Hyperspectral Data Noise Characterization using PCA: Application to AIRS

Dave Tobin, Hank Revercomb, Paolo Antonelli, Ken Vinson
CIMSS/SSEC/UW-Madison

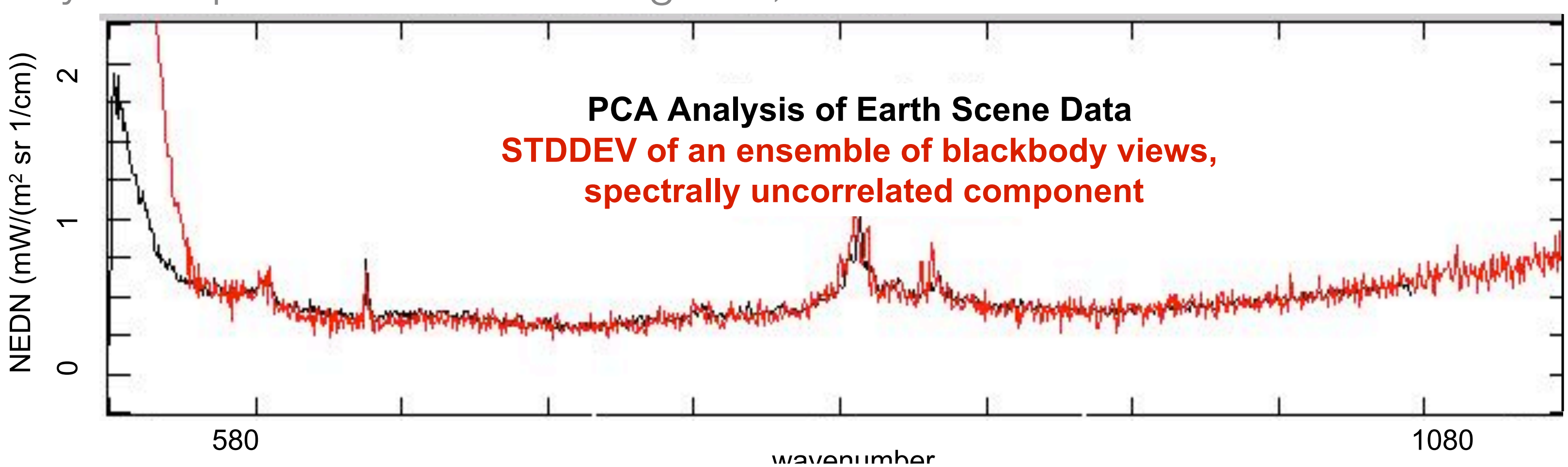
Third Annual Advanced High Spectral Resolution Infrared Observations Workshop
Madison, WI
26-28 April 2006

Outline

- Introduction
  - Approach: Dependent set PCA of AIRS Earth scene data
  - Results with simulated data
- NeDT investigations
  - Comparisons to blackbody estimates
  - Signal dependence
  - Spectrally correlated
- Non-gaussian behavior
  - Nσ events, “popping”, “striping”
- Inspection of PCs and individual spectra, etc.
  - Array correlated noise
  - A/B state artifacts
- Summary, Conclusions

Introduction

- Noise estimation/characterization using Earth scene data is made possible by PCA noise filtering
  - Huang et al., Application of Principle Component Analysis to High-Resolution Infrared Measurement Compression and Retrieval, *JAM*, 40, 365-388, 2001.
  - Antonelli et al., A principle component noise filter for high spectral resolution infrared measurements, *JGR*, 109, D23102, 2004.
  - Goldberg et al., Principle Component Analysis of AIRS Data, Workshop on Assimilation of high spectral resolution sounders in NWP, ECMWF, June 2004.
  - Rodgers, Application of Singular Value Decomposition to High Spectral Resolution Measurements, ASSFTS 2005.
  - Turner et al., Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis. *Journal of Atmospheric and Oceanic Technology*, accepted, 2006.
- An early example from the Scanning-HIS, our aircraft based FTS sensor:



Introduction, cont.

- One of the AIRS team “Early L1B verification assignments”

hha 1 November 01	**Earth scene based IR level 1b evaluation between launch+2 and launch+5 months**	**Concept defined**	**Initial Prototype evaluation using simulated data documented**	**Data input requirements documented and availability verified**	**Sensitivity analysis documente ation complete**	**Macro ready for real data**		**Launch+ 3 months Report on first real data**	**Launch+ 5 months Report**
5. Noise evaluation:									
	Verify level1b supplied noise estimates using using the statistics of adjacent footprint differences	Aumann	/hha/index.html						
	Noise evaluation using adjacent footprint difference under extended clear conditions (more than 2 footprints).	McMillin							
	Evaluate noise covariance and radiometric crosstalk.	McMillin							
	NeDT estimation using Earth scene data	Tobin							
	Evaluate noise covariance matrix using (ECMWF.calculated-observed).clear using fast RTA	Susskind							

Early_L1b_verification_assignments.ppt, HHA, 7 Nov 2001

Approach: Dependent Set PCA of AIRS L1B Earth Scene Data

Using an AIRS L1B granule (6 minutes of data, 90*135 FOVs)

1) Exclude bad channels using the AIRS team prescription

- Typically retains ~2120 of 2378 channels

2) Noise normalize the radiance spectra using an initial noise estimate (divide by NeDN)

- This initial estimate can also come from non-noise-normalized PCA

3) Generate principle components (PCs) of the covariance matrix of the noise normalized spectra

4) Reconstruct the spectra using a reduced number of PCs

- Using the method described in Turner et al. (previous presentation) to determine this number

5) Remove the noise normalization (multiply the reconstructed spectra by the initial noise estimate) and perform analyses of the reconstruction error to characterize and derive noise estimates.

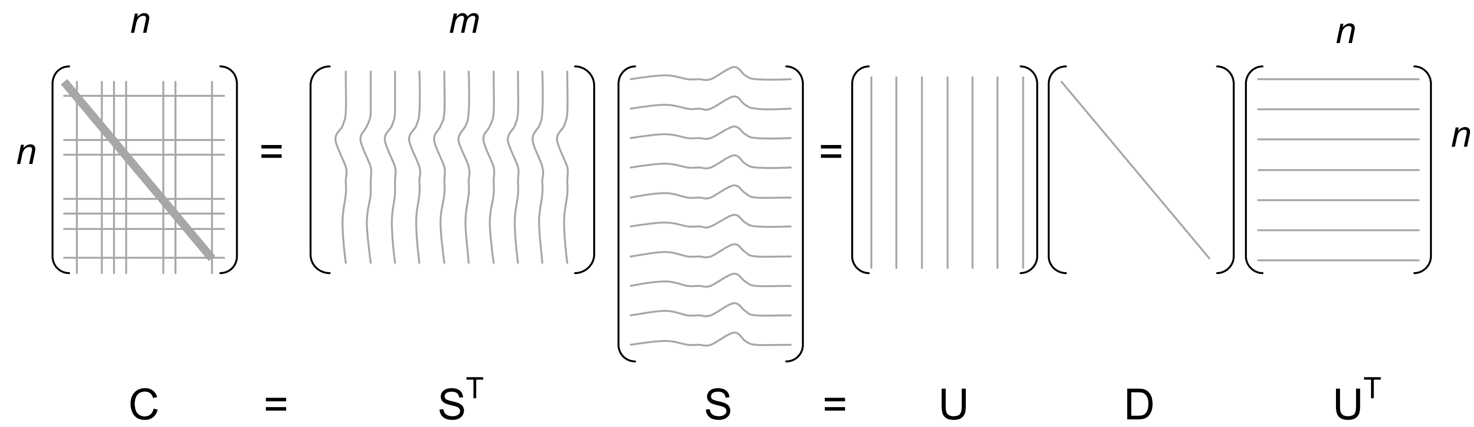
- e.g. NEDN = STDDEV(N_{orig}-N_{recon}) is an estimate of the *spectrally uncorrelated random noise*

Two caveats re: this presentation

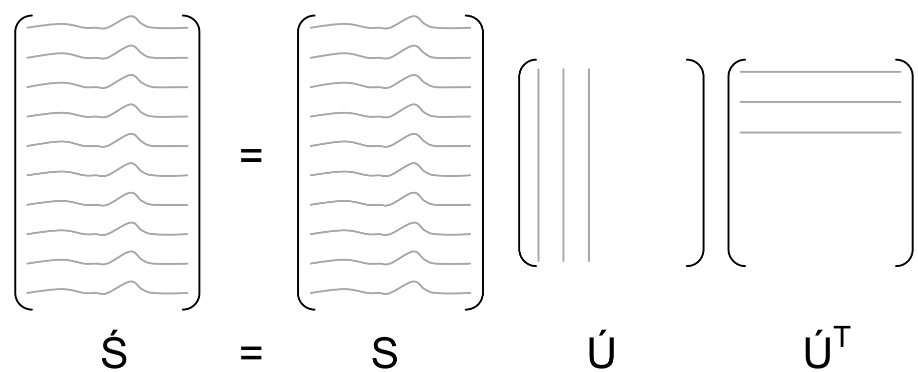
- Correction factor to PCA NEDN estimates to account for the random noise included in the n_T retained PCs
  - Not included in plots shown here
  - But the effect is small. $[n/(n-n_T)]^{1/2}$, e.g. $(2120/2060)^{1/2} = 1.015$

- Reconstruction with a reduced number of PCs introduces spectral correlation
  - This effect is also small, perhaps negligible (see Antonelli 2004, Turner 2006), but I have not quantified it here.

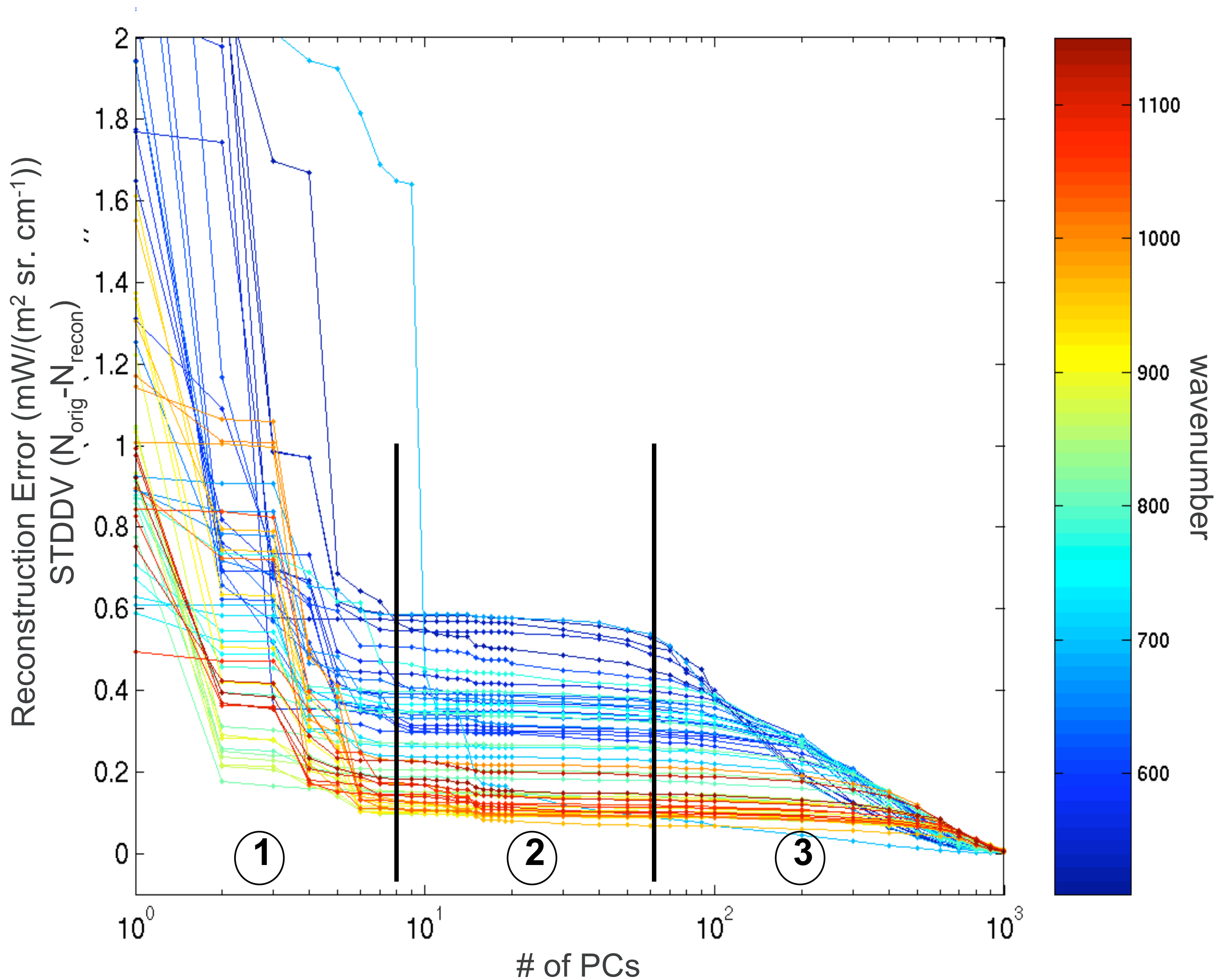
Terms, Equations



Variable	Description
n	number of channels
m	number of spectra
$S = N_{orig}/NEN_{init}$	noise normalized spectra
$C = S^T S = U D U^T$	covariance matrix of S
D	eigenvalues
U	eigenvectors
Ú	truncated eigenvector matrix
n_T	number of eigenvectors in Ú
$Ś = S\ Ú\ Ú^T\ NEN_{init}$	reconstructed spectra

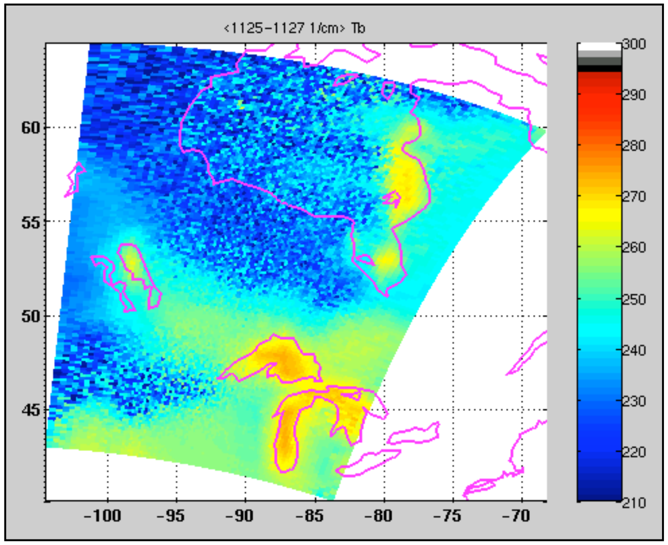


Reconstruction Error versus #PCs

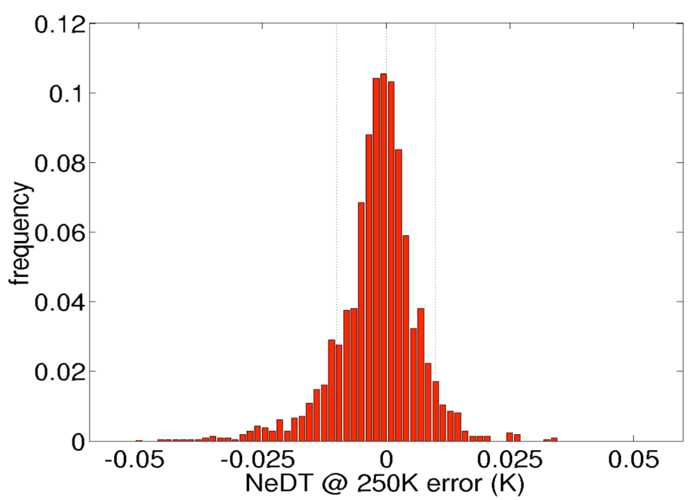


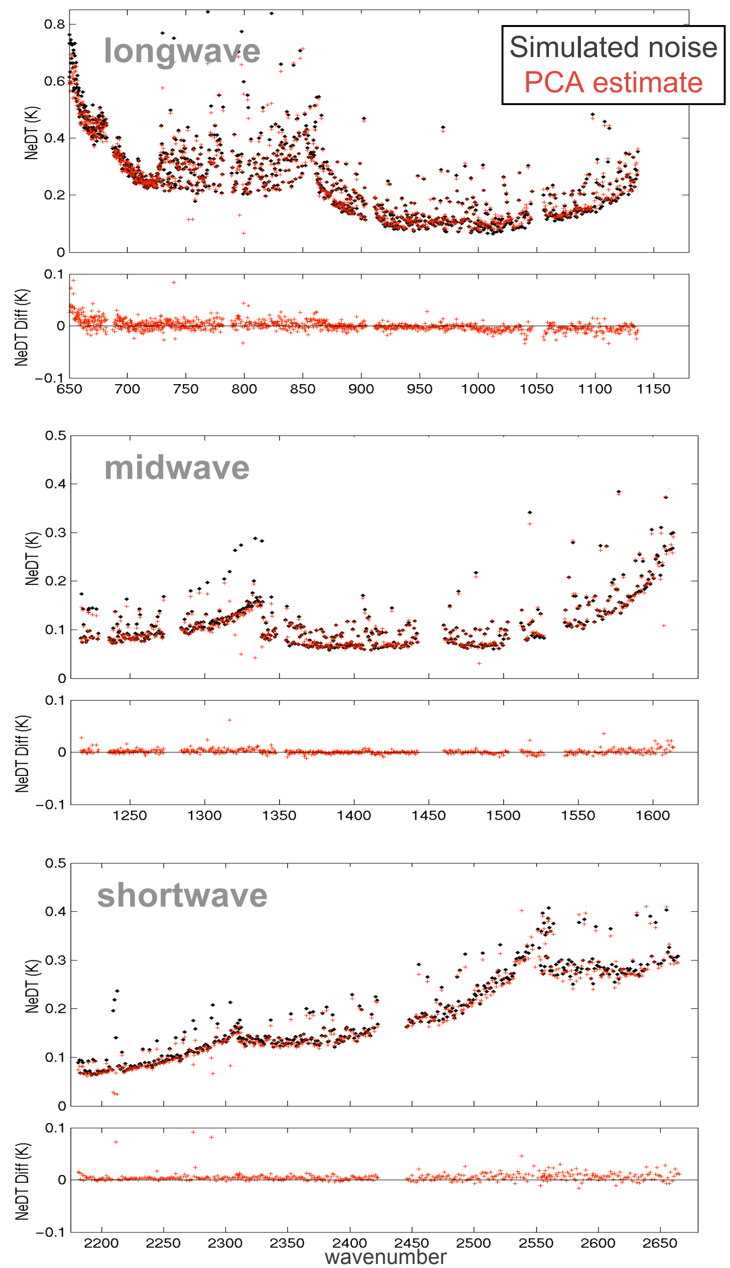
Example results for simulated data (purely Gaussian random noise)

15-Dec-2000 granule 081



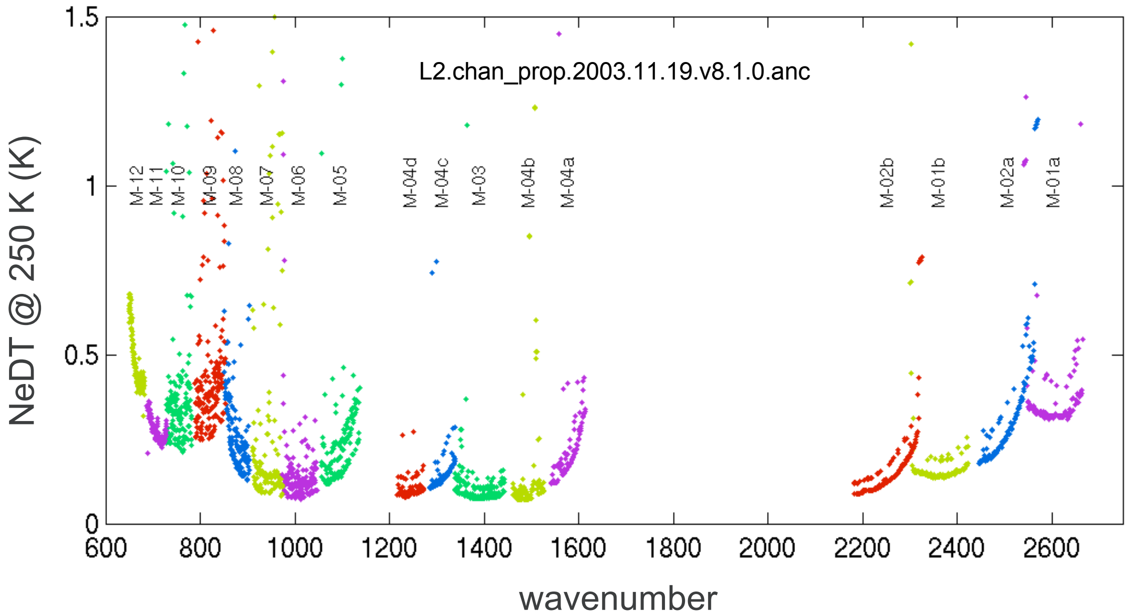
Distribution of differences



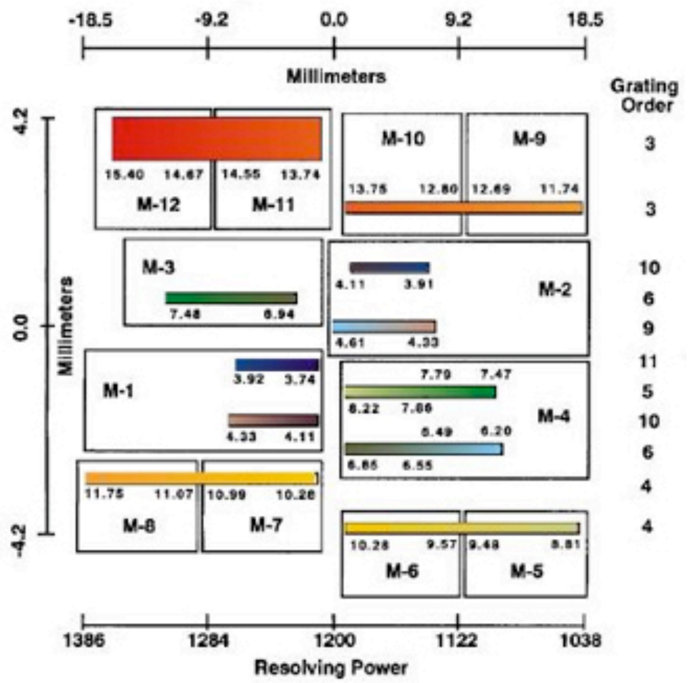


AIRS Noise

- *Total noise* estimates (NEN, NEDT@250K) derived from on-board blackbody and space views
  - provided in 1) Channel properties files and 2) L1B granule files



Focal plane array map



- Signal dependence AIRS L1B ATBD, l1bqa_changes.pdf @ GDAAC
- Array correlated noise Pagano, Weiler, AIRS Design Files #614, #620
- Striping e.g. Gaiser, Dec 2004 AIRS STM
- “Popping” M. Weiler, Nov 2002 AIRS TM; M. Weiler, SPIE Proc. 5882, 2005.

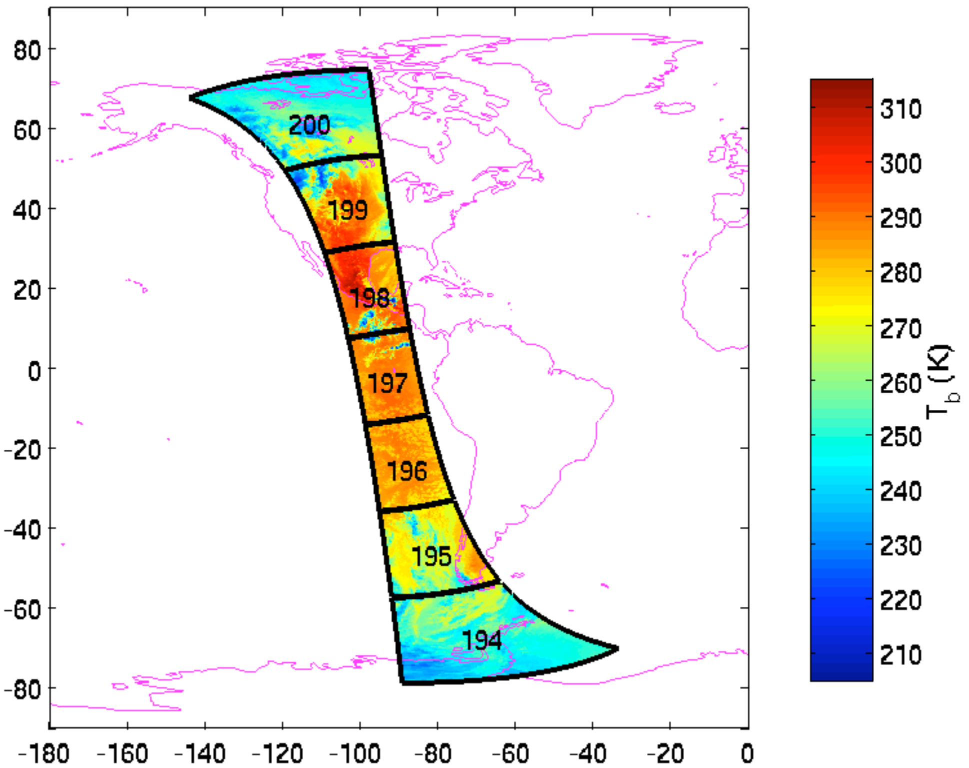
NEDT investigations

- STDDEV(N_{orig} - N_{recon})

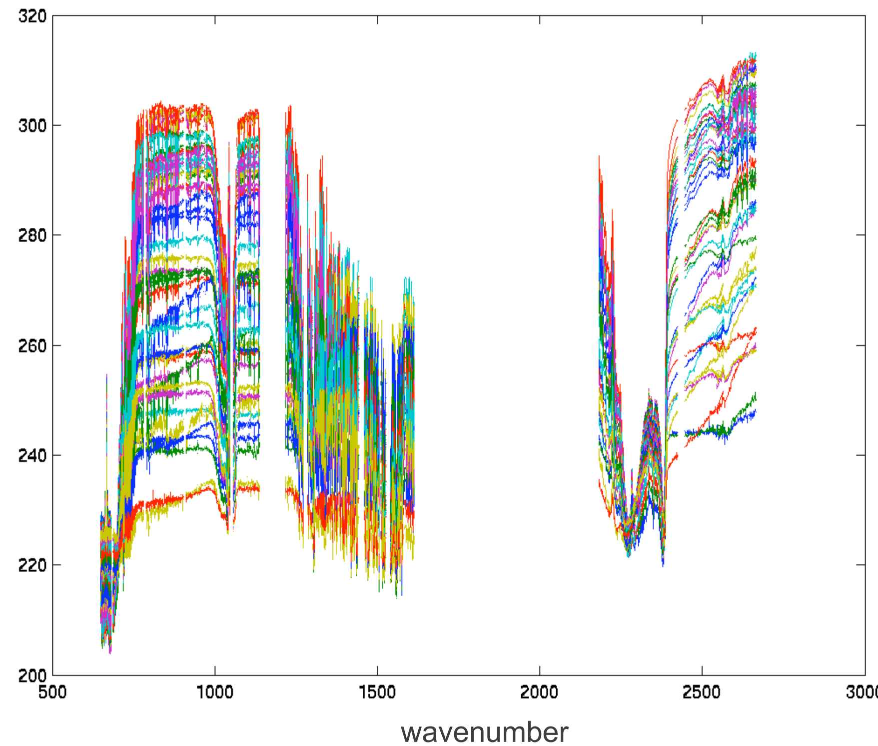
Analysis of 01 April 2005 data (v4.0.9 L1B)

7 ascending (daytime) granules

1000 cm^{-1} brightness temperatures



sample spectra

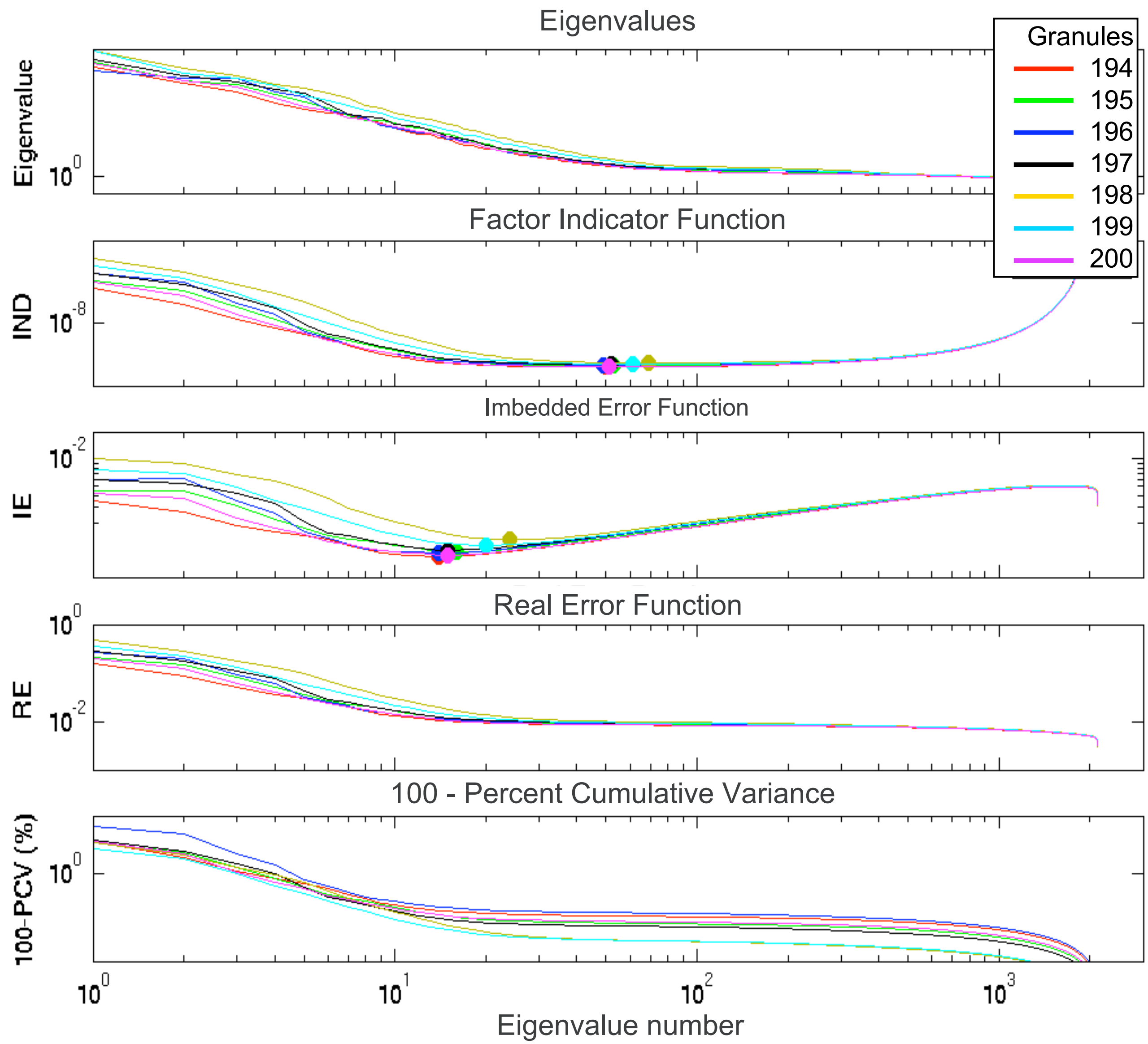


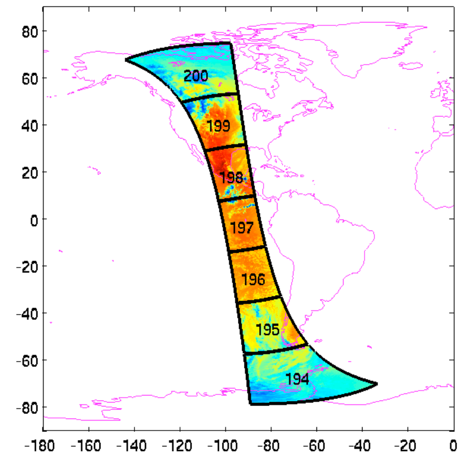
Channel Selection

Follows the guidelines provided by the AIRS project:

1. AB_State (in channel properties file) <= 2. (If AB_State > 2, the channel has known radiometric problems.)
2. NEDT@250K (computed from NeN in L1B granule file) <= 2K
3. Bits 6 (Anomaly in offset calculation), 5 (Anomaly in gain calculation), 4 (Pop detected), 3 (High Noise) of CalChanSummary (in L1B granule file) are not set.

