The use of principal components in processing AIRS data

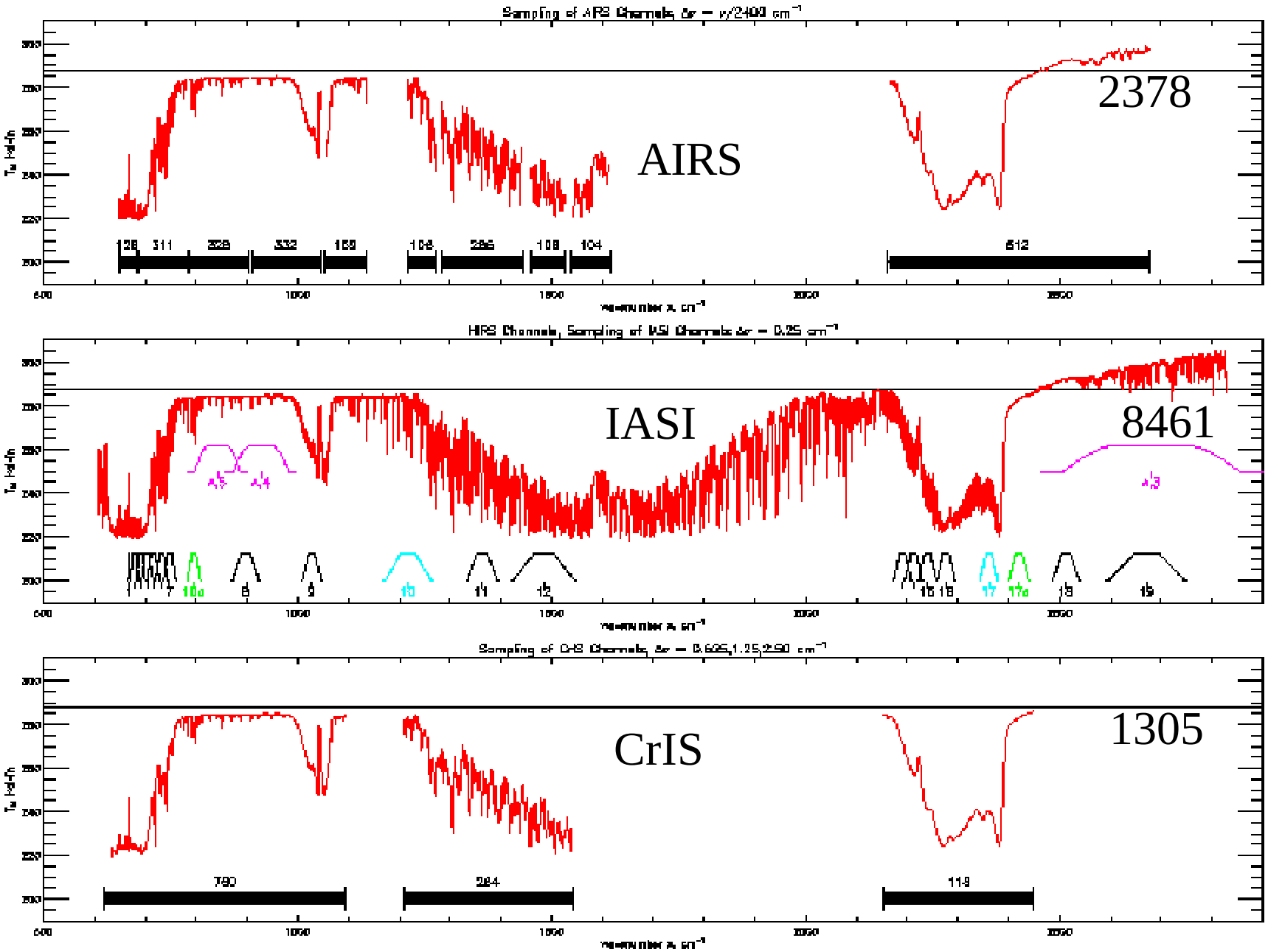
Mitchell D. Goldberg

NOAA/NESDIS/Office of Research and Applications

01/25/02

Applications of Eigenvectors to AIRS Data

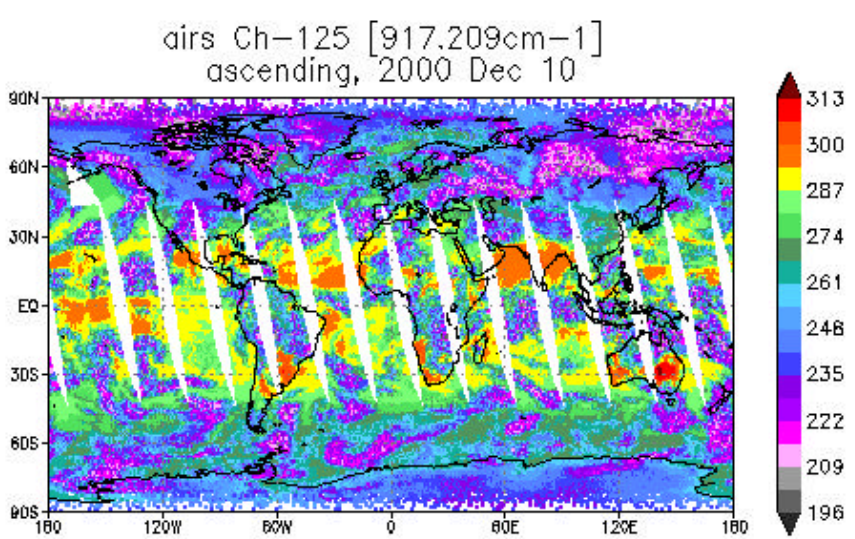
- NESDIS will provide AIRS data in near-real time to NWP centers.
- Data size is very large compared with current sounders (1 orbit ~ 2GB vs. 8 MB)
- Information is not independent. Eigenvectors provide an effective way to extract independent pieces of information.
- Eigenvector expansion coefficients (principal component scores) can be provided instead of the individual channels.
- Individual channels can be reconstructed with minimal signal loss.

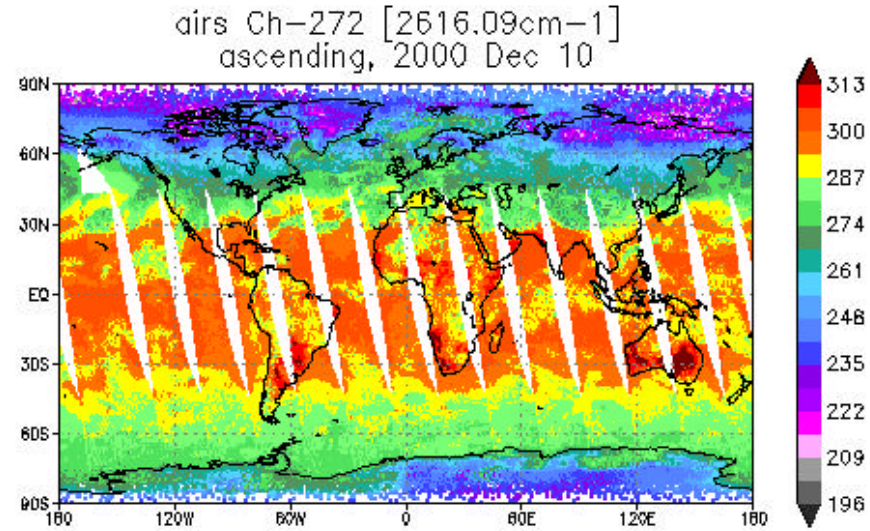

AIRS
2378
IASI
8461
CrIS
1305

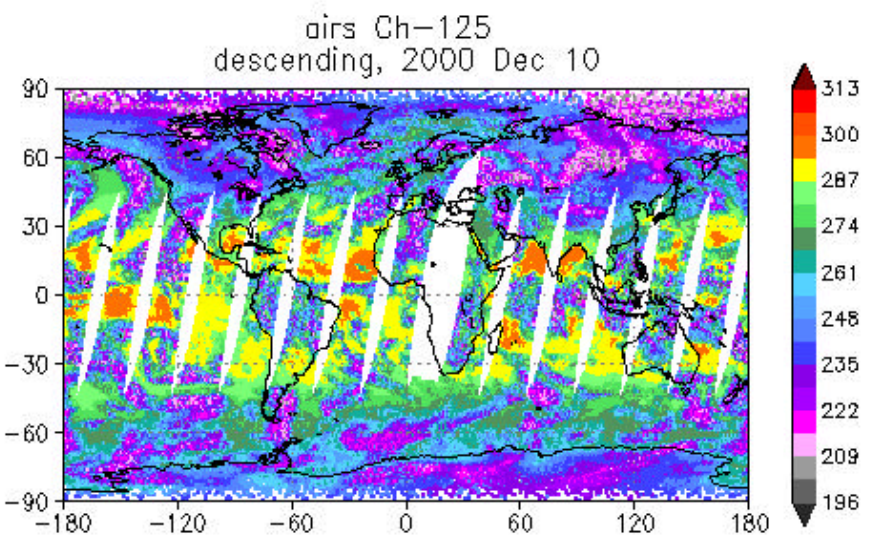
AIRS/AMSU/HSB Simulations

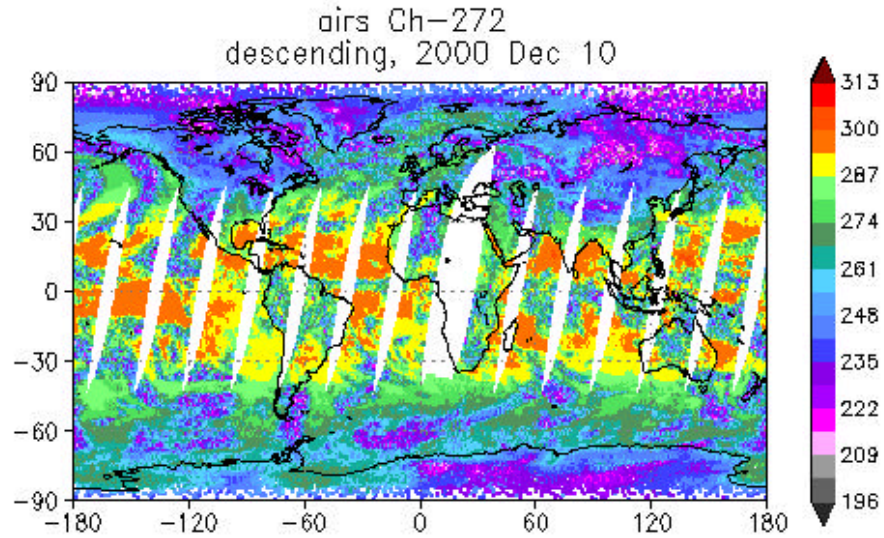
- Includes surface topography and variable surface pressure
- Daytime and nighttime conditions
- T(p),q(p),o3(p) from surface to .005 mb.
- cloud liquid water profiles
- multiple level cloudy conditions with spectrally varying cloud emissivity and reflectivity, consistent with atmospheric conditions (clouds from global model,but cloud amounts are randomized)
- variable surface skin temperature, surface emissivity and surface reflectivity
- variable land coverage with coastlines, lakes, etc.
- variable view and solar zenith angles

