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

Testing Normalization Schemes Of Uncalibratable Nir Spectral Images Of Mars Using Principal Component Analysis

Sarah Kessler
Physics & Astronomy
Rowan University
201 Mullica Hill Road
Glassboro, NJ 08028 USA

Faculty Advisor: Dr. David Klassen

Abstract

Our research is part of a larger program to measure the water content in Martian clouds over diurnal, seasonal, and interannual timescales. This requires recovering a surface spectral model independent of the atmospheric spectral response. This is done using Principle Component Analysis (PCA), which has been shown to be fairly uniform across all timescales. Much of the data used come from ground-based near infrared (NIR) imaging; however, some of the data were uncalibratable due to the absence of comparison standard-star measurements. To determine if the data were still useful, we tested three normalization schemes: data mean, data median, and spot-spectra. We then preformed PCA on the normalized data to look for trends in the eigenvectors. The primary question being: are the PCA results still uniform even in uncalibrated data? We present here the results of our consistency analysis. Most of it was done through simple observations of the graphs of the PCA eigenvectors of the different normalizations. We find the mean and median normalizations show too much variability across all time scales and so are considered less superior in comparison to the spot-spectra normalized data, which showed much greater uniformity over time. We will supplement our qualitative analysis using a quantitative measure of uniformity based on the average chi-squared value between eigenvectors and their median—the greater the average, the more non-uniform they are. Our research shows that data previously determined uncalibratable due the absence of a comparison star may still be useful in further research. This will allow us to extend the study of Martian clouds to days of less-than-ideal observing conditions.

Keywords: Astronomy, PCA, Mars

1. Data

The data for this project are ground-based spectral images taken at the NASA Infrared Telescope Facility (IRTF) atop Mauna Kea, Hawaii during the oppositions of 1997 and 2001. The instrument used was the NSFCAM, a 256×256 InSb near-infrared camera; we made use of its a circular variable filter (CVF) that allowed us to acquire narrow-band ($\Delta\lambda/\lambda \approx 1\%$) spectrophometric images. Table 1 presents some details of the data including the date the images were taken, the spectral sets from that used, the apparent size of Mars at the time in arcseconds, and the solar longitude of Mars. Solar longitude is just the position of Mars in its orbit and tells us what the local season is; $L_S = 0°$ is the northern spring equinox, $L_S = 90°$ is the northern summer solstice, etc.

In order to be able to model the data and recover the ice abundances in Martian clouds, the data must be calibrated so as to remove the effects of Earth's atmosphere on the spectra and to convert instrumental response values to actual flux. Normally Earth's atmosphere can be removed by comparing the images to those of a close-by star of known magnitude. The idea is that the Earth's atmosphere will distort the star's light in the same way it does the light from Mars, so by seeing how the light from the star is distorted we can correct the effect on the Martian data.

Unfortunately in some of the data there was no comparison star present, so the data from those dates is uncalibratable. The question we were addressing in this project is whether or not these data could be calibrated by other means and therefore still be useful in later analysis.

Table 1. Details of Mars observations used in this project.

Date	Set(s)	Size (″)	L_S
16 FEB 1997	1	12.2	79
17 FEB 1997	4	12.3	80
21 APR 1997	1–6	12.4	107
08 MAY 1997	1–4	10.9	115
30 JUN 1997	1–2	7.5	140
16 JUL 1997	1–2	6.9	149
18 MAY 2001	3–7	17.2	163
19 MAY 2001	1–7	17.4	164

2. Procedure

We chose to attempt data normalization, a relative calibration scheme. We normalized the uncalibratable data in three different ways: mean normalization, median normalization, and spot-spectra normalization. In the first case, we simply divided all areas of Mars by the average value of the visible disk. In the second case, we did the same but used the median value of the visible disk; we believed this would give a "better" idea of what a typical Mars spot was than the mean as it will be less affected by extremes caused by the Martian poles. In the third case we picked a particularly bright, non-cloud-covered, area of Mars and divided the spectra of every point by the data from this.

