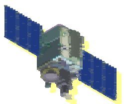
PCA based Noise Filter for High Spectral Resolution IR Observations

P. Antonelli, H. E. Revercomb, D. C. Tobin, L. Sromovsky, H. B. Howell, R. K. Garcia, A. Huang, F.A. Best, S. Dutcher

and

W. L. Smith

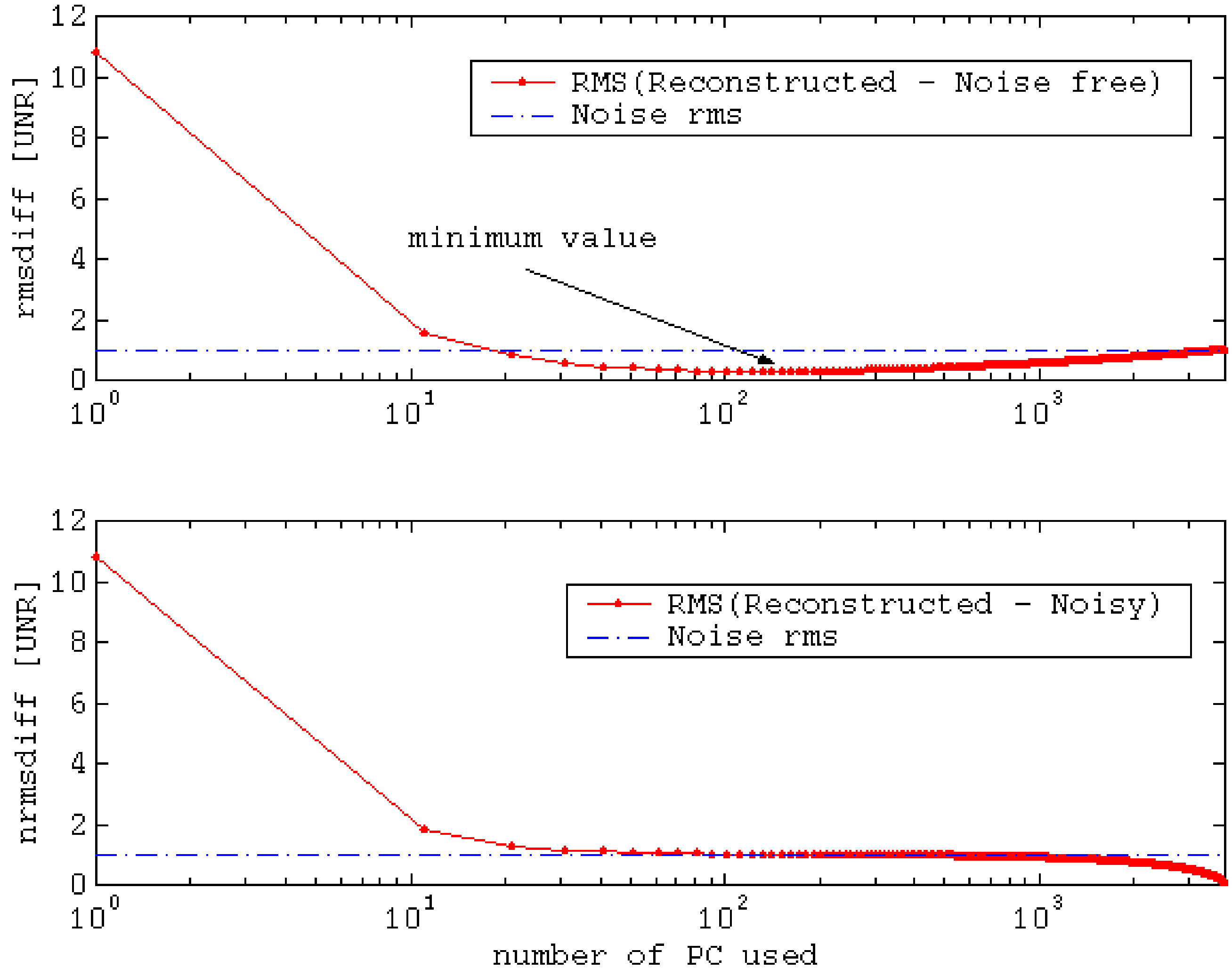
Outline

- Operational Pca based Noise Filter design (PNF)
  - Theory and Implementation
    - noise normalization, derivation of the principal components
- Application of PNF to high spectral resolution data
  - Simulated
    - Noise Reduction and Information Loss
    - relevance of Noise Normalization and Training Data representativeness
  - Aircraft (NAST-I, S-HIS) 
    - impact on rarely occurring observations (outliers)
    - relevance of Noise Normalization and Training Data Representativeness
  - Satellite (AIRS) 
    - Noise estimation
    - Noise Filtering

PNF Theory: PCA

- is a multivariate analysis technique commonly used to reduce the dimensionality of data-sets with large numbers of interdependent variables
- essentially performs a Singular Value Decomposition of the Covariance Matrix of the observations
- maps the spectrally correlated observations into de-correlated quantities (PC coefficients or Compressed Data)

Why PCA for Noise Filtering?

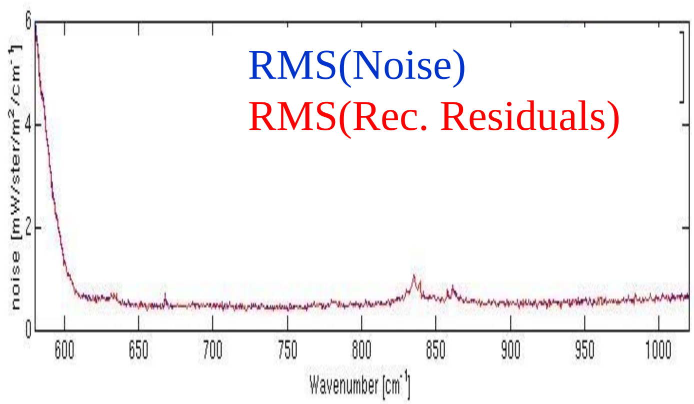


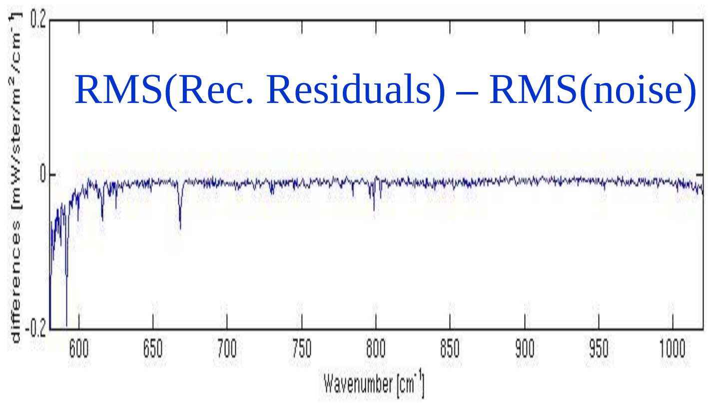
PNF Implementation: design

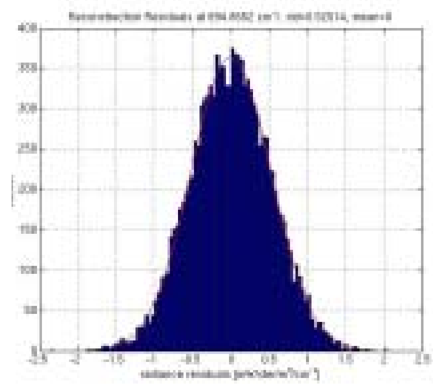
- Collect large set of spectra that has to be filtered
- Normalize spectra using initial noise estimate (divide by NeN)
- generate PCs of normalized spectra in dependent mode
- compress and reconstruct normalized spectra using reduced number of PCs
- remove normalization (multiply spectra by initial noise estimate)

Reconstruction Residual Properties

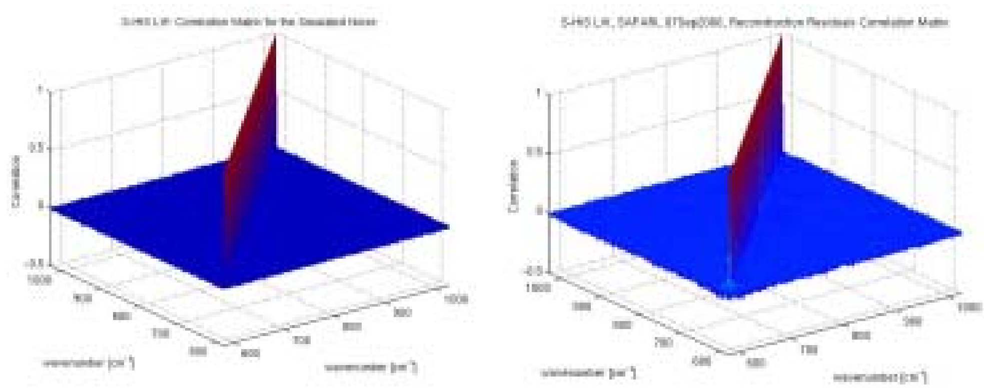
Rec. Residual RMS





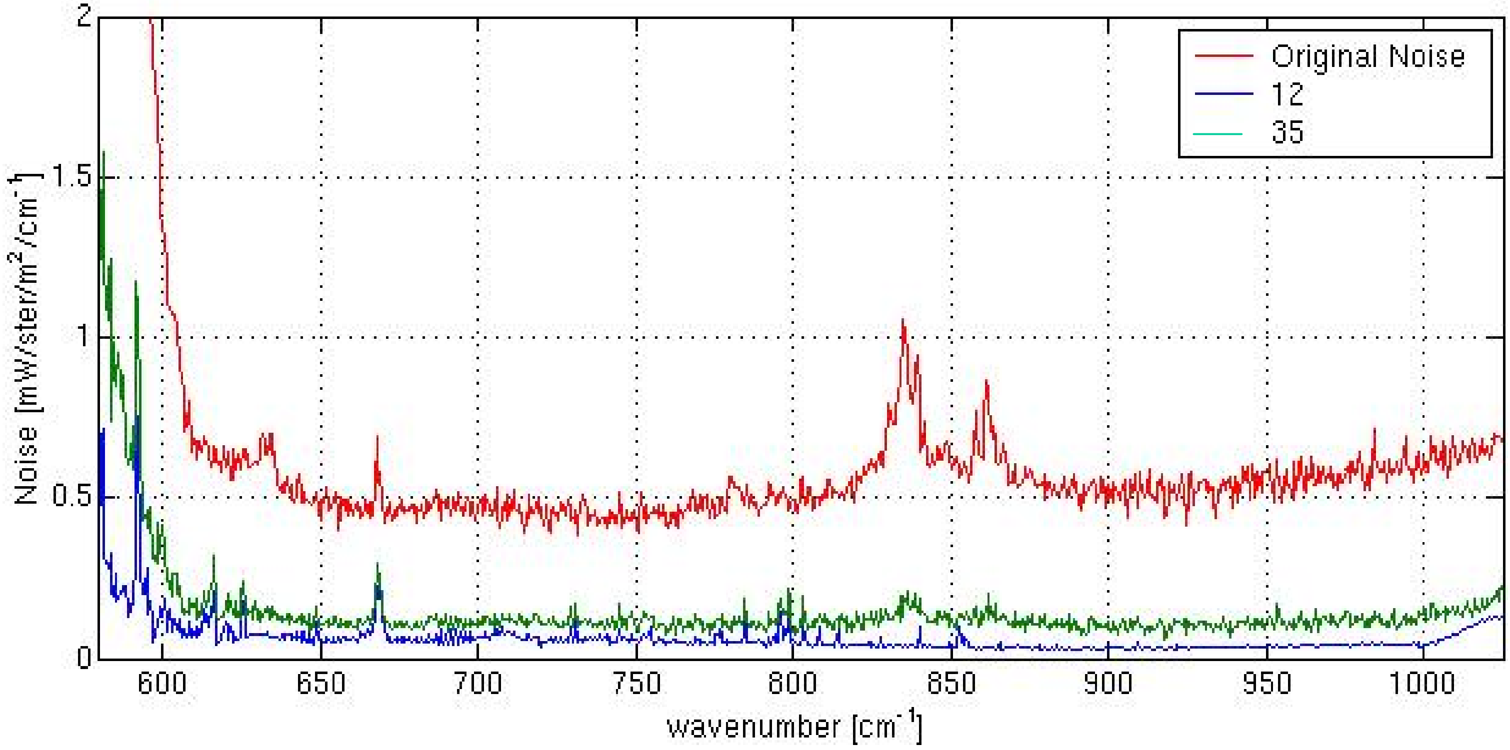


Single Channel Rec. Residual Distribution



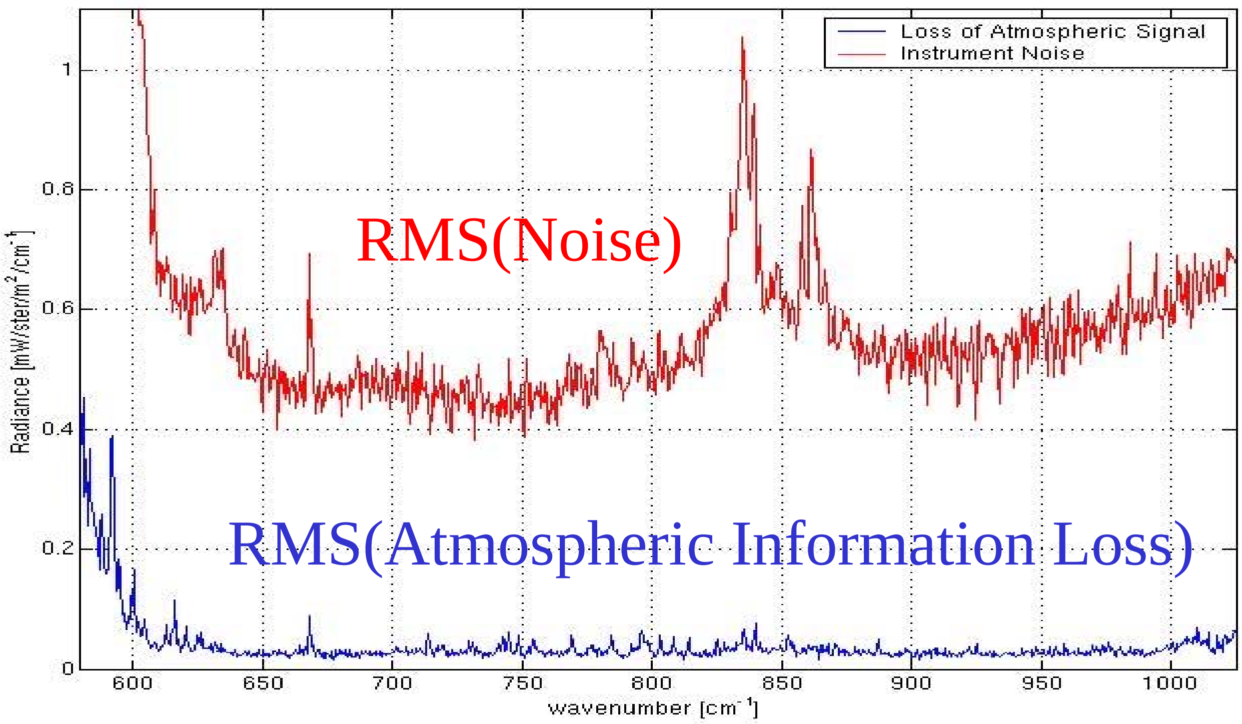
Rec. Residual Correlation Matrix

RMS Unfiltered Noise



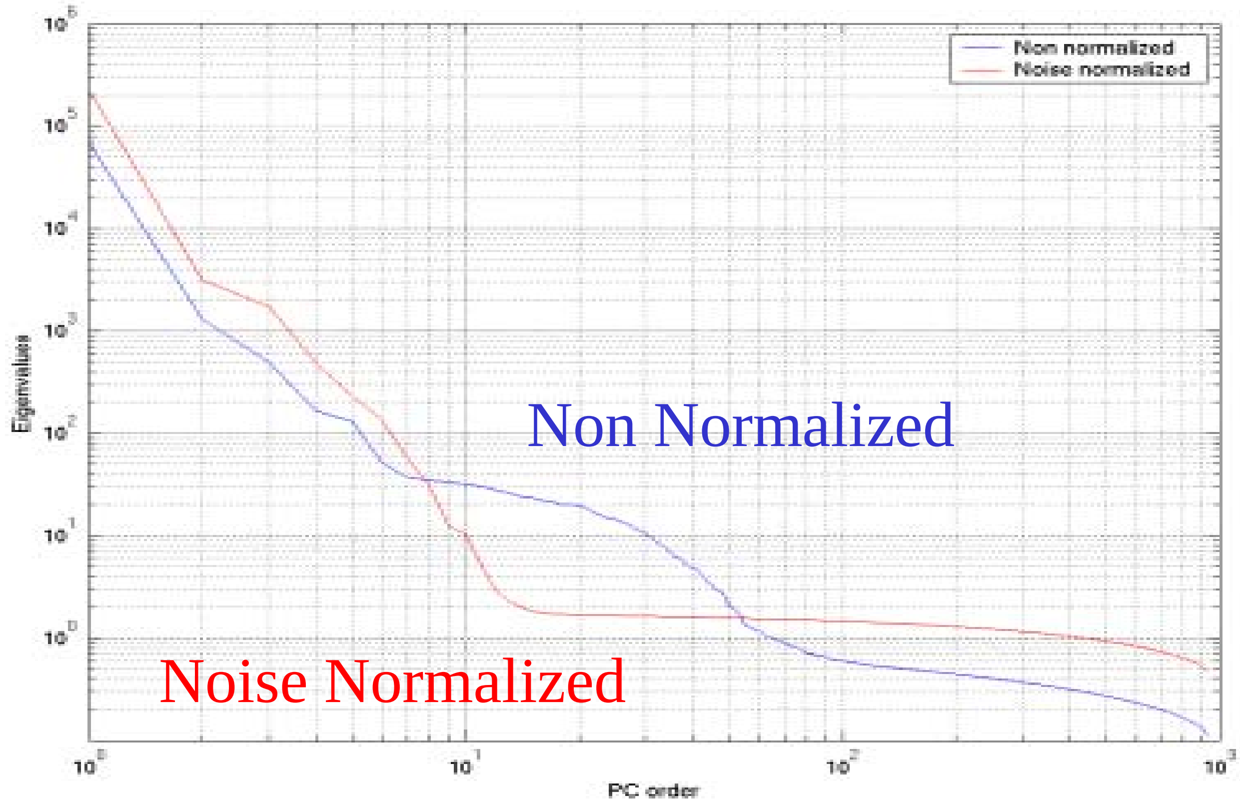
Unfiltered Noise=PCA(Original Noise)

RMS Atmospheric Information Loss



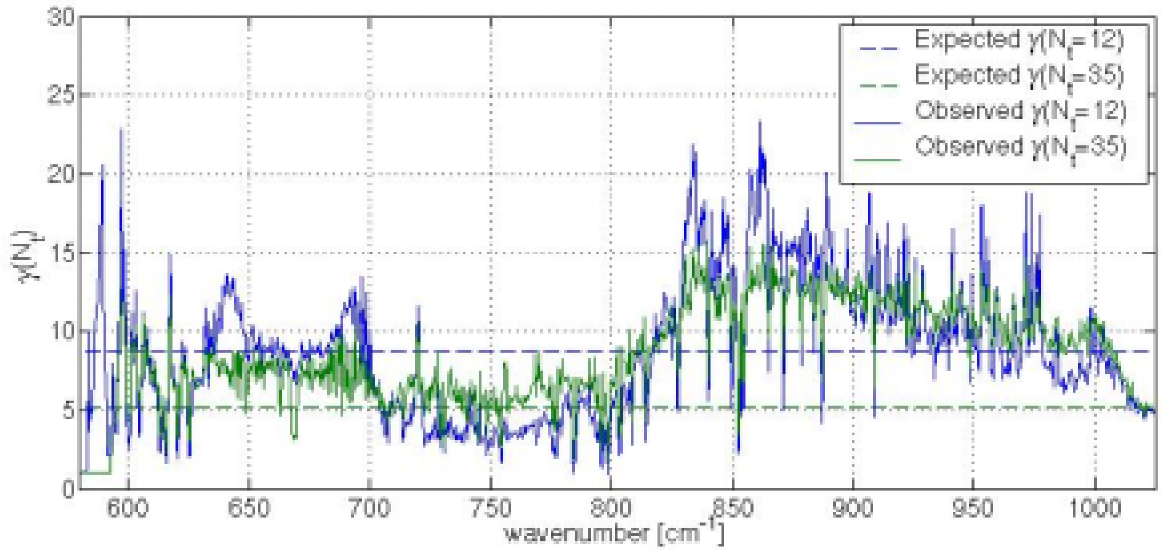
Atmospheric Information Loss =
Original Noise Free Obs – Filtered Noisy Obs

