Application of PCA to Hyperspectral Data for Noise Characterization, Noise Filtering and Instrument Monitoring

Paolo Antonelli, Dave Tobin, Hank Revercomb, Dave Turner, Steve Dutcher, Ben Howell, Bob Knuteson, Ken Vinson and William L. Smith

Space Science and Engineering Center, University of Wisconsin-Madison,
1225 West Dayton St., Madison, WI 53706, USA
paoloa@ssec.wisc.edu

Abstract. Exploiting the inherent redundancy in hyperspectral observations, Principle Component Analysis (PCA) is a simple yet very powerful tool not only for noise filtering and lossy compression, but also for the characterization of sensor noise and other variable artifacts using Earth scene data. After Antonelli et al. [1, 3], showed a substantial improvement of the signal to noise ratio can be achieved by taking advantage of the redundancy present in the spectral domain through the application of PCA, Tobin et al. [18] showed how PCA can be used to characterize the instrument noise and to monitor the instrument performances. This article is a compendium of the results described in those papers.

Keywords: Hyperspectral, Satellite, PCA, Noise, NEDT, Characterization.

1. Introduction

Principle Component Analysis (PCA) applied to hyperspectral satellite spectra is a useful tool for filtering random noise and to perform data compression [1-5]. PCA produces filtered radiance spectra that have a large majority of the random noise removed. However if, instead of retaining the low noise filtered spectra, we retain the reconstruction error it is possible to perform various investigations to diagnose the sensor noise. Proper characterization of sensor noise performance is essential for algorithm development and data usage. Inversion of infrared radiance observations to produce vertically resolved atmospheric soundings, in particular, is critically dependent on good noise performance and on the ability to properly characterize the sensor noise.

To improve retrieval accuracy, one can reduce noise by averaging the measurements over a number of adjacent Instantaneous Fields of View (I-FOV). This procedure, however, improves the signal to noise ratio in the spectral domain at the expense of the instrument spatial resolution, with significant degradation of the signal for non uniform scenes (for example in proximity to cloud edges or coastal regions).

Antonelli [1], and Huang and Antonelli [2], while investigating the use of Principall Component Analysis (PCA) to compress high spectral resolution IR data, showed that a substantial improvement of the signal to noise ratio can be achieved through PCA. However because PCA is essentially a lossy compression procedure, it is important to address the issue of information loss due to its application. In fact traditionally the lack of knowledge about the nature and magnitude of this information loss has been the main source of skepticism when using PCA. In particular, in the first part of this paper, we address this issue by investigating in detail the impact of PCA on simulated and real data from both a statistical point of view and on a case-by-case basis.

The investigations presented in the second part of the paper focus on the noise characterization obtained applying PCA to Atmospheric Infrared Sounder (AIRS) data. AIRS is the first of the new advanced hyperspectral infrared sounders designed to produce observations for improvements in medium range weather forecasting and climate applications. AIRS is a grating spectrometer with resolving power of ~1200. It utilizes 17 detector arrays to yield 2378 spectral channels covering the 3.7 to 15.4 μm region with good

noise performance. Aumann et al. [6] present a thorough overview of the science objectives, data products, retrieval algorithms, and ground-data processing concepts of AIRS.

2. Theoretical basis of PCA

PCA is a multivariate analysis technique that was first introduced by Pearson [15] and developed independently by Hotelling [16] at the beginning of the 20th century. It is commonly used to reduce the dimensionality of observations characterized by a large number, N (spectral channels), of interdependent variables. The reduction is achieved by projecting the deviation of the hyperspecrral IR observations from their mean onto the set of orthogonal vectors, spanning a subspace of the original observation space, which accounts for most of the input data variance. Hence the problem of dimensionality reduction is reduced to finding a linear transformation that maps the observations into a lower dimensional array of compressed (or reduced) observations with minimal loss of atmospheric information. Such transformation is obtained when the rows of the linear transformation are the eigenvectors, also referred to as Principal Components (PCs), of the observation covariance matrix. The first PC is the direction along which the variance of the input data has its maximum. The second PC is the vector in the orthogonal dimensional subspace complementary to the first PC, which explains most of the remaining variance and so on until the last PC is the direction of minimum variance. It can be easily shown that if all the PCs are retained, PCA linearly decorrelates the output variables, i.e., diagonalizes their covariance matrix. Hence PCA essentially performs a Singular Value Decomposition (SVD) of the observation covariance matrix.

3. Noise filtering properties of PCA

The value of PCA processing of high spectral resolution infrared data is its ability to take advantage of spectral redundancy to discriminate between the atmospheric signal and the random uncorrelated gaussian distributed component of the noise. To explain how PCA reduces the noise, let us assume that we have access to an ideal instrument characterized only by a random, spectrally uncorrelated noise of constant variance across the spectral region of interest. Under these assumptions the noise is homogeneously distributed in a N-dimensional sphere and its variance along the basis vector is independent of the basis selection. This implies that, after projecting the input data onto the PC space, if only a small number of PCs are retained to accurately represent the matrix of atmospheric (or noise-free) signals then the noise variance of the reconstructed spectra is significantly reduced by ratio of the number of channels divided the number of retained PCs

It is worth mentioning that for high spectral resolution instruments the uncorrelated component of the instrument noise is not uniform in the wavenumber space. Spectral regions of especially high noise can cause PCA to selectively fit the noise in these regions. To avoid fitting the noise instead of signal variations, the spectra are normalized to make the noise approximatively white, before applying PCA. This not only improves the ability of PCA to enhance the signal to noise ratio consistently throughout the spectrum, but it also facilitates a quantitative estimation of the noise reduction. After normalization the noise becomes uniform in the spectral domain, with the noise variance equal to one for each channel. Under these circumstances, by mapping an observation dimensional space onto a lower dimensional space, through PCA causes a reduction in the sum of the noise variances from the pre-compression value. This allows for a noise reduction factor (in the sense of total noise variance reduction) equal to the square root of the ratio between the original observation dimension and the reduced dimensionality of the subspace spanned by the retained PCs.

4. Applicaton to simulated S-HIS data

To provide an objective evaluation of the PNF performances, the rms of the atmospheric information loss (as defined by Antonelli et al. [3]) and the rms of the reconstructed noise, i.e. the noise embedded in the retained PCs, are compared their theoretical optimal values, in the framework of linear filters, provided by the Linear Estimation Theory. The two upper plots in fig. 1 respectively compare the reconstructed noise and the atmospheric information loss for the Minimum Mean Square Error algorithm (black) and for PCA (red). The plots have been obtained on S-HIS simulated data, using 15 PCs. The bottom plot compares the rms of the reconstructed noise (red circles) and the rms of the atmospheric information loss (black diamonds) to the MMSE estimates (red dashed and black solid lines). This figure shows that for the optimal number of PCs, the accuracy of PNF approaches the theoretical limit achievable by linear filters. Most importantly it shows that the noise that leaks through linear filters is in both cases correlated in the spectral domain and that this correlation is not simply an artifact introduced PCA. The correlation of this noise is clearly associated with the spectral distribution of the atmospheric information: absorption lines tend to carry more information than window channels. The higher the information content for a single channel, the smaller the angle(s) between the channel axis and the one or more of the retained PCs, and consequently the higher the noise that leaks through the filter for that channel. It is worth also emphasizing that the MMSE filter, is, in general, not applicable on real data as the covariance matrices of the noise free signal and of the instrument noise are not known.