Variance Metrics, following Turner et al.





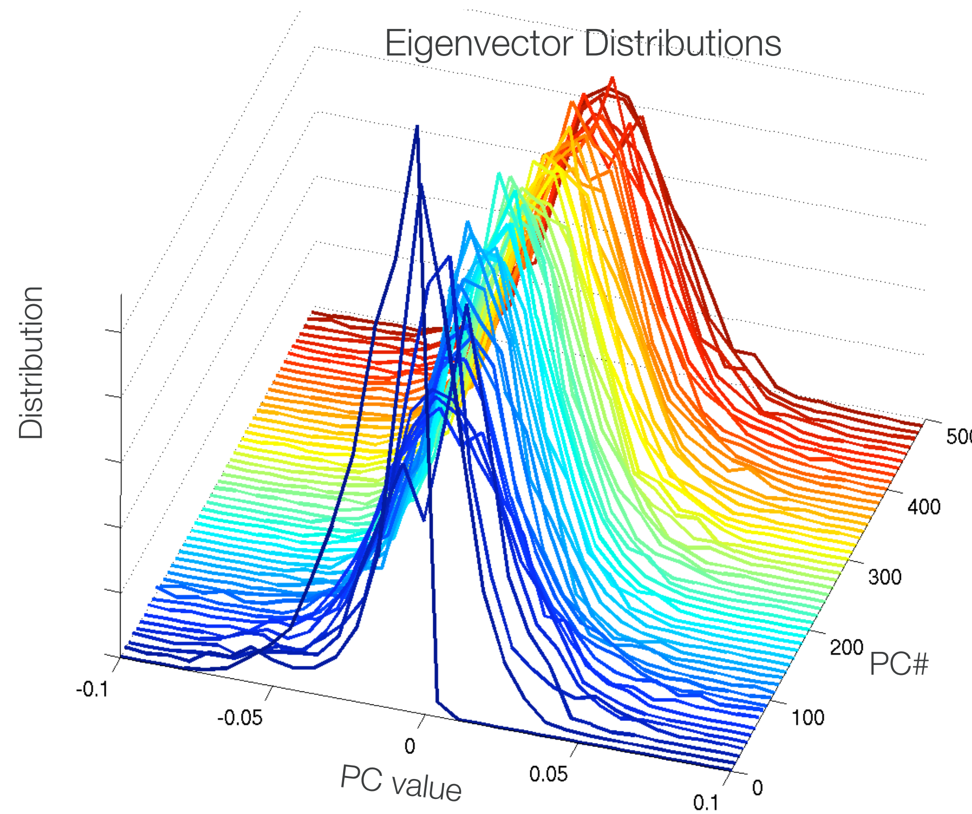
Gran.	IND	IE
194	50	14
195	53	16
196	49	14
197	52	15
198	69	24
199	61	20
200	51	15

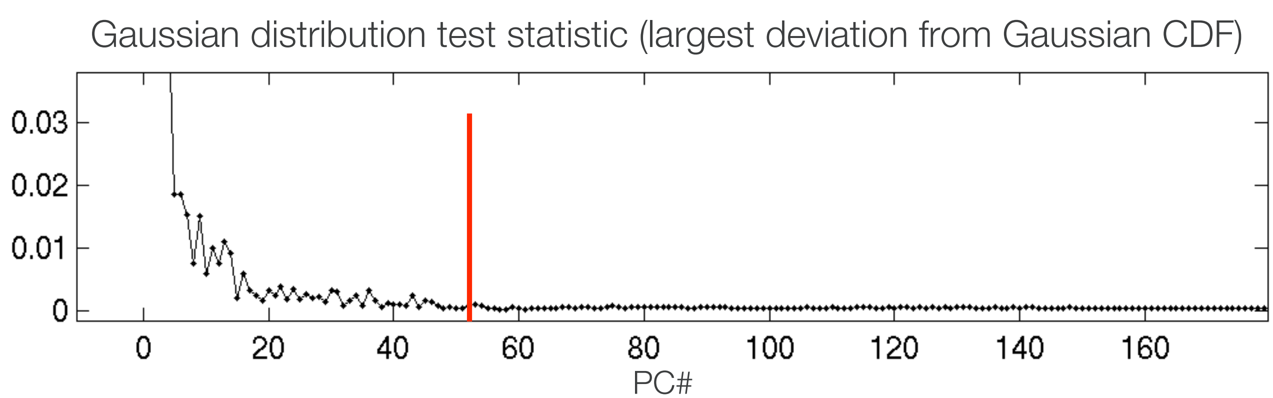
\# of PCs
used in the
reconstructions

Variance Metrics: “brute force” determinations

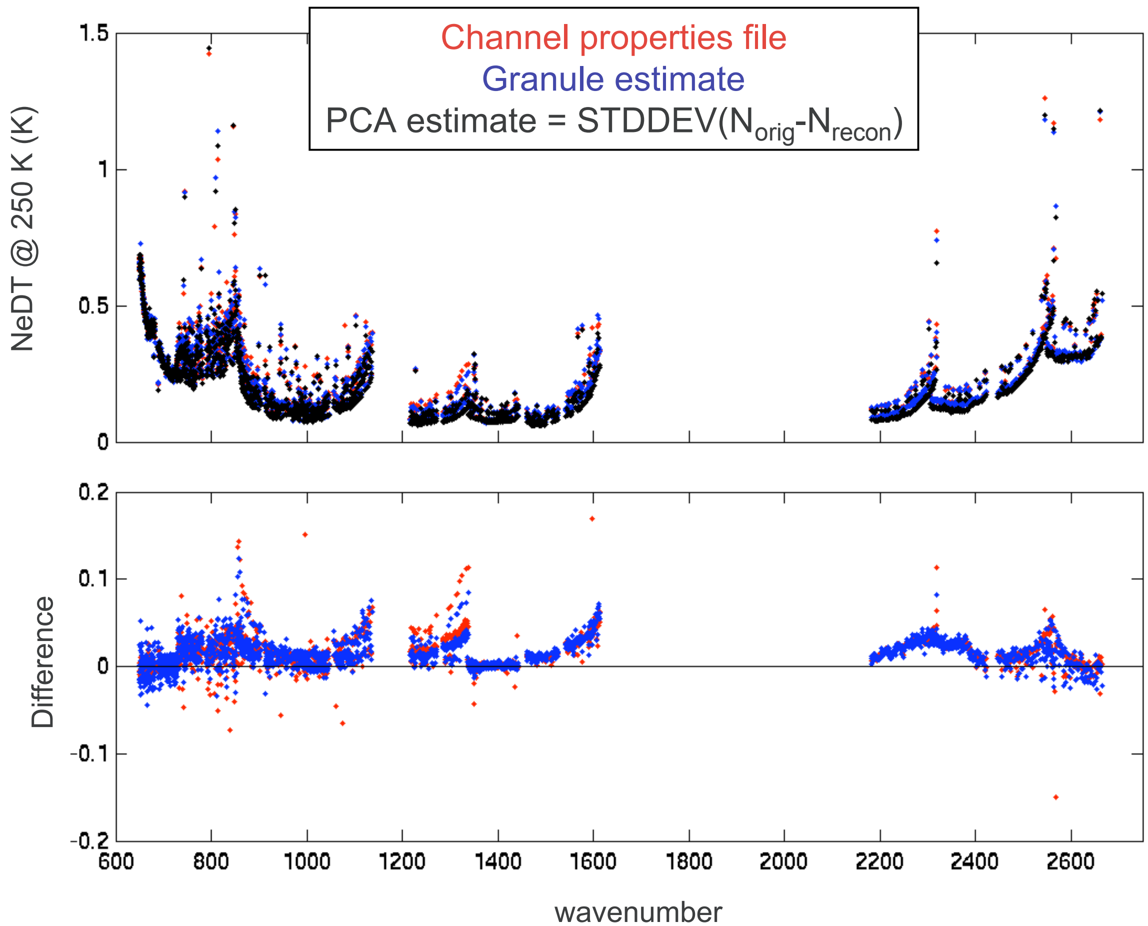
The high order (least significant) PCs should represent random white noise. Various normality tests performed on the PCs suggest that the number PCs used in the reconstructions (determined with the IND method) are reasonable.

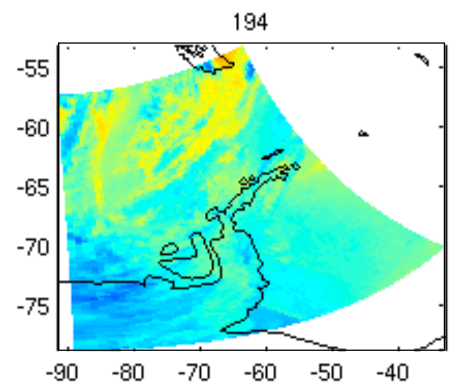
Example for granule 195 PCs:



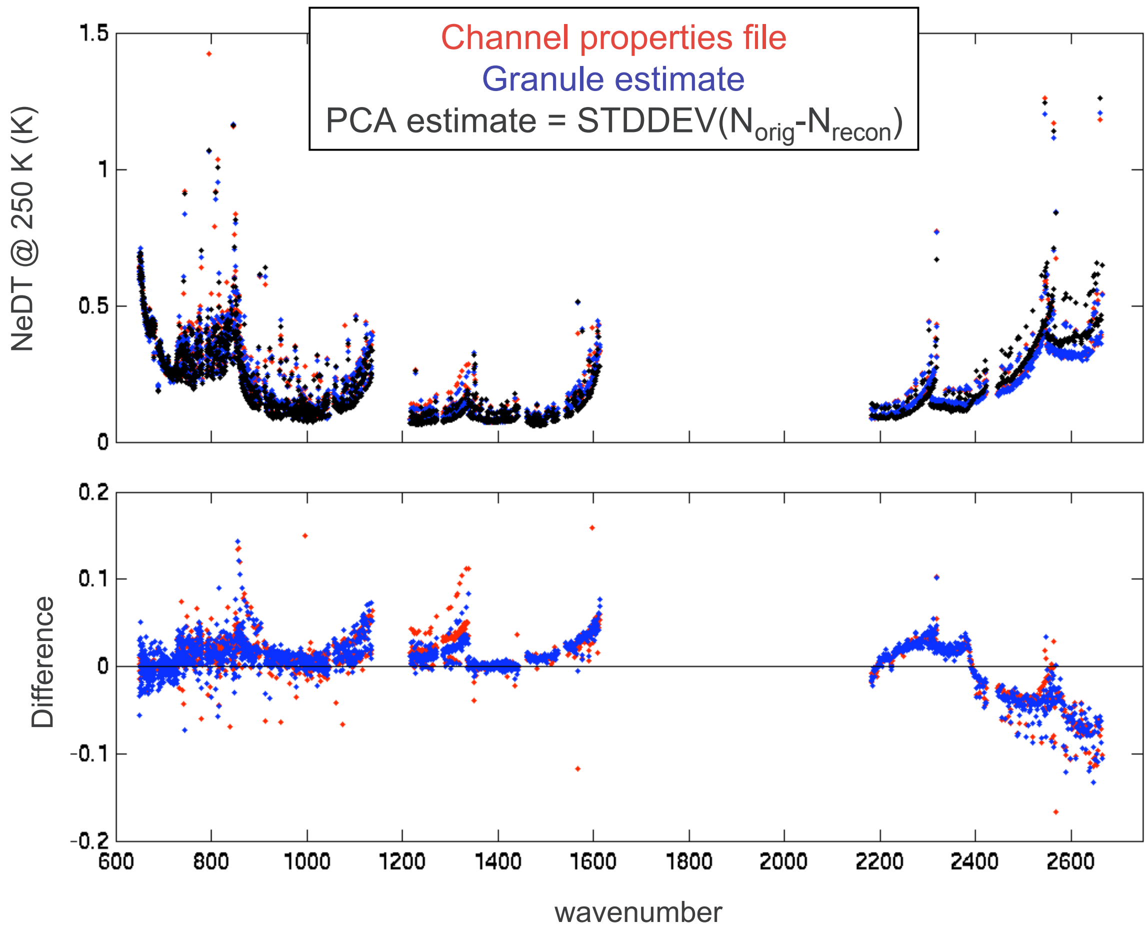


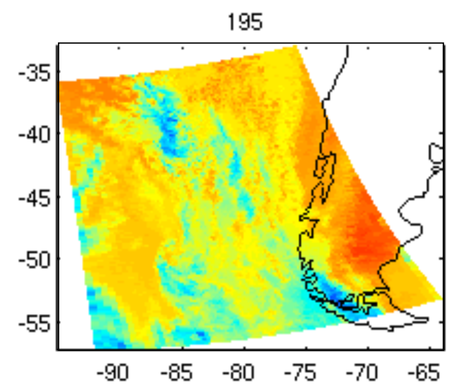
NeDT comparisons, granule 194



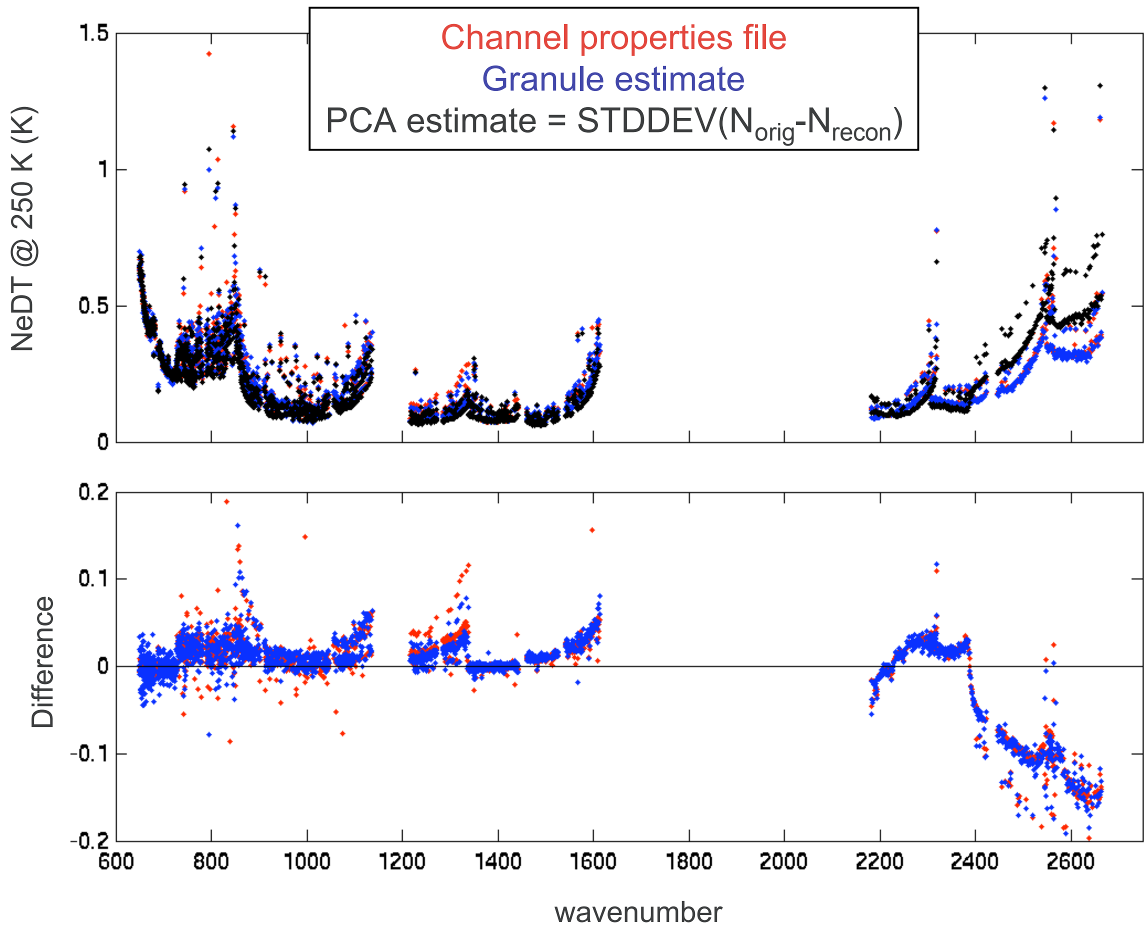


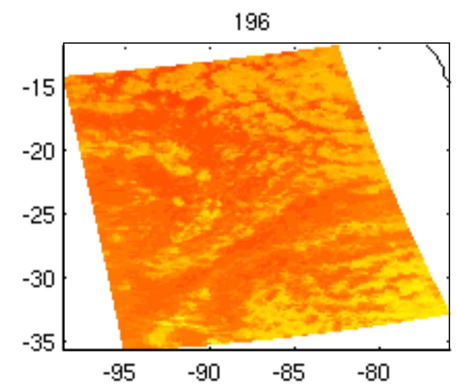
NeDT comparisons, granule 195



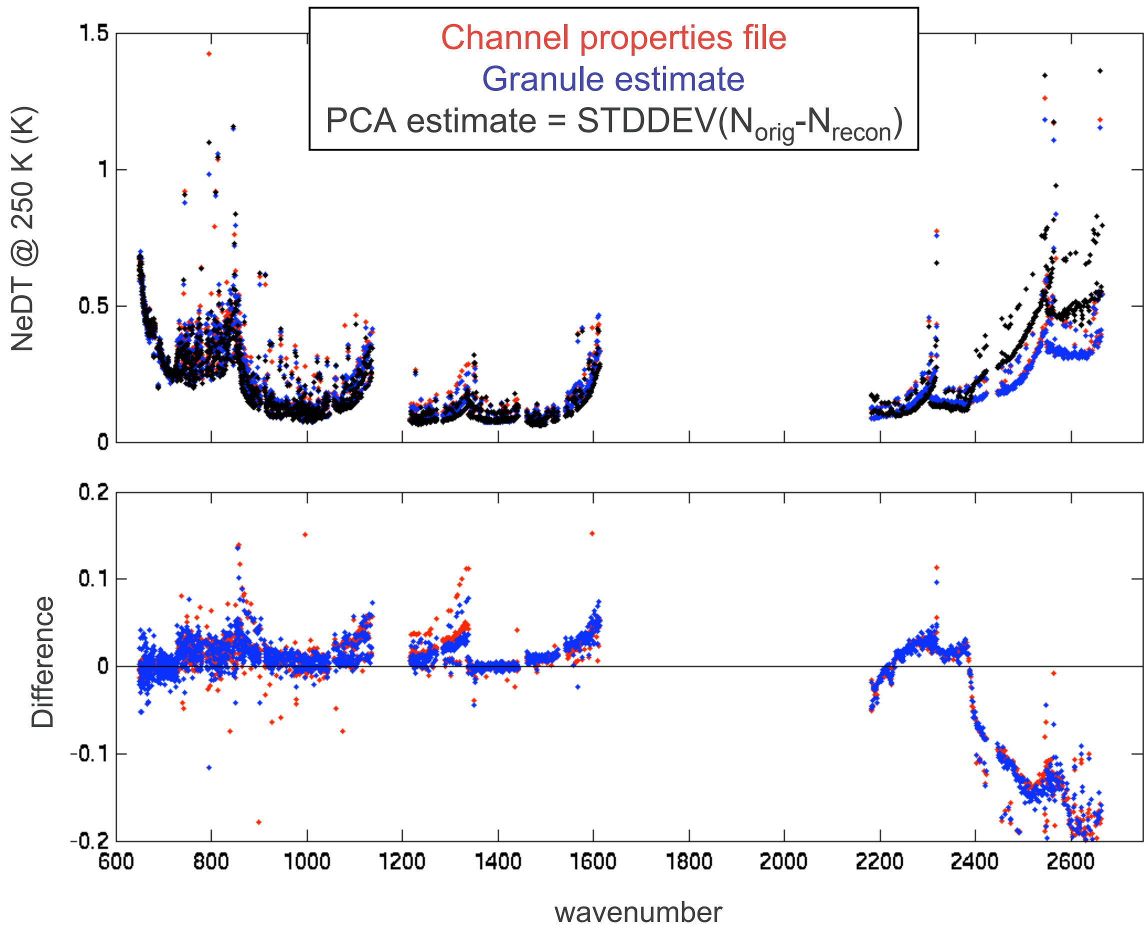


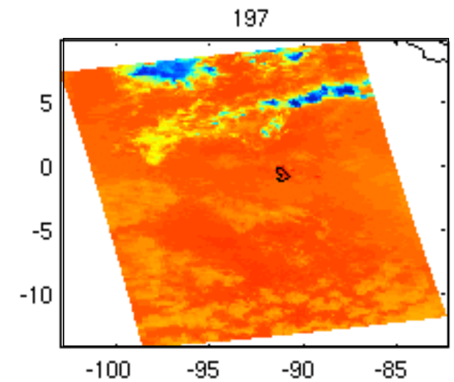
NeDT comparisons, granule 196



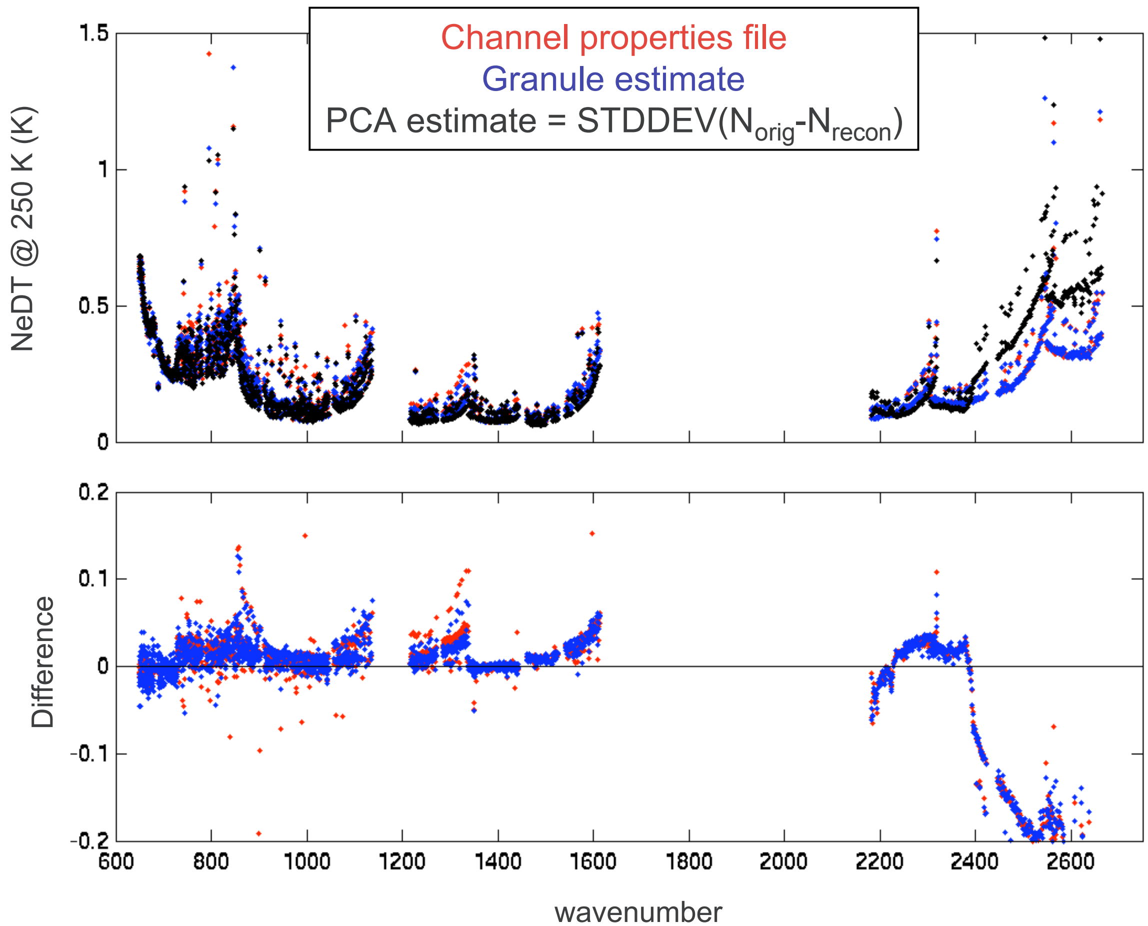


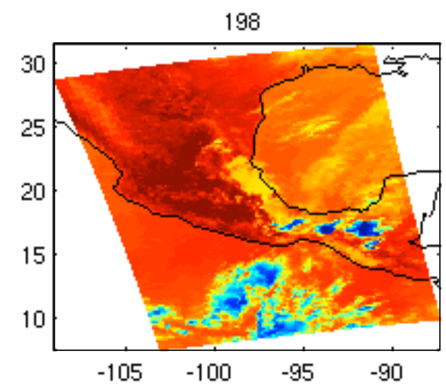
NeDT comparisons, granule 197



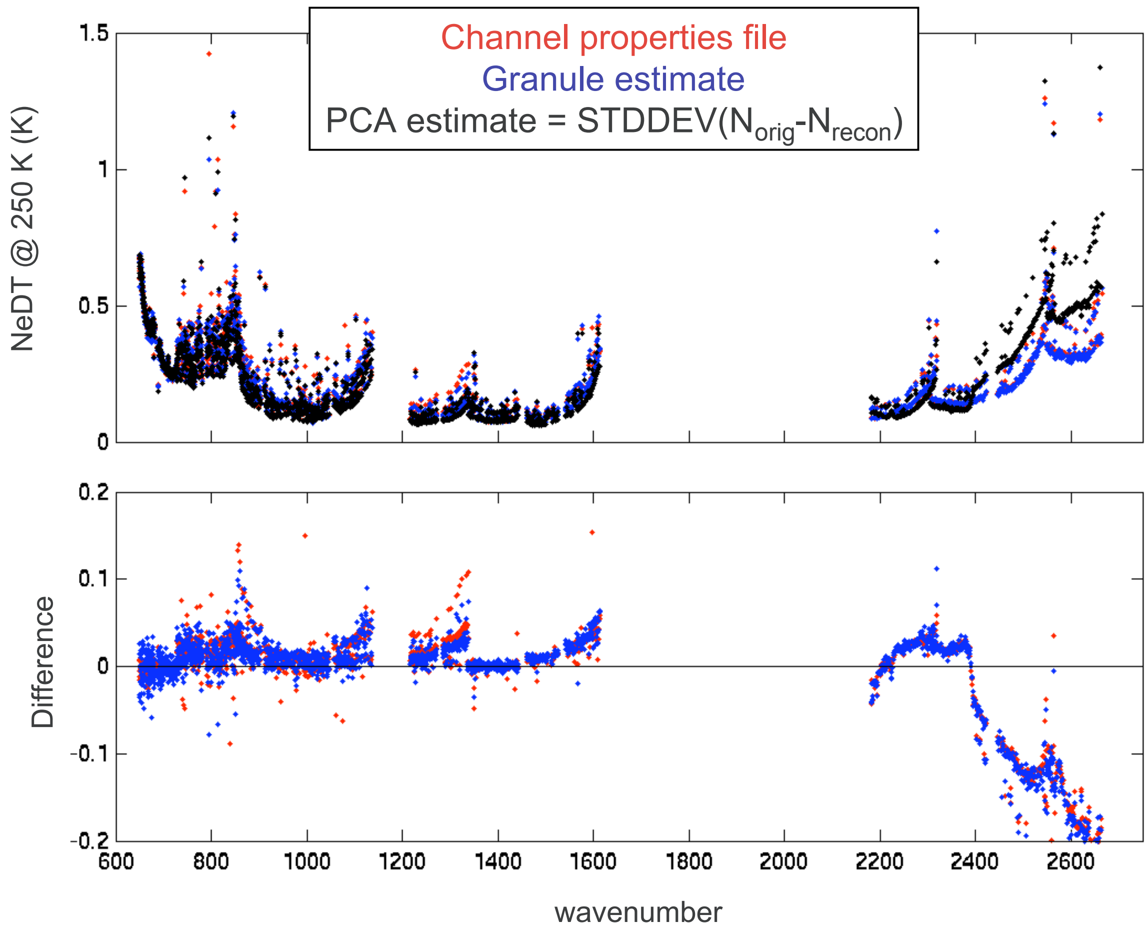


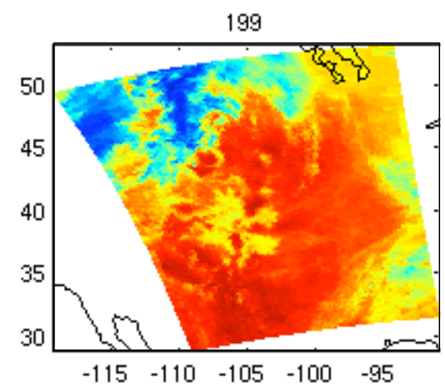
NeDT comparisons, granule 198



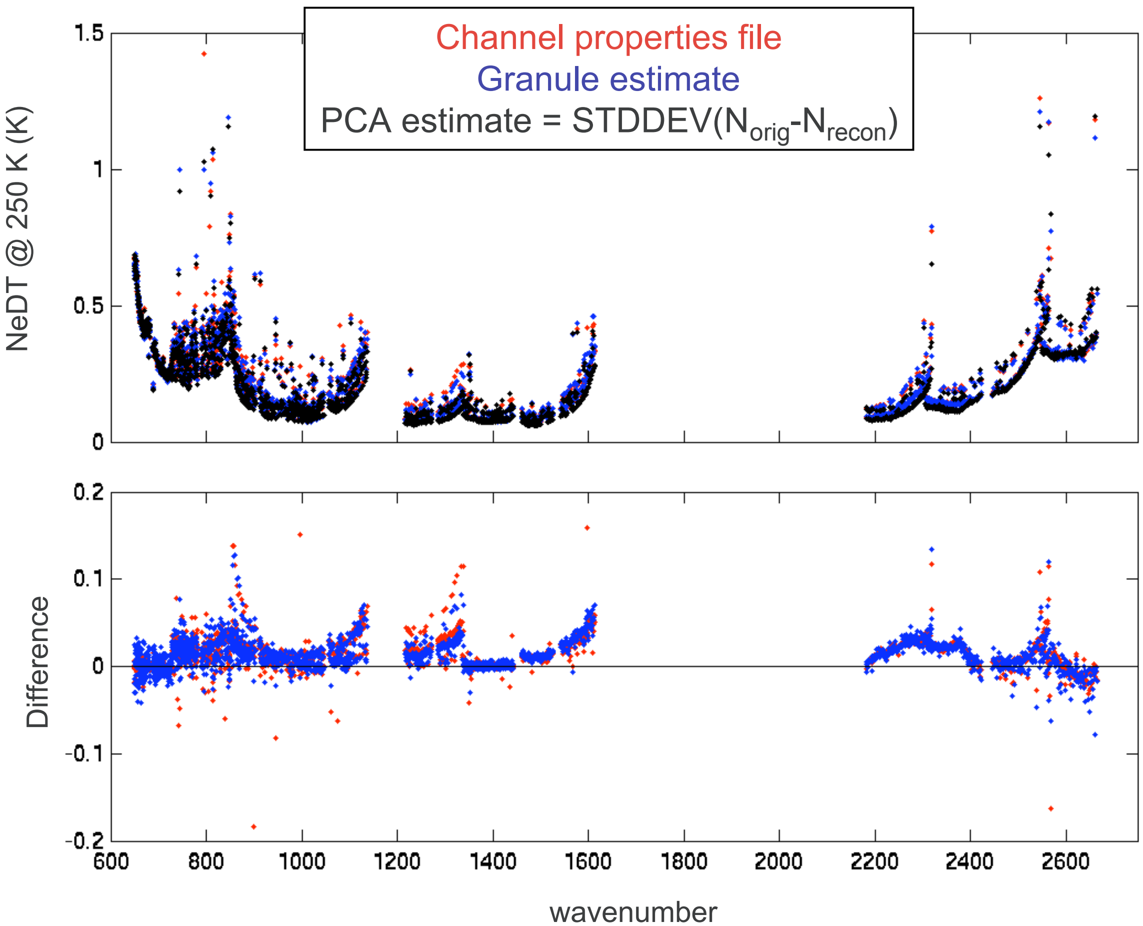


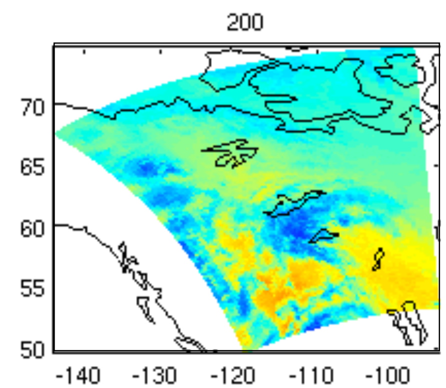
NeDT comparisons, granule 199



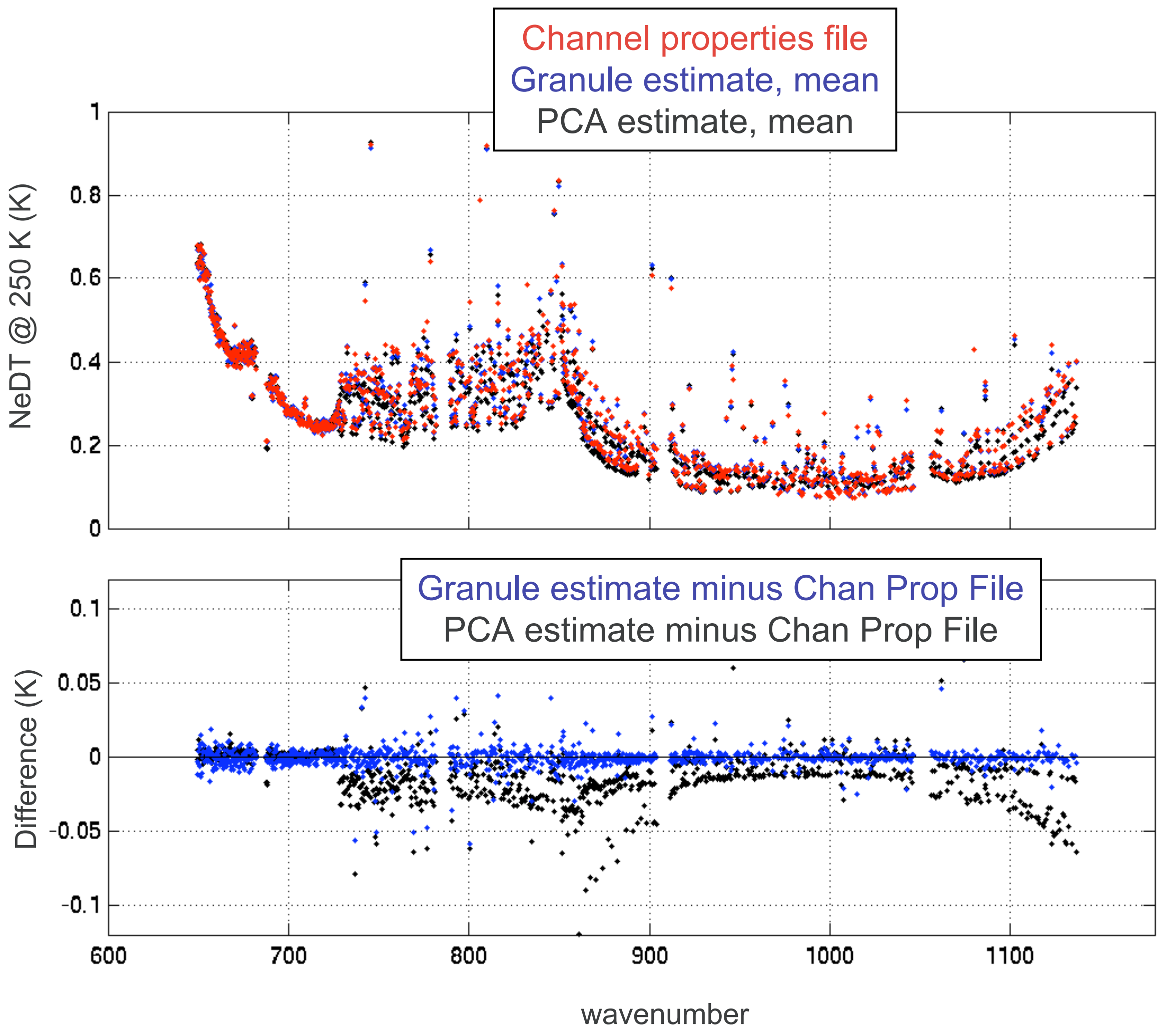


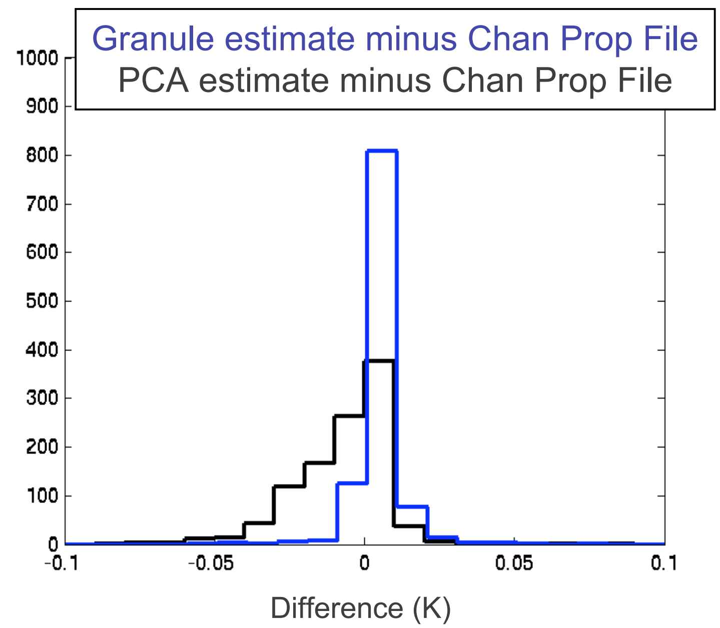
NeDT comparisons, granule 200





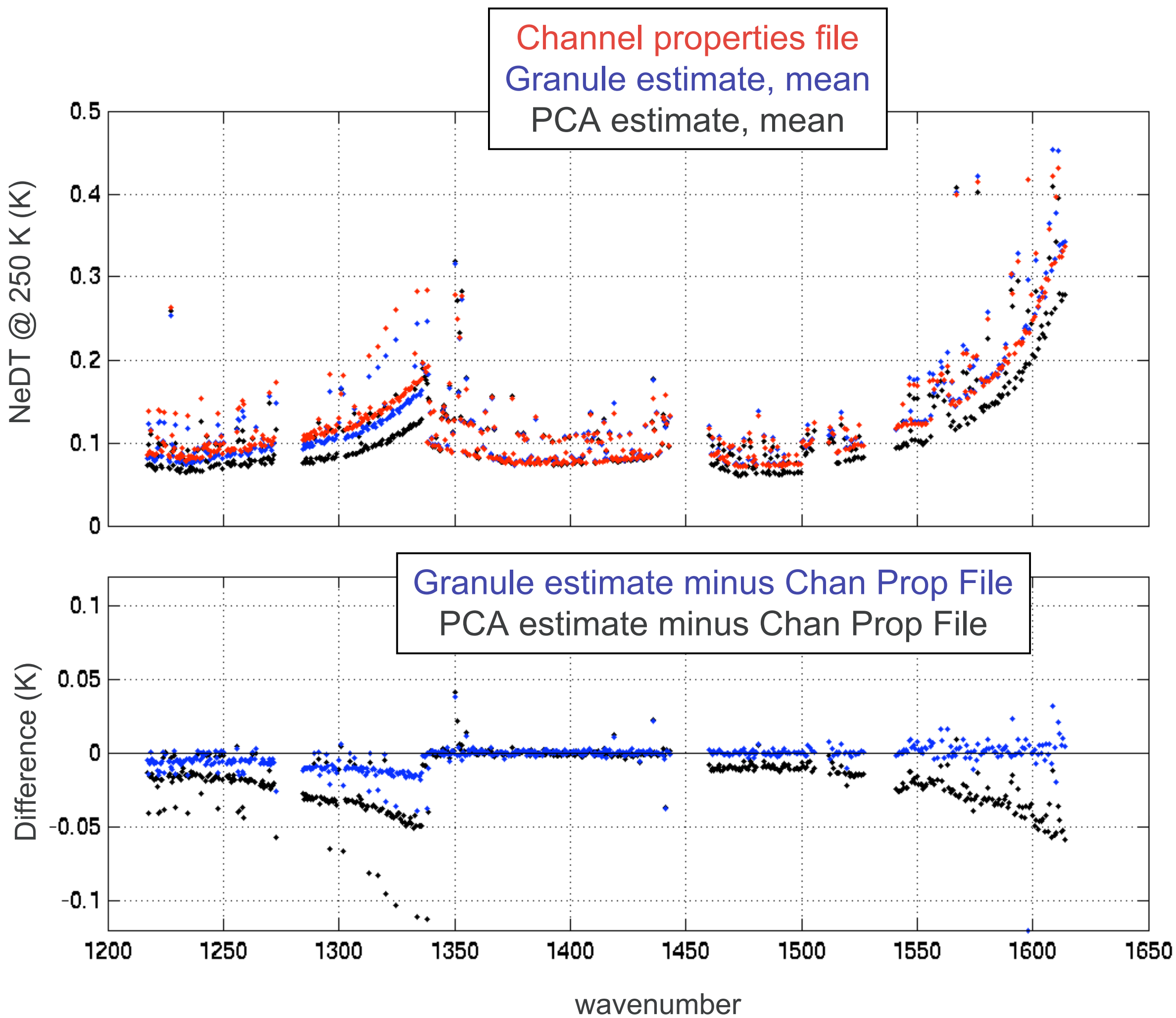
NeDT comparison summary, longwave

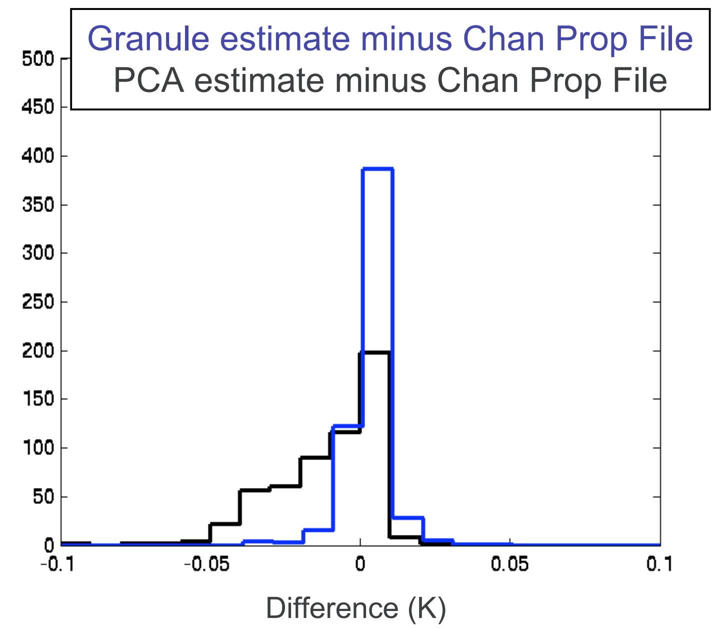




- Good agreement for the longwave PC arrays M-11 and M-12
- PCA estimates are slightly, but consistently, lower for the other arrays

NeDT comparison summary, midwave



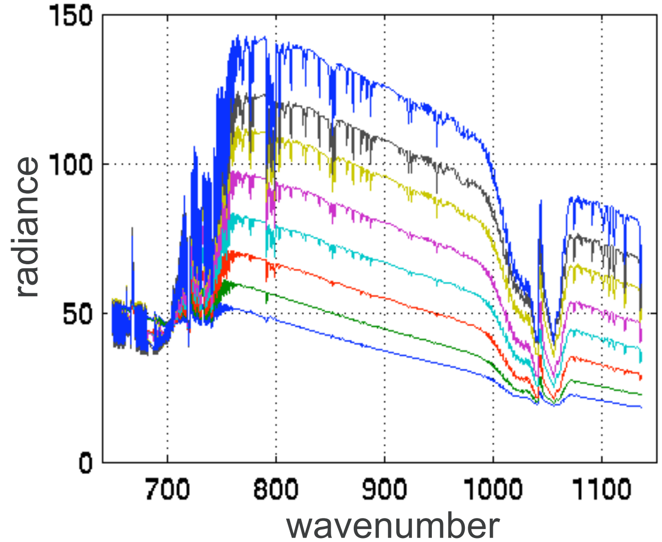


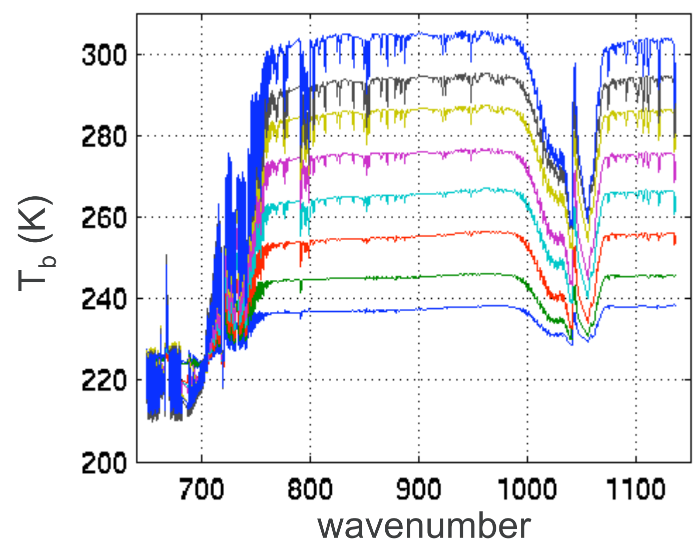
- Good agreement for array M-03. PCA estimates are slightly, but consistently, lower for the other arrays
- The granule-based estimate is also slightly lower than the channel properties estimate for arrays M-04c and M-04d

Scene dependence of NEN, longwave

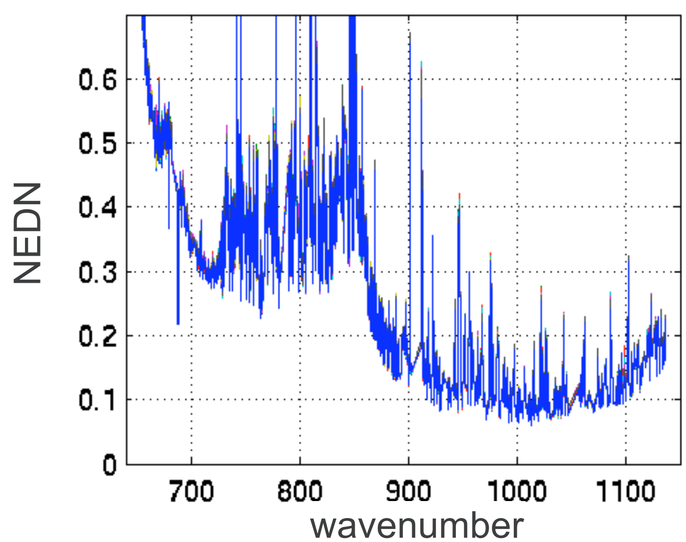
The data from all 7 granules is binned according to the SW window region signal levels. PCA is then performed for each bin and NEDN is reported as a function of signal level.

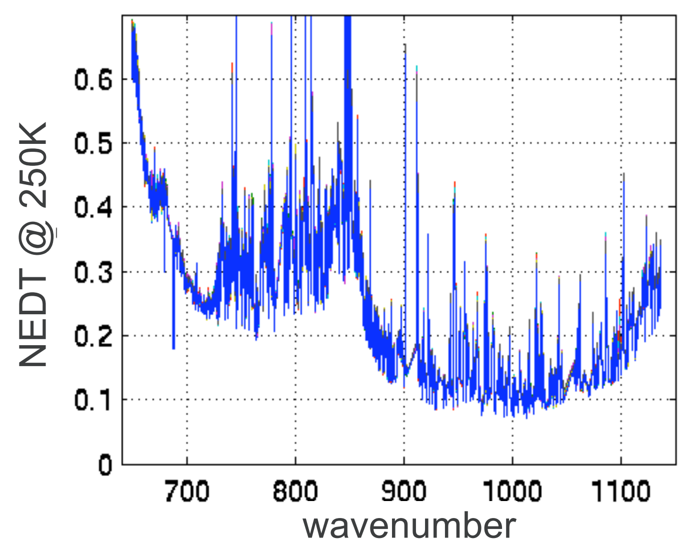
Mean radiance spectrum for each bin → converted to BT





NEDN for each bin → converted to NEDT@250K

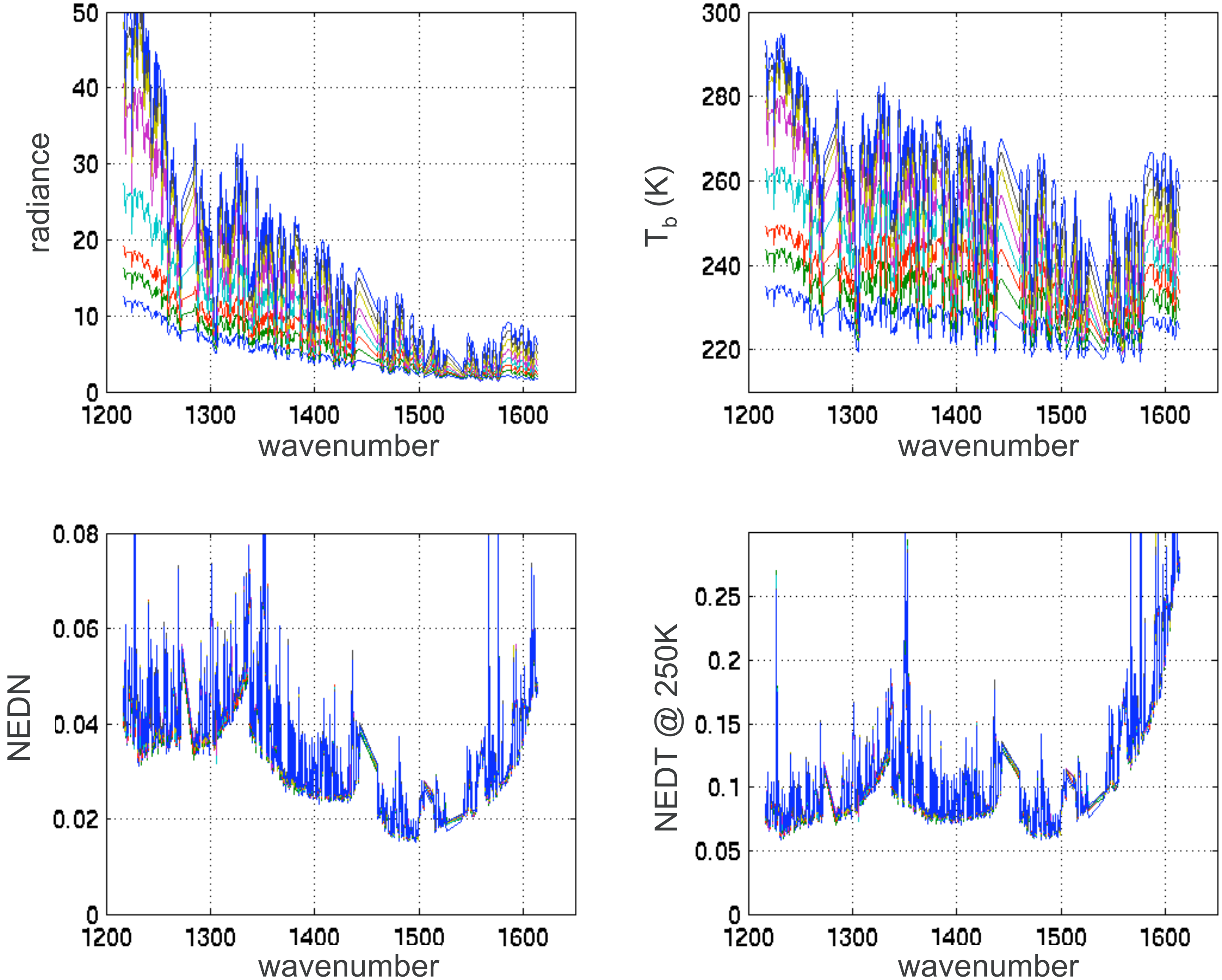




Bins:
245-255 K
255-265 K
265-275 K
275-285 K
285-295 K
295-305 K
305-315 K
315-325 K

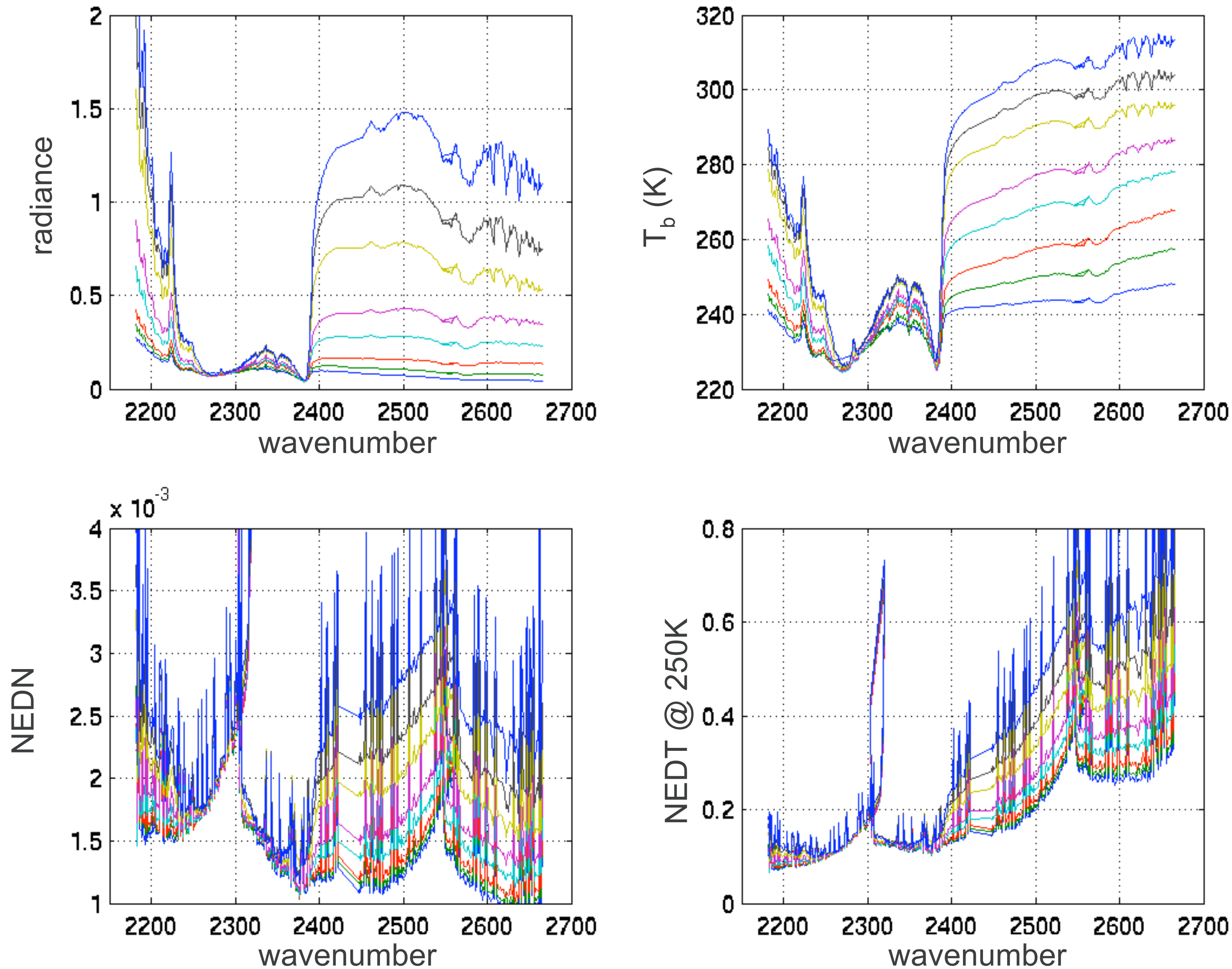
Scene dependence of NEN, midwave

Bins:
245-255 K
255-265 K
265-275 K
275-285 K
285-295 K
295-305 K
305-315 K
315-325 K



