Compression of AIRS data with principal components

Jonathan Smith

European Centre for Medium-Range Weather Forecasts

With acknowledgements to :

Tony McNally, Jean-Noël Thépaut,

Mitch Goldberg, Walter Wolf & Tony Lee

Aims

1. To compare existing encoded AIRS products with model data
2. To investigate different methods of eigenvector creation

Outline of talk

i. Encoding data with principal components
ii. Comparison with ECMWF model background
iii. Initial assimilation trial with ECMWF model
iv. Creation of PC sets: all-sky vs clear
v. Test of reconstruction error

Encoding of a spectra

- **Given a set of spectral eigenvectors, arranged as columns of a matrix U, an observation, obs, is coded into a vector of coefficients, c, by**

$$c = \mathbf{U}^T \; obs$$

  - Where T denotes the transpose of the matrix

- **The reconstructed spectra, $recon$, is calculated from c by :**

$$recon = \mathbf{U}\, c$$

- **Spectral features that cannot be represented by the given eigenvectors will not be included in $recon$.**
- **These missing features can be summarised into a Reconstruction Error, RE, calculated for each spectra from :**

$$RE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(obs_i - recon_i)^2}$$

  - where n is the number of channels (after Goldberg *et al. 2003)*
  - Subscript BT, RE_{BT}, will be used when obs & $recon$ are brightness temperatures
  - Subscript R, RE_R, will be used when obs & $recon$ are in noise normalised radiance units

Components of RE

$$RE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(obs_i - recon_i)^2}$$

Information that ends up in RE includes :

1.
Noise

Random noise in a particular channel not reconstructed - information from other channels

"de-noising"

2.

Spectral features due to the atmospheric state

i. **Useful information, such as a structure not in U**

ii. **Information outside NWP model, ie aerosol**

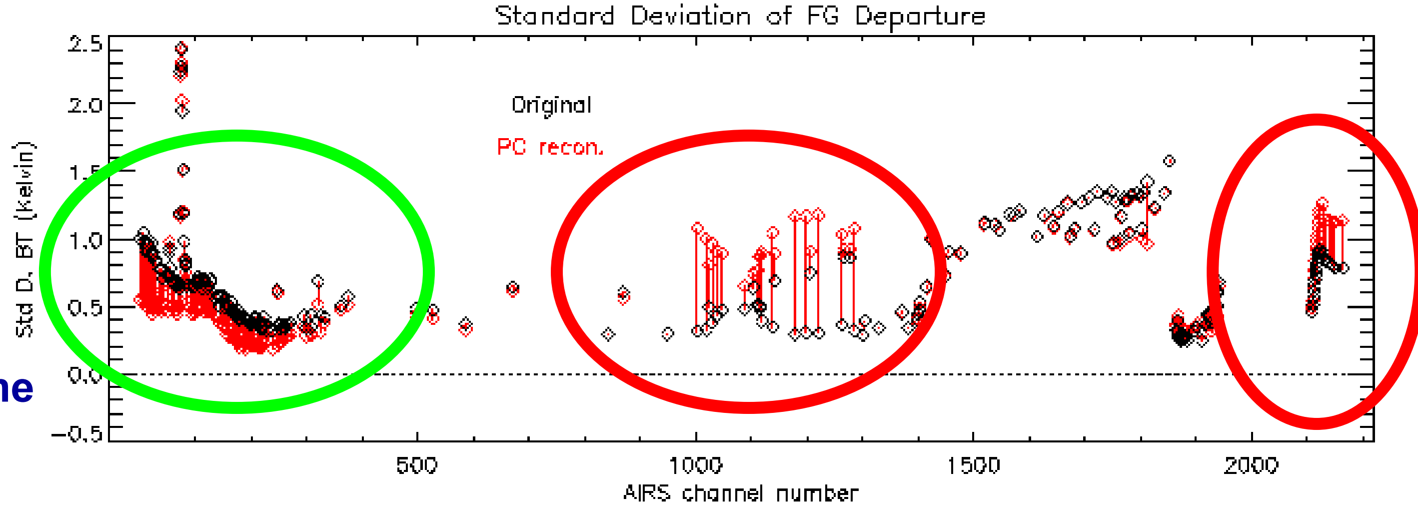
ii. Comparison with ECMWF model background

Comparing NESDIS PC data with ECMWF model

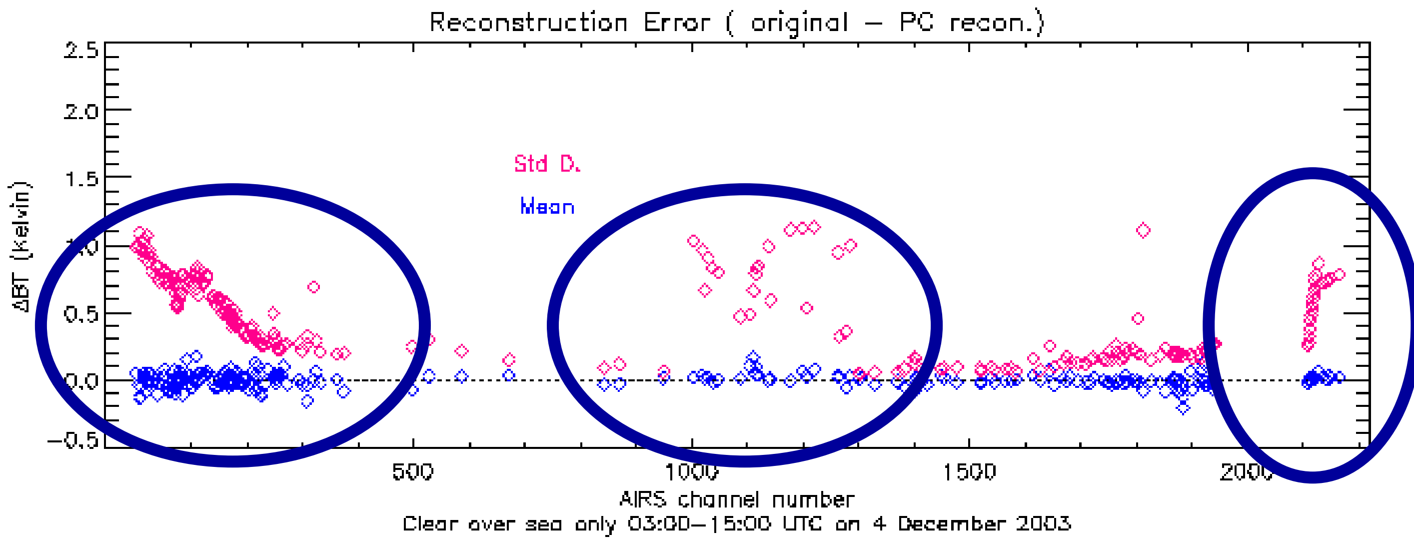
- **"1524" Principal Component set from NESDIS :**
  - Used 1,524 channels
  - created from data over 1 day, 20th December 2002 (thinned to reduce the volume only)
  - PC coefficients calculated for every central AIRS view from alternate "golfballs" - "U1" dataset of 1 / 18th data
  - first 200 PC coefficients transmitted for reconstruction using eigenvector set
- **Comparison over 12 hour cycle with:**
  - background from ECMWF operational model (CY26R3) and
  - standard 324 channel data

