Principal Component Analysis of Solar Spectral Irradiance Measurements

M. Rabbette and P. Pilewskie
NASA-Ames Research Center
Moffett Field, California

Introduction

The National Aeronautical and Space Administration (NASA) Ames Research Center Solar Spectral Flux Radiometer (SSFR) was deployed at the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) Cloud and Radiation Testbed (CART) facility during the 1997 Shortwave Intensive Observation Period (SWIOP). It measured the downwelling solar irradiance in the 360-nm to 2500-nm solar spectral region. Over 7000 spectra were acquired between September 17 and October 5 in atmospheric conditions that varied from pristine clear to very thick cloud, accompanied by a five-fold change in column-integrated water vapor (1 cm to 5 cm).

The primary focus of this paper is on the multivariate data analysis of the retrieved SSFR spectra, fundamental to all aspects of our research. Complications arising from spectral mixing make it important to discriminate individual components (coming from several sources) within the mixed signal. Principal Component Analysis (PCA) is used to determine the number of individual pieces of information (independent variables) that exist in the spectra. The technique can be summarized as a method to reduce a large set of measurement variables with complex interrelationships to a smaller set of uncorrelated variables, i.e., the principal components (PCs). Using the PCA technique, it is possible to determine the smallest number of parameters needed to characterize an irradiance spectrum (combinations of factors such as aerosol, cloud and molecular scattering and absorption, solar angle, surface albedo, etc.) and to establish in which regions of the spectrum these variables are strongest.

Instrument and Data Processing

The SSFR is sensitive to radiant energy between 300 nm and 2500 nm with varying resolution from 5 nm to 20 nm. The SSFR uses two thermo-electrically cooled solid-state detectors: Si for visible wavelengths and InGaAs for the near-infrared. The system has two modes of operation: 1) hemispheric (cosine response diffuser) or 2) narrow 1 mrad field of view (lens/mirror assembly). The SSFR is calibrated for wavelength, angular response, and absolute spectral irradiance. The absolute calibration was carried out using both a standard lamp (LI-COR system in the field) and a calibrated integrating sphere. The radiometer has a root mean square (rms) accuracy of 3%-5% and a precision of 0.5%.

In this paper, we present the PCA of the visible and near-infrared region of the spectra between 360 nm and 1000 nm. Interactive Data Language (IDL) procedures were written to incorporate the PCOMP and Singular Value Decomposition (SVD) routines used to compute the PCs of an m-column, n-row array (A), where m is the number of variables (flux at various wavelengths) and n is the number of observations (spectra). The Fall 1997 SWIOP input array is 640 (360 nm to 1000 nm range) by 7200 (total number of spectra) for which the eigenvalues and eigenvectors are computed. For further details on the physical and geometric aspects of vectors and matrices and principal components, see Twomey (1977).

Results

Figure 1 is a three-dimensional (3-D) surface plot (380 nm to 1000 nm) for clear-sky conditions during the entire 18-day experiment. The oxygen band at 762 nm is evident, as well as several water vapor bands in the visible and near-infrared, including the σ bands at 942 nm. At shorter wavelengths, the structure is due to Fraunhofer lines in the solar atmosphere. The daily repeatability of the maximum irradiance level near 500 nm is evidence of the stability of the instrument.

Figure 2 displays the first four PCs of the SWIOP visible solar irradiance data set. The first component, PC1, is the overall average variance. PC1 includes the solar angular response and systematic instrument noise. Each subsequent PC is orthogonal to all previous PCs and, therefore, is independent, in a least squares sense, of the variability due