Scene dependence of NEN, shortwave

Bins:
245-255 K
255-265 K
265-275 K
275-285 K
285-295 K
295-305 K
305-315 K
315-325 K

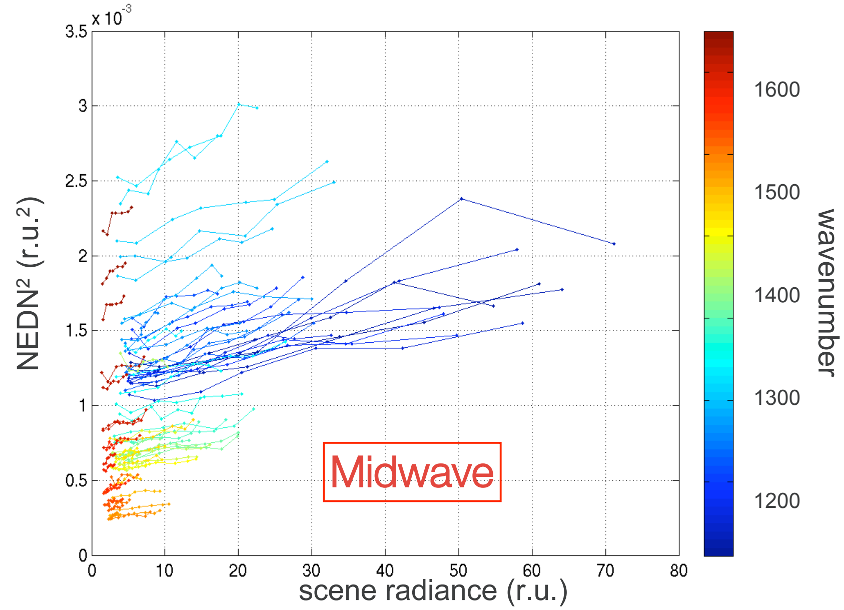


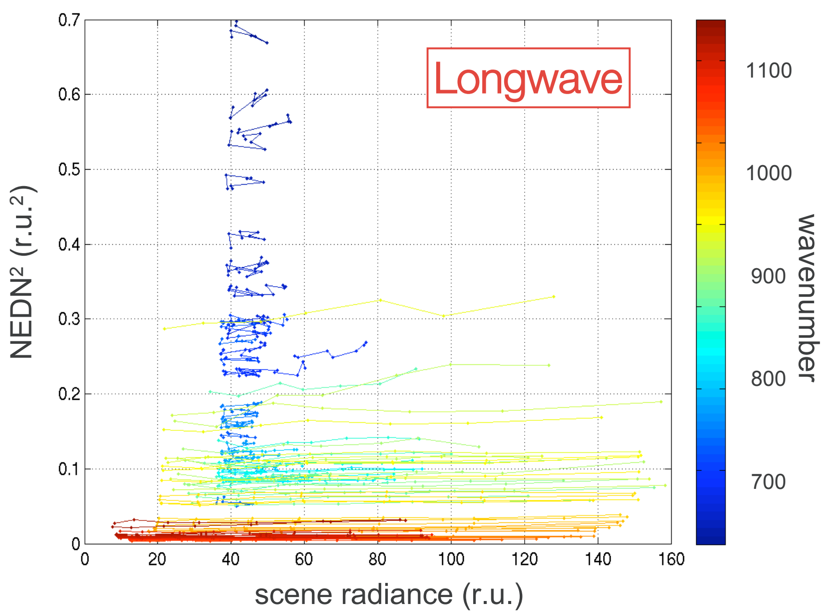
NEDN versus scene radiance

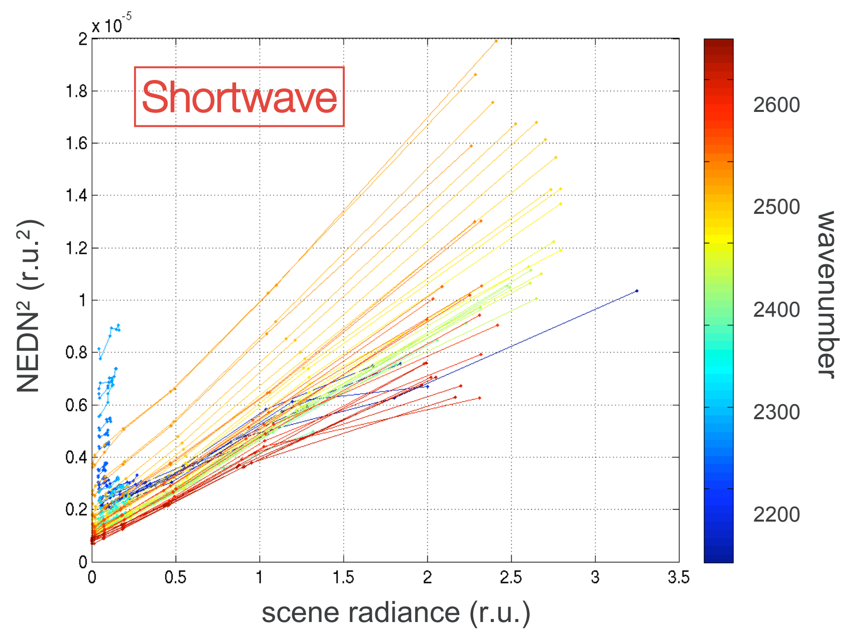
NEDN increases with sqrt{scene radiance}, consistent with photon noise. The total noise at scene temperature T is parameterized as

$$\mathrm{NEDN}(T) = [N(T)\,\gamma_{photon} + \mathrm{NEDN}_{thermal}^2]^{1/2}$$

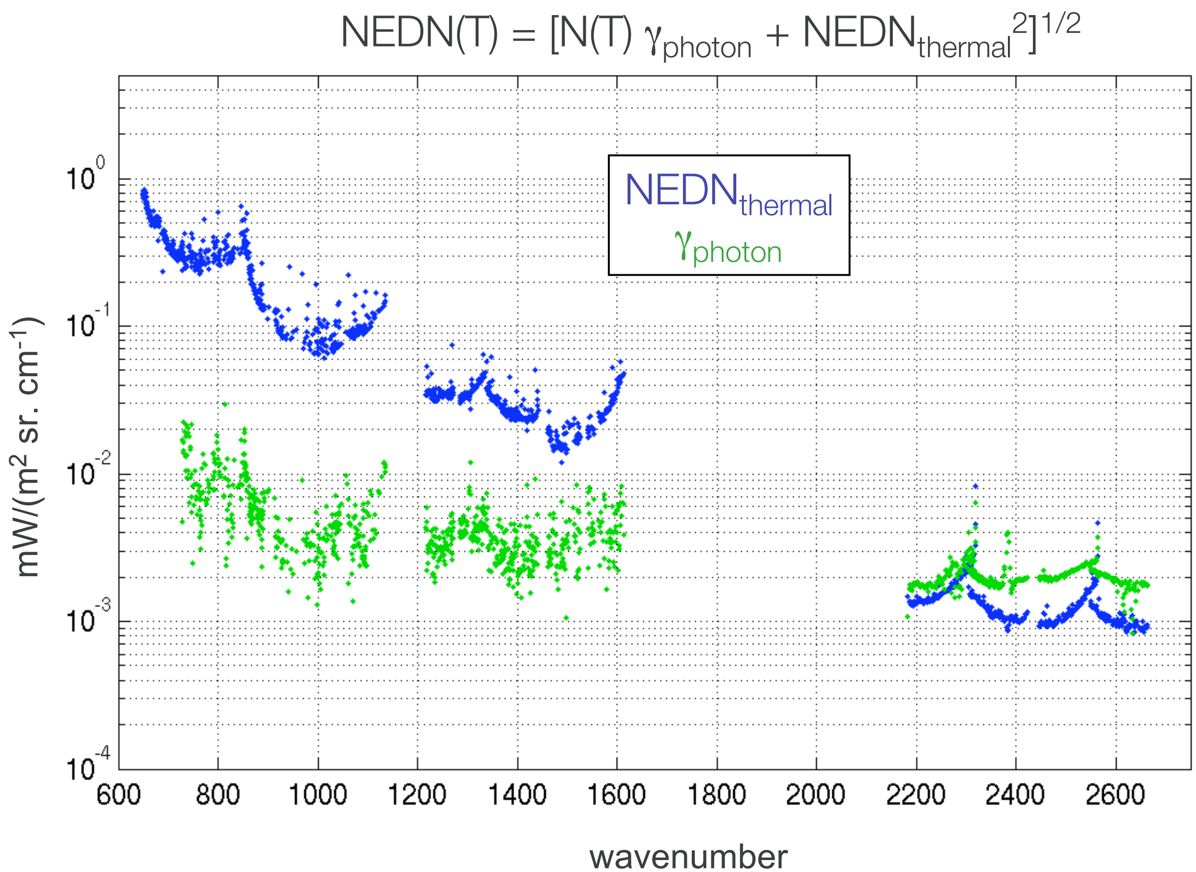
where $\mathrm{NEDN}_{thermal}^2$ (the y-intercepts) and γ_{photon} (the slopes) are determined for each channel.



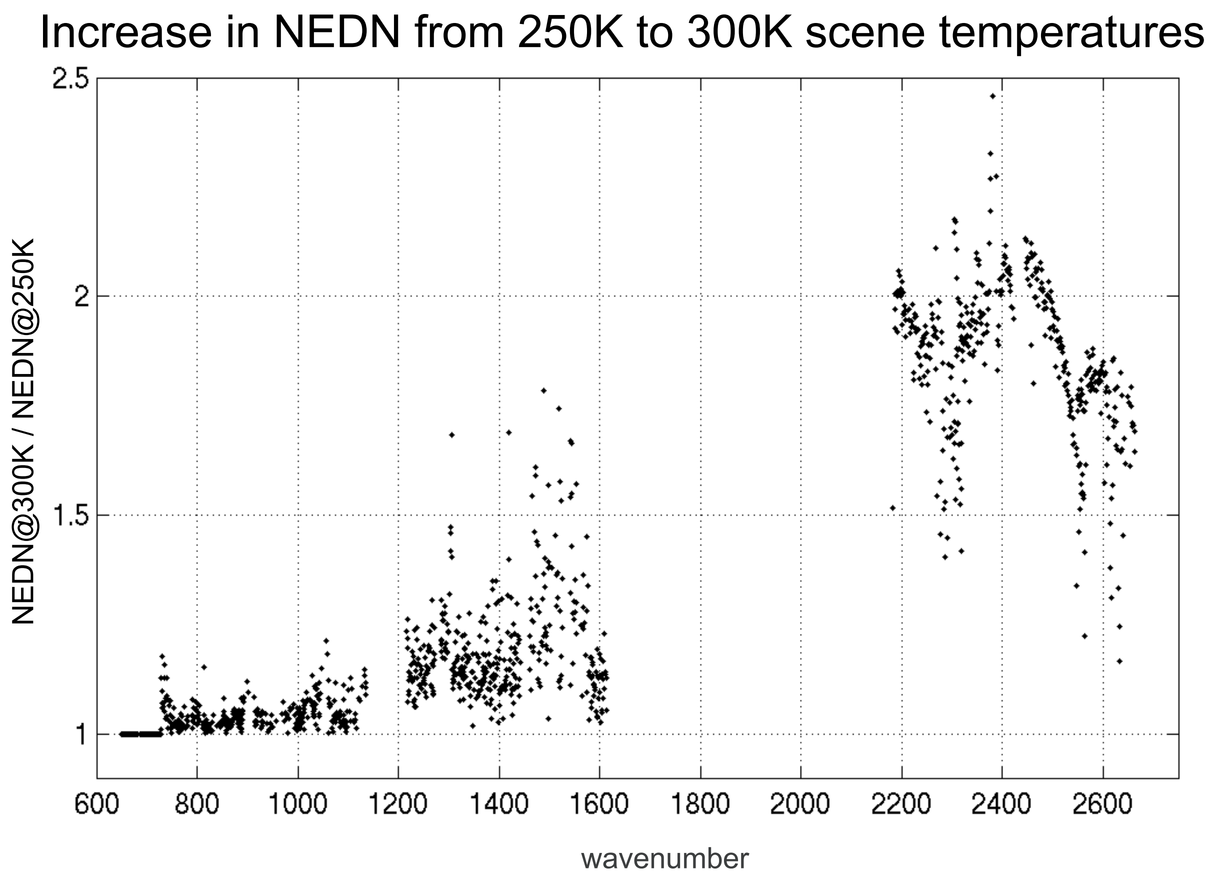




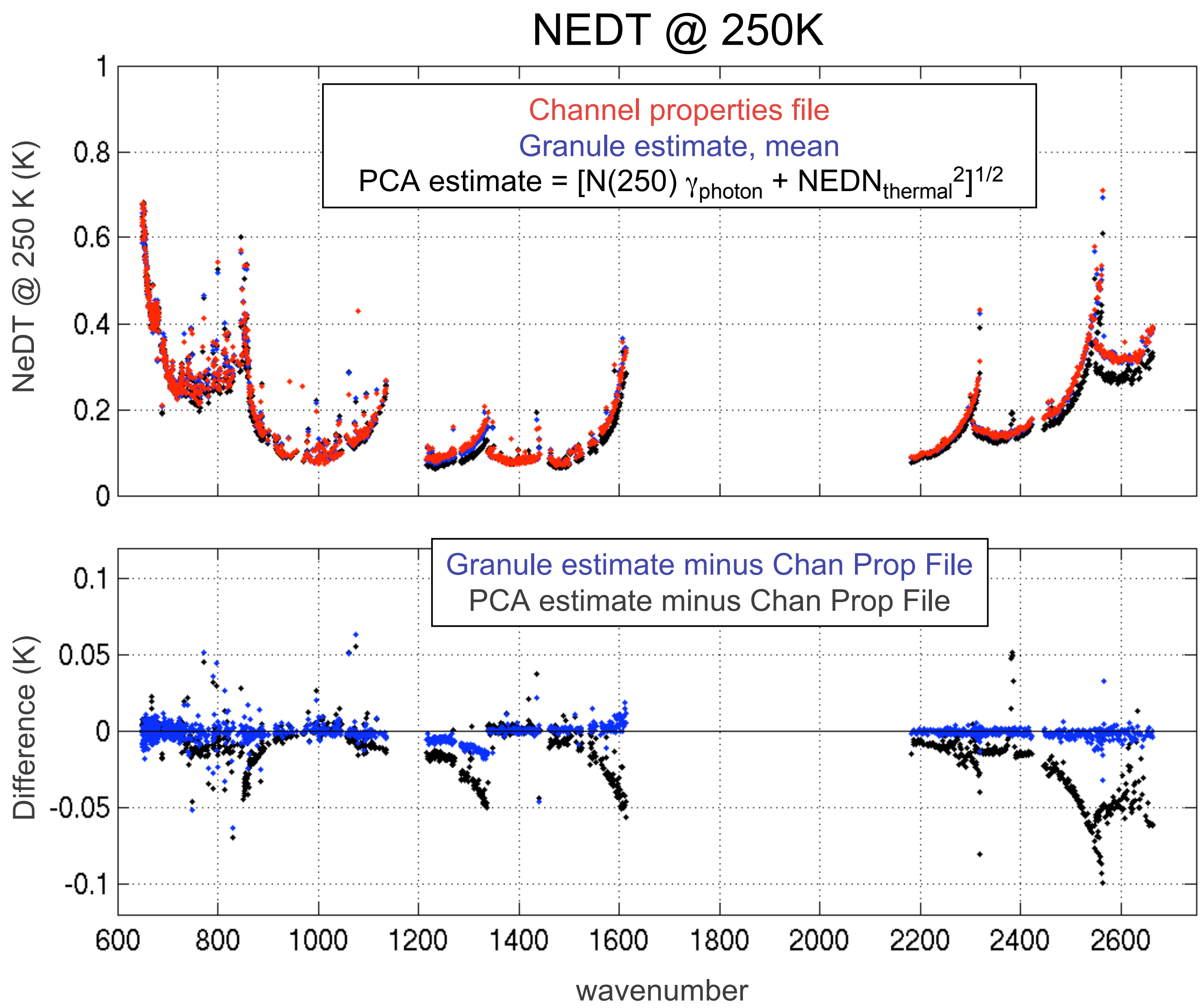
NEDN versus scene radiance parameterization



NEDN versus scene radiance

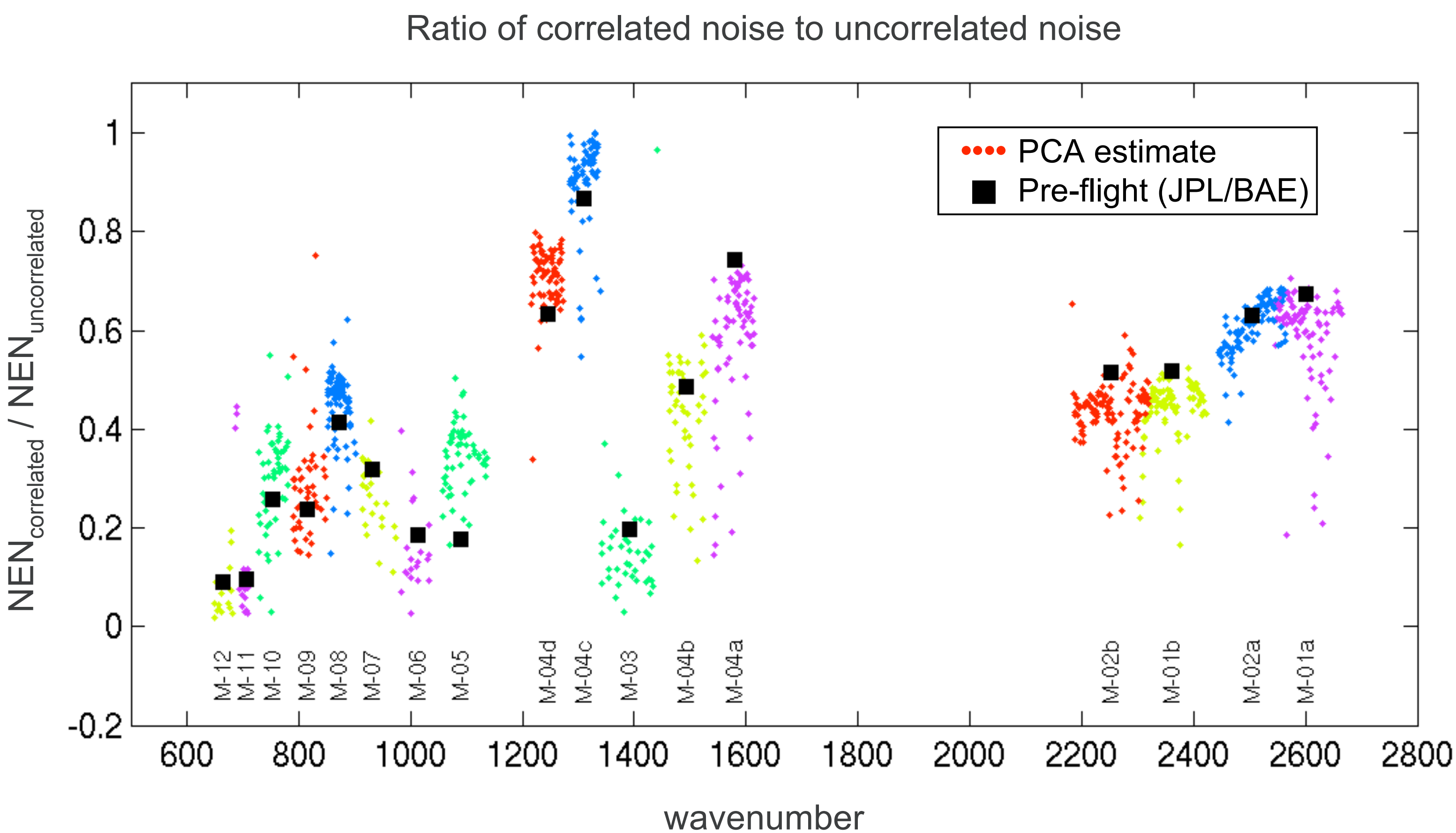


NeDT comparison summary, all bands



Spectrally Correlated Noise

The PCA estimate is of the spectrally uncorrelated noise; the spectrally correlated noise is computed as $[\text{total\_noise}^2 - \text{pca\_noise}^2]^{1/2}$ and compared to pre-flight determinations performed by JPL/BAE:



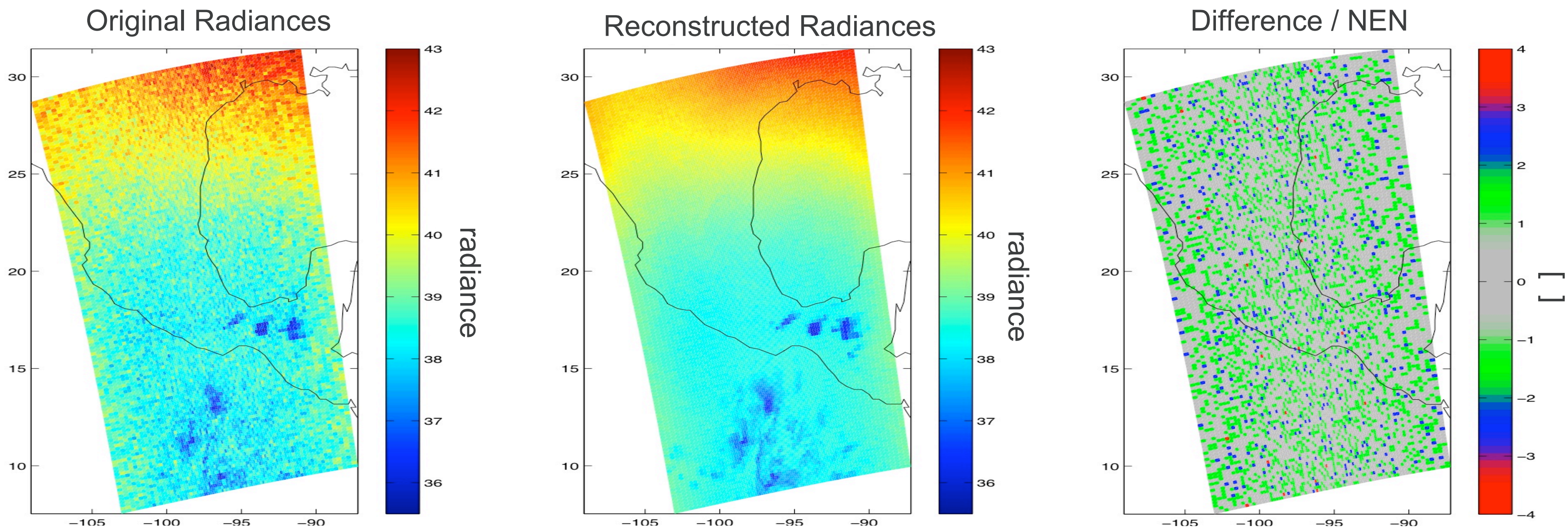
- Very good agreement between two very different and independent analyses.
- The correlated noise is a large fraction of the total noise for several arrays.

Non-Gaussian behavior

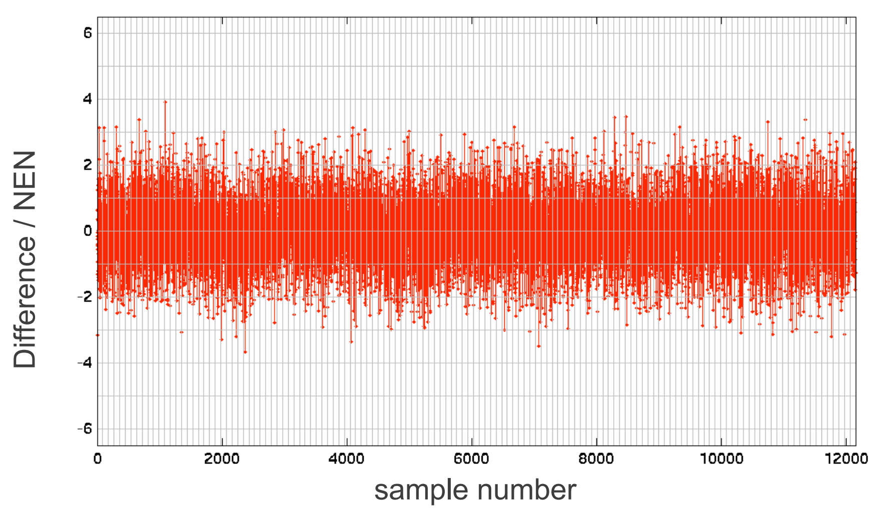
- Further investigation of N_{orig} - N_{recon}

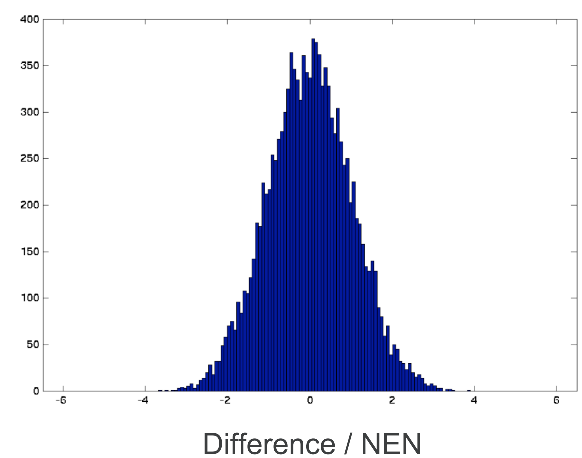
Gaussian channel example

Granule 198, Array M-11, Channel 171 @ 698.545 cm^{-1}



1σ events = 3897
2σ events = 566
3σ events = 23
1σ pops = 17
2σ pops = 0
3σ pops = 0
stripes = 0

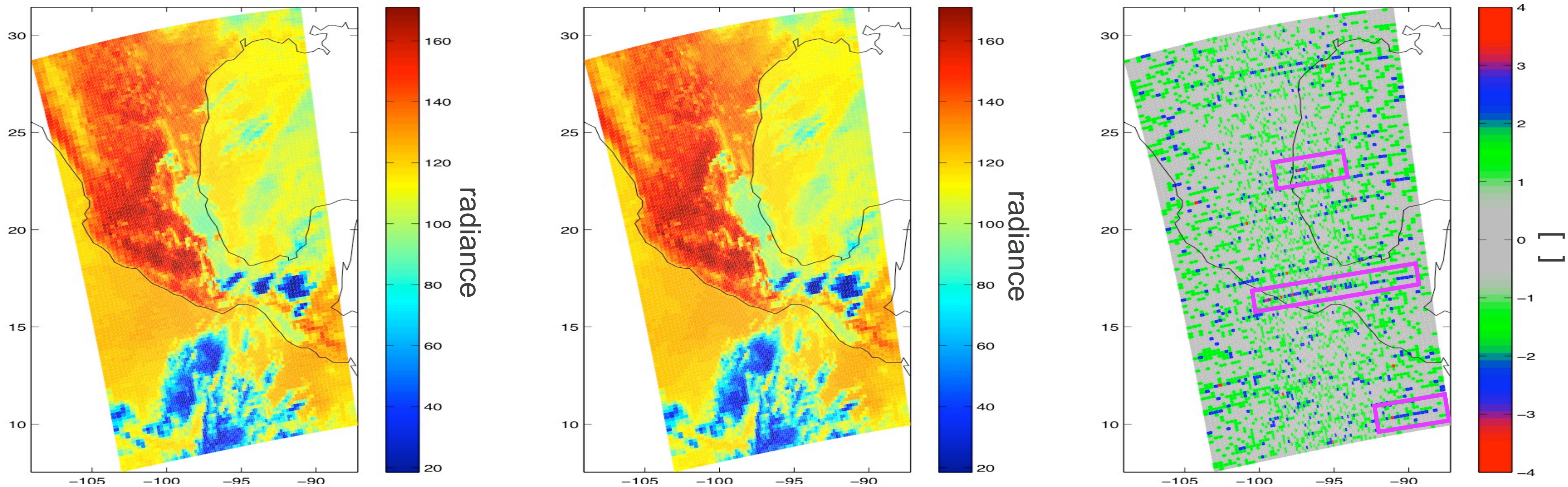




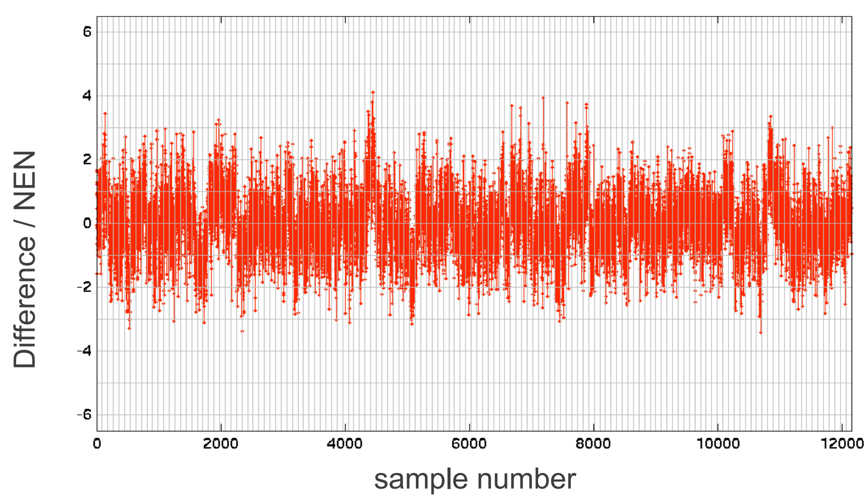
“Popping” channel example

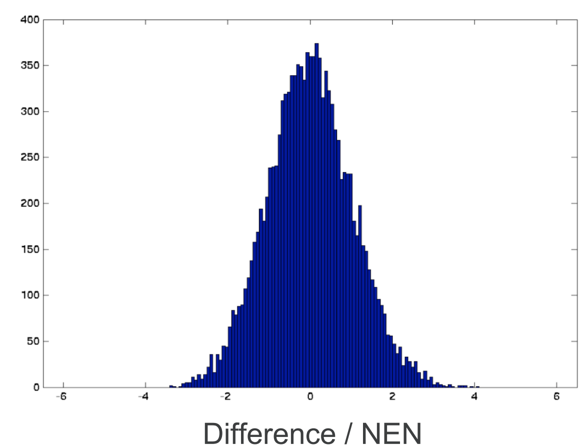
Granule 198, Array M-09, Channel 530 @ 821.597 cm^{-1}

