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

Testing Multiple PCA Techniques with Median Normalization to Analyze NIR Spectral Images of the Martian Surface

Dominic Payne
Physics and Astronomy
Rowan University
201 Mullica Hill Rd
Glassboro, New Jersey 08028 USA

Faculty Advisor: Dr. David Klassen

Abstract

The ultimate goal in this research is to better understand the ice clouds of the Martian atmosphere using near-infrared spectral images. To do this, a good, a priori spectral model of the surface is needed. To get that, we use a process known as principal component analysis (PCA); previous work has shown that the PCA eigenvectors can separate the surface and ice components of the spectra. Some of the data, however, could not be calibrated and thus it was not clear if PCA would have the same effect. To see if this data can still be useful, it was analyzed across three different types of normalization: data mean, data median, and spot-spectra. Collaborator results appear to indicate that the most useful method of normalization is the spot-spectra method, while data mean normalization was the least useful[1,2]. Even so, we were reluctant to completely disregard the median normalized data as the spot-spectra method requires a significantly bright region with no cloud coverage, and that is not always present in the images. Using only median normalized data, four different PCA tests were performed: using the full 1–4 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 and using the full image; omitting the 1.9–2.2 µm wavelengths form the spectrum and disregarding the Martian poles. The reason for omitting the 1.9–2.2 µm spectral region is that there is a major spectral absorption due to gaseous CO2 abundant in the Martian atmosphere—a non-linear process that could perhaps affect the linear PCA technique. The reason for omitting the poles is that they have a permanent cap of ground ice, which can confuse the modeling technique that assumes all ice spectral features are due to ice clouds. We are optimistic that these results will be complimentary to those from spot-spectra normalization, but without the limitations that spot normalizing presents. With this secondary method of normalization, trends from the data can potentially be analyzed that were either not present or not obvious through PCA on spot normalization. This will allow for meaningful data collection on not only days with less than ideal observing conditions, but also those when Mars is especially cloud covered.

Keywords: PCA, Mars, Normalization

1. Introduction

In spectra of the planet Mars, there is not only absorption from its atmosphere, but there also effects of its surface spectral response. This, of course, is due to the fact that the light in the near-infrared coming from Mars is mostly reflected sunlight. In order to study the atmospheric spectrum, it is necessary to eliminate the surface spectrum from the measured spectrum. However, in order to get the spectral response of the surface, all the atmospheric response would need to be eliminated. This seeming paradox can be resolved through a few possible techniques[3,4,5]. In this work the data itself is used, along with principal components analysis, to attempt to create an accurate surface spectral model that can then be eliminated.

The data used in this study is a subset of images from a multi-year observing program[5,6]; Mars was observed in the near-infrared over several weeks every opposition from 1995–2003 (approximately every 25 months). Data were collected with the NSFCAM, a 256×256 InSb camera, at the NASA Infrared Telescope Facility (IRTF) atop Mauna Kea, Hawaii. NSFCAM is equipped with a circular variable filter (CVF) that let us select specific wavelengths to create a spectral image set. The spectral resolution ($\Delta\lambda/\lambda$) of the CVF is about 1%. For details on the specific data used here, see table 1. Included are the dates the data were taken, which set on that date was used, the UT time range over which that set was collected, the apparent size of Mars, in arcseconds, and the solar longitude of Mars. The solar longitude is the position of Mars in its orbit and is a measure of the season, where $L_S = 0°$ is the northern spring equinox, $L_S = 90°$ is the northern summer solstice, etc. Thus, the data used here run from mid spring through late summer.

Table 1. Details on the dates used in this study.

