Catching up to the world:
The GMAO 4DVAR and its Adjoint-based Tools

Ricardo Todling & Yannick Trémolet*

Global Modeling and Assimilation Office, NASA

Mathematical Advancement in Geophysical Data Assimilation
Banff International Research Station, Canada, Feb 2008

* While on leave from ECMWF

Contacts:
Ricardo.Todling@nasa.gov
Yannick.Tremolet@ecmwf.int

Outline

- Overview
- Variational Data Assimilation
- Data Assimilation Illustrations:
  - Replacement of GSI Conjugate Gradient
  - Quick Eval: IAU, FGAT, and 4DVAR
- Forecast/Observation Sensitivities
- Observation Sensitivity Illustrations
  - Example in static analysis mode
  - Impacts on the 24-forecasts
- Summary

Four-dimensional Variational Approach

The general cost function of the variational formulation

$$
\begin{aligned}
J(\mathrm{x}) &= \frac{1}{2}(\mathrm{x}_0 - \mathrm{x}^b)^T \mathrm{B}^{-1}(\mathrm{x}_0 - \mathrm{x}^b) + J_c \\
&+ \frac{1}{2}\sum_{k=0}^{K}[\mathrm{h}(\mathrm{x}_k) - \mathrm{y}_k]^T \mathrm{R}_k^{-1}[\mathrm{h}(\mathrm{x}_k) - \mathrm{y}_k] \\
&+ \frac{1}{2}\sum_{k=1}^{K}[\mathrm{m}(\mathrm{x}_k) - \mathrm{x}_k]^T \mathrm{Q}_k^{-1}[\mathrm{m}(\mathrm{x}_k) - \mathrm{x}_k]
\end{aligned}
$$

where

- ▷ $\mathrm{x} \equiv [\mathrm{x}_0, \mathrm{x}_1, \cdots, \mathrm{x}_K]^T$ is a 4d state vector;
- ▷ h_k and m_k are the nonlinear observation and dynamical model operators, respectively;
- ▷ B, Q_k, and R_k are the background, model, and observation error covariances, respectively.
- ▷ Strong constraint formulation: $\mathrm{Q}_k \to \infty$;
- ▷ Weak constraint formulation, $\mathrm{Q} \neq 0$ accounts for imperfections in the model m;
- ▷ J_c represents a balance constraint.

Incremental Variational Formulation

The minimization of J is generally treated in the Gauss-Newton sense where an iterative procedure linearizes the cost function at each, so called, outer loop, turning the problem in to a quadratic minimization problem for the following function

$$\begin{aligned} J_j \equiv J(\delta \mathbf{x}_j) &= \frac{1}{2}(\delta \mathbf{x}_j - \delta \mathbf{x}_j^b)^T \mathbf{B}^{-1} (\delta \mathbf{x}_j - \delta \mathbf{x}_j^b) \\ &+ \frac{1}{2}(\mathbf{H}_j \mathbf{x}_j - \mathbf{d}_j)^T \mathbf{R}^{-1} (\mathbf{H}_j \mathbf{x}_j - \mathbf{d}_j) \end{aligned}$$

where $\mathbf{d}_j \equiv \mathbf{y} - \mathbf{h}(\mathbf{x}_{j-1})$, $\delta \mathbf{x}_j^b \equiv \mathbf{x}^b - \mathbf{x}_{j-1}$, and

- $\delta \mathbf{x}_j \equiv \mathbf{x}_j - \mathbf{x}_{j-1}$ is the control variable;
- $\mathbf{R}$ is a 4d matrix combining the matrices $\mathbf{R}_k$ and $\mathbf{Q}_k$;
- The inner loop minimization of J_j can be solved by
  - Conjugate gradient
  - Quasi-Newton (such as L-BFGS)
  - Lanczos
- Conditioning of the J_j minimization is determined by the Hessian $\nabla^2 J_j = \mathbf{B}^{-1} + \mathbf{H}_j^T \mathbf{R}^{-1} \mathbf{H}_j$, which spectrum is such that a good preconditioning is essential, particularly in 4dvar.

Preconditioning in the Variational Formulation

A useful preconditioning redefines the control variable at each outer loop to be $\chi = \mathrm{L}^{-1}\delta\mathrm{x}$, where L is a square-root factor of B, that is,

$$\mathrm{B} = \mathrm{L}\mathrm{L}^T .$$

The transformed cost function becomes

$$\hat{J} = \frac{1}{2}\chi^T\chi + \frac{1}{2}(\mathrm{H}\mathrm{L}\chi - \mathrm{d})^T\mathrm{R}^{-1}(\mathrm{H}\mathrm{L}\chi - \mathrm{d}) ,$$

and its corresponding Hessian

$$\nabla^2\hat{J} = \mathrm{I} + \mathrm{L}^T\mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}\mathrm{L} ,$$

indicates the overall minimization to be much batter conditioned than the original minimization since the smallest eigenvalue of he Hessian is now the unit.

The ideal preconditioning is give by the square-root of the inverse Hessian, $\sqrt{\mathrm{A}}^{-1}$, where $\mathrm{A} = \mathrm{B}^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}$:

- In practice this can be done using the CG-Lanczos connection, where the CG provides the Lanczos vectors of the Hessian;
- The cost of this modified CG is in storing the Lanczos vectors and in the re-orthogonalization needed to avoid degeneracy
- Use of the Lanczos-based CG is thus only justifiable in 4dvar, where fast convergence means avoid the costly integration of the model's TLM and ADM.

GEOS-5 DAS: G5AGCM & GSI

Superstructure: fvSetup, scripts

G5AGCM

Coupler

Analysis

PAQC

AltRTMs

GSI

Infrastructure: ODS, GFIO, Buffer, etc

Base Libraries: HDF, MPI, LAPACK, BLAS, ESMF, etc

Operating System

Main Additions to GEOS-5 DAS

- **Additions to GSI**
  - ESMF coupler interface
  - Observer capability
  - SQRT(B) preconditioning
  - Various minimization schemes (QNewton, L-BFGS, and Lanczos CG)
  - Adjoint GSI
- **Additions to Overall DAS**
  - TL/AD Dynamical Models (Forecast Sensitivity, Singular Vectors)
  - Observation Impact
  - FGAT
  - 4DVAR

GEOS-5 DAS: G5AGCM & GSI

Superstructure: fvSetup, scripts

TLGCM

G5AGCM

ADGCM

Coupler

Analysis

PAQC

AltRTMs

GSI+AD

Infrastructure: ODS, GFIO, Buffer, etc

Base Libraries: HDF, MPI, LAPACK, BLAS, ESMF, etc

Operating System

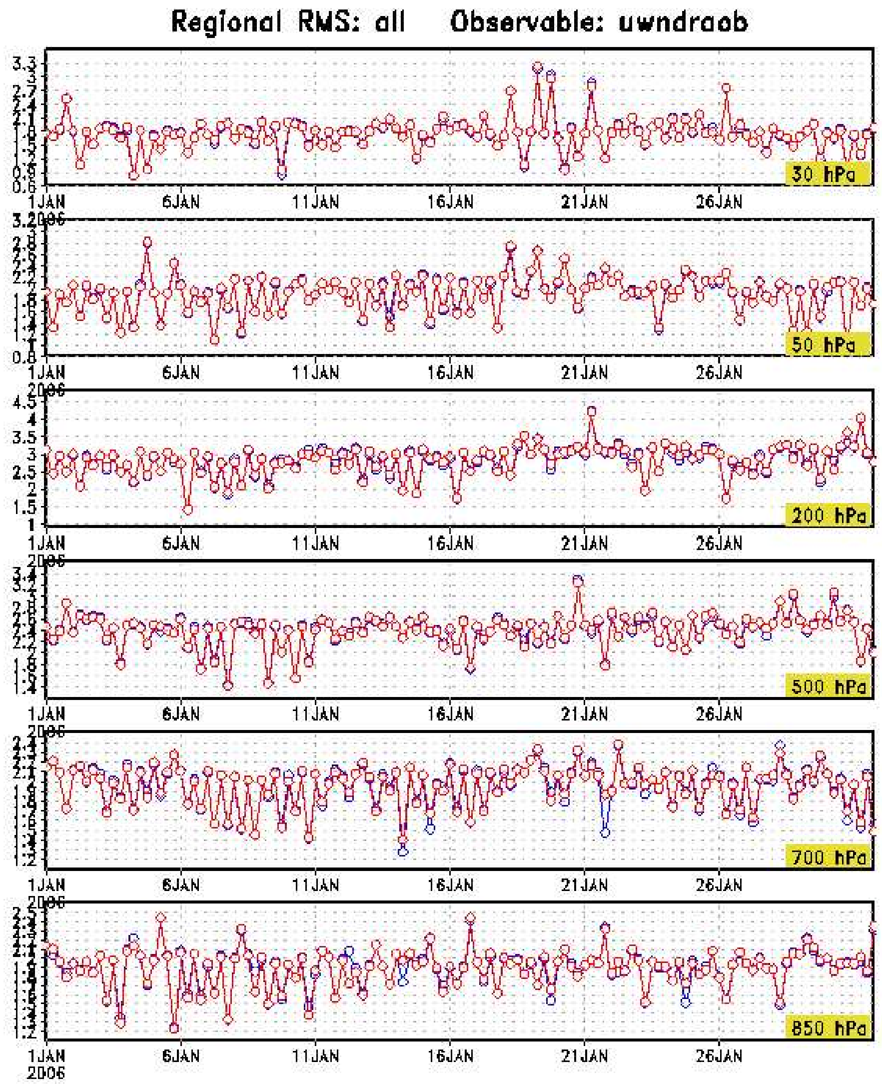
Notes: TL/ADGCM are separate libs from G5GCM
ADGSI is a knob in the GSI library

Illustration 1:
CG Evaluation: *NCEP vs Lanczos*

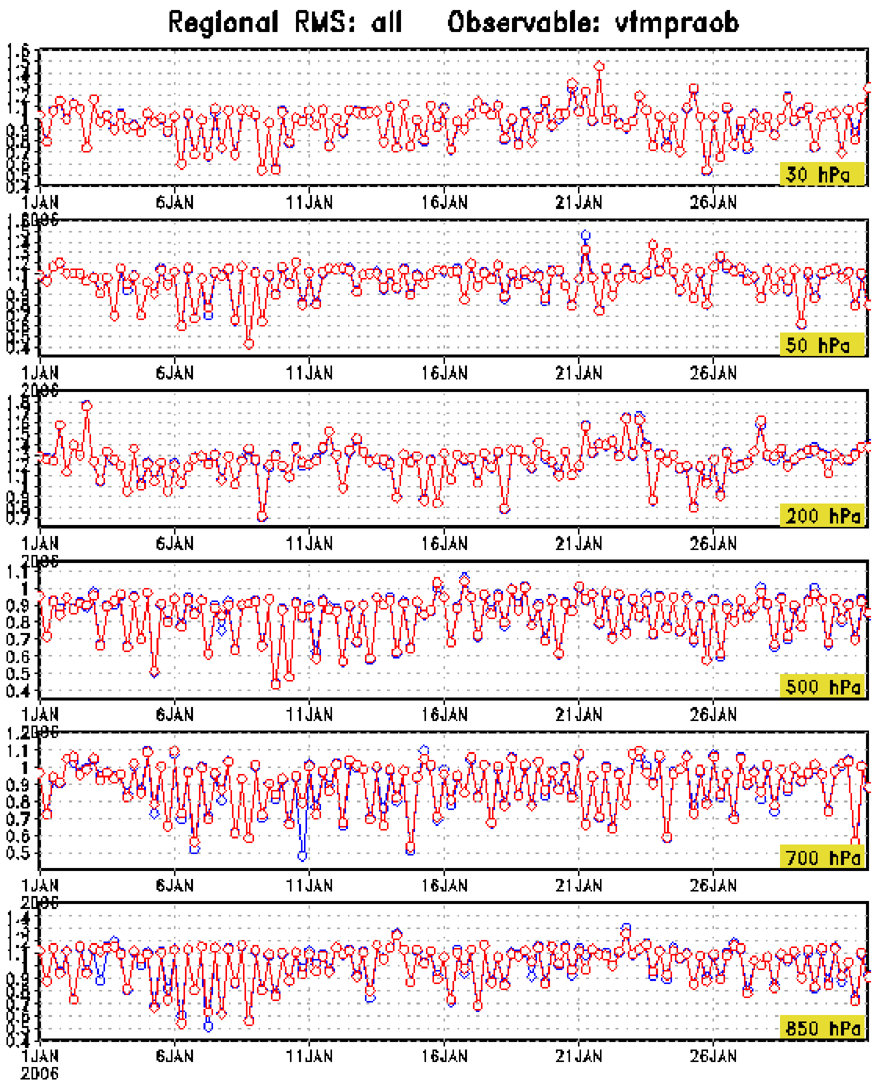
- 2x2.5x72 resolution
- Period: January 2006
- Comparison based on Residual Statistics and Monthly Means
- NCEP CG: 2 outer loops, 100/100 iterations
- Lanczos CG: 2 outer loops, 50/30 iterations

