

**Quantifying the Variability of Hyperspectral
Earth-reflected Radiation for Climate Studies**

by

Yolanda Roberts

B.S., Cornell University, 2007

M.S., University of Colorado at Boulder, 2011

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Atmospheric and Oceanic Science

2012

This thesis entitled:
Quantifying the Variability of Hyperspectral Earth-reflected Radiation for Climate Studies
written by Yolanda Roberts
has been approved for the Department of Atmospheric and Oceanic Science

Dr. Peter Pilewskie

Dr. Sebastian Schmidt

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Roberts, Yolanda (Ph.D., ATOC)

Quantifying the Variability of Hyperspectral Earth-reflected Radiation for Climate Studies

Thesis directed by Dr. Peter Pilewskie

An accurate assessment of the Earth's energy budget is essential to understanding how the Earth's climate is changing and what processes and feedbacks are causing those changes. This is difficult to achieve, in part, because reflected solar irradiance, and therefore albedo, is a challenging quantity to measure from space with sufficient accuracy to monitor climate changes. An alternative to irradiance or albedo is directly measured spectral radiance, which provides information about the Earth's atmospheric composition and surface properties that impact albedo variability. We have applied multivariate spectral decomposition techniques, such as principal component analysis (PCA), to quantify the variability of Earth-reflected hyperspectral solar radiance measured by the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) aboard ENVISAT. Using multivariate analysis we explored the potential for directly measured hyperspectral Earth-reflected solar radiance to provide sufficient information to study changes in Earth's climate based on the quantified variability of the data. The spectral signatures of the principal components (PCs) reveal that clouds, water vapor, vegetation, and sea ice are among the physical variables that explain the largest fraction of the SCIAMACHY data variance. The extraction of the spectral, spatial, and temporal variability in reflected shortwave hyperspectral radiance using multivariate analysis provides an alternate and complementary approach to applying inverse methods to space-based observations for climate studies.

Observation System Simulation Experiments (OSSEs) have been used to simulate solar radiation measurements during the twenty-first century for the NASA Climate and Absolute Radiance and Refractivity Observatory (CLARREO) hyperspectral shortwave instrument being designed. Comparing the spectral shapes of the OSSE and SCIAMACHY PCs shows

that the OSSE has a similar variance distribution to that observed by SCIAMACHY. We developed a quantitative comparison technique to quantify the degree to which the OSSE reproduces the variability within Earth's climate system relative to observations. These comparisons showed that the OSSE spectral variability is close to that observed by SCIAMACHY. In addition, for the first time, the near-decadal temporal variability of observed reflectance measured between 2002 and 2010 was quantified; the variance drivers in the nearly decadal variability of SCIAMACHY measurements exhibited temporal signals of physical variables such as the location of the Intertropical Convergence Zone and the annual cycle of the cryosphere. The intersection also allowed for the direct comparison between the temporal variability of SCIAMACHY and OSSE reflectance at the beginning of the twenty-first century. Finally, we quantified the centennial variability of OSSE output during the twenty-first century, demonstrating that the reflectance spectra simulated from the A2 emission scenario model output exhibited secular trends over the simulation period. Applying the multivariate techniques presented in this thesis to evaluate the OSSE centennial variability enables the development of trend detection methods to further study the temporal variability of reflected solar radiation.

Dedication

To my mother and sister for their love, support, and endless words of encouragement. And to the memory of my maternal grandfather for playing such an important role in my early education and personal development, in particular by teaching me how to read.

Acknowledgements

I am so thankful to all the people who have helped me during my graduate school career, particularly to the Atmospheric Radiation Group at CU: Peter Pilewskie, Sebastian Schmidt, Bruce Kindel, Odele Coddington, Patrick McBride, Scott Kittelman, Michael King, Samuel LeBlanc, Shi Song, and Andrew Kren. Your questions, suggestions, support, thirst for answers to intriguing questions, and desire for challenging and fun conversations have been so important to my development as a scientist and a person.

I also extend my appreciation to Constantine Lukashin for his time and patience in helping me to access the data needed to complete this thesis and become accustomed with the NASA Langley Research Center CLARREO computer.

To William Collins and Daniel Feldman, I extend my appreciation for their support in this work and for their generosity with sharing their results.

Contents

Chapter	
1	Introduction 1
2	Evaluating the Observed Variability in Hyperspectral Earth-reflected Solar Radiance 10
2.1	Introduction 11
2.2	SCIAMACHY 15
2.3	Multivariate Analysis Techniques 16
2.3.1	Historical Applications of Principal Component Analysis 16
2.3.2	Principal Component Analysis Description 18
2.3.3	Time Series PCA: Singular Spectrum Analysis 24
2.4	Hyperspectral Reflected Radiance Variability 27
2.4.1	Global Seasonal Variability 29
2.4.2	Hemispherical Seasonal Variability 43
2.4.3	Zonally Averaged Regional Variability 48
2.4.4	Case Study: Arctic Ocean Variability 51
2.5	Summary 55
3	Quantitative Comparison of the Variability in Observed and Simulated Shortwave Reflectance 59
3.1	Introduction 60
3.2	Data 63

3.2.1	Observed Reflectance: SCIAMACHY Measurements	63
3.2.2	Simulated Reflectance: Observation System Simulation Experiments	66
3.3	Quantitative Multivariate Methods	66
3.3.1	Principal Component Analysis	66
3.3.2	Quantitative Comparison Description	69
3.4	Hyperspectral Reflectance Variability	75
3.4.1	SCIAMACHY and OSSE Data Processing	75
3.4.2	Spectral Reflectance Variability	76
3.4.3	Quantitative Subspaces Comparison	81
3.5	Conclusions and Future Work	87
4	Temporal Variability of Observed and Simulated Hyperspectral Reflectance	90
4.1	Introduction	90
4.2	Data	93
4.2.1	Observed Reflectance: SCIAMACHY	93
4.2.2	Simulated Reflectance: OSSE Output	94
4.3	Multivariate Methods	95
4.3.1	Principal Component Analysis	95
4.3.2	Intersection Application	96
4.4	Results	97
4.4.1	SCIAMACHY Temporal Spectral Variability	98
4.4.2	Comparing Observed and Simulated Temporal Variability	113
4.4.3	Centennial Variability of OSSE Reflectance	118
4.5	Summary	122
5	Summary and Outlook	125

Bibliography

Tables

Table

Figures

Figure

- 2.1 a) Three sample SCIAMACHY radiance spectra each from pixels with different scene types from July 1, 2004: thick cloud (blue), clear sky over ocean (red), and clear sky over vegetation (green). The overlying black spectra are resampled at a spectral resolution of 15 nm and spectral sampling of 5 nm. b) These are the corresponding standardized spectra to those shown in a), standardized using equation 2.2. 20
- 2.2 The cumulative variance contribution for SCIAMACHY data resampled spectral resolutions constant over the spectral range between 3 nm and 100 nm FWHM. This shows that the variance distribution is insensitive to the spectral resolution between 3 and 25 nm, but not when the data are resampled to a drastically coarser spectral resolution of 100 nm fwhm, for which the dimensionality is reduced. 28
- 2.3 An example of the difference between a) full swath (black) and nadir (red) SCIAMACHY pixel locations shown over west central Africa and the eastern Atlantic Ocean. b) Cumulative variance contributions calculated from the PCA of full swath and nadir data, showing the similarity in variance distribution and identical dimensionalities of six and that information is not lost by reducing the data to nadir-only. 29
- 2.4 The cumulative variance contribution for the four monthly averaged cases. This shows that the first six PCs explain about 99.5% of the variance in each case. 30

- 2.5 The first six PCs calculated from the monthly averaged spectra from January (black), April (blue), July (green), and October (red). Even though each PC does not clearly represent physical variables, some of the PCs do clearly exhibit physical spectral identifiers (e.g. cloud reflectance in PC1, water vapor absorption in several PCs, and vegetation reflectance in PC5). 32
- 2.6 In addition to the spectral signature of vegetation in PC5, it is also evident where vegetation is present by looking at the spatial distribution of the PC5 scores. We see this here through the example of the PC5 scores and the MODIS NDVI image for April 2004. 34
- 2.7 This figure also exhibits how the PC5 scores show the locations of vegetated regions, but it is also clearer to see here, compared to Figure 2.6, where the drier regions are around the globe. There are negative scores in desert regions (southwest of the Amazon, the Sahara Desert, the Arabian Peninsula, and the Tibetan Plateau). Even in semi-arid regions, there are negative values closer to zero than those in the desert regions, such as the western United States, central Australia, and Africa south of the fertile sub-Saharan Africa region. 35
- 2.8 As confirmed by the NIR edge identifiable in PC6, part of this PC represents vegetation. This PC6 October scores plot also shows that vegetation is also partially explained by this PC. There are negative values in the Amazon, sub-Saharan Africa, and the southeastern United States, for example, and positive values over the Sahara Desert. This pattern of opposite sign between vegetated and desert/semi-arid regions is also seen in the PC5 scores. 36
- 2.9 PC4 scores plots from two months, April (a) and July (b). The expected meridional dependence is noticeable in these PCs, but there are also other notable spatial patterns. There is a hot spot of positive scores over Greenland in July, shown in b). High positive scores are also seen over Greenland in the April examples, but those positive values are comparable to scores magnitudes in other locations such as at the southernmost latitudes, the Sahara Desert, and in northeast Russia. 37

- 2.10 The PC2 scores from each of the four monthly averaged cases. Even though it is not clear which single physical variable is represented by this PC, it is clear to see that the Sahara Desert is seasonally insensitive to at least one physical variable this PC represents, seen by the consistently large positive scores over that region. Another pattern is that scores over dry, warm regions, such as the Sahara, central Australia, and the southern tip of Africa are all positive and have similar magnitude. 38
- 2.11 The PC3 scores for the October monthly averaged case show negative values over the Sahara and other dry regions (e.g. deserts in central Asia, the Arabian Peninsula, and the western United States). The highest positive scores are located just west South America and Africa in the southern hemisphere. 39
- 2.12 The rotated PCs for July show how the PCs change through and orthogonal rotation for the first six PCs. Particularly the last three PCs show how physical variables are emphasized after rotation, and how less significant PC loadings are driven toward zero (e.g. PC4). . . . 40
- 2.13 Comparisons of the July PC5 before rotation (a) and after rotation (b) to a green grass albedo spectrum from the John Hopkins ENVI spectra library. Although there are subtle differences between the two, the rotated PC exhibits a closer comparison to the green grass albedo than the original PC5 before rotation. 41
- 2.14 A comparison of the rotated component that mostly closely resembles a water vapor absorption spectrum from July (this happens to be PC6 after rotation) to percentage of the relative change in water vapor per decade, which also clearly shows the locations of water vapor absorption bands. 42
- 2.15 These are the data projected onto the rotated July PC6. The water vapor distribution can most clearly be seen over the oceans, but it appears that surface albedo dominates the patterns of the scores over land, obscuring the water vapor absorption pattern over most land regions. 43

- 2.16 The variance contributions for the first ten PCs for the northern hemisphere (NH) (left) and the southern hemisphere (SH) (right). In the SH, the spring/summer (October/January) and fall/winter (April/July) months are grouped by their variance contributions, this grouping is not as prominent in the NH. 44
- 2.17 The first ten eigenvalues for the southern hemisphere cases. Examining the individual eigenvalues helps to explain the differences in cumulative variance contribution seen in the SH in Figure 2.16b and how each PC dimension contributes to the dimensionality of each case. 45
- 2.18 The first six PCs for the northern hemisphere (left) and southern hemisphere (right). Spectral shapes of the NH PCs and the globally calculated PCs (shown in Figure 2.5) are more similar than the comparison between the SH and global PCs. Between the two hemispheres there are also differences among the PCs. The most notable difference is that the "vegetation PC" explains more of the data variance in April and July. We draw this conclusion because vegetation is represented by the fourth PC in April and July and by the fifth PC in January and October. In the NH vegetation is consistently represented by PC5 for all four months. 46
- 2.19 The SH scores corresponding to the PCs in each of the four month that exhibit a vegetation signal are shown here. In January and October, this is PC5 as in the global case, but in April and July, this is PC4, which implies that in April and July, vegetation explains more of the variance than in October and January. 48
- 2.20 The zonally averaged PCs for two regions: a) shows the PCs for a 10° region in the northern hemisphere mid-latitudes between 40°N and 50°N and b) shows the PCs for a 10° region in the southern hemisphere Tropics between 10°S and 20°S . We see similarities between the two regions in PC1, PC2, PC4, and PC5 , but there are also some aspects of each of the these PCs and the general shapes of the other two not listed above that exhibit differences between the regions as well. 49

2.21 The scores corresponding to the PCs shown in Figure 2.20 are shown here. The land and ocean scores were separated. a) shows the scores corresponding to the PCs calculated for the 40°N - 50°N region, and b) shows the scores corresponding to the 10°S - 20°S region. Here we can see differences between the land and ocean temporal behavior of the scores. The PC6 scores in a) and the PC4 scores in b) demonstrate that the PCs are also able to capture secular trends. 50

2.22 The PCs calculated over the 12 regions into which the Arctic Ocean was divided. Cloud reflectance (PC1), sea ice (PC2), and water vapor absorption bands (PC1 and PC3) are detectable from these components. 53

2.23 The spectral shape of PC2 is compared to that of a snow albedo spectrum from ARCTAS to show the close comparison between the two. Because we only included ocean pixels in the our analysis of the Arctic region, we assume that any snow or ice found is associate with sea ice cover. This also implies that the temporal scores from this component can be used to understand how the PC2 scores represent the temporal pattern of sea ice cover. . . 53

2.24 The Beaufort Sea PC2 scores (black) and the reconstruction of these scores (blue) are compared to the total Arctic Ocean sea ice extent from AMSR-E (red). We see the close similarity between the PC2 temporal behavior and that of the Arctic sea ice extent, exhibiting the benefit of using PC scores to study the temporal behavior of physical variables. 55

3.1 SCIAMACHY-measured reflectance spectra for three scene types: a thick cloud (black), cloud-free vegetation (blue), and cloud-free ocean (red). 65

3.2 The first 30 eigenvalues for the January, April, July, and October 2004 SCIAMACHY and OSSE Reflectance PCs. The difference in shape between each of the grey lines and the black lines shows how well the SCIAMACHY data resampling performed prior to PCA aligned the SCIAMACHY distribution of information with that of the OSSE reflectance spectra. 77

3.3	The cumulative variance fraction for each of the four cases for the first ten PC dimensions. For all four months, both SCIAMACHY (solid) and OSSE (dotted), six PC dimensions explain between 99.7% and 99.9% of the total data variance in both data sets.	78
3.4	The first nine October 2004 SCIAMACHY (black) and OSSE (red) principal components show a close comparison between the two data sets. Physical variables are identifiable within some PC spectral shapes including clouds (PC1), vegetation (PC4), and water absorption (PC1-4).	80
3.5	The spatial distribution of the October 2004 PC4 SCIAMACHY (a) and OSSE (b) scores and the PC5 SCIAMACHY (c) and OSSE (d) scores. Both PC4 and PC5 scores show evidence of vegetation, implying that this physical signal is distributed between at least these two components. Data set similarities seen in the overlap of PC shapes in Figure 3.3 are reinforced by the similarities seen in the spatial distribution of the scores.	81
3.6	The comparison of the eigenvalues from the spectral decomposition of the intersection to the maximum possible eigenvalue for each number of subspace dimensions. With more dimensions used to define the two pairs of subspaces, the similarity identified by the eigenvalues decreases.	83
3.7	The subspace distances between the SCIAMACHY and OSSE reflectance subspaces for ten subspaces (red) compared to the maximum possible distance between the two subspaces (black). The blue line shows the observed subspace distances relative to the maximum possible distance for each number of subspace dimensions, that is, the ratio of the values on the red line to the values on the black line.	84
3.8	The population distances between the OSSE-simulated and SCIAMACHY-observed reflectance spectra for January, April, July, and October 2004. The distances less than zero correspond to subspaces of k dimensions that are the same at the 95% confidence level. The vertical lines indicate the maximum number of dimensions the two data sets share at the 95% confidence level for each case.	86

3.9	The spectral shapes of the transformed eigenvectors for October 2004 in the shared intersection space. The first seven dimensions align well, and the first four dimensions share spectral characteristics with the first four original October 2004 PCs, shown in Figure 3.7.	87
4.1	The first 30 eigenvalues calculated from principal component analysis applied to the zonally averaged spectra from the six latitude bands. Three zonal bands (60°N-90°N, 30°N-60°N, and 60°S-90°S) have eigenvalue spectra that change slope at the fifth component, and the other three zonal bands have eigenvalue that change slope at the fourth component.	99
4.2	The variance contribution of the first 10 SCIAMACHY principal components for the six 30 degree zonal bands. For all the zonal bands, the first four components explain at least 99% of the total variance.	100
4.3	The comparison of the first principal component for the NH and SH tropics show their close similarity, which means that they are explaining very similar variables. The SH PC01 explains 88.98% of the data variance and the NH PC01 explains 95.04% of the variance.	102
4.4	PC01 land (black) and ocean (red) scores time series for the Northern (solid) and Southern Hemisphere (dotted) Tropics zonal bands, each between 0 and 30 degrees. While the two land time series are 180 degrees out of phase, the two ocean time series are about 90 degrees out of phase. The occurrence of the maxima and minima appear to correspond to the location of the Intertropical Convergence Zone.	103
4.5	The first principal component for the two zonal bands located in the polar regions, 60°N-90°N (black) and 60°S-90°S (red) compared to a snow albedo spectrum measured during the ARCTAS field experiment (blue).	105
4.6	The first PCs time series for the two polar regions compared to sea ice extent in the respective hemisphere.	106

4.7	The shaded region shows the boundaries of all the PC02 spectral shapes calculated from monthly averaged, gridded SCIAMACHY reflectance spectra for every month between August 2002 and December 2010. The black line in each panel is the second principal component for the four zonal bands outside the polar regions, and the blue line is a desert reflectance spectrum from the USGS spectral library [Clark <i>et al.</i> , 2007].	107
4.8	The PC2 scores calculated from four months in 2006: January, April, July, and October. The color scales in all four panels have the same data value range.	109
4.9	The land and ocean time series that correspond to the PC2 spectra shown in black in Figure 4.7.	110
4.10	The PC4 and PC5 spectral shapes and scores time series from the 30°N-60°N latitude band. These two PC time series exhibit seasonal variations.	111
4.11	Three examples of PCs and the corresponding scores time series show how an instrument anomaly is represented in the spectral and temporal variability.	112
4.12	The 0°-30°N comparison of SCIAMACHY and OSSE (control and experiment) eigenvalues (left) and principal components (right).	113
4.13	The transformed PCs calculated from the six-dimensional SCIAMACHY and OSSE A2 subspaces in the 0°-30°N zonal band. The correlation between each OSSE and SCIAMACHY PC is shown in the title.	115
4.14	The land (left) and ocean (right) intersecting scores time series that correspond to the transformed principal components shown in Figure 4.13 in the 0°-30°N zonal band. . . .	116
4.15	The transformed PCs calculated from the six-dimensional SCIAMACHY and OSSE A2 subspaces in the 0°-30°S zonal band. The correlation between each OSSE and SCIAMACHY PC is shown in the title.	117
4.16	The land (left) and ocean (right) intersecting scores time series that correspond to the transformed principal components shown in Figure 4.15 from the 0°-30°S zonal band. . .	118

4.17	The eigenvalues (left) and principal components (right) calculated from century time series of OSSE all sky constant CO ₂ concentration and A2 emission cases reflectance spectra in the 30°N-60°N zonal band.	119
4.18	On the left, the comparison of the control and experiment PC1 (top) and PC5 (bottom) land and ocean scores time series between 2000 and 2099 in the NH midlatitude zonal band. On the right, the A2 scores time series are shown to highlight the secular trends in the forced simulation PC time series.	120
4.19	The eigenvalues (left) and principal components (right) calculated from century time series of OSSE all sky constant CO ₂ concentration and A2 emission cases reflectance spectra in the 0°-30°S zonal band.	121
4.20	On the left, the comparison of the control and experiment PC1 (top) and PC4 (bottom) land and ocean scores time series between 2000 and 2099 in the SH Tropics zonal band. On the right, the A2 scores time series are shown to highlight the temporal trends in the forced simulation time series.	122

Chapter 1

Introduction

This thesis focuses on the information in the spectral variability of shortwave (solar) radiation reflected by the Earth system (atmosphere and surface) and how it can be used to monitor changes in climate. Here, shortwave radiation refers to radiation between 320 nm and 2500 nm, which accounts for approximately 95% of the solar radiation incident at the top-of-atmosphere (TOA). Reflected radiation at TOA contains information about Earth's surface features, such as vegetation, snow, and sea ice coverage, and atmospheric variables, such as clouds, aerosol particles, and greenhouse gases. This work uses directly measured spectrally resolved reflected radiation with high spectral resolution and sampling to address three main science questions:

- (1) As the Earth's climate changes, how do reflected solar radiation measurements change over space, time, and wavelength?
- (2) What are the physical causes of changes in Earth's climate?
- (3) How well do climate models reproduce variability observed in the climate system?

Radiative imbalances drive changes in Earth's weather and climate; however, over global and very long time scales, the planet can be assumed to be approximately in radiative equilibrium, which occurs when the absorbed and emitted radiation are equivalent. Earth's absorbed radiation is the difference between the incident TOA solar irradiance and reflected irradiance. Earth's emitted longwave irradiance is dependent upon the effective radiating

temperature, T_E , which is the temperature of a blackbody that emits the same amount of longwave radiation.

$$F = \frac{S_o}{4}(1 - \alpha) - \sigma T_E^4 \quad (1.1)$$

$S_o/4$ is the globally averaged incident TOA solar irradiance, α is the albedo, the ratio between the reflected and incident shortwave irradiance, and σ is the Stefan-Boltzmann constant.

Radiative imbalances can be caused by changes in greenhouse gases, such as CO_2 and CH_4 and other trace gases, land surface albedo, clouds, and incident solar radiation [IPCC, 2007]. These imbalances induce regional climate changes and drive changes in Earth's climate. For example, a positive radiative imbalance, $F > 0$, caused by the planet absorbing more radiation than it emits, may cause the planetary average effective temperature to increase, resulting in Earth emitting more longwave radiation to tend toward equilibrium. A globally averaged increase in temperature will then likely result in several other changes within the climate system.

The scientific community has developed several methods to estimate the magnitude of Earth's radiative imbalance. One method assumes that the excess radiative energy is stored in the ocean and that radiative imbalance can be determined by measuring the ocean heat content. *Willis et al.* [2004] used ocean temperature measurements to estimate the ocean heat content between 1993 and 2003, finding a warming rate during that period of $0.86 \pm 0.12 \text{ Wm}^{-2}$. *Hansen et al.* [2005] found a comparable value for 2003 of $0.85 \pm 0.15 \text{ Wm}^{-2}$ using the GISS global climate model (GCM) [Schmidt et al., 2006]. *Hansen et al.* [2011] also determined a positive energy imbalance of 0.58 Wm^{-2} for 2005-2010.

Satellite radiation budget measurements have also been used to estimate the Earth's radiative imbalance in addition to GCMs and ocean heat content measurements. The Earth Radiation Budget Experiment (ERBE) [Barkstrom, 1984] [Barkstrom, 1984] and the Clouds and Earth's Radiant Energy System (CERES) [Smith et al., 2004] missions have used satellite

instruments in Low Earth Orbit to measure reflected and emitted Earth radiation. CERES irradiance estimates found a positive 6.5 Wm^{-2} anomaly [Loeb *et al.*, 2009], which is much larger than the radiative imbalances calculated using GCM output and ocean heat content measurements. To estimate irradiance from radiance measurements, several assumptions are applied and model-based techniques, such as Angular Distribution Models, are used, which can introduce errors into the calculation. Loeb *et al.* [2009] applied corrections and adjustments to the CERES data that accounted for factors such as calibration uncertainty and an incorrect assumption of the total solar irradiance to make the CERES energy imbalance equivalent to the $+0.85 \text{ Wm}^{-2}$ value calculated by Hansen *et al.* [2005]. Charlson *et al.* [2005] showed that differences among measured and modeled albedo estimates are larger than the albedo change due to an increase in CO_2 concentration during the industrial era. The work in this thesis quantifies the variability within directly measured reflected radiance and reflectance to study climate change, rather than potentially introducing errors into data analysis by using radiance measurements to estimate Earth-reflected irradiance.

In addition to accurately determining Earth's radiative imbalance and identifying the underlying physical processes that cause those imbalances, there are several challenges to detecting and simulating changes in Earth's climate. Perhaps the greatest challenges are due to large uncertainties in cloud (largest), sea and ice albedo, and water vapor feedbacks [Stephens, 2005, Soden and Held, 2006]. Changes in cloud properties, for example, cloud height, can also result in either positive or negative feedbacks. Feedback uncertainties have contributed to differences among climate model projections in part because calculations of climate sensitivity are dependent upon the strength of climate feedbacks [Bony *et al.*, 2006]. Beyond the middle of the century, projections of changes in climate and their impacts are primarily model- and scenario-dependent [IPCC, 2007]. Feedbacks are not the only challenge; there are also uncertainties in forcing, with direct and indirect aerosol forcing being among the most uncertain [IPCC, 2007].

In order to detect changes in climate, highly accurate and stable measurements as well as the ability to monitor changes in instrument accuracy are essential. Measurements with high accuracy over large spatial scales and decadal temporal scales are needed to provide the sensitivity to detect small, but potentially climate-relevant, changes in Earth's climate. To determine instrument accuracy and stability requirements, the expected rate of change of the variables to be detected must be considered [Ohring *et al.*, 2005]. In addition to highly accurate data, measurements with high information content are needed for attribution, the identification of the physical causes underlying the observed changes. The utility of high information content measurements in climate change detection was demonstrated by *Feldman et al.* [2011b], which showed that the time needed to detect climate trends that are distinct from natural variability of the system was shorter for hyperspectral shortwave measurements than broadband shortwave measurements. *Goetz* [2009] defines *hyperspectral* measurements as spectrally resolved data with overlapping and contiguous spectral bands. The measurement requirements for climate change detection and attribution are among the main goals of the Climate Absolute Radiance and Refractivity Observatory (CLARREO) mission. Much of the work presented in this thesis was conducted in support of the CLARREO science definition team. It is, however, relevant to other climate observation missions as well.

Hyperspectral imaging measurements resolve spectral features caused by the interaction of radiation with atmospheric and surface variables and contain information about the variability of physical variables in Earth's climate system. Different surface materials, such as minerals and vegetation, have unique spectral signatures that are caused by the interaction of the specific atoms that make up the surface materials with electromagnetic radiation at different energy levels [Hunt, 1977]. Through radiative transfer theory, it is known that molecules, cloud particles, and aerosol particles in the atmosphere interact with solar radiation through wavelength-dependent scattering and absorption processes. Broadband satellite instruments have provided information about the global distribution of reflected and emitted radiation, and multi-spectral discrete-band satellite imagers have provided in-

formation about the global distribution of cloud, aerosol, and surface properties. However, hyperspectral measurements contain more information about the physical components in Earth's climate system, primarily because of the spectral contiguity of hyperspectral data, a distinguishing attribute from broadband and discrete-band measurements [Goetz, 2009].

The history of *hyperspectral imaging*, also known as *imaging spectrometry*, began in the early 1970s with the 1972 launch of Landsat-1. Landsat-1 was the first Earth-orbiting multi-spectral imager specifically designed to retrieve surface composition using the spectral variability of Earth's surface. Following the launch of Landsat-1, research imaging profilers [Chiu and Collins, 1978, W. Collins, 1982] and airborne imaging spectrometers, for example, the Airborne Imaging Spectrometer (AIS) [Vane et al., 1984] were developed, and were among the first imaging spectrometers. The Airborne Visible Infrared Imaging Spectrometer (AVIRIS) was first flown in 1987, and has been upgraded several times since its inception with a signal-to-noise ratio currently five to ten times greater than the original instrument [Green et al., 1998]. Finally in 2000, NASA launched the first spaceborne imaging spectrometer, HYPERION, as part of the EO-1 mission [Pearlman et al., 2003]. Land surface studies were the primary focus of these first imaging spectrometers, which had high spatial resolutions on the order of tens of meters. The launch of the Moderate Resolution Imaging Spectrometer (MODIS) [King et al., 1992] on Terra (1999) and Aqua (2002) finally enabled daily global monitoring of both land surface and atmospheric variables [King et al., 2003, Platnick et al., 2003], using multi-spectral imaging. In 2002, the European Space Agency launched the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) [Bovensmann et al., 1999] on ENVISAT to study atmospheric composition and its response to anthropogenic and natural forcings. SCIAMACHY measured the full solar spectrum reflected at the top of the atmosphere at a high spectral resolution with continuous spectral sampling. Because of its measurement characteristics, SCIAMACHY shortwave spectra were used in this thesis to quantify the spectral variability of Earth-reflected radiation.

A variety of data analysis techniques have been developed to extract relevant information from complex hyperspectral data. Among the earliest examples, was a spectral unmixing technique to reduce the dimensionality of multi-spectral reflectance to identify surface minerals [Smith *et al.*, 1985]. By *dimensionality* we mean the smallest number of variables needed to explain a sufficient amount of the variance; *unmixing* is the separation of unique spectral signatures in a data set. Smith *et al.* [1990] used Landsat data and unmixing to quantify fractions of vegetation amount in land surface scenes. In a related study, Gamon *et al.* [1993] used spectral unmixing and AVIRIS images to show that a linear relationship exists between vegetation abundance and vegetation canopy types. Boardman *et al.* [1993] developed an automated unmixing technique using convex geometry theory and applied this technique to AVIRIS data.

As more sophisticated multivariate analysis techniques were applied to hyperspectral remote sensing data, the need also arose for data analysis software to facilitate the application of these new techniques. The Interactive Data Language (IDL) was developed in the mid-1980s and was followed by ENVI (Environment for Visualizing Images), an imagery analysis tool written in IDL. The development and distribution of IDL and ENVI made processing multidimensional remote sensing data more accessible to the wider scientific community. In particular, the distribution of ENVI enabled spectral unmixing to become a more popular and recognized imaging spectrometry data analysis technique. These advancements also helped to communicate the value of hyperspectral measurements to the broader remote sensing science community.

Many spectral unmixing techniques used to study hyperspectral data employ principal component analysis (PCA) [Keshava, 2003], a multivariate spectral decomposition technique, as part of the process. The idea of PCA was first introduced by Pearson [1901] and Hotelling [1933] and has often been applied to problems in the social sciences. PCA has been used in atmospheric science applications as early as the 1940s [Obukhov, 1947]. It has been applied to the study of, for example, the spatio-temporal variability of sea level pressure [Hannachi

et al., 2007], geopotential heights [*Craddock and Flood*, 1969], and wind vectors [*Hardy and Walton*, 1978], as well as the relationships among different meteorological variables [*Kutzbach*, 1967]. PCA has been used to study multi-spectral and hyperspectral remote sensing data since the mid-1980s [*Smith et al.*, 1985]. Chapter 2 gives a detailed review of how PCA was used to quantify the spectral variability of shortwave and infrared spectral measurements. For the cases presented herein, the number of variables needed to explain the variability of spectral radiance or reflectance are significantly reduced by the application of PCA from several hundred or thousand spectral bands to on the order of tens of PCs or fewer. PCA is the primary spectral decomposition tool used in this thesis because it quantifies dominant modes of variability in terms of the information contained in the data without using external, model-based techniques or assumptions to do so. Chapter 2 (slightly modified from *Roberts et al.* [2011]) presents experiments in which PCA was applied to observations of reflected spectral radiance to quantify the data variability and identify climate variance drivers that could be extracted directly from hyperspectral radiance measurements. This technique provides an alternate and complementary approach to using inverse methods to retrieve physical variables from radiance measurements. A benefit of using exploratory methods to extract information from directly measured radiance rather than retrieved values is that the uncertainty and error that is introduced in the retrieval process is not included in the direct measurement analysis.

The principal components explaining the majority of the variance in a data set characterizes the variability of a multivariate data set, which can be used as an objective and quantitative measure of the similarity between two data sets. The spectral variability of longwave radiance has been used to evaluate the accuracy of model runs relative to observations [*Haskins et al.*, 1997, 1999]. *Goetz et al.* [1998] used principal component analysis to obtain modeled and measured shortwave hyperspectral irradiance subspaces. The intersection between these two subspaces was used to estimate how many dimensions the two data sets shared and could be used as a transformation between them. *Goetz et al.* [1998]

used this method to find the most appropriate atmospheric correction by determining which MODTRAN-simulated atmosphere spectrum best described the atmospheric features present in the measured spectrum. This automated atmospheric correction technique was used to retrieve surface reflectance spectra with better computational efficiency than previous techniques.

The spectral decomposition of the intersection between two subspaces can also be used to quantify their similarity [Crone and Crosby, 1995, Krzanowski, 1979, Roberts et al., 2012]. Subspace distance, the metric used as a measure of similarity, is a concept often employed in other fields, such as face recognition science [Wang et al., 2005, 2006, Sun and Cheng, 2006], but prior to the studies conducted in Chapter 3, it had not been used to quantify the similarity between two multivariate remote sensing data sets. Chapter 3 also presents a statistical significance test based on Crone and Crosby [1995] that quantifies how many intersecting dimensions are similar between two data sets at the 95% confidence level. The quantitative comparison technique and statistical significance test were used in Chapter 3 (slightly modified from Roberts et al. [2012]) to evaluate how well simulated CLARREO reflectance spectra [Feldman et al., 2011a] reproduced the variability in SCIAMACHY-observed spectra.

To address how hyperspectral shortwave measurements can be used for climate change detection and attribution, the temporal variability of the dominant spectral variables of SCIAMACHY hyperspectral reflectance was quantified between 2002 and 2010. Although the simulated decadal variability of hyperspectral shortwave reflectance has been studied by Feldman et al. [2011a], Feldman et al. [2011b], Chapter 4 presents the first results quantifying the observed decadal variability of the dominant spectral variables calculated from hyperspectral reflectance. In Chapter 4, the intersection was also used to compare the decadal variability of the OSSE reflectance to that observed by SCIAMACHY to understand how well the OSSE is able to simulate temporal changes during the first decade of the 21st century. Finally, the spectral, spatial, and centennial variability of dominant spectral variables was quantified using forced and unforced climate scenarios to illustrate how the multivariate

tools applied in this thesis can be used to identify changes in Earth's climate detected by shortwave hyperspectral reflectance.

Chapter 2

Evaluating the Observed Variability in Hyperspectral Earth-reflected Solar Radiance

We explore the potential for directly measured hyperspectral Earth-reflected solar radiances to provide sufficient information to study changes in Earth's climate based on the quantified variability of the data using principal component analysis (PCA) and singular spectrum analysis. To do this we used these two multivariate analysis techniques on Earth-reflected radiances between 300 and 1750 nm measured from space by the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) instrument. The spatial and temporal variability of hyperspectral reflected radiances over global, hemispherical, and regional scales was quantified. As few as six components were needed to explain over 99.5% in all cases studied, with the exception of an Arctic Ocean case in which only four components were needed. Both of these values represent large reductions in dimensionality of the input radiances from 291 spectral bands. PCA facilitated attribution of the dominant spectral patterns extracted to atmospheric and surface variables including water vapor, clouds, surface albedo, and sea ice. The second most dominant spectral variable, that is the second principal component, in the Arctic closely resembled sea ice reflectance, and followed the temporal behavior of sea ice extent determined from AMSR-E observations. The extraction of the spectral, spatial, and temporal variability in reflected shortwave hyperspectral radiance using multivariate analysis provides an alternate and complementary approach to inverse methods for applying space-based observations to climate studies.

2.1 Introduction

Imbalances between the incoming and outgoing global energy at the top of the atmosphere (TOA) drive weather and climate. Solar radiation is the primary source of external energy, exceeding the sum of all other contributors by over three orders of magnitude [Sellers, 1965]. Incident infrared radiation at wavelengths longer than about $4 \mu\text{m}$ and other energy sources make a negligible contribution to the external energy entering the Earth's system. The fraction of the incident solar energy that is reflected to the TOA by the atmosphere and surface is called the planetary albedo (α), with the remaining fraction, $(1-\alpha)$, absorbed by the atmosphere and surface. Radiative equilibrium, or zero net irradiance, F , is achieved when the absorbed solar radiation, $S_{abs} = (1 - \alpha) * S_o/4$ is exactly balanced by emitted infrared radiation, σT_E^4 , over sufficiently long temporal scales:

$$F = S_{abs} - \sigma T_E^4, \quad (2.1)$$

where $S_o/4$ is the global average of the incident solar irradiance at the top of the atmosphere, σ is the Stephan-Boltzmann constant, and T_E is the radiative effective temperature of the planet, the temperature of an ideal blackbody that emits at a rate equivalent to the rate that solar radiation is absorbed. Over the past fifty years, satellite measurements have improved our understanding of the radiation budget components on a global scale [House *et al.*, 1986]; however, we have yet to overcome the challenge of resolving changes in the reflected term [Loeb *et al.*, 2009], in part because of measurement and retrieval uncertainties.

Currently, the solar radiation incident at TOA is the most accurately measured component of the global radiative energy budget. The Total Irradiance Monitor (TIM) [Loeb *et al.*, 2009], which is part of the Solar Radiation and Climate Experiment (SORCE) [Rottman, 2005], measures the incident solar irradiance at TOA with an uncertainty of 0.5 Wm^{-2} [Kopp and Lean, 2011]. Since 1978, overlapping missions have provided an uninterrupted total solar irradiance record [Kopp *et al.*, 2005]. The incident solar irradiance has contributed to an estimated increase in the change in net irradiance at the tropopause, or radiative forcing,

of 0.12 Wm^{-2} since 1750 [IPCC, 2007]. The Earth Radiation Budget Experiment (ERBE) and the Clouds and Earth’s Radiant Energy System (CERES) have measured the globally distributed emitted longwave radiation and reflected shortwave radiation [Barkstrom, 1990]. ERBE included the first system of satellites designed to study the Earth’s radiation budget [Barkstrom, 1984]. CERES broadband instruments were designed to extend ERBE measurements to retrieve cloud properties with improved accuracy to better quantify TOA and surface fluxes [Smith *et al.*, 2004]. Measurements from ERBE provided details of feedback mechanisms and radiative forcings that refined estimates of global mean albedo to approximately 0.30 [Barkstrom, 1990]. Today, Earth’s mean albedo estimates determined from CERES range between 0.287 and 0.293 [Kandel and Viollier, 2010].

Radiative imbalances can occur through changes in greenhouse gas concentration, aerosol properties, land cover, and solar radiation [IPCC, 2007]. These imbalances result in regional climate responses, so accurately quantifying these imbalances is important to understanding climate changes. The attribution of these changes to specific forcing agents is difficult because of the complex manner in which individual climate variables influence the radiation field. As demonstrated by Loeb *et al.* [2009], measurement and retrieval accuracies must continue to improve in order to quantify and characterize radiative imbalances. Hansen *et al.* [2005] used the NASA Goddard Institute for Space Studies (GISS) global climate model [Schmidt *et al.*, 2006] to estimate a radiative imbalance of $+0.85 \pm 0.15 \text{ Wm}^{-2}$. Willis *et al.* [2004] estimated a similar value using ocean heat content measurements. CERES estimates result in a $+6.4 \text{ Wm}^{-2}$ anomaly [Fasullo and Trenberth, 2008]. Discrepancies remain in the derived radiation budget components despite the detailed analyses of Loeb *et al.* [2009] and Fasullo and Trenberth [2008] and improvements over time in measurements and models.

Differences in the retrieved planetary albedo, as discussed by Charlson *et al.* [2005], have led to diverse conclusions regarding the effect temporal changes in albedo anomalies have on Earth’s climate. Kato [2009] used CERES TOA irradiances and a description of the meridional transport of energy within the climate system to explain why the global

albedo anomaly has not been changing over time. While *Kato* [2009] presents a detailed explanation of how large-scale dynamics regulate interannual albedo variations, it is possible that small, but climatologically important, albedo anomalies exist but remain below our detection threshold. Using measurements of Earthshine to derive the Earth’s reflectance, *Pallé et al.* [2004] concluded that there was an increase in the albedo between 2000 and 2003 corresponding to a decrease in shortwave absorption of about 5 Wm^{-2} . *Wielicki et al.* [2005], however, used CERES observations to show that albedo decreased during this same period by an amount equivalent to a decrease in shortwave irradiance of 2 Wm^{-2} and concluded that the increase in albedo reported by *Pallé et al.* [2004] did not coincide with requisite changes in surface temperature, ocean heat content, or cloud amount. *Charlson et al.* [2005] showed that albedo estimates from various modeling and observational studies have uncertainties that exceed the greenhouse gas forcing of 2.4 Wm^{-2} . The uncertainty in observed albedo must be sufficiently reduced if it is to be used to monitor climate-relevant trends.

Clouds, aerosol particles, and molecules that make up the atmosphere, and surface characteristics such as vegetation, sea ice, desert terrain, and ocean all contribute to the outgoing TOA reflected radiation in unique ways through spectrally-dependent scattering and absorption processes. Changes in Earth’s climate are often studied by using satellite-based measurements of radiance to model terms in the energy budget and examining how changes in individual climate variables such as these cause those components to change over time. Here we present an alternative to this method that instead characterizes the variability of directly measured radiance to understand the spatial and temporal behavior of climate drivers. These methods use reflected hyperspectral radiance spanning the near ultraviolet (UV) to the near infrared (NIR) to identify a highly accurate representation of the current state of Earth’s climate system, extract climate change trends, and identify the underlying sources that drive the variance in the observed shortwave radiation. Here we adopt the *Goetz et al.* [1985] and *Goetz* [2009] definition of hyperspectral as referring to spectral data that have many narrow, contiguous, and overlapping wavelength channels.

The National Research Council's 2007 Decadal Survey for the Earth Sciences emphasizes the importance of understanding the information contained in Earth-reflected solar radiation pertaining to sea ice, aerosols, land use, and cloud properties and how this information can be used to study the resulting climate forcings [*National Research Council, 2007*]. The spectral oversampling characteristic of hyperspectral data enables the separation of unique spectral signatures from surface materials and atmospheric constituents, a process known as unmixing the signal. Hyperspectral measurements have been applied to a variety of environmental studies related to mineralogical mapping, vegetation abundance and canopy health, and ice and snow property identification [*Goetz, 2009*], among many others. Less common, however, are atmospheric and climate applications which employ the full span of the solar spectrum. In this chapter, we explore the potential for enhanced spectral resolution and range to provide the sufficient information needed to extract the small but significant changes in the Earth's climate.

