Response to Reviewer #1

Dear Prof. Huang,

Thank you for taking the time to review this manuscript. I believe your suggestions have helped to improve this document. I have responded to each of your comments below. Your comments are italicized, our responses are in normal text, and the edits that will be made in the manuscript text are in bold.

1. Line 12 on page 28306: "PCA is a multivariate spectral decomposition technique". "spectral" should be removed here. PCA is a multivariate decomposition technique, not replying on spectral decomposition such as Fourier transform and not being limited to study spectral radiances alone.

In the description of PCA, we will remove the word spectral.

2. The first paragraph of Section 1. This paragraph started with net radiative forcing. Note the radiative forcing defined in the IPCC is not the same as what described here: it is not merely the imbalance at the top of tropopause; the stratosphere must be adjusted back to equilibrium in such calculation of radiative forcing. Then when it comes to "Hansen et al. (2011) estimated the global radiative imbalance between 2005 and 2010 to be 0.58±0.15Wm−2...", it talks about the radiative imbalance at the top of atmosphere, not at tropopause. Moreover, the radiative forcing and radiative imbalance discussed in this paragraph has only remote connection with the theme of this study. I feel this paragraph is not that well delivered and not that related to the topic of this study. Maybe a rewrite would be better, or simply remove it. Actually even the section starts with the second paragraph, it seems to me totally fine.

We have decided to start the manuscript from the third paragraph (Page 28307, Line 22), where we start a discussion of the importance of reflected shortwave radiation, a more relevant introduction to the main topic of this paper.

3. Line 14 on page 28312, Anderson et al. 1999 is a reference for MODTRAN4 not MODTRAN5.3. MODTRAN5 has significant changes from the previous version. Appropriate references for MODTRAN5 can be found at http://modtran5.com/faqs/index.html#what_ref

The MODTRAN5 reference will be changed to [Berk et al., 2006]. This reference was obtained from the above MODTRAN5 website.

Berk, A., G.P. Anderson, P.K. Acharya, L.S. Bernstein, L. Muratov, J. Lee, M. Fox, S.M. Adler-Golden, J.H. Chetwynd, M.L. Hoke, R.B Lockwood, J.A. Gardner, T.W. Cooley, C.C. Borel, P.E. Lewis and E.P. Shettle, "MODTRAN5: 2006 Update," Proc. SPIE, Vol. 6233, 62331F, 2006.

4. Section 3.1.3, "Boundary between data signal and noise". The authors nicely summarized different methods for deciding the truncation of PCs. The orthogonal nature of PCA technique decides that, regardless which PC components to be examined, it always contains some noise components as well as some signals. Only difference is the percentage of signals in that particular PC. In another word, such global decomposition methods such as PCA and SVD cannot really definite a boundary between data signal and noise. The best they can do is to find a truncation that retains most signals (inevitably some noises, even small, will be retained as well). For the benefit of readers, it would be useful to spend one or two sentences to point out this caveat.

Yes, this is a good point. Please see the bolded text in the first paragraph below for the edit we will make in response to this comment.

5. Still in Section 3.1.3, Jerry North has a seminal paper (North et al., 1982, Month Weather Rev.) that gave a rule of thumb for the sampling errors in the estimation of empirical orthogonal functions. The rule can be used to decide the truncation as well. It has been widely used in the climate community. Might worth to include this method here as well.

To address comments #4 and #5, we plan to insert the following bold text in Section 3.1.3

3.1.3 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). **There is no clear, quantitative boundary between signal and noise. To maximize the variance explained by each principal component, information from both the signal and noise are included in each eigenmode. Unless the noise variance exceeds the signal variance, the signal in the data typically dominates the variance explained by the first few eigenmodes. The best estimate of the boundary between signal and noise is the dimension at which noise begins to consistently dominate the variance, which may be difficult to determine at times because the noise and signal variances are not always strictly anticorrelated with increasing dimension number. This can make it difficult to assess if a particular dimension is dominated by signal or by noise. There are methods that can be used to more clearly separate the signal and noise in a data set, such as the Minimum Noise Fraction Transform (Green et al., 1988), which is discussed briefly in Sect. 5.**

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), $\omega_k > \frac{1}{K}\sum_{j=k}^{K}\frac{1}{j}\omega_{\text{avg}}$ is the criterion for determining the location of the boundary.

Although the PCs are mathematically independent, it is still possible for the sampling distribution of each PC to be related to either of its neighboring PCs. One test of this separation is called the *North et al. Rule of Thumb* (North et al., 1982). This rule uses eigenvalue confidence intervals to determine if neighboring PCs are statistically separate from each other. If the neighboring eigenvalues fall outside the confidence interval of a given PC, then that PC is statistically different those on either side. The 95% confidence intervals for the eigenmodes can be calculated using $\Delta\omega = z(0.975) * \sqrt{\frac{2}{N}\hat{\omega}^2}$ **, where *N* is the number of observations (number of spectra in this study). This is another method that can be used to determine the approximate boundary between signal and noise; the boundary would be located where the eigenmodes are no longer statistically separated.**

The subjectivity of locating **the signal-noise** boundary is recognized by all the studies discussed here, but these suggested techniques often help to provide guidance in making this decision. 25 The information provided by these selection criteria help us to make sense of the sub- space 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 described in Sect. 3.2.4 into context as we will discuss in Sect. 4.