Default NCEP CG vs Lanczos-based CG: O-F time series

Radiosonde U-Wind

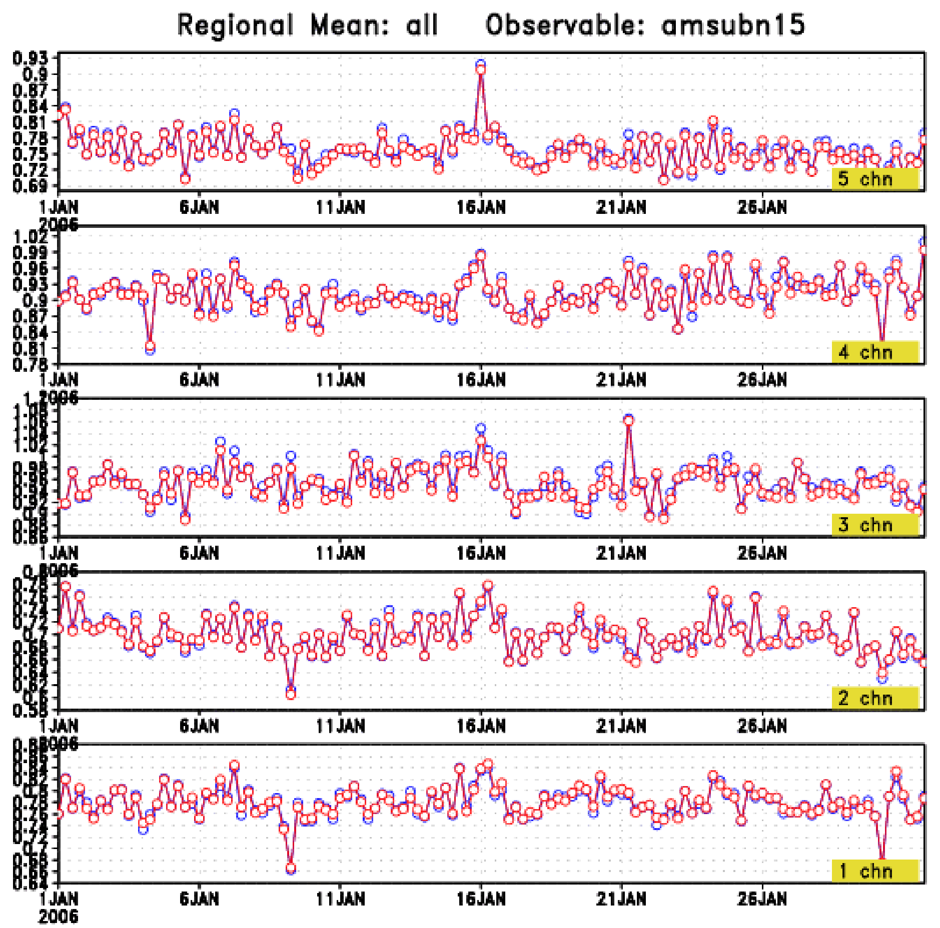


Radiosonde Temperature

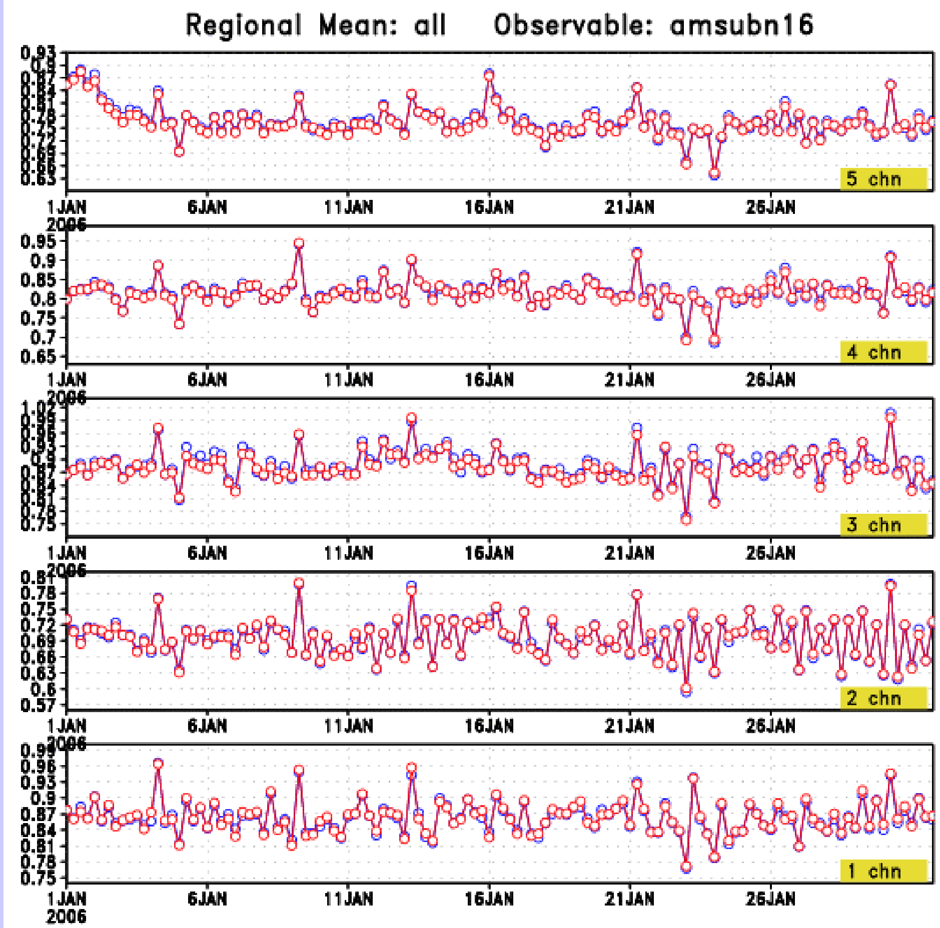


Default NCEP CG vs Lanczos-based CG: Jo/p time series

AMSUB NOAA-15

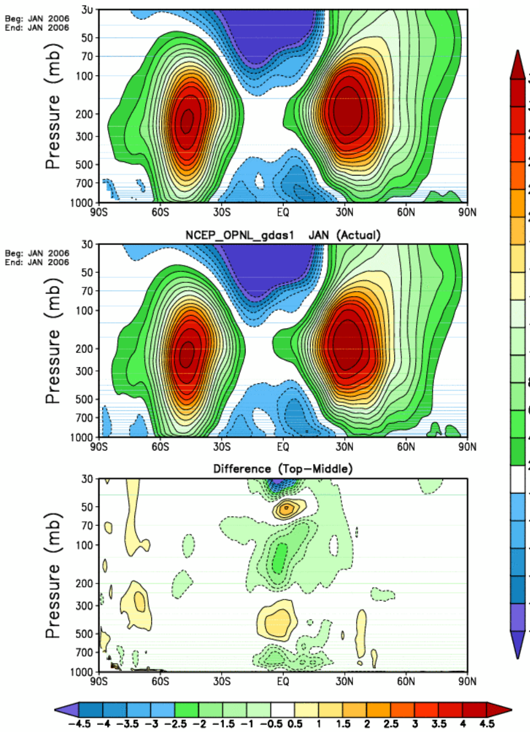


AMSUB NOAA-16

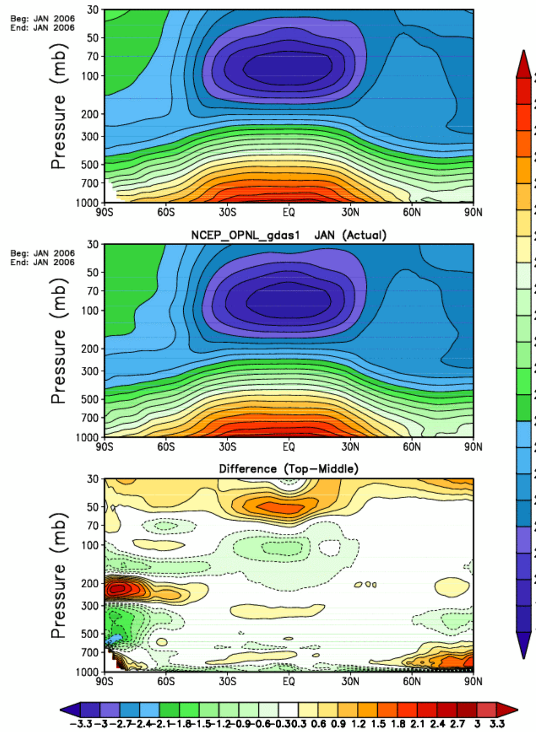


Default NCEP CG Minimization: Zonally Ave Monthly Means

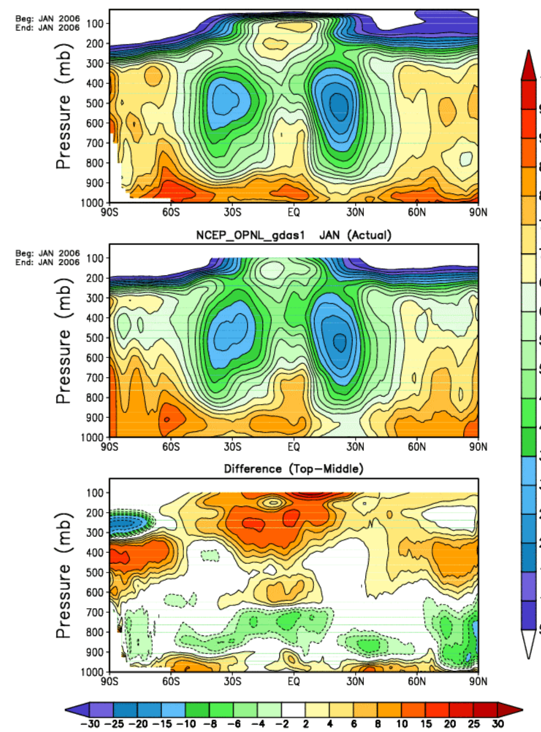
Zonal U-Wind (m/s)



Temperature (K)



Relative Humidity(%)