Hyperspectral sensors such as the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) [*Bovensmann et al., 1999*], Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [*Green et al., 1998*], and HYPERION [*Pearlman et al., 2003*] have at least an order of magnitude more spectral bands than their discrete band counterparts, requiring new approaches to data analysis and processing. A major benefit of using hyperspectral data over multispectral measurements is the multitude of information contained in the continuous spectral measurements, from which important spectral signals can be extracted directly without enlisting the assistance of assumptions prior to signal retrievals. Multivariate analysis methods lend themselves to the study of such complex and high-dimensional data sets. Principal component analysis (PCA) is frequently applied to simplify such high-dimensional data as hyperspectral radiances. In this study we use multivariate analysis to quantify the variance of space-borne measurements of hyperspectral Earth-reflected radiances over a continuous spectral range between the near UV and the near IR. This analysis extracts the dominant spectral patterns in the measured radiances.

This chapter illustrates the utility of using multivariate analysis techniques to quantify and attribute the variance in satellite-measured hyperspectral radiances.

The details of the SCIAMACHY instrument are described in Section 2.2. Section 2.3 presents an overview of how principal component analysis has been used historically with meteorological and radiation data and an explanation of its mathematical details. The results of PCA applied to SCIAMACHY radiances on global, hemispherical, and regional scales to quantify the variance in the data are shown in Section 2.4. Discussion, summary, and future work are presented in Section 2.5.

2.2 SCIAMACHY

This study employs reflected spectral radiances from the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY). SCIAMACHY is a hyperspectral imaging spectrometer on ENVISAT, a sun-synchronous satellite in near-polar orbit that measures across eight channels covering the spectral ranges 214 - 1773 nm, 1934 - 2044 nm and 2259 - 2386 nm with spectral resolutions between 0.22 nm and 1.48 nm [*Gottwald et al.*, 2011b]. Channels seven and eight (spanning 1934 - 2044 nm and 2259 - 2386 nm, respectively) do not have reliable optical throughput because of ice that deposits on the detectors [*Gottwald et al.*, 2011c]; therefore, analysis is restricted to the wavelength range 300 nm to 1750 nm. Nadir pixel size is dependent on the integration time and swath width setting causing sizes to vary between 26 km (along track) by 30 km (across track) and 32 km (along track) by 930 km (across track). With nadir sampling, SCIAMACHY has a scanning angular width of $\pm 32^\circ$ across track, which corresponds to a maximum nadir swath width of 960 km [*Gottwald et al.*, 2011b].

SCIAMACHY was designed to study how natural and anthropogenic activity affects global atmospheric composition [*Gottwald et al.*, 2011a]. Additional objectives include understanding the global distribution, chemistry, and physics of trace gases, aerosols, and clouds in the troposphere, stratosphere, and mesosphere [*Bovensmann et al.*, 1999]. SCIAMACHY's

moderate spectral resolution and global coverage resolve spectroscopic features of atmospheric trace gases [Gottwald *et al.*, 2011a]. Recent results have shown the utility of spectral measurements from SCIAMACHY in retrieving cloud properties [Kokhanovsky *et al.*, 2005], identifying regional statistically significant column-integrated water vapor trends Mieruch *et al.* [2008], and showing that annual increases of CO₂ can be monitored using satellite measurements [Buchwitz *et al.*, 2007]. These and other studies use SCIAMACHY spectral radiances to derive specific physical variables pertinent to climate studies. Alternatively, we examined the variability of hyperspectral reflected radiances over various temporal and spatial scales to extract the spectral features of variance drivers of Earth’s climate system.

2.3 Multivariate Analysis Techniques

2.3.1 Historical Applications of Principal Component Analysis

We used principal component analysis (PCA) to quantify the variability of hyperspectral reflected near ultraviolet, visible, and near infrared radiance between 300 and 1750 nm, identify spectral characteristics of this variability, and potentially link these characteristics to physical causes. PCA has historically been used in atmospheric science to study climate variability [Karoly *et al.*, 1996] and the spatiotemporal variability of meteorological variables, such as geopotential heights [Wallace and Gutzler, 1981] and temperature, [Kelly *et al.*, 1982, Kidson, 1975]. PCA is also a common decomposition technique used to study multispectral and hyperspectral radiation because it simplifies the analysis by reducing data dimensionality if spectral radiances, the variables in this application, are highly correlated. Additionally, it can extract the primary physical processes driving the variance in the data that may otherwise be difficult to identify. The dimensionality of a data set is the smallest number of variables needed to explain the majority of the data variance; the quantitative definition of majority is subjectively defined. The subset of variables used to define the dimensionality of a data set is also referred to as a subspace of the original data set.

Huang and Yung [2005] used PCA on Earth-emitted longwave radiation measured by the Atmospheric Infrared Sounder (AIRS) to identify temperature variations in the lower stratosphere and upper troposphere. They also compared the principal components (PCs), which are linear combinations of the spectral radiances used to quantify the distribution of variance, from observed and modeled radiances from the same time period and with comparable spatial coverage to quantify discrepancies and sensitivities present in the model output, a technique also used by others [*Haskins et al.*, 1997, 1999, *Huang et al.*, 2002]. *Huang et al.* [2002] calculated principal components from outgoing longwave radiation measured by the Infrared Interferometric Spectrometer (IRIS) and output from GCM and MODTRAN4 [*Anderson et al.*, 1999]. From these results they concluded, from differences in the variability of the first PC, that the variability of cloud properties was underestimated by 2 to 6 times in the model compared to the observations.

In addition to PCA applied to outgoing longwave radiation, there have also been spectral PCA studies done using reflected solar radiation in the visible and near infrared. AVIRIS is an airborne hyperspectral imaging spectrometer that began making routine measurements in 1987 of reflected solar radiance between 400 and 2500 nm [*Green et al.*, 1998]. *Boardman and Green* [2000] used PCA to analyze the dimensionality of AVIRIS data from individual flight lines and from the complete data set to demonstrate that the AVIRIS data had a high but significantly reduced dimensionality from 224 bands to 40 or 60 significant PCs. They warned against disregarding the information of the least significant PCs that explain only a fraction of a percent of the variance in the data because those PCs may contain relevant information about the signal, rather than being purely noise. *Goetz et al.* [1998] used PCA to quantify the relationship between observed and modeled radiances to improve a single-pixel atmospheric correction technique needed to analyze AVIRIS data. *Rabbette and Pilewskie* [2001] used spectral PCs of downwelling solar irradiance data measured with the Solar Spectral Flux Radiometer (SSFR) [*Pilewskie et al.*, 2003] and demonstrated that few PCs were needed to explain the majority of the variance in the data. Using these components, they

characterized the variance of solar downwelling irradiance with physical variables such as the liquid water content, water vapor, molecular scattering, and ozone absorption and concluded that an unknown absorber was not responsible for the underestimation of absorbed radiation by radiative transfer models. They also calculated the PCs from synthetic irradiances that included water vapor and ozone absorption and molecular scattering, illustrating the utility of understanding the PCs from data with known physical inputs. *Rabbette and Pilewskie* [2002] presented PCA results from a similar study with upwelling and downwelling SSFR irradiance and compared results from two different geographical areas.

2.3.2 Principal Component Analysis Description

Principal component analysis (PCA) is a spectral decomposition technique that uses second moment statistics to evaluate the distribution of variance within a data set. In this study, all PCA transformations were performed using the software package ENVI (Environment for Visualizing Images (Version 4.6), developed by I.T.T. Information Systems, 2009, Boulder, Colo., available at www.ittvis.com). At native resolutions hyperspectral data may contain hundreds or thousands of contiguous wavelength bands, such as AVIRIS with 224 bands or SCIAMACHY with over 8000 bands. PCA extracts the prominent linear combinations of original variables that explain the majority of the data variance, greatly reducing the large number of variables if the original variables are highly correlated. PCA examines the relationship between radiances (here, the original variables) in each wavelength band to those in every other wavelength band and determines orthogonal linear combinations that account for the greatest remaining variance.

The first step in our analysis was to determine if standardization of the data was required. Standardization is used when the variables span a large dynamic range. We standardized the SCIAMACHY radiances because covariance is sensitive to the magnitude of radiances. If unstandardized, spectral radiances with the largest dynamic range would dominate the PCs. By standardizing the data, the variables with larger radiances, and

therefore, covariances, do not dominate the directions of the primary PCs, and radiances in all wavelength bands are weighted equally. Standardized anomalies were calculated by subtracting the spectral mean and dividing by the spectral standard deviation using

$$\vec{z}_k = \frac{\vec{x}_k - \bar{x}_k}{s_k}, \quad (2.2)$$

where \vec{z}_k is a vector of all standardized radiances in a single band (indexed with k) for all pixels, \vec{x}_k is a vector of all radiances in one wavelength band, and \bar{x}_k and s_k are the mean and standard deviation of radiances in one wavelength band. The standardized data in \mathbf{Z} have zero mean and unit variance in each wavelength band.

Figure 2.1a shows example SCIAMACHY spectral radiance spectra [$Wm^{-2}nm^{-1}sr^{-1}$] for three different scene types. Figure 2.1b shows the corresponding unitless standardized spectra used for deriving the principal components. The colored lines in these figures represent SCIAMACHY data at the native spectral resolution of the instrument, and the overlying black spectra are resampled at a spectral resolution of 15 nm full-width at half maximum and a spectral sampling resolution of 5 nm. Standardization removes the spectral distribution of downwelling solar irradiance from the spectra, as evidenced by the differences between Figures 2.1a and 2.1b.

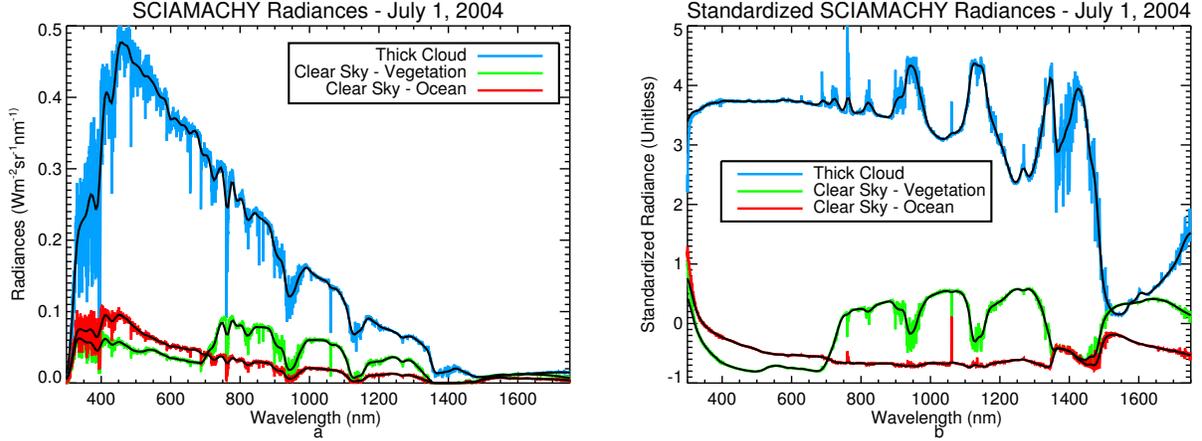


Figure 2.1: a) Three sample SCIAMACHY radiance spectra each from pixels with different scene types from July 1, 2004: thick cloud (blue), clear sky over ocean (red), and clear sky over vegetation (green). The overlying black spectra are resampled at a spectral resolution of 15 nm and spectral sampling of 5 nm. b) These are the corresponding standardized spectra to those shown in a), standardized using equation 2.2.

The next step was to calculate the $K \times K$ covariance matrix, \mathbf{C} , here equivalent to a correlation matrix because the data are mean-centered with unit variance.

$$\mathbf{C} = \frac{1}{N-1} \mathbf{Z}\mathbf{Z}^T \quad (2.3)$$

N is the number of observations (number of spectra), K is the number of variables (number of spectral radiances), and \mathbf{Z} is the $N \times K$ standardized data matrix. The eigenvectors and eigenvalues (together, the *eigenmodes*) of the covariance matrix are defined by the characteristic equation

$$\mathbf{C}\mathbf{E} = \lambda\mathbf{E}, \quad (2.4)$$

where \mathbf{E} is the $K \times K$ eigenvector matrix, and λ is a $1 \times K$ vector of eigenvalues. The covariance matrix can be spectrally decomposed using the singular value decomposition

$$\mathbf{C} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T, \quad (2.5)$$

where $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix containing the K eigenvalues in vector λ . Each eigenvalue represents the variance of the corresponding eigenvector, so the fractional variance contri-

bution of each eigenvector is $\lambda_k / \sum \lambda_k$. The eigenmodes are ordered by the magnitudes of the eigenvalues. The principal components are calculated by scaling the eigenvectors by the square root of the associated eigenvalues:

$$PC_k = \sqrt{\lambda_k} E_k. \quad (2.6)$$

The resulting individual values of each PC vector, the *loadings*, are also the correlation between each PC and the spectral radiances and are bound between -1 and 1. By definition, each PC is orthogonal to all others; that is, they are mathematically independent. Although the PCs are mathematically independent, it is still possible for the sampling distribution of each PC to be related to the neighboring PC. It is, therefore, helpful to know before attempting PC interpretation if the PCs are statistically significantly separated. One test of this separability is called the *North et al. Rule of Thumb* [North et al., 1982, Wilks, 2006]. This rule uses eigenvalue confidence intervals to determine if the PCs are statistically separate from each other. If the neighboring eigenvalues fall outside the confidence interval of a given PC, then they are statistically different from each other. We calculated 95% confidence intervals for the eigenmodes analyzed in this study using,

$$\Delta\lambda = z(0.975) * \sqrt{\frac{2}{N} \hat{\lambda}^2} \quad (2.7)$$

The dimensionality of the data can be determined using the fractional variance contribution of each PC. The resulting subspace can, ideally, be used to reconstruct the original data set. In the following section it will be shown that resampled SCIAMACHY spectra, each containing 291 spectral radiances, can be reduced to a set of only five or six principal components that explain 99.5% of the variance. Principal components with very small variance contribution (i.e. small eigenvalues) typically represent noise in the data set, and may therefore be eliminated from future analysis; however, PCA does not always successfully separate signal from noise, particularly when the signal explains a very small fraction of the data variance comparable to the noise variance. Because of this it is possible for less significant PCs to contain relevant information [Boardman and Green, 2000].

The eigenvector matrix defines a new coordinate system onto which the original data can be mapped. Projecting the standardized data matrix, \mathbf{Z} , onto this matrix of new axes, \mathbf{E} , gives the $N \times K$ PC scores matrix, \mathbf{D} ,

$$\mathbf{D} = \mathbf{Z} \mathbf{E}. \quad (2.8)$$

The K column vectors in \mathbf{D} are the PC scores that represent a weighted average of the Z_k vectors, for which the weights are the loadings of each eigenvector. Depending on whether the pixels are temporally or spatially distributed, these values represent the weightings of the PCs in time or space. The PCs represent new axes for which the directions define the greatest variance remaining in the data and represent new variables by which the variance is explained. Large absolute PC scores are strongly associated with the variables defined by each PC. As scores approach zero, they represent a less significant association with that PC.

Mathematical and Physical Orthogonality and PC Rotation Discerning the underlying physical meaning of any PC may prove difficult because it is a linear combination of all N input variables and is orthogonal to all other PCs. Also constraints in the mathematical formulation of the PCs do not necessarily represent physical constraints. The orthogonality requirement of PCA can suppress the obvious representation of physical variables by PCs explaining less of the variance than the first because various combinations of the physical variables are projected onto each PC in order to maximize the variance explained by each component, a concept well demonstrated by Figure 2 in *Huang and Yung* [2005]. In *Huang and Yung* [2005], the schematic shown in Figure 2 shows that there is some relationship between upper troposphere relative humidity and the thermal contrast between the Earth's surface and clouds, demonstrated by an acute angle between the two vectors representing each variable. The variable represented by the thermal contrast between cloud and surface explains a larger fraction of the variance than the upper troposphere relative humidity, which is illustrated by the thermal contrast vector being much longer than the

relative humidity vector. The first (most dominant) principal component in the schematic is drawn primarily in the same direction as the thermal contrast vector, but is between the thermal contrast vector and the relative humidity vector because that is the direction which allows for the largest amount of variance to be explained. Because of the orthogonality constraint that exists within the PCA algorithm, the second PC is perpendicular to the first PC and explains as much of the variance as it can in the plane orthogonal to the first PC. Because there are other physical variables that are present in the schematic, PC2 is a combination of physical variables as it explains the second greatest amount of variance not explained by the first PC.

A common tool that can be used to help clarify physical interpretation is rotation of the eigenvectors, which simplifies the components that define the dimensionality of the data set [Rabbette and Pilewskie, 2001, 2002]. The rotation of the primary PCs amplifies large absolute loadings and drives the intermediate and small loadings closer to zero, which defines a concept known as *simple structure* [Richman, 1986, Jolliffe, 2002]. This procedure and the achievement of simple structure often helps to clarify which of the spectral radiances are the most significant contributors to each PC, the absolute loadings of which will be noticeably higher than those of the less significant spectral radiances after rotation.

Either orthogonal or oblique rotations can be used. Oblique rotations allow for more flexibility because orthogonality between axes is not enforced, although PC independence is lost. However, as the orthogonality constraint often obscures physical information, oblique rotations may help to highlight the physical drivers explained by the group of PCs selected for rotation. For both types of rotation the total variance explained by the components used in the rotation is maintained, but the variance explained by each component is distributed more uniformly among the rotated components. While the dominance of each successive PC may be changed, the physical interpretation of the PCs may be clarified. Oblique and orthogonal rotations are performed using rotation criteria that are used to select the elements of the rotation matrix. Rotation criteria, such as the one used in this chapter, varimax, *Kaiser*

[1958, 1959] strive to achieve simple structure as defined above by optimizing some set of mathematical criteria. *Kaiser* [1958] and *Wilks* [2006] discuss in detail the mechanics of rotation.

2.3.3 Time Series PCA: Singular Spectrum Analysis

Singular spectrum analysis (SSA) is a spectral decomposition technique mathematically equivalent to PCA used to identify the dominant periodic patterns in univariate (single-variable) time series [*Elsner and Tsonis*, 1996]. SSA has historically been used in the geophysical sciences to detect dominant temporal modes in climate time series such as the El-Niño Southern Oscillation Index [*Rasmusson et al.*, 1990], global surface air temperature records [*Vautard et al.*, 1992], and geopotential heights [*Ghil and Mo*, 1991]. In this study, Multichannel Singular Spectrum Analysis (MSSA), the multivariate time series extension of Singular Spectrum Analysis, described in more detail below, was applied to extract the dominant temporal patterns from time series of PC scores over a spatial grid.

SSA is different from PCA only in that the variables are time-lagged snapshots of the univariate time series instead of spectral radiances. The length of each snapshot, the *window length*, M , is a user-defined parameter of SSA (and MSSA). $N/4$, where N is the length of the time series, is typically an acceptable window length, a selection also considered appropriate when deciding how many lagged snapshots of a time series to use with the autocorrelation function [*Elsner and Tsonis*, 1996]. $N/3$ is accepted as the upper limit for M selection [*Vautard et al.*, 1992]. *Plaut and Vautard* [1994] empirically determined that SSA tends to identify patterns between $M/5$ and M . Because each time series snapshot is M long, SSA cannot identify patterns longer than M . *Vautard et al.* [1992] suggest that an M bracketed by

$$\frac{1}{f} \leq M \leq \frac{1}{\delta f} \quad (2.9)$$

will isolate an oscillation with frequency f , although these bounds may not be determined

a priori. If there is prior knowledge about the oscillations of interest in the time series, f , may be known prior to conducting SSA (such as the case in Section 2.4.4). *Wilks* [2006] combines these criteria as $M/5 \leq 1/f \leq M \leq N/3$. Further discussion on the ideal selection of the window length may be found in *Elsner and Tsonis* [1996], *Vautard et al.* [1992], and *Plaut and Vautard* [1994].

The number of lagged snapshots is determined by $N' = N - M + 1$. The $N' \times M$ trajectory matrix, \mathbf{X} , is composed of these time series snapshots (column vectors), the variables from which the covariances will be calculated. The trajectory matrix is constructed using $X_{i,j} = x_{i+j-1}$ where $1 \leq i \leq N'$ and $1 \leq j \leq M$. Similar to PCA, the lagged-covariance matrix is

$$\mathbf{C} = \frac{1}{N' - 1} \mathbf{X}\mathbf{X}^T \quad (2.10)$$

and explains the relationships between all possible pairs of lagged snapshots of the time series in question. This construction of the trajectory matrix shows how the study of a univariate time series becomes a multivariate problem. The lagged-covariance matrix, \mathbf{C} , is spectrally decomposed into the eigenmodes (eigenvectors and eigenvalues), typically using the singular value decomposition shown in equation 2.5; here, the eigenvectors explain the primary temporal modes of variability. The temporal scores can be calculated in a similar manner to those determined in PCA by projecting the original data onto the eigenvectors using

$$D_i^k = \sum_{j=1}^m X_{i,j} E_j^k \quad (2.11)$$

where i and j are defined as above, and the eigenmodes are indexed by k . To reconstruct the original time series using a subset of the temporal principal components, the convolution of the eigenvectors and the temporal scores is determined. Thorough treatments of SSA can be found in *Vautard et al.* [1992], *Elsner and Tsonis* [1996], and *Golyandina et al.* [2001].

ysis (MSSA) is used to simultaneously identify temporal and spatial patterns in multivariate time series. This technique has been used for several geophysical science applications such as in the study of low-frequency atmospheric variability using 700 mb geopotential heights [Plaut and Vautard, 1994]. In this study, MSSA was used to analyze time series of PC scores over a spatial grid. The MSSA trajectory matrix is similar to the SSA trajectory matrix except that each row of the MSSA trajectory matrix contains lagged snapshots that cover the same time period for each of the L grid boxes, making the dimensions of the trajectory matrix $N' \times ML$. The $ML \times ML$ lagged-covariance matrix is calculated from this trajectory matrix in a manner similar to SSA. However, the MSSA covariance matrix is more complex because it is a collection of smaller covariance matrices, each of which explains the strength of the relationship between every possible pair of locations and lagged time series snapshots. If there are L grid boxes then the extended covariance (EC) matrix is

$$EC = \begin{pmatrix} C_{1,1} & \cdots & C_{1,mL} \\ \vdots & \ddots & \vdots \\ C_{mL,1} & \cdots & C_{mL,mL} \end{pmatrix} \quad (2.12)$$

Each of the covariance matrices on the diagonal ($C_{1,1}, C_{2,2}, \dots, C_{mL,mL}$) in the above equation are equivalent to the matrices that would be calculated for SSA of a time series at any of the L locations. The off-diagonal matrices are different from the diagonal matrices only in that they compare the lagged snapshots between two different grid boxes. The eigenvectors calculated from this extended covariance matrix are, therefore, space-time principal components (ST-PCs), representing patterns that explain decreasing amounts of variance, indicated by the corresponding eigenvalues.

Oscillations in the series that have the same period as the ST-PCs are identified using MSSA if the following three conditions are satisfied [Plaut and Vautard, 1994]:

- (1) Two consecutive modes have nearly equal variances, or, eigenvalues.
- (2) Those modes in (1) are represented by periodic eigenvectors that have the same

period.

- (3) The eigenvectors and corresponding scores are offset in frequency space by 90° , that is, they are orthogonal.

Equivalent to SSA and PCA, time series can be reconstructed using the output from MSSA. A reconstructed component is found by convolving the eigenvectors and temporal scores using two consecutive eigenmodes that satisfy (1), (2), and (3) above. A more detailed discussion of MSSA can be found in *Plaut and Vautard [1994]*. This method was used to compare the reconstructed time series of the PCA scores to time series of physical variables to demonstrate the utility of PCA in identifying key physical variables, not only spatially, but in the time domain as well. This will be discussed in more detail in Section 2.4.4.

2.4 Hyperspectral Reflected Radiance Variability

In the following sections a qualitative analysis of the PCA results of hyperspectral reflected radiances over different temporal and spatial scales is presented. There are several ways in which PCA can reveal the dominant physical variables in a data set, which are discussed in more detail below. Before performing PCA, we resampled SCIAMACHY spectra to coarser and spectrally constant spectral resolutions. Figure 2.2 shows the differences in variance contribution for the same set of SCIAMACHY radiances resampled to a variety of spectral resolutions, showing that data resampled to spectrally constant spectral resolutions between 3 and 25 nm have a variance distribution and dimensionality that do not change significantly. We also examined the shapes of the eigenvectors associated with these cases, and they were all essentially the same except for differences due to apparent smoothing resulting from the radiance spectral resampling. Only when a drastically coarser spectral resolution was used did the dimensionality notably decrease. This is shown consistently by six PCs explaining almost 100% of the variance for all the cases included. From this exercise we concluded that a spectral resolution could be selected between 3 and 25 nm without

losing a significant amount of information. For the studies presented here, we used a spectral resolution of 15 nm full width at half-maximum (FWHM) and spectral sampling of 5 nm, resulting in 291 spectral bands between 300 and 1750 nm. We analyzed only those pixels measured at line-of-sight (LOS) angles less than 5° , that is, near nadir, to remove azimuthal dependencies. The comparison between PCs of full swath and nadir SCIAMACHY spectra from April 8, 2004 showed that data dimensionality was conserved (Figure 2.3), justifying the generalization to nadir-only data. These two tests used to establish the angular restriction and spectral resolution for the SCIAMACHY radiances used here were motivated by a science definition study performed to understand requirements needed for future satellite missions to study the reflected shortwave radiation from Earth.

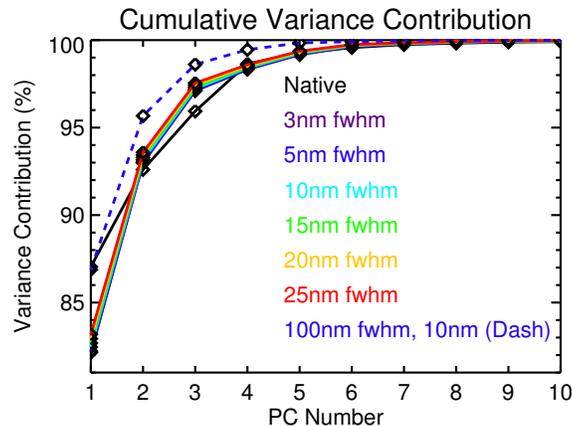


Figure 2.2: The cumulative variance contribution for SCIAMACHY data resampled spectral resolutions constant over the spectral range between 3 nm and 100 nm FWHM. This shows that the variance distribution is insensitive to the spectral resolution between 3 and 25 nm, but not when the data are resampled to a drastically coarser spectral resolution of 100 nm fwhm, for which the dimensionality is reduced.

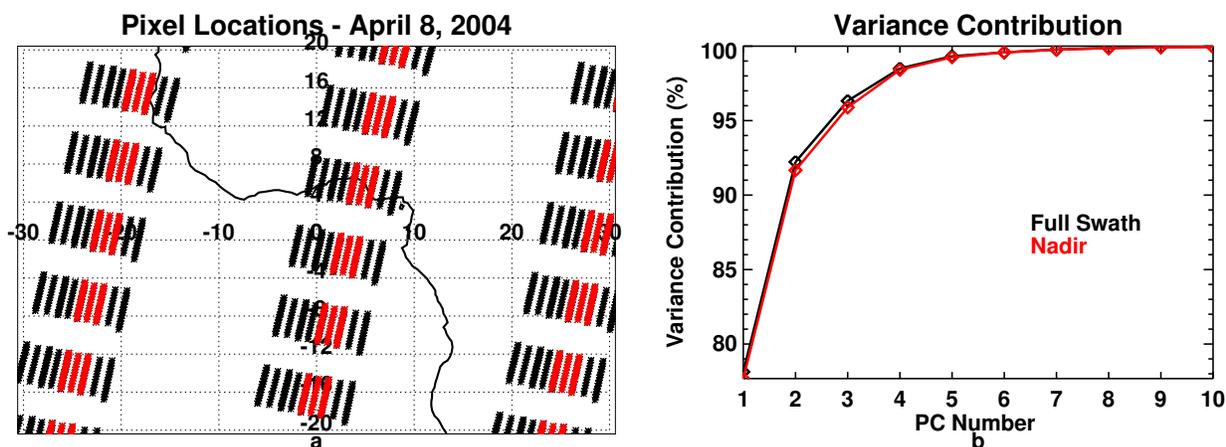


Figure 2.3: An example of the difference between a) full swath (black) and nadir (red) SCIAMACHY pixel locations shown over west central Africa and the eastern Atlantic Ocean. b) Cumulative variance contributions calculated from the PCA of full swath and nadir data, showing the similarity in variance distribution and identical dimensionalities of six and that information is not lost by reducing the data to nadir-only.

2.4.1 Global Seasonal Variability

To study how the spatial variability of the PCs changes with season, radiances of daily global spectra from four months (January, April, July, and October) in one year, 2004, were spatially averaged over a grid (4° latitude by 6° longitude). The PCs were then calculated from the monthly averaged, spatially gridded SCIAMACHY spectra, the results from which are discussed here. Figure 2.4 shows the cumulative variance contribution for the four cases. Although there are subtle differences in the distribution of variance among the less significant PCs between the four cases, their dimensionalities (here defined as the number of PCs explaining 99.5% of the variance) are identical. We also used the North et al. Rule of Thumb to evaluate the statistical significance of the differences of the PCs from one another, and found the first six PCs to be statistically different from each of their neighboring PCs at a confidence of 95%. For each case, we see that the dimensionality of the data is reduced from 291 wavelengths to six PCs. These six PCs are shown in Figure 2.5.

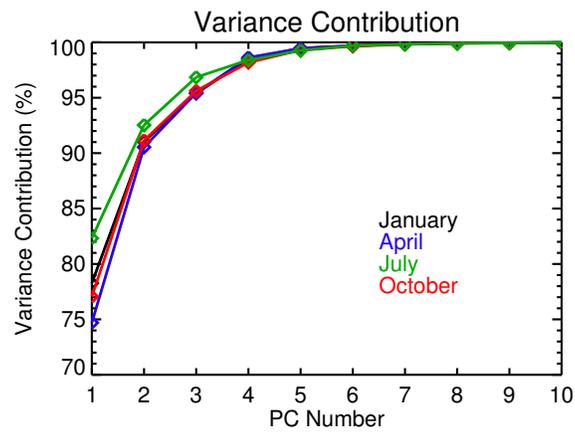


Figure 2.4: The cumulative variance contribution for the four monthly averaged cases. This shows that the first six PCs explain about 99.5% of the variance in each case.

These PCs resemble reflectance spectra because the spectra were standardized with equation 2.2, essentially removing the signature of incident solar spectral irradiance. Because the PCs are scaled by the eigenvalues, their loadings represent the correlation between each PC and the spectral radiances. To aid in PC interpretation we examined the spectral shapes of the PCs for similarity with known reflectance features and compared the spatial distribution of the scores with those of known spatial features. Even though it can be difficult to identify the physical variables explained by each spectral signature in all the PCs, some PCs have recognizable surface and atmospheric spectral features. For example, the first component exhibits a spectral signature indicative of cloud reflectance shown by PC loadings that are fairly constant for wavelengths throughout the visible. PC1 also exhibits the locations of near infrared water vapor absorption bands centered at 940, 1140, and 1350 nm.

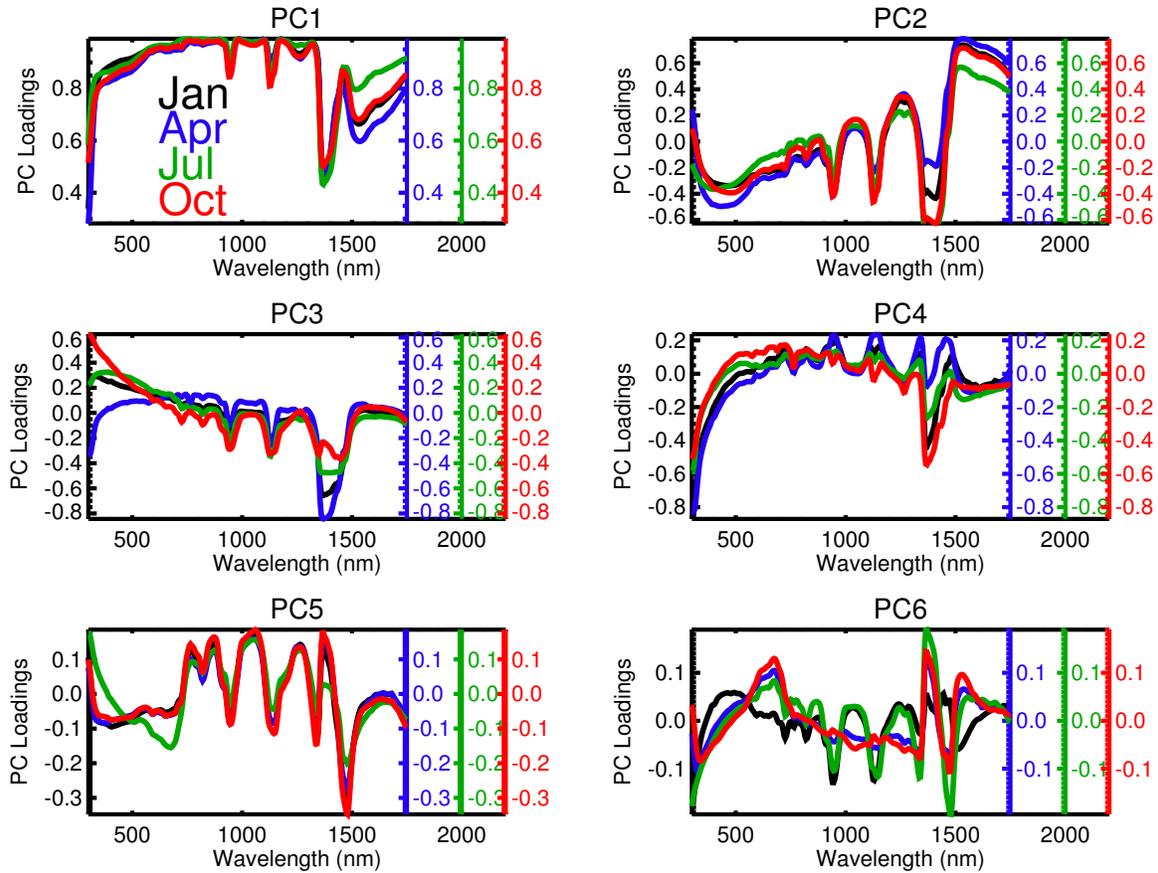


Figure 2.5: The first six PCs calculated from the monthly averaged spectra from January (black), April (blue), July (green), and October (red). Even though each PC does not clearly represent physical variables, some of the PCs do clearly exhibit physical spectral identifiers (e.g. cloud reflectance in PC1, water vapor absorption in several PCs, and vegetation reflectance in PC5).

The fifth component bears the spectral signature of reflectance over a vegetated surface, as in, for example, *Bowker et al.* [1985]; *Trishchenko et al.* [2004]; and *Michalsky et al.* [2003]. Although the water vapor absorption bands have higher correlations associated with other PCs, such as PC1 and PC2, PC5 also shows the locations of water vapor absorption bands. Vegetation has two distinct spectral properties in the visible and near infrared. Chlorophyll and carotenoid pigment absorption dominates at wavelengths shorter than 700 nm, with the absorption minima surrounding the "green peak" centered around 560 nm [*Bowker et al.*, 1985]. The near infrared edge, which occurs at wavelengths between about 700 nm and 750

nm, is caused by the boundary between cellular reflectance, prominent in the NIR wavelength bands, and chlorophyll absorption [Bowker *et al.*, 1985]. This difference between reflected radiation in the visible and near infrared has been used to calculate vegetation indices, which give a measure of the density of green vegetation. For example, a longer NIR edge over an area results in a larger vegetation index, and therefore implies denser vegetation in that area [Jordan, 1969, Elvidge and Chen, 1995]. Figure 2.5 shows that the length of the PC5 NIR edge for July is longer than that for the other three months implying that globally, there is a higher density of vegetation compared to the other months. The difference in vegetation amount between the different months is also observed by studying the PC5 scores.

The PC5 scores, two examples of which are shown in Figures 2.6 and 2.7, reinforce the link between reflectance and vegetation by comparing these scores from April and July 2004 with the MODIS Terra monthly averaged normalized difference vegetation index (NDVI) images from the same time periods. Positive scores are found in areas with abundant green vegetation, for example, in the Amazon, sub-Saharan Africa, and the southeastern United States. Scores close to zero are found over the oceans and in the Polar Regions, where there is no vegetation. The July scores in Figure 2.7 exhibit vegetation patterns typical of a northern hemisphere (NH) summer, so there are more locations the NH that have more positive scores than the April example, which is during the NH spring. In July, vegetated areas in the southern hemisphere (SH) are generally becoming less green, indicated both by the NDVI images and by scores that are still positive but closer to zero in the same regions that were very green in April. These figures also present examples of the differences in interpretation between negative and positive scores. The negative scores tend to be located in desert and semi-arid regions. Negative scores in Eurasia and Africa outline the locations of dry and vegetated regions in particular. In July and April, there are also regions just west of the western coasts of South America, North America, and southern Africa, for example, where there are concentrated areas of negative scores that do not indicate dry regions because they are over oceans, therefore indicating a different variable.

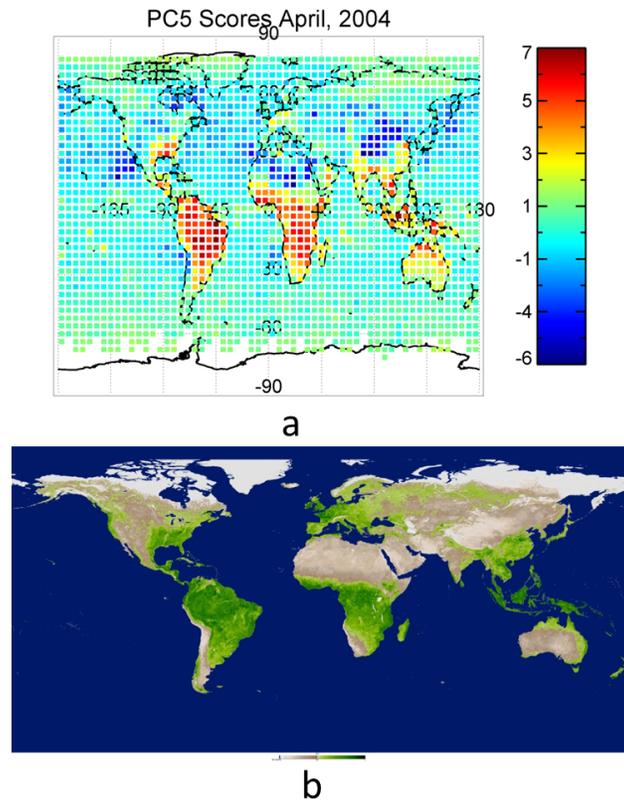


Figure 2.6: In addition to the spectral signature of vegetation in PC5, it is also evident where vegetation is present by looking at the spatial distribution of the PC5 scores. We see this here through the example of the PC5 scores and the MODIS NDVI image for April 2004.

Although we are not confident of all the physical variables described by PC6, some of the months exhibit a vegetated spectral signature. This is evidenced by the small NIR edge identifiable in the October PC6 spectral shape. The PC6 October scores in Figure 2.8 show negative values in the Amazon, southeastern United States, Southeast Asia, and sub-Saharan Africa and positive values over the Sahara and Arabian Deserts. The change in sign of the scores that we also see in the PC5 scores further represents the partial PC6 representation of vegetation.

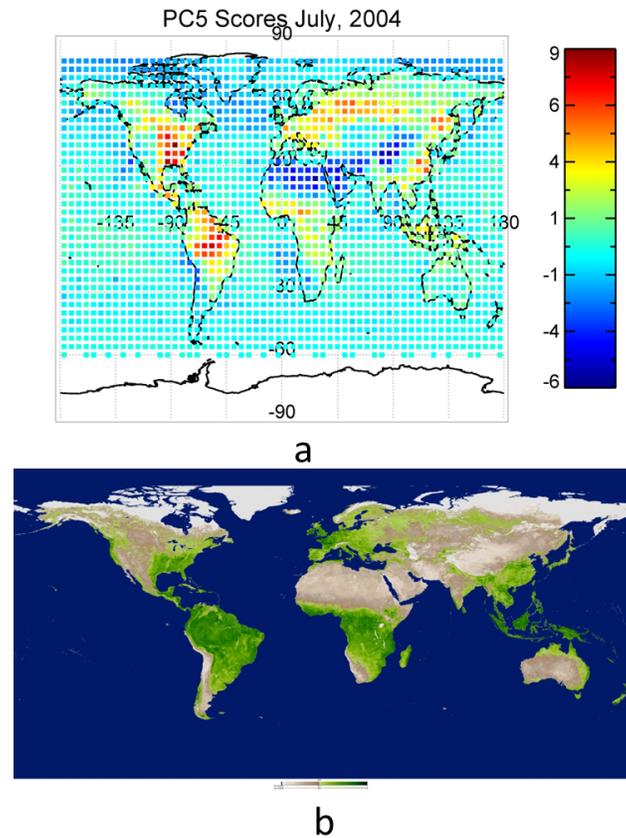


Figure 2.7: This figure also exhibits how the PC5 scores show the locations of vegetated regions, but it is also clearer to see here, compared to Figure 2.6, where the drier regions are around the globe. There are negative scores in desert regions (southwest of the Amazon, the Sahara Desert, the Arabian Peninsula, and the Tibetan Plateau). Even in semi-arid regions, there are negative values closer to zero than those in the desert regions, such as the western United States, central Australia, and Africa south of the fertile sub-Saharan Africa region.

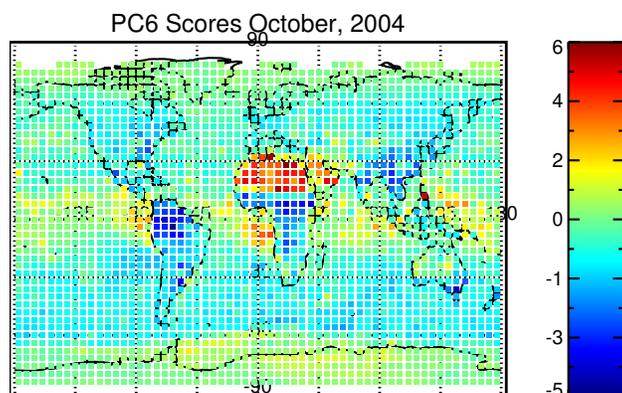


Figure 2.8: As confirmed by the NIR edge identifiable in PC6, part of this PC represents vegetation. This PC6 October scores plot also shows that vegetation is also partially explained by this PC. There are negative values in the Amazon, sub-Saharan Africa, and the southeastern United States, for example, and positive values over the Sahara Desert. This pattern of opposite sign between vegetated and desert/semi-arid regions is also seen in the PC5 scores.