The primary test of how well each normalization scheme worked is to run it through the initial analysis program and compare the consistency of the results. We expect, based on previous work[1, 2], that the results from this step should be relatively consistent from date to date and not vary across seasons.

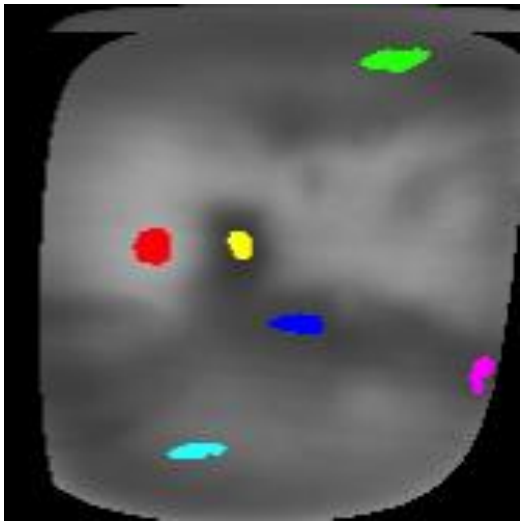
3. Principal Component Analysis

We observed Mars across 32 (or 105 in a few cases) individual wavelengths in the near-infrared (1.5–4.1 μm), however the brightness at each wavelength is not independent of the brightness at any of the others—they are correlated due to the fact that real substances typically have several characteristic absorption bands. These spectral data can be envisioned as a 32-dimensional plot, where each dimension is a wavelength, so every point on Mars would be plotted in this space based on its coordinate value (i.e. brightness) in each dimension. However, because the brightness may be correlated across wavelengths, this is a non-orthogonal data space. In order to make sense of this data cloud we use a linear algebra technique called principal component analysis (PCA) that takes our 32 dimensional non-orthogonal space and transforms it into a 32-dimensional orthogonal space. It does this by finding the "direction" of greatest variance through the data cloud and assigns that as a new dimension. It then repeats the process 31 more times with each subsequent direction having less variance than the previous one, and restricted to being orthogonal to all the previous ones. In linear algebra this is a classic eigenvalue problem; the new dimensions are the eigenvectors of the data variance/covariance matrix. Previous work[1, 2] found that the first four dimensions (eigenvectors 0–3) account for over 97% of the total variance in the data; in other words, the remaining eigenvectors 4–31 contain no valuable information and we can constrain our analysis to the first four. Thus, PCA reduces the complexity of the problem significantly.

Because the new dimensions describe the variation of the data, eigenvectors 0–3 each represent a "trait" of the Martian spectra—these traits appear to be consistent over time[1, 2]. Eigenvector 0 represents overall near-infrared brightness, eigenvector 1 represents cold/iciness, eigenvector 2 represents large-scale geology, and eigenvector 3 represents small-scale geology. These interpretations can be seen in figure 1; e.g. eigenvector 0 is bounded by the

brightest and darkest regions, eigenvector 1 is bounded by the cold, ice-covered, north polar region and a central, local-noon, dark, warm region.

If the normalized data shows similarly consistent graphs for these eigenvectors over time we can say we have obtained a true model of the Martian surface independent of the Martian atmosphere, and can then span the entire surface and disregard it in future research.



