Hybrid (3D-VAR /ETKF) Data Assimilation System

Meral Demirtas

National Center for Atmospheric Research

WRF Tutorial, WRF-Data Assimilation

21st July 2009, Boulder, Colorado

An outline of this presentation

- ***Why*** do we need a hybrid data assimilation system?
- ***What*** are the basic ingredients of a hybrid system?
- ***How*** have we implemented hybrid (3DVAR - ETKF) system at the Data Assimilation Testbed Center (DATC)?
- ***What*** we have found: Highlights of preliminary results
- Summary and conclusions
- An introduction for the hybrid practice session

Why do we need a hybrid system?

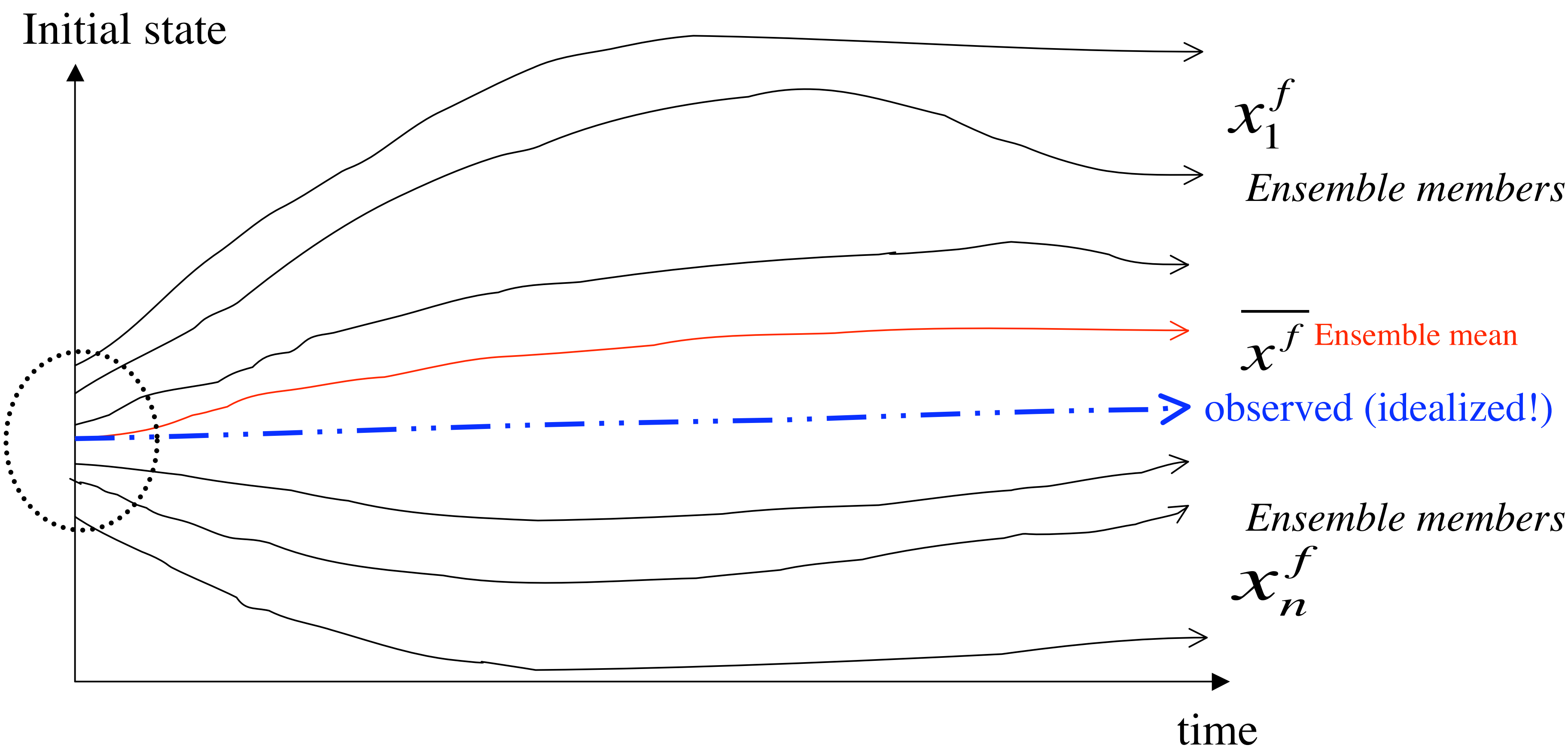
- The WRF 3D-VAR system uses only *climatological* (static) background error covariances.
- *Flow-dependent* covariance through ensemble is needed.
- Hybrid combines *climatological* and *flow-dependent* background error covariances.
- It can be adapted to an existing 3D-VAR system.
- Hybrid can be robust for small size ensembles.

What are the basic ingredients of a hybrid system?

1. Ensemble forecasts: *WRF-ensemble forecasts*
2. A mechanism to update ensemble perturbations: *Ensemble Transform Kalman Filter (ETKF)*
3. A data assimilation system: *WRF 3D-VAR*

It sounds simple…. :-)

Ensembles to address uncertainties in the initial state



Ensemble Formulation Basics

Assume the following ensemble forecasts:

$$X^f = (x_1^f, x_2^f, x_3^f, \ldots\ldots x_N^f)$$

Ensemble mean: $\overline{x}^f = \dfrac{1}{N}\sum_{i=1}^{N} x_n^f$

Ensemble perturbations: $\delta x_n^f = x_n^f - \overline{x^f}$

Ensemble perturbations in vector form:

$$\delta X^f = (\delta x_1^f, \delta x_2^f, \delta x_3^f, \ldots., \delta x_N^f) \quad n = 1, N$$

How to update ensemble perturbations?

ETKF technique updates ensemble perturbations by rescaling innovations with a transformation matrix (*Wang and Bishop 2003*).

$$x^a = x^f \mathrm{T}$$

T*ransformation matrix*
(solved by Kalman Filter Theory)

$$\mathrm{T} = \mathrm{C}(\Gamma + \mathrm{I})^{-1/2}\mathrm{C}^T$$

Bishop et al. (2001)

C: is the column matrix that contains the orthonormal eigenvectors

Γ: is the diagonal matrix that contains the eigenvalues

I: is the identity matrix

How to improve the under estimation of the analysis-error variance?

The background-error variance estimated from the spread of ensembles should be consistent with the background error-variance estimated from the differences between the ensemble mean and observations. Wang and Bishop (2003) has introduced an adaptive ***inflation factor***, $\prod$**,** to ameliorate this problem:

$$\mathbf{x}_i = \mathbf{x}_i^f \mathbf{T}_i \prod_i$$

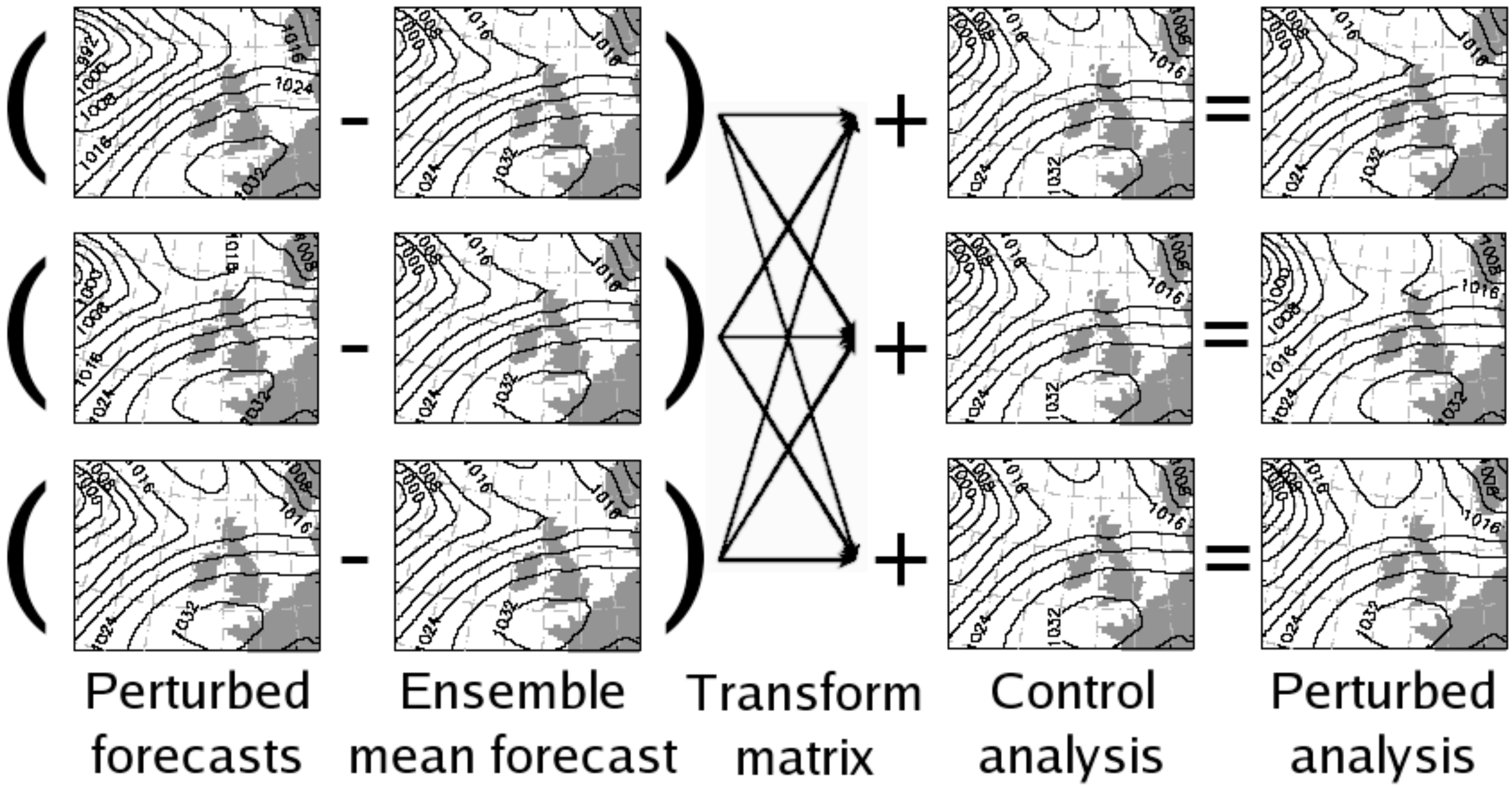
$$\prod_i = \sqrt{c_1, c_2, \ldots, c_i}$$

