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

Using PCA With the 2 μm Band Excluded Over Normalization Schemes to Analyze Uncalibratable NIR Spectral Images of Mars

Phillip Chandler
Department of Physics and Astronomy
Rowan University
Glassboro, New Jersey 08028 USA

Faculty Advisory: Dr. David Klassen

Abstract

In our research analyzing the water content of Martian clouds, much of the ground-based spectral imaging data collected was deemed uncalibratable due to the absence of standard star measurements to compare to our images of Mars. One of our goals in this research is to use principle component analysis (PCA) to produce a surface spectral model independent of the atmospheric spectral response. When the PCA was used and analyzed over the whole 1.5-4.1 μm spectrum, we noted that the absorption feature from 1.9-2.2 μm was showing up in all eigenvectors, implying it was major component of the surface spectral signature. However, we know that the atmospheric gas is really the dominant reason for this feature so the question was raised: if the 1.9-2.2 μm band were removed, would the resulting PCA eigenvectors show more consistent trends, leading to a better surface model? However, this spectral region is also where water ices are active, so the secondary question is: if the 1.9-2.2 μm band were removed, would it adversely affect our ability to model Martian clouds? In order to address the first question, we performed PCA tests across three different types of normalization excluding the 1.9-2.2 μm band. The three normalization schemes are: disc mean, disc median, and spot-sectra. We will present the results of these three tests: a comparison of the consistency of the eigenvectors across diurnal and seasonal time scale and to an overall median set of eigenvectors.

Keywords: Mars, Atmosphere, Water

1. Introduction

The primary goal of our research is to measure the water content in Martian clouds as a function of time. To do this we use near-infrared spectral images taken at the NASA Infrared Telescope Facility on Mauna Kea, Hawaii using the NSFCAM. The instrument is equipped with a circular variable filter (CVF) that lets us chose our own specific set of narrow-band ($\Delta\lambda/\lambda \approx 1\%$) wavelengths in which to image Mars. For this work, we used a set of 32 wavelengths across 1.5-4.1 μm. The data were acquired over several weeks during each opposition from 1995 through 2003 (approximately every 25 months).

As part of standard data reduction we have to correct for Earth's atmosphere and how it affects our data. Since the Earth's atmosphere changes dependent on various factors, we cannot rely on the atmosphere being constant from one observation to the next. To perform this correction if we happen to know the actual light received at Earth by a comparison star as a function of wavelength then at each wavelength, we can create a conversion factor from light we actually receive, as affected by passage through Earth's atmosphere, to the light we should be receiving. If Mars and the star were observed under similar enough conditions we can just use the conversion factors from the star on the Mars data to convert that into the light we should be getting from Mars. Unfortunately, there were several sets of data for which no standard star was observed—this project is an analysis of those data to see if any science can still be recovered from them. Table 1 gives some information on the subsets of data used in this project. Each observing season is the year in which the opposition took place; size is the range of apparent size in arcseconds of Mars, and L_S is the solar longitude of Mars. Solar longitude is simply the angular position of Mars in its orbit measured from its northern spring equinox; so L_S = 90º is the northern summer solstice, L_S = 180° is the northern autumnal equinox, etc. Therefore, the data used in this project span late notheren winter, through northern spring and summer— albeit, across several Martian years.

Table 1.

Observing Season	Size (“)	L_s (°)
1995	5.3-13.7	340-79
1997	6.8-14.2	48-149
1999	6.5-16.2	79-146
2001	11.0-17.5	142-206

Since the goal of this research is to find out the water content in the Martian clouds we need to make model spectra. By using a full radiative transfer code and adjusting the free parameters (ice cloud optical depth, dust optical depth) until we can get a best fit for our data. With this best fit we can calculate the water content in the Martian clouds. Unfortunately, this requires us to know the surface reflection, which we do not know with certainty. There are a few good methods once can use to acquire a surface reflectance, the two most common are: to assume a region nearby the region of interest is sufficiently "the same" and aerosol/cloud free that its reflectance can be used directly; to measure the spectrum of the region of interest at some other point in time where one can assume it is sufficiently aerosol/cloud free and use that as the reflectance. Neither of these is ideal. The technique we will use is to model the surface as a linear combination of "pure" spectral endmembers—then the coefficients of the linear combination become more free parameters for the best fit. Then all we have to do is find these endmembers.

