Atmospheric Data Assimilation at NCEP

John C. Derber
Environmental Modeling Center
NCEP/NWS/NOAA
With input from:
Daryl Kleist, Russ Treadon, Wan-shu Wu, David Parrish, Mike Lueken, Xu Li, Xiujuan Su, Min-Jeong Kim, Banhua Yan, Paul VanDelst, Lidia Curcurull, Manuel Pondeca, Mingjing Tong, David Groff, Dennis Keyser, George Gayno, Jim Purser, Jim Jung,Moidrag Rancic, Shun Liu, Yanqiu Zhu, Yoshiaki Sato

GSI analysis

- Grid-point Statistical Interpolation (GSI)
- Unified analysis system for all NCEP atmospheric applications
  - Global
  - Regional (NAM)
  - RTMA
  - GMAO
  - RR (GSD)
  - AFWA
- Developed for operational application

Operational context

- Forecasts must be complete within schedule
  - Trade-offs
    - More accurate formulation – higher resolution
    - Improved model – improved analysis
    - Enhanced physics – higher resolution
    - Etc.
- Must work everywhere – all the time
- Manual intervention should be minimal
- Both operational and research data used in systems

Operational context

- Over 1.43B observations received per day (most satellite data – Many more if you include Doppler radar data not currently used.
- Over 7M observations per day used.
- Current global operational model (T382L64) degrees of freedom – 47M
- Test global model (T878L91) degrees of freedom – 353M

Atmospheric analysis problem (theoretical)

$$J = J_b + J_o + J_c$$

$$J = (x-x_b)^T B_x^{-1}(x-x_b) + (K(x)-O)^T(E+F)^{-1}(K(x)-O) + J_C$$

J =background term+ observation term + constraint term

x	= Analysis
x_b	= Background
B_x	= Background error covariance
K	= Forward model (nonlinear)
O	= Observations
E+F	= R = Instrument error + Representativeness error
J_C	= Constraint term

Basic Assumptions (violated)

- Data (forecast and most observations) are unbiased
  - Radiosonde and others commonly biased
  - Satellite observations biased - but can be corrected?
  - All forecast models have significant biases.
- Observational errors normally distributed
  - Gross errors can make the observational errors non-normally distributed
  - Moisture errors not normally distributed because moisture cannot be < 0 or >> saturation.
- Background error uncorrelated to observational errors
  - May be true if not using retrievals
  - Representativeness error likely correlated

Solution Algorithm

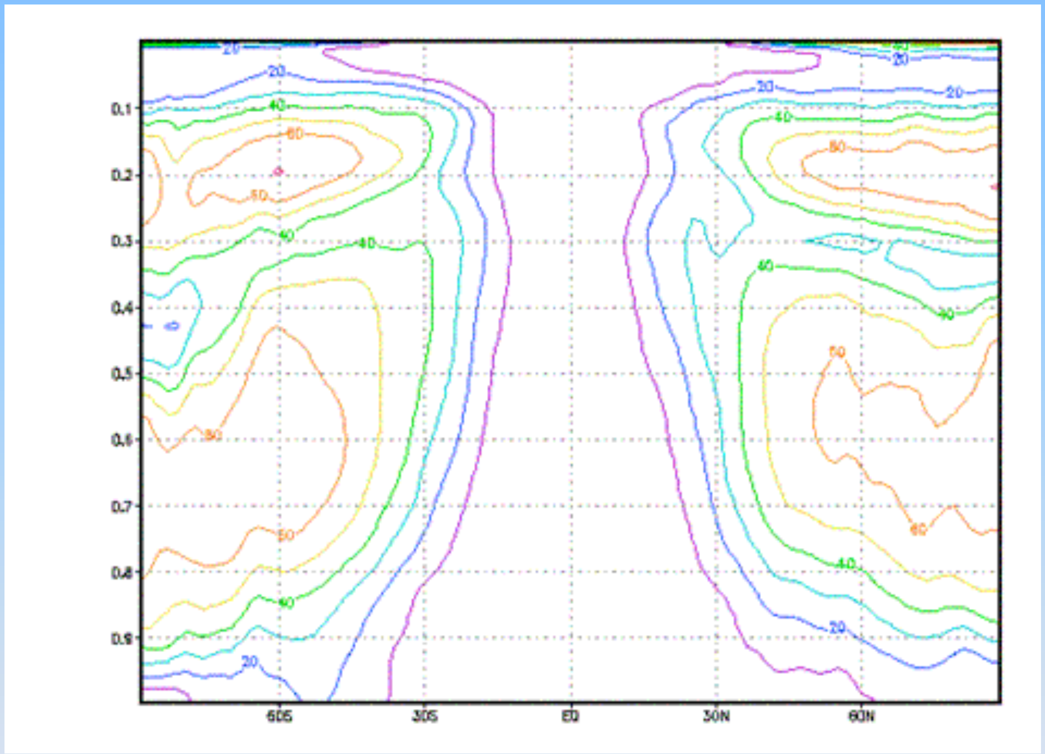
- Outer iteration, inner iteration structure
- Not guaranteed to converge to the correct solution
- Outer iteration
  - QC
  - More complete forward model
- Inner iteration
  - Preconditioned conjugate gradient
  - Often simpler forward model
  - Variational QC
  - Possibly lower resolution
  - Solution used to start next outer iteration

Analysis variables

- Streamfunction (Ψ)
- Unbalanced Velocity Potential (χ_u)
  - $\chi_B = C\,\Psi$
- Unbalanced Temperature (T_u)
  - $T_B = G\,\Psi$
- Unbalanced Surface Pressure (Ps_u)
  - $Ps_B = W\,\Psi$
- Normalized q
  - With (Holm et al.) or without (T,Ps) relationship
- Ozone
- CLW (only changed slightly)

Multivariate Variable Definition

$$T_b = \mathbf{G}\psi \quad ; \quad \chi_b = \mathbf{c}\psi \quad ; \quad Ps_b = \mathbf{W}\psi$$



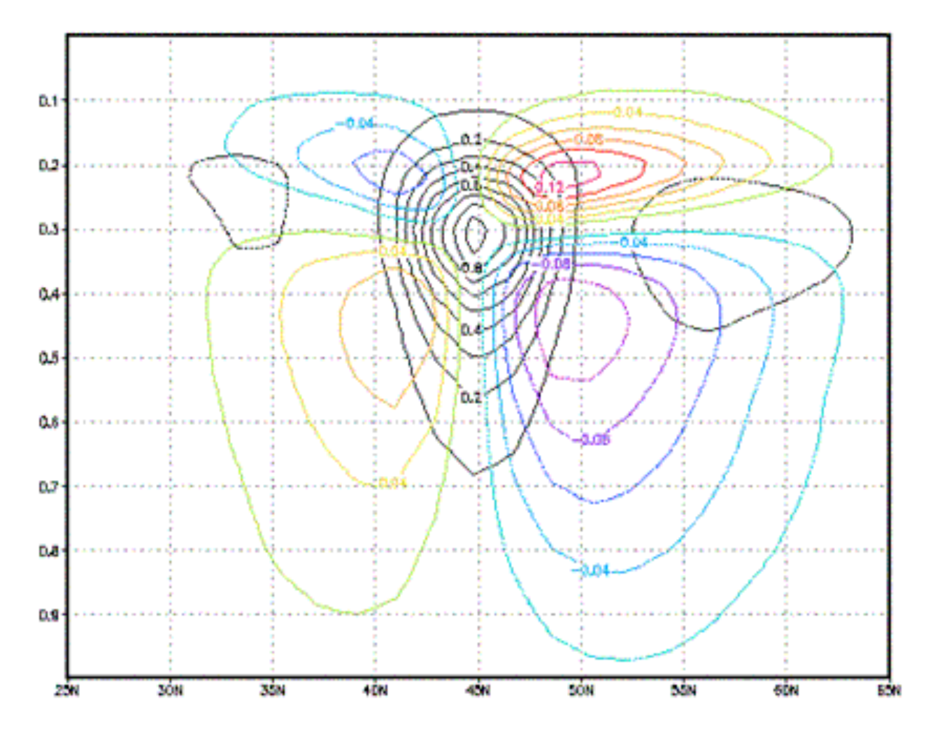
Percentage of full temperature variance explained by the balance projection

Gridpoint Statistical Interpolation Background term

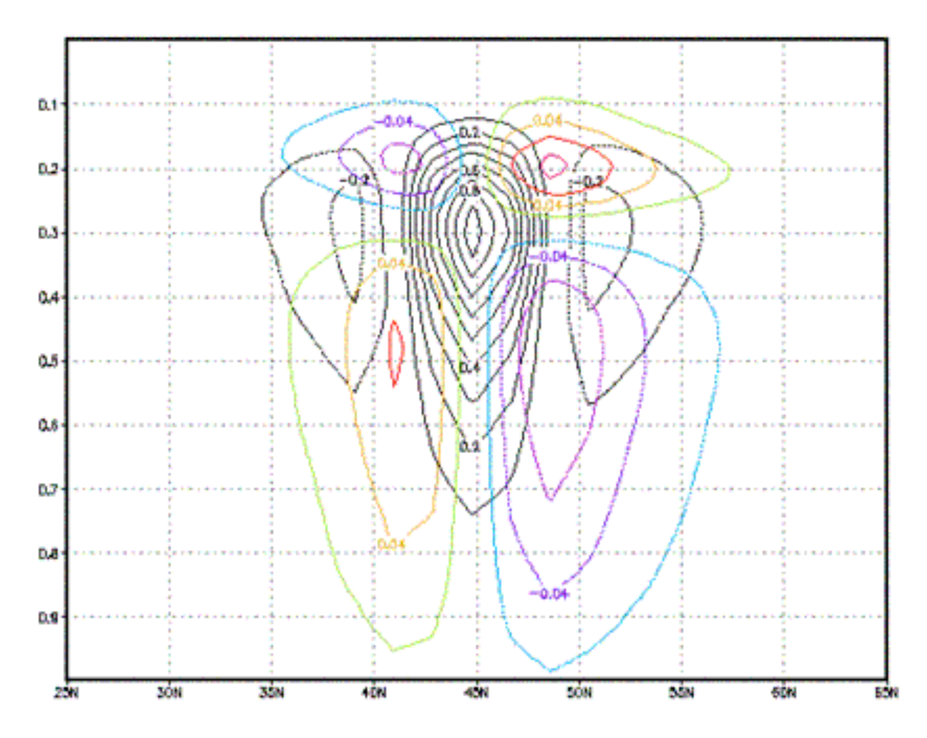
- Originated from SSI analysis system
  - Replace spectral definition of background errors with grid point representation, potential for:
    - Anisotropic, non-homogenous flow-dependent structures
    - Allows for application to various modeling systems
- Covariances Defined using Recursive Filters (Purser)
- Tangent Linear Normal Mode Constraint
- Situation Dependent Variances

Single Observation Analysis

Single zonal wind observation (1.0 ms^{-1} O-F and error)
Cross Section at 180^o

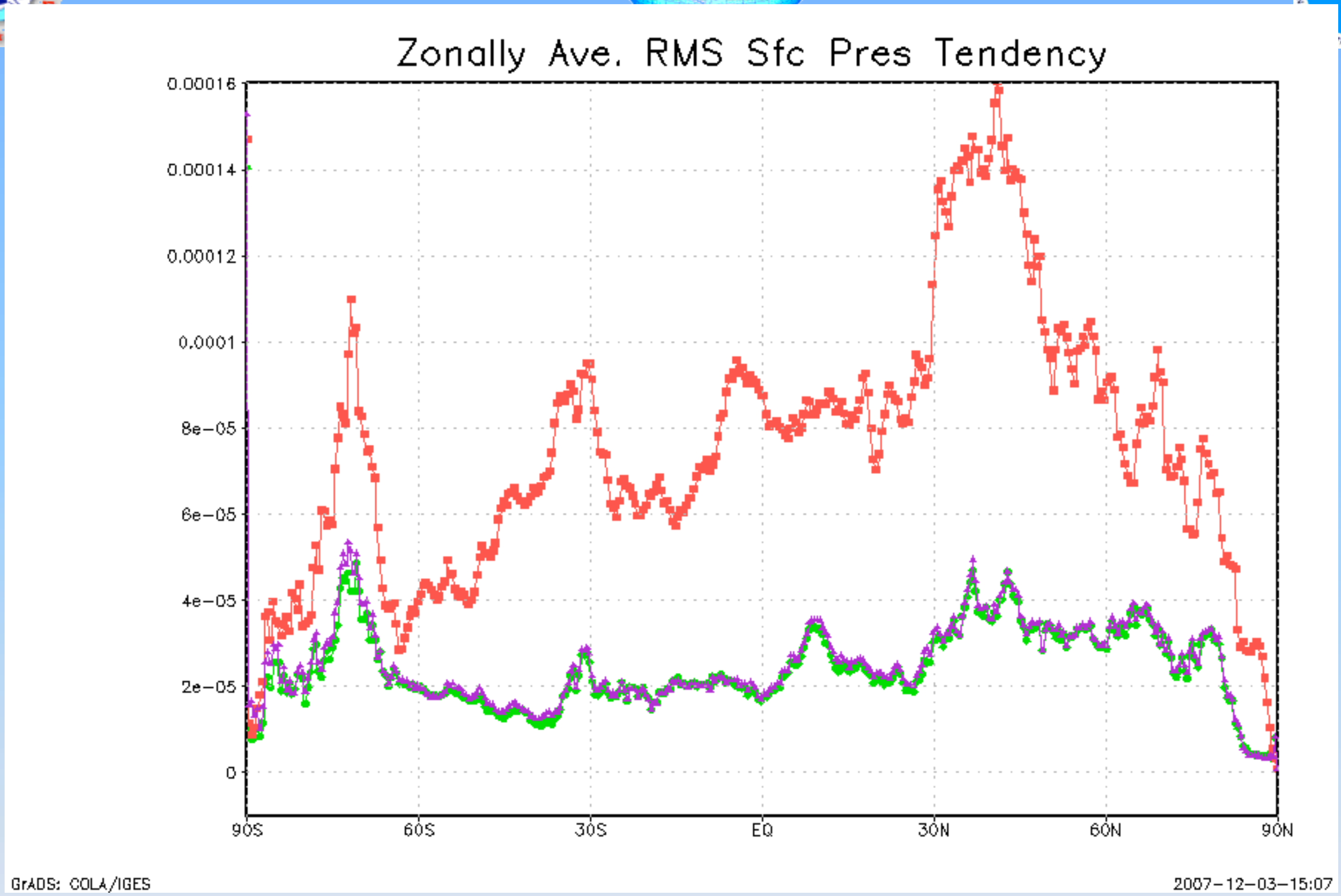


u increment (black,interval 0.1 ms^{-1}) and T increment (color, interval 0.02 K) from SSI



u increment at (black, interval 0.1 ms^{-1}) and T increment (color, interval 0.02K) from GSI (no constraint)

“balance” seems adequate at first glance



Zonal-average surface pressure tendency for guess (green), unconstrained GSI analysis (red), and GSI analysis with TLNMC (purple).

Tangent Linear Normal Mode Constraint

- Initial testing of GSI as part of GDAS showed a flaw relative to SSI
  - Statistical based multivariate coupling through variable definition and background error was deficient
    - Incremental balance was inadequate
    - Increment was often noisy
- Borrowing ideas from normal mode initialization,

$$J(x)=\frac{1}{2}\left\{x^T C^{-T} B^{-1} C^{-1} x+[H(x)-y]^T R^{-1}[H(x)-y]\right\}$$

$$x=\mathbf{C}u$$

* Similar idea developed and pursued independently by Fillion et al. (2007)

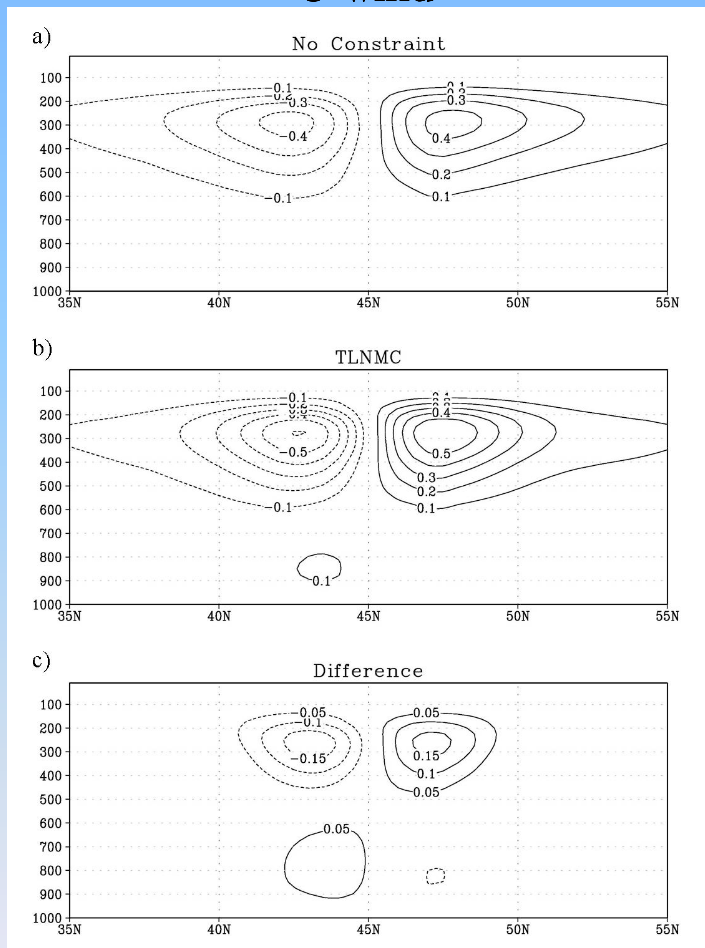
Tangent Linear Normal Mode Constraint

- Performs correction to *increment* to reduce gravity mode tendencies
- Applied during minimization to *increment*, not as post-processing of analysis fields
- Little impact on speed of minimization algorithm
- $\mathbf{CBC}^T$ becomes effective background error covariances for balanced increment
  - Adds implicit flow dependence
- Requires time tendencies of increment
  - Implemented dry, adiabatic, generalized coordinate tendency model (TL and AD)

