Principle Component Analysis of AIRS and CrIS data.

H. H. Aumann and Evan Manning
Jet Propulsion Laboratory, California Institute of Technology.
4800 Oak Grove Dr., Pasadena, CA 91109

Abstract

Synthetic Eigen Vectors (EV) used for the statistical analysis of the PC reconstruction residual of large ensembles of data are a novel tool for the analysis of data from hyperspectral infrared sounders like the Atmospheric Infrared Sounder (AIRS) on the EOS Aqua and the Cross-track Infrared Sounder (CrIS) on the SUOMI polar orbiting satellites. Unlike empirical EV, which are derived from the observed spectra, the synthetic EV are derived from a large ensemble of spectra which are calculated assuming that, given a state of the atmosphere, the spectra created by the instrument can be accurately calculated. The synthetic EV are then used to reconstruct the observed spectra. The analysis of the differences between the observed spectra and the reconstructed spectra for Simultaneous Nadir Overpasses of tropical oceans reveals unexpected differences at the more than 200 mK level under relatively clear conditions, particularly in the mid-wave water vapor channels of CrIS. The repeatability of these differences using independently trained SEV and results from different years appears to rule out inconsistences in the radiative transfer algorithm or the data simulation. The reasons for these discrepancies are under evaluation.

1. Introduction

Principle Component (PC) analysis has been used widely for the analysis of the information content of high spectral resolution spectra starting with Goody et al. [1]. Huang, H.-L. and P. Antonelli [2] used PC analysis for data compression of infrared spectra and temperature and water vapor profile retrievals. Goldberg et al. [3] used PC reconstruction for data quality control. In these applications the EigenVectors (EV) used for the PC reconstruction were derived from a large subset of observed spectra. We refer to these as Empirical EV. We can also derive EV generated from simulated spectral radiances. We refer to these as synthetic EV. The simulated spectra used for the derivation of the synthetic EV represent what is expected to be observed, in our case the upwelling spectral radiances from atmospheric temperature and moisture profiles and clouds. The simulation assumes that the forward algorithm which converts the state of the atmosphere and clouds to upwelling spectra is perfect, that the spectral response functions (SRF) are perfectly known and that the instrument is perfectly calibrated and stable. We then compare the observed radiances, "obs", with the reconstructed radiances, "PC.rec" using the synthetic EV and evaluate the statistical properties of (obs-PC.rec). The synthetic EV are derived for zero-mean-normalized spectra, i.e. the mean values (one number for each spectrum) is subtracted from all spectral elements of each spectrum before the reconstruction, and is then added back after the reconstruction. The mean of (obs-PC.rec) averaged over a large ensemble of spectra is expected to be zero and the standard deviation of (obs-PC.rec) is expected to be the instrument noise. Significant deviations from these expectations indicate that one or more of the assumptions made in the derivation of the SEV were incorrect. We used Radiative Transfer Algorithms (RTA) and scattering code to convert the state of the atmosphere and clouds defined by ECMWF and surface emissivity climatology to create several million simulated spectra for AIRS [4] and CrIS [5]. This was done independently for both instruments at the University of Maryland Baltimore Campus (UMBC) and NASA/Langley (LARC). From these spectra we selected maximum variance sets of typically 10,000 spectra to derive the SEV. We then evaluated the statistical properties of the PC reconstruction residuals for AIRS and CrIS spectra. AIRS and CrIS have comparable signal-to-noise, spectral range and spectral resolution. AIRS and CrIS both have a 12 km field of view at nadir, but AIRS scans in cross-track continuously, while CRIS employs a step and stare method requiring image motion compensation. AIRS and CrIS are specified to have 200 mK absolute accuracy, but both appear to be better than that [6, 7].

2. Data

AIRS and CrIS are nominally in 1:30 PM ascending node polar orbits, but at different altitudes, 705 km for AIRS, 825 km for CrIS. Approximately every three days the ground-tracks of the two orbits overlay nearly exactly. For these days there is a large number of spectra from AIRS and CrIS where both instruments observe nearly the same spot on the ground. These spectra are referred to Simultaneous Nadir Overpasses (SNO). An AIRS/CrIS SNO pair is defined as an AIRS and CrIS spectrum taken within 10 minutes, where the footprints on the ground are separated by no more than 8 km, both observations occurred within 3 degree of nadir and both spectra pass quality screening.