NESDIS PC departures from EC model

- Mean RE_{BT} ~ zero
- Some channels show higher RE_{BT} standard deviation (blue)



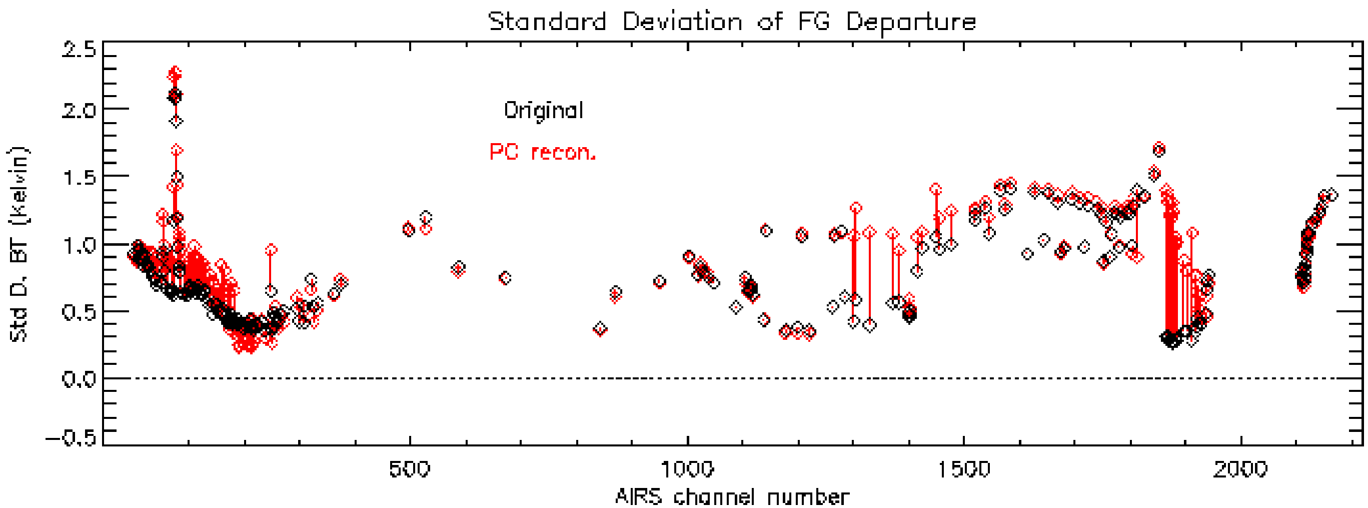
- some also show increased std deviation from model (red)
  - "tropical" daytime feature ?

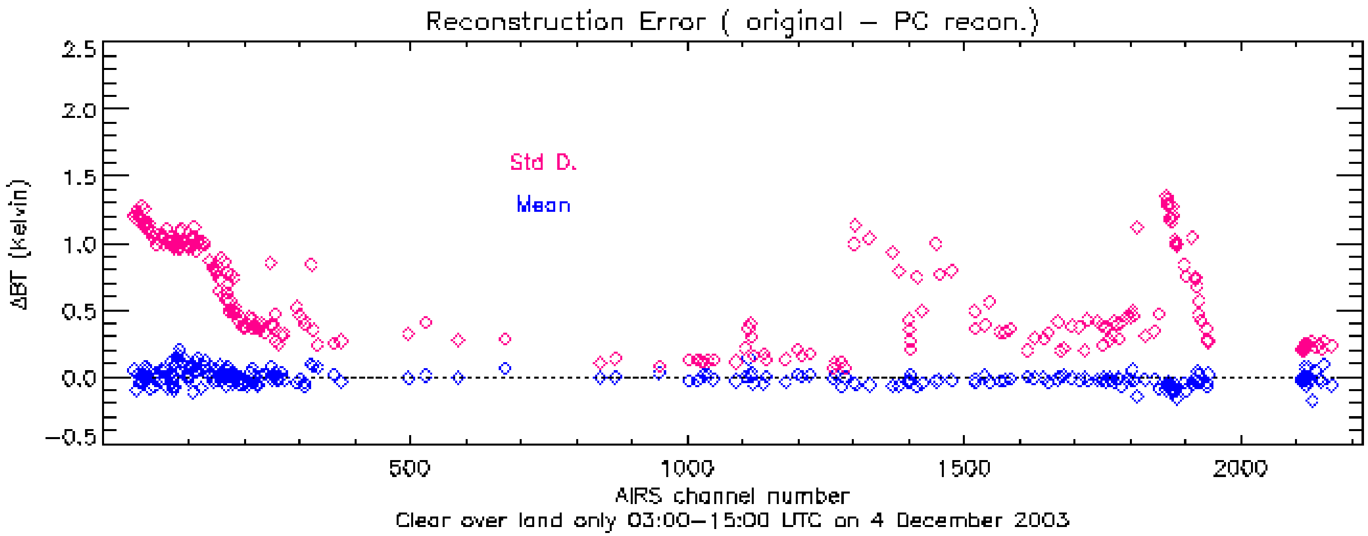


- Others - high, stratospheric, channels - show decrease in std deviation from model (green)
  - "de-noised"

Over land & at night …

- **At night increased SD went away**
- **Over land de-noising went away**
- **12 hours data in January**
  - still shows de-noising
  - Increased SD different





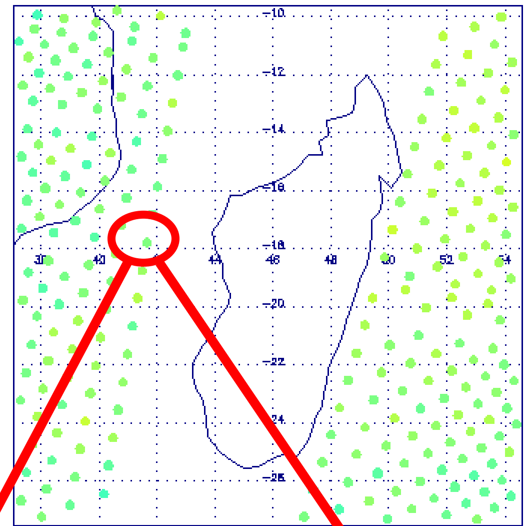
iii. Initial assimilation trial with ECMWF model