Date	**Set**	**UT Range**	**Size (″)**	L_S
27 DEC 1994	2	13:02–13:24	10.7	37°
16 FEB 1997	1	10:09–10:31	12.2	79°
18 MAY 2001	3	12:20–12:33	17.2	163°

2. Calibration

Since all of the data come from ground-based observations, the local, terrestrial, atmospheric effects, such as overall infrared transparency and seeing must also be considered. These effects can vary greatly depending on the direction the telescope is pointed and the night an observation is made. This problem is usually taken care of by calibrating with standard stars. On an observation run, along with the Mars data, data will also be taken from a star whose spectrum is already well known. As long as the standard star chosen is in close proximity to Mars in the sky, a comparison can be made between the star data to the actual spectrum of that star, so that a conversion factor can be used to eliminate Earth's atmospheric effects. This, however, is not always an option. Sometimes, there are no standard stars close enough to Mars, so different methods of calibration must be utilized. One of the methods used was to normalize each spectral image by the median of the entire visible disk of Mars—median normalization (NMD). This, in essence, gives an average brightness per wavelength that will not be heavily swayed by outliers in the data. While NMD provides a helpful substitute to the standard calibration method, it has its limitations. The normalization process uses a relative value (the median) as a reference point, as opposed to an absolute value like standard star data.

3. Principal Component Analysis

A spectrum is the measurement of brightness as a function of wavelength. In this research, data are usually collected across 32 wavelengths, ranging from 1.5–4.0 µm. This gives a 32-dimensional space in which the brightness per spot on Mars can be plotted, with each of the 32 axes representing one of the wavelengths. These axes, however, are not orthogonal, meaning that there are varying degrees of interdependence in the data, making it implausible to analyze directly. This is where principal component analysis (PCA) is necessary. PCA is a linear algebra technique that finds vectors through the data that follow the variance in the data. The 0^{th} vector points along the path of greatest variance, then the next vector points along the direction of next greatest variance, but restricted to being orthogonal to the 0^{th} direction, and so on. This process is illustrated in figure 1. These new dimension vectors are called eigenvectors and they are simply linear combinations of the original (wavelength) dimensions. It also turns out that nearly all of the variance in the data is accounted for after only the first 3–4 eigenvectors, which means PCA also reduces the dimensionality (and thus complexity) of our data. These few eigenvectors become the axes of a new space.

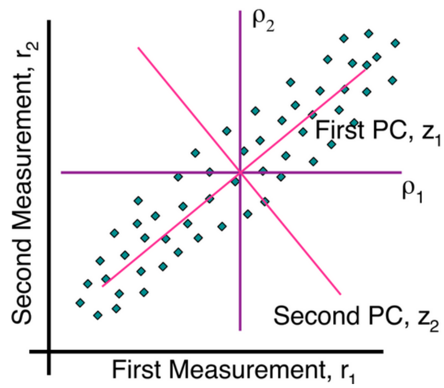

Figure 1. Eigenvectors going through the paths of greatest variance in a set of data

Since these eigenvectors represent the variance in the data, they can be used to identify the key traits responsible for the variation in the first place. Figure 2 is an example showing a PCA plot (first two dimensions only), along with a map of Mars; the points colored in the PCA plot correspond to the colored regions in the map. By using these, traits that the first few eigenvectors represent can be identified. We have found[5] that the first eigenvector, EV0, corresponds to brightness (bright areas have high values of EV0, dark areas have low values of EV0), EV1 corresponds to "coldness/iciness" (polar regions have high values, sub-solar dark regions have low values), and EV2 corresponds to regional geology.

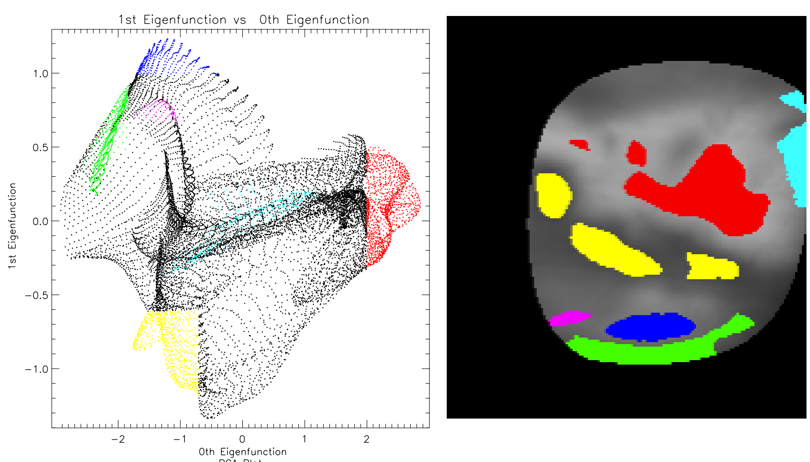

Figure 2. A 2-dimensional slice of a 3-dimensional PCA plot, along with a picture of Mars. The colored regions correspond to the colored data points in the PCA plot