Noise Normalization



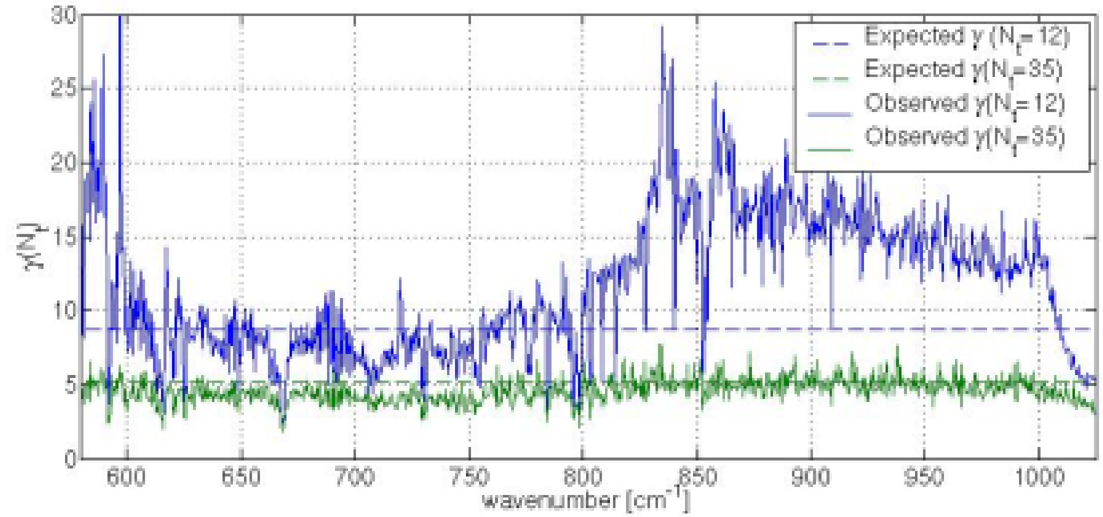
NRF=sqrt(N_c/N_t)

Noise Reduction Factor

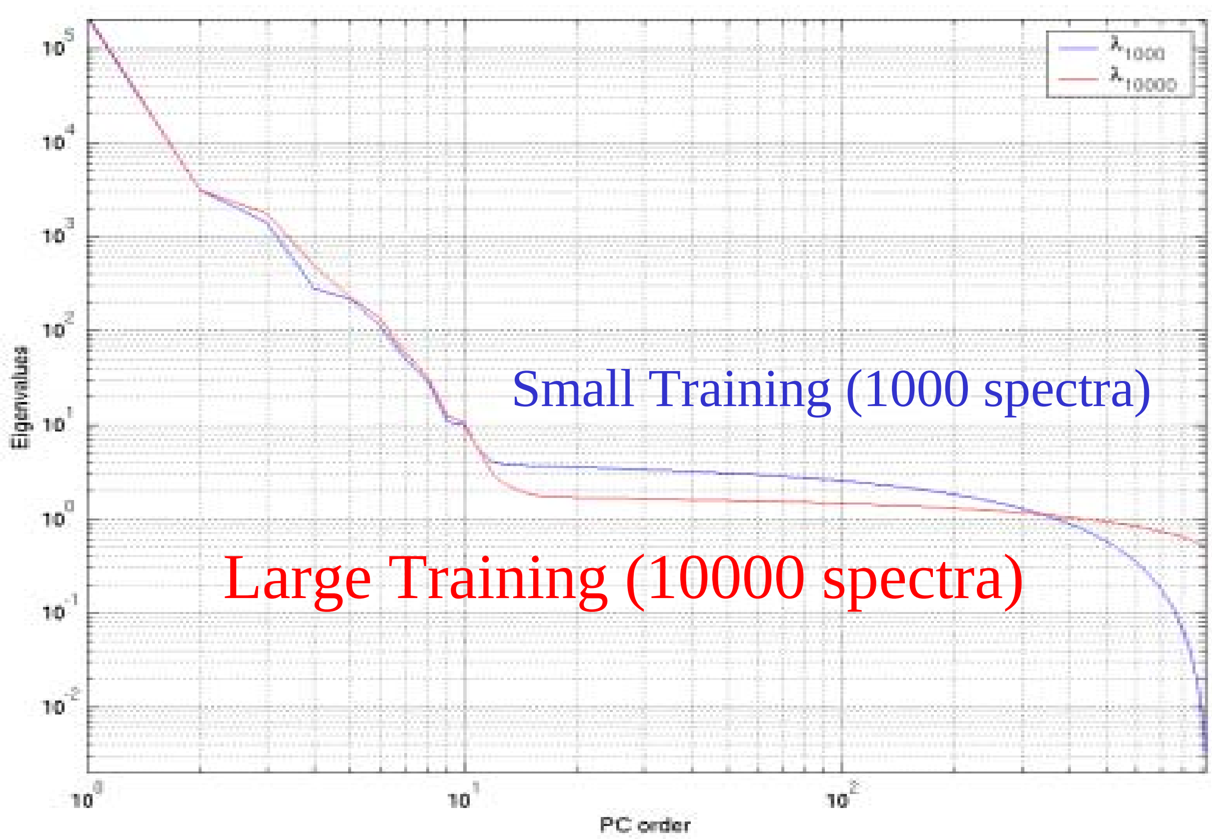


Non Normalized

Normalized



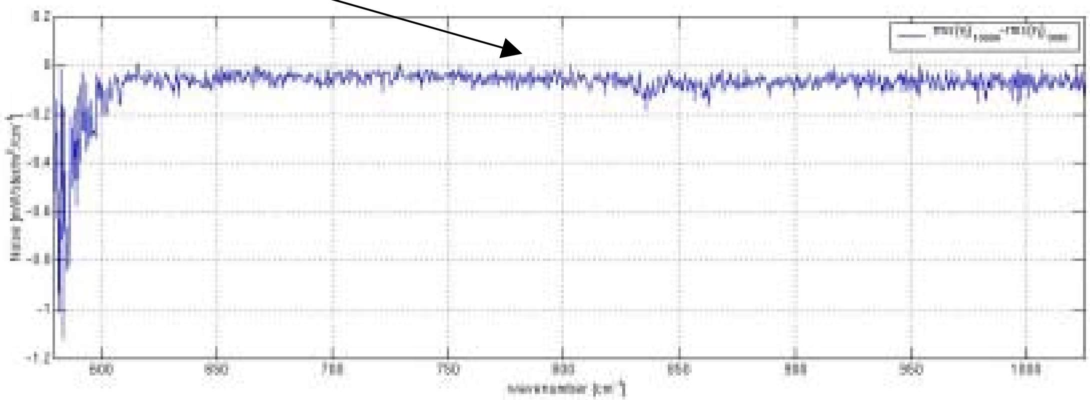
Training Set Size Study



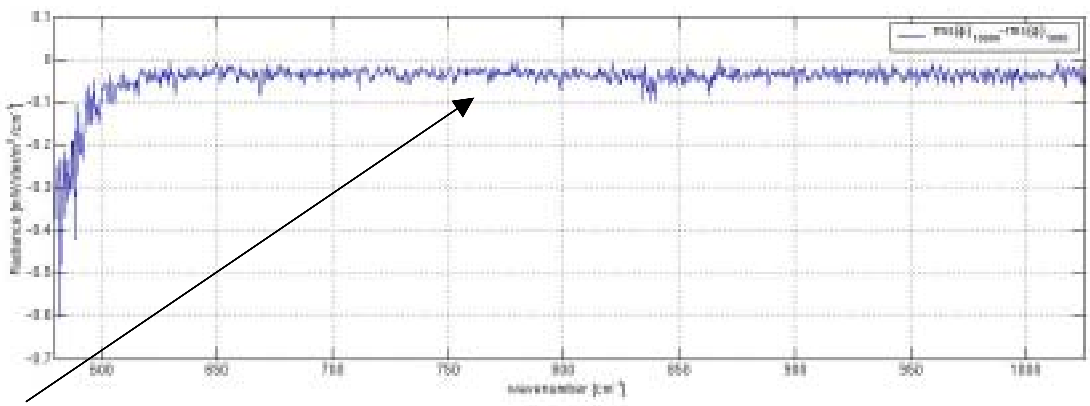
Large Samples

The smaller the training Set
The larger the reconstructed noise

RMS(Rec. Noise 10000) – RMS(Rec. Noise 1000)



The smaller the training Set
The larger the Atm. Info Loss



RMS(Atm. Info Loss 10000) – RMS(Atm. Info Loss 1000)

PNF on simulated Data: Conclusions

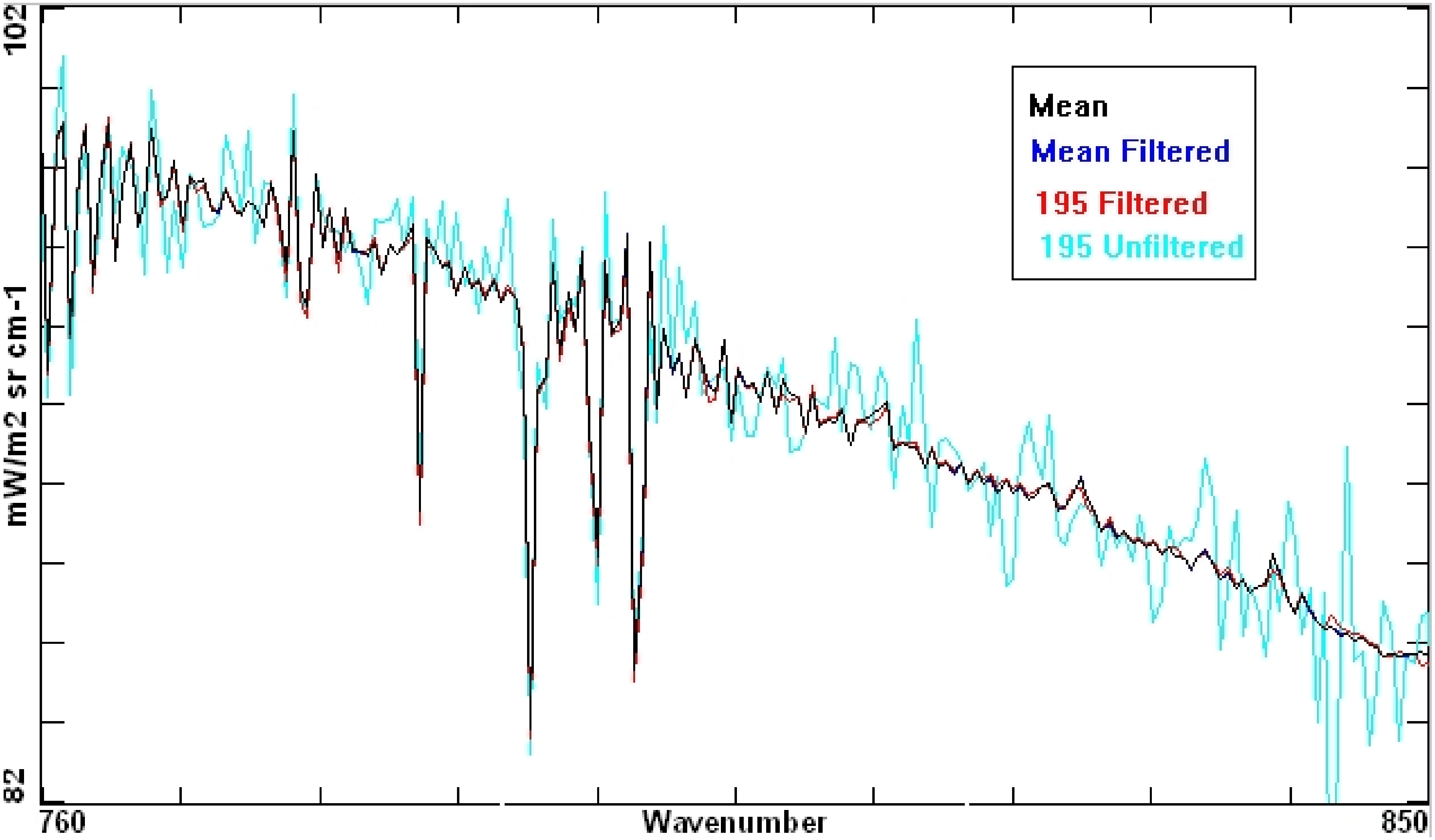
- Reconstruction Residuals show the same spectral properties of Simulated Random Component of Noise
- Noise Reduction Factor, NRF, (applying noise normalization) can be estimated as the square root of the ratio N_c/N_t (for 75 PCs and 2700 channels NRF=6)
- Atmospheric Information Loss can be estimated about 10 time smaller the removed noise (about .04% of atmospheric signal)

PNF on Real Data

- Relevance of Representativeness
- Impact of PNF on rarely occurring observations
- Information Loss: comparison between Line by Line Models and filtered observations
- Effects of PNF on (statistical) Retrievals

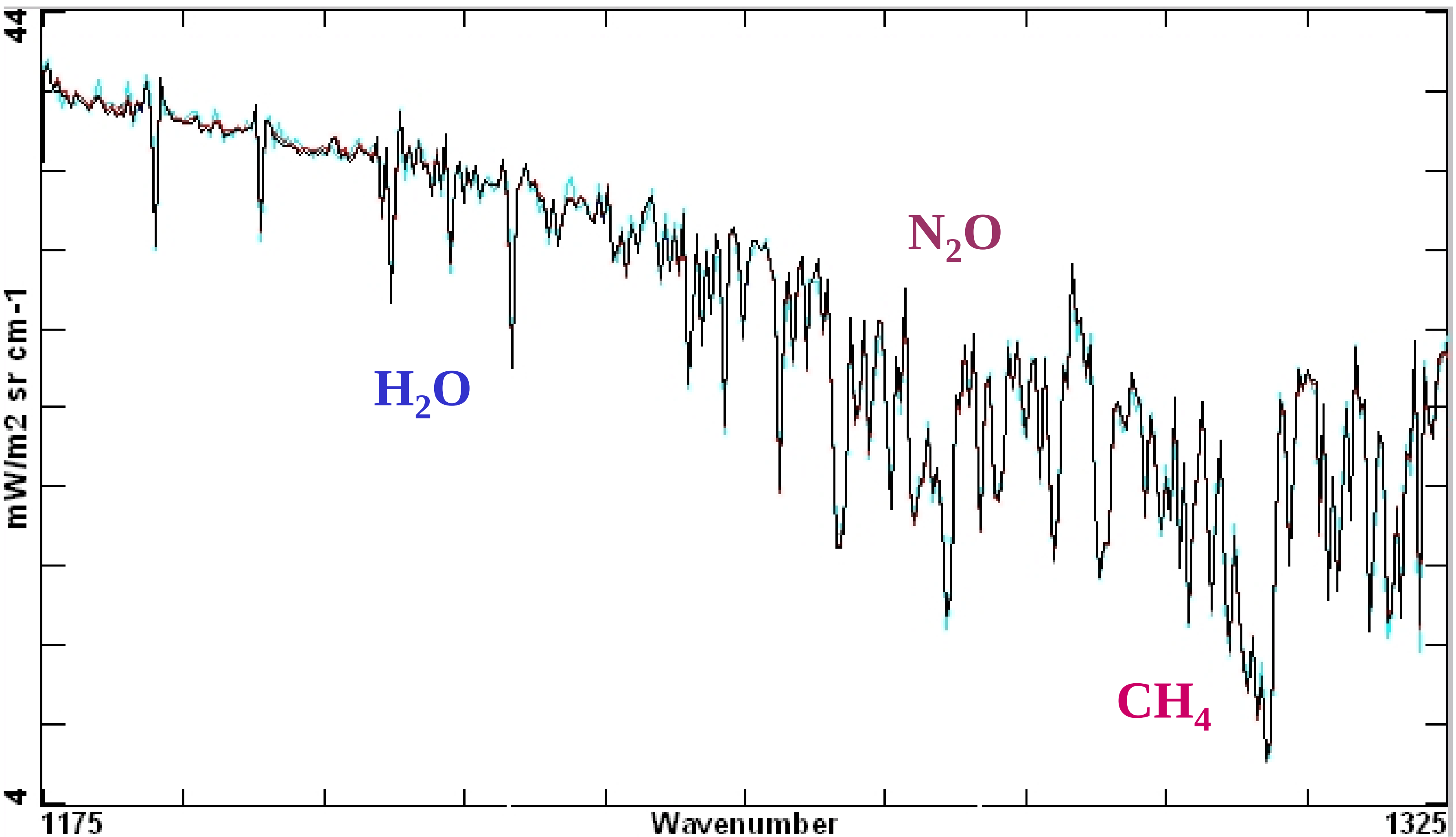
S-HIS PCA-Filtered Radiances

Longwave Window Region (760-850 cm^{-1})



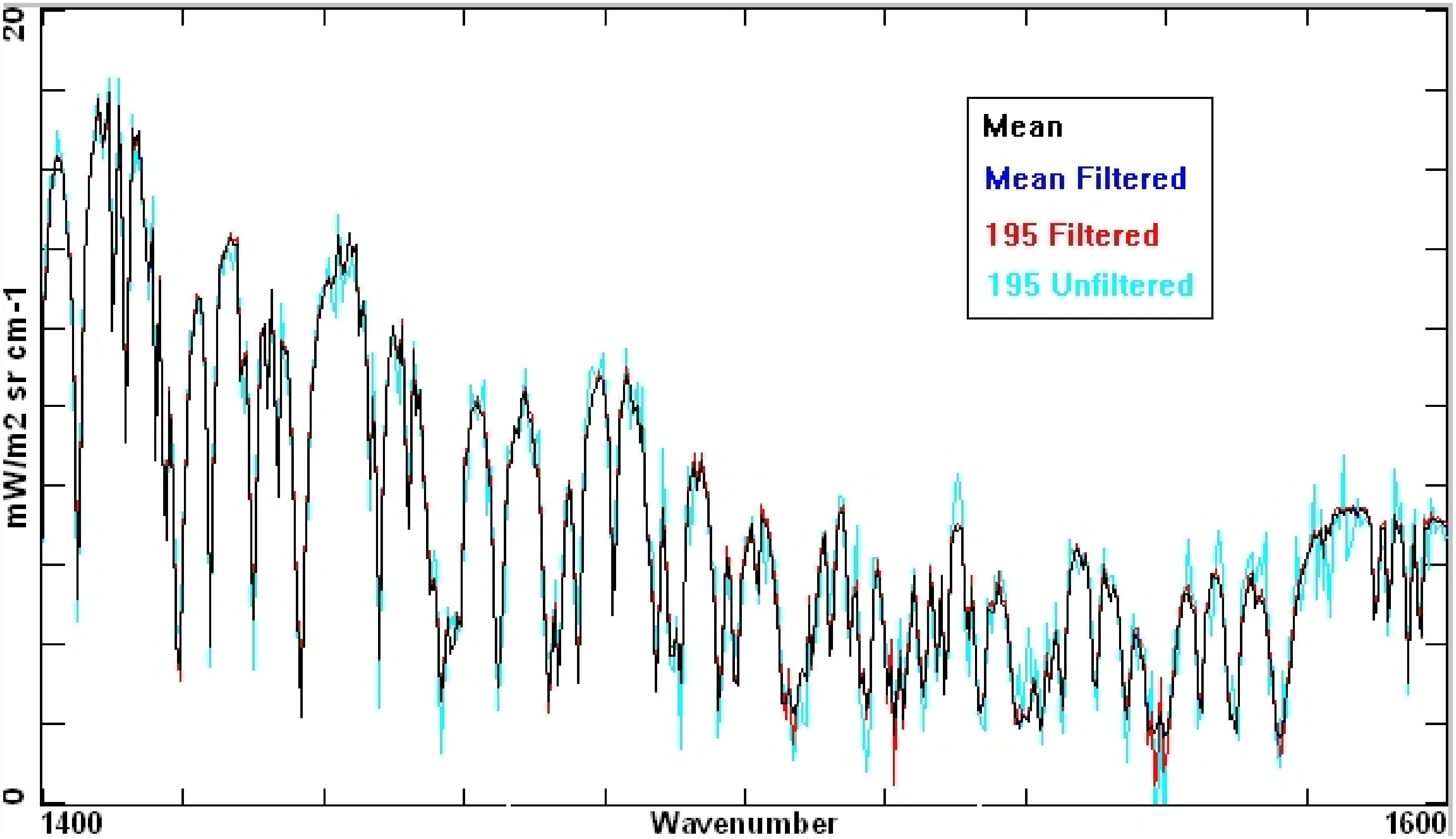
S-HIS PCA-Filtered Radiances

Midwave Window to Methane (1175-1325 cm^{-1})