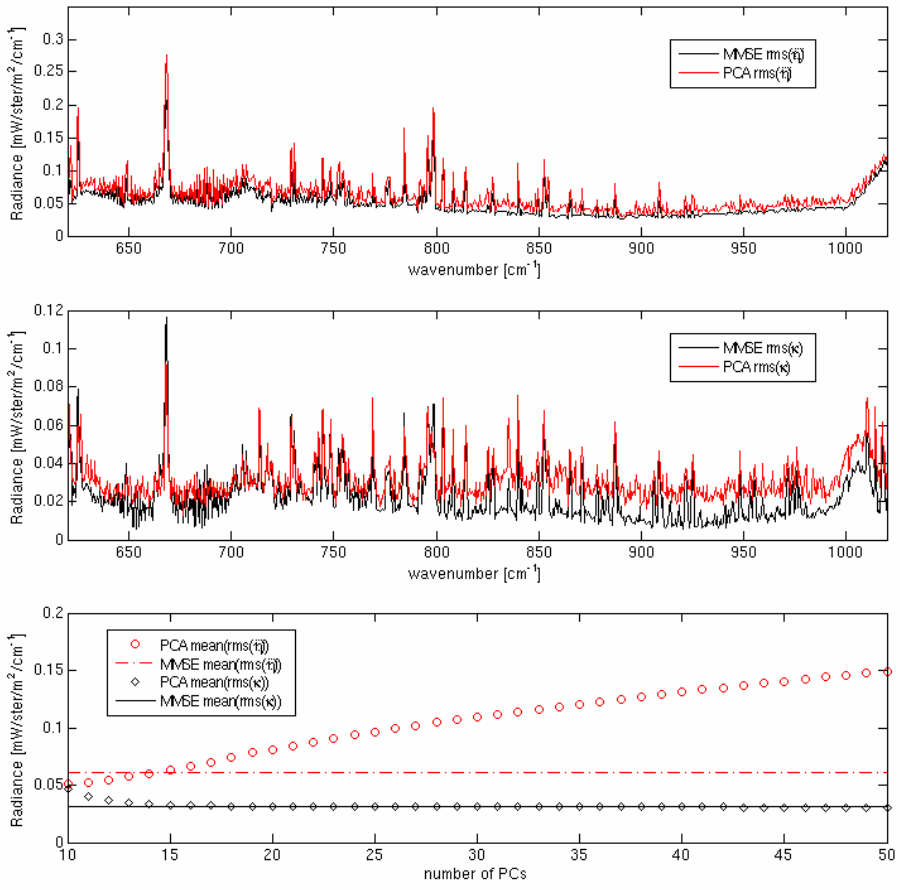

Figure 1: S-HIS simulated LW: (top) reconstruction residuals for the MMSE (black) and for PNF (red); (middle) comparison between atmospheric information loss for MMSE (black) and PNF (red). The plots have been obtained using 15 PCs. (bottom) PCA spectral mean of the reconstruction residuals (red circles) and of the atmospheric noise (black diamonds) compared to the MMSE estimates (red dashed and black solid line). For an optimal number of PCs, PCA approaches the theoretical limit achievable by linear filters.

5. Application to real S-HIS data

This section describes the application of PCA to real observations collected S-HIS in the SAFARI field experiment. S-HIS is an FTS instrument based on a Michelson flat mirror design with dynamic alignment and uses an internal laser (Helium-Neon) to sample the interferogram signal according to optical path delay. It uses numerical filtering to reduce the data volume in each of three spectral bands spanning the range of about 16.7 to 3.3 microns .The data set used for this case consists of about 15,000 spectra observed by S-HIS on 7 September 2000 during the SAFARI mission. For the selected flight the instrument operated in nadir mode only and the observations were collected over land and over ocean, in both clear and cloudy (low clouds only) conditions. In addition some observations were collected over ground fires.

To render visually the effect of PCA on real observations we show (fig. 2) 60 spectra collected over the Indian Ocean (from 10:03:00 to 10:05:00 UTC). The top plot shows the observations before filtering while the bottom plot shows the same observations after the application of PCA (retaining 30 PCs). The selected FOVs are very homogeneous, under clear sky condition, and with minimal surface and atmospheric variations. The bottom plot shows how PCA reduces the variability of the observations due to the random component of the noise.

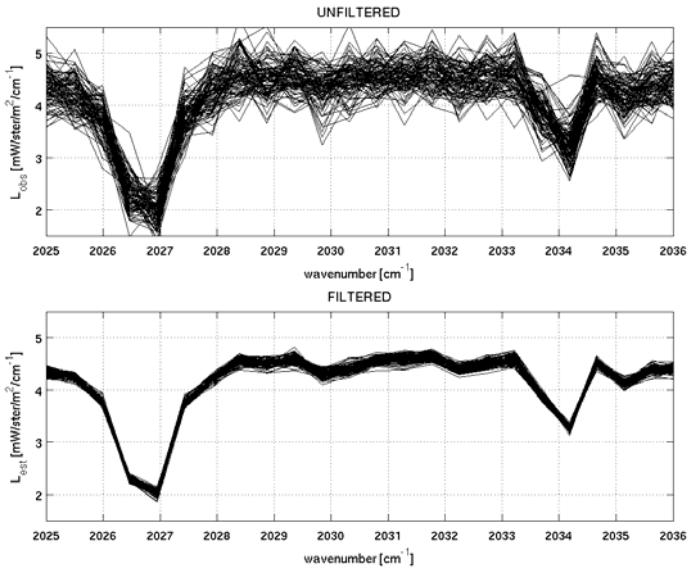

Figure 2: S-HIS observed shortvawe: clear sky observations over the ocean. (top) The unfiltered radiances (60 spectra observed between 10:03:00 and 10:05:00 UTC on September 7th, 2000) in the SW band observed over the ocean in clear sky condition. The lower plot on the left shows the same spectra after filtering with 30 PCs. The uncorrelated component of the noise is reduced by PCA.

The observations highly deviant from the mean and/or rarely occurring, commonly classified as outliers, often represent the most interesting cases. We verified that these observations, obtained from different instruments, and for a variety of special situations, are not irreversibly compromised by the lossy compression used by PCA.

For this study PCA was applied to observations collected in proximity to and over a fire. The distinctive feature under consideration is the Blue Spike [17] in the SW region at about 2397 cm^{-1} (left image in fig. 3); the observations selected are not well represented in dataset used to derive the principal components: among the 15,000 training spectra only about 80 show the blue spike feature, and these 80 spectra are relatively far from the mean training spectrum. The blue spike is characteristic of FOVs partially obstructed by hot (>350 K) atmospheric gases heated by the fires, above colder surface temperature. The results shown in figures 3a and 3b indicate that the reconstruction residuals for the FOV over the fire (8:43:20 UTC) are more sensitive

to the number of retained PCs. As expected, spectra labeled as outliers require more PCs than well represented spectra to be properly reconstructed. However, providing that the distinctive spectral features are represented in the PCs, they are properly filtered. The plots in figures 4 show unfiltered (left) and filtered (right) radiances observed over the fire between 8:43:12 and 8:43:22 UTC. The filter, not only preserves the information about spectral lines due to CH_4 and other gases, but it also makes it easier to identify them. This effect is even more evident in figure 5, which shows 50 spectra observed immediately after the fire overpass (between 8:43:26 and 8:44:03 UTC). These observations are characterized by blue spikes of smaller magnitude not easily detectable in the noisy observations (left plot) but clearly identifiable in the filtered spectra (right plot).

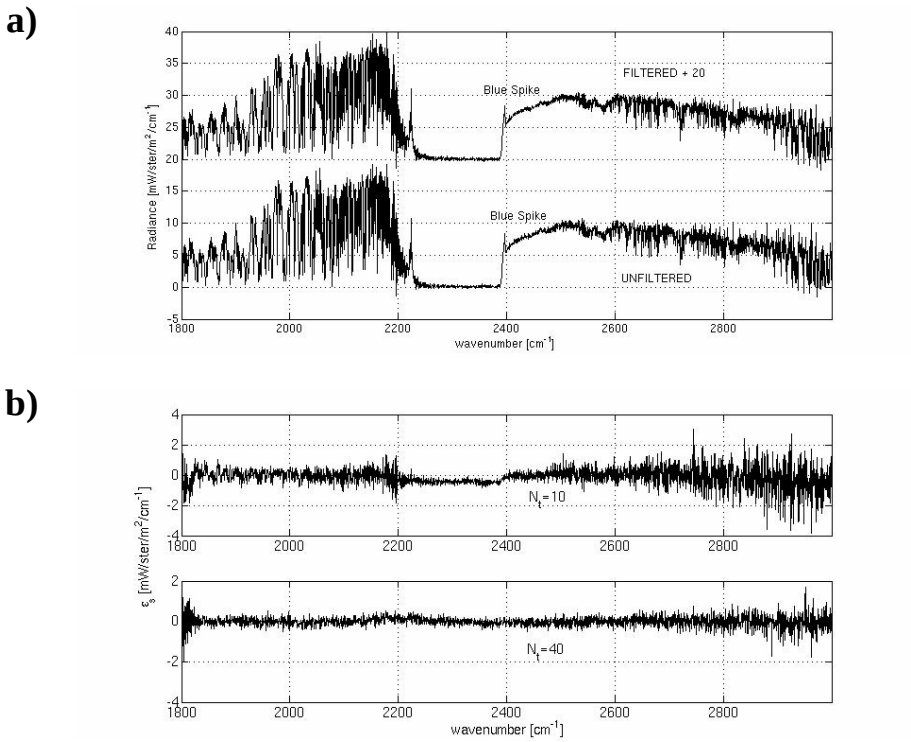

Figure 3: S-HIS observed SW: (a) shows the SW unfiltered and filtered radiances observed at 8:40:20 UTC. (b) shows the reconstruction residual for the same observation for 10 and 45 PCs. The results indicate that, in presence of outliers, the reconstruction residuals are very sensitive to the number of PCs and the optimal number of PCs, statistically determined, might introduce large correlated errors.