Original Radiances Reconstructed Radiances Difference / NEN



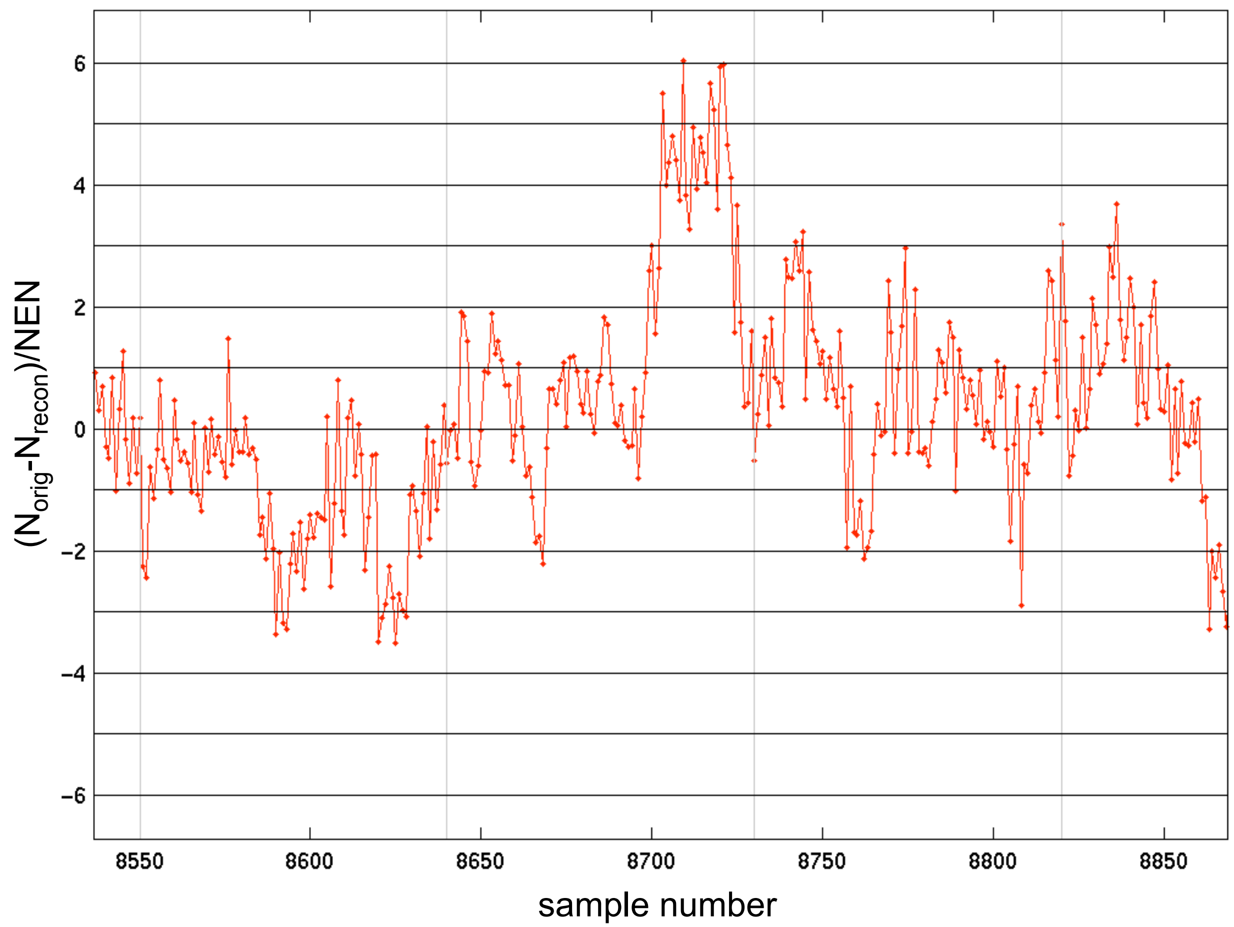
1σ events = 3825
2σ events = 563
3σ events = 26
1σ pops = 187
2σ pops = 15
3σ pops = 0
stripes = 0





“Popping” channel example

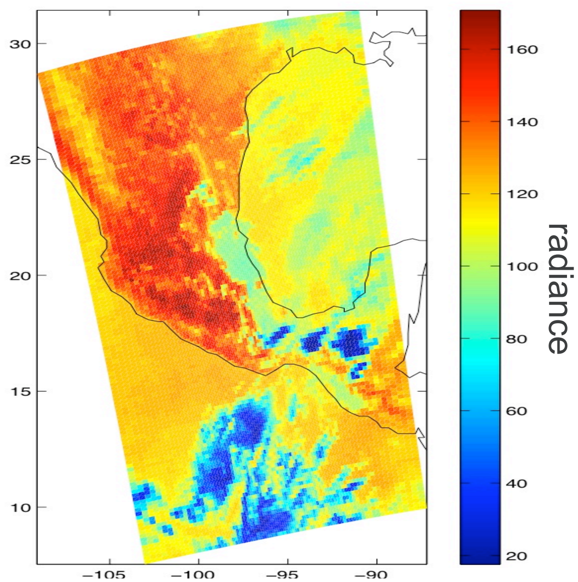
Granule 198, Array M-09, Channel 573 @ 838.106 cm^{-1}



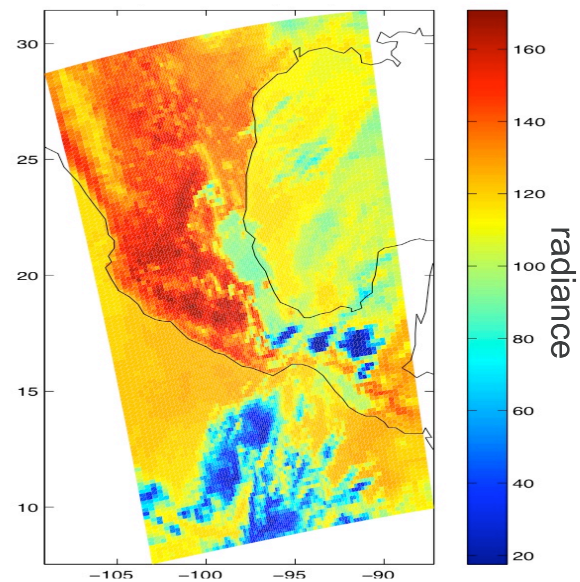
“Striping” channel example

Granule 198, Array M-09, Channel 564 @ 834.772 cm^{-1}

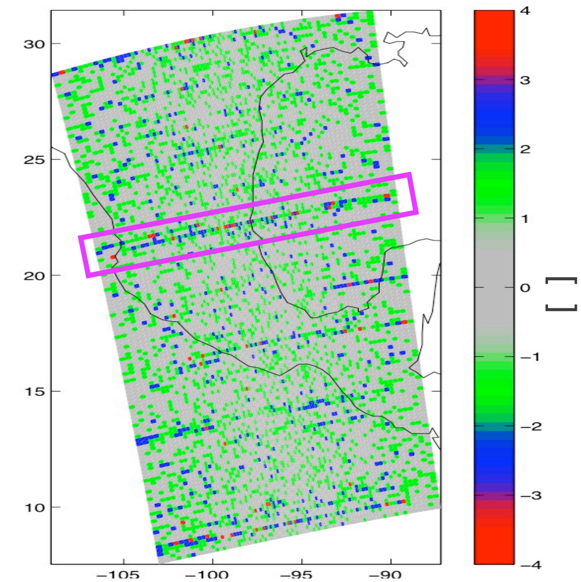
Original Radiances



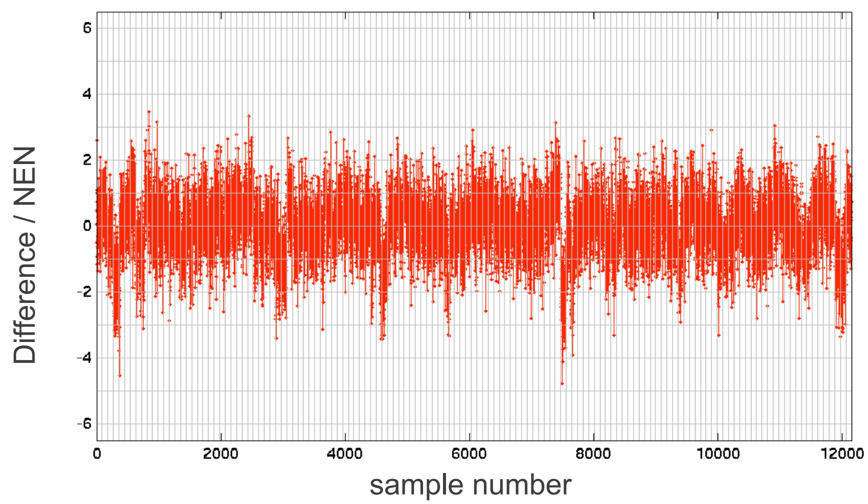
Reconstructed Radiances

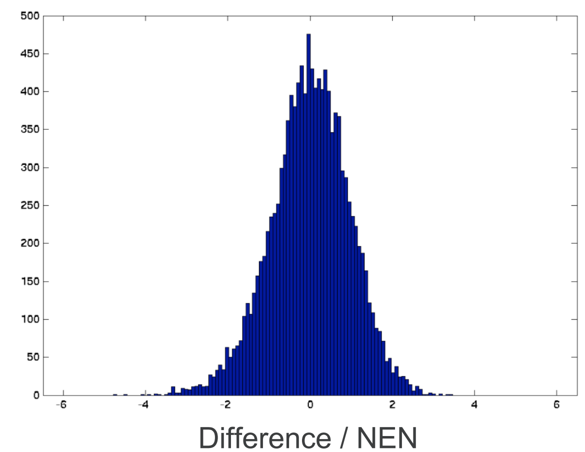


Difference / NEN

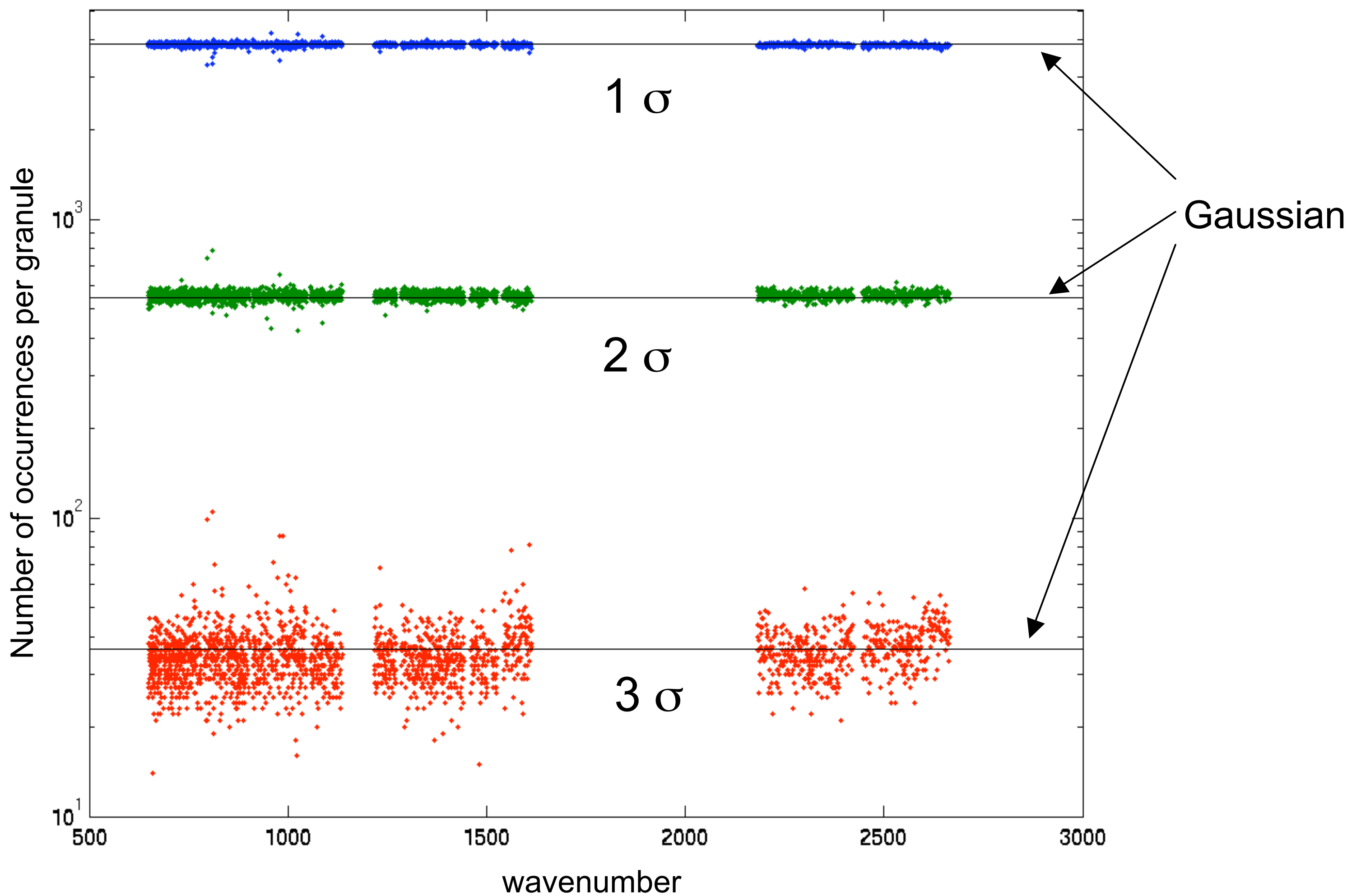


1σ events = 3713
2σ events = 593
3σ events = 61
1σ pops = 118
2σ pops = 12
3σ pops = 0
stripes = 1





Number of Nσ events detected



Number of events for pure Gaussian behavior:

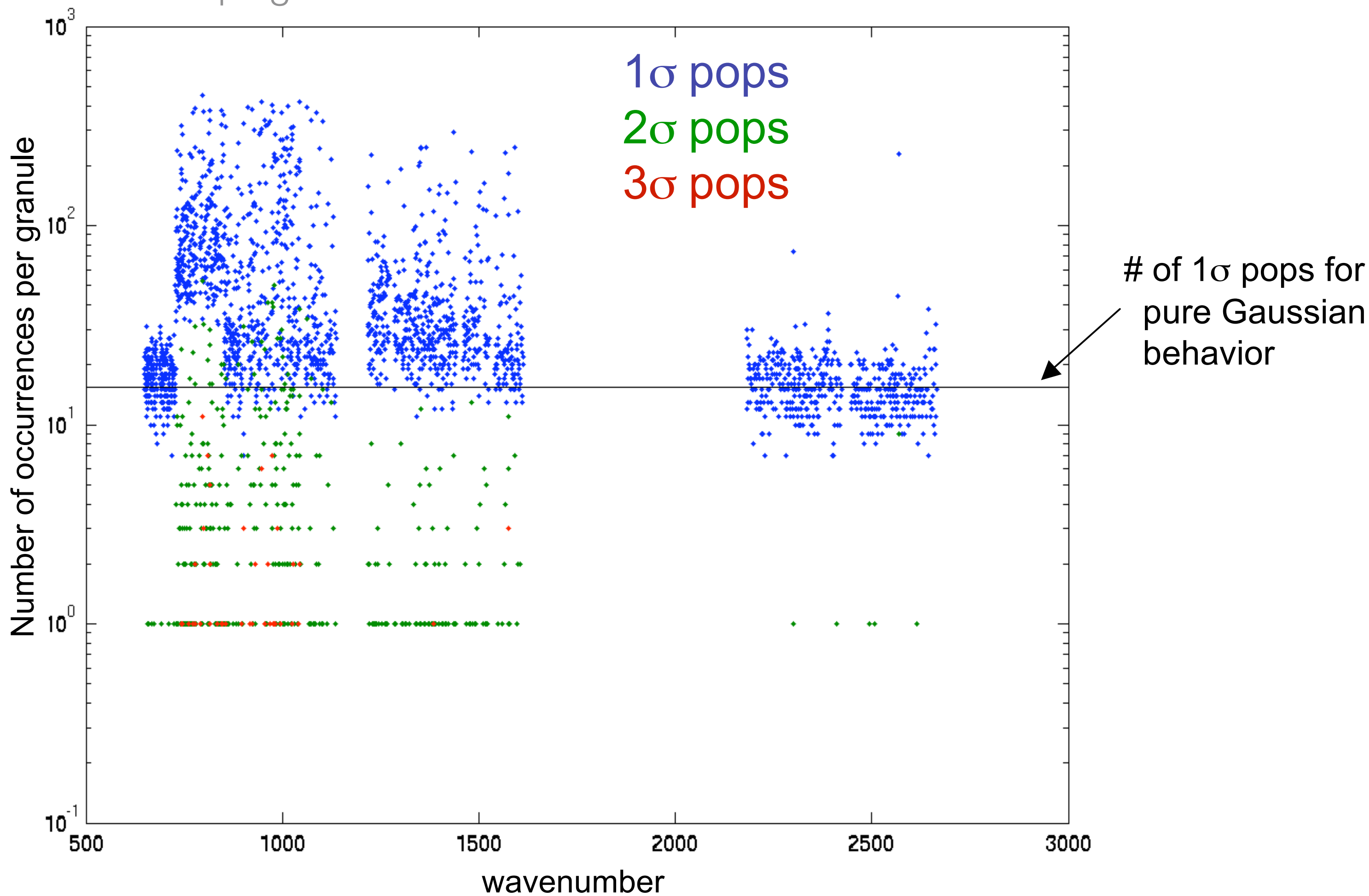
\# 1σ events per granule = 90*135*(1-0.683) = 3852

\# 2σ events per granule = 90*135*(1-0.955) = 547

\# 3σ events per granule = 90*135*(1-0.997) = 36

Number of Nσ Pops detected

- Nσ pop defined here as 4 or more consecutive Nσ events of same sign
- 875 / 482 / 42 channels exhibit 1σ / 2σ / 3σ popping significantly above Gaussian behavior
- 14 channels found to exhibit “striping”



Number of pops for pure Gaussian behavior:

\# 1σ pops per granule = $2*90*135*(0.5*(1-0.683))^4 = 15$

\# 2σ pops per granule = $2*90*135*(0.5*(1-0.955))^4 = 0$

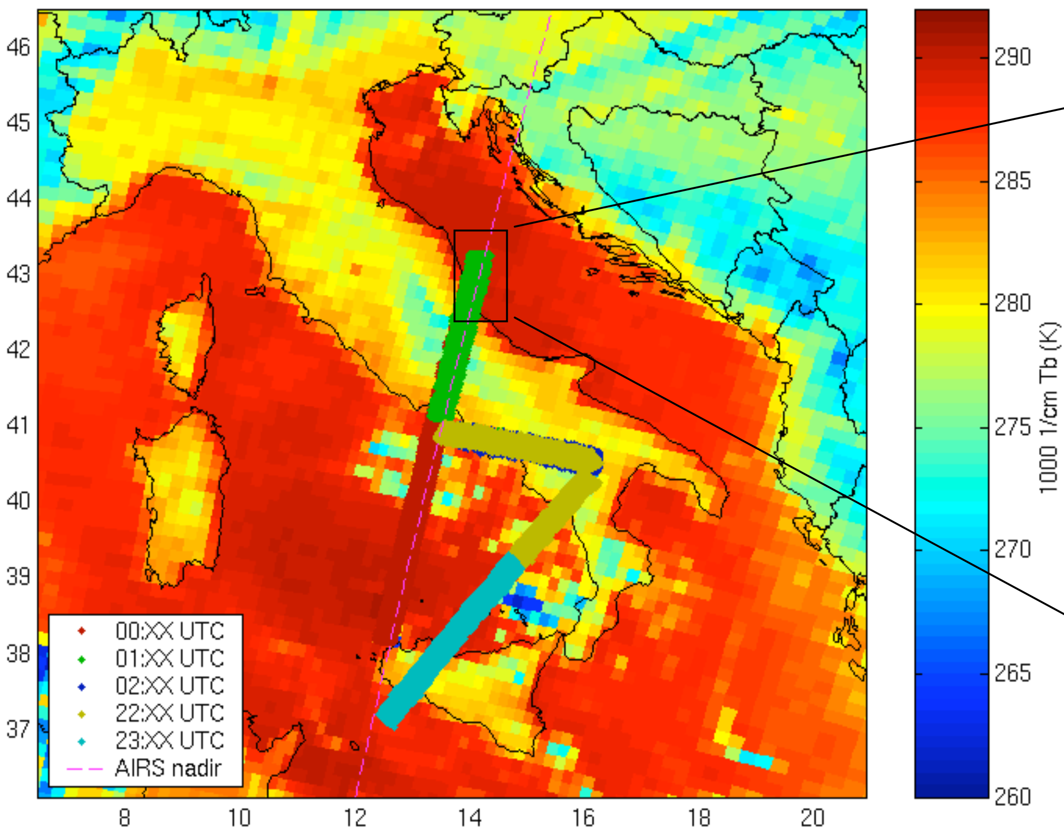
\# 3σ pops per granule = $2*90*135*(0.5*(1-0.997))^4 = 0$

Inspection of PCs and individual spectra

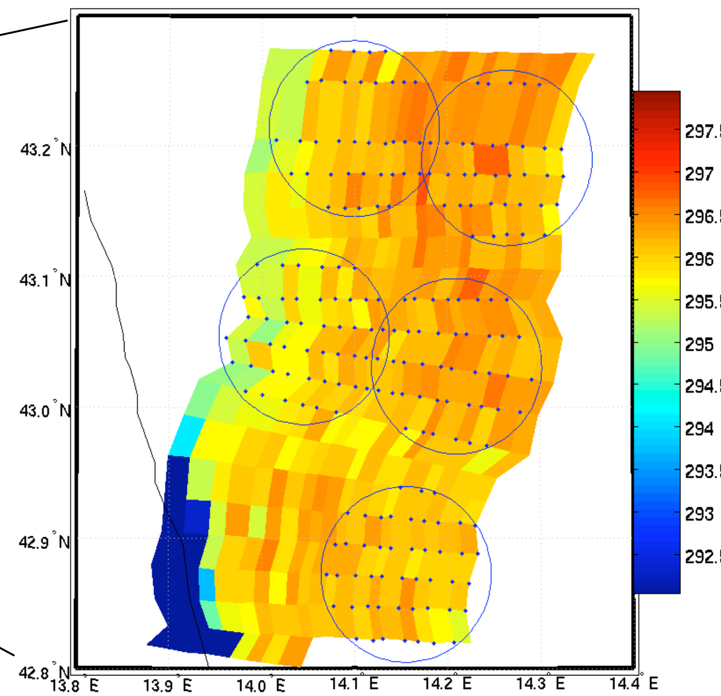
- Effects that lie in the less understood domain between calibration (long average) and spectrally random, repeatable noise.

AIRS underflight with the Scanning-HIS

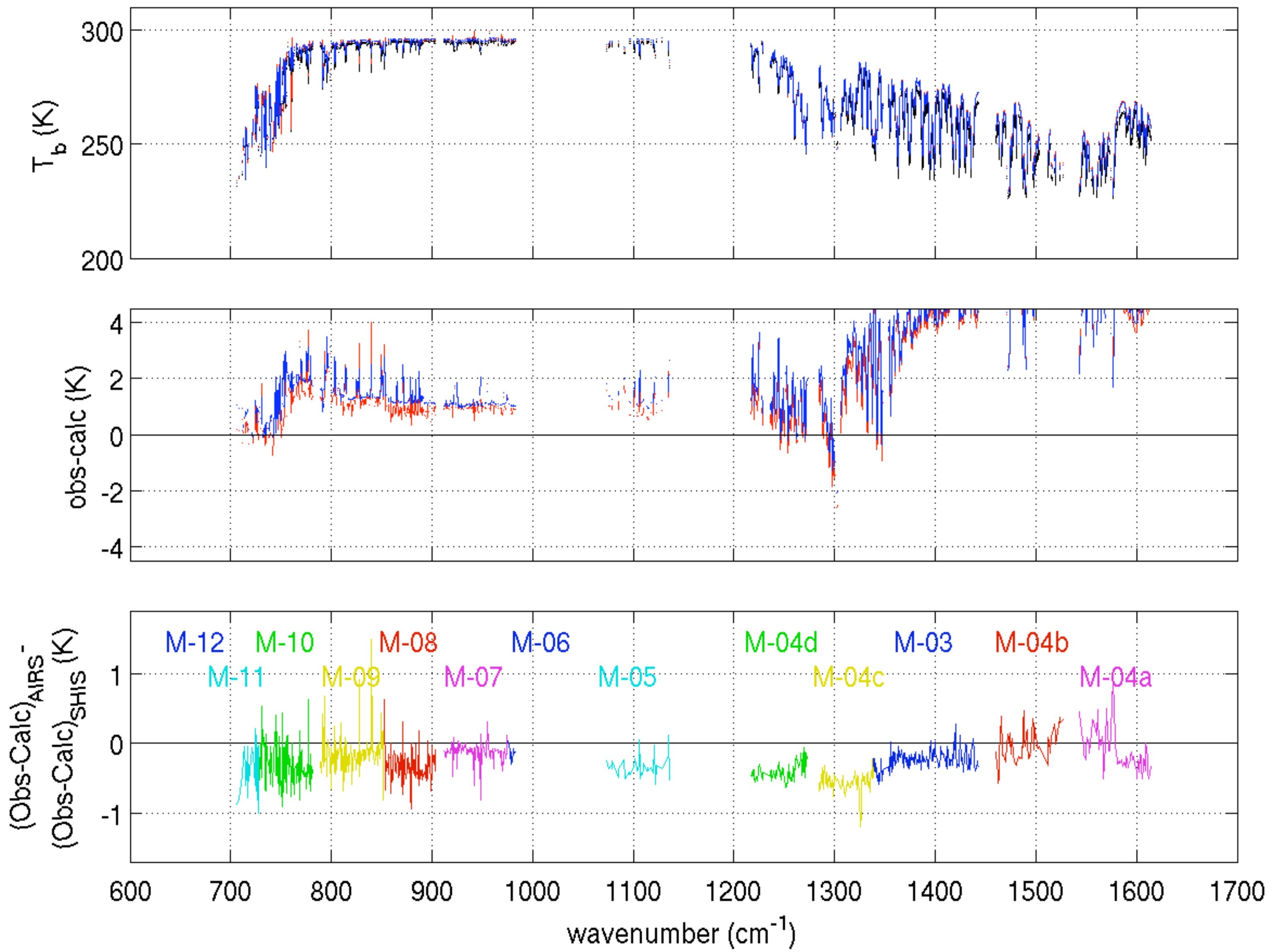
ADRIEX (EAQUATE) Campaign
S-HIS on Proteus aircraft @ 16km over Adriatic Sea
2004.09.08, 01:10 UTC