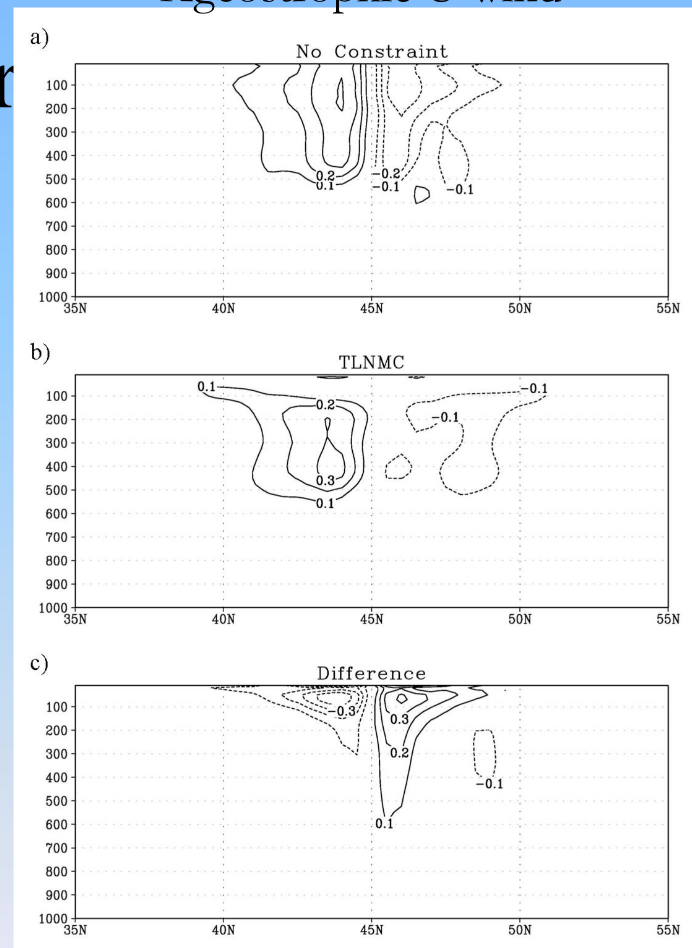
U wind

Ageostrophic U wind

From multivariate B

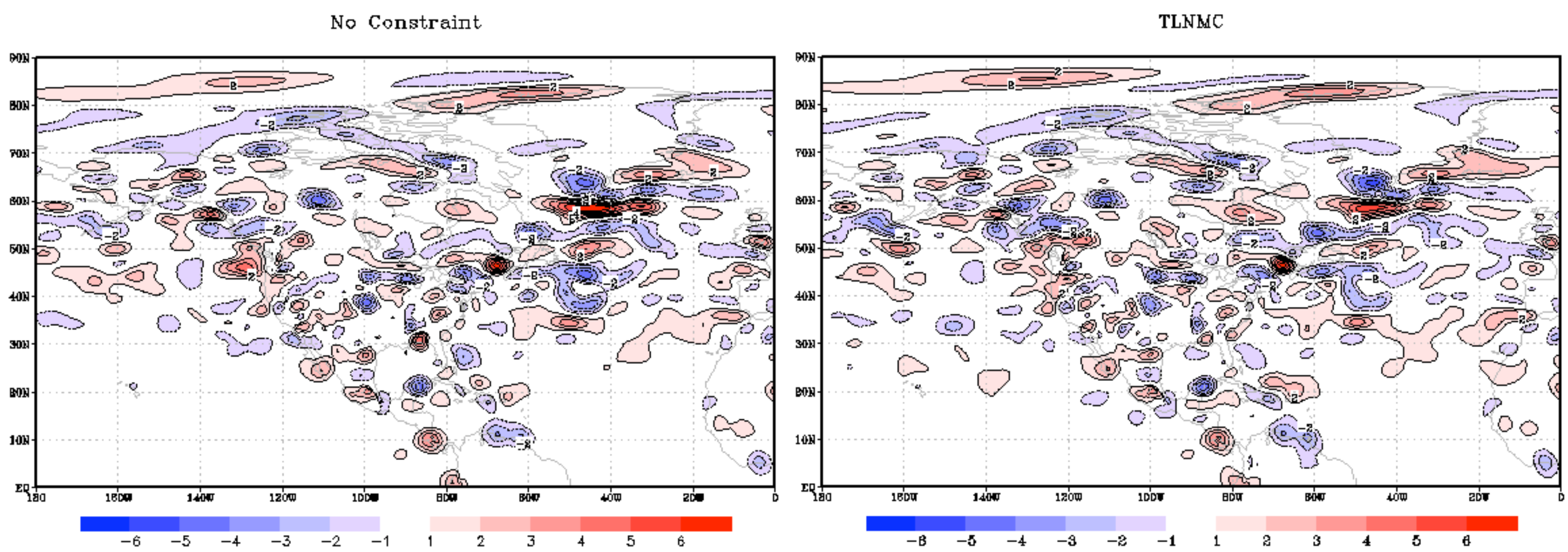


TLNMC corrects

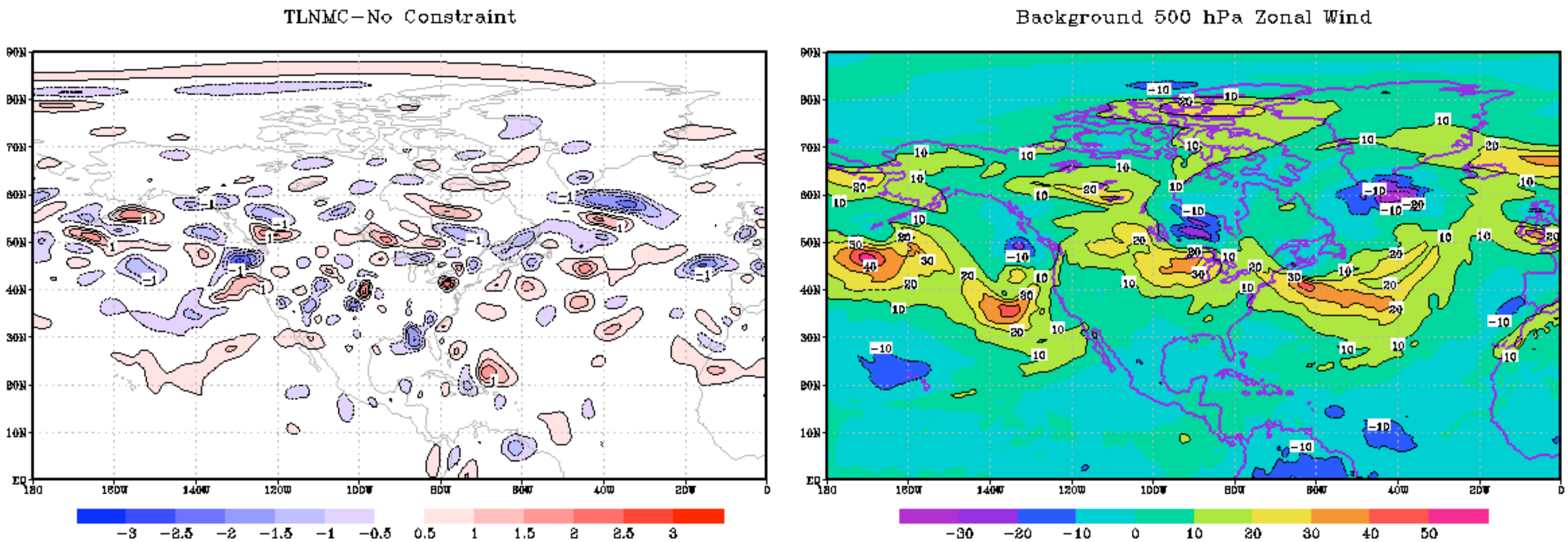


Smaller ageostrophic component

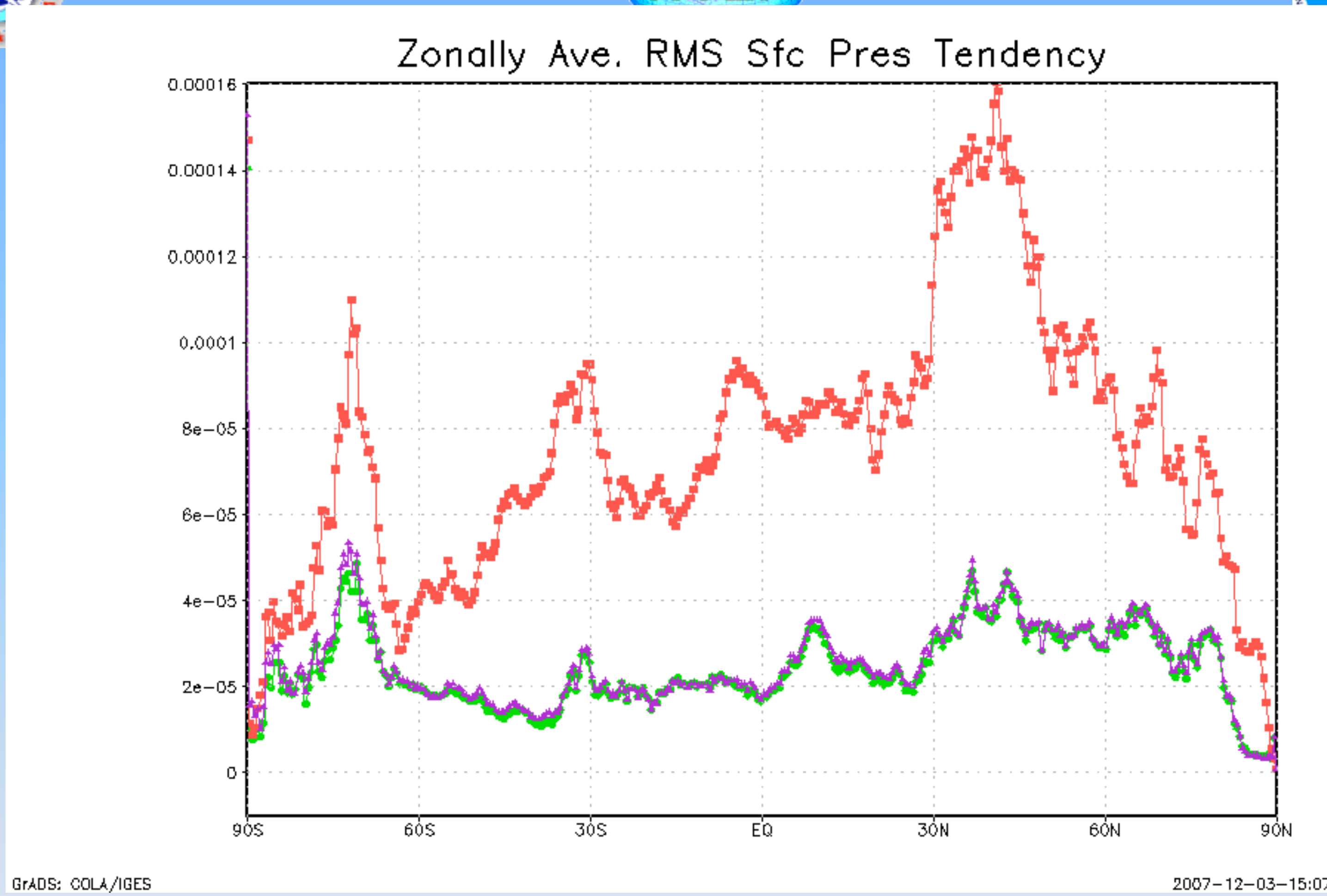
Cross section of zonal wind increment (and analysis difference) valid at 12Z 09 October 2007 for a single 500 hPa *temperature* observation (1K O-F and observation error)



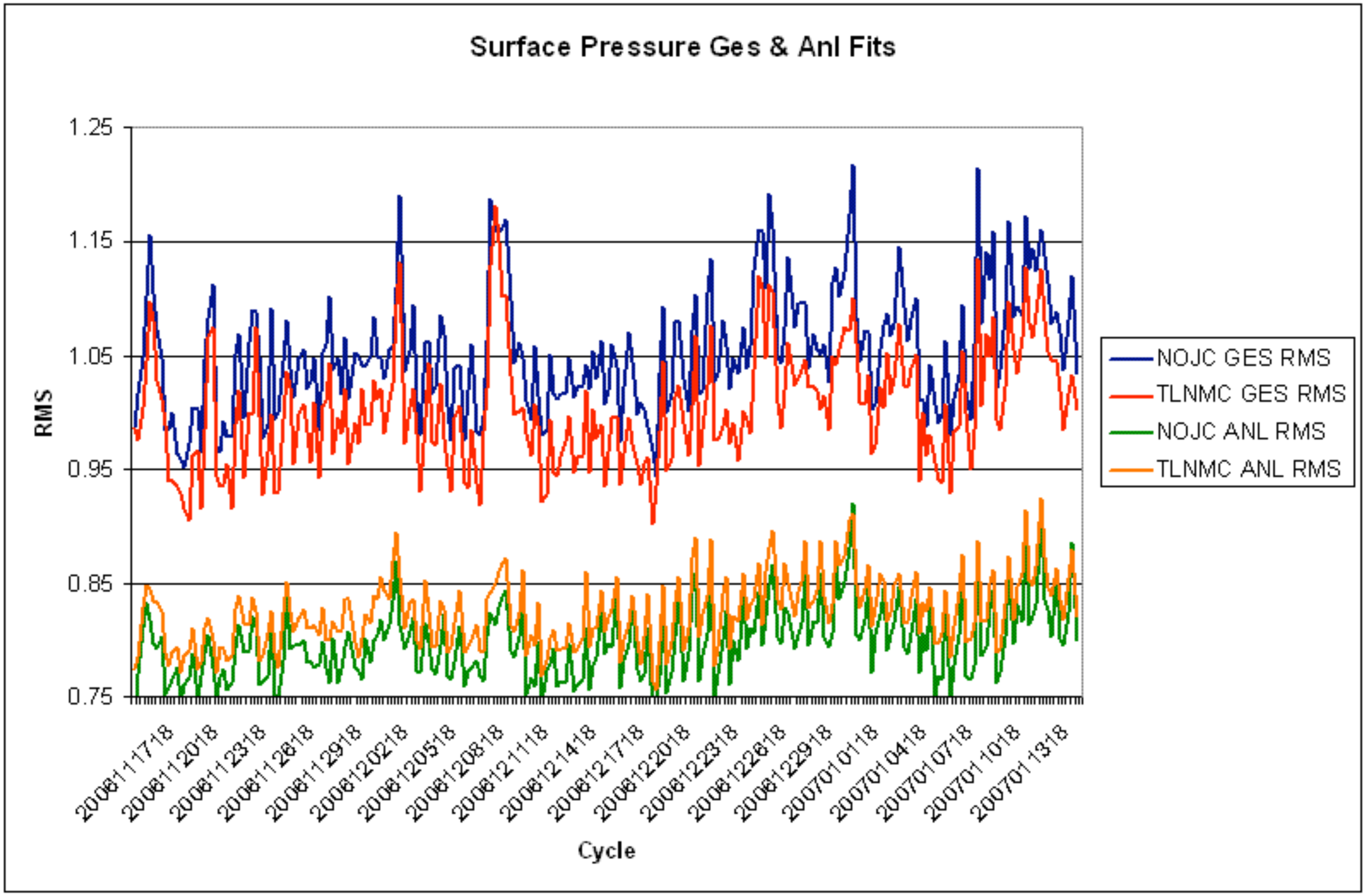
500 hPa zonal wind analysis increment valid at 12Z 09 October 2007 from the same background and set of available observations



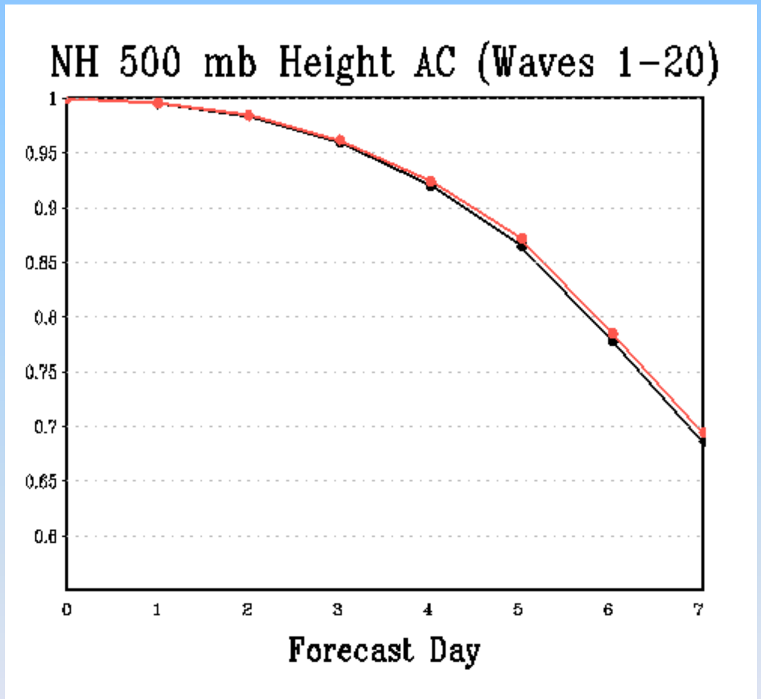
500 hPa zonal wind analysis difference (TLNMC-No Constraint ; left) and zonal wind background (right) valid at 12Z 09 October 2007

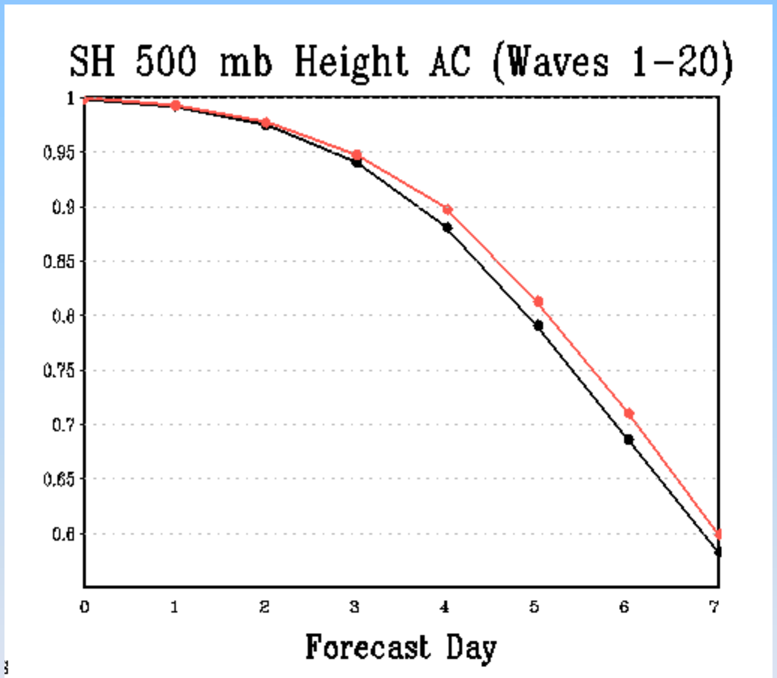


Zonal-average surface pressure tendency for guess (green), unconstrained GSI analysis (red), and GSI analysis with TLNMC (purple).


