Hybrid Variational/Ensemble Data Assimilation

Zhiquan Liu (liuz@ucar.edu)

NCAR/MMMM

Acknowledgements: D. M. Barker, M. Demirtas, A. P. Mizzi

Outline

- Motivation and differently proposed hybrid DA
- Elements of hybrid DA
- Preliminary results
- Introduction to practice

Why Hybrid?

- 3D-Var uses static ("climate") BE

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

- 4D-Var implicitly uses flow-dependent information, but still starts from static BE

$$\begin{aligned} J_j(\delta \mathbf{x}_j) &= \frac{1}{2}(\delta \mathbf{x}_j - \delta \mathbf{x}_j^b)^T \mathbf{B}^{-1}(\delta \mathbf{x}_j - \delta \mathbf{x}_j^b) \\ &+ \frac{1}{2}\sum_{k=0}^{K}(\mathbf{H}_{j,k}\mathbf{M}_{j,k}\delta \mathbf{x}_j - \mathbf{d}_{j,k})^T \mathbf{R}^{-1}(\mathbf{H}_{j,k}\mathbf{M}_{j,k}\delta \mathbf{x}_j - \mathbf{d}_{j,k}) \end{aligned}$$

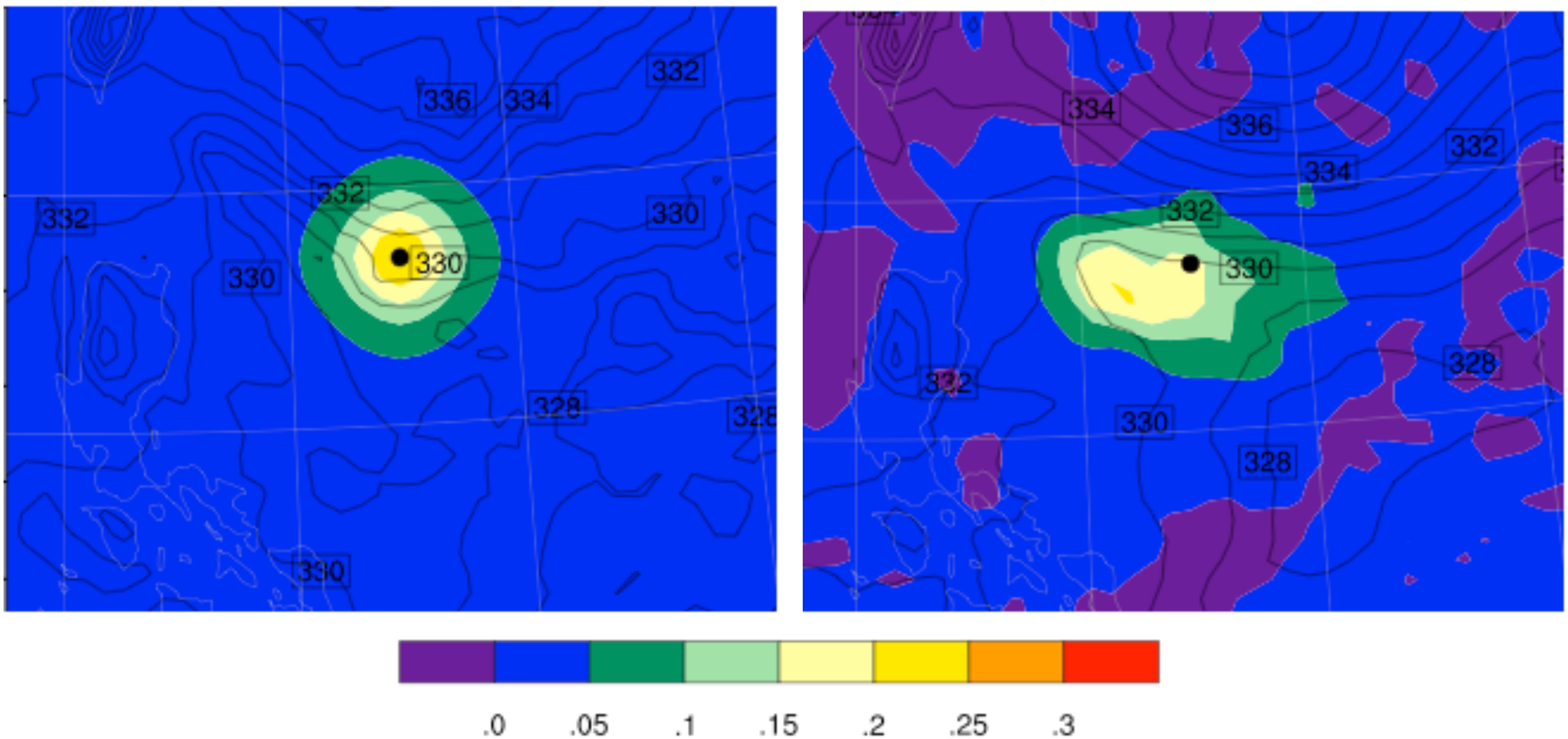
- Hybrid: using flow-dependent background error information from ensemble in a variational DA system

T Analysis increments from a single T obs

1K difference, 1K error

3DVAR

Hybrid (64 members)



What is the Hybrid DA?

- Combine ensemble and variational DA together

- Ensemble mean is analyzed by a variational algorithm (i.e., minimize a cost function). It combines the 3DVAR “climate” background error covariance and “error of the day” from ensemble.

- A system for updating ensemble
  - Could be (independent) ensemble forecasts already available from NWP centers
  - Could be an EnKF-based DA system
  - Could be an ETKF-based ensemble system

Hamill and Snyder, 2000

- 3DVAR cost function

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

- Idea: replace B by a weighted sum of 3DVAR B and the ensemble covariance

$$\mathrm{B} = \alpha_1 \mathrm{B}_1 + \alpha_2 \mathrm{B}_2, \ \alpha_1 = 1 - \alpha_2$$

Lorenc, 2003

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables (Lorenc, 2003)

$$J(x,\alpha) = \beta_1 J_b + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \overbrace{\beta_2 \frac{1}{2}\alpha^T A^{-1}\alpha}^{\textit{extended}\text{ control variables}} + \frac{1}{2}[y - H(x + x_e)]^T R^{-1}[y - H(x + x_e)]$$

- This is implemented in WRFDA (Wang et al., 2008)

- It is mathematically equivalent to Hamill and Snyder (2000).