Reconstruted Radiance assimilation

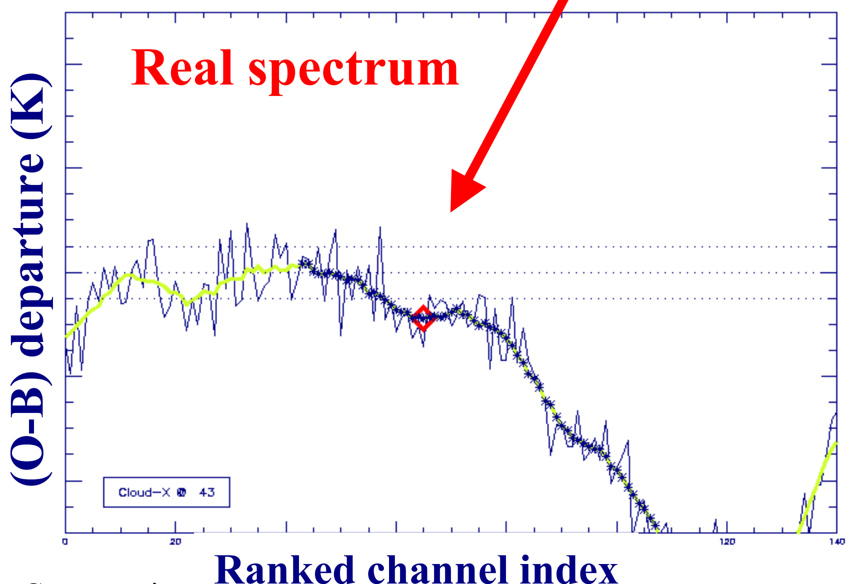
- **NESDIS "RR" data**
  - as the NESDIS "1524" PC data, except based on 2,047 channels PC's
  - Delivered as BT's over 322 channels
  - 200 coefficients used in reconstruction
- **Assimilated into EC operational model**
- **For a first attempt, error characteristics unchanged**
  - **… two following slides from Tony McNally**

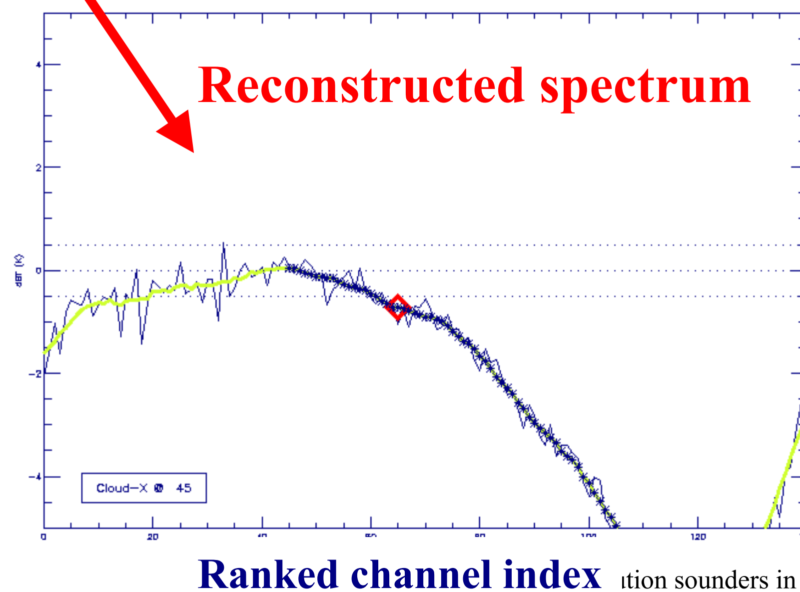
De-noising with 200 NESDIS principal components

The PC de-noising influences the filtering used to reduce noise during the cloud detection process.
(2 ranked channel change)



Channel to channel departure reduced



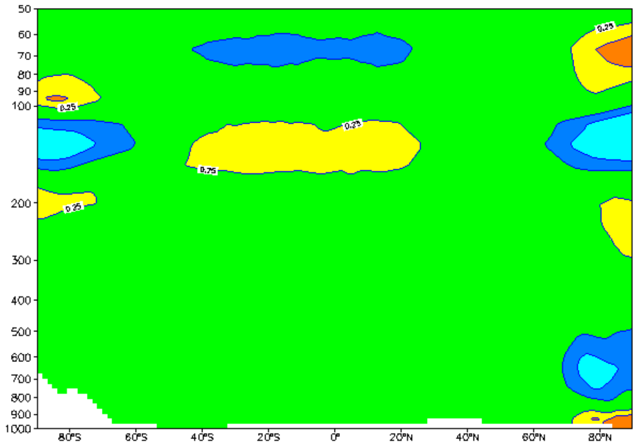


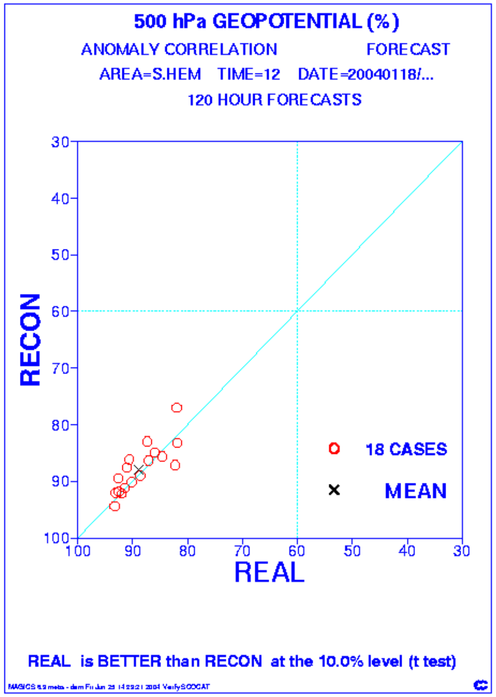
Assimilation experiments with NESDIS PC reconstructed radiances

Slightly larger stratospheric analysis increments obtained with the reconstructed (de-noised) radiances

possible *organisation* of radiance signal ?

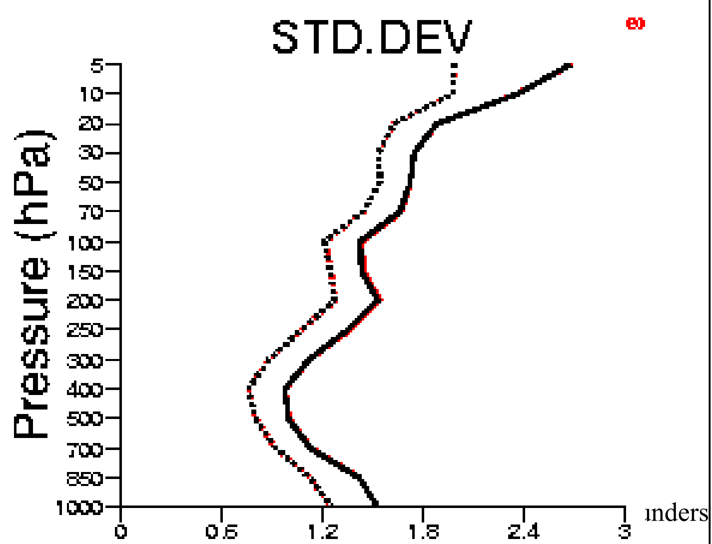
Zonal mean of RMS analysis increments





No impact on forecast quality or analysis fit to other data

Fit to radiosonde temperature data



al components.