We developed a process using principal component analysis (PCA)[1] and target transformation (TT)[2] to find surface spectral endmembers. PCA transforms the data from a measurement space (brightness as a function of wavelength) into a spectral trait space defined by the variance within the data itself. These spectral traits are just the eigenvectors of the data variance and the most significant of them are used in TT to create the candidate endmember spectra. This project investigates the ability to use the uncalibratable data in these processes by looking at the results of PCA on them.

2. Normalization

Our solution to having no standard star for absolute calibration was to normalize our data. Normalization allows gives us relative brightness values which then allows us to compare the images of Mars across all wavelengths in a meaningful way. We chose to investigate three methods of normalization: mean, median, and spot-spectra.

Mean and median normalization divides out an "average spectrum" of the light from Mars filtered by Earth's atmosphere. These two methods should preserve most of the Mars point-to-point variations and give us a better chance at seeing clouds on Mars, which will prove useful when determining the water content in later steps. The trade off to these two methods is that we could be washing out clouds if they are extensive enough to affect that "average spectrum". The use of a mean is because it has a more meaningful interpretation; the use of the median is because it will be less affected by extreme outlier spectra.

For spot-spectra normalization we choose a "reference point" on the Martian surface and normalize to that spot so that all spectra become relative to that spot chosen. If we choose a location on Mars that is non-dusty, non-cloudy, and we get rid of the overall local geology we will not wash out anything other than the location chosen. Even if by chance there are clouds or dust over our reference location, the places with less or no cloud or dust cover will have inverse absorption features, which we can still attempt to model—or use this fact to choose a new reference location and start again. This allows us not to actually lose any clouds as we do with the mean and median normalization. The only downsides to this method of normalization is that we could have absorption and anti absorption in the same spectra in which differentiating between the two absorption types becomes difficult. Also if the spot is not "consistent" over time the spots chosen would not be intrinsically close enough to each other causing the results to have less consistency.

3. Principle Component Analysis

Once we normalize our data we then put it through our first analysis step, a process called principal component analysis (PCA). If we think of our spectral image cube as simply a collection of spectra, we see that the brightness at each wavelength can be thought of as a coordinate in "direction" of that wavelength dimension. Thus, we have a 32-dimensional space where you can plot each location on Mars—we create a 32-dimensional "data cloud" of points. However, because brightnesses of substances at multiple wavelengths are not independent of one another (e.g. something green is green because of a correlated set of reflected colors at many specific wavelengths), the axes of this space are not orthogonal; they are not independent.

The job of PCA is to identify a new set of axes through this data cloud that are orthogonal. To do this, PCA identifies the direction through the data cloud with the greatest variance (i.e. longest "distance" through the cloud) and make that the initial new dimension. It will repeat this process with the next greatest variance, but is restricted to move only in any direction that is orthogonal to all previous dimensions found. Mathematically, this is just a classic eigenvalue problem and the new dimensions are the eigenvectors of the

variance/covariance matrix of the data. The eigenvalues are a measure of how much data variance is explained by that dimension. Previous work (Klassen, et al., 1999) has noted that over 90% of the data variance can be explained by the first 3-4 eigenvectors and the rest of the dimensions are indistinguishable from noise. Thus, PCA not only transforms the data into a orthogonal space, but it reduces the dimensionality (and thus complexity) of the analysis.

For this particular project, we made an initial restriction on the input spectral image cubes: we removed the 1.9-2.1 µm images. The reason for this is because the atmospheric CO2 gas has a strong absorption feature that takes up most of this and. Atmospheric absorption is a very non-linear process but PCA is, by definition, a linear technique and other work by our group indicates that removal of this band improves the PCA results.

Since our eigenvectors are created from following the variations in the data, we are finding dimensions that describe intrinsic "traits" in the data itself. This means that each eigenvector produced corresponds to a variation in physical properties. This allows us to intrinsically link eigenvectors to actual physical properties on Mars. Figure 1 shows a plot of Mars in the first two dimensions: eigenvector 1 vs. eigenvector 0. The colored regions are vertices of the data cloud which are also colored in the map of Mars, showing the correspondence between PCA space and regions. We see that eigenvector 0 corresponds to overall near-infrared brightness—bight regions (red) of Mars have high values in this dimension and the darkest regions (purple) have low values. Eigenvector 1 corresponds to coldness/iciness—high values (green and cyan) correspond to the polar region or areas with ice-cloud cover and low values correspond to relatively dark areas at local noon (yellow and blue). Eigenvector 2 represents the large scale geology and eigenvector 3 represents small scale geology.