Example of simulated AIRS window channels: LW, SW



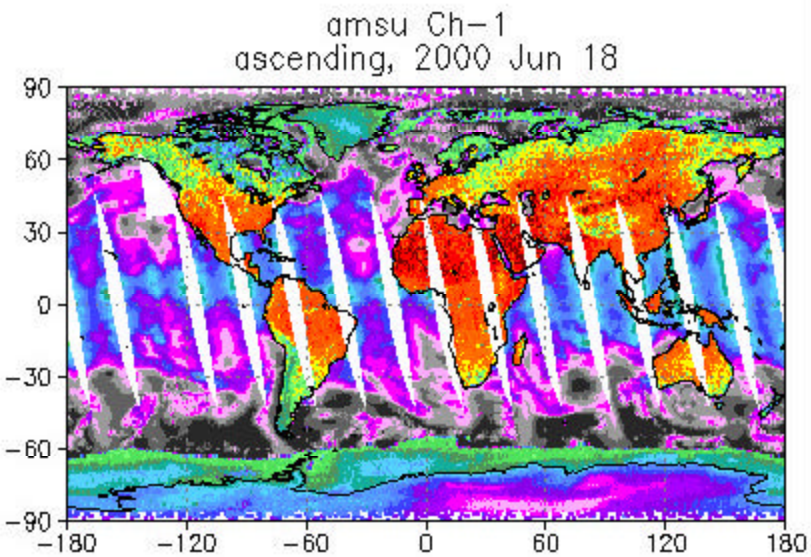


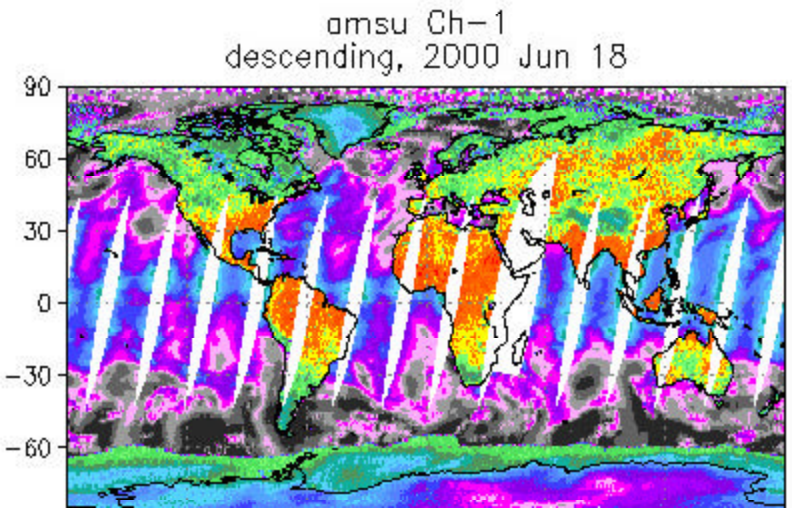


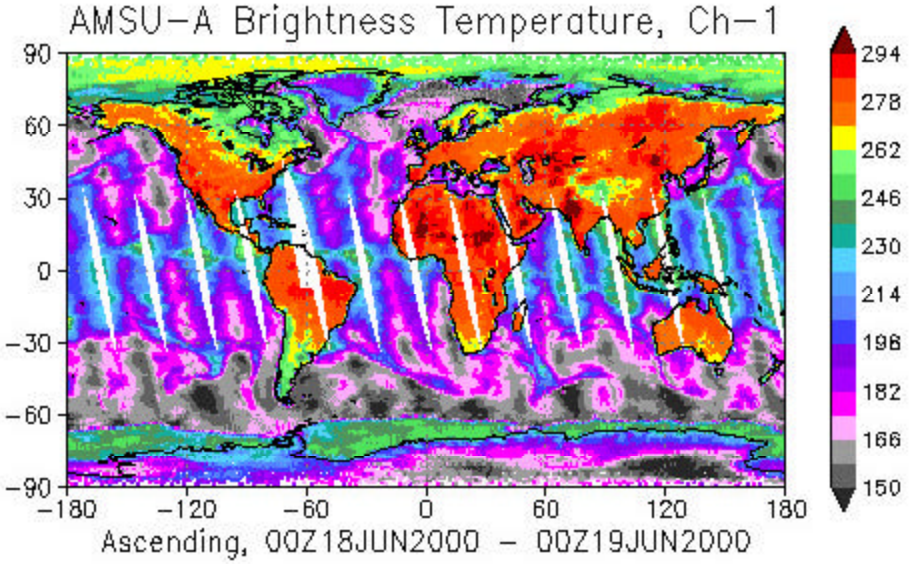


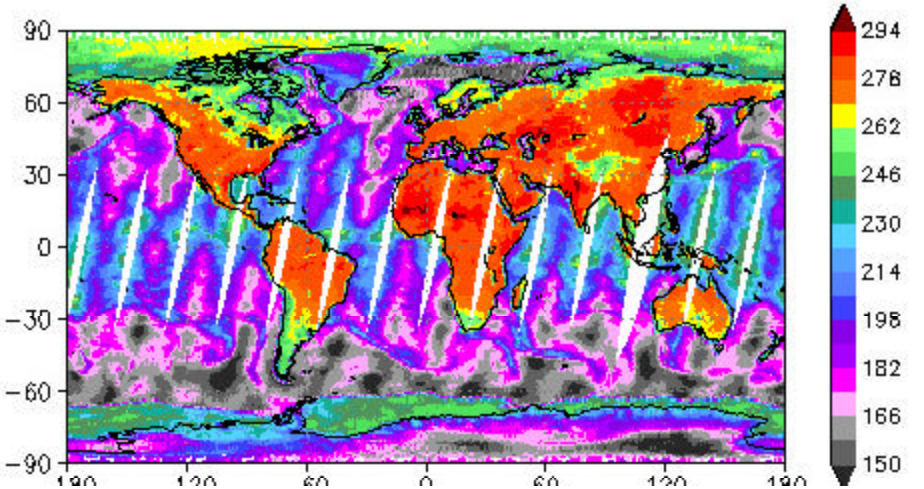
Simulated AMSU

Real AMSU









Principal Component Analysis

- Principal component analysis (PCA) is often used to reduce data vectors with many components to a different set of data vectors with much fewer components that still retains most of the variability and information of the original data

- $\mathbf{R} = r_1.\mathbf{i}_1 + r_2.\mathbf{i}_2 + r_3.\mathbf{i}_3 + \ldots + r_n.\mathbf{i}_n$

  where $\mathbf{i}_1 = (1,0,0,0,0\ldots.0_n)$; $\mathbf{i}_2 = (0,1,0,0,0\ldots.0_n)$

- Data are rotated onto a new set of axes, such that the first few axes have the most explained variance.

- $\mathbf{R} = p_1.\mathbf{E}_1 + p_2.\mathbf{E}_2 + p_3.\mathbf{E}_3 + \ldots + p_n.\mathbf{E}_n$
  where $\mathbf{E}$ are eigenvectors and $p_1 = \mathbf{R}\ \mathbf{E}_1$

- So instead of **R** vectors of length n, we can have a truncated **P** vectors of length m, where m << n

Eigenvectors are used for

- Data compression
- Quality control
- Retrievals of geophysical parameters
  - » Atmospheric temperature, moisture and ozone
  - » Surface temperature and emissivity

Topics

- Generation of eigenvectors
- Applications of eigenvectors in AIRS data processing.
- Monitoring of representativeness of eigenvectors

Generating AIRS eigenvectors

- Each AIRS data vector has 2103 radiance values.
- The radiances are normalized by expected instrumental noise (signal to noise)
- Compute the covariance matrix S
- Compute the eigenvectors E and eigenvalues Λ

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

- E = matrix of orthonormal eigenvectors (2103x2103)

  Λ = vector of eigenvalues (explained variance)

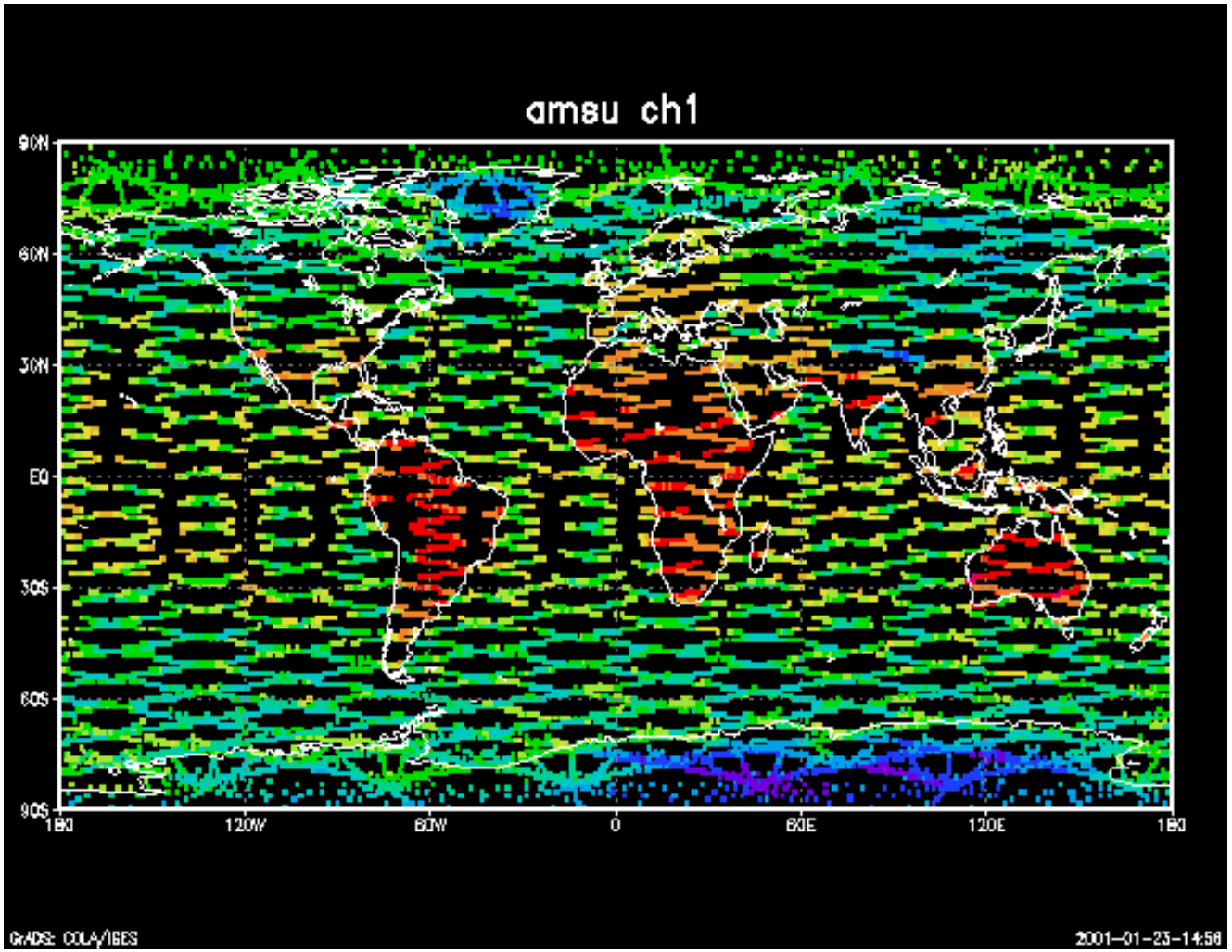
Applying AIRS eigenvectors

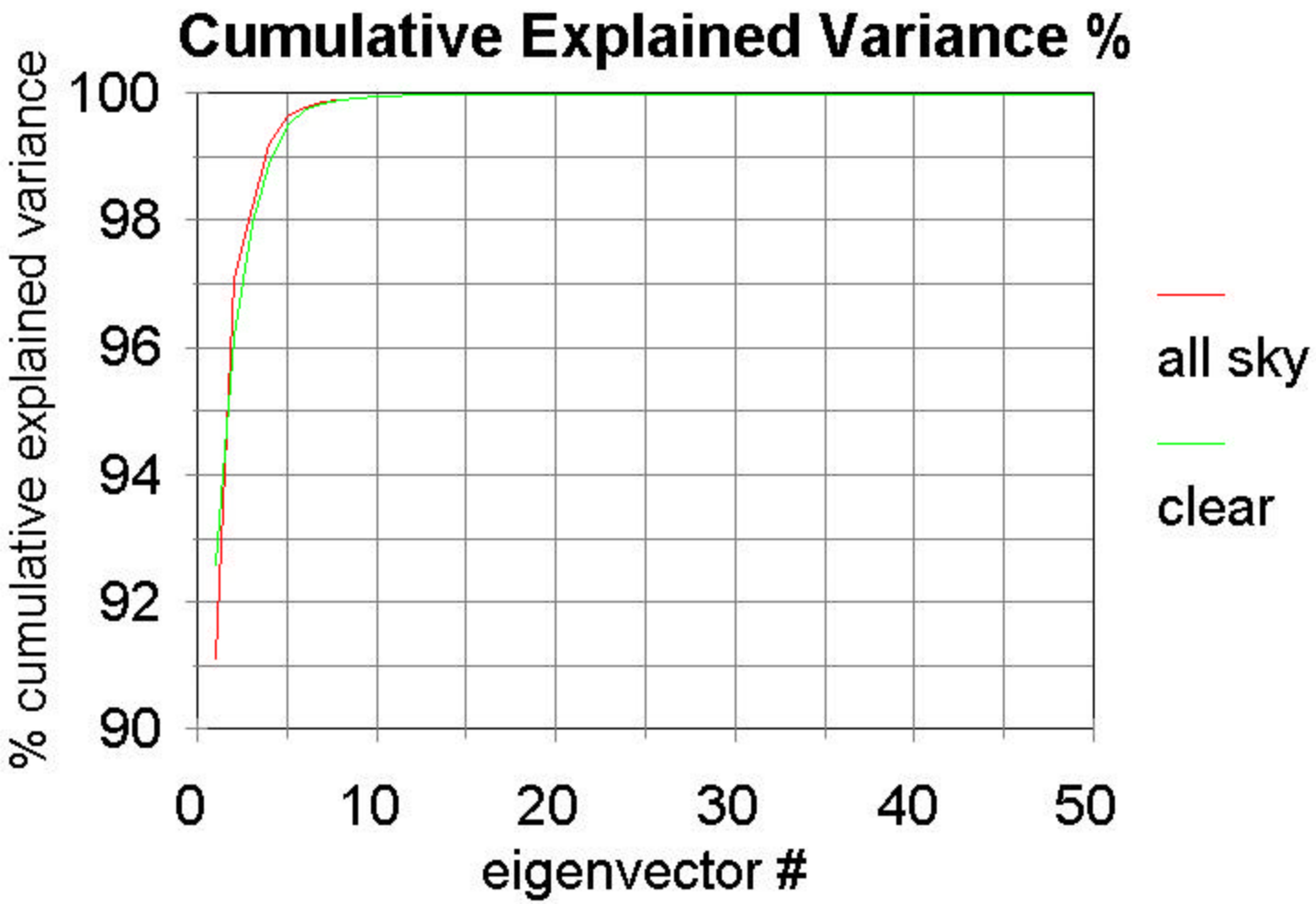
- On independent data – compute principal component scores.
- $P = E^T R$; elements of $R = (r_i - \overline{r_i}) / n_i$
- Invert equation and compute reconstructed radiances R^*.
- $R^* = E P$
- Reconstructed radiances are used for quality control.
- Reconstruction score $= [\, 1/N \sum_{i=1 \ldots N \text{ channels}} (R^*_i - R_i)^2 \,]^{1/2}$

AIRS applications

- AIRS is currently being simulated in near-real time.
- Each day a spatial subset of AIRS data is produced (200 mbytes vs 30 GB full data)
- Eigenvectors are generated daily.
- A static set of eigenvectors is used, but the ensemble is occasionally updated with new structures.
- When the ensemble is updated a new set of eigenvectors is also updated.