PC4 in the visible part of the spectrum follows an inverse fourth power law as a function of wavelength, similar to that of molecular scattering. This is expected from highly variable molecular scattering paths that result from changing sun geometry with latitude. Figure 2.9, showing the PC4 scores from April and July, provides evidence for other factors influencing this PC. The latitudinal variability in both figures can be explained by sun position: there are large, negative scores in the tropics, and high, positive scores over the poles. This pattern is most noticeable over the oceans. It is also clear that there are differences between land and ocean. In the July plot, scores over land are typically closer to zero than the scores over the surrounding bodies of water, indicating that surface albedo likely contributes to PC4. A clear feature present in July but not as much in April is a "hot spot" of positive scores isolated over the Greenland Ice Sheet. High positive scores are also seen over Greenland in the April scores, but those positive values are comparable to the magnitudes of scores in other locations such as the southernmost latitudes, the Sahara Desert, and northeast Russia.

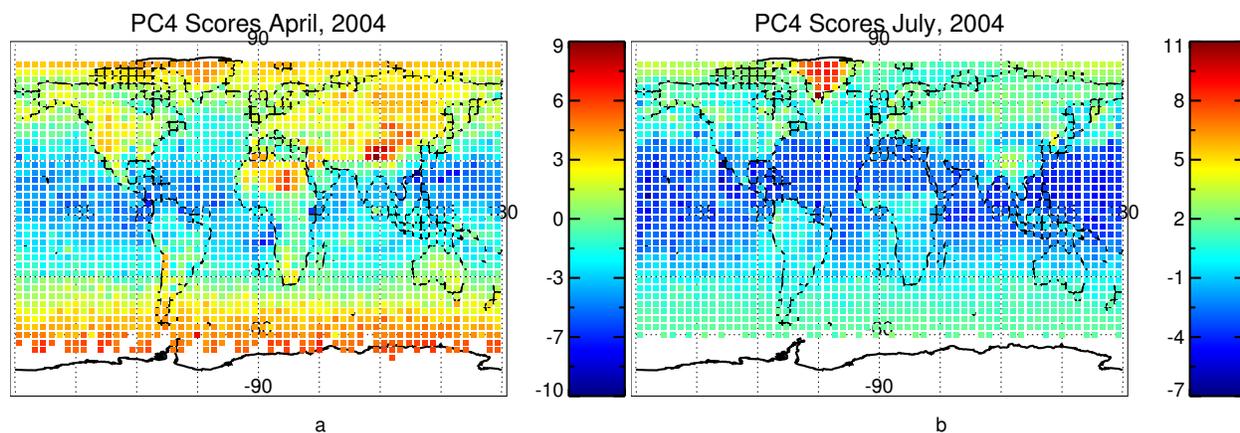


Figure 2.9: PC4 scores plots from two months, April (a) and July (b). The expected meridional dependence is noticeable in these PCs, but there are also other notable spatial patterns. There is a hot spot of positive scores over Greenland in July, shown in b). High positive scores are also seen over Greenland in the April examples, but those positive values are comparable to scores magnitudes in other locations such as at the southernmost latitudes, the Sahara Desert, and in northeast Russia.

Although PC2 also shows the near-infrared water vapor absorption bands centered at 940, 1140, and 1350 nm, there are other spectral shapes that do not directly link to physical variables. Figure 2.10 shows the PC2 scores for all four cases. Over widespread snow-covered areas such as the Polar Regions in the spring or winter, there are negative scores particularly in January and April. Positive scores are then seen over dry, warm regions, such as the Sahara and Arabian Deserts, central Australia, and the southern tip of Africa. The third consistent spatial pattern related to this PC is that there are negative values over the equator that are opposite in sign to the scores over the warm, dry regions. It may be that these three patterns are not explaining a single physical variable, but that three (or more) physical variables are represented by this PC but are difficult to separate as individual physical processes, as discussed in Section 2.3.2.

Similarly PC3 cannot be associated with a unique spectral characterization; although, the water vapor bands are again evident. The October PC3 scores in Figure 2.11 show that there are negative values over the Earth's dry regions (for example, the western United

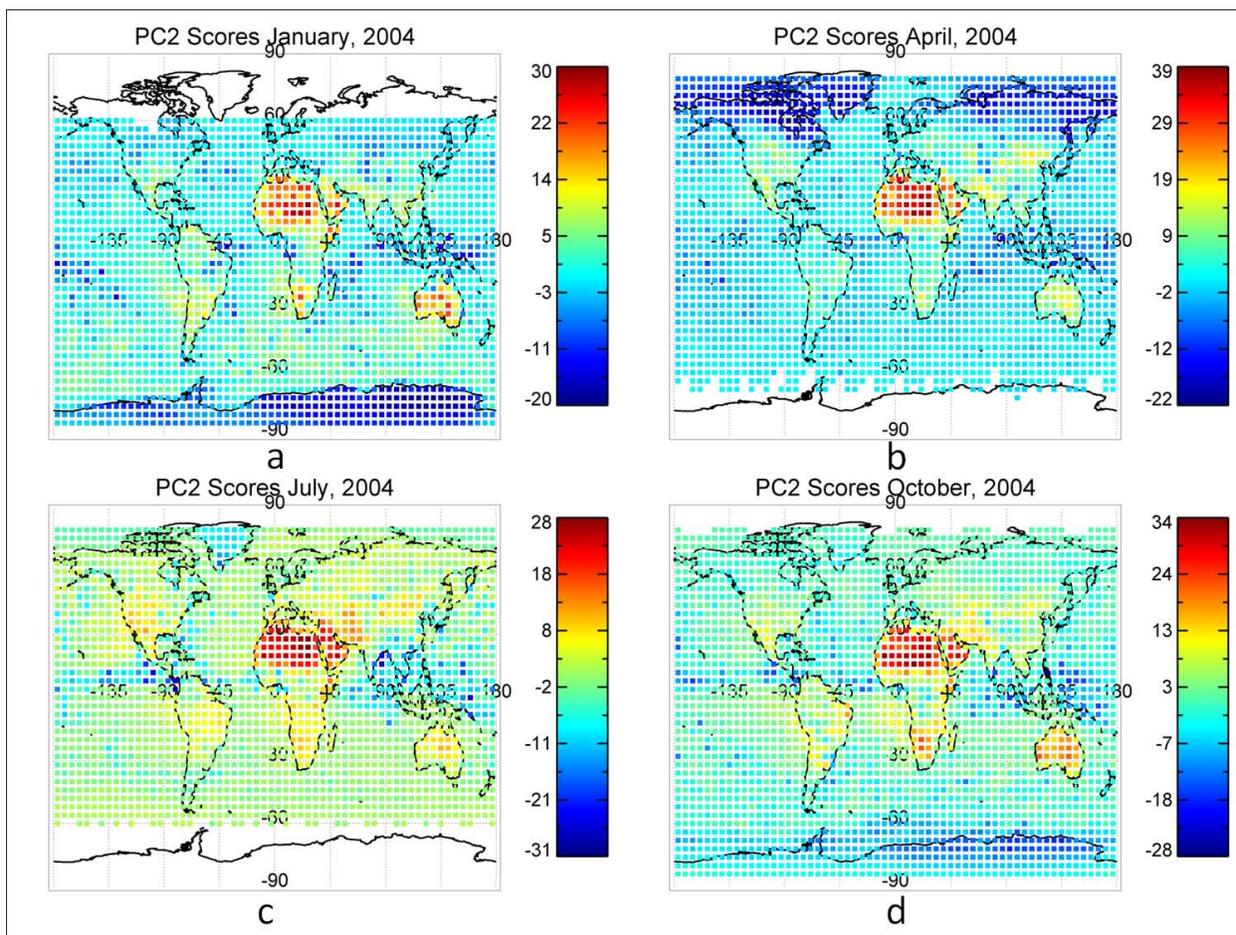


Figure 2.10: The PC2 scores from each of the four monthly averaged cases. Even though it is not clear which single physical variable is represented by this PC, it is clear to see that the Sahara Desert is seasonally insensitive to at least one physical variable this PC represents, seen by the consistently large positive scores over that region. Another pattern is that scores over dry, warm regions, such as the Sahara, central Australia, and the southern tip of Africa are all positive and have similar magnitude.

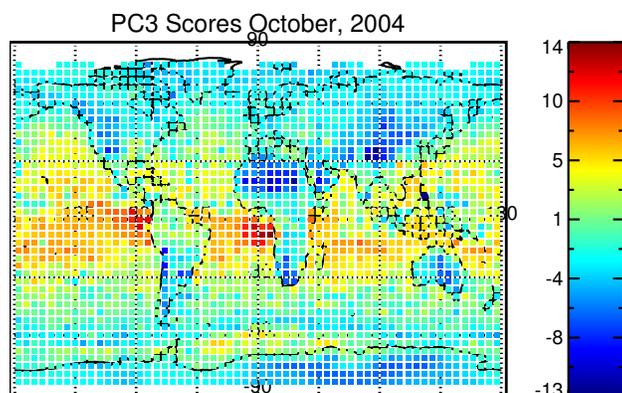


Figure 2.11: The PC3 scores for the October monthly averaged case show negative values over the Sahara and other dry regions (e.g. deserts in central Asia, the Arabian Peninsula, and the western United States). The highest positive scores are located just west South America and Africa in the southern hemisphere.

States, Sahara Desert, western South America, and the Tibetan Plateau). In general, large positive values are found only over the oceans, not over land. There are also large positive values at and just south of the equator over ocean. There are maxima over ocean regions west of the Amazon, sub-Saharan Africa, and northern Australia.

Principal Component Rotation To attempt to improve physical interpretation the first six July PCs, shown in Figure 2.5 were orthogonally rotated (Section 2.3.2) using the varimax rotation criterion of *Kaiser* [1958, 1959]. Figure 2.12 shows the six rotated components. Recall from Section 2.3.2 that the rotated PCs have nearly equal variance, unlike the original PCs, for which the order is indicative of variance contribution. This also means that a change in PC order has no implications on their relative importance. The main objective of rotation is to drive the intermediate PC loadings toward zero and amplify large absolute amplitude loadings, a concept referred to as simple structure. The fourth rotated component exhibits this improvement toward simple structure: the molecular scattering spectral signature is still evident in the visible, but the other spectral features present in the original PC4 have been muted such that the rest of the loadings on this PC are now very close to zero.

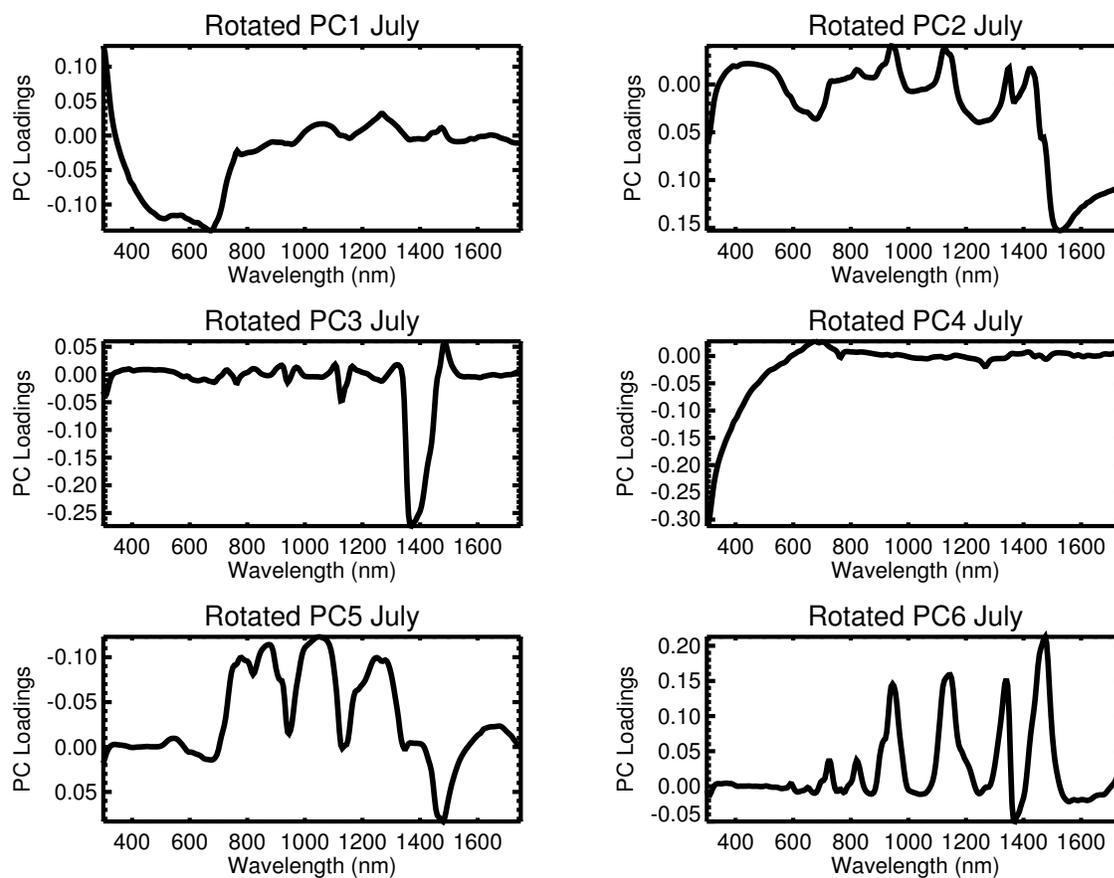


Figure 2.12: The rotated PCs for July show how the PCs change through and orthogonal rotation for the first six PCs. Particularly the last three PCs show how physical variables are emphasized after rotation, and how less significant PC loadings are driven toward zero (e.g. PC4).

An improvement toward a pure vegetation reflectance spectrum occurs following the rotation of PC5. Figure 2.13 compares PC5, rotated PC5, and a green grass albedo spectrum from the John Hopkins University ENVI spectral library [Baldrige *et al.*, 2009]. Before rotation there is a sharp slope in the short wavelength end of the visible spectrum that is not representative of typical vegetation reflectance. After rotation the loadings at those wavelengths are close to zero. In addition, the green peak and NIR edge more closely match those in the green grass albedo spectrum from reference. Although the PC5 spectrum does not likely derive solely from green grass reflectance, this spectrum more closely resembles a vegetation spectral signature than the unrotated PC5.

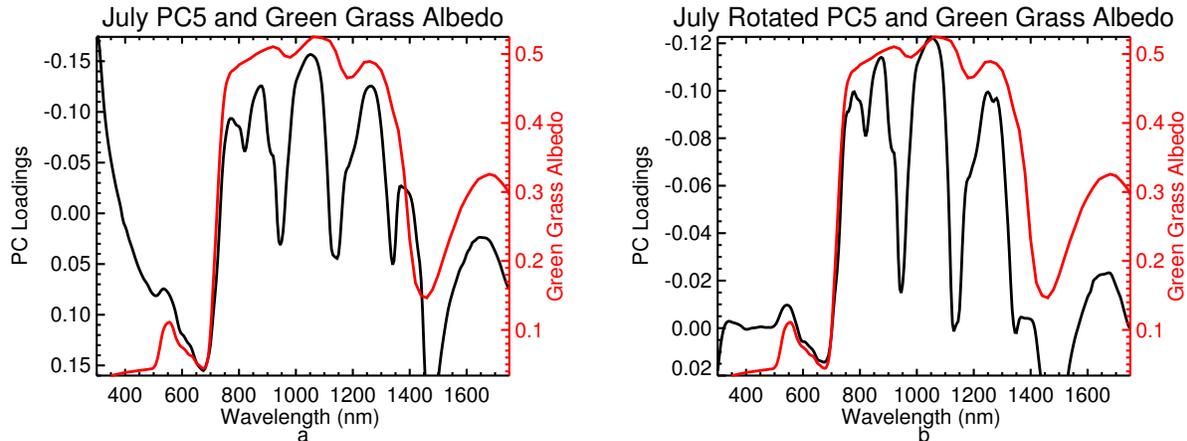


Figure 2.13: Comparisons of the July PC5 before rotation (a) and after rotation (b) to a green grass albedo spectrum from the John Hopkins ENVI spectra library. Although there are subtle differences between the two, the rotated PC exhibits a closer comparison to the green grass albedo than the original PC5 before rotation.

Figure 2.14 shows the July rotated component that most closely resembles a water vapor absorption spectrum, which in this case is PC6. Recall from Section 2.3.2 that after rotation the order of the PCs is no longer linked to their individual variance contributions and that the variance is distributed more evenly among the PCs included in the rotation. The red spectrum in Figure 2.14 was calculated using the change per decade in globally

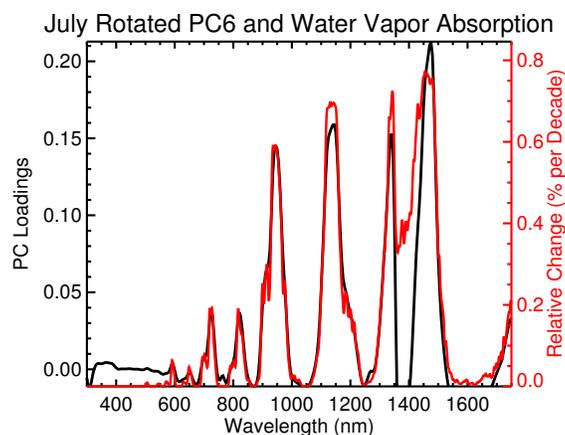


Figure 2.14: A comparison of the rotated component that mostly closely resembles a water vapor absorption spectrum from July (this happens to be PC6 after rotation) to percentage of the relative change in water vapor per decade, which also clearly shows the locations of water vapor absorption bands.

averaged water vapor over ocean, 0.41 kgm^{-3} per decade, from *Santer et al.* [2007] using data from the Special Sensor Microwave Imager (SSM/I). Using MODTRAN4, the change in radiance due to this water vapor trend was derived. The red spectrum in Figure 2.14 is the relative change in radiance in percent per decade. The lower boundary of the rotated PC6 y-axis was restricted to illuminate the similarity of the spectral features between the two spectra. The close comparison of this rotated PC6 (black line) and the water vapor absorption spectrum (red line) suggests that water vapor is the primary physical variable linked to this rotated PC.

Scores can also be projected onto the rotated PCs so they can be expressed in the rotated PC space. This helps to illustrate how the spatial PC distribution is improved by rotation. We expect to see a water vapor spatial pattern through the scores of the July rotated PC6, which are shown in Figure 2.15. Over the oceans the spatial distribution of scores is as expected, with larger scores in the tropics and scores closer to zero closer to the poles, but over land it is more difficult to see the spatial distribution of water vapor. This may be because there are surface effects that mask the spatial distribution of water vapor absorption even after rotation. It is encouraging, however, to see the latitudinal dependence

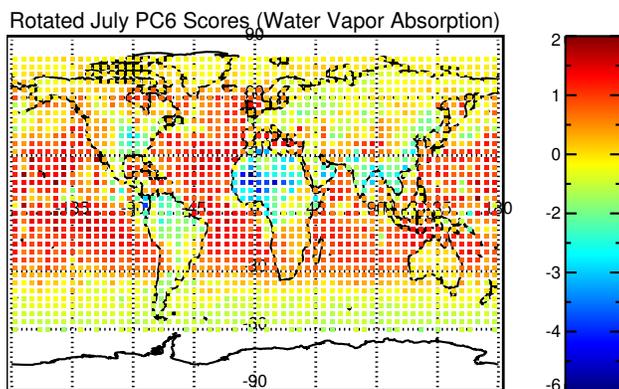


Figure 2.15: These are the data projected onto the rotated July PC6. The water vapor distribution can most clearly be seen over the oceans, but it appears that surface albedo dominates the patterns of the scores over land, obscuring the water vapor absorption pattern over most land regions.

of scores over the oceans. This example illustrates the potential of rotated eigenvectors to more closely align with reflectance signatures from unique atmospheric and surface sources.

2.4.2 Hemispherical Seasonal Variability

Although examples in the previous section revealed advantages of PCA applied to global scenes, seasonal variability may have been masked due to the inclusion of data from both hemispheres. To better emphasize seasonal changes, we applied PCA separately to scenes from both hemispheres (NH and SH). The cumulative variance contributions for the first ten PCs for both hemispheres are shown in Figure 2.16. The cumulative variance contribution of the first six PCs represents, on average, 99.7% of the variance for both the NH and SH, but there are differences in the cumulative variance contributions from the less significant PCs. In the SH, the first four PCs for April and July (fall/winter) explain more of the variance than the PCs for January and October (spring/summer). This can be seen in Figure 2.16 by the way the lines corresponding to each pair of months are grouped. To better understand why this is the case, the variance contributions of each individual SH PC were examined. The major differences between the two groups (fall/winter and spring/summer)

are due to differences between the variance contributions of the first, second, fourth, and fifth PCs, which can be seen in Figure 2.17. The seasonal grouping seen in the SH variance distributions of Figure 2.16b and Figure 2.17 can also be seen by how well the shapes of the PCs from the pairs of seasons listed above match each other in Figure 2.18b. Comparing differences between the two hemispheres, there is often a difference in slope in the visible, such as in the first and second PCs. There are also differences in the visible between different months within each hemisphere. Eigenvalue confidence intervals revealed that the first six PCs are statistically different from one another at a confidence of 95%.

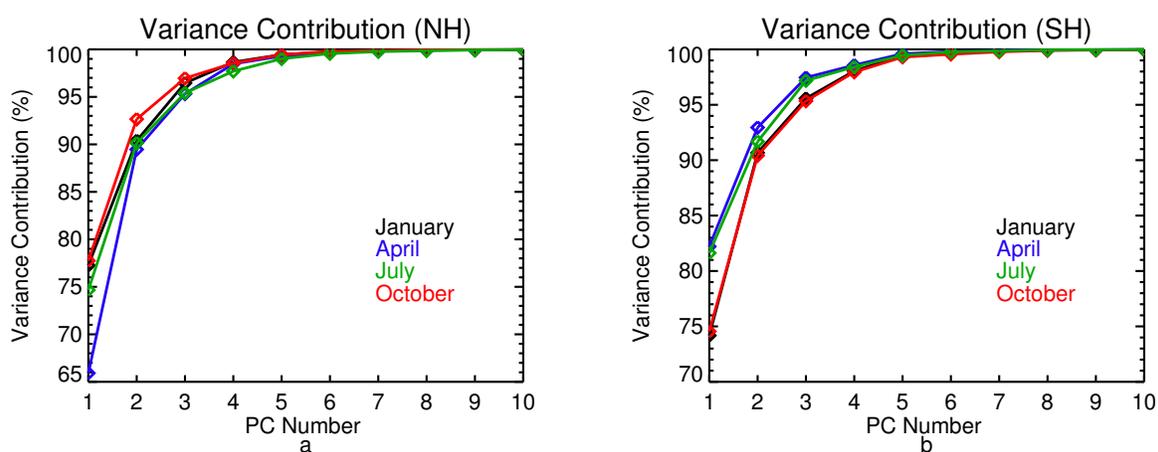


Figure 2.16: The variance contributions for the first ten PCs for the northern hemisphere (NH) (left) and the southern hemisphere (SH) (right). In the SH, the spring/summer (October/January) and fall/winter (April/July) months are grouped by their variance contributions, this grouping is not as prominent in the NH.

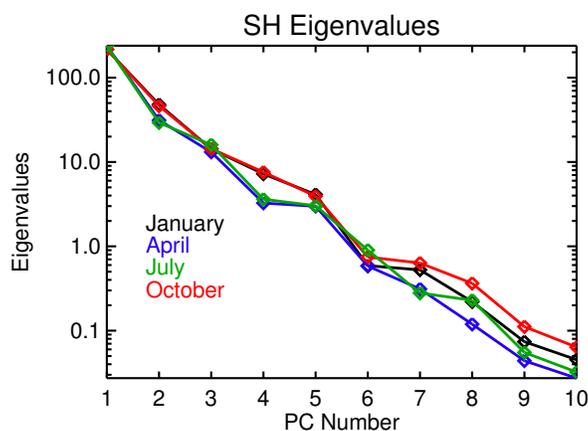


Figure 2.17: The first ten eigenvalues for the southern hemisphere cases. Examining the individual eigenvalues helps to explain the differences in cumulative variance contribution seen in the SH in Figure 2.16b and how each PC dimension contributes to the dimensionality of each case.

The so-called vegetation PCs, identified using both the PC spectral shape and scores, in the SH exhibit how PCA of hemispherical radiances amplifies seasonal differences. In Figure 2.18b these PCs are shown in the lower left panel. The vegetation PC for April and July is PC4 and for January and October is PC5. Figure 2.19 shows the scores for each of these four PCs. During the Amazonian dry season, the months of which included here are April and July, vegetated surface spectral features (the green peak and near infrared edge) contribute more to the variance than during the wet season, the months of which included here are October and January, as evidenced by the differences in PC order. The significance of the PCs can be determined by their rank and therefore by their individual variance contribution relative to the other PCs in each separate case, not necessarily by the differences in the individual variance contributions among the cases shown in Figure 2.17. How these PCs vary seasonally can be seen by the changes of the distribution of the scores in Figure 2.19 and the spectral shapes of the PCs, shown in Figure 2.18b.

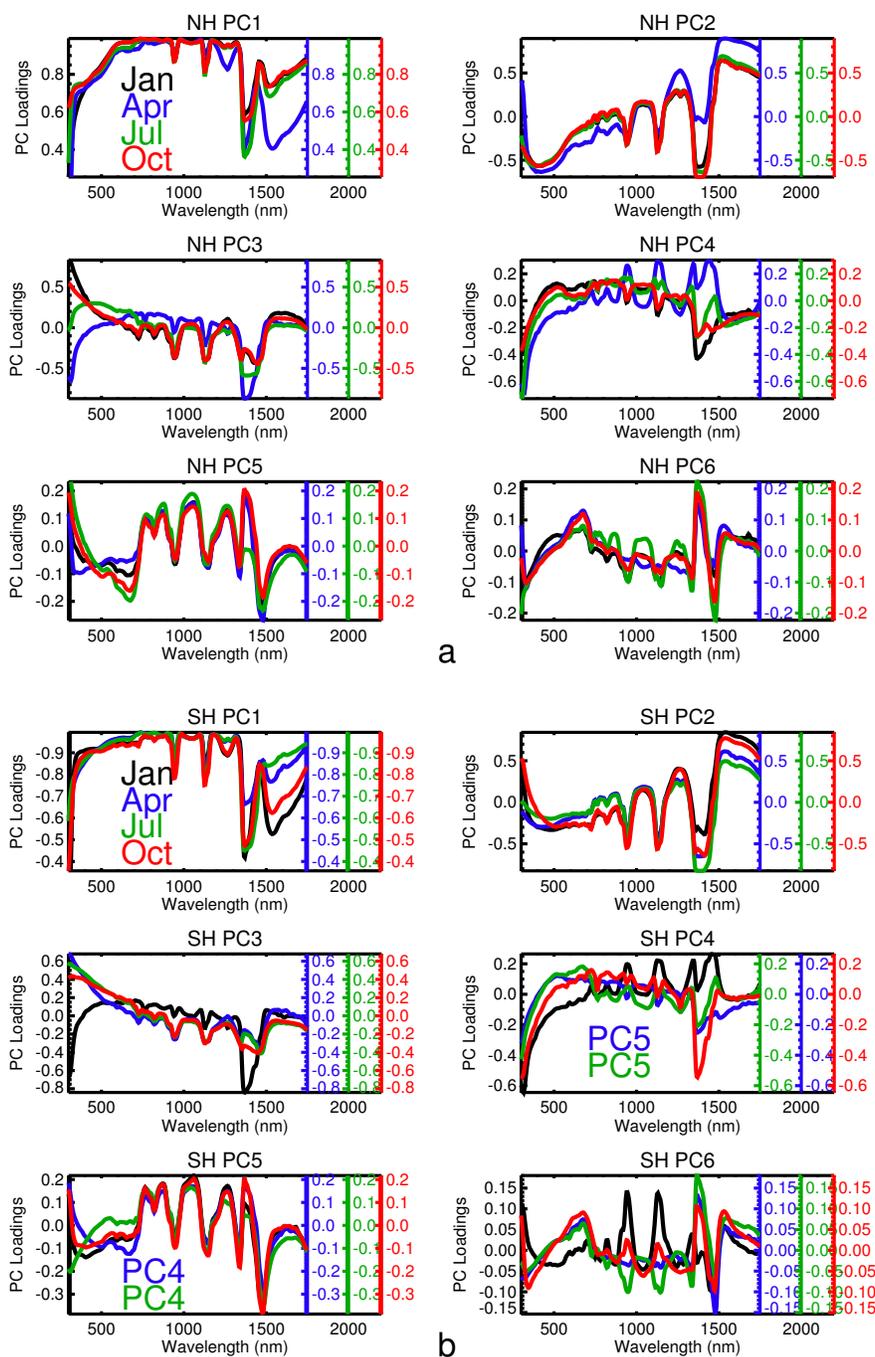


Figure 2.18: The first six PCs for the northern hemisphere (left) and southern hemisphere (right). Spectral shapes of the NH PCs and the globally calculated PCs (shown in Figure 2.5) are more similar than the comparison between the SH and global PCs. Between the two hemispheres there are also differences among the PCs. The most notable difference is that the "vegetation PC" explains more of the data variance in April and July. We draw this conclusion because vegetation is represented by the fourth PC in April and July and by the fifth PC in January and October. In the NH vegetation is consistently represented by PC5 for all four months.

That vegetation explains more of the variance in April and July is supported by recent results reporting greening observed in the dry season, possibly linked to an increase in sunlight [*Huete et al.*, 2006]. Further evidence of vegetation is found in the spectral shape of the fourth and fifth PCs for the months not discussed directly above (middle-right panel in Figure 2.18b) and PC6 where the NIR edge is evident (for example, October PC6), but the spatial and spectral vegetation identifiers are most dominant in the April and July PC4s and October and January PC5s. We see evidence of this physical variable in several PCs likely because of the orthogonality constraint in PCA discussed in Section 2.3.2.

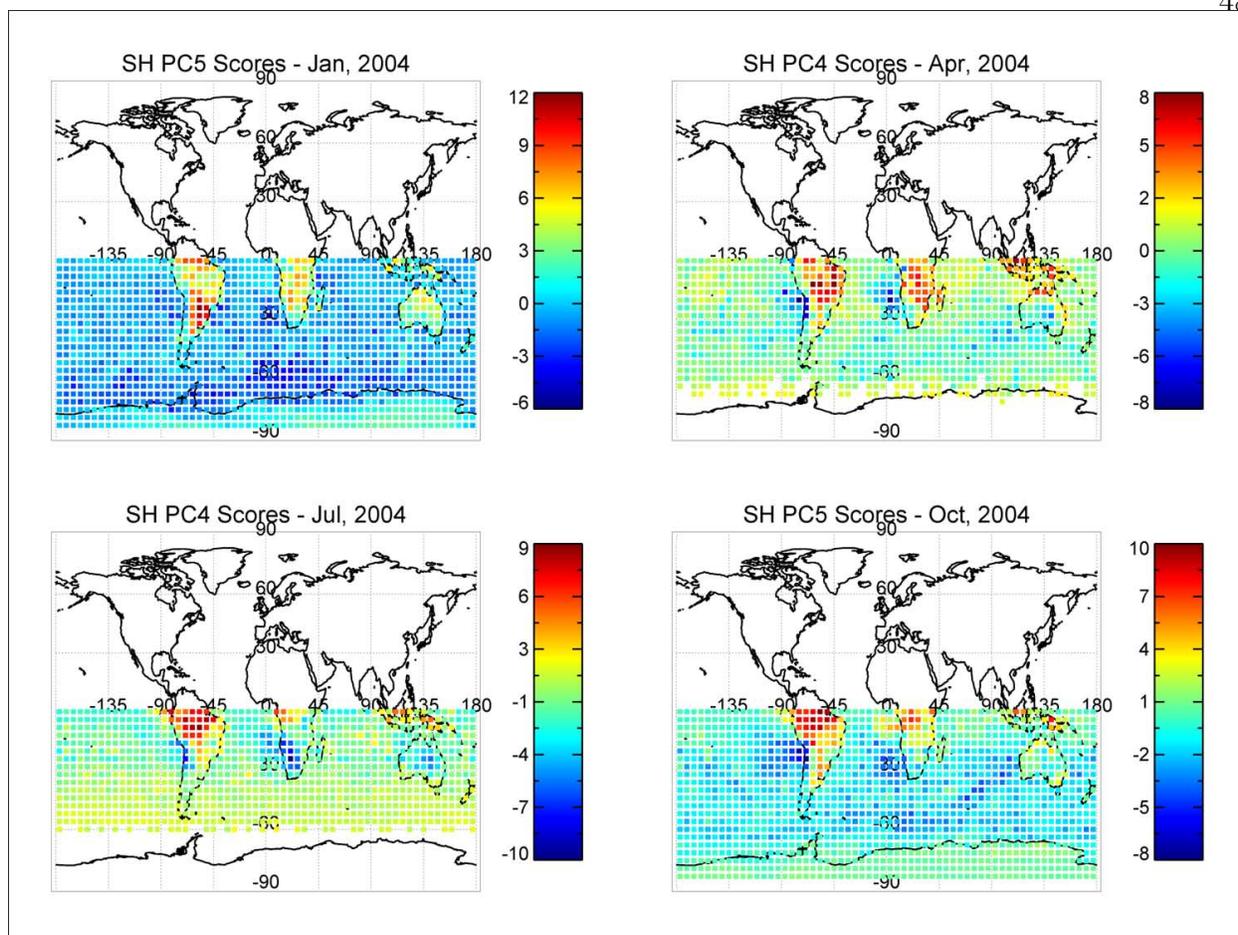


Figure 2.19: The SH scores corresponding to the PCs in each of the four month that exhibit a vegetation signal are shown here. In January and October, this is PC5 as in the global case, but in April and July, this is PC4, which implies that in April and July, vegetation explains more of the variance than in October and January.

2.4.3 Zonally Averaged Regional Variability

In Sections 2.4.1 and 2.4.2 we have shown the utility of studying the spectral and spatial variability of hyperspectral radiance; PCA of hyperspectral radiance can also be used to study the spectral and temporal characteristics of a data set, which is the topic of the following two sections. We averaged land and ocean radiances separately within 10° zonal bands to study the temporal variability of hyperspectral radiances within each zonal band. Two examples will be presented here: 40°N – 50°N and 10°S – 20°S , representing a northern hemi-

sphere mid-latitude region and a southern hemisphere tropical region, respectively. The PCs were calculated for these two zonal regions individually and are shown in Figure 2.20. The corresponding scores time series for land and ocean are shown for each region in Figure 2.21.

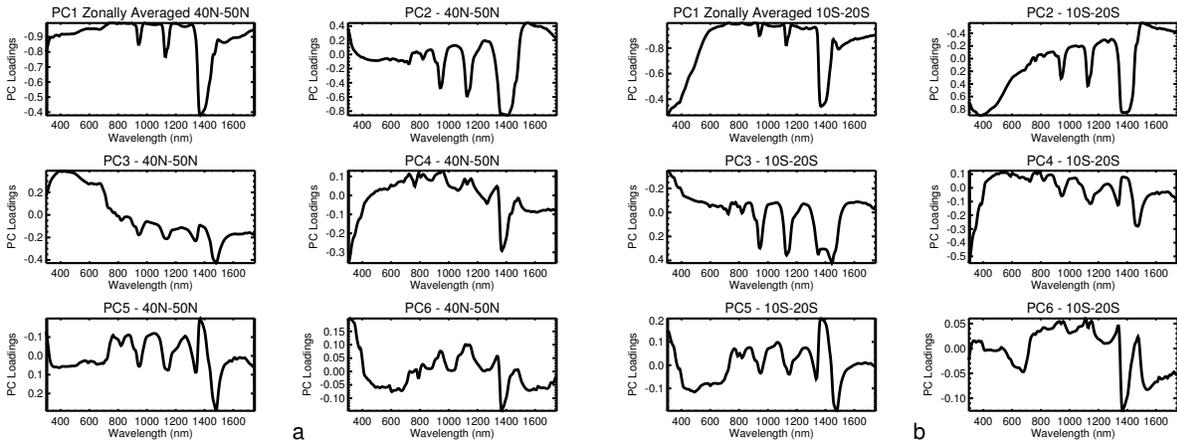


Figure 2.20: The zonally averaged PCs for two regions: a) shows the PCs for a 10° region in the northern hemisphere mid-latitudes between 40°N and 50°N and b) shows the PCs for a 10° region in the southern hemisphere Tropics between 10°S and 20°S . We see similarities between the two regions in PC1, PC2, PC4, and PC5, but there are also some aspects of each of these PCs and the general shapes of the other two not listed above that exhibit differences between the regions as well.

Figure 2.20 shows that the subspaces defined by the first six PCs from each zonal band are characterized by different PCs. Some PCs retain the same general shape in both regions: PC1 resembles a cloud reflectance spectrum, water vapor bands are still identifiable in most PCs, and vegetation still has a clear spectral signal in PC5. There are other PCs, however, that have different spectral shapes between the two regions, which implies that some of the dominant spectral variables in the NH mid-latitudes and the SH tropics are different. We further examine these PCs by studying their associating PC time series. The time series shown in Figure 2.21 were smoothed to emphasize the general patterns.

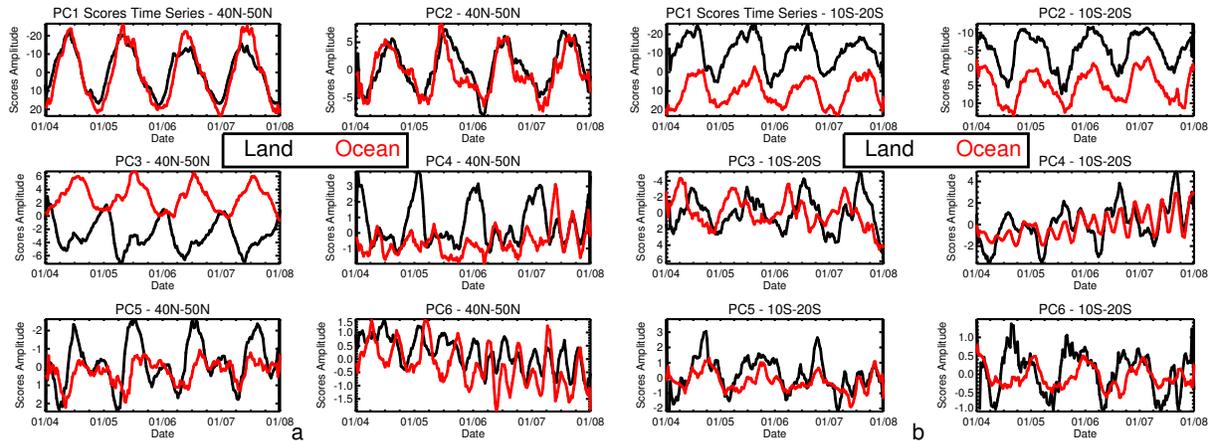


Figure 2.21: The scores corresponding to the PCs shown in Figure 2.20 are shown here. The land and ocean scores were separated. a) shows the scores corresponding to the PCs calculated for the $40^{\circ}\text{N} - 50^{\circ}\text{N}$ region, and b) shows the scores corresponding to the $10^{\circ}\text{S} - 20^{\circ}\text{S}$ region. Here we can see differences between the land and ocean temporal behavior of the scores. The PC6 scores in a) and the PC4 scores in b) demonstrate that the PCs are also able to capture secular trends.

The land and ocean time series are shown separately for each PC because for some components, we expect there to be differences in how the scores for each surface type behave over time. As an example of examining the land and ocean time series separately, we see that the annual variation in the NH mid-latitude PC5 time series has a larger amplitude over land than over ocean. This is expected because we do not expect to see a vegetation-related seasonal signal over ocean. We do still see annual and seasonal variability in the PC5 ocean time series indicating that there are likely other physical variables represented by this PC that exist over ocean as well, such as water vapor absorption. These time series also demonstrate that the PCs are able to extract annual variability (PC1 in both regions), seasonal variability (PC4 in both regions), and secular trends (PC6 in mid-latitudes and PC4 in the tropics). Even though we may not be able to decipher the meaning of these time series from these plots, this exercise shows that using PCs to study the temporal variability of spectral variables over time can also provide insight into studying how climate is changing using hyperspectral radiances. A more specific example of this technique is presented in the

following section.

2.4.4 Case Study: Arctic Ocean Variability

Using SCIAMACHY radiances over the Arctic Ocean we extend the temporal work showcased in Section 2.4.3 to also show that multivariate analysis techniques can provide insight into interpreting regional temporal variability. Northern hemisphere sea ice extent is a harbinger for climate change in the Arctic [*Francis et al.*, 2005]. Climate change in the Arctic also signals global change, as evidenced by the correlations between anomalies in surface skin temperature (the equivalent blackbody temperature of the Arctic surface as it radiates directly out to space [*Minnis and Khaiyer*, 2000]), cloud fraction, and precipitable water with various climate indices, such as ENSO [*Wang and Key*, 2005]. More work needs to be done, however, to understand the extent to which changes in the Arctic are caused by local or global climate changes [*Wang and Key*, 2005]. In 2007, sea ice extent in the Arctic reached a minimum since global measurements of sea ice extent from satellite were initiated in 1966 [*Vinnikov et al.*, 1999]. The Arctic has not been seasonally sea ice-free for about 800,000 years [*Overpeck et al.*, 2005], leading scientists to believe that the recent minima are among the first signs of anthropogenic climate effects. It is therefore important to continue to study the magnitude and rate of this climate signal. PCA was used to examine the variance in hyperspectral radiances in the Arctic and their utility as indicators of climate change.

We examined reflected radiances over the Arctic to study how sea ice is changing in this region. This study was restricted to the Arctic Ocean north of 70°N, where sea ice melts and refreezes each year. The seasonal behavior of sea ice in this region was determined using the sea ice coverage maps derived from measurements made by the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) [*Spreeen et al.*, 2008]. Radiances were averaged from one day each week during sunlit period of the year between mid-spring (April) and late summer (September), 25 weeks per year between 2004 and 2007. The data from each day were averaged within twelve grid boxes that contained near-equal ocean areas (on average

1.6 million km²). The three-dimensional data set was comprised of 12 regions, 100 time steps, and 291 wavelengths (15 nm FWHM and 5 nm spectral sampling).