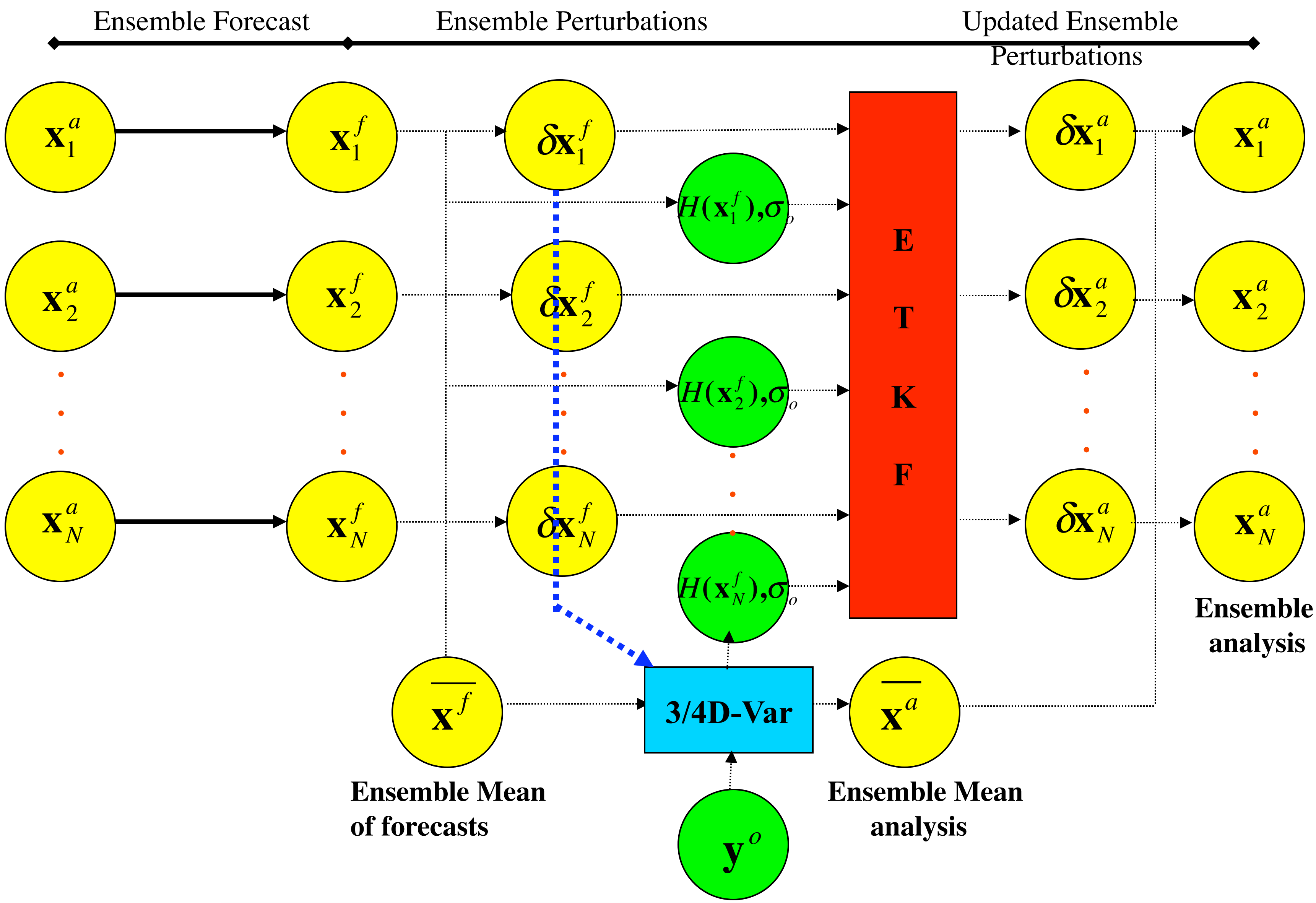
Advantages of the Hybrid DA?

- Hybrid DA system can be more robust than a pure EnKF-based DA
  - For some observations type, e.g., radiances, localization is not well defined in observation space, bias correction issues
  - Localization is in model space in a variational framework.
  - For small-size ensemble since can adjust amount of 3DVAR and ensemble covariances.