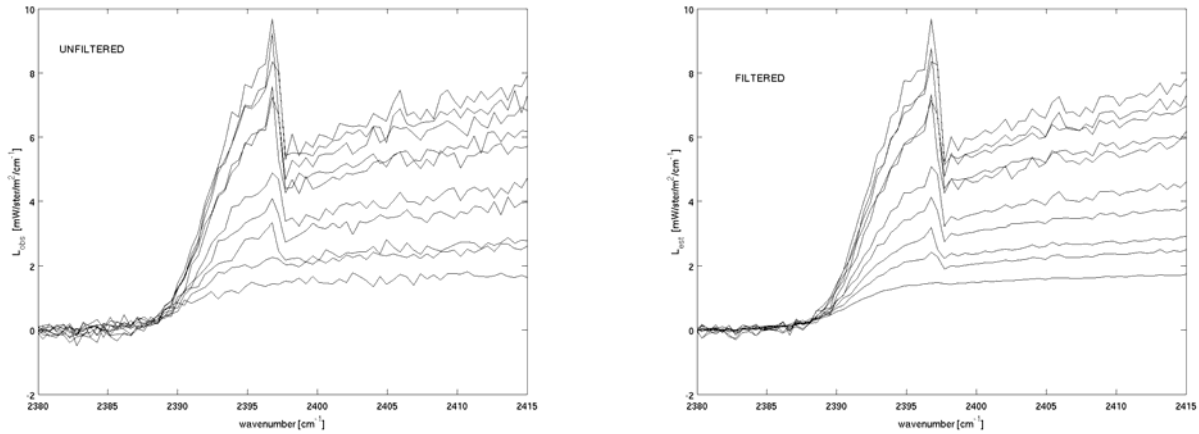

Figure 4: S-HIS observed SW: The plots show a series of 9 unfiltered (left) and filtered (right) spectra collected consecutively over fires on 7 September 2000 at about 8:43:20 UTC. The blue spike is not affected by the noise filter and, more important, some of the absorption lines (CH4,NH3,OH, etc ...) between 2400 and 2410 cm-1 are actually more evident in the filtered data than in the unfiltered ones.

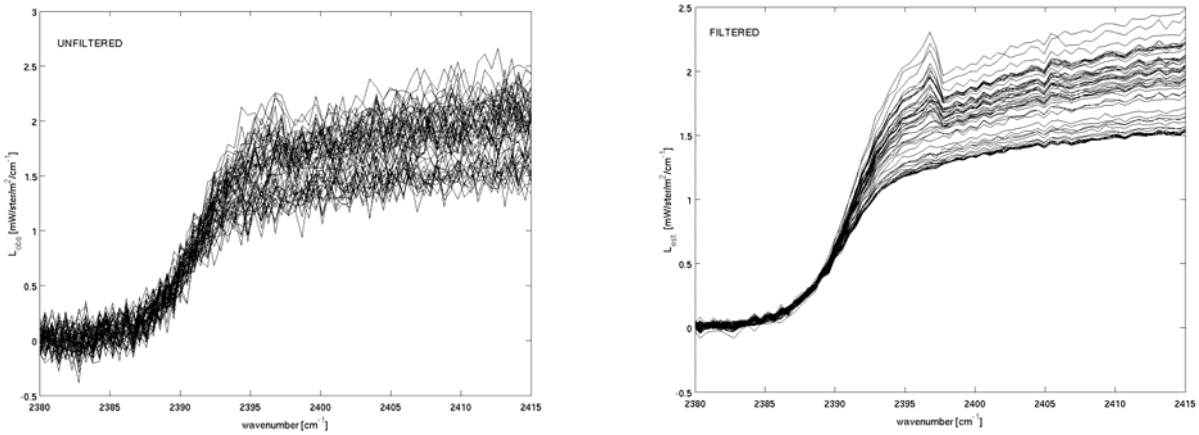

Figure 5: S-HIS observed SW: The plots show a series of 50 unfiltered (left) and filtered (right) spectra collected consecutively immediately after the fires (after 8:43:25 UTC). The blue spike signal is about 20 time smaller than the peak value, and is not clearly identifiable in the unfiltered data, but it becomes visible in the filtered data along with some of the absorption lines (CH_4,NH_3,OH, etc ...) between 2400 and 2410 cm-^{1}.

The relevance of PCA relies on the importance of reducing the random component of the instrument noise to allow for more accurate retrievals of the atmospheric parameters and for more accurate validations of radiative transfer software. However PCA can also be used to estimate instrument noise and monitoring on instruments performances as showed in the following sections.

6. PCA applied to AIRS data

This paragraph describes details of the approach particular to the application of dependent set PCA, as presented in Antonelli et al. [3], to AIRS data. The analysis is performed separately for each AIRS Level 1B granule, which is 6 minutes of data including 135 cross track scans with 90 footprints per scan (12150 footprints total), providing a relatively large ensemble of spectra compared to the number of spectral channels. AIRS contains spectral channels that are not recommended for use in radiometric analyses and these channels are not included in the analyses presented here. Channels are excluded following the AIRS team prescription [7] using quality flags contained within the Level 1B granule file and an ancillary channel properties file. Typically, roughly 2120 of the original 2378 channels are retained. Following [3], the PCA uses "noise normalization" where each original radiance value is divided by the nominal noise of each channel. (This can be performed using the noise values provided by the AIRS team, or with values derived from PCA without using the noise normalization). For an optimal method of retaining signal variance and removing noise, we use the prescription provided by Turner et al. [5] to determine the number of PCs used in the observation reconstructions.

Various analyses are performed to diagnose the sensor noise. The reconstructed spectra have a large majority of the spectrally random noise removed. Noise that is not spectrally random will generally be retained. For example, if L_{ORIG} and L_{RECON} are the original and reconstructed radiances, respectively, then the standard deviation of L_{ORIG} -L_{RECON} taken over the whole granule is an estimate of the spectrally uncorrelated random noise removed.

This is a robust method for determining sensor noise. This is demonstrated in Fig. 6, which shows the reconstruction error for sample channels plotted versus the number of PCs used in the reconstructions. There are three regimes observed in the figure. In regime 1, the signal variance is reconstructed and the reconstruction error decreases rapidly as the number of PCs increases. For very high numbers of PCs (regime 3), the full variance including real signal and random noise is reconstructed. In regime 2, most of the signal

has been reconstructed and yet most of the random noise is excluded, and the reconstruction error is relatively insensitive to the number of PCs. It is in this region where the reconstructions are performed.