Lanczos-based CG Minimization: Zonally Ave Monthly Means

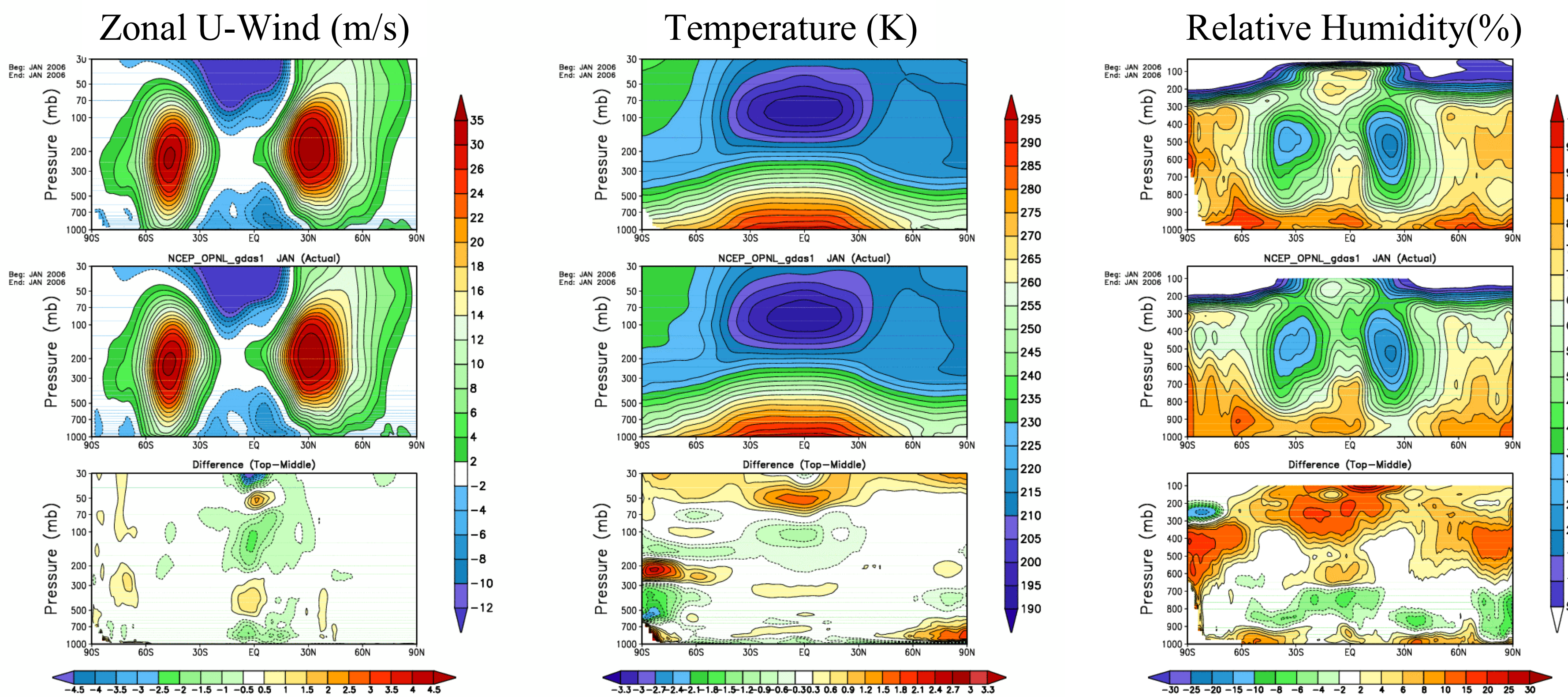


Illustration 1: *Summary*

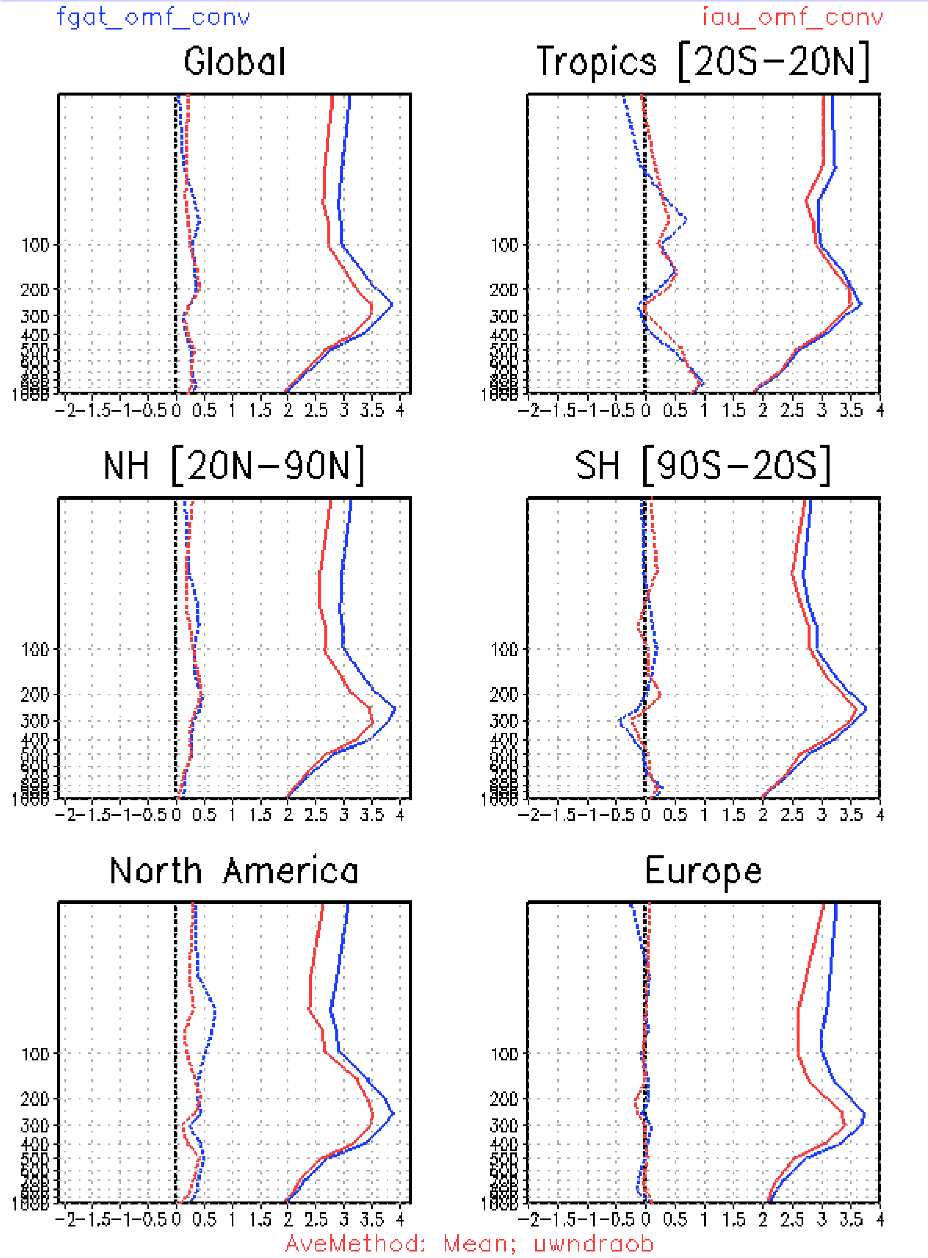
- The objective of this exercise is not to show one CG approach is better than the other, but rather to serve as a check for the implementation of the LCG aimed to be used in 4dvar
- Preliminary results using 70/50 iterations for LCG show indistinguishable results in monthly means and in residual statistics time series (*not* shown here)
- With the more plausible choice of 50/30 iterations results (shown here) remain scientifically reliable
- Impact on forecast skills will be examined soon

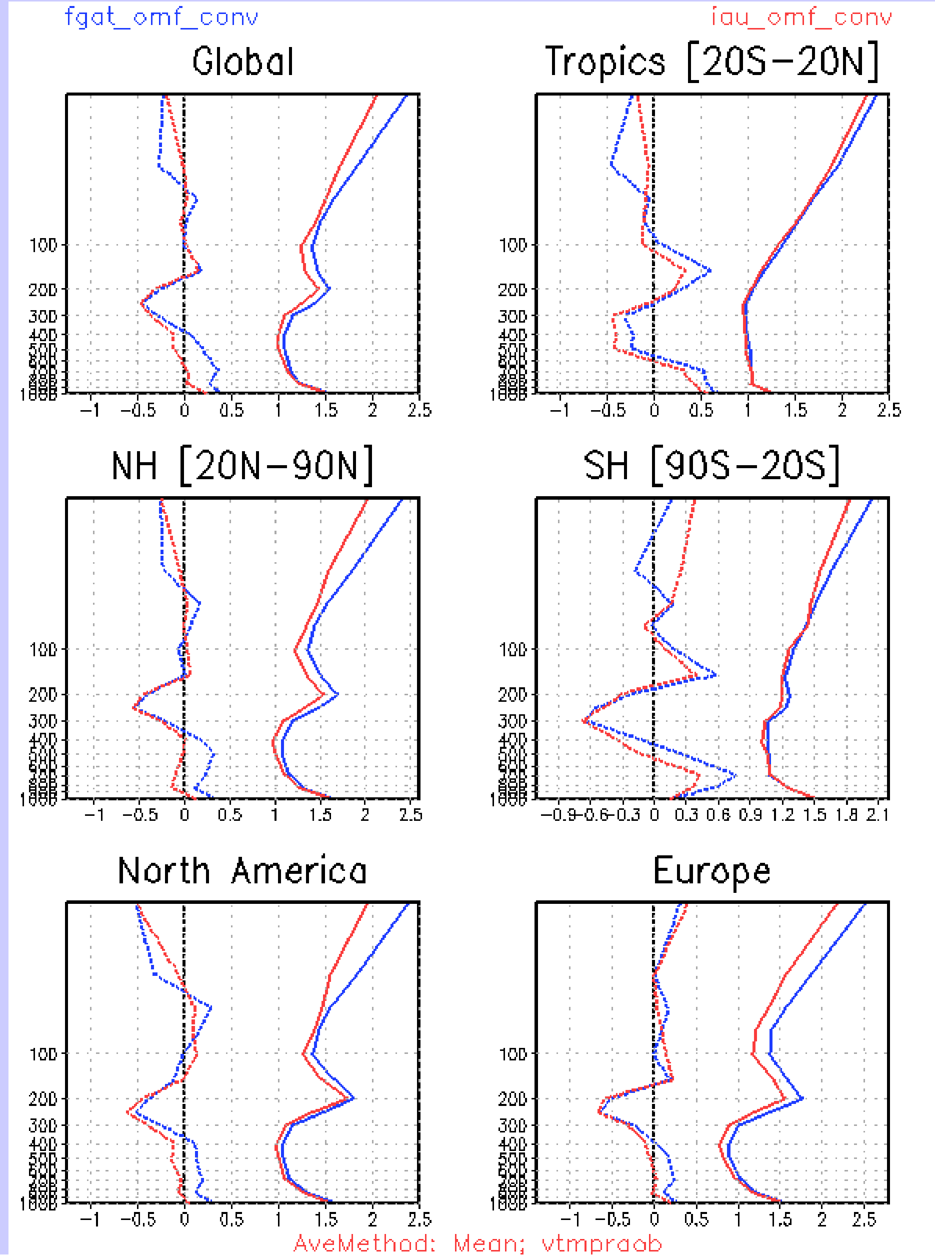
Illustration 2:
IAU, FGAT and 4DVAR in GEOS-5

- 2x2.5x72 resolution
- Period: January 2006
- Comparison based on Monthly Means and Residual Statistics
- All using Lanczos-based CG; 50/30 iterations
- Lanczos CG: 2 outer loops, 50/30 iterations

DAS Comparison: IAU vs FGAT

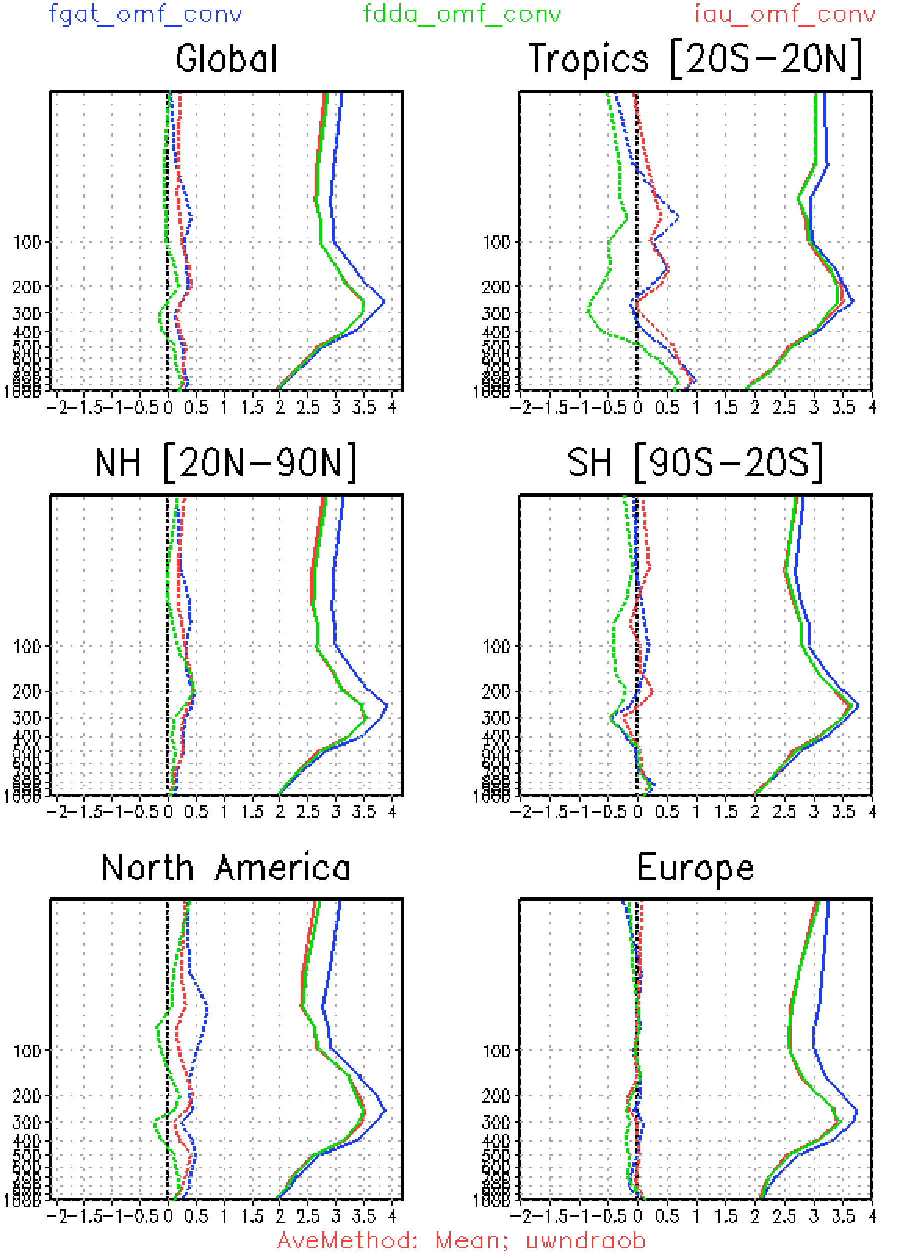
Mean (dashed) & Std Dev for U (left) and T **Raob** O-F