AIRS has 2378 independent spectral channels, with Noise Equivalent Delta Temperature (NEdT) of typically 0.2K. We focus on the 2206 “good” channels of the 2378 channels which have NEdT less than 1K, and the noise is Gaussian. Of the remaining 172 channels about 60 have been dead since before launch in 2002, the others have non-gaussian noise. Most of these channels are in the 800-1150 cm^{-1} region of the spectrum. These channels are replaced in the L1c process using the high information redundancy in the spectra [9]. The statistical results presented in the following analysis are based on the “good” channels.

CrIS has 1305 channels, produced in 3 bands using 9 independent detector per band in a 3x3 pattern. The center detector is on-axis, the other eight detectors are off-axis. The CrIS data used the IDPS (NOAA) calibration and are presented on a fixed frequency grid. The off-axis detectors are shifted on-axis in the ground calibration software. While the data identify which of the 9 CrIS detectors matched the AIRS footprint, in the following analysis we do not distinguish between the detectors. We apodize the CrIS data using the Hamming apodization, which is a running average with weights (0.22825, 0.5435, 0.22825) and assumed that the true CrIS SRF is the SRF given by the Hamming function. The Hamming SRF is also used for the data simulation. The quality filtered used the prescription by Han et al. [8].

There were 28650 SNO from the tropical oceans from all days of April 2014. In order to simplify the analysis we selected spectra with no or only very little cloud based on the magnitude of the cloud effect. We define the cloud effect as the difference between the brightness temperature in the 900 cm^{-1} atmospheric window channel and the known surface temperature from NOAA. We then focused on the 9520 (34%) of the spectra where the cloud effect is less than 5K in the AIRS and CrIS data. Of this 5 K difference typically 3 K is due to water vapor absorption in a clear tropical atmosphere. The “clear“ data can’t be analyzed in terms of the mean and the standard deviation of the difference of pairs, because of the spatial mismatch, and differences in the spectral sampling and the spectral resolution. The data can be analyzed by comparing PC reconstruction residuals.

3. Results

Figures 1a and 1b show a typical CrIS and AIRS reconstructed night-time spectrum using 200 EV overlaid on the observed (L1c) spectra in brightness temperature units. On this large scale the reconstruction appears to be perfect. Figure 1c and 1d show the corresponding reconstruction residuals. The reconstruction residual for a single spectrum is dominated by the noise in each spectral channel. Figure 2a shows the nominal NEdT at 250K from AIRS and CrIS. Except for the shortwave channel, CrIS is less noisy than AIRS.