where

$$c_i = \frac{\mathrm{tr}\left\langle \tilde{\mathsf{d}}_i \tilde{\mathsf{d}}_i^T \right\rangle - N_{obs}}{\mathrm{tr}(\boldsymbol{\Gamma})}$$

$$\tilde{\mathbf{d}}_i = \mathbf{R}^{-1/2}\left(\mathbf{y}_i - \mathbf{H}\overline{\mathrm{x}}_i^b\right)$$

Schematic illustration of the ETKF technique



Courtesy of Bowler et al. (2008)

Pros and Cons of ETKF Technique

- Desirable aspects:
  - ETKF is fast (computations are done in model ensemble perturbation subspace).
  - It is suitable for generating ensemble initial conditions.
  - It updates initial condition perturbations.

- Less desirable aspects:
  - ETKF does not localize, therefore it does not represent sampling error efficiently. It may need very high inflation factors.

Hybrid: Combine 3D-VAR and ETKF

- Flow-dependent covariance through ensembles.
- Coupling wind, temperature and moisture fields.
- Hybrid can be more robust for small size ensembles.
- It can be adapted to an existing 3D-VAR system.
- It is less computationally expensive compared to other ensemble filters.

The hybrid formulation….

Ensemble covariance is implemented into the 3D-VAR cost function via ***extended control variables***: *(Wang et. al. 2008)*

$$J(x_1^{'},\alpha)=\beta_1\frac{1}{2}x_1^{'T}B^{-1}x_1^{'}+\beta_2\frac{1}{2}\alpha^T C^{-1}\alpha+\frac{1}{2}(y^{o'}-Hx^{'})^T R^{-1}(y^{o'}-Hx^{'})$$

$$x^{'}=x_1^{'}+\sum_{k=1}^{K}(\alpha_k\circ x_k^e)$$

Conserving total variance requires:

$$\frac{1}{\beta_1}+\frac{1}{\beta_2}=1$$

C: correlation matrix for ensemble covariance localization

$x_1^{'}$ 3D-VAR increment

$x^{'}$ Total increment including hybrid

α Extended control variable

β_1 Weighting coefficient for static 3D-VAR covariance

β_2 Weighting coefficient for ensemble covariance

Hybrid Horizontal Localization

Ensemble covariance horizontal localization is done through recursive filters. Preconditioning designed as: *(Wang et. al. 2008)*

- 3D-VAR part

$x'_1 = U_1 v_1$ where $U_1 \approx B^{1/2}$ **B** *is the 3D-VAR static covariance matrix.*

- Ensemble part

$\alpha = U_2 v_2$ where $U_2 \approx C^{1/2}$ **C** *is the ensemble correlation matrix which defines ensemble covariance localization.*

Hybrid Vertical Localization

- Spurious sampling error are not confined to horizontal error correlations, it affects vertical too.
- Hybrid alpha control variables can be made 3D to alleviate spurious vertical correlations.
- Two approaches are considered for specifying vertical localization function:
  i. *Empirical function (as in horizontal).*
  ii. *Use vertical background error covariances to define localization.*
- Initial studies use "*empirical function (as in horizontal)*".
- An EOF decomposition of the vertical component of the localization matrix is performed to reduce size of alpha CV.

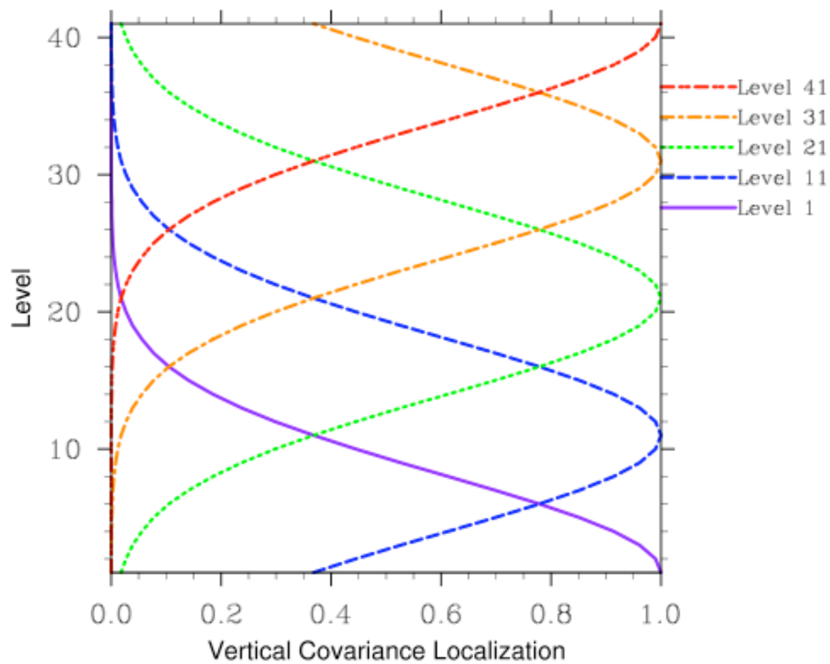
Courtesy of Dale Barker

Empirical Vertical Covariance Localization

Apply Gaussian Vertical Covariance Localization function:

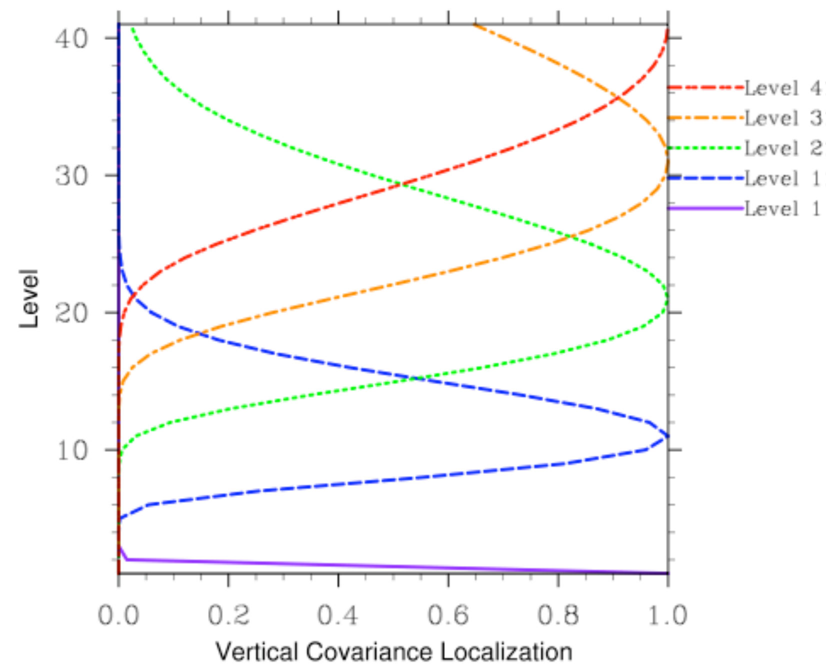
$$\rho(k-k_c)=\exp\left[-(k-k_c)^2/L_c^2\right]$$