Surface Pressure Ges & Anl Fits
1.25
1.15
1.05
0.95
0.85
0.75
RMS
2006111718
2006112018
2006112318
2006112618
2006112918
2006120218
2006120518
2006120818
2006121118
2006121418
2006121718
2006122018
2006122318
2006122618
2006122918
2007010118
2007010418
2007010718
2007011018
2007011318
Cycle
NOJC GES RMS
TLNMC GES RMS
NOJC ANL RMS
TLNMC ANL RMS

Impact of TLNMC on 500 hPa AC Scores





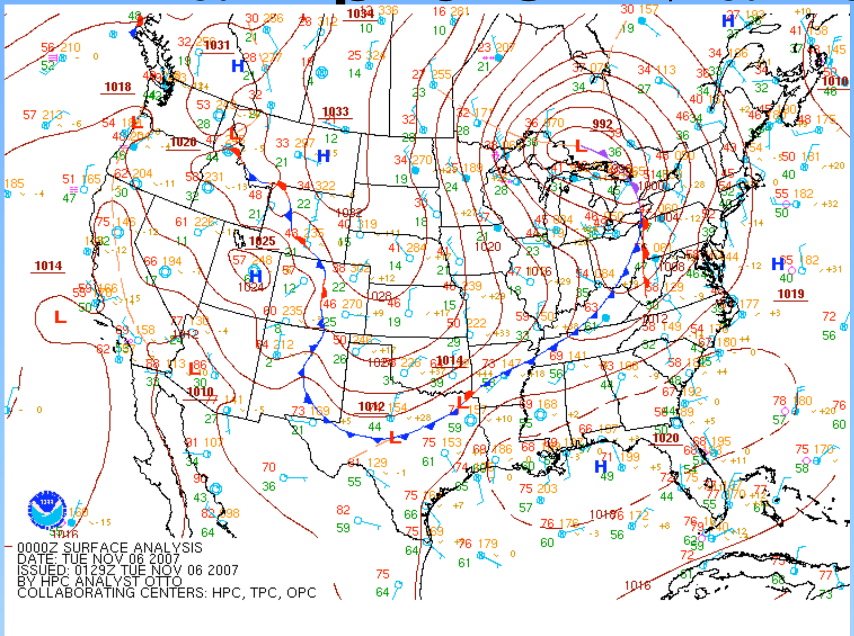
Control: Black TLNMC: Red

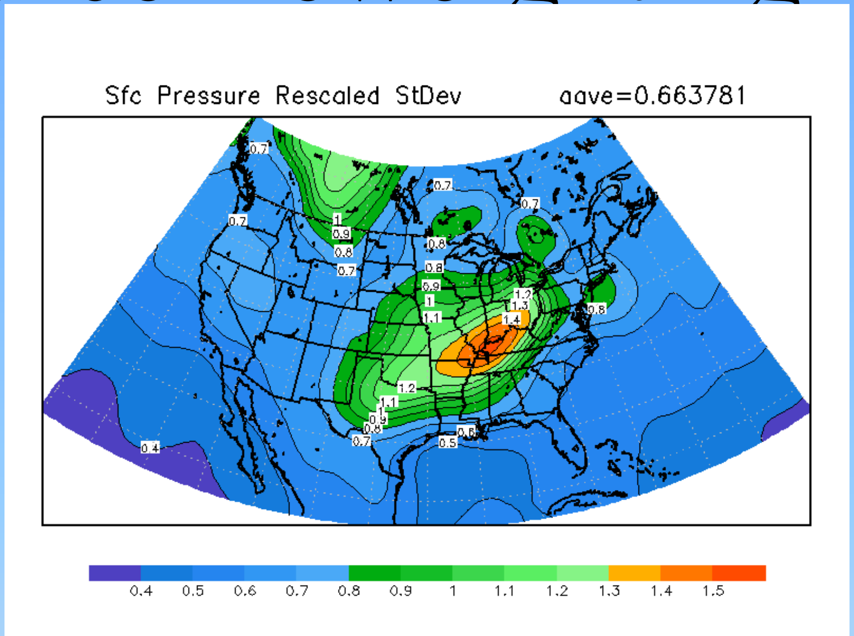
500 hPa Geo. Height AC Scores for period 01 Dec. 2006 to 14 Jan. 2007

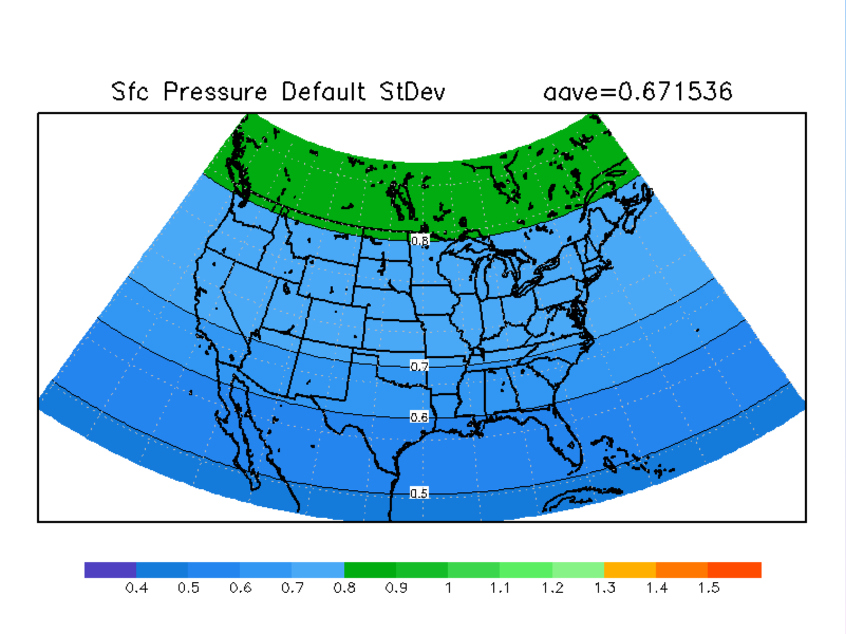
Flow Dependent **B** (variances only)

- One motivation for GSI was to permit flow dependent variability in background error
- Take advantage of FGAT (guess at multiple times) to modify variances based on 9h-3h differences
  - Variance increased in regions of rapid change
  - Variance decreased in “calm” regions
  - Global mean variance ~ preserved
- Perform reweighting on streamfunction, velocity potential, virtual temperature, and surface pressure only (for now)

Example of Variance Reweighting







Surface pressure background error standard deviation fields

a) with flow dependent re-scaling
b) without re-scaling

Valid: 00 UTC November 2007

Observation Term

- Include wide variety of observations
- Most of differences in inner and outer loop are in this term
- Our philosophy is to use data as close a possible to the observed quantity

Input data - Conventional

- Radiosondes
- Pibal winds
- Synthetic tropical cyclone winds
- wind profilers
- conventional aircraft reports
- ASDAR aircraft reports
- MDCARS aircraft reports
- dropsondes
- MODIS IR and water vapor winds
- GMS, METEOSAT and GOES cloud drift IR and visible winds
- GOES water vapor cloud top winds
- Surface land observations
- Surface ship and buoy observation
- SSM/I wind speeds
- QuikScat wind speed and direction SSM/I precipitable water
- SSM/I and TRMM TMI precipitation estimates
- Doppler radial velocities
- VAD (NEXRAD) winds
- GPS precipitable water estimates
- GPS Radio occultation refractivity profiles
- SBUV ozone profiles (other ozone data under test)
- TAMDAR aircraft data (under final testing in regional model

Input data - Satellite

- Regional
  - GOES-11 and 12 Sounders – thinned to 120km
    - Channels 1-15
    - Individual fields of view
    - 4 Detectors treated separately
    - Over ocean only
  - AMSU-A – thinned to 60km
    - NOAA-15 Channels 1-10, 12, 15
    - NOAA-18 Channels 1-8, 10-11, 15
  - AMSU-B/MHS – thinned to 60km
    - NOAA-15 Channels 1-3, 5
    - NOAA-16 Channels 1-5
    - NOAA-17 Channels 1-5
    - NOAA-18 Channels 1-5
  - HIRS – thinned to 120km
    - NOAA-17 Channels 2-15
  - METOP AMSU-A, HIRS, MHS under final pre operational testing
- Global
  - GOES-11 and 12 Sounders –thinned to 180km
    - Channels 1-15
    - Individual fields of view
    - 4 Detectors treated separately
    - Over ocean only
  - AMSU-A – thinned to 145km
    - NOAA-15 Channels 1-10, 12-13, 15
    - NOAA-18 Channels 1-8, 10-13, 15
    - METOP Channels 1-13, 15
    - AQUA Channels 1-6, 8-13, 15
  - AMSU-B/MHS – thinned to 240km
    - NOAA-15 Channels 1-3, 5
    - NOAA-16 Channels 1-5
    - NOAA-17 Channels 1-5
    - NOAA-18 Channels 1-5
    - METOP Channels 1-5
  - HIRS - thinned to 180km
    - NOAA-17 Channels 2-15
    - METOP Channels 2-15
  - AIRS – thinned to 180km
    - AQUA 148 Channels
  - METOP – IASI (longwave channels) under final pre-operational testing

Atmospheric analysis problem
Outer (K) and Inner (L) iteration operators

Variable	*K* operator	L operator
Temperature – surface obs. at 2m	3-D sigma interpolation adjustment to different orography	3-D sigma interpolation Below bottom sigma assumed at bottom sigma
Wind – surface obs. at 10m over land, 20m over ocean, except scatt.	3-D sigma interpolation reduction below bottom level using model factor	3-D sigma interpolation reduction below bottom level using model factor
Ozone – used as layers	Integrated layers from forecast model	Integrated layers from forecast model
Surface pressure	2-D interpolation plus orography correction	2-D interpolation
Precipitation	Full model physics	Linearized model physics
Radiances	Full radiative transfer	Linearized radiative transfer

Operational data requirements

- Requirements for operational use of observations
  - Available in real time in acceptable format
  - Assurance of stable data source
  - Quality control procedures defined (conservative)
  - Observational errors defined (and bias removed if necessary)
  - Accurate forward model (and adjoint) available
  - Integration into data monitoring
  - Evaluation and testing to ensure neutral/positive impact

Note!

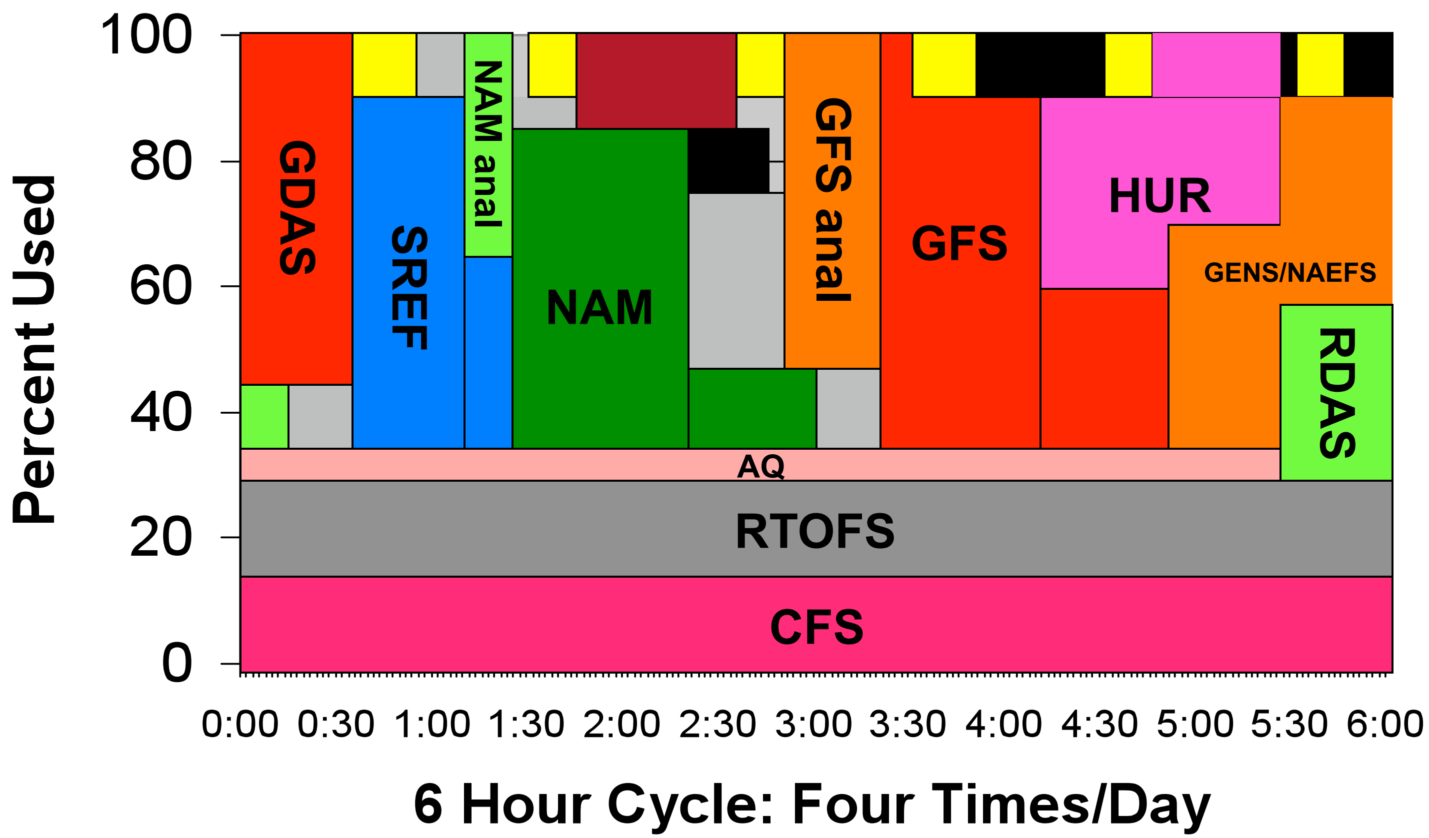
- I will be using satellite radiances as an example of the use of observations in data assimilation
- Other types of data can be as complex as the radiances – but with a different set of problems

Data available in real time in acceptable format

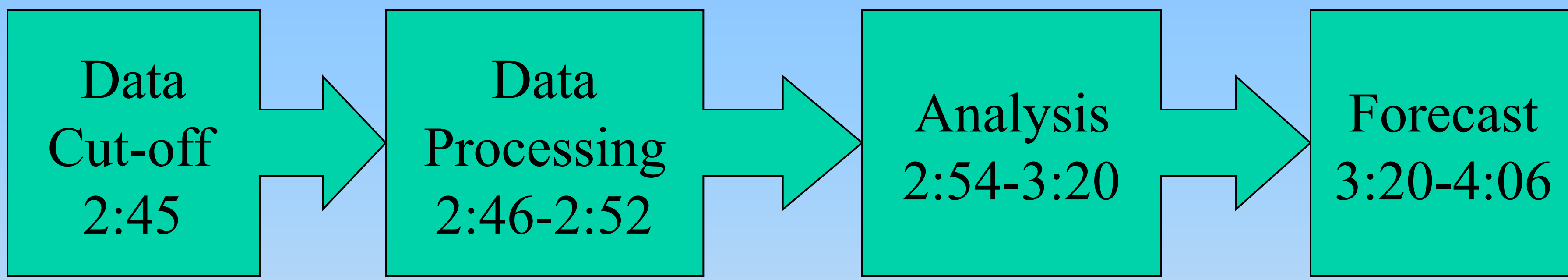
- Data formats
  - WMO acceptable formats – BUFR – CREX (not really relevant) – used by most NWP centers
  - Almost every satellite program uses a different format
  - Significant time and resources used understanding/ converting/developing formats
- If data is not available in time for use in data assimilation system – not useful

