The Use of Principal Component Analysis (PCA) in Processing AIRS Data

Lihang Zhou[2], Mitchell D. Goldberg[1], Walter W. Wolf[2], Chris Barnet[1] and Yanni Qu[3]
[1]NOAA/NESDIS/ORA, Camp Springs, MD 20746; [2]QSS Group, Inc, Lanham, MD 20706; [3]ITT

1. INTRODUCTION

Atmospheric Infrared Sounder (AIRS) on the NASA EOS AQUA platform is providing much improved atmospheric temperature and moisture profiles when compared to soundings from current operational sounders. Principal Components Analysis (PCA) provides an effective way to approximate the AIRS observations. This poster presents the application of PCA to AIRS data, including the generation and application of eigenvectors, and the use of PCA for data compression, high noise channel detection and reconstruction, noise filtering and estimation, and regression retrieval**.**

2. METHOD

In our application the data vectors are radiance spectra that are divided by the expected instrumental noise, which we refer to as normalized radiances. The eigenvectors are related to the covariance matrix by:

$$S = E \Lambda E^T$$

where S is a N x N covariance matrix. E is the matrix of eigenvectors, and Λ is the diagonal matrix of eigenvalues. The principal component scores P are computed from:

$$P = E^T R;$$

where R is the vector of centered normalized radiances.

An overall estimate of how well the principal component scores can reconstruct the original data is provided by Reconstruction Score (RS), which is defined as:

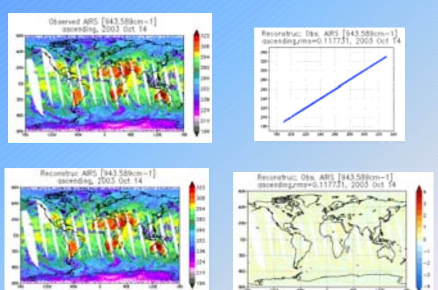
$$RS = \left[1/N \sum_{i=1}^{N} (O_i - R_i)^2 \right]^{1/2}$$

i=1…N channels, O and R are noise scaled observed and reconstructed radiances for ith channel.

Application 1: Reconstruct Radiance

- Our studies indicate that AIRS brightness temperature can be reconstructed within the noise level, using about 80 eigen-vectors.

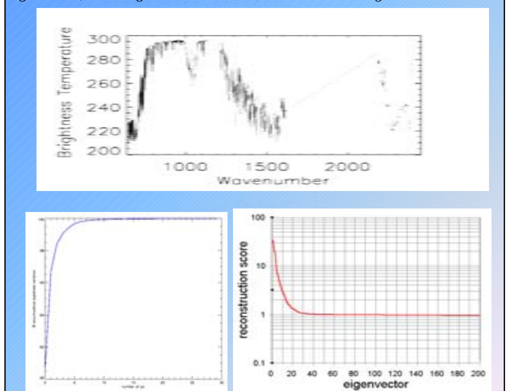
Figure 2: Example of Reconstructed and Observed Brightness Temperature for an AIRS window channel



3. APPLICATION

- We use the PCA to generate reconstructed radiances and to compute the RMS between the reconstructed radiances and the observed radiances.
- It appears for most cases that the RMS is comparable with the instrument noise level. It implies that the AIRS observations can be reconstructed very accurately by much less number of PCS.

Fig 1 Top: A typical AIRS Spectrum; Lower left: Explained Variance vs. number of eigenvectors; Lower right: Reconstruction score vs. number of eigenvectors



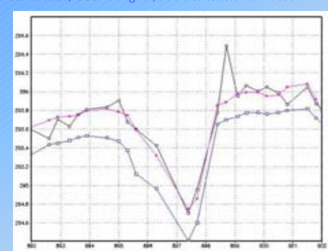
4. CONCLUSION

- Principal component regression of near real time AIRS data provides a very computational efficient retrieval of atmospheric temperature, moisture, and ozone, etc.
- Principal component scores have been a very effective tool for data compression, detector monitoring and bad channel handling, as well as the noise estimation.

Application 3: Noise Estimate/Filtering

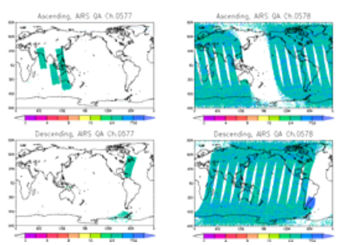
Figure 3: shows impact of reconstructing radiance. This is a expanded view from a spectrally boring region 892 to 902. A retrieval would benefit from operating on the difference between purple and blue, instead of black and blue.