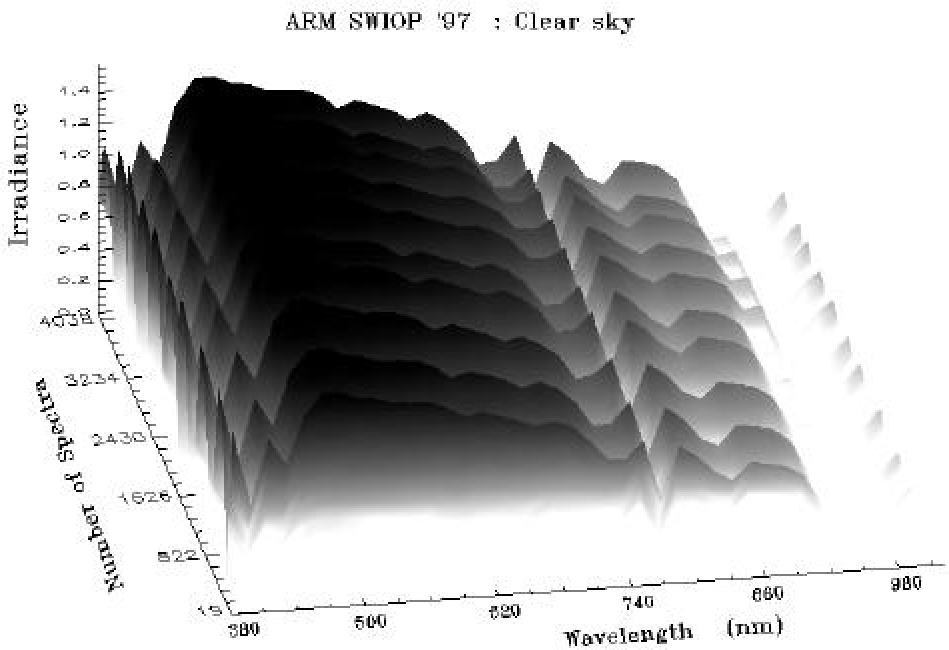

Figure 1. Clear sky 3-D surface plot (380 nm to 1000 nm) of entire SWIOP period from September 17 through October 5, 1997.

to the features accounted for in all previous PCs. The second component, PC2, has absolute amplitude maxima in the water absorption bands, but there is also a broad visible feature associated with this component. The third component has a broad spectral feature heavily weighted in the visible region, but of opposite sign to the previous component, as expected. The fourth component proved very useful as a diagnostic tool for these SSFR spectra. The relatively large weighting at 700 nm was the result of saturation in just a few of the 7200 sample spectra. Without the PC analysis, isolating these saturated spectra would have proven difficult.

Using the theory of SVD the *m*-column, *n*-row input array (A) can be expressed as A=YwX where X is the matrix of eigenvectors (orthogonal functions of *space*), Y is a matrix of eigenvectors (orthogonal functions of *time*), and w contains the eigenvalues. An eigenvector may display a predictable temporal behavior such as trends and cycles or it may simply display random variations in time. Therefore, the above matrix Y can be very useful in the physical interpretation of the PCs. Figure 3 is the time series for PC1. The diurnal influence of the solar angular response is obvious, especially on the clear days.

Summary

Although the independent variables are complex linear combinations of the SSFR channels, we can make the following general conclusions based on the results of the multivariate analysis:

- More than 99% of the information is contained in only four components.

- There is a broad visible component correlated with water vapor and another (opposing) component independent of water vapor.

- PCA is a useful diagnostic tool for SSFR spectra.

The main purpose of multivariate analysis is to derive easily interpretable independent variables (PCs). However, it is often difficult to extract an accurate physical interpretation from these components. The application of a simple structure rotation to the PC output loading matrix will uniquely cluster and simplify the variables in the PC loading