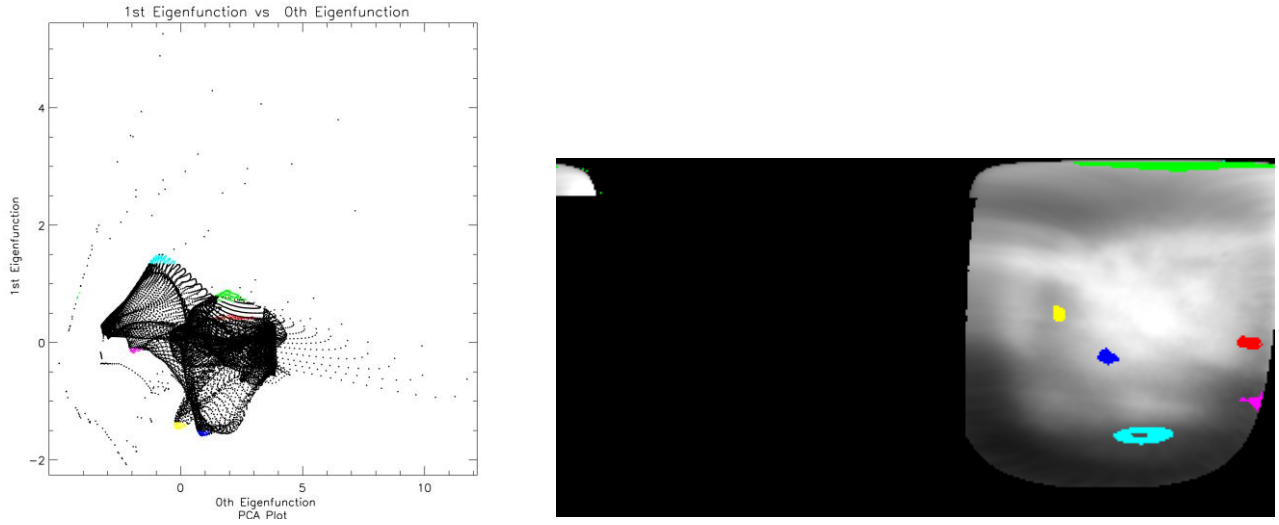


Figure 1. This colored data cloud graph corresponds to the adjacent map of Mars. As stated before this shows the correspondence between PCA space and regions.

4. Results

To determine which version of normalization is "best" we compared the eigenvector spectra across several dates for each type. Previous results with calibrated data 2,3 indicate that the significant eigenvectors tend to be fairly constant across all dates so in our comparisons we will assess the degree of consistency of each eigenvector, in turn, for each of the normalization schemes.

The eigenvector 0 normalizations (figure 2) look fairly consistent for all normalizations. The two best were the mean and median normalizations as seen by the relative compactness and similarity of their spectral shapes.