unders in NWP

iv. Creation of PC sets: all-sky vs clear

Motivations for study

- **De-noising to suit NWP application**
- **Investigate different "training" sets for PC**
  - Can PC based only on clear views better reconstruct clear views ?
  - Clear based PC allow addition of cloud signal eigenvectors :
    - **Gather residual spectra from several cloudy views**

$$residual = obs - \mathbf{U}c$$

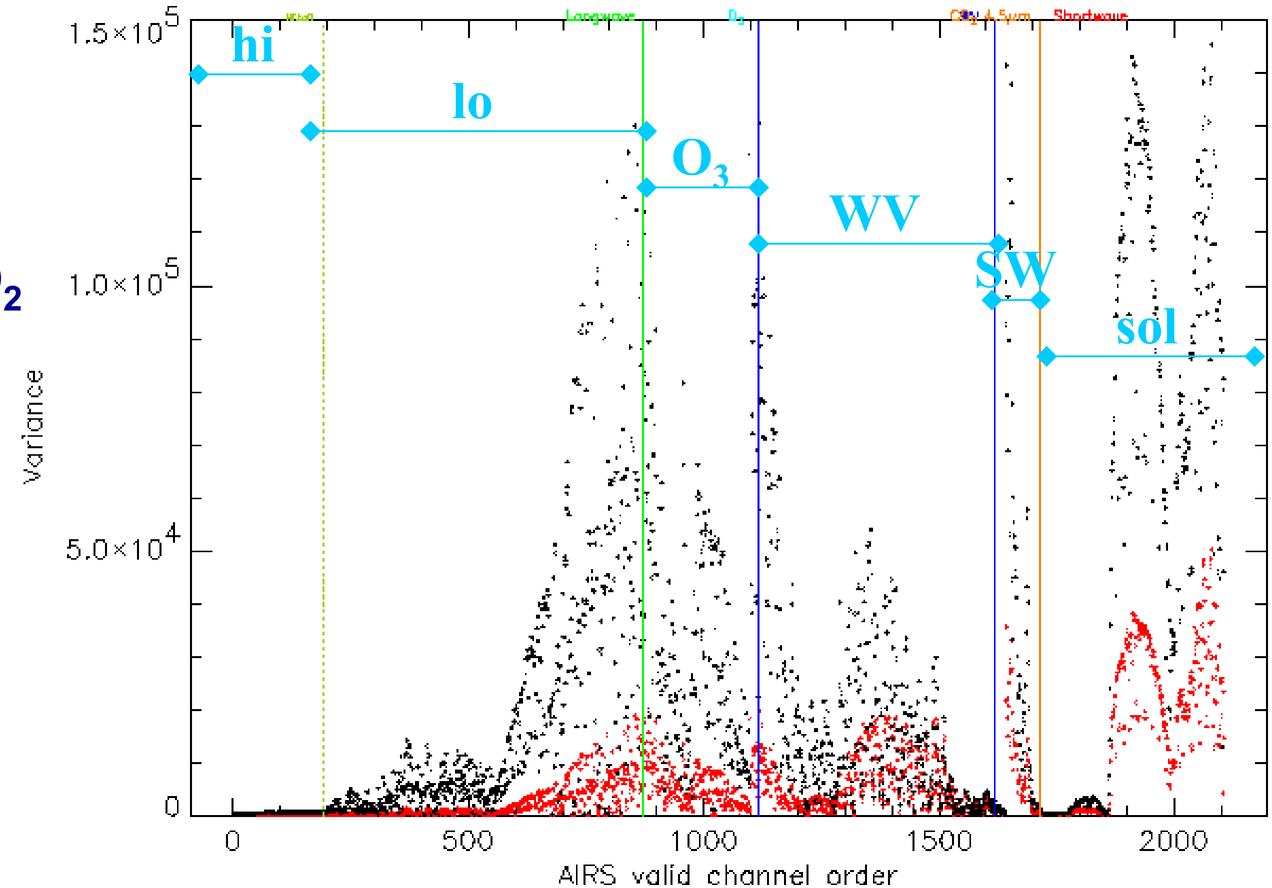
    - **Singular value decomposition to create eigenvectors**
    - **Concatenated into a new U with new coefficients added to end of *c***
    - **Can be repeated for other scenes … .. .**

Principal Component test sets

- **Two sets from data in July 2003**
  - "All" set
    - one day (15th), all views, all angles, land & sea, 1/9th thinned (324,000 spectra)
  - "Clears" set
    - one day (16th), over sea, clear at AIRS Level 2, unthinned (gave 85,000 spectra, *~3% clear*)
    - Added further high latitude (> 40°) clears from 15th
- **For both**
  - 2,107 channels used
    - channels valid for AIRS Level 2
  - Radiance data, noise normalised
    - instrument noise taken from channel properties file
  - calculate spectral covariance matrix and derive eigenvectors from that

PC sets' variance

- Diagonal of covariance matrices
  - In noise normalised radiance
  - Black - "All"
  - Red - "Clears"
- Spectral Bands
  1. high peak LW CO_2
  2. low peak LW CO_2
  3. ozone
  4. water vapour
  5. SW CO_2
  6. solar end