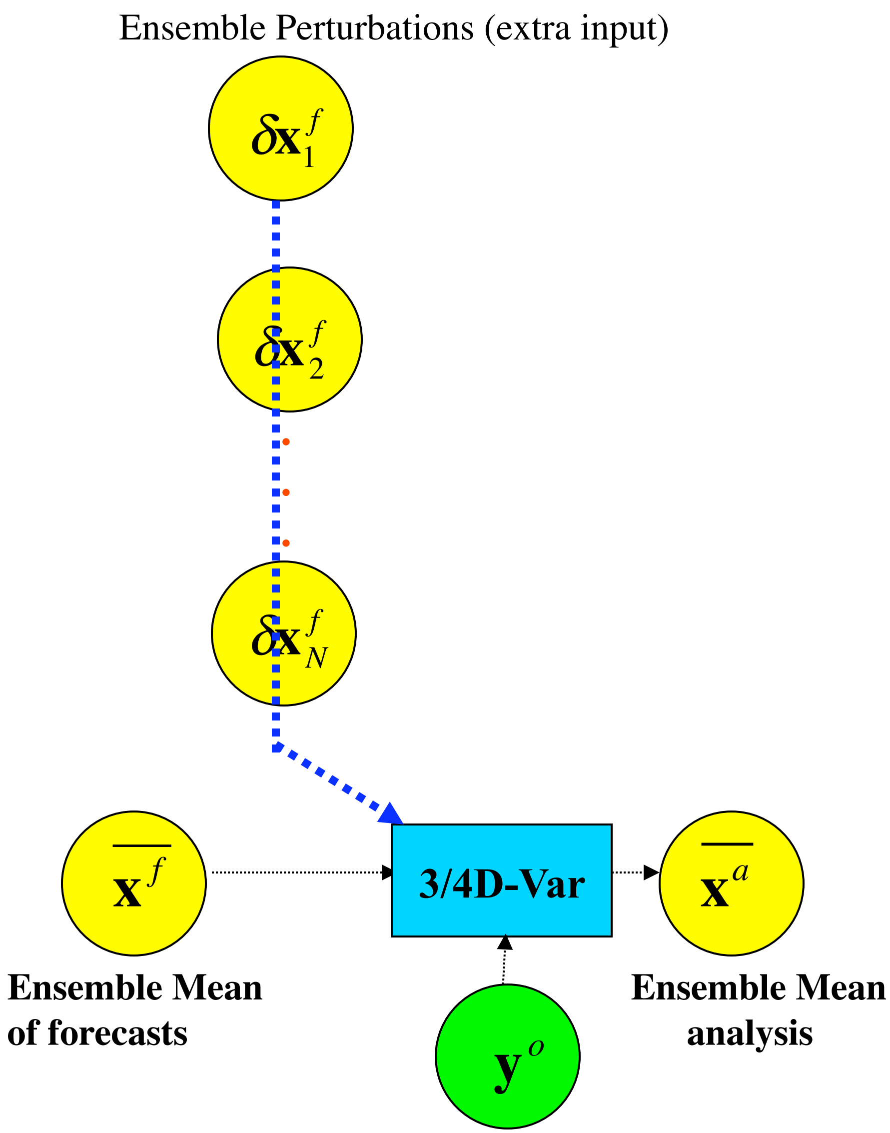
Elements of Hybrid DA

- Ensemble forecasts: *WRF-ensemble forecasts*
- Ensemble Transform Kalman Filter (ETKF):
  - Update forecast/background ensemble perturbations to analysis ensemble perturbations
- A Variational DA to update ensemble mean.

Cycling WRF/WRFDA/ETKF System (Hybrid DA)



Hybrid DA: Variational Part



Hybrid 3DVAR formulation

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables (Lorenc, 2003; Wang et al., 2008)

$$J(x,\alpha) = \beta_1 J_b + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \overbrace{\beta_2 \frac{1}{2}\alpha^T A^{-1}\alpha}^{\textit{extended}\text{ control variables}} + \frac{1}{2}[y - H(x + x_e)]^T R^{-1}[y - H(x + x_e)]$$

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1,\ \ x_e = \frac{1}{\sqrt{\mathrm{N-1}}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \alpha_i \bullet x_i^{'},\ \ \text{where } x_i^{'} \text{ is the ensemble perturbation for the member i.}$$

extended control variable $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)$ has dimension of

M(dimension of x)$\times$N (ensemble size)

the matrix A plays the role for ensemble covariance localization.

$$\mathrm{A} = \begin{pmatrix} S & & \\ & \ddots & \\ & & S \end{pmatrix},\ \mathrm{S} = \langle (\alpha_i)(\alpha_i)^{\mathrm{T}} \rangle$$

Hybrid 3DVAR formulation

- Equivalently can write in another form (Wang et al., 2008)

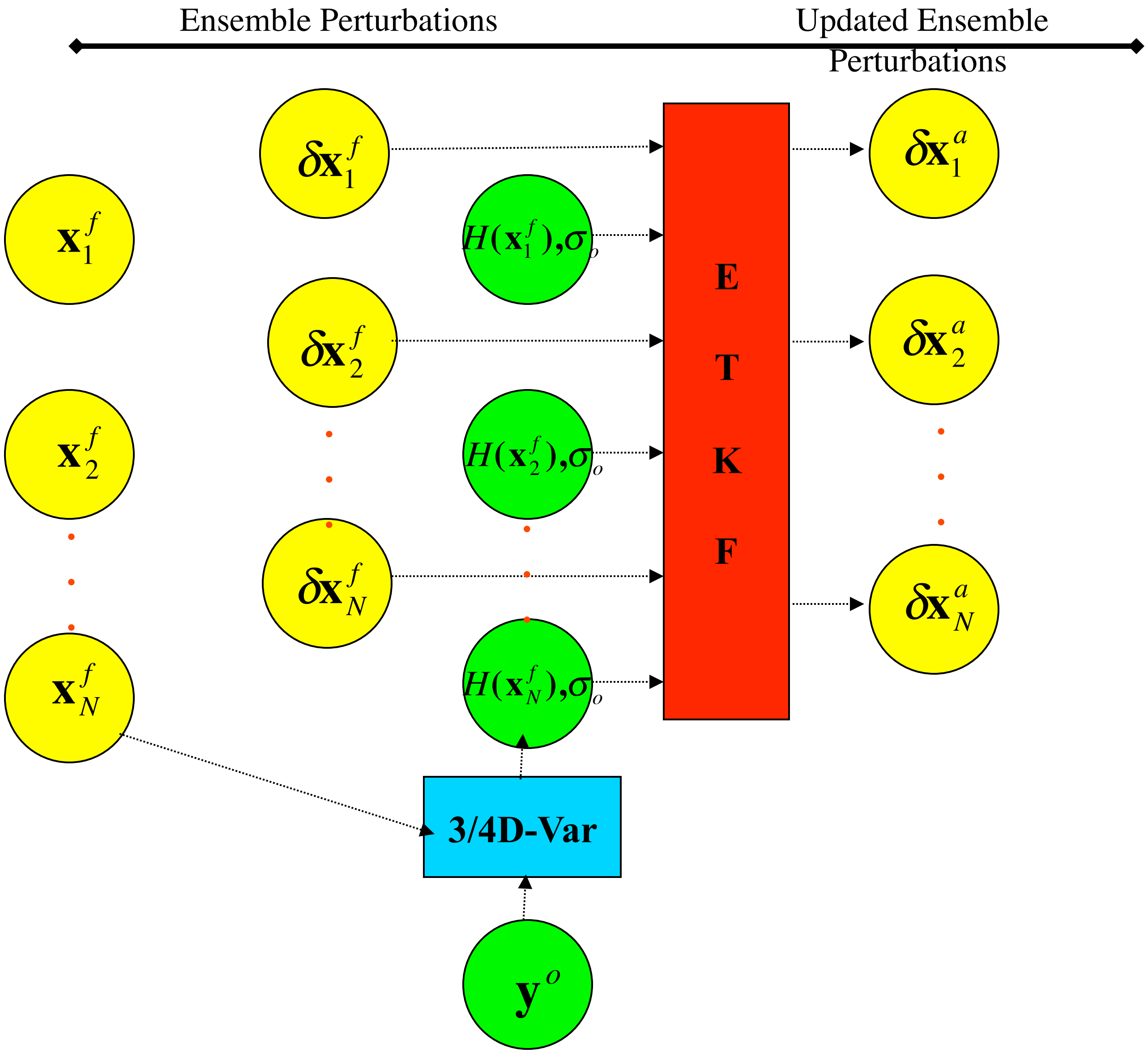
$$J(x,\alpha) = \frac{1}{2}(x + x_e - x_b)^T (\frac{1}{\beta_1} B + \frac{1}{\beta_2} P^e \circ S)^{-1} (x + x_e - x_b)$$
$$+ \frac{1}{2}[y - H(x + x_e)]^T R^{-1} [y - H(x + x_e)]$$

$P^e = \frac{1}{N-1}(x')(x')^T$ is the sample ensemble covariances.

