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

The Analysis of the Effects of Principle Component Analysis on Mars Image Mosaics

Sean D. Hoyt
Department of Physics & Astronomy
Rowan University
201 Mullica Hill Road
Glassboro, NJ 08028 USA

Faculty Advisor: Dr. David R. Klassen

Abstract

Our overall goal is to measure the water content of clouds in the Martian atmosphere as a function of time over diurnal, seasonal, and interannual scales using ground-based, near-infrared spectral images. To do this, the surface reflectance must be understood well enough so that its spectral signature can be removed. Principle Component Analysis (PCA) is used to reduce the dimensionality of the data in order to recover the smallest number of surface endmember spectra needed to model the surface reflectance. During the 1999 opposition, the angular size of Mars was so large that its projected size was larger than the camera detector. Therefore, the images of Mars were taken in four quadrants. Before PCA could be performed on the data, the images needed to be stitched together. PCA was then done on the mosaic images. The mosaic PCA results are compared to previous results, which show PCA eigenvectors are fairly consistent across all timescales. We present here the results of that task and show that the overall effect on the PCA of the residual mosaic stitch lines is minimal.

Keywords: Mars, PCA, Mosaics

1. Introduction

The overall goal of this research group is to measure the ice abundance in clouds on Mars over diurnal, seasonal, and interannual time-scales. To do this, spectral imaging data of Mars was collected over every opposition from 1995–2003; oppositions occur roughly every 25 months. During the 1999 opposition, Mars was sufficiently large that, if reduced in size to fit entirely in the frame, the light collected would saturate the detector pixels at even the shortest possible integration times throughout half of the spectrum. The solution to this problem was to image Mars in quarters (Figure 1).

An internal camera lens that changes the detector plate scale was used to enlarge Mars so that the light was sufficiently spread out across the chip so that no pixels would saturate. This made the image of Mars about 50% larger than the detector itself meaning that each quarter-image has a significant amount of overlap with the other three quarters.

The downside to using this technique is that the images would have to be stitched together as a post-processing, pre-analysis step. Considering the ubiquity of such a process in planetary science, it was believed at the time of data collection that this would be a “simple” operation. Unfortunately, it turned out that for the most part, the process of making image mosaics appears to be done in an ad hoc manner. Therefore, the goal of this specific project was to work out such a process for these data and to investigate the effect of any residual mosaicing artifacts in the images on the standard analyses.