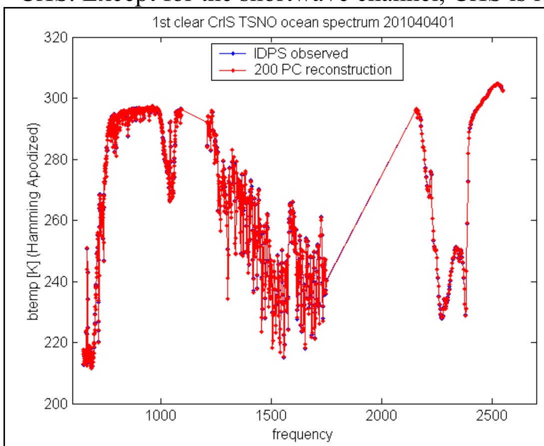

Figure 1a. Typical CrIS spectraum

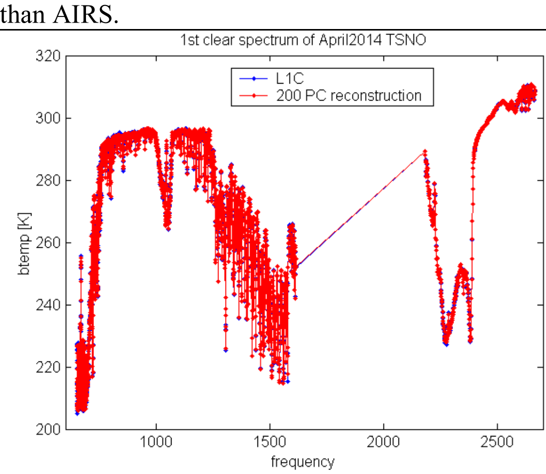

Figure 1b. Typical AIRS L1c spectrum

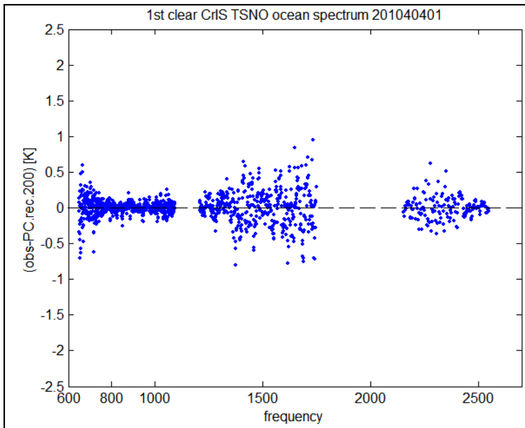

Figure 1c. CrIS single spectrum reconstruction residual

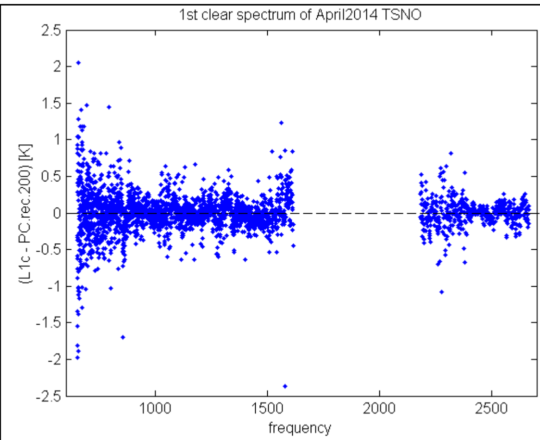

Figure 1d. AIRS single spectrum reconstruction residual

We now calculate the mean and standard deviation of the difference between the observed spectra (obs) and the PC reconstructed spectra, PCrec, for the 9520 clear tropical ocean spectra. The number of PC required for the reconstruction is related to the information content of the data. Based on numerical experiments we found that 20 PC reconstructions achieve an amazingly good fit on a 200 to 320K scale, but there are differences of several degree K at the center of lines. Figure 2b shows the stdev (obs-PC.rec) from AIRS and CrIS as function of frequency for the clear ocean data using a 100 PC reconstruction trained on UMBC spectra. We show the difference in brightness temperatures, since the accuracy of the absolute calibration, which is relevant to temperature and water vapor retrievals, is better visualized in terms of brightness temperatures. As expected, the standard deviation of the PC reconstruction residual mimics the NEdT except in the very cold shortwave region between 2250 and 2400 cm^{-1}, where the non-linearity of the Planck function makes the comparison with the NEdT, which is referenced on a 250K scene, irrelevant. Results with between 100 and 200 PC appear to be optimum. Beyond 200 PC the reconstruction starts to fit the noise. The information content of the CrIS and AIRS spectra is captured with about 150 PC.

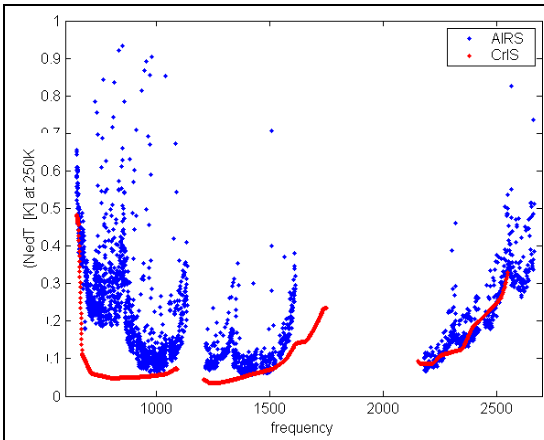

Figure 2a. AIRS and CrIS NeDT at 250K

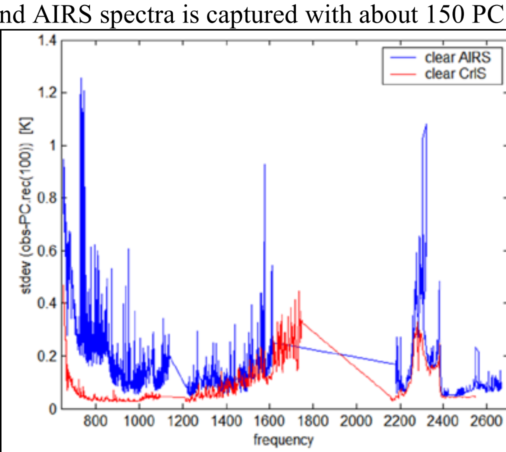

Figure 2b. AIRS and CrIS stdev(obs-PC.rec)