- This explains why S is for localization.

- This is also equivalent to Hamill and Snyder (2000).

Hybrid DA: Ensemble Part (ETKF-based)



ETKF formulation

- The ETKF (Wang et al. 2007) finds the transformation matrix **T** to update forecast/background perturbations to analysis perturbations

$$\delta \mathbf{x}^a = \delta \mathbf{x}^f \mathbf{T}$$

$$\mathbf{T} = r\mathbf{E}(\rho\lambda + \mathbf{I})^{\mathbf{-1/2}}\mathbf{E}^{\mathbf{T}}$$

- Where **E and λ** contain eigenvectors and eigenvalues of a NxN (N is ensemble size) matrix

$$[\mathbf{H}(\delta \mathbf{x}^{\mathbf{f}})]^{\mathbf{T}}\mathbf{R}^{\mathbf{-1}}[\mathbf{H}(\delta \mathbf{x}^{\mathbf{f}})]/(N-1)$$

$$\mathbf{H}(\delta \mathbf{x}_k^f) = H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}_k^f)}$$

Inflation

- r inflation factor, ρ accounts for the fraction of the forecast-error variance projected onto the ensemble subspace.
  - Both factors are adaptively calculated for each DA cycle by using innovation statistics.

- Inflation is to ensure that on average the background error variance estimated from the spread of ensembles is consistent with innovation statistics, i.e.,

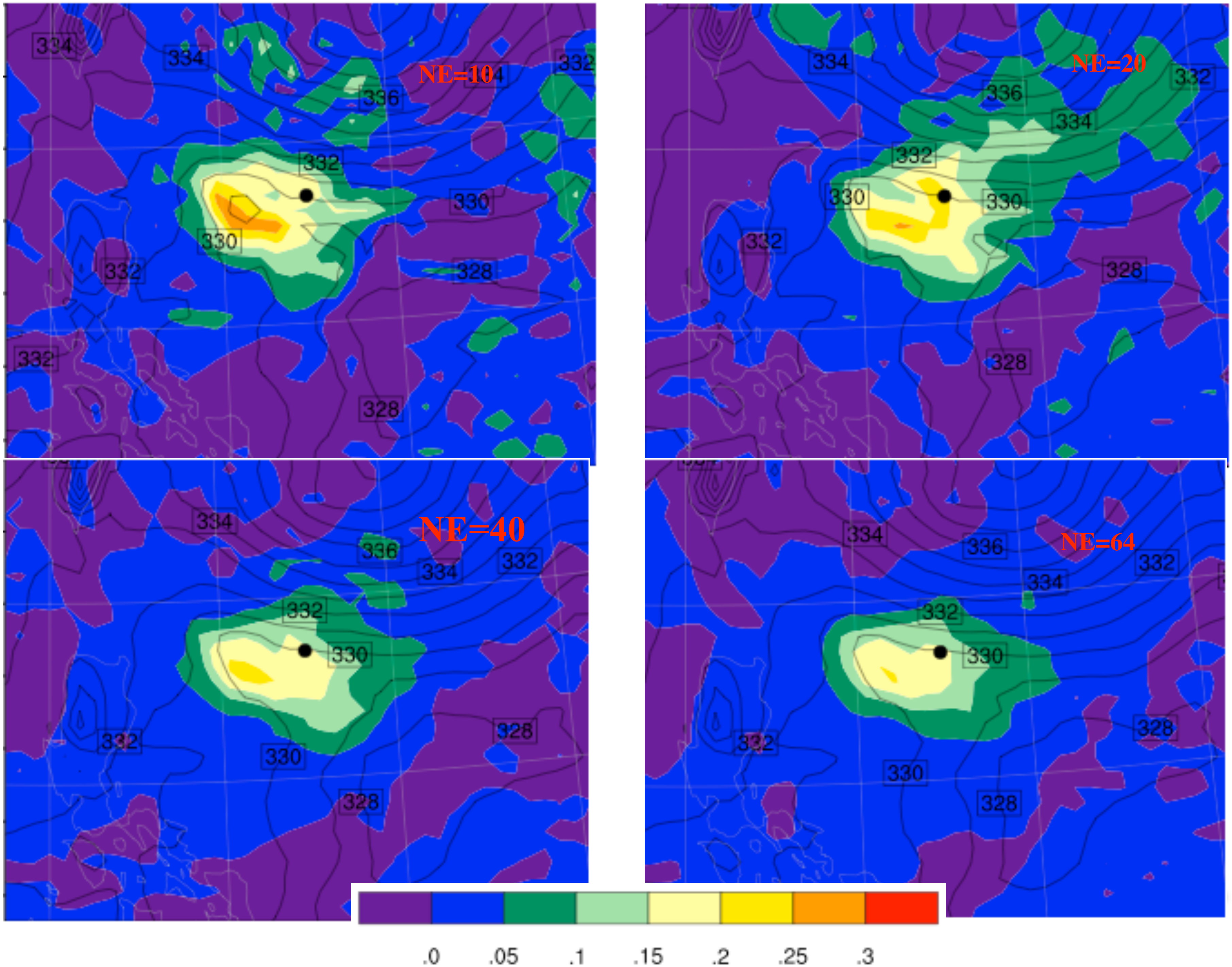
$$d^{\mathrm{T}}\mathrm{R}^{-1}d \approx trace\left(\frac{1}{N-1}\sum_{i=1}^{N}[H(x_i)-H(\bar{x})]\mathrm{R}^{-1}[H(x_i)-H(\bar{x})]+\mathrm{I}\right)$$