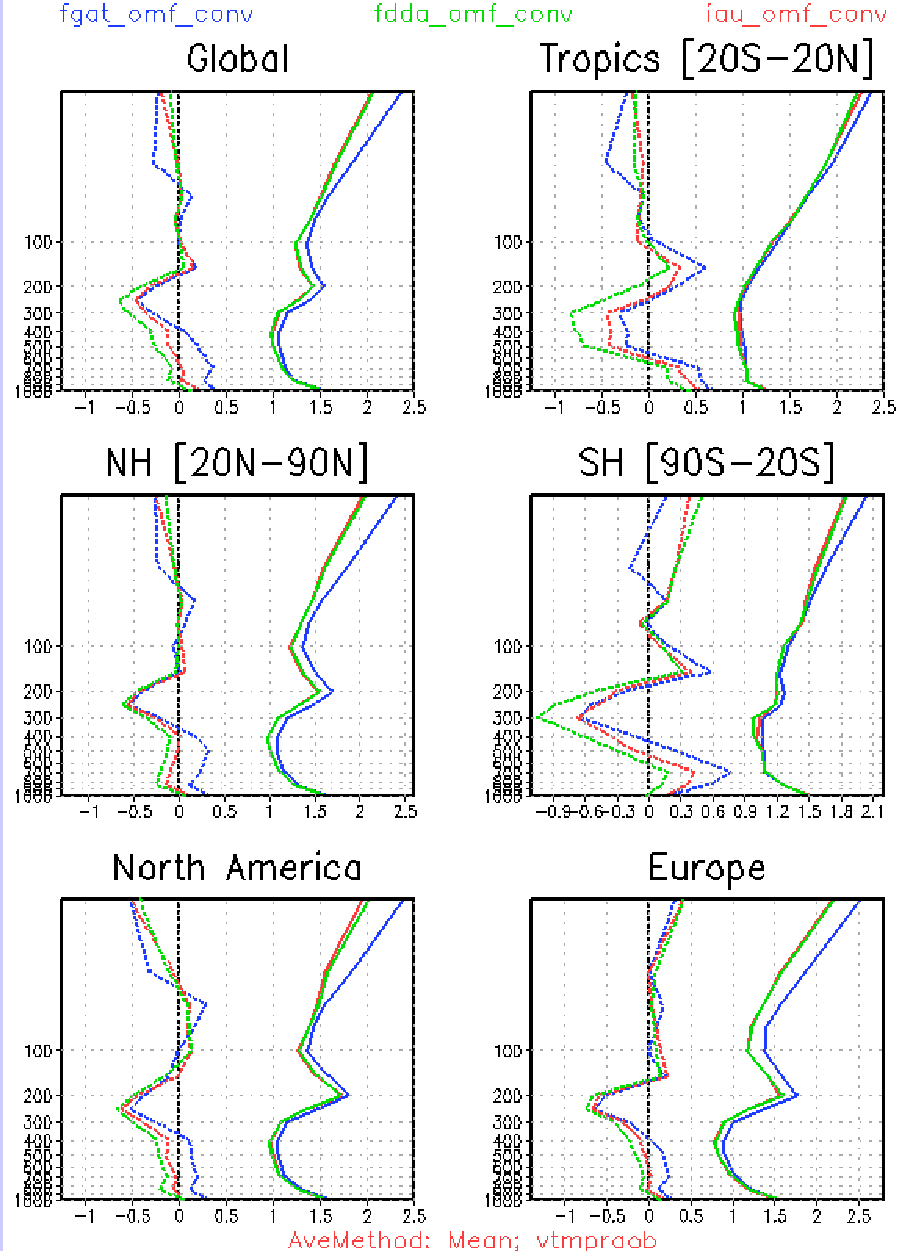




DAS Comparison: IAU vs FGAT vs 4DVAR

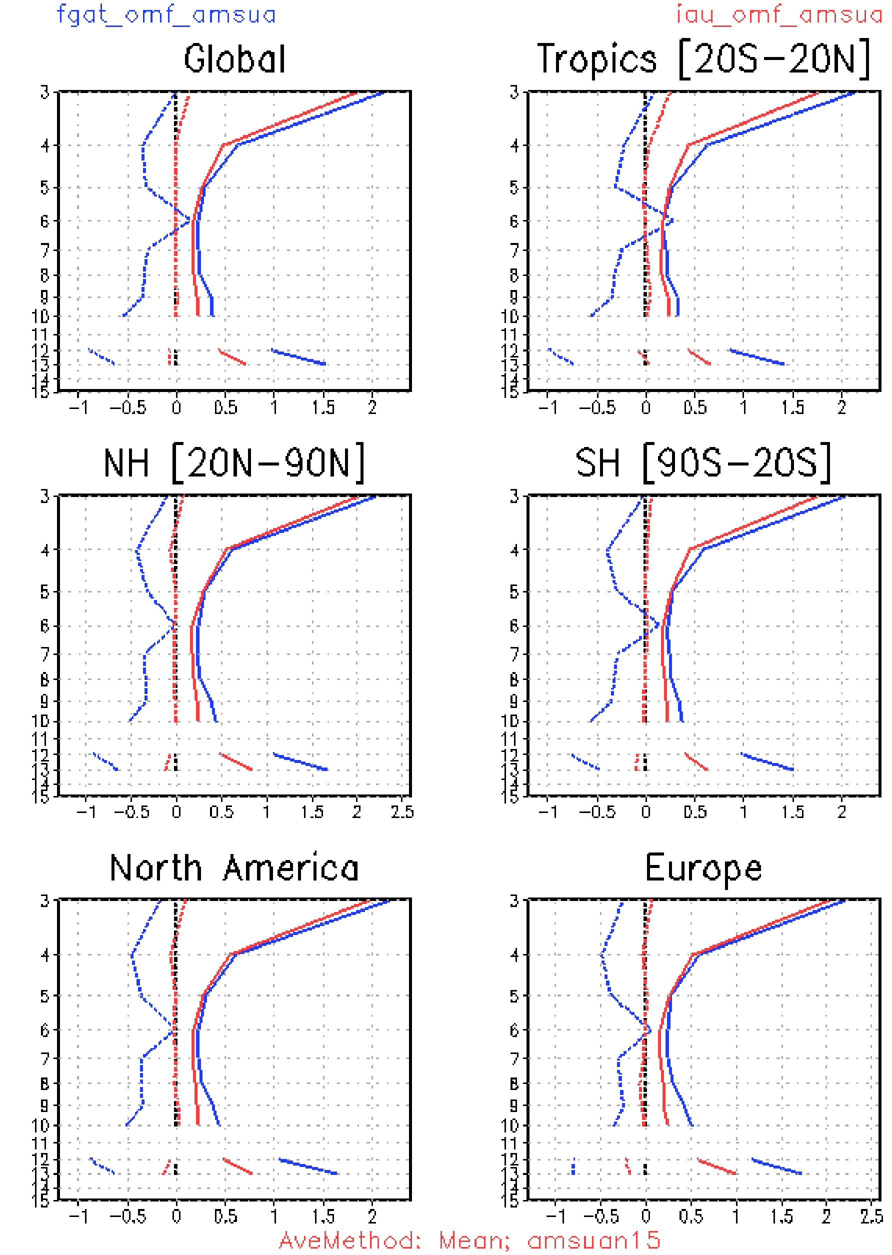
Mean (dashed) & Std Dev for U (left) and T **Raob** O-F

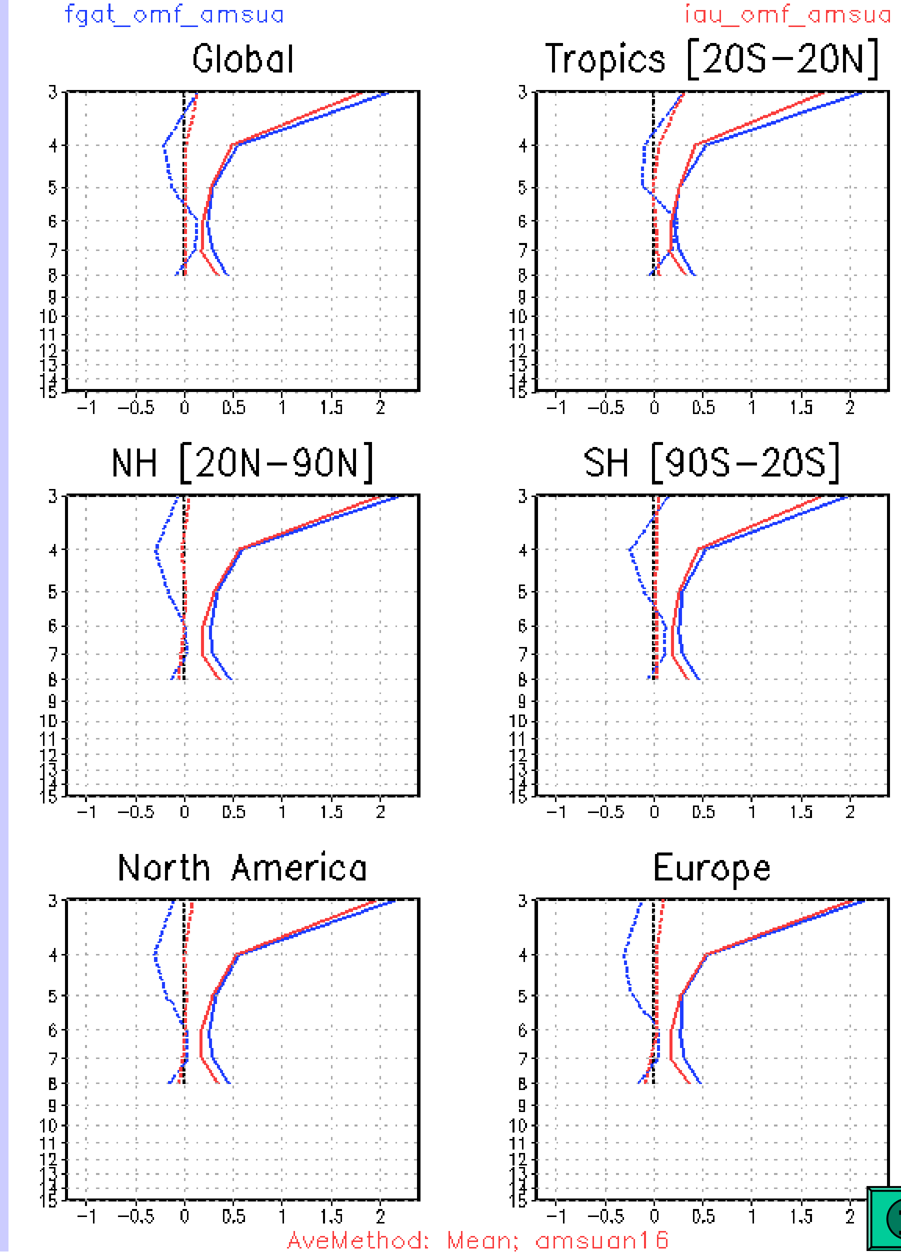




DAS Comparison: IAU vs FGAT

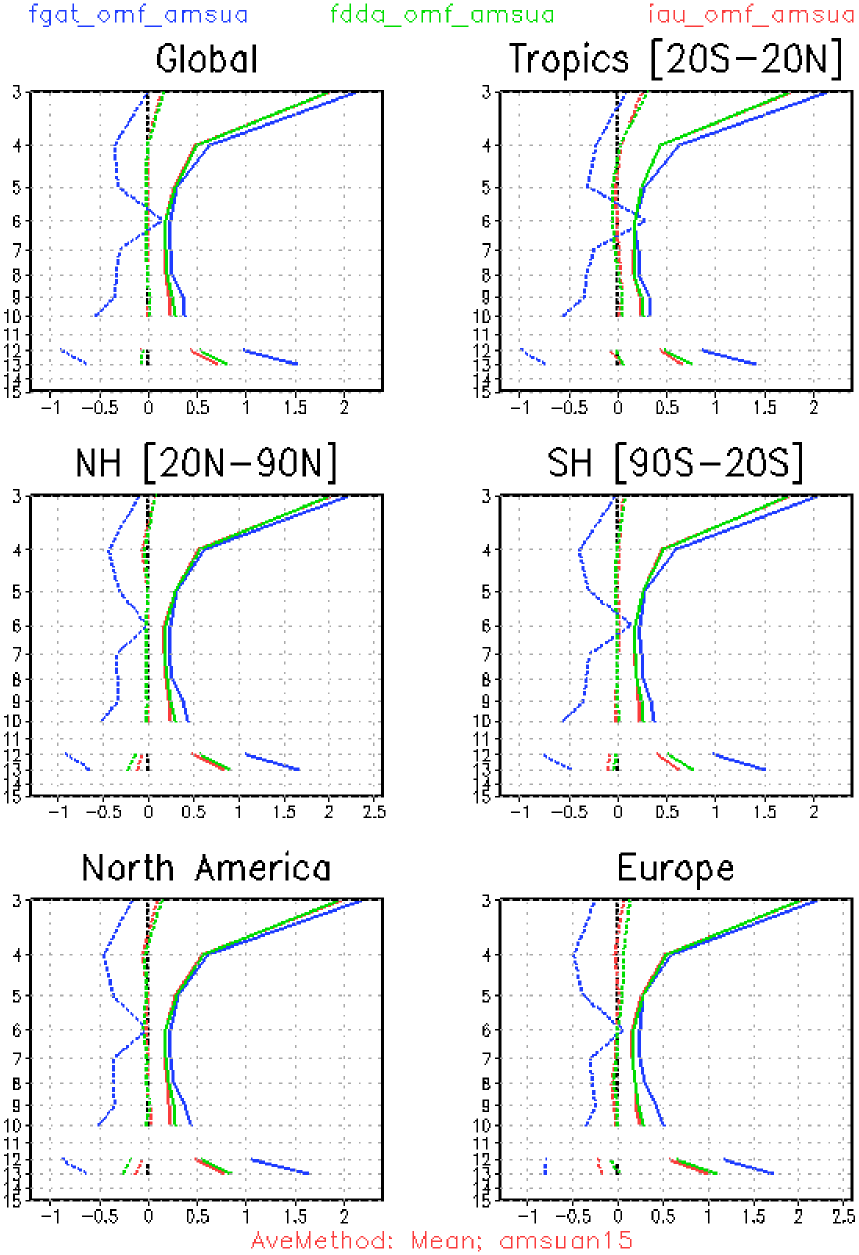
Mean (dashed) & Std Dev for NOAA-15 (left) and NOAA-16 **AMSUA** O-F