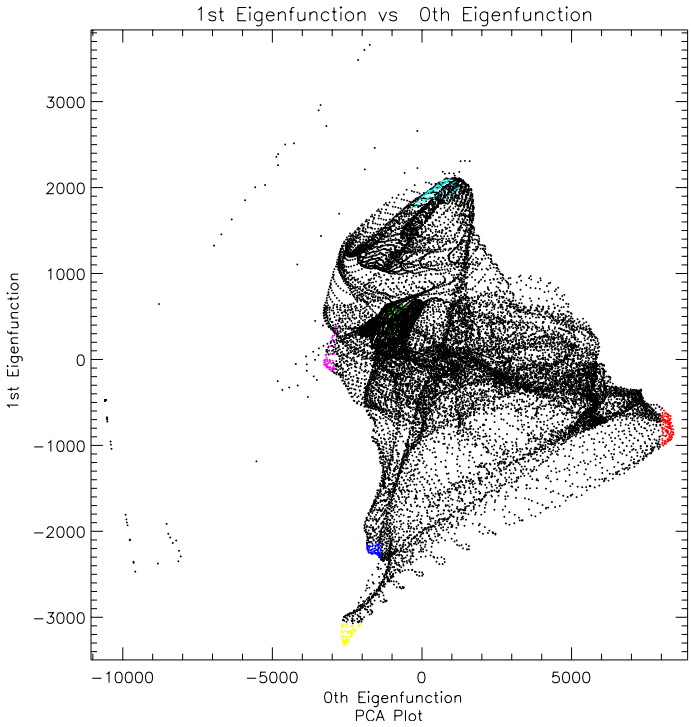


Figure 1. Areas of Mars (top) represented on a plot of eigenvector 0 vs eigenvector 1 (bottom) showing the correlations of specific regions and PCA cloud vertices/endpoints implying the trait interpretations noted in the text.

4. Results

In the mean normalization the 2001 data usually grouped very closely together, but when comparing the 2001 data to the 1997 data within the same eigenvector it is difficult to say that they are consistent. This is best shown with the mean normalized graphs of eigenvector 2 (figure 2). There is not even a hint of similarity when comparing the 2001 dates to the 1997 dates; so we can say that the eigenvectors are not consistent over time and therefore the data has no real meaning. We can say the 1997 dates are inconsistent because even though they are spread out over several months—as opposed to a single day in the 2001 dates—previous research has shown[1,2] that the eigenvectors should not vary with the seasons.

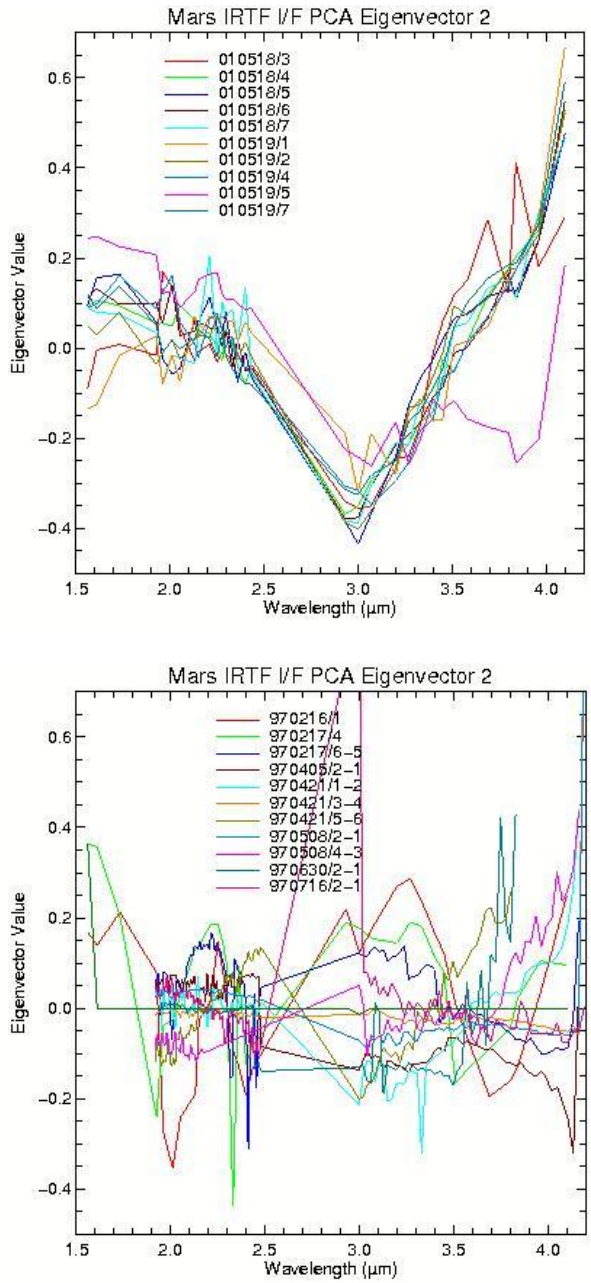


Figure 2. Mean normalized eigenvector 2 for 2001 (top) and 1997 (bottom)