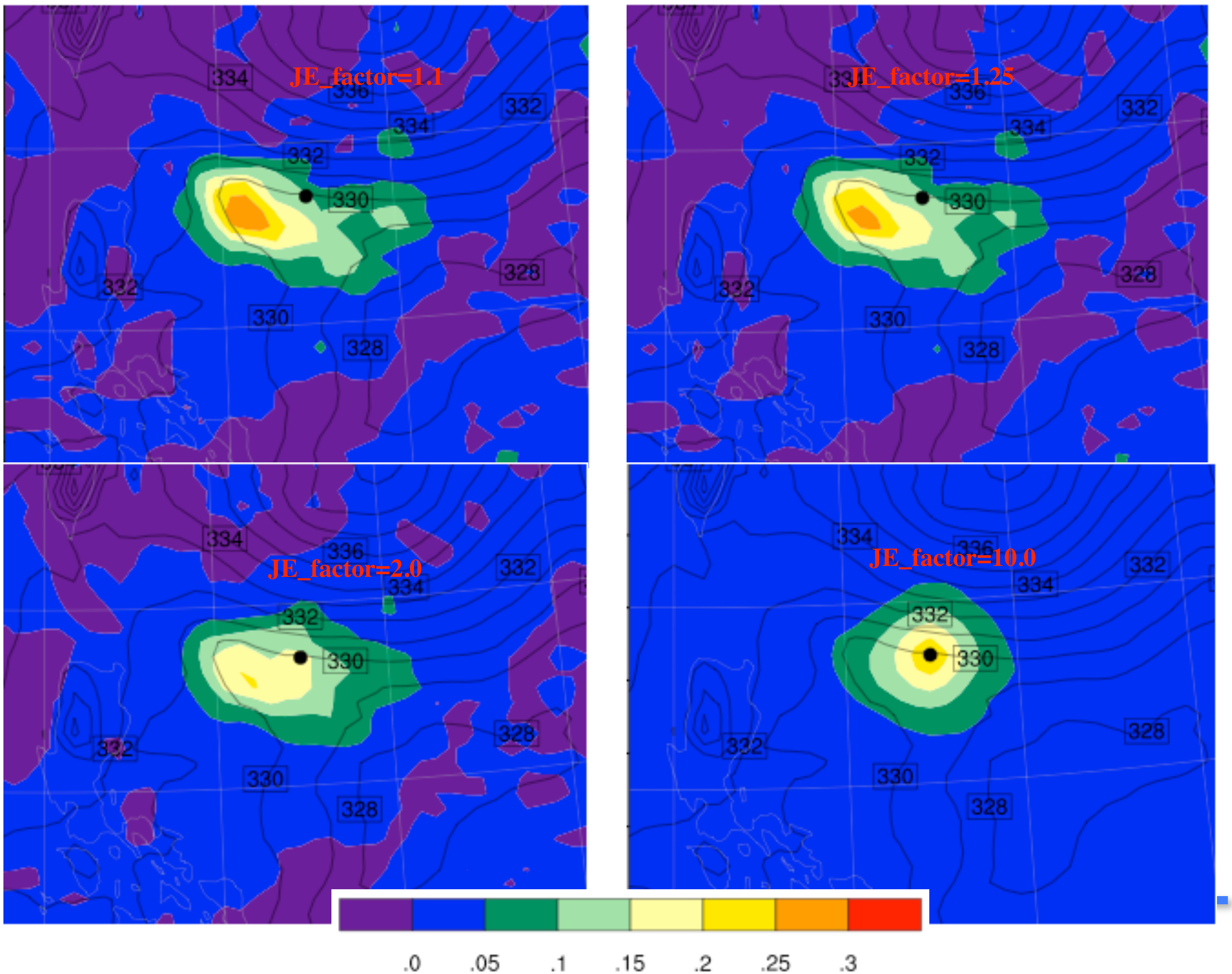
Pros and Cons of ETKF

- Desirable aspects:
  - ETKF is fast (computations are done in model ensemble perturbation subspace).
  - It directly updates perturbations.

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

- Alternatives for ensemble part
  - EnKF, Perturbed obs, LETKF


NE=10
NE=20
NE=40
NE=64
334
336
332
330
328
.0
.05
.1
.15
.2
.25
.3

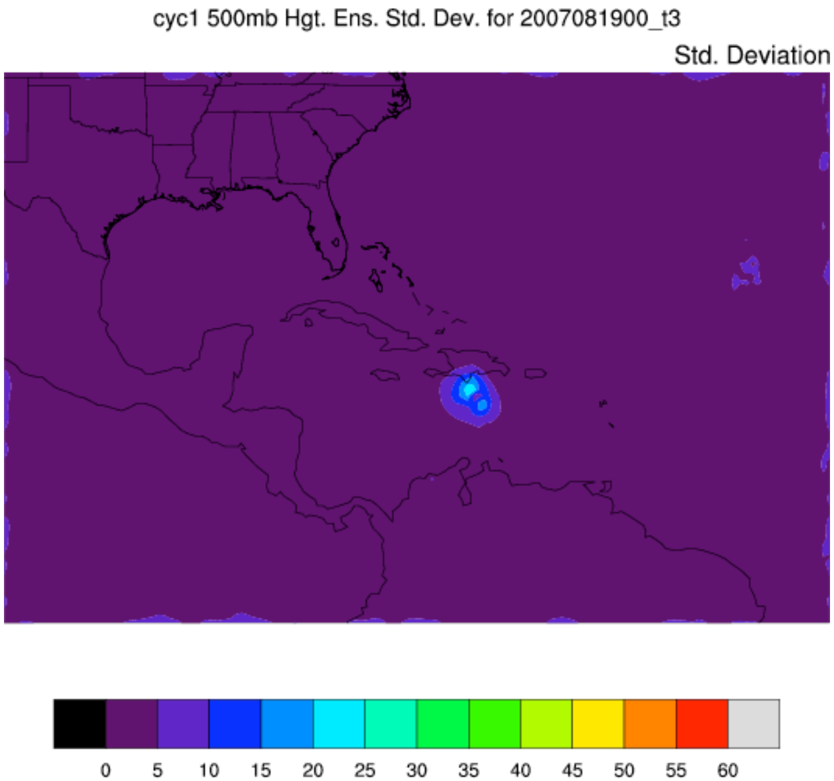

JE_factor=1.1
JE_factor=1.25
JE_factor=2.0
JE_factor=10.0
334
336
332
334
332
330
328
332
330
328
.0
.05
.1
.15
.2
.25
.3

Old results (need to update in the future)

