The Use of Principal Components Analysis (PCA) in Processing Simulated IASI Data

Zhaohui Cheng[1], Lihang Zhou[1], Thomas King[1], Walter W. Wolf[1], Haibing Sun[1], Wen Zhou[1], Chris Barnet[2], Mitchell D. Goldberg[2]

[1]QSS Group, Inc, Lanham, MD 20706; [2]NOAA/NESDIS/STAR, Camp Springs, MD 20746

Abstract

● Principal component analysis (PCA) is a useful technique in analyzing high spectral infrared radiance data (such as AIRS, IASI) due to the high correlation among the different spectral channels. Currently NOAA NESDIS STAR uses PCA to process the simulated IASI data for the data monitoring, quality control and regression retrievals of geophysical parameters.

● Eigenvectors and the corresponding reconstruction scores have been computed for all 8461 IASI channels as well as the three IASI bands. Large reconstruction errors can be used to identify the suspicious channel/band. A web page was built to monitor these reconstruction errors in both near real-time and offline.

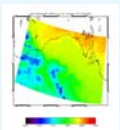
● The AIRS regression retrieval algorithm has been used for processing IASI simulated data. The STAR uses 85 principal components (still retaining most of the variability of the information of the original data. Goldberg et al., 2003) in the regression for atmospheric temperature, moisture, ozone, surface temperature and surface emissivity. A clear field of view detection algorithm is being developed to select the best regression retrievals with good global distribution.

PCA Processing

- The algorithm to generate and apply the IASI eigenvectors is the same as AIRS.
- Simulated IASI data training set

  For each granule we use:

  * 2 scan lines
  * 4 IASI FOV
  * 8461 channels



- Computed Eigenvectors for all 8461 channels
- For easy computation, divided the 8461 channels into three bands:

  band 1: 2261 channels 645cm-1 ~ 1210cm-1

  band 2: 3160 channels 1210.25cm-1 ~ 2000cm-1

  band 3: 3040 channels 2000.25cm-1 ~ 2760cm-1
- Computed 200 principal components for each band.
- Reconstruct the radiances by using principal components.
- Compute the reconstruction scores.

PCA Application

- Data Monitoring and Quality Control
- Regression retrievals of geophysical parameters

  Atmosphere temperature, moisture and ozone

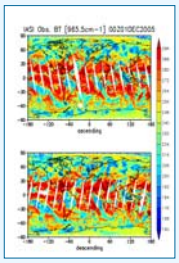
  Surface temperature and emissivity

Data Monitoring and Quality Control

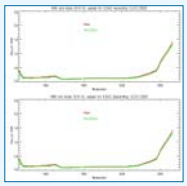
- Near Real-Time – Granule Data Monitoring (Granule data comes in about every 3-minute and there are about 491 Granules/day)
- Offline data monitoring – Global Grids Data Monitoring

Global Grids Dataset: 0.5 deg lat x 2 deg long 616-channel Single IASI FOV

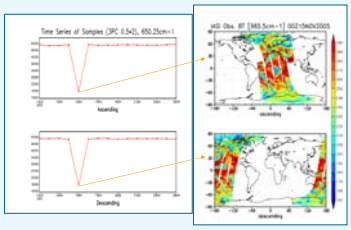
Global Map (Daily Monitoring)



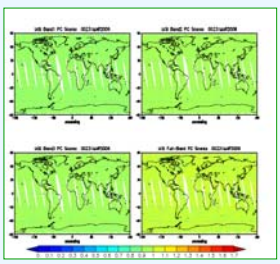
Comparison of Reconstructed RMS with Noise (Daily Monitoring)



Weekly Updated Statistics



Monitoring of Reconstruction Scores for each band and all channels. Large reconstruction error can be used to identify suspicious data



Clear Field of View Detection

Algorithm:

- The brightness temperatures (BT) of AMSU channels 4, 5 and 6 are used to predict IASI BT at 2390cm^{-1} (lower tropospheric infrared channel)
- **The predicted IASI BT = 44.007-0.272*amsu4+1.837*amsu5-0.738*amsu6+0.004*csza+6.391*(1-cview)**

  Where csza: cosine solar zenith angle

  cview: cosine view angle (scan angle)

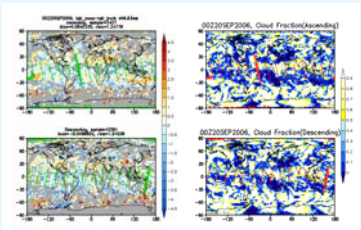
  amsu4, amsu5 and amsu6 are AMSU BTs of channel 4, 5 and 6
- The coefficients are obtained by the regression AMSU data against the simulated IASI data under clear condition.
- FOV is labeled "mostly clear"

  * if predicted IASI – observed IASI < 2
  * if the standard deviation of the 2390cm^{-1} radiance (4 fovs within AMSU fov) is less than 3 times the instrument noise level (3*0.31).

Test Results:

Global map of bias between regression retrieval of temperature using AMSU clear test and truth at 496.63 hPa vs. cloud fraction distribution

AMSU Clear Test **Cloud Fraction**



Applying clear detection algorithm can:

- screen out the clouds with large cloud fraction
- reduce the error of the regression
- keep relatively good global distribution (the number of samples is twice more than that of samples of the cloud fraction less than 0.1)

Regression Retrieval

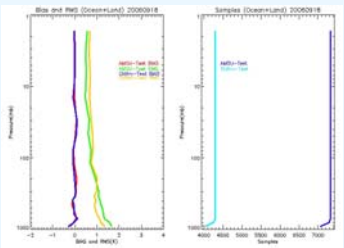
The simulated noisy radiances are regressed against geophysical truth.

$$G(L,k) = a(0) + \Sigma_{n=1,N}\, a(n)\cdot R(n,k)$$

- G(L,k) is geophysical truth.
- a(0)…a(n) are regression coefficients.
- L=1,N_L and k=1,K: L is the number of atmospheric levels and K is the number of individual cases.
- R(n,k) are the predictors. N is number of the predictors.
- For temperature, moisture and ozone regression retrieval, we use 85 principal component scores and other two view angle related items as the predictors

Statistics of regression retrieval (AMSU clear test vs. cloud fraction<0.1) Total samples: 17160

Bias and RMS **Samples**



Summary and Future Plan

- A data validation system has been built using principal components (PC) for IASI data monitoring and quality control.
- To get more accurate regression retrievals, a clear detection algorithm has been developed to remove the fovs with large cloud fraction.
- Based upon the AIRS algorithm, the PC regression had been performed for IASI simulated dataset.
- Will create post launch eigenvector set for IASI data to produce a stable eigenvector set.
- Will continue to test clear detection algorithm and the regression retrievals of the geophysical parameters.