The median normalization eigenvectors were also inconsistent over time, as seen when comparing eigenvector 0 plots. As in the case of the mean normalized data, the 2001 dates are consistent within themselves, but when compared to data from 1997 there is no true distinguishable trend (figure 3). However the median data still follows a general trend in eigenvectors 2 and 3 (figures 4), so we are hesitant to rule it out completely.

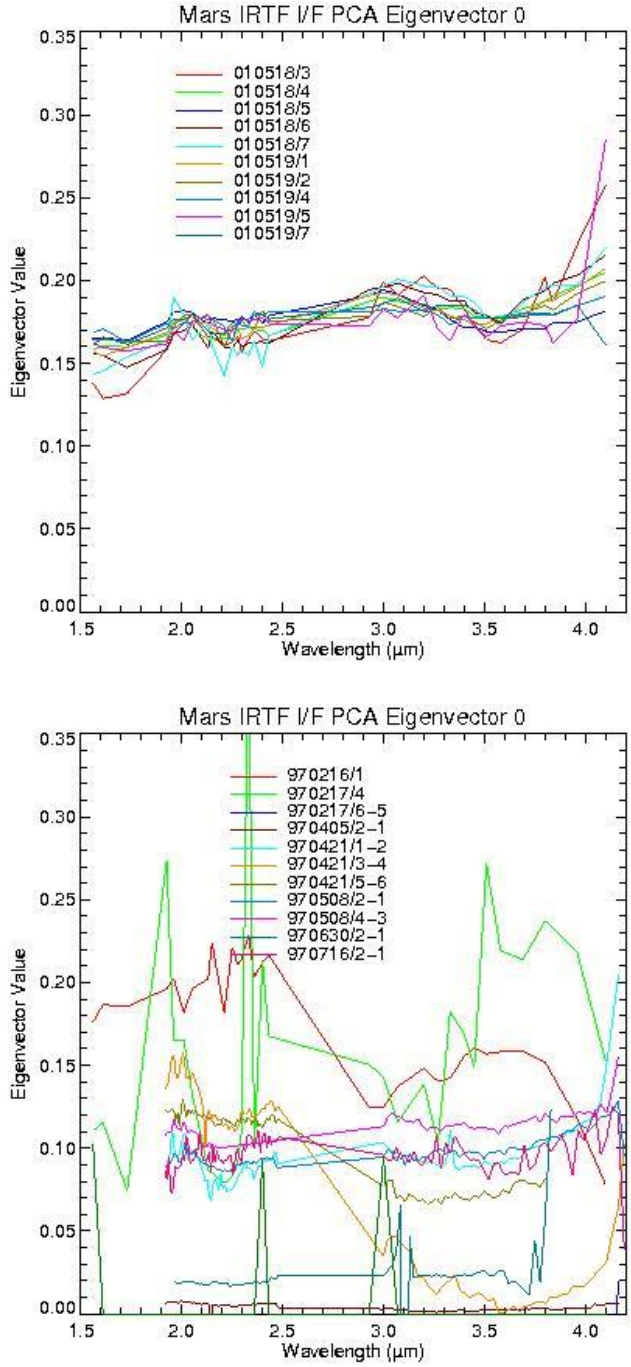


Figure 3. Median normalized eigenvector 0 for 2001 (top) and 1997 (bottom)