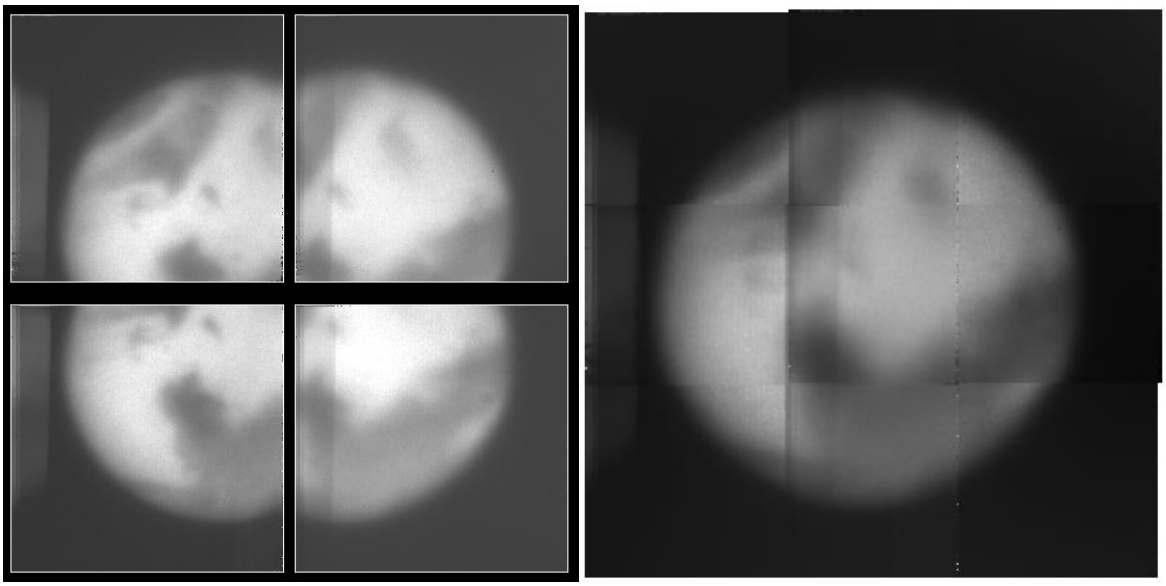

Figure 1. Mars was imaged in quadrants and run through a mosaic program

2. Data

The data used in this study were taken on 24 April 1999 at the NASA Infrared Telescope Facility (IRTF) atop Mauna Kea, Hawaii using the NSFCAM. NSFCAM has a circular variable filter (CVF) that let us select specific near-infrared (NIR) wavelengths. In the case of this data, the images were stepped through the filter at Nyquist sampling (half the filter spectral resolution, which for the CVF is $\Delta\lambda/\lambda \sim 1\%$). As a result, there were 105 mosaic quarters that needed stitching. Manually stitching this number of images would be an inefficient solution that could be facilitated by automating the process.

3. Making Mosaics

It was originally thought that finding a mosaic program to work with the data would be an easy task, however it turned out to be far more difficult. Although making mosaics from patchworks of images is a common affair in imaging—it is a basic function in even bare-bones camera software packages—the restriction of flux-preservation limits commercial application use. The goal of commercial software is to have a "pretty picture" so what happens to pixel values is a secondary consideration, so long is it looks good (enough) to the human eye. But with scientific data, pixel value *is* the data; whatever transformations one does to an image, it must not alter pixel values differently from one another. But even excluding commercial applications, there are plenty of scientific groups creating mosaics of images (e.g. Hubble Space Telescope and the Mars rovers). Even here, the science groups are not usually using the mosaic images; the mosaics are, again, for public consumption and so only have to be of the "pretty picture" variety. And for those groups that do create mosaics for scientific research, the techniques appear to have been created in ad hoc manners and created for their specific data acquired under very specific conditions.

After an extensive search, a small number of generalizable programs were found. Each was analyzed for both their ease of use, ease of modification to the data (or, rather, ease of modification of the data for use with the program), and overall results. In the end, a package called Montage[1] developed by NASA and Caltech was selected to use with the data.

Several programs were written in order to make use of the package. The first problem was to make the images compatible with the Montage program. Montage works by using the World Coordinate System (WCS) that is a

standard created for FITS images and stored in their headers. However, the data does not make use of this system so the data had to be modified to add this information to the headers of all the images.

Since operation must be done on an entire image cube—32 images in a set—the process needed to be automated to work on multiple sets of mosaic inputs. Montage works by having all the parts of the image that must be stitched together in a directory, then the user runs the Montage program, pointing it to that directory. After this, the final mosaic image is sent to a new directory. It was not practical to go through this process manually for each wavelength set in the image cube. The program that was written automatically takes four corresponding images that must be mosaiced, puts them into the directory, runs the Montage package, then takes the final mosaic from its output directly, and moves it into a new mosaics directory. This process repeats until all of the image sets have created a mosaic.

In order to run the data through the analysis programs, the images must be coregistered—so that a particular pixel coordinate in one image is the same location on Mars in all the images. This is complicated by the fact that Mars is rotating as the data were being collected. The only way to coregister them then is to remap each image, individually, to a cylindrical projection map. When Montage runs, it is unconcerned with the final size of the mosaic image, but the remapping program requires that each input image have the same image size (so that "center of image" has the same meaning for each one). Therefore a final program was written that automatically goes through all of the mosaiced images and trims them to equal size.

Finally, the image mosaics were calibrated to radiance factor and were ready for the analysis testing. The primary question was whether the mosaic process would have any negative effects on the subsequent analysis. The most prominent issue was the presence of residual "stitch lines" which are easily seen in the final mosaic image (see Figure 1).

