

Analysis of a Series of Images with PCA and PARAFAC

Barry M. Wise and Paul Geladi

The Images



The Data

These images were taken by the Eigenvector Research web camera on March 9, 2000 in half hour intervals from 11:33 to 13:33. Each of the 5 images is 240 by 352 pixels with three colors (rgb), thus the original data array is 240 by 352 by 3 by 5. Because the images would not be expected to be bilinear in the image plane, application of PARAFAC model directly would not be "natural."

Analysis with PCA

For the PCA the images were matricized to 84,480 (240*352) by 15 (3*5). In this arrangement all of the data for each pixel (all the rgb values over the 5 images) are a single sample in the analysis. The variance captured by the PCA model is given in Table 1. The loadings and scores in the image plane are shown below, along with the residuals based on a 2 PC model.

Table 1. Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	3.30e+04	89.52	89.52
2	2.71e+03	7.35	96.86
3	5.59e+02	1.52	98.38
4	1.92e+02	0.52	98.90
5	1.88e+02	0.51	99.41

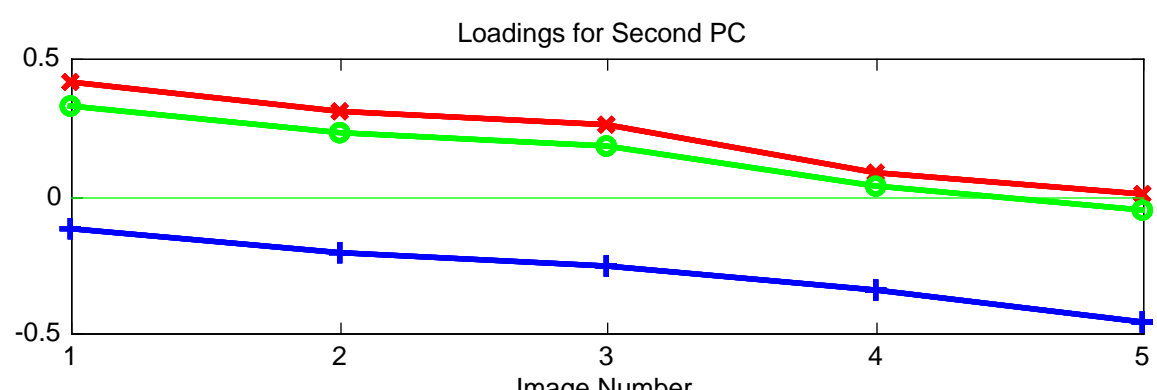
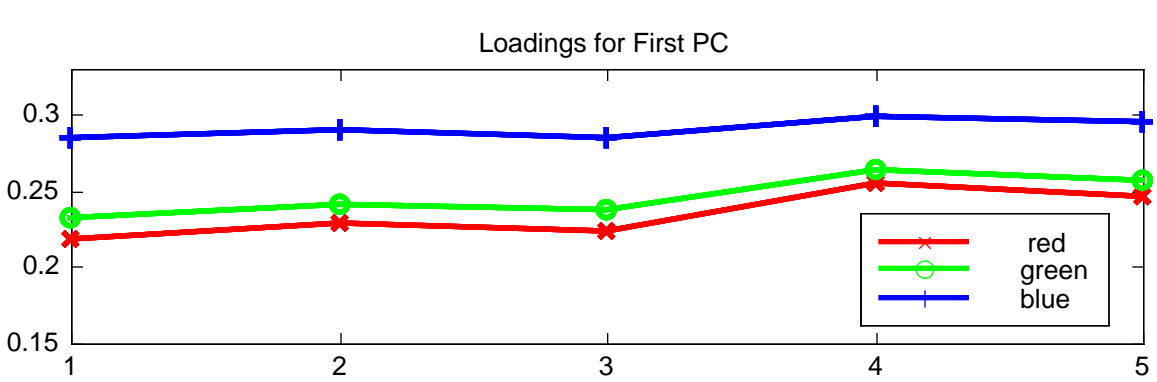
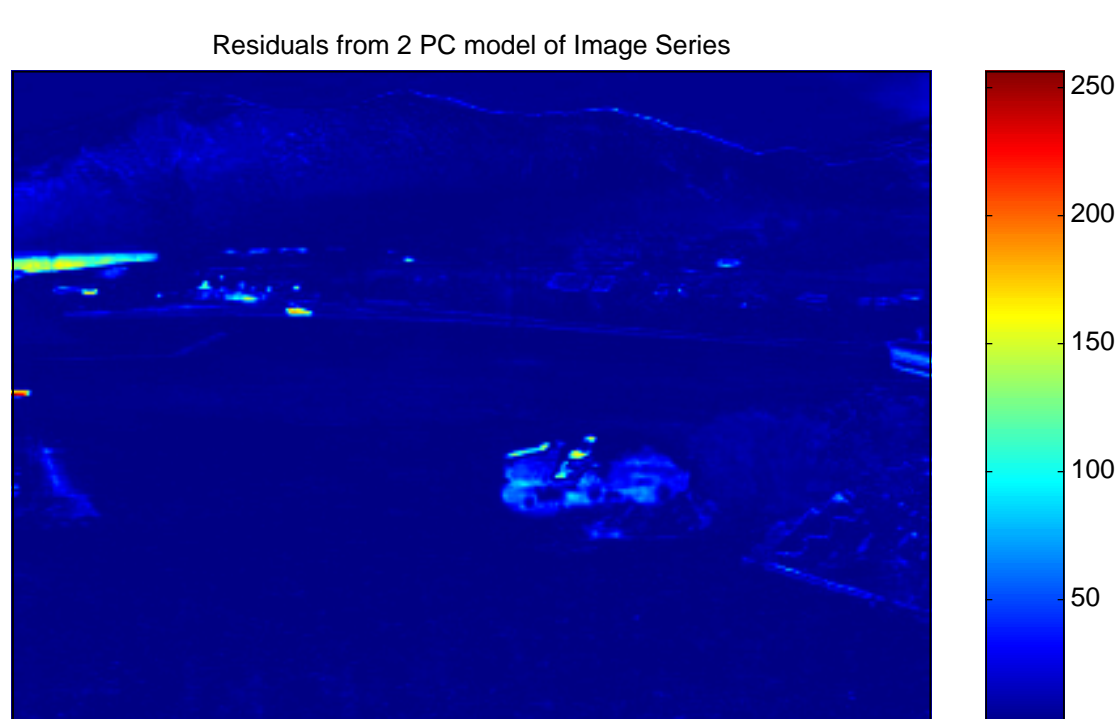
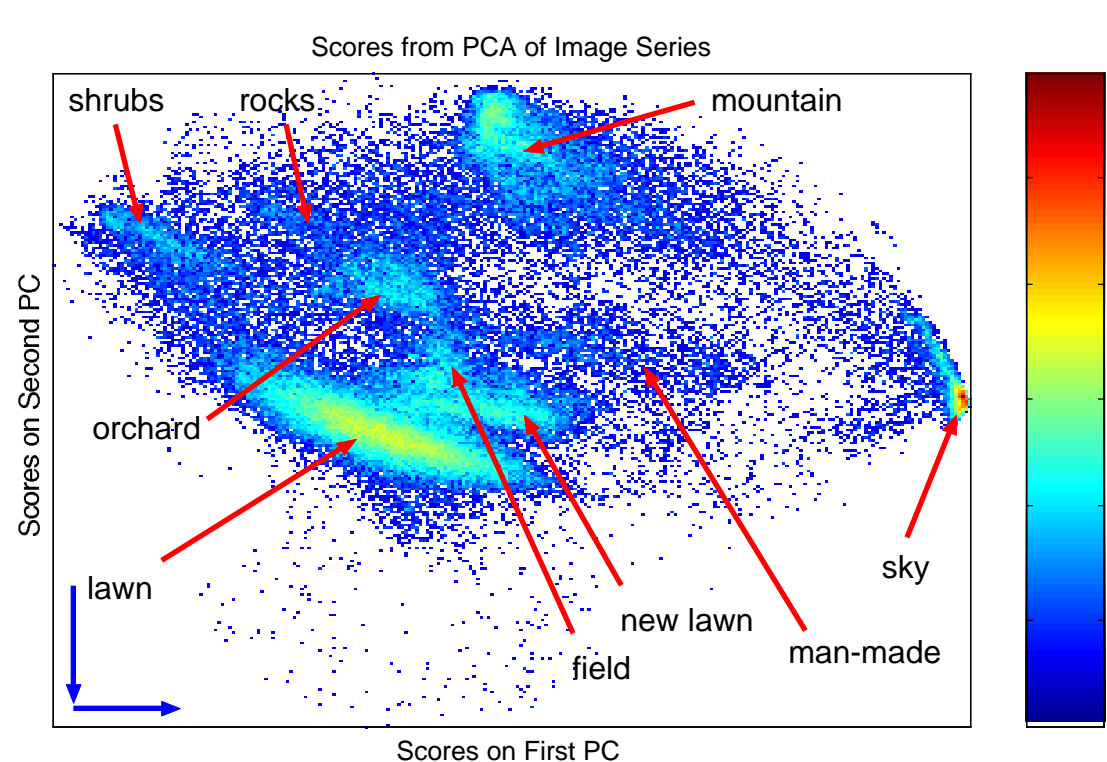
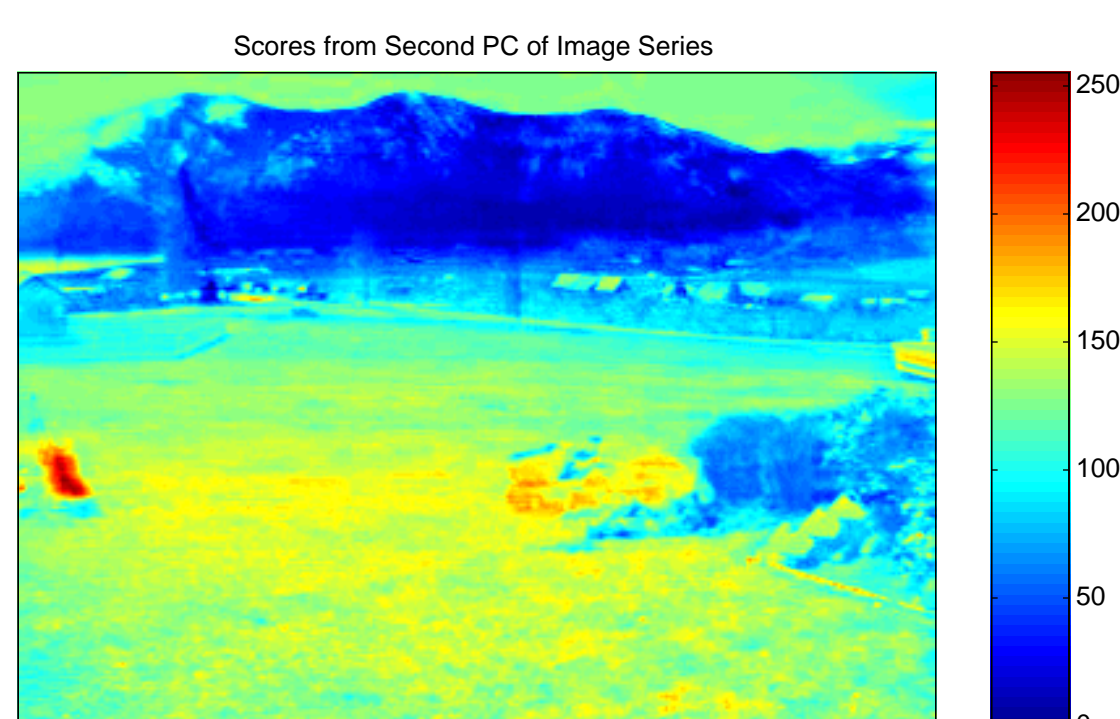
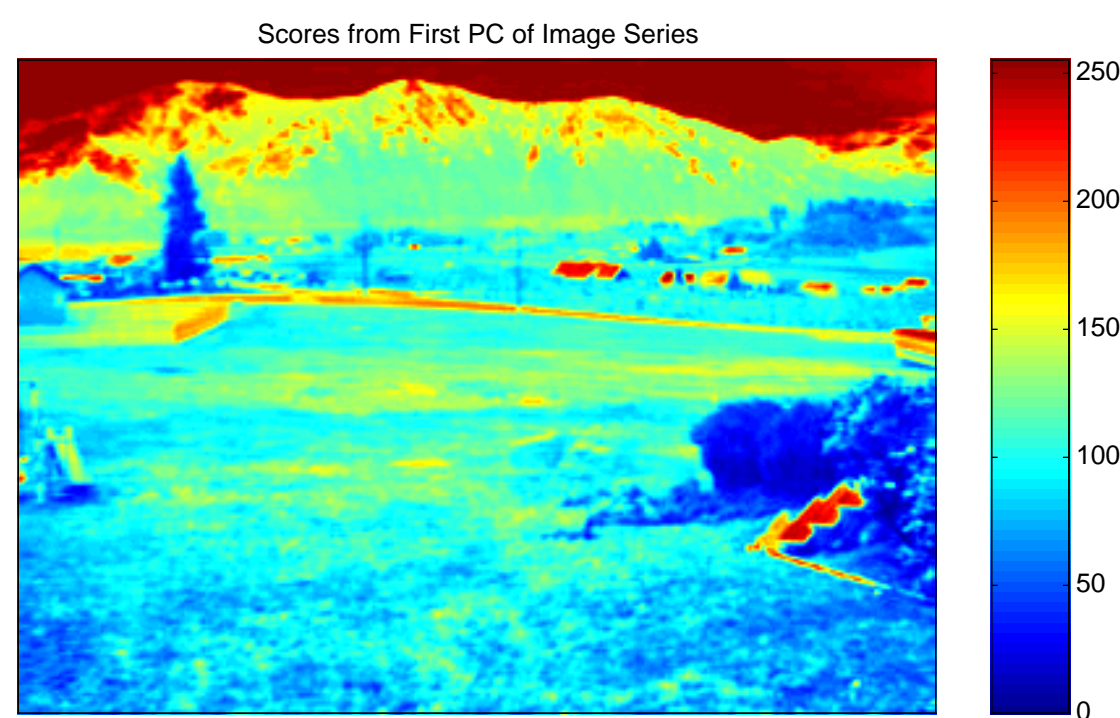
1