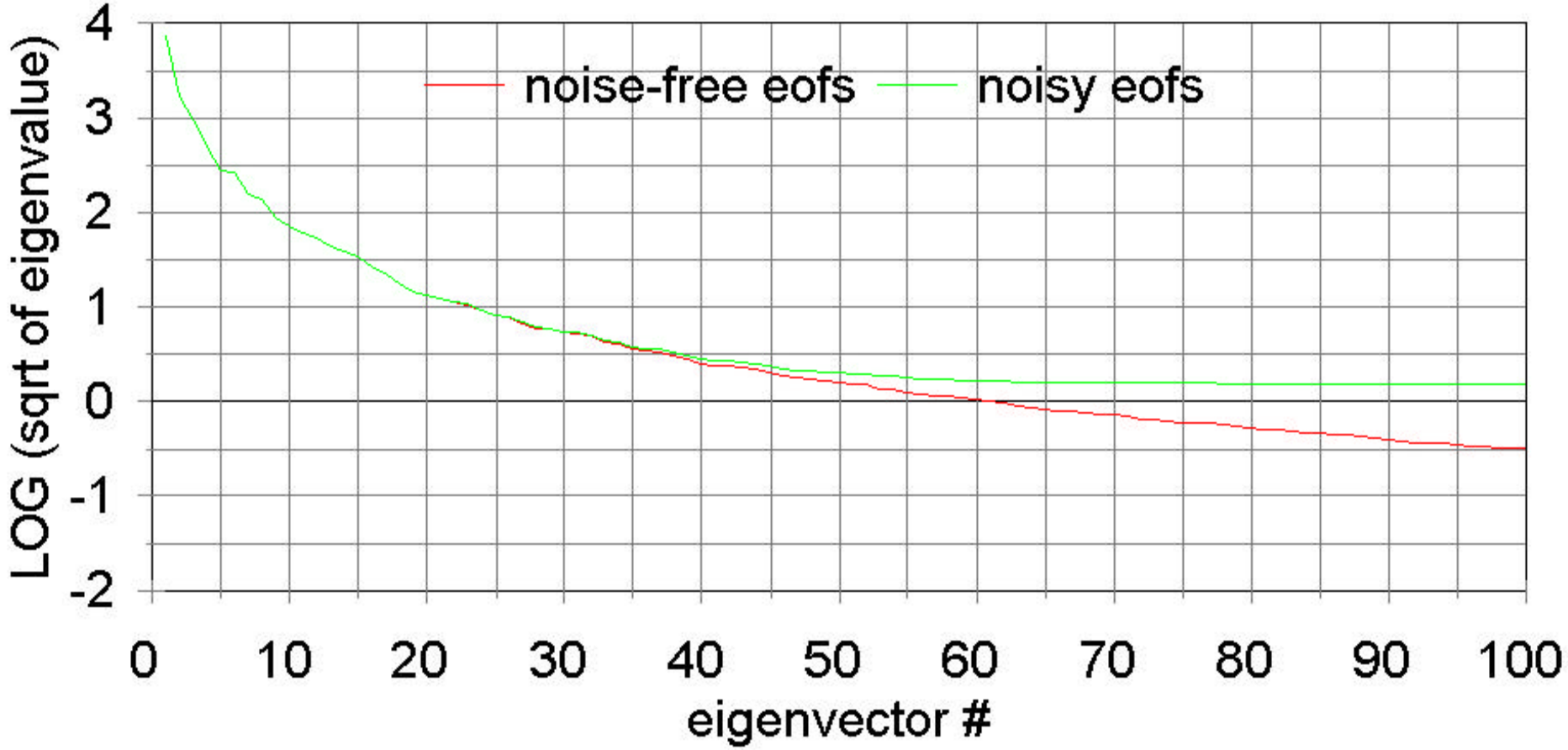
Locations used in generating eigenvectors




Cumulative Explained Variance %
% cumulative explained variance
100
98
96
94
92
90
0
10
20
30
40
50
eigenvector #
all sky
clear

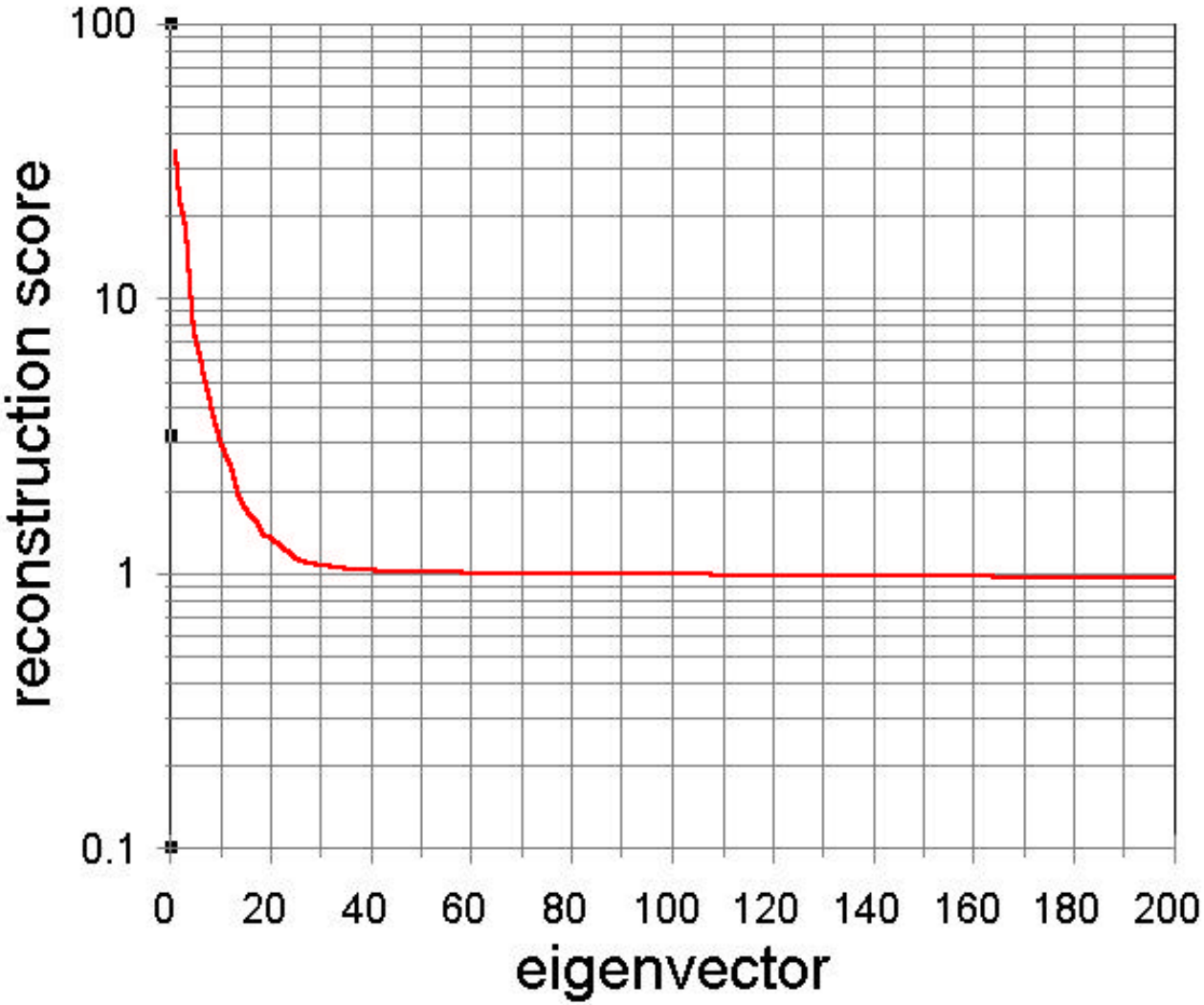
Square root of the eigenvalues

1	7497.60	19	14.68	37	3.38	55	1.25
2	1670.40	20	13.49	38	3.11	56	1.19
3	945.52	21	12.28	39	2.82	57	1.16
4	496.01	22	11.32	40	2.53	58	1.15
5	284.01	23	10.70	41	2.41	59	1.09
6	266.30	24	9.08	42	2.39	60	1.05
7	156.95	25	8.24	43	2.34	61	1.02
8	139.67	26	7.85	44	2.24	62	0.98
9	88.27	27	6.77	45	2.03	63	0.90
10	72.83	28	5.98	46	1.86	64	0.86
11	60.03	29	5.83	47	1.78	65	0.81
12	53.42	30	5.39	48	1.71	66	0.80
13	45.01	31	5.34	49	1.65	67	0.78
14	39.72	32	4.98	50	1.61	68	0.77
15	34.54	33	4.34	51	1.54	69	0.73
16	26.57	34	4.09	52	1.52	70	0.72
17	22.62	35	3.62	53	1.35	71	0.70
18	17.60	36	3.48	54	1.34	72	0.66


noise-free eofs
noisy eofs
LOG (sqrt of eigenvalue)
4
3
2
1
0
-1
-2
0
10
20
30
40
50
60
70
80
90
100
eigenvector #

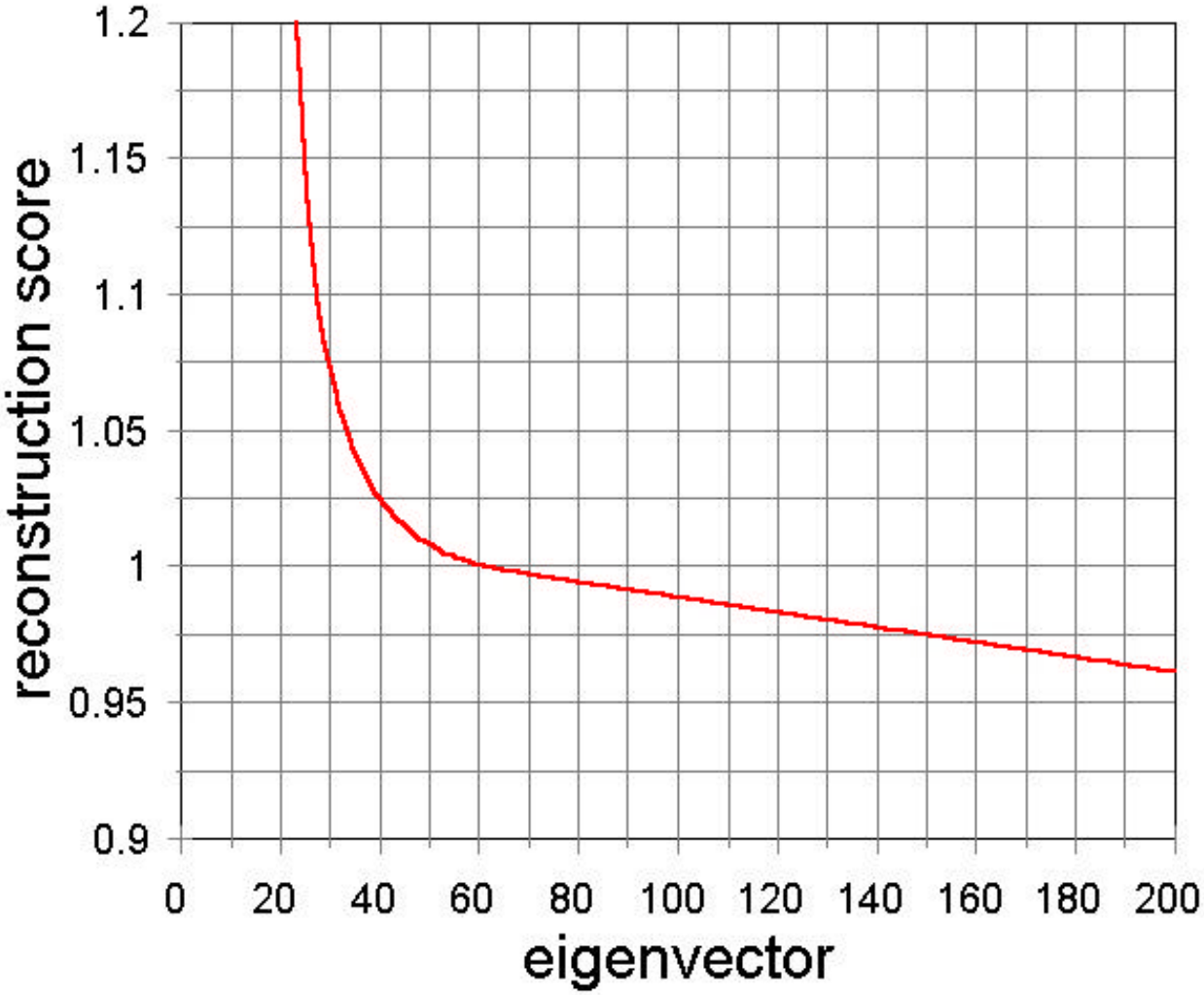
- Reconstruction score $= \left[1/N \sum (R^*_i - R_i)^2 \right]^{1/2}$

  $i = 1 \ldots N$ channels

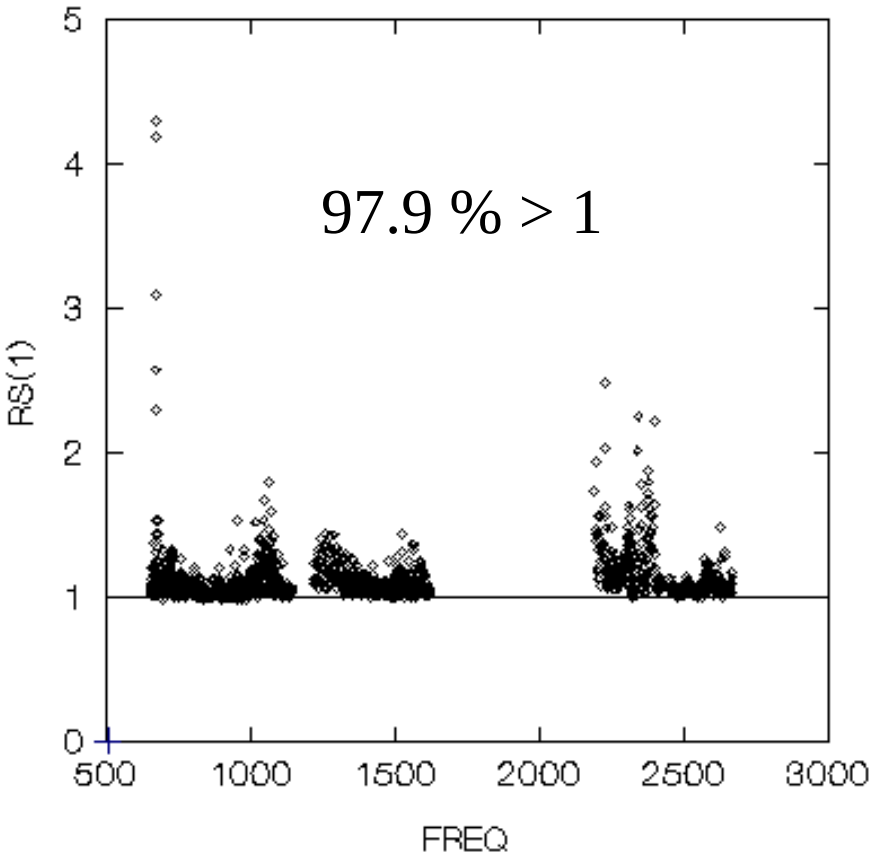


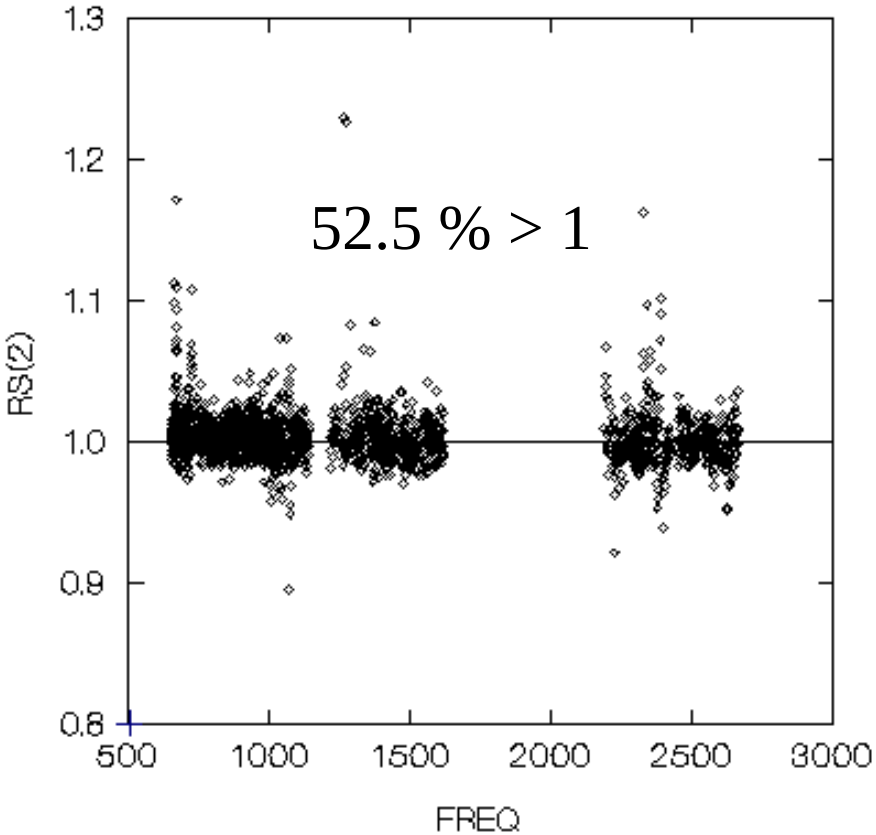
- Reconstruction score $= \left[1/N \sum (R^*_i - R_i)^2 \right]^{1/2}$

