the sample and the mean vectors of the training samples. Moreover, to emphasize the similarities between the spectral profiles of the samples while de-emphasizing their absolute values, the spectral angle mapping (SAM) distance could be used.

2.3.7.3 Classification Result

The various algorithms discussed above based on likelihood or distance measures have all consistently generated similar classification results. One of the classification results is shown in **Figure 14**. This result was obtained by PCA-based feature extraction and Mahalanobis distance discussed above. It is clearly seen that the entire Thaumasia Highlands region covered by the TES data is classified into the plain region (green), and the mountains (blue). In particular, the volcano inside the white circle to the right of the center of the image is clearly distinguished from the surrounding plain area.

2.3.8 Comparison of Unsupervised and Supervised Results

Since the machine learning techniques used employ different internal models, we cannot expect duplicate results *a priori*. The fact that the results from all the

23

unsupervised algorithms applied to both the raw TES emissivity spectra and the endmember abundance data are consistent with each other, coupled with the strong similarity to the supervised classification results, indicates that the mountains and plains regions of the Thaumasia Highlands are strongly separable. In the absence of ground truth data, we cannot precisely quantify the accuracy and robustness of the models, however we can be confident in their qualitative interpretation.

2.3.9 Data Quality

When analyzing data from the TES instrument, it is advisable that uncertainties and biases in the data be quantified as accurately as is feasible. During our initial investigations, a strong bias in the spectral data manifested itself in our clustering and classification results such that adjacent and overlapping observations were placed into opposing classes with high confidence. This unintuitive result prompted further investigation into the quality of the observed data.

The bias correlates strongly with the estimated surface brightness, which corresponds to the time of observations. It is hypothesized that the variability in the martian atmosphere and incident solar radiation between day and night observations is sufficient to bias the spectral data such that a naïve application of Machine Leaning techniques may lead to unsupported conclusions. Since our application is seeking to separate large spatial regions, the exact bias of a particular observation is not critical so long as the bias is consistent across all observations.

3. Results

Application of the linear unmixing and machine learning algorithms led to similar results which together indicate qualitative and quantitative differences between the compositional types within the plains- and mountain-forming materials. The similarity of these results lends confidence to the assessment that the machine learning algorithm can be used to exploit subtle spectral differences and permit rapid classification of hyperspectral data from spacecraft.

3.1 Linear Unmixing of Surface Mineralogy

The results of the first linear unmixing using only canonical endmembers indicated that atmospheric components, while present throughout the dataset, dominated only in small areas. Atmospheric dust accounted for as much as 40 to 80% of the emissivity measured at the sensor in some places. Water ice clouds were quite sparse but reached levels as high as 29-40% of the spectral signal. The Syrtis Type 1 dominated the plains units, while the Acidalia Type 2 was found in both mountains and plains, with stronger abundances in the plains.

Similar results were found in the second unmixing, which combined the canonical endmembers with the minerals, although exact proportions varied considerably. The ranges for atmospheric endmembers, on the one hand, tended to be quite similar, which lends confidence to the results of both algorithms. The results for the second unmixing are shown in **Figure 15**. Proportions of atmospheric endmembers were quite similar for both investigations, but surface abundances varied significantly. The low-CO_2 level dust endmember was found to be more abundant than the high-CO2 level dust endmember in both simulations, while water ice clouds were found in only a few places. These stand out best in the high-latitude water ice cloud image in **Figure 15**. Acidalia, Syrtis, and Surface Dust type endmembers were comparatively lacking in this

simulation, largely because the algorithm was free to select other minerals. In many cases, the ability to alter the proportions of component minerals leads to much better spectral matches. This is indicated by the low levels of residual, and by the fact that the "other" class was hardly invoked at all.

It is notable that both simulations invoked a large contribution from the "Unity" endmember of the mountainous provinces. While meant to correct for blackbody radiation contributions, particularly in areas of high shading or porosity (Bandfield et al., 2000) the "Unity" endmember also tends to be selected in unmixing algorithms to account for higher emissivity levels than can be matched with available input minerals. This suggests that the Thaumasia highlands may contain minerals which are not included in our initial selection of endmember minerals.

The unmixing results invoked higher levels of augite, (a clinopyroxene) and both fayalite and forsterite (olivines) in the plains than in the highlands. This is consistent with a more basaltic composition in the plains-forming materials. Albite and labradorite, both major components of andesitic to granitic rocks, were surprisingly sparse in both mountains and plains. Like fayalite, microcline, bytownite, muscovite, microcline, oligoclase and quartz were found in both plains and mountains, but with higher levels in the plains. Since these are found in both basaltic and granitic rock assemblages, this is consistent with either composition. Detailed examination of endmember proportions for both regions may help shed further light on the type of rock making up the mountain- and plains-forming materials.

3.2 Quantification of Spectral and Geologic Information.

To associate both the Machine Learning "untrained" classified TES type 1 and 2 information and "trained" classified mountain- and plains-forming TES spectral

information with the Thaumasia ancient mountain terrain and the younger surrounding plains, we co-located all the datasets into a Geographic Information System (GIS) for analysis. The Thaumasia Geologic map, USGS publication I-2605 (Dohm 2000) was used to select mountain- and plains-forming pixel types at prime locales, as well as to segregate geologic map units into two types for comparative analysis with the Machine Learning-based results: (1) ancient mountain-forming map units (mostly Stage 1, Early to Late Noachian), and (2) younger map units (Stage 2-4, mostly Late Noachian to Early Hesperian). Areas calculations for the two types of geologic units were computed using an Equal-area Sinusoidal projection. The TES-based information of **Figures 16 and 17** were then read into our GIS as simple point locations. Because the region for this TES-based investigation was smaller in extent than the extent of the total area of the I-2605 geologic map, the geology was clipped to match the study region (matching the the limits of the TES data). Lastly, we then used a simple process of intersecting the TES points with the two geologic unit types to create the density calculations (**Figures 16 and 17 and corresponding Tables 1 and 2**).

The density results clearly indicate that the mountain-forming materials are distinct from the plains-forming materials with more dominant Type 2 and mountain-type for the ancient mountain-forming materials of the Thaumasia highlands mountain range and Type 1 and plains-forming type for the plains-forming materials, respectively. This is consistent with Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b), which indicated that the mountains could be comprised of diverse rock materials such as basement complex and the plains-forming materials mostly basaltic lava flows. Importantly, those pixels that do not show such a correspondence could be explained by multiple factors, including noisy pixels (e.g., atmospheric conditions and dust loading) and geology (e.g., materials shed from the prominent

mountains may form alluvial fan materials comprised of both mountain-forming materials and plains-forming materials, lateral and vertical variations in the mantles that may obscure bedrock, and younger volcanoes and lava flows that occur in the mountain range; see Dohm et al., 2001b).

4. Discussion.

We have developed an intelligent software system (referred to here as Machine Learning) for robust analysis of hyper-spectral Thermal Emission Spectrometer (TES) data (e.g., Christensen et al., 2000, 2001a,b; Bandfield et al., 2000) with the impetus to determine whether the rock materials of the ancient Thaumasia highlands comprise rock materials other than just basalt/basaltic andesite as the published TES maps have portrayed (Christensen et al. 1999, 2001a; Bandfeld et al., 2000). This is in part based on ground-truthing experiences on Earth such as was previously reported for the structurally controlled basin of the Atacama Desert where a significant part of the geologic record could not be identified from satellite image data because of mantling and secondary weathering processes. Through Machine Learning Systems we can perform expeditious comparative analysis among any type of multispectral image information. Using Geographic Information Systems, we can also readily separate out ancient mountain-forming materials from relatively young plains-forming materials based on published geologic information and couple the geologic information with the Machine Learning-based results to determine the density of pixels of both end member types in both the mountain-forming and plains-forming rock materials.

The results clearly indicate that there is a spectral distinction between the two types of materials. But what does the distinction mean? Clearly from the results at least we can say that the Thaumasia highlands mountain range comprises a greater diversity

of rocks than just basalt-basaltic andesite. But can we properly identify the rock types that compose the mountain-forming materials with existing orbital platforms? For example, will CRISM be able to accurately identify ancient rock materials that may be poking up through Tharsis-era mantles?

Determining the composition of the Thaumasia highlands mountain range is of first order importance since the range records an ancient part of Mars' evolution, which includes possible Earth-like evolutional phases (Baker et al., 2002) such as plate tectonism (Sleep, 1994; Márquez et al., 2004), particularly during its embryonic stages of evolution (Dohm et al., 2002; Fairén et al., 2002; Fairén and Dohm, 2004). In addition to the ancient Thaumasia highlands mountain range, there are other markers of an ancient Mars that are difficult to explain from a process other than plate tectonism, including the Coprates rise mountain range (Scott and Tanaka, 1986; Schultz and Tanaka, 1994; Dohm et al., 2001a,b) and other tens to thousands-km-long structures such as Tempe Mareotis Fossae and Phlegra Montes (**Figure 1**; also see Dohm et al., 2002). Many of the features (a) are interpreted to be the result of compressional deformation such as thrust faulting (e.g., Schultz and Tanaka, 1994; Dohm et al., 2001a), (b) occur among highly degraded promontories (interpreted to be silicate-rich constructs or intrusives; e.g., Scott and Tanaka, 1986; Hodges and Moore, 1994; Dohm et al., 2001a), (c) are embayed by relatively younger rock materials (e.g., Scott and Tanaka, 1986, Greeley and Guest, 1987; Tanaka and Scott, 1987), and (d) form the margins of elongated basins and "banded" magnetic anomalies (Connerney, 1999) (such as those in the Terra Cimmeria and Terra Sirenum regions), similar to what is observed in geologic terrains of Earth that have recorded plate tectonism (e.g., for further details on why the banded anomalies in Terra Cimmeria and Arabia Terra may be explained by plate tectonism, see Fairén et al., 2002 and Connerney et al., 2005, respectively). Other

seemingly anomalous observations in regard to the geological evolution of Mars include a martian crust, which shows major variations from thin beneath the northern plains (~30 km) to thick (~60 km) beneath the southern highlands and Tharsis (Zuber et al., 2000), and linear crustal magnetization anomalies of remarkable intensity that occur in the southern highlands (Acuña et al., 1999, 2001; Connerney et al., 1999, 2005; Arkani-Hamed, 2003).