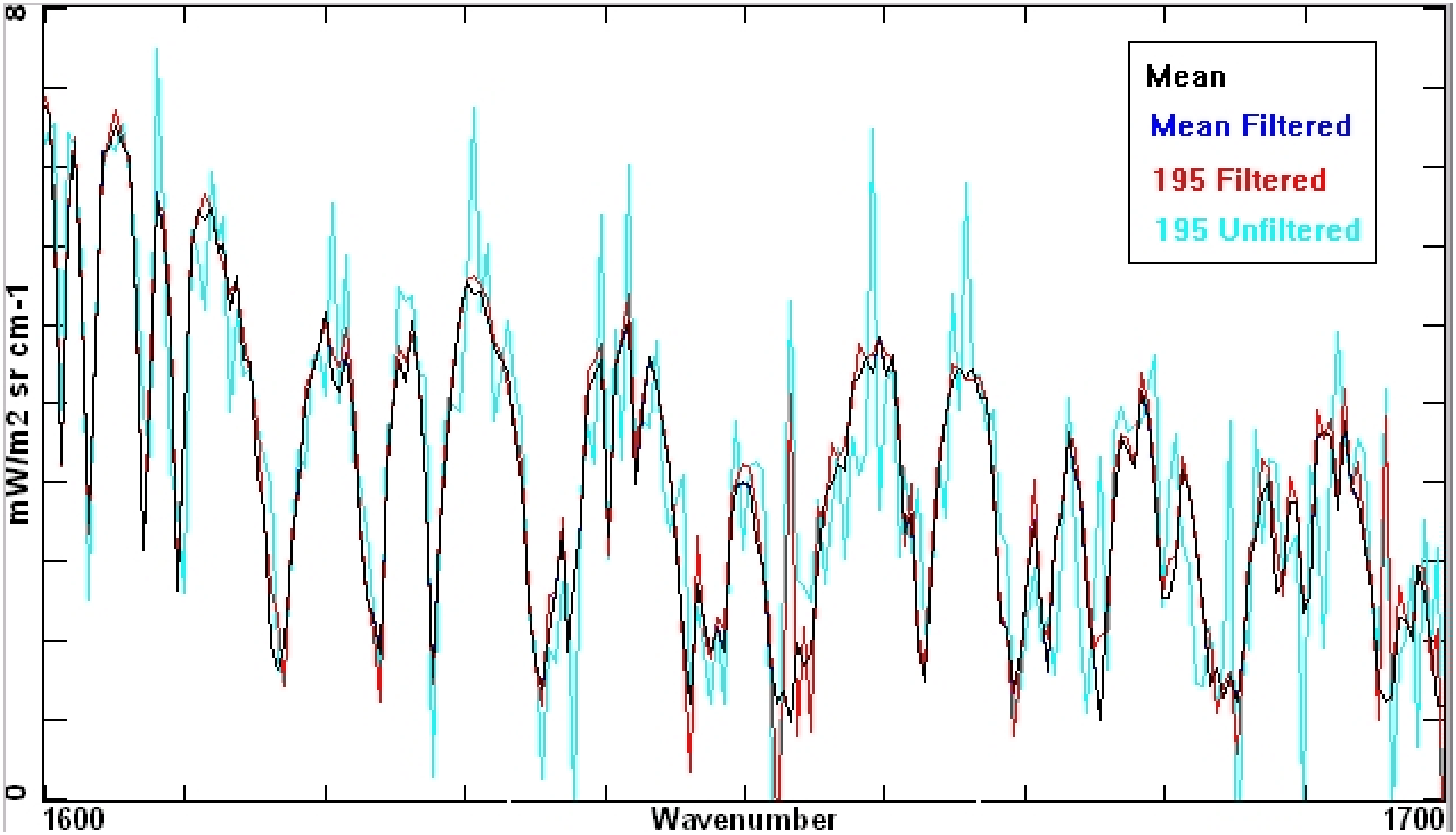
S-HIS PCA-Filtered Radiances

6.3 micron Water Vapor (1400-1600 cm^{-1})

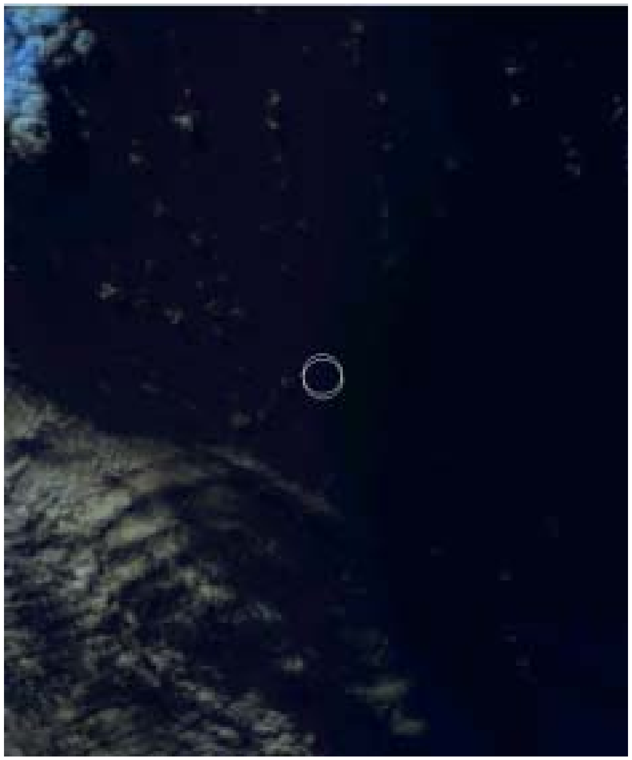


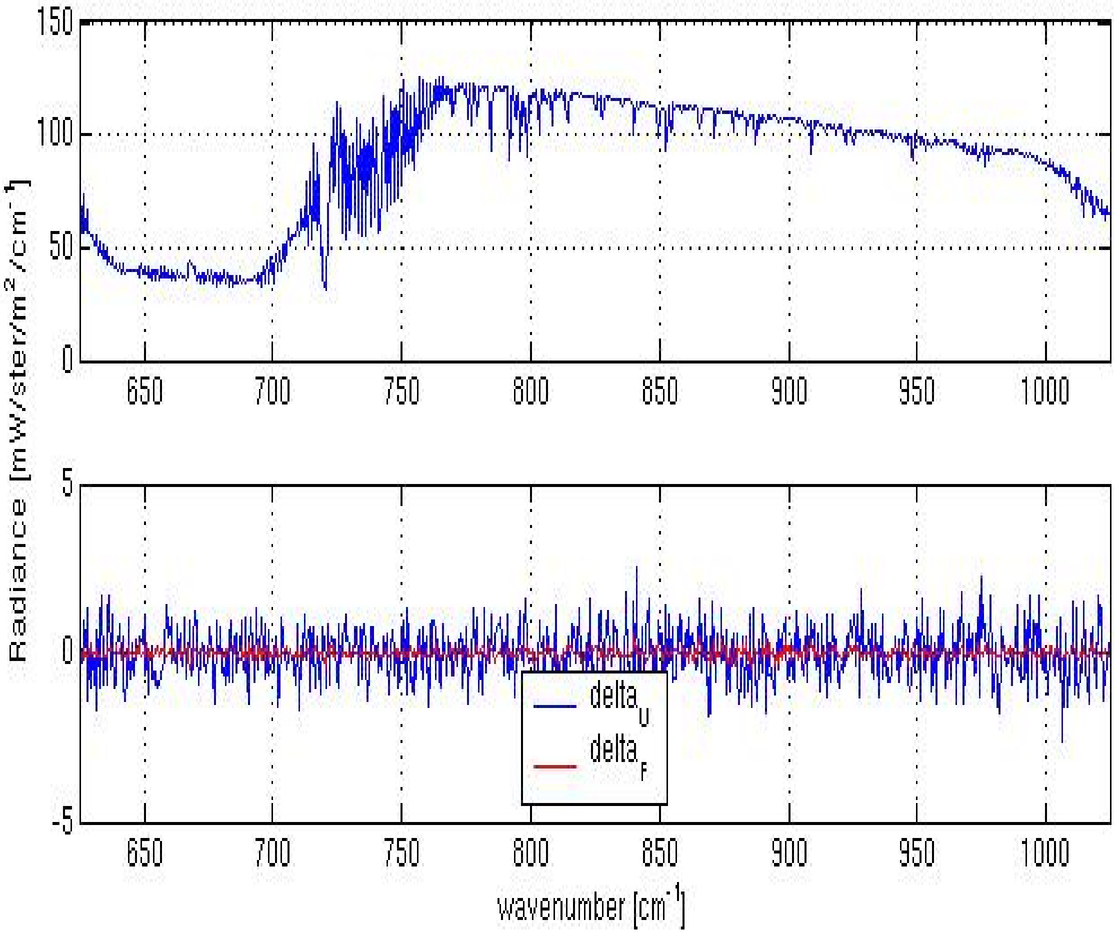
S-HIS PCA-Filtered Radiances

6.3 micron Water Vapor (1600-1700 cm^{-1})

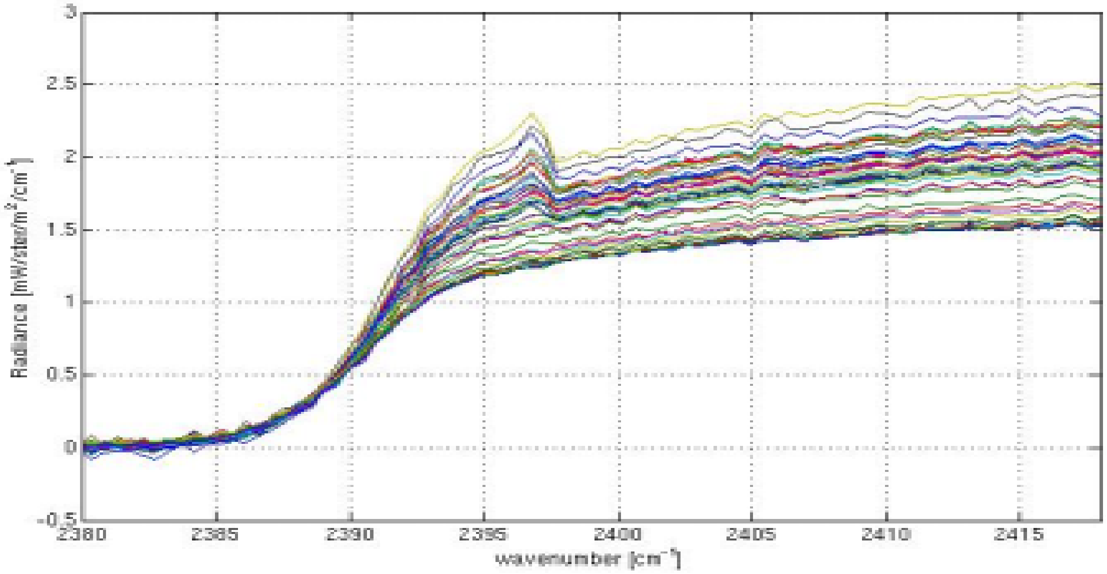


S-HIS: Noise Filter





S-HIS: Noise Filter

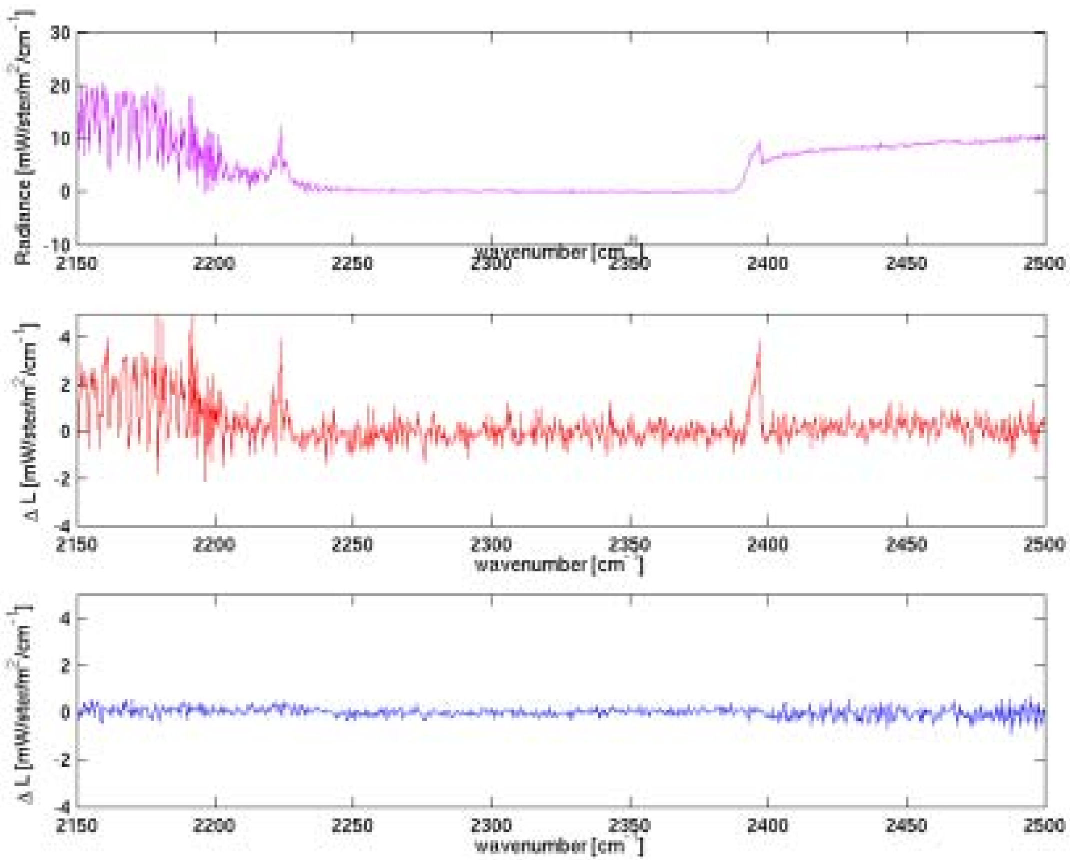


Training representativeness and rarely occurring observations

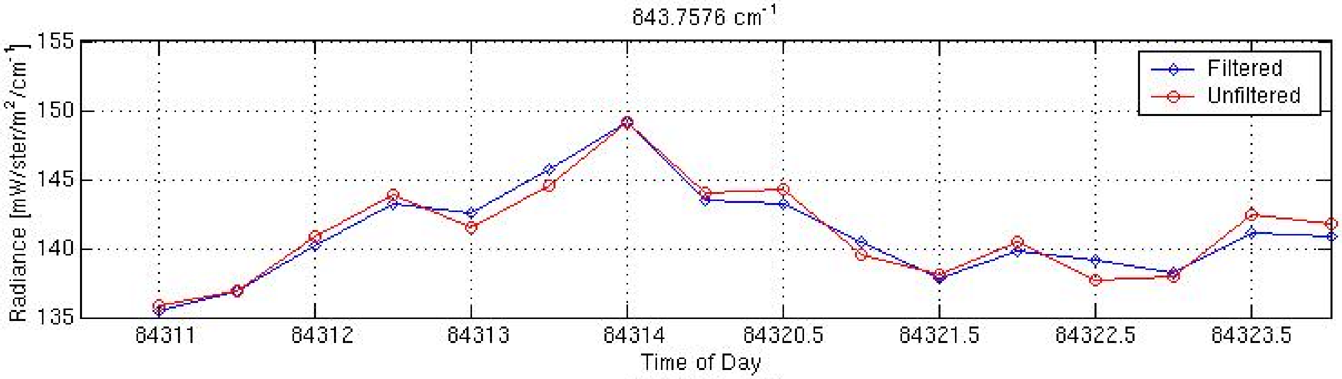
Original

Original-Filtered (A)

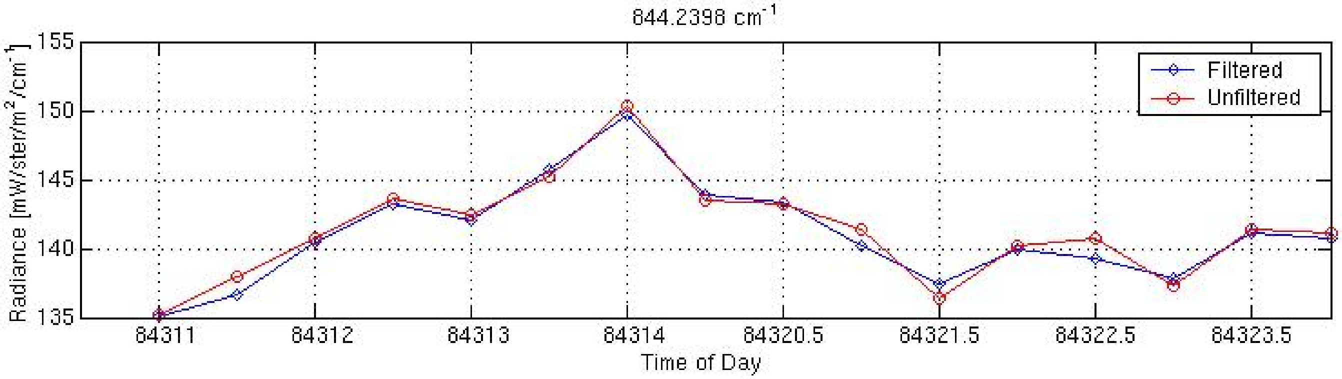
Original-Filtered (B)