  $i = 1 \ldots N$ channels

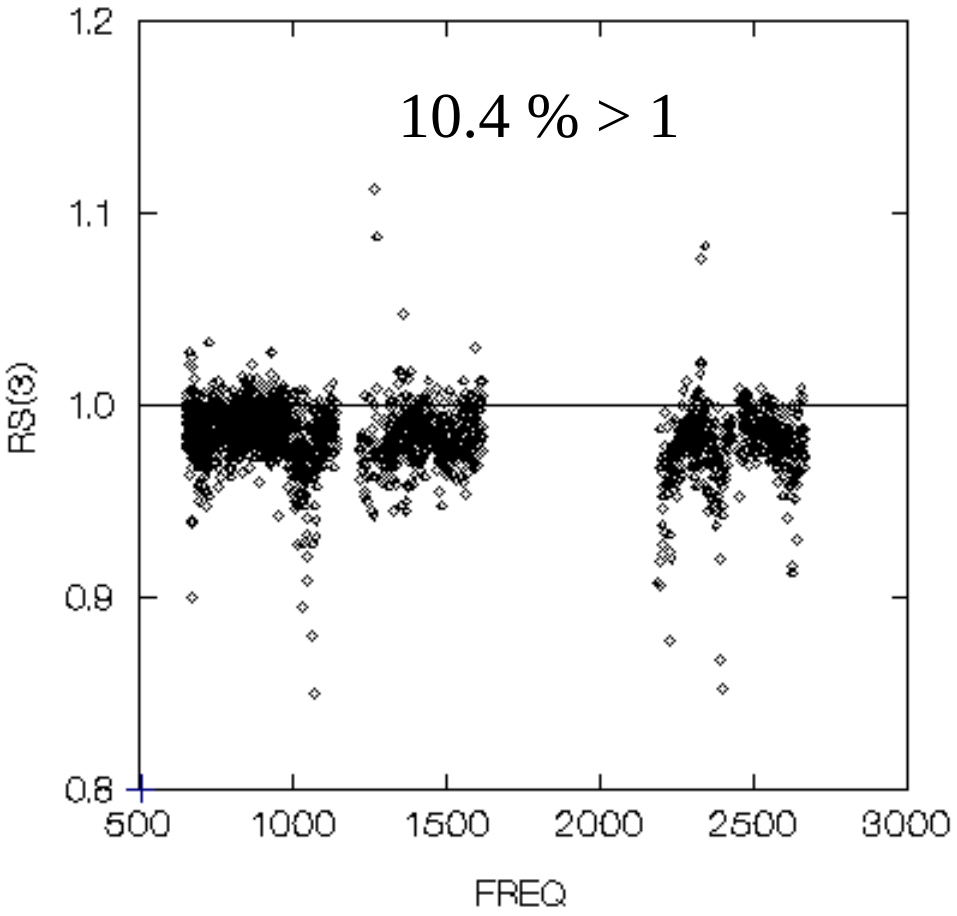


Channel reconstruction scores for 25 and 50 pcs

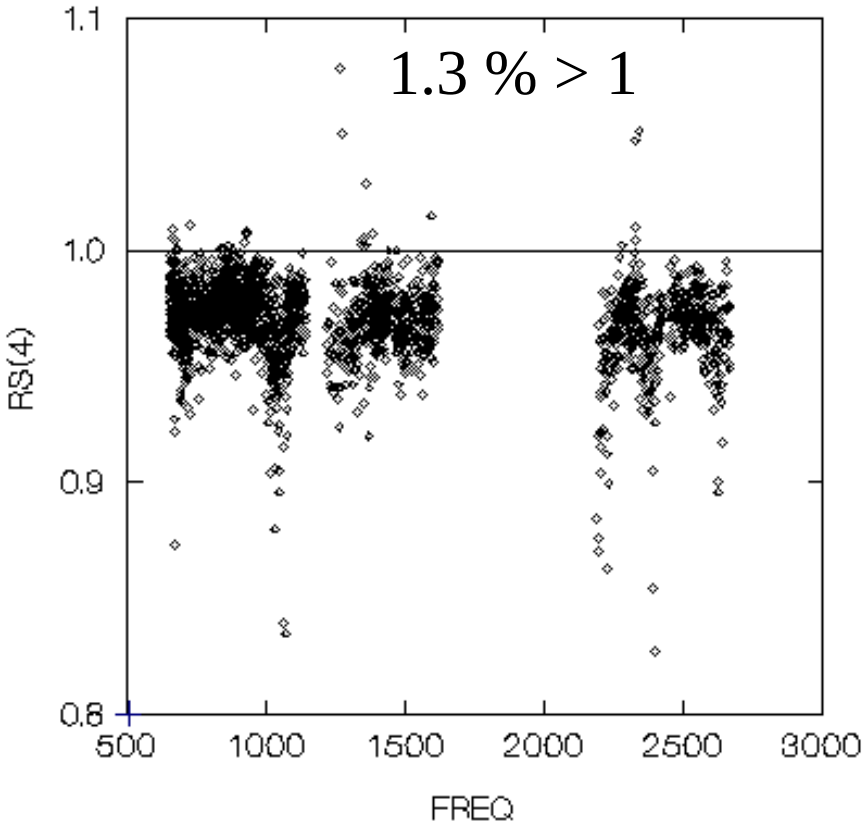


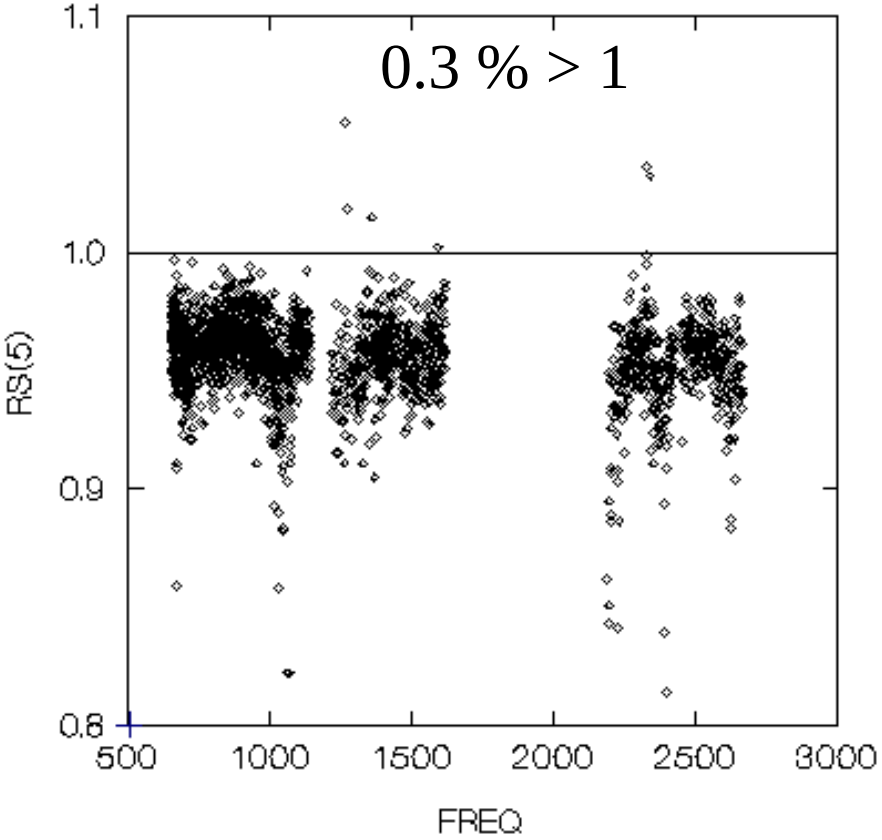


Channel reconstruction scores for 100 pcs



Channel reconstruction scores for 150 and 200 pcs



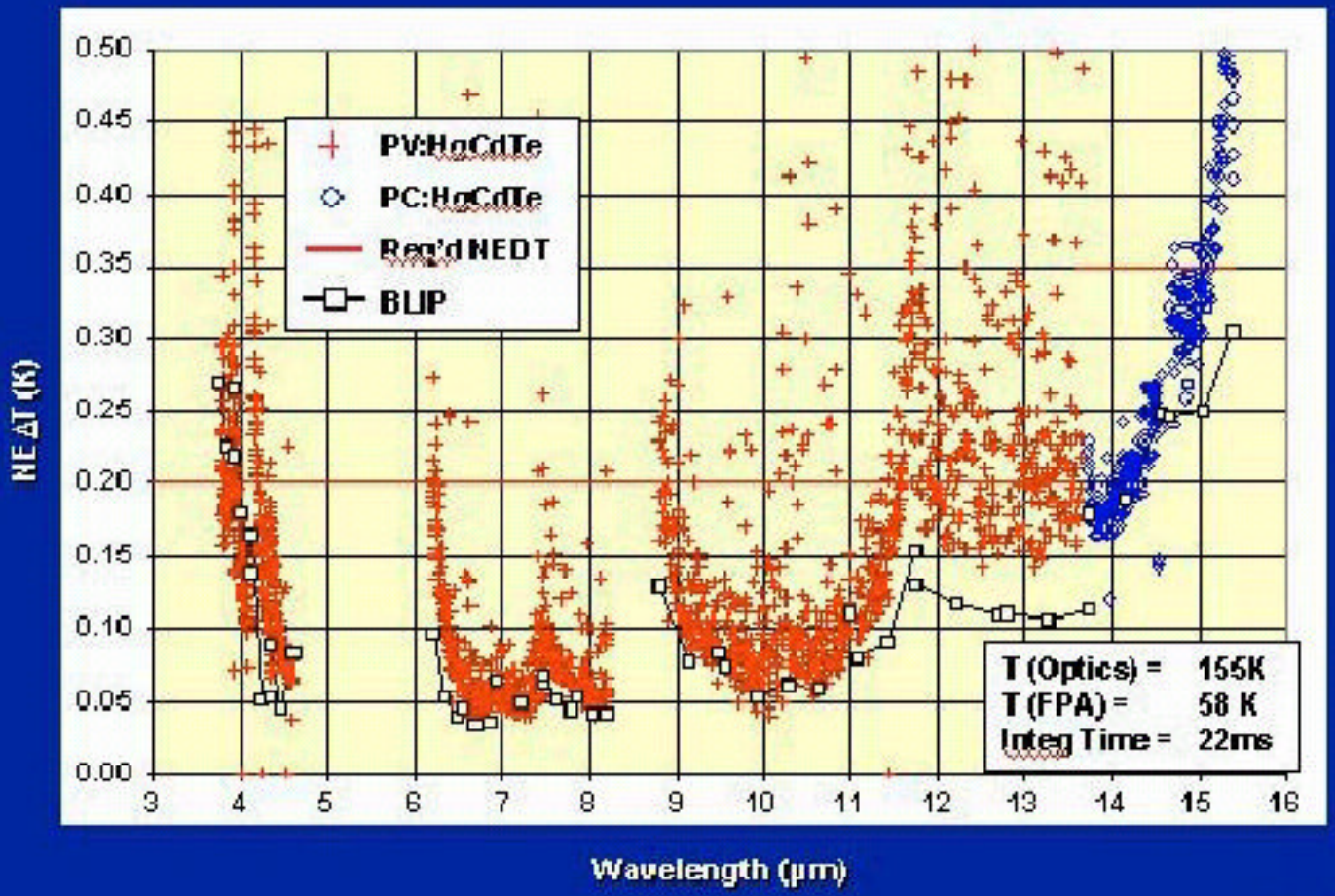


Noise filter

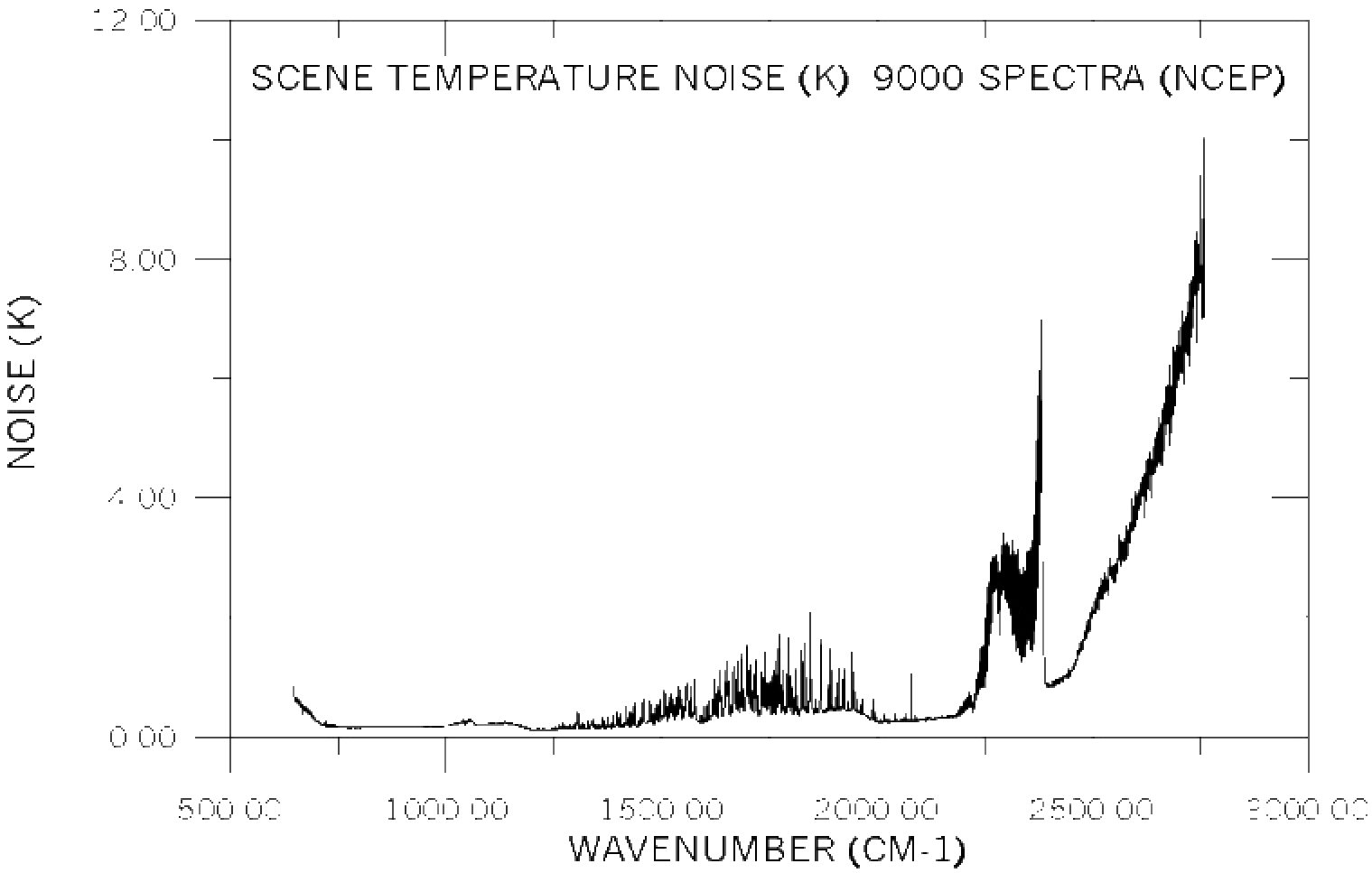
- Truncating eigenvectors also results in noise filtering.
- AIRS has very low noise -- so reconstructing the noise is not an issue.
- But for other sounders, noise filtering may be very desirable.