Mars is commonly viewed as a one plate planet since its incipient development based largely on geophysical and geochemical arguments mainly from analysis of SNC meteorites (e.g., Halliday et al., 2001; Nimmo and Tanaka, 2005). Other investigations such as those largely based on Gamma Ray Spectrometer (GRS) information are shedding further light; the GRS instrument on the Mars Odyssey (MO) spacecraft can record elemental abundances for such elements as hydrogen (H), chlorine (Cl), silicon (Si), potassium (K), thorium (Th), and iron (Fe) in rock materials up to 1/3 m depth (Boynton et al., 2002, 2004, JGR-Planets—in progress; Taylor et al., JGR-Planets— accepted). For example, detailed analysis of the GRS-based K, Th, and Fe concentrations on the martian surface reveal significant results (Taylor et al., JGR Planets—accepted). For example, detailed analysis of the GRS-based K, Th, and Fe concentrations on the martian surface reveal significant results (Taylor et al., JGR Planets—accepted), including: (1) bulk Mars is enriched in moderately volatile elements compared to Earth, but has a much lower K/Th ratio than CI chondrites, (2) Mars is enriched in FeO compared to Earth (suggesting that terrestrial planets formed from relatively narrow accretion zones), (3) surface concentrations of K and Th measured by GRS are systematically higher than in SNC meteorites (thus the meteorites are not representative of martian surface rocks), and (4) GRS data for K and Th do not seem consistent with widespread recycling of the crust, long-acting plate tectonics, or

extensive formation of a tertiary crust on Mars. Yet another GRS-based detailed investigation of the TES-based "surface type 1" and "surface type 2" materials also show why the GRS data appears to be inconsistent with plate tectonics involving subducting slabs rich in fluids and hydrously altered basaltic rocks as they are on Earth, based largely on the spatial distribution of K and Th (Suniti Karunatillake, JGR Planets—in press).

The distinction in TES spectral signature among the mountain- and plains-forming materials is interestingly consistent with the Gamma Ray Spectrometer data, which shows that the ancient mountain forming materials of the Thaumasia highlands and Coprates rise mountain ranges are elevated in Si when compared to the younger Tharsis volcanic materials (Stages 4-5; see Dohm et al., 2001b, 2005) (**Figure 6**). In addition, Mars Odyssey, Mars Express, and the Mars Exploration Rovers are increasingly showing a greater diversity of rock types for Mars such as hematite, andesite, sulfates, layered sedimentary deposits, , and even quartz-bearing granitoids, etc. (e.g., Malin and Edgett, 2000; Bandfield et al., 2000; Christensen et al., 2001a,b, 2004, 2005; Rieder et al., 2004; Gendrin et al., 2005). Do the distinct TES and GRS signatures, the ancient relative age of the martian mountain ranges, which is based on stratigraphic and crosscutting relations among materials and structure (Dohm et al., 2001a,b; Anderson et al., 2001), impact crater statistics (Scott and Tanaka, 1986; Tanaka, 1986; Neukum et al., 2001), and magnetic data (e.g., Acuna et al., 1999), and a more mineralogically diverse Mars based on MGS, MO, and Mars Express results, collectively point to a diverse ancient rock record in the Thaumasia highlands mountain range and other ancient features on Mars? And if so, what does this mean?

Importantly, similar to the potential problem of obscuration of an ancient rock record through the mantling and secondary weathering processes from a satellite-based

perspective, is it possible that the SNC—associated sampling of Mars is only revealing a partial sampling of the geological and geochemical histories of Mars, as indicated by the GRS-based investigation of Taylor et al. (JGR-Planets—in press)? Furthermore, if there is a limited percentage of an ancient rock record exposed at or near the surface (GRS is the first spectrometer that can sample below a < 1 cm-thin skin of materials) is the GRS instrument capable of identifying such a limited exposure of ancient rock record at resolution? If so, this rock record may have contributed to the formation of flow materials such as alluvial fans--outcrops of ancient rocks being shed off the range that contribute to alluvial fan materials. In simple terms, is there an ancient rock record other than basalt-basaltic andesite, sulfates, and hematite such as basement complex that we only have hints of through existing spaceborne platforms, or does it simply not exist? With resolutions as high as 25 centimeters per pixel, CRISM should shed further light on the potential for such an ancient rock record.

5. Summary

The results of our investigation coupled with recent MGS-, MO-, and Mars Express-based findings, which indicate a greater diversity of rock types through recent data acquisitions, are promising, and designing an effective system to perform expeditious analysis of huge data sets such as TES and THEMIS will ultimately lead to an improved understanding into the geological evolution of Mars. The results presented here indicate that the ancient mountain-forming materials record a distinct TES signature from the plains-forming materials.   This is consistent with Viking-era, geological mapping-based interpretations that the mountain-forming materials could be comprised of a diversity of rock types, which includes basement complex, whereas the plains-forming materials are mostly volcanic. Such a potential rock record and possible

implications on Mars' embryonic stages of planetary evolution underscore the necessity for further investigation through MGS, MO, Mars Reconnaissance Orbiter (MRO), and future science-driven reconnaissance missions.

REFERENCES.

Acuña, M.H., Connerney, J.E.P., Ness, N.F., Lin, R.P., Mitchell, D., Carlson, C.W., Mcfadden, J., Anderson, K.A., Reme, H., Mazelle, C., Vignes, D., Wasilewski, P., Cloutier, P., 1999. Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. Science 284, 790-793.

Acuña, M.H., Connerney, J.E.P., Wasilewski, P., Lin, R.P., Mitchell, D., Anderson, K.A., Carlson, C.W., McFadden, J., Réme, H., Mazelle, C., Vignes, D., Bauer, S.J., Cloutier, P., Ness, N.F., 2001. Magnetic field of Mars: Summary of results from the aerobraking and mapping orbits. J. Geophys. Res. 106, 23,403-23,417.

Anderson, R. C., Dohm, J. M., Golombek, M. P., Haldemann, A., Franklin, B. J., Tanaka, K. L., Lias, J., Peer, B., 2001. Significant centers of tectonic activity through time for the western hemisphere of Mars. J. Geophys. Res. 106, 20,563-20,585.

Anderson, R.C., Dohm, J.M., Haldemann, A.F.C., Hare, T.M., Baker, V.R., 2004. Tectonic histories between Alba Patera and Syria Planum, Mars, Mars. Icarus 171, 31-38.

Arkani-Hamed, J. (2003), Thermoremanent magnetization of the martian lithosphere, J. Geophys. Res., 108, 10.1029/2003JE002049.

Baker, V.R., Strom, R.G., Gulick, V.C., Kargel, J.S., Komatsu, G., and Kale, V.S., 1991. Ancient oceans, ice sheets and the hydrological cycle on Mars, *Nature*, 352, 589-594.

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 (http://lin.springer.de/service/journals/10069/free/conferen/superplu/).

Bandfield, J.L., P.R. Christensen, and M.D. Smith, Spectral data set factor analysis and end-member recovery: Application to analysis of Martian atmospheric particulates, J. Geophys. Res., 105(E4) 2000a.

Bandfield, J.L., V.E. Hamilton, and P.R. Christensen, A global view of Martian surface compositions from MGS-TES, Science 287, 1626, 2000b.

Bandfield, J.L., and M.D. Smith, Multiple emission angle surface-atmosphere separations of Thermal Emission Spectrometer data, Icarus 161, 47-65, 2003

Bilmes, J., 1997, A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, ICSI-TR-97-021, University of Berkeley

Bishop, C., 1995, Neural Networks and Pattern Recognition, Oxford University Press.

Boardman, J. W., Kruse, F. A., and Green, R. O., 1995, Mapping target signatures via partial unmixing of AVIRIS data: in Summaries, Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1, v. 1, p. 23-26.

Boynton, W. V., Feldman, W. C., Squyres, S. W., Prettyman, T. H., Brückner, J., Evans, L. G., Reedy, R. C., Starr, R., Arnold, J. R., Drake, D. M., Englert, P. A. J., Metzger, A. E., Mitrofanov, I., Trombka, J. I., d'Uston, C., Wänke, H., Gasnault, O., Hamara, D. K., Janes, D. M., Marcialis, R. L., Maurice, S., Mikheeva, I., Taylor, G. J., Tokar, R., Shinohara, C., 2002. Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. Science 297, 81-85.

Boynton, W.V., Feldman, W.C., Mitrofanov, I.G., Evans, L.G., Reedy, R.C., Squyres, S.W., Starr, R., Trombka, J.I., d'Uston, C., Arnold, J.R., Englert, P.A.J., Metzger, A.E., Wänke, H., Brückner, J., Drake, D.M., Shinohara, C., Fellows, C., Hamara, D.K., Harshman, K., Kerry, K., Turner, C., Ward, M., Barthe, H., Fuller, K.R., Storms, S.A., Thornton, G.W., Longmire, J.L., Litvak, M.L., and Ton' Chev, A.K., 2004. The Mars Odyssey Gamma-Ray Spectrometer Instrument Suite. Space Science Reviews 110, 37-83.

Chong Diaz, G., Mendoza, M., Garcia-Veigas, J., Pueyo, J.J., and Turner, P., 1999. Evolution and geochemical signatures in a Neogene forearc evaporitic basin: the Salar Grande (Central Andes of Chile). Paleography, Paleoclimatology and Paleoecolgy 151, 39-54.

Christensen, P.R., Bandfield, J.L., Smith, M.D., Hamilton, V.E., Clark, R.N., 2000. Identification of a basaltic component on the martian surface from Thermal Emission Spectrometer data. J. Geophys. Res. 105, 9609-9621.

Christensen, P.R., Bandfield, J.L., Hamilton, V.E., Ruff, S.W., Kieffer, H.H., Titus, T.N., Malin, M.C., Morris, R.V., Lane, M.D., Clark, R.L., Jakosky, B.M., Mellon, M.T., Pearl, J.C., Conrath, B.J., Smith, M.D., Clancy, R.T., Kuzmin, R.O. Roush, T., Mehall, G.L., Gorelick, N., Bender, K., Murray, K., Dason, S., Greene, E., Silverman, S., Greenfield, M., 2001a. The Mars Global Surveyor Thermal Emission Spectrometer experiment: Investigation description and surface science results. J. Geophys. Res. 106, 23,823-23,871.

Christensen, P.R., Morris, R.V., Lane, M.D., Banfield, J.L., Malin, M.C., 2001b. Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. J. Geophys. Res. 106, 23,873-23,885.

Christensen, P. R. McSween, H.Y. Jr., Bandfield, J.L., Ruff, S.W., Rogers, A.D., Hamilton, V.E., Gorelick, N., , Wyatt, M.B., Jakosky, B.M., Kieffer, H.H., Malin, M.C., Moersch, J.E., 2005. The Igneous Diversity of Mars: Evidence for Magmatic Evolution Analogous to Earth. Lunar Planet. Sci. Conf., XXXVI, #1273 (abstract) [CD-ROM], 2005.

Christensen, P.R., J.L. Bandfield, V.E. Hamilton, D.A. Howard, M.D. Lane, J.L. Piatek, S.W. Ruff, and W.L. Stefanov, A thermal emission spectral library of rock-forming minerals, J. Geophys. Res., 105,9735-9739, 2000.

Clifford, S.M. and Parker, T.J., 2001. The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. Icarus 154, 40–79.

Connerney, J.E.P., Acuña, M.H., Wasilewski, P.J., Kletetschka, G., Ness, N.F., Rème, H., Lin, R.P., Mitchell D.L., 1999. The global magnetic field of Mars and implications for crustal evolution. Science 284, 790-793.