2

3

4

5

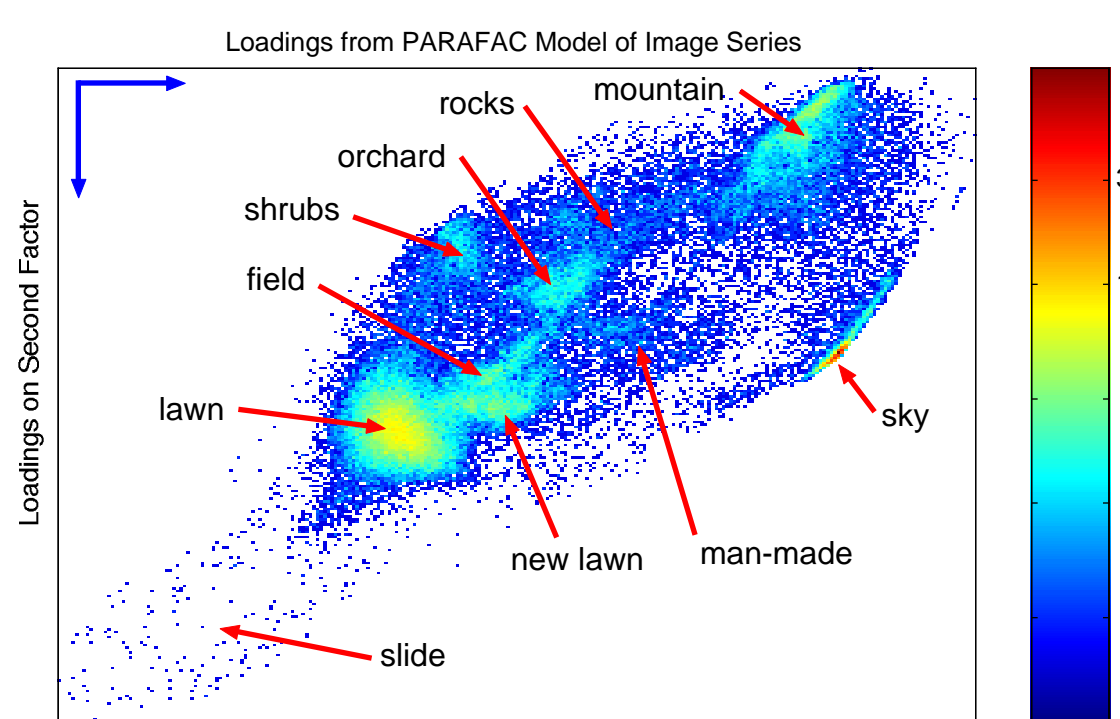
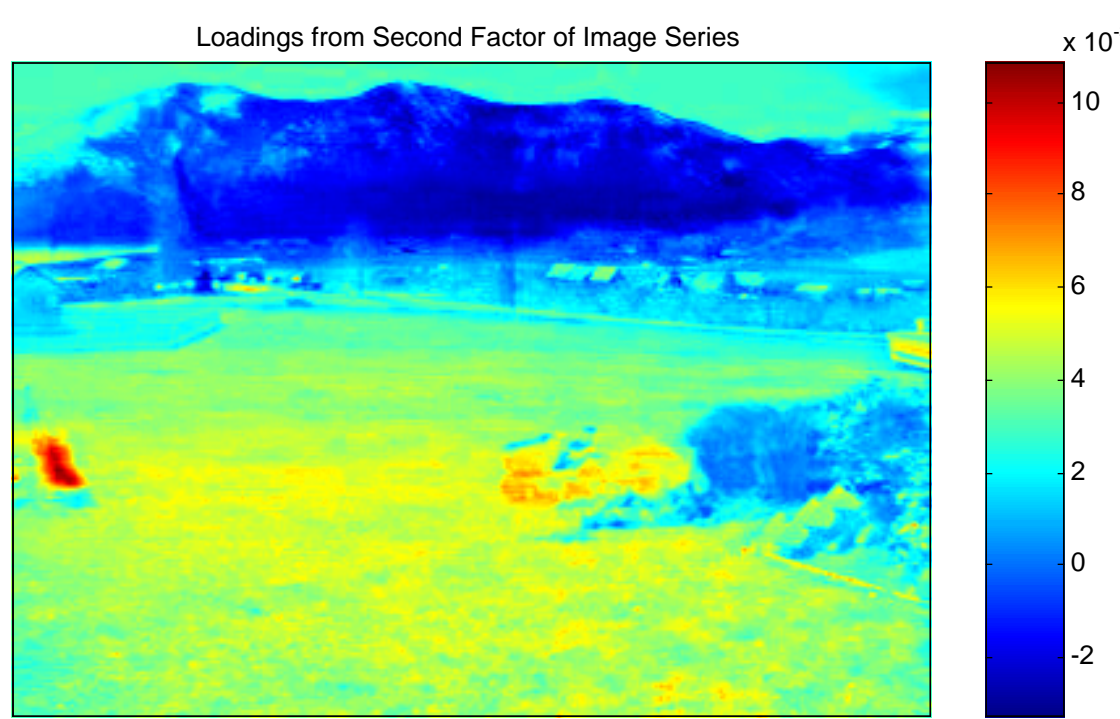