NCEP Production Suite
Weather, Ocean, Land & Climate Forecast Systems

2007

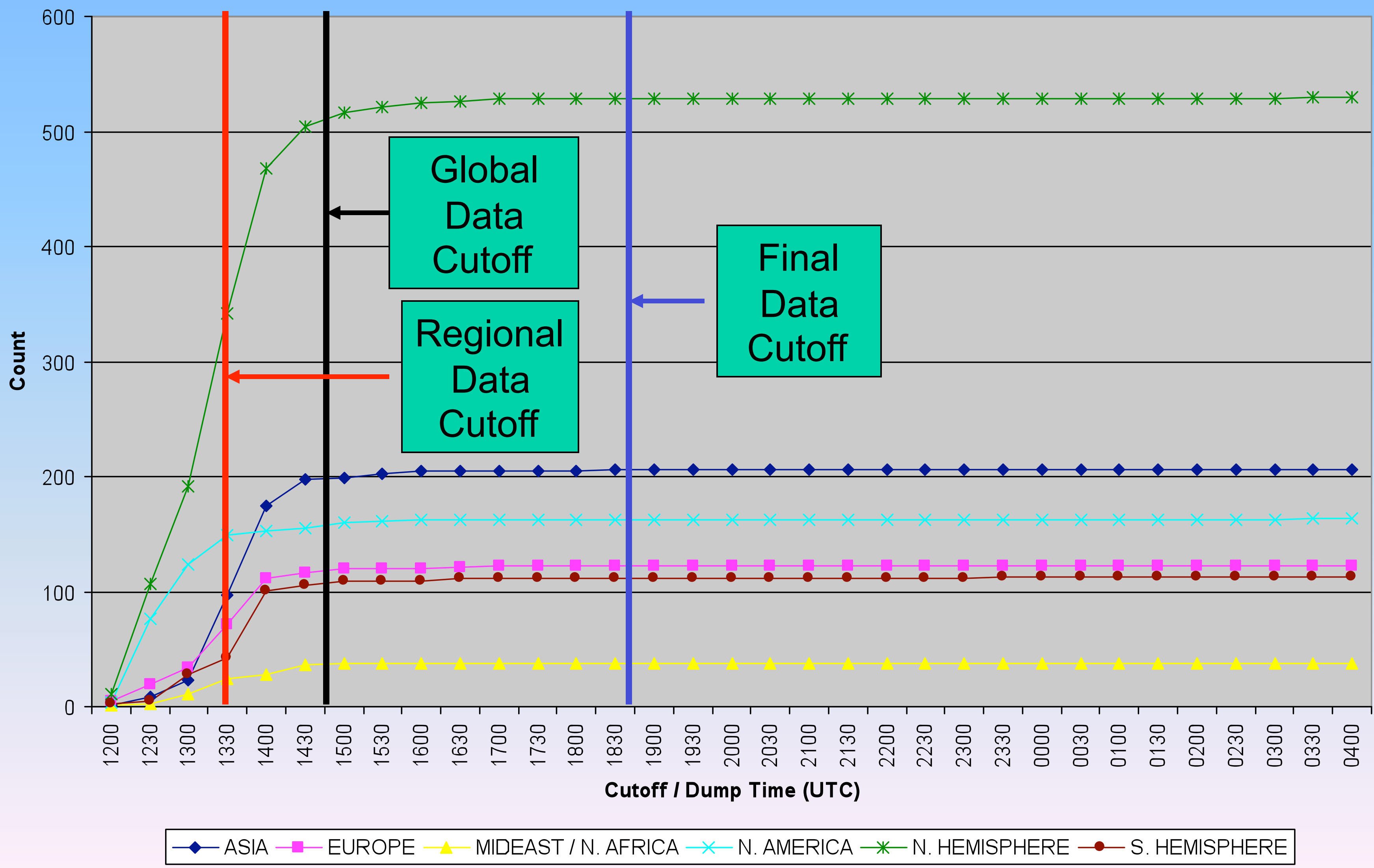


GFS analysis/forecast cycle

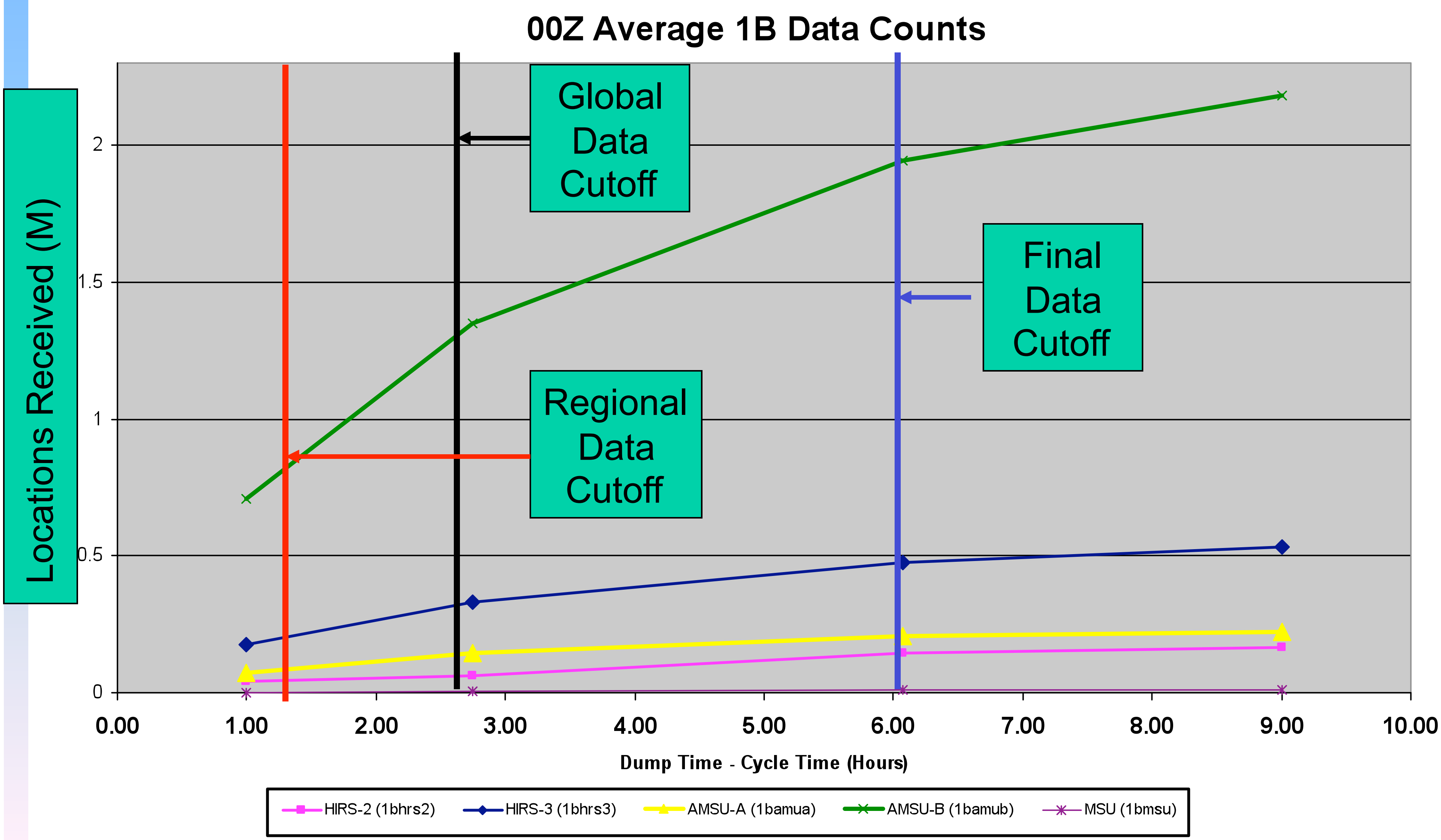


- Any data not available by Cut-off will not be used
- Later catch up cycle at +6:00

Rawinsonde Delivery



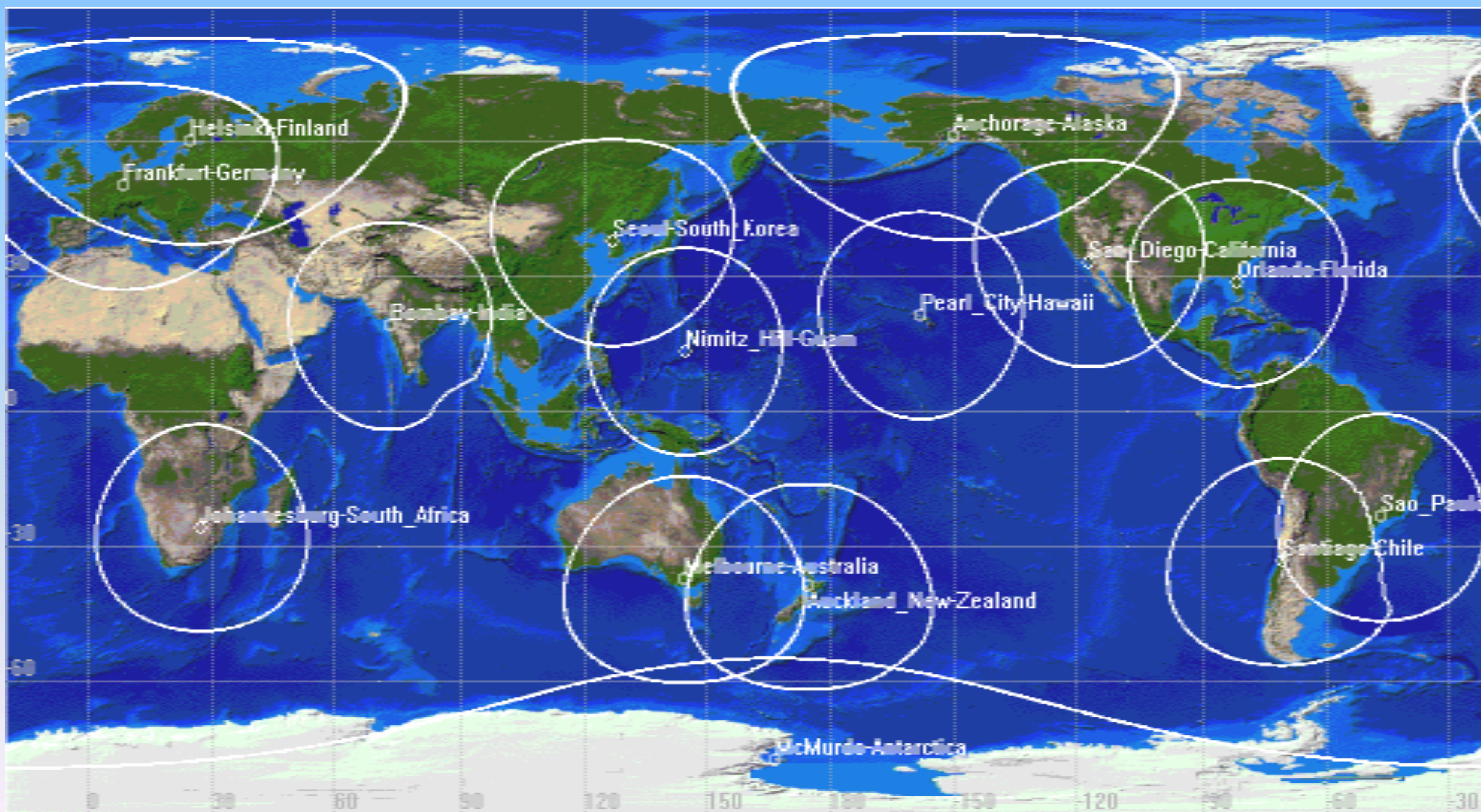
POES Data Delivery



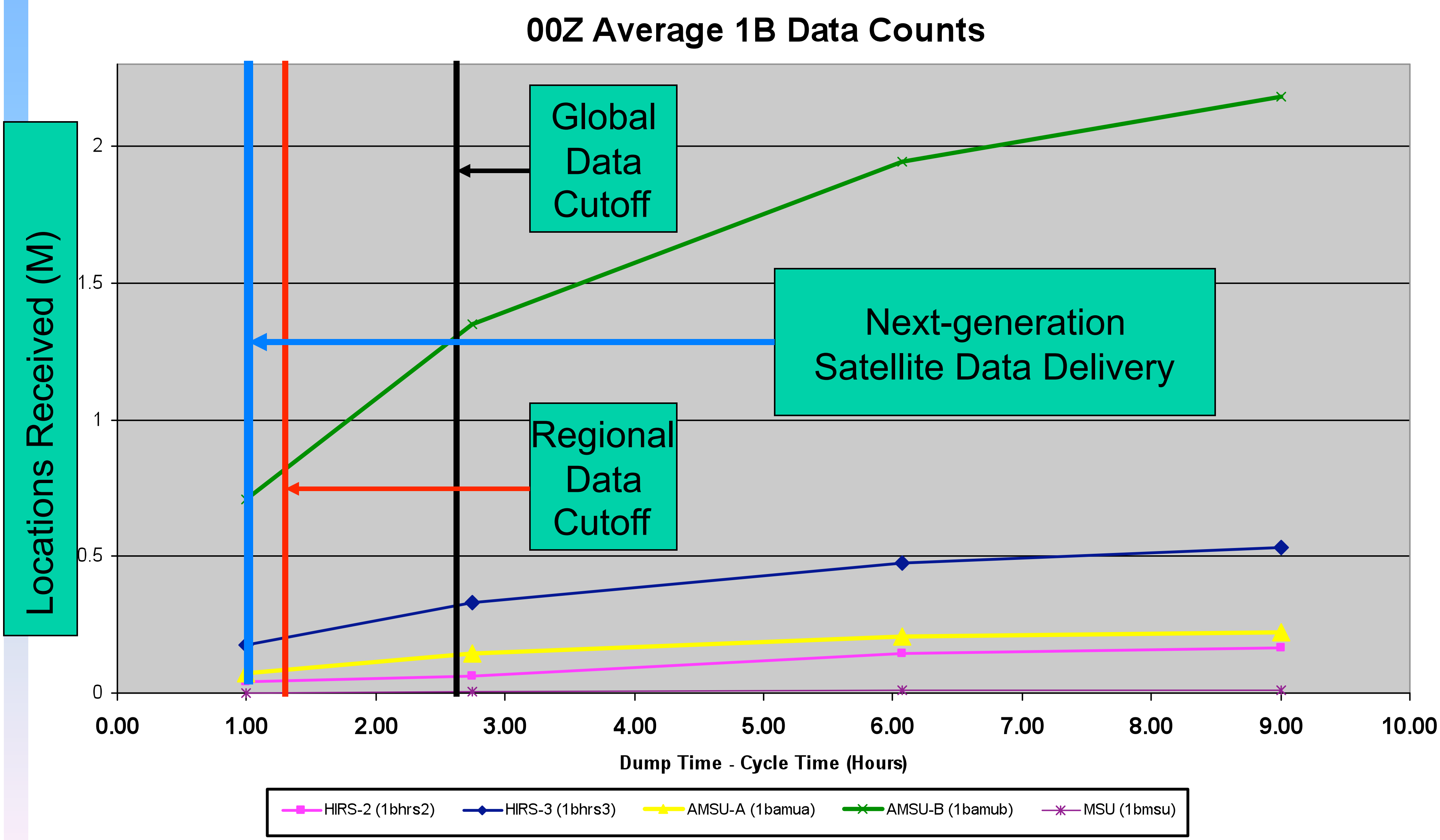
Satellite data delivery

- Satellite data must wait until ground station within sight to download – Blind orbits
- Conflicts between satellites
- Proposed NPOESS ground system (METOP currently left out)
  - **SafetyNet is a system of 15 globally distributed receptors linked to the centrals via commercial fiber, it enables low data latency and high data availability**

NPOESS SafetyNetTM Architecture



POES Data Delivery



Assurance of stable data source

- Changes in data processing can result in changes in observation error characteristics
- Notification, testing and provision of test data sets essential prior to changes
- For operational satellites – situation OK
- For research satellites – means loss of control by instrument/program scientists

Accurate forward model

- One of the biggest data assimilation developments in the last 15 years was allowing the observations to be different from the analysis variables
  - In variational schemes this is done through the *K* operator
  - In OI, the same thing could be done – but was only rarely done.
  - The development allows us to use the observations as they were observed AND allows the use of analysis variables with nice properties.

Satellite data

- Example –
  - y are radiance observations,
  - x are profiles of temperature, moisture and ozone.
  - *K* is the radiative transfer equation and
  - z are unknown parameters such as the surface emissivity (dependent on soil type, soil moisture, etc.), CO2 profile, methane profile, etc.
- In general, *K* is not invertable – thus retrievals.
  - Physical retrievals – usually very similar to 1D variational problems (with different background fields)
  - Statistical retrievals – given y predict x using regression

Satellite data context

- 3-4 D variational analysis can be thought of as a generalization of “physical retrieval” to include all types of data and spatial and temporal variability.
- To use data in 2 steps – retrieval and then analysis-- can be done consistently if K is linear and if one is very careful – but is generally suboptimal.

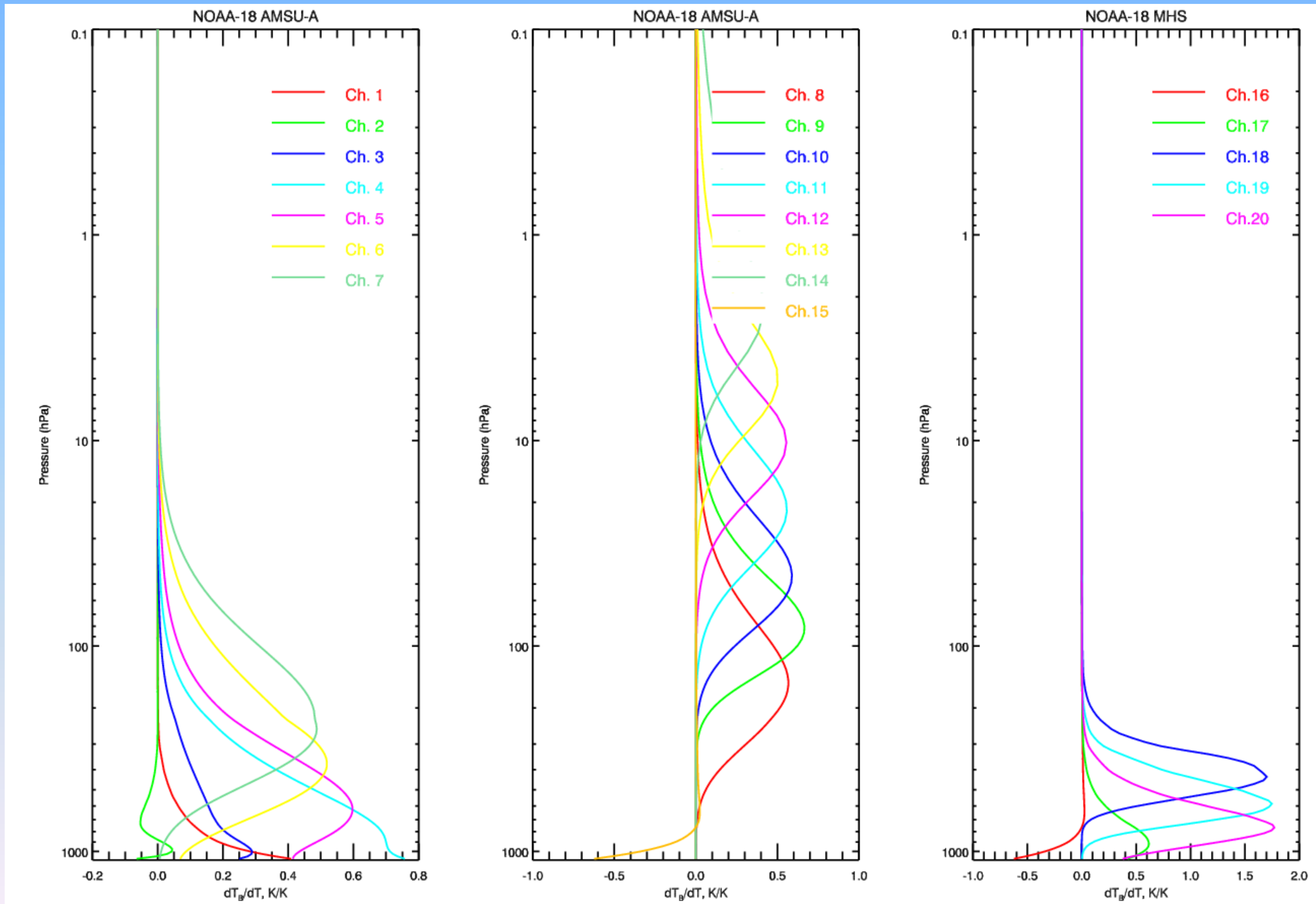
Satellite Radiance Observations

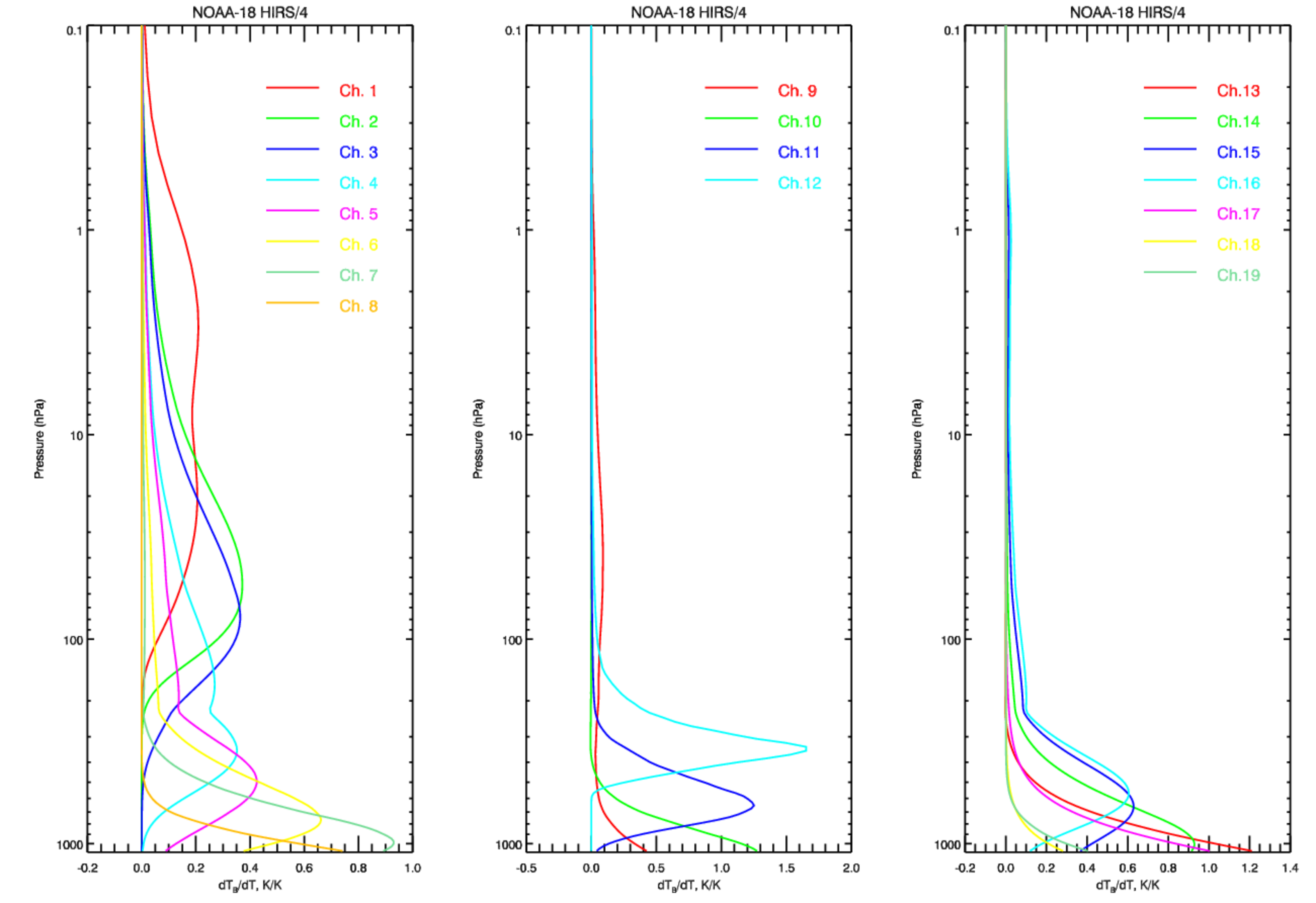
- Measure upwelling radiation at top of atmosphere
- Measure deep layers
  - IR not quite as deep as microwave
  - New IR instruments (AIRS, IASI, GIFTS) narrower, but still quite deep layers
  - Deep layers generally implies large horizontal scale