Example 1:
Constant Localization Scale



$L_c = 10$

Example 2:
Vertically-Dependent Localization Scale



$L_c = 20k_c / 41$

Courtesy of Dale Barker

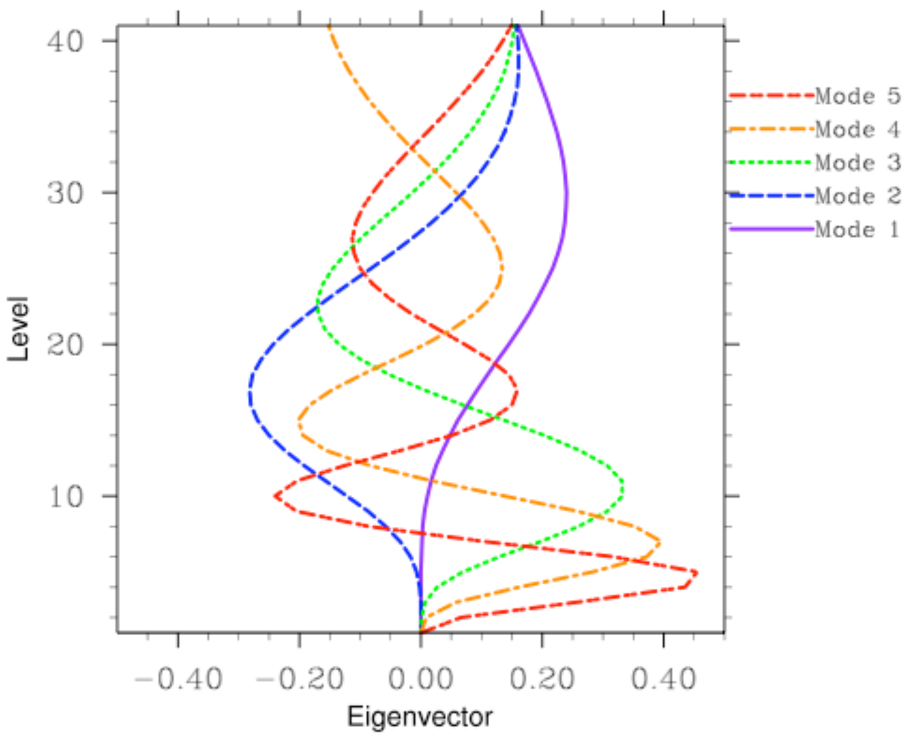
Covariance Localization Decomposition

Example: Gaussian Localization
with variable localization scale:

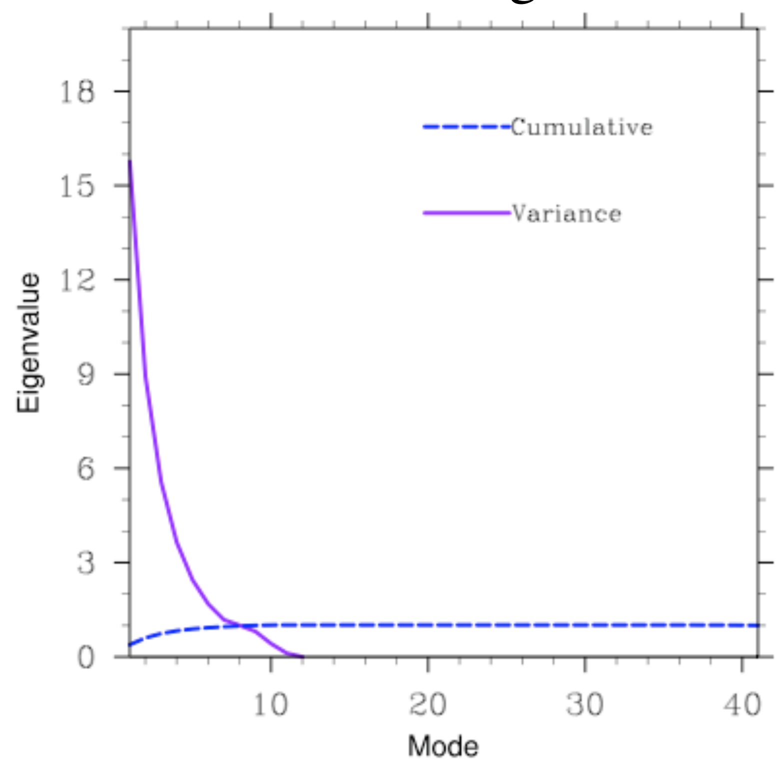
$$\rho(k - k_c) = \exp\left[-\left(k - k_c\right)^2 / L_c^2\right]$$

$$L_c = 20k_c / 41$$

Eigenvectors



Eigenvalues

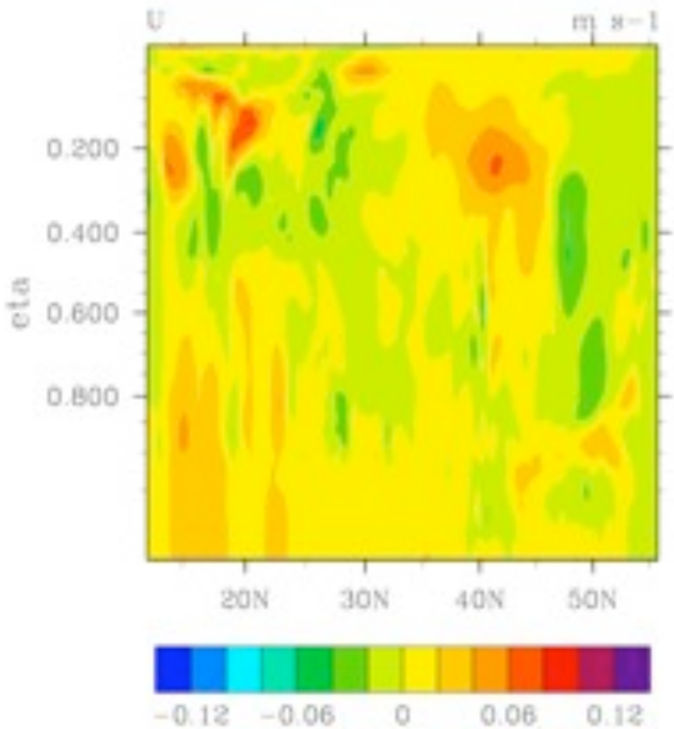


75% data compression via use of EOFs for covariance localization.

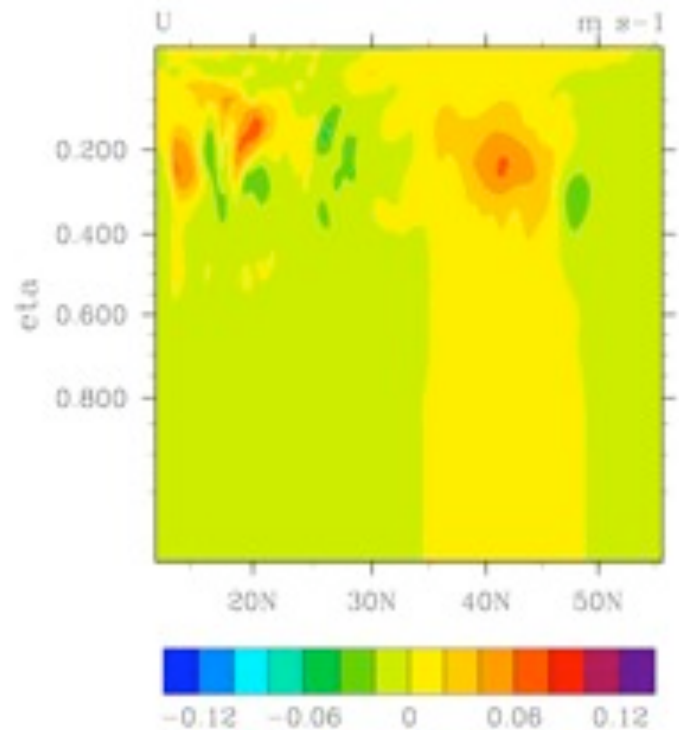
Courtesy of Dale Barker

Vertical cross-section of analysis increments (single-observation based)