Measured Sensitivity (NEΔT)

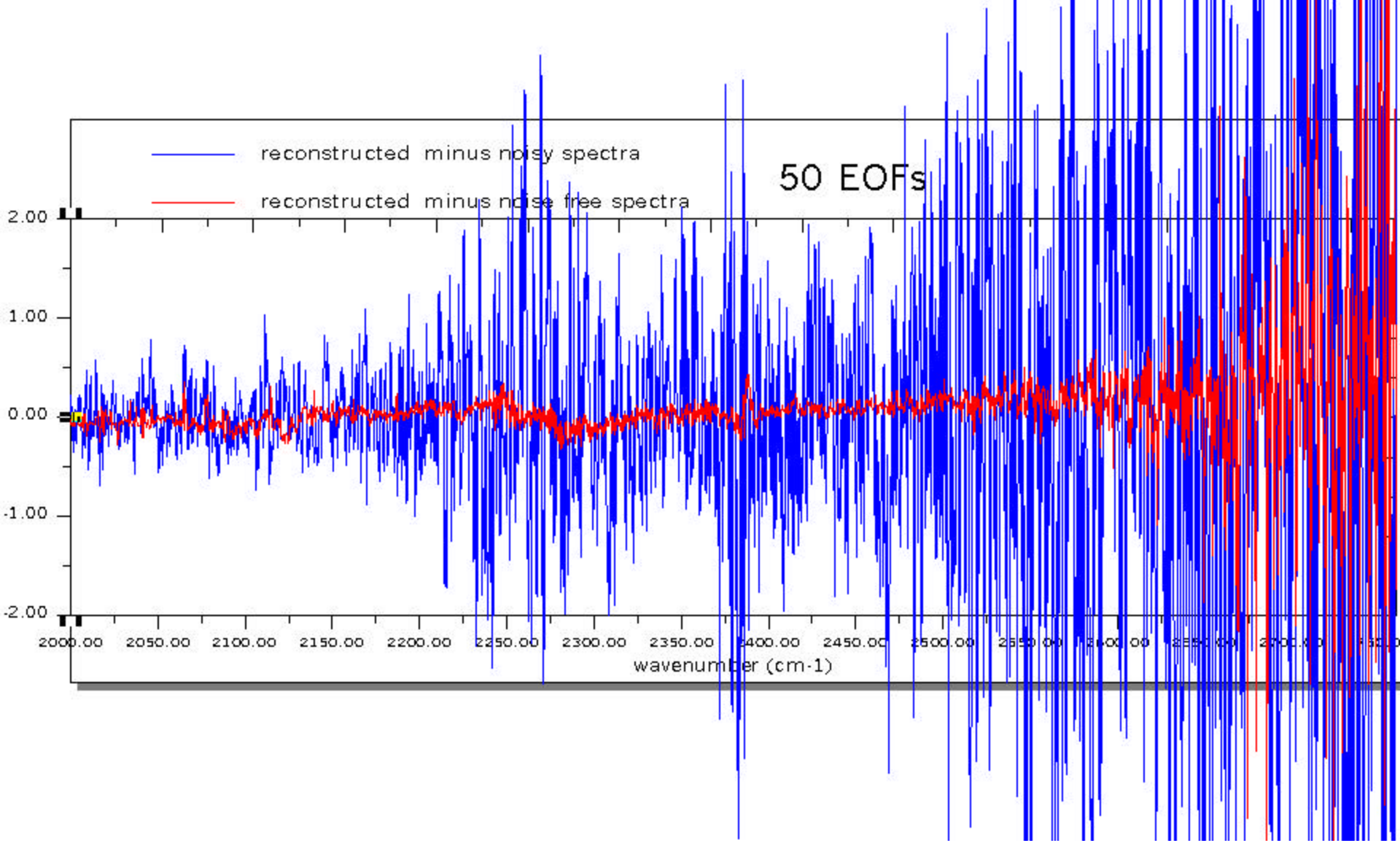
Single Look (1.1° x 0.6° IFOV) - 250K Scene



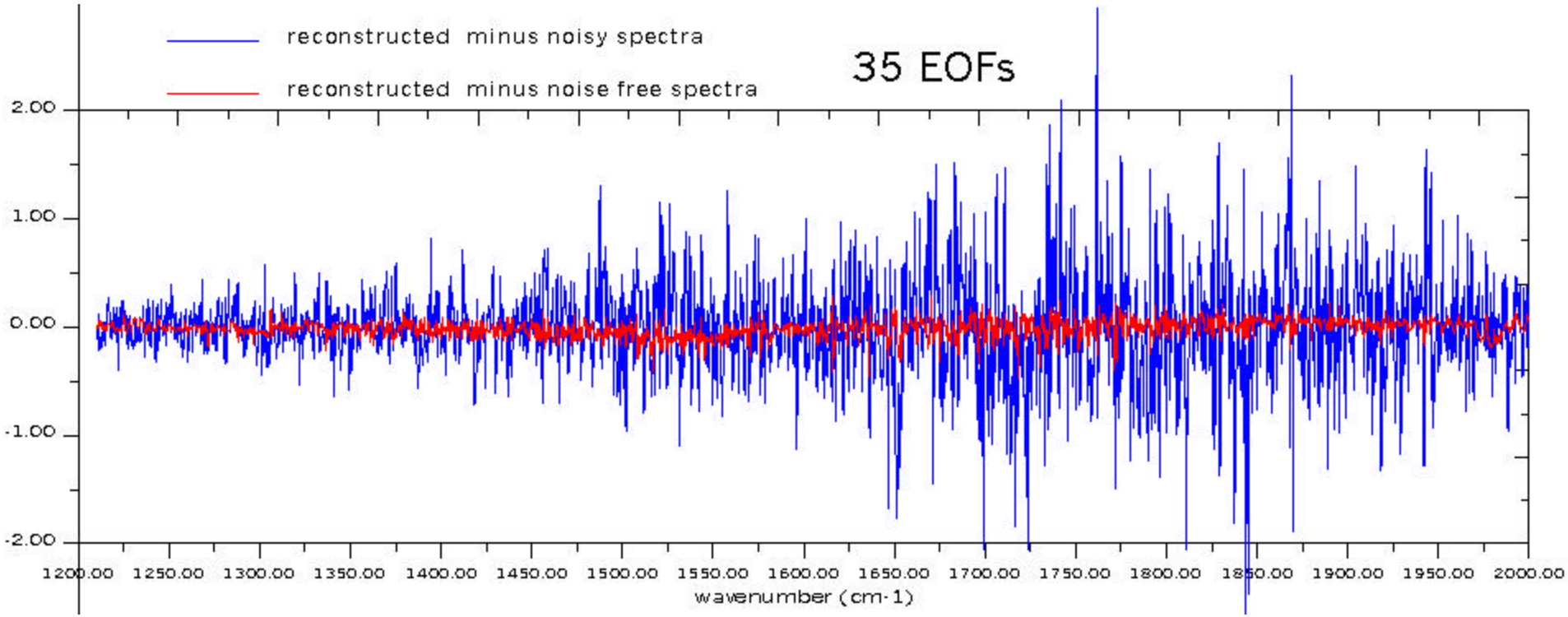
Noise filtering IASI



BAND 3 RECONSTRUCTION- SINGLE CASE



BAND 2 RECONSTRUCTION- SINGLE CASE

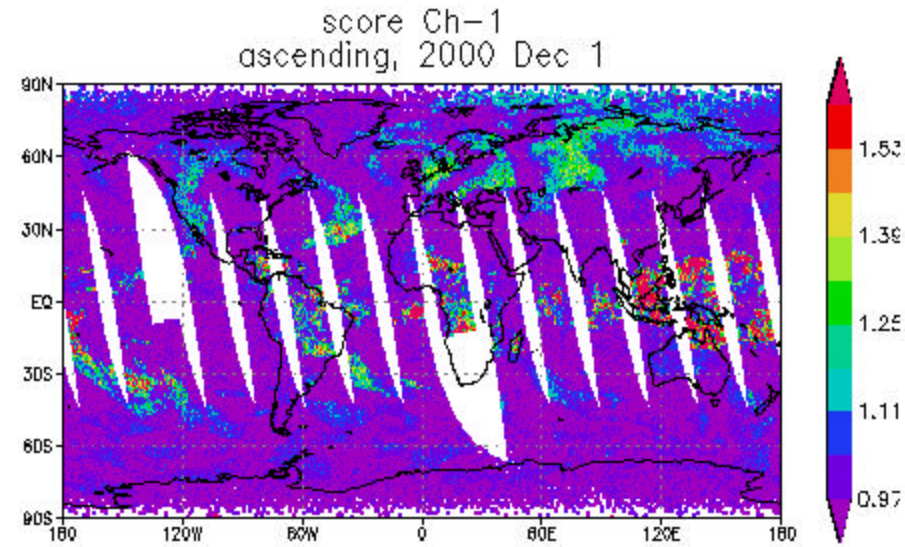


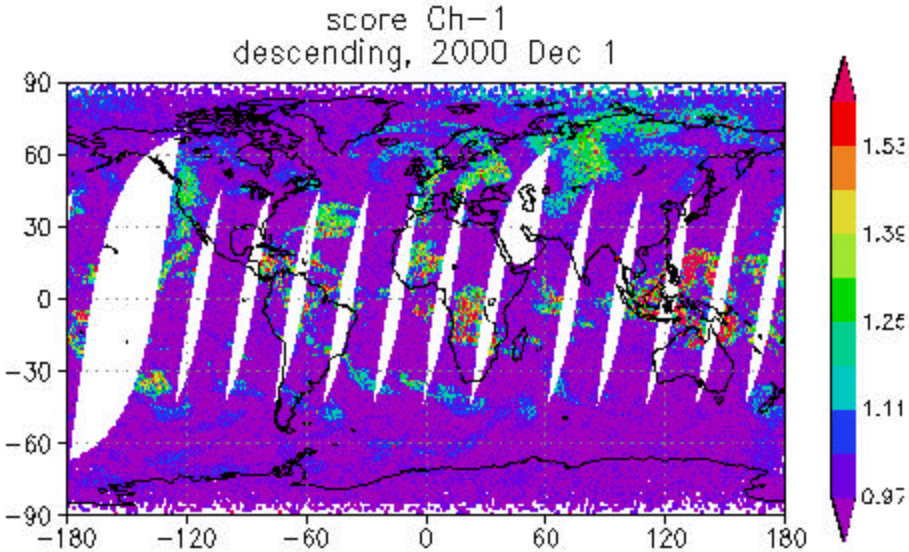
Monitoring Eigenvectors

- Monitoring eigenvectors is critical
- Eigenvectors may need to be updated due to new structures that were not in the original ensemble

AIRS Near-Real-Time Simulations

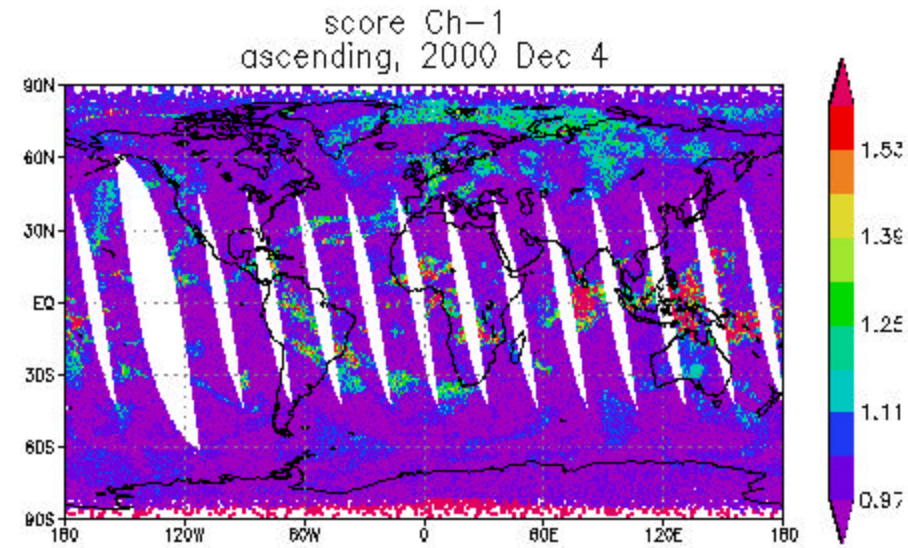
Thinned Radiance Data
EOF Scores
Radiance File
Rad. vs. Freq.
Level2 regression retrieval
At 100 Levels
At 25 Layers
Profiles
Error Estimate
At 100 Levels
At 25 Layers
Level2 Truth
At 100 Levels
At 25 Layers
Related Info.
Background
AIRS at JPL
Other Links

