Proceedings of The National Conference
On Undergraduate Research (NCUR) 2014
University of Kentucky, Lexington, KY
April 3–5, 2014

Analysis of Martian Spectral Modeling by Spot-Spectra Normalization

Mark Pultrone
Department of Physics & Astronomy
Rowan University
Glassboro, NJ 08028 USA

Faculty Advisor: Dr. David Klassen

Abstract

We used a method called principal component analysis (PCA) to try to find meaningful surface spectral models in our Martian spectral image data that were uncalibratabledue to the absence of a comparison star. The data are ground-based, near-infrared spectra images taken during the 19xx opposition. In an attempt to get something useful out of them, we normalized the uncalibratable data in three different ways: disc mean, disc median, and spot-spectra. Our research has found that the method that produced the best results was spot-spectra normalization. Building off these findings, we decided to analyze this method further. Using only the spot-spectra normalized data, we performed four different PCA tests: using the full 1.5–4.1 μm wavelength spectrum and the full image; using the full spectrum and disregarding the Martian poles; omitting the 1.9–2.2 μm wavelengths from the spectrum but using the full image; and omitting the 1.9–2.2 μm wavelengths from the spectrum and disregarding the Martian poles. The 2 μm region omission tests were performed because that spectral region is dominated by the non-linear absorption features of the atmospheric gas, making the linear PCA modeling inaccurate. The polar region omission tests were performed because they contain large amounts of surface ices, which could affect the creation of a “standard surface’’ model spectrum. We will present here our analysis of these tests and our assessment of which should be used in the next phase of the project—creation of a surface model in order to measure the water ice abundance in the Martian clouds.

Keywords: Mars, Atmosphere, PCA

1. Introduction

The overall goal of this research is to measure the water ice abundance in Martian clouds. Essentially, light from the Sun is going through the atmosphere of Mars, interacting with the surface, going back up through the atmosphere, going through the atmosphere of Earth, and to the detector. In order to determine exactly what the Martian atmosphere is doing to the light, it must be known how the Earth’s atmosphere affected the light and how the Martian surface affected the light. Therefore, those effects must be eliminated, leaving only the effects from the Martian atmosphere. In this portion of the research, the goal is to create a “standard surface” model spectrum to be used in future analyses.

The data being using for this project are a subset of a long-term near-infrared monitoring program. The images were taken at the NASA Infrared Telescope Facility (IRTF) using the NSFCAM instrument. This camera is equipped with a circular variable filter (CVF) that allows the choice of a set of any wavelengths in the 1.5–4.1 μm region; a set of 32 wavelengths were chosen based on the spectral features of mineral and ice species expected to be found on Mars. Table 1 provides some details on the dates of the three spectral image set used in this project including: the date the image set was acquired, which set of the night is was, the UT time range over which the set was imaged, the apparent size of Mars on that night in arcseconds, and the solar longitude of Mars on that night. Solar longitude is just the position of Mars in its orbit and is an indication of season, where L_S = 0° is the northern spring equinox, L_S = 90° is northern summer solstice, etc.

Table 1. Some observing details on the data used in this project

Date	Set	UT Range	Size (″)	L_S (°)
27 DEC 1994	2	13:01–13:24	10.7	37
16 FEB 1997	1	09:41–09:59	12.2	79
18 MAY 2001	3	12:20–12:33	17.2	163

2. Data Calibration

In general, the first step in analyzing the data is to calibrate it, which is the conversion of the light seen from instrumental response numbers to actual flux (or, in this case, radiance factor). To do this, images of a star of known brightness that is near Mars during the data collection are used. Since it is known exactly what the spectrum of the star should look like, it can be compared to the spectra received from the star during the observing night to determine how the Earth's atmosphere is affecting the data. It is then assumed that these same effects are present in the Martian data, which allows the removal Earth's atmospheric effects from the data and gives the actual light coming from Mars.