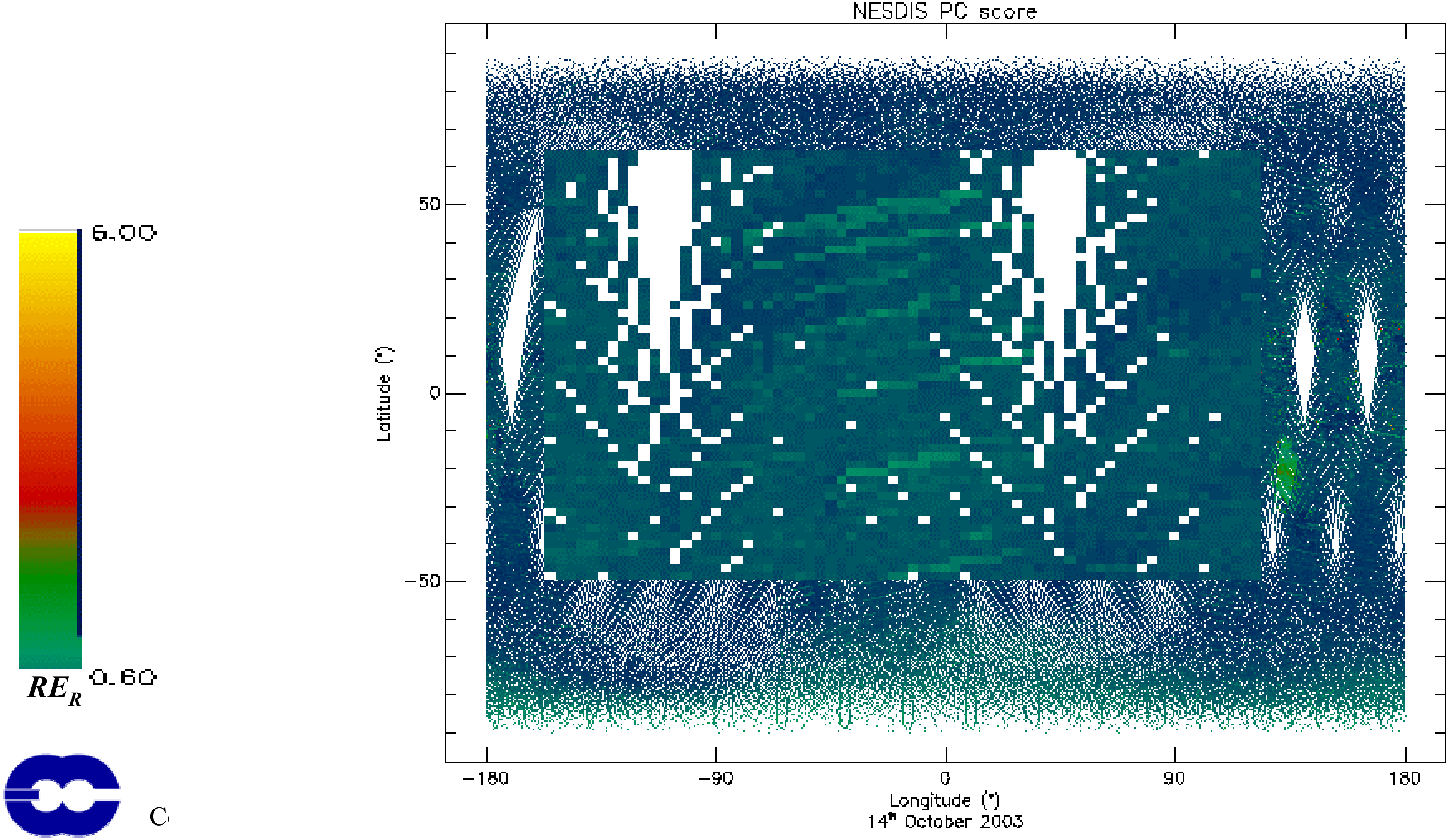
v. Test of reconstruction error

Full spectra comparison data

- **Illustrative data from 1 in nine spectra from one day - 14th October 2003**
  - (processed granules 1 to 188)
- **"All" view based PC compared with "Clears" PC**
- **Maps of RE_R**

Mapped RE_R

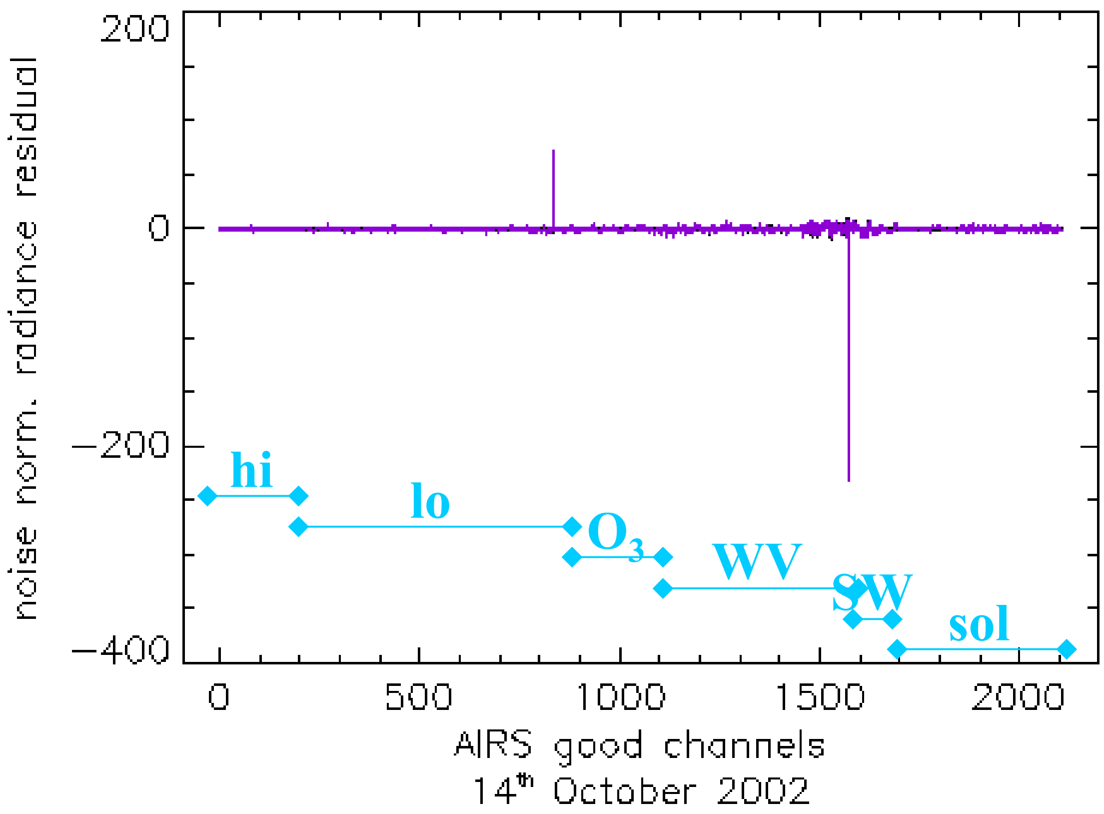
- clear views from Level 2 - mainly between 40°N to 40°S
- 63 % of clear spectra have lower RE_R with “Clears” PC
- for average RE_R, 4% lower than NESDIS PC
- for comparison RE_r from “Clears” PC, 26% lower than NESDIS PC