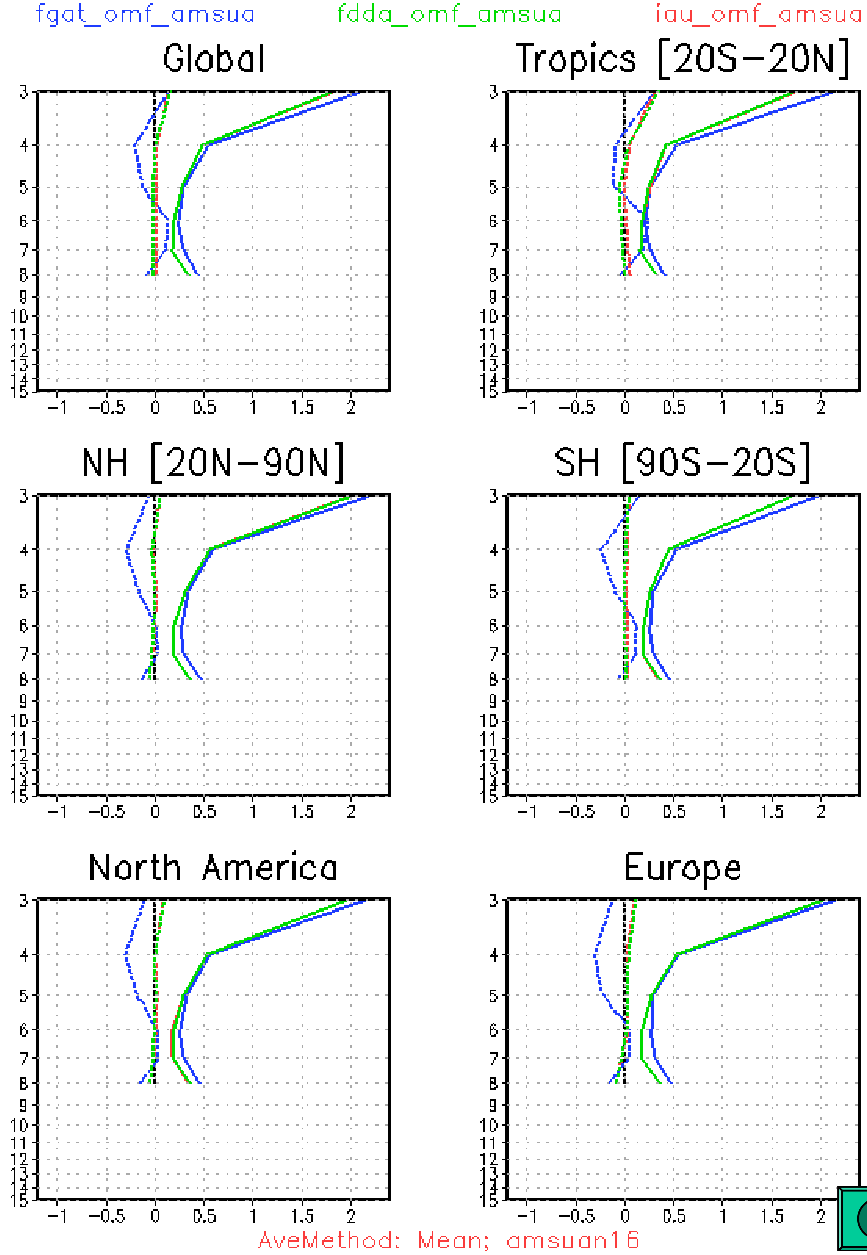




DAS Comparison: IAU vs FGAT vs 4DVAR

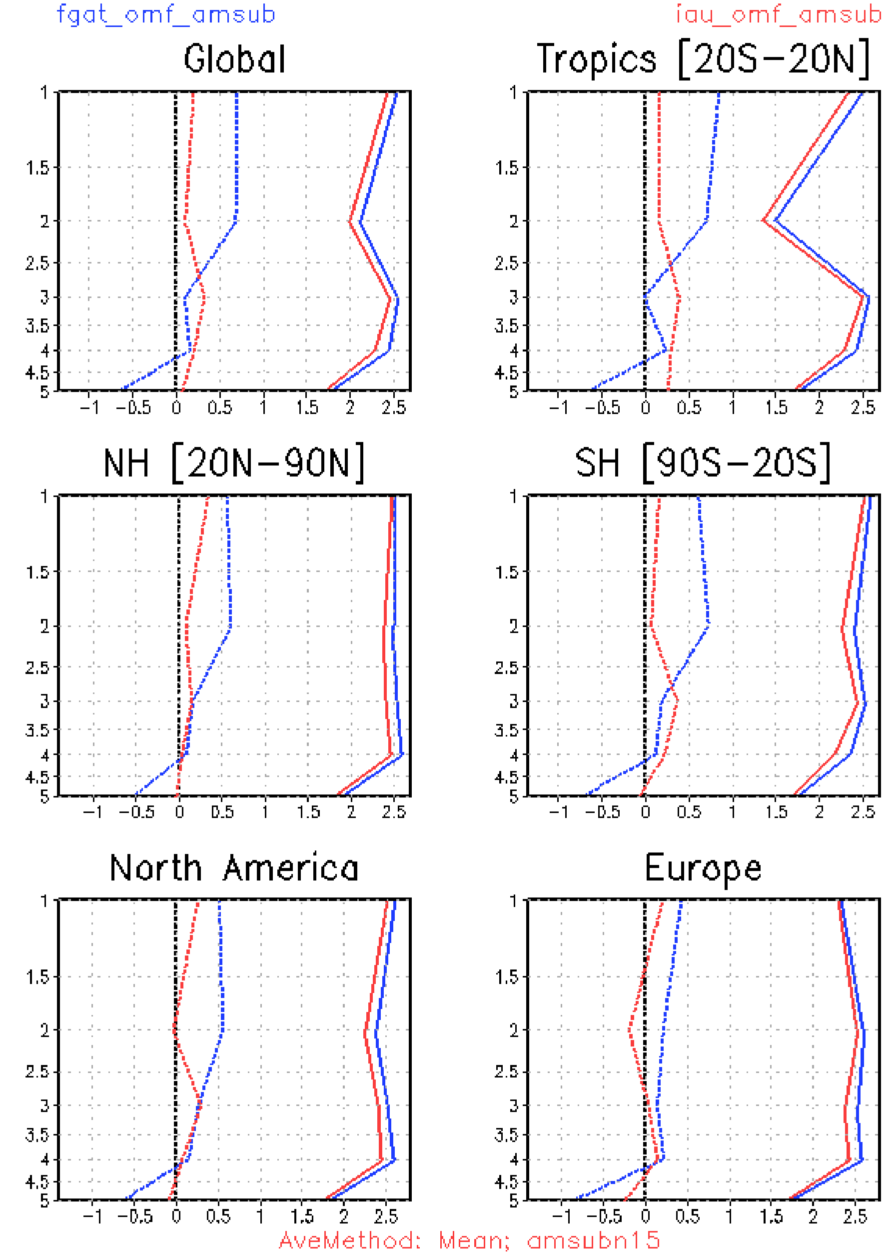
Mean (dashed) & Std Dev for NOAA-15 (left) and NOAA-16 **AMSUA** O-F

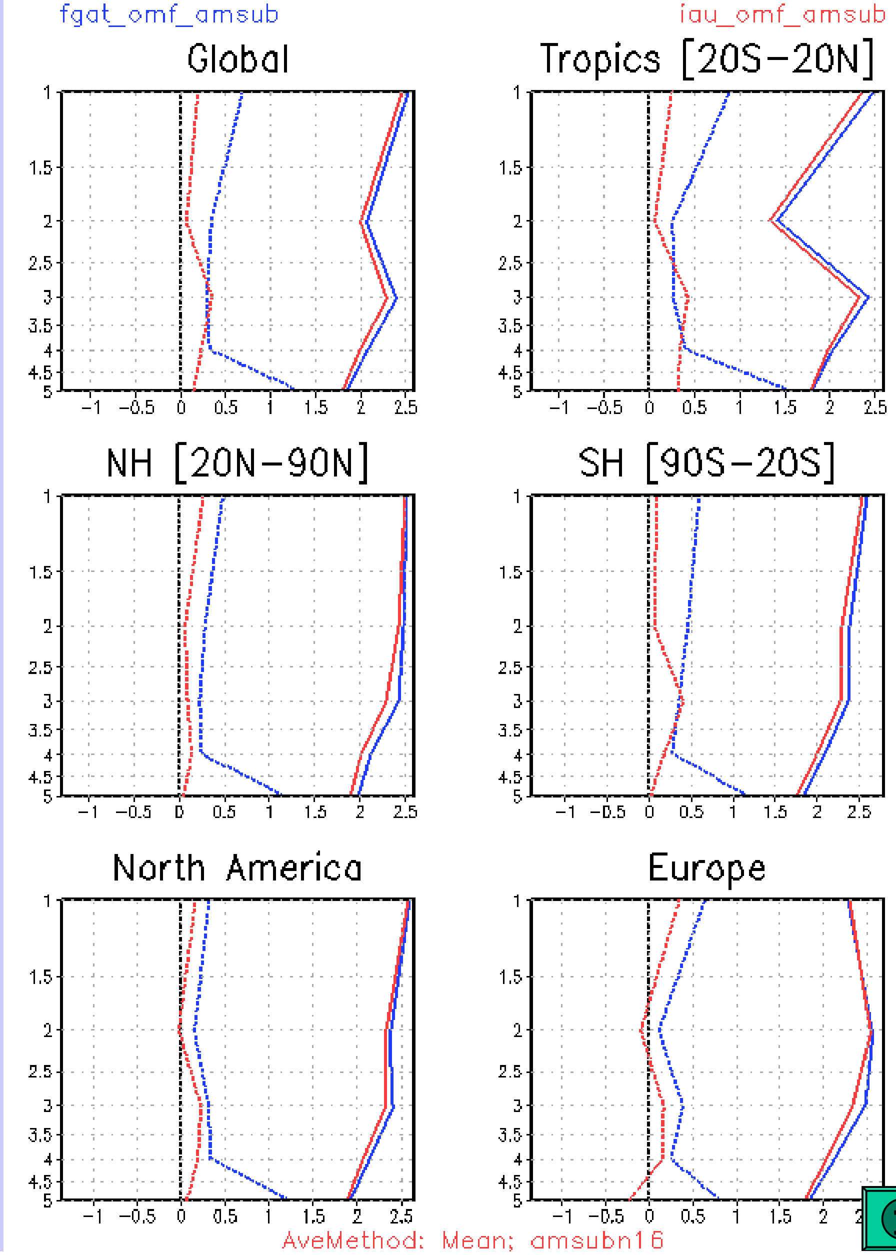




DAS Comparison: IAU vs FGAT

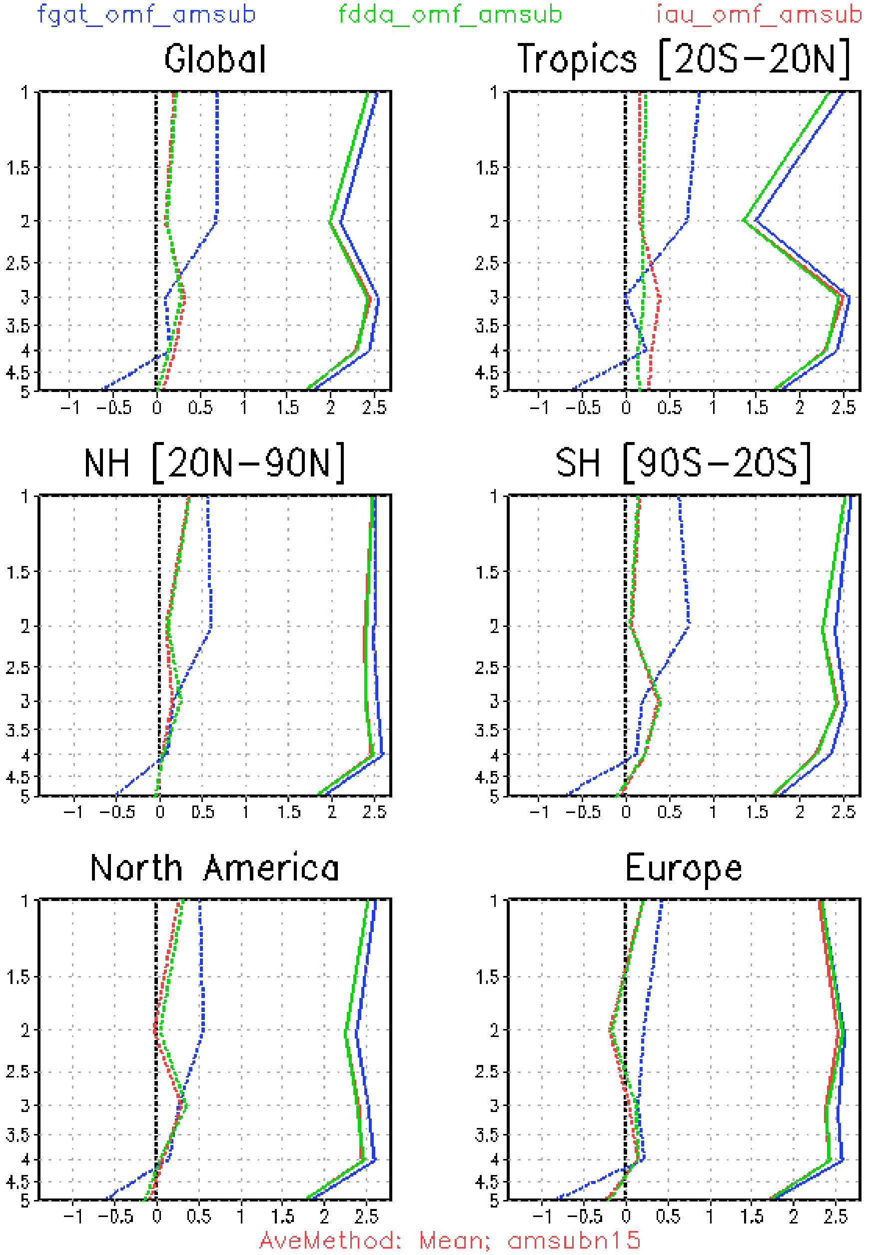
Mean (dashed) & Std Dev for NOAA-15 (left) and NOAA-16 **AMSUB** O-F