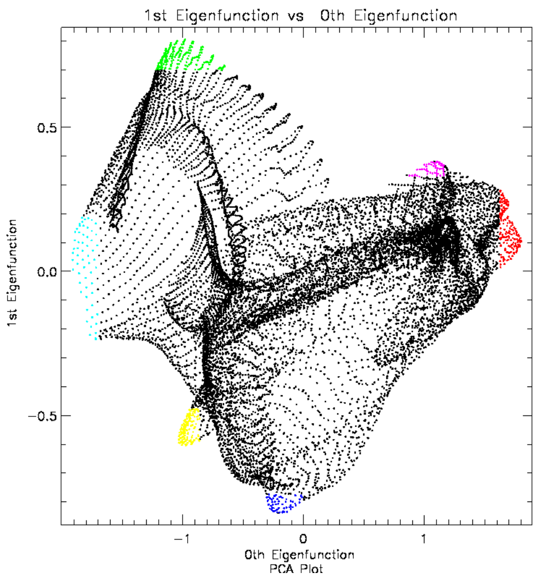
Unfortunately, some of the data did not come with comparison star images, so this process cannot be used to determine the actual light coming from Mars. Thus, other means of calibrating the data were necessary. In this research, the use of normalizing images to one another was investigated. This still does not give absolute data, but may still allow analysis of the data in a relative, at least semi-quantitative, manner. Three types of normalization schemes were investigated to see which type would give the most reliable results. The normalization scheme focused on in this paper is spot-spectra normalization.

Spot-spectra normalization involves choosing a region of Mars and dividing the spectrum of each region by the spectrum of the chosen region. The advantage of using this type of normalization is that it yields data for each pixel relative to one spot. Thus, if a pixel with no surface ices and no cloud cover is chosen, then once the images are normalized, pixels with surface ices and cloud cover will have relative spectra that demonstrate the presence of surface ice or clouds. Because the light from every spot on Mars is affected in the same way by Earth's atmosphere, this normalization automatically corrects out any of those effects.

3. Principal Component Analysis

Once the images are normalized, the process towards making a "standard surface" model can continue. To do this, it must be known what different aspects of the Martian surface cause changes in brightness at each wavelength. However, the brightnesses of a given pixel at a given wavelength may not be independent of the brightnesses at other wavelengths—mainly because many substances have multiple absorption features in the near-infrared. Thus, for example, the brightness of a region at 3 μm could be related to its brightness at 3.5 μm. One can imagine each region as a point in a 32-dimensional space with each dimension represented by a wavelength; the coordinate in each dimension is simply the brightness at each wavelength. But because of the interdependence between wavelengths, these 32 dimensions are not orthogonal to one another. As a result, the brightness at each wavelength cannot be separated to determine any effects of surface features.

There is, however, a linear algebra technique called principal component analysis (PCA) that plots each pixel in a different 32-dimensional space where all the directions are orthogonal. Essentially, PCA plots each pixel in its original 32-dimensional space then looks for the direction through that data cloud with the greatest variation (i.e. distance) and assigns that direction as the first axis in the new space it is creating. Then, it looks in all orthogonal directions to that first direction it found and picks the direction with the next greatest variation and makes that the next new axis. It does this process another 30 times, giving a new space in which each axis is orthogonal to the others, and calculates the coordinates of each regions in this new space. Mathematically, this is a classic eigenvalue problem and the new coordinates are just the eigenvectors of the variance/covariance matrix of the data. Figure 1 show Mars with the pixels plotted onto a graph of the 1st eigenfunction vs. the 0th eigenfunction. The colored pixels in the graph correspond to the colored pixels in the image of Mars.



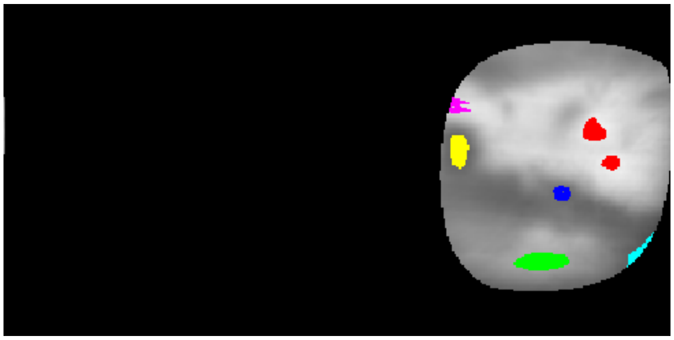


Figure 1. Mars plotted in the first two dimensions (1st Eigenfunction vs 0th Eigenfunction) of PCA space (left). Colored pixels correspond to the same colored regions in the map of Mars (right).