To determine the dominant spectral variables in the data set, the principal components were calculated using all the averaged spectral radiances in all 12 regions (1200 spectra total). These results are shown in Figure 2.22. The first PC explains 78.6% of the data variance and appears to be dominated by cloud variability, identical to the analysis shown in Figure 2.4. The second PC explains 12.7% of the data variance and is dominated by sea ice reflectance, illustrated in Figure 2.23 where PC2 is compared to a snow albedo spectrum from Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) [Schmidt *et al.*, 2009]. We conclude that PC2 is likely dominated by sea ice reflectance because we only included ocean pixels in our Arctic analysis. The third PC does not have spectral signature of any distinct single physical variable, but water vapor is evident in this component. The fourth component has no clear explanation either. The third and fourth components explain 5.9% and 1.9% of the data variance, respectively. These four components together explain 99.1% of the variance in the Arctic Ocean data set and were found to be statistically significantly different from each other by the North *et al.* Rule of Thumb.

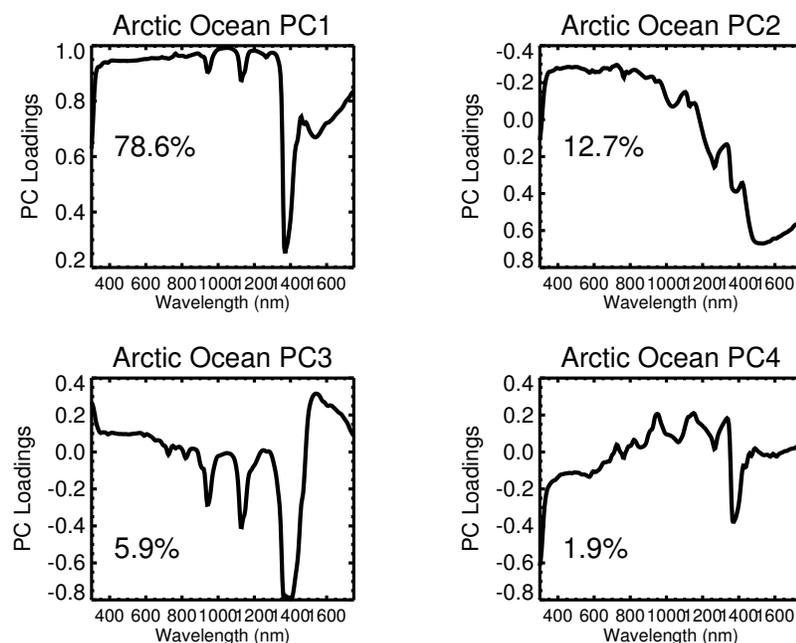


Figure 2.22: The PCs calculated over the 12 regions into which the Arctic Ocean was divided. Cloud reflectance (PC1), sea ice (PC2), and water vapor absorption bands (PC1 and PC3) are detectable from these components.

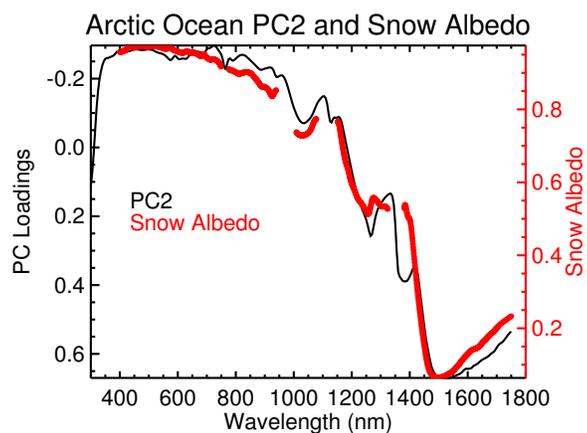


Figure 2.23: The spectral shape of PC2 is compared to that of a snow albedo spectrum from ARCTAS to show the close comparison between the two. Because we only included ocean pixels in our analysis of the Arctic region, we assume that any snow or ice found is associated with sea ice cover. This also implies that the temporal scores from this component can be used to understand how the PC2 scores represent the temporal pattern of sea ice cover.

To provide an additional link between PC2 and sea ice variability we compared the temporal behavior of PC2 to that of sea ice extent from AMSR-E. We applied multichannel singular spectrum analysis (MSSA) to all twelve PC2 time series to determine the dominant temporal patterns of this component. There was data for 100 time periods total, with 25 samples per year, so the window length was set to 30 in order to capture the expected annual period, $1/f$, of 25 time steps. The first two temporal eigenmodes exhibited an annual pattern, so the scores from an area roughly equivalent to the Beaufort Sea region were reconstructed using these modes. This is shown with the blue line in Figure 2.24. Any of the other eleven PC2 time series could have been reconstructed with the first two temporal modes as well, but this region is presented here as an example. Figure 2.24 also shows the original time series of PC2 scores from the Beaufort Sea area (black dots) and the total Arctic sea ice extent from AMSR-E (red line) [Spreen *et al.*, 2008]. Differences between the end of one annual time series and the start of the next occur because there was insufficient SCIAMACHY data during the six-month Arctic night, between the last data point at the end of one year's summer and the first data point for the following year in mid-spring. The missing data during this time period likely also contributes to the phase differences seen between the PC2 and AMSR-E time series. Nevertheless, similarities can be seen between the Beaufort Sea PC2 scores, and the AMSR-E sea ice extent. The maxima and minima of the PC2 scores occur at similar times of year as the sea ice extent maxima and minima. The sea ice melt in the summer is also represented well by the black dots. The MSSA reconstruction shows a smoothed version of the PC2 scores because the MSSA dominant patterns were the annual periods present in the PC2 scores. Although differences occur in part because the AMSR-E sea ice extent data used sea ice concentrations over the entire Arctic Ocean region, not just the Beaufort Sea area, this allows us to more clearly see how the maxima, minima, and slopes of the scores each year align with the sea ice extent periods over this time period.

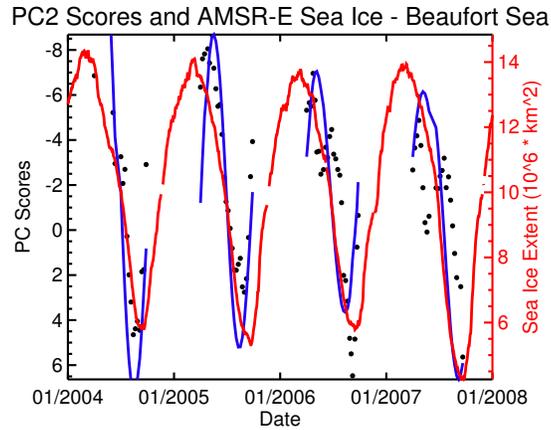


Figure 2.24: The Beaufort Sea PC2 scores (black) and the reconstruction of these scores (blue) are compared to the total Arctic Ocean sea ice extent from AMSR-E (red). We see the close similarity between the PC2 temporal behavior and that of the Arctic sea ice extent, exhibiting the benefit of using PC scores to study the temporal behavior of physical variables.

2.5 Summary

In this chapter, we quantified the variance in SCIAMACHY measurements of reflected hyperspectral solar radiance to determine the dominant spectral contributions to the measured variance and to detect trends that can be used as sufficiently sensitive, confident indicators of climate change, potentially linking spectral signatures and trends to physical causes. The principal components of SCIAMACHY shortwave (near ultraviolet through near infrared) hyperspectral radiances were derived from spatially gridded monthly averaged global and hemispherical data, zonally averaged data in the northern hemisphere mid-latitudes and southern hemisphere mid-latitudes, and ocean-gridded data over the Arctic Ocean. As few as six components were needed to explain over 99.5% of the variance in all cases with the exception of the Arctic Ocean case where only four components were needed, both of which are a large variable reduction from the 291 spectral radiances prior to the PCA transformation. There were similarities in the spectral shapes of the PCs and their relative variance contributions exhibited among all of these regions. Although physical variables are easily identified in some of these PCs because of common spectral features, we also identified dif-

ferences among the spectral shapes of these PCs in different cases. We see an example of this in Figures 2.20a and 2.20b where, PC1 exhibits a difference in slope in the near UV and short visible wavelengths. The physical variables identified from some of these PCs are discussed in more detail below.

Attributing physical processes to the principal components has begun to provide an understanding of the factors that drive changes in Earth's climate through the study of hyperspectral radiance data variability. Cloud reflectance and water vapor absorption were repeatedly represented by the first principal component for all of the cases considered here. Water vapor absorption bands were also consistently present in several PCs in addition to the first. Vegetation reflectance was also consistently present in the PCs for those cases over land. In the southern hemisphere vegetation makes a larger contribution to the variance during the Amazon dry season than during the wet season, in agreement with the results from *Huete et al.* [2006]. Molecular scattering was evident in the global monthly PCs. While the origin of this feature is a response to solar geometry and thus not a climatically significant variable, it reveals the success of the algorithms in separating independent sources of variability. In addition to linking some of the PCs to the physical variables by their spectral shapes, the spatial and temporal distributions of component scores were well correlated with known physical variables, providing confidence that these methods can be used to determine how these variables are changing over time. We also illustrated the benefit of rotating the dominant PCs to further isolate the physical variables represented in that subspace. Attributing physical variables to the dominant spectral variables determined is a key part of understanding the role these variables play in Earth's climate.

In the Arctic Ocean case cloud reflectance and sea ice reflectance were clearly separable. It is difficult to discriminate between snow and ice-covered surfaces and overlying clouds using passive remote sensing in the visible because of the spectral similarity between clouds and snow or ice-covered surfaces [*Gao et al.*, 1998]. Comparing the shapes of the Arctic Ocean PC1 and PC2 in the visible we can see this similarity. The linear combinations of

hyperspectral radiance determined by PCA made this independent distinction between cloud and the sea ice surface possible.

These results highlight the importance of using hyperspectral data to extract dominant spectral climate variables. Even though only four to six PCs are typically needed to explain over 99% of the variance in the data, each PC is a linear combination of all 291 spectral radiances between 300 and 1750 nm, the spectral range used in this study. That is, the combinations of radiances in all 291 wavelength bands define the primary PCs, which is not identical to four to six wavelength bands yielding the four to six PCs extracted in each of the cases presented here. The extraction of the spectral signals of known climate drivers shows that PCA of hyperspectral radiances can identify these signals without the need to make assumptions about the data prior to applying PCA, as is necessary when studying processes using multispectral data. This also shows that climate variables are detectable in the directly observed hyperspectral radiances and may indicate that these radiances can be used to study climate change.

As we have shown in this chapter, multivariate analysis can be used to extract climate-relevant information from hyperspectral reflected radiances in the visible and near infrared. One important application of the analysis techniques presented in this chapter is the quantitative comparison of the spectral variability of modeled hyperspectral radiances to observed hyperspectral radiances. Climate models have helped us to understand the relationship between different components within our climate system and give us the flexibility to study how climate variables are changing over longer periods of time than for which we have measurements. *Feldman et al.* [2011a] conducted Observation System Simulation Experiments (OSSEs) to simulate the measurements from a hyperspectral climate-observing shortwave imager, demonstrating that twenty-first century climate projections are detectable in reflected shortwave spectral radiance. Quantitative comparisons between the output from climate and radiative transfer models, such as from the OSSEs described by *Feldman et al.* [2011a], and observed hyperspectral data using multivariate analysis techniques such as those

applied in this chapter aid in evaluating how well these models reproduce the spectral variability within Earth's climate system. *Haskins et al.* [1997], *Haskins et al.* [1999], and *Huang et al.* [2002] have used PCA to evaluate the validity of climate models relative to longwave radiation measurements; similar techniques may be applied to shortwave hyperspectral measurements. One main difference between the PCA applied in this study and the PCA results used for such quantitative comparisons is that the radiances from both data sets must remain unstandardized prior to PCA [*Preisendorfer and Mobley*, 1988]. PCA can be applied to hyperspectral radiances from climate model output to study how the spatial variability of the dominant spectral variables changes over time. PCA combined with SSA/MSSA will improve our understanding of the temporal variability of these variables as well. Linking these multivariate analysis results to climate model inputs provides the additional benefit of relating the dominant spectral variables to the individual inputs and can be used to develop future retrieval algorithms. We will further explore this application in a forthcoming study.

Chapter 3

Quantitative Comparison of the Variability in Observed and Simulated Shortwave Reflectance

The Climate Absolute Radiance and Refractivity Observatory (CLARREO) is a climate observation system that has been designed to monitor the Earth's climate with unprecedented absolute radiometric accuracy and SI traceability. Climate Observation System Simulation Experiments (OSSEs) have been generated to simulate CLARREO hyperspectral shortwave imager measurements to help define the measurement characteristics needed for CLARREO to achieve its objectives. To evaluate how well the OSSE-simulated reflectance spectra reproduce the Earth's climate variability at the beginning of the 21st century, we compared the variability of the OSSE reflectance spectra to that of the reflectance spectra measured by the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY). Principal component analysis (PCA) is a multivariate spectral decomposition technique used to represent and study the variability of hyperspectral radiation measurements. Using PCA, between 99.7% and 99.9% of the total variance the OSSE and SCIAMACHY data sets can be explained by subspaces defined by six principal components (PCs). To quantify how much information is shared between the simulated and observed data sets, we spectrally decomposed the intersection of the two data set subspaces. The results from four cases in 2004 showed that the two data sets share eight (January and October) and seven (April and July) dimensions, which correspond to about 99.9% of the total SCIAMACHY variance for each month. The spectral nature of these shared spaces, under-

stood by examining the transformed eigenvectors calculated from the subspace intersections, exhibit similar physical characteristics to the original PCs calculated from each data sets, such as water vapor absorption, vegetation reflectance, and cloud reflectance.

3.1 Introduction

Spatial and temporal local variations in the net radiative forcing (the difference between the absorbed solar irradiance and the emitted longwave irradiance at the top of the tropopause) drive the Earth’s weather and climate. The Earth’s climate system is approximately in planetary radiative equilibrium when the net radiative forcing is zero over global and annual scales. *Hansen et al.* [2011] estimated the global radiative imbalance between 2005 and 2010 to be $0.58 \pm 0.15 \text{ Wm}^{-2}$, which will cause the Earth’s temperature to increase until the Earth returns to radiative equilibrium. A global mean radiative imbalance can be caused by changes in several climate variables, such as greenhouse gases, land surface type, aerosol properties, and incoming solar irradiance [*IPCC*, 2007]. Identifying radiative imbalances, accurately assessing their magnitude, and attributing those imbalances to atmospheric and surface variables will aid in the understanding of how and why the Earth’s climate is changing.

Absorbed solar radiation, one of the primary components in the global radiative energy budget is the difference between the incident top-of-atmosphere solar irradiance and reflected solar irradiance. Therefore, in addition to monitoring the downwelling top of the atmosphere solar irradiance at high accuracy, currently with the Solar Radiation and Climate Experiment (SORCE) [*Rottman*, 2005] Total Irradiance Monitor [*Kopp et al.*, 2005], monitoring and studying the spatial, temporal, and spectral variability of the Earth-reflected solar radiation is a key component to understanding changes in Earth’s climate.

Reflected solar radiation from Earth contains information about several variables relevant to changes in Earth’s climate, including cloud properties, aerosols, land surface albedo, and sea ice [*National Research Council*, 2007, *Loeb et al.*, 2007, *Roberts et al.*, 2011, *Wielicki*

et al., Submitted]. Changes in these and other atmospheric and surface variables impact the spectral, spatial, and temporal variability of reflected solar radiation through spectrally dependent scattering and absorption processes. Monitoring solar reflectance from space to study climate requires highly accurate, hyperspectral measurements [*Wielicki et al.*, Submitted]. In this context hyperspectral refers to spectrally contiguous, overlapping spectral radiation measurements [*Goetz et al.*, 1985, *Goetz*, 2009]; solar (shortwave) radiation includes wavelengths ranging from the near ultraviolet to the near infrared, 300-2500 nm, accounting for approximately 95% of the solar radiation incident at the top of the atmosphere. Since the 1970s, the information in shortwave hyperspectral measurements has facilitated the identification of individual surface materials and the application of sophisticated atmospheric correction techniques to obtain surface spectral reflectance [*Goetz*, 2009]. The information about Earth's surface and atmospheric properties in space-based hyperspectral shortwave measurements can also be used in climate change detection and attribution studies. This information in spectrally resolved shortwave radiation can be used to understand the variability of the climate system using spectral decomposition techniques such as principal component analysis (PCA) [*Rabbette and Pilewskie*, 2001, 2002, *Grenfell and Perovich*, 2008, *Roberts et al.*, 2011].

Roberts et al. [2011] quantified the spectral variability of Earth-reflected hyperspectral solar radiance to study the information contained in direct satellite measurements for climate change detection and attribution. Although complete separation of all atmospheric and surface variables represented in a reflectance spectrum is challenging even using information-rich hyperspectral measurements, occasionally it is possible to spectrally identify the physical variance drivers such as clouds, sea ice, and vegetation using spectral decomposition techniques [*Rabbette and Pilewskie*, 2001, 2002, *Huang and Yung*, 2005, *Roberts et al.*, 2011]. For example, *Roberts et al.* [2011] applied PCA to Arctic Ocean radiance spectra, separating contributions to the data variance from clouds and sea ice. These results demonstrated that hyperspectral reflected radiation contains physical information about the Earth's climate

system that can be extracted with multivariate spectral decomposition techniques.

Highly accurate climate observation systems are being designed that will include spectrally resolved measurements in the visible and near infrared. Such systems include the Climate Absolute Radiance and Refractivity Observatory (CLARREO) [*National Research Council, 2007, Wielicki et al., Submitted*] and the Traceable Radiometry Underpinning Terrestrial- and Helio- Studies (TRUTHS) [*Fox et al., 2003, 2011*]. The shortwave instruments proposed by both of these projects will provide high spectral resolution measurements with unprecedented absolute radiometric accuracy and SI traceability.

Feldman et al. [2011a] designed a climate Observation System Simulation Experiment (OSSE) as a CLARREO shortwave instrument emulator used to derive measurement and mission requirements. For the OSSE, *Feldman et al.* [2011a] used global climate model output with a radiative transfer model to simulate CLARREO shortwave instrument reflectance measurements. By comparing OSSE output from forced and unforced scenarios, changes in variables such as clouds, aerosols, sea ice, and snow cover were evident in zonally averaged spectra, implying that spectrally resolved reflectance may be capable of detecting changes in key climate variables by the middle and end of the 21st century [*Feldman et al., 2011a*]. Using the OSSE, *Feldman et al.* [2011b] also found that spectrally resolved reflectance improves time-to-detection over broadband shortwave measurements. The results from the climate OSSE studies further support the need for a highly accurate climate observation system that includes hyperspectral shortwave reflectance measurements [*Feldman et al., 2011a, Feldman et al., 2011b, Wielicki et al., Submitted*].

The climate OSSE is a powerful tool; however, we need to evaluate how realistic the variability of these simulated spectral reflectance spectra is relative to observations of spectral reflectance. The ability of the OSSE to reproduce present-day climate variability is necessary to use OSSE simulations to make confident statements about climate change detection and attribution. There is still the possibility, however, that even if the OSSE is able to meet this necessary condition of reproducing present day climate variability, its twenty-first century

climate change predictions may not be realistic depending on how well the underlying climate model simulates future changes in climate. The spectral variability of simulated and observed hyperspectral reflectance can be compared both qualitatively [*Feldman et al.*, 2011b], and quantitatively, by the methods presented here.

In this chapter, we evaluate how well simulated shortwave hyperspectral reflectance reproduces the variability in satellite-measured reflectance. To address this question we will explore the utility of the variability of shortwave reflectance to serve as an appropriate measure of the similarity between two data sets. *Roberts et al.* [2011] showed that it is possible to extract physical variables from directly measured radiance using principal component analysis rather than inverse modeling techniques or any other model-based analysis. Therefore, we compare the variability of measured and simulated reflectance using PCA and other multivariate analysis techniques to quantify their similarity.

The next section is an overview of the observed and simulated reflectance spectra used in this chapter. Section 3.3 details the multivariate techniques used in the comparisons. In Section 3.4, we present an example that exhibits the quantitative comparison techniques with data, and Section 3.5 provides a summary of the study, conclusions, and a discussion of future work.

3.2 Data

3.2.1 Observed Reflectance: SCIAMACHY Measurements

This chapter uses hyperspectral reflectance measured by the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) [*Bovensmann et al.*, 1999]. SCIAMACHY flew on the European Space Agency’s ENVISAT (Environmental Satellite), a sun-synchronous satellite in near polar orbit that operated from March 2002 to April 2012. In May 2012, the European Space Agency declared the official end of the ENVISAT mission after a spacecraft failure in April 2012. SCIAMACHY was designed to study the effect

of natural and anthropogenic sources on global atmospheric composition [Gottwald *et al.*, 2011a]. Additional objectives included understanding the global distribution, chemistry, and physics of trace gases, aerosols, and clouds in the troposphere, stratosphere, and mesosphere [Bovensmann *et al.*, 1999]. SCIAMACHY measured across eight channels covering the spectral ranges 214-1773 nm, 1934-2044 nm, and 2259-2386 nm with spectral resolutions ranging between 0.22 nm and 1.48 nm [Gottwald *et al.*, 2011b]. Ice deposited on channels seven and eight (spanning 1934-2386 nm), interfering with the optical throughput [Gottwald *et al.*, 2011c]. For the present study, analysis is restricted to the wavelength range 300 nm to 1750 nm. Nadir pixel size is dependent on the integration time and swath width, causing footprint sizes to vary between 26 km (along track) by 30 km (across track) and 32 km (along track) by 930 km (across track). For nadir sampling, SCIAMACHY has a scanning angular width of $\pm 32^\circ$ across track, which corresponds to a maximum nadir swath width of 960 km [Gottwald *et al.*, 2011b]. The measurement characteristics of SCIAMACHY make it the best candidate to compare space-based observations of shortwave reflectance with climate OSSE-simulated reflectance spectra.

SCIAMACHY Reflectance Spectra Figure 3.1 presents three examples of SCIAMACHY reflectance spectra for different scene types: a thick cloud (black), cloud-free green vegetation (blue), and cloud-free ocean (red). Spectral reflectance, R_λ , is defined by,

$$R_\lambda = \frac{\pi I_\lambda^\uparrow}{\cos(\theta) F_\lambda^\downarrow} \quad (3.1)$$

where I_λ^\uparrow is reflected spectral radiance, F_λ^\downarrow is the incident spectral solar irradiance at the top of the atmosphere, and θ the solar zenith angle. Similarities among the three spectra in Figure 3.1 include absorption features such as the Oxygen-A band centered around 762 nm and water absorption bands centered at 940 nm, 1140 nm, and 1350 nm. There are also differences among these reflectance spectra that are characteristic of the different scenes. Throughout the visible and much of the near infrared outside water absorption

bands, the reflectance values for the thick cloud spectrum are generally higher than the two surface reflectance spectra shown. For a cloud-free ocean spectrum measured at nadir, the reflectance values are low throughout the spectral range except between 300 nm to 400 nm where atmospheric molecular and aerosol scattering increase reflectance. The spectrum measured over green vegetation has low reflectance in the visible with a local maximum known as the "green peak", centered around 550 nm. The vegetation "red edge" is the increase in reflectance between 700 and 750 nm. These examples of reflectance serve as a frame of reference when examining the spectral shapes of principal components, such as those shown in Figure 3.4, which often resemble reflectance spectra, albeit often with a linear and nonlinear mixture of source signatures.

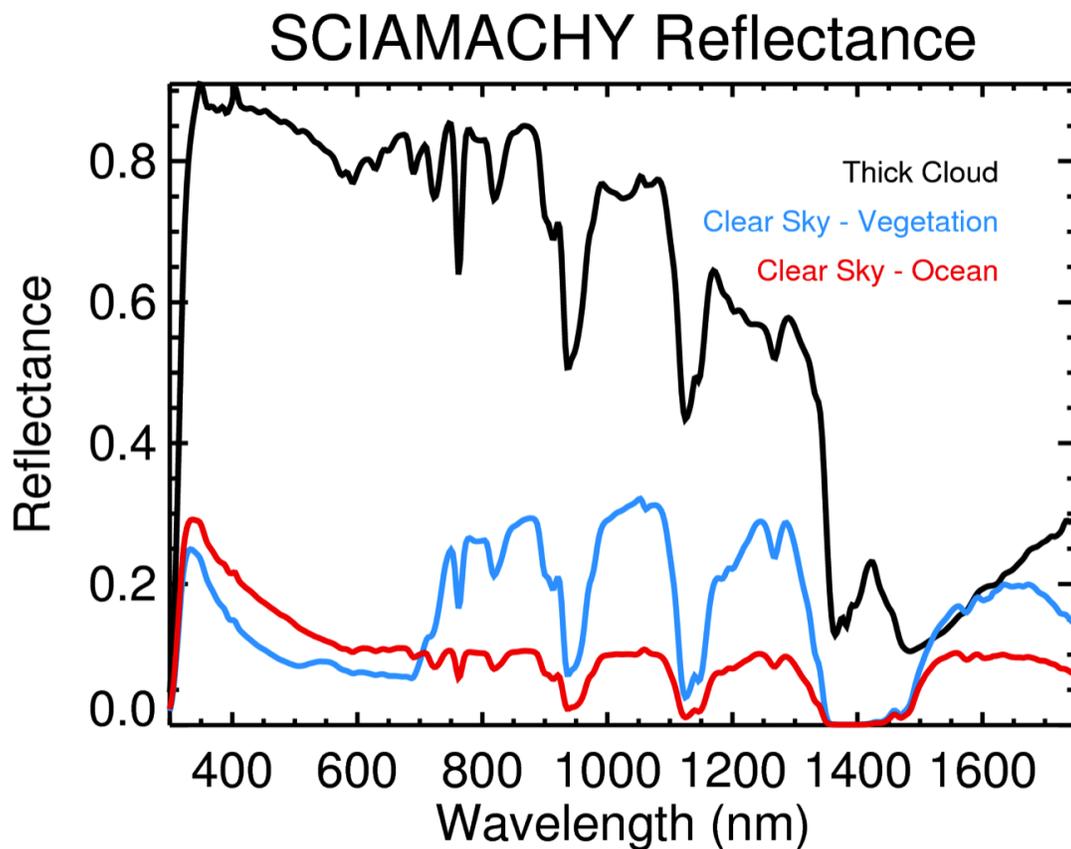


Figure 3.1: SCIAMACHY-measured reflectance spectra for three scene types: a thick cloud (black), cloud-free vegetation (blue), and cloud-free ocean (red).

3.2.2 Simulated Reflectance: Observation System Simulation Experiments

Feldman et al. [2011a] constructed OSSEs using input from the Community Climate System Model version 3.0 (CCSM) (Collins et al., 2006) Global Climate Model and using MODTRAN 5.3 [Anderson et al., 1999] to simulate CLARREO shortwave spectral reflectance measurements during the twenty-first century. Monthly averaged fields from two IPCC AR4 emission scenarios were used to produce the OSSE results. The all-sky (cloud-inclusive scenes) reflectance spectra used in this study were simulated using the unforced constant CO_2 emission scenario model results, in which well-mixed radiatively active atmospheric greenhouse gases and aerosols were held constant at levels observed in the year 2000 throughout the model run [Meehl et al., 2005, 2007]. Results from the forced A2 emission scenario [IPCC, 2007], in which concentrations of well-mixed greenhouse gases were steadily increased to the year 2050 then reduced to 1900 levels by the year 2100, were also used to simulate reflectance using the OSSE [Feldman et al., 2011a]. Changes in the climate system over the course of the century include a tripling of CO_2 relative to pre-industrial levels, surface warming, and decreases in snow and ice cover. In the present study we have used the unforced scenario results because we are comparing individual months at the beginning of the 21st century, when differences between the two scenarios are minimal.

3.3 Quantitative Multivariate Methods

3.3.1 Principal Component Analysis

Principal component analysis (PCA) is a spectral decomposition technique used to quantify the variance distribution in a multivariate data set. The principal components (PCs) are linear combinations of the original variables, in this case the spectral reflectance values. If the original variables are correlated (as is generally the case with reflectance spectra), PCA can significantly reduce the number of variables needed to explain the majority of the variance in a data set. *Roberts et al.* [2011] presented a literature review of how PCA

has been used in various atmospheric science applications, specifically to understand the variability of spectral radiation in the longwave (longer than $4\ \mu\text{m}$) and the shortwave (350 nm - 2500 nm).

PCA Input Shortwave reflected radiation provides a number of options for what form of the data to use in PCA, each of which has its own advantages. *Roberts et al.* [2011] used standardized spectral radiance rather than unstandardized radiance because of the large differences in spectral variance. Wavelength bands with the most variability in spectral radiance (largest spectral variance) dominate the spectral shapes of the PCs. By standardizing the data, principal components typically represent physical variables more prominently than PCs calculated from unstandardized data [*Preisendorfer and Mobley*, 1988]. Standardized radiances are calculated by spectrally mean-centering the radiances and normalizing them by the spectral standard deviation. The standardized PCs can be used when comparing two data sets, but not to quantify how much information is shared between them. Unstandardized data must be used in quantitative comparisons because the standard deviation contains pertinent information about each data set. *Haskins et al.* [1999] compared both standardized and unstandardized infrared radiance PCs, taking advantage of both the qualities of the standardized and unstandardized principal components. When the information in the standard deviation is removed, it becomes difficult to identify genuine differences and similarities between two data sets, the main purpose of this study.

Using reflectance rather than radiance is one way to avoid using standardized radiance while retaining the information in the standard deviation. Normalization by incident solar irradiance removes this known and dominant spectral shape from the PCs, leaving the influences due to atmosphere and surface properties. In the case of SCIAMACHY, the calculation of reflectance may also remove instrument anomalies because the downwelling TOA solar irradiance and Earth-reflected radiance are measured using the same sensors, in part diminishing the impact of spectrally dependent noise on the PCA results.

PCA Description The first step in principal component analysis is calculating the covariance matrix, \mathbf{C} , from the spectrally mean-centered reflectance spectra. Using a spectral decomposition technique, the eigenvalues (Λ) and eigenvectors (\mathbf{E}) of the covariance matrix are determined such that they satisfy the characteristic equation, $\mathbf{C}\mathbf{E} = \Lambda\mathbf{E}$, where \mathbf{E} is a $K \times K$ matrix, and Λ is a $K \times K$ diagonal matrix composed of the eigenvalues (λ). K is the number of variables, or spectral bands, contained in each spectrum. Each eigenvalue, λ_k , is the variance explained by each eigenvector. In this study we define the principal components (PCs) to be the eigenvectors scaled by the square root of their corresponding eigenvalues:

$$PC_k = \sqrt{\lambda}E_k \quad (3.2)$$

The spectral shapes of the principal components may provide insight into which physical variables are explained by each PC dimension. Projection of the mean-centered data onto the eigenvectors are the PC scores, the weighted averages of the input data with the eigenvectors as the weights. Depending on the spatial or temporal distribution of the data, the scores can be used to evaluate how the principal components vary in space or time. The physical significance of PCs is often difficult to determine because the PCs are linear combinations of the original variables. The original variables may have nonlinear dependencies on the physical variance drivers; however, some of the dominant physical variables can occasionally be identified. For a more detailed and mathematically rigorous description of PCA, consult *Roberts et al.* [2011] and *Jolliffe* [2002].

Boundary Between Data Signal and Noise There are several methods that can be applied to estimate the number of dimensions that define the boundary between signal and noise in a data set [*Jolliffe*, 2002]. *Cattell* [1966] suggested using a plot of the eigenvalues on a linear scale to determine the location of this boundary. It is identified graphically by visually locating the initial change in slope in the eigenvalue spectrum. *Craddock and Flood* [1969] presented a similar technique, but instead used a logarithmic eigenvalue scale,

a method that has been justified with the PCA of simulated data with known variance structures [Farmer, 1971]. In studying the eigenvalue spectrum calculated from PCA of solar radiation, we have found that the logarithmic plot of the eigenvalues is one of the best tools to identify how many PCs explain signal in the data. Kaiser [1960] suggests that all principal components associated with eigenvalues larger than the average eigenvalue explain the signal. A more liberal criterion can be used in which some fraction (typically 0.7) of the average eigenvalue is used as the cut-off, in an attempt to account for possible sampling variations [Jolliffe, 1972]. In the Broken Stick Method [Jolliffe, 2002], $\lambda_k > \frac{1}{p} \sum_{j=k}^p \frac{1}{j} \lambda_{avg}$ is the criterion for determining the location of the boundary, where λ_k is the k^{th} eigenvalue, p is the total number of variables, and $1 \leq k \leq p$. The subjectivity of locating this boundary is recognized by all the studies discussed here, but these suggested techniques often help to provide guidance in making this decision. The information provided by these selection criteria help us to make sense of the subspace comparison techniques applied to the two data sets below. An approximation of the boundary between signal and noise in these two data sets puts the results from the significance test into context as we will discuss in Section 3.4.

3.3.2 Quantitative Comparison Description

Comparing Spectral Variability in the Infrared Comparing the variability between two data sets of radiance data using second-moment statistics (statistics derived from the squared values of the data, such as variance) as an objective test of climate models is a technique that has been used by several others [Goody *et al.*, 1998]. For example, second-moment statistics have been used to evaluate climate model variability in temperature [Polyak, 1996] and longwave emission spectra [Haskins *et al.*, 1997, 1999, Huang *et al.*, 2002]. All of these studies state that the model needs to exhibit correct second-moment statistics with respect to sufficiently accurate measurements to be considered rigorously validated. Haskins *et al.* [1997] compared the variance contribution and the spectral shapes of

principal components to evaluate how well simulated infrared spectra reproduced the variability observed in Infrared Interferometer Spectrometer (IRIS) measurements. In a subsequent study, *Haskins et al.* [1999] inverted IRIS radiance principal components to derive cloud fraction, relative humidity, and temperature. Those principal component inversions quantified the constraints imposed upon climate models by infrared radiance measurements confirming that clouds are a dominant driver of the climate system and explain the largest fraction of the variance in the measured data [*Haskins et al.*, 1999]. *Haskins et al.* [1999] also concluded that if a model is unable to reproduce the observed cloud variability represented in the most dominant principal component, it is unlikely that it would simulate realistic changes in climate. *Huang et al.* [2002] combined principal component analysis with statistical regression techniques to quantify how well a GCM represented cloud variability relative to IRIS measurements and found that the model underestimated cloud variations by 2 to 6 times compared to measurements. The latter two studies converted the PCs of observed radiance to physical quantities using inversions and regression techniques to quantitatively evaluate the performance of climate models. In the present study, we only use the information provided by the principal components that explain nearly all the variance in the data set to evaluate simulated reflectance.

Quantitative Comparison using Subspace Intersection There are several methods that can be used to compare the variability of two data sets. Similar to variability analysis in the infrared introduced above, one method is to compare the spectral shapes of the components. This comparison can be helpful in that it is a preliminary, qualitative, representation of the relationship between the two data sets. Here, we examine the information shared by two data sets by applying a method similar to that used by *Goetz et al.* [1998] to develop a novel atmospheric correction lookup table method to retrieve AVIRIS surface reflectance. This method compared subspaces of the measured AVIRIS radiance spectra with that simulated by MODTRAN under a variety of atmospheric conditions. The

spectral decomposition of the intersection between these subspaces determined how many dimensions the two data sets shared. The intersection was used as a transformation between the two data sets providing the means to relate the simulated atmospheric conditions with those observed in the AVIRIS spectra. The *Goetz et al.* [1998] primary objective was to develop a computationally efficient method of atmospheric correction for surface reflectance studies. In the present study, a similar mathematical framework is applied to determine how much of the total variance is shared between two data sets as a quantitative measure of their similarity.

Quantitative methods similar to those presented in this chapter have been used in other areas of atmospheric science and other scientific fields to evaluate multivariate data [Krzanowski, 1979, Crone and Crosby, 1995]. For example, *Crone and Crosby* [1995] used the spectral decomposition of the intersection between subspaces of independent satellite measurements to determine their similarity. Determining the significance of the difference between two subspaces is instrumental for principal component regression because this decision gives guidance for determining if a subspace defined by one set of principal components is appropriate to explain the variability of another [Crone and Crosby, 1995, Jolliffe, 2002]. The ultimate goal of the present study is to evaluate one data set based on its relationship to another, so we employ similar spectral decomposition analysis techniques.

Mathematical Details of Intersection Decomposition The intersection comparison method described here is largely derived from a technique described by *Krzanowski* [1979] for comparing groups of principal components. First, we calculate the principal components. The following process is repeated p times, where $1 \leq k \leq p$, and p is some number less than the total number of PCs. Using the eigenvectors calculated from PCA, we calculate the intersection (**I**) between the two data sets using:

$$I = E_A E_B^T E_B E_A^T \quad (3.3)$$

The intersection will be a $k \times k$ square matrix. The eigenvector matrices (\mathbf{E}_A and \mathbf{E}_B) used to calculate the intersection are composed only of the k eigenvectors used to define the subspace. Singular value decomposition determines the eigenvalues (Γ) and eigenvectors (\mathbf{Y}) of \mathbf{I} . Because \mathbf{I} is a symmetrical matrix, the two sets of eigenvectors calculated in this decomposition are equivalent:

$$I = Y\Gamma Y^T \quad (3.4)$$

The eigenvalues on the diagonal of the $k \times k$ diagonal eigenvalue matrix (Γ) can also be represented as a vector, γ , k elements long.

The spectral decomposition provides information from which we can understand the amount of shared variance between the two subspaces. The eigenvector matrix, \mathbf{Y} , is used to determine the transformed eigenvector matrices for each data set in the shared intersecting space:

$$\begin{aligned} A &= E_A^T Y \\ B &= E_B^T E_B A \end{aligned} \quad (3.5)$$

Each of the k vectors in \mathbf{A} and \mathbf{B} are mutually orthogonal and are used to understand the spectral nature of the overlap between the two data sets.

The eigenvalues in γ provide a measure of similarity between each pair of subspaces. If the sum of all the eigenvalues (also the trace of the intersection matrix in Equation 3.4) is equivalent to the number of dimensions, k , included in the analysis, the two subspaces are equivalent. If the sum of all k eigenvalues is zero, then the two data sets are completely orthogonal and do not share any information.

Subspace Similarity Significance We adopt the *Crone and Crosby* [1995] method for determining if two subspaces are significantly close at the 95% confidence level using their distance. That is, if the distance between the two subspaces is significantly small, they are similar. The result from this significance test provides an upper limit for the number of

dimensions that two data sets share and can be used as a guideline. This significance test determines how much of the total variance of each data set is shared between the two data sets. Determining if two subspaces are equivalent is not the same as concluding that the individual PCs are the same, nor is it equivalent to concluding that the covariance matrices calculated in the PCA process are equal to each other; rather, this analysis helps to determine to what degree the subspaces spanned by the k PCs are similar.

To address this question, we use a metric called the subspace distance, which is defined using the intersection eigenvalues:

$$D(\widehat{Obs}\widehat{Sim}) = \sqrt{k - \sum_1^k \gamma_k} \quad (3.6)$$

where k is the number of PCs used to define the subspace and \widehat{Obs} and \widehat{Sim} represent the original observed and simulated reflectance data sets, respectively. The distance defined in Equation 3.7 is the sample distance calculated from the spectral decomposition of the intersection between the two data sets. We use this sample distance to test the null hypothesis that the population distance between the two subspaces is zero, $D(Obs, Sim)_k = 0$, against the alternative hypothesis that $D(Obs, Sim)_k > 0$.

The distance metric is used in the triangle inequality to construct a confidence interval that tests the null hypothesis [Crone and Crosby, 1995]. The form of the triangle inequality we use is:

$$D(\widehat{Obs}, \widehat{Sim})_k \leq D(\widehat{Obs}, Obs)_k + D(Obs, Sim)_k + D(\widehat{Sim}, Sim)_k \quad (3.7)$$

and is rearranged to give:

$$D(Obs, Sim)_k \geq D(\widehat{Obs}, \widehat{Sim})_k - D(\widehat{Obs}, Obs)_k - D(\widehat{Sim}, Sim)_k \quad (3.8)$$

where Obs and Sim are bootstrap-generated observed and simulated reflectance data sets. This equation allows us to estimate a one-sided confidence interval for the true param-

eter distance between the observed and simulated reflectance. To estimate the distributions of $D(\widehat{Obs}, Obs)_k$ and $D(\widehat{Sim}, Sim)_k$, we generate new reflectance data sets with the same number of spectra, N , as the original observed and simulated data sets by using *bootstrap with replacement*. The bootstrapped data sets are formed by randomly selecting reflectance spectra in the observed and simulated sets of spectra until the newly formed data sets are the same size as the originals. This is done *with replacement*, meaning that it is possible for each spectrum to be chosen more than once. We perform this procedure using a random number generator.

We calculate the principal components for the bootstrap-generated data sets and find the intersections between the bootstrap-generated and original observed and simulated data sets. Then we calculate the distances between the generated and original data sets for each number of k components used to define the subspaces. We repeat this process 500 times to estimate distributions of $D(\widehat{Obs}, Obs)_k$ and $D(\widehat{Sim}, Sim)_k$, using those distributions to find the distances in the 97.5 percentile to use in Equation 3.8 for the estimation of $D(Obs, Sim)_k$. 500 repetitions was the value used by *Crone and Crosby* [1995], but we also investigated the impact different numbers of repetitions had on the results, finding that at least 500 repetitions create continuous distance distributions. Creating these distributions with 1000 repetitions resulted in equivalent estimations of $D(Obs, Sim)_k$ compared to using 500 repetitions. If $D(Obs, Sim)_k > 0$, then the null hypothesis is rejected; otherwise, we fail to reject the null hypothesis. We can also think of these population distances as 95% confidence intervals. If $D(Obs, Sim)_k > 0$, then the confidence interval does not include zero; otherwise, the confidence interval does include zero, and it is possible for the population distance to be zero.

3.4 Hyperspectral Reflectance Variability

3.4.1 SCIAMACHY and OSSE Data Processing

Because we are quantitatively evaluating the similarity between the SCIAMACHY and OSSE reflectance spectra, it is important that the spectral, spatial, and temporal resolution and sampling of the two data sets are comparable. We created identically sized reflectance data sets by spectrally, spatially, and temporally resampling both of them and by including only averaged spectra located in grid boxes with data in both the SCIAMACHY and OSSE resampled data sets. Both sets of reflectance spectra were resampled to 10 nm full-width at half maximum spectral resolution and 3 nm sampling resolution. The OSSE spectra are produced using monthly averaged data on a 1.25° grid (Section 3.2.2). To ensure that SCIAMACHY pixels from at least every three days (the approximate time over which SCIAMACHY obtains near global coverage) throughout each month were represented in the monthly average of each grid box, we expanded the grid to 5.625° , four times the size of the original OSSE grid. We temporally aligned the data by calculating monthly averages of the SCIAMACHY reflectance by linearly averaging the SCIAMACHY pixels falling into each 5.625° grid box within each month.