No vertical localization applied

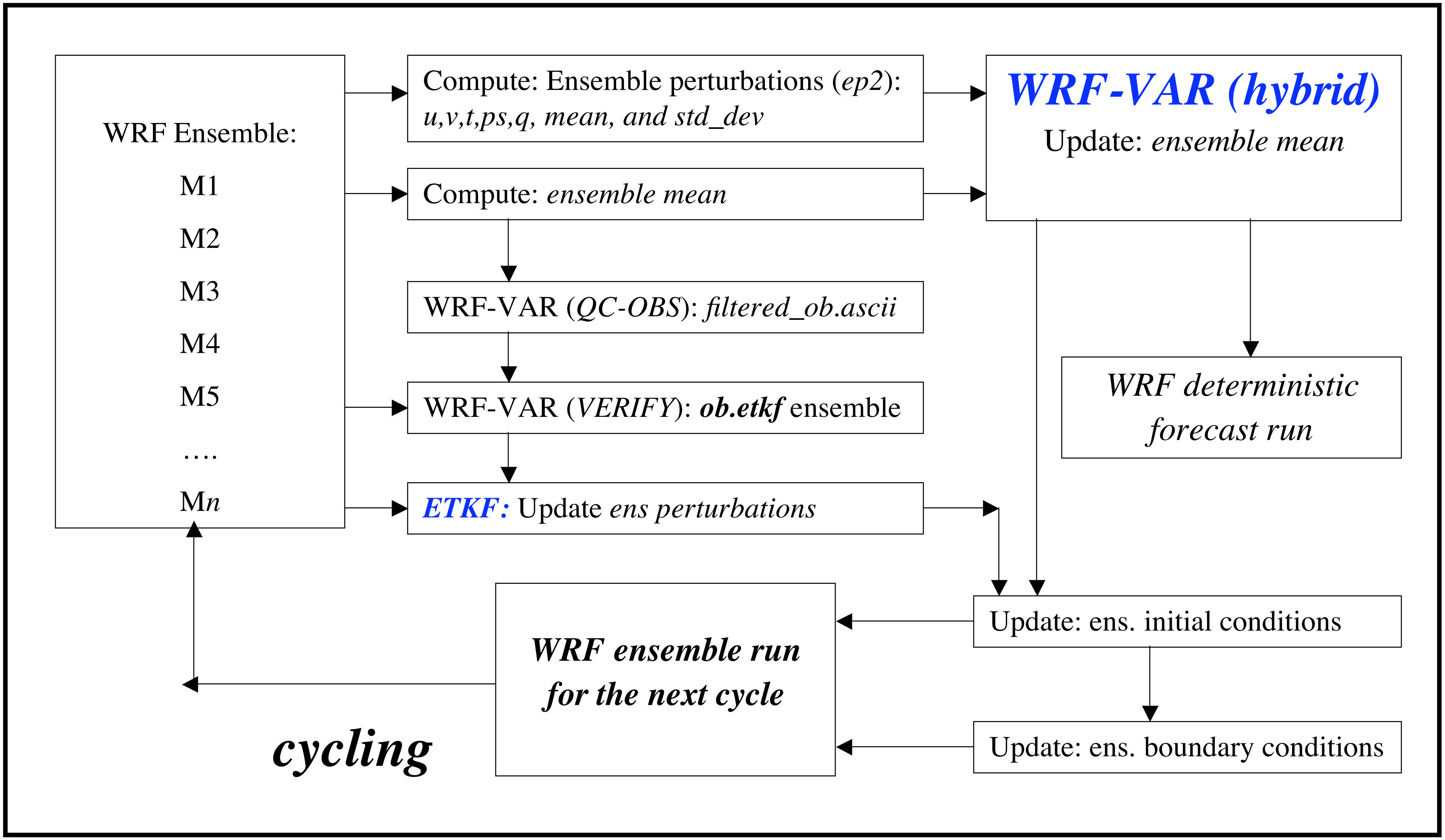


Vertical localization applied



Courtesy of Dale Barker

How have we implemented hybrid (3DVAR -ETKF) system at the (DATC)?



(Demirtas et al. 2009)

A few notes on hybrid settings

- alpha_corr_scale=1500km (Default)
- je_factor (β_1)=2.0
- jb_factor (β_2)=je_factor/(je_factor -1)=2.0
- alphacv_method=2 (ensemble perturbations on model space)
- ensdim_alpha=10 (ensemble size)

N.B. Conservation of total variance requires: $\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1$

The DATC Hybrid System Application Experiment Set-up

- Ensemble size: 10
- Test Period: 15th August - 15th September 2007
- Cycle frequency: 3 hours
- Observations: GTS conventional observations
- Deterministic ICs/BCs: Down-scaled GFS forecasts
- Ensemble ICs/BCs: Produced by adding spatially correlated Gaussian noise to GFS forecasts (*Torn et al. 2006*).

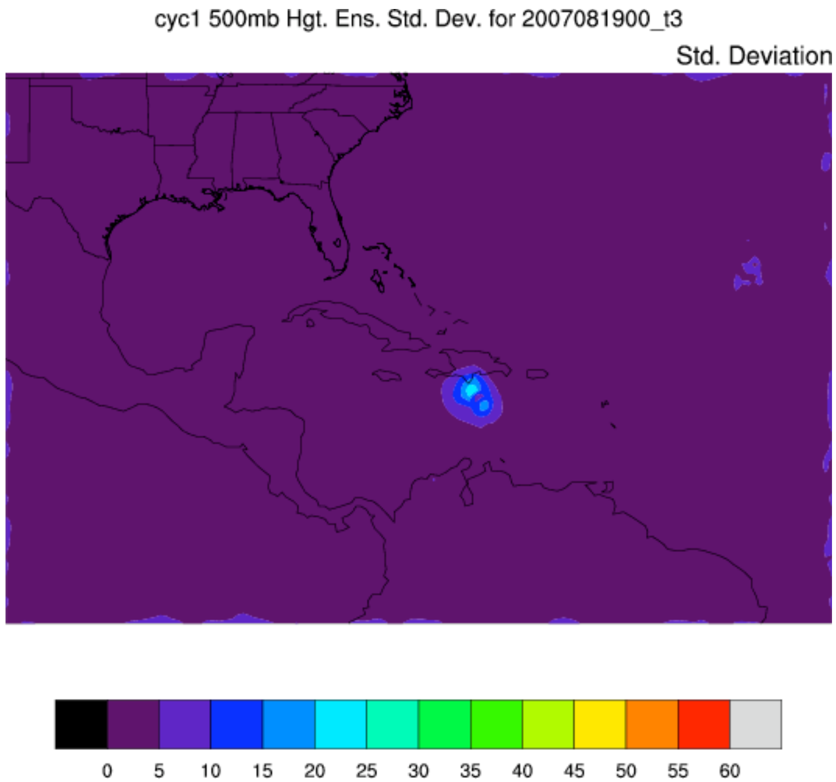
  (WRF-VAR and some additional tools.)
- Horizontal resolution: 45km
- Number of vertical levels: 57
- Model top: 50 hPa

(For details see: Demirtas et al. 2009)

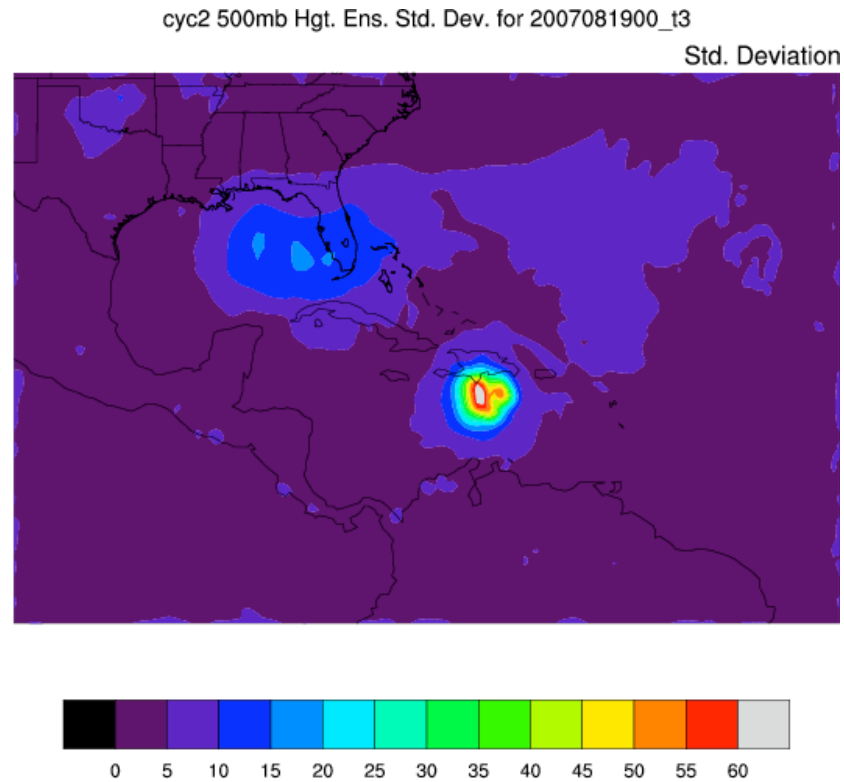
What we have found:
Highlights of preliminary results

Ensemble spread: 500 hPa height (m) std. dev.