4. Results

Using NMD, four different versions of PCA were tested. The first version was simply ordinary PCA; the second was PCA without the 2µm wavelengths; the third was PCA without the Martian poles; the fourth without the 2µm band and the Martian poles. The 2µm wavelengths in some of the PCA's have been excluded due to heavy CO_2 absorption in the 2µm band[8], and the Martian poles have been excluded because of the abundant ground ice present[9] there that could confuse the modeling technique. The "best" method is determined by finding those that produce the most consistent trends, as previous work with calibrated data[10] has shown should be the case. The graphs in figures 3–5 are the results from three different data sets for each eigenvector. In each graph, the colored lines are the eigenvector spectra from each date and the bold line is the median of those eigenvectors.

As can be seen in figure 3, the best trends for EV0 (the brightness eigenvector) show up in each PCA that excludes the Martian poles, as the general trend is much closer to the median line than in the other 2 types of PCA. It is difficult to say whether or not the removal of the 2 µm band is an improvement over simply removing the pole alone. Note the relative flatness of the graphs in the figure—this feature is a product of the normalization process. By normalizing the data to the median brightness, we should expect a relatively flat line in EV0, since it corresponds to overall near-infrared brightness.

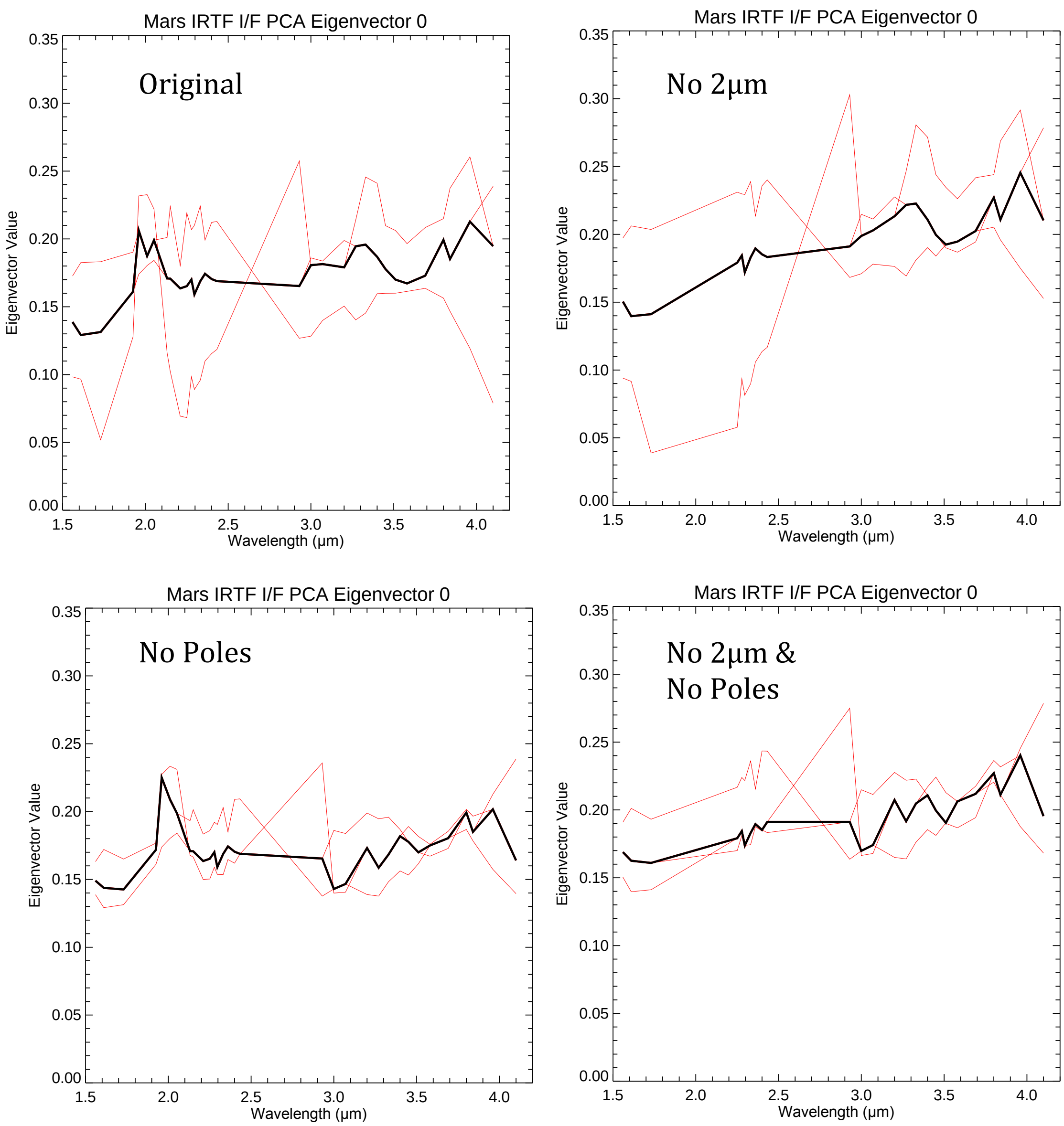

Figure 3. Results of the four PCA tests for eigenvector 0. The median line is in bold.

As can be seen in figure 4, the best trends for EV1 (the "iciness" eigenvector) show up in the original PCA as well as the PCA that excludes the 2μm band. Removal of the polar region appears to worsen the overall consistency.