4. Principle Component Analysis

The primary goal of retrieving water abundances in the clouds of Mars requires removal of the surface reflectance from the spectral data. The technique made use of to do this is principal component analysis (PCA). PCA is a technique that projects the data into a new (orthogonal) space defined by the variances in the data[2]. The new dimensions are ordered by their contribution to the total data variance. PCA is done by maximizing the variance/covariance matrix of the data, which becomes a standard eigenvalue-eigenvector problem. The eigenvalues of this matrix are then proportional to the variances of the data along the new dimensions and the eigenvectors form the new basis vectors. As can be seen in figure 2, over 95% of the total data variance can be explained by the first three dimensions—this means that there are really no more than 3 or 4 intrinsic dimensions that describe these data. Thus, PCA serves to reduce the dimensionality of the data from 32 down to 3–4. Additionally, since these new dimensions are related to intrinsic differences within the data, these new dimensions are "traits"that describe the data. And since the dimensions are orthogonal, these traits are also "orthogonal", meaning we can separate out various properties of the data.

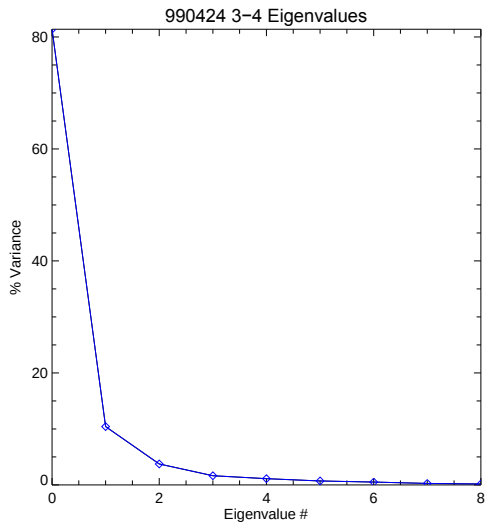

Figure 2. After first three eigenvalues the variance drops off to almost zero showing that only the first few eigenvectors are important

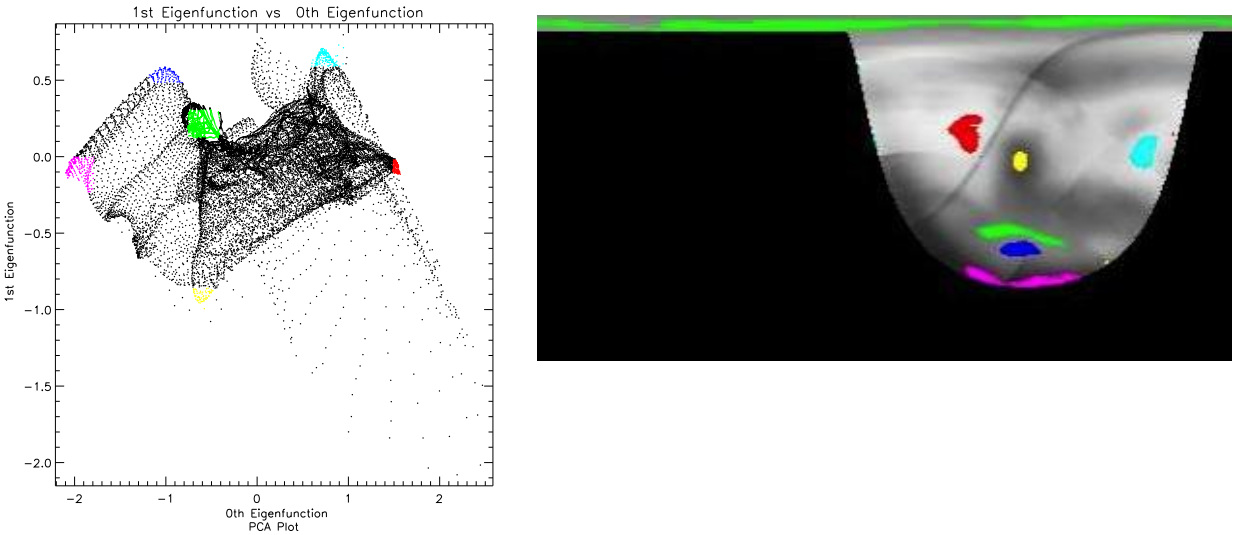

Figure 3. PCA maps areas of the Martian surface corresponding to different eigenvectors