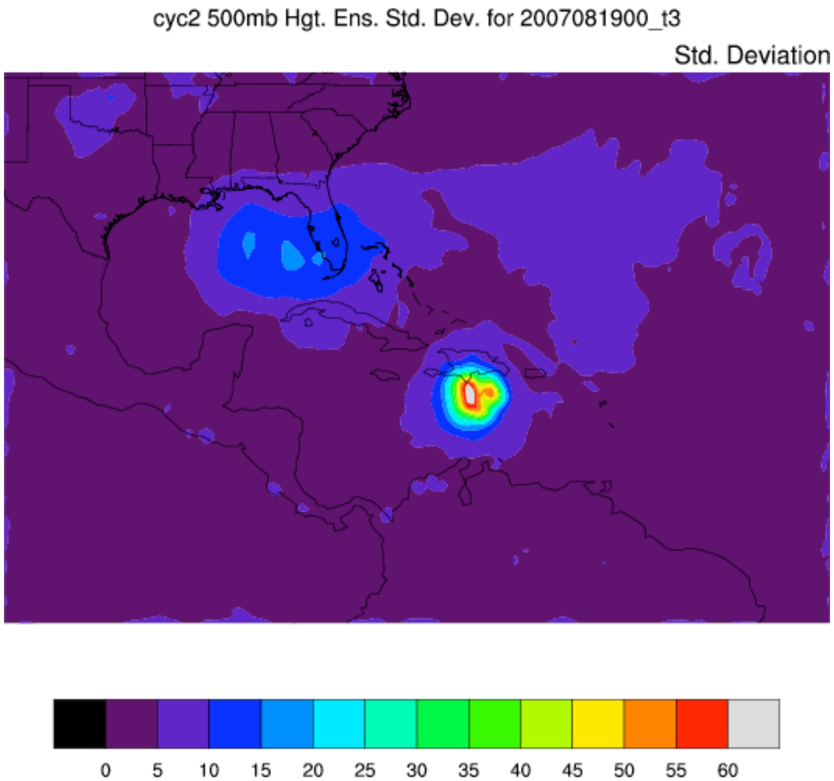
- Ensemble size: 10
- Test Period: 15th August - 15th September 2007
- Cycle frequency: 3 hours
- Observations: GTS conventional observations
- Deterministic ICs/BCs: Down-scaled GFS forecasts
- Ensemble ICs/BCs: Produced by adding spatially correlated Gaussian noise to GFS forecasts.
- Horizontal resolution: 45km
- Number of vertical levels: 57
- Model top: 50 hPa