WRF t+3 valid at 2007081900

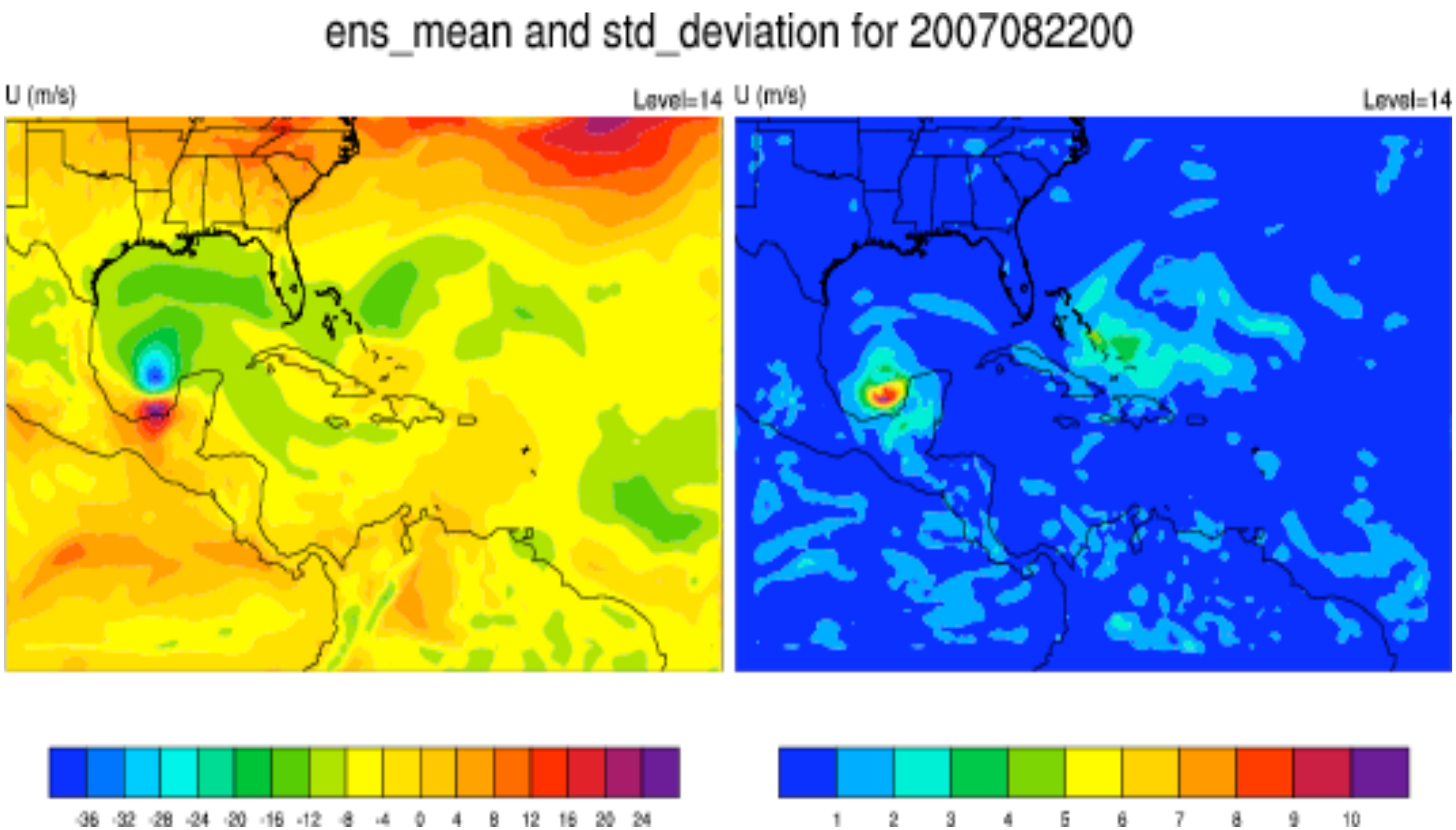


Modest inflations factors used

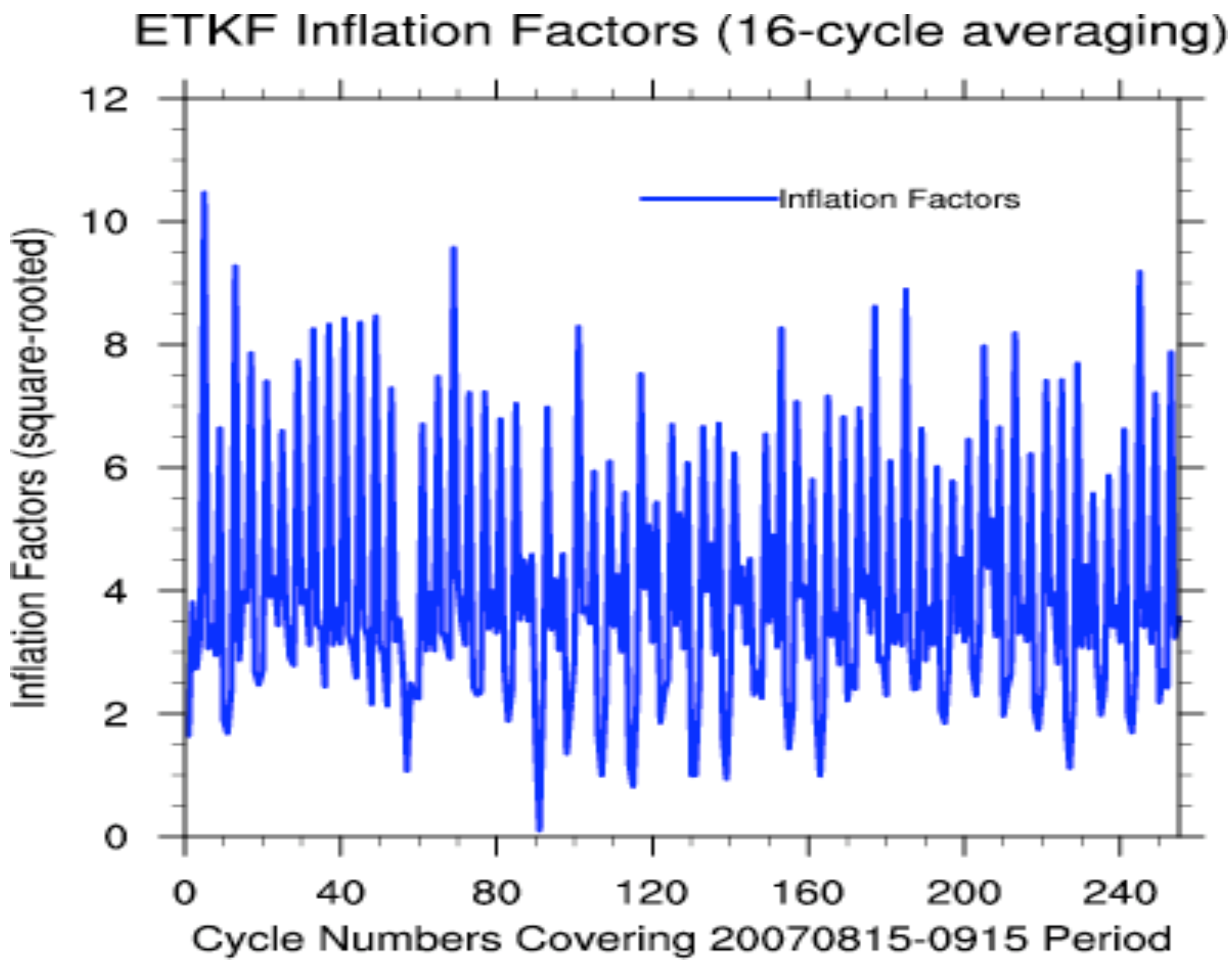


Higher inflations factors used

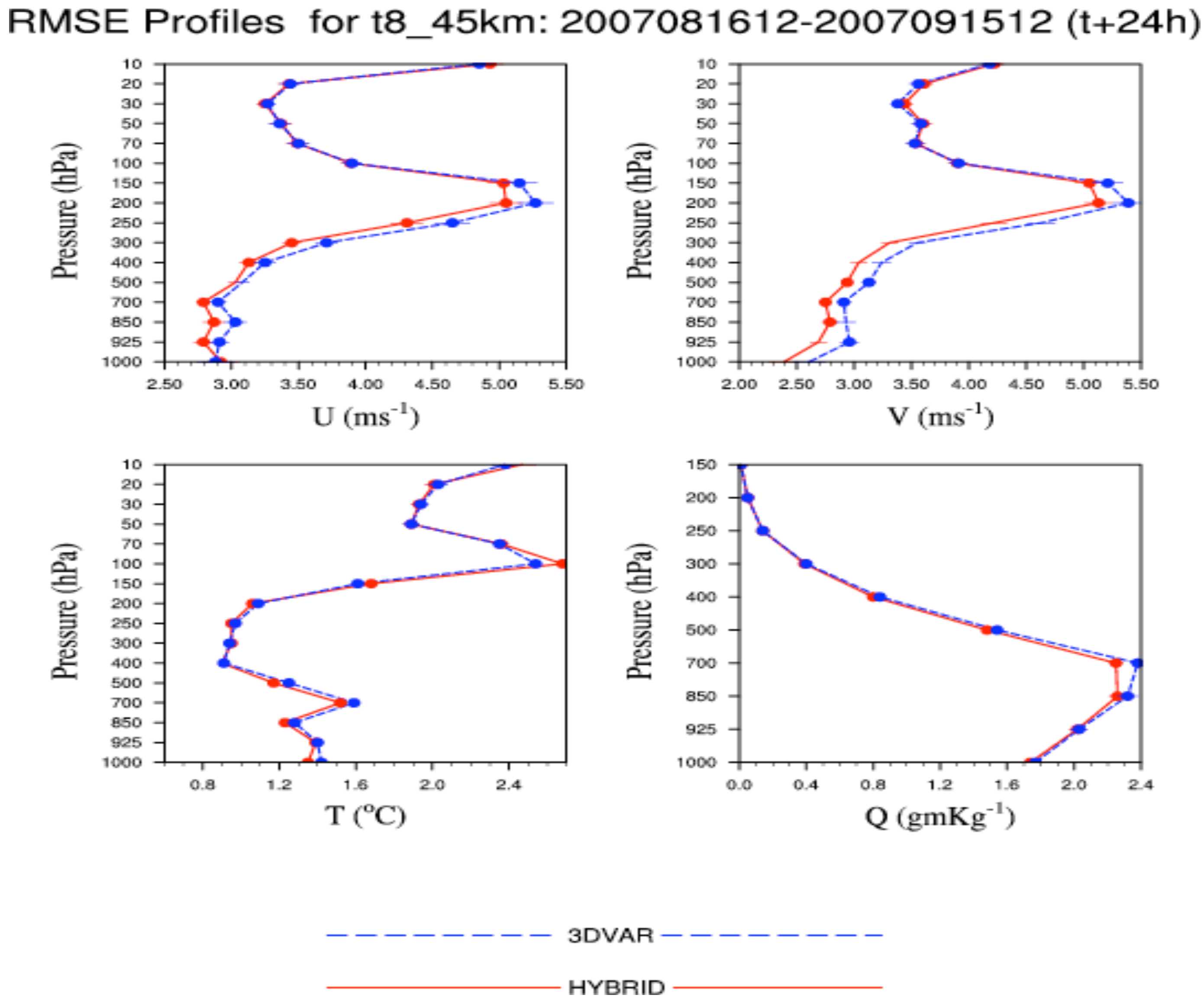