Because each new dimension is built from the variation in the data, each axis in the new space represents an independent trait of the data. In Figure 1, the pixels with the highest values in eigenvector 0 (red) correspond to the brightest region(s) of Mars and the pixels with the lowest values in eigenvector 0 (cyan) are the darkest regions of Mars—in this case dark shade and noise. These findings are consistent through every image, so it has been concluded that eigenvector 0 represents overall near-infrared brightness.

Looking at eigenvector 1, the pixels with the highest values (green) are near the south pole. The pixels with the lowest values (blue) are near the equator, dark, and sub-solar (i.e. near local noon). This also holds true for every image, so it has been concluded that eigenvector 1 represents the iciness/coldness (where "warm" is just "not cold"). It is believed eigenvectors 2 and 3 represent large-scale and small-scale geologic traits of the surface.

Since each subsequent eigenvector created represents less variation than the previous eigenvectors, the importance of each eigenvector decreases with respect to the overall variation in the data. Beyond eigenvector 3, the variation in the data becomes so small that it is indistinguishable from noise in the data, so those eigenvectors are being ignored completely. Eigenvector 0 accounts for approximately 85% of the variation in the data and eigenvector 1 accounts for about another 10% of the variation. The other 30 eigenvectors account for the last 5% of variance, so relative to eigenvectors 0 and 1, they are rather insignificant. Thus, not only does PCA recover a trait-based, orthogonal space, it also significantly reduces the dimensionality of the data to something far more manageable—from 32 to 3–4.

Once PCA was performed on all of the normalized data, it was decided there may have still been some factors that were negatively affecting the results. Since the goal is to make a "standard surface" model for the entire Martian surface, using the poles could skew the model because there is so much more surface ice at the poles than anywhere else. And because of this, it would be impossible to model ice clouds over an icy surface. Therefore, it was decided to try PCA again using the same images, but with the poles eliminated. Another potential problem is that CO_2 gas absorbs light at wavelengths around 2 μm and this atmospheric absorption is a non-linear process. Therefore, when PCA, a linear algebra technique, was performed, it can give skewed results from attempting to make a linear plot of a non-linear function. Thus, PCA was tried again using the whole images, but leaving out the data from 1.9–2.2 μm. Finally, PCA was performed one last time taking both factors into account—eliminating the images from 1.9–2.2 μm and removing the poles from the remaining images.

4. Results

The results will be discussed by comparing the four versions of PCA over each eigenvector in turn. Ideally, the results should show that the eigenvectors have a consistent spectral shape over all three dates; previous results using calibratable data tend to have very consistent eigenvector spectral shapes[1, 2]. It will be determined which version of PCA is “best” by which one has the most consistent sets of eigenvectors.

4.1. Eigenvector 0

The best version of PCA for eigenvector 0 (Figure 2), is the version without poles but all wavelengths. The graph shows the tightest spread of data that follow the same general pattern. After this, the no-poles-no-2-µm-band version was second, followed by all-regions-all-wavelengths; the least consistent was the all-regions-no-2-µm-band. It is clear that for eigenvector 0, taking away the poles yielded better results.