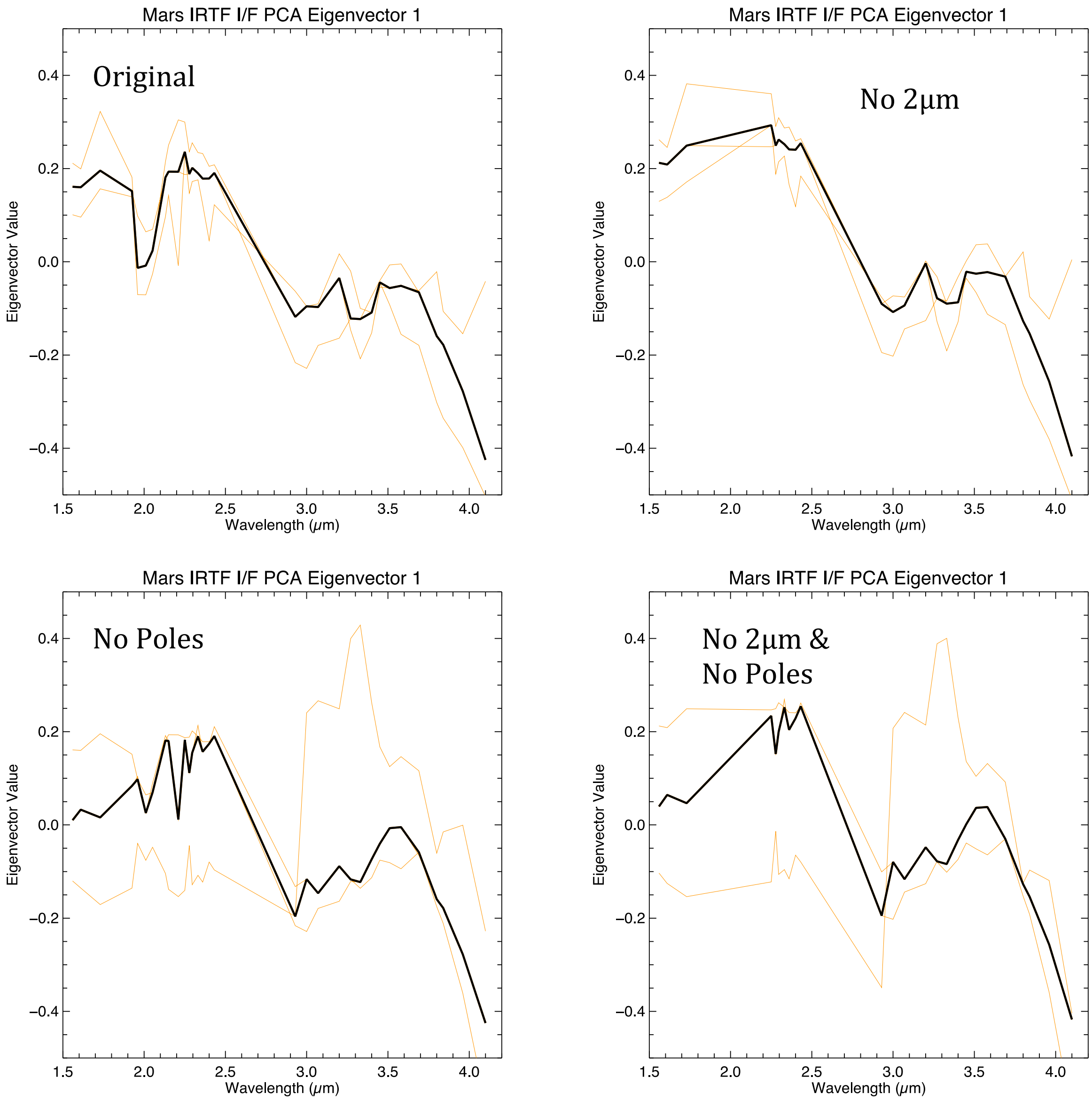

Figure 4. Results of the four PCA tests for eigenvector 1. The median line is in bold

As can be seen in figure 5, the best trends for EV2 (the geology eigenvector) show up in each PCA that excludes the Martian poles. In this case, removal of the 2 μm band does not make the constancy any better, but it also does not make it significantly worse.