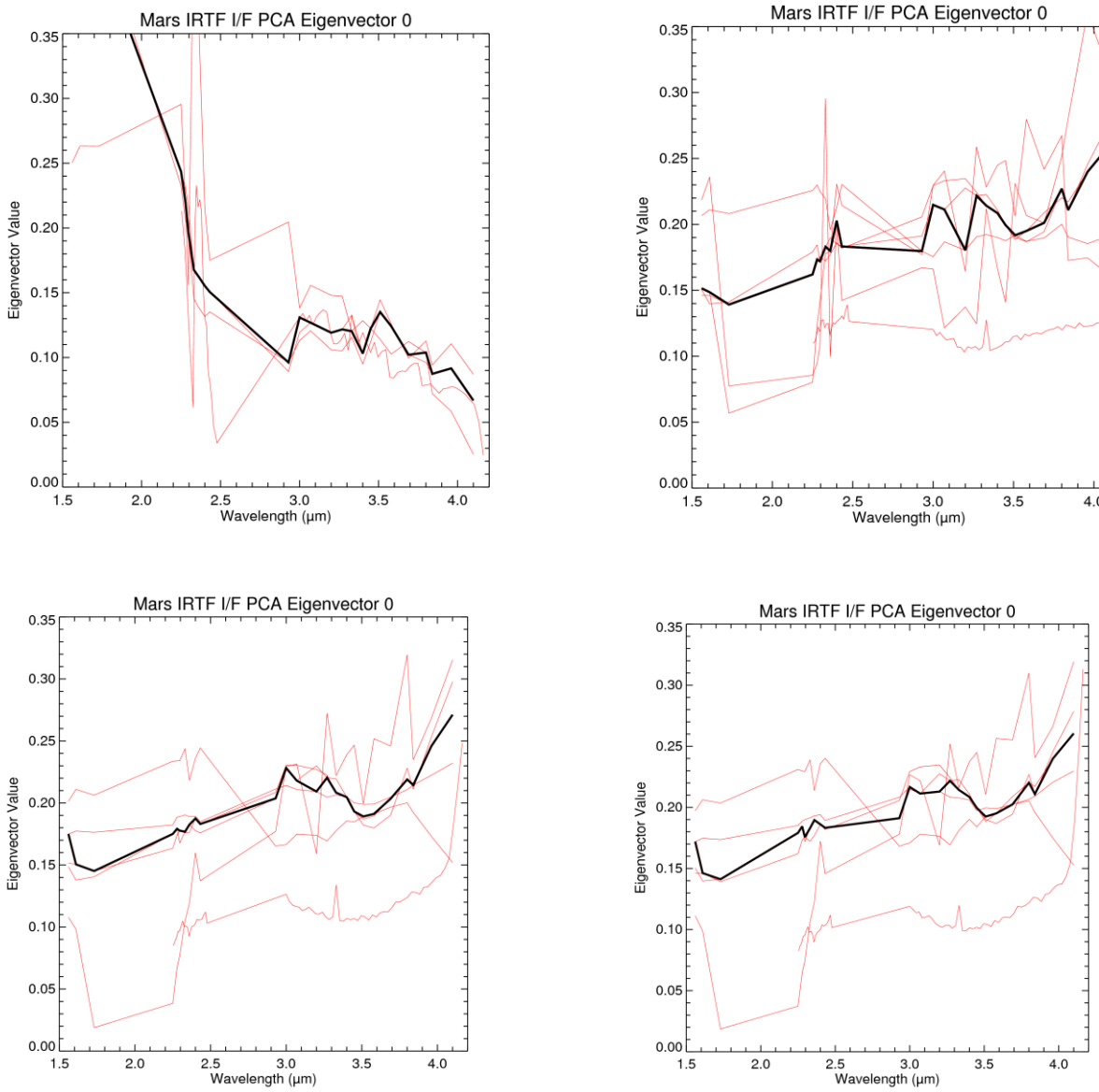


Figure 2: Eigenvector 0 for unnormalized (top left), spot-spectra normalized (top right), mean normalized (bottom left) and median normalized (bottom right)

For eigenvector 1, (figure 3) the spectra, again, look fairly consistent for all normalizations. The two best were the mean and median normalization once again. This is evident because of the the relative compactness and similarity of their spectral shapes.