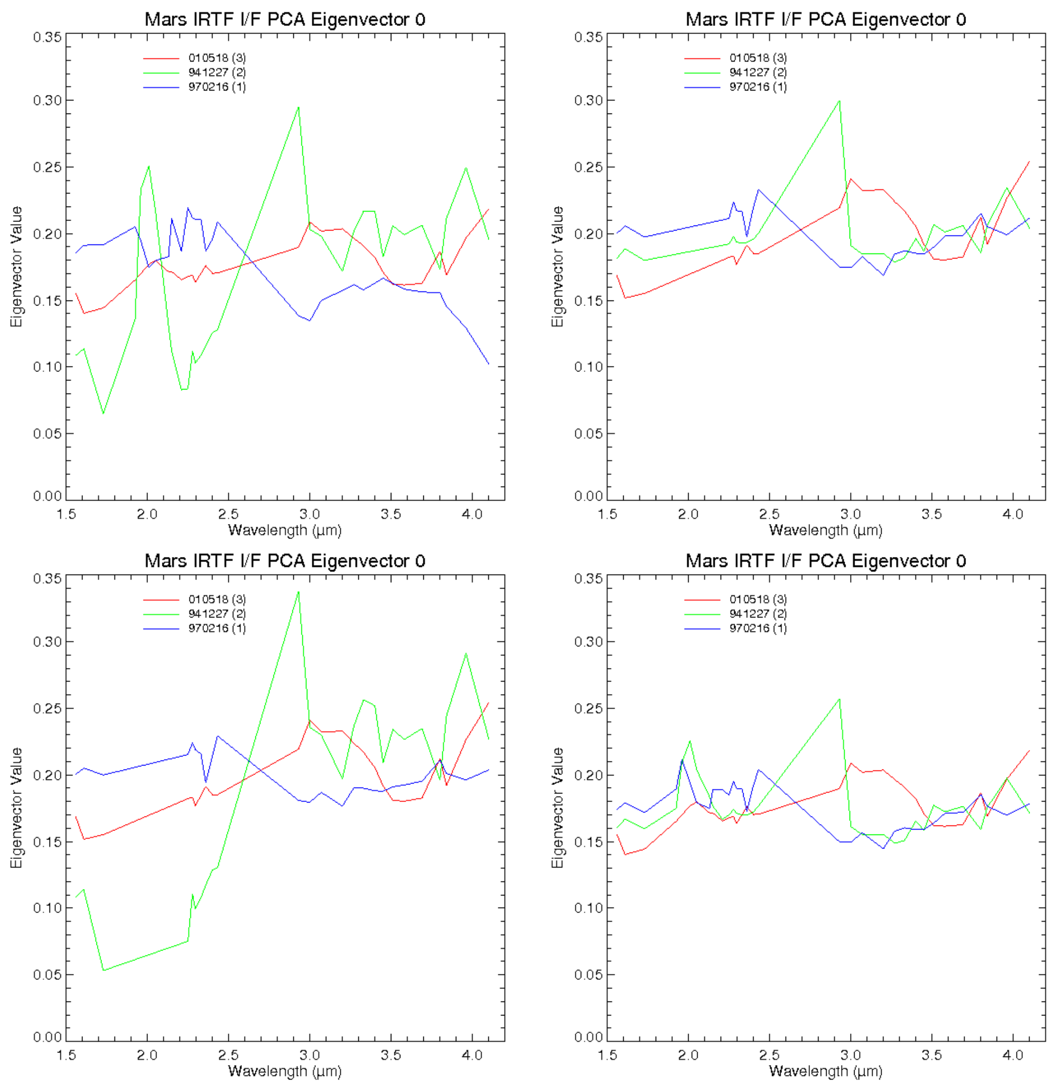


Figure 2: Graphs of eigenvector 0.

(top left) all regions and all wavelengths; (top right) without poles and without the 2 µm band; (bottom left) without the 2 µm wavelength band but all regions; (bottom right) all wavelengths but without the poles.

4.2. Eigenvector 1

The best version of PCA for eigenvector 1 (figure 3), is the version with all regions, but no 2 µm band. After this, the all-regions-all-wavelengths version was second, followed by no-poles-no-2-µm-band; the least consistent was the no-poles-all-wavelengths. It is clear that for eigenvector 1, taking away the poles yielded poorer results.