Ensemble Mean and Std. Deviation (spread)



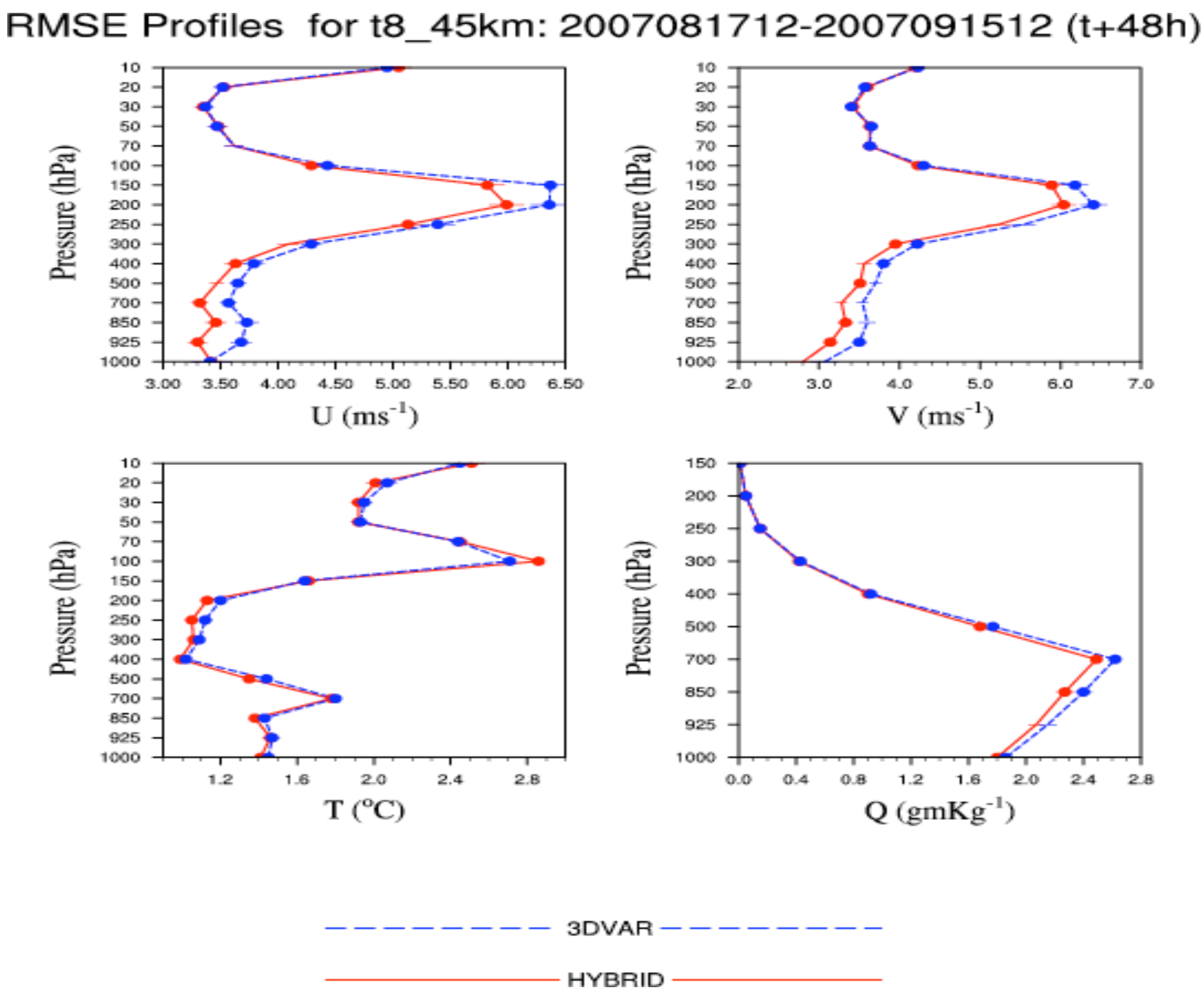
Inflation Factors (from 3-hourly cycling)



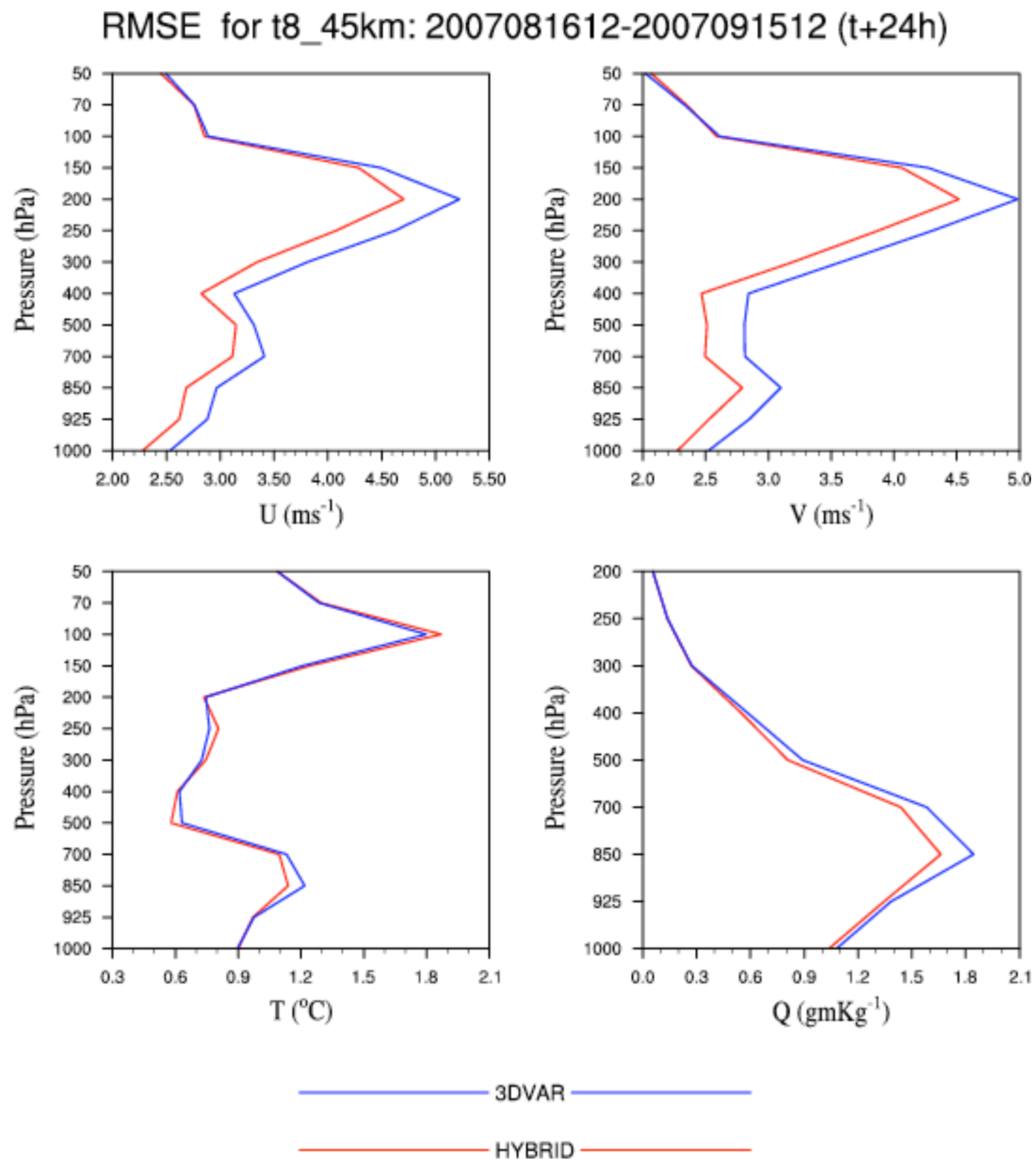
Stable, no-smoothing has been applied yet.