Figure 3 shows a plot of the spectral image cube in the first two PCA dimensions with several different regions that were selected as regions of interest—typically because they are vertices or endpoints of the data cloud in the new dimensions. The red area marks a typical bright region, the yellow marks the darkest/warmest region, the green marks the polar region, the purple marks the dark/noise, and the cyan and blue mark areas usually covered by condensate clouds. This indicates that eigenvector 0 represents the trait of "overall near-infrared brightness" while eigenvector 1 represents the trait of "cold/icy vs warm". While it is not shown here, the typical interpretation, based on previous work[3], represents regional geology differences. Because ices are only seen in clouds (or ground ice in the polar regions), this shows that PCA can separate out ground spectral signatures.

4. Analyzing the Effects of Principle Component Analysis

From past analysis results, there are the eigenvectors shown in Figure 4 of non mosaic data, which allow for a calculation of median eigenvectors to compare to the mosaic eigenvectors.

To determine the effect of residual stitch lines on the Principle Component Analysis, the eigenvectors of the image mosaics are compared to the eigenvectors from previous work[3, 4, 5]. Figure 4 shows the eigenvector spectra from several dates and sets from previous work, as well as the median spectrum for each eigenvector. Figure 5 shows plots of the mosaic eigenvectors alongside the median eigenvectors spectra from the previous work.