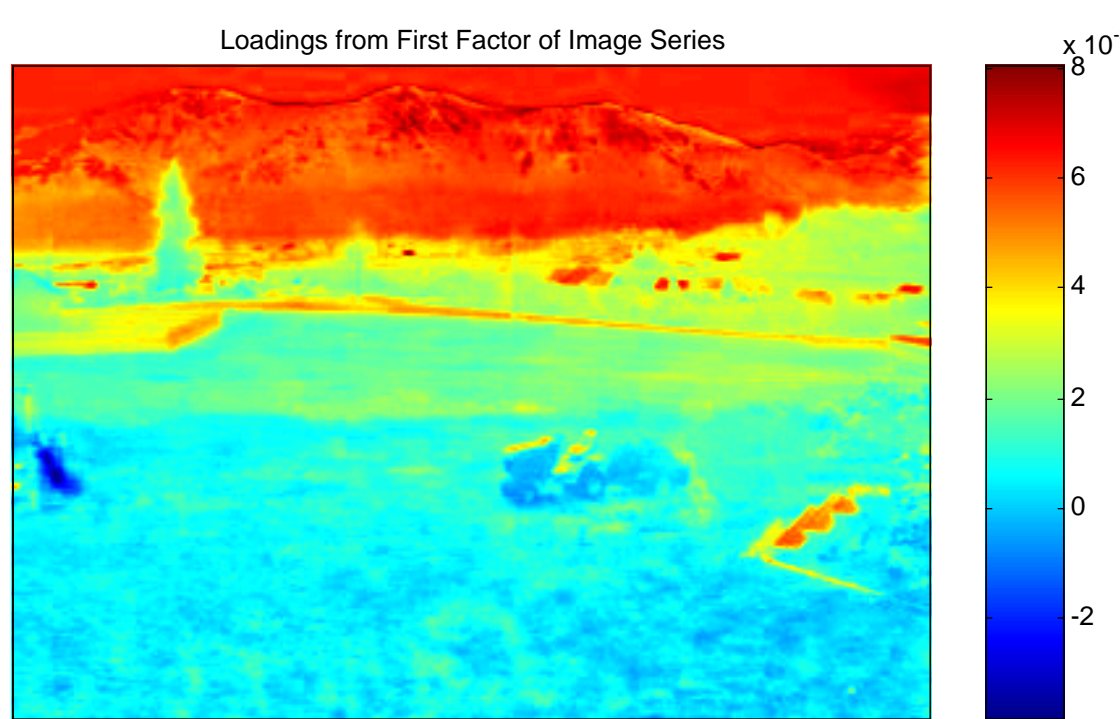
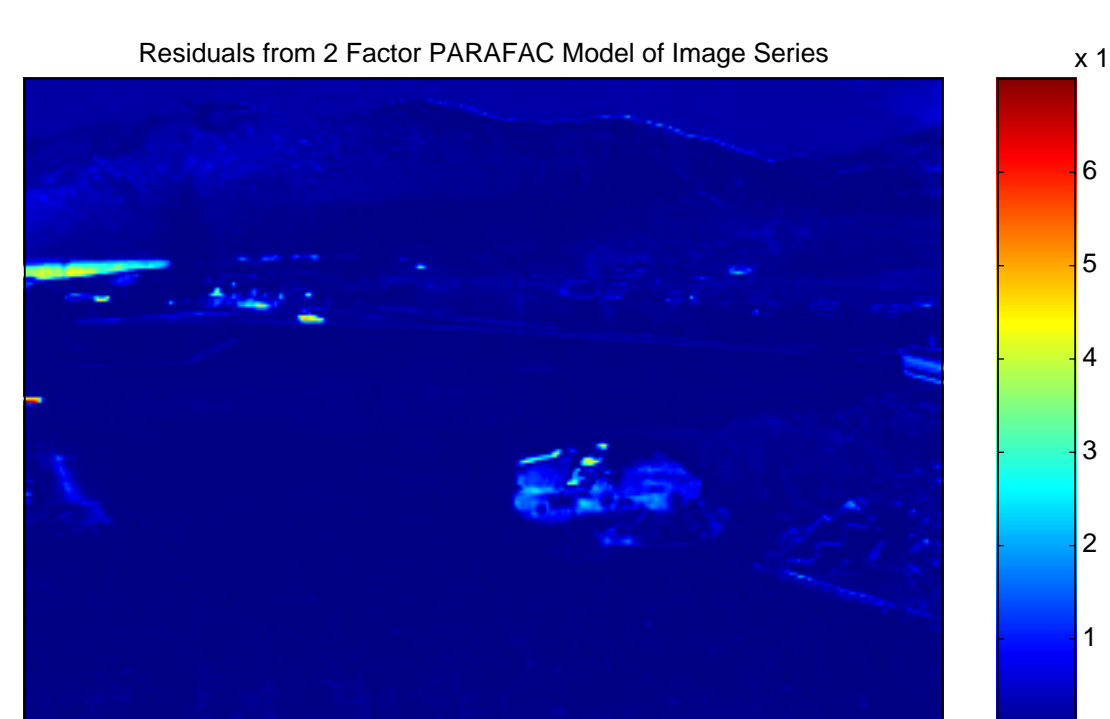