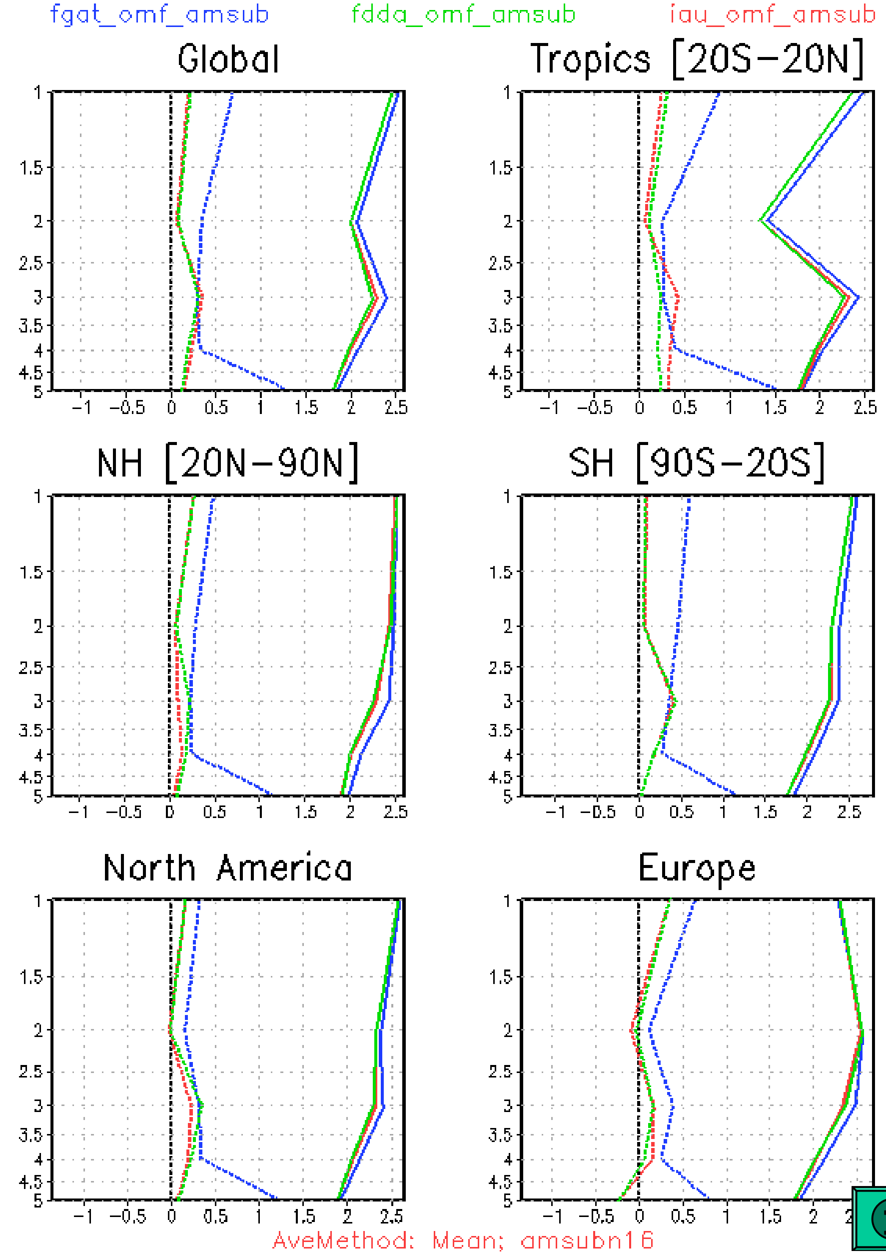




DAS Comparison: IAU vs FGAT vs 4DVAR

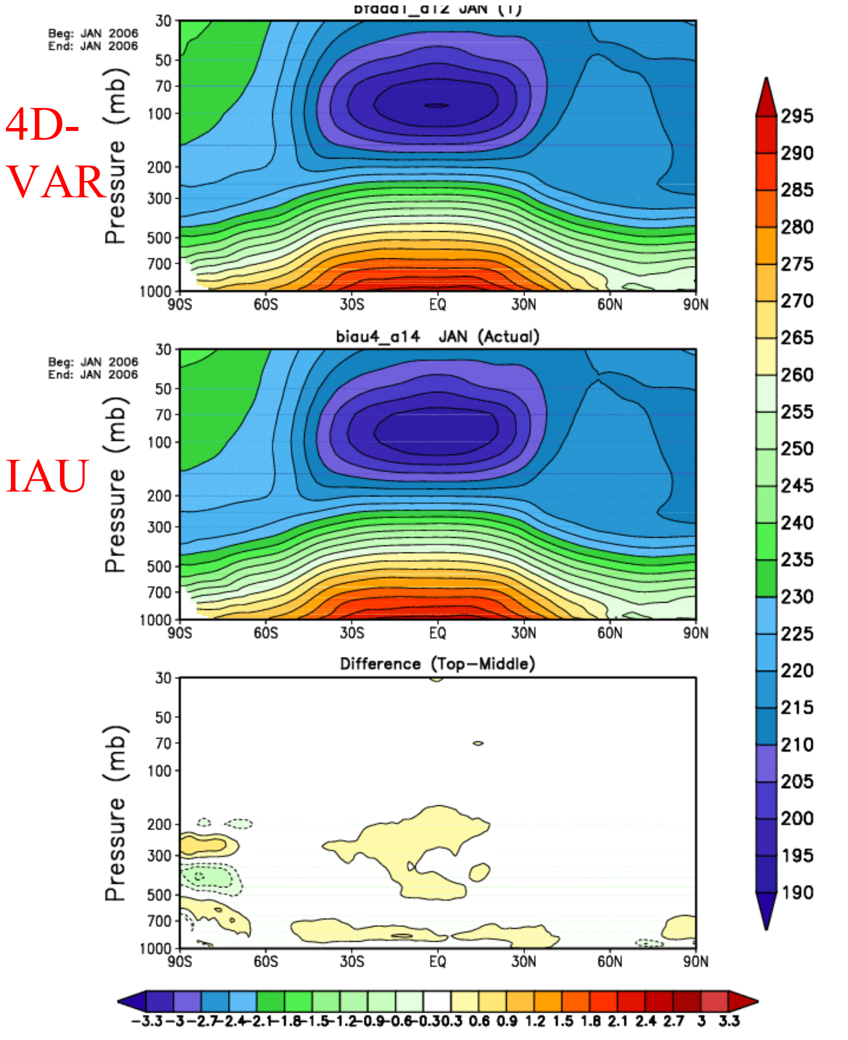
Mean (dashed) & Std Dev for NOAA-15 (left) and NOAA-16 **AMSUB** O-F



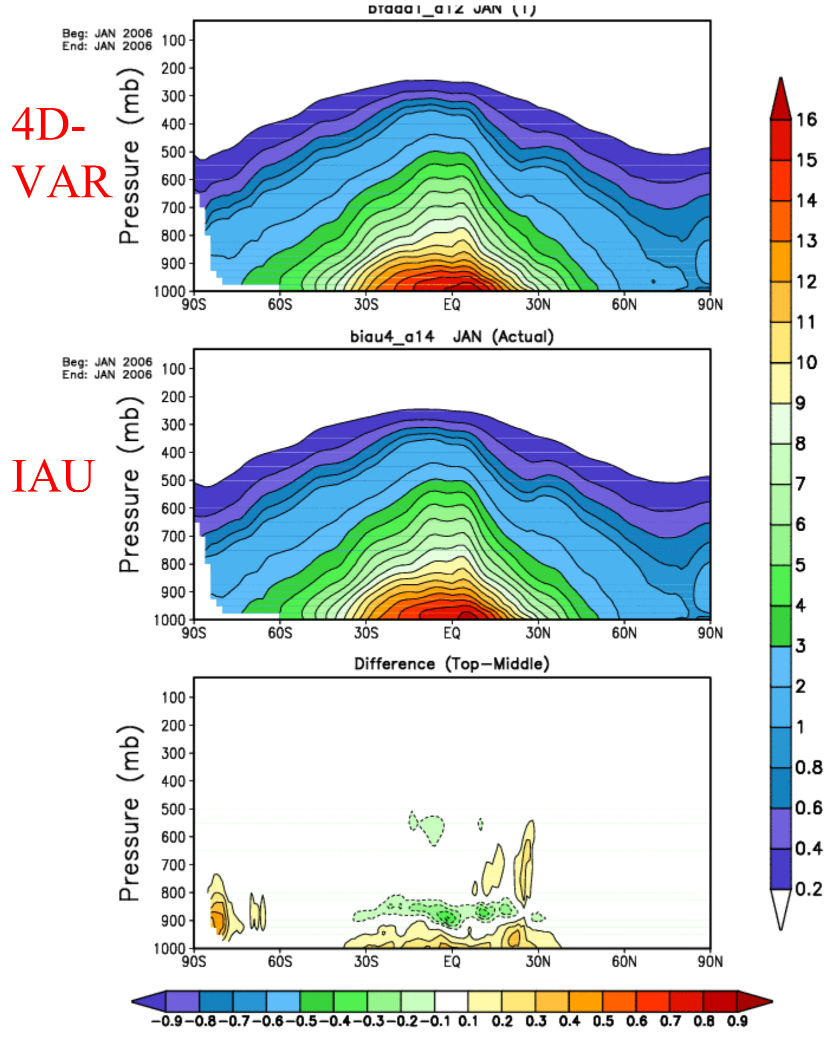


Zonally Ave Monthly Means: IAU vs 4DVAR

Temperature (K)

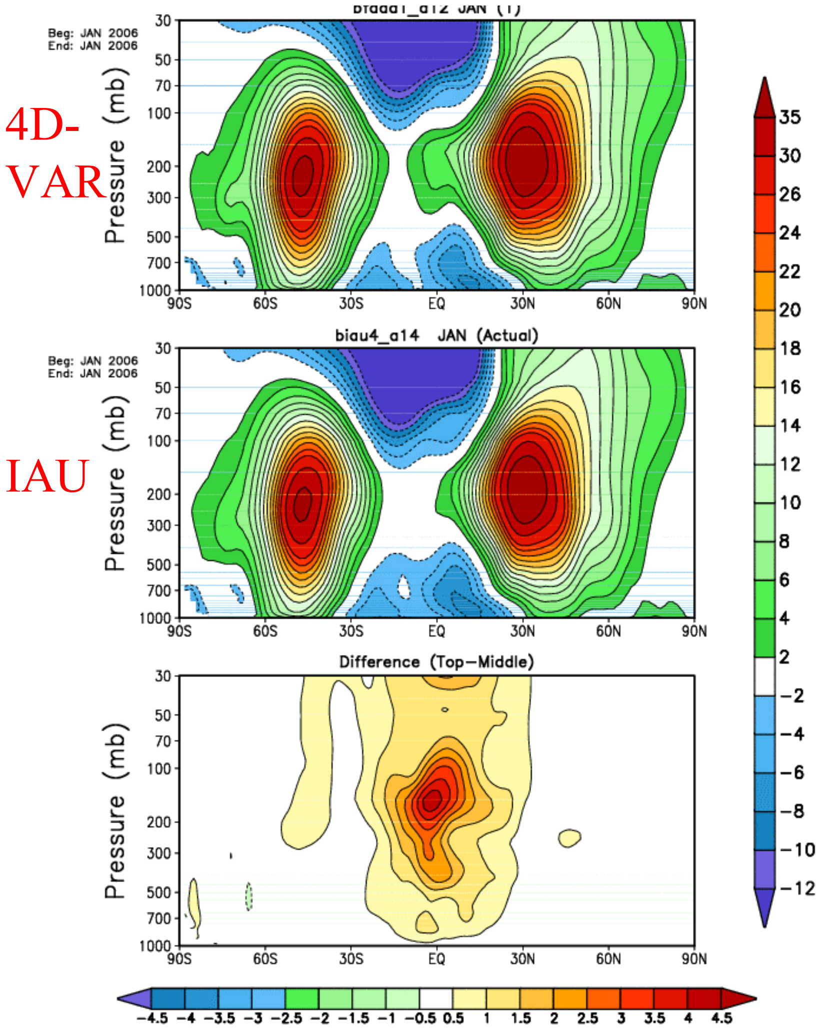


Specific Humidity (g/Kg)

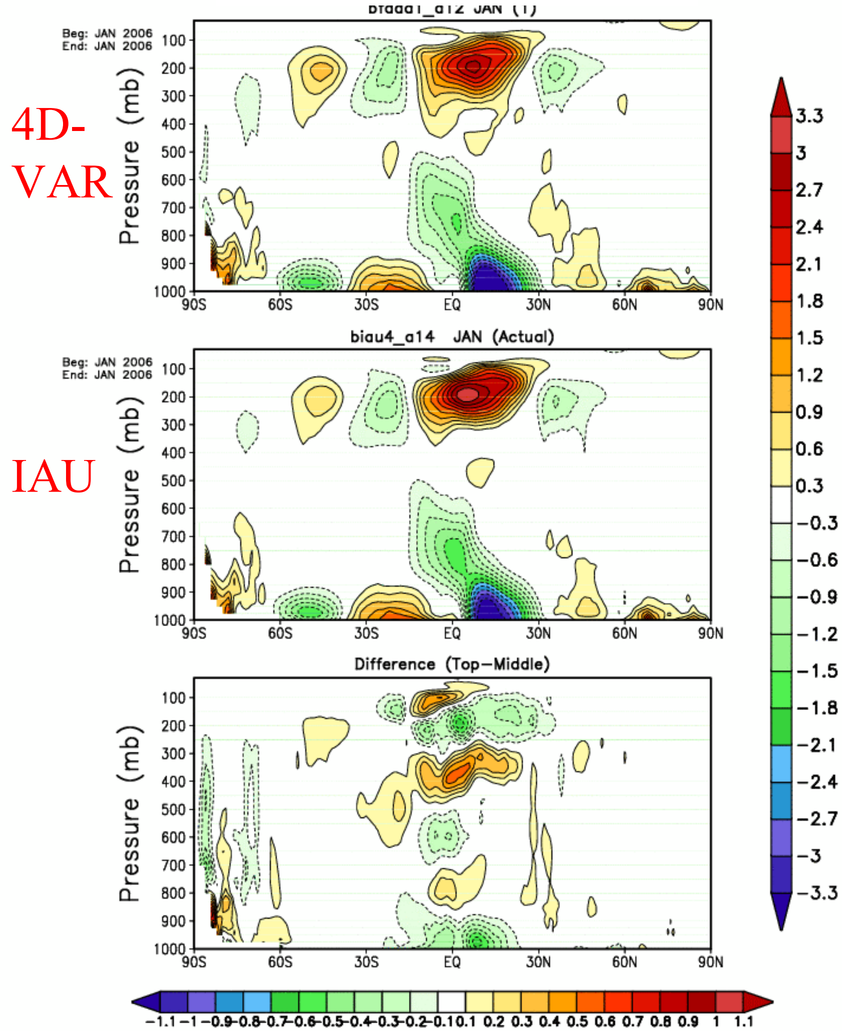


Zonally Ave Monthly Means: IAU vs 4DVAR

Zonal U-Wind (m/s)

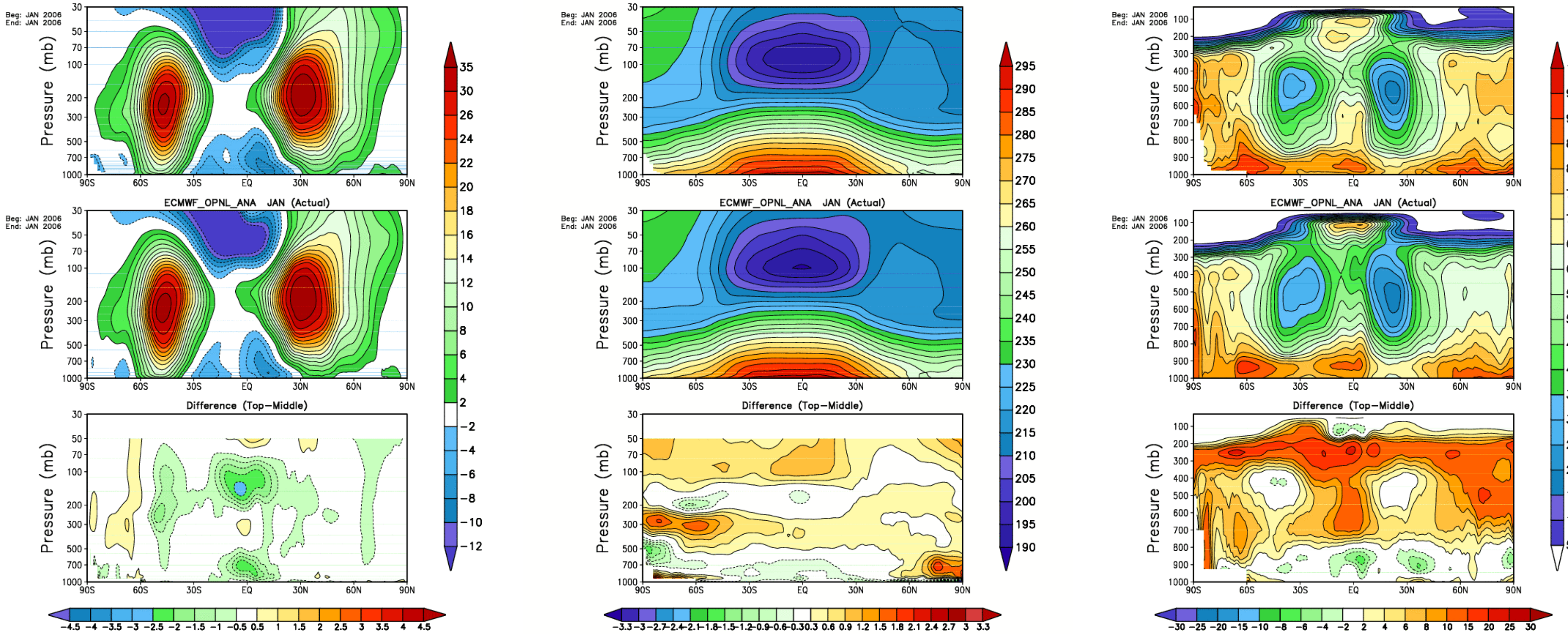


Meridional V-Wind (m/s)