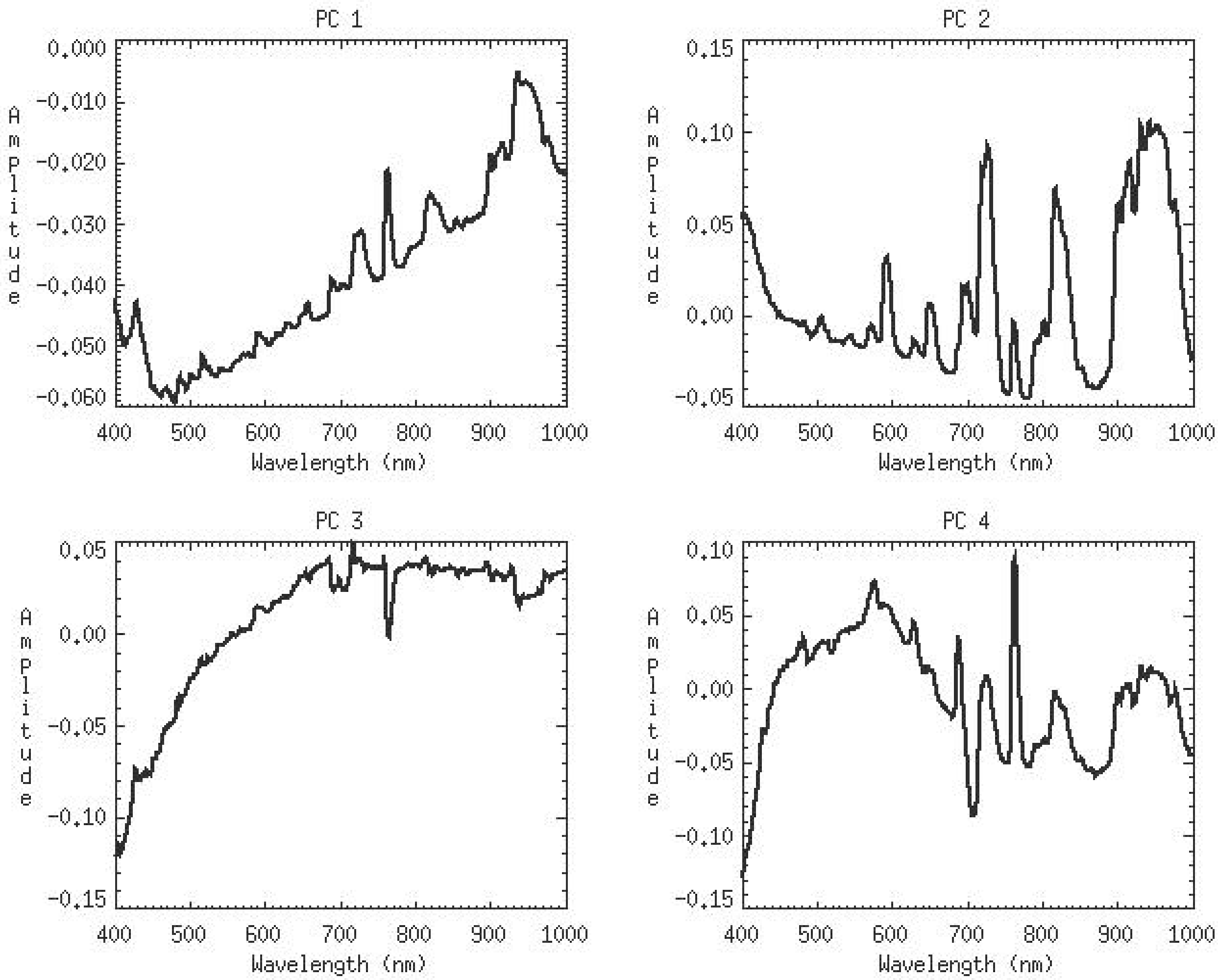

Figure 2. The first four PCs of the SSFR array.

matrix. Richman (1986) compares and discusses various available analytic simple structure rotations (e.g., orthogonal rotations, oblique rotations, and Procrustes rotations) and finds that each brings out different relationships among the data. We are in the process of computing an orthogonal rotation of the PC-loading matrix using a generalized orthomax criterion (including quartimax, varimax, and equamax rotations). Comparisons of these outputs to alternative rotational solutions as well as to the unrotated PCs should yield results that are physically meaningful. Future work will also incorporate comparisons between these PCA results and synthetic (computed) PCA results.

Acknowledgments

We wish to thank Larry Pezzolo and Warren Gore for engineering and technical assistance. This work was supported under U.S. Department of Energy Agreement ITF 353796-A-Q5.

References

Richman, M. B., 1986: Rotation of Principal Components. *J. of Climatology*, **6**, 293.