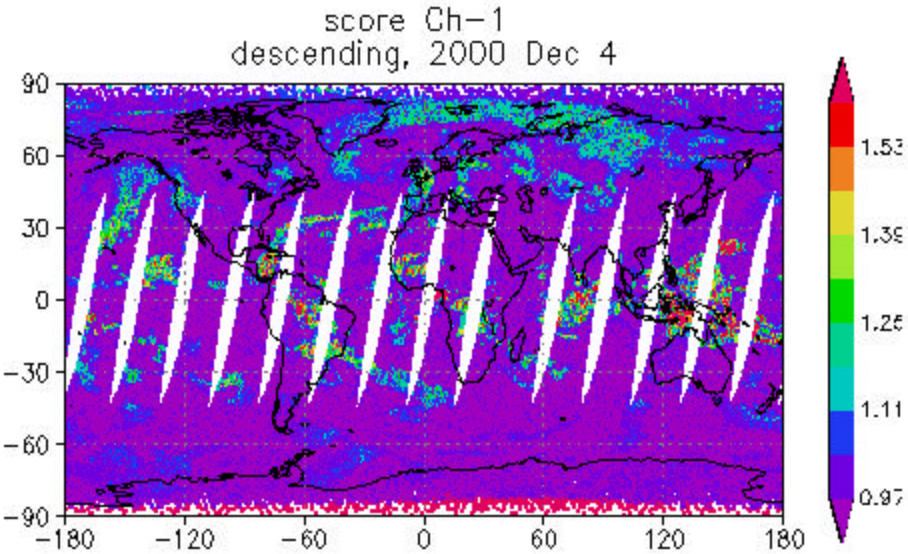




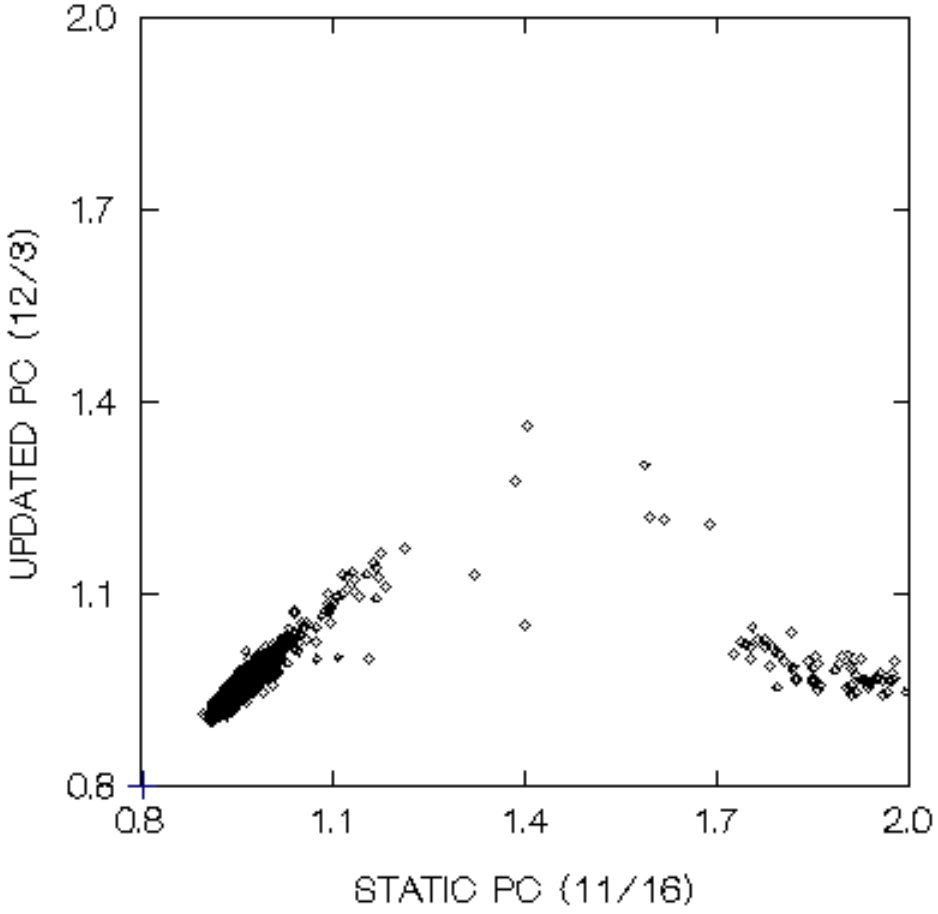
AIRS Near-Real-Time Simulations

Thinned Radiance Data
EOF Scores
Radiance File
Rad. vs. Freq.
Level2 regression retrieval
At 100 Levels
At 25 Layers
Profiles
Error Estimate
At 100 Levels
At 25 Layers
Level2 Truth
At 100 Levels
At 25 Layers
Related Info.
Background
AIRS at JPL
Other Links

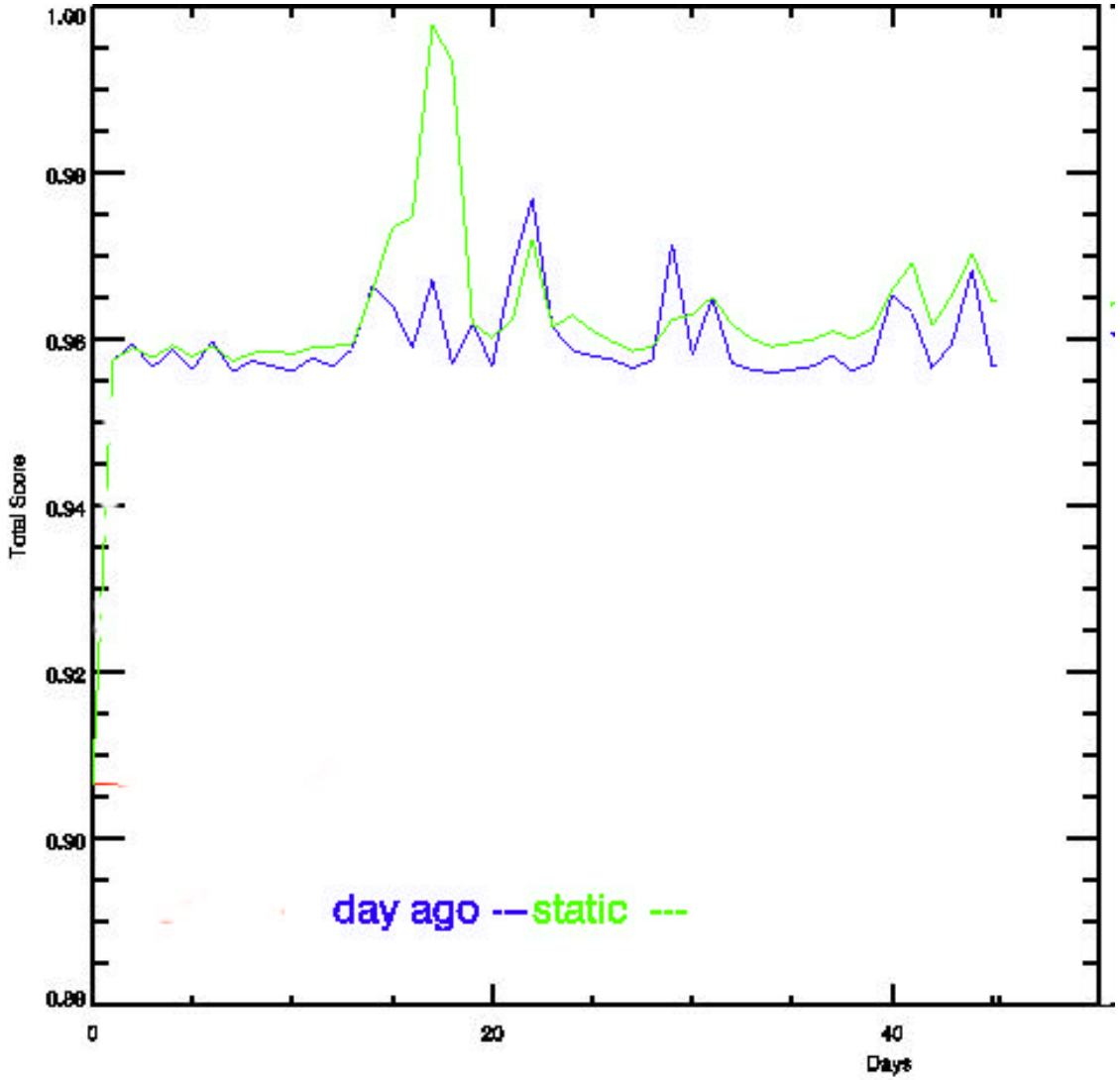




12/4/00 reconstruction scores

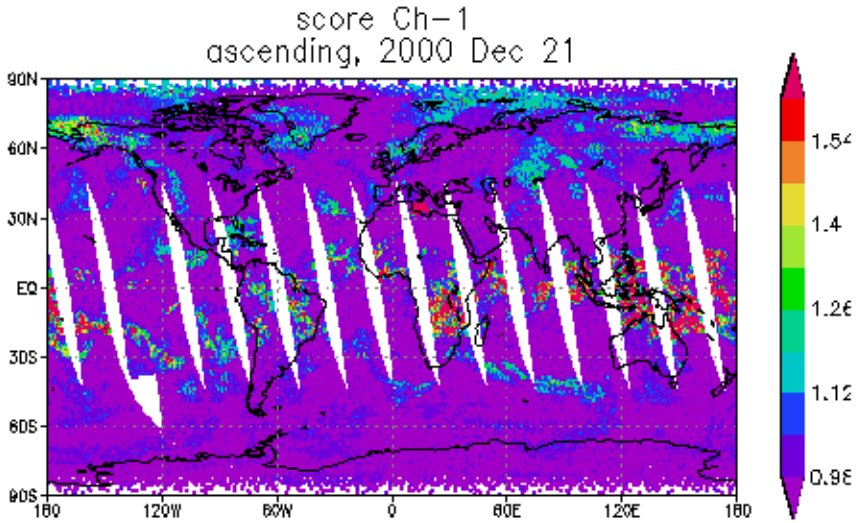


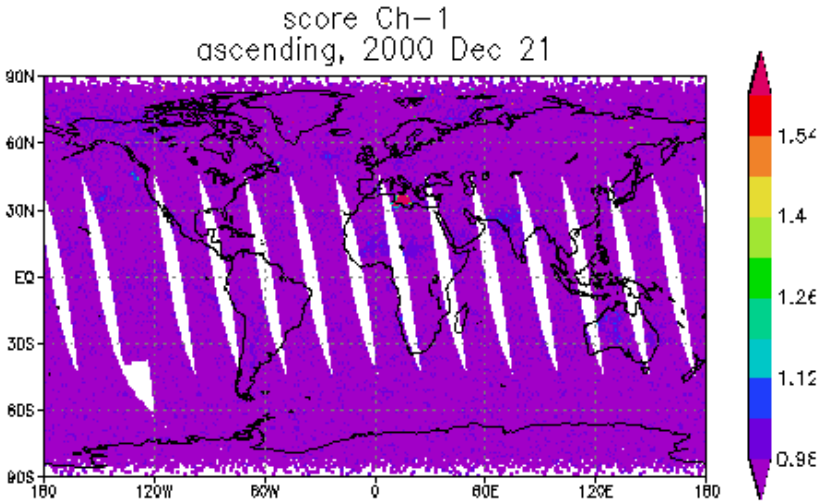
Monitoring reconstruction score is important