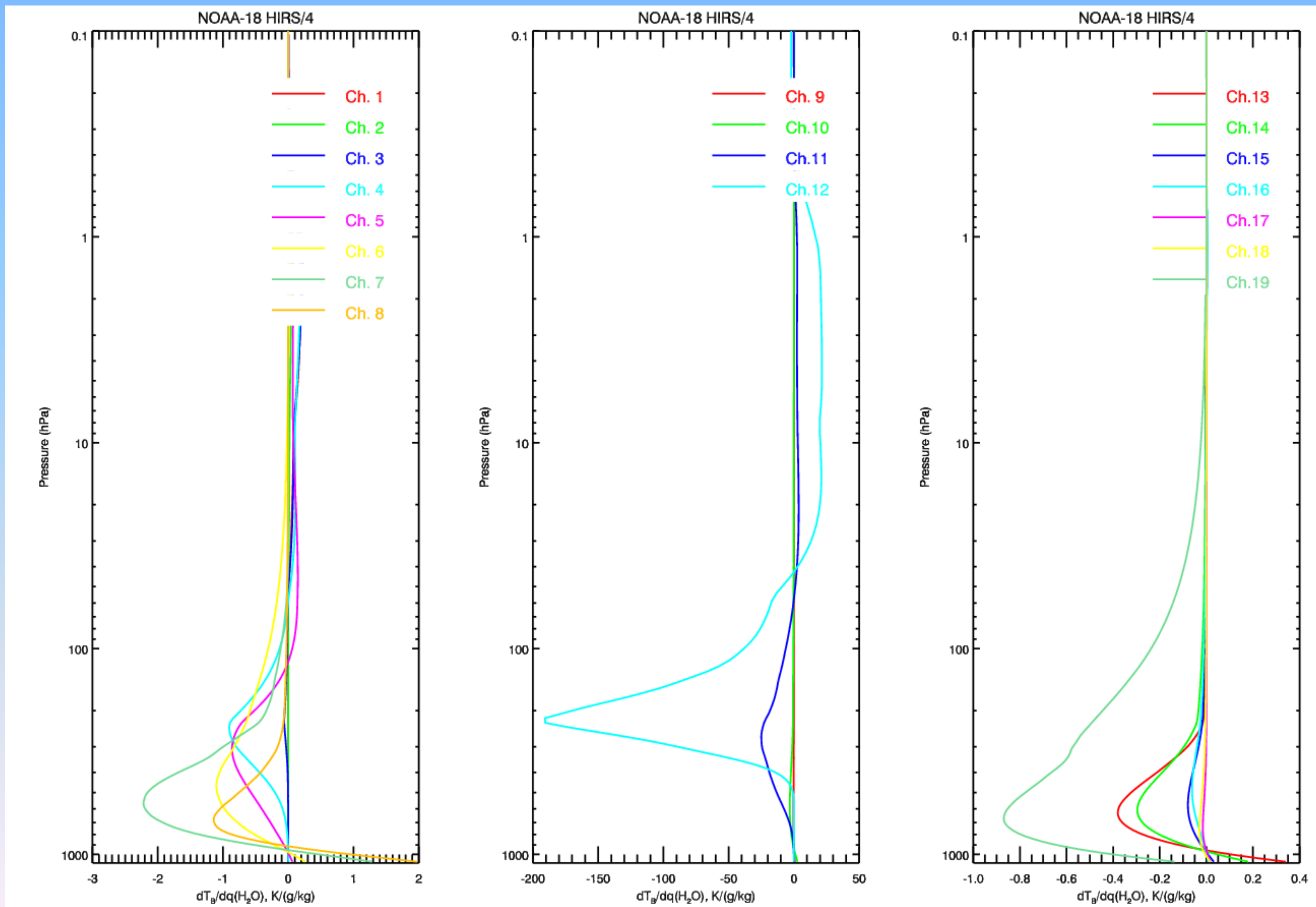
Forward model for RT

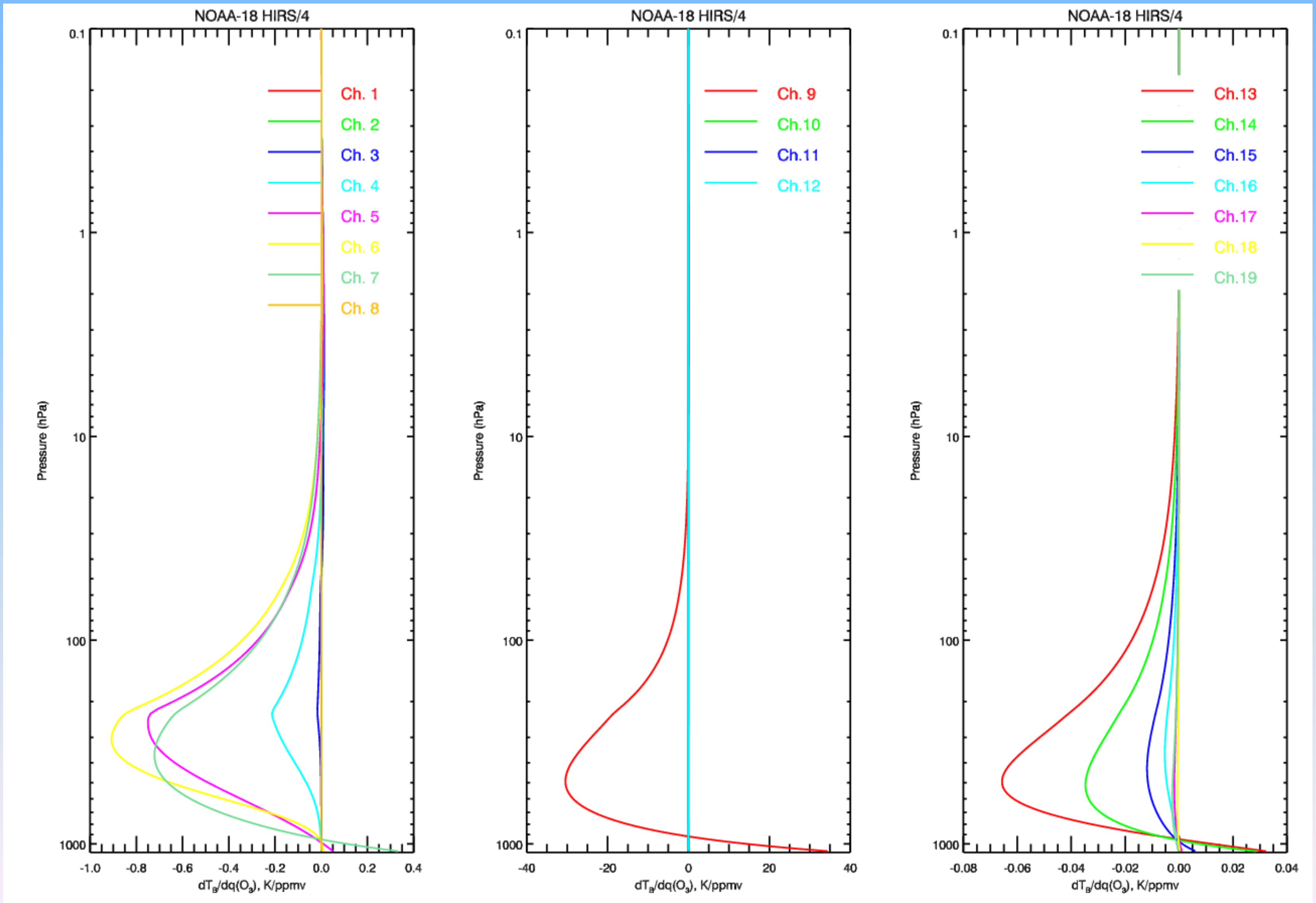
- RTTOV – CRTM two examples of fast forward models
- From CRTM get both simulated radiance and

$$\frac{\partial R}{\partial T}, \frac{\partial R}{\partial q}, \frac{\partial R}{\partial q}, \frac{\partial R}{\partial O_3}, \ldots$$

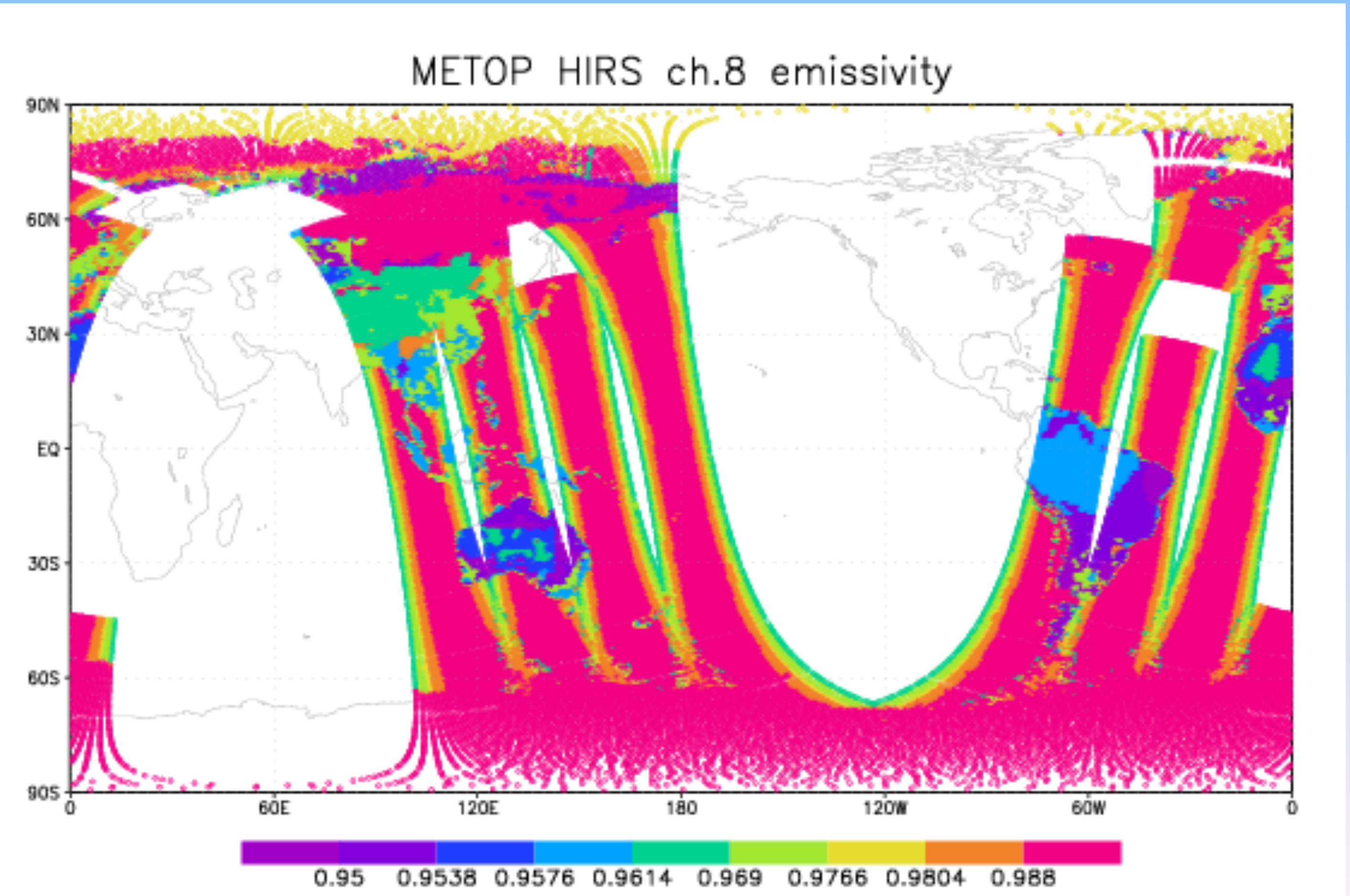

NOAA-18 AMSU-A
Ch. 1
Ch. 2
Ch. 3
Ch. 4
Ch. 5
Ch. 6
Ch. 7
Pressure (hPa)
0.1
1
10
100
1000
-0.2
0.0
0.2
0.4
0.6
0.8
dT_B/dT, K/K
NOAA-18 AMSU-A
Ch. 8
Ch. 9
Ch.10
Ch.11
Ch.12
Ch.13
Ch.14
Ch.15
Pressure (hPa)
0.1
1
10
100
1000
-1.0
-0.5
0.0
0.5
1.0
dT_B/dT, K/K
NOAA-18 MHS
Ch.16
Ch.17
Ch.18
Ch.19
Ch.20
Pressure (hPa)
0.1
1
10
100
1000
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
dT_B/dT, K/K


NOAA-18 HIRS/4
Pressure (hPa)
0.1
1
10
100
1000
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
dT_B/dT, K/K
Ch. 1
Ch. 2
Ch. 3
Ch. 4
Ch. 5
Ch. 6
Ch. 7
Ch. 8
NOAA-18 HIRS/4
Pressure (hPa)
-0.5
0.0
0.5
1.0
1.5
2.0
dT_B/dT, K/K
Ch. 9
Ch.10
Ch.11
Ch.12
NOAA-18 HIRS/4
Pressure (hPa)
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
dT_B/dT, K/K
Ch.13
Ch.14
Ch.15
Ch.16
Ch.17
Ch.18
Ch.19

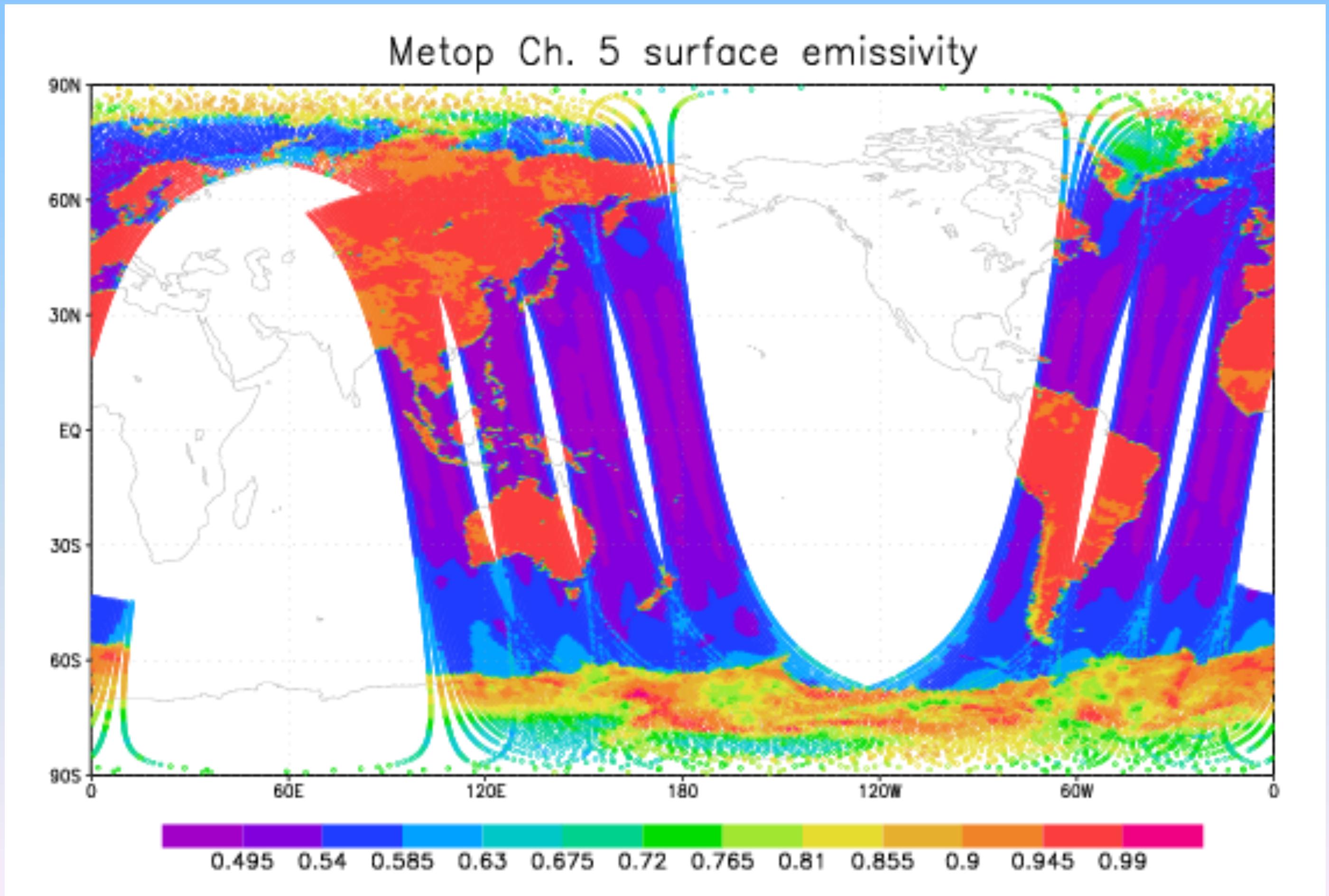

NOAA-18 HIRS/4
Pressure (hPa)
0.1
1
10
100
1000
-3
-2
-1
0
1
2
dT$_B$/dq(H_2O), K/(g/kg)
Ch. 1
Ch. 2
Ch. 3
Ch. 4
Ch. 5
Ch. 6
Ch. 7
Ch. 8
NOAA-18 HIRS/4
Pressure (hPa)
-200
-150
-100
-50
0
50
dT$_B$/dq(H_2O), K/(g/kg)
Ch. 9
Ch.10
Ch.11
Ch.12
NOAA-18 HIRS/4
Pressure (hPa)
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
dT$_B$/dq(H_2O), K/(g/kg)
Ch.13
Ch.14
Ch.15
Ch.16
Ch.17
Ch.18
Ch.19


NOAA-18 HIRS/4
Pressure (hPa)
$dT_B/dq(O_3)$, K/ppmv
Ch. 1
Ch. 2
Ch. 3
Ch. 4
Ch. 5
Ch. 6
Ch. 7
Ch. 8
NOAA-18 HIRS/4
Pressure (hPa)
$dT_B/dq(O_3)$, K/ppmv
Ch. 9
Ch.10
Ch.11
Ch.12
NOAA-18 HIRS/4
Pressure (hPa)
$dT_B/dq(O_3)$, K/ppmv
Ch.13
Ch.14
Ch.15
Ch.16
Ch.17
Ch.18
Ch.19