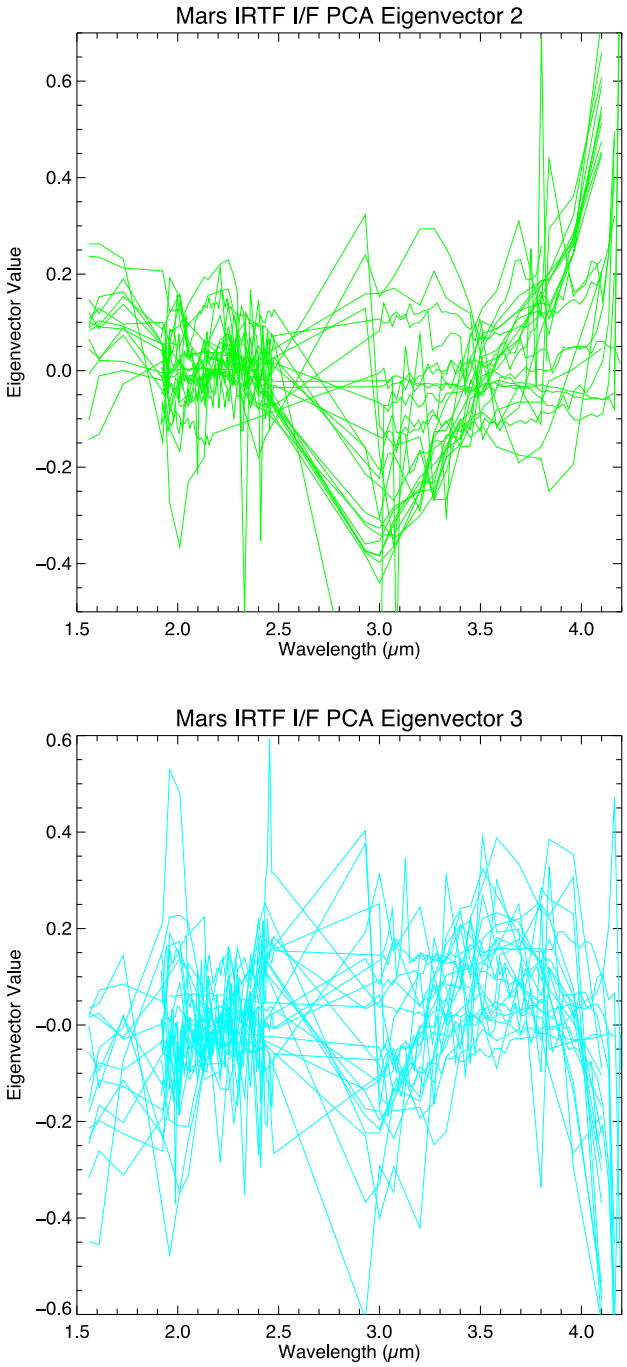


Figure 4. Median normalization in eigenvector 2 (top) and eigenvector 3 (bottom). While there are obvious outliers the graphs follow a general trend.

Finally, we look at the spot-spectra normalization. In graphs of eigenvector 1 we can see that data follows the same general trend in both sets of dates (figure 5). There a few clear outliers in the 1997 data, but the overall shapes are similar. The pattern is better seen when graphed over all dates in figure 6.