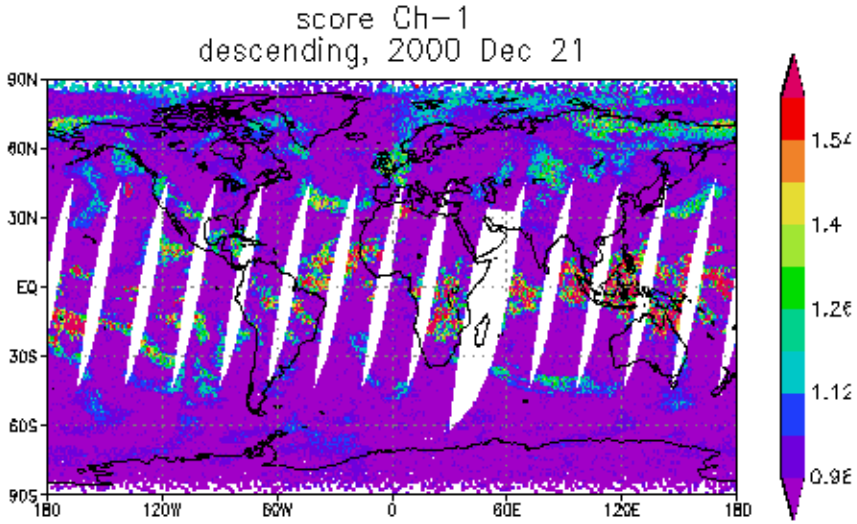


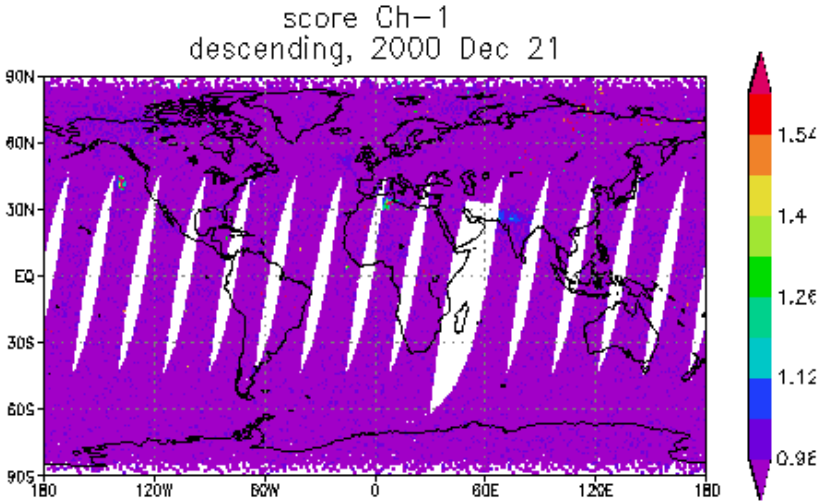
Clear eigenvectors

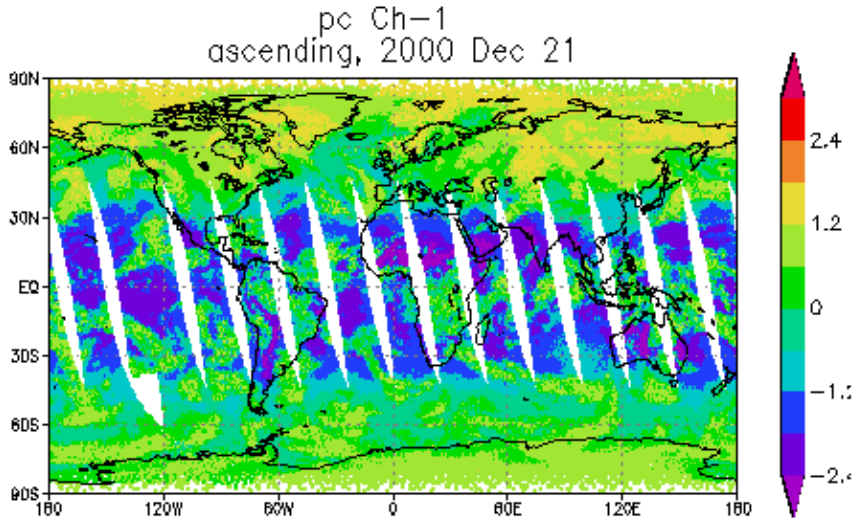
all-sky eigenvectors

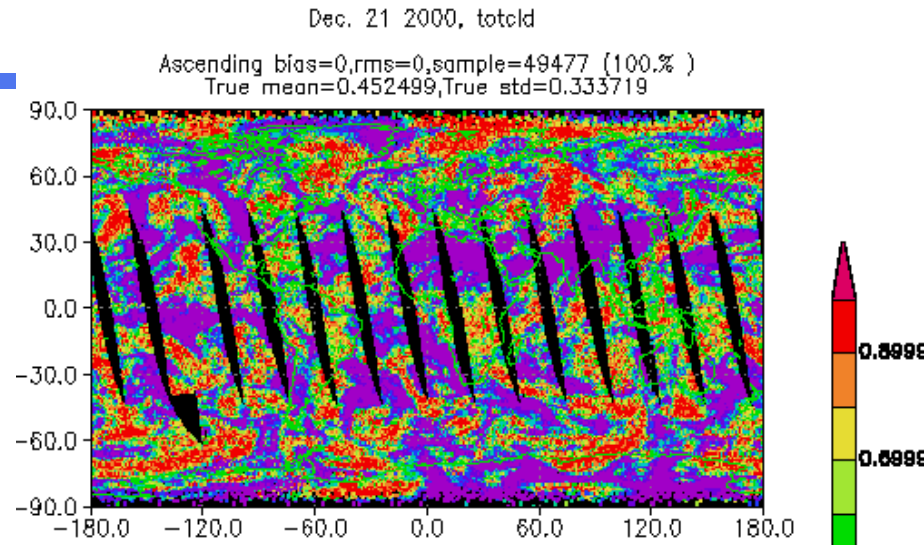


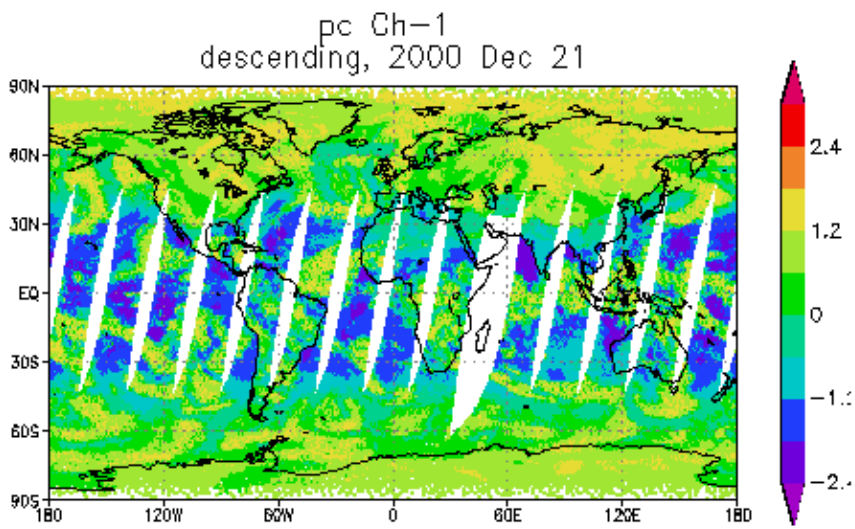


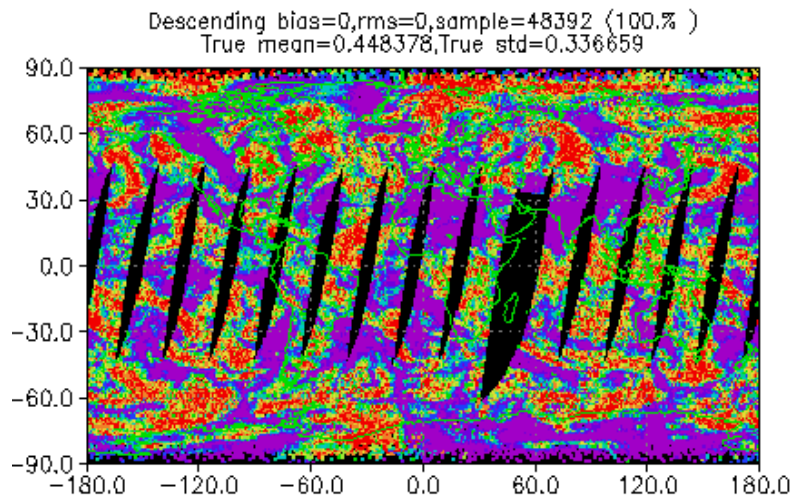


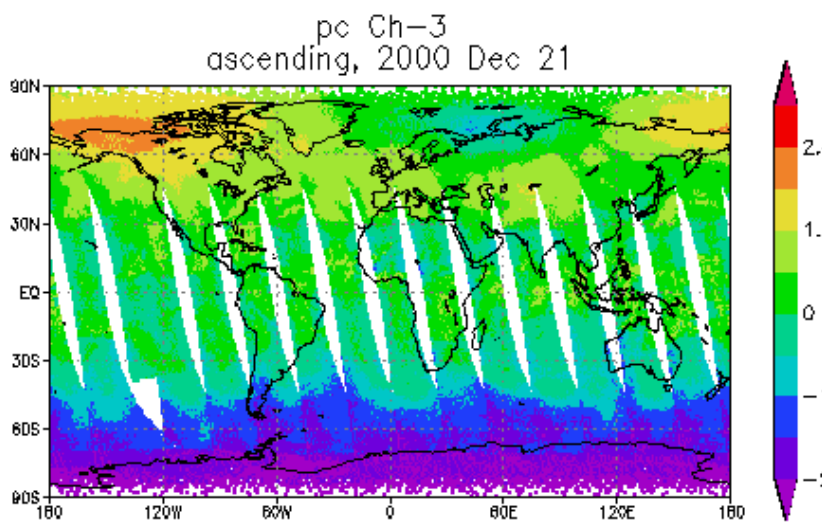


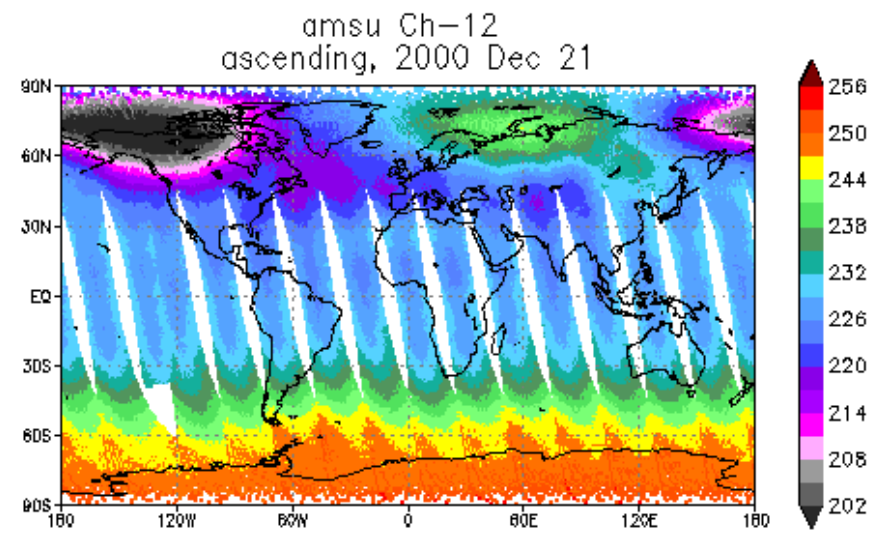

pc Ch−1
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0
−1.2
−2.4

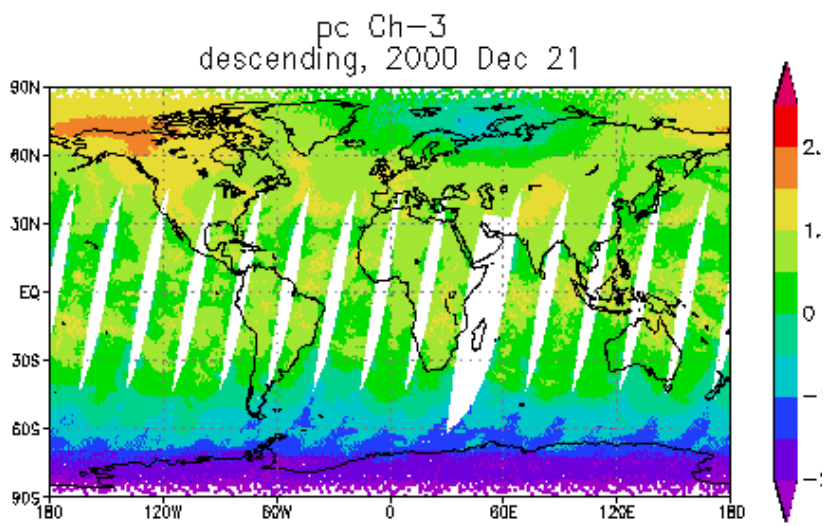

Dec. 21 2000, totcld
Ascending bias=0,rms=0,sample=49477 (100.%)
True mean=0.452499,True std=0.333719
90.0
60.0
30.0
0.0
−30.0
−60.0
−90.0
−180.0
−120.0
−60.0
0.0
60.0
120.0
180.0
0.8999
0.6999
0.4999
0.2999
0.0999

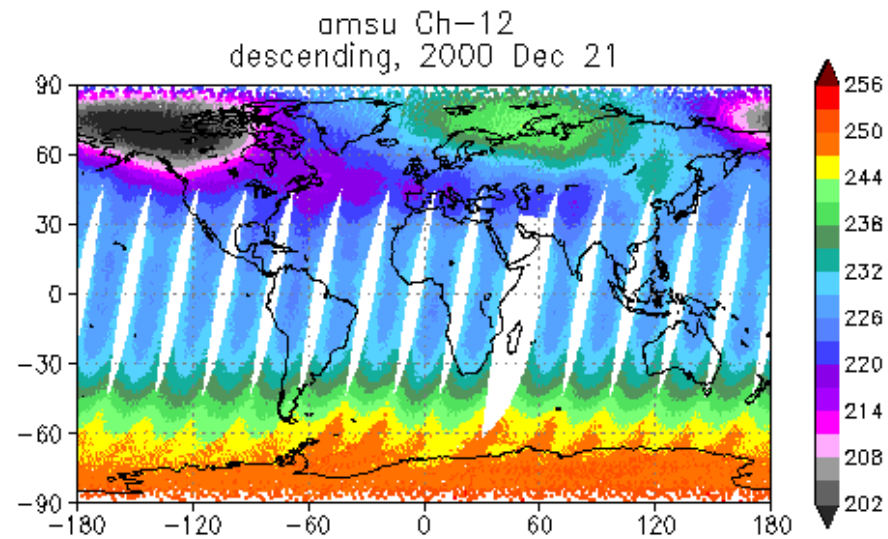

pc Ch−1
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0
−1.2
−2.4

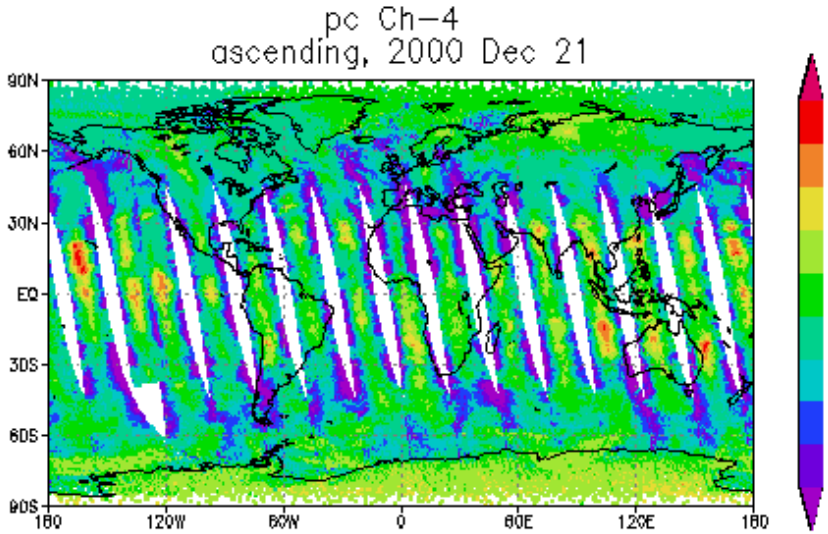

Descending bias=0,rms=0,sample=48392 (100.%)
True mean=0.448378,True std=0.336659
90.0
60.0
30.0
0.0
−30.0
−60.0
−90.0
−180.0
−120.0
−60.0
0.0
60.0
120.0
180.0