Connerney, J. E. P., M. H. Acuña, N. F. Ness, G. Kletetschka, D. L. Mitchell, R. P. Lin, and H. Reme (2005), Tectonic implications of Mars crustal magnetism, *Science*, *102*, 14970–14975.

Dalton, J.B., Sutter, B., Kramer, M.G., Stockstill, K.R., Moersch, J., and Moore, J.M., Search for evaporite minerals in Flaugergues Basin, Mars, Lunar Planet. Sci. XXXV, 1869, 2004.

Dempster, A., Laird, N., Rubin, D., 1977, Maximum likelihood from incomplete data via the EM algorithm J. R. Stat. Soc. Ser. B, 39, 1–38

Dohm, J.M., Tanaka, K.L., 1999. Geology of the Thaumasia region, Mars: plateau development, valley origins, and magmatic evolution. Planetary and Space Science 47, 411-431.

Dohm, J. M., Ferris, J. C., Baker, V. R., Anderson, R. C., Hare, T. M., Strom, R. G., Barlow, N. G., Tanaka, K. L., Klemaszewski, J. E., Scott, D. H., 2001a. Ancient drainage basin of the Tharsis region, Mars: Potential source for outflow channel systems and putative oceans or paleolakes. J. Geophys. Res. 106, 32,943-32,958.

Dohm, J.M., Tanaka, K.L., Hare, T.M., 2001b. Geologic map of the Thaumasia region of Mars. US Geol. Survey Map I-2650.

Dohm, J. M. et al., 2001c. Latent outflow activity for western Tharsis, Mars: Significant flood record exposed. J. Geophys. Res. Planets, 106, 12301-12314.

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. Lunar Planet. Sci. XXXIII, 1639 (abstract).

Dohm, J.M., K. Kerry, J.M. Keller, V.R. Baker, S. Maruyama, and R.C. Anderson, J.C. Ferris, and T.M. Hare, 2005. Mars geological province designations for the interpretation of GRS data. Lunar Planet. Sci. Conf., XXXVI, #1567 (abstract) [CD-ROM].

Fairén, A.G. Dohm, J.M., 2004. Age and origin of the lowlands of Mars. Icarus 168, 277-284.

Fairén, A. 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.

Fairén, A.G., Dohm, J.M., Baker, V.R., de Pablo, M.A., Ruiz, J., Ferris, J.C., and Anderson, R.C., 2003. Episodic flood inundations of the northern plains of Mars. Icarus 165, 53-67, 2003.

Fairén, A.G., Fernández-Remolar, D., Dohm, J.M., Baker, V.R., and Amils, R., 2004. Inhibition of carbonate synthesis in acidic oceans on early Mars. Nature, 431, 423-426.

Fink W., Dohm J.M., Tarbell M.A., Hare T.M., Baker V.R., 2005. Next-Generation Robotic Planetary Reconnaissance Missions: A Paradigm Shift. Planetary and Space Science 53, 1419-1426.

Frey, H., 1979. Thaumasia: A Fossilized Early Forming Tharsis Uplift. J. Geophys. Res. 84, 1009-1023.

Green, A. A., Berman, M., Switzer, P, and Craig, M. D., 1988, A transformation for ordering multispectral data in terms of image quality with implications for noise

removal: IEEE Transactions on Geoscience and Remote Sensing, v. 26, no. 1, p. 65-74.

Halliday A.N., Wanke H., Birck J-L ., Clayton, R.N., 2001. The accretion, composition andearly differentiation of Mars. Space Sci. Rev. 96:197–230

Hodges, C.A., and Moore, H.J., 1994. Atlas of volcanic landforms on Mars.: U.S. Geological Survey Professional Paper 1534.

Kohonen, T, 1989, Self-Organization and Associative Memory, Springer-Verlag, Berlin, 3rd Edition.

MacKay, D., 1998, Introduction to gaussian processes," in Neural Networks and Machine Learning, ser. NATO Advanced Study Institute, C. M. Bishop, Ed. Berlin: Springer, vol. 168, pp. 133-165.

MacQueen, J. B., 1967, Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297

Márquez, A., Fernández, C., Anguita, F., Farelo, A., Anguita, J., de la Casa, M-A., 2004. New evidence for a volcanically, tectonically, and climatically active Mars, Icarus 172, 573-581.

Nimmo, F., Tanaka, K.L., 2005. Early crustal evolution of Mars. Ann. Rev. Earth Planet. Sci. 33133-161.

Peel, D. and McLachlan, G.J., 2000, Robust mixture modeling using the t distribution, Statistics and Computing, Vol. 10 (4), 339-348.

Quinonero-Candela, J., 2004, Learning with Uncertainty - Gaussian Processes and Relevance Vector Machines. PhD Thesis, Technical University of Denmark, ISSN 0909-3192.

Ruff, S. W., 2003, Basaltic Andesite or Weathered Basalt: A New Assessment, paper presented at Sixth International Conference on Mars, Lunar and Planetary Institute, Pasadena, California, July 20-25 2003

Ruff, S.W. and P.R. Christensen, Bright and dark regions on Mars: Particle size and mineralogical characteristics based on Thermal Emission Spectrometer data, J. Geophys. Res. 107(E12), 5127, doi:10.1029/2001JE001580, 2002.

Schulze-Makuch, D., Dohm, J.M., Fairén, A.G., Baker, V.R., Fink, W., Strom, R.G., 2005. Venus, Mars, and the ices on Mercury and the Moon: astrobiological implications and proposed mission designs. Astrobiology 5, 778-795.

Scott, D.H., Tanaka, K.L., 1986. Geologic map of the western equatorial region of Mars. USGS Misc. Inv. Ser. Map I-1802-A (1:15,000,000).

Scott, D.H., Dohm, J.M., Rice, J.W.Jr., 1995. Map of Mars showing channels and possible paleolake basins. USGS Misc. Inv. Ser. Map I-2461 (1:30,000,000).

Sleep, N. H., 1994. Martian plate tectonics. J. Geophys. Res. 99, 5639-5655.

Smith, M.D., J.L. Bandfield, and P.R. Christensen, Separation of atmospheric and surface spectral features in Mars Global Surveyor Thermal Emission Spectrometer (TES) spectra, J. Geophys. Res., 104 (E4), 9589-9608, 2000

Squyres, S. W. et al., 2004. The Spirit Rover's Athena Science Investigation at Gusev Crater, Mars, *Science*, *305*, 794-799.

Squyres, S.W., Grotzinger, J.P., Arvidson, R.E., Bell III, J.F., Calvin, W., Christensen, P.R., Clark, B.C., Crisp, J.A., Farrand, W.H., Herkenhoff, K.E., Johnson, J.R., Klingelhöfer, G., Knoll, A.H., McLenna, S.M., McSween, Jr., H.Y., Morris, R.V., Rice Jr., J.W., Rieder, R., Soderblom, L.A., 2004. In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars, Science, 306, 1709.

Stockstill, K.R., J.E. Moersch, S.W. Ruff, A. Baldridge and J. Farmer, TES Hyperspectral Analyses of Proposed Paleolake Basins on Mars: No Evidence for In-Place Carbonates, J. Geophys. Res. Planets, 110, doi:10.1029/2004JE002353, 2005.

Tanaka, K.L., Skinner, J.A., Hare, T.M., Joyal, T., Wenker, A., 2003. Resurfacing history of the northern plains of Mars based on geologic mapping of Mars Global Surveyor data, J. Geophys. Res. Planets, 108, doi: 10.1029/2002JE001908.

Tanaka, K.L., Skinner, J.A., Hare, T.M., 2005. Geologic map of the northern plains of Mars, USGS Misc. Inv. Ser. Map 2888 (1:15,000,000 scale).

Vapnik, V., 1995, The Nature of Statistical Learning Theory, Springer.

Vélez-Reyes, M. and Rosario, S., Proceedings IEEE International Geosciences and Remote Sensing Symposium, Alaska, 2004, "Solving Adundance Estimation in Hyperspectral Unmixing as a Least Distance Problem.

Wänke, H., Bruckner, J., Dreibus, G., Rieder, R., and Ryabchikov, I., 2001. Chemical composition of rocks and soils at the Pathfinder site, Space Science Reviews, 96, 317-330.

Wyatt, M. B., and McSween, H. Y., 2002, Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars, Nature, *417*, 263-266

Zuber, M.T., Solomon, S.C., Phillips, R.J., Smith, D.E., Tyler, G.L., Aharonson, O., Balmino, G., Banerdt, W.B., Head, J.W., Johnson, C.L., Lemoine, F.G., McGovern, P.J., Neumann, G.A., Rowlands, D.D., Zhong, S. 2000. Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. Science 287, 1788-1793.

TABLES.

Table 1. Corresponding with **Figure 15**, the TES data was classified by applying Machine Learning to Type-1 and Type-2 end member information of Bandfield et al. (2002) by selecting pixels of quality TES strips that occur in the ancient mountain-building materials and younger plains-forming materials. The Geology has been reclassified into two units: Ancient (mostly Early and Middle Noachian—Stage 1 map units) and Young (Late Noachian and younger—Stages 2-4 map units) based on Dohm et al. (2001b). The density results clearly indicate that the mountain-forming materials are distinct from the plains-forming materials with more dominant Type-2 and Type-1 signatures, respectively. This is consistent with Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b).

Geologic Type	Total Area km^2	Total TES-Type1 Pixels	Total TES-Type 2 Pixels	Density-TES Type 1	Density-TES Type 2
Ancient	631718.9552	7095	19017	0.01123	0.03010
Young	819600.7087	23165	10862	0.02826	0.01325

Table 2. Corresponding with **Figure 16**, the TES data was classified through Machine Learning by selecting pixels of quality TES strips that occur in the ancient mountain-building materials and younger plains-forming materials. The Geology has been reclassified into two units: Ancient (mostly Early and Middle Noachian—Stage 1 map units) and Young (Late Noachian and younger—Stages 2-4 map units) based on Dohm et al. (2001b). The density results clearly indicate that the mountain-forming materials are distinct from the plains-forming materials with more dominant mountain-type and plains-forming-type signatures, respectively. This is consistent with Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b), interpreted to be largely basement complex and basaltic lava flows, respectively, based on Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b)

Geologic Type	Total Area km^2	Total plains-forming type pixels	Total moutain-forming type pixels	Density- total plains-forming type pixels	Density- total moutain-forming type pixels
Ancient	631718.9552	2505	22537	0.00397	0.03568
Young	819600.7087	16260	15886	0.01984	0.01938