Zonally Ave Monthly Means: IAU (top) vs ECMWF (middle)

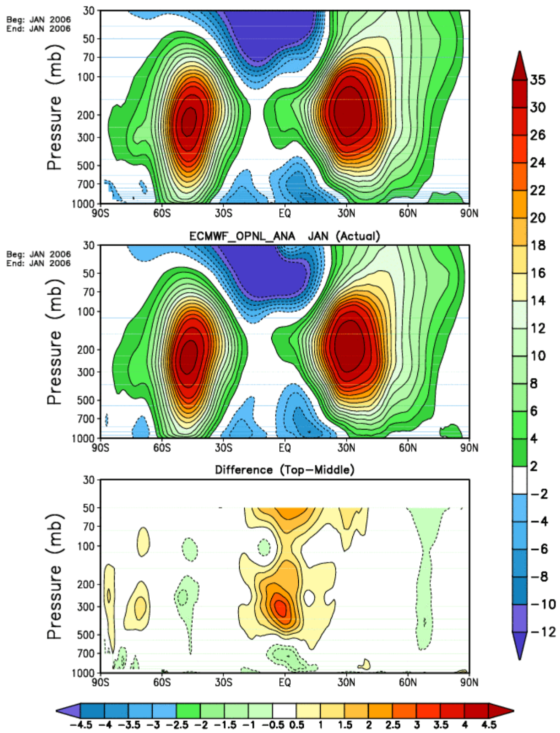
Zonal U-Wind (m/s) | Temperature (K) | Relative Humidity(%)



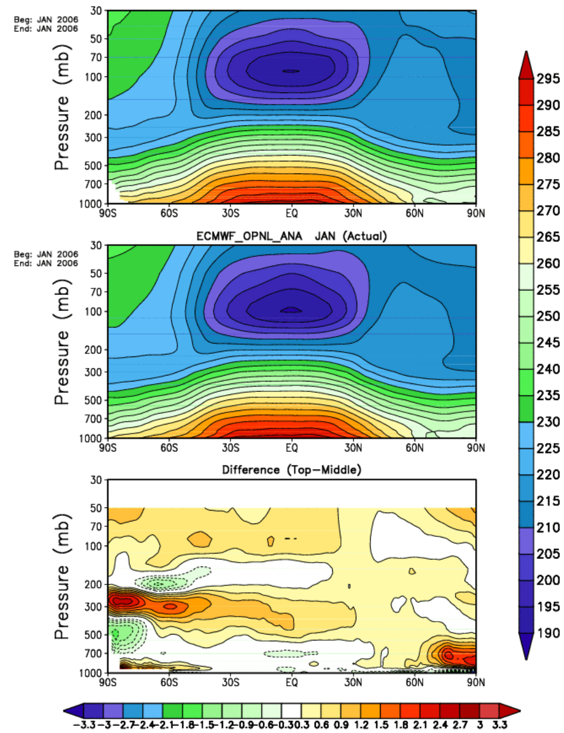
Note: GOES-5 results displayed here are from a 2-degree experiment and are not completely representative of full resolution results.

Zonally Ave Monthly Means:
4DVAR (top) vs ECMWF (middle)

Zonal U-Wind (m/s)



Temperature (K)



Relative Humidity(%)

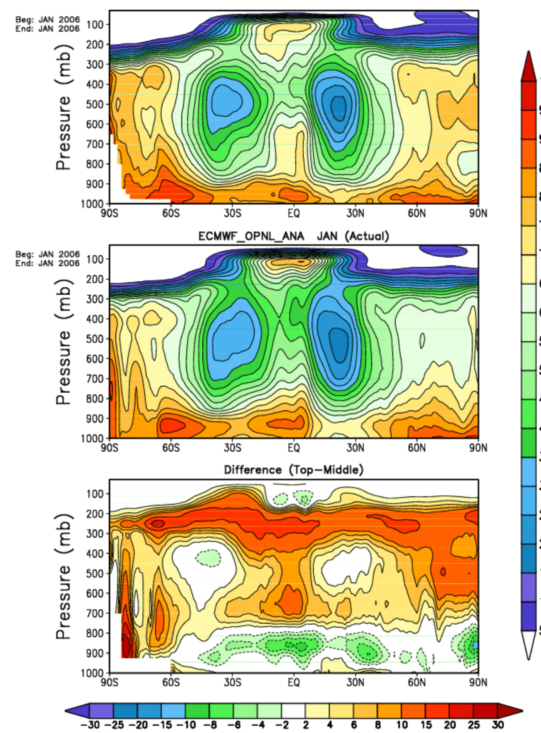


Illustration 2: *Summary*

- Though much remains to be done, first 4dvar results are quite encouraging
- Much work still to take place:
  - Nested resolution outer loop
  - Various details in observer
  - Apply and tune Jc for FGAT
  - Apply and tune DFI for 4DVAR
  - Replace model TLM/ADM by cube-sphere's
  - Model change to take weak constraint
  - Eventually, retune **B**
- Exercise weak constraint (work on model error **Q**)

Observation Sensitivity and Impact

Take $\mathcal{F}$ to be the measure of an aspect of the forecast one wishes to examine. Since the forecast if a function of the background field x^b and the observations y, this measure is a convolution of operations:

$$\mathcal{F}(\mathrm{x}^b, \mathrm{y}) = F \bullet M \bullet G(\mathrm{x}^b, \mathrm{y})$$

where M and G represent the forecast model and the data assimilation system, respectively.

The sensitivity of the forecast to observations is

$$\frac{\partial \mathcal{F}}{\partial \mathrm{y}} = \mathrm{G}^T \mathrm{M}^T \frac{\partial F}{\partial \mathrm{x}^f} = \mathrm{G}^T \frac{\partial F}{\partial \mathrm{x}^a}$$

which in general requires second order adjoint information (Le Dimet et al 2002).

The observation impact, defined as the change in the forecast aspect $\mathcal{F}$ due to a set of observations can be approximated to first order by

$$I_1 \equiv < \frac{\partial F}{\partial \mathrm{x}}, \delta \mathrm{x} >$$

where $\delta \mathrm{x}$ represents an analysis increment.

Observation Sensitivity and Impact (cont.)

For a linear DAS, $\delta\mathbf{x} = \mathbf{K}\mathbf{d}$, here,

$$I_1 = < \frac{\partial F}{\partial \mathbf{x}}, \mathbf{K}\mathbf{d} > = < \mathbf{K}^T \frac{\partial F}{\partial \mathbf{x}}, \mathbf{d} > = < \frac{\partial F}{\partial \mathbf{y}}, \mathbf{d} > .$$

For a nonlinear DAS, the increment is a successive correction of the linear-type increment, that is,

$$\delta\mathbf{x}_j = \mathbf{K}_j\mathbf{d}_j + (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)(\mathbf{x}^b - \mathbf{x}_{j-1}) .$$

The final (total) increment is

$$\delta\mathbf{x} = \sum_{j=1}^{m} \mathbf{K}_m\mathbf{H}_m \cdots \mathbf{K}_{j+1}\mathbf{H}_{j+1}\mathbf{K}_j\mathbf{d}_j .$$

Therefore, for a nonlinear DAS, the first order impact is approximately

$$I_1 = \sum_{j=1}^{m} < \mathbf{K}_m\mathbf{H}_m \cdots \mathbf{K}_{j+1}\mathbf{H}_{j+1}\mathbf{K}_j \frac{\partial F}{\partial \mathbf{x}}, \mathbf{d}_j > .$$

Higher order terms for linear DAS have been derived and discussed in Errico (2007) and Gelaro et. (2007); Tremolet (2008) gives a comprehensive discussion and derivation for nonlinear DAS.

The Adjoint of a Variational Analysis System

Observation sensitivity and impact studies require the adjoint of the underlying data assimilation system. The model adjoint provides the model sensitivity as the input to the analysis adjoint for calculation of the observation sensitivities.

Concentrating on the analysis adjoint, there are at least three ways to obtain the adjoint of a variational analysis system:

- Direct, line-by-line, adjoint (Zhu & Gelaro 2007)
- Operator manipulation:
  - Observation space (Baker & Daley 2000):
    $\mathrm{K}^T \partial F / \partial \mathrm{x} = \delta \mathrm{z}$

$$(\mathrm{HBH}^T + \mathrm{R})\delta \mathrm{z} = \mathrm{HB}\frac{\partial F}{\partial \mathrm{x}}$$

Full DAS

  - In physical/spectral space (Tremolet 2008):
    $\mathrm{K}^T \partial F / \partial \mathrm{x} = \mathrm{R}^{-1}\mathrm{H}\delta \mathrm{x}$

$$(\mathrm{B}^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H})\delta \mathrm{x} = \frac{\partial F}{\partial \mathrm{x}}$$

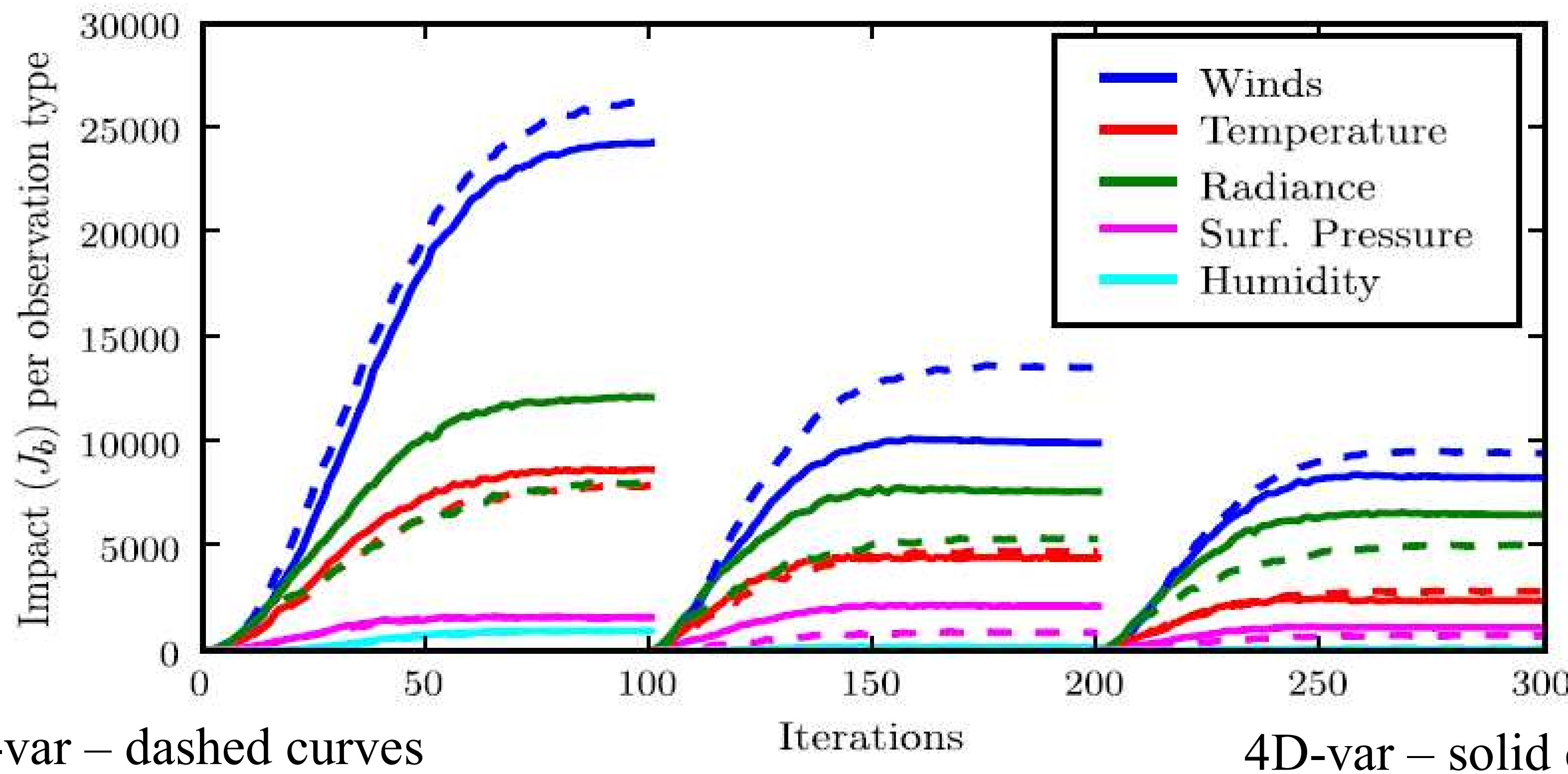
Static analysis

- Approximating the Hessian (Cardinali et al. 2004):
  $\mathrm{K}^T \partial F / \partial \mathrm{x} = \mathrm{R}^{-1}\mathrm{H}\tilde{\mathrm{A}} \partial F / \partial \mathrm{x}$

$$\tilde{\mathrm{A}} \sim \sqrt{\mathrm{B}^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}}$$

Illustration 1: Static Analysis

Observation Impacts within each inner loop for various outer loops: 3D-var vs 4D-var