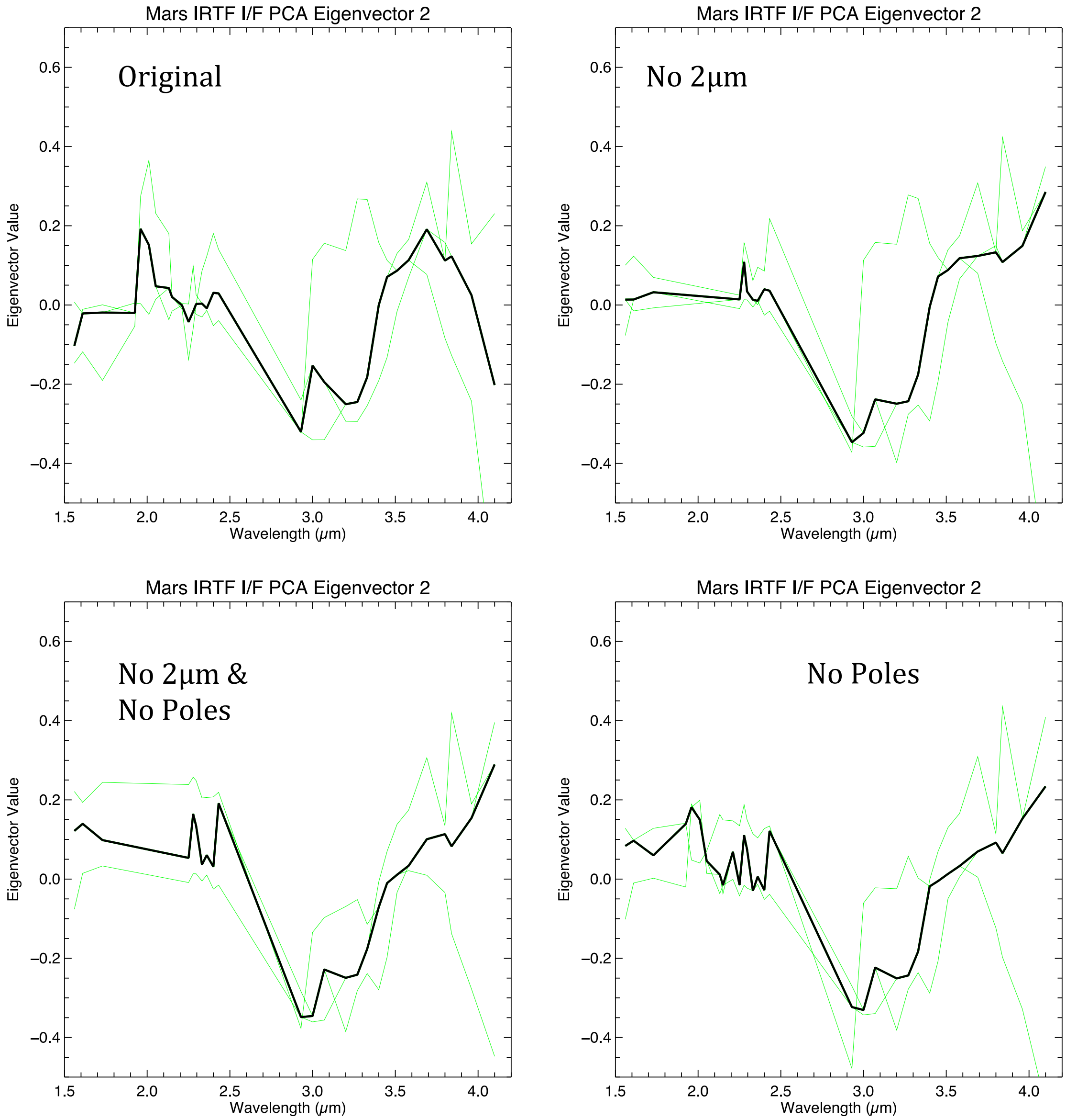

Figure 5. Results of the four PCA tests for eigenvector 2. The median line is in bold.

5. Conclusions

The exclusion of the Martian poles seems to yield more consistent trends for both EV0 and EV2, the brightness and geography eigenvectors. One explanation, at least for EV0, may be that the ground ice at the Martian poles may influence overall brightness due to its high reflectivity. EV1 seems to have more consistent trends with the use of either the original PCA or the PCA without the 2μm band. Although normalization is not as reliable as standard star calibration, it may be able to serve as a substitute method in the absence of a standard star or during less-than-ideal conditions on the planet. These results, however, came from a small sample of data (three sets), so it will be necessary to investigate further with more data before any solid conclusions can be made.

6. Acknowledgements

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

7. References

1. Kessler S., Testing Normalization Schemes of Uncalibratable NIR Spectral Images of Mars Using Principal Component Analysis, *this proceedings*
2. Pultrone M., Analysis of Martian Spectral Modeling by Spot-Spectra Normalization, *this proceedings*