C

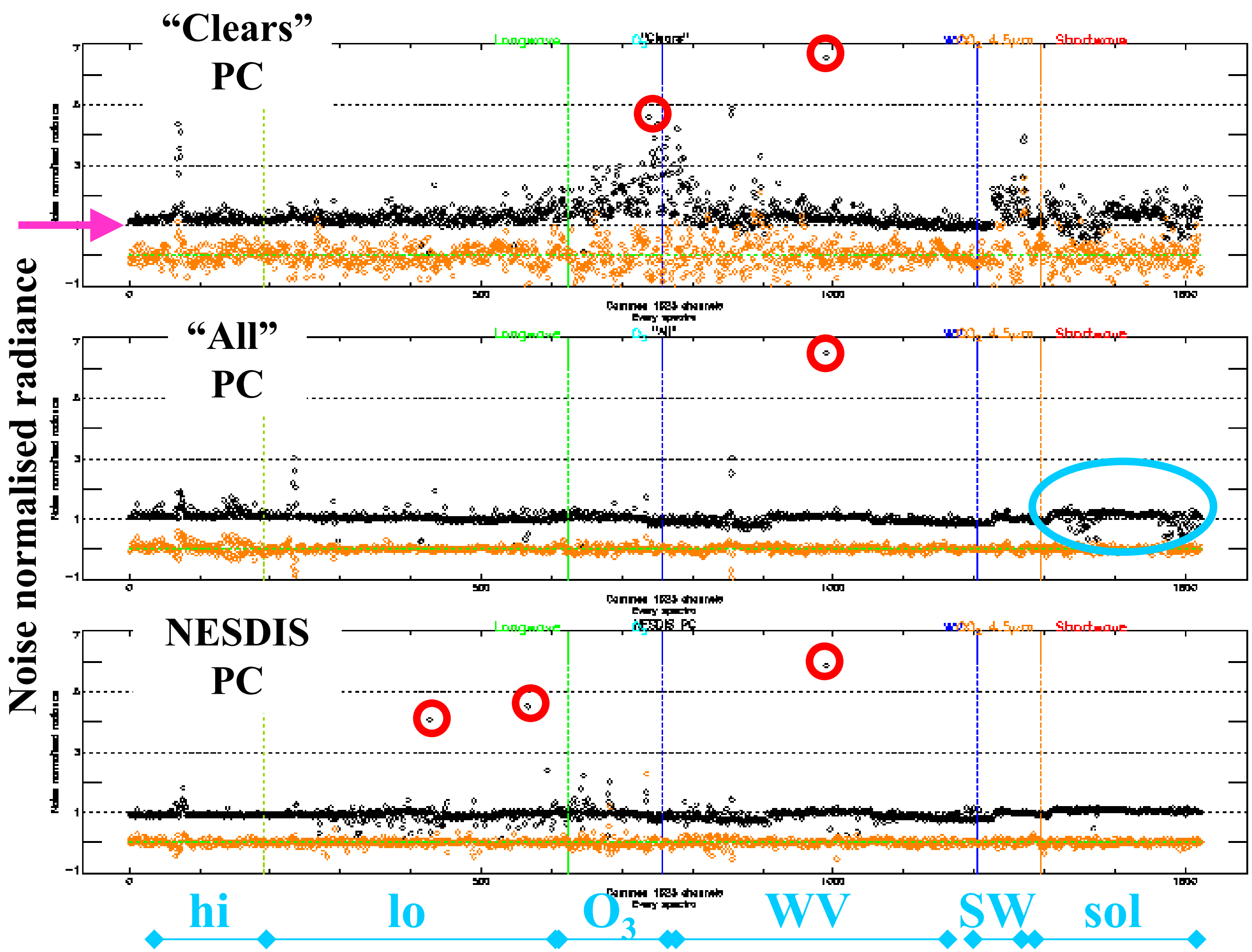
Results - time consistency

- ***RE* compromised by channel changes**
  - Plot shows RE_r from October 2002 using “All” & “Clears” PC sets (derived from July 2003 data)
  - Spikes from channels : 957 **(982.0 cm^{-1})** & 1,791 **(1561.6 cm^{-1})** (pop & noise)



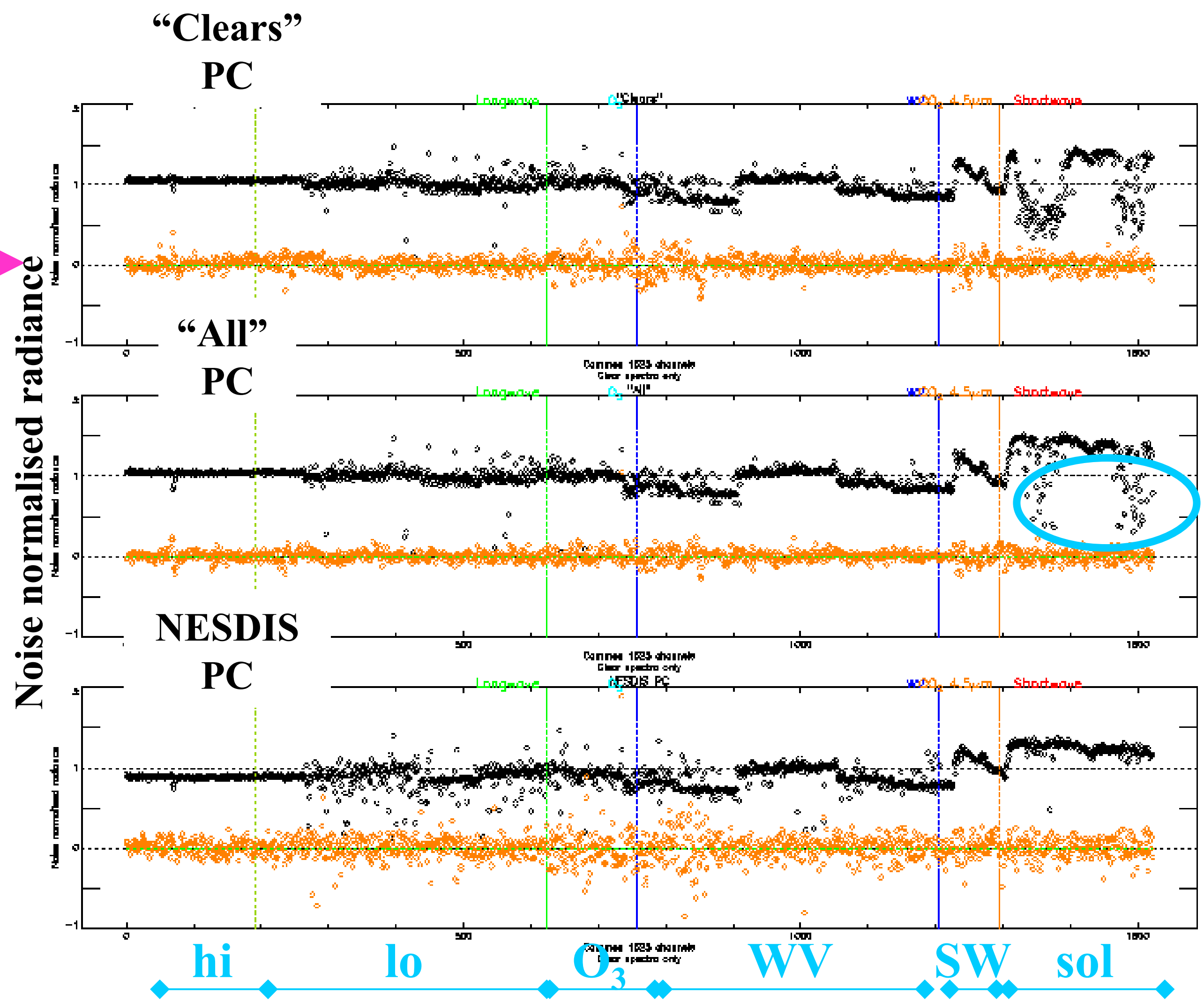
All views

- Analysis per channel of radiance reconstruction error
- Std Deviation in black and mean in orange
  - Over the 1,523 common channels in “All”, “Clears” & NESDIS PC
  - Data from 14th October 2003 from 254,800 spectra