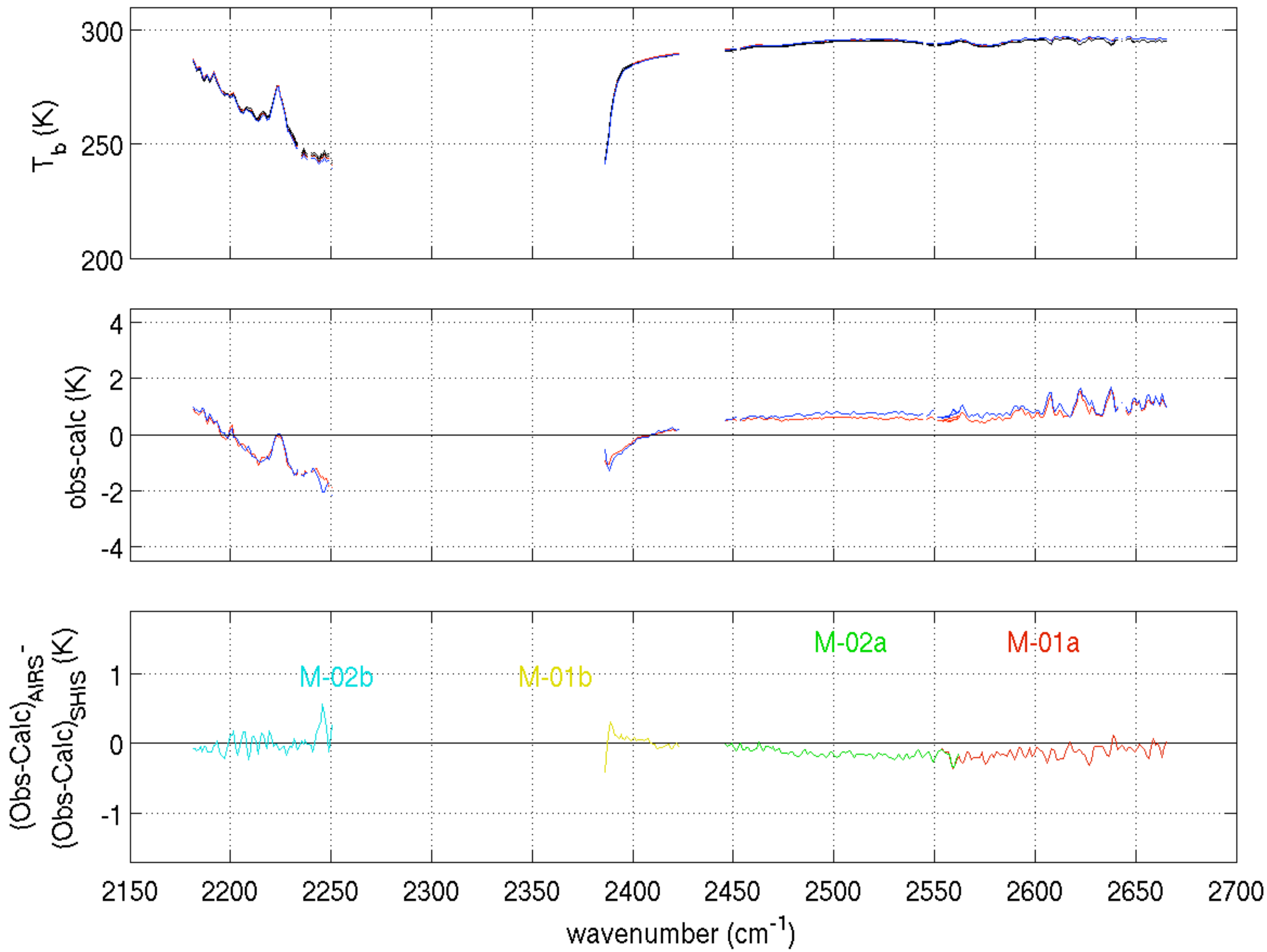
S-HIS 12 μm brightness temperatures
and AIRS FOV locations



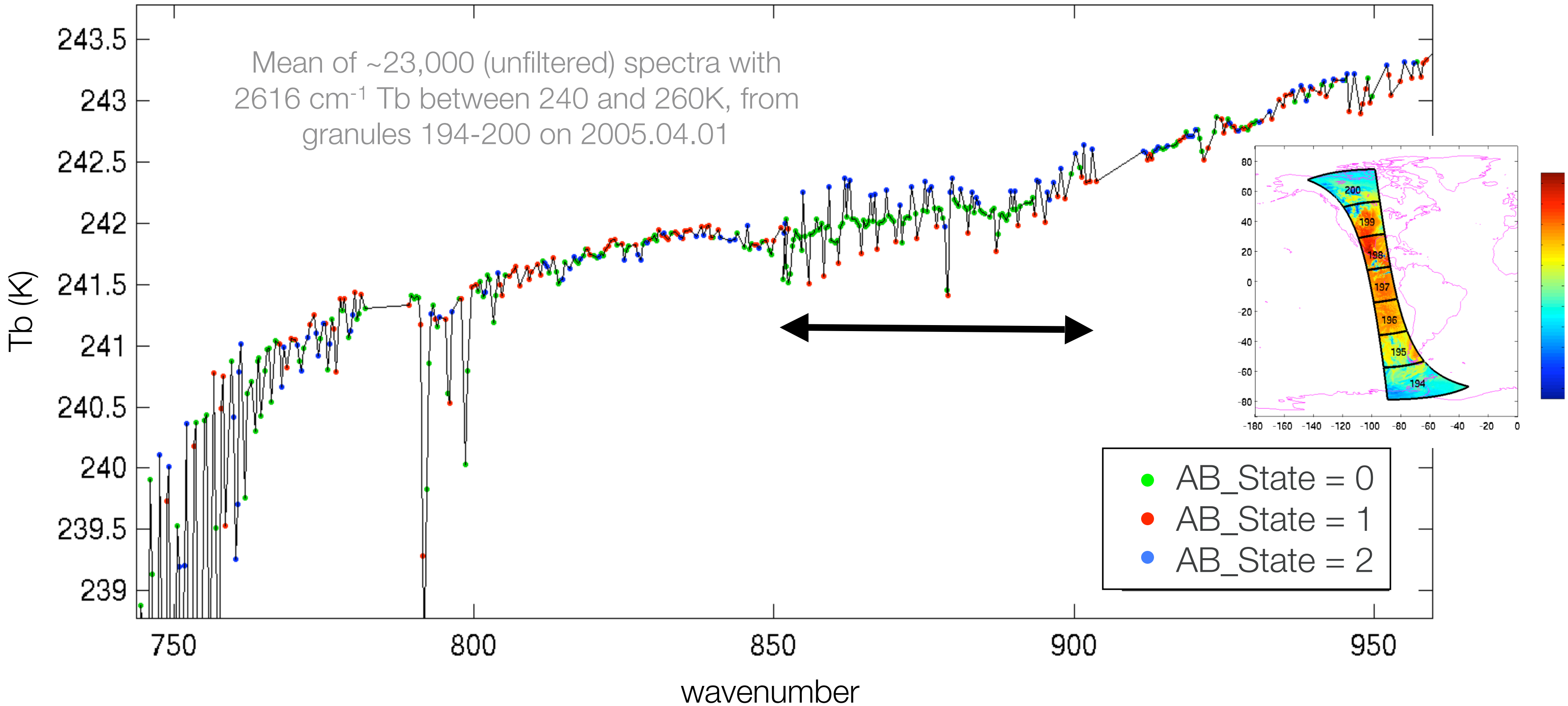
AIRS radiance validation, longwave and midwave



AIRS radiance validation, shortwave

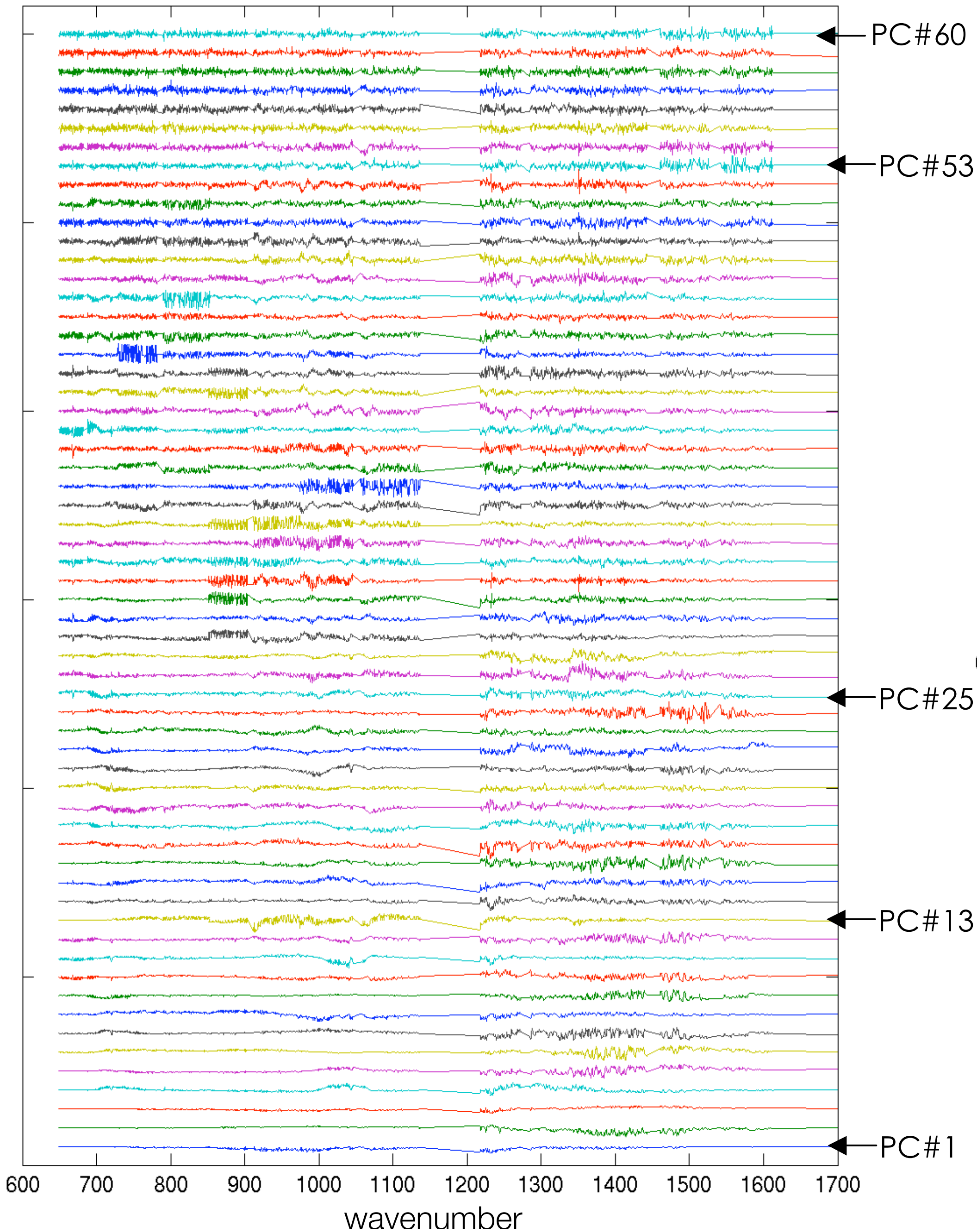


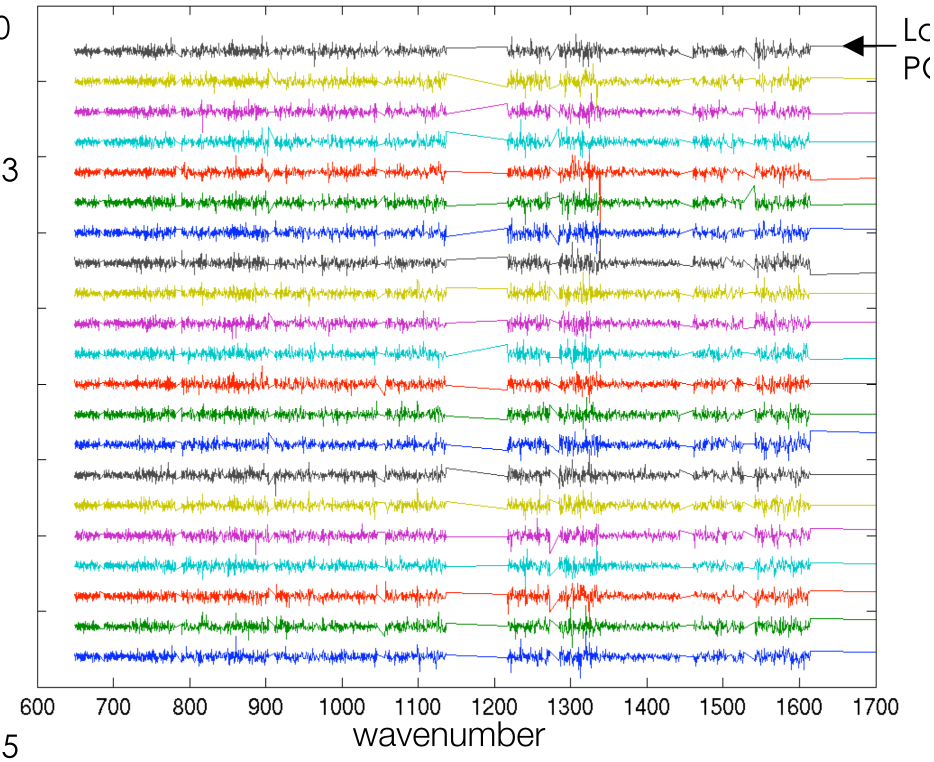
A / B State dependent calibration in M-08



- Similar behavior observed for similar scenes throughout the mission
- Less evident in mean spectra at colder (e.g. Antarctica) and warmer (e.g. clear ocean) scenes.

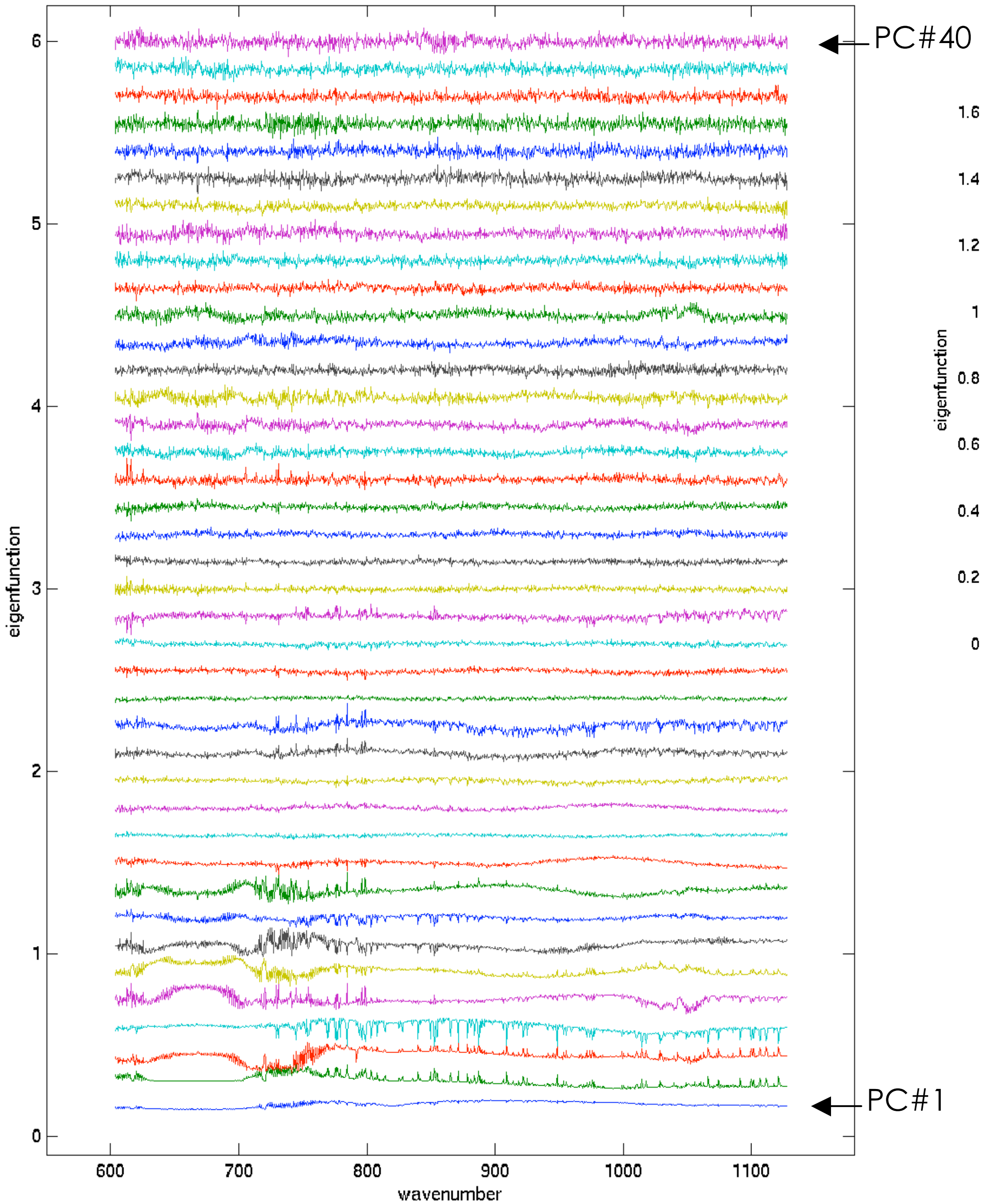
AIRS Principle Components, granule 195

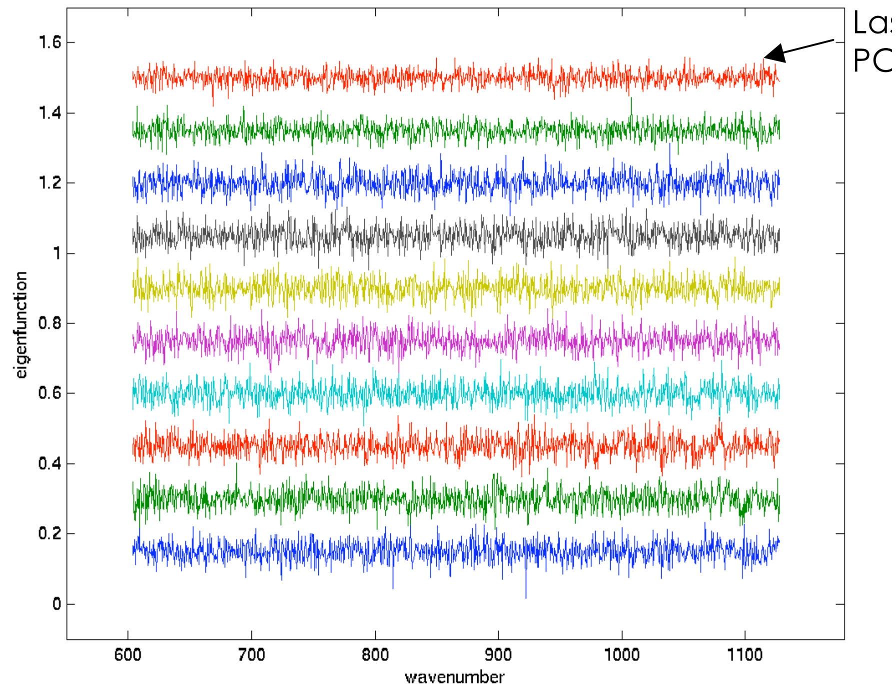




- Earliest PCs look clean - of realistic atmospheric, cloud, and surface variations
- Array module-to-module biases & noise variations are apparent at lower levels (higher PCs)
- Last (least significant) PCs are of random noise

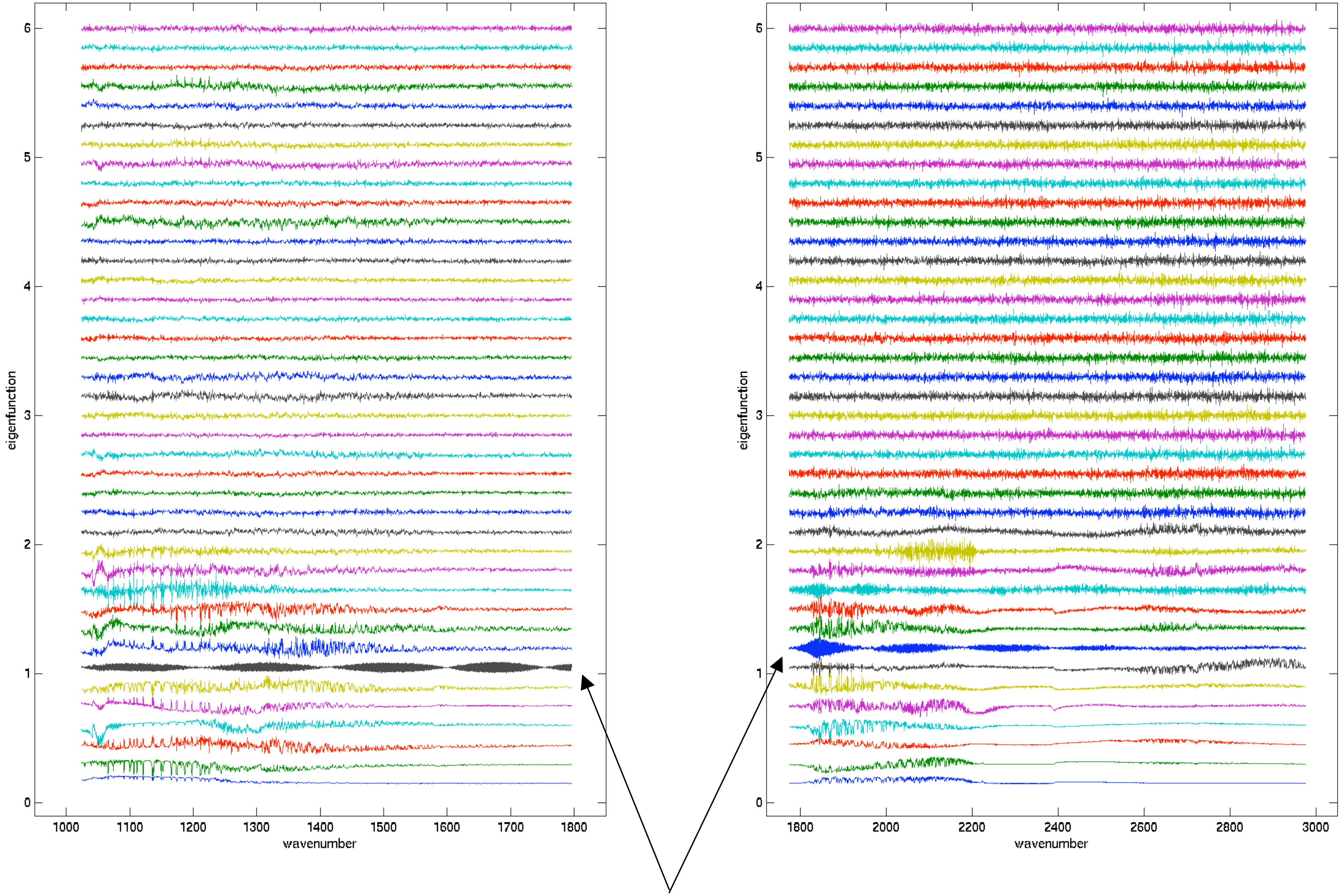
Scanning-HIS Principle Components, Longwave





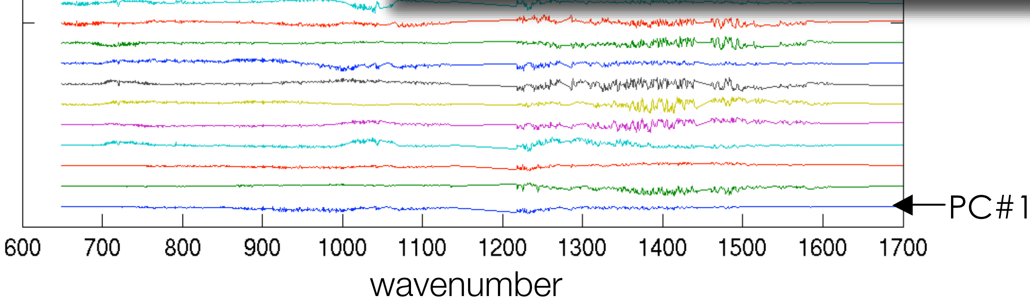
Spectral signatures in low order PCs are clean, indicative of real atmospheric characteristics

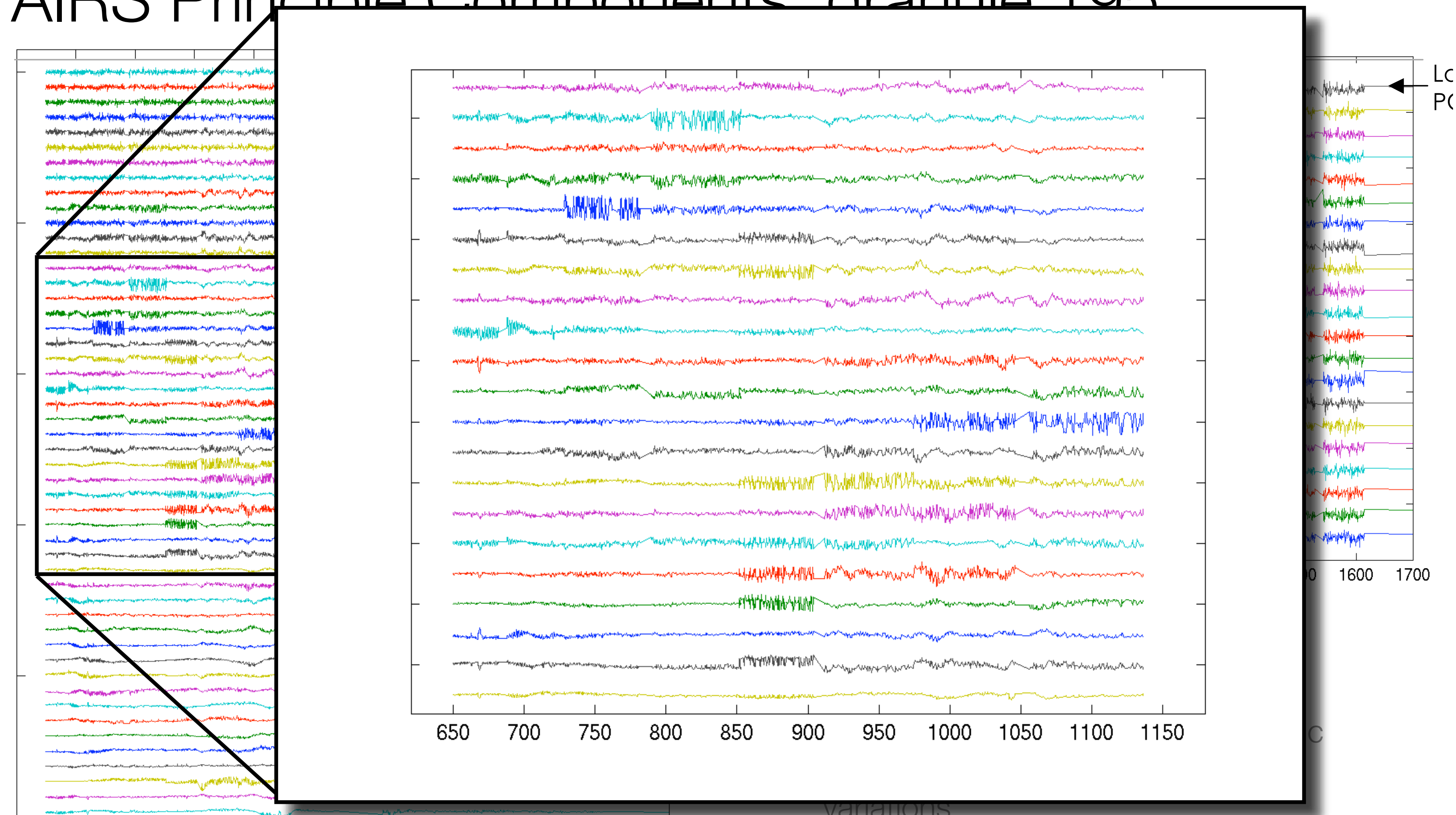
Scanning-HIS PCs, Midwave and Shortwave



Ringing is indicative of spectral resampling processing artifact- easily fixed by band guard position mod