3. Combe, J.-P., Le Mou´elic, S., Sotin, C., Gendrin, A., Mustard, J. F., Le Deit, L., Launeau, P., Bibring, J.-P., Gondet, B., Langevin, Y., Pinet, P., the OMEGA Science team (2008), Analysis of OMEGA/Mars Ex- press data hyperspectral data using a Multiple-Endmember Linear Spectral Unmixing Model (MELSUM): Methodology and first results, *Planet. Space Sci.*, *56*, 951–975.
4. Madeleine, J.-B., Forget, F., Spiga, A., Wolff, M. J., Montmessin, F., Vincendon, M., Jouglet, D., Gondet, B., Bibring, J.-P., Langevin, Y., Schmitt, B. (2012), Aphelion water-ice cloud mapping and property retrieval using the OMEGA imaging spectrometer onboard Mars Express, *Journal of Geophysical Research (Planets)*, *117*, E00J07.
5. 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.
6. Klassen, D. R., Bell, J. F. (2003), Radiance factor calibration of near-infrared spectral images of Mars, *Icarus*, *163*, 66–77.
7. Klassen, D. R., Bell, J. F., III (2001), Principal components analysis studies of martian clouds, In: AAS/Division for Planetary Sciences Meeting Abstracts #33, Volume 33 of *Bulletin of the American Astronomical Society*, pp. 1069–1070.
8. Moroz, V. I. (1964), The Infrared Spectrum of Mars (λ1.1—4.1μ)., *Soviet Astronomy*, *8*, 273–281.
9. Larson, H. P., Fink, U. (1972), Identification of Carbon Dioxide Frost on the Martian Polar Caps, *ApJ*, *171*, L91–L95.
10. Klassen, D. R. (2009), Principal components analysis of Mars in the near-infrared, *Icarus*, *204*, 32–47.