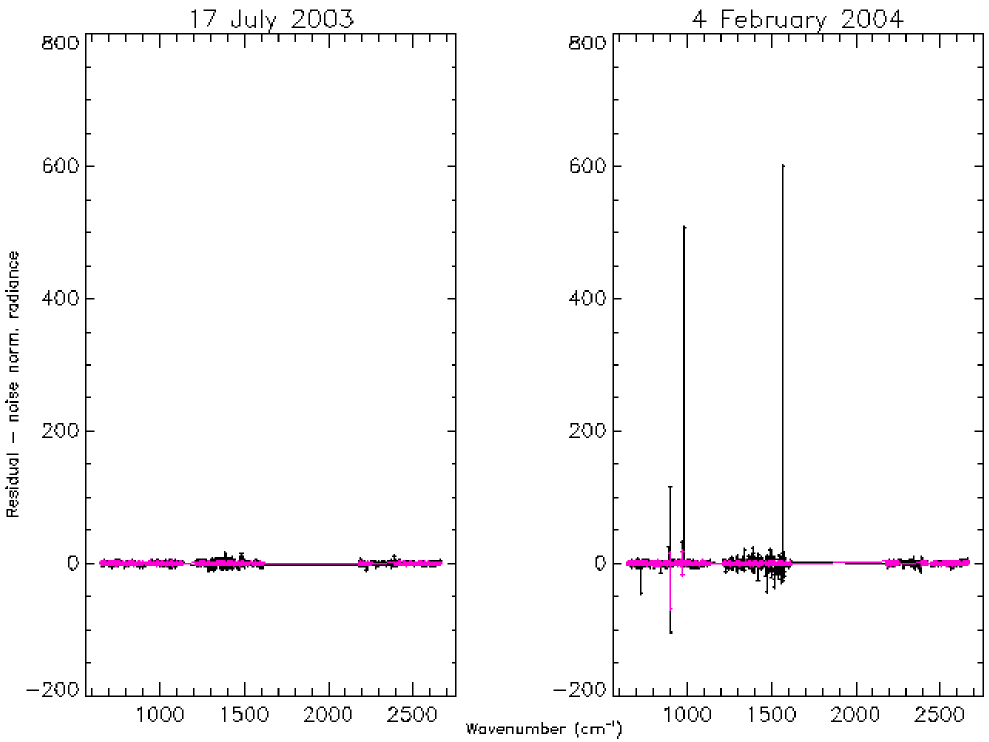
Clear views

- **Less noise in "clears"**
  - NESDIS Std Dev. less for high LW CO_2
  - (outliers still present - offscale)



Solar shutdown changes

- **Several detectors not the same on power up**
- **Show up in RE_r**
- **Comparison of July ‘03 and February ’04**
  - Spikes in February data for channels : 756 **(899.3 cm^{-1})**, 765 **(902.4 cm^{-1})**, 957 **(982.0 cm^{-1})** & 1,802 **(1569.3 cm^{-1})** none in the 324 operational channels
  - Operational channels changed were : 318 **(741.3 cm^{-1})**, 1,883 **(2197.9 cm^{-1})** & 1,884 **(2198.8 cm^{-1})**(FG dep. > 3 x Std Err.)



Conclusions

- **De-noising**
  - NESDIS PC compared with model first guess
    - Seen for stratospheric channels
    - Other channels the same or noisier
  - NESDIS RR assimilated, as above plus
    - Increased stratospheric increments
  - Need not "de-noise" all channels if residual available
  - Monitoring needs RE and FG departure

- **"Clears" PC**
  - Improvement found for majority of clear views
    - ✗ It was only a small majority
    - ? Improvement big enough ?
    - ? Tests so far only over limited number of clears from Level 2
    - ? Too many bad channels

Conclusions (continued)

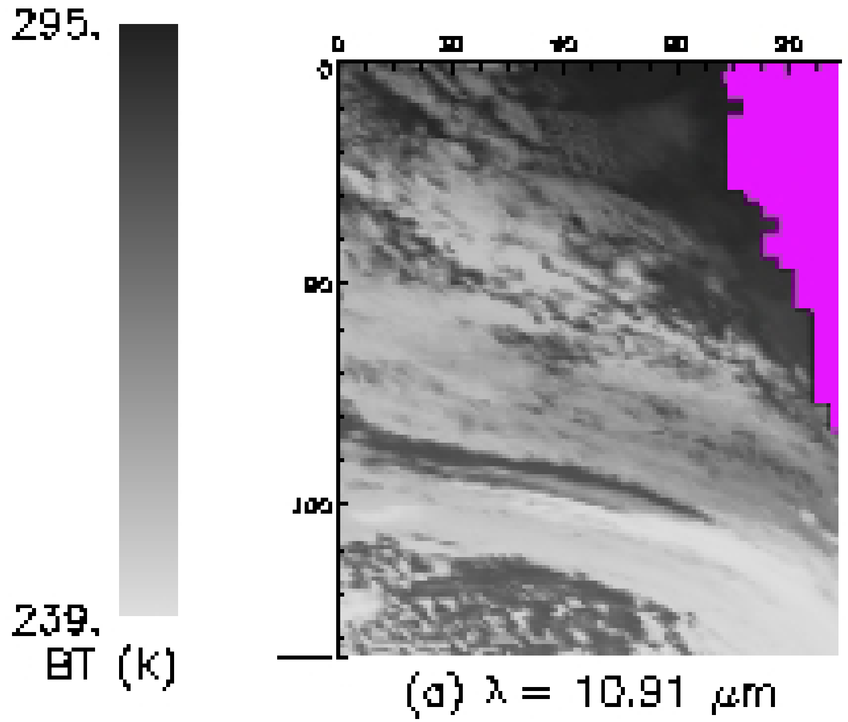
- **In particular for AIRS :**
  - channels that go bad are a long term problem
  - line by line offset significant when looking for the one clear view for NWP

?. but ….