It is a challenge to entirely eliminate the sampling differences between the two data sets. The OSSE spectra were generated using gridded input data from monthly averaged GCM output. The SCIAMACHY reflectance spectra, on the other hand, were instantaneous measurements from a satellite in sun-synchronous, near polar orbit. Even with the data resampling, inherent differences between satellite-measured and model-generated reflectance may remain. The objective of the steps presented above is to mitigate the impact on the quantitative comparison due to sampling differences.

To understand the effect of computing comparable spatial grids and monthly averages, we performed PCA on all SCIAMACHY reflectance spectra measured in January, April, July, and October 2004 for each month separately. The eigenvalue spectra from these

all-inclusive PCA results (Figure 3.2: gray) show that the variance of the SCIAMACHY dominant modes is higher than when the spatial and temporal averages are computed. The shapes of the eigenvalue spectra in black, calculated from the resampled SCIAMACHY data are much closer to the shape of the red OSSE eigenvalue spectra, implying that the distribution of information is also more comparable after resampling. Despite the sampling differences between the SCIAMACHY and OSSE data sets, the spectral, temporal, and spatial resampling performed here aligns the distribution of the variability within the data sets, lending confidence to the appropriateness of the applied resampling.

3.4.2 Spectral Reflectance Variability

To illustrate the quantitative intersection methods, this chapter will focus on the four months for which we had daily SCIAMACHY data in 2004, January, April, July, and October as an initial evaluation of the OSSE performance at the beginning of the twenty-first century. Before employing the quantitative comparison tools described above, we first calculate the principal components from the unstandardized OSSE and SCIAMACHY reflectance spectra. The eigenvalues (i.e. the variance of each PC dimension) for each of the four cases are shown in Figure 3.2. The shapes of the eigenvalue spectra show that the general distributions of the variance for both data sets is similar, at least for, approximately, the first 15 or 16 dimensions. The cumulative variance contribution in Figure 3.3 shows some differences in variance in the first few PCs, but for both data sets and all four months, six PC dimensions explain between 99.7% and 99.9% of the total data variance.

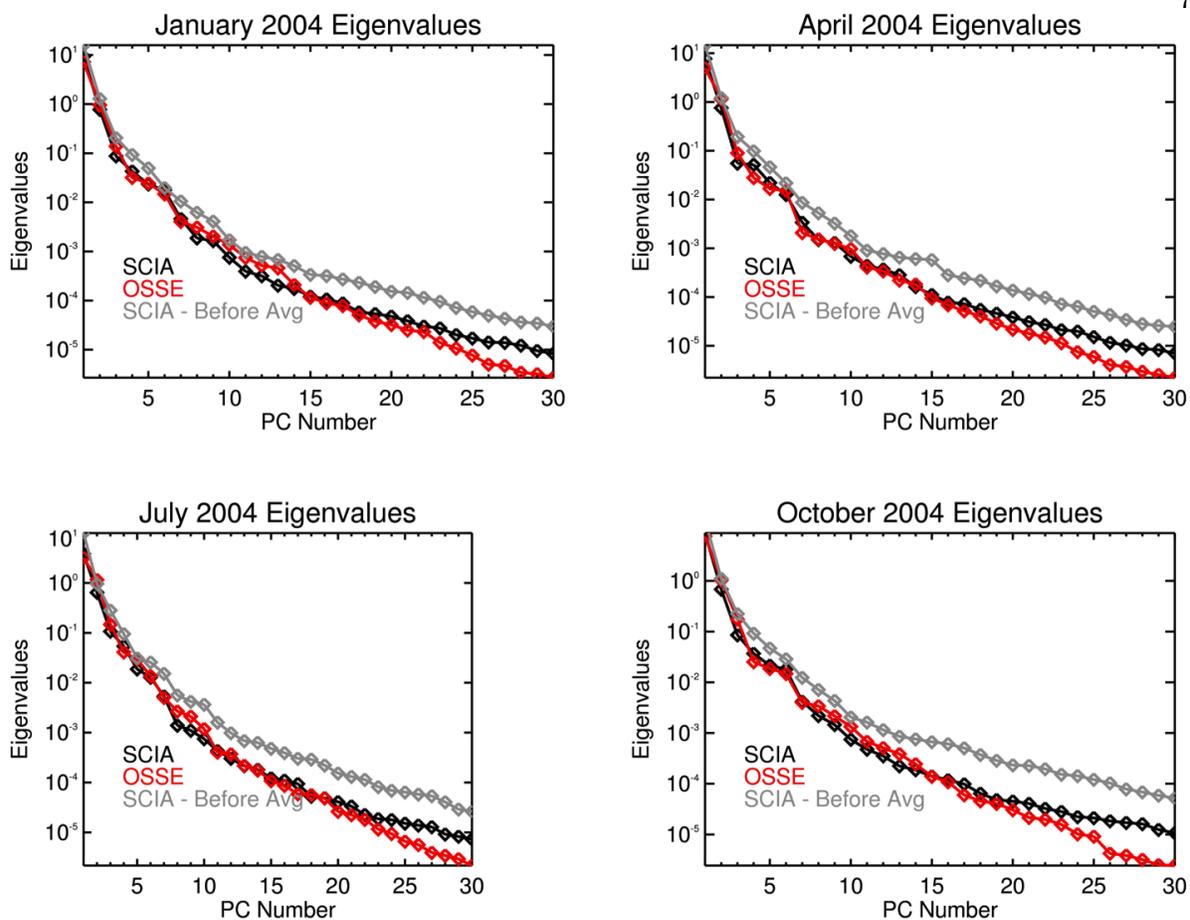


Figure 3.2: The first 30 eigenvalues for the January, April, July, and October 2004 SCIAMACHY and OSSE Reflectance PCs. The difference in shape between each of the grey lines and the black lines shows how well the SCIAMACHY data resampling performed prior to PCA aligned the SCIAMACHY distribution of information with that of the OSSE reflectance spectra.

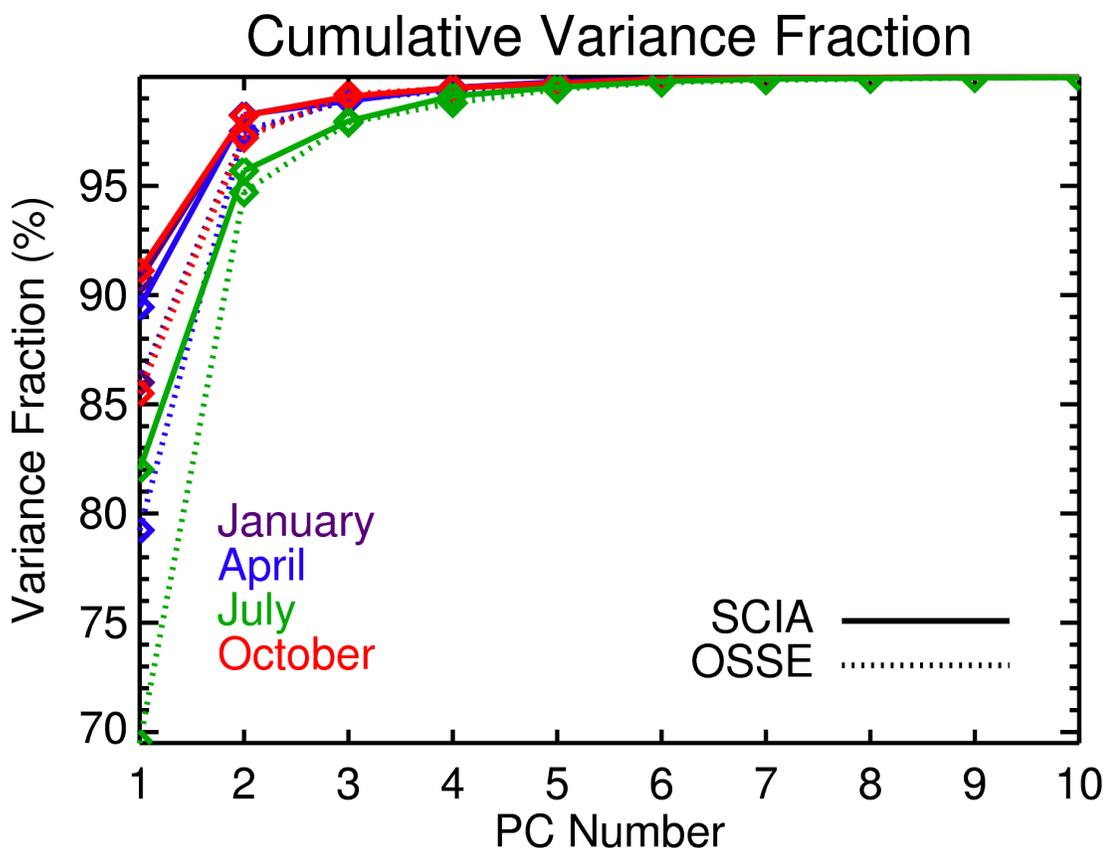


Figure 3.3: The cumulative variance fraction for each of the four cases for the first ten PC dimensions. For all four months, both SCIAMACHY (solid) and OSSE (dotted), six PC dimensions explain between 99.7% and 99.9% of the total data variance in both data sets.

In addition to studying the distribution of variance for the two data sets, we also examine the spectral shapes of the first several components that dominate the data variability. Figure 3.4 compares the first nine October 2004 SCIAMACHY and OSSE PCs. Generally among the four cases the spectral shapes of the SCIAMACHY and OSSE components are very similar. The spectral shapes of the first two components are nearly identical for the four cases and together explain 94.7%-97.5% (OSSE) and 95.7%-98.3% (SCIA) of the data variance. In addition to there being similarities between the PCs from the two data sets, there are spectral features that are indicative of physical variables. Water absorption bands are evident in at least the first four PCs for both data sets. The first PC resembles a cloud re-

fectance spectrum, and PC4 resembles a green vegetation reflectance spectrum (Figure 3.1). It is likely that the other PCs explain physical variables but they cannot be uniquely identified. An illustration of this point is presented in Figure 3.5, which shows the October 2004 OSSE and SCIAMACHY scores for PC4 and PC5. The spectral shape of PC4 is indicative of vegetation. This is confirmed in the spatial distribution of the scores by the relatively high scores over regions that are green in October such as the Amazon, the southeastern US, sub-Saharan Africa, and southern Asia. Moreover, negative scores are seen over areas typically devoid of green vegetation such as the oceans, polar regions, and semi-arid regions. Although similar spatial patterns are also observed in the PC5 scores, evidence that PC5 is partly explained by vegetation is not apparent by the spectral shape of PC5. This point also helps to support the importance of comparing entire subspaces when evaluating the data set similarity rather than solely relying on one-to-one PC comparisons.

There are some cases, however, in which individual comparisons of the PCs can reveal important differences between data sets. For example, although the first October 2004 SCIAMACHY PC contains some aspects of a cloud reflectance spectrum (Figure 3.4), its spectral shape also contains characteristics of a frozen surface, such as ice clouds or ice or snow at the surface. The local maximum that occurs between 1400 and 1450 nm occurs because of its position between the water vapor absorption band centered at 1350 nm and the ice absorption band centered at approximately 1500 nm. Although it appears that the first OSSE PC does not have the same ice spectral feature at 1400 nm as the first SCIAMACHY PC, this feature is in the OSSE PC, but it is broadened so that the peak occurs at a longer wavelength. We also see this difference in the January and April 2004 PC1 cases. It is likely that we do not see this ice feature and difference between the data sets in the July PC1 because of the reduction in Arctic snow and ice in July and the Antarctic night that occurs during this time. The way in which this feature is manifested in the PC may be representative of how snow reflectance values from MODIS are used as input for MODTRAN within the OSSE. The BRDF under snowy conditions was determined from snow-covered

and snow-free MODIS surface reflectance and was created by linearly interpolating over the MODIS channels to obtain an estimate of the spectral BRDF function for each grid box. This estimate was input into MODTRAN. The necessary linear interpolation over the coarse band coverage in the near infrared may be the cause of the broadened ice feature around 1400, which is visible in the PC1 comparison in Figure 3.4.

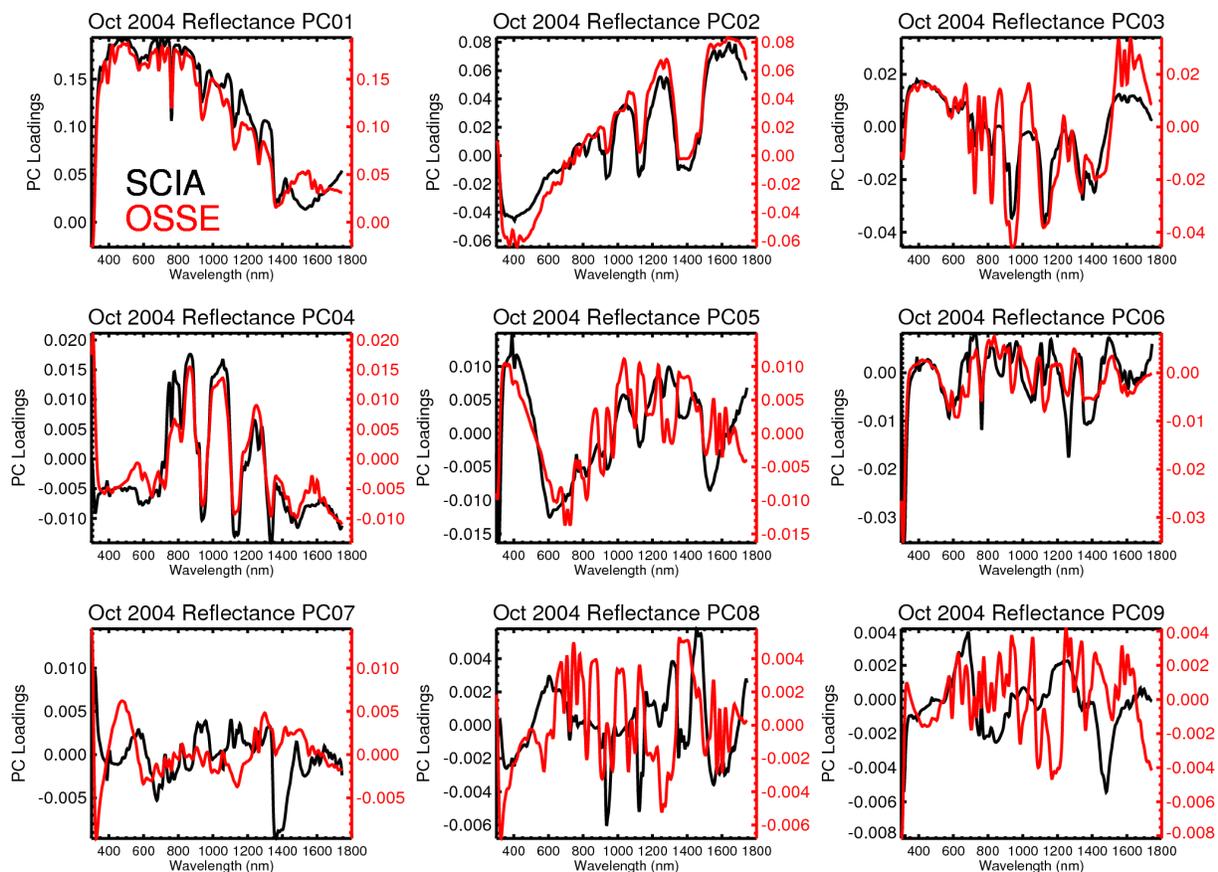


Figure 3.4: The first nine October 2004 SCIAMACHY (black) and OSSE (red) principal components show a close comparison between the two data sets. Physical variables are identifiable within some PC spectral shapes including clouds (PC1), vegetation (PC4), and water absorption (PC1-4).

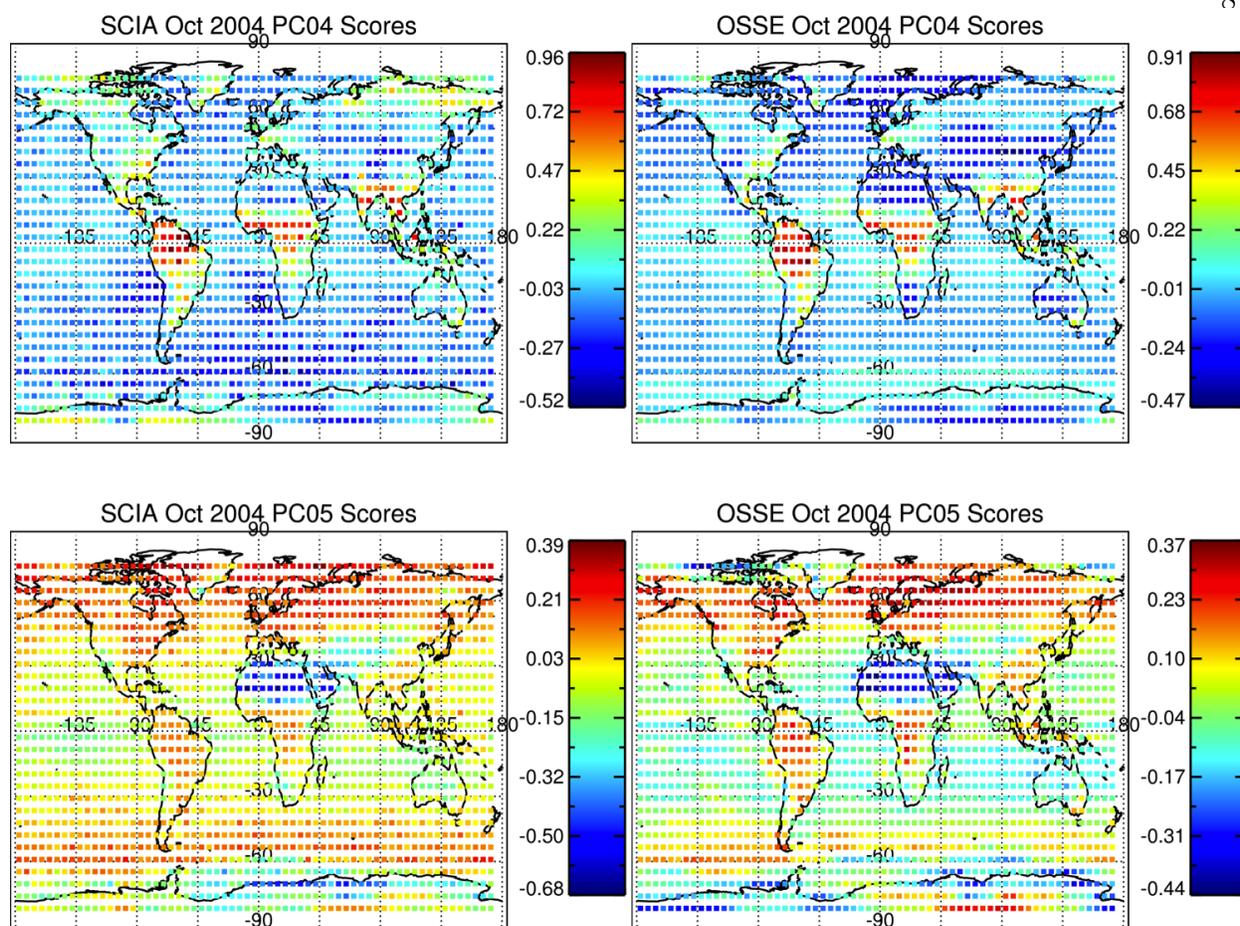


Figure 3.5: The spatial distribution of the October 2004 PC4 SCIAMACHY (a) and OSSE (b) scores and the PC5 SCIAMACHY (c) and OSSE (d) scores. Both PC4 and PC5 scores show evidence of vegetation, implying that this physical signal is distributed between at least these two components. Data set similarities seen in the overlap of PC shapes in Figure 3.3 are reinforced by the similarities seen in the spatial distribution of the scores.

3.4.3 Quantitative Subspaces Comparison

Initial evaluation of the comparison of the eigenvalues and PC spectral shapes suggests that the variance distribution between these data sets is similar. To quantify how much of the variance is shared between the observed and simulated reflectance, we begin by using PC selection criteria to estimate the number of dimensions that define the boundaries between signal and noise. For example, using the October 2004 logarithmic eigenvalue plot suggested

by *Craddock and Flood* [1969] (Figure 3.2d) it appears that six dimensions may be sufficient to represent the signal explained by the variability in the data set, which is also how many dimensions the fractional Kaiser method [*Jolliffe*, 1972] suggested. The dip between the sixth and seventh eigenvalues and the change in slope before and after these eigenvalues likely indicates that the first six dimensions explain most of the variance in the data sets. This is also supported by the increasing amount of noise in the PC shapes after PC6 in Figure 3.4. The Broken Stick Method [*Jolliffe*, 2002] suggested 14 dimensions, but the Broken Stick Method typically suggests the largest number of dimensions among the PC selection criteria described above. Using these criteria for the other three months as well, we have estimated that seven dimensions explain the reflectance signal for January and April, and eight dimensions explain the signal for July. Because the Broken Stick Method suggested that between 14 and 16 dimensions were above the noise level, we calculated intersections using the first twenty eigenvectors. Even though we estimated that fewer dimensions than define the boundary between signal and noise, we calculated 20 different intersections with between one and twenty eigenvectors to find the number of dimensions at which the two data sets are different at the 95% confidence level.

For each month, the twenty intersections were computed using the subspaces spanned by $1 \leq k \leq 20$ eigenvectors of the SCIAMACHY and OSSE data, and the spectral decompositions of each of these intersections were performed. Recall that the eigenvalues from the intersection decomposition are measures of similarity between the subspaces. The eigenvalues for each of these subspaces are shown in Figure 3.6 for each month, with the maximum possible similarity, k , shown in black. As k , the number of dimensions used to define each subspace, increases between one and twenty dimensions, the observed similarity between the two data sets decreases. This is illustrated by the increasing difference between the red lines and the black lines in Figure 3.6. To quantify the largest number of dimensions that the two data sets share, we first calculate the distance between each set of subspaces. The calculated subspace distances, the maximum possible distances (\sqrt{k}) and the ratio between the calcu-

lated distance and the maximum distance are shown in Figure 3.7. The subspace distances also confirm the result shown by the eigenvalues in Figure 3.6, most clearly demonstrated by the relative distances, which generally increase with the number of dimensions used to define each subspace.

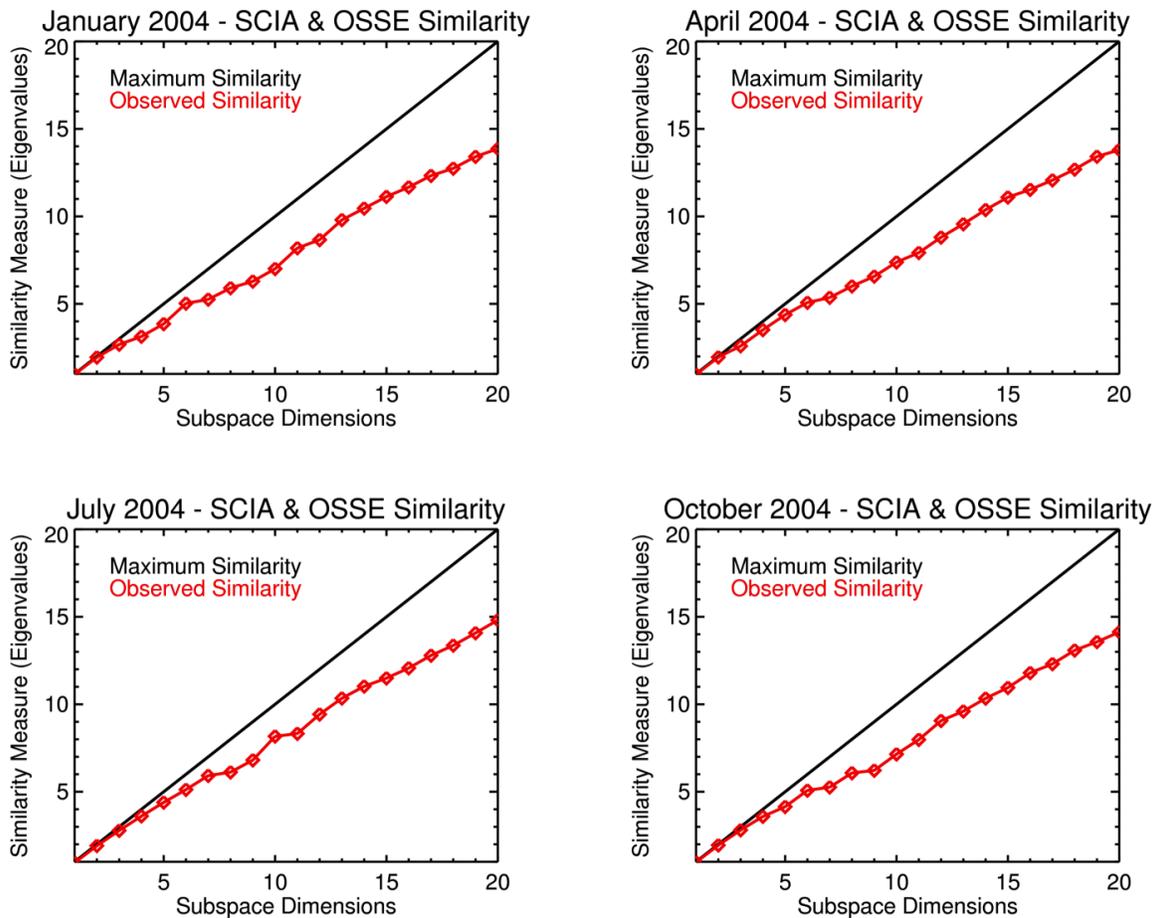


Figure 3.6: The comparison of the eigenvalues from the spectral decomposition of the intersection to the maximum possible eigenvalue for each number of subspace dimensions. With more dimensions used to define the two pairs of subspaces, the similarity identified by the eigenvalues decreases.

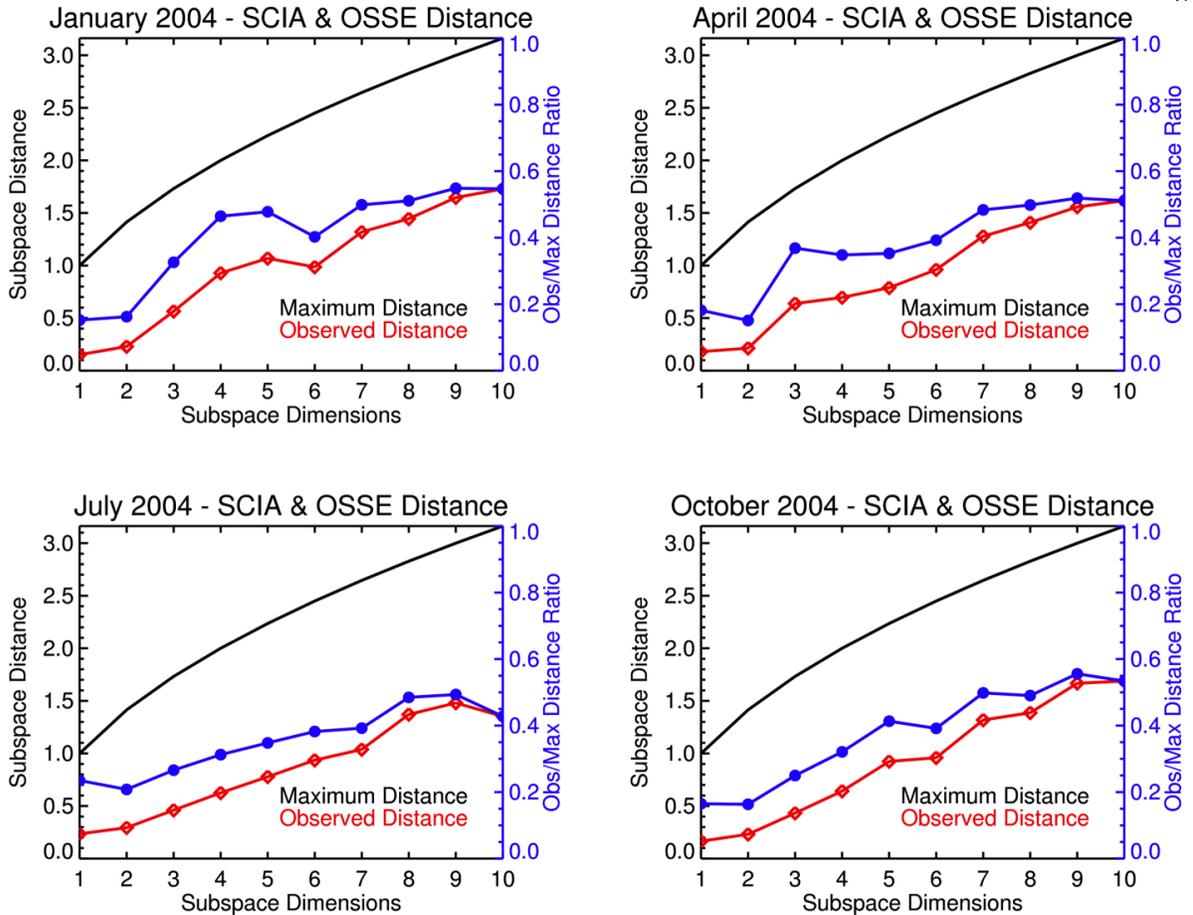


Figure 3.7: The subspace distances between the SCIAMACHY and OSSE reflectance subspaces for ten subspaces (red) compared to the maximum possible distance between the two subspaces (black). The blue line shows the observed subspace distances relative to the maximum possible distance for each number of subspace dimensions, that is, the ratio of the values on the red line to the values on the black line.

The subspace distances shown in Figure 3.7 are the observed distances on the left side of Equation 3.8. Continuing with the process, the triangle inequality is used to estimate the population distances, shown in Figure 3.8. This statistical significance test shows how many k -dimensional subspaces are the same at a 95% confidence level, and vertical lines indicate the largest k -dimensional subspace for which this is true for each case in Figure 3.8. We also note that the selection criteria results using the logarithmic eigenvalue plot shown in Figure 3.2 give similar values to those determined by the statistical significance test. The statistical significance test found that the two data sets agree over seven dimensions in April and July

and eight dimensions in January and October. This alignment demonstrates that the two data sets are generally similar at the signal to noise boundary, discussed at the beginning of this section and estimated to be located at seven (January), seven (April), eight (July), and six (October) dimensions. Using the cumulative variance explained (Figure 3.3) by the number of dimensions indicated by the vertical lines in Figure 3.8 we can determine how much OSSE and SCIAMACHY variance is explained in the k-dimensional space in which they are similar at the 95% confidence level. The results in Figure 3.3 and 3.8 show that for the number of dimensions over which the two data sets agree in January, April, July, and October approximately 99.9% of the SCIAMACHY and OSSE data variance is explained.

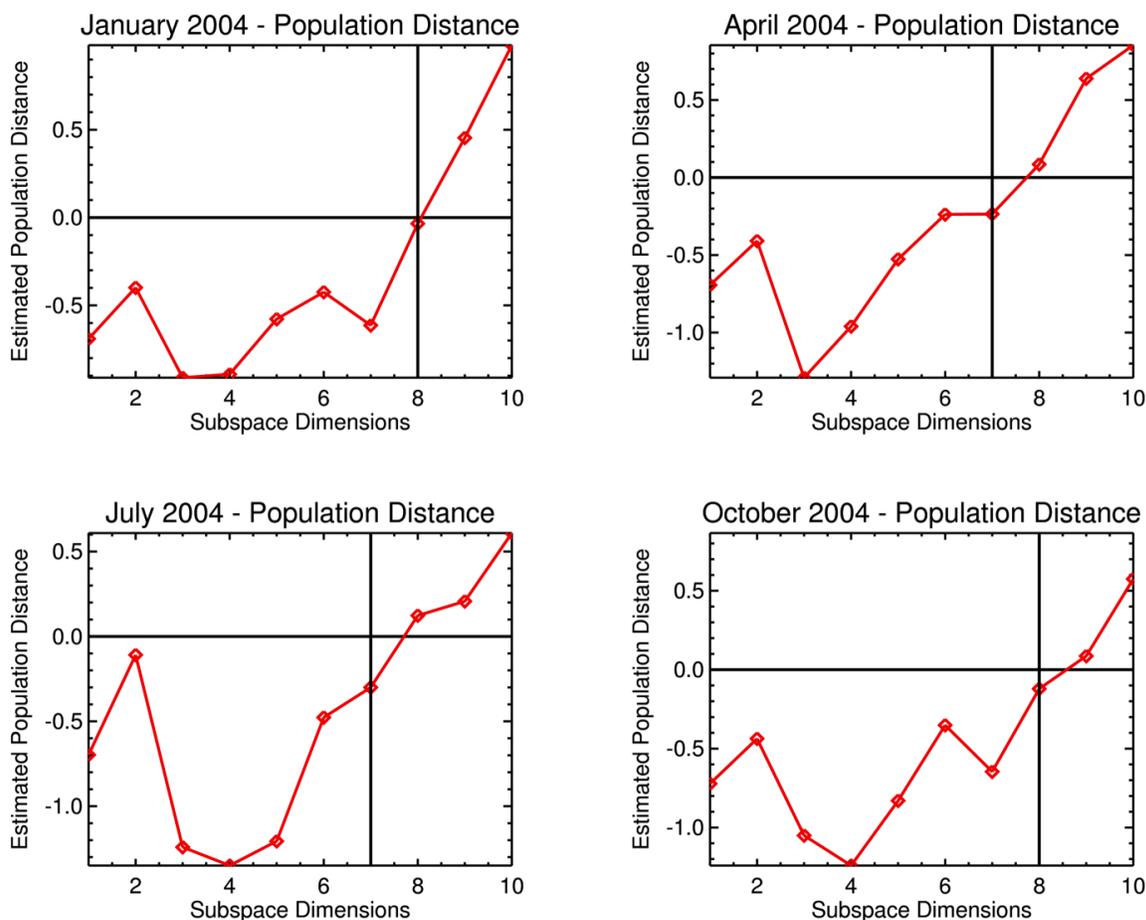


Figure 3.8: The population distances between the OSSE-simulated and SCIAMACHY-observed reflectance spectra for January, April, July, and October 2004. The distances less than zero correspond to subspaces of k dimensions that are the same at the 95% confidence level. The vertical lines indicate the maximum number of dimensions the two data sets share at the 95% confidence level for each case.

It is also informative to inspect the spectral shapes of each pair of transformed eigenvectors (Figure 3.9). Using the October 2004 results from the statistical significance test, we show the transformed vectors of the eight-dimensional shared space between the SCIAMACHY and OSSE reflectance data. The first three eigenvectors exhibit several spectral characteristics that are also present in the original PCs in Figure 3.4. The fourth transformed eigenvector in part resembles the original PC4, but the others contain only segments of recognizable spectral features, if any. Because some of the transformed vectors resemble the original PCs, this means that overlapping information between the two data sets is

very similar to that of the original dominant modes of observed variability. By applying the intersection decomposition and the statistical significance technique, we have presented an objective method with which to quantitatively compare two multivariate subspaces, a technique which has several other applications, as described below.

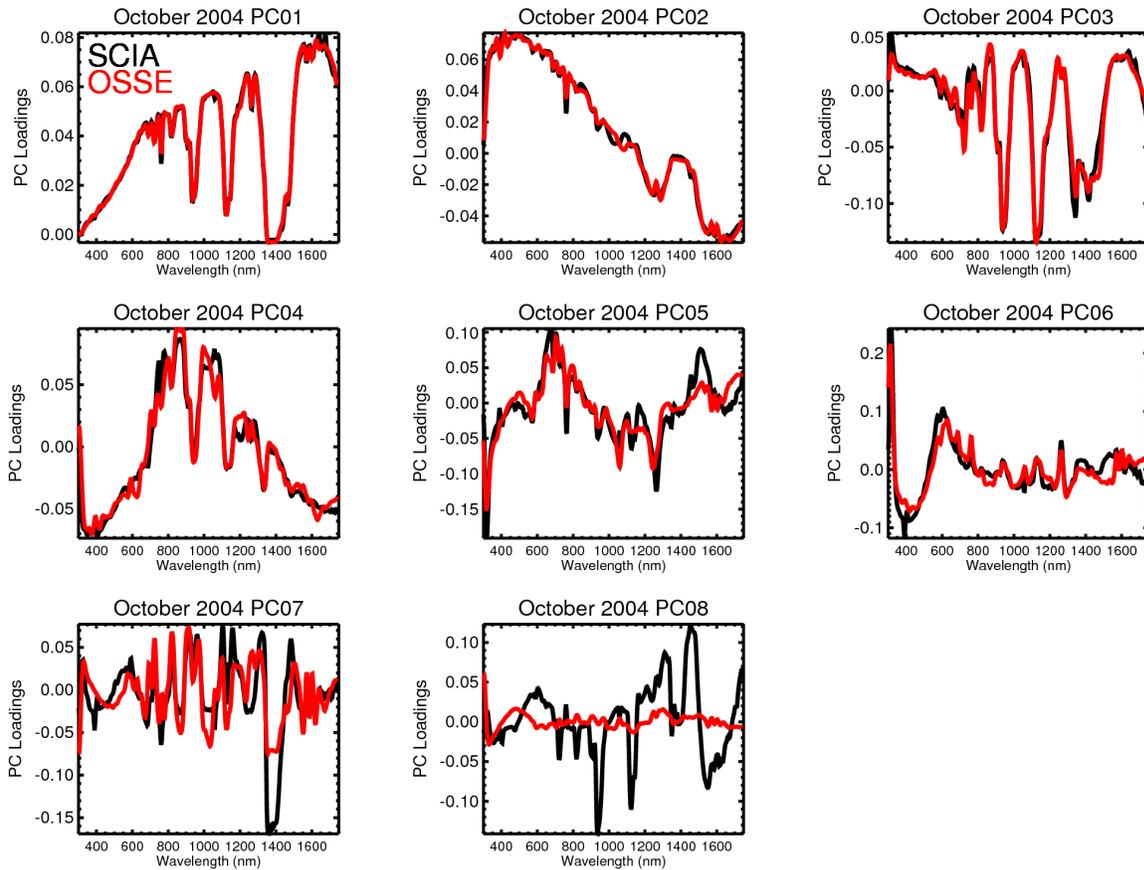


Figure 3.9: The spectral shapes of the transformed eigenvectors for October 2004 in the shared intersection space. The first seven dimensions align well, and the first four dimensions share spectral characteristics with the first four original October 2004 PCs, shown in Figure 3.7.

3.5 Conclusions and Future Work

In this chapter, we used SCIAMACHY-measured hyperspectral solar reflectance to evaluate how well OSSE-simulated hyperspectral reflectance captures variability in the Earth's climate system. We presented two primary ways in which the information between two data

sets can be compared. First, we qualitatively compared the most dominant principal components that explained the majority of the variance in both data sets and found that the two data sets appear to share similar variance distributions. We also found that linear interpolation of surface reflectance in the OSSE manifests as a difference in the first principal component for the January, April, and October cases. Second, we quantitatively compared the spectral variability of the two data sets using their principal components. This analysis showed that the OSSE and SCIAMACHY reflectance spectra share a large fraction of their spectral variability and that this variability shares spectral characteristics with the original PC transformation of the measured data set. From these results we conclude that at the beginning of the century, the OSSE appears to give a realistic representation of the Earth's variability relative to SCIAMACHY-measured reflectance. These findings provide a necessary, initial condition that helps us to understand the predictive potential of the OSSE for understanding how Earth's variability may change during the 21st century.

There are several other research questions that the quantitative comparison method applied in this study could be used to address. For the OSSE simulations used in this study, CCSM3 output was used, but other climate model results could also be used for the OSSE simulations. The comparison method presented here can provide rigorous objective testing of these different climate models to determine which model best reproduces Earth's present-day climate variability and is likely better to study future changes in Earth's climate.

Feldman et al. [2011a] used the OSSE to compare the shortwave reflectance signal observed between two different emission scenarios simulated using CCSM3: the constant CO₂ and the A2 emission scenarios. The quantitative method described in this chapter can be used to understand how changes in different climate forcing scenarios are manifested in the variability of hyperspectral reflectance. These results can be studied during the first decade of the 21st century, for comparison to SCIAMACHY reflectance, and during the entire 21st century, to attempt to understand how changes in climate contribute to changes in reflected shortwave spectral variability on a centennial time scale. This may provide insight

into which variables contribute to changes in the measured reflectance over different time scales. In a subsequent chapter, we will evaluate how well the OSSE reproduces the temporal variability of the Earth’s climate system over the decade for which we have SCIAMACHY measurements.

In addition to the ideas presented above, there are other ways in which to improve and expand upon the analysis presented in this chapter. We focused on the similarities between the observed and simulated data, but it may also be useful to investigate the nature of the differences between the two data sets. The Minimum Noise Fraction (MNF) transform [Green *et al.*, 1988] is a method that can be used in conjunction with the comparison method described in this chapter. The MNF is a two-part PCA that whitens or decorrelates the noise in the data set, so if a well-defined noise characterization is available from noise-equivalent dark spectra during an instrument’s lifetime, this transformation can be applied to the radiances before the quantitative comparison method is used. One of the benefits of the MNF transform is that it typically provides a clearer boundary between the signal and noise levels using the eigenvalue spectrum.

Another improvement to this work involves the method used to spatially and temporally resample the sun-synchronous satellite-measured reflectance. Although the methods aligned the spectral, temporal, and spatial sampling of the two data sets, it would be beneficial to establish more appropriate methods for gridding sun-synchronous satellite data to minimize the potential sampling differences in comparisons such as these. As climate observation systems are deployed, we will be able to apply the techniques described here to further improve the development of climate OSSEs as future instruments are designed. The results presented in this chapter provide a foundation for how these quantitative comparisons between two hyperspectral data sets can be made. These results also provide the community with a measure of how well the OSSEs are able to reproduce the variability of the Earth’s climate system.