More on the Fires

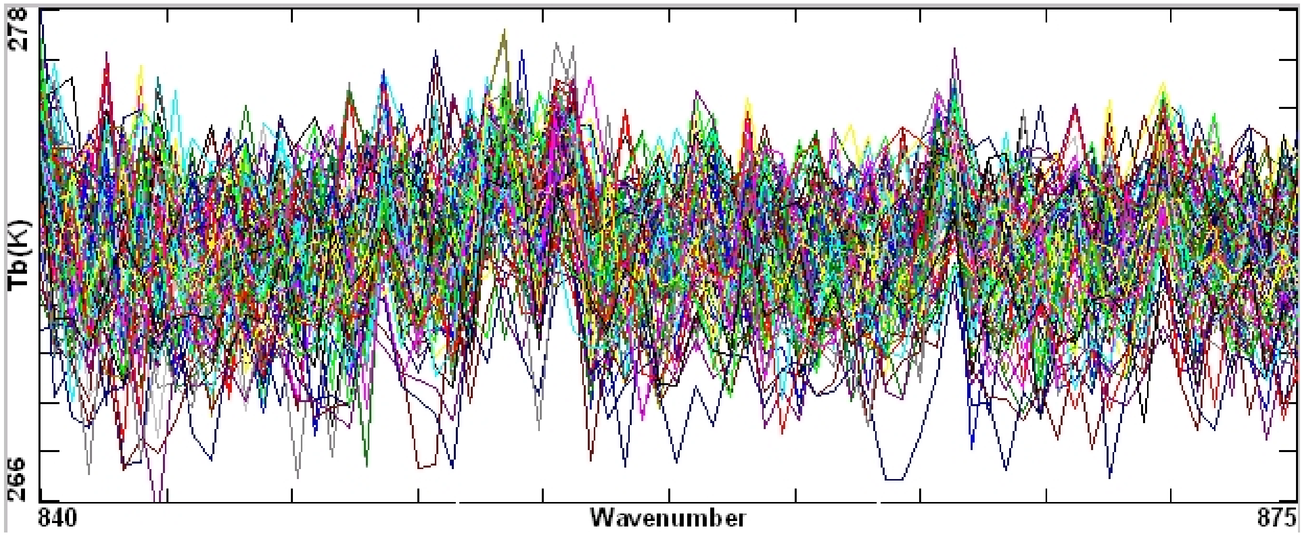


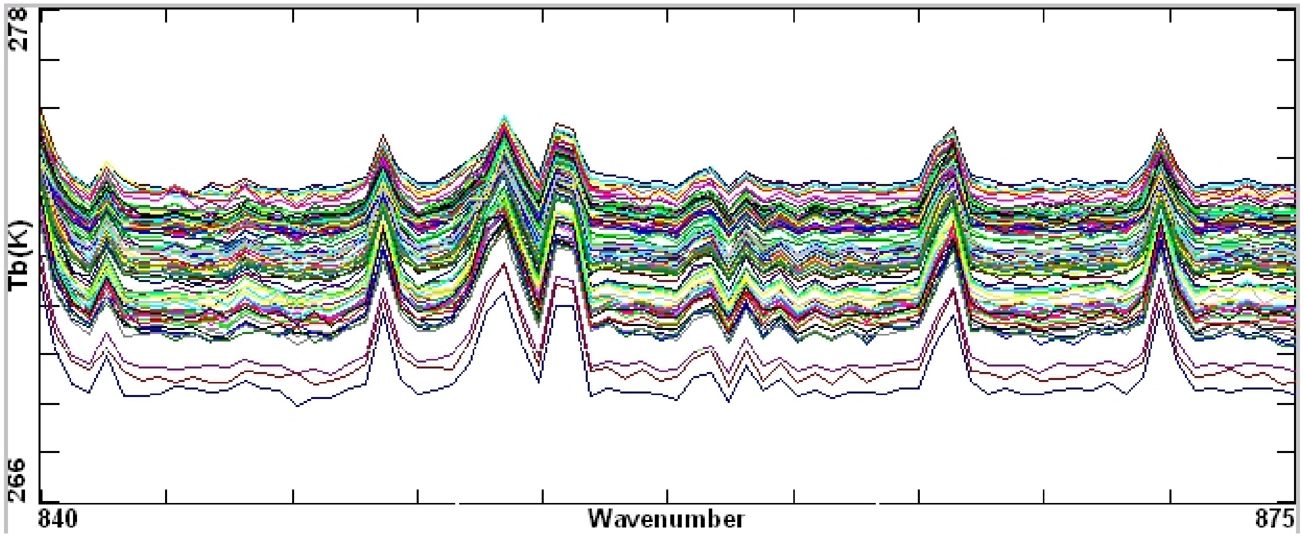
843.7576 cm^{-1}



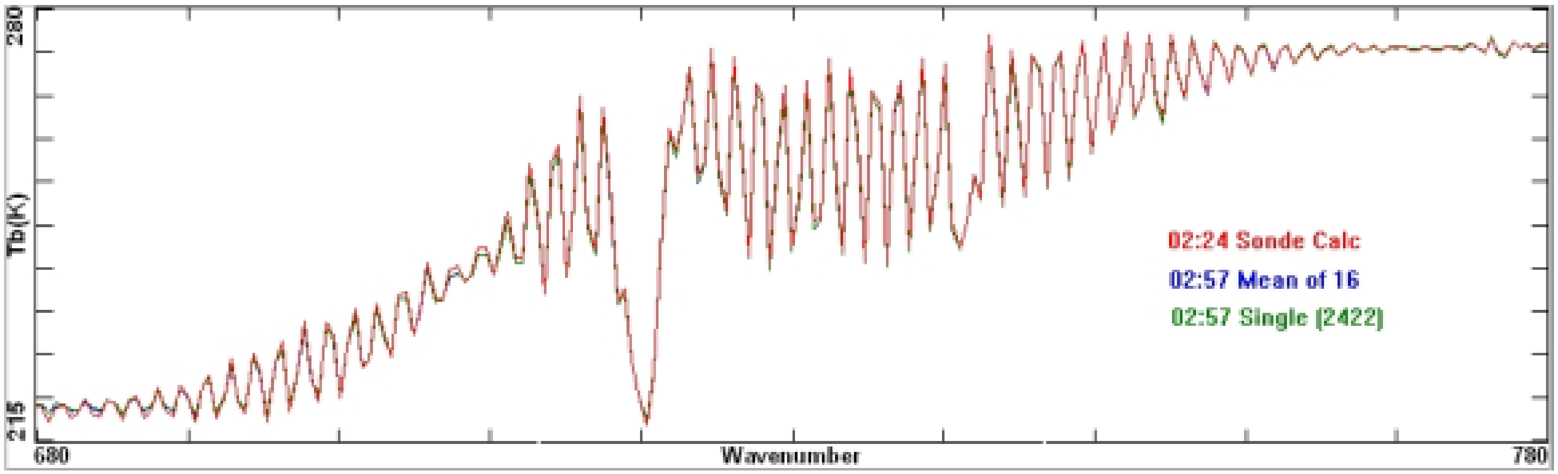
844.2398 cm^{-1}

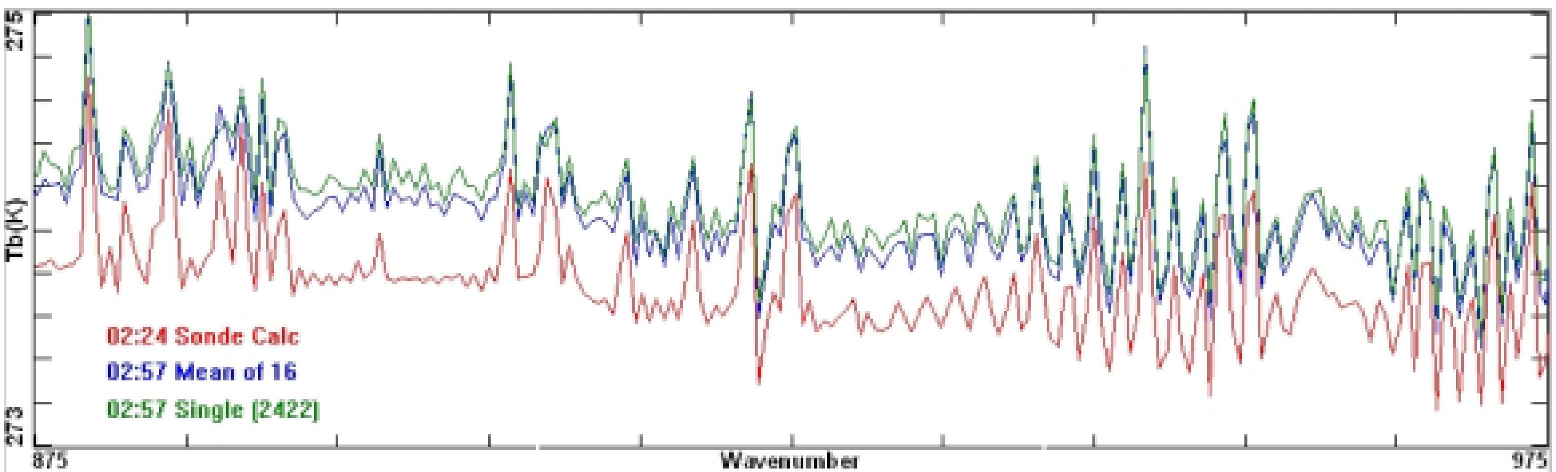
Unfiltered versus Filtered: DC8





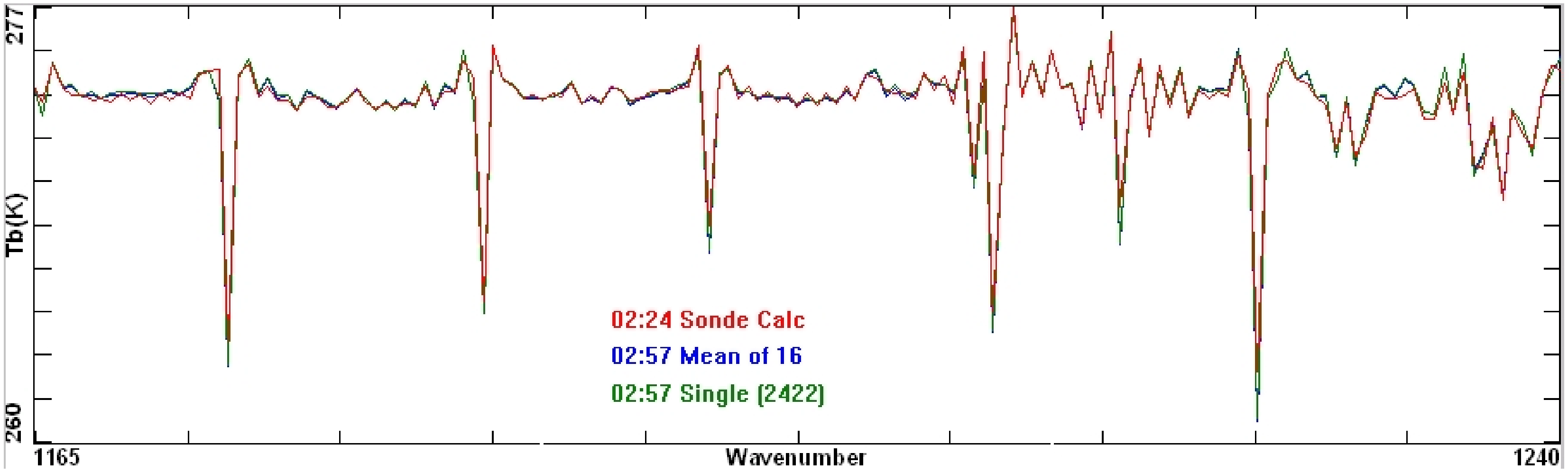
Filtered Radiances and Line by Line

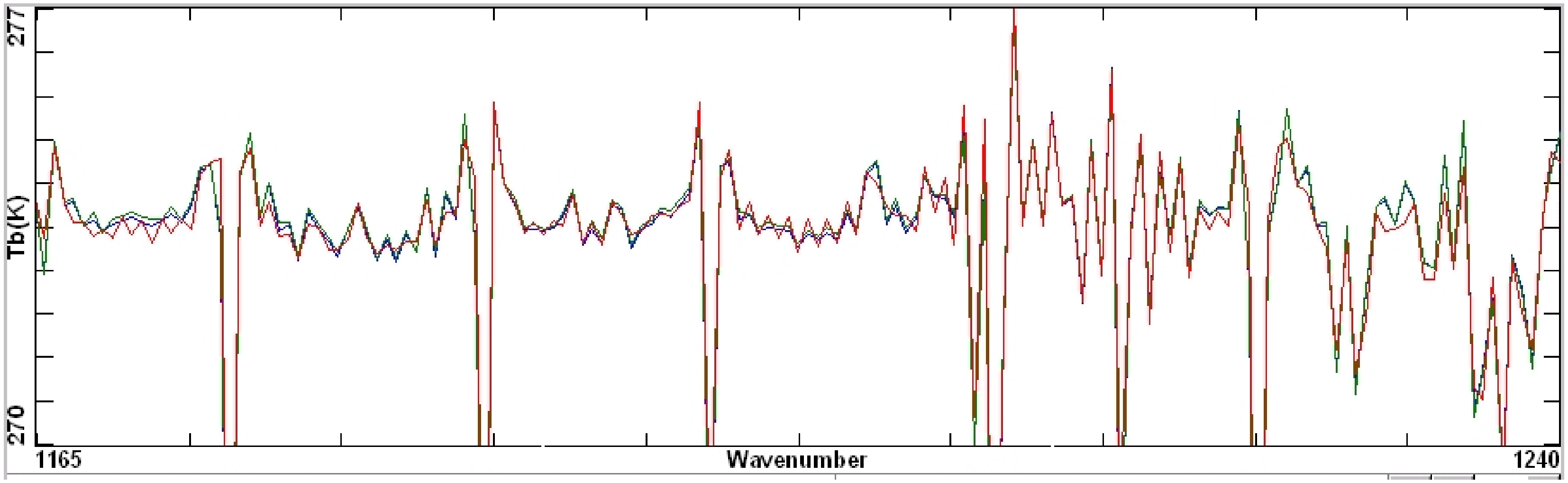




12-10-00 AFWEX

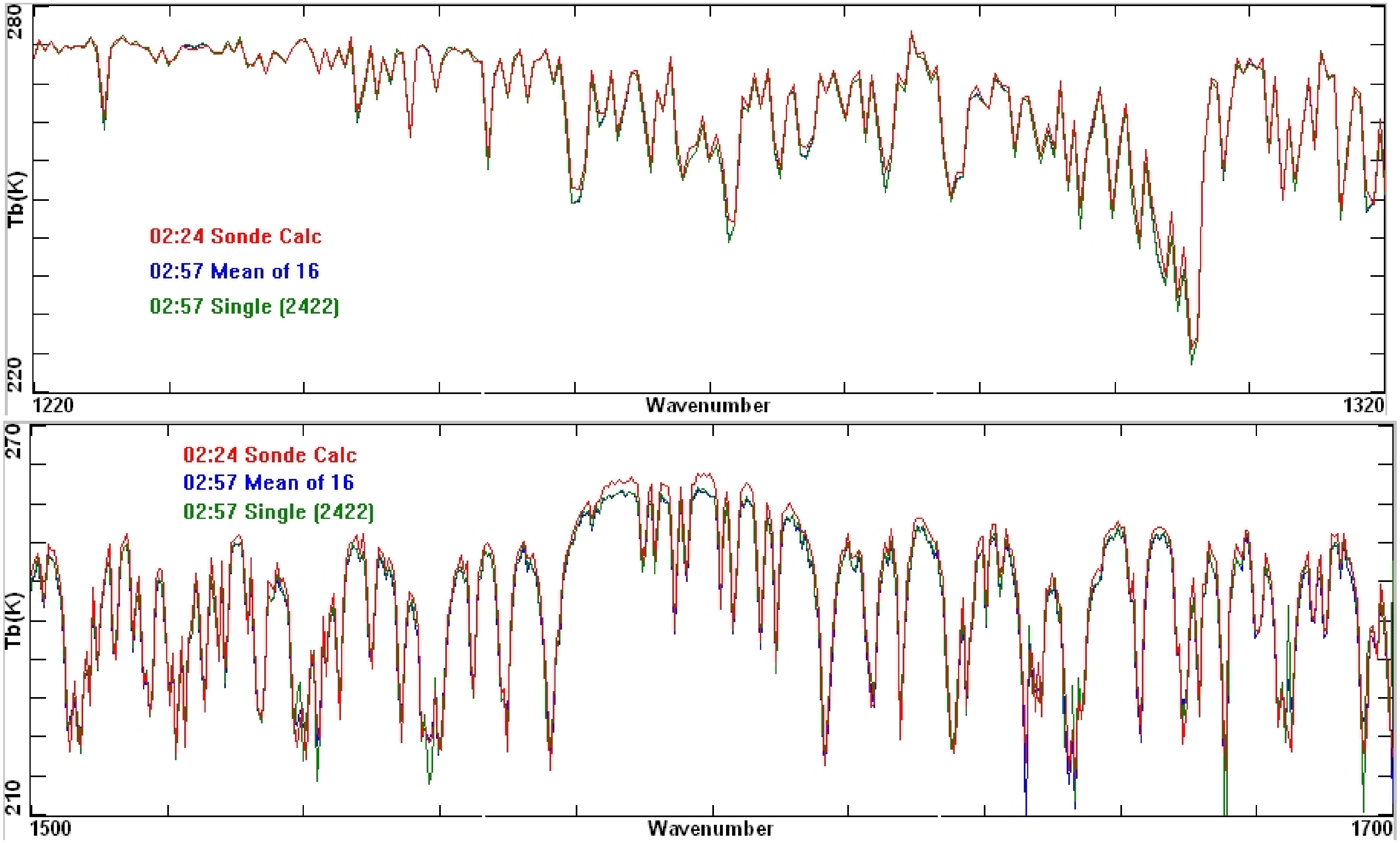
Calculation from Sonde Compared to S-HIS Brightness T Spectra





12-10-00 AFWEX

Calculation from Sonde Compared to S-HIS Brightness T Spectra



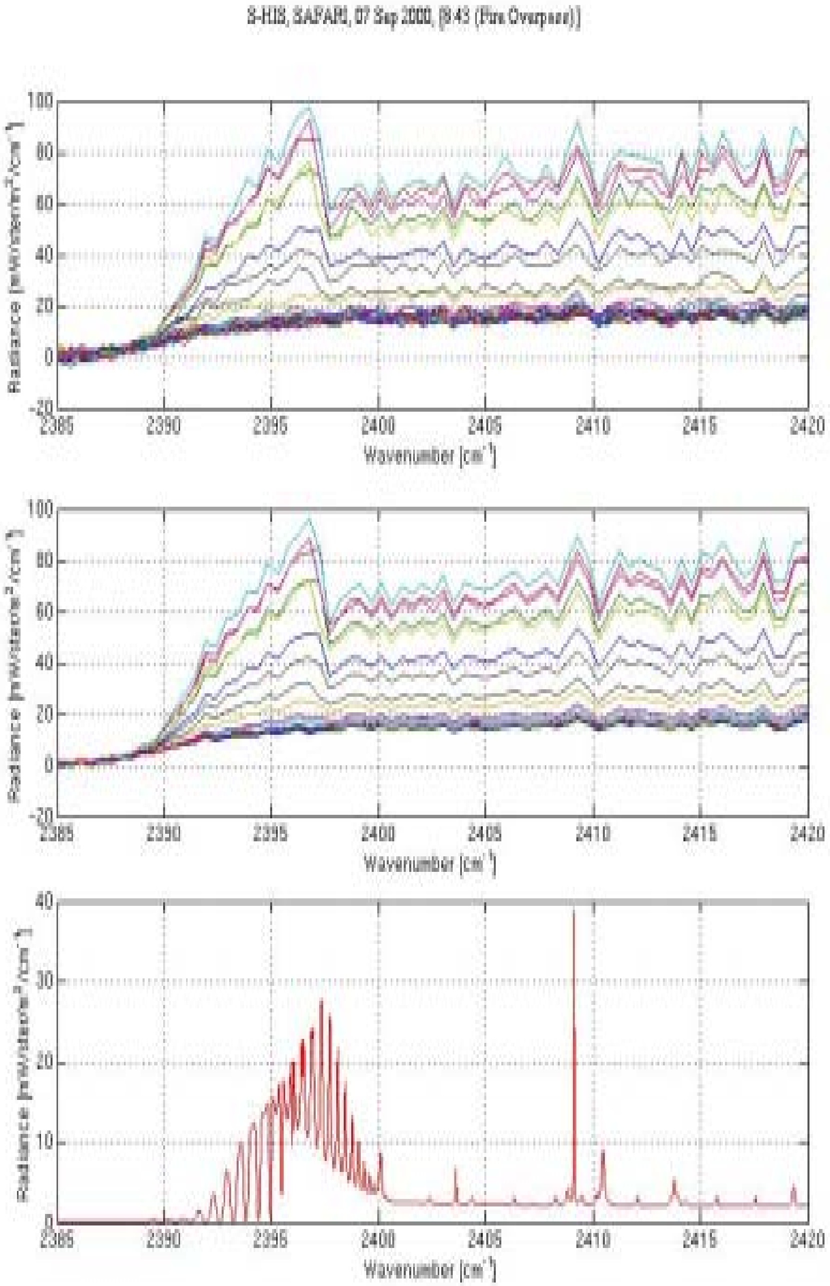
12-10-00 AFWEX

Line by Line versus Filtered and Unfiltered Data

Unfiltered

Filtered

Line by Line



PNF on Aircraft data Conclusions

- Noise Normalization is relevant when dealing with rarely occurring observations
- Representativeness of Training set is relevant when dealing with rarely occurring observations
- PNF preserves the atmospheric information also for rarely occurring spectra

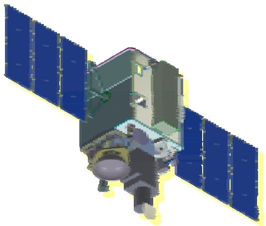
Open Questions

- Can we define an optimal number of PCs?
- If this number can be defined it is going to be a function of:
  - Instrument Characteristics
  - Observation Variability
  - What else?
- How does the noise reduction impact the Regression and the Physical Retrievals?
- How does the information loss impact them?
- What other tests can we run to address the reliability of PNF?