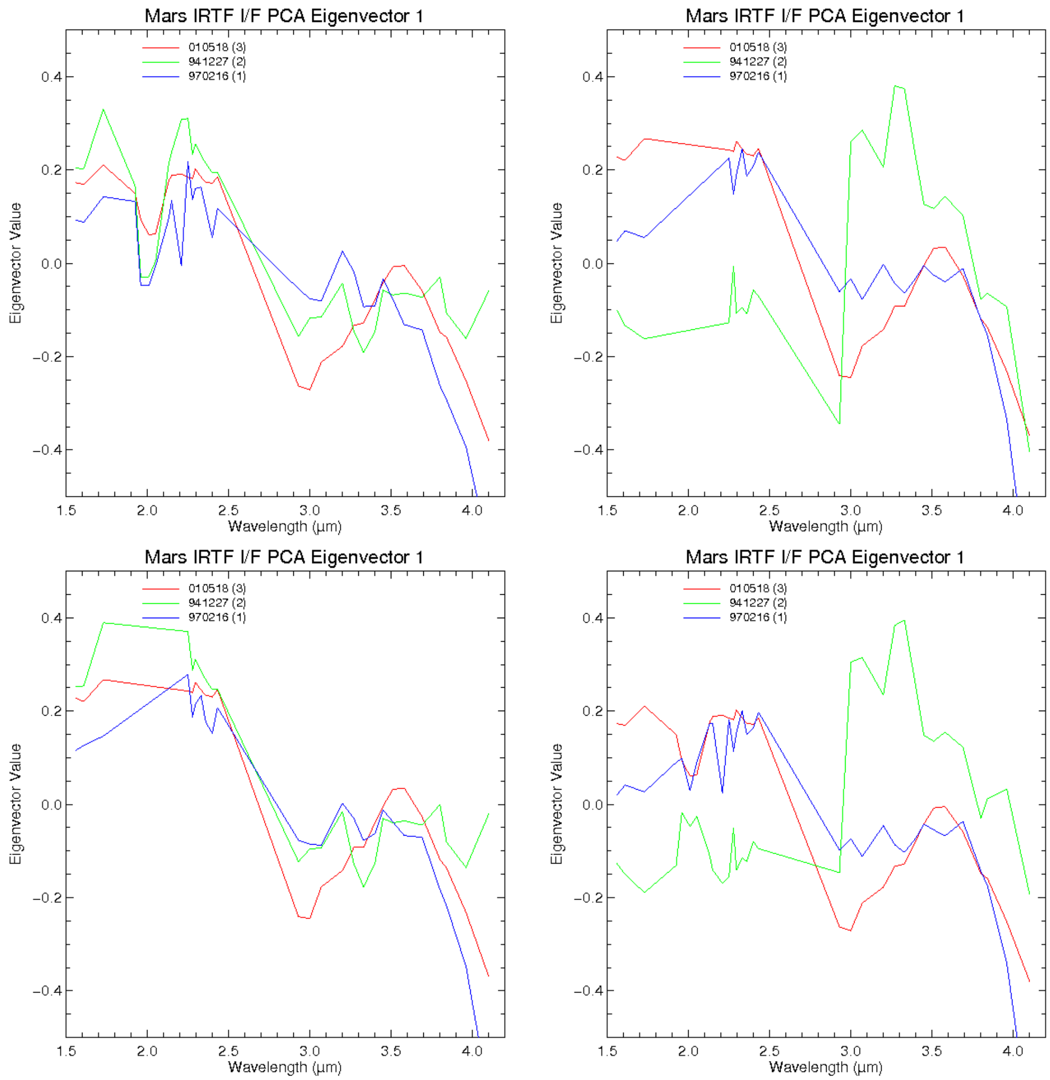


Figure 3: Graphs of eigenvector 1

(top left) all regions and all wavelengths; (top right) without poles and without the 2 µm band; (bottom left) without the 2 µm wavelength band but all regions; (bottom right) all wavelengths but without the poles.

4.3. Eigenvector 2

The best version of PCA for eigenvector 2 (figure 4), is the all-regions-no-2-µm-band. After this, the no-poles-no-2-µm-band version was second, followed by all-regions-all-wavelengths; the least consistent was the no-poles-all-wavelengths. It is clear that for eigenvector 2, taking away the 2 µm band yielded better results.

