Quantitative comparison of the variability in observed and simulated shortwave reflectance

Joheen Chakraborty (jc5110)

May 2, 2022

1 Introduction

Radiation from the Sun which reflects off Earth encodes vital information about atmospheric composition and climate variables. From analyzing a reflected spectrum, one can obtain information about cloud properties and cover, trace gas abundances, aerosol chemical composition, and albedo on the Earth's surface (Loeb et al., 2007). Spatial and temporal variability of atmospheric and surface components thus depend on a complex range of physical and chemical processes, such as wavelength-dependent scattering and absorption processes and atmospheric abundances.

For my project, I followed along closely with a previous study by Roberts et al. (2013), which quantitatively compared spectral decompositions of variability in observed and simulated reflectance spectra to serve as a testing ground for state-of-the-art climate simulation models. I adopt similar methodology, though with slightly different datasets so as to produce original results and test a different type of simulation product. In Sec. 1.1 and 1.2, I describe the datasets used for my analysis. In section 2.1, I describe how physical variables are extracted from reflectance spectra using the DOAS algorithm. In section 2.2, I describe my approach to characterizing spatial reflectance variability using PCA, and in section 2.3 I compare the results between observed and simulated datasets. Finally, in section 3 I make concluding remarks and suggest promising future directions for further work.

1.1 Observational data: SCIAMACHY spectrograph

Atoms and molecules in the atmosphere primarily interact with incoming radiation from the sun in the following ways: (I) photoelectric absorption followed by re-emission, or (II) one of various scattering processes, which slightly modulates the radiation wavelength after the interaction. To first order, the incoming solar radiation is well-described by a 5800 K blackbody spectrum modified by solar Fraunhofer lines (spectral absorption lines from the surface of the Sun) as well as atomic absorption lines from constituents of Earth's atmosphere. Thus, by studying and understanding the reflected solar radiation, as well as its interactions with trace gases during the process of travelling through the atmosphere, we can understand the chemistry at play within Earth's climate system.

Accordingly, there has been historical interest in ground- and satellite-based monitoring of the reflected short-wave radiation since around the 1970s, with a recent installment being the **Sc**anning **I**maging **A**bsorption Spectro**m**eter for **A**tmospheric **C**artograp**hy**, or SCIAMACHY spectrograph onboard the European Space Agency's ENVISAT satellite (ESA, 2013). SCIAMACHY was launched with the following stated goals:

1. Improve understanding of physical and chemical processes determining the behavior of the atmosphere;
2. Demonstrate and assess the capability of remote sensing from space for Earth System and Atmospheric Science; and

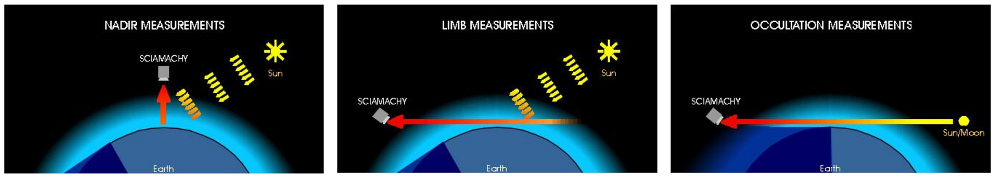

Figure 1: The 3 SCIAMACHY viewing modes taken at each point in its sun-synchronous orbit. For this analysis, I restrict to only using nadir measurements to avoid the nontrivial problem of viewing-angle correction and disentangling directly transmitted solar radiation from atmosphere-processed reflected radiation.

3. Move towards a global observing system sufficient for Earth System Science and provide the global data necessary for policymakers

In particular, data from SCIAMACHY would allow climate researchers to study the effect of natural and anthropogenic sources on global atmospheric composition Gottwald et al. (2011). In order to use SCIAMACHY data products to extract critical information about these very important goals, it is first important to ensure a detailed understanding of the pre-processing steps in collecting and analyzing the data.