PNF applied to AIRS data

- Noise Characterization using Earth scene data
- Noise Filter
- PNF effect on Popping Noise

Noise Estimation



- **Approach:**

  Using all spectra from a granule:

  1) generate dependent PCs within the granule.

  2) reconstruct spectra using reduced number of PCs

  3) use statistics of reconstruction residual to derive initial noise

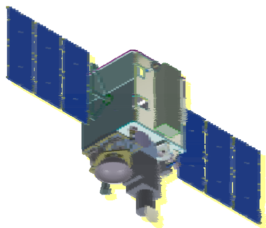
  4) Normalize spectra using initial noise estimate (divide by NeN)

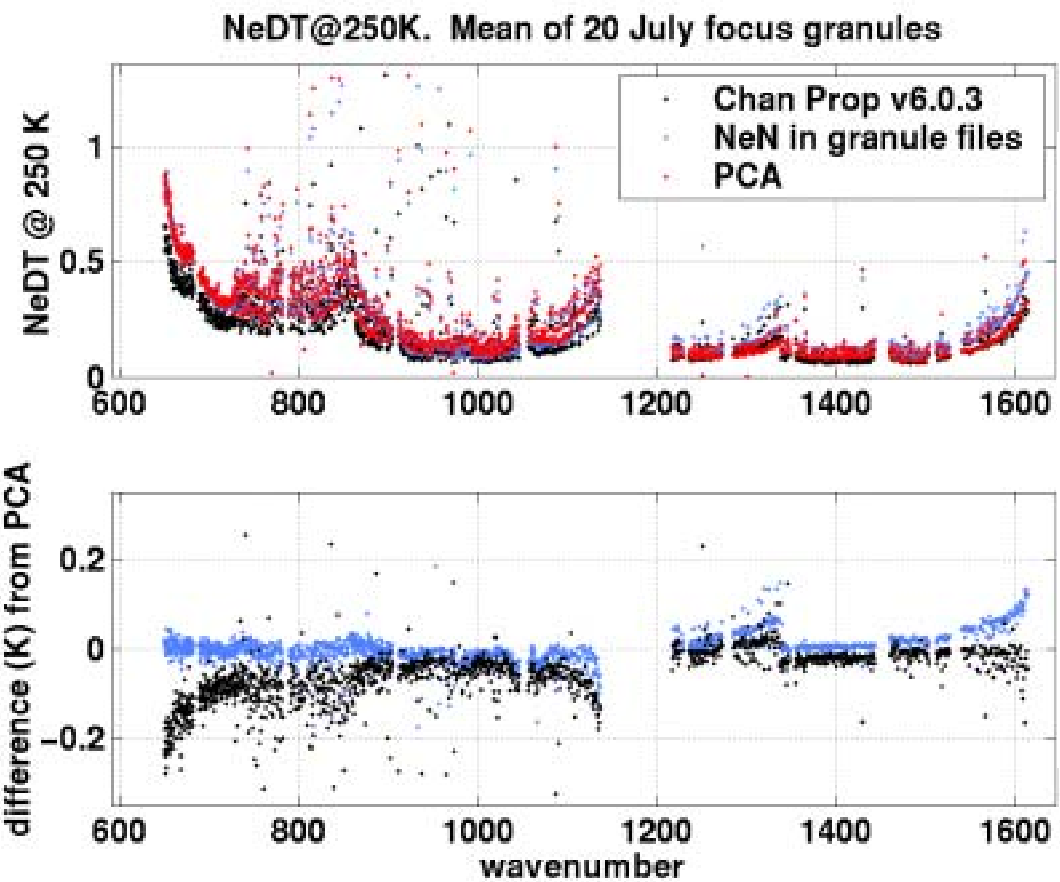
  5) generate PCs for normalized spectra within the granule.

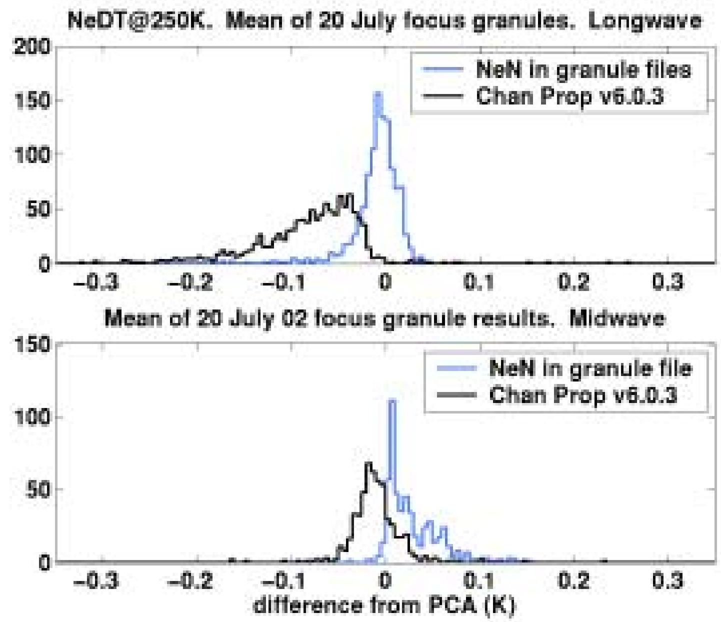
  6) reconstruct normalized spectra using reduced number of PCs,

  7) Remove normalization (multiply spectra by initial noise estimate) and use statistics of reconstruction error to derive noise estimates

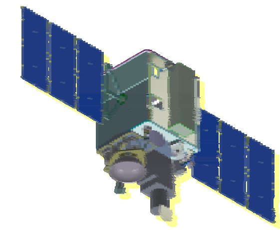
Noise Characterization using PNF

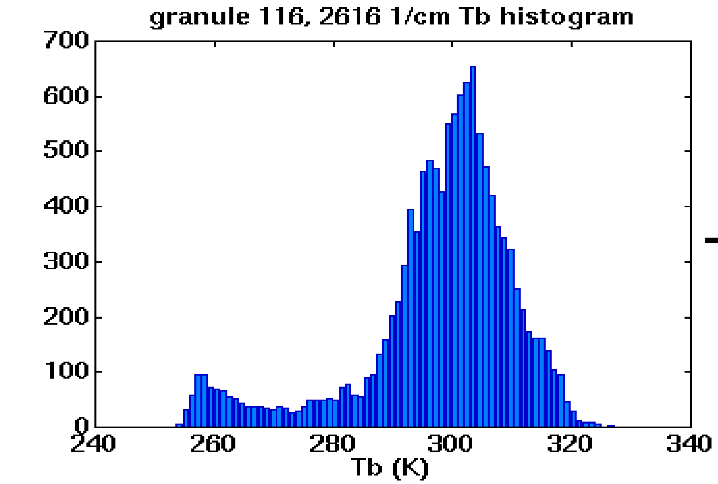


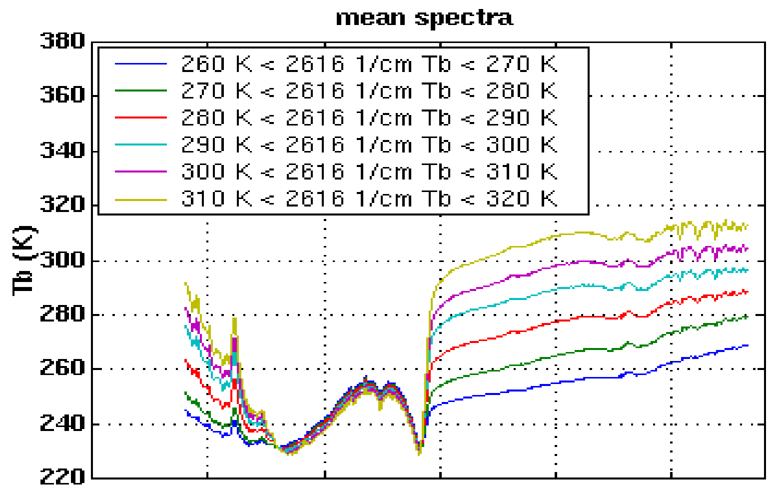


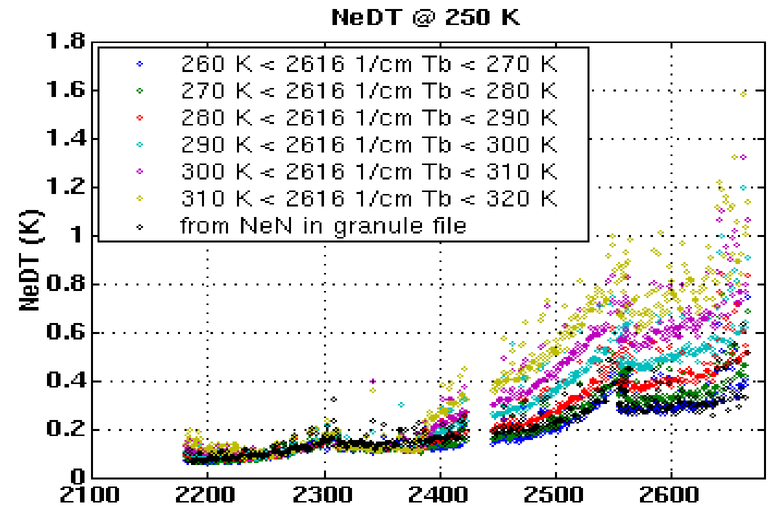


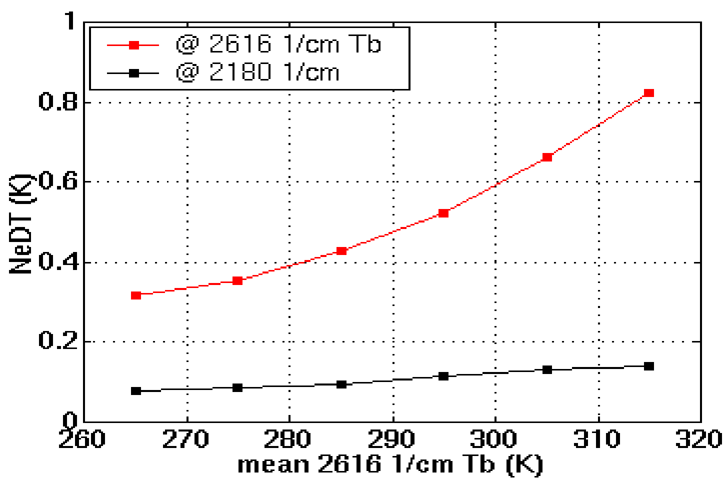
Characterization of photon limited SW noise



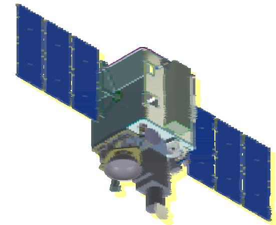






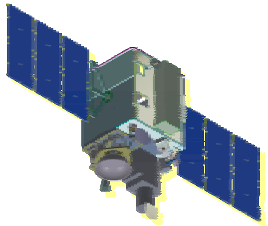


NeDT Estimation using Earth Scene Data



- **Summary**
  - LongWave NeDT@250K:
    - PNF results yield best agreement with “NeN” from granule files (order ~0.02 K)
    - PNF results are slightly higher than values provided in v6.0.3 channel properties file (order 0.05 K)
  - MidWave NeDT@250K:
    - PNF results yield good agreement with “NeN” from granule files, but are slightly higher by ~0.03 K.
    - PNF results yield best agreement with values provided in v6.0.3 channel properties file, but are slightly lower by ~0.02 K.
  - ShortWave:
    - photon-limited noise characterized

Striping/Popping Analysis



- Residual calibration noise and transient detector behavior (“popping”) both contribute to spatial “striping”.

- Contributions due to the calibration in the 6/14/2002 data have been greatly reduced by use of the L1B ATBD algorithm (using granule mean gains versus scan-line by scan-line gains).

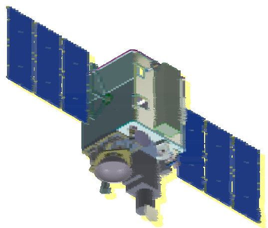
- Residual striping due to detector “popping” is still present. Order 1K can be seen for LW PV detectors, with nearly all demonstrating at least some low level of popping.

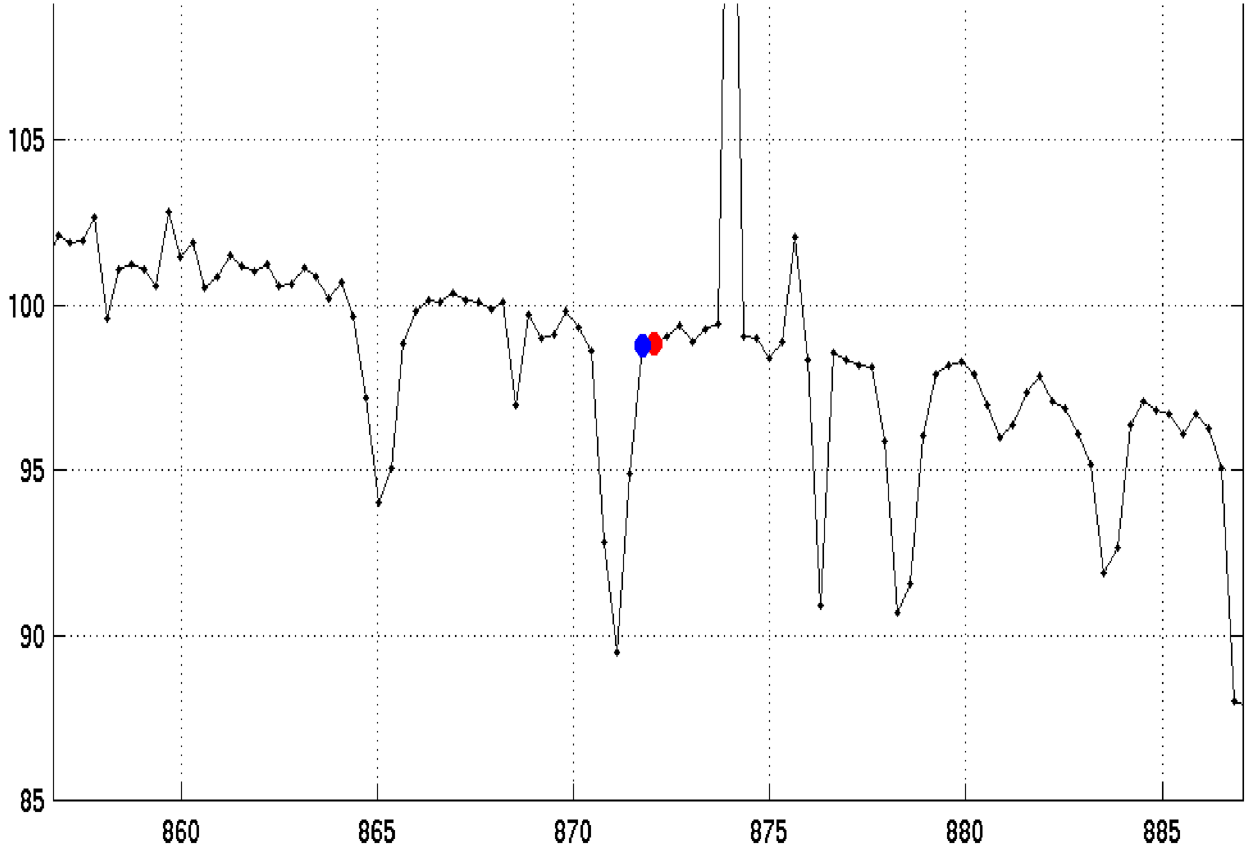
- This presentation:

    - Uses channel differences and PNF used to characterize residual “striping” in 7/20/2002 granules.

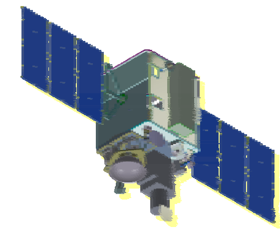
    - Demonstrates that PNF is effective in removing the majority of the residual striping, as well as most of the spatially uncorrelated noise.

Granule 016
channels 674, 675
AB_State = [0,0], Rad_Qual = [0,0],
Bad_Flag = [0,0]

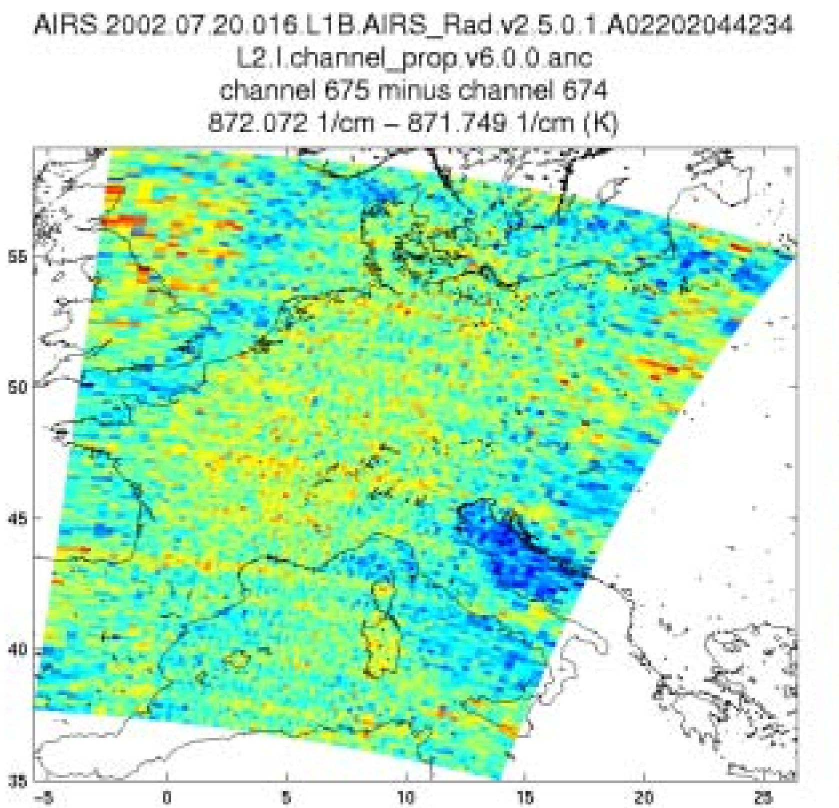




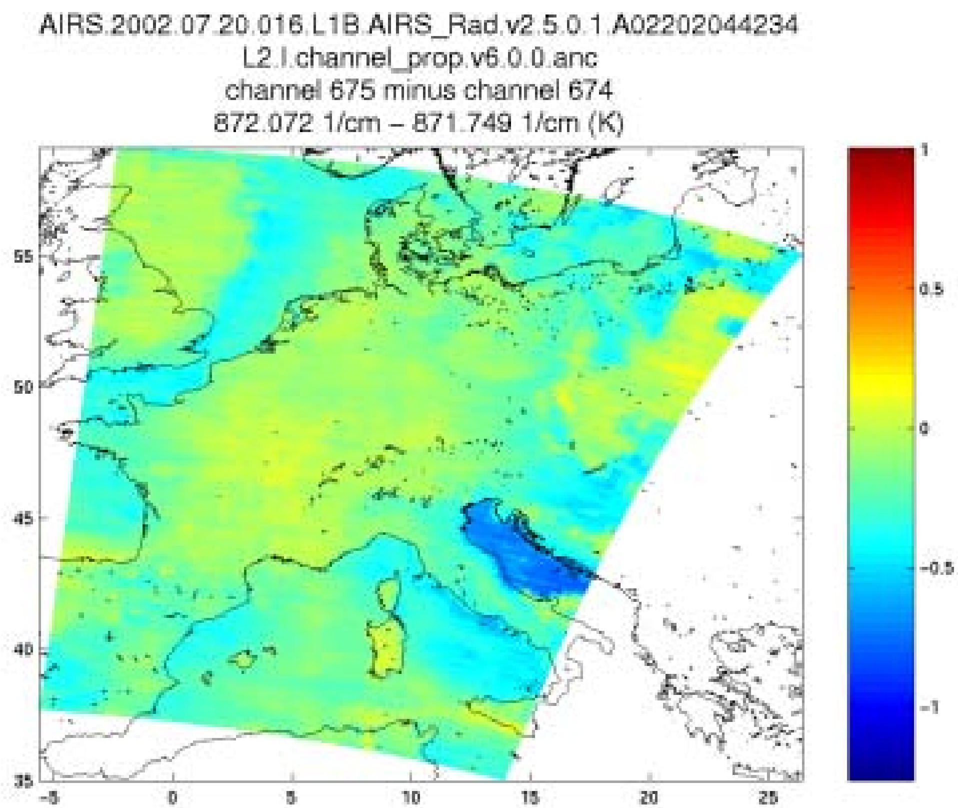
Channel 675 - Channel 674
Brightness Temperature Differences