FIGURES.
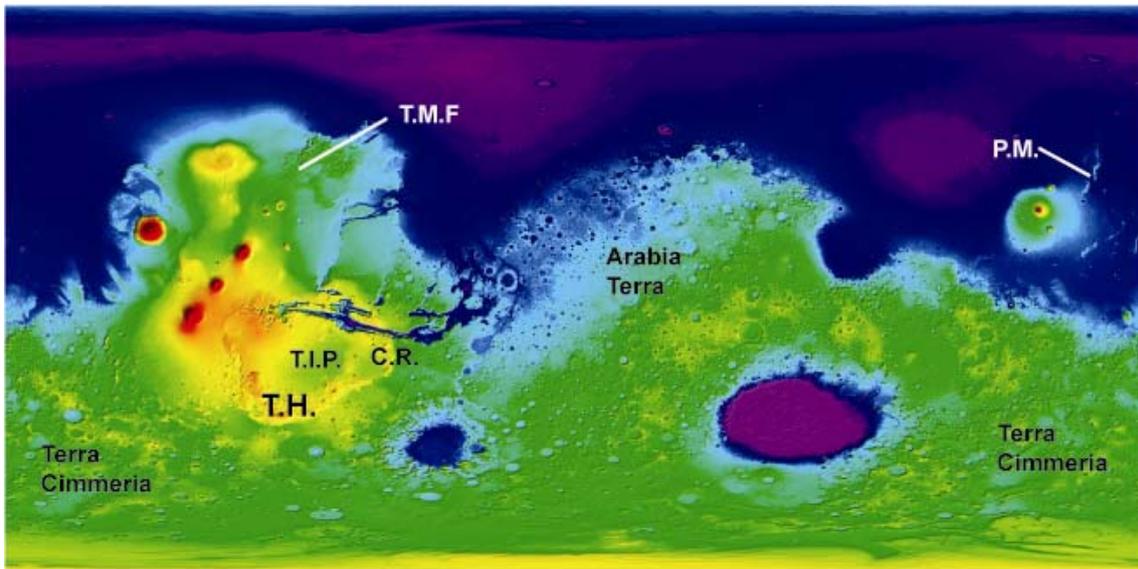
Fig. 1. MOLA-based map showing ancient features such as Thaumasia highlands (T.H.) and Coprates rise (C.R.) mountain ranges, Phlegra Montes, Tempe Mareotis Fossae (T.M.F), Thaumasia igneous plateau (T.I.P) and ancient geologic provinces such as Terra Cimmeria and Arabia Terra, all of which are considered as ancient markers of the early evolution of Mars. The "T" denoting Thaumasia highlands is approximately located at the central part of the region under investigation.
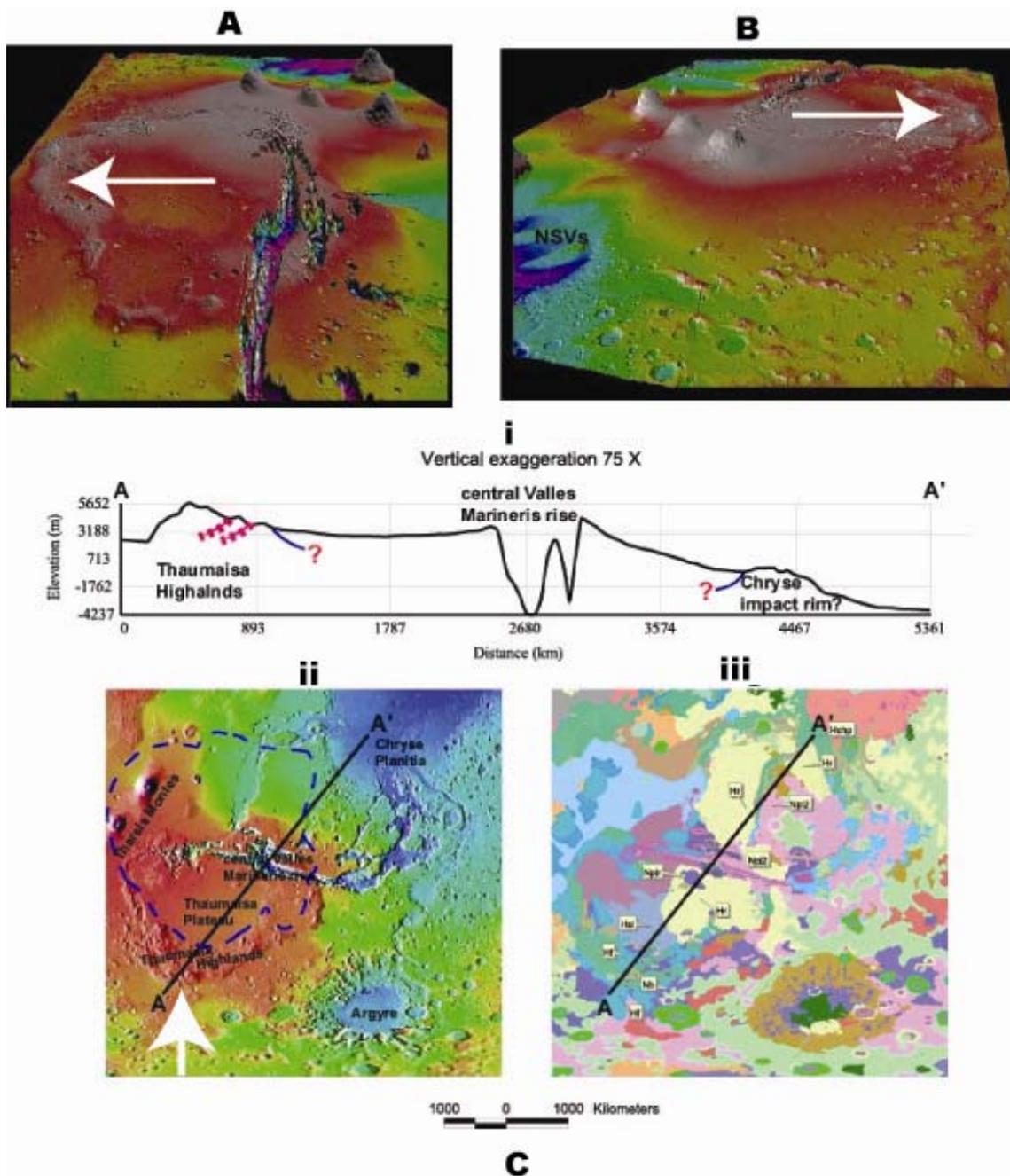
Fig. 2. A. MOLA-based 3D projection looking to the west obliquely across the Thaumasia highlands mountain range (white arrows). **B.** MOLA-based 3D projection looking to the east obliquely across the Thaumasia highlands mountain range (white arrows). **C.** Based on Dohm et al. (2001b), (**i**) present-day MOLA profile (Transect A-A') across the west-central part of Thaumasia highlands mountain range (Dohm and Tanaka, 1999; Dohm et al., 2001a), central part of a putative Noachian drainage basin (queried blue line represents uncertain basin extent) (Dohm et al., 2001b), including west-central Valles Marineris rise (center of tectonic activity, interpreted to be the result of magmatic-driven uplift (Dohm et al., 1998, 2001a-c), and Tempe Terra igneous plateau (Frey, 1979), (**ii**) MOLA shaded relief map showing features of interest, including the approximated boundary of the Noachian drainage basin (dashed blue line) and the central Valles Marineris rise, and (**iii**) part of the geologic map of the western equatorial region of Mars (representative map units are shown—[Scott and Tanaka, 1986]).