AIRS Principle Components, granule 195

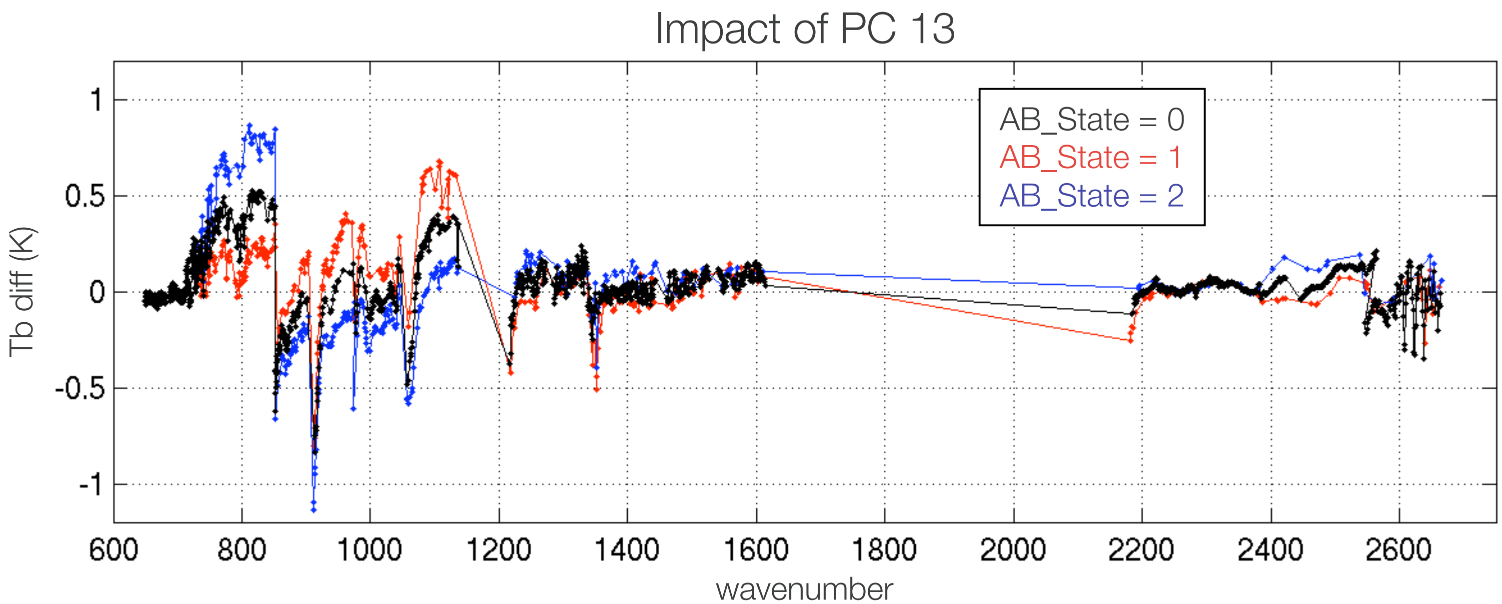


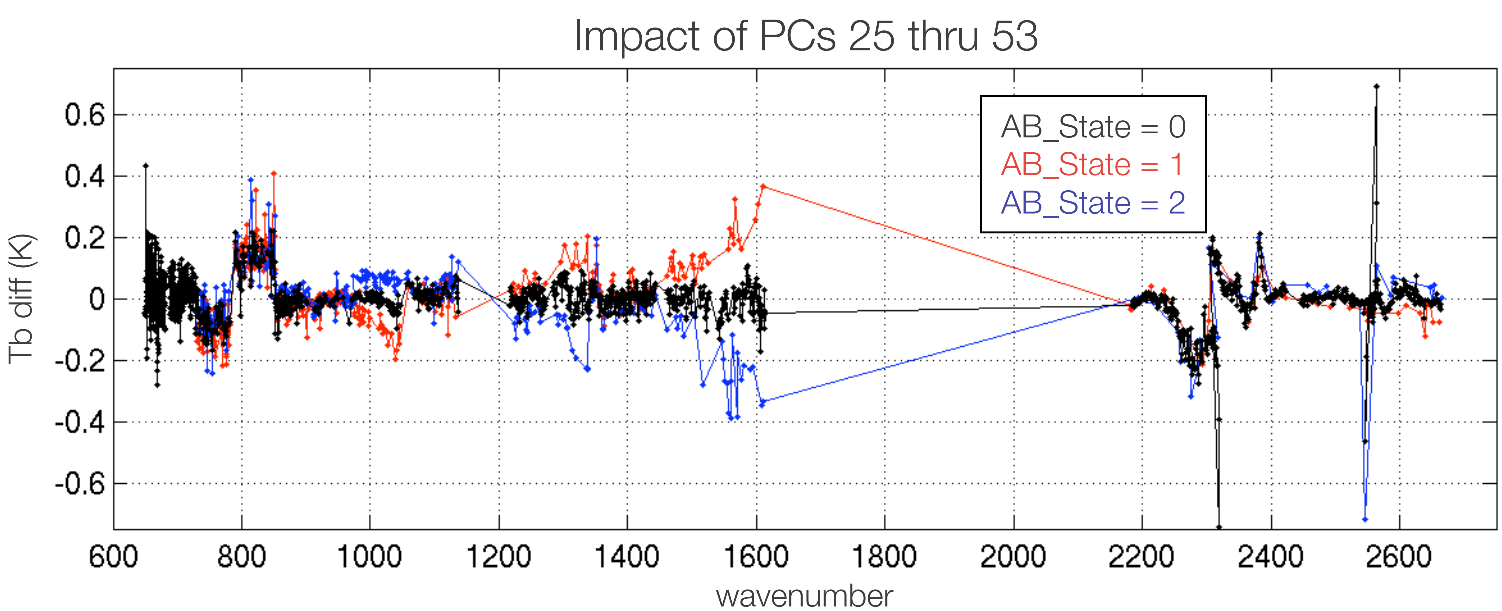


variations

- Array module-to-module biases & noise variations are apparent at lower levels (higher PCs)
- Last (least significant) PCs are of random noise

Impact of PC 13 and 25-53 on an example spectrum

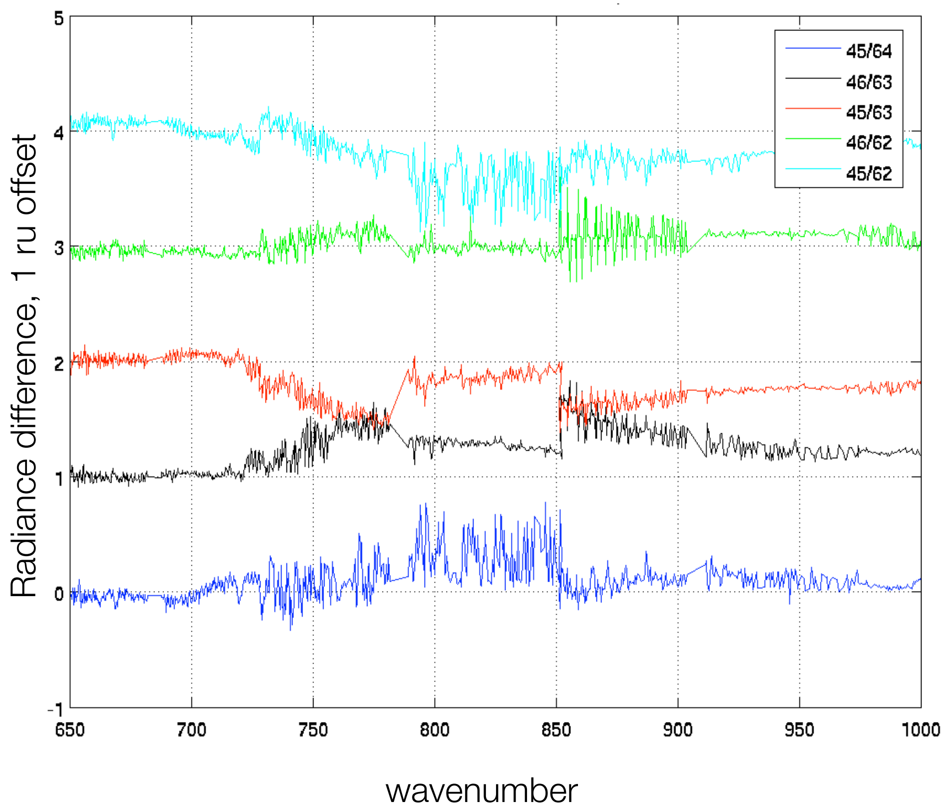




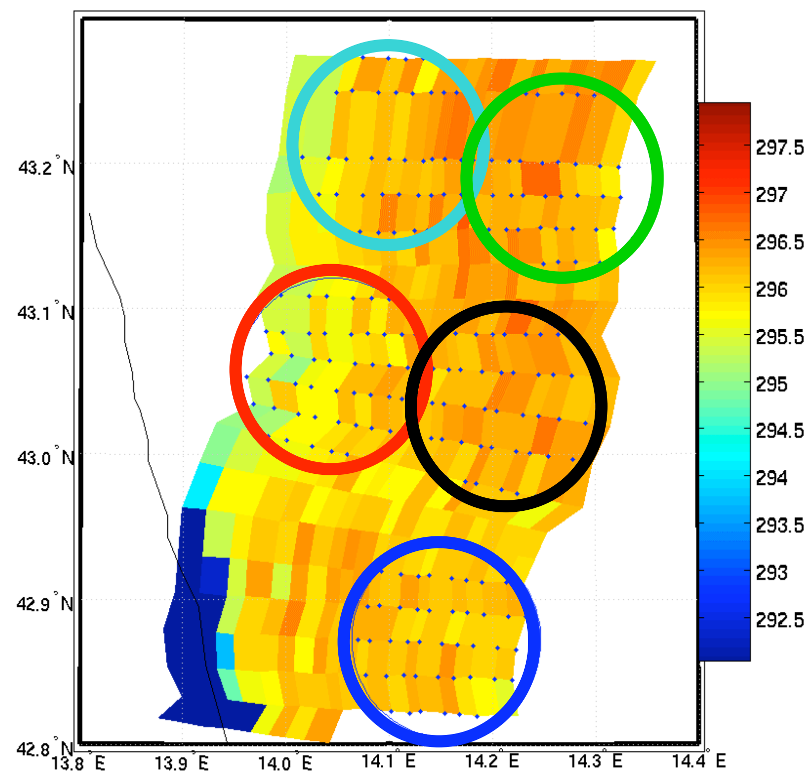
Variations in individual spectra. 2004.09.07 EAQUATE cas

AIRS

Differences from mean spectrum



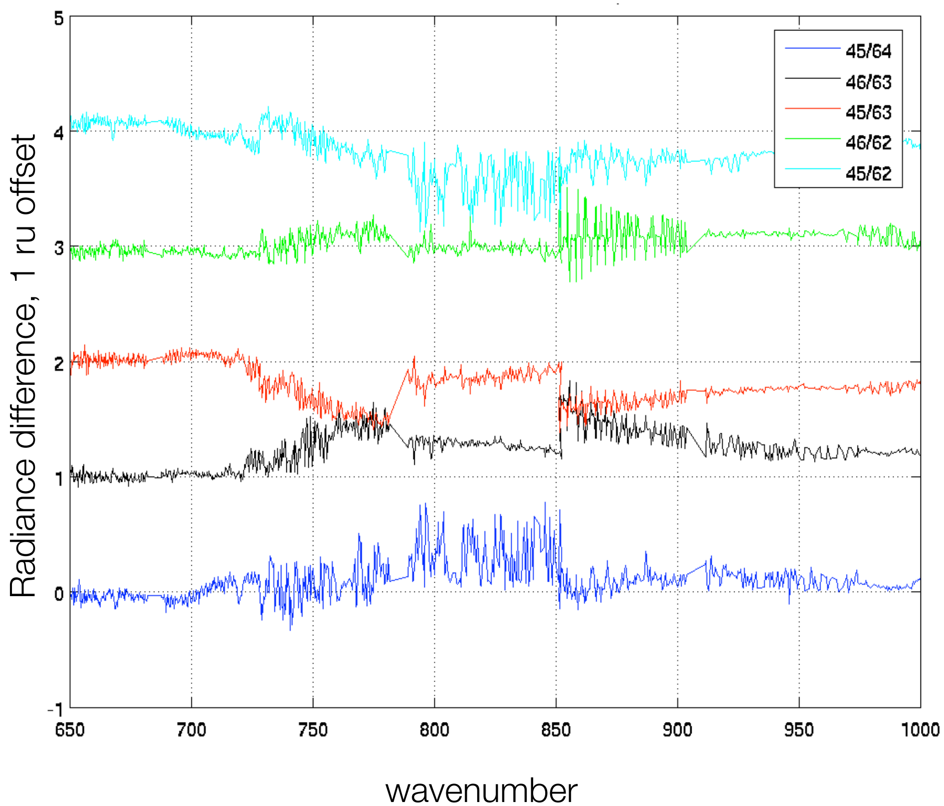
S-HIS brightness temperatures and AIRS FOV locations



Variations in individual spectra. 2004.09.07 EAQUATE cas

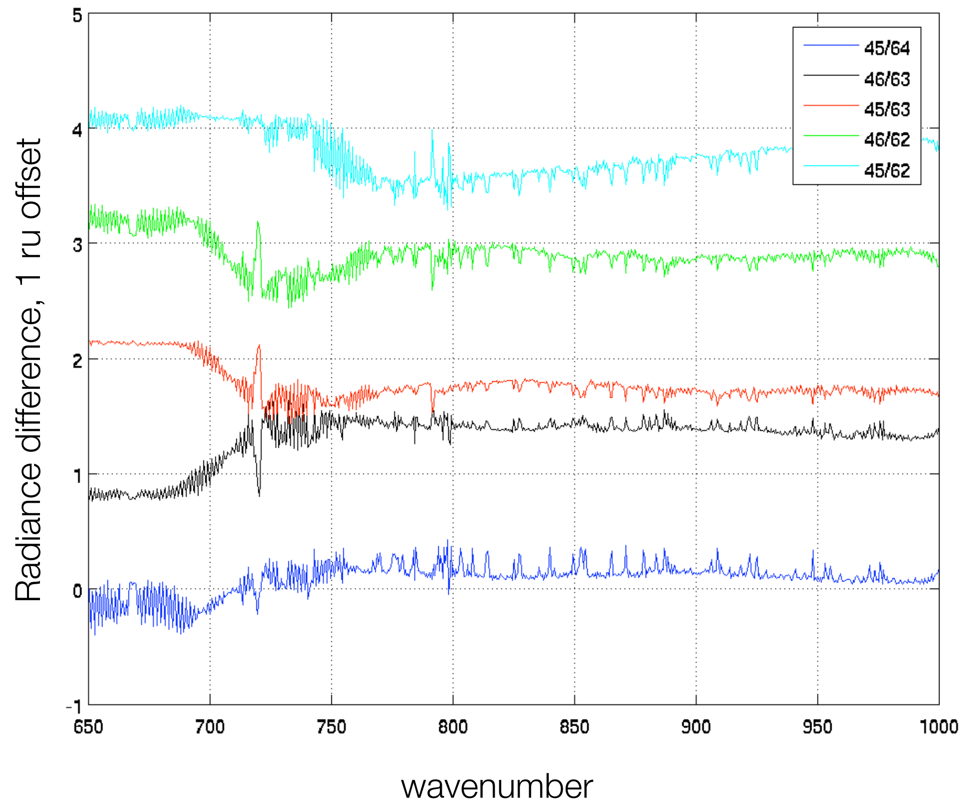
AIRS

Differences from mean spectrum



S-HIS

Differences from mean spectrum



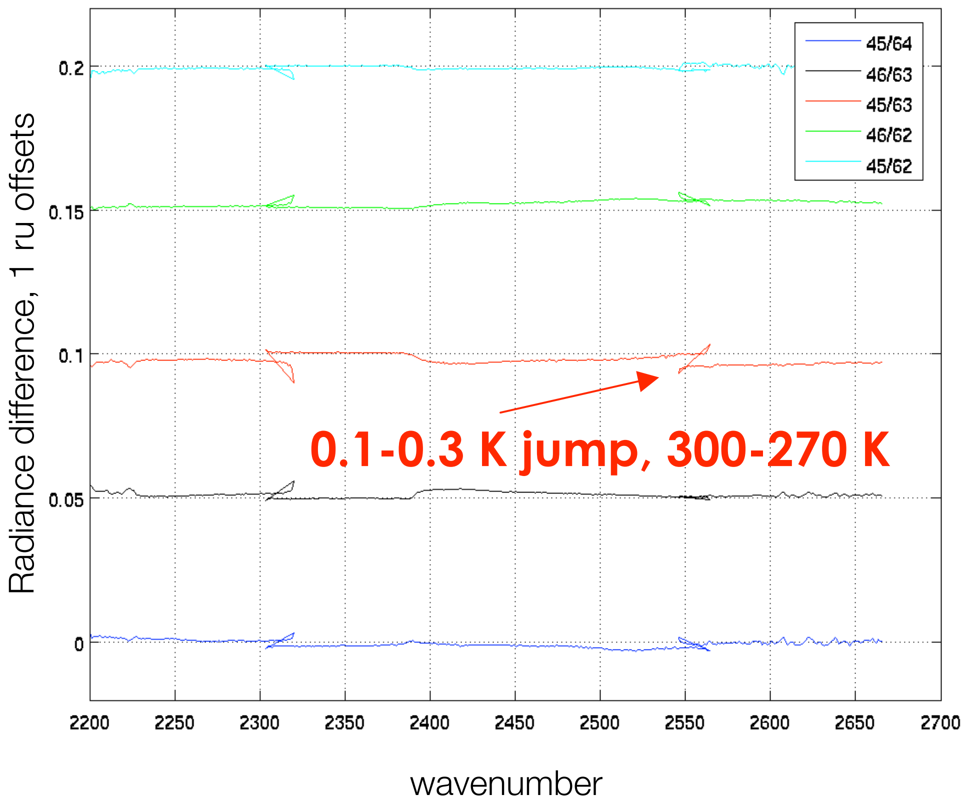
A/B state and array correlated artifacts

Smooth and physically reasonable

Variations in individual spectra. 2004.09.07 EAQUATE cas

AIRS

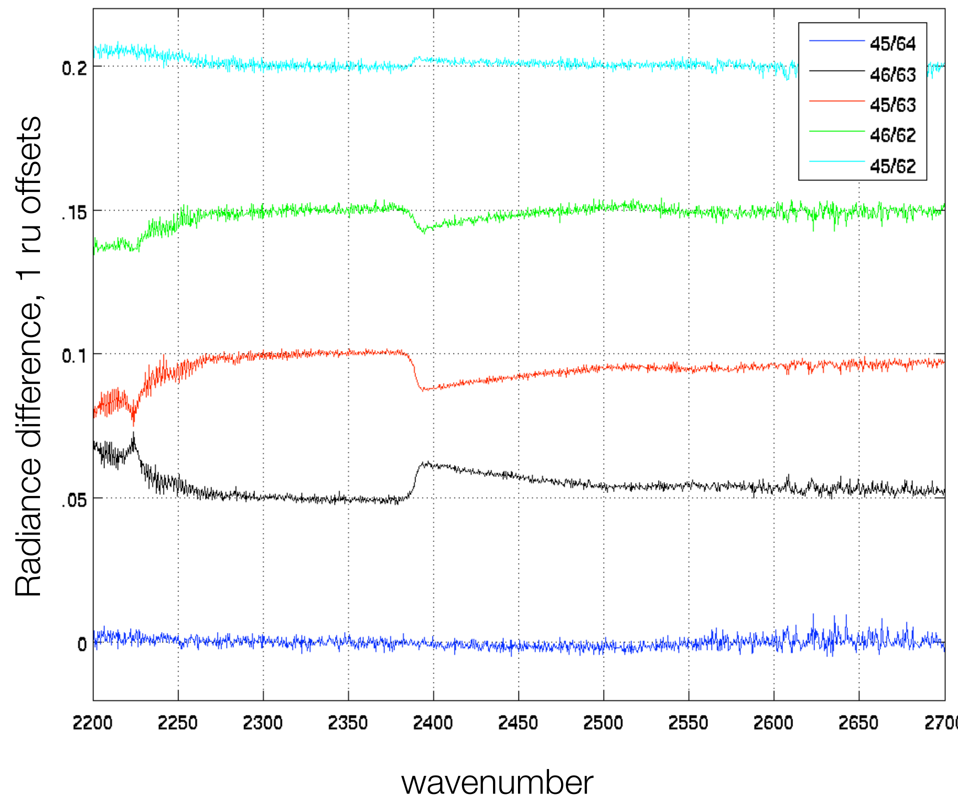
Differences from mean spectrum



Low random noise, but significant array-to-array jumps

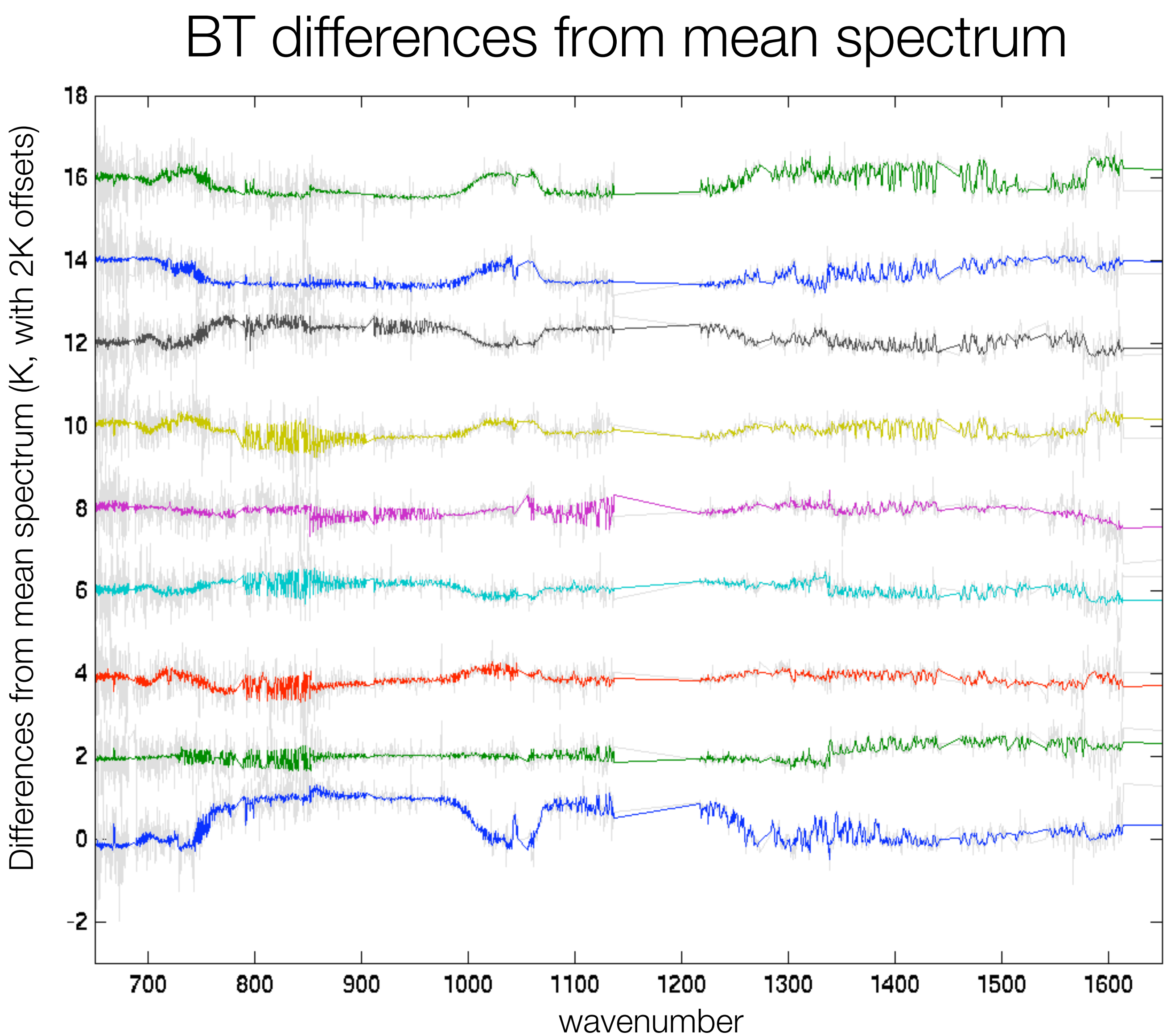
S-HIS

Differences from mean spectrum

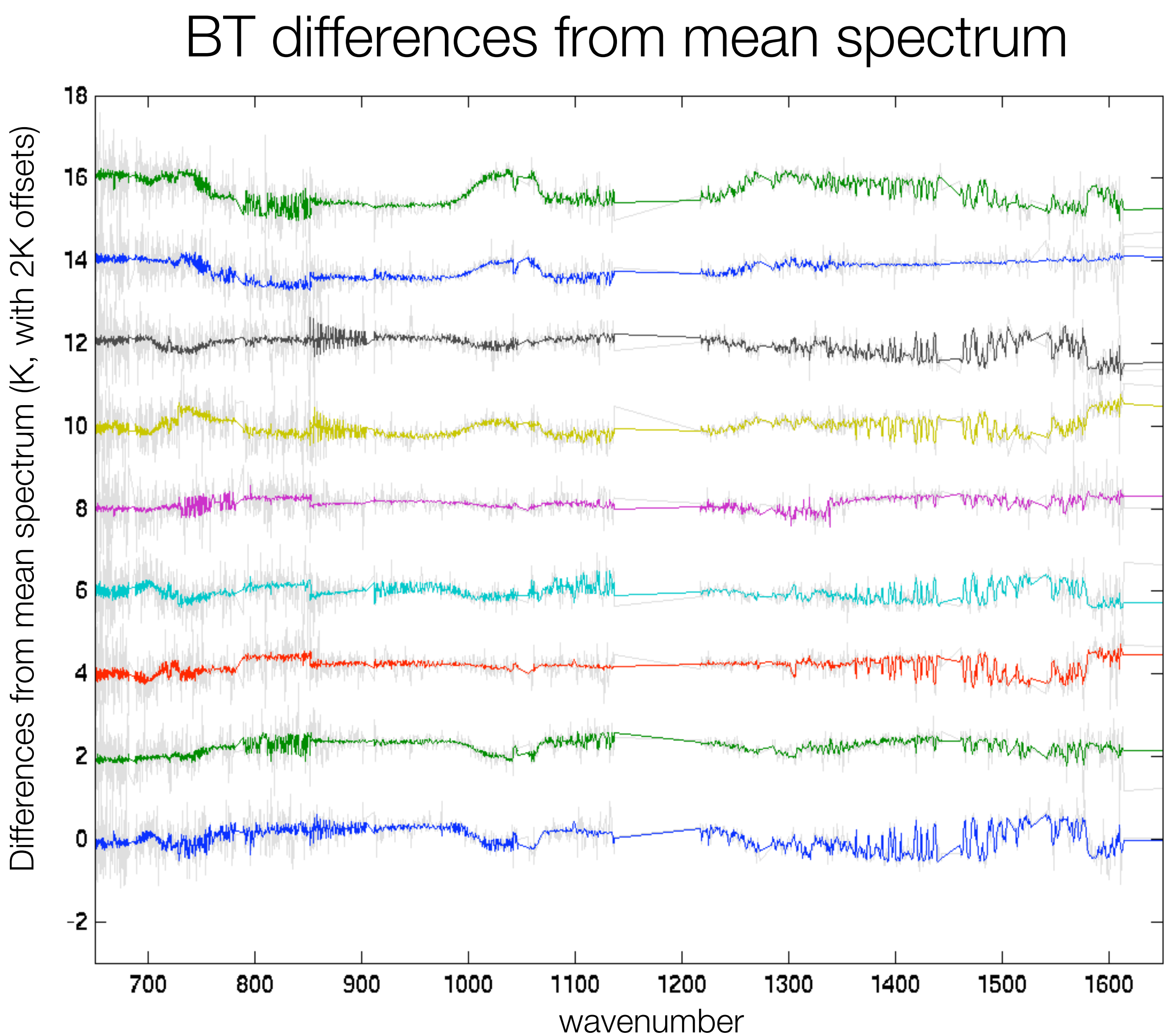


Higher random noise, but 4x spectral resolution and spectrally smooth

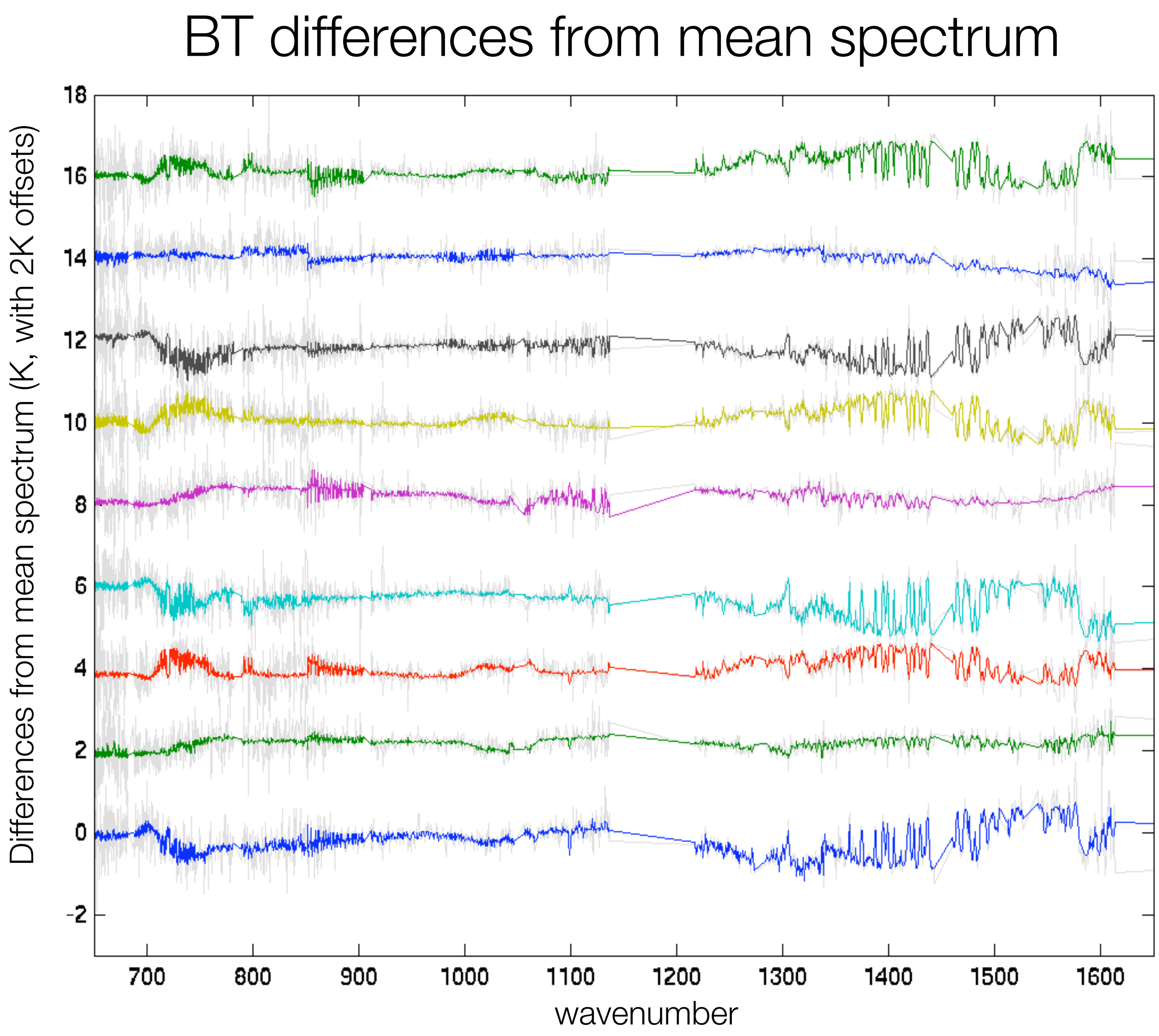
More individual AIRS spectra, 2005.04.01 granule 20(