ENVISAT adopted a near-polar, sun-synchronous orbit, from which SCIAMACHY made various types of spectral measurements from 2002 until a satellite failure in 2012. It was sensitive to radiation in the range 214-2386 nanometers, corresponding from near-ultraviolet to near-infrared, and capturing about 95% of the incoming radiation from the Sun. There are three primary sources of radiation measured in a typical SCIAMACHY spectrum:

- Scattered and reflected spectral radiance in nadir and limb geometry
- Spectral radiance transmitted through the atmosphere in the solar and lunar occultation geometry
- Extraterrestrial solar irradiance and the lunar radiance

Each measured spectrum will be an admixture of these components, with varying extents dependent on occultation and viewing geometry. SCIAMACHY has three viewing modes: nadir, limb, and occultation, corresponding to various degrees of viewing angle (from face-on to edge-on; see Fig.1). To avoid needing to disentangle this admixture, I resrict this analysis only to nadir observations, and make the standard assumption that all observed radiation is reflected radiation, originally emitted from the Sun and reprocessed by trace gases and aerosols within Earth's atmosphere.

For this project, I used hyperspectral reflectance measurements from 2004, following the lead of Roberts et al. (2013). Data was restricted to the range 300-1750 nanometers in order to avoid the issue of dealing with ice deposited on channels corresponding to the highest-wavelength/lowest-energy infrared light. Nadir measurement pixel size is heavily dependent on the length of integration time and swath width, varying between 26 km (along track observations) by 30 km (across track observations) and 32 km (along track) by 930 km (across track). For nadir sampling, SCIAMACHY has a scanning angular width of ±32° across track, which corresponds to a maximum nadir swath width of 960 km (Gottwald et al., 2011). The spatial resolution of SCIAMACHY measurements compares favorably to other global-coverage, space-based atmospheric spectrometers, making it a good candidate to compare products of shortwave reflectance spectroscopy with climate models.

1.2 Simulated data: CCSM 3.0

Feldman et al. (2011) constructed a grid of Observation System Simulation Experiments (OSSEs) using input from the Community Climate System Model (CCSM 3.0, NSF) Global Climate Model. The all-sky reflectance spectra are generated using unforced constant carbon dioxide emission, in which well-mixed radiatively active atmospheric greenhouse gases and aerosols

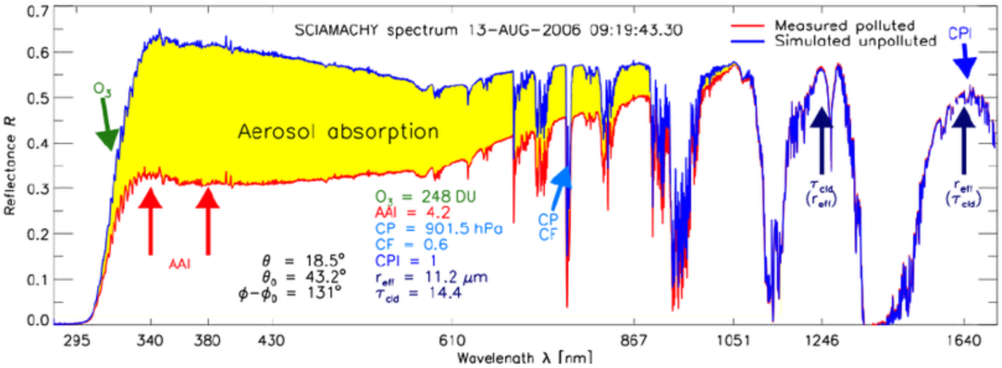

Figure 2

are held at constant levels in the year 2000 throughout the model run. Results from the forced A2 emission scenario, 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. Changes in the climate system over the course of the century include a tripling of CO2 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 data at the beginning of the 21st century, when differences between the two scenarios are minimal (Roberts et al., 2013).

2 Methods

2.1 Extraction of trace gas data

Roberts et al. (2013) directly used infrared-to-ultraviolet reflectance spectra (both simulated and observed) as input for a Principal Component Analysis (PCA) approach. I adopt a different approach, instead using data products from the next stage of processing provided by the ESA SCIAMACHY pipeline.