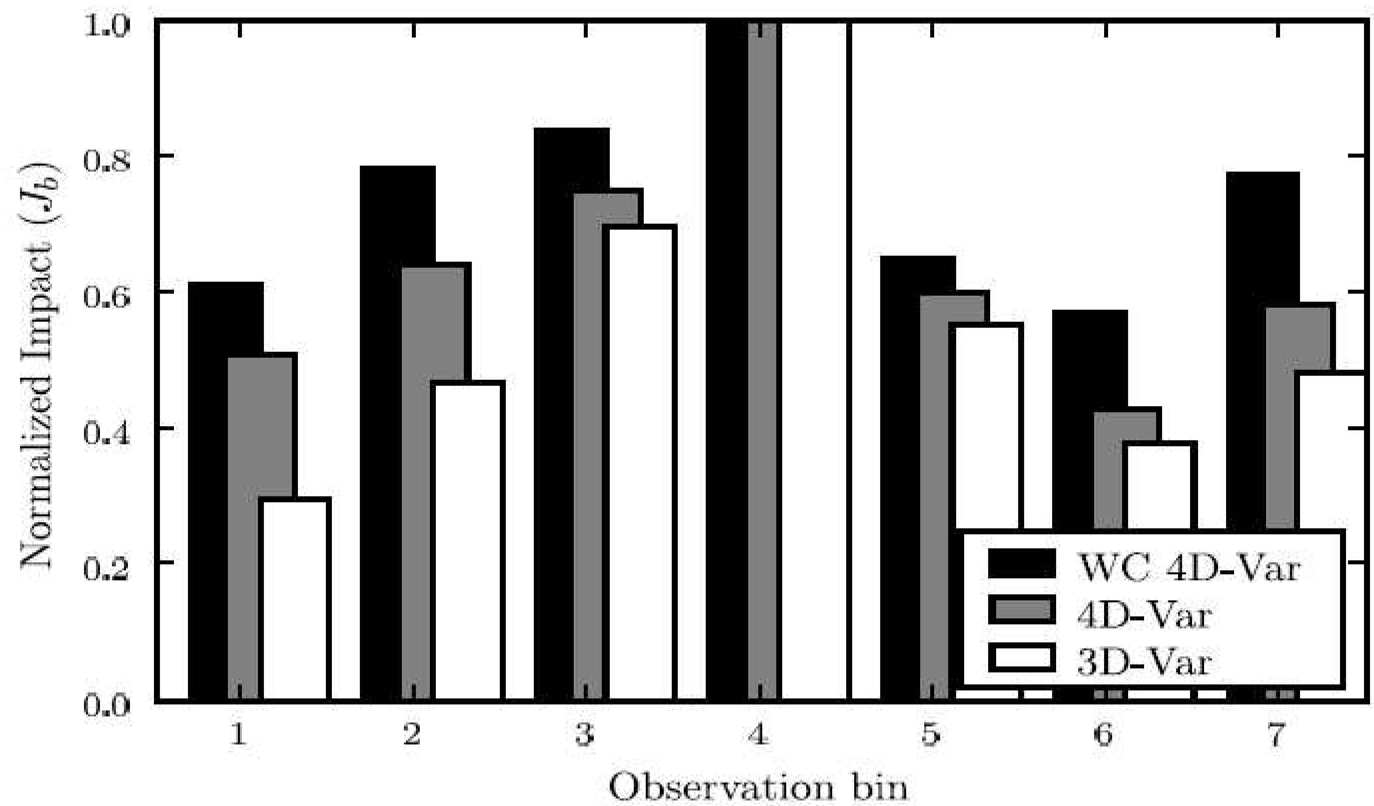
3D-var – dashed curves

4D-var – solid curves

Remark: Impacts shown above are for a single 6-hr interval for M=I

Illustration 1: Static Analysis (cont.)

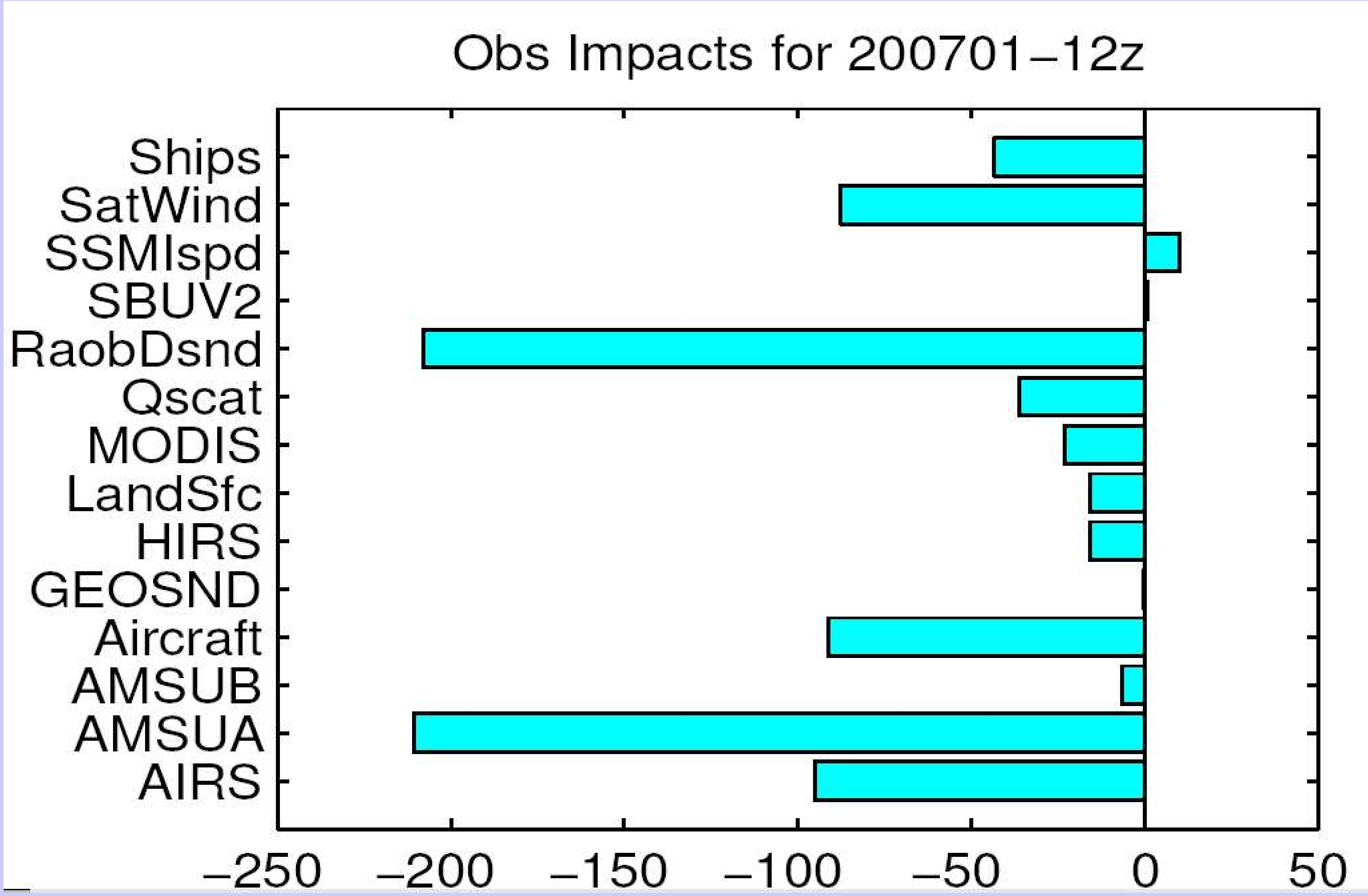
Observation Impacts vs Time within a 6-hr Assimilation Window



Remarks: - Impacts shown above are for a single 6-hr interval for M=I
- Model error covariance taken as a multiple of background error covariance

Illustration 2: Full DAS & Forecasts

Accumulated forecast error reduction due to various observing instruments for the 24-forecasts from 12Z for January 2007



Summary and Conclusions

- NASA GEOS DAS has entered the 4D-world (again!)
- Various adjoint tools are now available in GEOSDAS, capable of performing studies in forecast sensitivities, singular vectors, analysis sensitivity and observations impact
- First exercise including some of these tools will be the Observations Impact Inter-comparison Study (NASA, NRL, ECMWF, and Env. Canada)
- Hooks for weak constraint are in place in GSI and soon will be in place in the GEOS-5 GCM
- Work is on way to update the GCM TLM/ADM with cube-sphere core
- Soon we will be able to compare 4DVAR with NCEP's approximate 4D-scheme; First Order Time-interpolation to Observations (FOTO)

The implementations done thus far benefited greatly from the incredible infrastructure of GSI.

Let the fun begin …

References of Explicitly Cited Works

Cardinali, C., S. Pezzulli, and E. Andersson, 2004: Influence-matrix diagnostic of a data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **130**, 2767-2786.

Errico, R. M., 2007: Interpretations of an adjoint-derived observation impact measure. *Tellus,* , **59A**, 273-276.

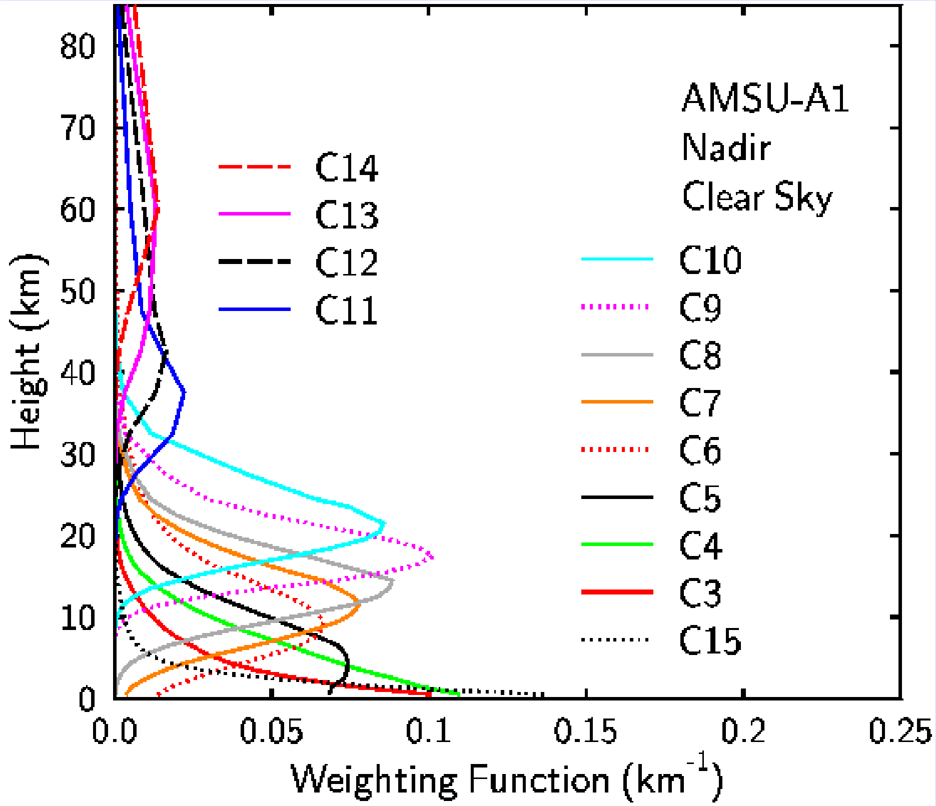
Gelaro, R., Y. Zhu, R. M. Errico, 2007: Examination of various-order adjoint-based approximations of observation impact. *Meteorologische Zeitschrift*, in press.

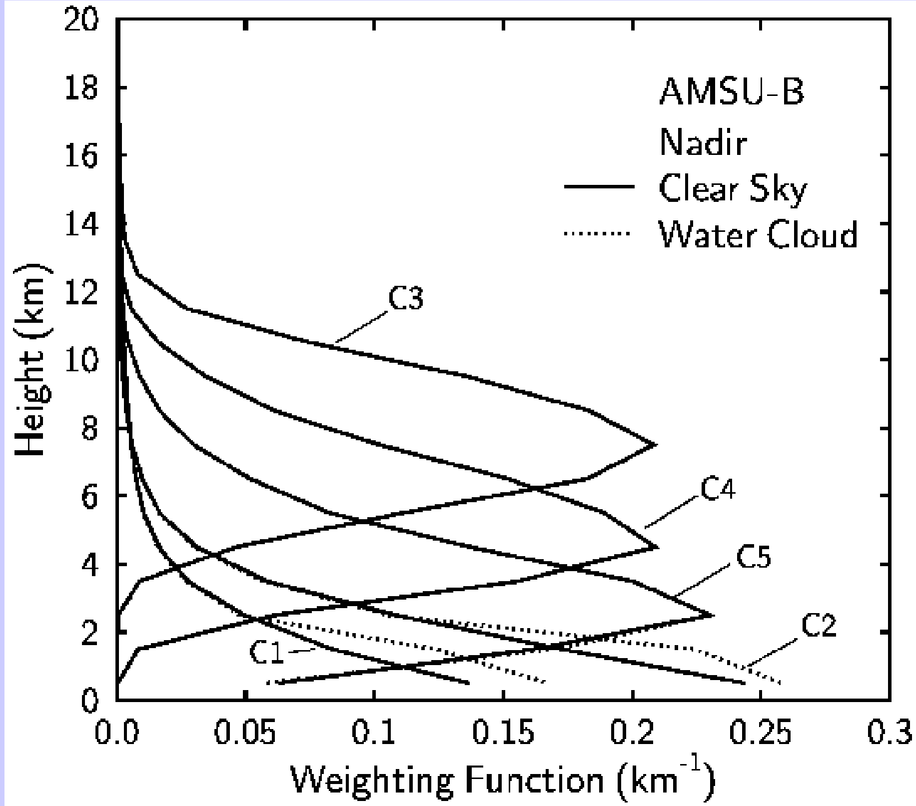
Le Dimet, F.-X., I.M. Navon, D. N. Daescu, 2002: Second-order information in data assimilation. *Mon. Wea. Rev.,* , **130**, 629-648.

Tremolet, 2008: Observation sensitivity and observation impact in incremental variational data assimilation. *In preparation.*

Zhu, Y., and R. Gelaro, 2007: Observation sensitivity calculations using the adjoint of the grid-point statistical interpolation (GSI) analysis system. *Mon. Wea. Rev.*, **135**, in press.

AMSUA & AMSUB Weighting Functions





From http://amsu.cira.colostate.edu/weights.html