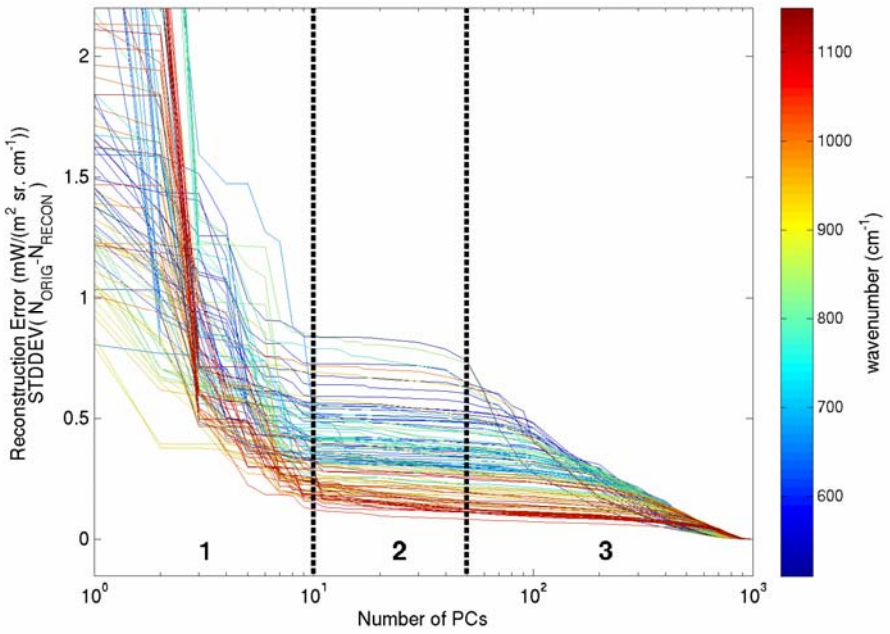

Figure 6: Reconstruction error versus the number of PCs used in the PCA reconstruction for sample longwave AIRS channels.

7. NEDT Investigations

Following the approach described above, PCA is performed on 7 ascending (daytime) granules collected on April 1, 2005. This is version 4.0.9 Level 1B AIRS radiance data. These data are shown in Fig. 7, which shows geographic maps of window region brightness temperatures and sample brightness temperature spectra. The PCA is performed separately on each granule. For granules 194 through 200, the numbers of PCs used in the reconstructions are 50, 53, 49, 52, 69, 61, and 51, which correlate strongly with the amount of scene variability observed in each granule.

An estimate of the spectrally uncorrelated random radiance noise is computed as NEDN = STDDEV(L_{ORIG}-L_{RECON}). A small correction factor (typically ~1.015) is also applied to the PCA NEDN values to account for the random noise included in the PCs retained in the reconstructions. Figure 8 shows the comparison of the PCA noise estimate for the AIRS longwave channels compared to values determined using on-board blackbody and space views provided in the ancillary channel properties file and in the Level 1B granule files. The comparisons are shown in terms of Noise Equivalent Brightness Temperature (NEDT) where the NEDN values have been converted to brightness temperature at a scene temperature of 250K. The figure shows relatively good agreement between the PCA and blackbody estimates. However, where they differ, the PCA estimates are fractionally lower than the blackbody estimates. This is related to the fact that the PCA estimate is of the spectrally uncorrelated random noise and the blackbody analysis represents the total noise, and is discussed further below.

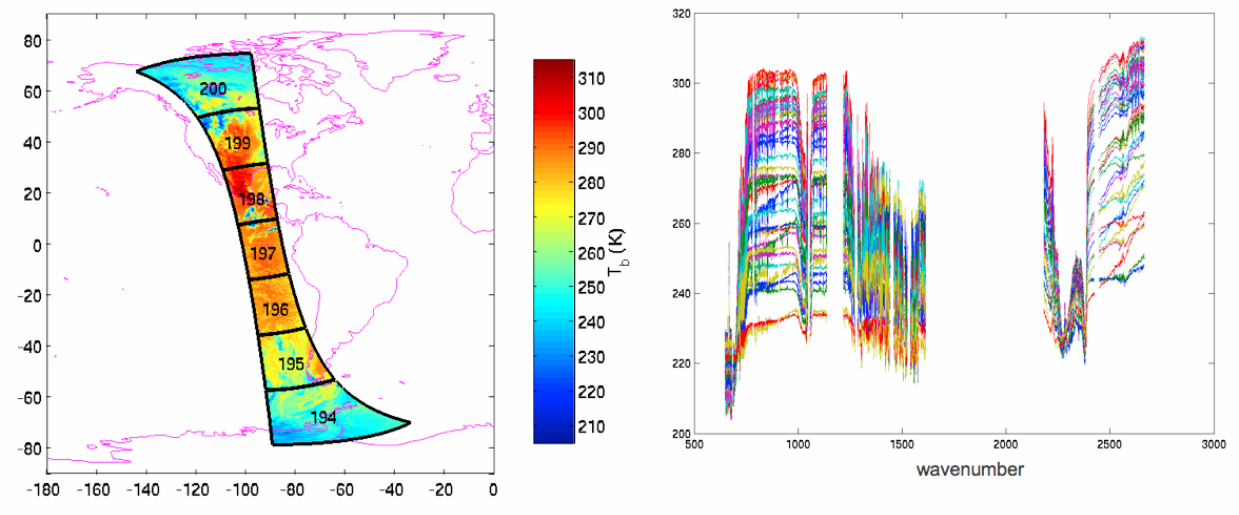

Figure 7: AIRS data from April 1, 2005 used in this analysis. Left: ~1000 cm-1 brightness temperature maps for each granule. Right: Sample brightness temperature spectra.

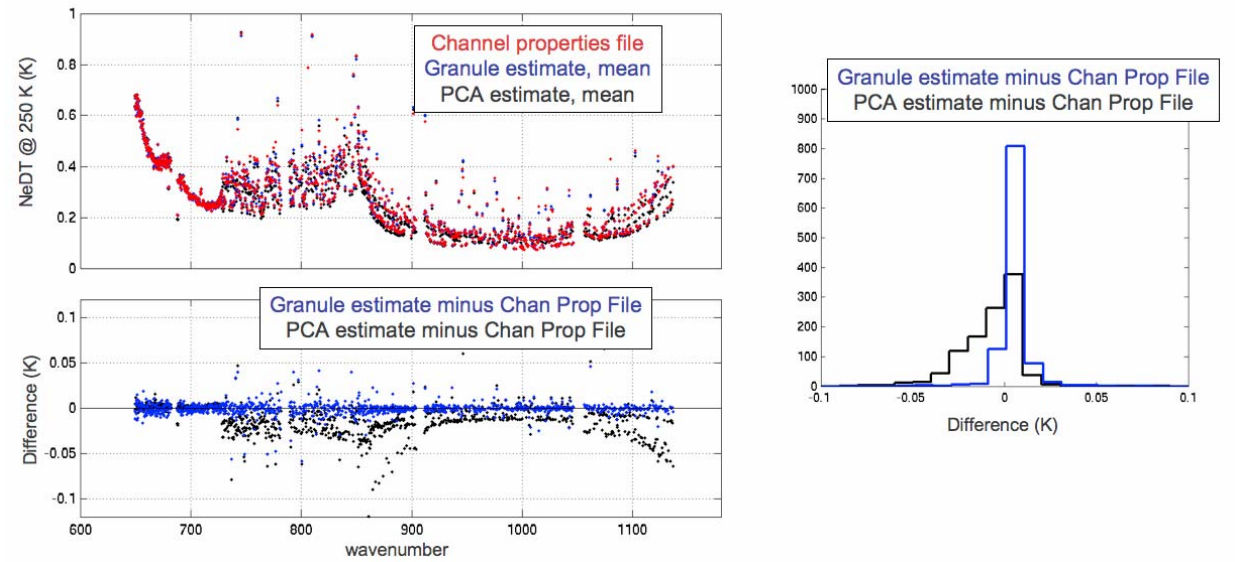

Figure 8: Comparison of PCA spectrally uncorrelated noise estimates for the AIRS longwave channels to total noise determined using on-board blackbody and space views provided in the ancillary channel properties file and in the Level 1B granule file.