Following acquisition of the raw spectral reflectances from nadir observations, the data processing pipeline designed for SCIAMACHY uses the Differential Optical Absorption Spectroscopy (DOAS) approach (Burrows, 1999) for retrieval of trace gas column densities from the spectral information (ESA, 2013). In Fig. 2, which contains a sample collected SCIAMACHY optical spectrum, we can see in addition to the continuous spectral energy distribution (SED) at lower wavelengths, that there are several discrete “dips" in the SED at wavelengths from about 700-1600 nanometers. These correspond to absorption lines of various atmospheric gases. I wanted to test the hypothesis that the data encoded by these absorption lines are the primary data of interest in isolating trace gas column densities. If so, we don’t need to use the entire spectrum, which lives in a much higher-dimensional space than the products we can extract from DOAS, allowing for more efficient data storage and analysis.

The DOAS approach used by SCIAMACHY is adopted from Burrows (1999), and relies on the main ideas of isolating high-frequency structure of molecular absorbers via Rayleigh and Mie scattering; and the separation of spectroscopic retrievals and radiative transfer calculations. DOAS finds the total amount of absorption and scattering by dividing the Earthshine radiance by the direct solar irradiance. The latter provides the absorption free background. The molecular absorption cross section together with a polynomial is then fitted to the logarithm of this ratio, yielding the trace gas concentration along the light path (slant column concentration). Finally, the average light path through the atmosphere is calculated using a radiative transfer model. Using the DOAS algorithm, atmospheric columns of a number of species can be determined, including O3, NO2, SO2, HCHO, BrO, and OClO. I can then directly acquire these trace gas column densities as input for data analysis.

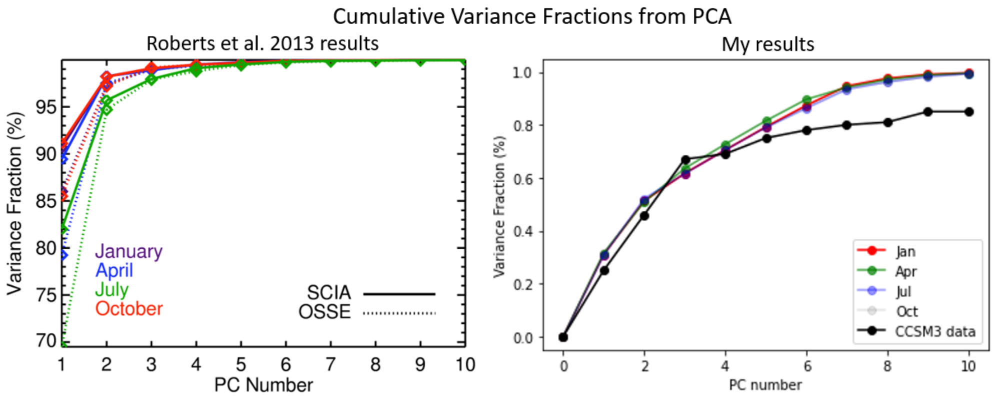

Figure 3

2.2 Principal Component Analysis

I used trace gas abundances, derived as described in Sec. 3.1 and obtained from the European Space Agency online data portal[1], from observed SCIAMACHY spectra, as input to a Principal Component Analysis (PCA). Prior to running the algorithm, I standardized the data sampling (spatially and temporally) between the observed and simulated datasets. Because I am quantitatively evaluating their similarity, it is important that the spatial and temporal resolution are similar so as not to introduce unwanted sampling artifacts. Both datasets were resampled to a grid box of $5.625°$ (corresponding to 32 latitude × 62 longitude bins), four times the size of the original CCSM3 simulation grid, to ensure that SCIAMACHY pixels were sampled at least every three days (the approximate time over which SCIAMACHY obtains global coverage in its orbit). I aligned the data in temporal space by averaging measurements/simulation results over a month, and linearly averaging the values falling into each $5.625°$ grid box over the whole month.