Surface Emissivity
Infrared

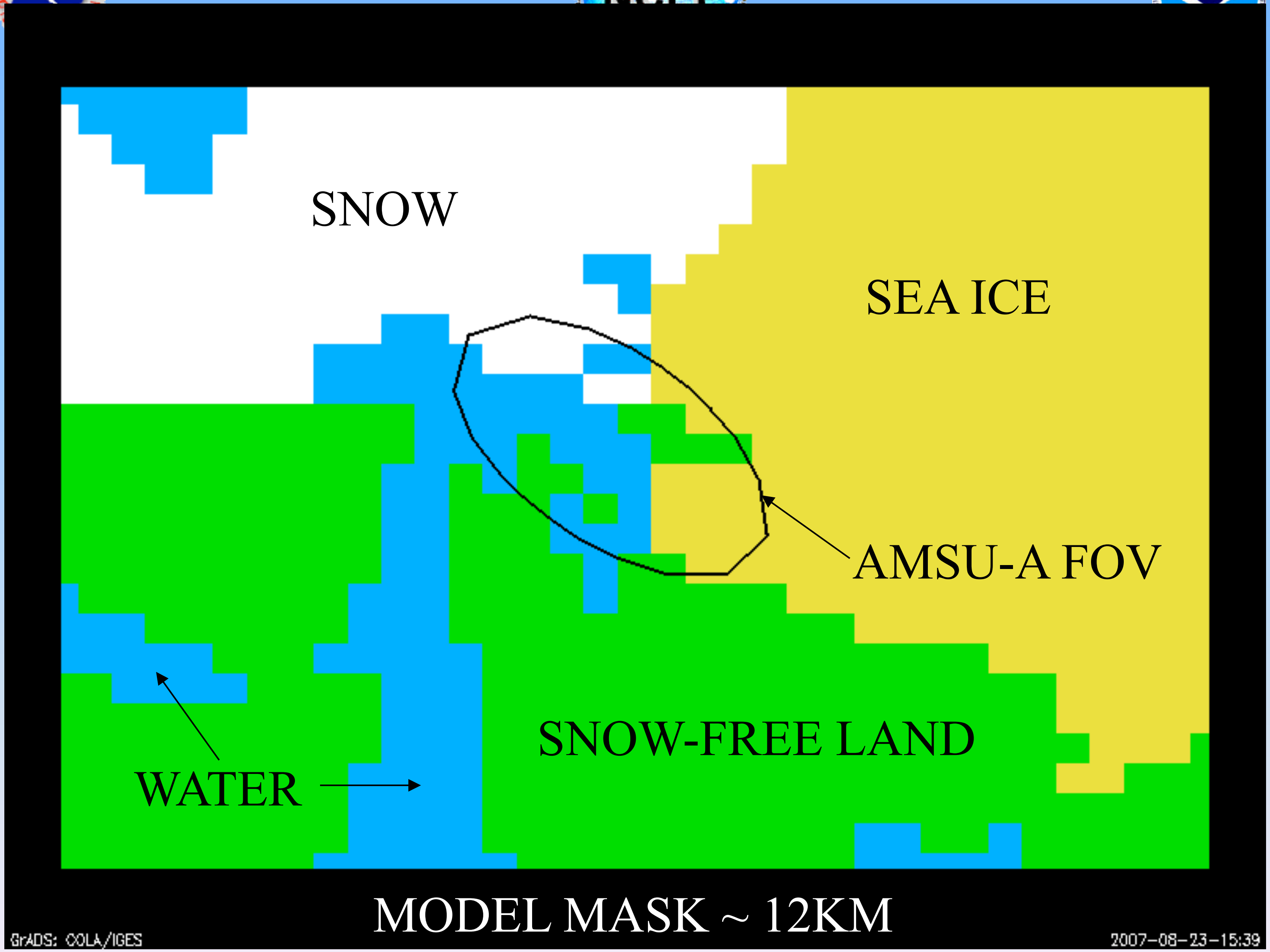


Surface Emissivity Microwave



Radiance Assimilation Problem is not done

- FOV
- Antenna Correction
- Slant path
- RT enhancements
  - Surface Emissivity
  - Additional absorbers/scatterers
  - Etc.
- Cloud Assimilation
- Aerosol Assimilation
- Trace Gas Assimilation

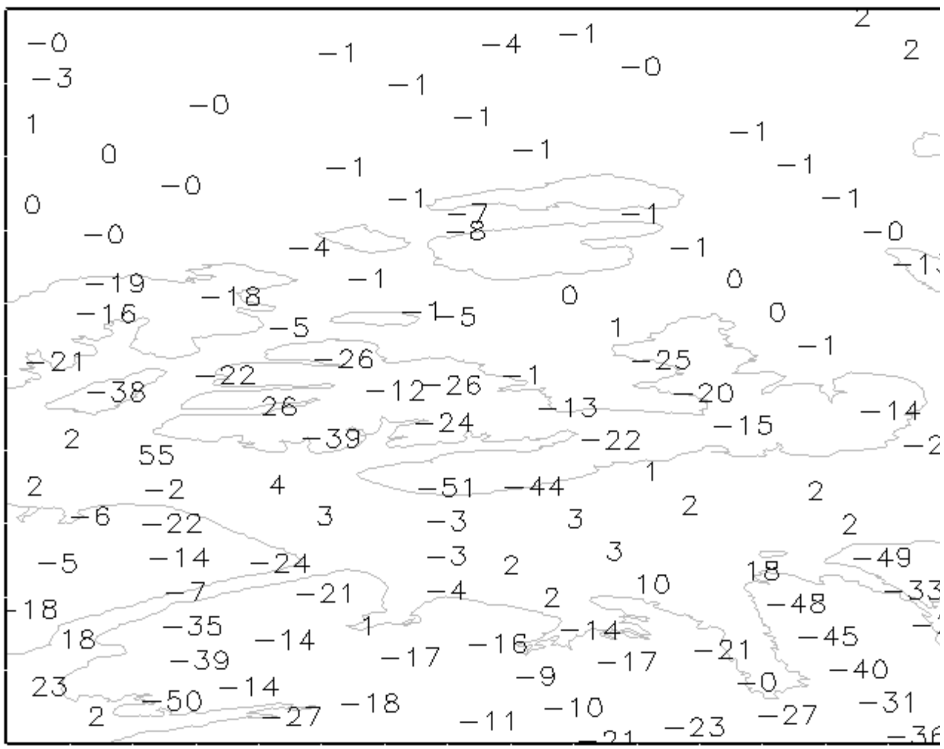

SNOW
SEA ICE
AMSU-A FOV
SNOW-FREE LAND
WATER
MODEL MASK ~ 12KM
GrADS: COLA/IGES
2007-08-23-15:39

IMPACT: ACCOUNTING FOR FOV

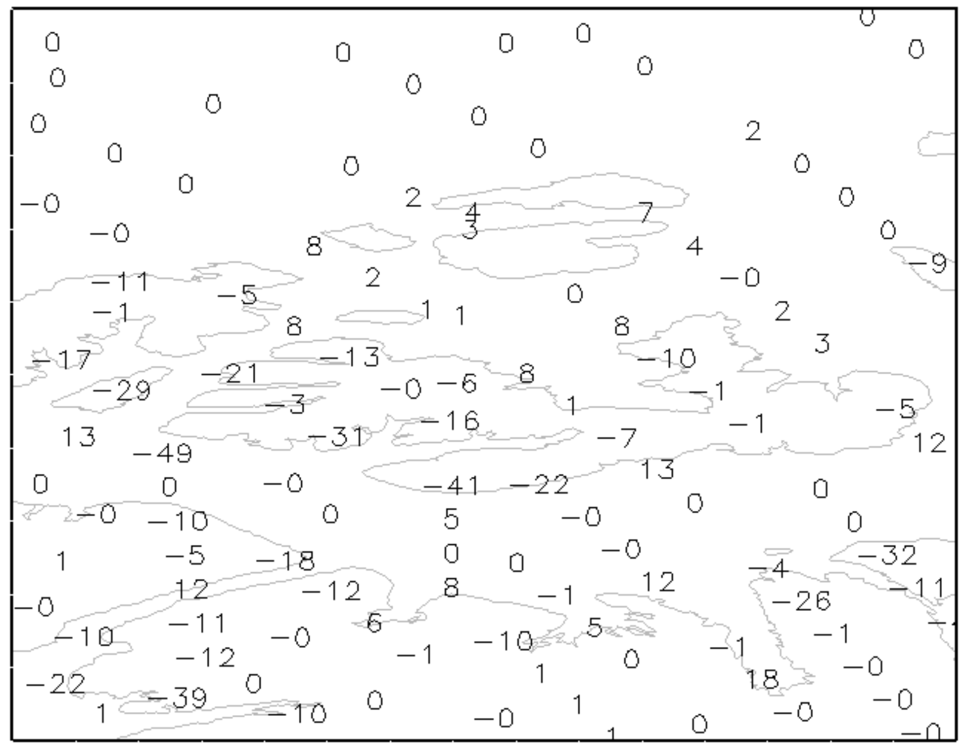
EX: NOAA-15 AMSU-A, CHANNEL 2

CONTROL: OBS. MINUS GUESS T_b

IMPACT: CHANGE IN OBS. MINUS GUESS T_b



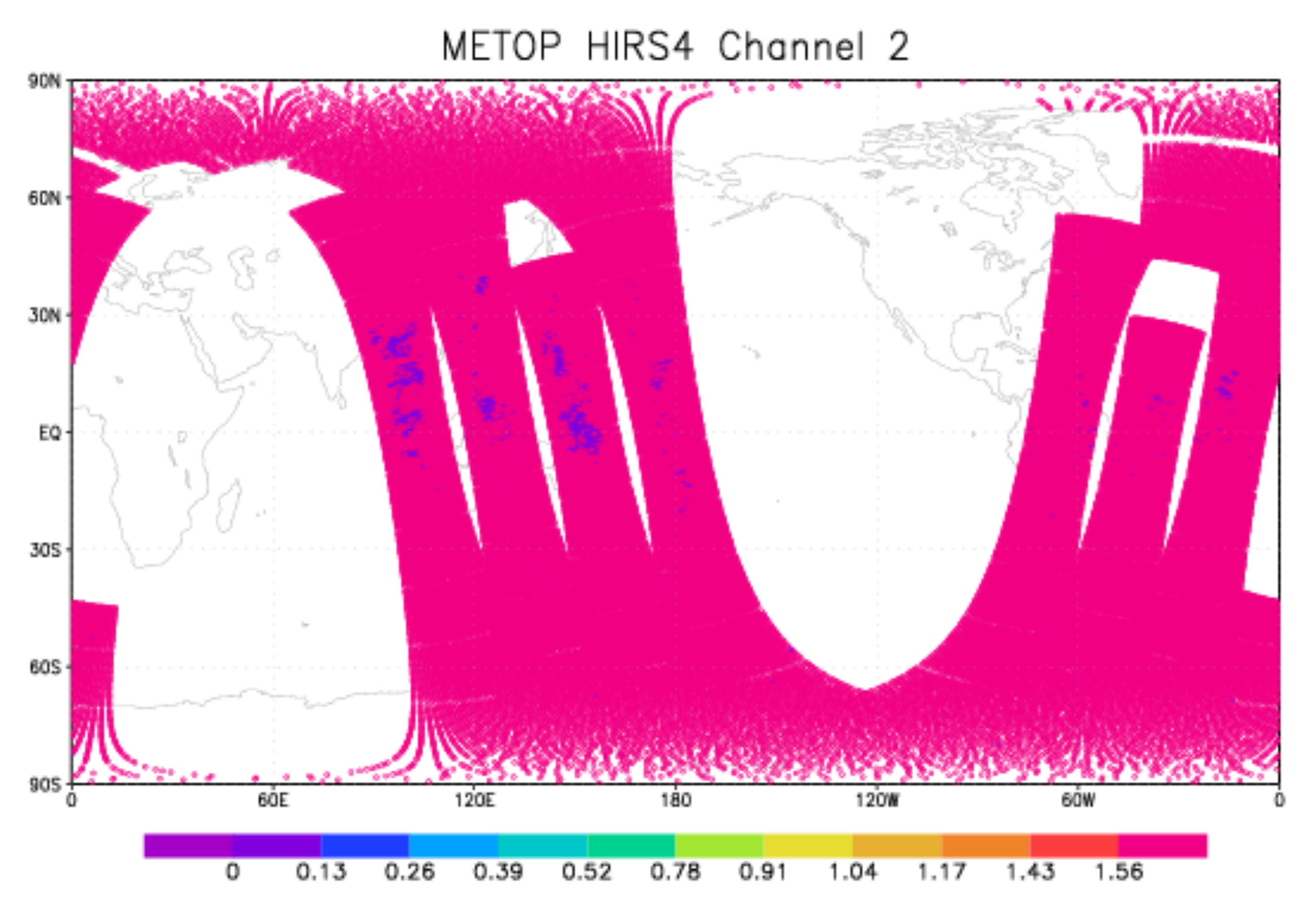
NORTHERN CANADA



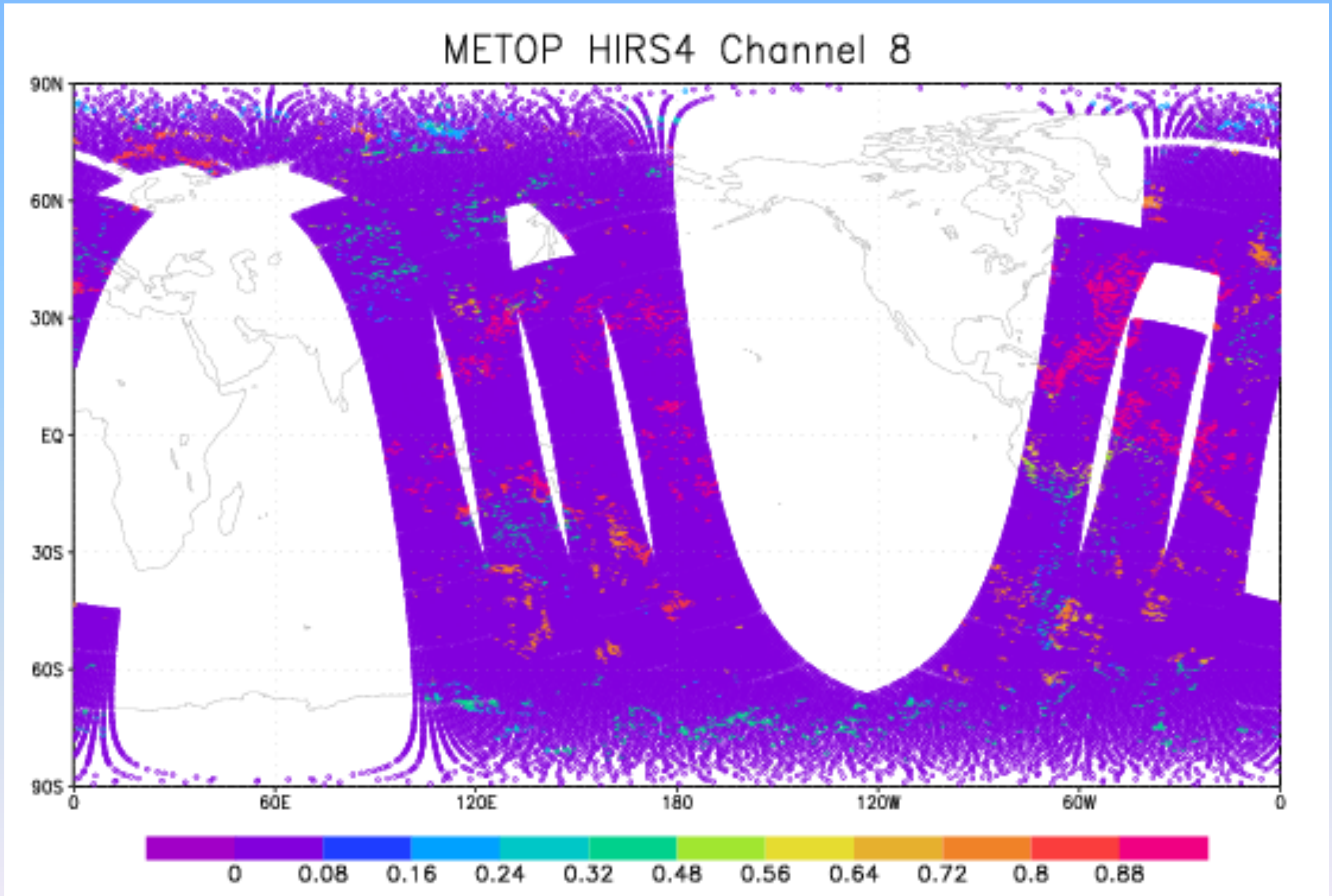
NEGATIVE IS IMPROVEMENT

Quality control procedures

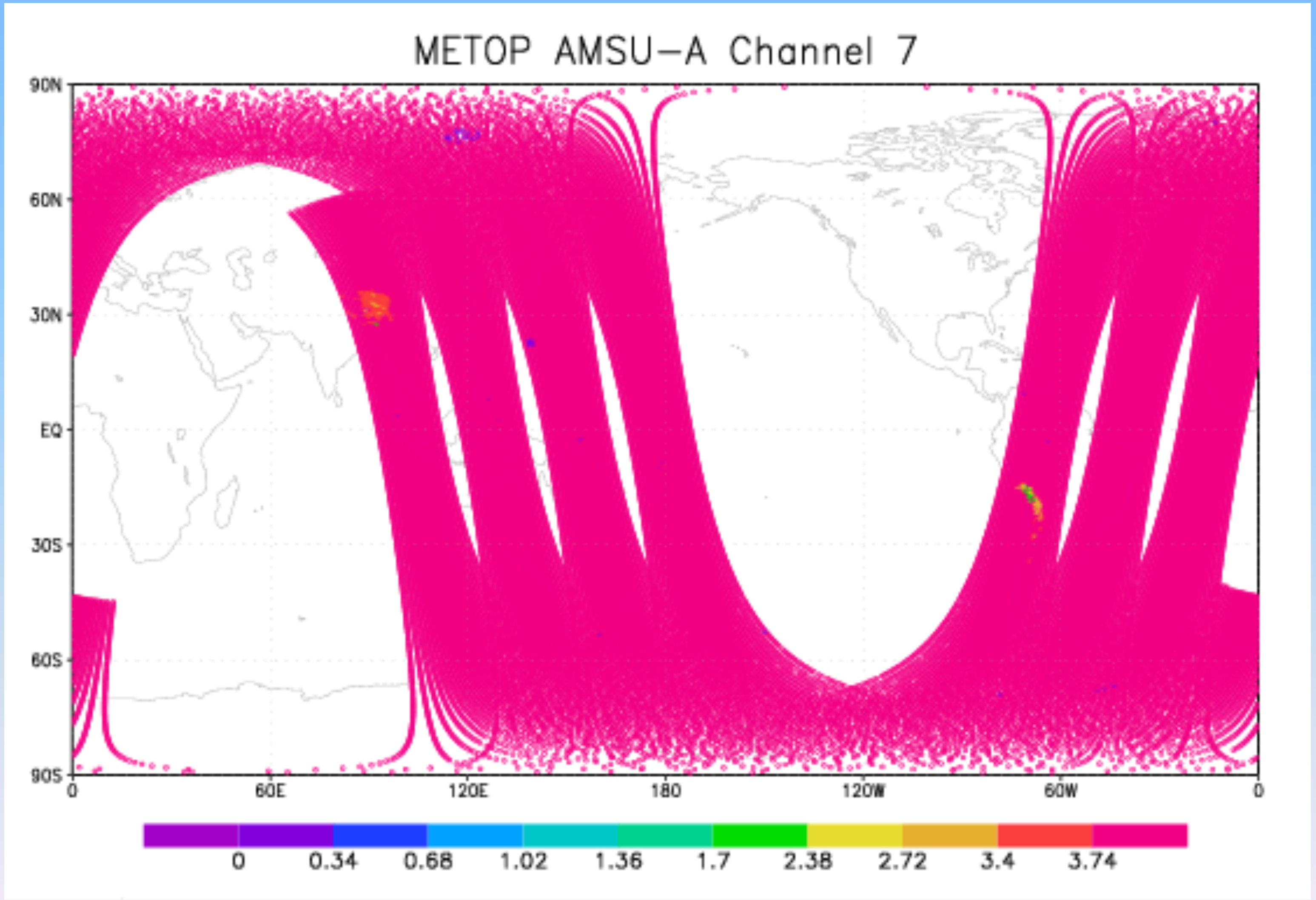
- The quality control step may be the most important aspect of data assimilation
- Data must be removed which has gross errors or which cannot be properly simulated by forward model
- Most problems with satellite data come from 3 sources
  - Instrument problems
  - Clouds and precipitation simulation errors
  - Surface emissivity simulation errors