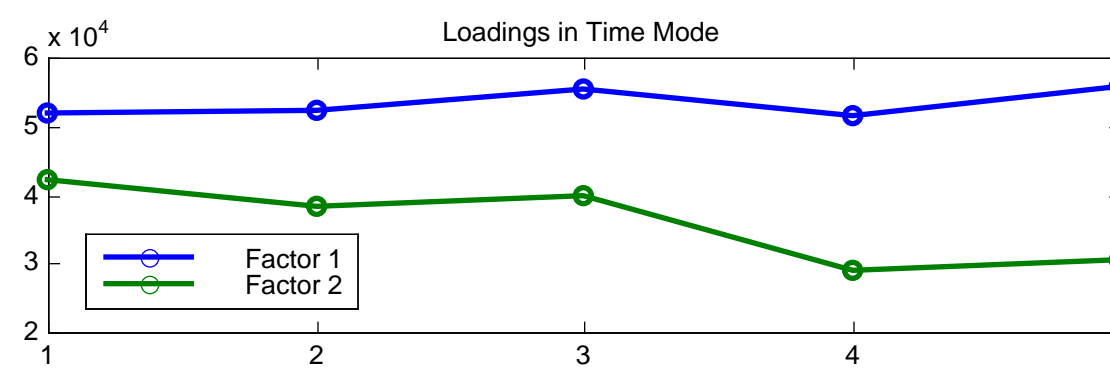
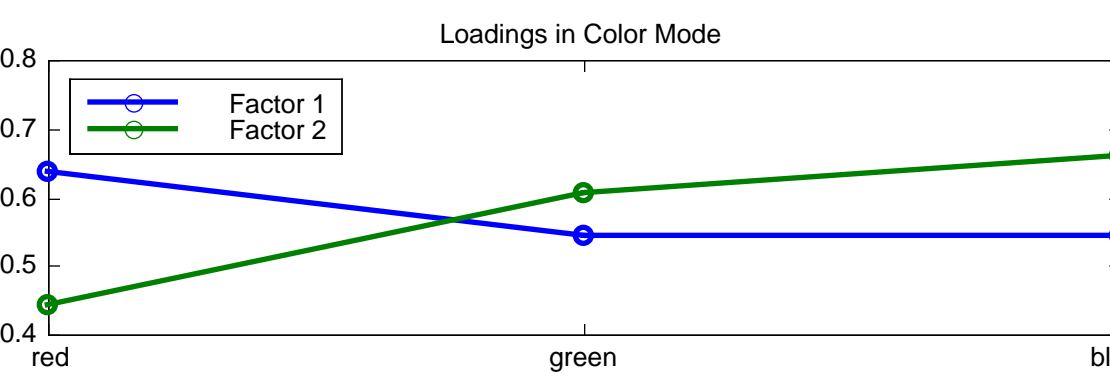