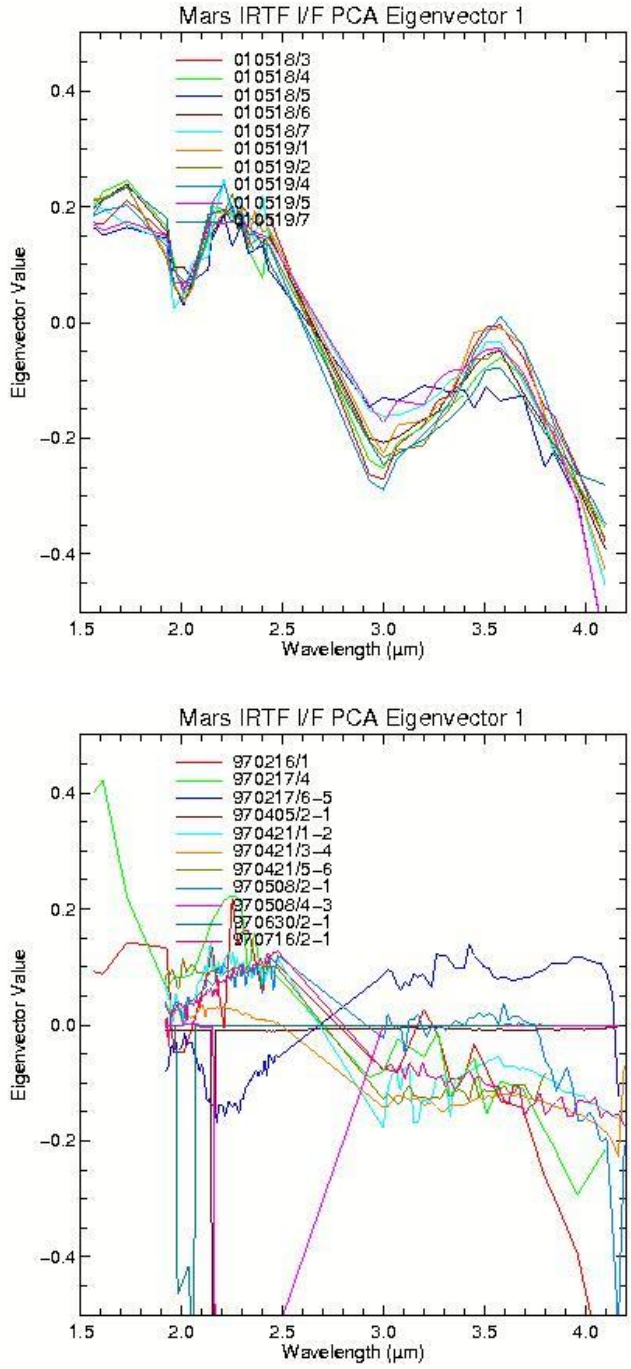


Figure 5. Spot-spectra normalized eigenvector 1 for 2001 (top) and 1997 (bottom)

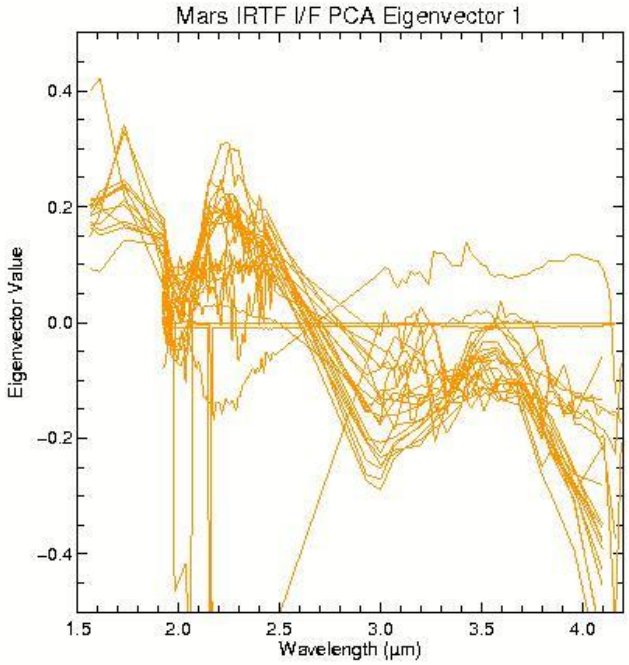


Figure 6: spot-spectra normalized eigenvector 1 graphed over all dates. There is an easily recognized trend with only a few outliers

While the spot-spectra normalized data was the best of all three there were several inconsistencies. For example, while the graphs had similar overall patterns across dates they still tended to group more closely within the year they were taken. This can be seen in the graphs of eigenvector 2 (figure 7). The graph on the top shows eigenvector 2 graphed over all dates, and though there are several very prominent outliers, it is easy to see an overall trend. However, when we use the same graph, but group each set of dates by color, we can see that the graph really trends differently based on the year the data was taken.