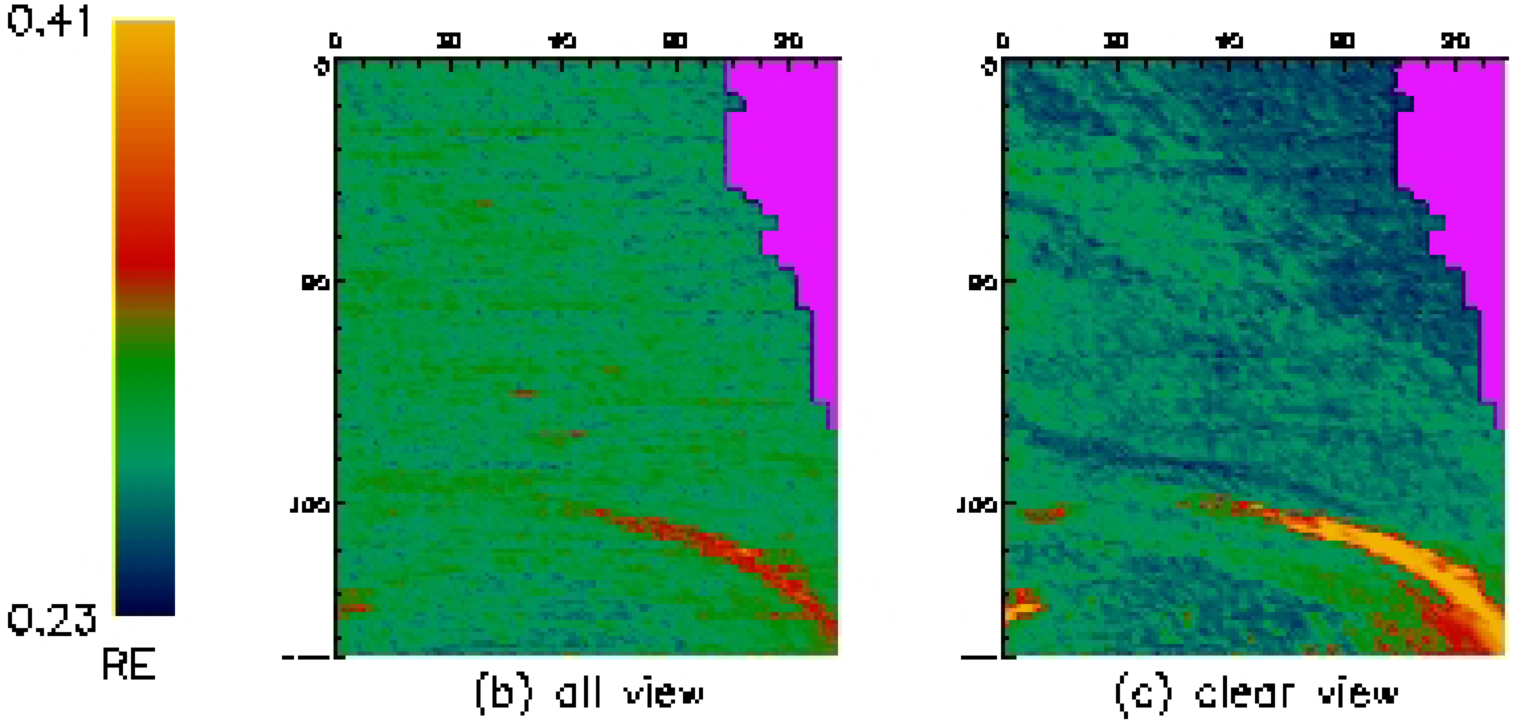
Better clear PC set

- PC set similar to "clear-sky"
- But 75 % of spectra not normalised by noise
- Decreased mean
- Overall increase in variance, less at shortwave

Example: Western Australia (purple), Granule 180, 15th October 2003



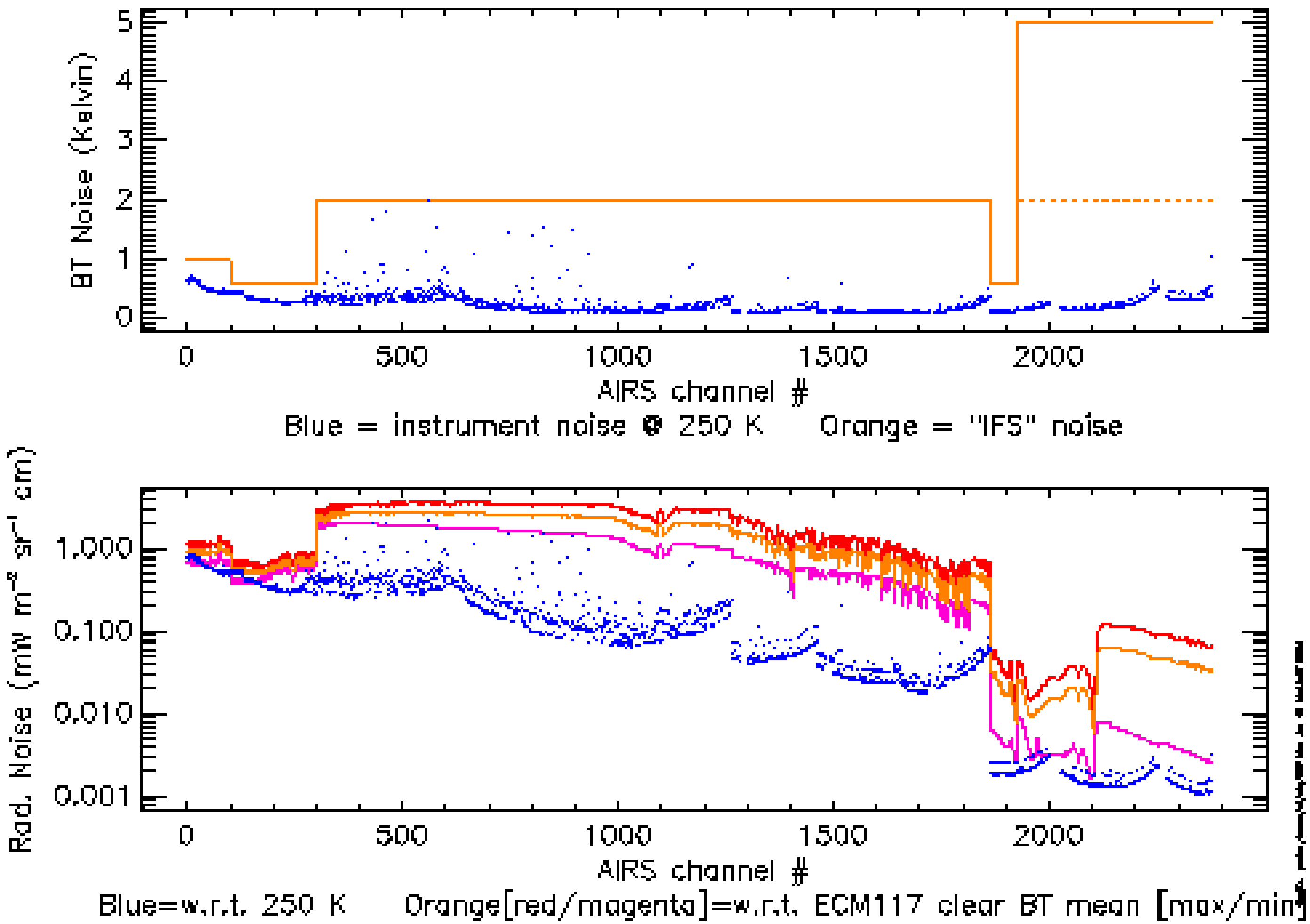
Lower RE_{BT}



- **83% of clears with lower RE_{BT} than NESDIS PC set**

extra … .. .

Noise



BT Noise (Kelvin)
5
4
3
2
1
0
0
500
1000
1500
2000
AIRS channel #
Blue = instrument noise @ 250 K Orange = "IFS" noise
Rad. Noise (mW m-2 sr-1 cm)
1.000
0.100
0.010
0.001
0
500
1000
1500
2000
AIRS channel #
Blue=w.r.t. 250 K Orange[red/magenta]=w.r.t. ECM117 clear BT mean [max/min]

Lines across RE_r

- **Due to offset calibration at end of each line**
  - details of "striping" in Steve Gaiser's presentation March '04 Science Team meeting
  - Overall bias zero - but adds 5% noise overall
- **Line to line variability significant**
- **Example …..**