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

WRF t+3 valid at 2007081900

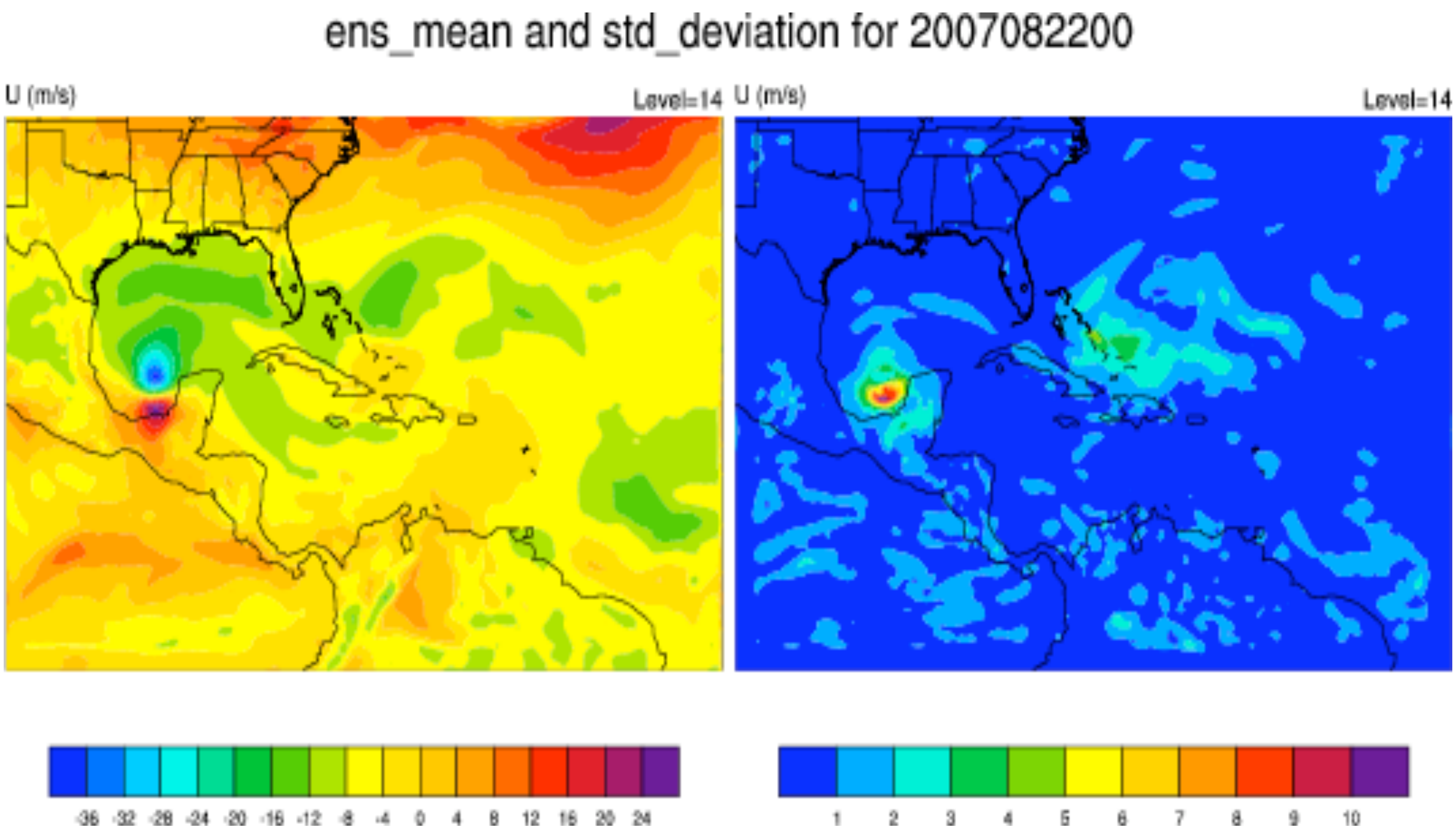


Modest inflations factors used

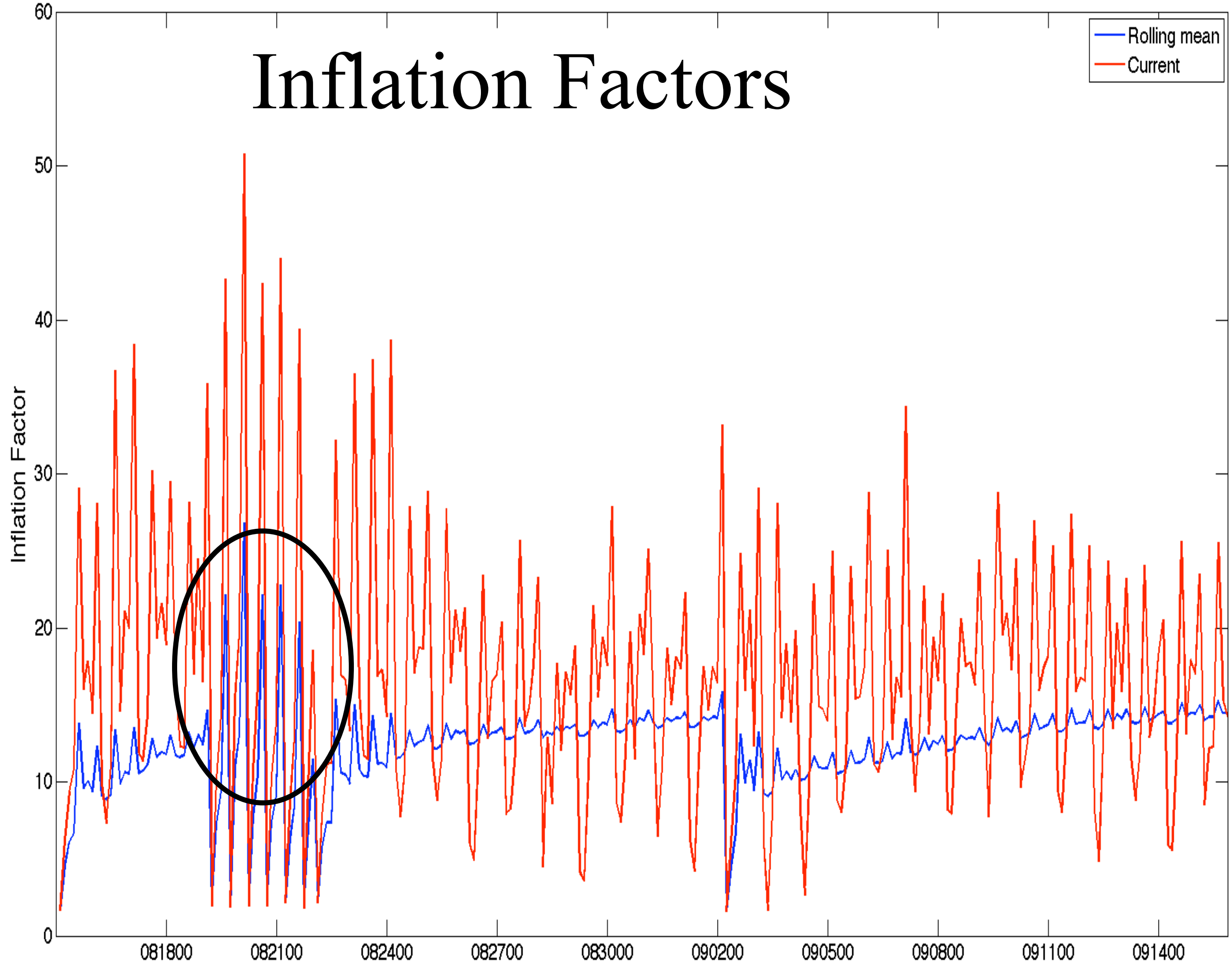


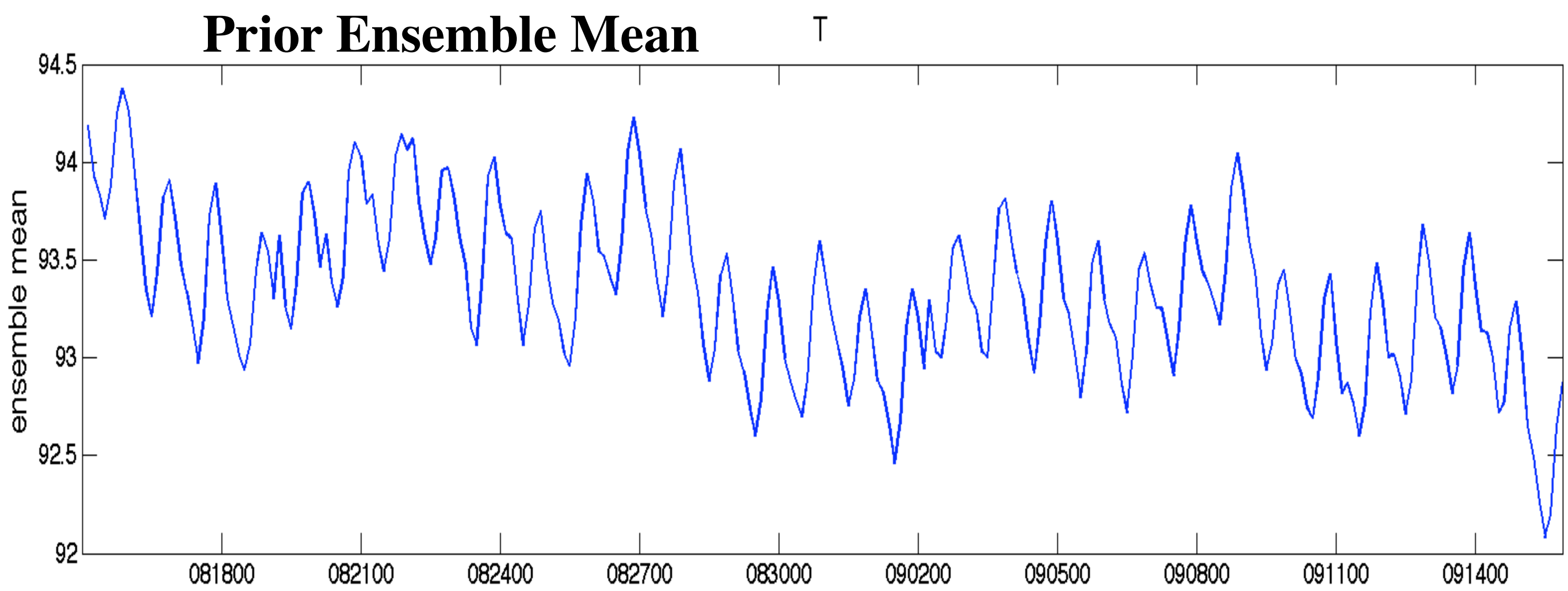
Higher inflations factors used

Ensemble Mean and Std. Deviation (spread)