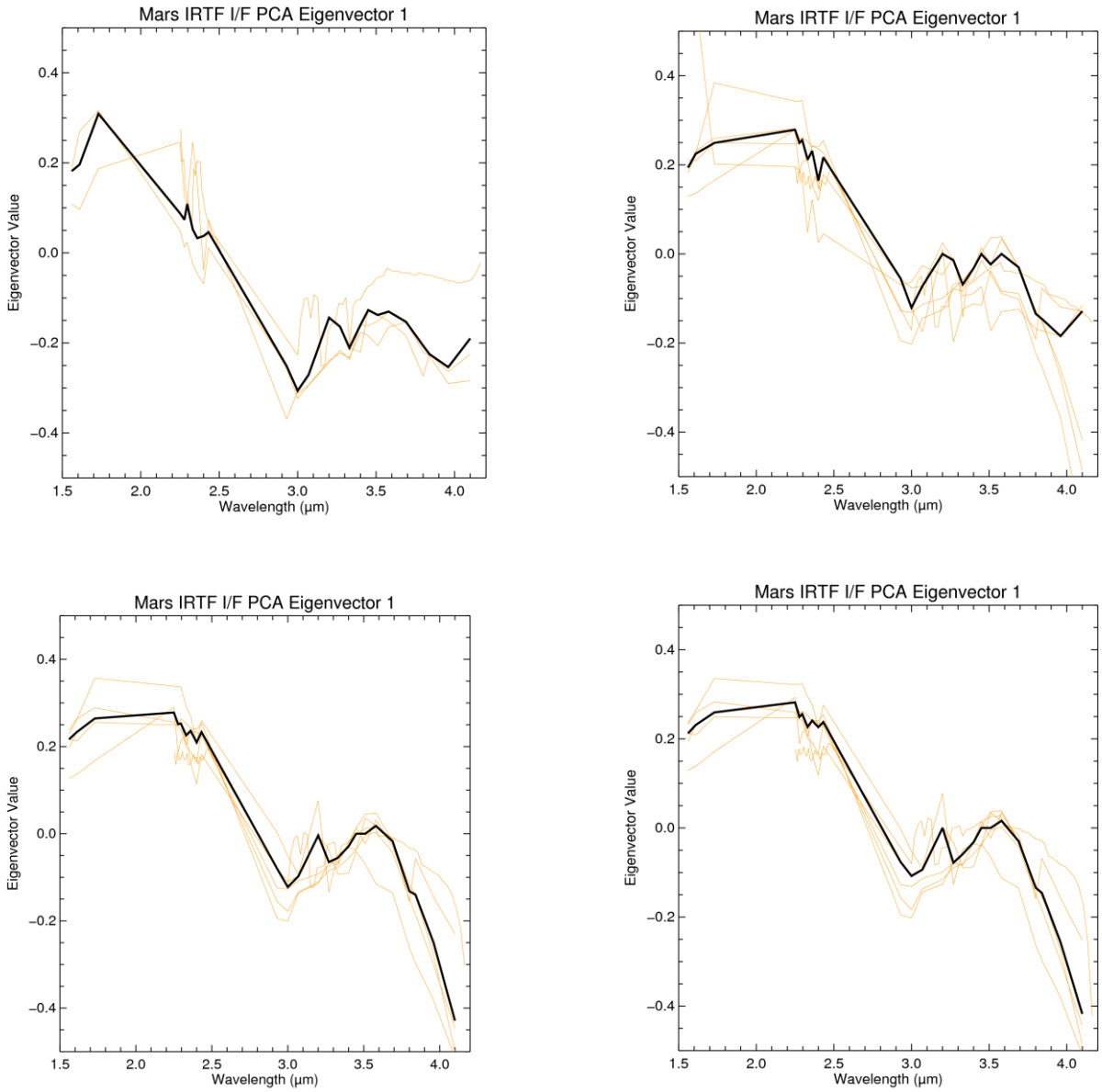


Figure 3: Eigenvector 1 for unnormalized (top left), spot-spectra normalized (top right), mean normalized (bottom left) and median normalized (bottom right)

In eigenvector 2 (figure 4) we see a lot of variation in the non-normalized plot and also in the spot-spectra plot. The mean normalization shows the clearest overall trend while the median normalization has almost as clear of a trend as the mean normalization.