Comparisons similar to Fig. 3 for the midwave and shortwave spectral regions do not show as good agreement. This is due to the signal dependence (i.e. photon limited) AIRS noise performance in these spectral regions. To account for this in the analysis, data from the 7 granules is divided into various scene level range ensembles, and the PCA is performed separately for each ensemble according to the shortwave window region radiance values. NEDN is then computed for each ensemble and the scene dependence of NEDN is examined as a function of scene radiance. The results are shown in Fig. 9, where $NEDN^2$ is plotted versus scene radiance for selected channels. As seen in the figure, the shortwave channel noise has a significant linear dependence on scene radiance, as expected. The midvave channels have a weak dependence on scene radiance and the longwave channels are independent of scene radiance. Based on this study, a simple parameterization of NEDN, which includes thermal noise and photon limited noise components, has been developed in terms of scene radiance. This parameterization produces results that are in good agreement with the parameterization based on Level-0 counts [8-9]. The scene dependence of the

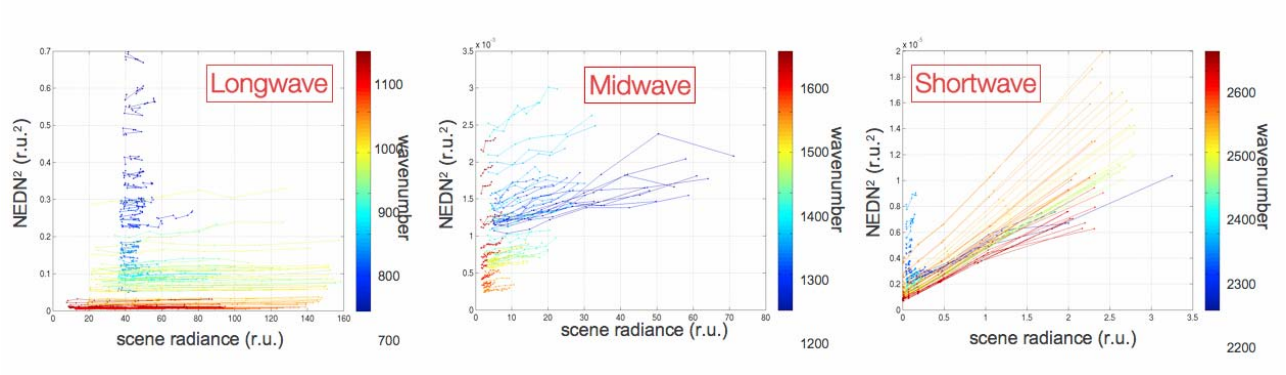

Figure 9: Dependence of NEDN^2 on scene radiance for sample channels in the longwave, midwave, and shortwave spectral regions. Radiance units (r.u.) are mW/(m^2 sr. cm^{-1}).

AIRS noise is significant for shortwave channels; NEDN for a 300 K scene is roughly two times larger than that for a 250K scene. The AIRS noise is typically reported for a 250 K scene. Incorporating this information regarding the scene dependence of NEDN, the comparison of PCA noise estimates with on-board blackbody/space view estimates is shown in Fig. 10. The figure shows good agreement between the three noise estimates. As before, however, where the PCA estimates differ, the PCA estimates are fractionally lower than the blackbody estimates, which are related to the fact that the PCA estimate does not include spectrally correlated noise components.

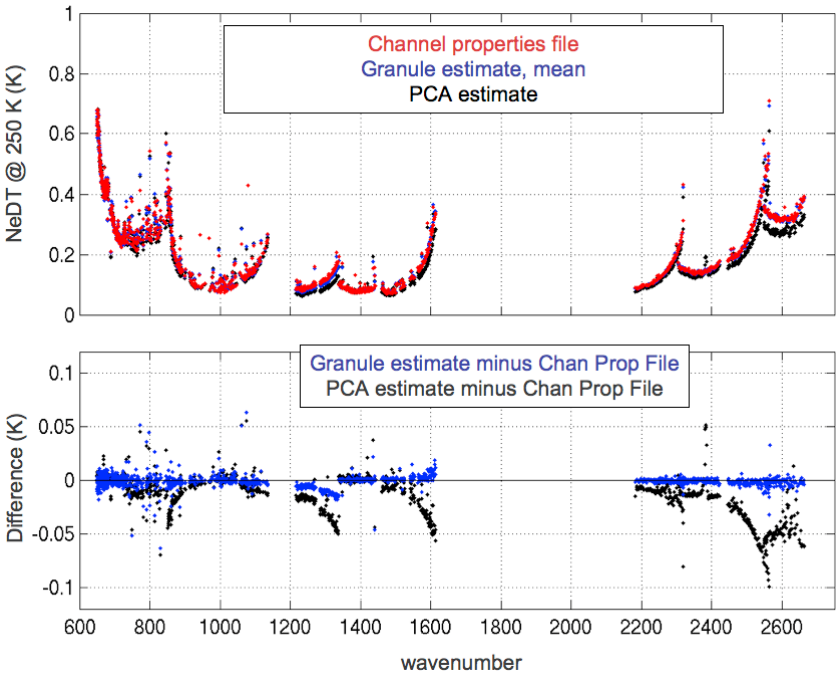

Figure 10: Comparison of PCA spectrally uncorrelated noise estimates to total noise values determined using on-board blackbody and space views provided in the ancillary channel properties file and in the Level 1B granule files.

Spectrally correlated noise (i.e. noise that is not random from one spectral channel to another) is not included in the PCA noise estimates because the PCs which represent the spectral correlation generally account for a larger fraction of the total ensemble variability, are therefore included in the low order PCs, and are therefore included in the reconstructed radiances. Since the on-board blackbody/space view noise estimates provided by the AIRS project are of the total noise, we can therefore deduce the amount of spectrally correlated noise.

The spectrally correlated noise for AIRS is attributed to detector array readouts and is therefore, to a good approximation, independent of single channel detector noise. We therefore compute the total noise as the Root Sum Square of the spectrally correlated noise and the spectrally uncorrelated noise. Figure 11 shows the PCA estimate of the spectrally correlated noise for AIRS (using the total noise from the blackbody/space views and the spectrally uncorrelated noise from the PCA estimate) compared to the spectrally correlated noise determined for AIRS in the pre-launch phase [10-11]. The pre-launch estimates are detector array means derived from the covariance of ensembles of blackbody views. The comparison shows very good agreement between two very different and independent analyses, providing good confidence in the PCA approach. The results also show that the detector array correlated noise for AIRS can be a significant fraction of the total noise, particularly for the M-04 arrays in the water vapor sounding regions. Note that the correlated noise for the longwave photo-conductive arrays M-11 and M-12 is small.

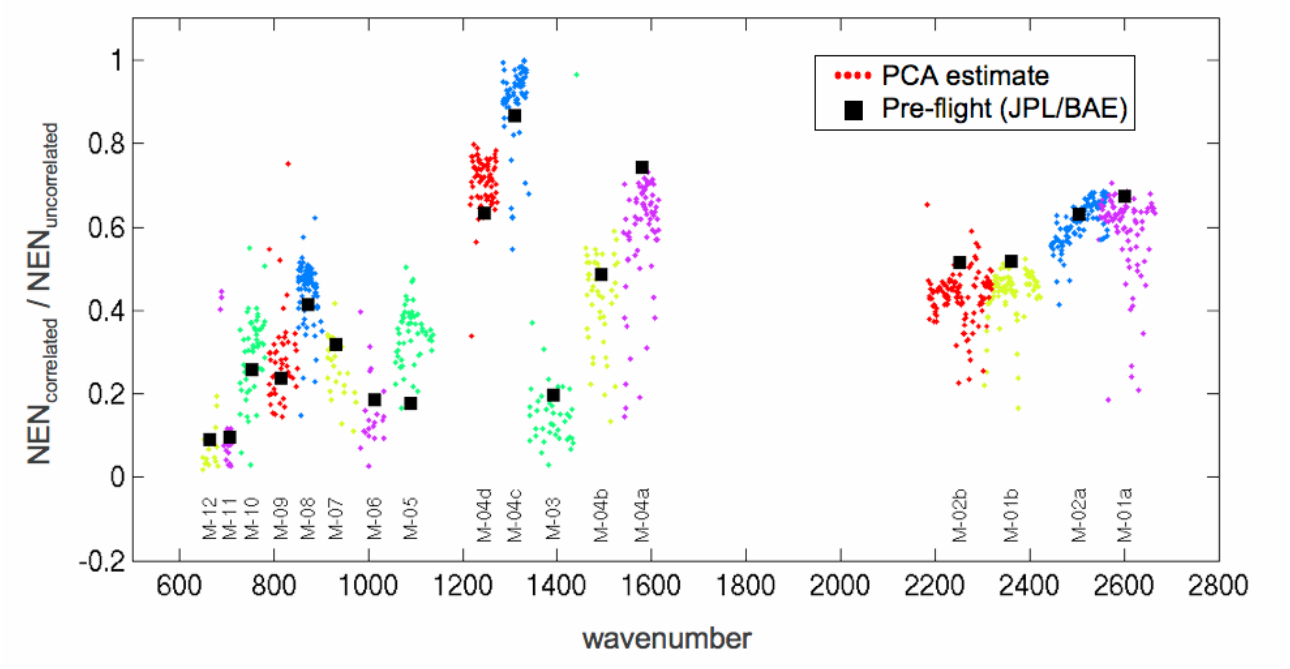

Figure 11: Ratio of spectrally correlated noise to spectrally uncorrelated noise derived from PCA and from pre-flight analysis of blackbody view ensembles. Spectral channels are color-coded according to the AIRS detector arrays (M-01 through M-12).