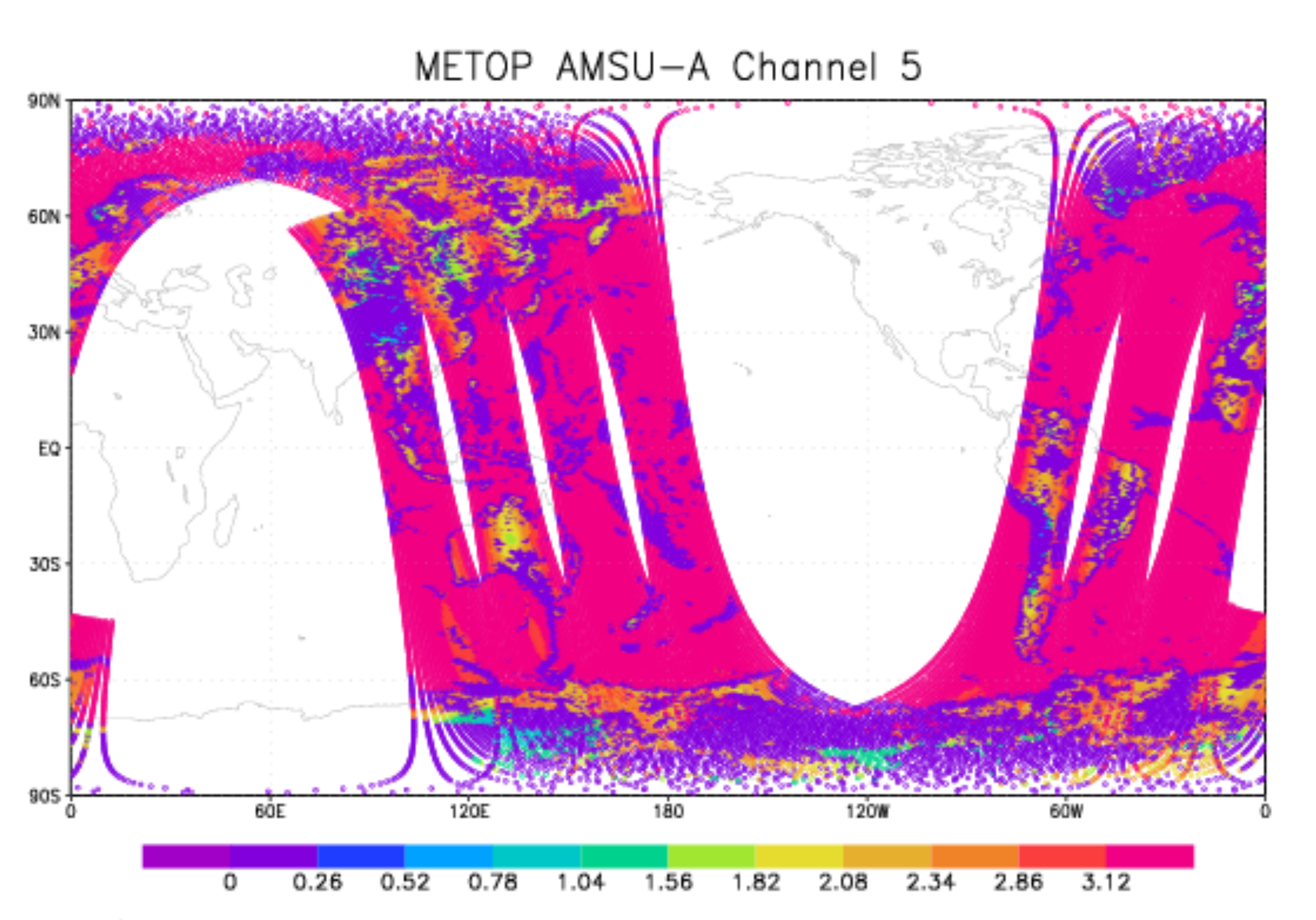
METOP HIRS4 Channel 2
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
0
0.13
0.26
0.39
0.52
0.78
0.91
1.04
1.17
1.43
1.56



METOP HIRS4 Channel 8
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
0
0.08
0.16
0.24
0.32
0.48
0.56
0.64
0.72
0.8
0.88



METOP AMSU-A Channel 7
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
0
0.34
0.68
1.02
1.36
1.7
2.38
2.72
3.4
3.74



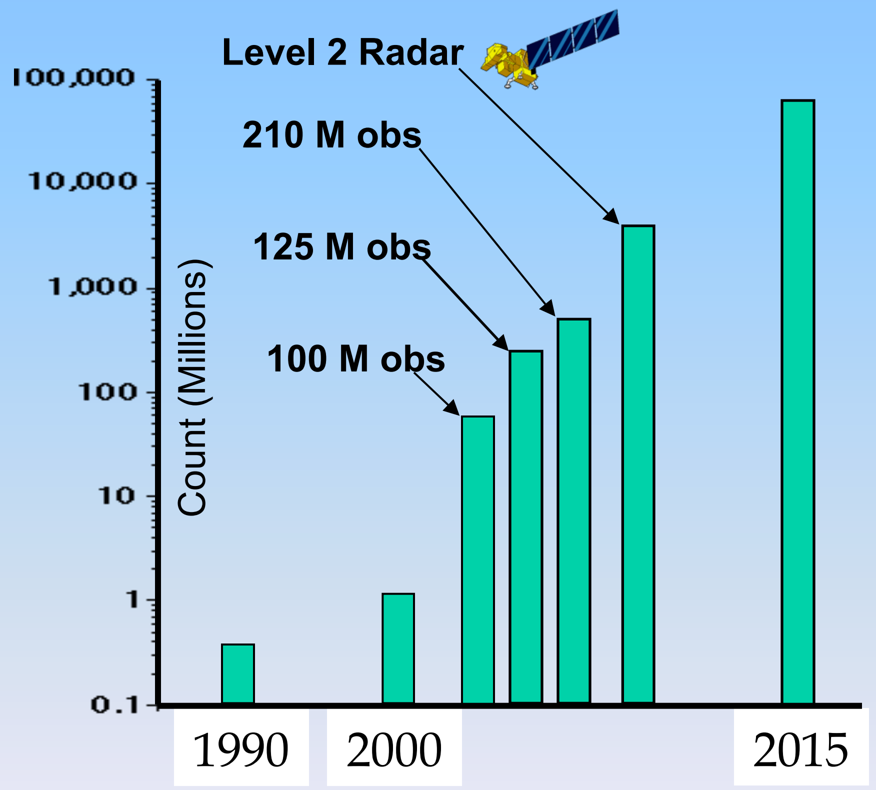
METOP AMSU-A Channel 5
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
0
0.26
0.52
0.78
1.04
1.56
1.82
2.08
2.34
2.86
3.12

Quality control procedures (thinning)

- Some data is thinned prior to using
- Three reasons
  - Redundancy in data
    - Radiances
    - AMWs (feature - tracked winds)
  - Reduce correlated error
    - AMWs
  - Computational expense
    - Radiances

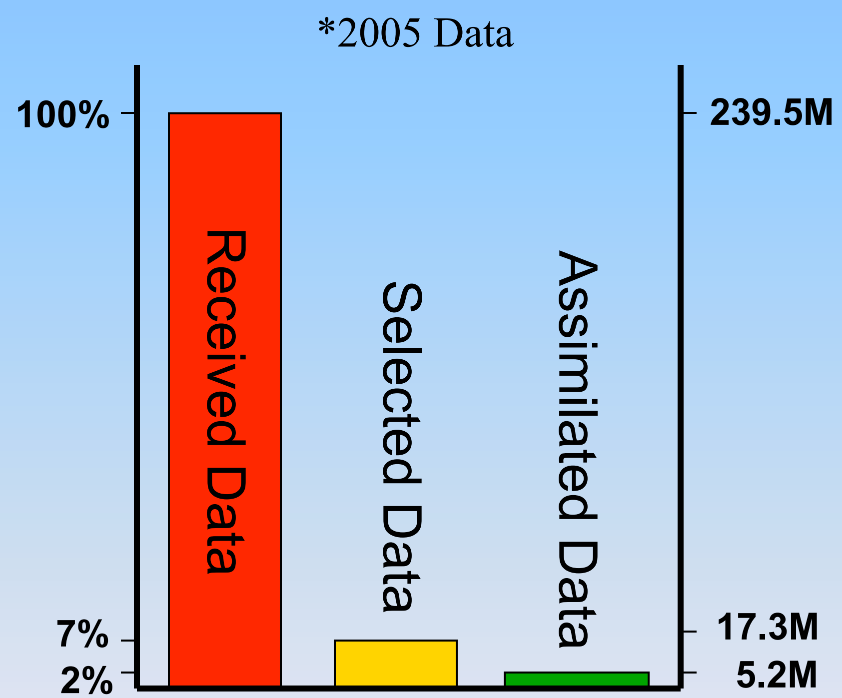
Satellite Data Ingest

Daily Satellite & Radar Observation Count



Five Order of Magnitude Increases in Satellite Data Over Fifteen Years (2000-2015)

Daily Percentage of Data Ingested into Models



Received = All observations received operationally from providers
Selected = Observations selected as suitable for use
Assimilated = Observations actually used by models

Observational errors

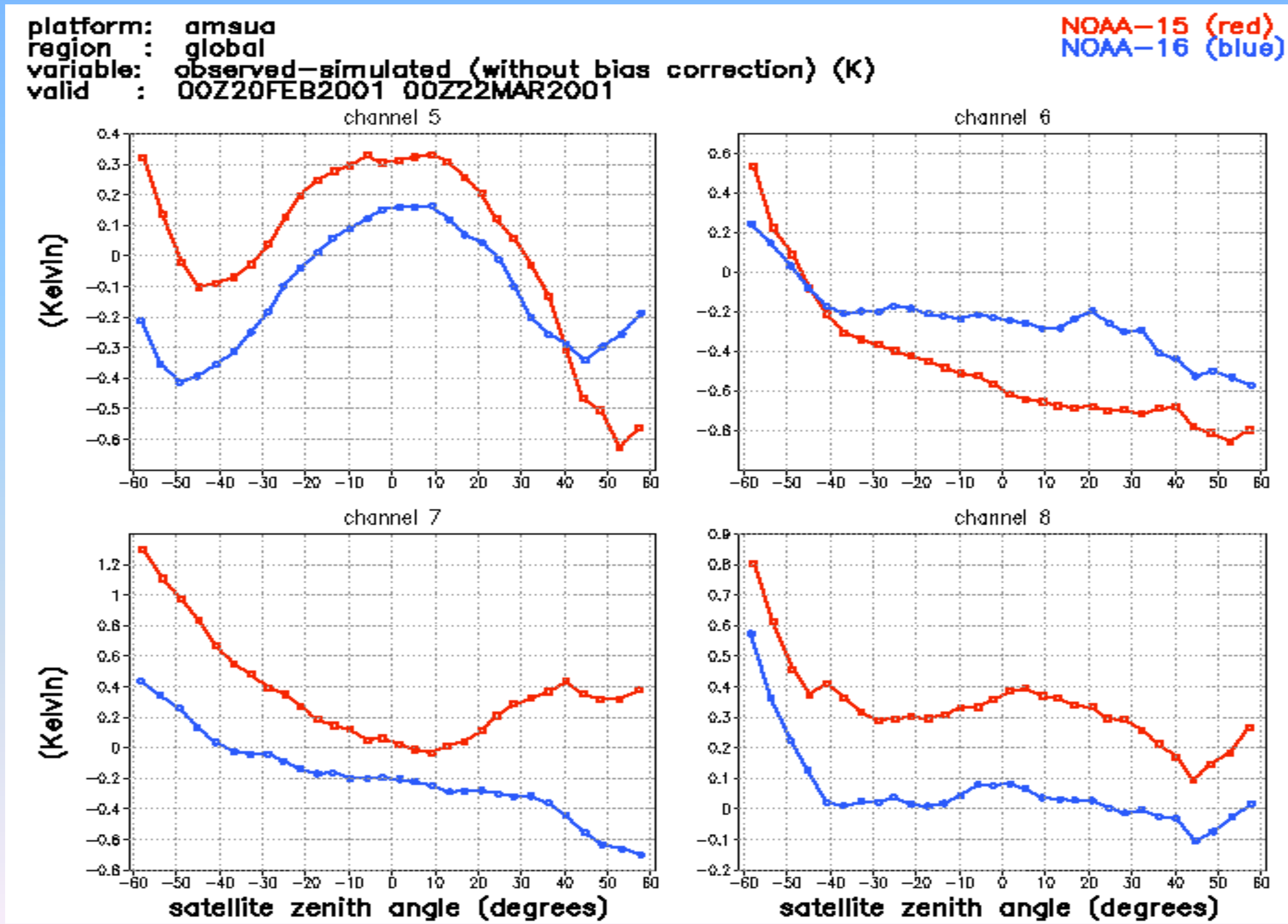
- Observation errors specified based on instrument errors and o-b statistics.
- Generally for satellite data errors are specified a bit large since the correlated errors are not well known.
- Bias must be accounted for since it is often larger than signal
- The source of the bias can come from
  - Biased observations
  - Inadequacies in the characterization of the instruments
  - Deficiencies in the forward models
  - Biases in the background

Satellite observations

- Different observation and error characteristics
  - Type of data (cloud track winds, radiances, etc.)
  - Version of instrument type (e.g., IR sounders - AIRS, HIRS, IASI, GOES, GIFTS, etc.)
  - Different models of same instrument (e.g., NOAA-15 AMSU-A, NOAA-16 AMSU-A)

Bias Correction

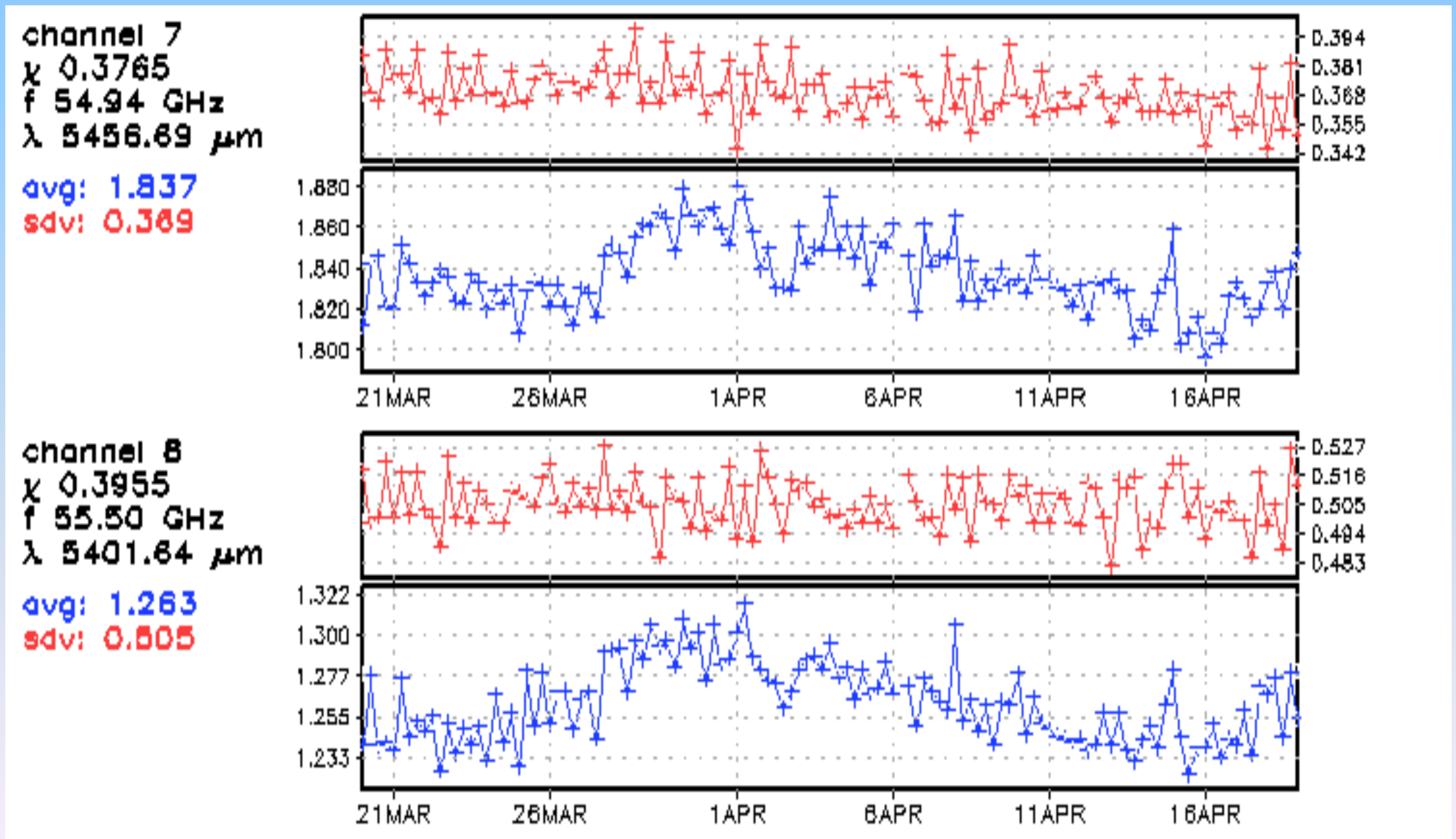
- Currently we are only bias correcting, the radiances and the radiosonde data (radiation correction)
- For radiances, biases can be much larger than signal. Essential to bias correct the data
- NCEP uses a 2 step process for radiances (others are similar)
  - Angle correction (very slowly evolving – different correction for each scan position)
  - Air Mass correction (slowly evolving based on predictors)


platform: amsua
region : global
variable: observed-simulated (without bias correction) (K)
valid : 00Z20FEB2001 00Z22MAR2001
NOAA-15 (red)
NOAA-16 (blue)
channel 5
channel 6
channel 7
channel 8
(Kelvin)
(Kelvin)
satellite zenith angle (degrees)
satellite zenith angle (degrees)