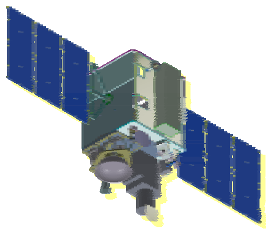
Original



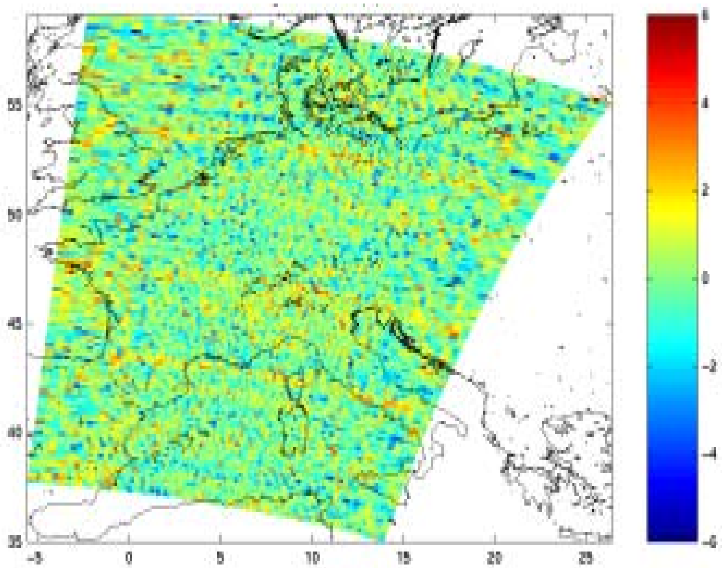
Filtered

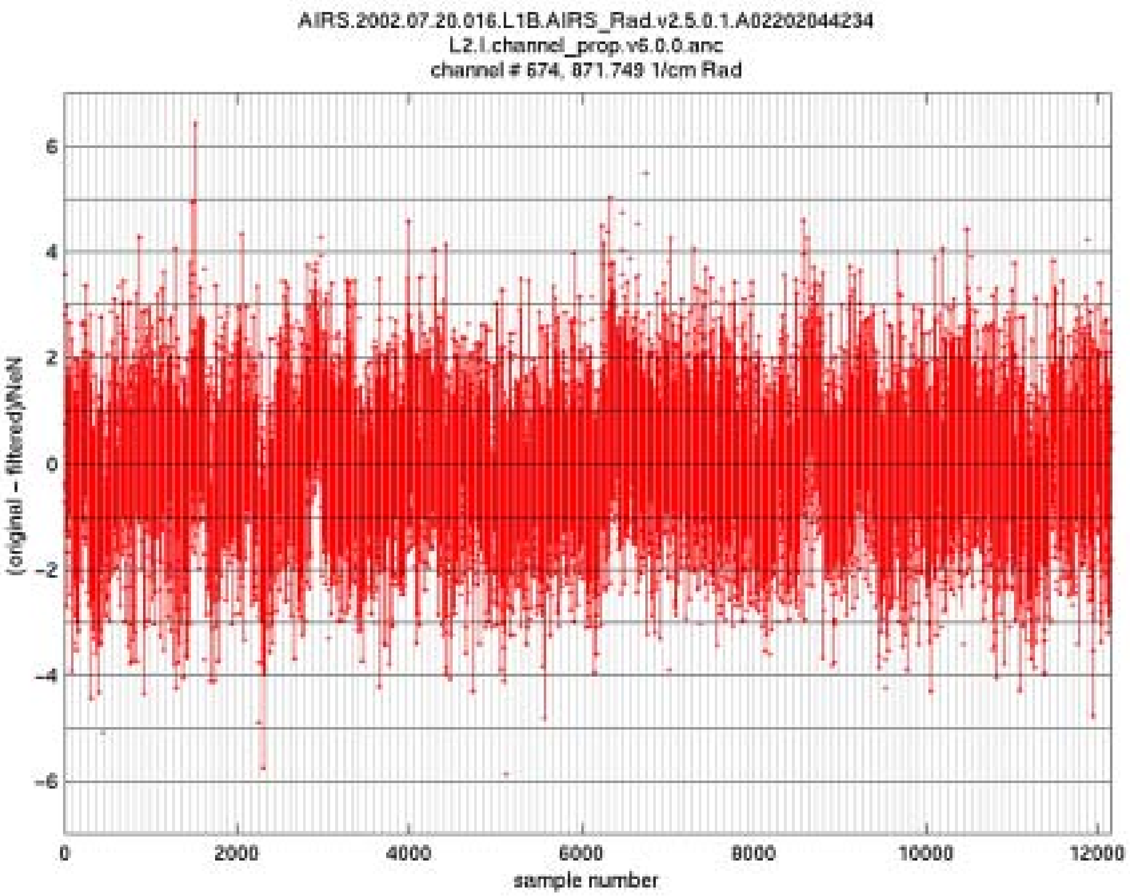


Channel # 674



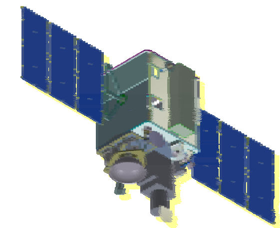
(Original-Filtered)/NeN



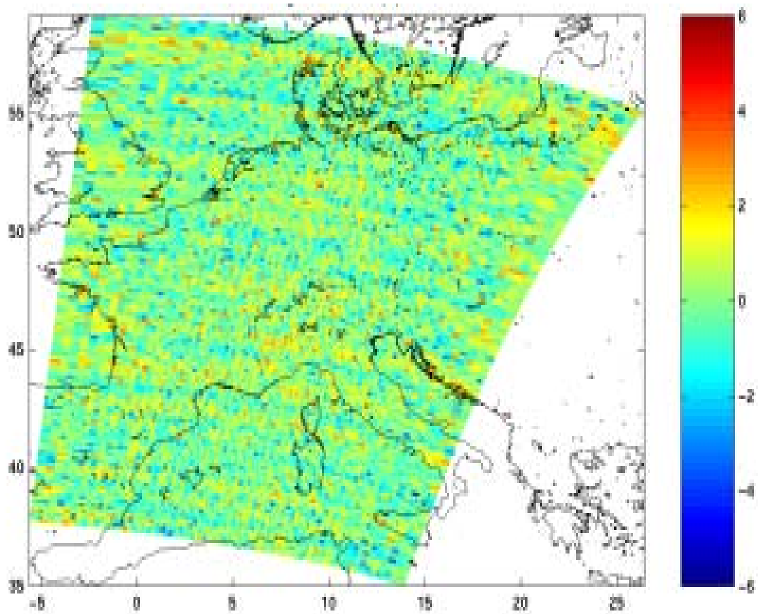


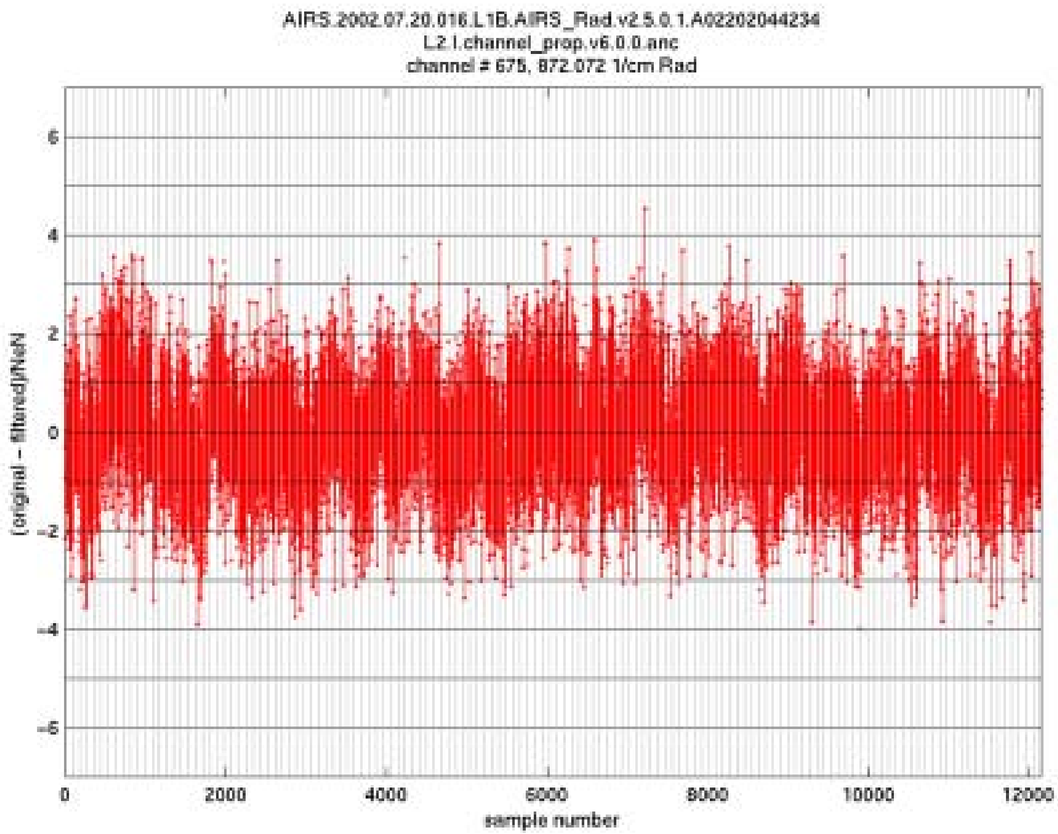
pops: 0 (5σ), 0 (4σ), 0 (3σ), 7 (2σ), 139 (1σ)

Channel # 675



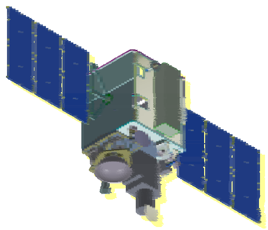
(Original-Filtered)/NeN

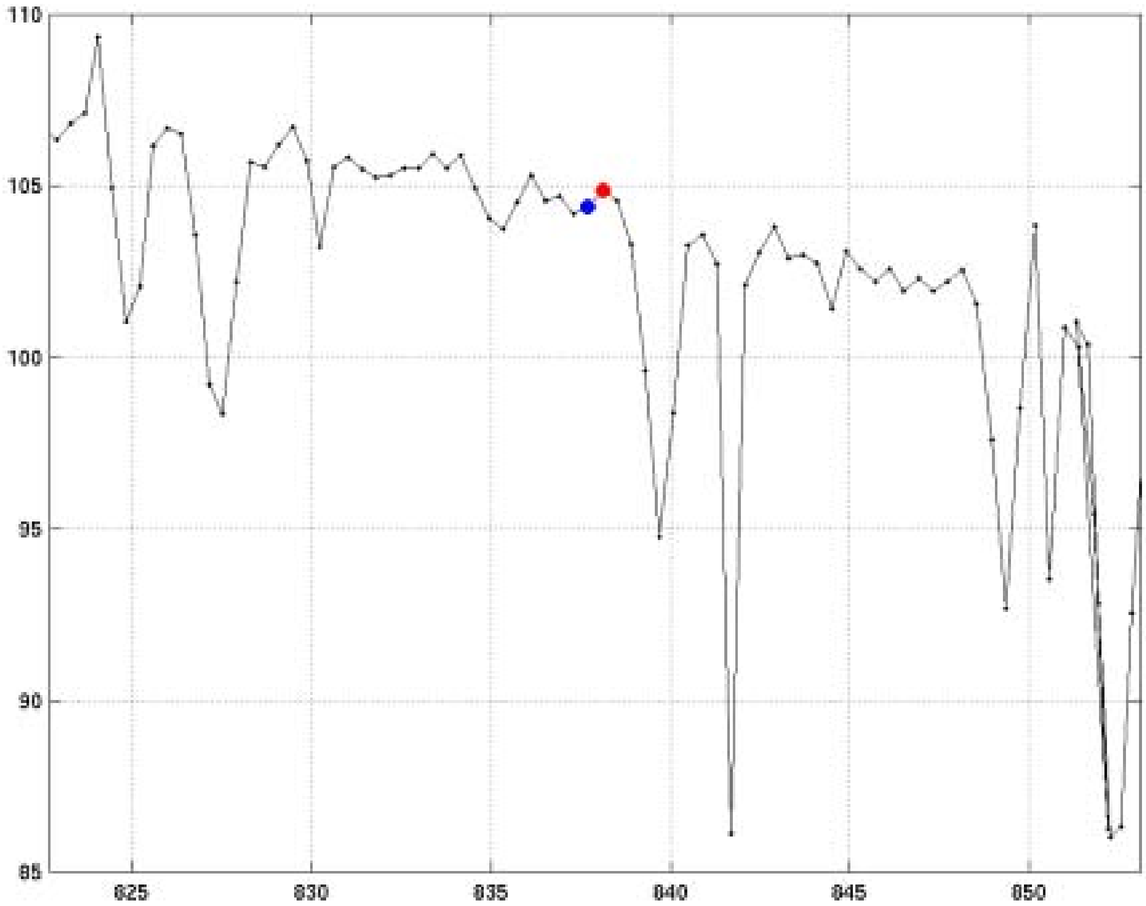




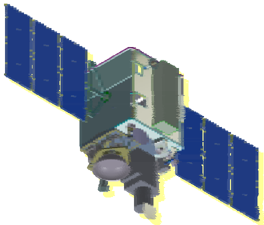
pops: 0 (5σ), 0 (4σ), 0 (3σ), 0 (2σ), 88 (1σ)

Granule 016
channels 572, 573
AB_State = [0,0], Rad_Qual = [0,0],
Bad_Flag = [0,0]

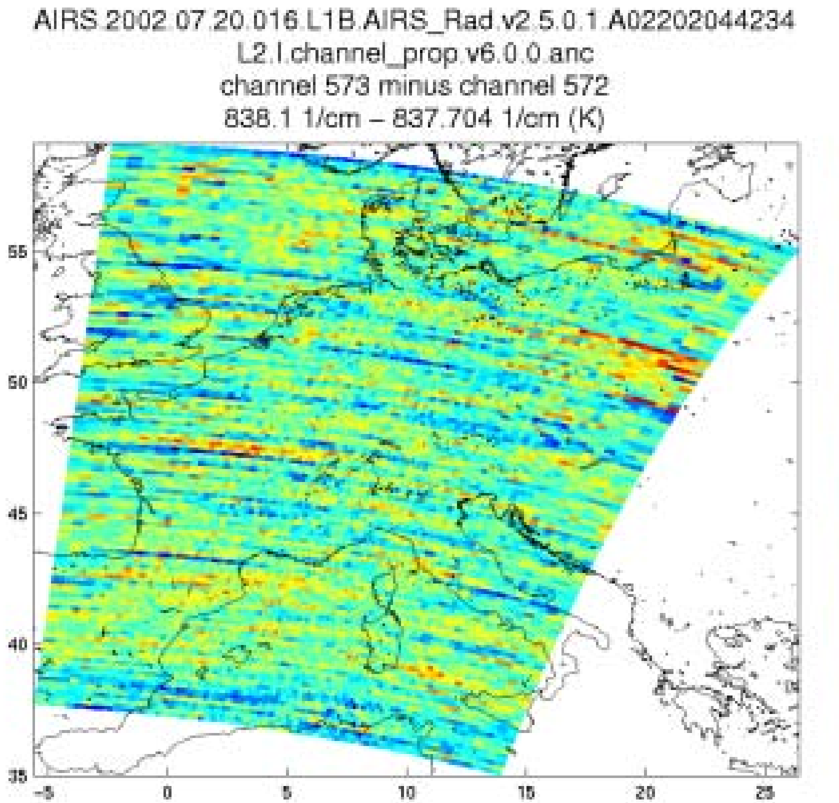




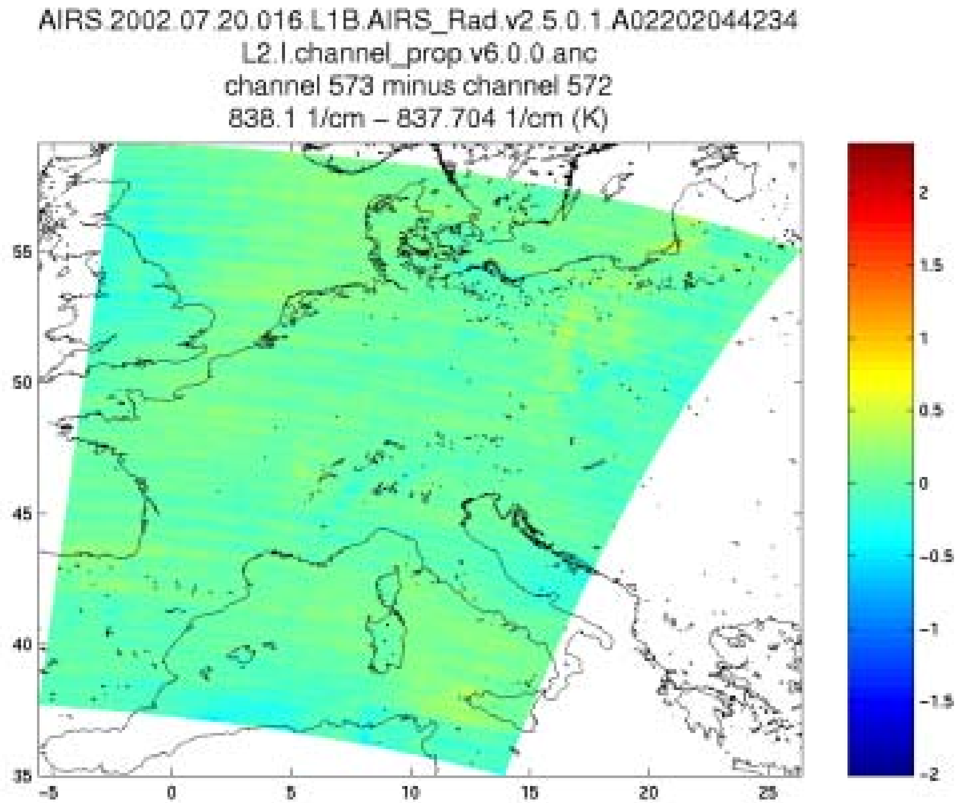
Channel 573 - Channel 572
Brightness Temperature Differences