8. Non-gaussian behavior

Instead of using statistics, here we examine individual differences between the original and reconstructed radiances as a function of time or as a spatial map to investigate non-Gaussian behavior. A particular phenomenon of interest for AIRS is referred to as "popping," as described by Weiler et al. [12]. An example for AIRS channel number 530 in the M-09 detector array at ~821.597 cm^{-1} for granule 198 is shown in Fig. 12. This channel is shown to exhibit a relatively small amount of popping. The reconstruction error, L_{ORIG}-L_{RECON}, when viewed as a time series or a map shows differences that are non-Gaussian. Assuming Gaussian statistics, we can compute the number of expected N-sigma events within a granule, giving 3852 1-sigma events, 547 2-sigma events, and 36 3-sigma events expected to occur within one granule of data. For this example, the number of N-sigma events is in very good agreement with the Gaussian expectations. Further defining a "pop" as M consecutive N-sigma events (all of the same sign), we can also compute the number of pops expected within a granule. The actual definition of a pop and the values of M and N are subjective. Using M=4, we expect 15 1-sigma pops per granule and no 2- or 3-sigma pops per granule with pure Gaussian behavior. Counting the number of M=4 pops in the AIRS data, we find 187 1-sigma pops, 15 2-sigma pops, and 0 3-sigma pops.

In general, we find the number of N-sigma events for all AIRS channels to be very close to that expected from Gaussian behavior, but that many channels exhibit popping beyond what is expected from Gaussian behavior. This result is demonstrated in Fig. 13. The number of channels found to exhibit M=4 popping at the 1-, 2-, and 3-sigma level are 875, 482, and 42, respectively. The popping behavior is found in many of the longwave and midwave photo voltaic detector channels; the longwave photo-conductive and shortwave channels do not exhibit this behavior. In addition, we find a small number of channels (14) which exhibit "striping," where the non-Gaussian behavior extends across a full cross track scan of 90 footprints.

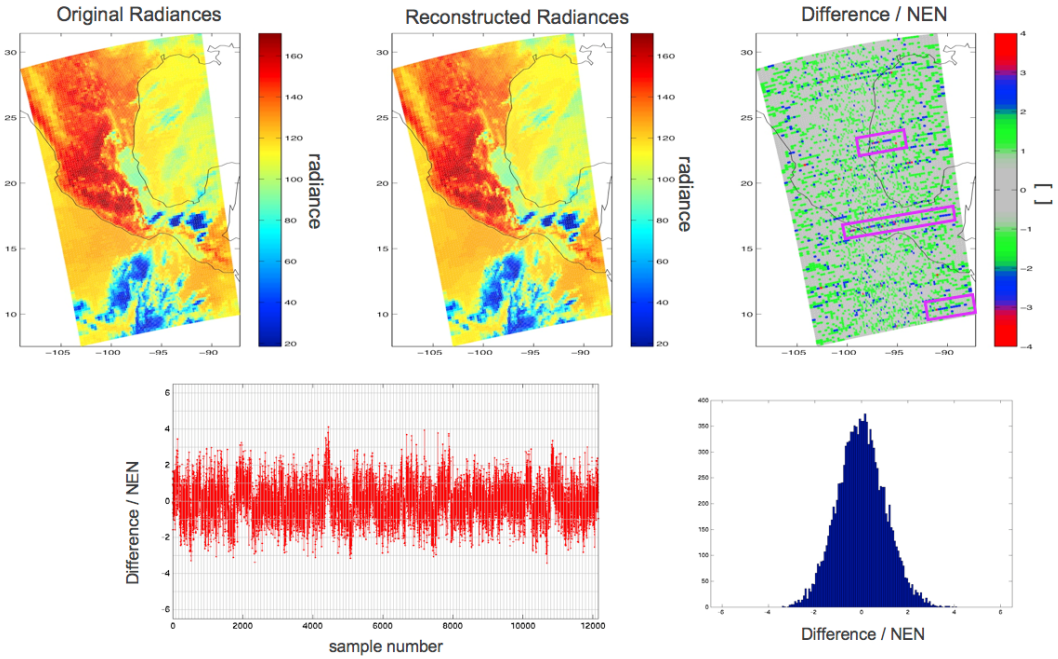

Figure 12: Original and reconstructed radiances and their differences for channel number 530 in the M-09 detector array at ~821.597 cm-1.

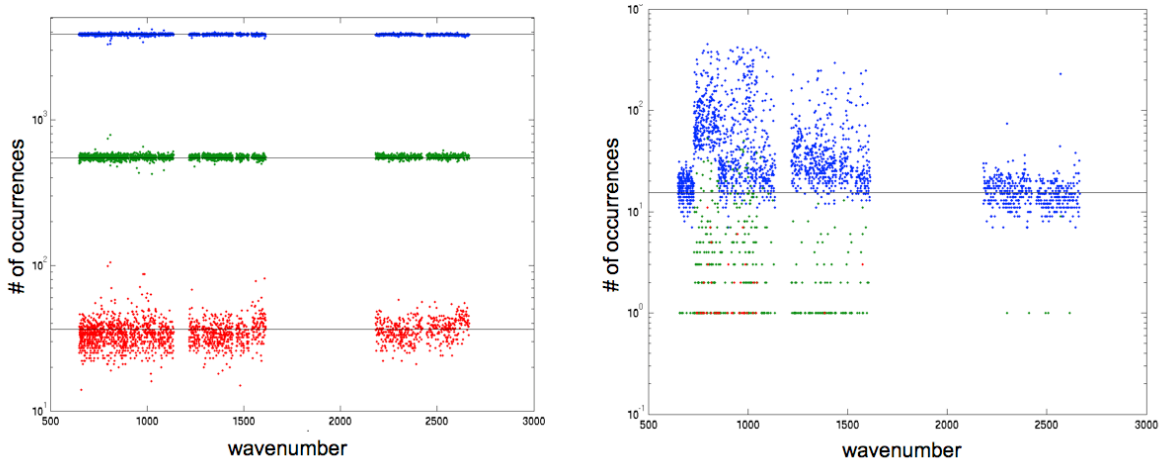

Figure 13: PCA analysis of non-Gaussian behavior. Left: Number of 1- (blue), 2- (green), and 3- (red) sigma events found per granule compared to the pure-Gaussian expectations (black lines). Right: Number of 1- (blue), 2- (green), and 3- (red) M=4 pops found per granule compared to the pure-Gaussian expectations (black-line for 1-sigma pops, 0 otherwise).

9. Inspection of pcs and individual spectra

When performing PCA on Earth scene data, inspection of the PCs and individual reconstructed radiances is a powerful diagnostic for sensor performance evaluation and monitoring. The low order (most relevant) PCs represent real variations in the scene within the ensemble and are most often due to clouds, surface, water

vapor, and temperature variations. The highest order (least significant) PCs most often represent random white noise. Therefore, the spectral characteristics of the PCs can be inspected to look for artifacts due to sensor noise and/or calibration. Such studies have been performed, for example, by Rodgers [13] for the Tropospheric Emission Spectrometer (TES) on EOS Aura. We have also recently used such techniques to diagnose processing artifacts in the University of Wisconsin Scanning-HIS [14], an aircraft based infrared spectro-radiometer.