Following Roberts et al. (2013), I performed PCA on SCIAMACHY data from the months of January, April, July and October 2004 separately. Additionally, I used annually-averaged data from CCSM3 because of the significant computational cost of going to sub-year time resolution on the simulated data, and the relative lack of available computing resources at my disposal.

Briefly, PCA involves calculating the covariance matrix from the mean-centered trace gas column densities. Using a spectral decomposition technique, the eigenvalues and eigenvectors of the covariance are determined in order to satisfy a characteristic equation; each eigenvalue is the variacne explained by each eigenvector. The PCs provide insight into which physical variables are captured by the atmospheric data.

In my analysis results, the first several PCs from both observed and simulated data match closely in terms of variance fraction captured (Fig. 3, right), as was the case also for Roberts et al. (2013). However, unlike their study, I see significant divergence beyond the first 5-6 PCs, implying that the trace gas datasets—as opposed to the full reflectance spectra—fall off below the signal/noise boundary somewhat more quickly due to the lesser amount of information/dimensions contained. For 10 PC dimensions, I capture 98% of the variability in SCIAMACHY trace gas components and 82% within CCSM3 atmospheric indicators.

Qualitatively, there is noticeable similarity between select PCs of Roberts et al. (2013) and my study. For example, the most significant PC I found is a common indicator within atmospheric studies, indicative of vegetation. We can see this from the spatial distribution of the PC (. 4), which is high over the Amazon, sub-Saharan Africa, and Southern/South-east Asia. Moreover, negative scores are seen over areas typically devoid of green vegetation such as the oceans, polar regions, and semi-arid regions. The recovery of this important feature is one sanity check that

[1] https://earth.esa.int/eogateway/instruments/sciamachy

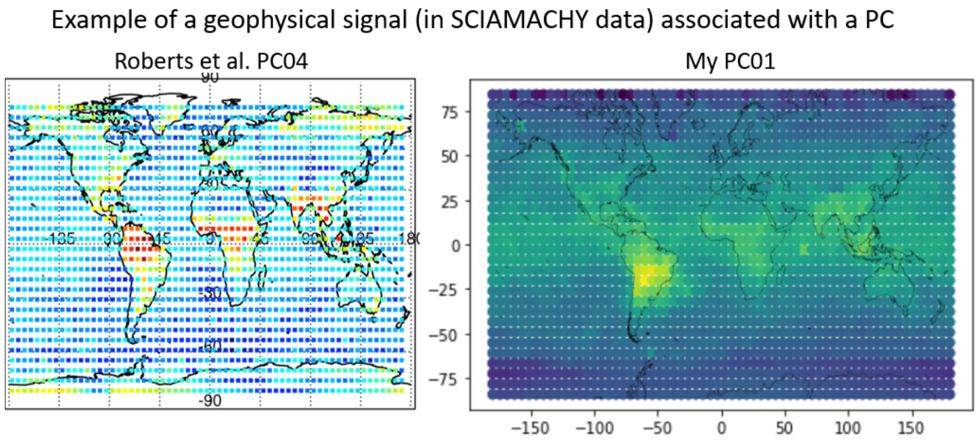

Figure 4

the data gathering, regridding, and analysis steps I have taken do not excessively dilute key physical information.

2.3 Quantitative subspaces comparison

At first-order, we can note the qualitative similarity between PCs of simulated and observed data by their geographic distribution. Beyond this, however, a quantitative comparison of similarity provides a more robust method of comparison to check the amount of shared information between the two datasets. As the goal of my study is to evaluate one dataset based on its similarity and relationship to another, I employ spectral decomposition analysis techniques similar to what authors in the field have done in the past.