47
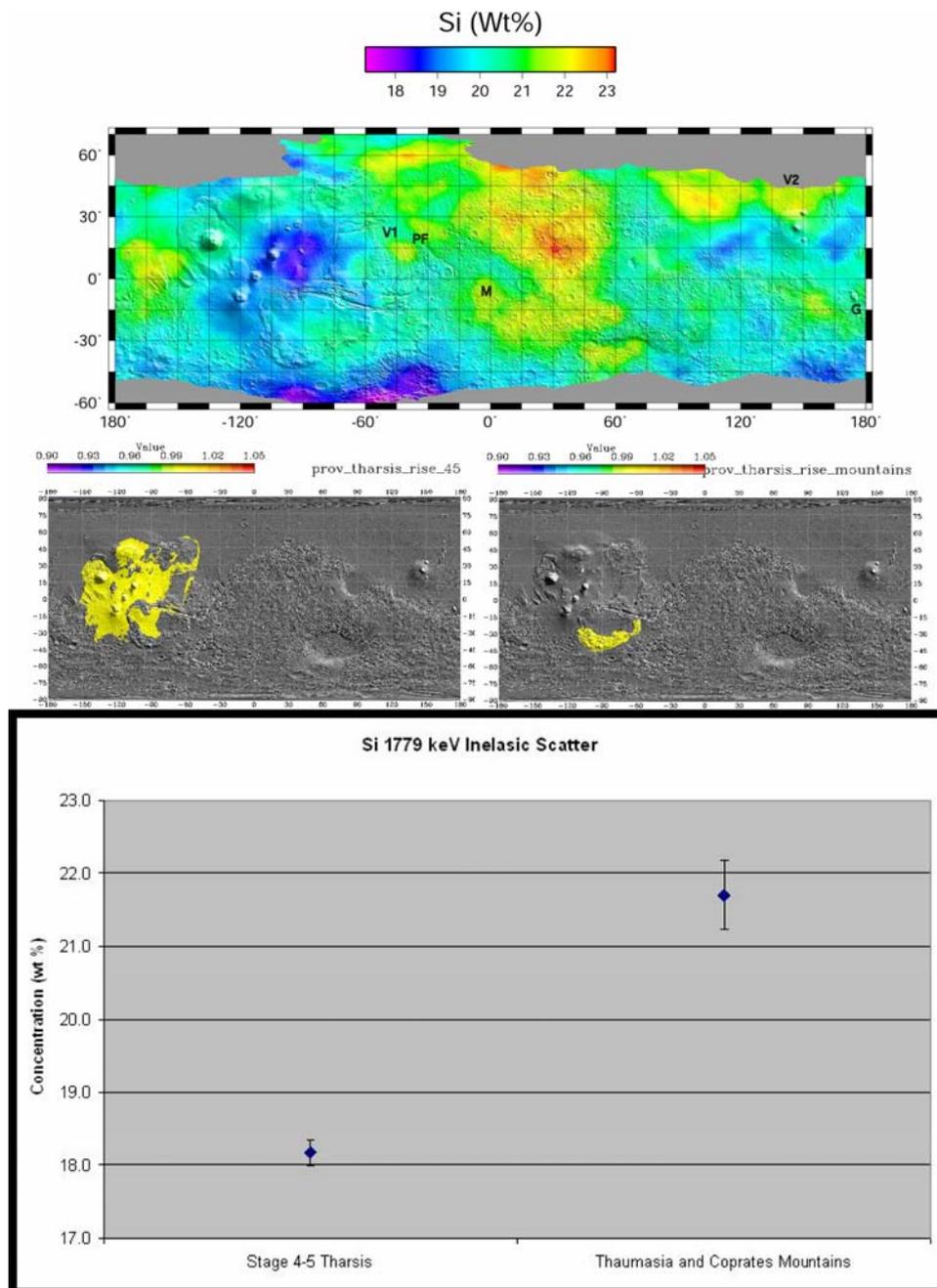
Fig. 3. (top) Preliminary unpublished GRS-based silicon map (e.g., Boynton et al., 2004), (middle) the two geologic provinces used in the comparison (younger, Stage 4-5 Tharsis vs. Thaumasia highlands and Coprates rise mountain ranges; based on Dohm et al., 1001a,b, 2005), and (bottom) a scatter plot showing concentrations of Si 1779 keV for both geologic provinces (note that this information will be updated with the latest GRS map information).
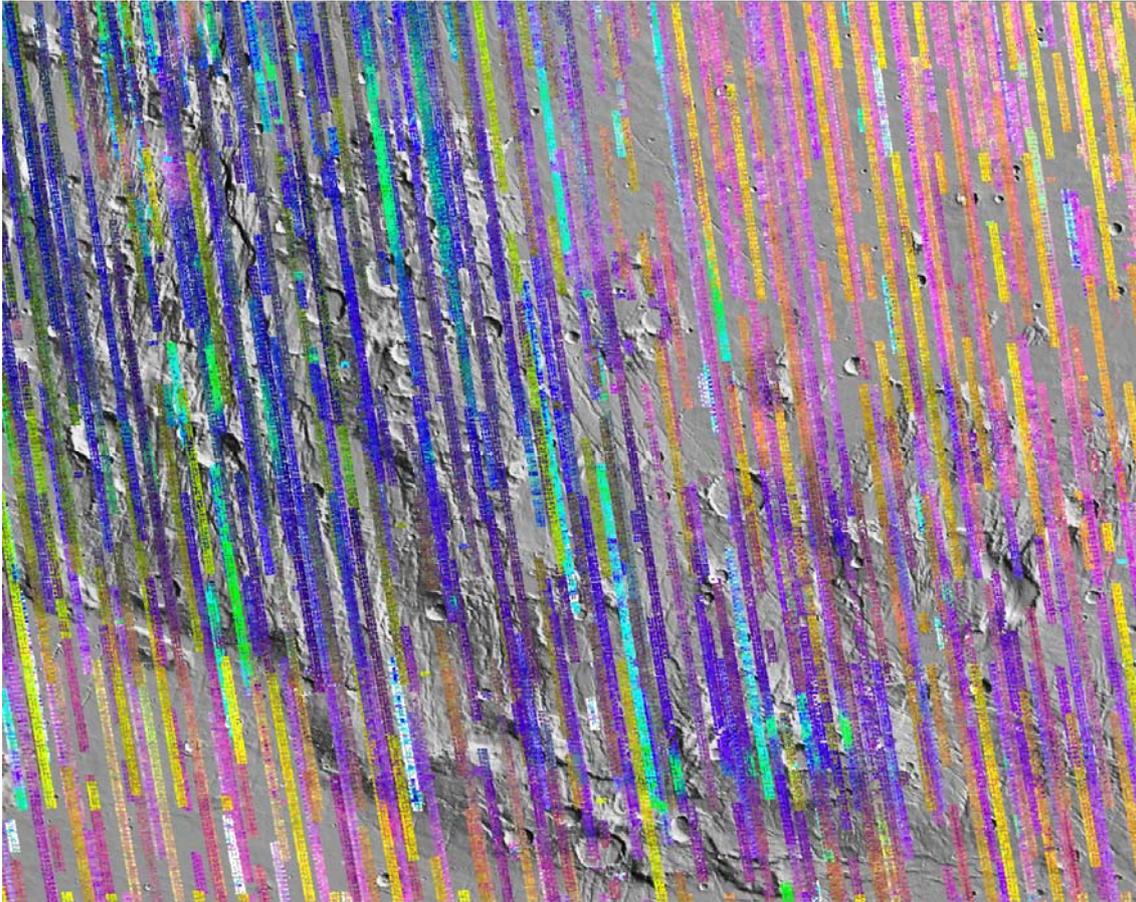
Fig 4. Minimum Noise Fraction (MNF)-transformed TES emissivity observations registered to map base. Color variations represent information content of individual orbits. Orbital tracks whose color variations stand out from the rest typically exhibit greater levels of atmospheric and surface dust loading, or influences from water ice clouds. Adjacent tracks with color variations that correlate with surface features (but not necessarily the same colors) tend to contain more information about surface composition.
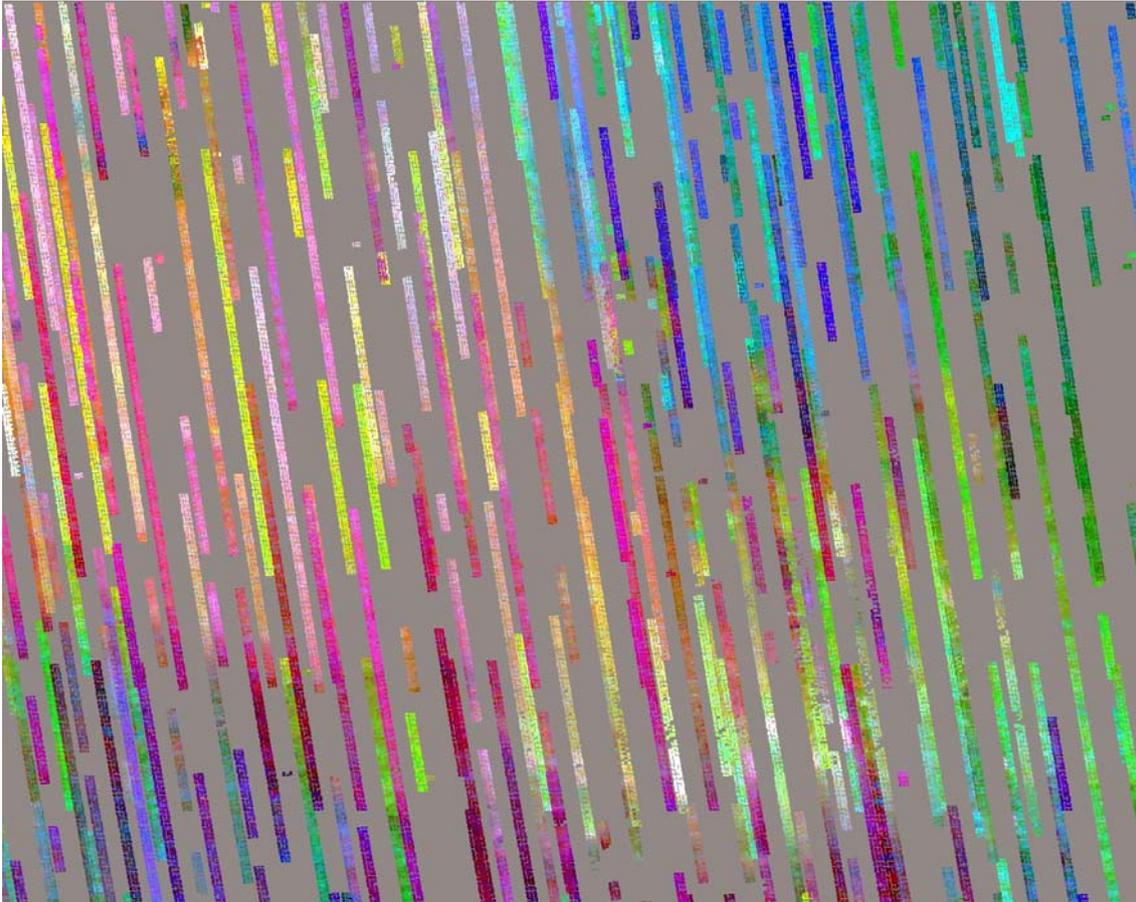
Fig 5. MNF-transformed TES emissivity hypercube after removal of orbital tracks most strongly compromised by atmospheric dust and water, ice clouds, surface dust, instrument artifacts, and low signal-to-noise levels. Color variations of these remaining tracks are more closely correlated to surface composition than in the original data set, though some atmospheric influence remains.
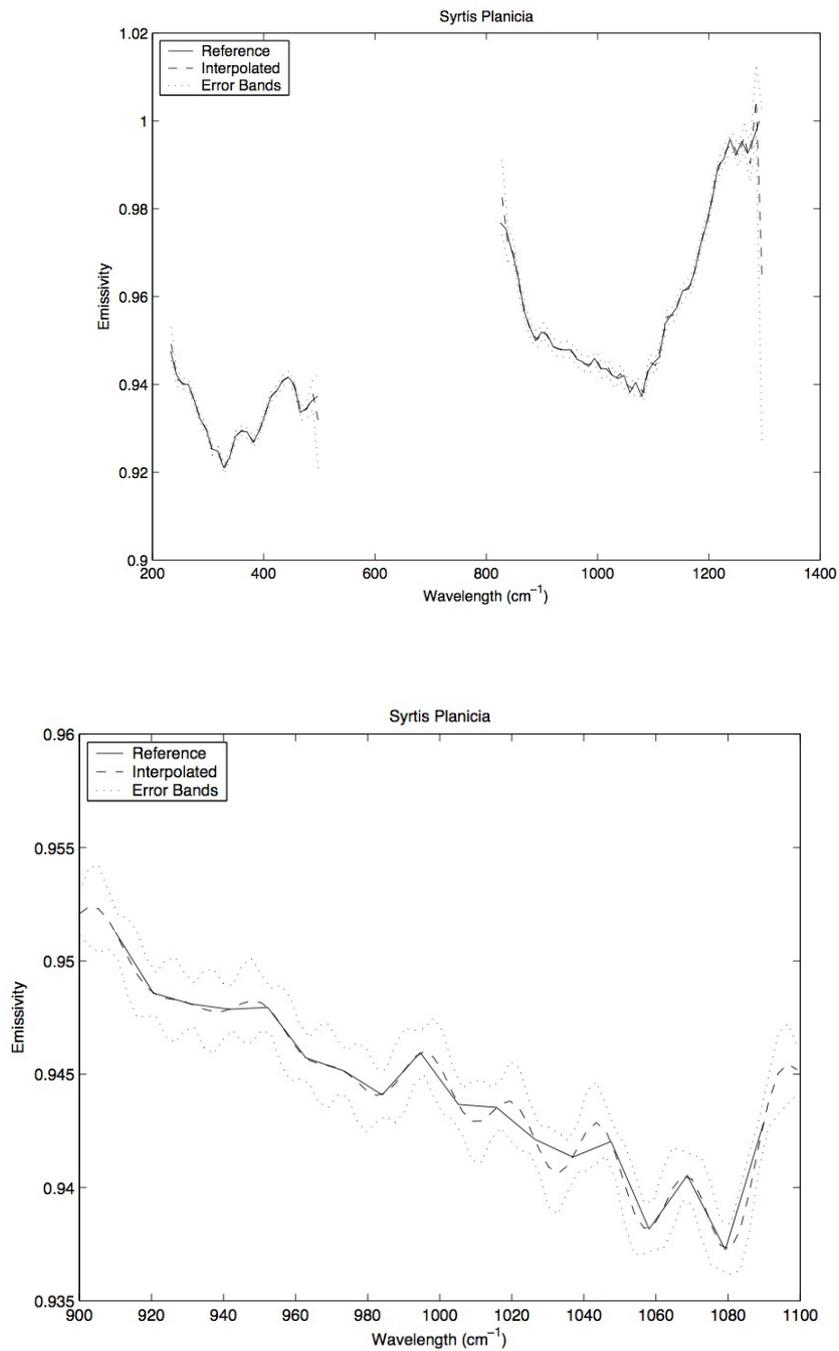
Fig 6. Example of the interpolation of spectral library endmembers in order to align with the TES observation wavelengths. The top figure shows the regression over the entire spectrum of interest. The bottom figure is a close up of the region about 1um. The solid line is the given endmember, the dashed line is the Gaussian Process regression mean, and the dotted line represents the 1-sigma confidence bounds.