Figure 14 shows the first 50 (most significant) PCs for AIRS granule number 196. Each PC is normalized to unit variance; the variance explained by each PC is the PC multiplied by the square root of the corresponding eigenvalue. The most significant PCs show spectral structure that is explained by cloud, surface, and atmospheric effects, as expected. The least significant PCs (not shown) are characteristics of white detector noise. PC numbers ~25 to ~45 show effects which are not explained by surface or atmospheric effects or by random detector noise. Most noticeable are the spectral signatures in the longwave spectral regions which are isolated within individual detector arrays, such as, for example, PC number 28 (black curve) in the ~851-903 cm^{-1} region (array M-08), or PC number 43 (blue curve) in the ~728-781 cm^{-1} (array M-10) region. These signatures are indicative of noise that is correlated for channels within individual detector arrays. The photovoltaic channels each have two detectors (labeled A-side and B-side) and inspection of the PCs shows spectral correlation among A-side channels and among B-side channels but less between A-side and B-side channels. These results are qualitatively consistent with Weiler [11] and with the magnitude of spectrally correlated noise presented in Section 7 and shown in Fig. 11.

To characterize the effects of sensor noise it is also useful to examine the spectral differences between original and reconstructed radiances. An example is shown in Fig. 15. This figure shows the differences from the mean spectra for five AIRS footprints collected over a spatially uniform scene selected from granule 200. The differences are shown for the original spectral radiances and for the reconstructed spectral radiances. With the majority of the random noise removed (center panel), the spectrally correlated nature of the radiances is evident in the discontinuities between adjacent detector arrays. This is also evident in the original radiances (left panel) but is more difficult to identify due to the presence of the random noise. Also shown is the difference between the original and reconstructed radiances (right panel) which, for this example, demonstrates that the PC filtering is removing spectrally random noise and not introducing the spectrally correlated artifacts. The mean differences over each detector array are shown in magenta in each panel for both the original and filtered spectra to highlight the fact that the PCA is not introducing the spectrally correlated artifacts. Also, in the right panel, the standard deviation of the original minus filtered spectra for each detector array is compared to the mean total noise values (provided from the L1B granule file) for each detector array to highlight the fact that the PCA is removing spectrally random noise. Other useful techniques include the reconstruction of radiances using selected PCs and/or with selected PCs excluded, to see the impact of those PCs on the resulting radiance spectra.

10. Summary

Dependent set PCA using Earth scene data is a simple yet very powerful tool not only for noise filtering, lossy compression, and for the characterization of hyperspectral sensor noise and other variable artifacts. This is possible due to the inherent redundancy of the high spectral resolution observations. We have presented our approach for dependent set PCA of S-HIS and AIRS Earth scene data. Aspects of the analyses include 1) comparisons of PCA and MMSE; 2) impact of PCA on outliers; 3) estimation of NEDT using PCA and comparisons to values derived from on-board blackbodies, 4) estimation of the scene dependence of NEDN, 5) estimation of the spectrally correlated component of NEDT and comparison to pre-launch analyses using blackbody views, 6) investigation of non-Gaussian noise behavior, and 7) inspection of

individual PCs. The results of the PCA analyses are generally consistent with results obtained pre-launch and on-orbit using blackbody and/or space view data, giving confidence in the approach and methodology. Specific findings include: 1) PCA performances on simulated S-HIS data approach those of MMSE; 2) dependent PCA noise filtering is robust even in presence of outliers; 3) PCA estimates of AIRS spectrally random and spectrally correlated NEDN compare well with estimates computed from on-board blackbody and space views, 4) the signal dependence of AIRS NEDN is accurately parameterized in terms of the scene radiance, 5) examination of the reconstruction error allows non-Gaussian phenomenon such as popping to be characterized, and 6) inspection of the PCs and individual PC filtered radiance spectra is a powerful technique for diagnosing low level artifacts in hyperspectral data. Dependent set PCA is a useful technique for evaluation and monitoring of present and future hyperspectral sensors. Future plans include continued analysis of AIRS noise performance and evaluation of forthcoming data from the Infrared Atmospheric Sounding Interferometer (IASI) on the METOP-1 platform and the Cross-track Infrared Sounder (CrIS) on the NPP platform.

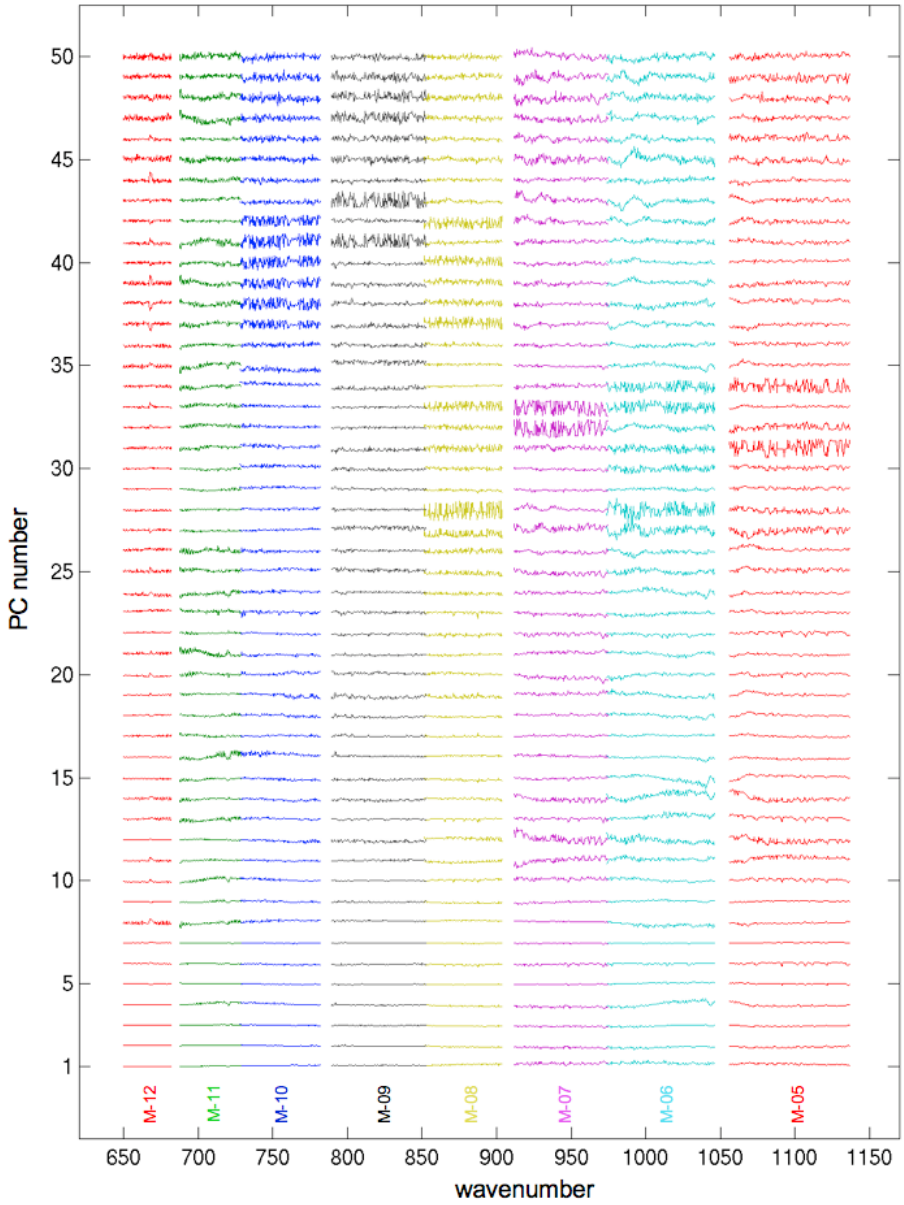

Figure 14: The first 50 PCs for AIRS granule 196. The PCs are plotted with the more significant PCs toward the bottom of the plot, with a fixed offset between consecutive PCs, and color-coded in the spectral domain according to the AIRS detector arrays (M-05 through M-12 shown).