In Figure 3 we compare the mean of the reconstruction residuals (averaged over all clear data) from AIRS and CrIS for SNO using 150 PC reconstructions in the brightness temperature domain. For channels below 1200 cm^{-1} and between above 1700 cm^{-1} the reconstruction residuals of AIRS and CrIS are of comparable magnitude. In the mid-wave-band (1200-1700 cm^{-1}), the CrIS mean(obs-PC.rec) is much larger than the corresponding AIRS mean, with deviation from zero as large as 0.5K. The mean reconstruction residuals fluctuates about zero. The average over all spectral channels is zero by definition, since the PC reconstruction uses zero-mean-normalized spectra. We use the magnitude of this fluctuation about zero to define a Reconstruction Figure of Merit, RFM, for the longwave (<1200 cm-1), midwave (1200 – 1700 cm-1) and shortwave (>1700 cm-1) regions. The RFW is the standard deviation of the mean PC reconstruction residual evaluated for CrIS channels and the "good " AIRS channels in each of the three regions. The results are summarized in Table 1.

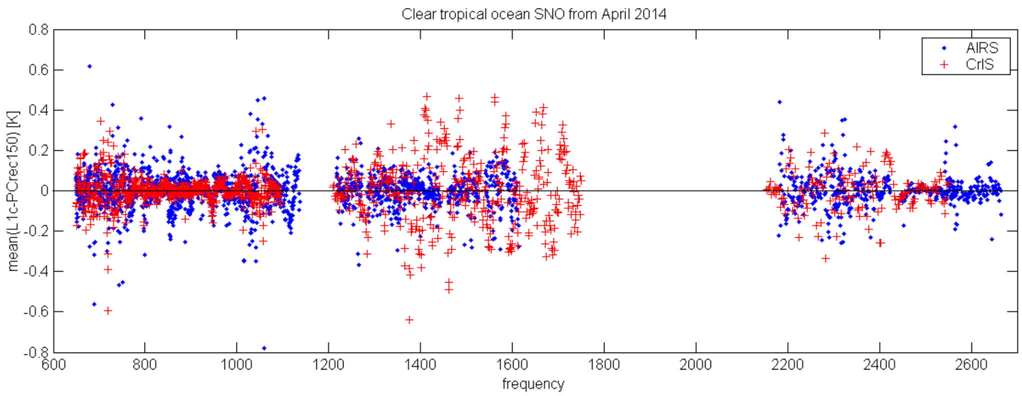

Figure 3. Mean reconstruction residuals from AIRS and CrIS for the 9520 clear SNO using 150 PC

	AIRS	CrIS
longwave	0.092K (1141 channels)	0.070K (713 channels)
midwave	0.074K (577 channels)	0.175K (392 channels)
shortwave	0.070K (488 channels)	0.109K (199 channels)

Table 1. AIRS and CrIS Reconstruction Figure of Merit for clear tropical ocean SNO spectra from April 2014.

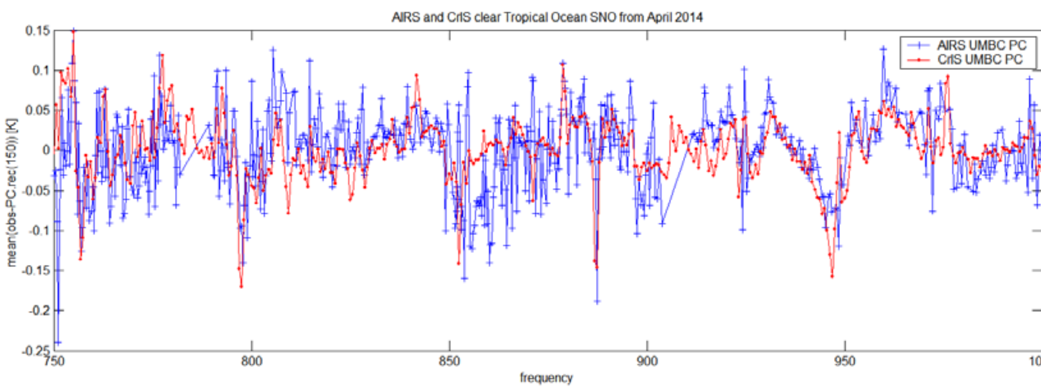

Figure 4. Mean reconstruction residuals from AIRS and CrIS for the 9520 clear SNO using 150 PC