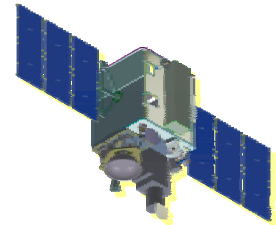


Original



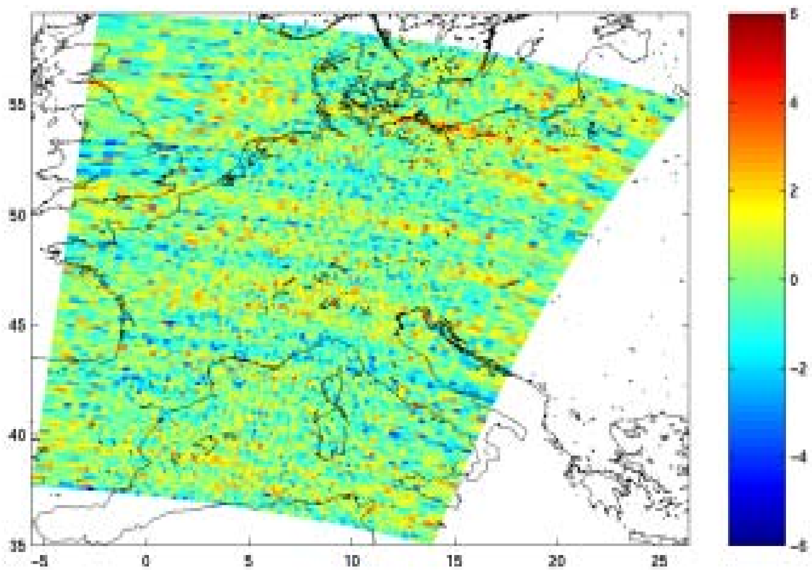
Filtered

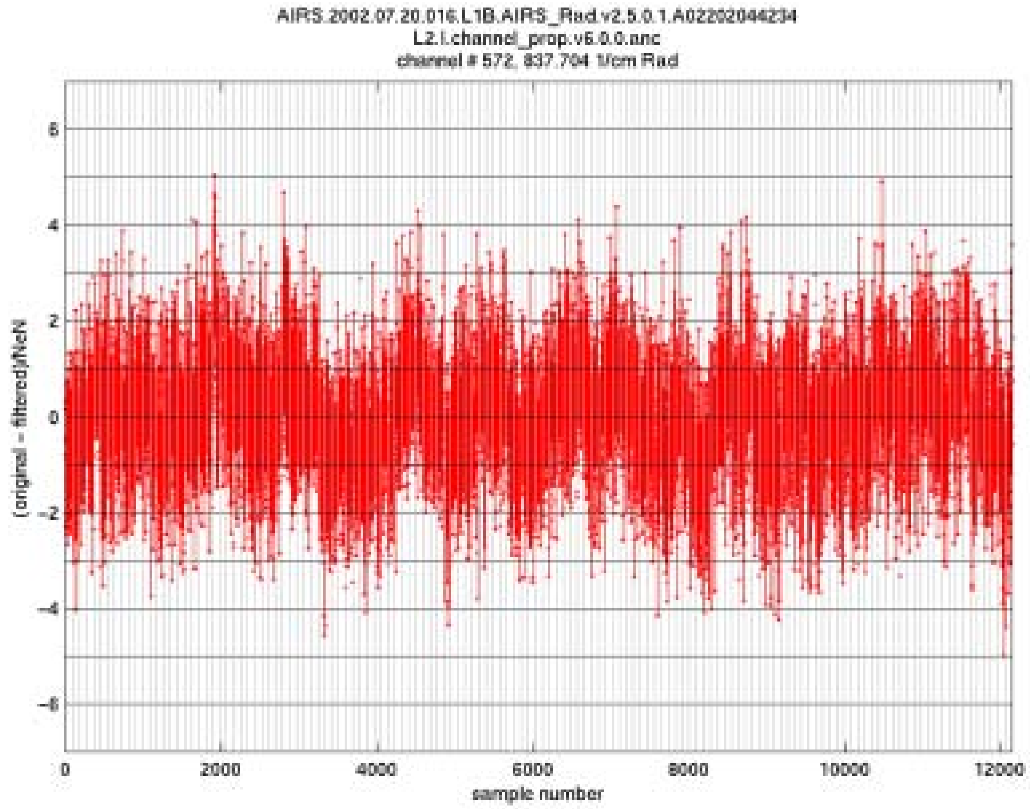




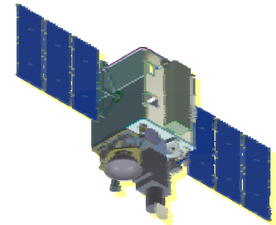
Channel # 572

(Original-Filtered)/NeN



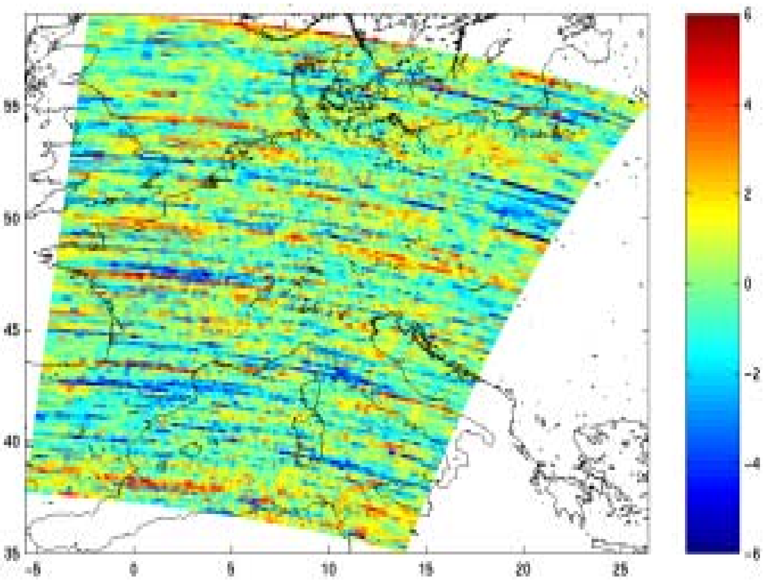


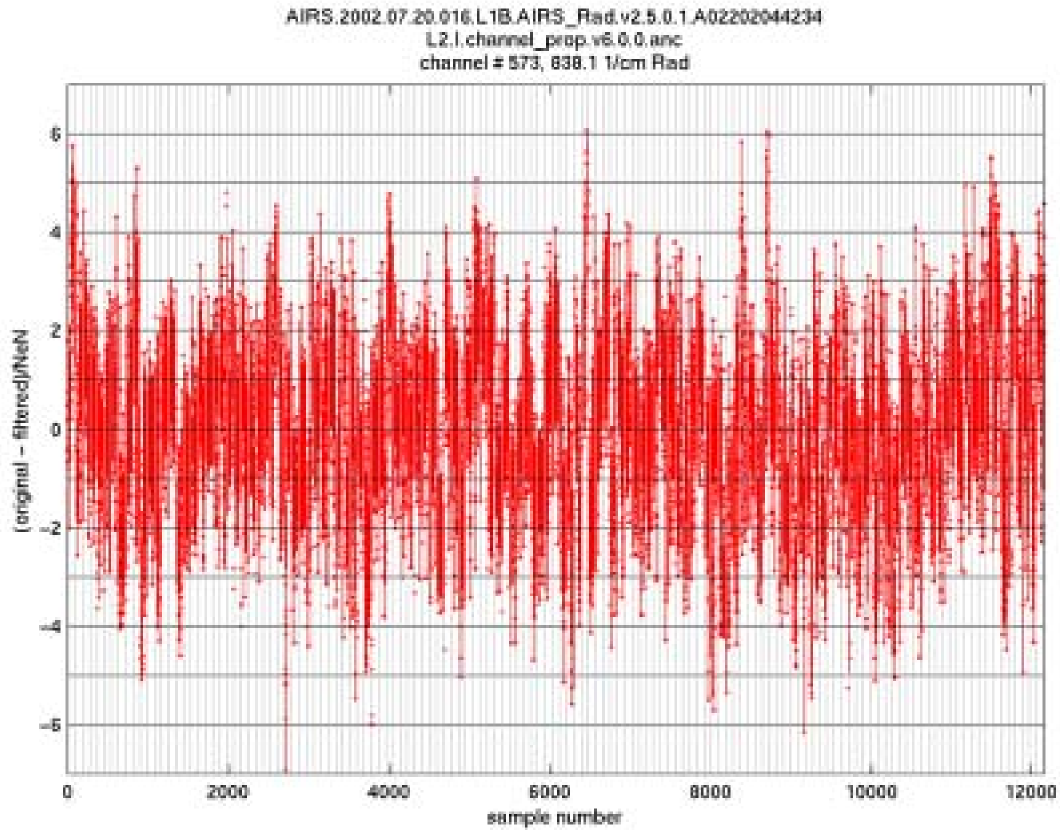
pops: 0 (5σ), 0 (4σ), 0 (3σ), 15 (2σ), 225 (1σ)



Channel # 573

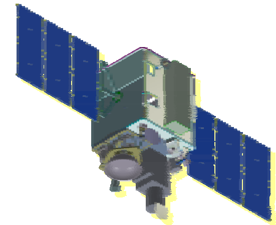
(Original-Filtered)/NeN



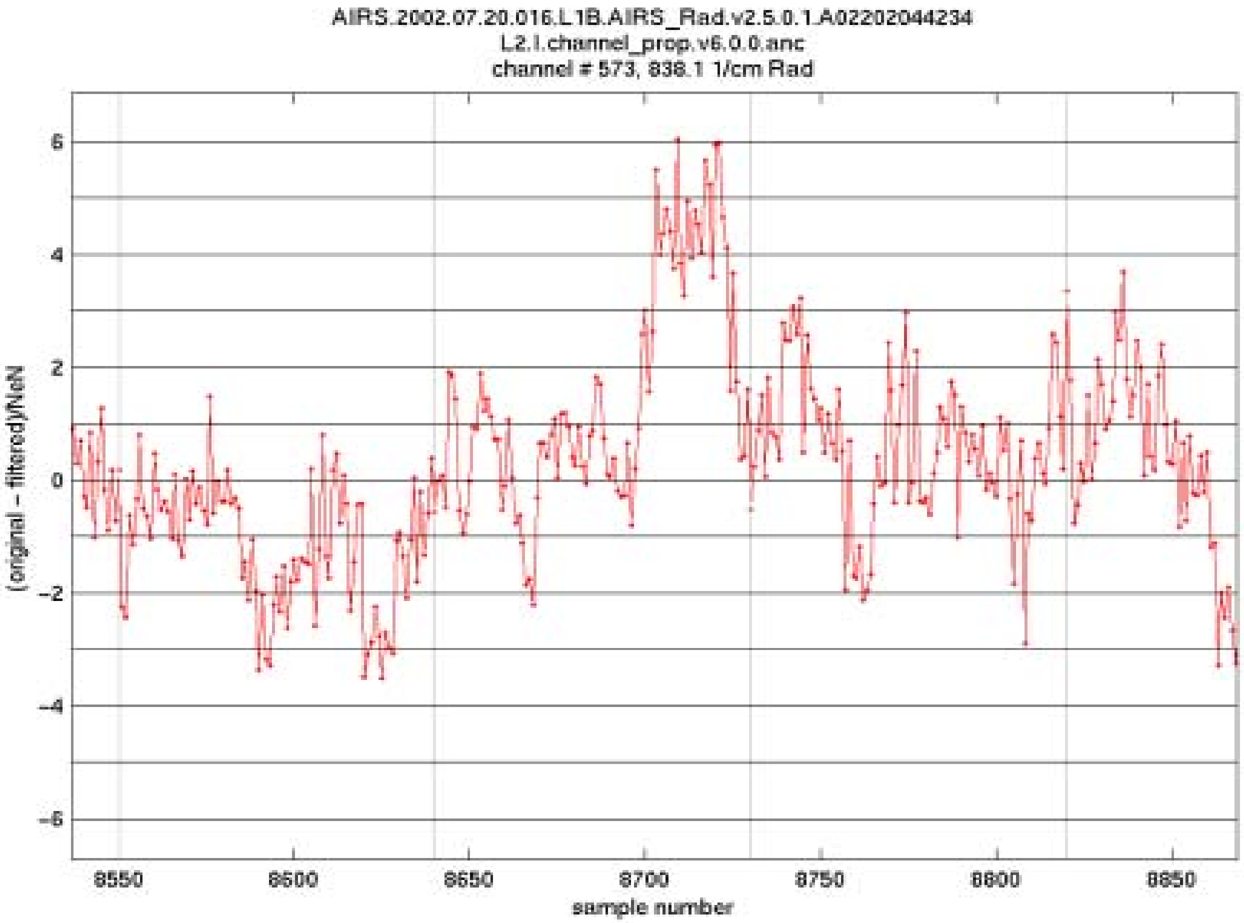


pops: 2 (5σ), 16 (4σ), 91 (3σ), 351 (2σ), 1025 (1σ)