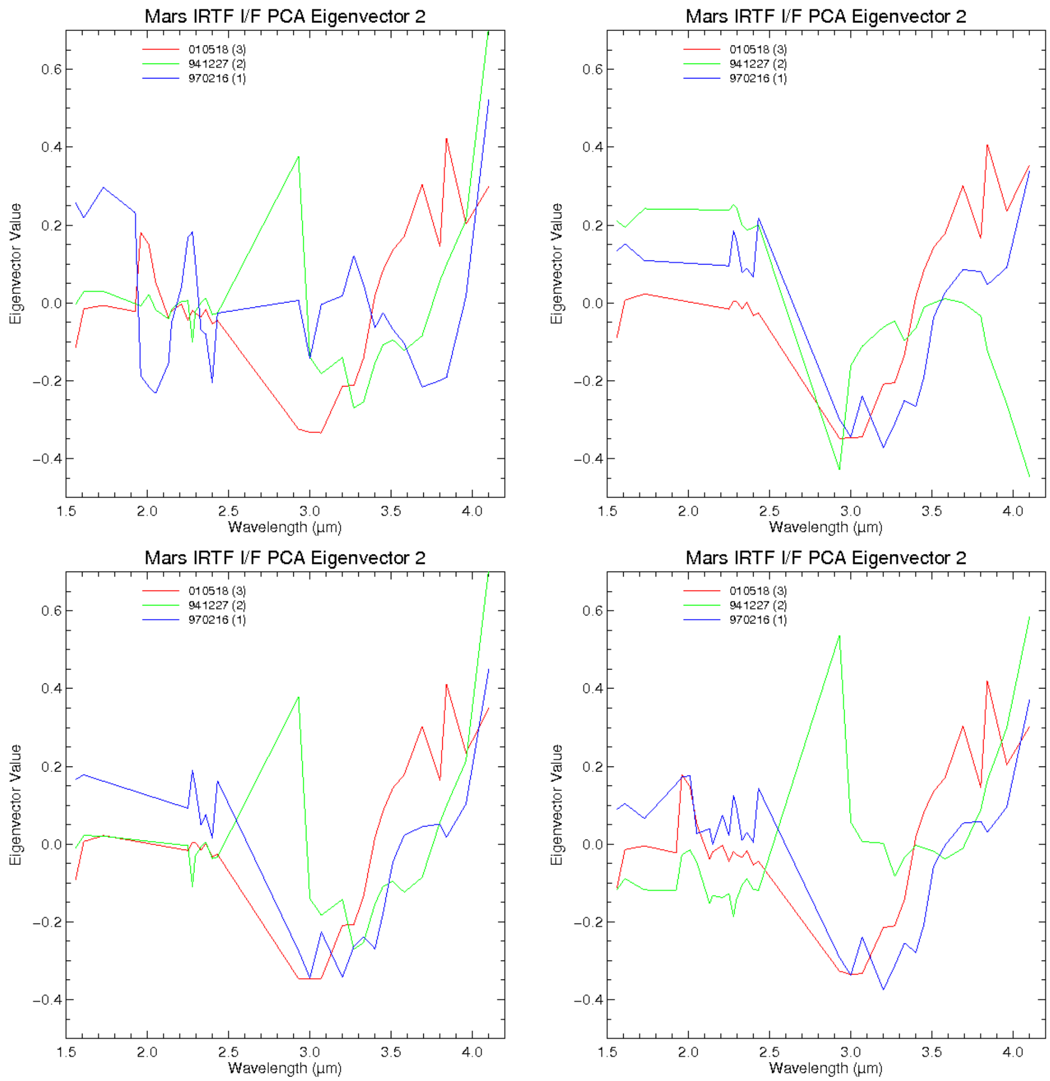


Figure 4: Graphs of eigenvector 2

(top left) all regions and all wavelengths; (top right) without poles and without the 2 µm band; (bottom left) without the 2 µm wavelength band but all regions; (bottom right) all wavelengths but without the poles.

4.4. Eigenvector 3

In eigenvector 3 (figure 5), it is clear that noise has begun to overtake the data itself. The best version of PCA for eigenvector 3 is the all-regions-all-wavelengths. After this, the all-regions-no-2-µm-band version was second, followed by no-poles-all-wavelengths; the least consistent was the no-poles-no-2-µm-band. There is too much variation in these graphs to give any further information.