Hybrid gives better RMSE scores for wind compared to 3D-VAR.

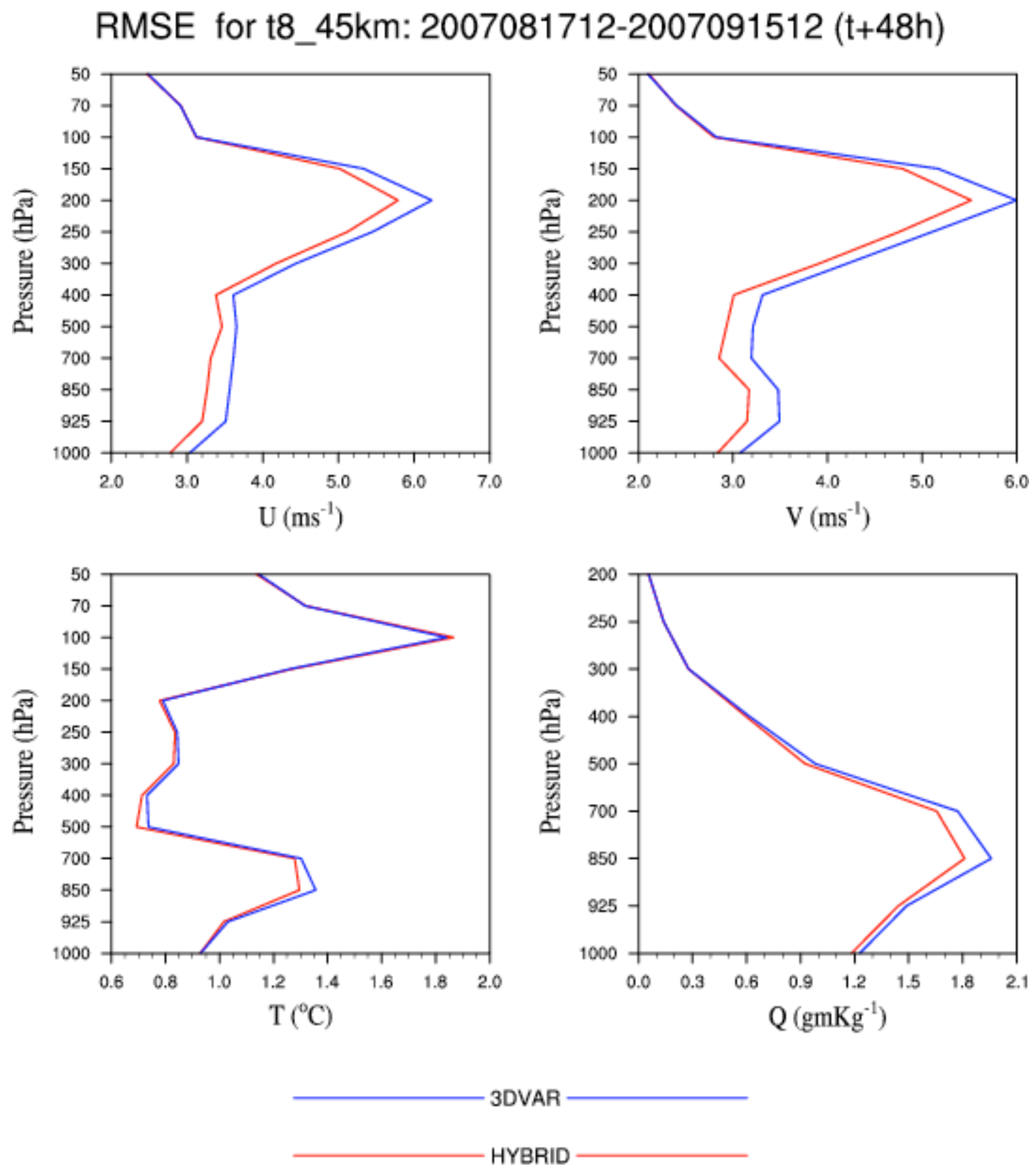


Hybrid gives better RMSE scores for wind compared to 3D-VAR.



ECMWF analysis data (T106) used.

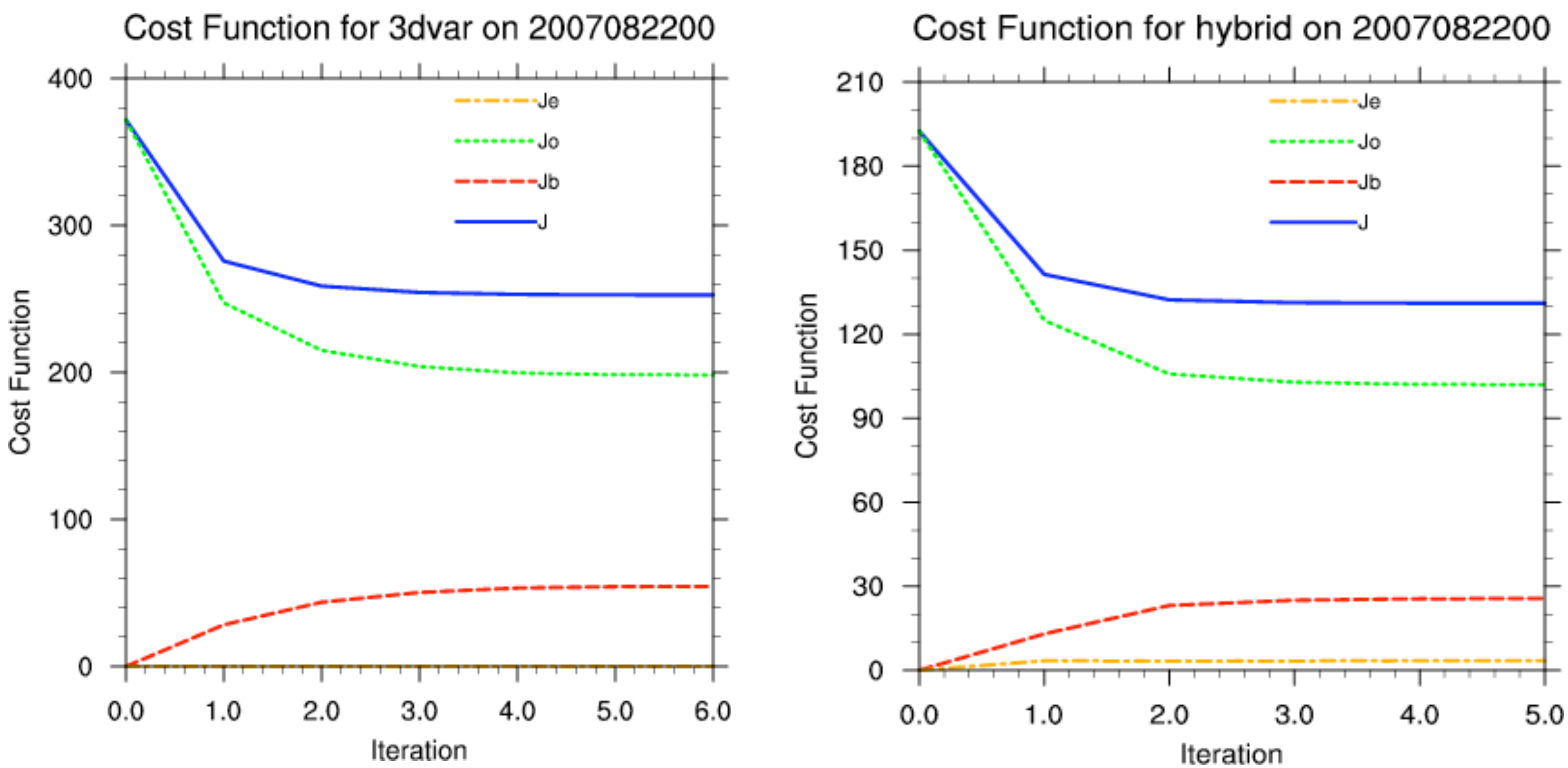
Hybrid gives better RMSE scores for wind compared to 3D-VAR.



ECMWF analysis data (T106) used.

Hybrid gives better RMSE scores for wind compared to 3D-VAR.