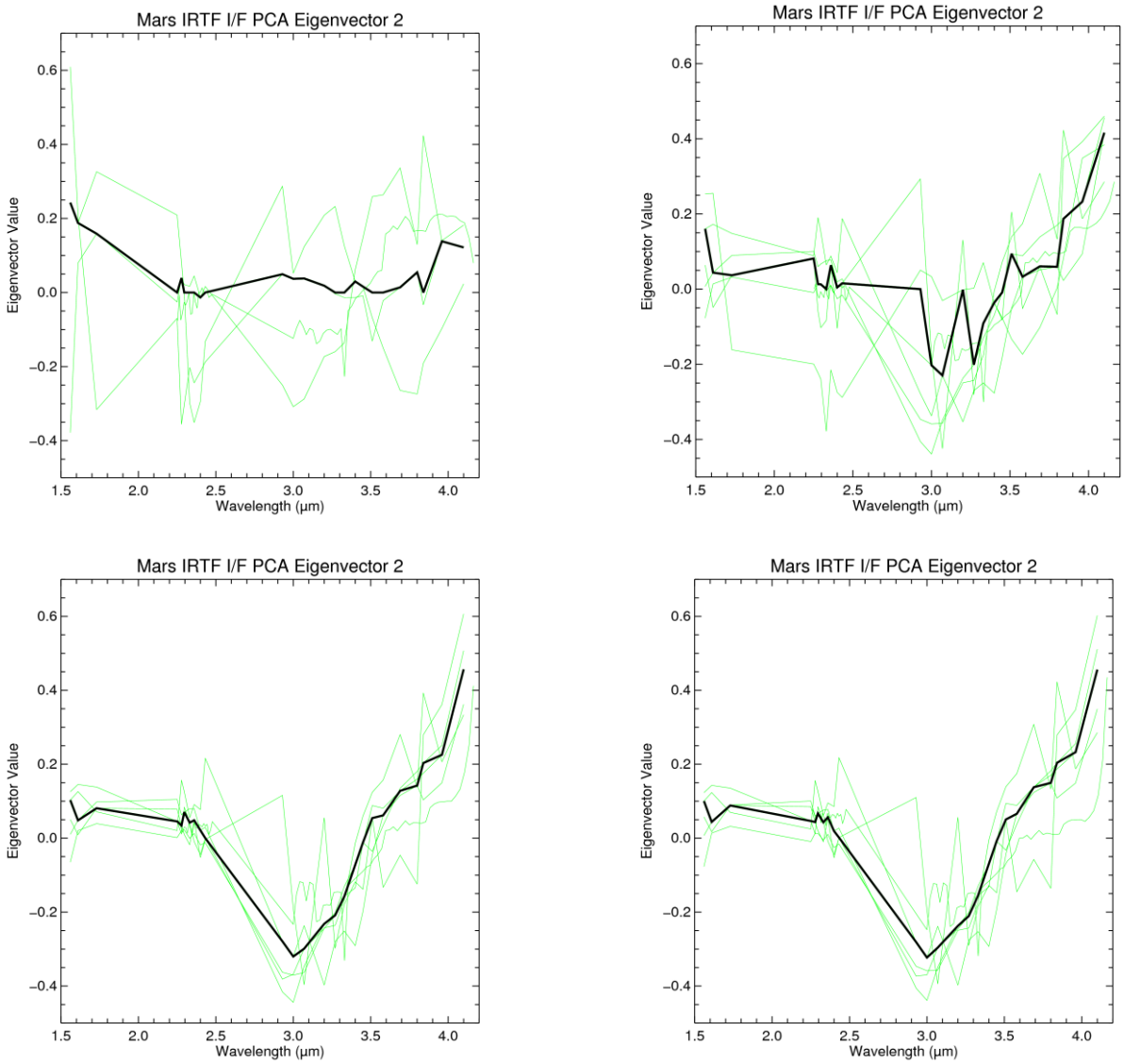
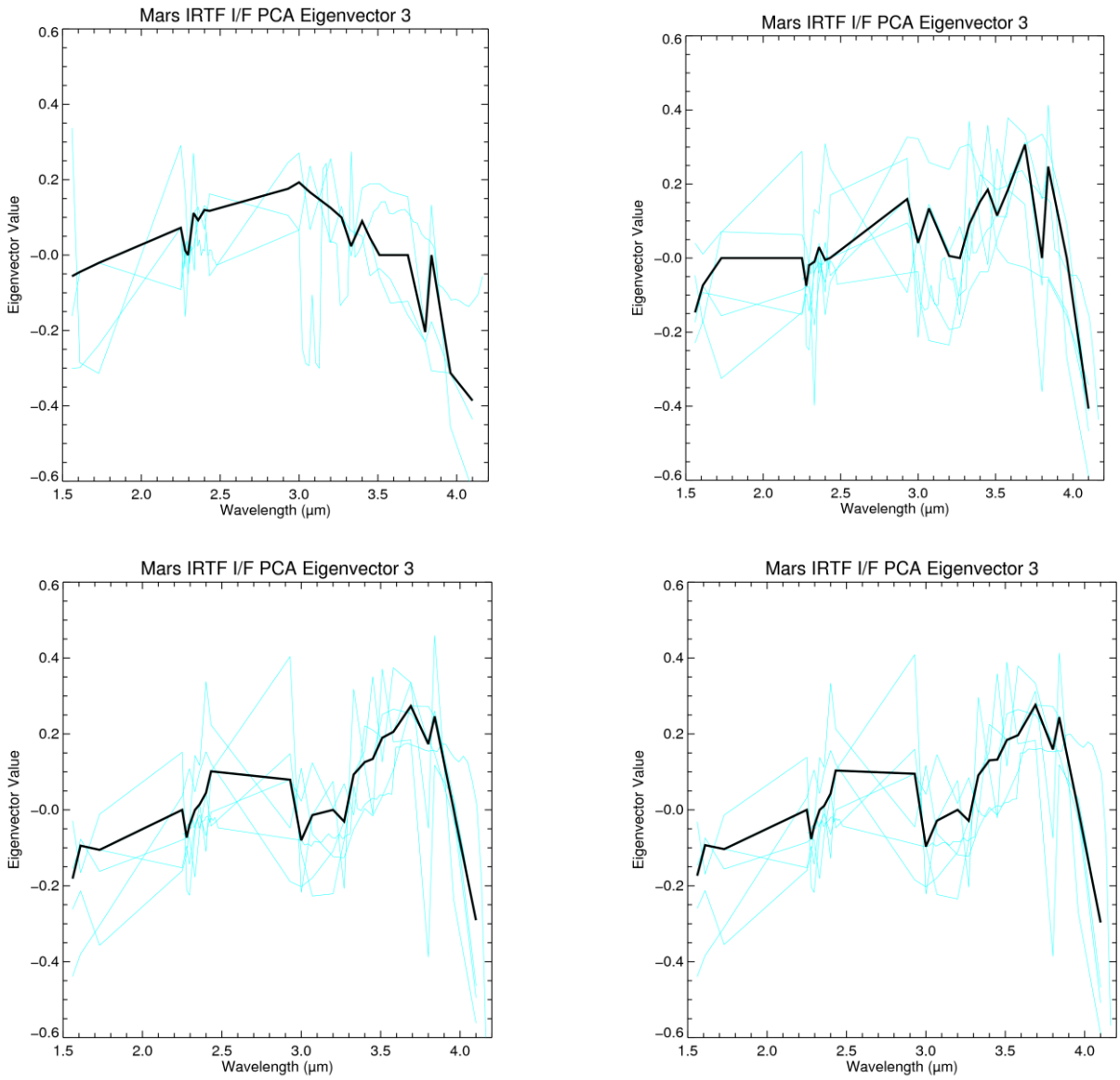