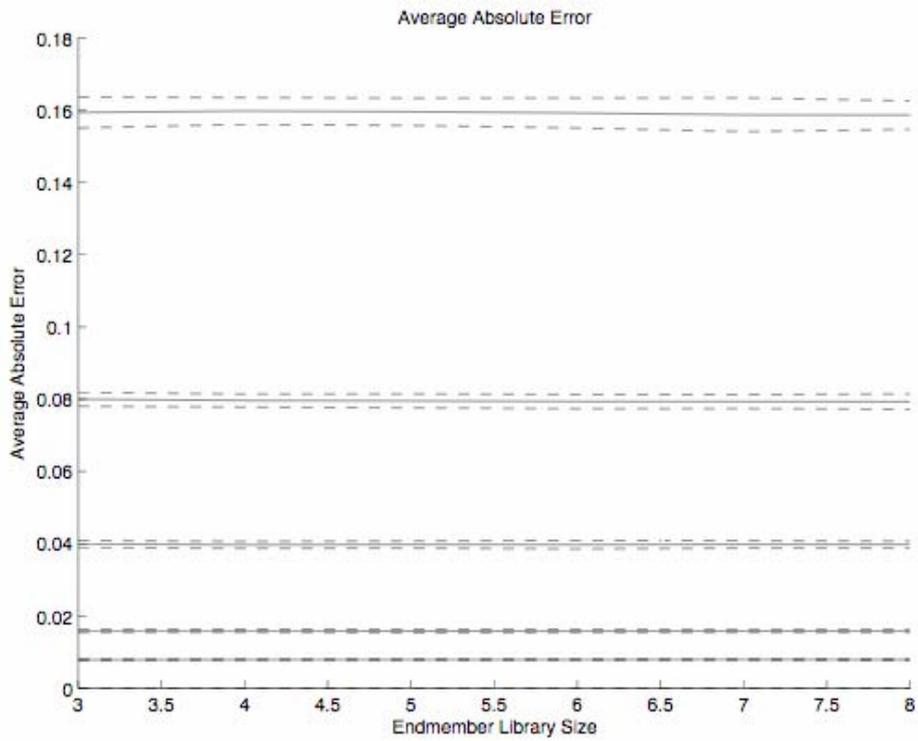
Fig 7. Plot of the average absolute error of the spectral reconstruction. Each horizontal line represents a different amount of Gaussian noise introduced to a synthetic spectra composed of *k* end members. The spectral unmixing algorithm is robust as the amount of error is proportional to the amount of introduced noise.
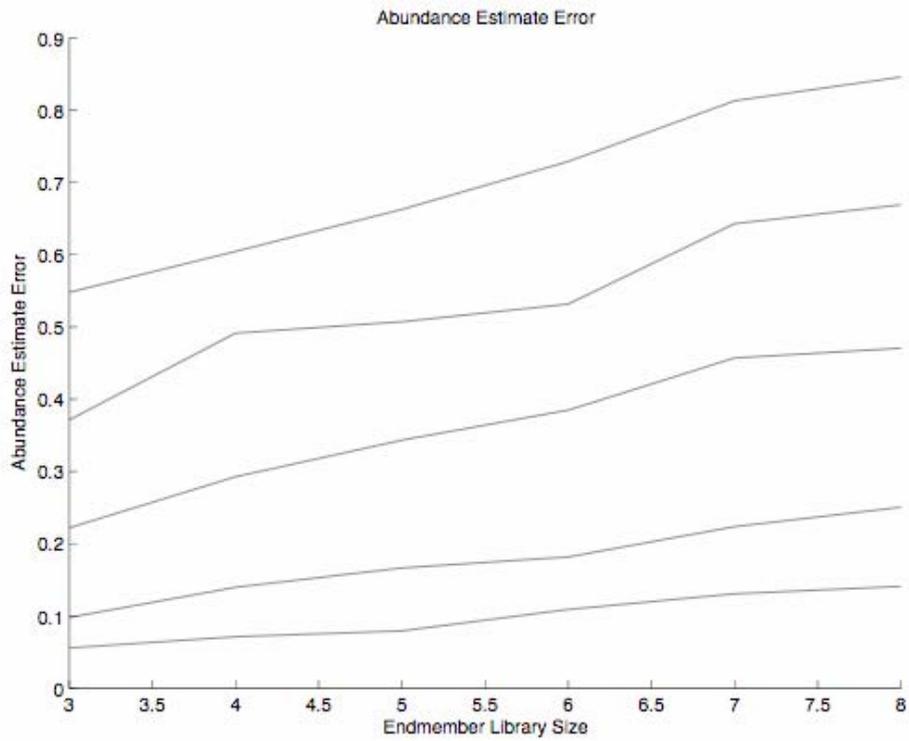
Fig 8. Plot of the abundance estimate error for different levels of Gaussian noise versus the number of endmembers used to create the synthetic spectra. The abundance estimation error increases with the amount of noise and as the number of end members increases.
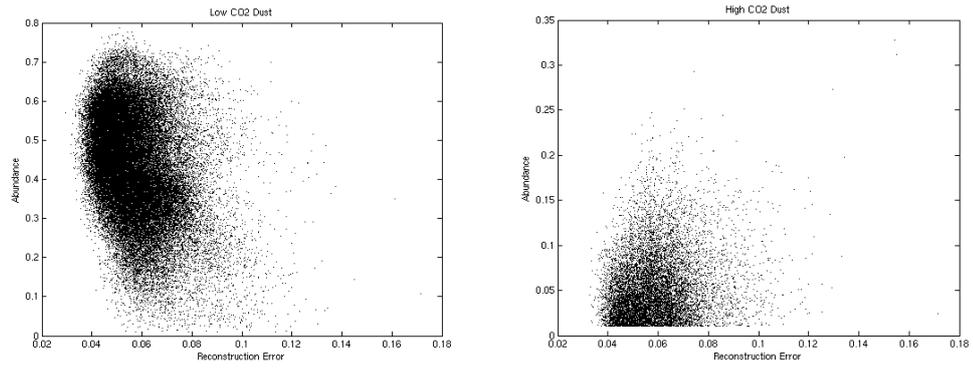
Fig 9. A typical plot of reconstruction error versus abundance estimation. The figures show that there is not a significant correlation between the amount of CO_2 and the reconstruction error of the spectra by the linear unmixing algorithm.
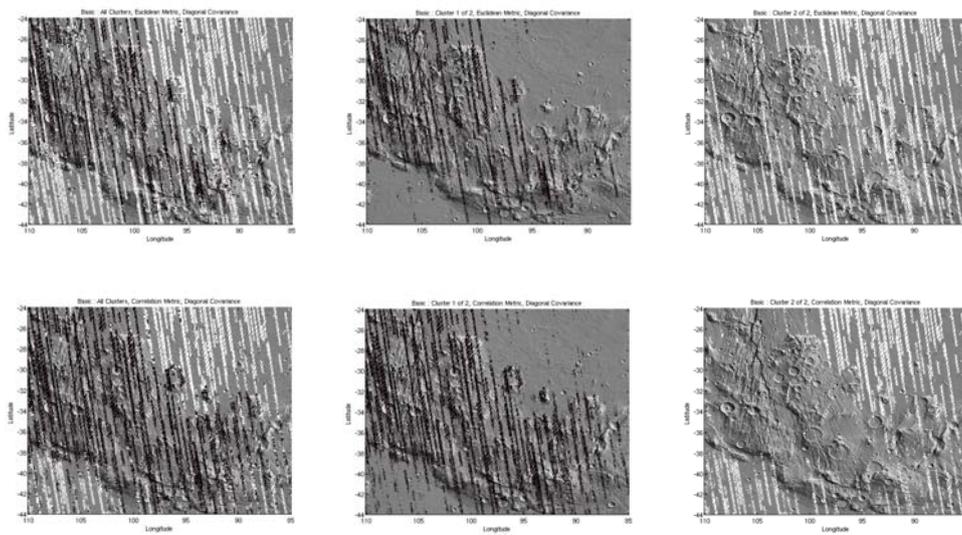
Fig 10. A series of plots showing the separability of the region using only two clusters. The top row shows the results using a Euclidean distance metric and the bottom row uses a Correlation metric. Both methods are able to separate out the plains regions (middle column), but the correlation metric shows a much better classification of the mountainous regions (lower-right).
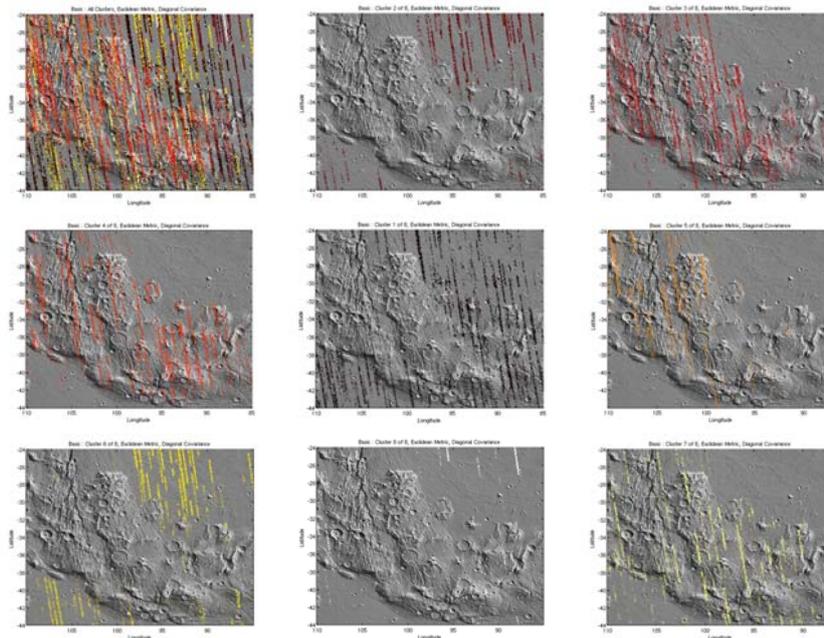
Fig 11. A series of plots showing the classification of the region based on eight underlying clusters. While the majority of the clusters show a strong selectivity towards either the plain-forming or mountainous regions, by including more clusters, indeterminate components (top-right, bottom-right) are identified and may be discarded.
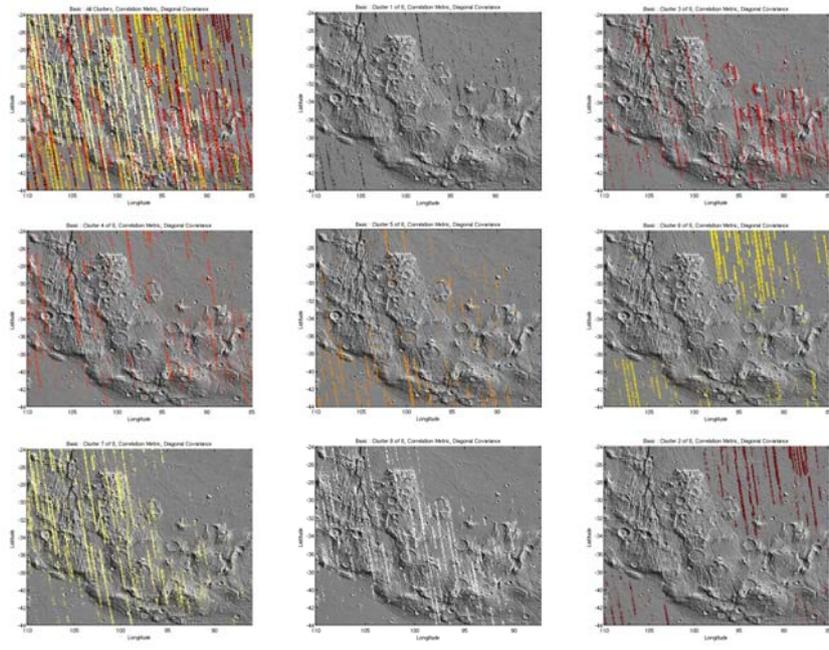