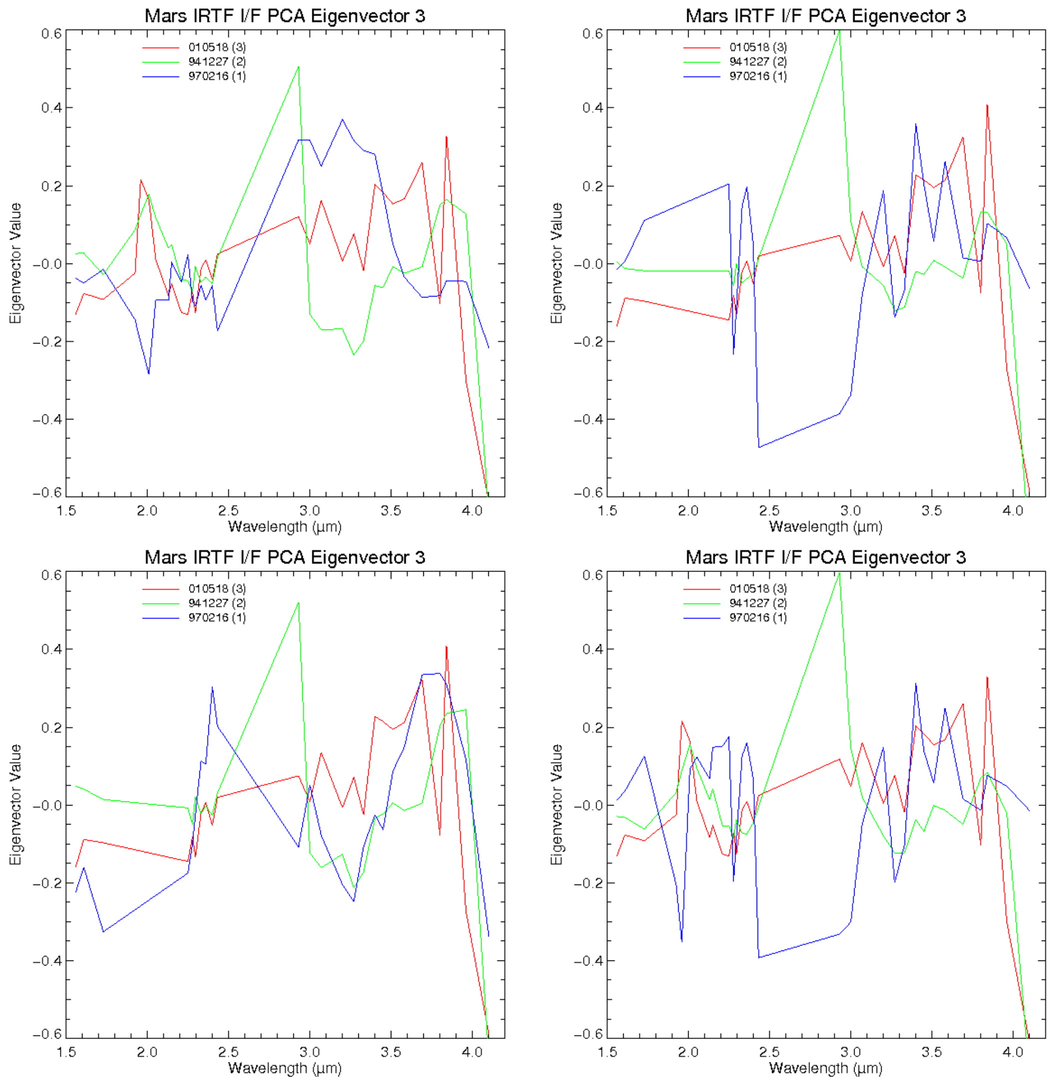


Figure 5: Graphs of eigenvector 2

(top left) all regions and all wavelengths; (top right) without poles and without the 2 μm band; (bottom left) without the 2 μm wavelength band but all regions; (bottom right) all wavelengths but without the poles.

5. Conclusions

Based on this work, it appears that PCA is improved when the 2 μm band is removed. This confirms the expectation that the non-linear absorption effects of the atmospheric CO_2 interferes with the intrinsically linear analysis technique of PCA. This type of PCA was consistently best in both eigenvector 1 and eigenvector 2 and a close second in eigenvector 3. It is also believed that the inclusion of the polar regions when performing PCA would inhibit the ability to create a “standard surface” model of Mars. It is for this reason that removal of the poles was considered, in addition to the removal of the 2 μm wavelength band. It is not clear that removal of the polar region improves the consistency overall. It improves eigenvector 0, has a neutral effect on eigenvector 2 (although is an improvement when combined with removal of the 2 μm band), but it worsens eigenvector 1.

It should be noted that while this analysis covered a significant portion of a Martian year (and thus the data span seasonal changes), only three image sets were looked at. This is not really enough data to fully determine which type of PCA is best in general and so this result is considered preliminary. Image sets will continue to be added to this analysis in future work to see if the trends noted here continue.

6. Acknowledgements

This work has been supported by NASA Mars Data Analysis Program (NAG5-11076) and by both the National Science Foundation Planetary Astronomy and Research at Undergraduate Institutions programs (0307924 and 1008744).

7. References

1. Klassen, D. R. (2007), Martian cloud modeling using NIR spectral images, In: AAS/Division for Planetary Sciences Meeting Abstracts #39, Volume 39 of *Bulletin of the American Astronomical Society*, pp. 456.

2. Klassen, D. R. (2009), Principal components analysis of Mars in the near-infrared, *Icarus, 204,* 32–47.