Purple is reconstructed, black is original, blue is calculated from model



Application 2: Channel Monitoring and Bad Channel Handling

Indx	freq.	Nedt	aveob/avere	s<1 asc	s<1 desc	1<s<1.5 a	1<s<1.5 d	1.5<s<2 a	1.5 <s<2 d	s>2 a	s>2 d
418	[illegible]	0.296	[illegible]	0.226	0.256	0.211	0.223	0.250	0.242	0.425	0.447
419	[illegible]	0.285	[illegible]	0.300	0.300	0.214	0.218	0.275	0.271	0.285	0.298
420	[illegible]	0.349	[illegible]	0.330	0.328	0.298	0.311	0.321	0.333	0.414	0.433
421	[illegible]	0.214	[illegible]	0.218	0.221	0.294	0.309	0.325	0.256	0.508	0.507
422	[illegible]	0.338	[illegible]	0.376	0.382	0.396	0.396	0.487	0.508	0.599	0.616
423	[illegible]	0.469	[illegible]	0.585	0.580	0.658	0.628	0.814	0.796	0.745	0.788
424	[illegible]	0.281	[illegible]	0.277	0.278	0.255	0.262	0.278	0.281	0.403	0.428
425	[illegible]	0.426	[illegible]	0.437	0.441	0.397	0.405	0.391	0.422	0.465	0.455
426	[illegible]	0.296	[illegible]	0.344	0.315	3.346	0.408	9.860	4.562	15.296	5.227
427	[illegible]	0.305	[illegible]	2.204	2.211	5.578	6.470	10.692	14.757	22.821	24.518
428	[illegible]	0.392	[illegible]	0.379	0.379	0.351	0.375	0.393	0.440	0.605	0.643
429	[illegible]	0.482	[illegible]	0.195	0.198	0.226	0.212	0.225	0.262	0.428	0.435
430	[illegible]	0.542	[illegible]	0.592	0.589	0.595	0.590	0.683	0.641	0.691	0.804
431	[illegible]	0.295	[illegible]	0.289	0.291	0.267	0.276	0.312	0.329	0.408	0.423
432	[illegible]	0.498	[illegible]	0.491	0.495	0.453	0.455	0.482	0.508	0.622	0.627
433	[illegible]	0.287	[illegible]	0.292	0.291	0.264	0.274	0.296	0.309	0.295	0.358
434	[illegible]	0.418	[illegible]	0.412	0.412	0.378	0.386	0.405	0.399	0.652	0.529



- In operational AIRS processing, when a channel has a bad radiance value, it has to be replaced by something useful since information from each channel is required to reconstruct the radiances

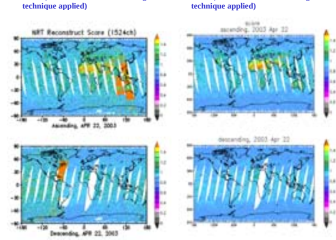
- Bad channel handling technique: when a bad channel is detected, use the neighboring channels to compute the pcs, and reconstruct the bad channel.

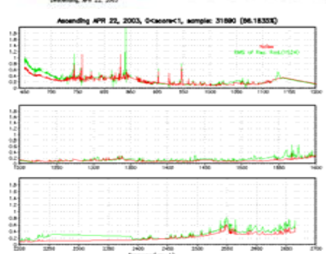
Quality Flag Definitions:

0 – Good Quality Data; 2 – Reserved; 4 – Undesirable Channel Properties; 8 – Bad Offset for Channel ; 16 – Bad Gain for Channel; 32 – Pop Detected in Channel; 64 – Bad Telemetry ; 128 – High Noise in Channel; 256 – The Instrument was not in science mode; 512 – No data for this channel

Score (before bad channel handling technique applied)

Score (after bad channel handling technique applied)



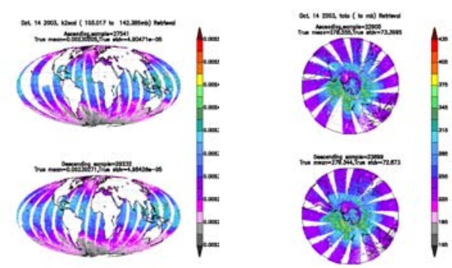


5. FUTURE WORK

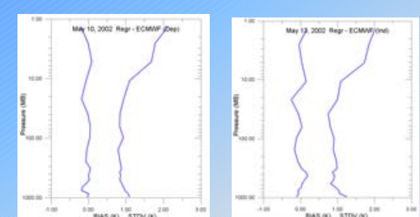
- The knowledge that we are gaining from the AIRS data will be useful for high spectral infrared radiance data processing and compression for future satellite observations.
- Additional work to study noise filtering feature of eofs is needed.
- To create a stable PCA operational system using as many AIRS channels as possible.
- Use most correlated channels instead of neighboring channels for bad channel reconstruction.

Application 4: Regression Retrieval

- Use the first 85 PCS as predictors, generate regression retrievals of atmospheric temperature, moisture, ozone, and skin temperature
- Regression coefficients generated from June have been stable.



Statistics from pcs regression retrieval: Left: May 10 2003 (dependent), Right May 13 2003 (independent)



For more information, please visit
http://orbit35i.nesdis.noaa.gov/crad2/airsgrid/AIRS_NRT

International TOVS Study Conference, 13th, TOVS-13, Sainte Adele, Quebec, Canada, 29 October-4 November 2003. Madison, WI, University of Wisconsin-Madison, Space Science and Engineering Center, Cooperative Institute for Meteorological Satellite Studies, 2003.