Cost function: 3DVAR and Hybrid



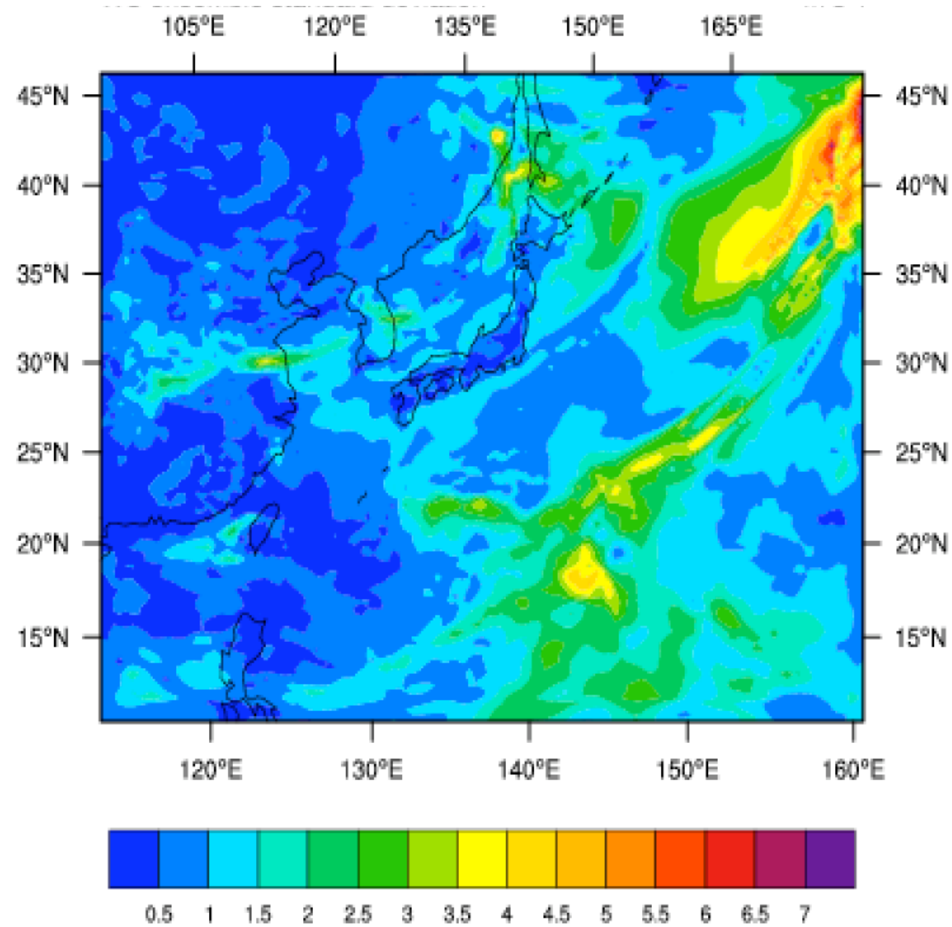
Cost function is smaller in hybrid.

Preliminary results from JME applications (snapshots)

Joint Mesoscale Ensemble (JME) Applications

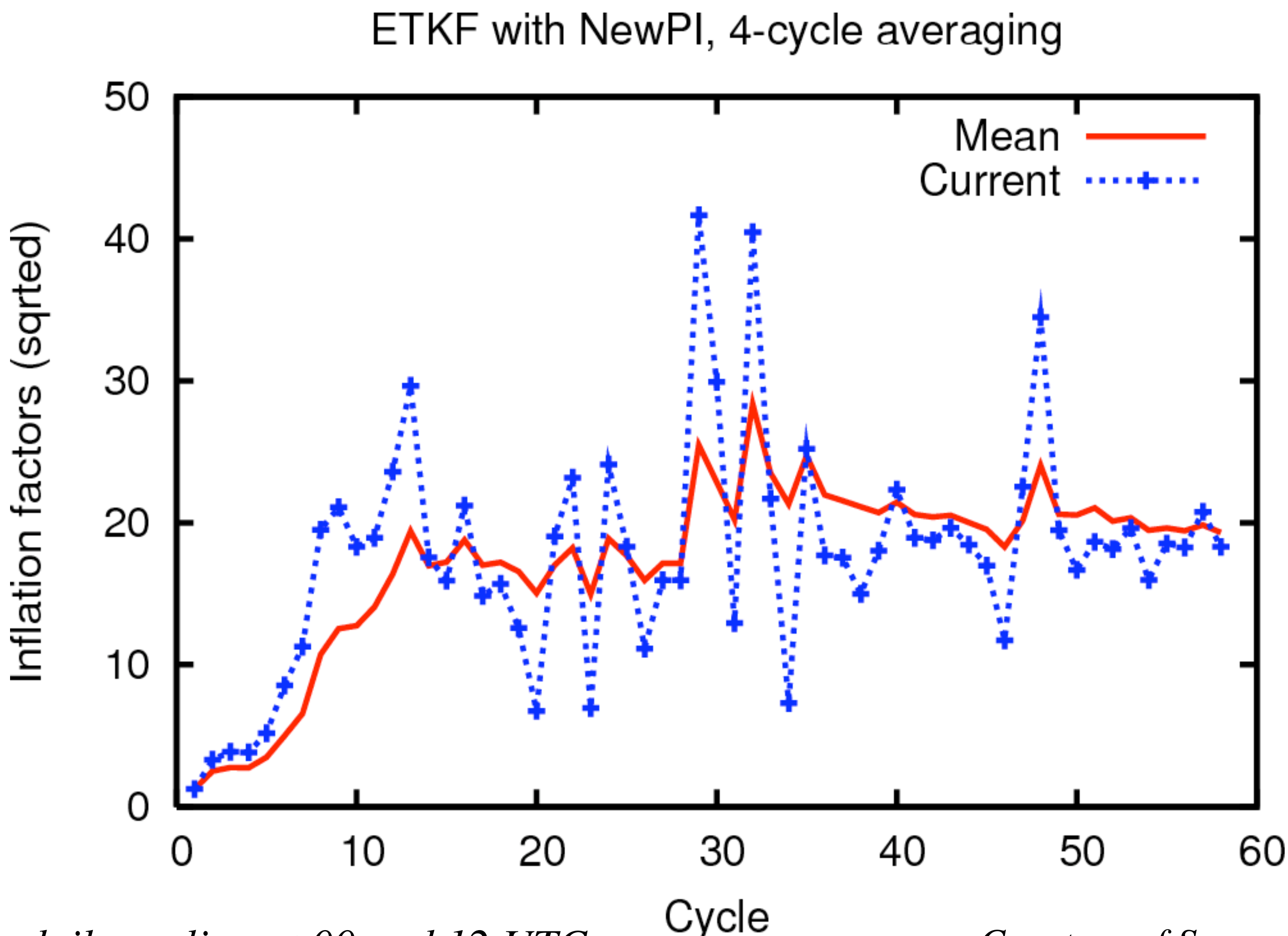
- 10 to 20 WRF physics configurations
- Capability for WRF-Var to update mean and/or individual members
- Capability for ETKF perturbations
- Lateral boundary conditions from global ensemble (GFS)
- Research on multi-parameter and stochastic approaches
- WRF-Var used to compute innovations.

10-m wind speed ensemble standard deviation (spread)



Courtesy of Josh Hacker

Inflation Factors Generated by JME



Twice daily cycling at 00 and 12 UTC

Courtesy of Soyoung Ha

Referred references

Bishop, C. H., B. J. Etherton, S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Weather Rev., 129, 420–436.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. Q. J. R. Meteorol. Soc. 134, 703–722.

Demirtas, M., D. Barker, Y. Chen, J. Hacker, X-Y. Huang, C. Snyder, and X. Wang, 2009: A Hybrid Data Assimilation System (Ensemble Transform Kalman Filter and WRF-VAR) Based Retrospective Tests With Real Observations. Preprints, the AMS 23rd WAF/19th NWP Conference, Omaha, Nebraska.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited area ensemble Kalman filters. *Mon. Wea. Rev.,* **134**, 2490-2502.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.

Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, 136, 5116-5131.

What is on the menu for the hybrid practice session

- **Computation:**
  - Computing ensemble mean.
  - Extracting ensemble perturbations (EP).
  - Running WRF-VAR in “hybrid” mode.
  - Displaying results for: ens_mean, std_dev, ensemble perturbations, hybrid increments, cost function and, etc.
  - If time permits, tailor your own test by changing hybrid settings; testing different values of “je_factor” and “alpha_corr_scale” parameters.

- **Scripts to use:**
  - Some NCL scripts to display results.

Brief information for the chosen case

Ensemble size: 10

Domain info:

- time_step=240,
- e_we=122,
- e_sn=110,
- e_vert=42,
- dx=45000,
- dy=45000,

Input data provided (courtesy of JME Group):

- WRF ensemble forecasts valid at 2006102800
- Observation data (ob.ascii) for 2006102800
- 3D-VAR “be.dat” file

Thanks for attending…..