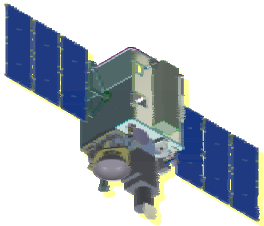
Channel # 573

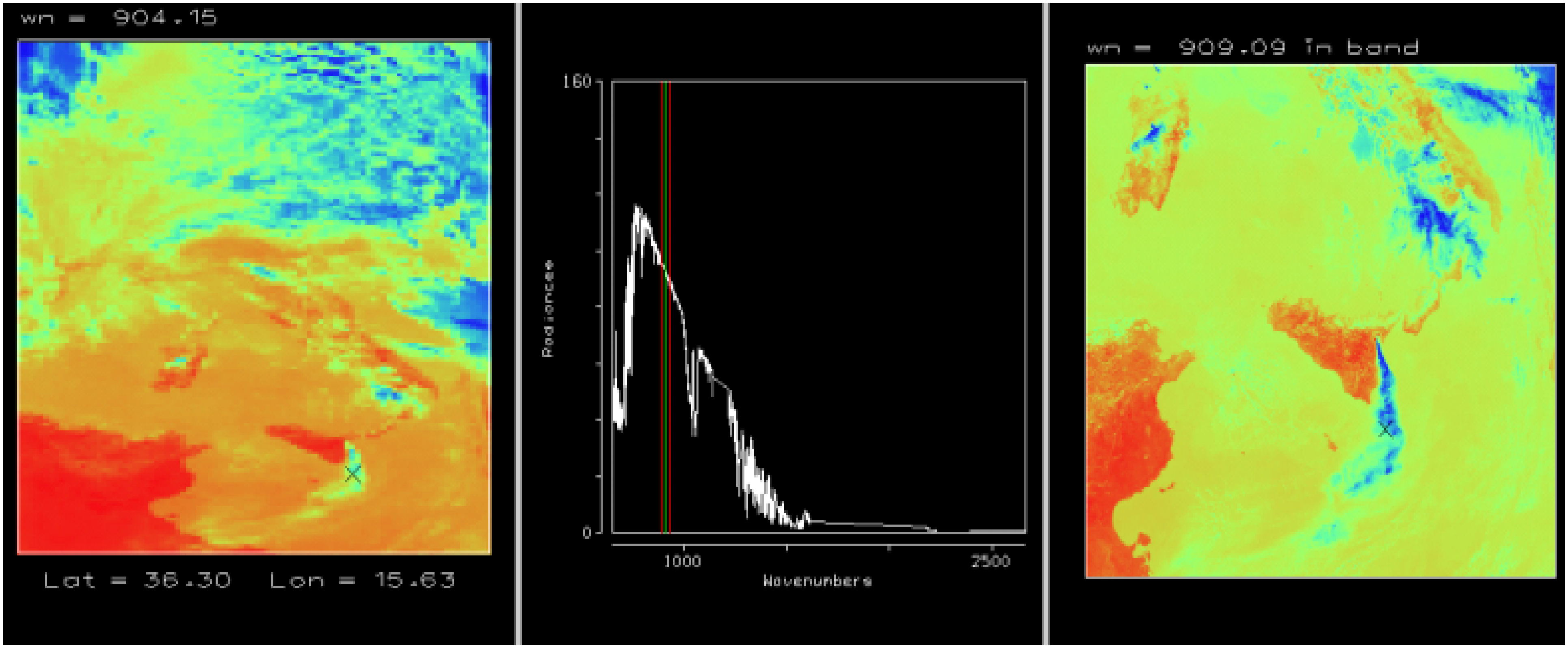


(Original-Filtered)/NeN



Mount Etna Eruption: 28 Oct 2002

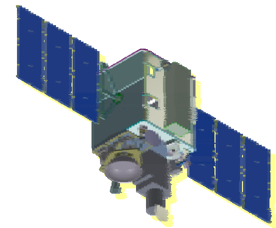


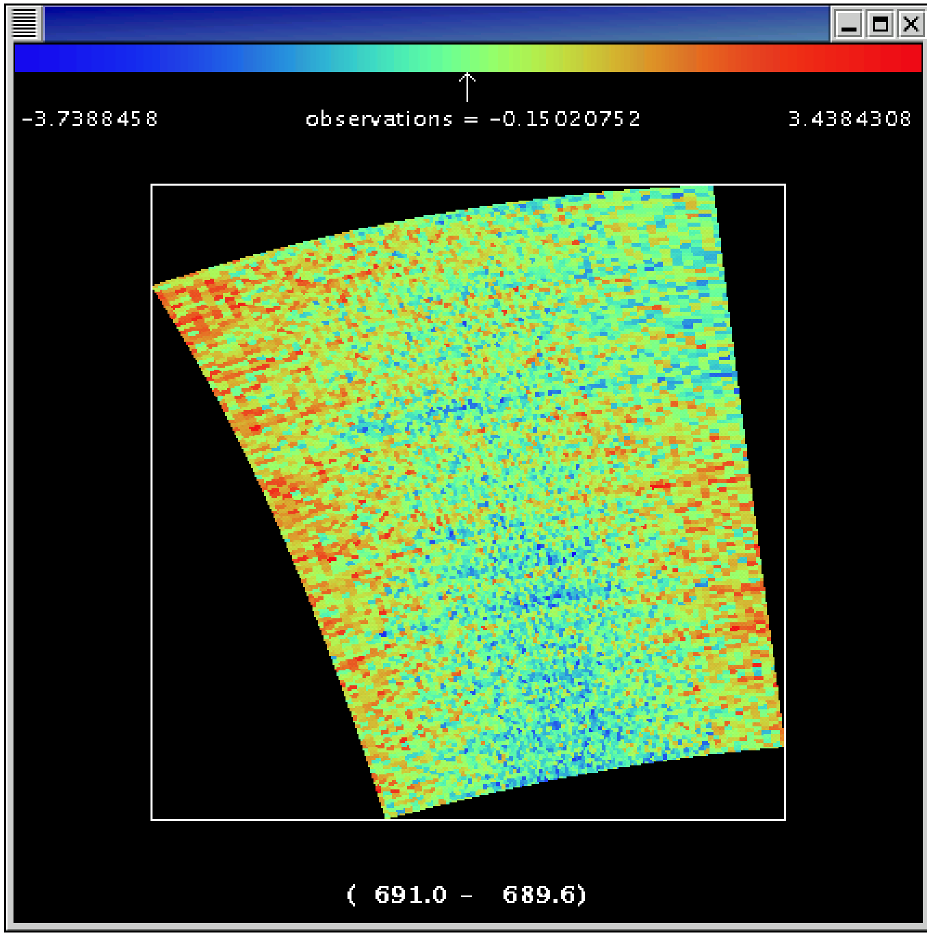


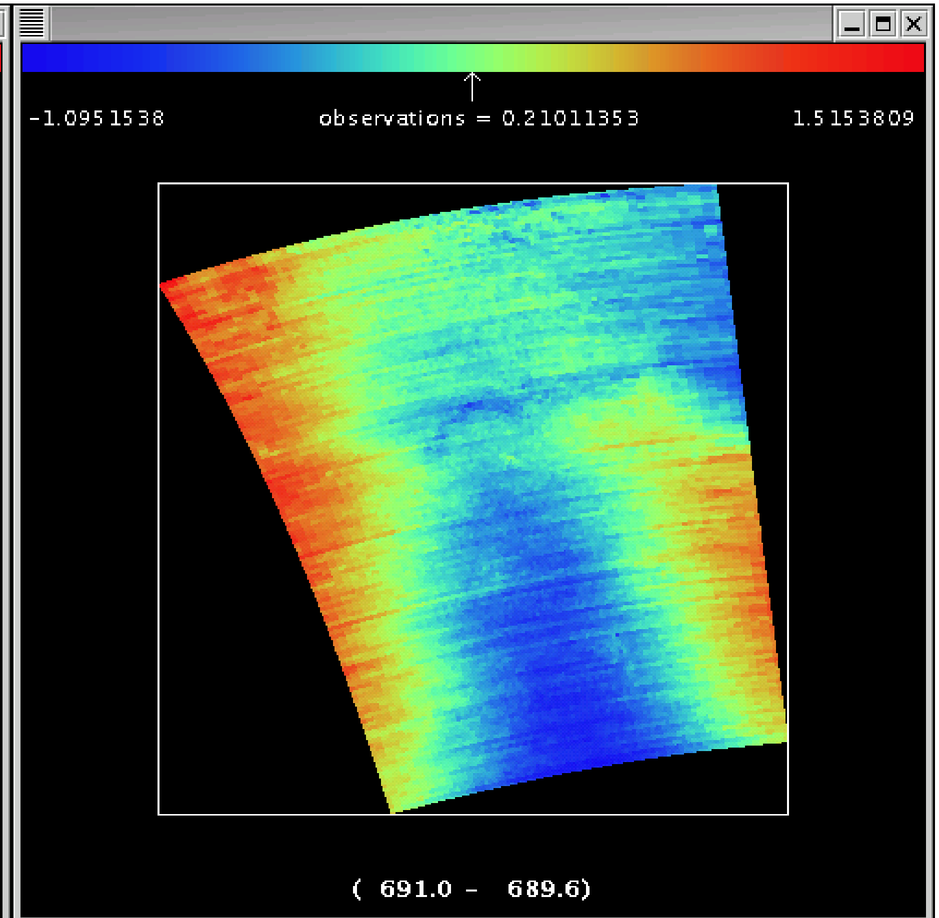
AIRS

MODIS

691.0 – 689.6 cm^{-1}
Brightness Temperature Differences



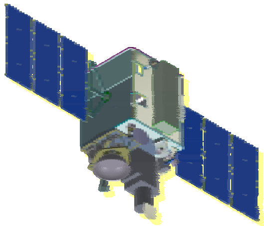


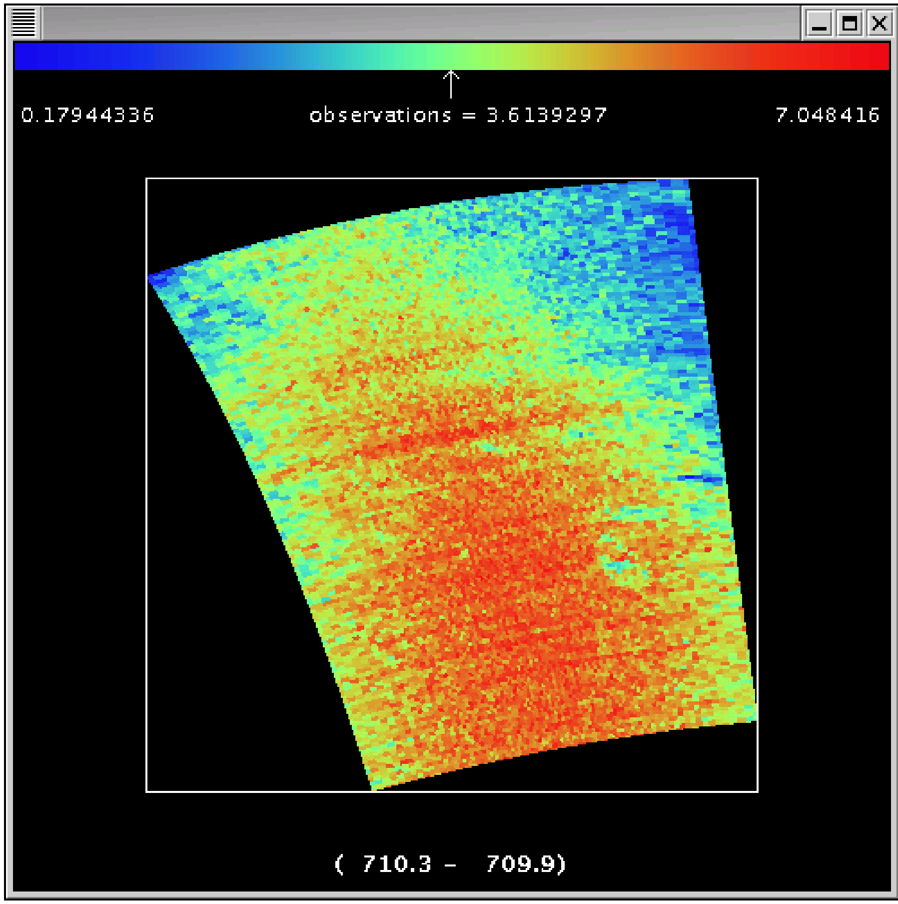


Original (691.0 cm^{-1}- 689.6 cm^{-1})

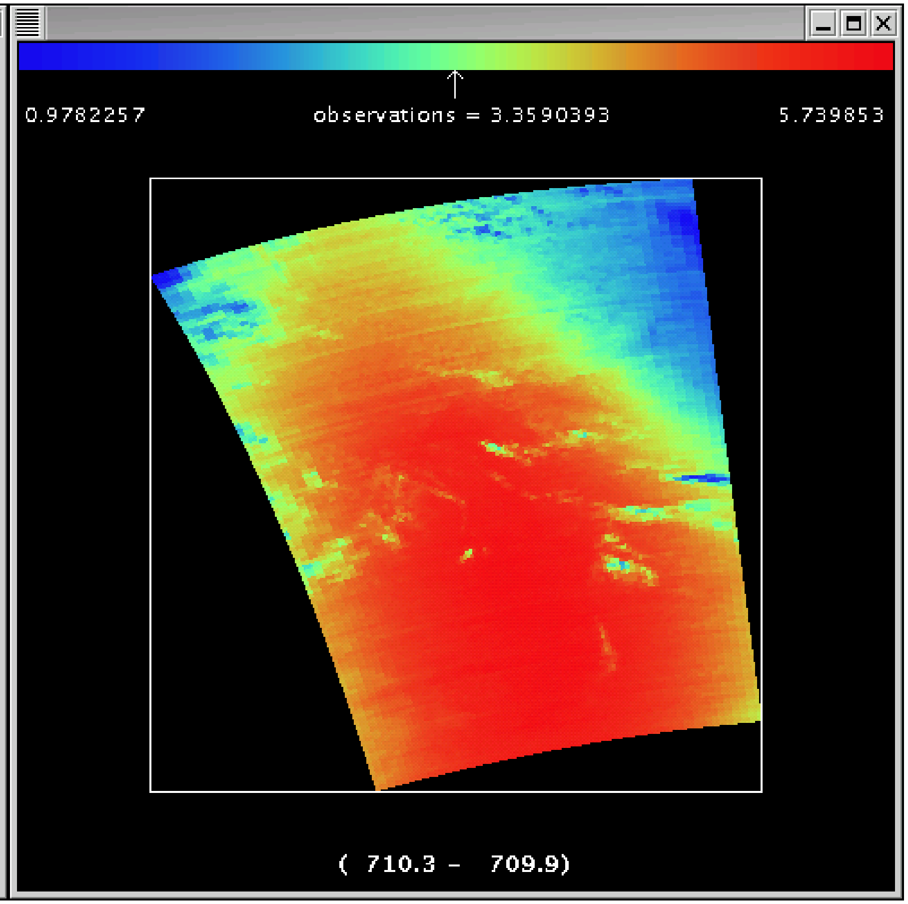
Filtered (691.0 cm^{-1}- 689.6 cm^{-1})

710.3 – 709.9 cm^{-1}
Brightness Temperature Differences



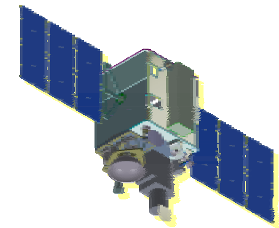


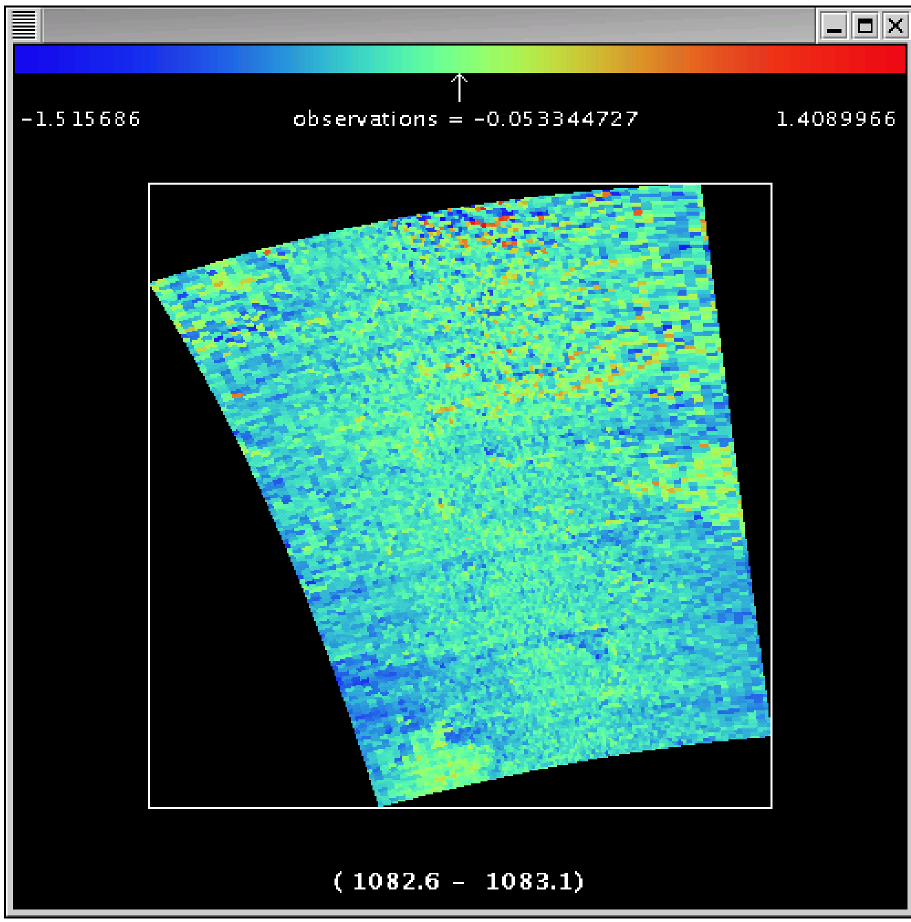
Original (710.3 cm^{-1}- 709.9 cm^{-1})



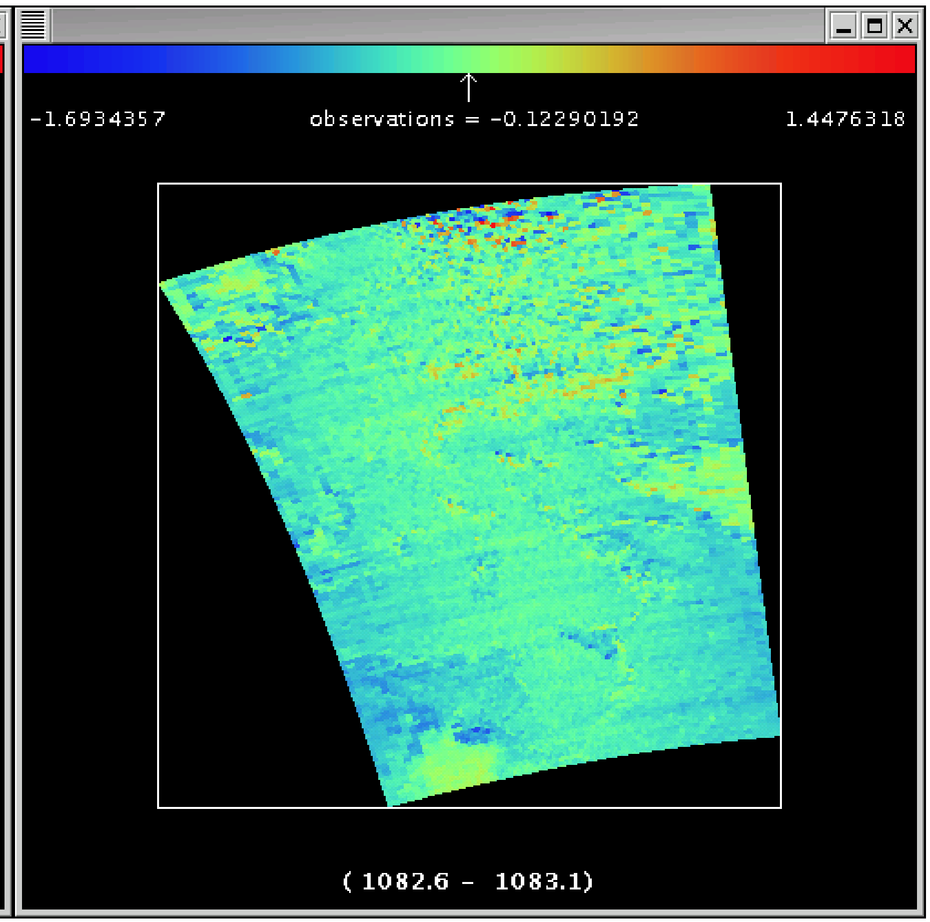
Filtered (710.3 cm^{-1}- 709.9 cm^{-1})

1082.6 – 1083.1 cm^{-1}
Brightness Temperature Differences



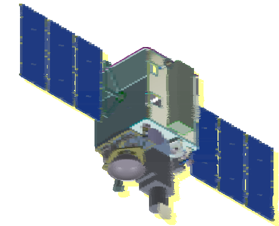


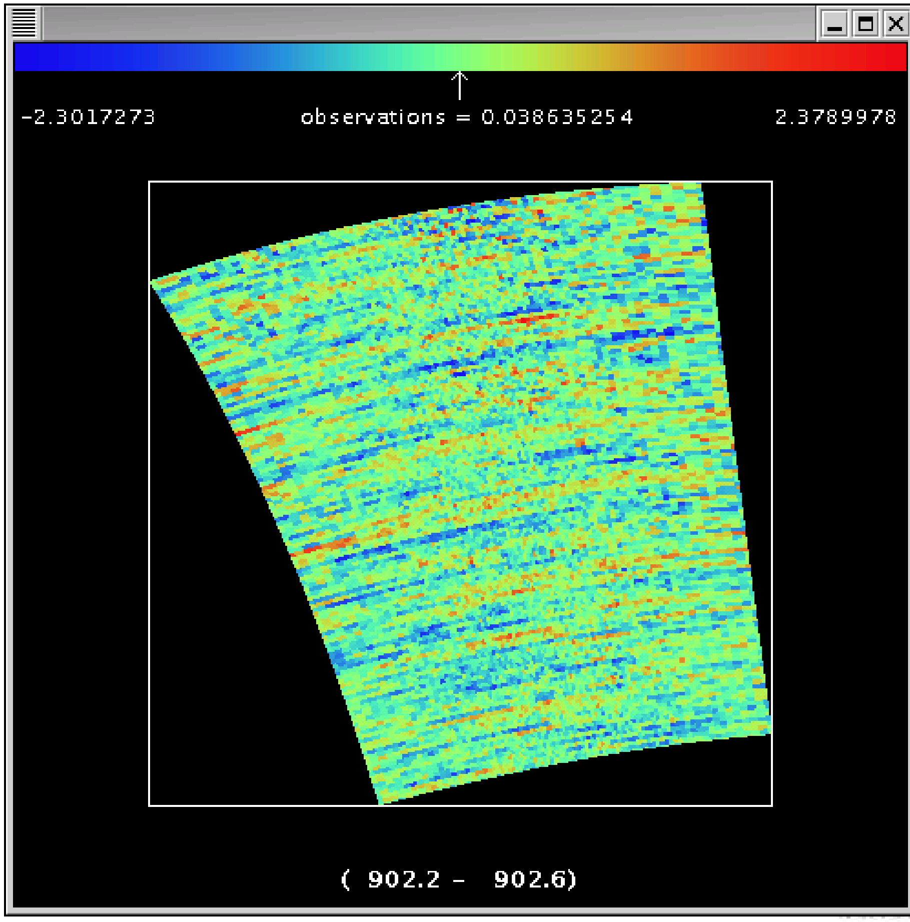
Original (1082.6 cm^{-1}- 1083.1 cm^{-1})

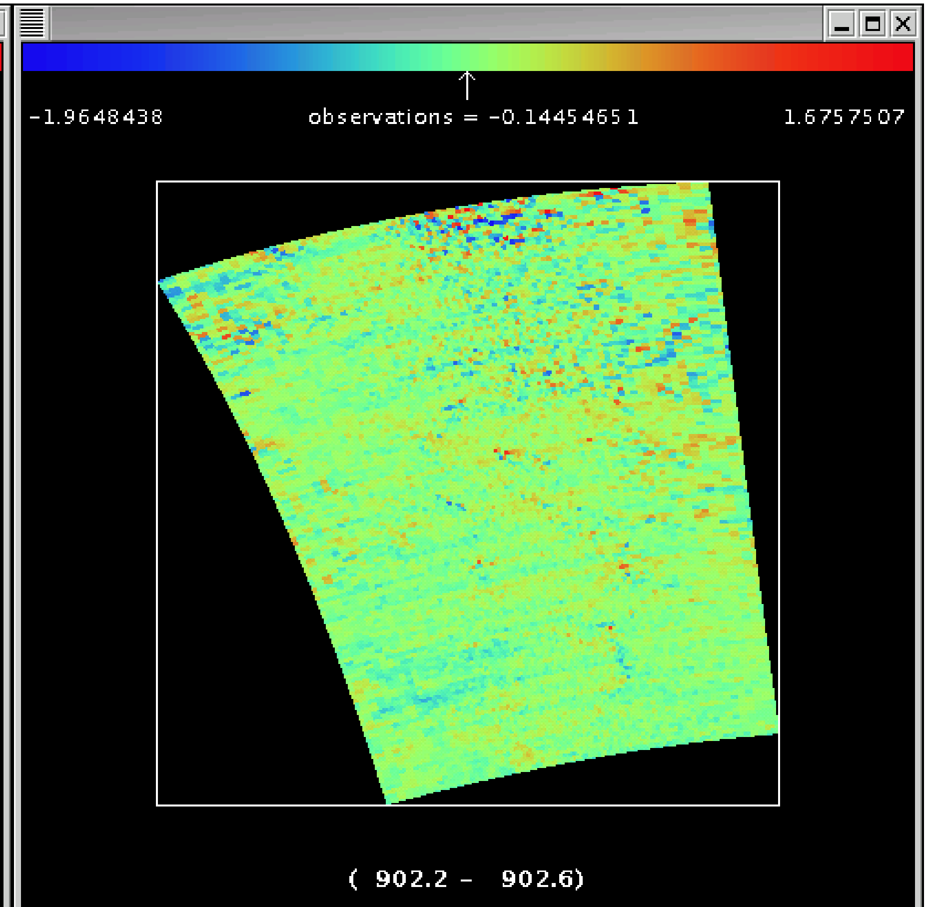


Filtered (1082.6 cm^{-1}- 1083.1 cm^{-1})

$902.2 – 902.6\ cm^{-1}$
Brightness Temperature Differences

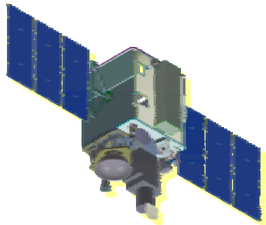






Original ($902.2\ cm^{-1}$- $902.6\ cm^{-1}$) Filtered ($902.2\ cm^{-1}$- $902.6\ cm^{-1}$)

PNF on AIRS data: Conclusions



- PNF
- allows for estimation of random component of instrument noise (using Earth Scene Data)
- filters out random component of instrument noise
- filters out part of popping noise