Chapter 4

Temporal Variability of Observed and Simulated Hyperspectral Reflectance

4.1 Introduction

Earth's weather and climate are driven by the imbalance of incoming and outgoing radiation. Over global spatial scales and very long temporal scales, Earth is in approximate radiative equilibrium, although several studies have estimated positive annual radiative imbalances over the past two decades that range between 0.58 Wm^{-2} and 0.85 Wm^{-2} [*Hansen et al.*, 2005, *Hansen et al.*, 2011, *Willis et al.*, 2004]. Earth's incoming (absorbed) radiation is the difference between the solar radiation incident at the top of the atmosphere (TOA) and the solar (shortwave) radiation reflected at TOA, a value dependent upon Earth's albedo. Albedo is the ratio of the reflected irradiance and the incident TOA solar irradiance. The solar spectral range generally considered important for climate processes is between about $.320 \mu\text{m}$ and $4 \mu\text{m}$, while the spectral range of Earth's emitted radiation is between $4 \mu\text{m}$ and $100 \mu\text{m}$. Radiative imbalances may exist for a number of reasons, one of which may be the increase of greenhouse gas concentrations from the combustion of fossil fuels. In addition to greenhouse gases, other forcing agents of Earth's climate include incident solar radiation, aerosols, and land surface changes [*IPCC*, 2007]. The response of the climate system to a positive energy imbalance occurs over a wide range of spatial and temporal scales, but a globally averaged response requires temperature to increase for the system to reach radiative equilibrium.

Satellite observations provide a unique view of Earth because they give us the oppor-

tunity to obtain global distributions of shortwave and longwave radiation and the physical variables that drive and respond to changes in Earth's climate. Satellite systems that have been operational for the last decade or longer have confirmed some *in situ* observed trends in climate variables. For example, *Davies and Molloy* [2012] found a decrease in annual mean cloud top height between 2000 and 2010 using cloud top height retrievals from the Multiangle Imaging Spectroradiometer (MISR) [*Diner et al.*, 1998]. In another example, passive microwave satellite measurements have monitored Arctic sea ice extent from space since 1979, and verify an increase in length of the Arctic melt season, a general thinning of ice, and a downward trend in the minimum sea ice extent observed at the end of the melt season [*Markus et al.*, 2009]. Measurements from the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) [*Bovensmann et al.*, 1999] measurements between 2003 and 2009 have verified that the globally averaged CO₂ concentration increased throughout that time period [*Schneising et al.*, 2011]. Also, *Loeb et al.* [2012] used Earth Observing System and A-Train satellite observations to demonstrate how a decade of satellite measurements are contributing to our understanding of the underlying physical processes related to Earth's top of atmosphere radiation budget.

Loeb et al. [2012] emphasized the importance of extending current decadal observations over even longer time scales with improved measurement accuracy to monitor changes in Earth's climate. Obtaining highly accurate measurements with the sensitivity to detect changes in Earth's climate is one of the goals of the Climate Absolute Radiance and Refractivity Observatory (CLARREO) mission. In addition to high accuracy, high information content is also a goal of the CLARREO mission. High spectral resolution reflectance measurements with contiguous overlapping spectral bands (i.e. hyperspectral measurements) contain more information about surface and atmospheric variables in the Earth's climate system than broadband or discrete band measurements.

Feldman et al. [2011a] developed a climate Observation System Simulation Experiment (OSSE) to simulate hyperspectral CLARREO reflectance spectra as a tool for refining mea-

surement characteristics of the CLARREO shortwave instrument. *Feldman et al.* [2011a] studied the response of these OSSE-simulated reflectance spectra to climate forcings prescribed in a global climate model during the 21st century. Their results indicated that spectrally resolved reflectance may be able to detect decadal and centennial scale changes in clouds, aerosols, and land surface albedo, implying that shortwave reflectance contains information about climate-relevant forcings and feedbacks. *Feldman et al.* [2011a] calculated the algebraic differences of the spectral reflectance decadal averages between the beginning and middle and between the beginning and end of the century to draw conclusions about how the CCSM-simulated climate system changes were manifested in the shortwave signal.

In this study, we quantify the temporal variability of shortwave reflectance using principal component analysis (PCA). PCA is a multivariate spectral decomposition technique that calculates independent variables that are linear combinations of the original variables (here, spectral reflectances) and that are ranked by the amount of variance explained in the data set. A subset of the PC variables that dominate the variance in the data set can be used as a representation of the entire data set. Several studies have used PCA to evaluate the spectral variability of shortwave radiation [*Rabette and Pilewskie*, 2001, 2002, *Grenfell and Perovich*, 2008, *Roberts et al.*, 2011, 2012]. *Roberts et al.* [2011] applied PCA to hyperspectral reflected radiance to quantify the variability in Earth's climate system and found that the dominant spectral variables exhibited spatial, seasonal, and periodic temporal patterns, and that physical variables were discernible in the dominant PCs. Here, in a novel study we quantify the nearly decadal variability of observed hyperspectral reflectance and the centennial variability of simulated hyperspectral reflectance using spectral decomposition techniques.

Using measured and simulated reflectance spectra we address several objectives. First, we quantify the near-decadal variability of measured hyperspectral reflectance between 2002 and 2010, which is the time period over which we have access to SCIAMACHY reflectance data. *Roberts et al.* [2012] used the intersection of the spectral variability of measured and

simulated hyperspectral reflectance to quantify how well the simulated reflectance reproduces observed spectral variability at the beginning of the century. The second objective of this study is to use the intersection to evaluate the near-decadal variability of simulated reflectance relative to observed reflectance. We also use multivariate analysis techniques to quantify the centennial variability of the OSSE-simulated 21st century reflectance spectra and develop methods of trend detection to identify annual, seasonal, and interannual variability and secular trends.

4.2 Data

4.2.1 Observed Reflectance: SCIAMACHY

The observed reflectance spectra used in this study were measured by the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) [*Bovensmann et al.*, 1999]. SCIAMACHY was a hyperspectral imaging spectrometer mounted on the European Space Agency's (ESA) Environmental Satellite (ENVISAT). This instrument was launched on ENVISAT in March 2002 and was operational until April 2012. ESA lost contact with ENVISAT in April 2012 and declared the end of the mission in May 2012 after several unsuccessful attempts to regain contact with the satellite. SCIAMACHY made continuous spectral measurements of reflected radiation between the ultraviolet and the near infrared. Eight detectors sampled the spectral ranges spanning 214 - 1773 nm, 1934 - 2044 nm, and 2259 - 2386 nm with spectral resolutions between 0.22 nm and 1.48 nm. SCIAMACHY made nadir measurements with an angular full swath width of $\pm 32^\circ$. Nadir pixel sizes were dictated by integration time with sizes ranging between 26 km (along track) by 30 km (across track) and 32 km by 930 km for wide swath settings.

Because the focus of this study is on the temporal variability over much of the SCIAMACHY lifetime, it is important to consider the optical degradation of the instrument. The SCIAMACHY team applied correction factors to the data to account for instrument optical

degradation [Noel *et al.*, 2003]. These correction factors were applied during data calibration and were calculated using a reference light source (often a SCIAMACHY-measured solar spectrum) and account for the optical degradation relative to the first measurements made in August 2002.

4.2.2 Simulated Reflectance: OSSE Output

Feldman et al. [2011a] developed a climate Observation System Simulation Experiment (OSSE) to emulate the CLARREO shortwave instrument and to help refine mission and configuration design requirements. Earth's climate was simulated during the 21st century using the NCAR Community Climate System Model Version 3 (CCSM3) [Collins *et al.*, 2006] and two IPCC Fourth Assessment Report emission scenarios. The Moderate Resolution Atmospheric Transmission (MODTRAN) [Anderson *et al.*, 1999] radiative transfer model used the model output from CCSM to simulate CLARREO shortwave instrument hyperspectral reflectance measurements. For each climate scenario, all-sky (cloud inclusive) and clear-sky (cloud free) reflectance spectra were calculated. The constant CO₂ concentration scenario was the control case [Meehl *et al.*, 2005, 2007] in which greenhouse gases and aerosols were fixed to year 2000 levels throughout the model run. The A2 emission scenario, the experiment, simulated the Earth system's response to increasing levels of well-mixed greenhouse gases such as CO₂, CH₄, N₂O, and chlorofluorocarbons throughout the century. Sulfate and carbonaceous aerosol loadings were increased through 2050, after which they were reduced through the end of the century to 1990 levels. The A2-prescribed changes to the atmospheric composition resulted in several changes to the climate system including a tripling CO₂ concentration relative to pre-industrial levels, considerable surface warming, and an increase in snow and sea ice melting over time. *Feldman et al.* [2011a] compared the reflectance spectra simulated with these two climate change scenarios and demonstrated how the prescribed changes in the experiment affected the reflectance compared to the control. The climate OSSE was the first model-based study conducted to evaluate the solar spectral reflectance

response to decadal and centennial scale changes to Earth's climate system [Feldman *et al.*, 2011a]. Feldman *et al.* [2011b] also used the CLARREO climate OSSE to show that the time needed for climate change detection is shorter using spectrally resolved reflectance than broadband shortwave measurements.

4.3 Multivariate Methods

4.3.1 Principal Component Analysis

Principal component analysis is a multivariate analysis technique used to simplify multivariate data sets into a smaller number of variables that explain a large fraction of the data variance. PCA has been applied to a multitude of meteorological data such as measurements of chemical constituents [Chan *et al.*, 2006], sea level pressure [Hannachi *et al.*, 2006], and geopotential heights [Craddock and Flood, 1969]. In remote sensing, it has been used to study the spectral, spatial, and temporal variability of broadband and spectrally resolved measurements of emitted longwave radiation [Haskins *et al.*, 1997, 1999, Huang *et al.*, 2002], and spectrally resolved shortwave radiance, irradiance, and reflectance [Rabette and Pilewskie, 2001, 2002, Grenfell and Perovich, 2008, Roberts *et al.*, 2011, 2012]. In this study we applied PCA to measured and simulated reflectance spectra.

The principal components (PCs) are calculated by spectrally decomposing the covariance matrix and are linear combinations of the original variables, in this case the spectral reflectances. The PCs are ordered by their variances, which are the eigenvalues from this decomposition. The principal components are the eigenvectors scaled by the square root of the corresponding eigenvalues (that is, the standard deviations of the corresponding dimensions). When the original variables are correlated, which is typically true of spectral reflectances, PCA results in a large reduction in the number of variables needed to explain most of the variance in the data set. Although they are mutually orthogonal, the PCs result from non-linear radiative responses to the underlying physical variables in the Earth's

climate system, so they are not necessarily physically independent. The reflectance spectra projected onto the principal components are the scores. In this study, we are evaluating the spectral and temporal variability of reflectance time series, so we will use the scores time series to evaluate how the dominant principal components vary over time.

4.3.2 Intersection Application

The eigenvectors that explain almost all the data variance span a subspace that can be used to represent the data set. The intersection between subspaces calculated from the PCA of the two data sets can be used to quantify their similarity. *Roberts et al.* [2012] used the intersection to evaluate how similar the variability of simulated reflectance was to that of observed reflectance. In this study, we use the intersection to directly compare the temporal patterns of the shared spectral variables. The spectral decomposition of the intersection, explained in mathematical detail in *Roberts et al.* [2012], allows for the transformation of the PCs for each data set into the intersecting space. We can then project each data set's PC scores onto that shared space and use the intersection scores to directly compare the temporal variability of the two data sets. Using the same notation found in *Roberts et al.* [2012], the intersection scores are calculated using

$$\begin{aligned} S_A &= D_A Y \\ S_B &= D_B Y \end{aligned} \tag{4.1}$$

where subscripts A and B refer to the two data sets, the \mathbf{S} matrices are the intersection scores, the \mathbf{D} matrices are the original PCA scores calculated from the two reflectance data sets, and \mathbf{Y} contains the eigenvectors calculated from the spectral decomposition of the intersection. Because we are studying the temporal variability of reflectance, the original PC scores and the transformed scores are time series that describe the temporal variability of each of the PCs. Using the eigenvalues from the subspace intersection decomposition, we can determine how many transformed PC dimensions to use for a direct comparison of the temporal variability between the two data sets. This analysis will quantify the temporal

variability of simulated reflectance spectra and compare it to the temporal variability of observed reflectance during the same time period.

4.4 Results

To quantify the near-decadal temporal variability of observed hyperspectral reflectance, we used SCIAMACHY reflectance spectra between August 2002 and December 2010 representing the range of SCIAMACHY data currently available on the CLARREO cluster at the NASA Langley Research Center. The SCIAMACHY Operations Support Team (SOST) Data Quality log [*Bramstedt et al.*, 2012] includes a list of time periods when decontamination occurred, during which detectors 7 and 8 (covering 1934 nm - 2386 nm) were heated to melt ice deposits that had formed on the detectors and reduced throughput. The decontamination process affected detectors 7 and 8 the most out of the eight because their stabilization times (i.e. the time for the temperature of the detectors to return to thermal equilibrium) were the longest. During decontamination periods, ultraviolet and visible measurements continued, but data quality was compromised, according to the SOST [*Bramstedt et al.*, 2012]. Therefore, we did not include reflectance spectra measured during the nine decontamination operations that were performed between 2002 and 2010.

Figure 3.1 presents three sample SCIAMACHY reflectance spectra, showing the spectral differences discernible in reflectance among a thick cloud, a cloud-free vegetation scene, and a cloud-free ocean scene. The SCIAMACHY and OSSE spectra were both resampled to a 10 nm spectral resolution and 3 nm spectral sampling because for our studies we did not need SCIAMACHY's native high spectral resolution, and this coarser spectral resolution is sufficient to detect climate variables [*Roberts et al.*, 2011]. SCIAMACHY reflectance spectra were also limited to those measured at nadir (i.e. line of sight angles less than 5°). We divided the globe into six 30° zonal bands and calculated the monthly averaged land and ocean reflectance spectra in each zonal band. The International Geosphere-Biosphere Programme (IGBP) land cover indices [*Loveland and Belward*, 1997, *Loveland et al.*, 2000] were used to

classify the SCIAMACHY and OSSE pixels as either land or ocean. Monthly averages were used because the OSSE spectra were calculated from monthly averaged model output.

4.4.1 SCIAMACHY Temporal Spectral Variability

First, the observed temporal variability of SCIAMACHY hyperspectral reflectance was quantified. For each 30° zonal band, principal component analysis was applied to all of the monthly and zonally averaged SCIAMACHY land and ocean reflectance spectra. This resulted in six sets of principal components describing the spectral variability of the zonal bands and two scores time series for each PC describing the land-atmosphere and ocean-atmosphere temporal variability of each PC. Because the data sets from which we calculate the PCs vary spectrally, spatially (over land and ocean), and temporally, the PCs explain the spectral, spatial, and temporal variability within each zonal band.

There are several qualitative and quantitative methods that can be used to determine how many PCs are sufficient to explain nearly all the data variance [Jolliffe, 2002, Roberts *et al.*, 2012]. The Broken Stick Method [Jolliffe, 2002] and the log-eigenvalues method [Craddock and Flood, 1969] are among the best methods to determine number of dimensions that define the boundary between signal and noise in spectral variability studies of hyperspectral solar radiation. The *Broken Stick* criterion states that $\omega_k > \frac{1}{p} \sum_{j=k}^p \frac{1}{j} \omega_{avg}$ defines how many PCs define the signal-noise boundary, where ω_k is the k^{th} eigenvalue, p is the total number of variables, and $1 \leq k \leq p$. Craddock and Flood [1969] found that the dimension number at which the plot of the logarithm of the eigenvalues first changes slope tends to define the signal-noise boundary. We have applied both of these methods to the PC eigenvalues, shown for the six zonal band cases in Figure 4.1. These two methods determined that four PCs define the signal-noise boundary in the three bands between 30°N and 60°S (0°-30°N, 0°-30°S, and 30°S-60°S); 99.6% of the signal in these zonal bands is explained by the first four principal components (Figure 4.2). These two methods also show that for the other three zonal bands (60°N-90°N, 30°N-60°N, and 60°S-90°S) the first five dimensions, which explain

at least 99.5% of the signal (Figure 4.2), define the signal-noise boundary.

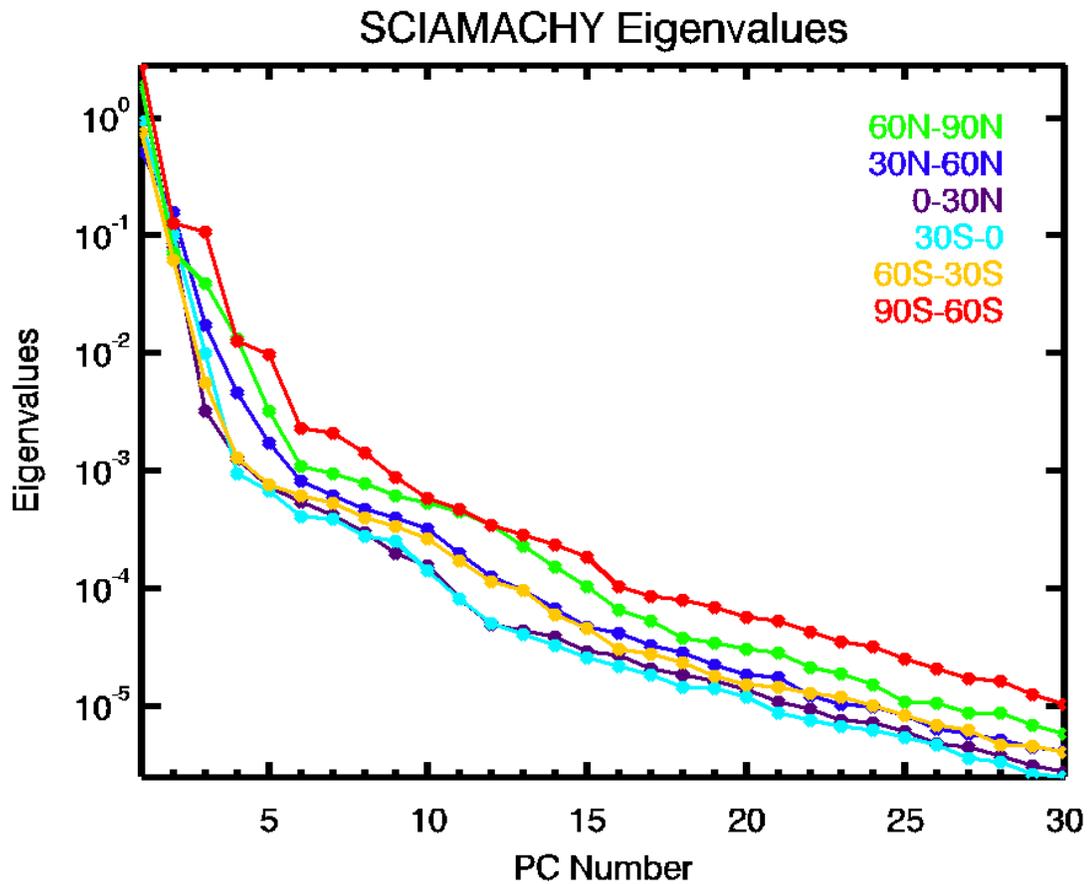


Figure 4.1: The first 30 eigenvalues calculated from principal component analysis applied to the zonally averaged spectra from the six latitude bands. Three zonal bands (60°N - 90°N , 30°N - 60°N , and 60°S - 90°S) have eigenvalue spectra that change slope at the fifth component, and the other three zonal bands have eigenvalue that change slope at the fourth component.

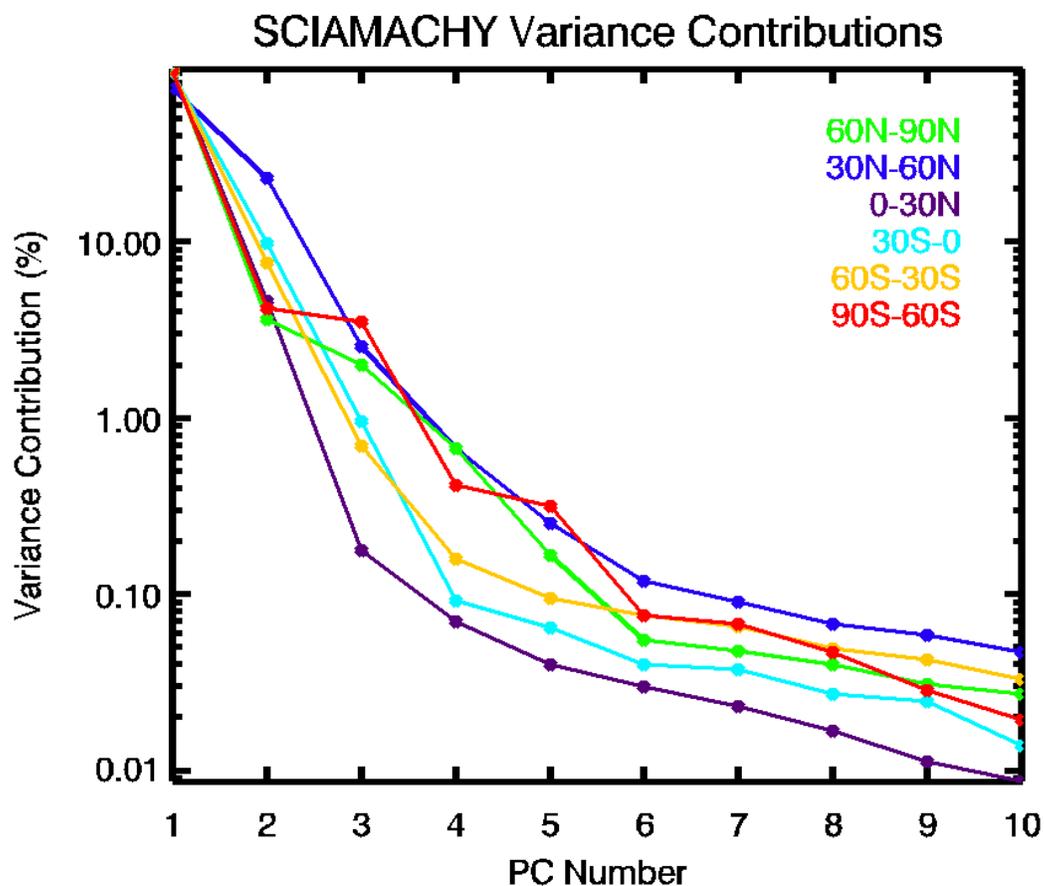


Figure 4.2: The variance contribution of the first 10 SCIAMACHY principal components for the six 30 degree zonal bands. For all the zonal bands, the first four components explain at least 99% of the total variance.

The spectral shapes of the PCs and the PC scores time series can be examined to determine the physical causes of variability. The first principal components in the two tropical bands (0° - 30° N and 0° - 30° S) are compared in Figure 4.3. These two components have a very similar spectral shape indicating that the first PCs in the northern hemisphere (NH) and southern hemisphere (SH) tropics zonal band, which explain 88.98% and 95.04% of the variance respectively, represent nearly the same variable. Water vapor and other gas absorption bands, the near infrared edge, and cloud reflectance are among the spectral signatures represented at least in part in its spectral shape. In Figure 4.4, we compare the

time series that correspond to the NH and SH Tropical PC1s. One similarity among the four time series shown in Figure 4.4 is that they all exhibit annual variability. However, there are differences between the land and ocean time series and also between the NH and SH time series. In both zonal bands, the land scores are all positive, and the ocean scores are all negative, implying that the land and ocean spectra explain different aspects of the PC1 spectral variable. Also, there are phase differences between the NH and SH time series. The land time series are 180° out of phase, with the NH maxima (minima) and SH minima (maxima) occurring in July or August (January or February). The extrema in each time series correspond to the average location of the Intertropical Convergence Zone (ITCZ), which tends toward the summer hemisphere. The NH and SH ocean time series are also out of phase, but by about 90° , where the SH ocean maximum (minimum) occurs in the SH spring (fall). The NH ocean and land time series have approximately the same period.

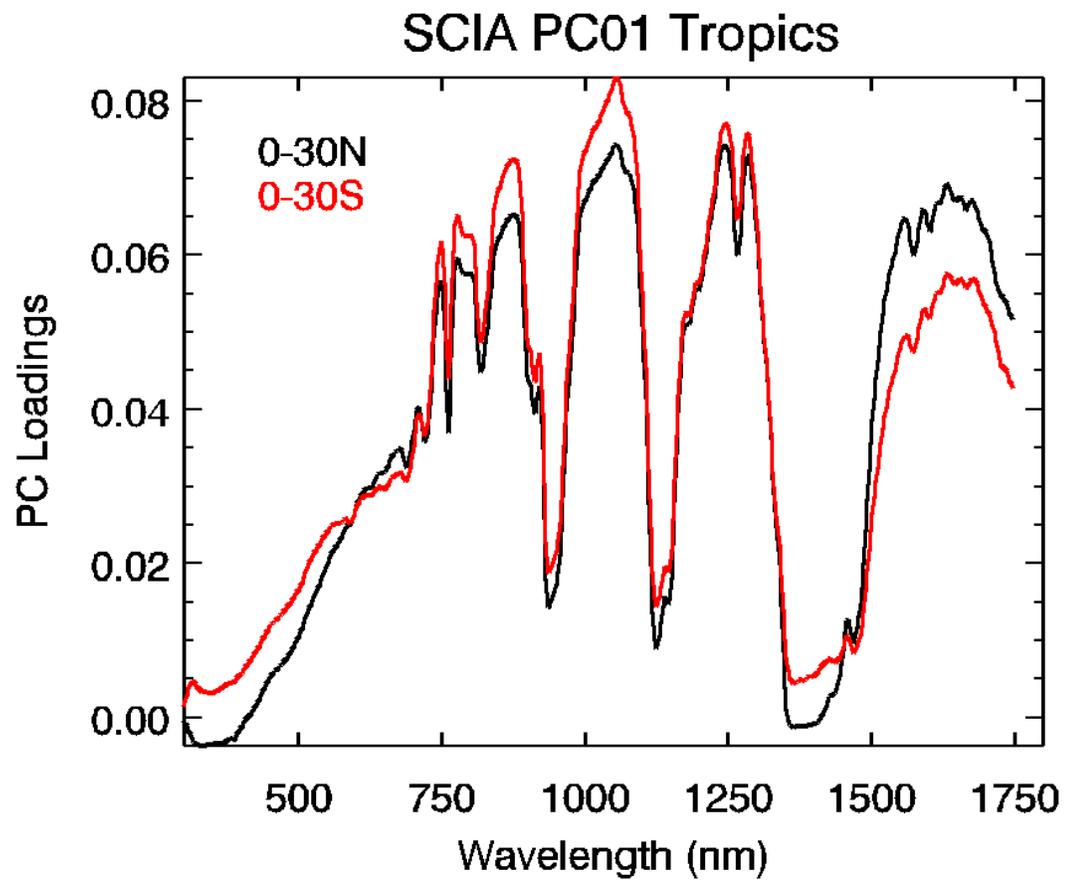


Figure 4.3: The comparison of the first principal component for the NH and SH tropics show their close similarity, which means that they are explaining very similar variables. The SH PC01 explains 88.98% of the data variance and the NH PC01 explains 95.04% of the variance.

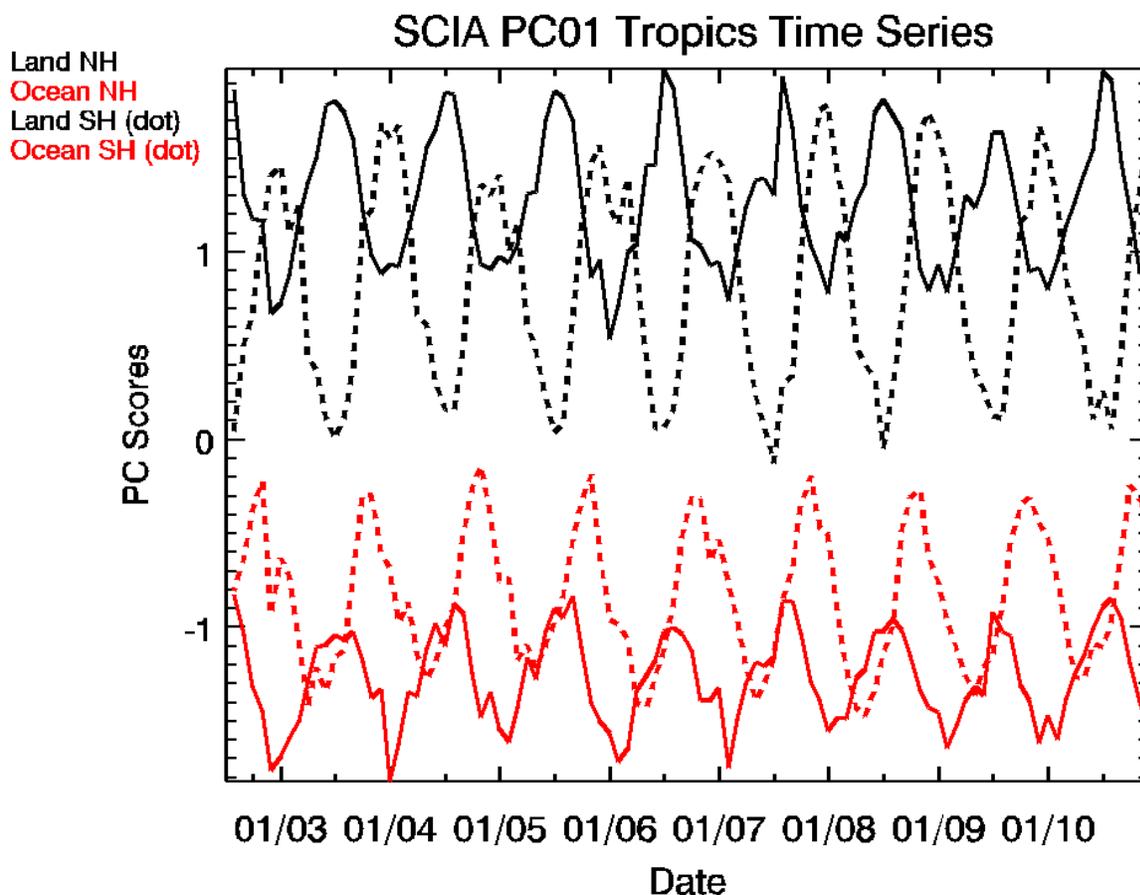


Figure 4.4: PC01 land (black) and ocean (red) scores time series for the Northern (solid) and Southern Hemisphere (dotted) Tropics zonal bands, each between 0 and 30 degrees. While the two land time series are 180 degrees out of phase, the two ocean time series are about 90 degrees out of phase. The occurrence of the maxima and minima appear to correspond to the location of the Intertropical Convergence Zone.

In the two polar regions, 60°S - 90°S and 60°N - 90°N , the first components have similar spectral shapes that resemble the spectral signatures of the cryosphere, such as snow and sea ice, confirmed by comparing these PCs in Figure 4.5 to a snow albedo spectrum [Schmidt *et al.*, 2009]. PC1 in the NH (SH) explains 93.32% (91.38%) of the data variance. There are spectral differences between the PCs and the snow albedo spectrum because there are other physical variables that contribute to the spectral signature of the PCs including gas absorption (e.g. ozone, oxygen, and water vapor) and other components of the cryosphere

in addition to snow, for example, sea ice. Figure 4.6 shows a comparison between the sea ice extent [Fetterer *et al.*, 2002] in the northern and southern hemispheres to the NH and SH PC1 time series. The PC1 in both polar regions does not explain the variability of snow or sea ice alone, but also contains information about the variability of gas absorbers, frozen surfaces, and other physical variables; however, the general match in spectral shape between PC and snow albedo suggests that snow reflectance is the dominant contributor to this PC. In the top panel of Figure 4.6, which shows the NH sea ice extent and PC1, the melt season is evident in the PC1 time series. It is likely that the scores time series extrema do not align well with the NH sea ice extent because of the limited incident sunlight in the Arctic during the NH spring and autumn. In the southern hemisphere, the time series are not aligned as well, but that may also be due to the limited sampling of Antarctica and its immediate surroundings between the SH fall and spring. In the NH polar zonal band, there is snow-covered land and sea ice that extend south to 60°N , whereas in the SH between Antarctica and the 60°S boundary, there is primarily open ocean. In the southern hemisphere, the only land in the 60°S - 90°S band is Antarctica, and the only sea ice is located directly around the continent. These differences in the location of snow and ice covered surfaces in the NH and SH polar bands may be the reason for the differences in how well sea ice extent compares to the PC1 time series.

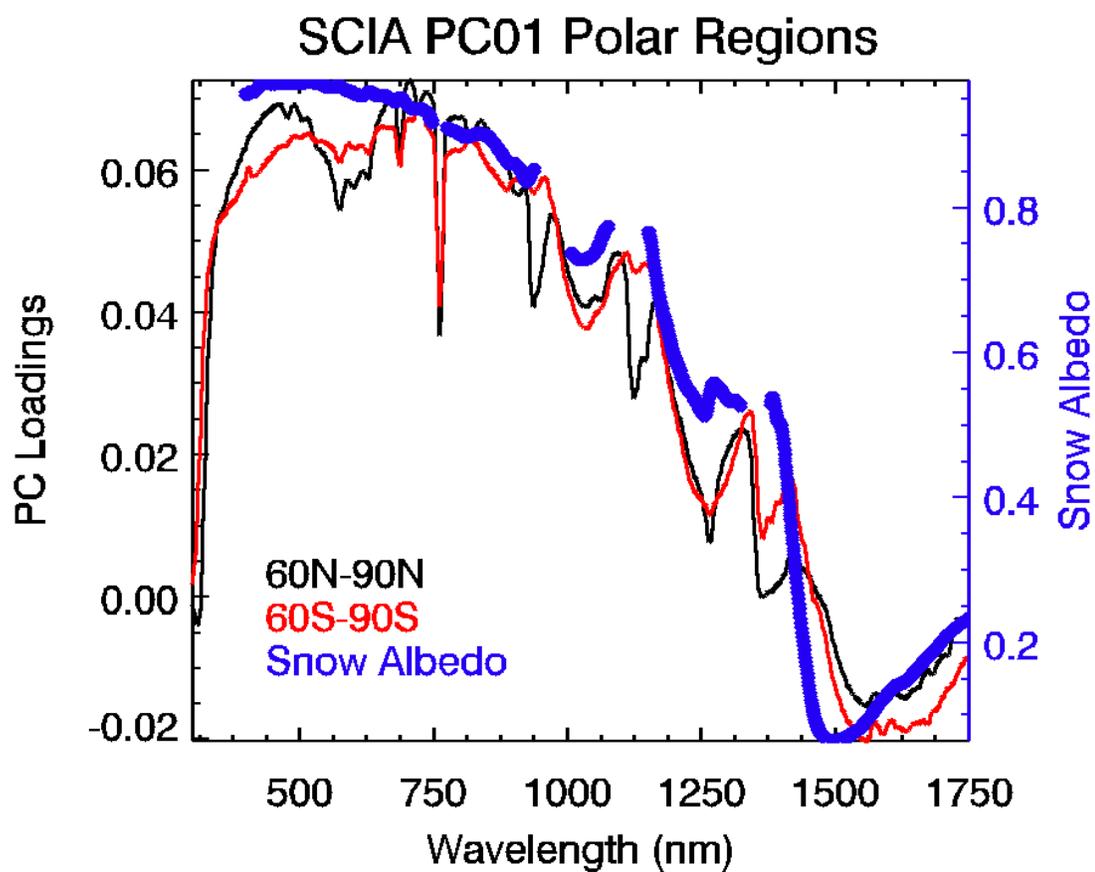


Figure 4.5: The first principal component for the two zonal bands located in the polar regions, 60°N-90°N (black) and 60°S-90°S (red) compared to a snow albedo spectrum measured during the ARCTAS field experiment (blue).

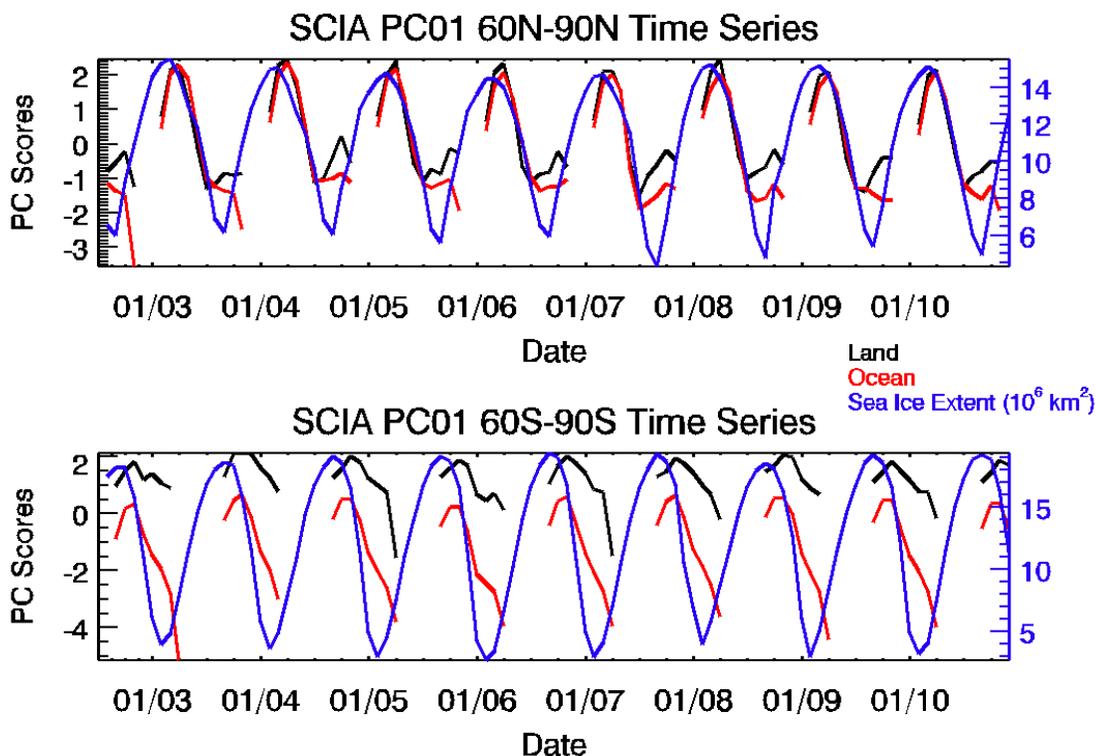


Figure 4.6: The first PCs time series for the two polar regions compared to sea ice extent in the respective hemisphere.

In the NH and SH tropics and midlatitudes bands, the second PC has nearly the same spectral shape. These four PCs also have similar spectral shapes to the second PC calculated from monthly averaged reflectance spectra with global coverage on a 5.625° grid. Figure 4.7 shows the four tropical and midlatitude PC2 spectra compared to the PC2 calculated from monthly averaged global SCIAMACHY reflectance from August 2002 to December 2010. These PCs generally have similar spectral shapes to desert reflectance, an example of which is overlaid in Figure 4.7 in blue [Clark *et al.*, 2007], although other physical variables are evident in the spectral shapes of the different PC2s, including gas absorption and vegetation. The agreement between the PC2s calculated from these zonal bands with those calculated from the monthly global data sets, suggests that PC scores from the zonally averaged and global coverage data sets are good indicators of how PC2 changes over time and space.

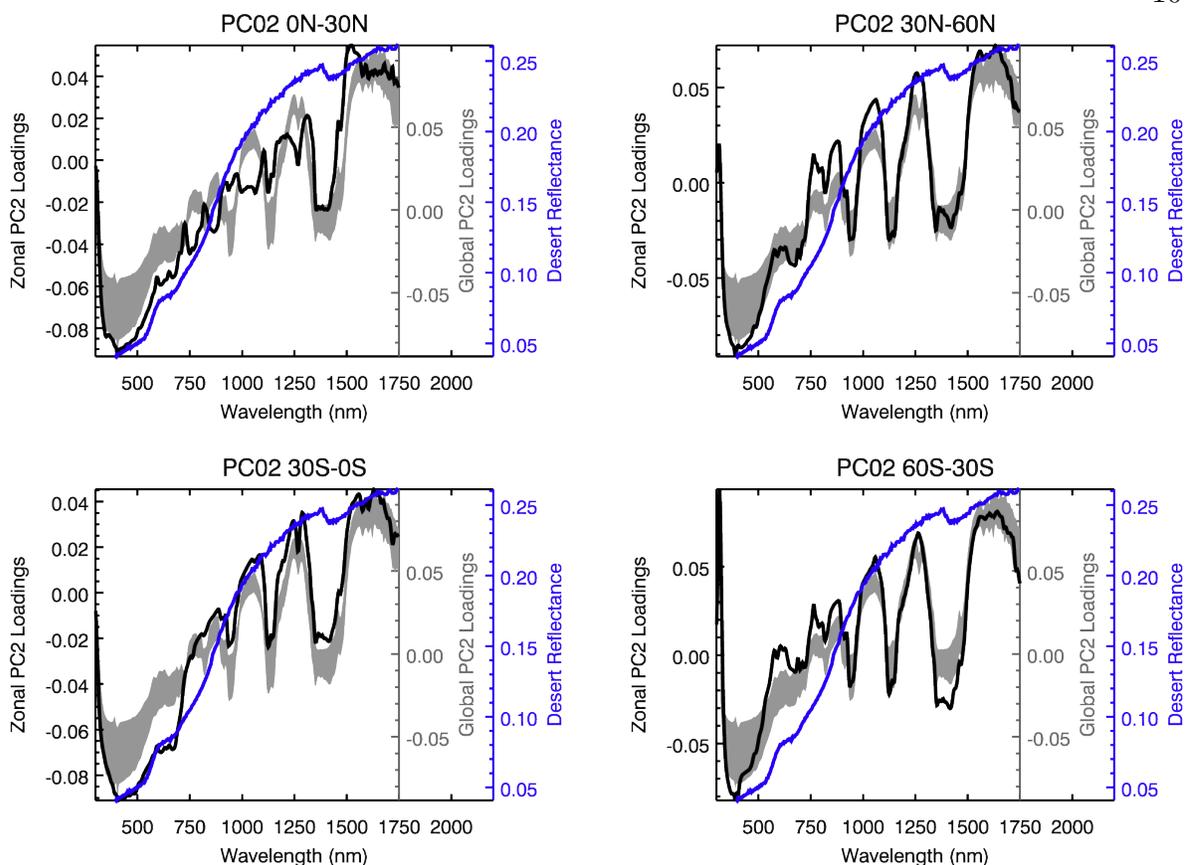


Figure 4.7: The shaded region shows the boundaries of all the PC02 spectral shapes calculated from monthly averaged, gridded SCIAMACHY reflectance spectra for every month between August 2002 and December 2010. The black line in each panel is the second principal component for the four zonal bands outside the polar regions, and the blue line is a desert reflectance spectrum from the USGS spectral library [Clark *et al.*, 2007].