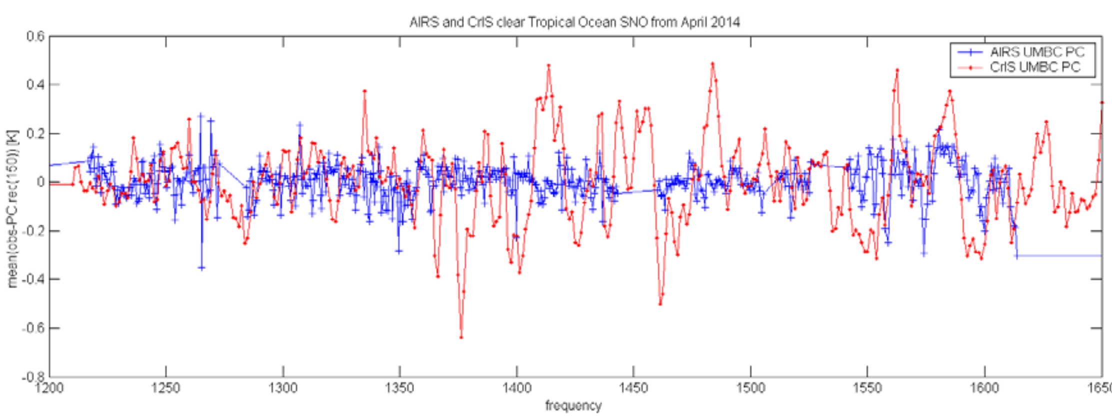

Figure 5. Mean reconstruction residuals from AIRS and CrIS for the 9520 clear SNO using 150 PC

In order to show these differences more clearly we zoom in (Figure 4) on the 750 to 1000cm^{-1} region of the spectrum. There is a high correlation between the AIRS and CrIS reconstruction residual. Many of the deviations of the reconstruction residual from zero seen by AIRS are also seen by CrIS. The cold 0.15K feature near 792 cm^{-1} seen in AIRS and CrIS is a CO_2 Q-branch. The UMBC training used profiles from 2009 for AIRS and CrIS, while the data are from almost 4 years later, i.e. about 8ppmv more CO_2. This makes the observed radiances colder than the reconstructed radiances. A 0.1K cold broad dip in the mean residual spectra centered near 949 cm^{-1} appears to due to the presence of detectable amounts of C_2H_4, which is not included in the UMBC simulation of AIRS and CrIS. A 0.15K deep absorption features near 888 cm^{-1} is seen by AIRS and CrIS has no good explanation. There is a shift in the AIRS calibration of the order of 0.05K in the transitions between detector modules in the 855 cm^{-1} region.

Figure 5 zooms in the 1200-1650 cm^{-1} region of the spectrum. In this mid-wave band the residuals from CrIS are much larger than those from AIRS. While the AIRS residual show a number of isolated channels with a larger than normal reconstruction residual, likely due to Spectral Response Function (SRF) errors in "good" channels, CrIS shows a spectrally highly correlated bias for many channels.

4. Discussion

We selected clear tropical ocean SNO for the comparison of CrIS and AIRS to have a sample of highly similar atmospheric conditions from both instruments and. At least for a start, avoid complex issues with instrument performance and data simulation under cloudy conditions. While there are many similarities, there are curious systematic differences exceeding the 200 mK levels, the nominal absolute calibration accuracy of AIRS and CrIS.