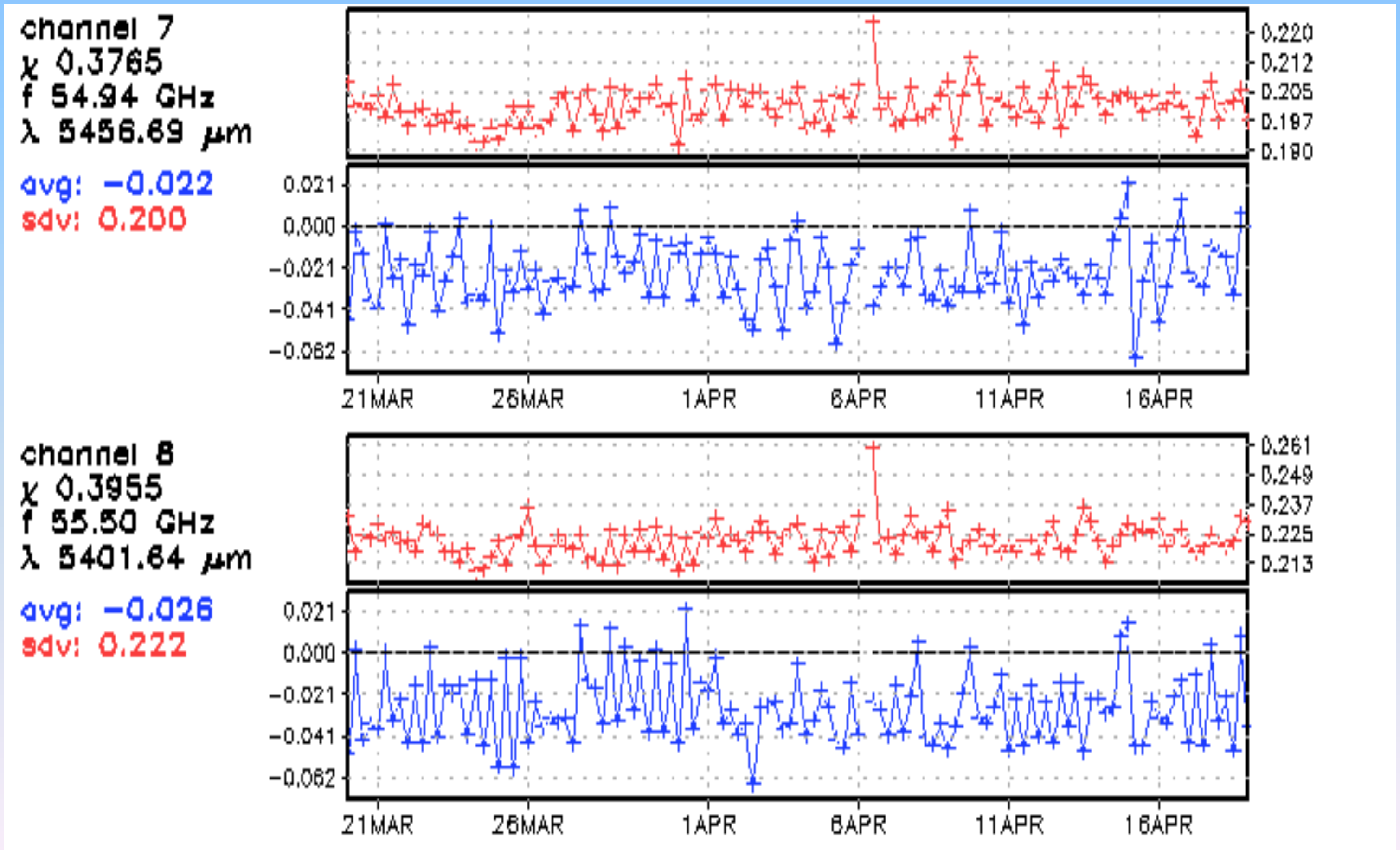
Satellite radiance observations Bias correction

- Air Mass prediction equation for bias
  - Coefficients in equation analysis variable w/ background (previous analysis) values
  - Predictors
    - mean
    - path length (local zenith angle determined)
    - integrated lapse rate
    - integrated lapse rate ** 2
    - cloud liquid water

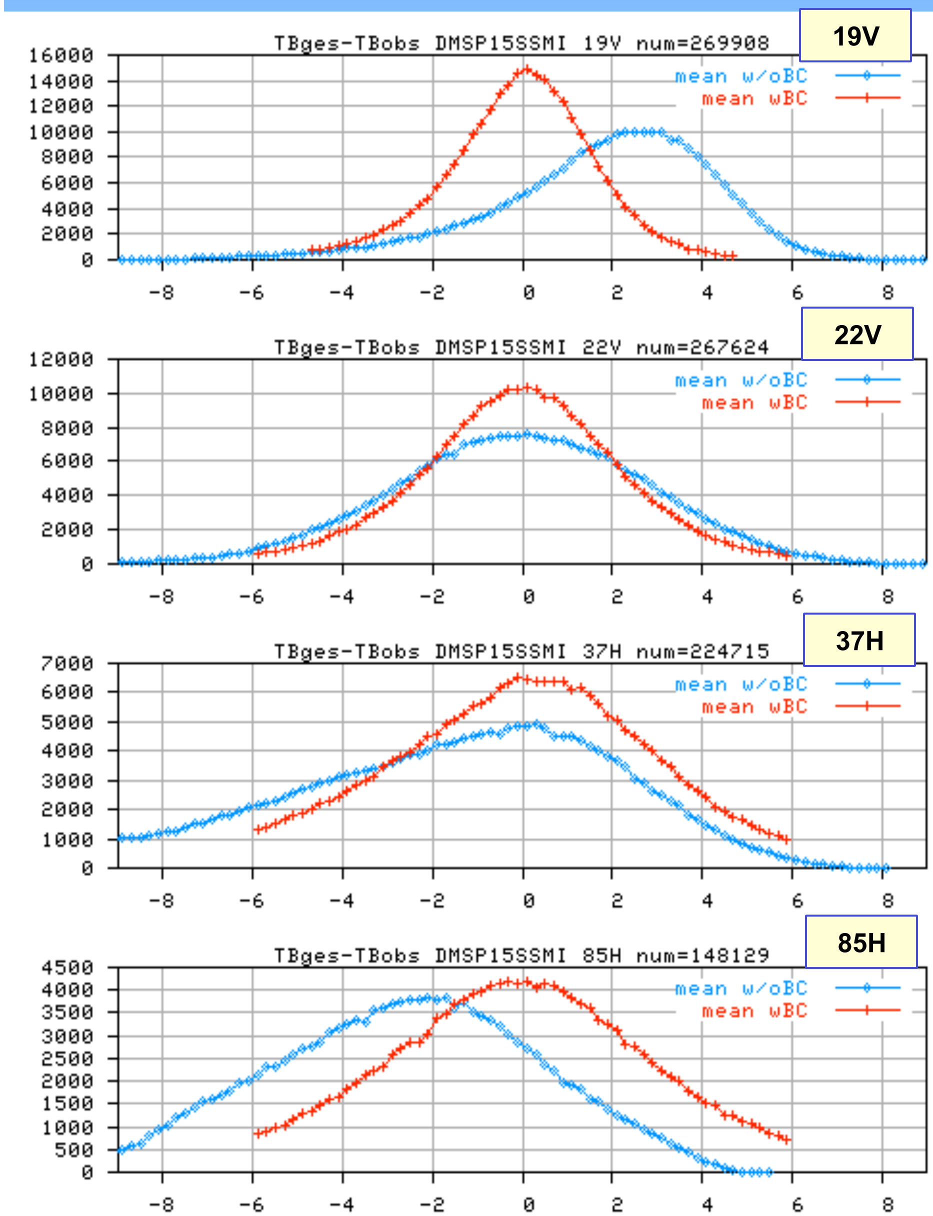
NOAA 18 AMSU-A
No Bias Correction

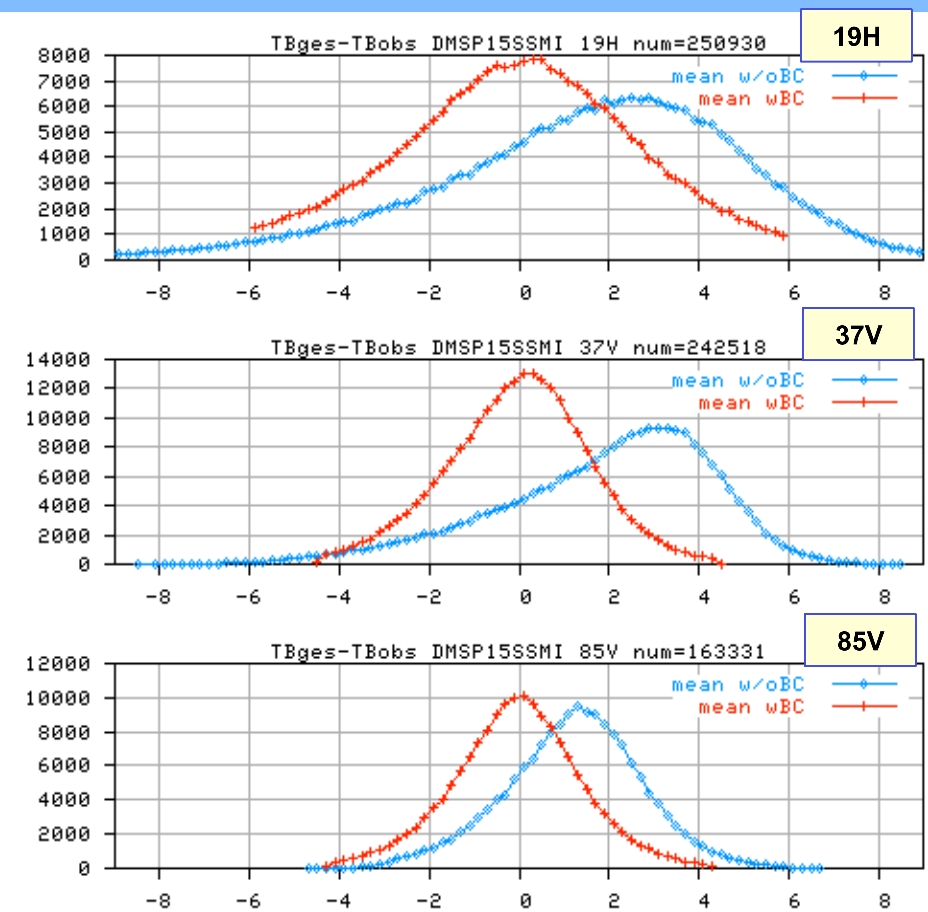


NOAA 18 AMSU-A
Bias Corrected



G-O histogram





DMSP15 July2004 : 1month

before bias correction

after bias correction

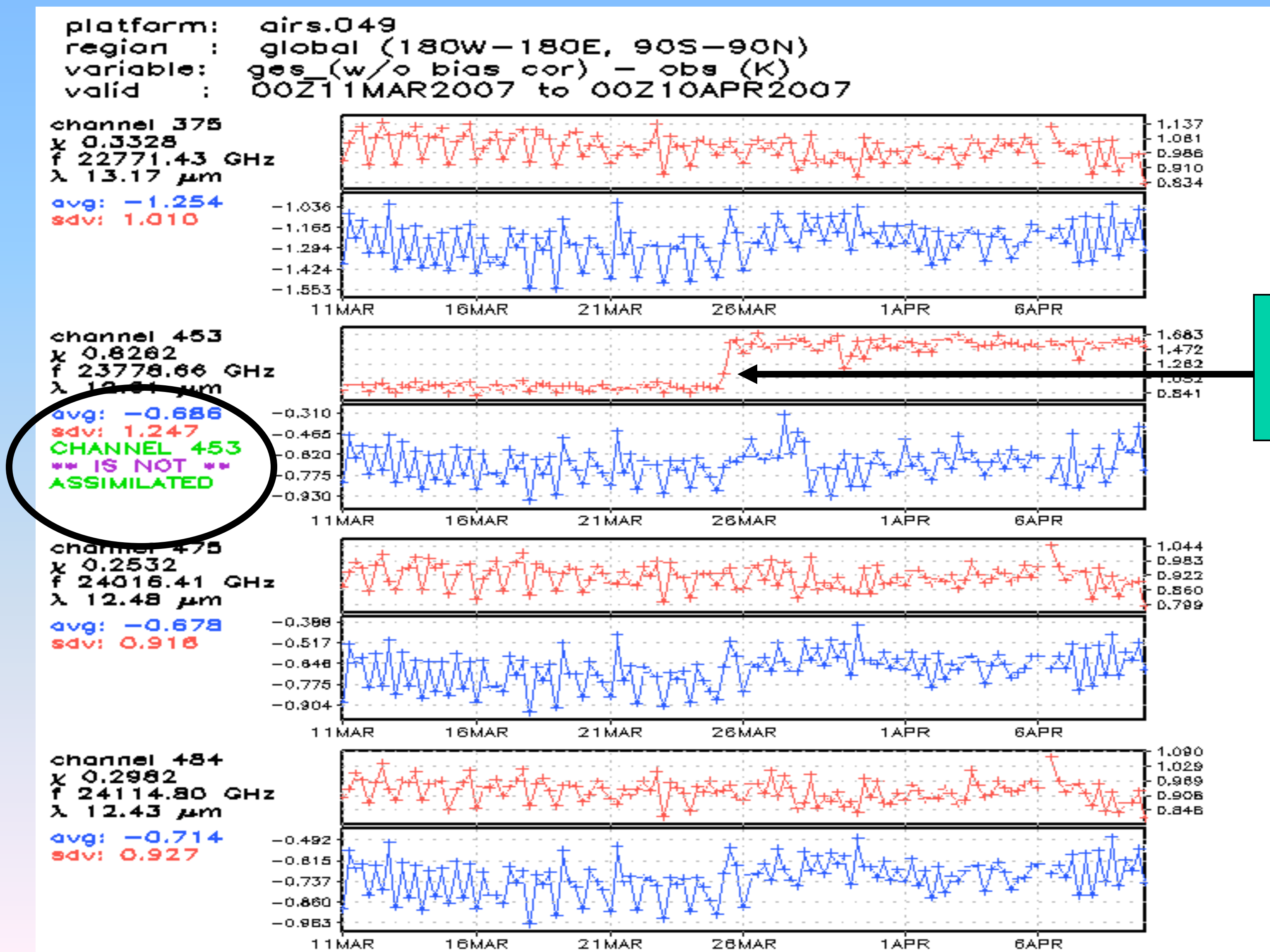
gl use 2004070200-073118(63cases)

Data Monitoring

- It is essential to have good data monitoring.
- Usually the NWP centres see problems with instruments prior to notification by provider (Met Office especially)
- The data monitoring can also show problems with the assimilation systems
- Needs to be ongoing/real time

Quality Monitoring of Satellite Data

AIRS Channel 453 26 March 2007



Increase in SD Fits to Guess

Data impact

- Much of the improvement in forecast skill can be attributed to the improved data and the improved use of the data
- Must be measured relative to rest of observing system – not as stand alone data sets
- Extremely important for planning ($$$$)

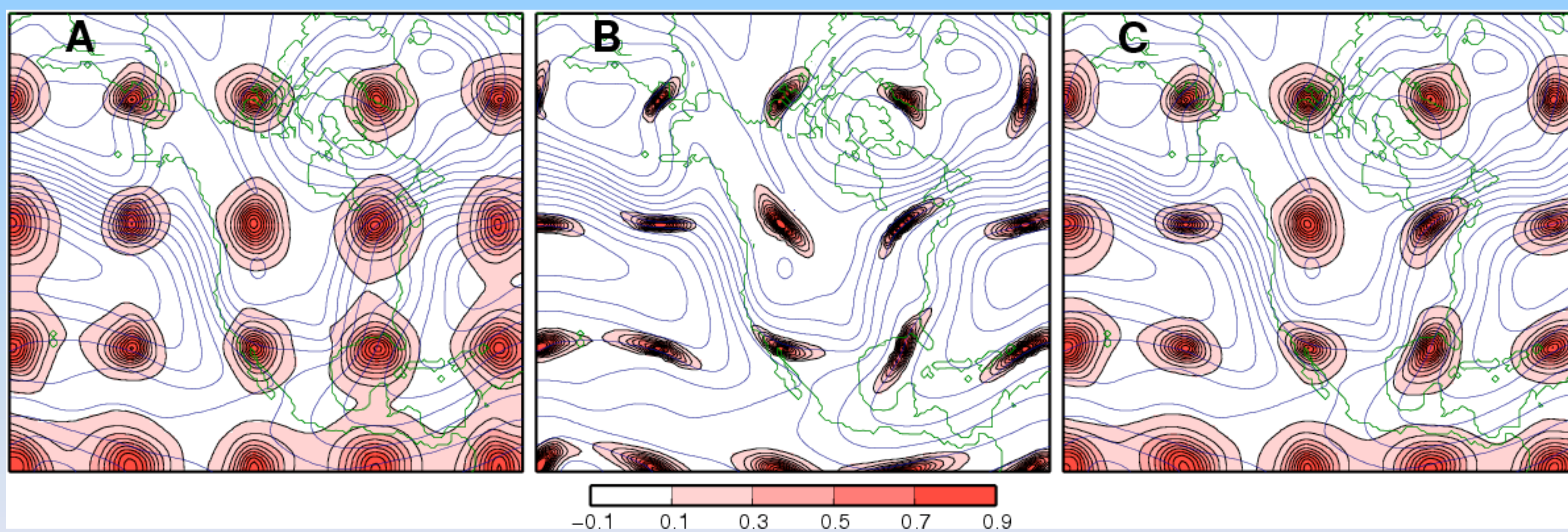
Constraint term

- Includes weak (inexact) constraints
- Only one currently used
  - Moisture ($0 < q \sim< q_{sat}$)
- Could include non-negative constraints on cloud liquid water and ozone.

Data assimilation plans

- Improve Assimilation Techniques
  - – Merger with GMAO code
  - – Situation Dependent Background Errors
  - – Inclusion of 4d-var
    - Increment model/adjoint developed
    - Testing has begun
  - – Hybrid 4d-var – KF to improve definition of background error
- Analyze additional variables
  - – Clouds and Precipitation
  - – Aerosols
  - – Atmospheric constituents

Situation Dependent Background Errors



Data assimilation plans

- Use new data
  - NOAA-19
  - NPOESS – NPP – GOES-R
  - GRAS – GRACE – GPS RO (Lidia)
  - Hurricane central pressure estimat
  - Level 2 Doppler radar (refractivity and radial velocities)
  - Mesonets
  - Etc.
- Use current data better
  - Improve RT
  - Account for addition factors in forward models
    - Slant path
    - Atmospheric boundary layer
    - Cloudy Radiances
    - Etc.
  - Improve QC

Final Comments

- The details of the assimilation system determine the quality of the assimilation system.
- The forecast models are a part of the assimilation system and any improvement in the forecast models should translate into improved analyses and assimilation
- Anticipate large expansion in scope of data assimilation
  - Assimilation of clouds
  - Assimilation of aerosols
  - Assimilation of atmospheric constituents
  - Ocean, land surface, ice surface, ecosystem assimilation