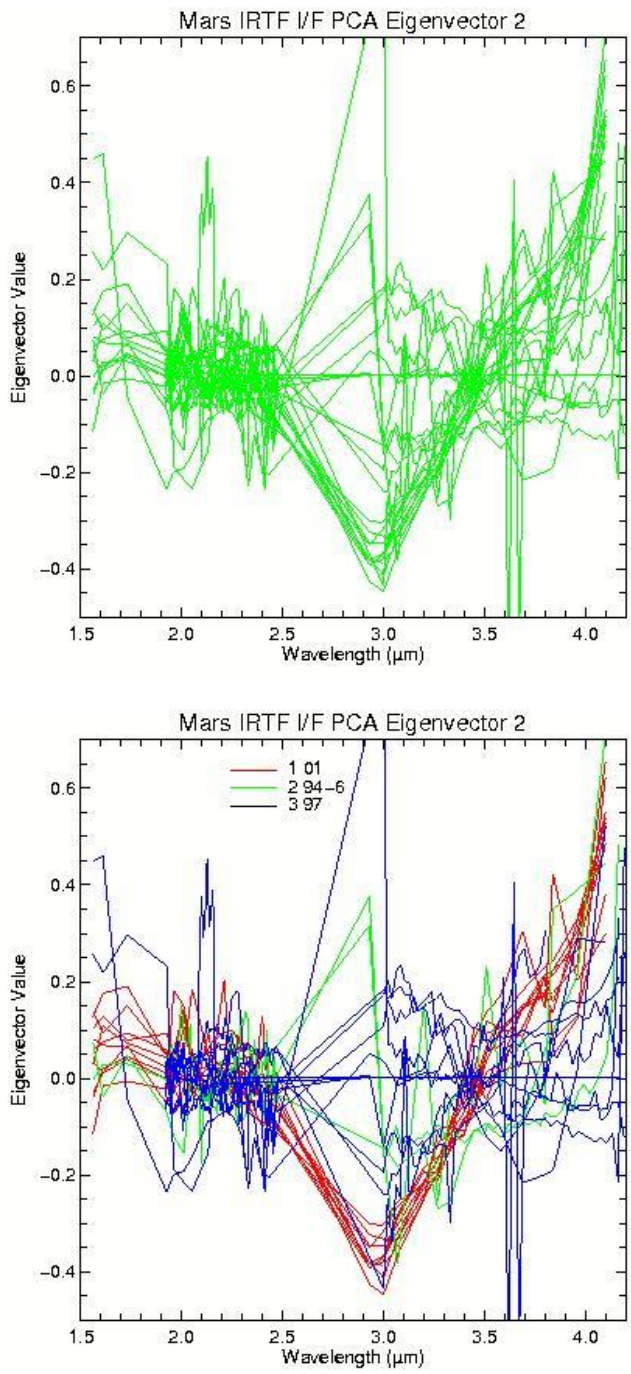


Figure 7. Spot-spectra normalized eigenvector 2 graphed over all dates in one color (top) and with 2001, 1997, and 94-96 dates separated by color (bottom)

5. Future Work

The outlying data are a concern—if these are truly outliers, as in there is some reasonable explanation for the differences based on, say, observing conditions, it is less problematic. But if they truly represent real differences, it will mean that there is no way to calibrate the data in a relative sense and still analyze them. One possible explanation is the inclusion of the Martian poles in our data. The north and south poles of Mars are very icy, which

means they reflect a lot of light and are therefore very bright. Their brightness, however, is not representative of the rest of the Martian surface and might be detrimentally affecting the analysis. Another issue throughout all of our normalizations was the inclusion of data from 1.9–2.4 µm; this is because Mars has a mostly carbon dioxide atmosphere, and carbon dioxide has a strong 2 µm absorption band. Atmospheric gas absorption is a non-linear process, and all of our work up to this point has been based of the fact that PCA is a linear technique. As we move forward with our research we will try to reduce the outliers by removing either or both the poles and the 2 µm band.

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.