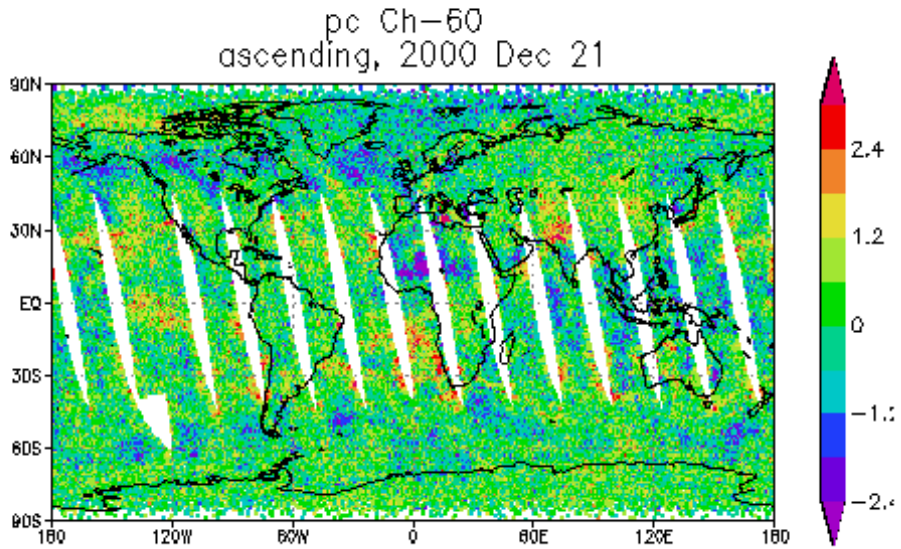

pc Ch-3
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.
1.
0

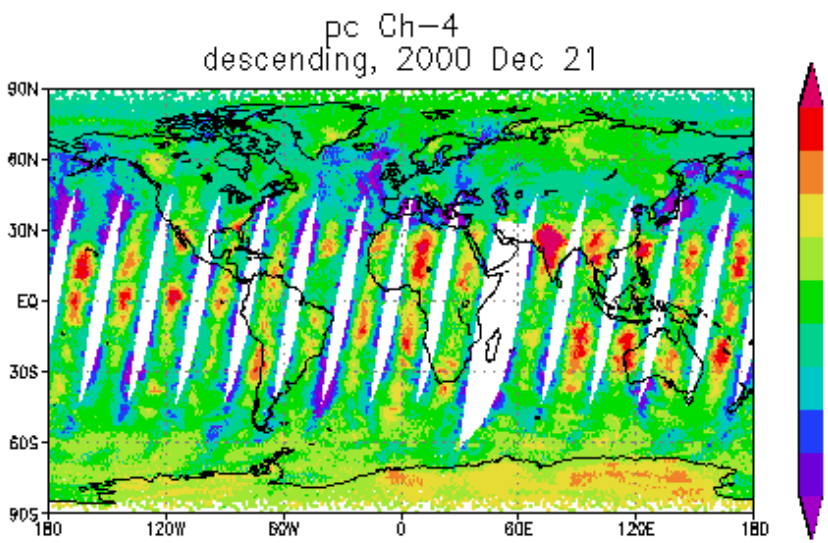

amsu Ch-12
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
256
250
244
238
232
226
220
214
208
202

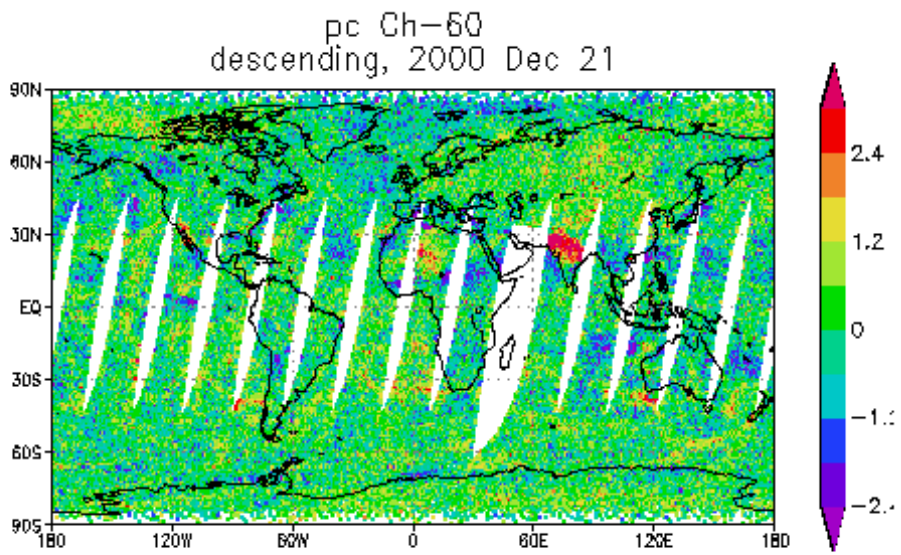

pc Ch-3
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.
1.
0

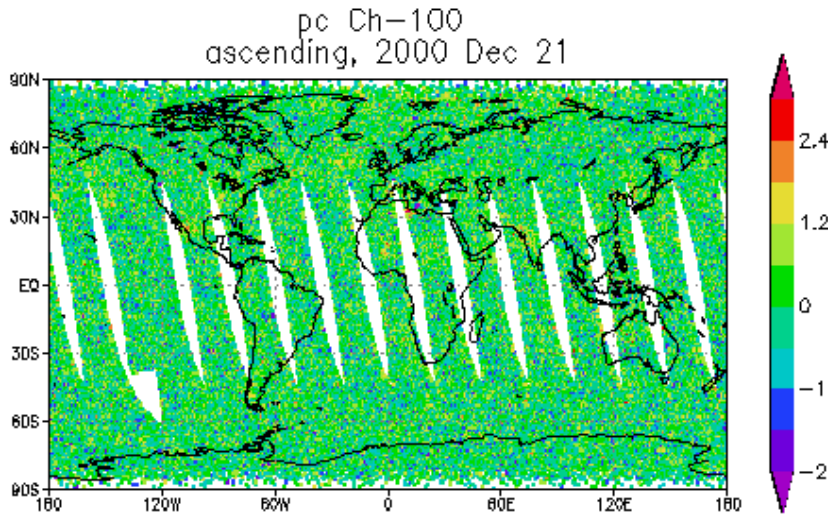

amsu Ch-12
descending, 2000 Dec 21
90
60
30
0
-30
-60
-90
-180
-120
-60
0
60
120
180
256
250
244
238
232
226
220
214
208
202

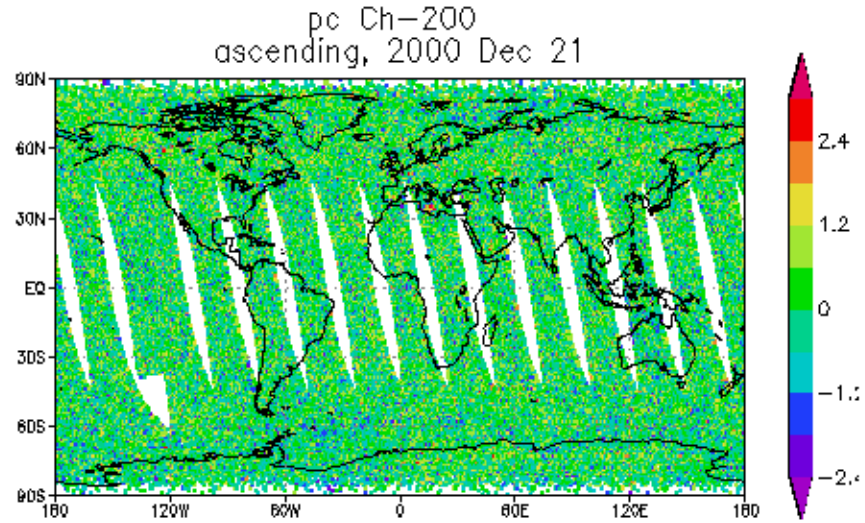

pc Ch−4
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180

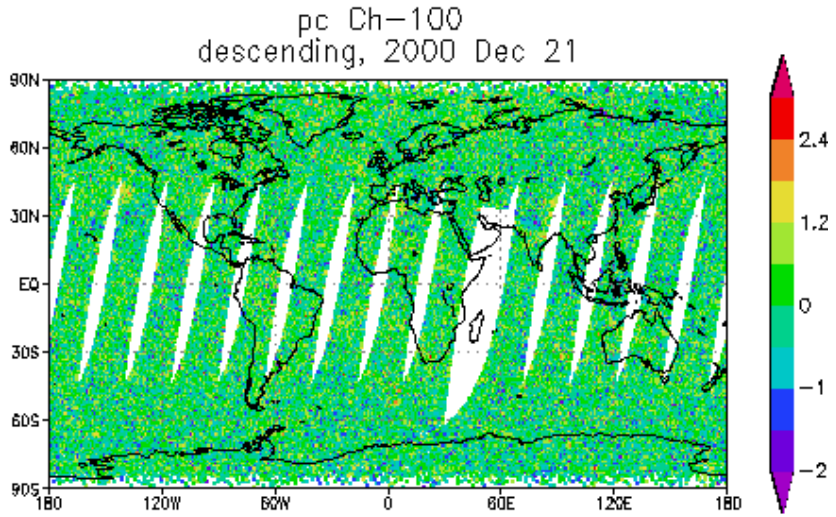

pc Ch−60
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0

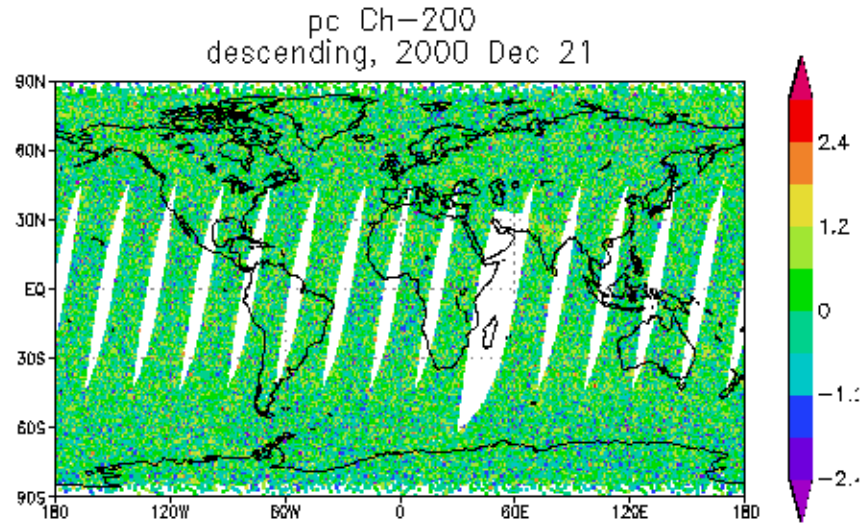

pc Ch−4
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180


pc Ch−60
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0

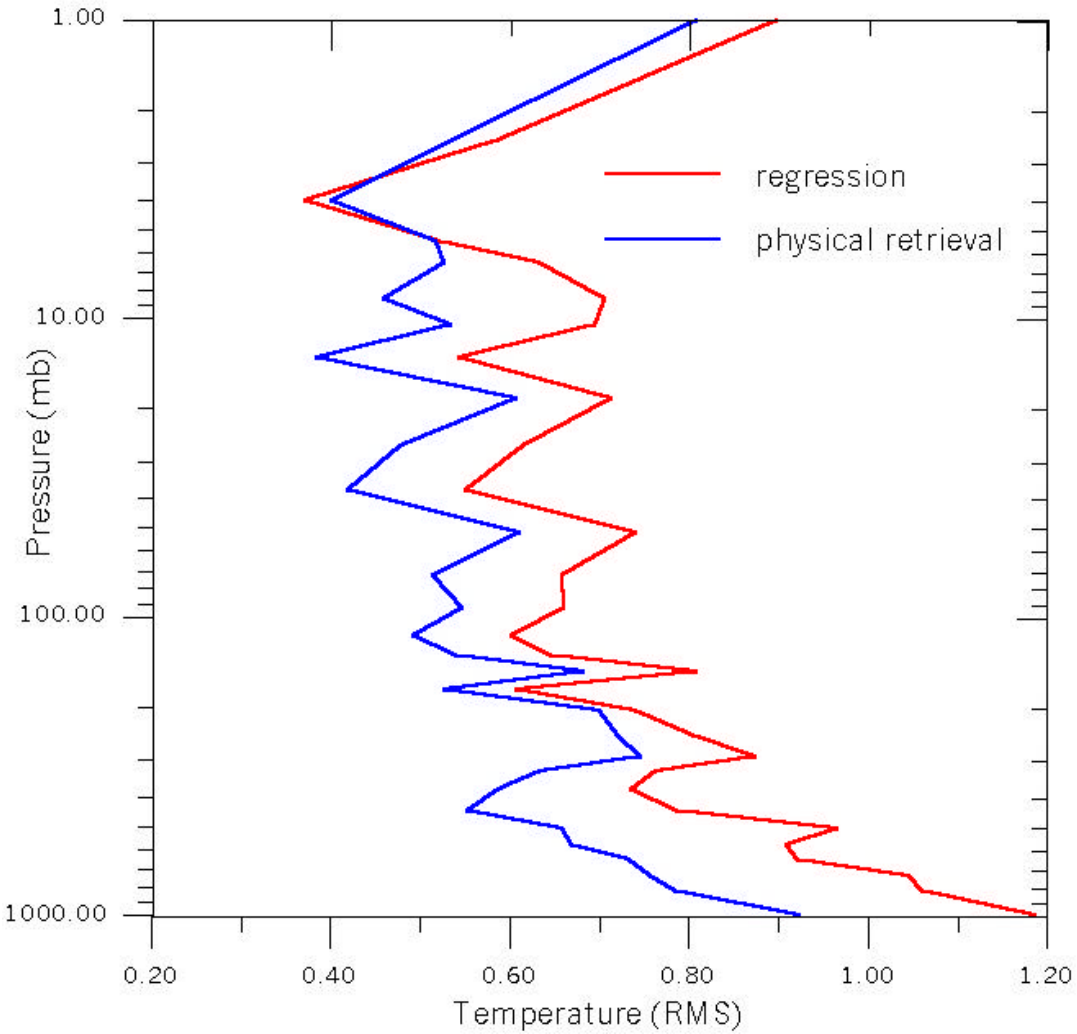

pc Ch−100
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0
−1
−2


pc Ch−200
ascending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0


pc Ch−100
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0
−1
−2


pc Ch−200
descending, 2000 Dec 21
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.4
1.2
0

PC Regression retrievals



Compression Factors

- 200 pc scores ~ 9.5 % (lossy compression).
- For near lossless compression. Store 100 pc scores plus the residual. The residual can be stored as 1 byte as long as residual is +- 1.28 K.
- Not all channels will be within +- 1.28 K, but no more than 20 channels per fov.
- Full data set 2100 x 4 bytes = 9400
- Compression 100 pc x 4 bytes + 2100 ch x 1 byte + 20 outliers x 4 = 2580 ~ 27 %
- Bit trimming can provide even greater lossless data compression (perhaps 10%).

Summary

- All sky eigenvectors
- Monitor reconstruction scores
- Update eigenvectors periodically during first 6 months.
- Provide to NWP centers 200 pc scores (normalized by sqrt(eigenvalues)).
- Provide subset of individual channels.
- Eigenvector regression is quite accurate from the middle troposphere to upper stratosphere.
- Need to study noise filtering feature of eofs.