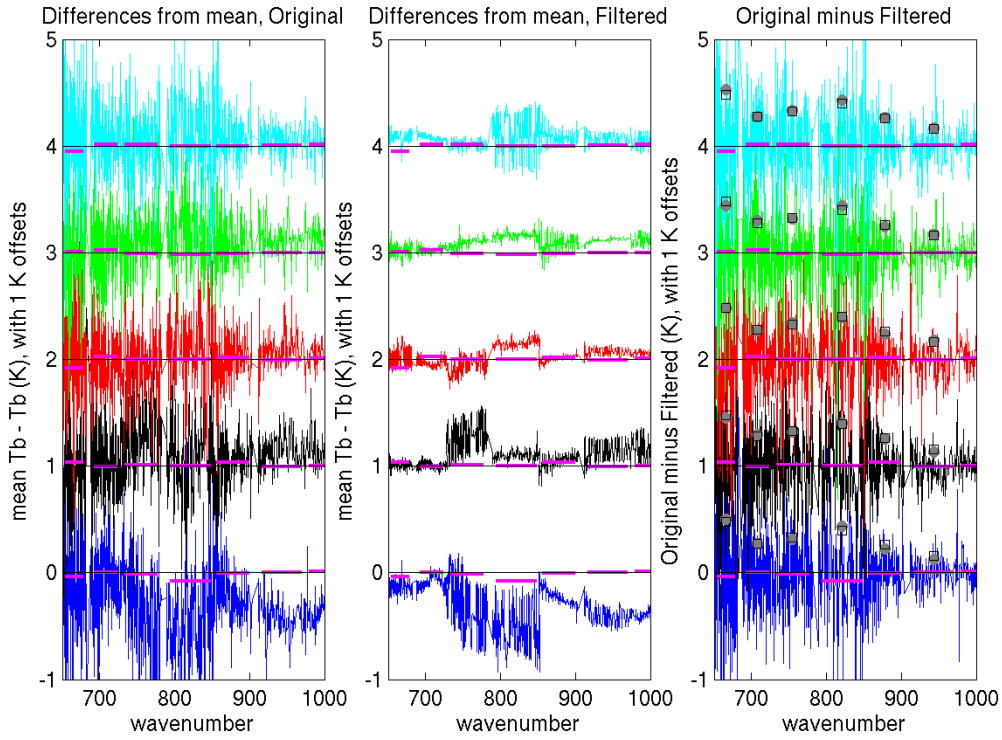

Figure 15: Spectral differences from the mean spectrum for five AIRS spectra collected over a spatially uniform scene from granule 200. Differences are shown for the original spectra (left panel) and the PC filtered spectra (center panel). The right hand panel shows the differences between the original and filtered spectra. The magenta lines are the mean of original minus filtered values for each detector array. In the right hand panel, the solid grey circles are the standard deviation of the original minus filtered spectra for each detector array and the black squares are the mean total noise values (provided in the L1B granule file) for each detector array.

11. References

P. Antonelli, 2000: Principle component analysis: A tool for processing hyperspectral infra-red data. *Ph.D. thesis, Univ. of Wisc*., Madison, Wisc.

H.-L. Huang and P. Antonelli, 2001: Application of principle component analysis to high-resolution infrared measurement, compression and retrieval. *J. Appl. Meteor.* **40**, 365-388 [doi:10.1175/1520-0450(2001)040<0365:AOPCAT>2.0.CO;2]

P. Antonelli, H. E. Revercomb, L. A. Sromovsky, et al., 2004: A principle component noise filter for high spectral resolution infrared measurements. *J. Geophys. Res.* **109**, D23102. [doi:10.1029/2004JD004862]

M. Goldberg, et al., 2004: Principle component analysis of AIRS data. In *Proc. of the Workshop on Assimilation of high spectral resolution sounders in NW*P, ECMWF. Available from: http://www.ecmwf.int/publications/library/do/references/list/17444

D. D. Turner, R. O. Knuteson, H.E. Revercomb, C. Lo and R. G. Dedecker, 2006: Noise reduction of atmospheric emitted radiance interferometer (AERI) observations using principal component analysis. *J.Atmos. Oceanic Technol.* **23**, 1223-1238 [doi:10.1175/JTECH1906.1].

H. H Aumann, M. T. Chanine, C. Gautier, et al., 2003: IRS/AMSU/HSB on the Aqua mission: design, science objectives, data Products, and processing systems. *IEEE Tran. Geosci. Rem. Sens.* **41**, 253-264 [doi:10.1109/TGRS.2002.808356].

S. Gaiser, 2004: Calibration status of the atmospheric infrared sounder (AIRS) on Aqua. AIRS Science Team Meeting, November 2004.
http://airs.jpl.nasa.gov/Science/ResearcherResources/MeetingArchives/TeamMeeting20041130/

T. S. Pagano, H. H. Aumann, S. E. Broberg, S. L. Gaiser, D. E. Hagan, T. J. Hearty and M. D. Hofstadter, 2002: On-Board calibration techniques and test results for the atmospheric infrared sounder (AIRS). In *Earth Observing Systems VII*, W. L. Barnes, Ed., Proc. SPIE **4814**, 243-254. [doi:10.1117/12.451547]

T. Pagano, H. H. Aumann, S. L. Gaiser and D. T. Gregorich, 2003: Early calibration results from the atmospheric infrared sounder (AIRS) on Aqua. In *Optical Remote Sensing of the Atmosphere and Clouds III*, H.-L. Huang, D. Lu and Y. Sasano, Eds, Proc. SPIE **4891**, 76-83. [doi: 10.1117/12.465869]

T. Pagano, 2002: Estimation of AIRS correlated noise. AIRS Design File Memo. ADFM-614, Available through the AIRS Project Office.

M. Weiler, 2003: Investigation of AIRS correlated noise. AIRS Design File Memo, ADFM-620, Available through the AIRS Project Office.

M. Weiler, K. R. Overoye, J. A. Stobie, P. B. O'Sullivan, S. L. Gaiser, S. E. Broberg and D. A. Elliott, 2005: Performance of the atmospheric infrared sounder (AIRS) in the radiation environment of low-Earth orbit. In *Earth Observing Systems X*, J. J. Butler, Ed., Proc. SPIE **5882**, 588210. [doi:10.1117/12.615244]

C. Rodgers, 2005: Application of singular value decomposition to high spectral resolution measurements," *12th Atmospheric Science from Space using Fourier Transform Spectrometry* (ASSFTS) (2005). http://bernath.uwaterloo.ca/ASSFTS/

D. C. Tobin, H. E. Revercomb, R. O. Knuteson, et al. 2006: Radiometric and spectral validation of atmospheric infrared sounder observations with the aircraft-based scanning high-resolutioniInterferometer sounder. *J. Geophys. Res.* **111**, D09S02. [doi:10.1029/2005JD006094].

K. Pearson, 1901: On lines and planes of closest fit to systems of points in space" *London Edimburgh Dublin Philos. Mag. J. Sci.*, 2, 527.

H. Hotteling, 1933: Analysis of acomplex of statistical variables into principal component. *J. Educ. Psychol.*, 24, 417-441.

J. Heland and K. Schafer, 1997: Analysis of aircraft exhausts with Fourier-transform infrared emission spectroscopy. *Appl. Opt*. 36, 4922-4931.

D. Tobin, P. Antonelli, H. Revercomb, S. Dutcher, D. Turner, J. Taylor, R. Knuteson, and K. Vinson, 2007: Hyperspectral Data Noise Characterization using Principle Component Analysis: Application to the Atmospheric Infrared Sounder".*Journal of Applied Remote Sensing.* 1, 013515 [doi: 10.1117/1.2757707]

Paolo Antonelli, is a researcher at the Space Science and Engineering Center Center at the University of Wisconsin-Madison. He received Laurea degree, and PhD degree in physics from the University "La Sapienza" of Rome in 1994, and from University of Wisconsin – Madison in 2001, respectively. His research focuses on: development of information content and data inversion algorithm for high spectral resolution infrared data; convection detection and convective rain estimation from microwave and infrared data; and international training in remote sensing.

David Tobin is an associate scientist at the Space Science and Engineering Center at the University of Wisconsin-Madison. He received BS, MS, and PhD degrees in physics from the University of Maryland Baltimore County in 1991, 1993 and 1996, respectively. His current research interests include infrared molecular spectroscopy and atmospheric radiative transfer, atmospheric water vapor, infrared spectro-radiometer calibration and validation, infrared remote sensing, and climate change.