Fig 12. As in **Figure 7**, except using a correlation metric. As before, the overall quality of the clustering is improved, but indeterminate clusters are still identified (center, middle-left).
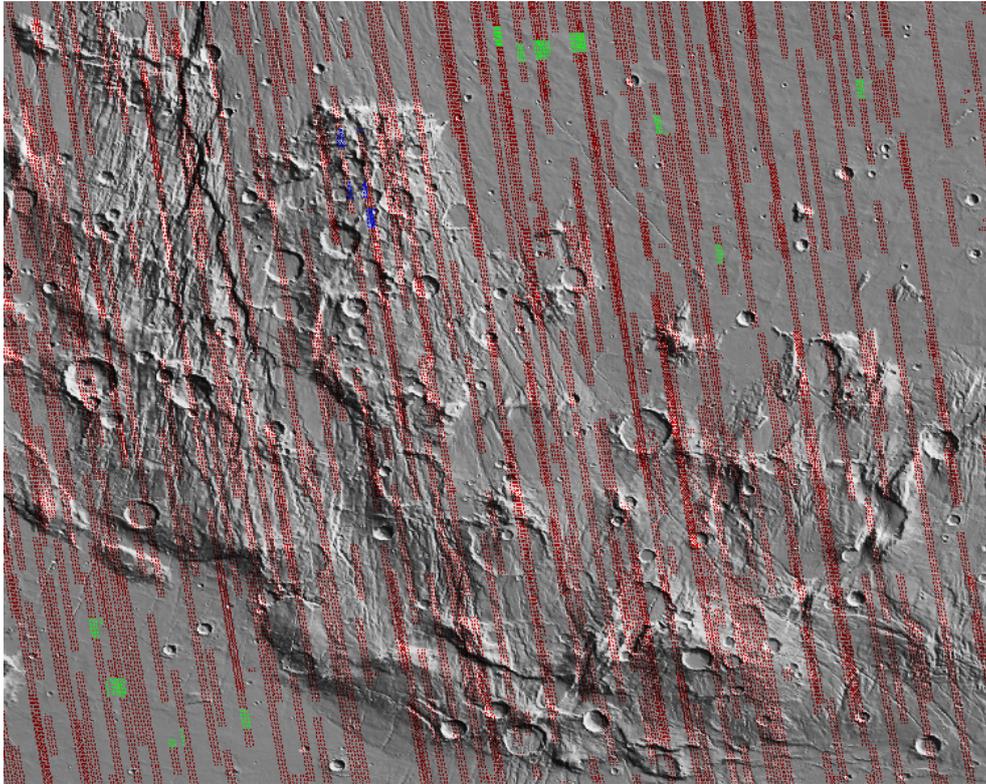
Fig 13. The samples colored in green were chosen to represent the plain (upper middle and lower left), while the samples colored in blue were chosen to represent the mountains (top-left of the center).
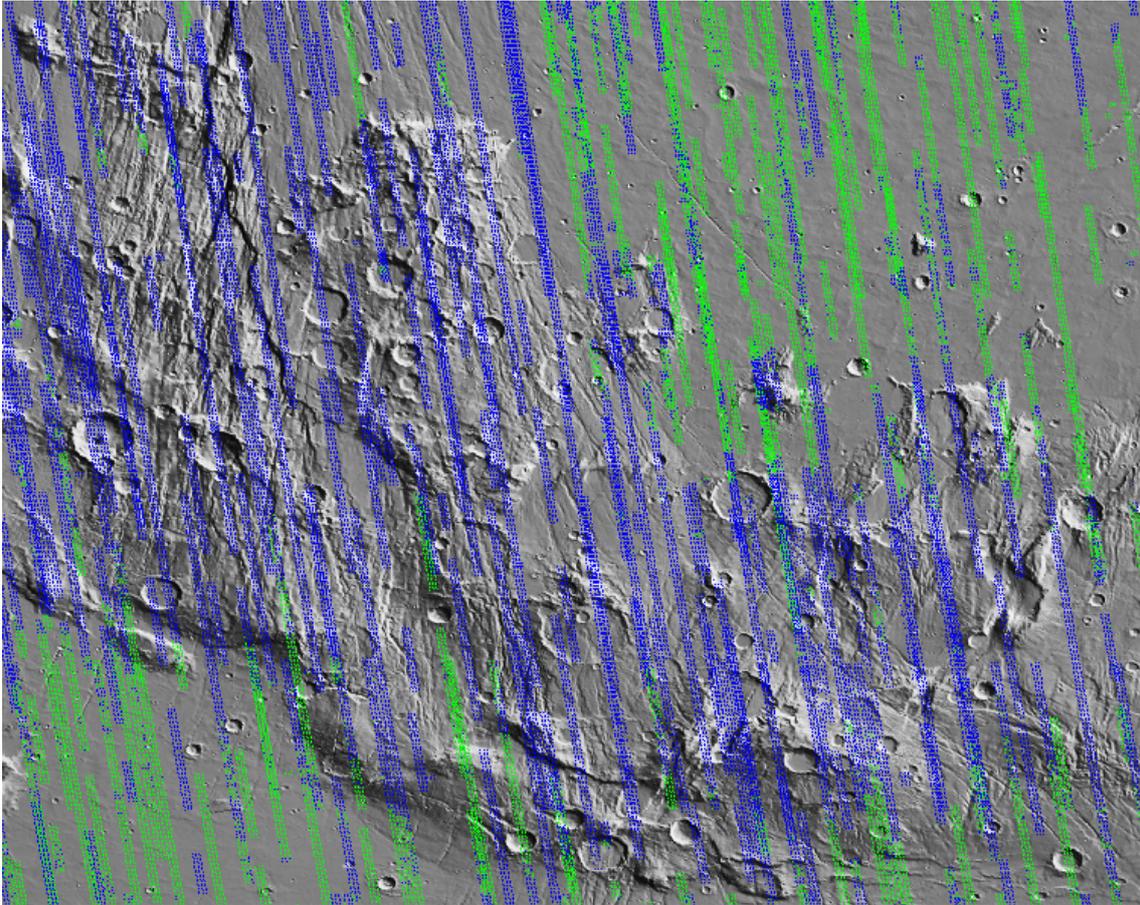
Fig 14. The region under investigation in part of the Thaumasia highlands mountain range and surrounding plains is classified into plains- (green) and mountain- (blue) forming type regions using a supervised classifier and the training data from **Figure 9**.
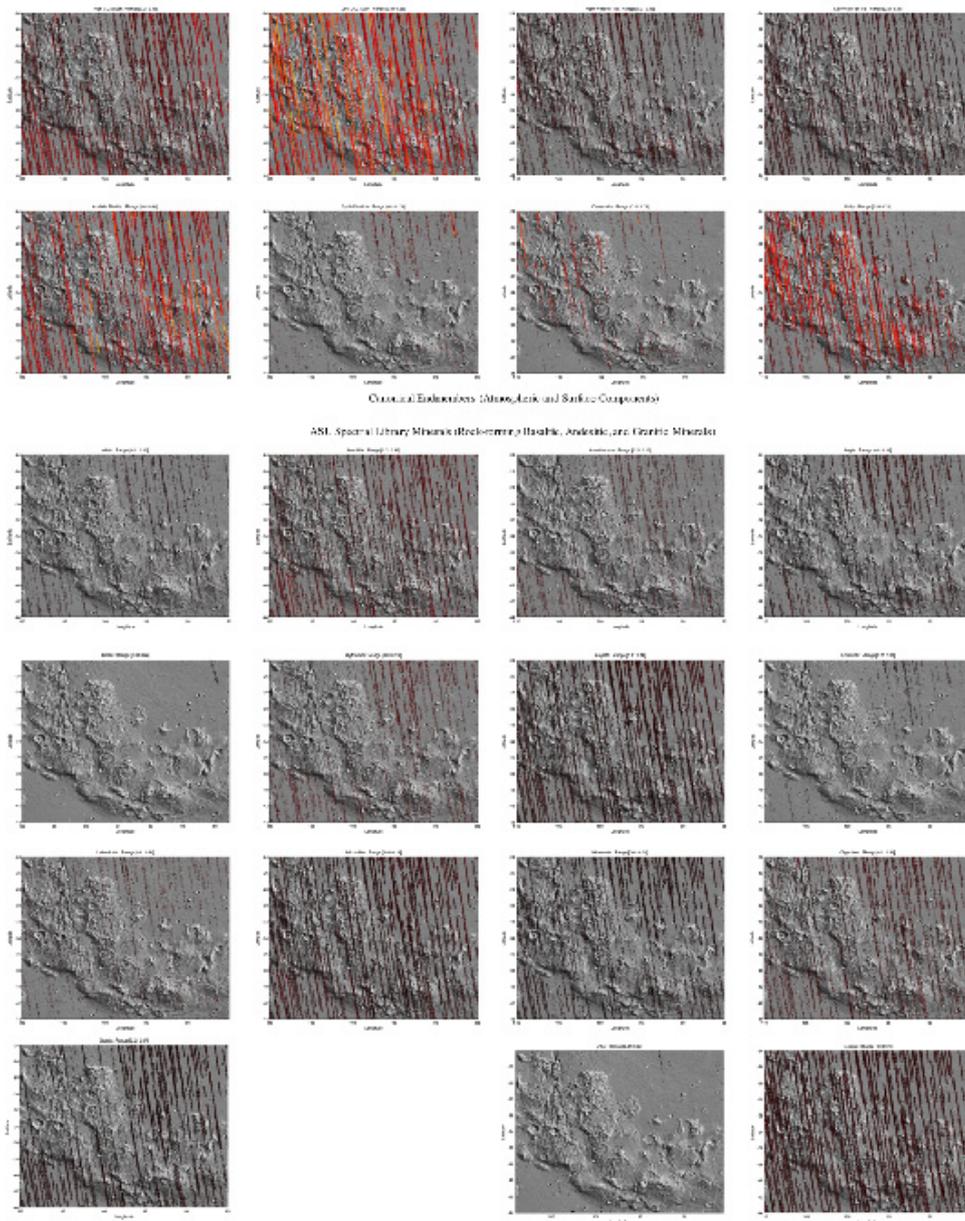
Fig. 15. Results of linear spectral unmixing using 8 canonical endmembers plus 13 spectra of rock-forming minerals from the ASU Spectral Library. The canonical endmembers are shown at top. TES pixels in each image are colored to denote the contribution of that particular endmember to the apparent emissivity measured at the sensor.
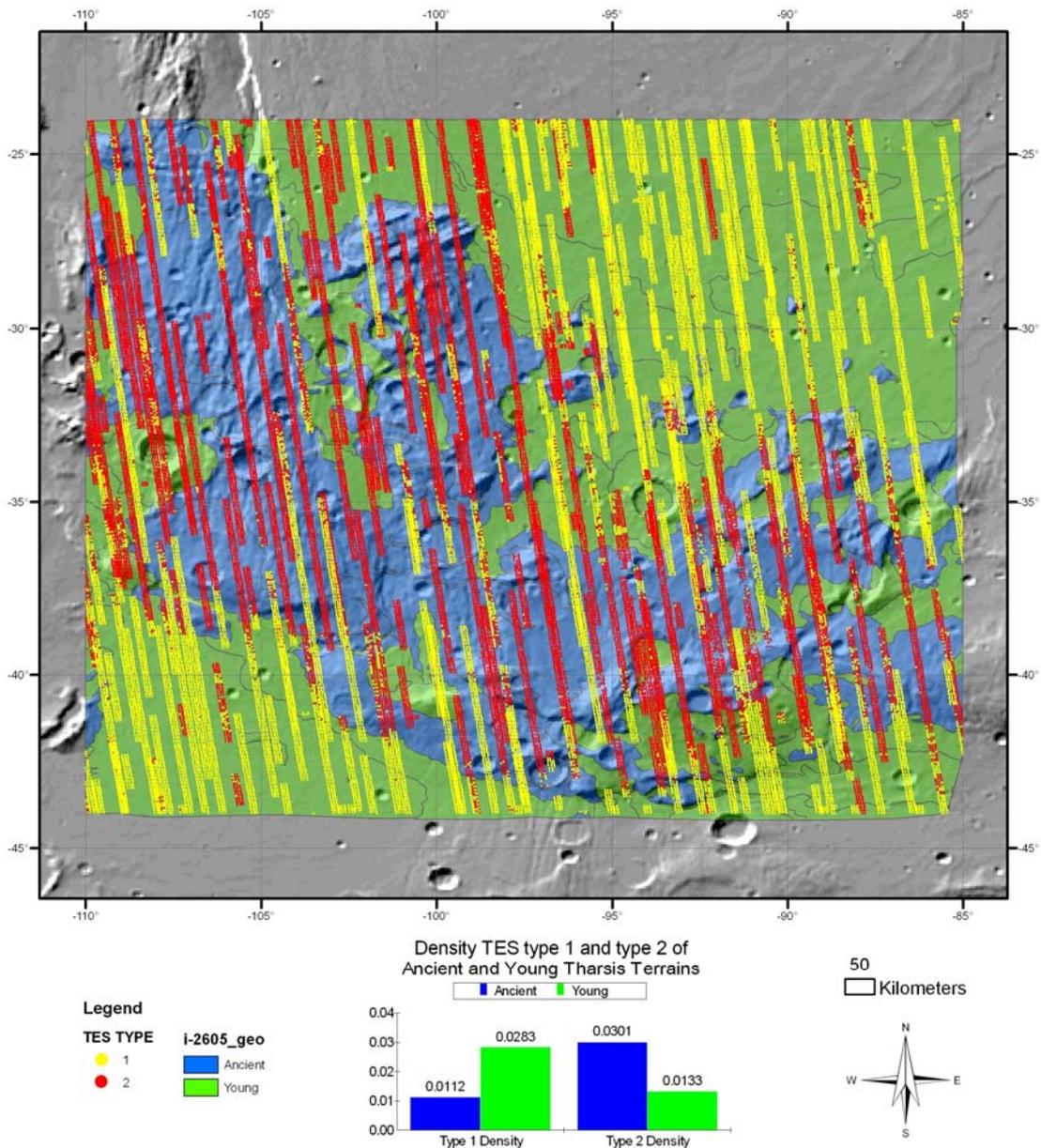
Fig. 16. Map, which corresponds with **Table 1**, shows TES type 1 and 2 plotted over a MOLA hillshade and a portion of the published USGS I-2605 Thaumasis Geologic map of Dohm et al. (2001b). The TES data was classified by applying Machine Learning "untrained" techniques to Type 1 and Type 2 information (e.g., Bandfield et al., 2000; Ruff, 2003; Wyatt and McSween, 2002) using quality TES strips that cover part of the ancient Thaumaisa highlands mountain range and younger plains-forming materials to the north and south of the mountain range. The Geology has been reclassified into two units: Ancient (mostly Early and Middle Noachian—Stage 1 map units) and Young (Late Noachian and younger—Stages 2-4 map units) based on Dohm et