4.1. AIRS Residuals:

The reconstruction residuals from AIRS are largely a consequence of the way an A/B redundancy was implemented throughout the AIRS electronics and the grating array implementation. The signal from each AIRS channel is actually derived from an A/B redundant pair of detectors. Normally the onboard data processing combines the signal from the A-side and the B-side, A+B. If the A-side fails, the B-side is selected (by ground command) and vice versa. Each AIRS channel has its own Spectral Response Function (SRF), which was determined prelaunch. The nominal SRF values for the A+B configuration are given in a 2378 x100 tabulation in the Channel Properties file. However, the difference between the nominal SRF and the true SRF is slightly A/B dependent. This dependence and measurement error in the pre-launch SRF measurements creates a small bias unique to each spectral channel. These differences between A+B/A/B are systematic. The AIRS ground calibration assumes and the AIRS data simulation assumes the A+B state for all channels. These differences, which is particularly pronounced in the 800 – 1000 cm^{-1} region of the spectrum, were first noted years ago under very cold conditions, but is seen here for typically 300 K clear spectra. Figure 6 shows the AIRS reconstruction bias as function of the A/B state of the channel. Since the simulations used for the creation of the EV assume a perfect calibration, residual A+B/A/B differences are revealed as reconstruction bias.

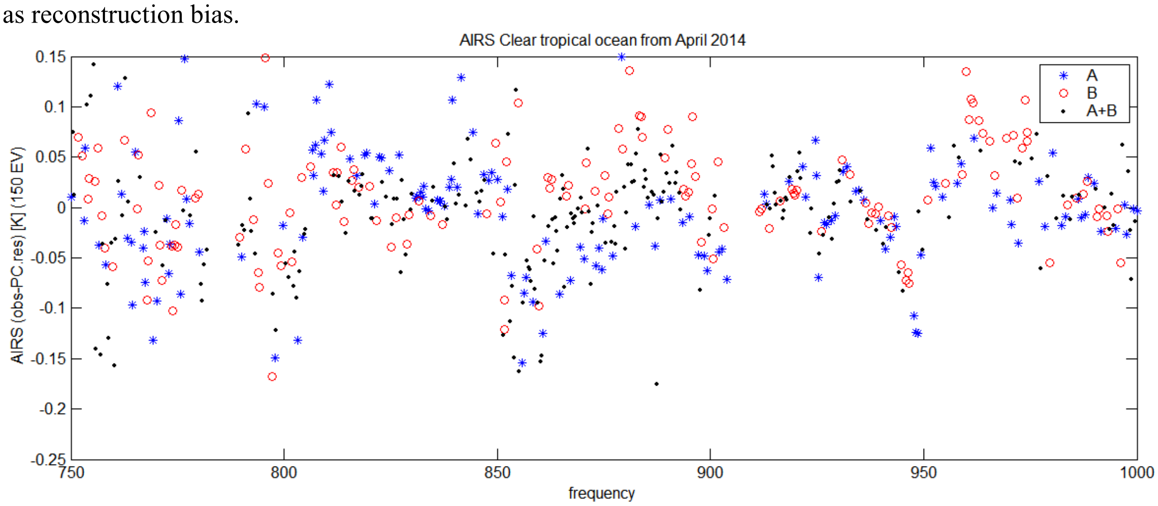

Figure 6. Mean reconstruction residuals from AIRS as function of A/B/A+B state.

4.2. CrIS Residuals:

We evaluate to potential effects: EV Training set and SRF uncertainty.

4.2.1 EV training set:

We have CrIS training sets for clear and cloudy data from UMBC and LARC. The results shown in Figures 3 through 6 used the EV derived from the UMBC simulations. Figure 7 compares the CrIS clear tropical ocean reconstruction residual bias from UMBC and LARC derived synthetic EV. The systematic residuals in the CrIS mid-wave band seen in the UMBC based EV is nearly identical to the residuals from the LARC based EV, particularly in the mid-wave band (Figure 8).