I used an approach of intersection decomposition drawing heavily from the techniques described in Roberts et al. (2013) and Krzanowski (1979). First, we calculate principal components as described in Sec. 3.2. Then, using the eigenvectors of the PCA we calculate the intersection matrix I of the simulated-data and observed-data PCA spaces:

$$\mathbf{I} = \mathbf{E}_{\text{obs}}\mathbf{E}_{\text{sim}}^T\mathbf{E}_{\text{sim}}\mathbf{E}_{\text{obs}}^T$$

where $\mathbf{E_{obs}}$ and $\mathbf{E_{sim}}$ represent the k-eigenvector matrices from the observed and simulated PCs, respectively. $\mathbf{I}$ will be a $k \times k$ matrix where k is the dimension of the PCA decomposition. Singular Value Decomposition (SVD) determines eigenvalues, Γ, and eigenvectors, $\mathbf{Y}$, of $\mathbf{I}$:

$$\mathbf{I} = \mathbf{Y}\Gamma\mathbf{Y}^T$$

The spectral decomposition provides information to understand the amount of “shared" variance between the two PC subspaces. $\mathbf{Y}$ is used to determine the transformed eigenvector matricese for each data set in the shared, intersecting space:

$$\mathbf{A} = \mathbf{E}_{\text{obs}}^T\mathbf{Y}$$

$$\mathbf{A} = \mathbf{E}_{\text{sim}}^T\mathbf{E}_{\text{sim}}\mathbf{Y}$$

The vectors of eigenvalues $\boldsymbol{\gamma}$ can be used to construct a similarity measure between the PC subspaces. If the sum of the eigenvalues in $\boldsymbol{\gamma}$, equivalently the trace of $\mathbf{I}$, is equal to the number of PC dimensions, k, then the two data sets share minimal information. Conversely, a lower eigenvalue sum corresponds to a greater shared information.

Then, following the convention of Crone and Crosby (1995) to determine whether two subspaces are close in a statistically significant sense (i.e. their distance is sufficiently minimal), we use the following “subspace distance" metric:

$$D(\text{obs}, \text{sim})_k = \sqrt{k - \sum_{1}^{k} \boldsymbol{\gamma}_k}$$

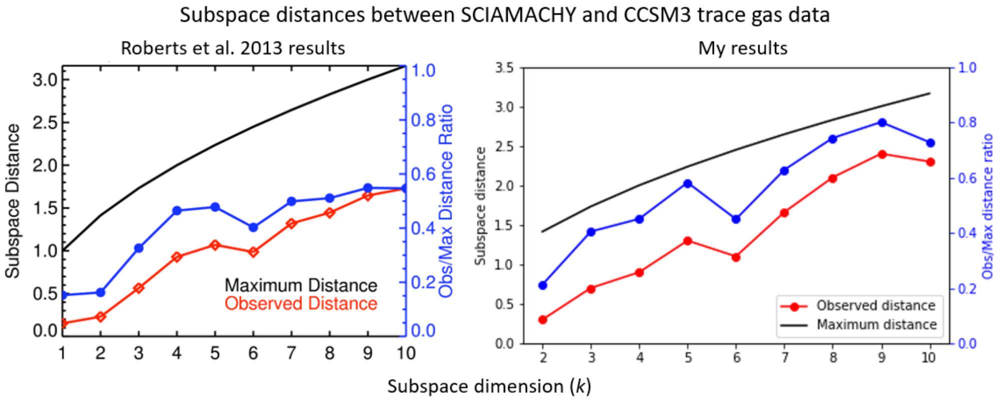

Figure 5

Determining if two subspaces are equivalent is not necessarily equivalent to concluding the individual PCs or covariance matrices calculated by the PCA are the same: instead, the distance metric captures the similarity between the subspaces spanned by the sets of k PCs.

When using this distance measure to calculate similarity of the PCs between observed and simulated reflected spectra, Roberts et al. (2013) found the ratio of actual distance to maximum possible distance (i.e. no shared information) ran from ~ 0.1 for a subspace of dimension of $k = 1$, to ~ 0.55 for a subspace dimension of $k = 10$, roughly increasing linearly for the intervening k. While the relatively large distance for higher PC dimension can likely be attributed to instrumental measurement noise, as well as numerical error and fundamental limitations of the simulation forecasting, the agreement between the two data sets for low PC dimension is remarkably good, suggesting that simulations are able to capture the most important/significant spatial variability patterns.