al. (2001b). The histogram chart shows the density of the type 1 and 2 TES data. This was calculated by dividing the individual TES points by the area of the ancient and young terrains. The longitudes are positive East. The density results clearly indicate that the mountain-forming materials are distinct from the plains-forming materials with more dominant Type-2 and Type-1 signatures, respectively. This is consistent with Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b).
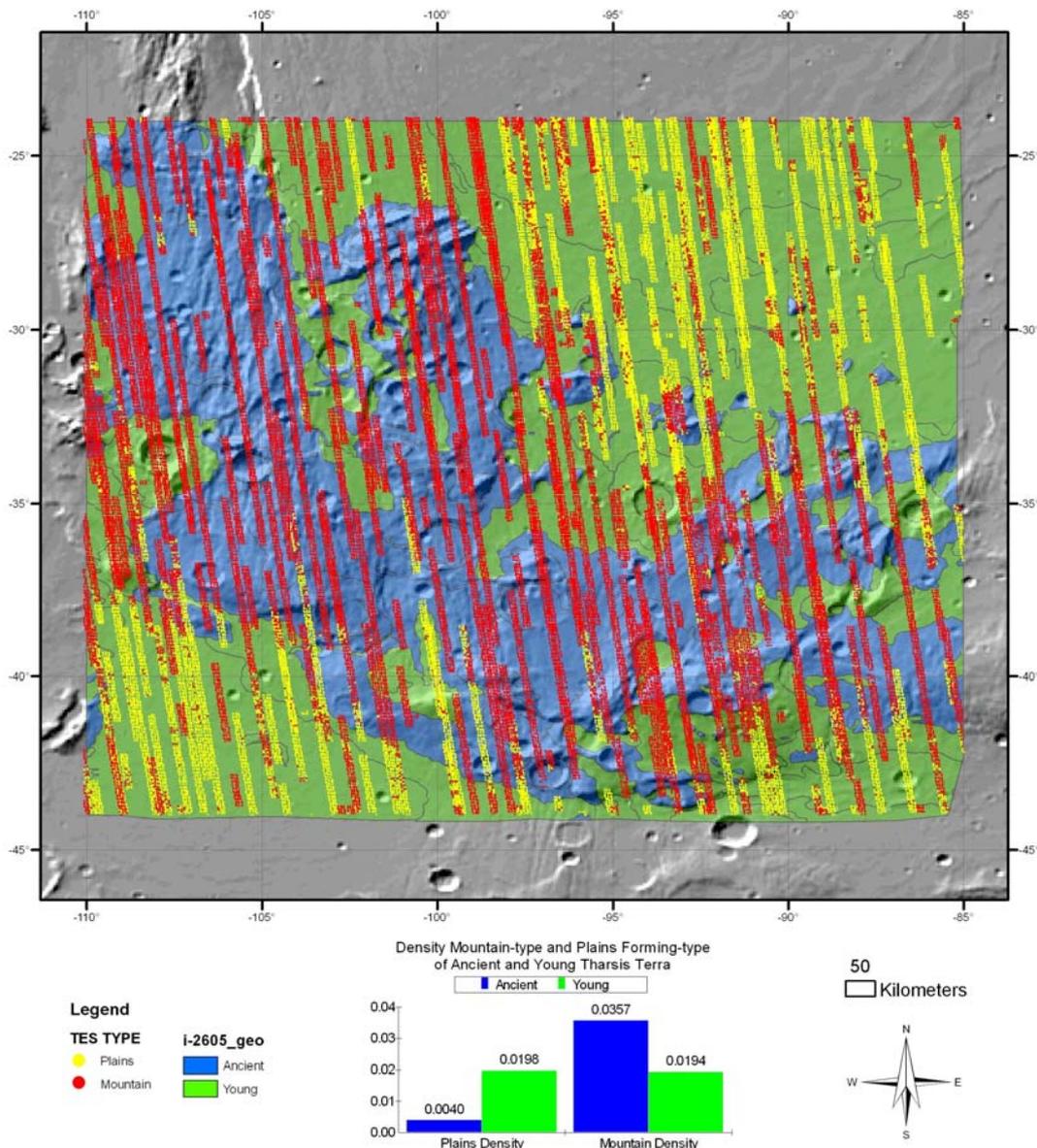
Fig. 17. Map, which corresponds with **Table 2**, shows mountain- and plains-forming-pixel types plotted over a MOLA hillshade and a portion of the published USGS I-2605 Thaumasis Geologic map of Dohm et al. (2001b) (the pixel location for each type was also based on the USGS I-2605 map). The TES data was classified by applying Machine Learning "trained" techniques to identify mountain-forming pixels from plains-forming pixels using quality TES strips that cover part of the ancient Thaumaisa highlands mountain range and younger plains-forming materials to the north and south of the mountain range. The Geology has been reclassified into two units: Ancient (mostly Early and Middle Noachian—Stage 1 map units) and Young (Late Noachian and younger—Stages 2-4 map units) based on Dohm et al. (2001b). The histogram chart shows the density of the mountain-forming and plains-forming TES-based information. This was calculated by dividing the individual TES pixels by the area of the ancient and young terrains. The longitudes are positive East. The density results clearly indicate that the mountain-forming materials are distinct from the plains-forming materials with corresponding mountain-forming and plains-type signatures, respectively. This is consistent with Viking-era geologic investigations (Scott and Tanaka, 1986; Dohm et al., 2001b).