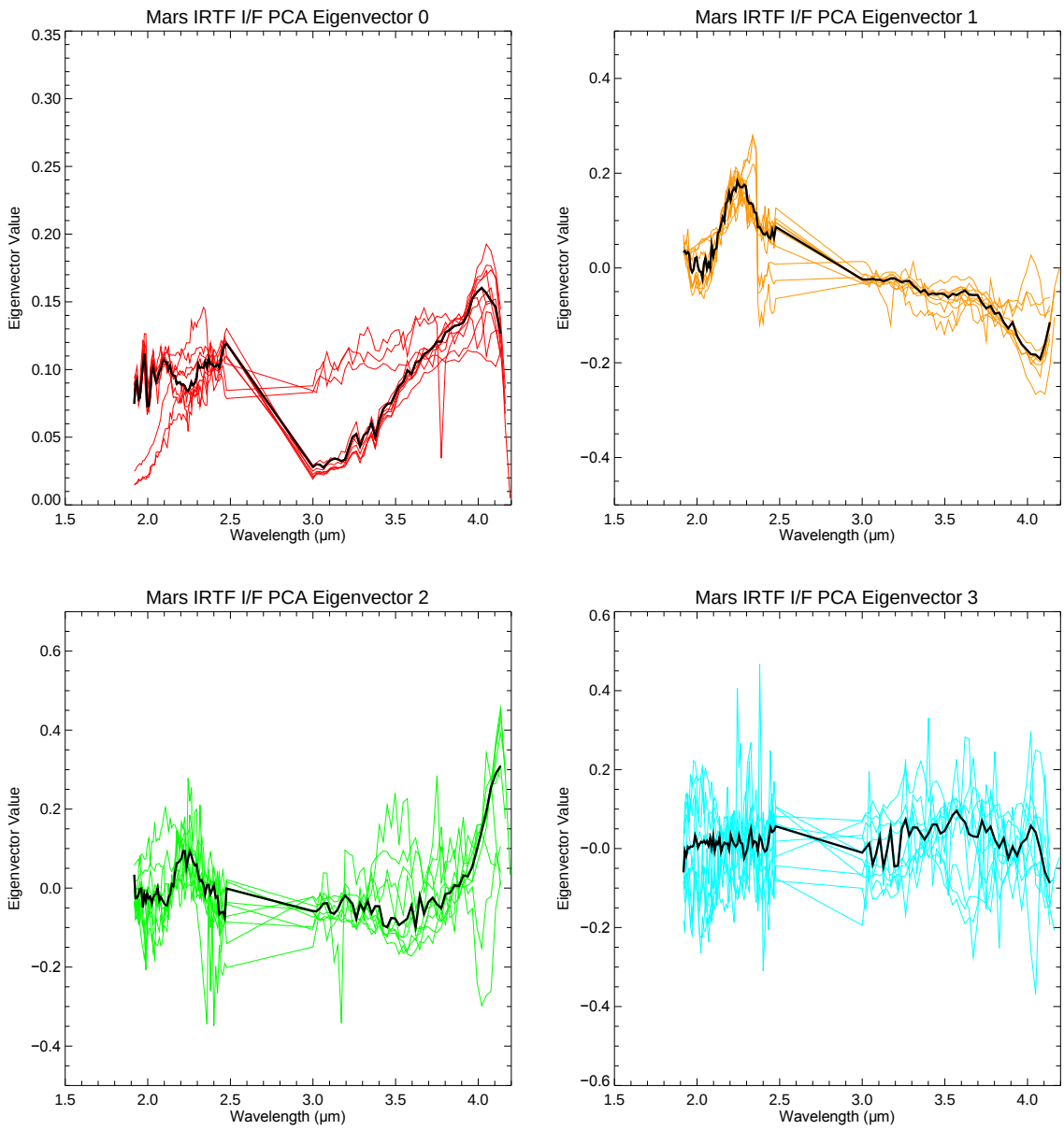

Figure 4. Median eigenvector plots

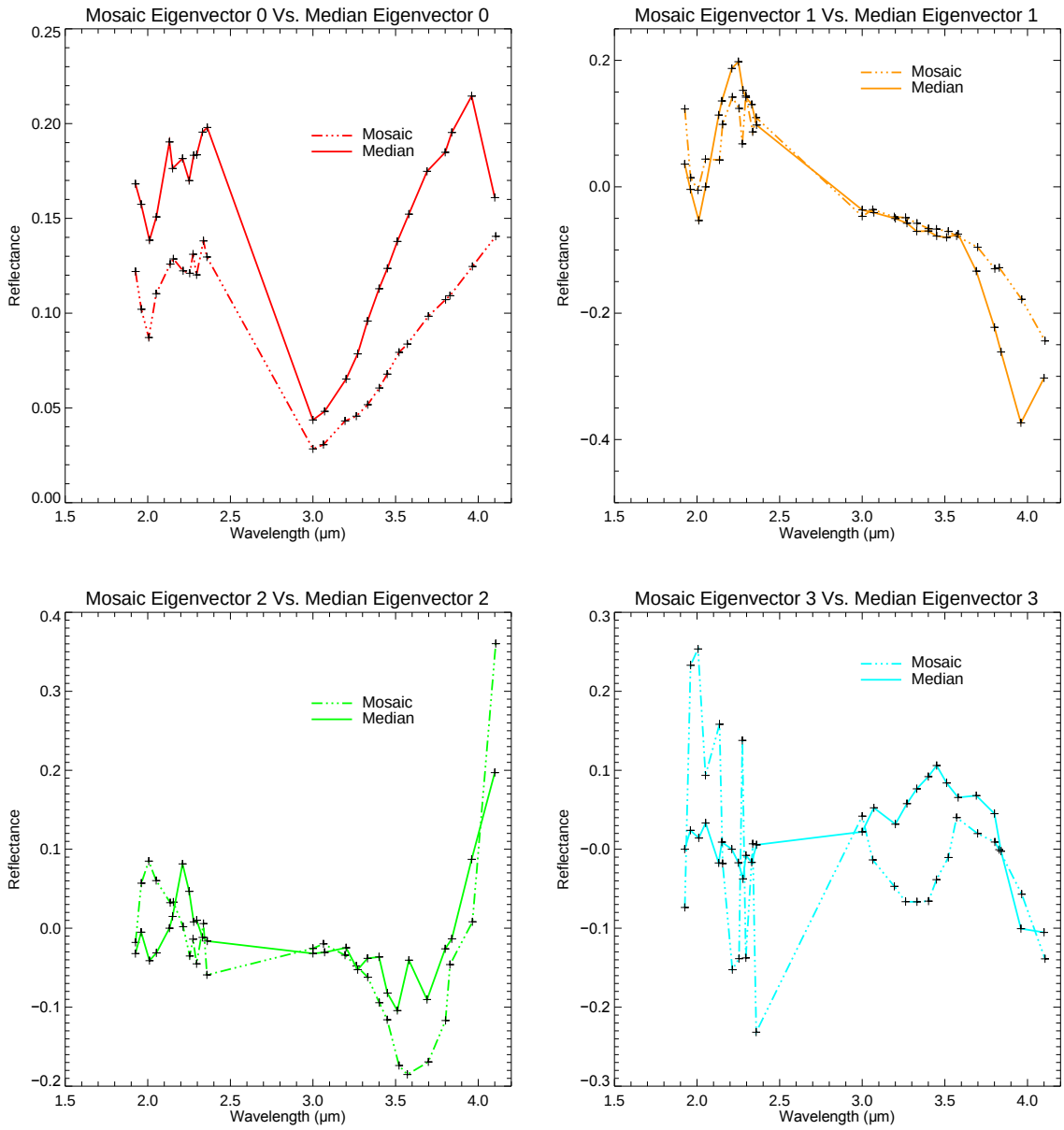

Figure 5. Mosaic eigenvectors compared to median eigenvectors

From these plots we see that there is a strong consistency between the mosaic eigenvectors and the median eigenvectors. The fact that the shapes are so similar indicates that the residual stitch lines do not appear to affect eigenvectors resulting from PCA. To make a more quantitative comparison, the root mean square (RMS) differences between the mosaic eigenvectors and the previous work median were found, as well as the average RMS differences between the previous work eigenvectors and their median. These RMS values are shown in Table 1.

Table 1. RMS differences between eigenvectors and the median eigenvector

Eigenvector	Mosaic RMS	Previous Work RMS
0	0.0528	0.985
1	0.0443	0.114
2	0.0557	0.163
3	0.102	0.169
4	0.0802	0.172
5	0.104	0.174

From the calculations, the RMS of the mosaic eigenvectors is less than the average RMS of the previous work which, again, indicates that the residual stitch lines do not affect the PCA results and the mosaic data is able to be used in future analyses, along with the non mosaic data.

5. Acknowledgments

This research made use of Montage, funded by the NASA Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and CalTech. Montage is maintained by the NASA/IPAC Infrared Science Archive. This work was supported in part by the NSF Planetary Astronomy and Research at Undergraduate Institutions programs (1008744)

6. References

1. Berriman, G. B., Good, J. C., Laity, A. C., Jacob, J. C., Katz, D. S., Deelman, E., Singh, G., Su, M.-H. (2007), Chapter 19: Web-based Tools - Montage: An Astronomical Image Mosaic Engine, In: Graham, M. J., Fitzpatrick, M. J., McGlynn, T. A. (Eds.), Astronomical Society of the Pacific Conference Series, Volume 382 of *Astronomical Society of the Pacific Conference Series*, pp. 179.

2. Salmon, G. (1927), *A Treatise on the Analytic Geometry of Three Dimensions*. Chelsea Publishing Company, New York.

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

4. Klassen, D. R., Wark, T. J., Cugliotta, C. G. (2005), Principle components analysis of Mars at the 2003 opposition, In: AAS/Division for Planetary Sciences Meeting Abstracts #37, Volume 37 of *Bulletin of the American Astronomical Society*, pp. 693–4.

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.