Twomey, S., 1977: Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements. Dover Publications, Inc., New York.

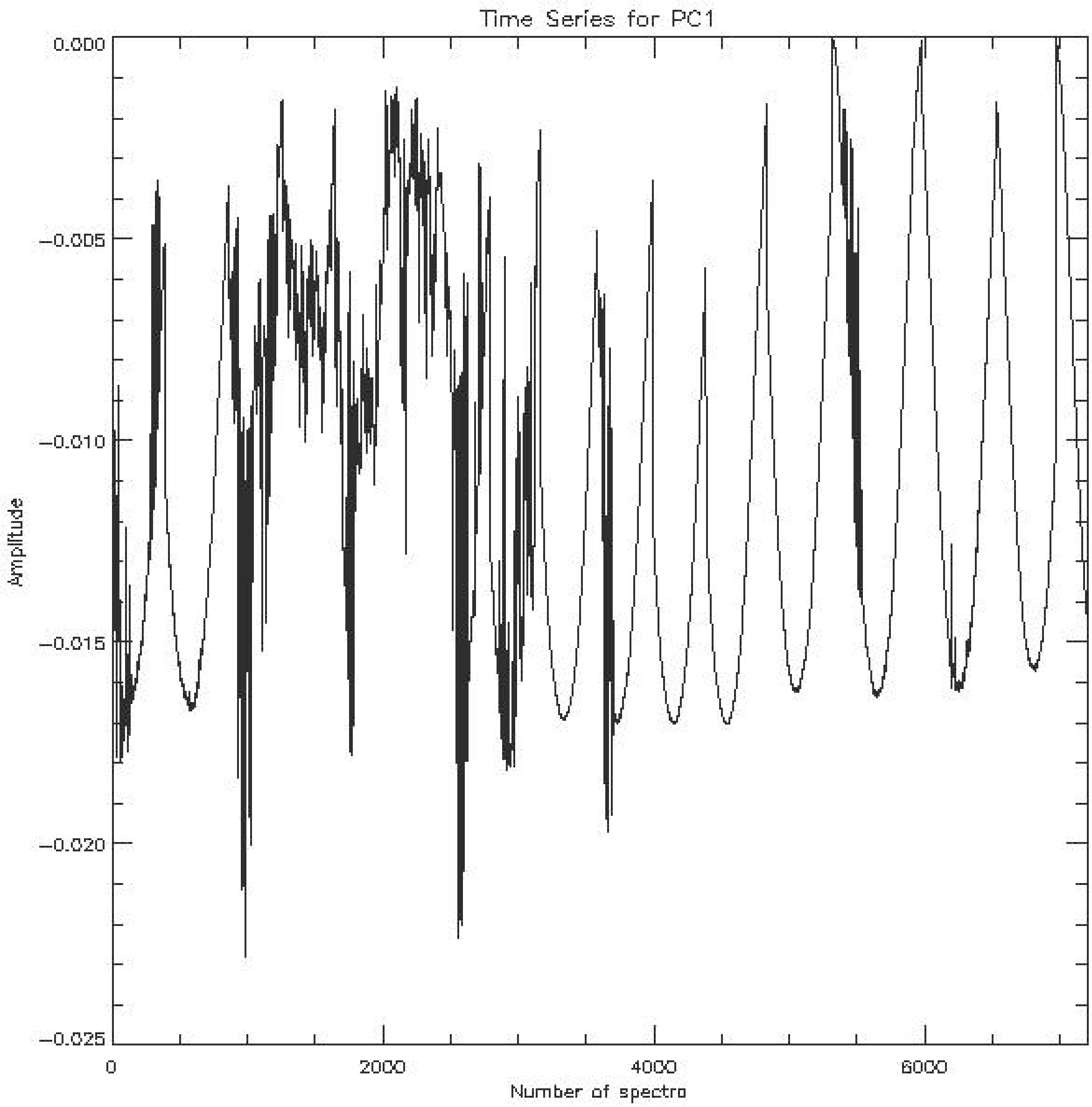

Figure 3. Time series for the first PC.