My results (Fig. 5) are similar in trend (observed:maximum distance ratio running from 0.2 for $k = 2$ to 0.7 for $k = 10$) but somewhat worse than those of Roberts et al. (2013). The decrease in quality from their study is likely to be mainly attributed to the decreased information being considered by my analysis: as discussed, there is nontrivial information (e.g. albedo, land composition) encoded in atmospheric spectra which is *not* captured by the column densities produced with DOAS. However, considering the order-of-magnitude decrease in computational storage space/dataset size in our approaches, these results are not necessarily negative. Further interesting work could lie in identifying whether these spatial variance measurements show similar subspace distance trends across further years in the simulation forecasting, as well as analyzing variance trends in temporal space.

3 Conclusion and future work

In this study, I have used trace gas column densities derived from SCIAMACHY shortwave spectral reflectance measurements and compared them to similar data from state-of-the-art global climate models as a way to check how well simulations are able to reproduce the real data collected by atmospheric spectrographs. I followed closely with a similar study, Roberts et al. (2013), while modifying the data approach slightly to check that the comparison between real and simulated data is robust across various levels of data product. My results were fair, with a quantitative measure for the subspace distance yielding results that imply 20-40% shared information for low (2-5) PCs, but quickly fall off for higher subspace dimension. However, some of the most important physical indicators, such as vegetation abundance and ice cover, were captured qualitatively by my analysis. More fine-grained structure was likely lost, however; possible contaminant factors include insufficient temporal resolution of the simulated data used, which was a bottleneck due to the availability of computing resources at my disposal.

This preliminary study opens itself up to many interesting future directions. For one, I used only a small fraction of the spatial resolution and temporal coverage of SCIAMACHY and CCSM

3.0 data, and compared only spatial variability; expanding to higher resolution, and analyzing trends of temporal variability, would both be a natural augmentation of data size/quality to test for further connections.

A potentially more original avenue of research, however, would be to more closely examine explanatory variables for the spatial PC trends we observe in this study and others. Limited work has been done thus far in training supervised machine learning models on various selections of spatially resolved, independent variables (both natural *and* anthropogenic in nature) to test which factors may account most for localized trends in atmospheric trace gas/aerosol distribution. Such knowledge could directly inform lifestyle and policy making, providing a critical link from the most sophisticated satellite data directly to the human level.

References

Burrows, J.P., Weber, M., Buchwitz, M., Rozanov, V., Ladstatter-Weissenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K., and Eisinger, M.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, Journal of the Atmospheric Sciences, 151-175, 1999.

Crone, L. and Crosby, D.: Statistical applications of a metric on subspaces to satellite meteorology, Technometrics, 37, 324–328, 1995.

European Space Agency: Envisat SCIAMACHY Product Handbook, 2013. earth.esa.int

Feldman, D. R., Algieri, C. A., Ong, J. R., and Collins, W. D.: CLARREO shortwave observing system simulation experiments of the twenty-first century: simulator design and implementation, J. Geophys. Res., 116, D10107, 2011.

Gottwald, M., Hoogeveen, R., Chlebek, C., Bovensmann, H., Carpay, J., Lichtenberg, G., Krieg, E., Lutzow-Wentzky, P., and Watts, T.: SCIAMACHY – Exploring the Earth’s Changing Atmosphere, Springer, New York, chap. 3, 29–46, 2011.

Krzanowski, W.: Between-groups comparison of principal components, J. Am. Stat. Assoc., 74, 703–707, 1979.

Loeb, N., Wielicki, B., Su, W., Loukachine, K., Sun, W., Wong, T., Priestley, K., Matthews, G., Miller, W., and Davies, R.: Multi-instrument comparison of top-of-atmosphere reflected solar radiation, J. Climate, 20, 575–591, 2007.

Roberts, Y.L, Pilewski, P., Kindel, B.C., Feldman, D.R., and Collins, W.D.: Quantitative comparison of the variability in observed and simulated shortwave reflectance, Atmos. Chem. Phys., 13, 3133-3147, 2013.