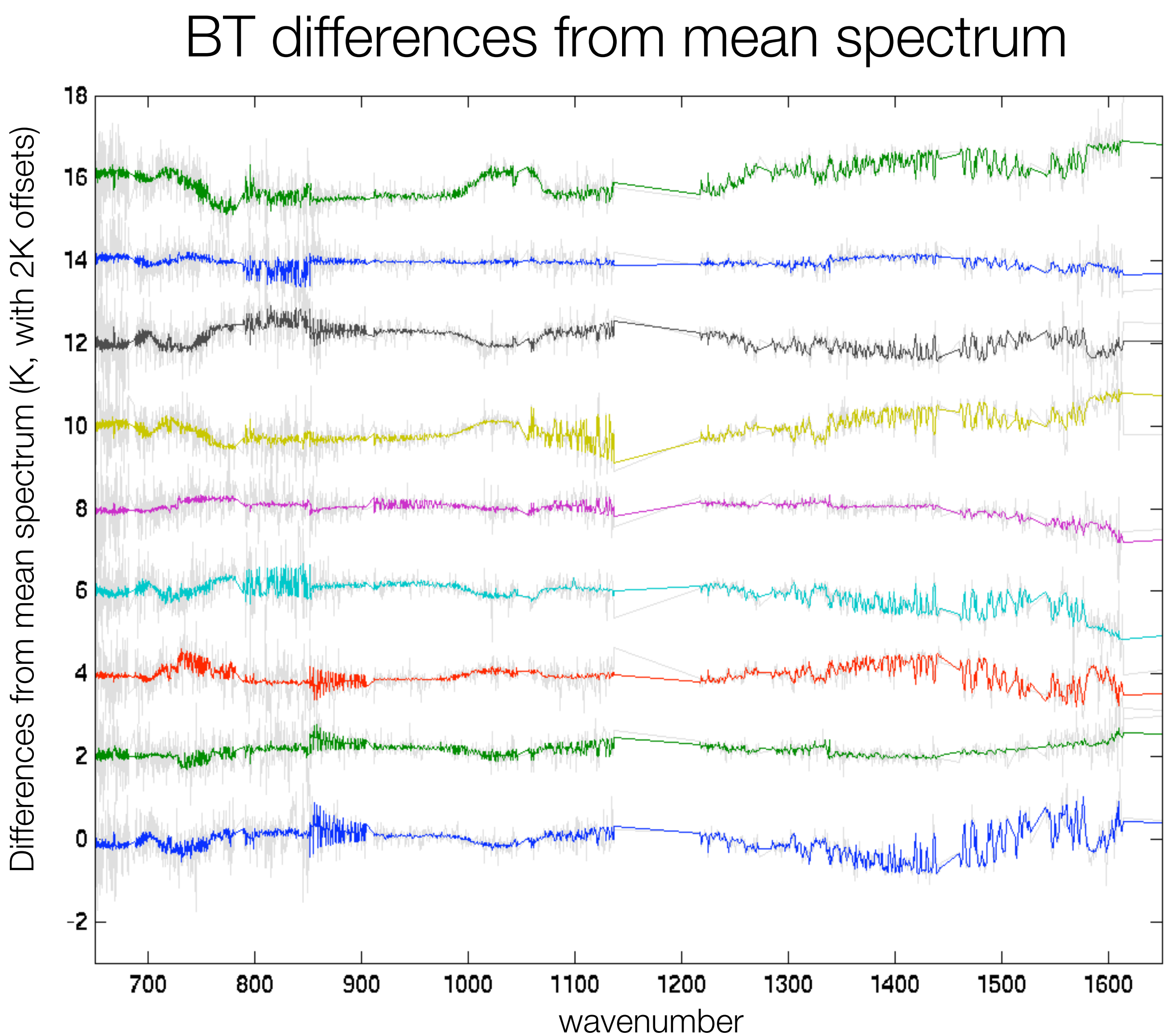
More individual AIRS spectra, 2005.04.01 granule 20(



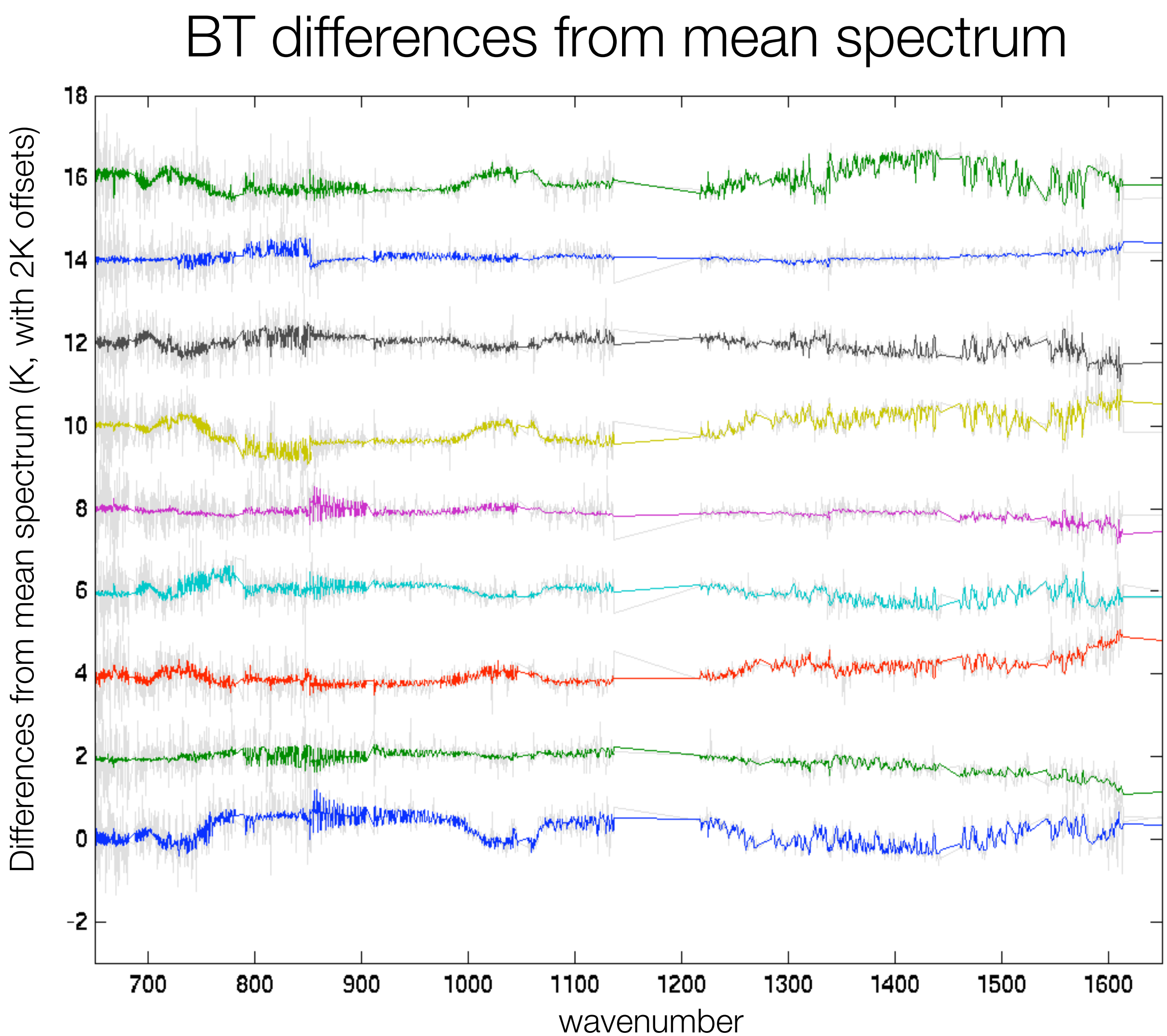
More individual AIRS spectra, 2005.04.01 granule 20(



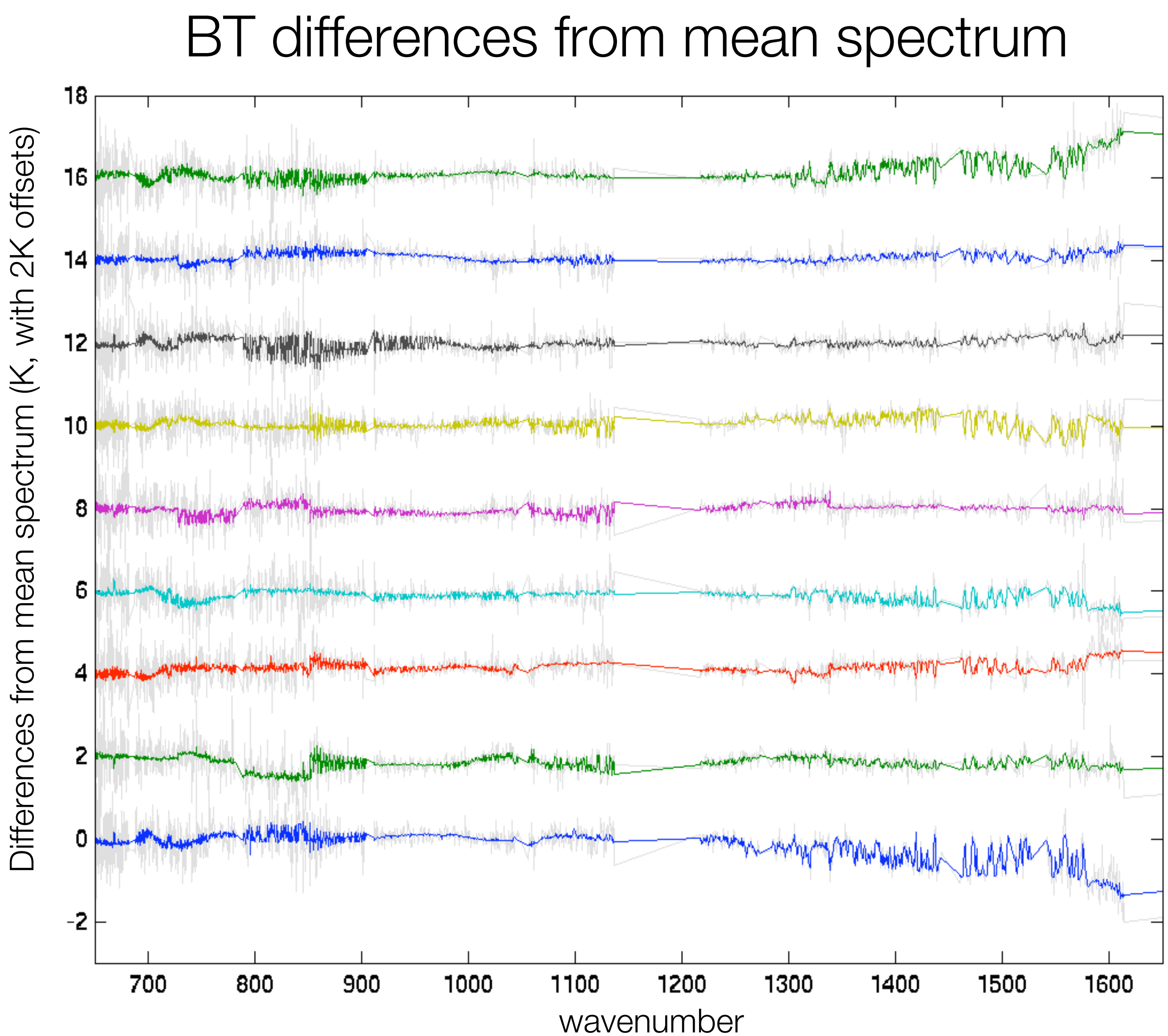
More individual AIRS spectra, 2005.04.01 granule 20(



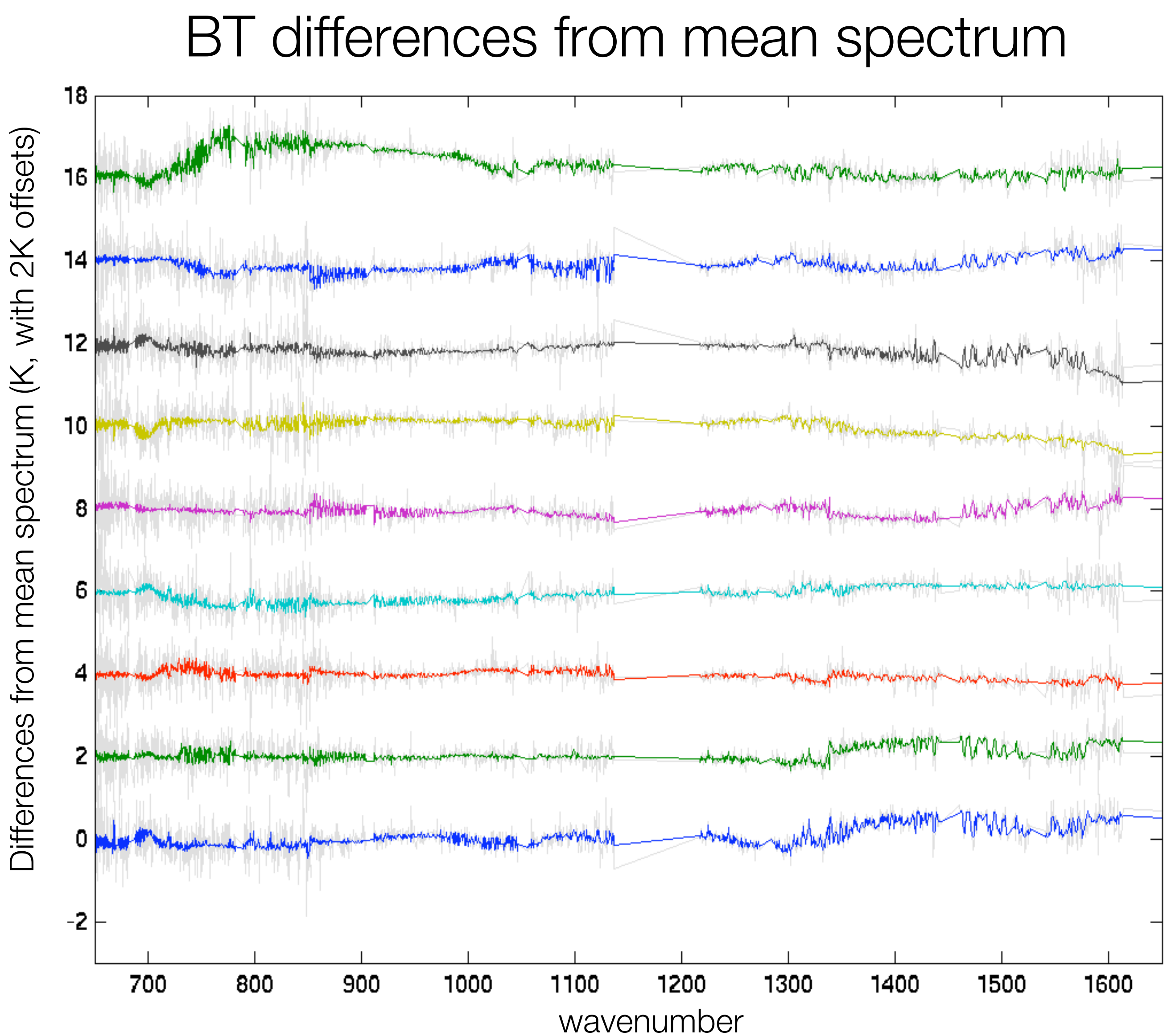
More individual AIRS spectra, 2005.04.01 granule 20(



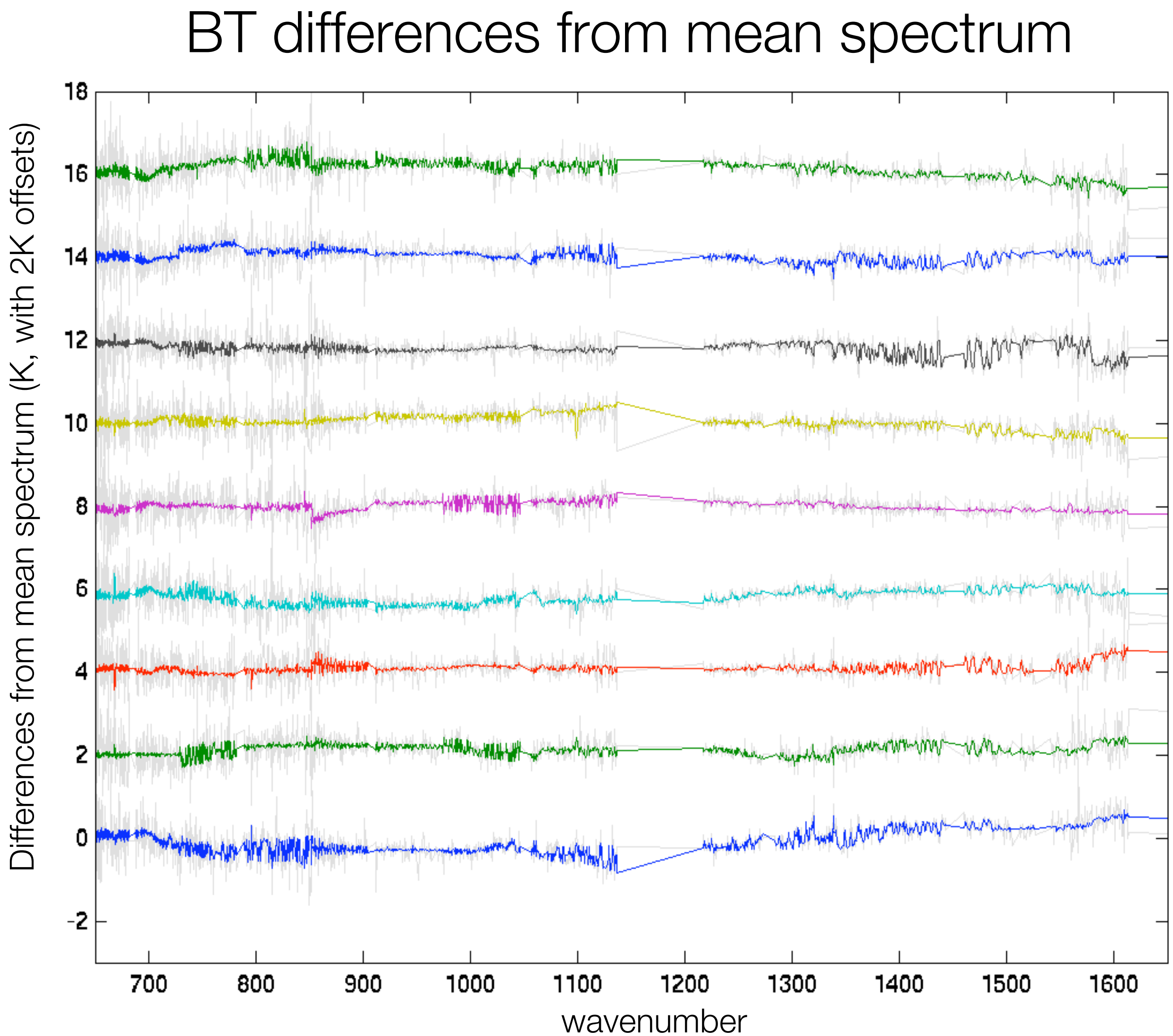
More individual AIRS spectra, 2005.04.01 granule 20(



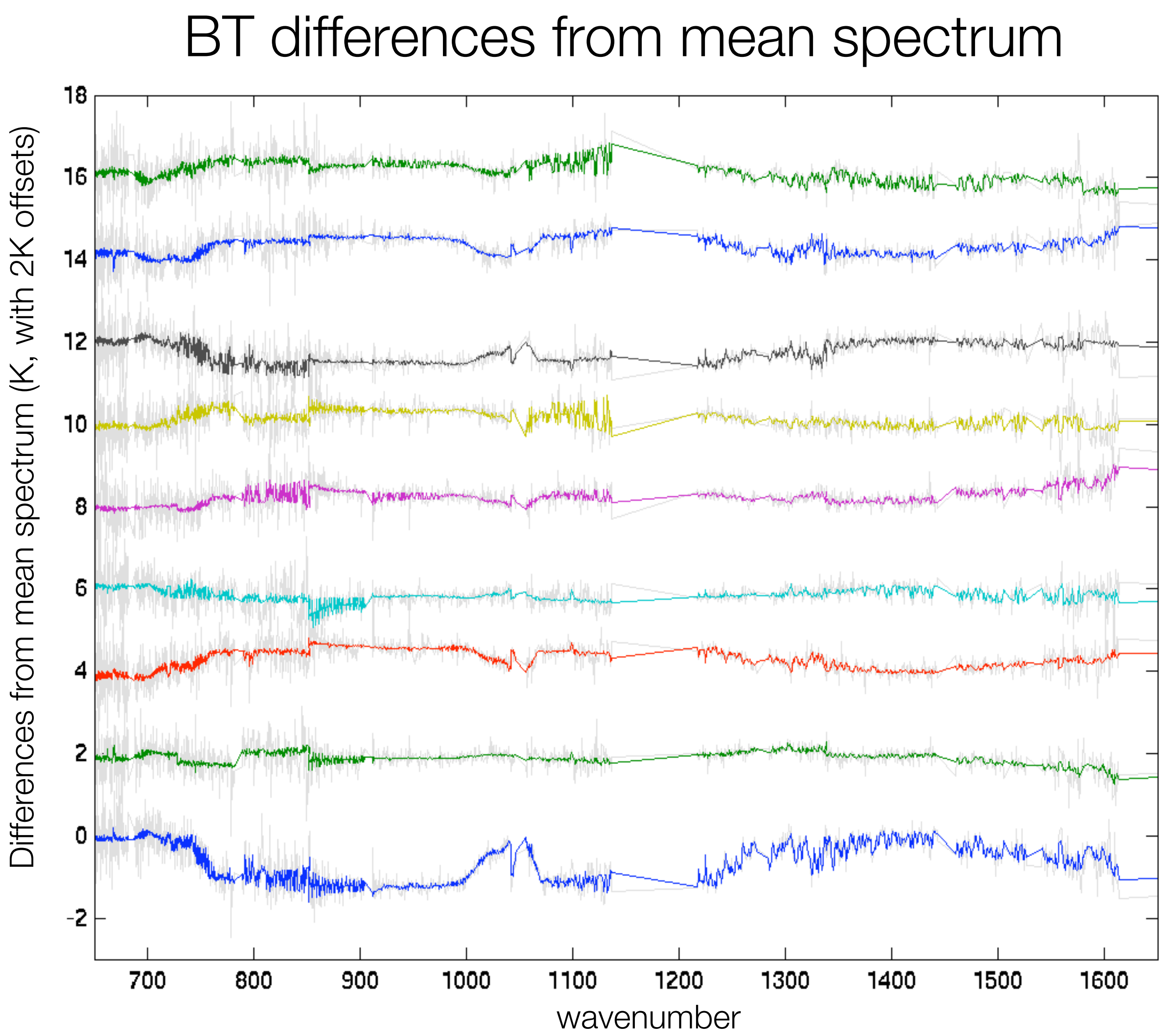
More individual AIRS spectra, 2005.04.01 granule 20(



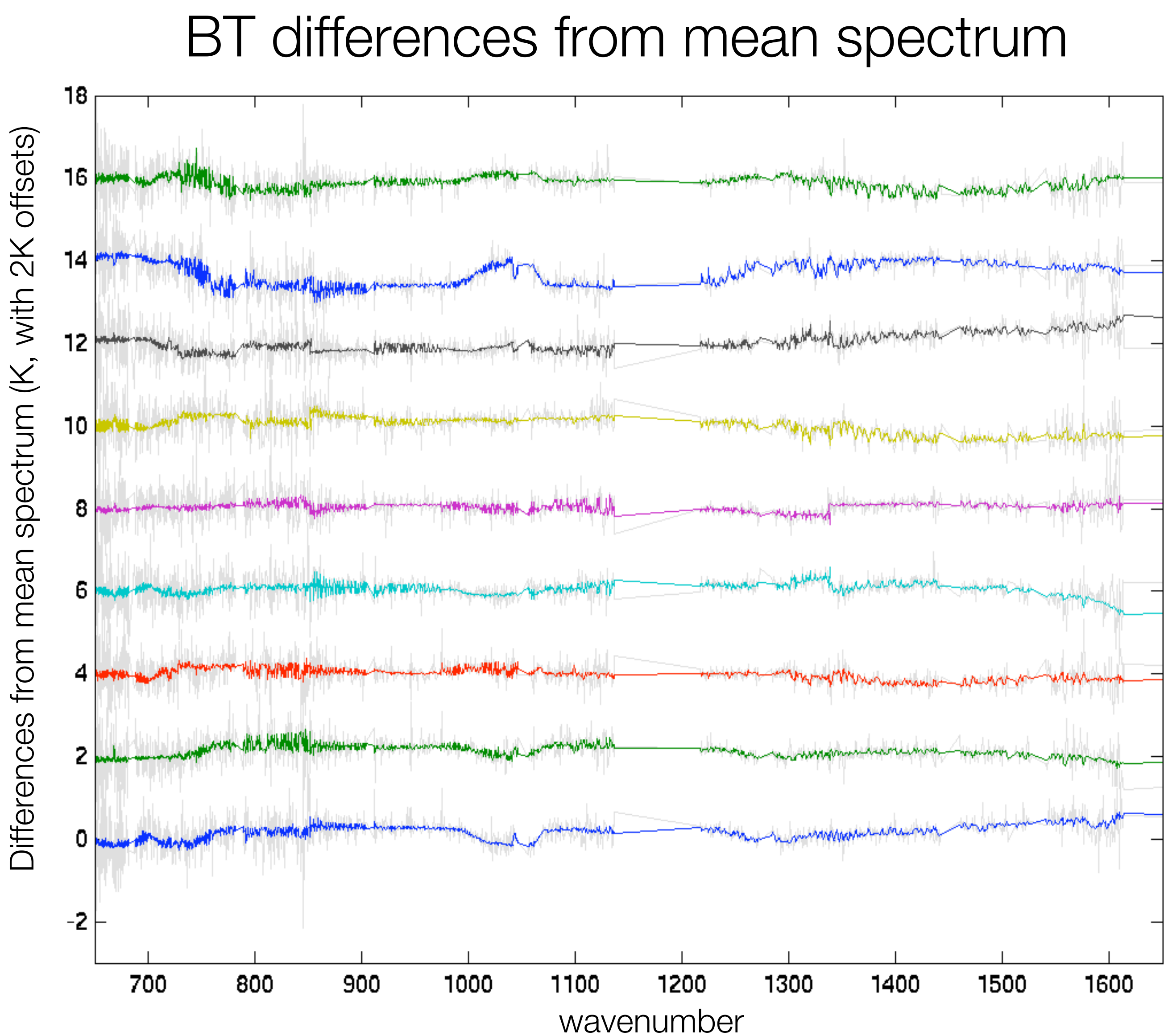
More individual AIRS spectra, 2005.04.01 granule 20(



More individual AIRS spectra, 2005.04.01 granule 20(



More individual AIRS spectra, 2005.04.01 granule 20(



Summary, Conclusions

Exploiting the redundancy in high spectral resolution observations, 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. Many of the findings presented here are consistent with analyses performed using pre-flight and on-orbit blackbody and space views, providing strong evidence for the validity of the PCA approach and results.

Specific findings:

- PCA estimates of AIRS spectrally random and spectrally correlated NEDN compare well with estimates computed from the on-board blackbody and space views.
  - The signal dependence of AIRS NEDN can be accurately parameterized in terms of the scene radiance (e.g. γ_{photon}, $NEDN_{Thermal}$). AIRS shortwave NEDN for a 300K scene is ~2x larger than for a 250K scene.
  - The PCA estimate is of the spectrally random noise; estimates of the AIRS spectrally correlated noise using [total_noise2 - pca_noise2]1/2 agree very well with preflight determinations using blackbody data. The spectrally correlated noise is a large fraction of the total noise for several detector arrays.
- Examination of N_{orig}-N_{recon} allows other non-Gaussian phenomenon to be characterized.
  - Many longwave and midwave PV detectors exhibit “popping” behavior above that expected from pure Gaussian behavior; shortwave channels do not exhibit this “popping”.
  - For v4 L1B data, only a small percentage (14 out of 2378) of channels exhibit “striping”.
- Inspection of the PCs and individual PC filtered radiance spectra reveal effects that lie in the less understood domain between calibration (long average) and spectrally random, repeatable noise.
  - The radiometric performance of AIRS at the level of the NEN contains artifacts not described by spectrally and temporally random noise, or by long-term calibration uncertainty.
  - Large ensemble averages show A / B state dependent artifacts on the order of ±0.4K for M-08.
  - For several arrays, the spectrally correlated noise is large, and dependent on A / B state (i.e. A-A and B-B channel correlation, but not A-B), generally consistent with Weiler ADFM#620.

Next

- Further characterization and parameterization of the AIRS array-correlated noise
- Study the impact of these findings on AIRS retrievals
- Continue with other sensors

The End

- Thank you