Figure 4.8 shows the spatial distributions of the PC2 scores from four months in 2006: January, April, July, and October. All four maps of the scores in Figure 4.8 have the same scores range, so the colors in the four panels correspond to the same scores values. There are large positive PC2 scores in arid and semi-arid regions throughout the year including over the Sahara Desert, the Arabian Desert, and western and central Australia. Over most of the ocean the scores are consistently negative except for much of the Southern Ocean, a spatial pattern that is visible in the January, April, and October scores maps. Also, over

most land surfaces in the tropics and midlatitudes the scores are consistently positive, with the exception of the small fraction of land in the SH midlatitude band. Figure 4.9 shows the land and ocean time series of the four PC2s shown in Figure 4.7. The spatial patterns described above are also exhibited in the time series shown in Figure 4.9. For example, in the time series the ocean scores in the 30°N-60°N (Figure 4.9b) and 0°-30°S (Figure 4.9c) bands are negative throughout the year, an observation confirmed by the spatial distribution of the scores in Figure 4.8. In the 0°-30°N band, where the Saharan and Arabian Deserts have large positive scores throughout the year, there is still an annual pattern to the scores (Figure 4.9a). With the exception of the 0°-30°S band, the land and ocean time series have the same annual periods. There are also phase differences among the zonal bands. The NH and SH tropics land (ocean) time series are about 180° (90°) out of phase, two patterns that were also observed in the tropical PC1 time series (Figure 4.4). In the midlatitudes, both land and ocean time series are about 180° out of phase, with the maxima occurring in the summer and the minima occurring in the winter. The SH midlatitudes region is also the only band in which the ocean time series has a greater amplitude than the land (Figure 4.9d); this may be because this zonal band has the largest fraction of ocean among the four zonal bands.

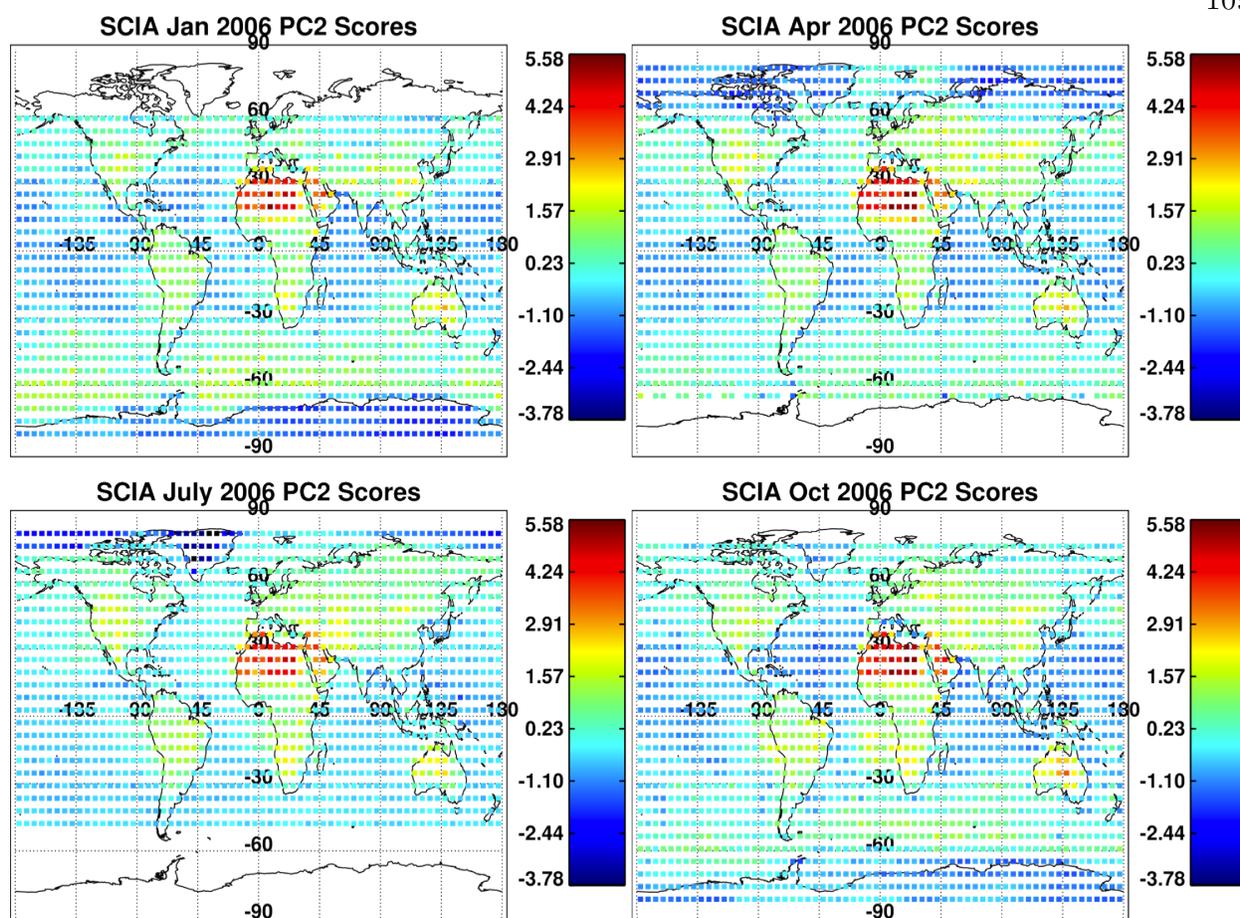


Figure 4.8: The PC2 scores calculated from four months in 2006: January, April, July, and October. The color scales in all four panels have the same data value range.

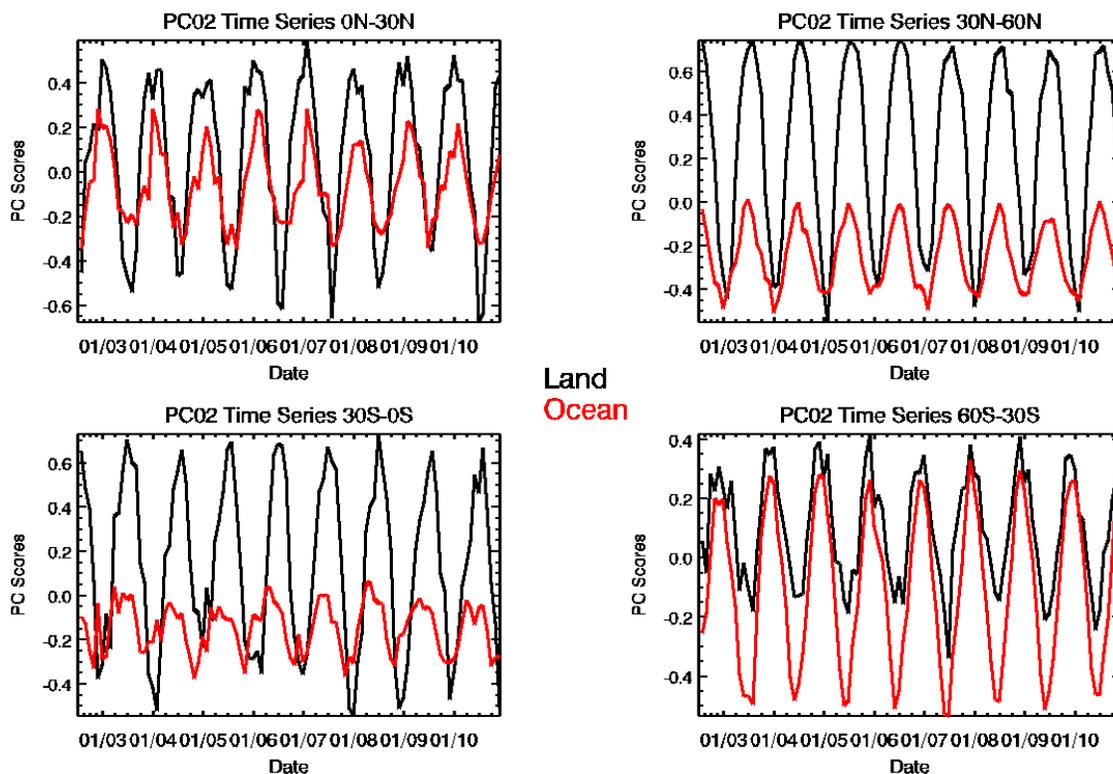


Figure 4.9: The land and ocean time series that correspond to the PC2 spectra shown in black in Figure 4.7.

In addition to annual variability, some of the PC time series also exhibit seasonal variability. Figure 4.10 shows PC4 and PC5 spectra calculated from the zonally averaged SCIAMACHY reflectance spectra in the 30°N-60°N band, along with their scores time series. The green peak, the near infrared edge, and water vapor absorption bands are distinguishable in the spectral shape of PC4. In the top right panel of Figure 4.10 there is a seasonal pattern in the land and ocean PC4 time series, where annually two maxima and minima occur. In the land time series the maxima occur in January or February and July, and the winter maximum is typically larger than the summer maximum. The minima occur in the fall and spring, and the fall minimum is always more negative than the spring maximum. The ocean time series exhibit a small phase difference relative to the land time series, but the maxima and minima occur in the same seasons as the land time series. The PC5 spectral shape

also shows evidence of vegetation reflectance and gas absorption. The PC5 time series has seasonal variability similar to PC4, but land and ocean are in phase. The land and ocean winter maxima and fall minima have nearly the same magnitude, but the amplitude between the spring minima and summer maxima are much larger for land than for ocean.

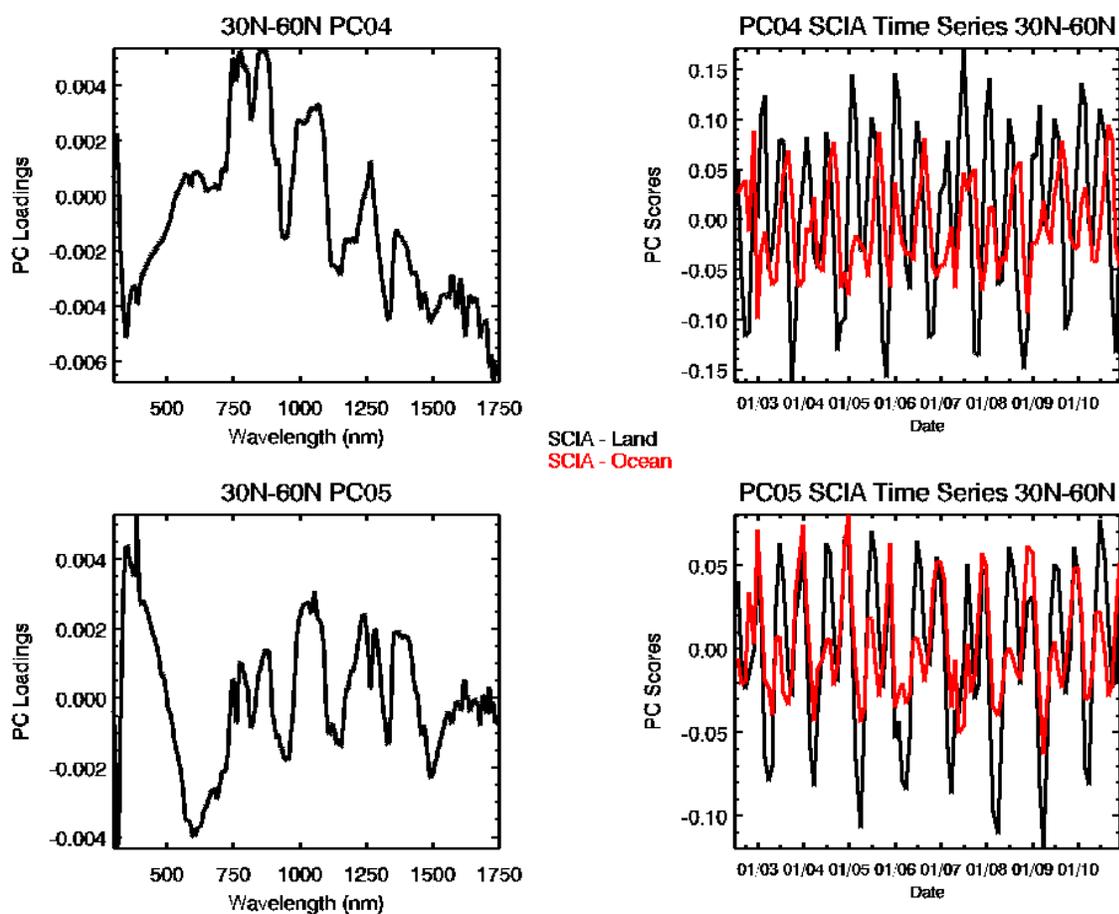


Figure 4.10: The PC4 and PC5 spectral shapes and scores time series from the 30°N-60°N latitude band. These two PC time series exhibit seasonal variations.

In each zonal band, with the exception of 60°N-90°N, there are one or two PC time series in which two peaks occur at the same times of year, January 2006 and June 2007. Figure 4.11 shows three of these occurrences in three different zonal bands. In the spectral shapes of the corresponding PCs shown in Figure 4.11, there are noisy spectral features

between 1600 nm and 1750 nm. According to the SOST data quality record, there were two instrument operation procedures that occurred during these two months, but both events affected a small number of orbits. These operations were also performed during other months when such peaks in the scores time do not occur. Because the PCs were calculated from monthly averages, instrument anomalies would need to be significant or systematic to contribute to the PC spectral signals and time series.

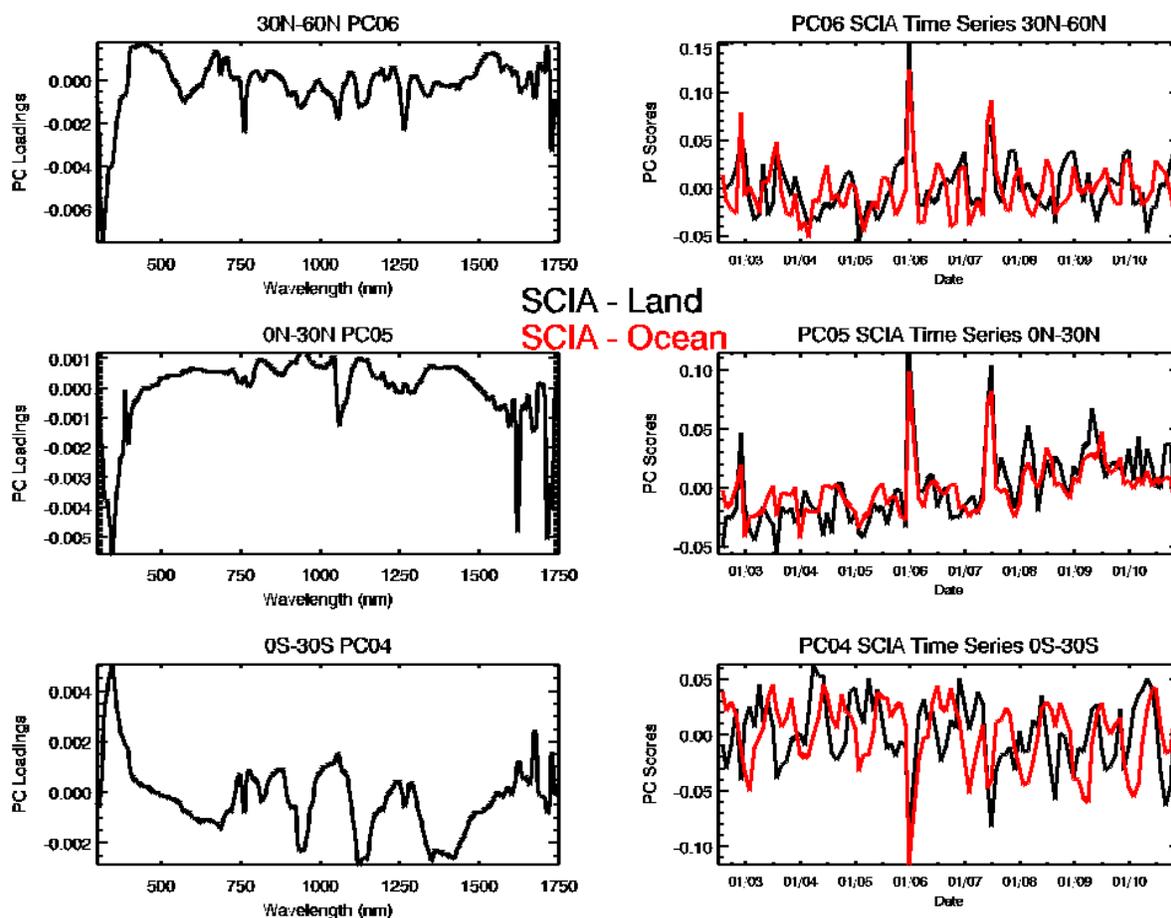


Figure 4.11: Three examples of PCs and the corresponding scores time series show how an instrument anomaly is represented in the spectral and temporal variability.

4.4.2 Comparing Observed and Simulated Temporal Variability

Using the availability of a near-decade of SCIAMACHY measurements on the CLARREO cluster, we evaluated how well the OSSE simulates Earth's temporal variability at the beginning of the century. Figure 4.12 compares the PCA results for both OSSE cases to the SCIAMACHY PCA results in the 0° - 30° N zonal band. The first three eigenvalues are nearly the same (Figure 4.12a), but diverge at larger dimensions. In the PCA results for all six zonal bands the SCIAMACHY eigenvalue slope is smaller than the OSSE slope at dimensions larger than those that define the boundary between signal and noise, perhaps because of SCIAMACHY instrument noise. Further evidence of the similarity of the first three components is shown in PCs in Figure 4.12b. In PC4 there are similarities between the SCIAMACHY and OSSE PC spectra in the near infrared, but there are stark differences as well, primarily in the visible. In the fourth, fifth, and sixth PCs, there is evidence of noise not present in the OSSE PCs, and likely specific to the SCIAMACHY instrument.

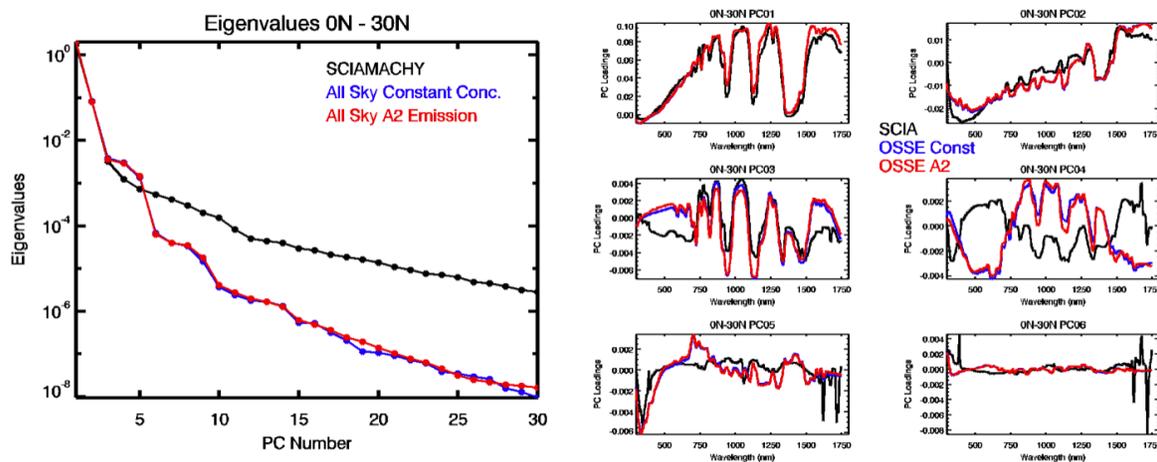


Figure 4.12: The 0° - 30° N comparison of SCIAMACHY and OSSE (control and experiment) eigenvalues (left) and principal components (right).

The SCIAMACHY and OSSE PC time series cannot be directly compared because, except for the first two PCs in Figure 4.12, the spectral shapes are not the same. This means

that the time series from the SCIAMACHY and OSSE PCs do not explain the spectral, spatial, and temporal variability of the same variables. To compare the temporal variability of the same spectral variables, we calculate the intersection between six-dimensional OSSE and SCIAMACHY subspaces [Roberts *et al.*, 2012] and project the original PC scores time series onto the Transformed PCs (TPCs). Figure 4.13 shows the 0°-30°N OSSE A2 and SCIAMACHY TPCs calculated from the intersection between the two data sets. The correlations between each pair of PCs are in the title of each subfigure. The first four TPCs have correlations that are greater than 0.95. Based on the correlations, the first three TPCs and possibly the fourth are nearly the same, indicating that the intersection may be used to evaluate the similarity of their temporal variability. The comparisons in Figure 4.14 show that the first three SCIAMACHY and OSSE land and ocean TPC time series exhibit identical periods with some differences in amplitude. There are greater differences between the two PC4 time series. In addition, the TPC4 correlation, 0.9513, is lower than the first three, perhaps indicating that the time series for this component are driven by different sources.

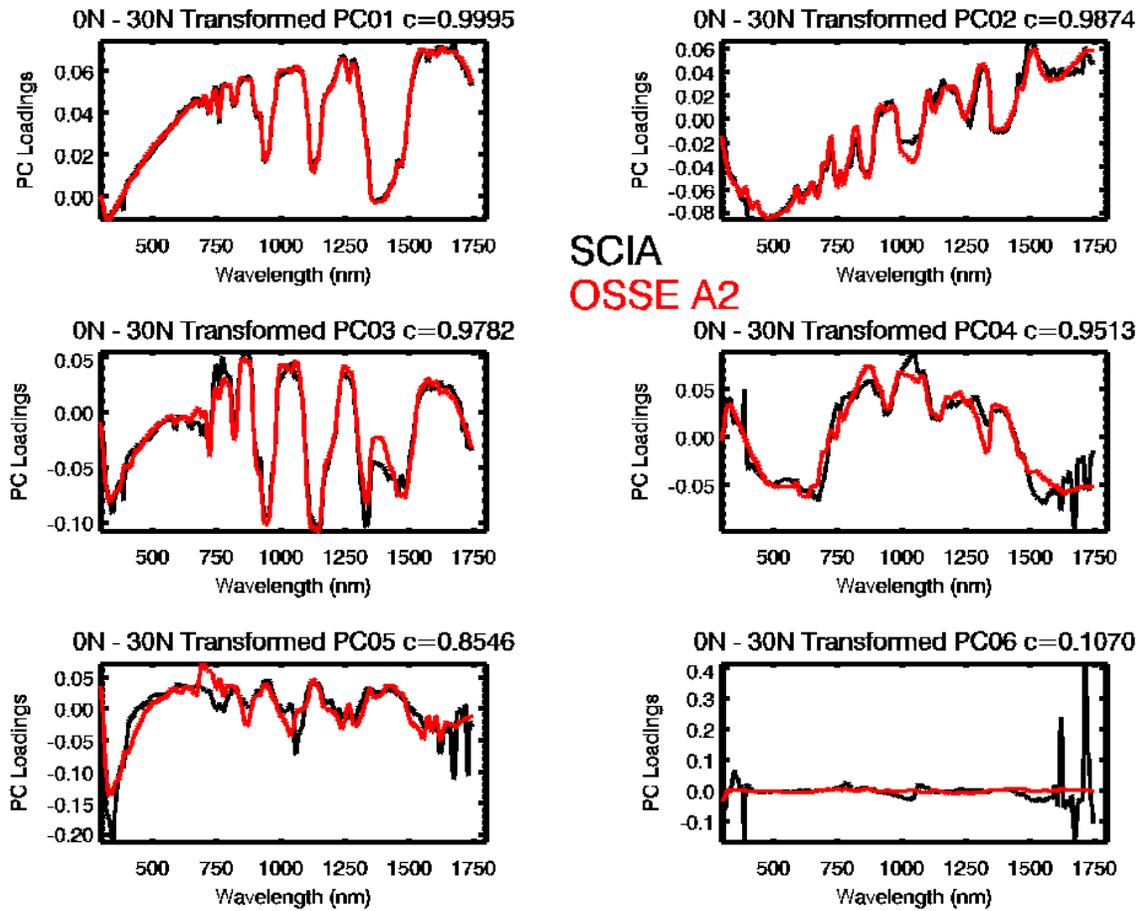


Figure 4.13: The transformed PCs calculated from the six-dimensional SCIAMACHY and OSSE A2 subspaces in the 0° - 30° N zonal band. The correlation between each OSSE and SCIAMACHY PC is shown in the title.

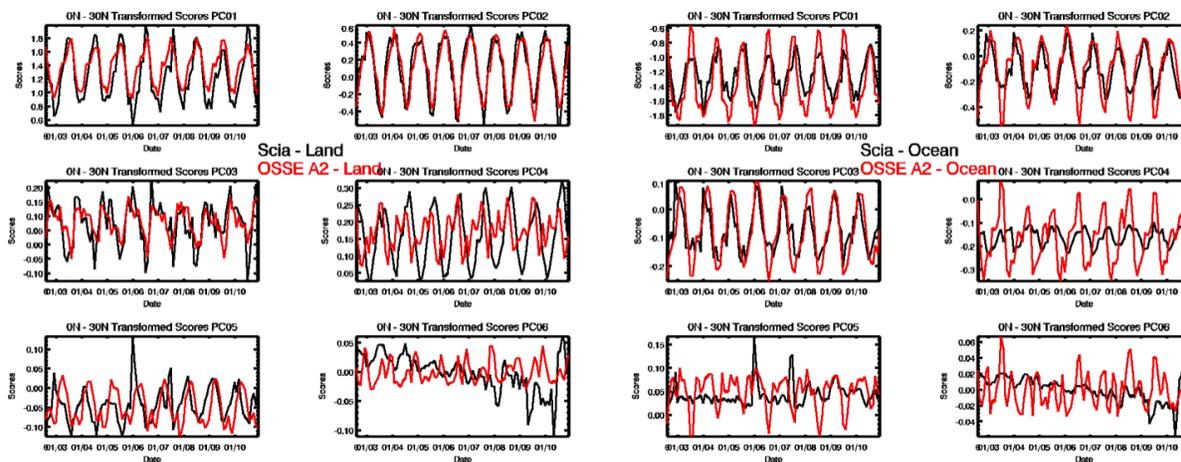


Figure 4.14: The land (left) and ocean (right) intersecting scores time series that correspond to the transformed principal components shown in Figure 4.13 in the 0° - 30° N zonal band.

Time series for other components provided further evidence that the spikes in the time series in Figure 4.11 are the result of instrument artifacts. Noisy spectral features are evident between 1600 nm and 1750 nm in the fourth, fifth, and sixth SCIAMACHY TPCs. Their absence from the OSSE TPCs imply that within the six dimensional subspace used to calculate the intersection, variables that are responsible for these noisy spectral features do not exist in the OSSE data. There is also a negative trend in the SCIAMACHY PC6 time series that is not present in the OSSE PC6 time series. Deriving secular trends in the observed signal relies on robust noise reduction in the observed signal; this analysis provides one method for doing so.

In Figure 4.15 we use the 0° - 30° S PC results to show another example of using the intersection for a direct comparison of the temporal variability of SCIAMACHY and OSSE PCs. In this case only the first two TPCs have correlations above 0.95; TPC3 and TPC4 have correlations above 0.90. The differences between the data sets due to instrument noise are manifested in the last four TPCs shown in Figure 4.15. The first two land time series differ only in amplitude (Figure 4.16). The first pair of TPC ocean scores differs in phase, while TPC2 ocean scores differ in both phase and amplitude. The lower correlations for

TPC3-6 imply that only the first two TPCs are sufficient to use for a direct comparison of the temporal variability in this data set. In the fifth and sixth TPCs, the two spikes in 2006 and 2007 are again visible in the SCIAMACHY time series.

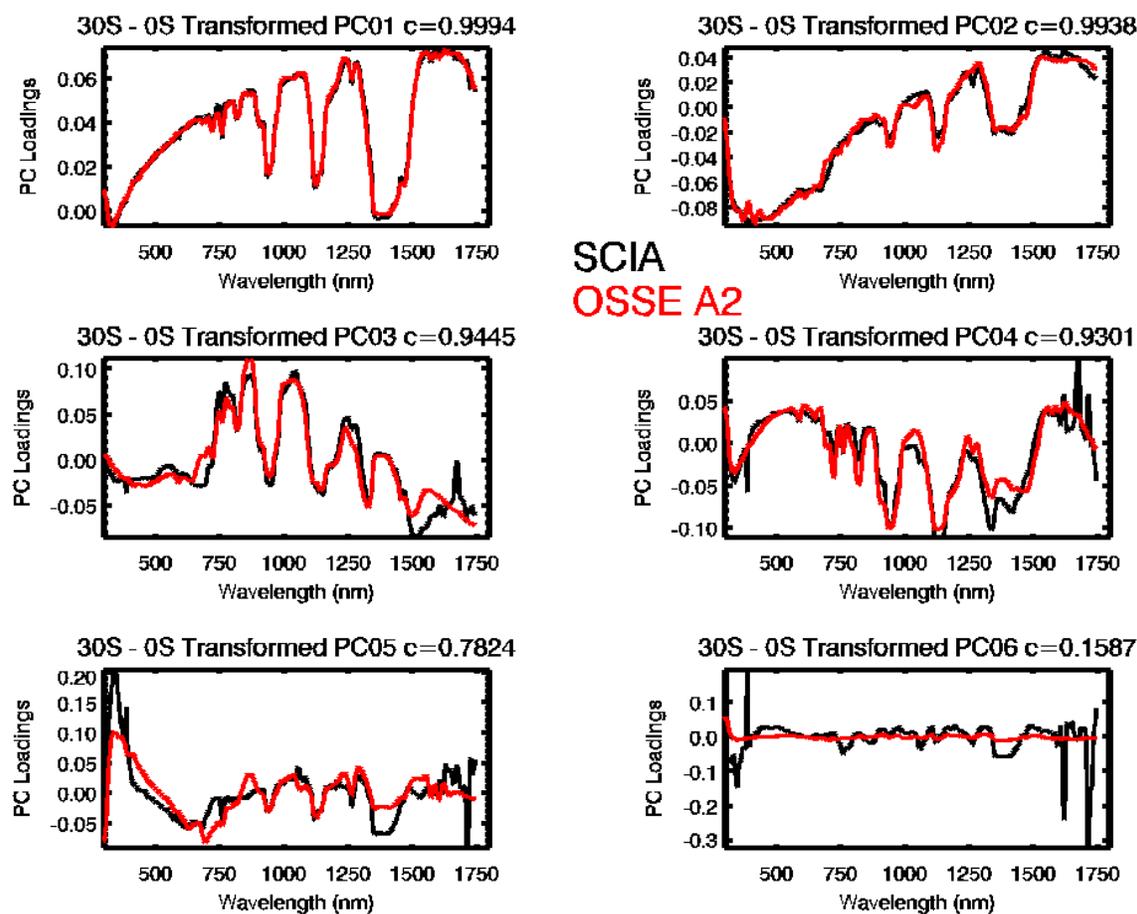


Figure 4.15: The transformed PCs calculated from the six-dimensional SCIAMACHY and OSSE A2 subspaces in the 0° - 30° S zonal band. The correlation between each OSSE and SCIAMACHY PC is shown in the title.

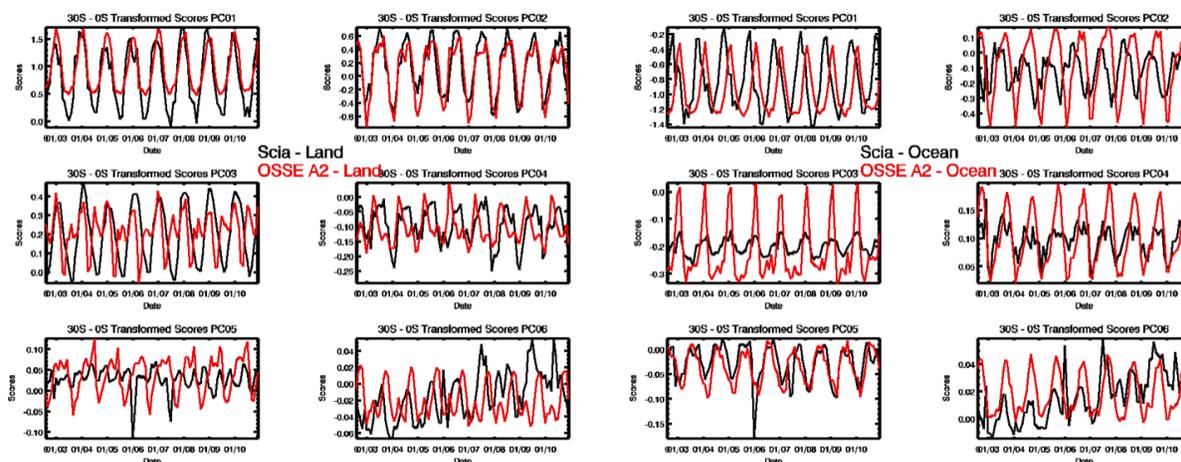


Figure 4.16: The land (left) and ocean (right) intersecting scores time series that correspond to the transformed principal components shown in Figure 4.15 from the 0° - 30° S zonal band.

4.4.3 Centennial Variability of OSSE Reflectance

To evaluate the centennial variability within the OSSE-simulated reflectance spectra, we calculated the PCs for each of the six 30° zonal bands, which included the land and ocean monthly and zonally averaged reflectance spectra between 2000 and 2099. Similar to the PCs shown in Sections 4.4.1 and 4.4.2, the PCs calculated here include information about the spatial, spectral, and temporal (i.e. centennial) variability of the two OSSE cases. Here we show the PCA results in two zonal bands: 30° N- 60° N and 0° - 30° S. Figure 4.17 shows the results from the NH midlatitude band (30° N- 60° N). The eigenvalues for the two cases, the control (constant CO_2 concentration) and experiment (A2), compare well except for the PC2 eigenvalues, and the slopes differ after the sixth dimension. There are some spectral similarities in PC2, PC4, and PC5 for the two cases, but the other PCs have spectral shapes that are very different (Figure 4.17b). The differences between these two sets of components imply that the centennial variability between these two cases is very different, an implication further confirmed by the scores time series. Atmospheric constituents remained at levels from the year 2000 throughout the unforced simulation (constant CO_2), but greenhouse gas and aerosol amounts were altered throughout the century in the forced simulation (A2). We see

the impact of these differences between the simulations manifested in the temporal patterns of the scores. Figure 4.18a and 4.18c compare the first and fifth PC land and ocean time series for both the control and experiment cases. The control time series (labeled "Const") do not exhibit secular trends at any point during the century, although they contain annual and seasonal variability. Figure 4.18b and 4.18d provide expanded views of the A2 scores to highlight the temporal trends. The A2 PC1 time series annual amplitudes increase between the beginning and end of the century. The maxima in both land and ocean time series do not change, but the minima become increasingly negative over time. The A2 PC5 land and ocean scores increase slightly between 2000 and 2050, and decrease between 2050 and 2099, but the land annual amplitude decreases while the ocean annual amplitude increases between the beginning and end of the century. In 2000, the land time series had larger amplitudes than ocean, but because of the amplitude trend, this is reversed by 2099.

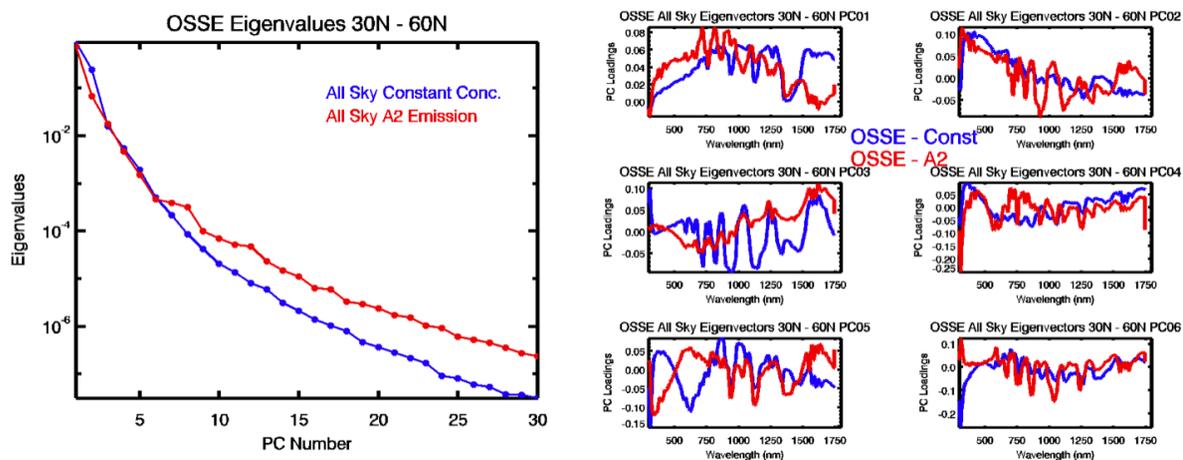


Figure 4.17: The eigenvalues (left) and principal components (right) calculated from century time series of OSSE all sky constant CO_2 concentration and A2 emission cases reflectance spectra in the 30°N - 60°N zonal band.

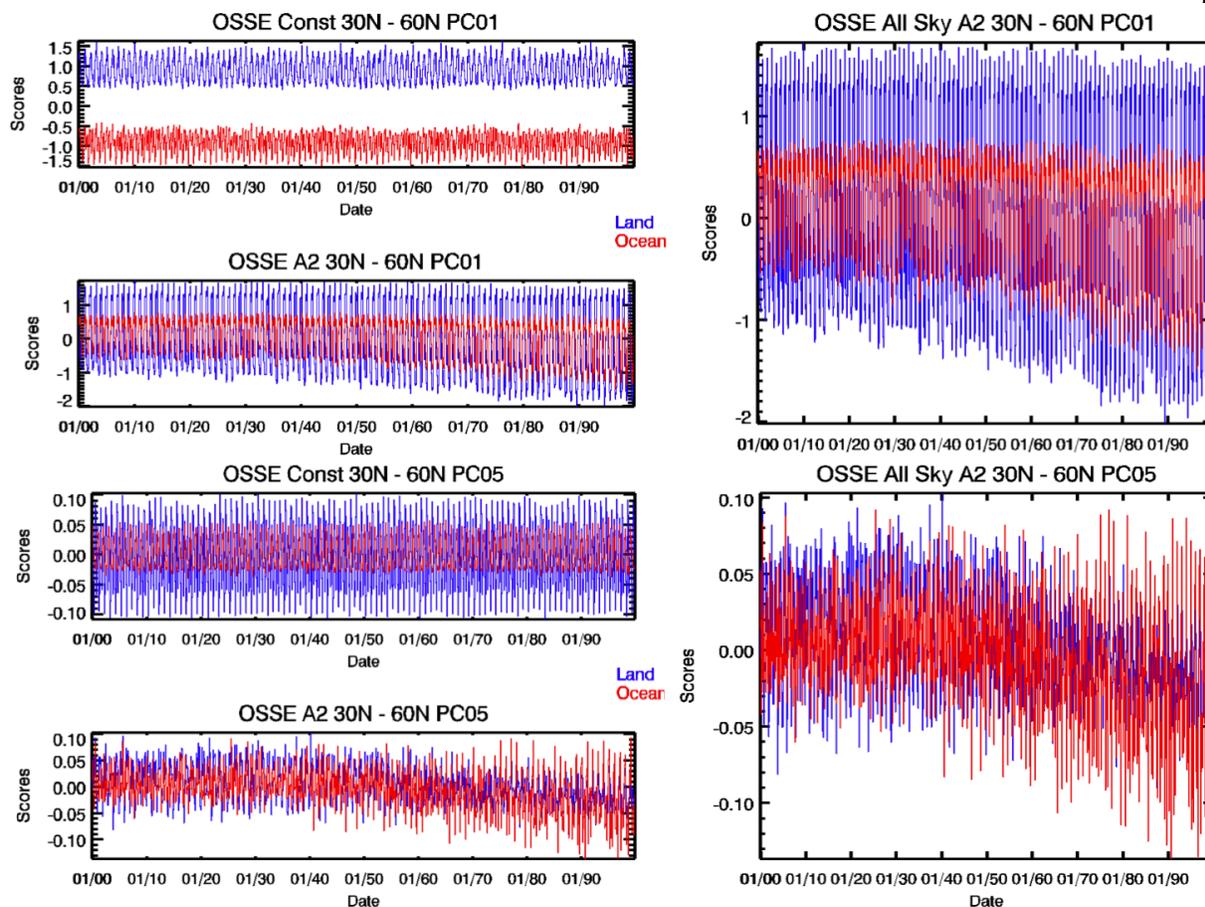


Figure 4.18: On the left, the comparison of the control and experiment PC1 (top) and PC5 (bottom) land and ocean scores time series between 2000 and 2099 in the NH midlatitude zonal band. On the right, the A2 scores time series are shown to highlight the secular trends in the forced simulation PC time series.

In the second example, we examine the PCA results from the SH tropics zonal band, 0° - 30° S. The two simulations have different variance distributions, although the spectral shapes of the first six PCs have some similarities (Figure 4.19). The PC1 and PC4 scores are shown in Figure 4.20, with comparisons between the constant CO_2 and A2 cases and expanded versions of the A2 scores. The control case exhibits annual and seasonal variability but no secular trends; however, the A2 PC time series do exhibit secular trends. The A2 PC1 trend aligns with prescribed aerosol loadings changes in the A2 model simulation [Feldman *et al.*, 2011b]. Aerosol loadings in the A2 emission scenario were increased between 2000 and

2050, after which they were decreased to 1990 levels by the end of century. In Figure 4.20b, the land and ocean time series increase between 2000 and about 2040 and decrease from that point until the end of the century. The PC4 spectral shape clearly shows water absorption bands, which is manifested in the time series trends. In the first half of the century, the A2 PC4 time series decreases, but after 2050, there is a positive trend that increases to the end of the century. This is similar to the change in water vapor prescribed in the A2 emission model simulation [Feldman *et al.*, 2011b].

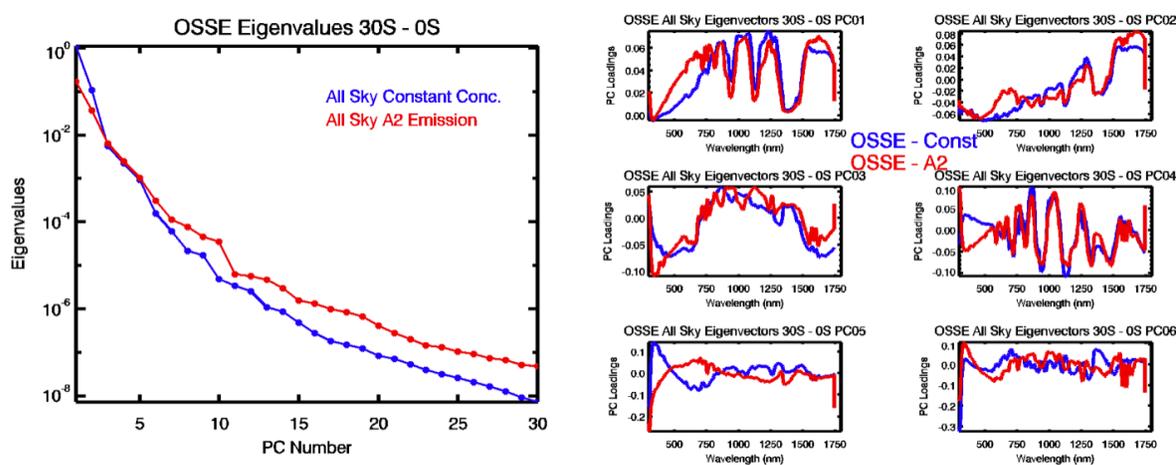


Figure 4.19: The eigenvalues (left) and principal components (right) calculated from century time series of OSSE all sky constant CO_2 concentration and A2 emission cases reflectance spectra in the 0° - 30°S zonal band.