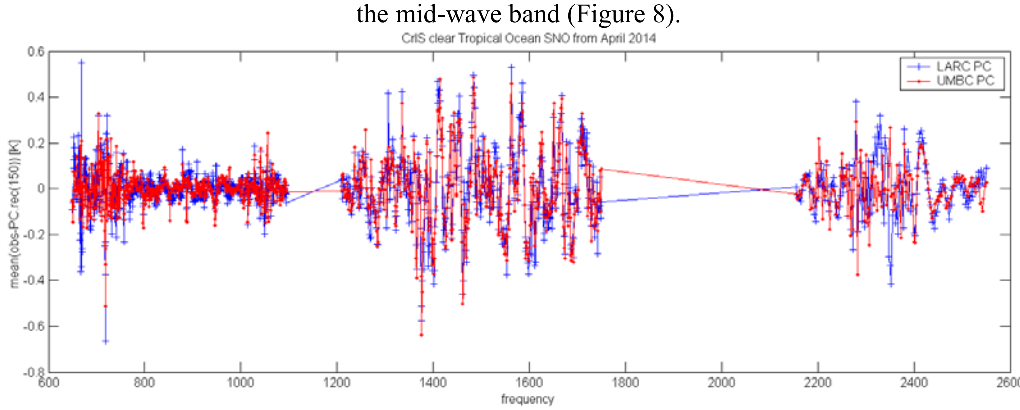

Figure 7. CrIS clear tropical ocean reconstruction residual bias.

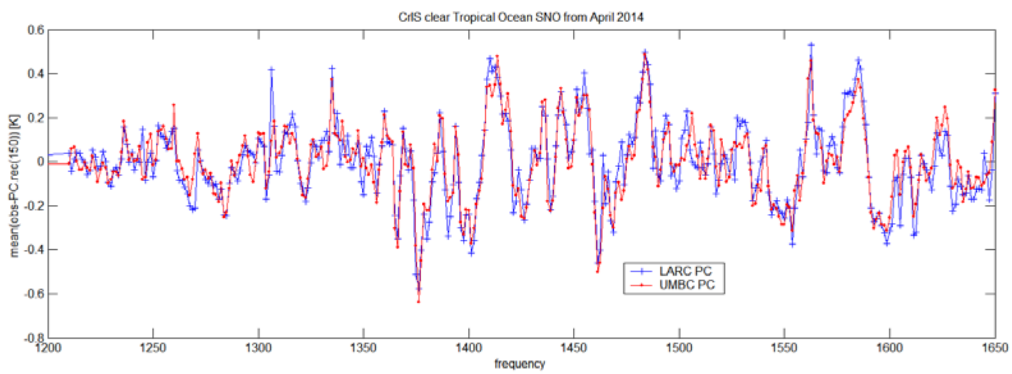

Figure 8 CrIS clear tropical ocean reconstruction residual bias between 1200 and 1650 cm^{-1}.

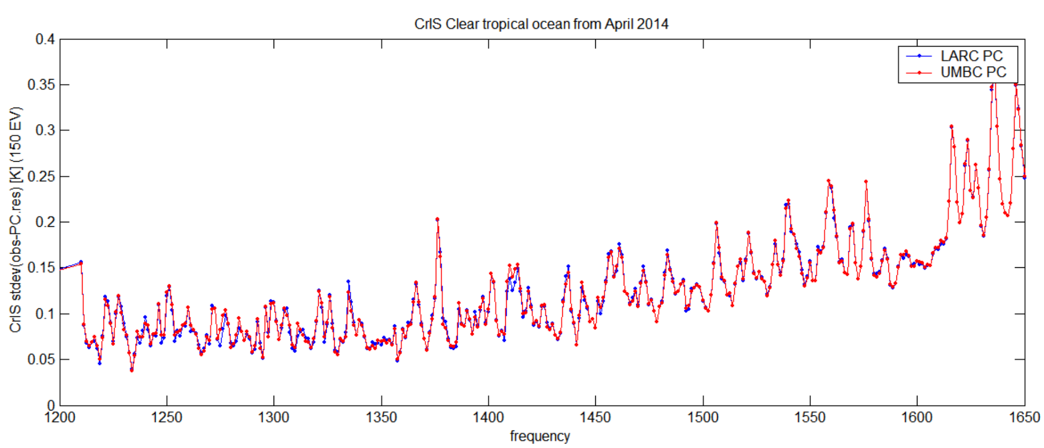

Figure 9. Standard deviation of the CrIS reconstruction residuals.

Figure 9 shows the standard deviation of the reconstruction residual for CrIS mid-wave band. The standard deviation should mimic the smooth spectral dependence of the noise. Instead we see unexpected factor of two and larger excursions from the LARC and UMBC EV based reconstructions.