PCA Results

Different types of vegetation and other landscape elements cluster in the scatter plot of the scores. The residual image highlights elements that are not common to all images such as the lawn tractor (image 5), a car moving along the road (image 2), and changes in the surface of the lake. The loadings in the convolved time and color modes are difficult to interpret. It appears that the first factor captures overall brightness. The second factor has negative loadings for all of the blue variables and mostly positive for red and green. A trend with time is apparent.

Analysis with PARAFAC

For the PCA the images were matricized to 84,480 (240*352) by 3 by 5. In this arrangement the data would be expected to be approximately tri-linear. A two factor model captured 99.6% of the total sum of squares. The loadings in the image plane, color mode and time mode are shown, along with the residuals based on a 2 factor model.



PARAFAC Results

The PARAFAC results are very similar to the PCA results. Different elements of the image series cluster similarly and the residual image is nearly identical. The main difference is that the loadings in the time and color mode are separate which greatly simplifies the interpretation. The first factor is higher in red than green and blue and is approximately constant in time. Thus, in the scatter plot, movement to the right is associated with elements that are brighter, particularly in red, and relatively constant in time. The second factor is strongest in blue and depleted in red, and shows an overall decrease with time. Thus movement down in the plot is associated with elements decrease in brightness over time, particularly in their blue content. There are also fewer parameters in the color and time modes for the PARAFAC model compared to the PCA model (16 vs. 30).

Conclusion

The PARAFAC and PCA models give very similar results. The main advantage of PARAFAC is ease of interpretation and parsimony. The main advantage of PCA is computation time (18s versus about an hour).