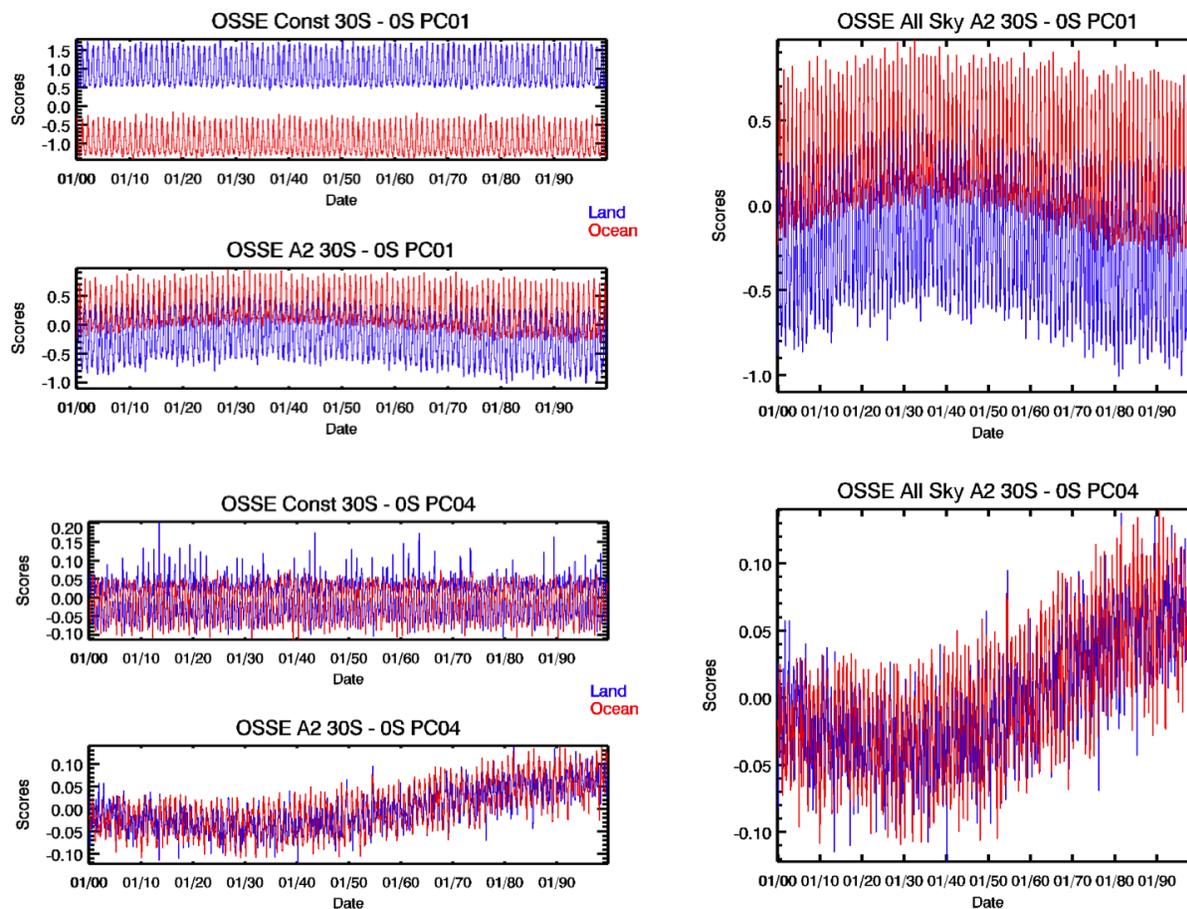


Figure 4.20: On the left, the comparison of the control and experiment PC1 (top) and PC4 (bottom) land and ocean scores time series between 2000 and 2099 in the SH Tropics zonal band. On the right, the A2 scores time series are shown to highlight the temporal trends in the forced simulation time series.

4.5 Summary

In this chapter, the temporal variability of observed and simulated hyperspectral reflectance was quantified in six 30° zonal bands that encircled the Earth. The first focus of this study was on the observed spectral variability between 2002 and 2010 using SCIAMACHY reflectance spectra. The temporal variability of SCIAMACHY reflectance exhibited temporal patterns of the seasonal movement of the Intertropical Convergence Zone (ITCZ) and the annual variability of the cryosphere. The second most dominant PC in all zonal bands except the two polar regions had an evident desert spectral signal that was also apparent in

PCs calculated with globally-distributed monthly averaged 5.625° gridded reflectance spectra. The spatial and temporal variability of this PC was observed using the scores from both the global-gridded and zonally averaged PC results. There were also dominant PCs that exhibited seasonal variations over land and ocean surfaces. Finally, noise features were identified in five out of the six zonal bands that were not evident in the simulated dominant spectral variables. In the PC time series that corresponded to the PCs containing these noisy spectral features, two extrema dominated the signal in January 2006 and July 2007. Instrument anomalies were likely responsible for these spectral features, rather than climate scale events, such as La Niña.

To evaluate how well the OSSE-simulated reflectance spectra reproduced the temporal variability in the observed spectra, we calculated the intersection between the two data sets and presented two examples: the NH and SH Tropical zonal bands. In the NH Tropics, among the intersection dimensions with high correlations, at least above 0.95, the OSSE and SCIAMACHY intersection scores time series were very similar, with the phases and amplitudes well aligned. In the SH Tropical band, the first two PCs had correlations above 0.99. For these two transformed dimensions, the OSSE and SCIAMACHY land time series overlap, but the OSSE ocean time series exhibit phase differences relative to the SCIAMACHY ocean time series. Further investigation is needed to determine what is causing the differences between these time series, but it demonstrates that the OSSE may not be simulating the ocean-atmosphere temporal variability accurately in the SH Tropics. Using the intersection to evaluate the simulated temporal variability also highlights the instrument artifacts evident in January 2006 and July 2007. The noisy spectral features were evident in SCIAMACHY transformed PCs that had low correlations with the OSSE TPCs.

Finally, we quantified the centennial variability of the OSSE reflectance spectra simulated using the IPCC AR4 constant CO_2 concentration (the control) and A2 (the experiment) emission scenarios. The PCA results clearly showed the differences between the two cases in the distribution of variance, the spectral shapes of the PCs, and the scores time

series. The constant CO₂ concentration scores time series in all zonal bands exhibited annual and seasonal variability but did not contain secular trends. The A2 scores time series contained annual and seasonal patterns as well but also had clear secular trends. The PCA results quantifying the OSSE centennial variability demonstrate how multivariate analysis techniques can be used to evaluate temporal changes in Earth's climate. Another benefit of applying these analysis techniques to simulated results is that we are able to link the changes seen in the spectral, spatial, and temporal variability to changes prescribed in the model simulations.

Chapter 5

Summary and Outlook

In this thesis, multivariate analysis techniques were used to quantify the spectral, spatial, and temporal variability of hyperspectral shortwave radiance and reflectance. The observed hyperspectral reflectance measurements were from the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY), and simulations were from the Observing System Simulation Experiments (OSSE) developed for the Climate Absolute Radiance and Refractivity Observatory (CLARREO) science definition studies. Principal component analysis (PCA) and singular spectrum analysis (SSA) were applied to these observed and simulated data sets to quantify the spectral, spatial, and temporal variability.

We presented several multivariate analysis tools to extract and analyze information from hyperspectral shortwave radiation for climate studies. Dominant spectral variables calculated from PCA of reflected radiance described the spatial and temporal distribution of spectral variability; using PCA, only six dimensions were needed to explain over 99.5% of the variance in the data. The dominant principal components exhibited global, hemispheric, regional, and seasonal variability, and reflectance signals from water vapor, land surface albedo (e.g. vegetation, snow, and sea ice), clouds, and molecular scattering were detectable in the dominant PCs. The Arctic Ocean case study showed that hyperspectral shortwave radiance could be used to distinguish between clouds and underlying snow and sea ice surfaces. Comparing a northern hemisphere sea ice extent time series to that of the Arctic Ocean PC with a spectral shape resembling snow albedo demonstrated how the dominant PCs could be used

to evaluate the temporal variability of physical climate variables. This work represents the first time that the variability of Earth-reflected hyperspectral radiance measured from space was quantified. The results from these studies contributed to the CLARREO project by demonstrating that some of the dominant spectral variables calculated directly from hyperspectral radiance could be attributed to unique sources. These results contributed in part to the decision to design the CLARREO shortwave imager with high spectral resolution, continuous spectral sampling, and full range spectral measurements characteristics.

We also developed a quantitative comparison technique to evaluate how well simulated reflectance reproduced Earth's variability relative to that observed in SCIAMACHY measurements at the beginning of the 21st century. These quantitative, objective comparisons concluded that in four months during 2004 the SCIAMACHY and OSSE data sets agreed over seven or eight transformed PC dimensions, which together explained over 99% of the variance in each data set. This was the first time the subspace intersection and distance has been applied to evaluate model output relative to observations. This technique can also be applied to compare other types of multivariate data in the atmospheric sciences and other fields. This work contributed to the CLARREO project by quantifying the comparison of spectral variability in the CLARREO-simulated reflectance spectra from the OSSE to observations of reflectance made by an operational instrument, SCIAMACHY.

Using SCIAMACHY reflectance measured between 2002 and 2010, the observed nearly-decadal temporal variability in hyperspectral reflectance measurements was quantified using 30° zonal bands encircling the Earth. Some dominant spectral variables demonstrated the temporal variability of recognized variables such as the seasonal location of the Intertropical Convergence Zone and the annual variation in snow and sea ice. Other dominant temporal variables also exhibited seasonal variability and evidence of instrument artifacts or noise. The intersection method was used to directly compare the temporal variability within SCIAMACHY measurements and OSSE simulations during the same time period.

Using the OSSE output generated between 2000 to 2009, the centennial variability of

the forced and unforced cases was quantified. The time series corresponding to the principal components calculated from century-long OSSE simulations differed between the forced and unforced scenarios. The time series corresponding to the PCs calculated from the unforced OSSE case exhibited annual and seasonal patterns, but no secular trends; however, the time series corresponding to the PCs calculated from the forced OSSE scenario exhibited secular trends during the 21st century. The centennial variability of some forced PC time series demonstrated the temporal variability of the change in aerosol amount and total column water vapor during the simulation period.

In future work, simulated reflectance spectra calculated from different climate models will be evaluated using the intersection method as a way to determine which models most accurately reproduce the observed variability relative to observations. If an appropriate noise characterization of SCIAMACHY data can be obtained, the Minimum Noise Fraction transform will be applied to SCIAMACHY measurements to spectrally whiten the noise in the data set. It is expected that using noise-whitened SCIAMACHY data to evaluate OSSE output will be more appropriate because the OSSE reflectance spectra were calculated without measurement uncertainty [*Feldman et al.*, 2011b].

One new topic related to this research involves more detailed analysis of temporal variability of hyperspectral signals. Narrower zonal bands that have longitude boundaries that do not encircle the entire globe will enable more focused study on specific climate processes. For example, changes to the reflected solar signal over the Tropical Pacific Ocean due to La Niña and El Niño events may be resolved. Additionally, Singular Spectrum Analysis will be used to quantify the dominant temporal patterns in the PC scores time series. The Temporal PCs will be used to calculate the intersection to quantify the similarity between observed and simulated dominant PC time series and to quantify how well the OSSE is able to reproduce the temporal variability observed relative to SCIAMACHY reflectance. Finally, the SCIAMACHY data currently available on the CLARREO cluster included measurements only through the end of 2010. Extending the record to the full decade of measurements will

enable the quantification of the observed decadal variability in SCIAMACHY reflectance and quantitative comparison between OSSE and SCIAMACHY data through April 2012.

In addition to the future work discussed above, another new project will contribute to the CLARREO Science Definition Team by evaluating the decadal and global scale accuracy of operational cloud retrievals and provide newly developed climate-specific algorithms. Current measurements of reflected solar radiation are used to retrieve precipitable water, cloud fraction, optical thickness, effective radius, and thermodynamic phase. Because typical process-specific studies that use these cloud properties focus on small spatial (e.g. local or regional) and short time (e.g. hourly, diurnal, synoptic, or seasonal) scales, cloud retrieval algorithms have been developed to minimize pixel-level errors. However, the accuracy of these retrievals has not been tested on global and decadal scales relevant for climate studies. In fact, certain aspects of current satellite retrieval algorithms at the pixel-level may cause aliasing (cloud property distortions) in the decadal-length time series of retrieved cloud properties. Contributions to aliasing include averaging errors from inherent non-linearities, three-dimensional inhomogeneities, and discretized lookup tables. This novel study will be an investigation of how these potential contributions to aliasing errors impact long-term and large scale averages often used in climate studies.

Bibliography

- Anderson, G. P., A. Berk, P. K. Acharya, M. W. Matthew, L. S. Bernstein, J. H. Chetwynd, H. Dothe, S. M. Adler-Golden, A. J. Ratkowski, G. W. Felde, J. A. Gardner, M. L. Hoke, S. C. Richtsmeier, B. Pukall, J. Mello, and L. S. Jeong, MODTRAN4: Radiative transfer modeling for remote sensing, *Optics in Atmospheric Propagation and Adaptive Systems III*, 3866, 2–10, 1999.
- Baldrige, A. M., S. J. Hook, C. I. Grove, and G. Rivera, The ASTER spectral library version 2.0, *Remote Sens. Environ.*, 113(4), 711–715, 2009.
- Barkstrom, B. R., The earth radiation budget experiment ERBE, *Bull. Am. Meteorol. Soc.*, 65, 1170–1185, 1984.
- Barkstrom, B. R., Earth radiation budget measurements: pre-ERBE, ERBE, and CERES, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, vol. 1299, edited by B. R. Barkstrom, pp. 52–60, 1990.
- Boardman, J., and R. Green, Exploring the spectral variability of the earth as measured by AVIRIS in 1999, in *Proc. Ninth Airborne Earth Science Workshop*, Jet Propulsion Laboratory, 2000.
- Boardman, J., et al., Automating spectral unmixing of aviris data using convex geometry concepts, in *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1, pp. 11–14, JPL Publication 93–26, 1993.
- Bony, S., R. Colman, V. Kattsov, R. Allan, C. Bretherton, J. Dufresne, A. Hall, S. Hallegatte, M. Holland, W. Ingram, et al., How well do we understand and evaluate climate change feedback processes?, *Journal of Climate*, 19(15), 3445–3482, 2006.
- Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance, and A. P. H. Goede, SCIAMACHY: Mission objectives and measurement modes, *J. Atmos. Sci.*, 56(2), 127–150, 1999.
- Bowker, D., R. Davis, D. Myrick, K. Stacy, and W. Jones, Spectral reflectances of natural targets for use in remote sensing studies, *Tech. rep.*, National Aeronautics and Space Administration, Scientific and Technical Information Branch, 1985.

- Bramstedt, K., M. Gottwald, E. Krieg, and S. Noël, Sciamachy operations support, 2012.
- Buchwitz, M., O. Schneising, J. Burrows, H. Bovensmann, M. Reuter, and J. Notholt, First direct observation of the atmospheric CO₂ year-to-year increase from space, *Atmos. Chem. Phys.*, 7(16), 4249–4256, 2007.
- Cattell, R., The scree test for the number of factors, *Multivariate behavioral research*, 1(2), 245–276, 1966.
- Chan, T., M. Mozurkewich, et al., Application of absolute principal component analysis to size distribution data: identification of particle origins, *Atmospheric Chemistry and Physics Discussions*, 6(5), 10,493–10,522, 2006.
- Charlson, R. J., F. P. J. Valero, and J. H. Seinfeld, In search of balance, *Science*, 308(5723), 806–807, doi:10.1126/science.1108162, 2005.
- Chiu, H., and W. Collins, A spectroradiometer for airborne remote sensing, *Photogrammetric Engineering and Remote Sensing*, 44, 507–517, 1978.
- Clark, R., G. Swayze, R. Wise, E. Livo, T. Hoefen, R. Kokaly, and S. Sutley, Usgs digital spectral library splib06a, U.S. Geological Survey, Digital Data Series 231, 2007.
- Collins, W., C. Bitz, M. Blackmon, G. Bonan, C. Bretherton, J. Carton, P. Chang, S. Doney, J. Hack, T. Henderson, et al., The community climate system model version 3 (ccsm3), *Journal of Climate*, 19(11), 2122–2143, 2006.
- Craddock, J., and C. Flood, Eigenvectors for representing the 500 mb geopotential surface over the northern hemisphere, *Quarterly Journal of the Royal Meteorological Society*, 95(405), 576–593, 1969.
- Crone, L., and D. Crosby, Statistical applications of a metric on subspaces to satellite meteorology, *Technometrics*, pp. 324–328, 1995.
- Davies, R., and M. Molloy, Global cloud height fluctuations measured by misr on terra from 2000 to 2010, *Geophysical Research Letters*, 39(3), L03,701, 2012.
- Diner, D., J. Beckert, T. Reilly, C. Bruegge, J. Conel, R. Kahn, J. Martonchik, T. Ackerman, R. Davies, S. Gerstl, et al., Multi-angle imaging spectroradiometer (misr) instrument description and experiment overview, *Geoscience and Remote Sensing, IEEE Transactions on*, 36(4), 1072–1087, 1998.
- Elsner, J., and A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*, Plenum Publishing Corporation, 1996.
- Elvidge, C., and Z. Chen, Comparison of broad-band and narrow-band red and near-infrared vegetation indices, *Remote Sens. Environ.*, 54(1), 38–48, 1995.
- Farmer, S., An investigation into the results of principal component analysis of data derived from random numbers, *The Statistician*, pp. 63–72, 1971.

- Fasullo, J. T., and K. E. Trenberth, The annual cycle of the energy budget. Part I: Global mean and land ocean exchanges, *J. Climate*, *21*, 2297–2311, doi:10.1175/2007JCLI1935.1, 2008.
- Feldman, D., C. Algieri, W. Collins, Y. Roberts, and P. Pilewskie, Simulation studies for the detection of changes in broadband albedo and shortwave nadir reflectance spectra under a climate change scenario, *Journal of Geophysical Research*, *116*(D24), D24,103, 2011b.
- Feldman, D. R., C. A. Algieri, J. R. Ong, and W. D. Collins, CLARREO shortwave observing system simulation experiments of the twenty-first century: Simulator design and implementation, *J. Geophys. Res.*, *116*(D15), D10,107, doi:10.1029/2010JD015350, 2011a.
- Fetterer, F., K. Knowles, W. Meier, and M. Savoie, Sea ice index, Boulder, Colorado USA: National Snow and Ice Data Center. Digital media., updated 2009, 2002.
- Fox, N., J. Aiken, J. Barnett, X. Briottet, R. Carvell, C. Frohlich, S. Groom, O. Hagolle, J. Haigh, H. Kieffer, et al., Traceable radiometry underpinning terrestrial-and helio-studies (truths), *Advances in Space Research*, *32*(11), 2253–2261, 2003.
- Fox, N., A. Kaiser-Weiss, W. Schmutz, K. Thome, D. Young, B. Wielicki, R. Winkler, and E. Woolliams, Accurate radiometry from space: an essential tool for climate studies, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1953), 4028–4063, 2011.
- Francis, J. A., E. Hunter, J. R. Key, and X. Wang, Clues to variability in Arctic minimum sea ice extent, *Geophys. Res. Lett.*, *32*, L21,501, doi:10.1029/2005GL024376, 2005.
- Gamon, J., C. Field, D. Roberts, S. Ustin, and R. Valentini, Functional patterns in an annual grassland during an aviris overflight, *Remote Sensing of Environment*, *44*(2), 239–253, 1993.
- Gao, B., W. Han, S. Tsay, and N. Larsen, Cloud detection over the Arctic region using airborne imaging spectrometer data during the daytime, *J. Appl. Meteorol.*, *37*(11), 1421–1429, 1998.
- Ghil, M., and K. Mo, Intraseasonal oscillations in the global atmosphere. Part II: Southern hemisphere, *J. Atmos. Sci.*, *48*(5), 780–790, 1991.
- Goetz, A., Three decades of hyperspectral remote sensing of the Earth: A personal view, *Remote Sens. Environ.*, *113*, S5–S16, doi:10.1016/j.rse.2007.12.014, 2009.
- Goetz, A., G. Vane, J. Solomon, and B. Rock, Imaging spectrometry for Earth remote sensing, *Science*, *228*(4704), 1147–1153, 1985.
- Goetz, A. F. H., K. B. Heidebrecht, B. Kindel, and J. W. Boardman, Using ground spectral irradiance for model correction of AVIRIS data, in *Summaries of the Seventh JPL Airborne Earth Science Workshop January 12-16, 1998*, 1998.

- Golyandina, N., V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and related techniques*, CRC Press, 2001.
- Goody, R., J. Anderson, and G. North, Testing climate models: An approach, *BULLETIN-AMERICAN METEOROLOGICAL SOCIETY*, 79, 2541–2549, 1998.
- Gottwald, M., R. Hoogeveen, C. Chlebek, H. Bovensmann, J. Carpay, G. Lichtenberg, E. Krieg, P. Lützow-Wentzky, and T. Watts, *SCIAMACHY - Exploring the Earth's Changing Atmosphere*, chap. 10, pp. 175–216, Springer, 2011a.
- Gottwald, M., R. Hoogeveen, C. Chlebek, H. Bovensmann, J. Carpay, G. Lichtenberg, E. Krieg, P. Lützow-Wentzky, and T. Watts, *SCIAMACHY - Exploring the Earth's Changing Atmosphere*, chap. 3, pp. 29–46, Springer, 2011b.
- Gottwald, M., R. Hoogeveen, C. Chlebek, H. Bovensmann, J. Carpay, G. Lichtenberg, E. Krieg, P. Lützow-Wentzky, and T. Watts, *SCIAMACHY - Exploring the Earth's Changing Atmosphere*, chap. 6, pp. 77–97, Springer, 2011c.
- Green, A., M. Berman, P. Switzer, and M. Craig, A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *Geoscience and Remote Sensing, IEEE Transactions on*, 26(1), 65–74, 1988.
- Green, R., M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis, et al., Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS), *Remote Sens. Environ.*, 65(3), 227–248, 1998.
- Grenfell, T. C., and D. K. Perovich, Incident spectral irradiance in the Arctic Basin during the summer and fall, *J. Geophys. Res.*, 113(D12117), doi:10.1029/2007JD009418, 2008.
- Hannachi, A., I. Jolliffe, D. Stephenson, and N. Trendafilov, In search of simple structures in climate: simplifying eofs, *International journal of climatology*, 26(1), 7–28, 2006.
- Hannachi, A., I. Jolliffe, and D. Stephenson, Empirical orthogonal functions and related techniques in atmospheric science: A review, *International Journal of Climatology*, 27(9), 1119–1152, 2007.
- Hansen, J., L. Nazarenko, R. Ruedy, M. Sato, J. Willis, A. Del Genio, D. Koch, A. Lacis, K. Lo, S. Menon, T. Novakov, J. Perlwitz, G. Russell, G. A. Schmidt, and N. Tausnev, Earth's energy imbalance: Confirmation and implications, *Science*, 308, 1431–1435, doi: 10.1126/science.1110252, 2005.
- Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann, Earth's energy imbalance and implications, *Atmos. Chem. Phys*, 11, 13,421–13,449, 2011.
- Hardy, D., and J. Walton, Principal components analysis of vector wind measurements., *Journal of Applied Meteorology*, 17, 1153–1162, 1978.

- Haskins, R. D., R. M. Goody, and L. Chen, A statistical method for testing a general circulation model with spectrally resolved satellite data, *J. Geophys. Res.*, *102*(D14), 16,563 – 16,581, 1997.
- Haskins, R. D., R. M. Goody, and L. Chen, Radiance covariance and climate models, *J. Climate*, *12*, 1409 – 1422, 1999.
- Hotelling, H., Analysis of a complex of statistical variables into principal components., *Journal of educational psychology*, *24*(6), 417, 1933.
- House, F., A. Gruber, G. Hunt, and A. Mecherikunnel, History of satellite missions and measurements of the Earth radiation budget (1957-1984), *Rev. Geophys.*, *24*(2), 357–377, 1986.
- Huang, X., and Y. L. Yung, Spatial and spectral variability of the outgoing thermal IR spectra from AIRS: A case study of July 2003, *J. Geophys. Res.*, *110*(D9), D12,102, doi: 10.1029/2004JD005530, 2005.
- Huang, X., J. Farrara, S. S. Leroy, Y. L. Yung, and R. M. Goody, Cloud variability as revealed in outgoing infrared spectra: Comparing model to observation with spectral EOF analysis, *Geophys. Res. Lett.*, *29*(8), 1270, doi:10.1029/2001GL014176, 2002.
- Huete, A. R., K. Didan, Y. E. Shimabukuro, P. Ratana, S. R. Saleska, L. R. Hutyrá, W. Yang, R. R. Nemani, and R. Myneni, Amazon rainforests green-up with sunlight in dry season, *Geophys. Res. Lett.*, *33*, L06,405, doi:10.1029/2005GL025583, 2006.
- Hunt, G., Spectral signatures of particulate minerals in the visible and near infrared, *Geophysics*, *42*(3), 501–513, 1977.
- IPCC, Synthesis report: Contribution of working groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change Core Writing Team, *IPCC, Geneva, Switzerland*, p. 104, 2007.
- Jolliffe, I., Discarding variables in a principal component analysis. i: Artificial data, *Applied statistics*, pp. 160–173, 1972.
- Jolliffe, I. T., *Principal Component Analysis*, Springer-Verlag, New York, 2002.
- Jordan, C., Derivation of leaf-area index from quality of light on the forest floor, *Ecology*, pp. 663–666, 1969.
- Kaiser, H., The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, *23*(3), 187–200, 1958.
- Kaiser, H., Computer program for varimax rotation in factor analysis, *Educational and Psychological Measurement*, 1959.
- Kaiser, H., The application of electronic computers to factor analysis., *Educational and psychological measurement*, 1960.

- Kandel, R., and M. Viollier, Observation of the Earth's radiation budget from space, *Comptes Rendus Geoscience*, 342(4-5), 286–300, doi:10.1016/j.crte.2010.01.005, 2010.
- Karoly, D. J., P. Hope, and P. D. Jones, Decadal variations of the southern hemisphere circulation, *Int. J. Climatol.*, 16(7), 723–738, 1996.
- Kato, S., Interannual Variability of the Global Radiation Budget, *Journal of Climate*, 22, 4893–4907, doi:10.1175/2009JCLI2795.1, 2009.
- Kelly, P., P. Jones, C. Sear, B. Cherry, and R. Tavakol, Variations in surface air temperatures: Part 2. Arctic Regions, 1881-1980, *Month. Weath. Rev.*, 110(2), 71–83, 1982.
- Keshava, N., A survey of spectral unmixing algorithms, *Lincoln Laboratory Journal*, 14(1), 55–78, 2003.
- Kidson, J., Eigenvector analysis of monthly mean surface data, *Month. Weath. Rev.*, 103(3), 177–186, 1975.
- King, M., Y. Kaufman, W. Menzel, and D. Tanre, Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (modis), *Geoscience and Remote Sensing, IEEE Transactions on*, 30(1), 2–27, 1992.
- King, M., W. Menzel, Y. Kaufman, D. Tanré, B. Gao, S. Platnick, S. Ackerman, L. Remer, R. Pincus, and P. Hubanks, Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from modis, *Geoscience and Remote Sensing, IEEE Transactions on*, 41(2), 442–458, 2003.
- Kokhanovsky, A. A., V. V. Rozanov, J. P. Burrows, K.-U. Eichmann, W. Lotz, and M. Vountas, The SCIAMACHY cloud products: Algorithms and examples from ENVISAT, *Advances in Space Research*, 36, 789–799, doi:10.1016/j.asr.2005.03.026, 2005.
- Kopp, G., and J. L. Lean, A new, lower value of total solar irradiance: Evidence and climate significance, *Geophys. Res. Lett.*, 38, L01,706, doi:10.1029/2010GL045777, 2011.
- Kopp, G., G. Lawrence, and G. Rottman, The total irradiance monitor (tim): Science results, *Sol. Phys.*, 230, 129–139, doi:10.1007/s11207-005-7433-9, 2005.
- Krzanowski, W., Between-groups comparison of principal components, *Journal of the American Statistical Association*, pp. 703–707, 1979.
- Kutzbach, J., Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over north america, *J. Appl. Meteor.*, 6(5), 791–802, 1967.
- Loeb, N., B. Wielicki, W. Su, K. Loukachine, W. Sun, T. Wong, K. Priestley, G. Matthews, W. Miller, and R. Davies, Multi-instrument comparison of top-of-atmosphere reflected solar radiation, *Journal of climate*, 20(3), 575–591, 2007.

- Loeb, N., S. Kato, W. Su, T. Wong, F. Rose, D. Doelling, J. Norris, and X. Huang, Advances in understanding top-of-atmosphere radiation variability from satellite observations, *Surveys in geophysics*, pp. 1–27, 2012.
- Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, Toward Optimal Closure of the Earth’s Top-of-Atmosphere Radiation Budget, *Journal of Climate*, *22*, 748–766, doi:10.1175/2008JCLI2637.1, 2009.
- Loveland, T., and A. Belward, The igbp-dis global 1km land cover data set, discover: first results, *International Journal of Remote Sensing*, *18*(15), 3289–3295, 1997.
- Loveland, T., B. Reed, J. Brown, D. Ohlen, Z. Zhu, L. Yang, and J. Merchant, Development of a global land cover characteristics database and igbp discover from 1 km avhrr data, *International Journal of Remote Sensing*, *21*(6-7), 1303–1330, 2000.
- Markus, T., J. Stroeve, and J. Miller, Recent changes in arctic sea ice melt onset, freezeup, and melt season length, *Journal of Geophysical Research*, *114*(C12), C12,024, 2009.
- Meehl, G., C. Covey, B. McAvaney, M. Latif, and R. Stouffer, Overview of the coupled model intercomparison project (cmip), *Bulletin of the American Meteorological Society*, *86*, 89–93, 2005.
- Meehl, G., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. Mitchell, R. Stouffer, and K. Taylor, The wcrp cmip3 multi-model dataset: A new era in climate change research, *Bulletin of the American Meteorological Society*, *88*, 1383–1394, 2007.
- Michalsky, J., Q. Min, J. Barnard, R. Marchand, and P. Pilewskie, Simultaneous spectral albedo measurements near the atmospheric radiation measurement southern great plains (ARM SGP) central facility, *J. Geophys. Res.*, *108*, 4254, doi:10.1029/2002JD002906, 2003.
- Mieruch, S., S. Noël, H. Bovensmann, and J. P. Burrows, Analysis of global water vapour trends from satellite measurements in the visible spectral range, *Atmospheric Chemistry and Physics*, *8*(3), 491–504, doi:10.5194/acp-8-491-2008, 2008.
- Minnis, P., and M. Khaiyer, Anisotropy of land surface skin temperature derived from satellite data, *J. Appl. Meteorol.*, *39*(7), 1117–1129, 2000.
- National Research Council, N., *Earth science and applications from space: National imperatives for the next decade and beyond*, National Academy Press, 2007.
- Noel, S., H. Bovensmann, J. Skupin, M. Wuttke, J. Burrows, M. Gottwald, and E. Krieg, The sciamachy calibration/monitoring concept and first results, *Advances in Space Research*, *32*(11), 2123–2128, 2003.
- North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions, *Mon. Weather Rev.*, *110*(7), 699–706, 1982.
- Obukhov, A., Statistically homogeneous fields on a sphere, *Usp. Mat. Nauk*, *2*(2), 196–198, 1947.

- Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, Satellite instrument calibration for measuring global climate change, *Bull. Amer. Meteor. Soc.*, *86*(9), 1303–1313, 2005.
- Overpeck, J. T., M. Sturm, J. A. Francis, D. K. Perovich, M. C. Serreze, R. Benner, E. C. Carmack, F. S. Chapin, III, S. C. Gerlach, L. C. Hamilton, L. D. Hinzman, M. Holland, H. P. Huntington, J. R. Key, A. H. Lloyd, G. M. McDonald, J. McFadden, D. Noone, T. D. Prowse, P. Schlosser, and C. Vörösmarty, Arctic system on trajectory to new, seasonally ice-free state, *EOS Transactions*, *86*, 309–313, doi:10.1029/2005EO340001, 2005.
- Pallé, E., P. R. Goode, P. Montañés-Rodríguez, and S. E. Koonin, Changes in Earth's Reflectance over the Past Two Decades, *Science*, *304*, 1299–1301, doi:10.1126/science.1094070, 2004.
- Pearlman, J. S., P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, Hyperion, a space-based imaging spectrometer, *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1160–1173, doi:10.1109/TGRS.2003.815018, 2003.
- Pearson, K., Liii. on lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572, 1901.
- Pilewskie, P., J. Pommier, R. Bergstrom, W. Gore, S. Howard, M. Rabbette, B. Schmid, P. V. Hobbs, and S. C. Tsay, Solar spectral radiative forcing during the Southern African Regional Science Initiative, *J. Geophys. Res.*, *108*, 8486, doi:10.1029/2002JD002411, 2003.
- Platnick, S., M. King, S. Ackerman, W. Menzel, B. Baum, J. Riédi, and R. Frey, The modis cloud products: Algorithms and examples from terra, *Geoscience and Remote Sensing, IEEE Transactions on*, *41*(2), 459–473, 2003.
- Plaut, G., and R. Vautard, Spells of low-frequency oscillations and weather regimes in the northern hemisphere, *J. Atmos. Sci.*, *51*, 210–236, 1994.
- Polyak, I., Observed versus simulated second-moment climate statistics in gcm verification problems, *Journal of the atmospheric sciences*, *53*(5), 677–694, 1996.
- Preisendorfer, R. W., and C. D. Mobley, *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, 1988.
- Rabbette, M., and P. Pilewskie, Multivariate analysis of solar spectral irradiance measurements, *J. Geophys. Res.*, *106*, 9685–9696, doi:10.1029/2000JD900582, 2001.
- Rabbette, M., and P. Pilewskie, Principal component analysis of Arctic solar irradiance spectra, *J. Geophys. Res.*, *107*, 8049, doi:10.1029/2000JC000566, 2002.
- Rasmusson, E., X. Wang, and C. Ropelewski, The biennial component of ENSO variability, *J. Mar. Syst.*, *1*(1-2), 71–96, 1990.
- Richman, M., Rotation of principal components, *Int. J. Climatol.*, *6*(3), 293–335, 1986.

- Roberts, Y., P. Pilewskie, and B. Kindel, Evaluating the observed variability in hyperspectral earth-reflected solar radiance, *Journal of Geophysical Research*, 116(D24), D24,119, 2011.
- Roberts, Y. L., P. Pilewskie, B. C. Kindel, D. R. Feldman, and W. D. Collins, Quantitative comparison of the variability in observed and simulated shortwave reflectance, *Atmos. Chem. and Phys. Discuss.*, 12(10), 28,305–28,341, doi:10.5194/acpd-12-28305-2012, 2012.
- Rottman, G., The SORCE Mission, *Sol. Phys.*, 230, 7–25, doi:10.1007/s11207-005-8112-6, 2005.
- Santer, B. D., C. Mears, F. J. Wentz, K. E. Taylor, P. J. Gleckler, T. M. L. Wigley, T. P. Barnett, J. S. Boyle, W. Bruggemann, N. P. Gillett, S. A. Klein, G. A. Meehl, T. Nozawa, D. W. Pierce, P. A. Stott, W. M. Washington, and M. F. Wehner, Identification of human-induced changes in atmospheric moisture content, *Proceedings of the National Academy of Science*, 104, 15,248–15,253, doi:10.1073/pnas.0702872104, 2007.
- Schmidt, G., R. Ruedy, J. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V. Canuto, Y. Cheng, A. D. Genio, G. Faluvegi, A. Friend, T. Hall, Y. Hu, M. Kelley, N. Kiang, D. Koch, A. Lacis, J. Lerner, K. Lo, R. Miller, L. Nazarenko, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, G. Russell, M. Sato, D. Shindell, P. Stone, S. Sun, N. Tausnev, D. Thresher, and M.-S. Yao, Present-day atmospheric simulations using giss modele: Comparison to in situ, satellite, and reanalysis data, *Journal of Climate*, 19(2), 153–192, 2006.
- Schmidt, K. S., E. Bierwirth, P. Pilewskie, J. Redemann, R. E. Brandt, A. Lyapustin, C. Gatebe, C. Schaaf, and R. Kahn, Airborne measurements of surface albedo in alaska, in *Fall Meet. Suppl., Abstract A43A-0159*, vol. 90, 2009.
- Schneising, O., M. Buchwitz, M. Reuter, J. Heymann, H. Bovensmann, and J. Burrows, Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from sciamachy, *Atmospheric Chemistry and Physics*, 11(6), 2863–2880, 2011.
- Sellers, W., *Physical climatology*, University of Chicago Press, 1965.
- Smith, G., B. Wielicki, B. Barkstrom, R. Lee, K. Priestley, T. Charlock, P. Minnis, D. Kratz, N. Loeb, and D. Young, Clouds and Earth radiant energy system: An overview, *Adv. Space Res.*, 33, 1125–1131, doi:10.1016/S0273-1177(03)00739-7, 2004.
- Smith, M., P. Johnson, and J. Adams, Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis, *J. Geophys. Res.*, 90, C797–C804, 1985.
- Smith, M., S. Ustin, J. Adams, and A. Gillespie, Vegetation in deserts: I. a regional measure of abundance from multispectral images, *Remote sensing of Environment*, 31(1), 1–26, 1990.
- Soden, B., and I. Held, An assessment of climate feedbacks in coupled ocean-atmosphere models, *Journal of Climate*, 19(14), 3354–3360, 2006.

- Spreen, G., L. Kaleschke, and G. Heygster, Sea ice remote sensing using AMSR-E 89 GHz channels, *J. Geophys. Res.*, *113*, C02S03, doi:10.1029/2005JC003384, 2008.
- Stephens, G., Cloud feedbacks in the climate system: A critical review, *J. Climate*, *18*(2), 237–273, doi:10.1175/JCLI-3243.1, 2005.
- Sun, X., and Q. Cheng, On subspace distance, *Image Analysis and Recognition*, pp. 81–89, 2006.
- Trishchenko, A. P., Y. Luo, K. Khlopenkov, and Z. Li, Dynamics of the surface albedo over the ARM SGP area during spring 2003 Aerosol IOP, in *Fourteenth ARM Science Team Meeting Proc. Albuquerque, New Mexico, March 22-26, 2004*, 2004.
- Vane, G., A. Goetz, and J. Wellman, Airborne imaging spectrometer: A new tool for remote sensing., *IEEE Transactions on Geoscience and Remote Sensing*, *22*(6), 546–549, 1984.
- Vautard, R., P. Yiou, and M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, *58*(1-4), 95–126, 1992.
- Vinnikov, K., A. Robock, R. Stouffer, J. Walsh, C. Parkinson, D. Cavalieri, J. Mitchell, D. Garrett, and V. Zakharov, Global warming and northern hemisphere sea ice extent, *Science*, *286*(5446), 1934, 1999.
- W. Collins, G. R., S.H. Chang, Mineralogical mapping of sites near death valley, california and crossman peak, arizona, using airborne near-infrared spectral measurements, in *Proc. Intl. Symp. on Remote Sens. of Environ., 2nd Thematic Conference on Remote Sensing for Exploration Geology, ERIM, Fort Worth, TX*, 1982.
- Wallace, J., and D. Gutzler, Teleconnections in the geopotential height field during the northern hemisphere winter, *Month. Weath. Rev.*, *109*(4), 784–812, 1981.
- Wang, L., X. Wang, and J. Feng, Intrapersonal subspace analysis with application to adaptive bayesian face recognition, *Pattern recognition*, *38*(4), 617–621, 2005.
- Wang, L., X. Wang, and J. Feng, Subspace distance analysis with application to adaptive bayesian algorithm for face recognition, *Pattern recognition*, *39*(3), 456–464, 2006.
- Wang, X., and J. R. Key, Arctic surface, cloud, and radiation properties based on the AVHRR polar pathfinder dataset. Part II: Recent trends, *J. Climate*, *18*(14), 2575–2593, doi:10.1175/JCLI3439.1, 2005.
- Wielicki, B. A., T. Wong, N. G. Loeb, P. Minnis, K. Priestley, and R. Kandel, Changes in Earth’s albedo measured by satellite, *Science*, *308*, 825, doi:10.1126/science.1106484, 2005.
- Wielicki, B. A., D. F. Young, M. G. Mlynczak, K. J. Thome, S. Leroy, J. Corliss, J. G. Anderson, C. Ao, R. Bantges, F. Best, K. Bowman, H. Brindley, W. Collins, J. A. Dykema,

D. R. Doelling, D. R. Feldman, N. Fox, X. Huang, R. Holz, Y. Huang, Z. Jin, D. Jennings, D. G. Johnson, K. Jucks, S. Kato, R. Knuteson, G. Kopp, D. P. Kratz, X. Liu, C. Lukashin, A. J. Mannucci, N. Phojanamongkolkij, P. Pilewskie, S. Platnick, V. Ramaswamy, H. Revercomb, J. Rice, Y. Roberts, C. M. Roithmayr, F. Rose, S. Sandford, E. L. Shirley, S. W.L. Smith, B. Soden, P. W. Speth, L. Strow, W. Sun, P. Taylor, D. Tobin, and X. Xiong, Climate absolute radiance and refractivity observatory (CLARREO): Achieving climate change absolute accuracy in orbit, *B. Am. Meteorol. Soc.*, Submitted.

Wilks, D., *Statistical Methods in the Atmospheric Sciences, vol. 91*, vol. 630, Elsevier, 2006.

Willis, J., D. Roemmich, and B. Cornuelle, Interannual variability in upper ocean heat content, temperature, and thermocline expansion on global scales, *J. Geophys. Res.*, 109(C18), C12,036, doi:10.1029/2003JC002260, 2004.