4.2.2. SRF uncertainty.

We noted previously that CrIS has 9 detectors for each band in a 3x3 configuration. The center detector is on-axis, the other eight detectors are off-axis. The off-axis detectors are shifted from their nominal positions to the on-axis position in the ground calibration software. This processes has residual errors, which result in an uncertainty in the SRF. We tested the effect of an SRF uncertainty by using the Hanning apodization for the CrIS data (a 0.25, 0.5, 0.25 sliding average), which creates a 10% broader SRF than the Hamming apodization which was used in the EV training set. The resulting RFM increases by nearly a factor of three. This indicates that the CrIS SRF created by convolving the un-apodized calibrated radiances with the Hamming function is actually slightly broader than the Hamming function used to create the simulated data. This suggests that the shifting the eight detectors to the on-axis position or other disturbances create errors which effectively broaden the CrIS SRF.

Uncorrected residuals of the observed magnitude will degrade retrieval and direct assimilation results for weather forecasting and will frustrate the application of the data to climate change analysis. The material shown above has been presented to the AIRS and CrIS calibration teams. An improved CrIS calibration algorithm, CCAST, has been developed. The difference between the March 2015 release of CCAST and the IDPS algorithm are very small. Further analysis is required, including separating the results from CrIS into individual detectors and looking at cloudy data. One approach which accomplished both is to create images of the residuals as function of frequency from data granules (8 minutes of data). Each CrIS data granule contains 90 x 180 spectra covering about 1600 x 3200 km on the ground. An example of this approach is found in [10]. Analysis of these images will reveal spatial correlations and possibly identify outlier detectors.

5. Conclusions

Synthetic EV used for the statistical analysis of the PC reconstruction residual of large ensembles of spectra are a novel tool for the analysis of spectra actually observed and spectra calculated based on imperfect understanding of an instrument. The analysis of clear tropical ocean SNO reveals unexpected differences in the reconstruction bias between AIRS and CrIS at the larger than 200 mK level under relatively clear conditions, particularly in the mid-wave water vapor channels. The reasons for these discrepancy are under evaluation.

Acknowledgments

We thank Drs. Sergio Machado (UMBC) and Daniel Zhou (LARC) for generating the EV training sets for AIRS and CrIS. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] Goody, R. M., J. Anderson, and G. North, "Testing Climate Models: An Approach", *Bull. Amer. Meteor. Soc.,* 79, 2541–2549 (1998).

[2] Huang, H.-L., and P. Antonelli, "Application Of Principal Component Analysis To High-Resolution Infrared Measurement Compression And Retrieval," *J. Appl. Meteor.,* 40, 365–388 (2001).

[3] Goldberg, M. D., Y. Qu, L. M. Mcmillin, W. Wolf, L. Zhou, and M. Divakarla, "AIRS Near-Real-Time Products And Algorithms In Support Of Operational Numerical Weather Prediction," IEEE *Trans. Geosci. Remote Sens.,* 41, 379–389 (2003).

[4] Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. and Susskind, J., "AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products and Processing Systems," *IEEE Trans. Geosci. Remote Sensing*, 41, 253-264 (2003).

[5] Glumb, R. J., Williams, F. L., Funk, N., Chateauneuf, F., Roney, A., and Allard,R., "Cross-track Infrared Sounder (CrIS) development status," Proc. SPIE 5152, (2003).

[6] Pagano, T. S., H. Aumann, R. Schindler, D. Elliott, S. Broberg , K. Overoye, M. Weiler, "Absolute Radiometric Calibration Accuracy of the Atmospheric Infrared Sounder," Proc SPIE 7081-46, (2008)

[7] Tobin, D. et al. "The Cross-track Infrared Sounder (CrIS) on SUOMI NPP: Intercalibration with AIRS, IASI and VIIRS," 93rd AMS Annual Meeting. Austin, TX, January (2013).

[8] Han, Yong et al. "Suomi NPP CrIS Measurements, Sensor Data Record Algorithm, Calibration and Validation Activities, and Record Data Quality," Special Issue of AGU JGR Atmospheres on Suomi NPP Cal/Val Science Results, (2014)

[9] Evan M. Manning, H. H. Aumann, and A. Behrangi, "AIRS Level-1C and applications to cross-calibration with MODIS and CrIS,",, Proc. SPIE "Earth Observing Systems XIX", (2014)

[10] Denis A. Elliott and H. H. Aumann,"Calibration and data analysis issues in modern infrared spectrometers using large data arrays," Proc. SPIE "Earth Observing Systems XX", (2015)