Figure 4: Eigenvector 2 for unnormalized (top right), spot-spectra normalized (top left), mean normalized (bottom left) and median normalized (bottom right)



Eigenvector 3 is inconsistent for all normalizations, which is to be expected due to its small-scale nature, and was expected to hold little results. Regardless, the mean normalization shows the most symmetric trend.

Figure 5: Eigenvector 3 for unnormalized (top right), spot-spectra normalized (top left), mean normalized (bottom left) and median normalized (bottom right)

Overall, we see that, after we exclude the 2 μm band from the data, the best normalization to use going forward, is the mean normalization or the median normalization. The spot-spectra and unnormalized versions simply do not have consistent enough trends compared to previous work.

5. Conclusions

While we did not expect any consistency in the unnormalized data (as the spectra are almost entirely uncorrelated) we originally felt that the spot-spectra normalization would be more useful—a well-chosen reference point has less chance of washing out ice and dust aerosol spectral signatures. It is no surprise that mean and median normalization schemes yield similar results—each simply references each point to an "average" spectrum and so long as whatever data is being analyzed does not have any major outliers, the two concepts are very similar to one another. That they give the best PCA results implies either that we did not have "well chosen" reference spots, our references spots (which were not all the same) were too inconsistent, or that there is no such thing as a "well

chosen" reference spot. If the latter two ideas are correct, then spot-spectra normalization is not usable. We will investigate this further in future work.

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., Bell, J. F., Howell, R. R., Johnson, P. E., Golisch, W., Kaminski, C. D., Griep, D. (1999), Infrared spectral imaging of martian clouds and ices, *Icarus, 138*, 36-48.

2. Bandfield, J. L., Christensen, P. R., Smith, M. D. (2000), Spectral data set factor analysis and end-member recovery: Application to analysis of Martian atmospheric particulates, *J. Geophys. Res., 105*, 9573-9588.

3. 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.

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