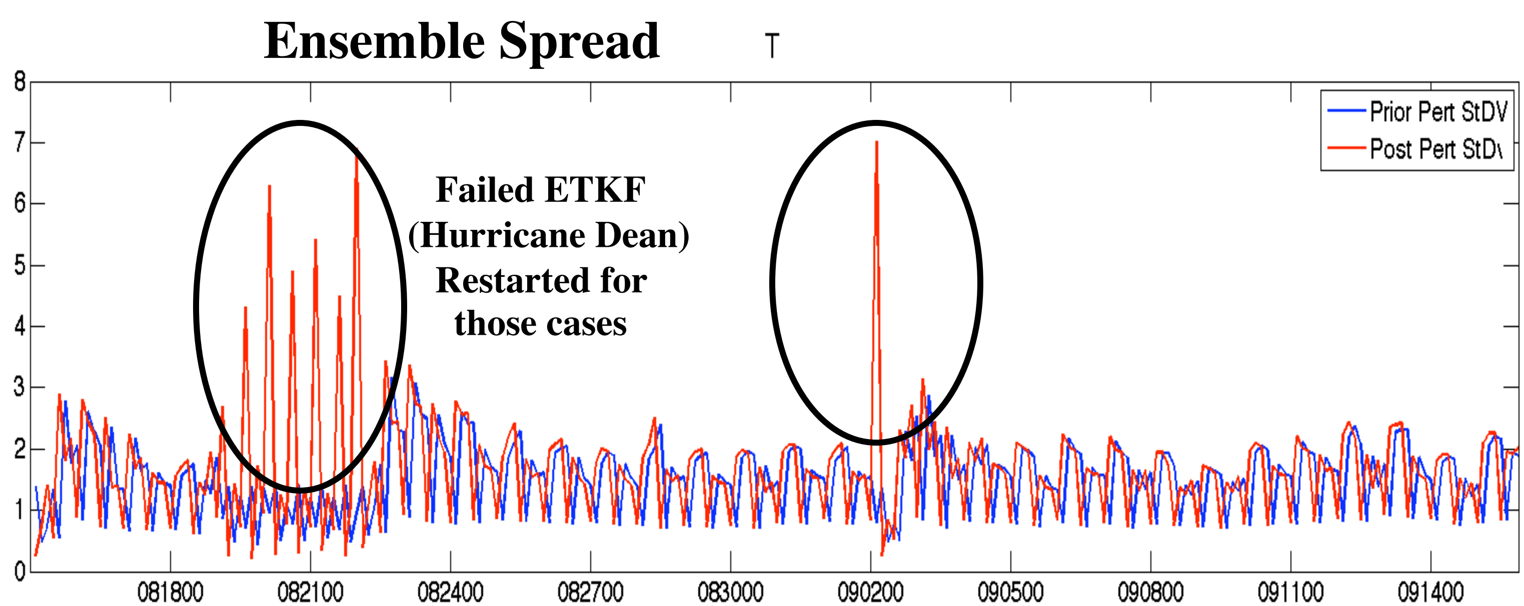


Inflation Factors

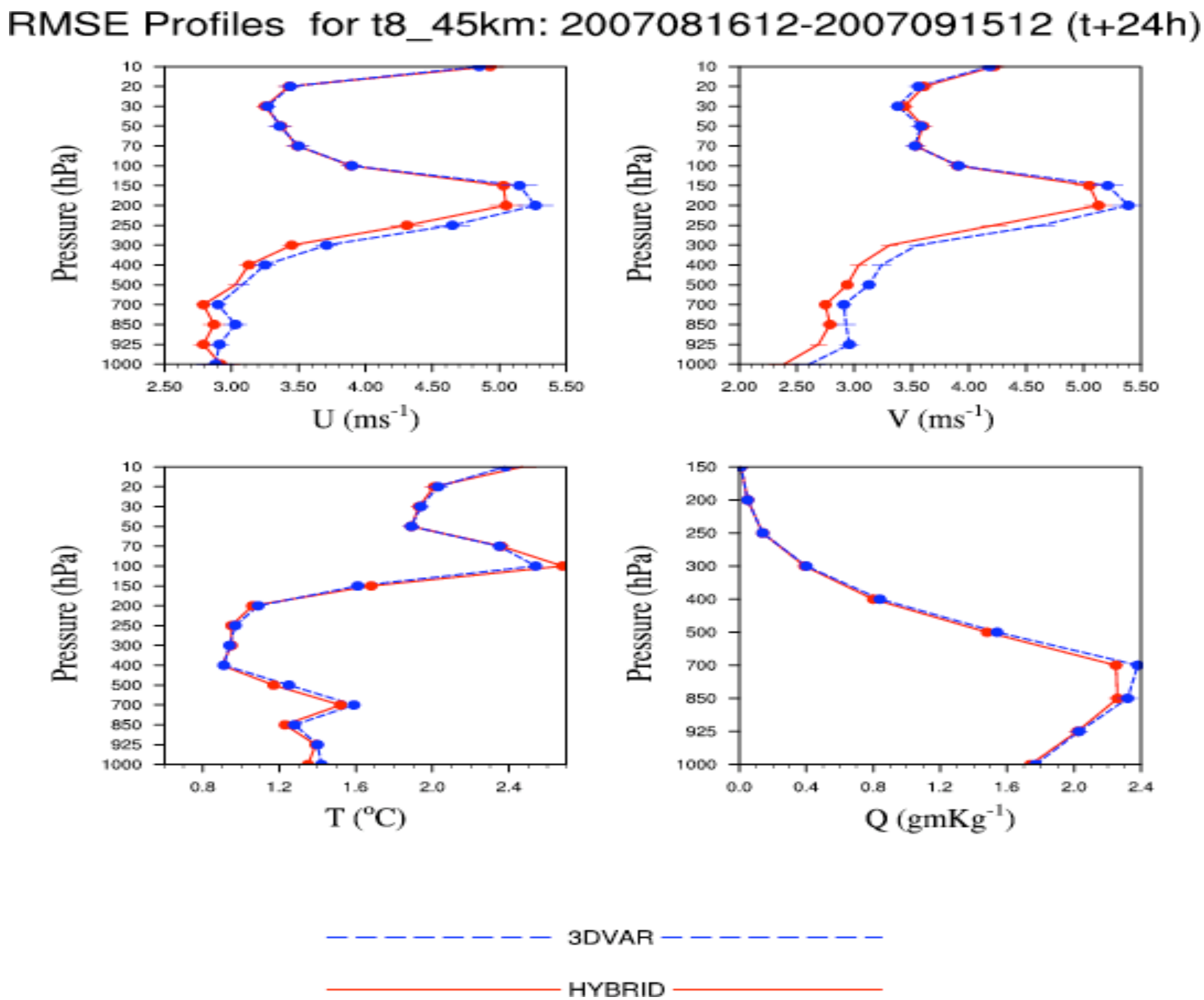




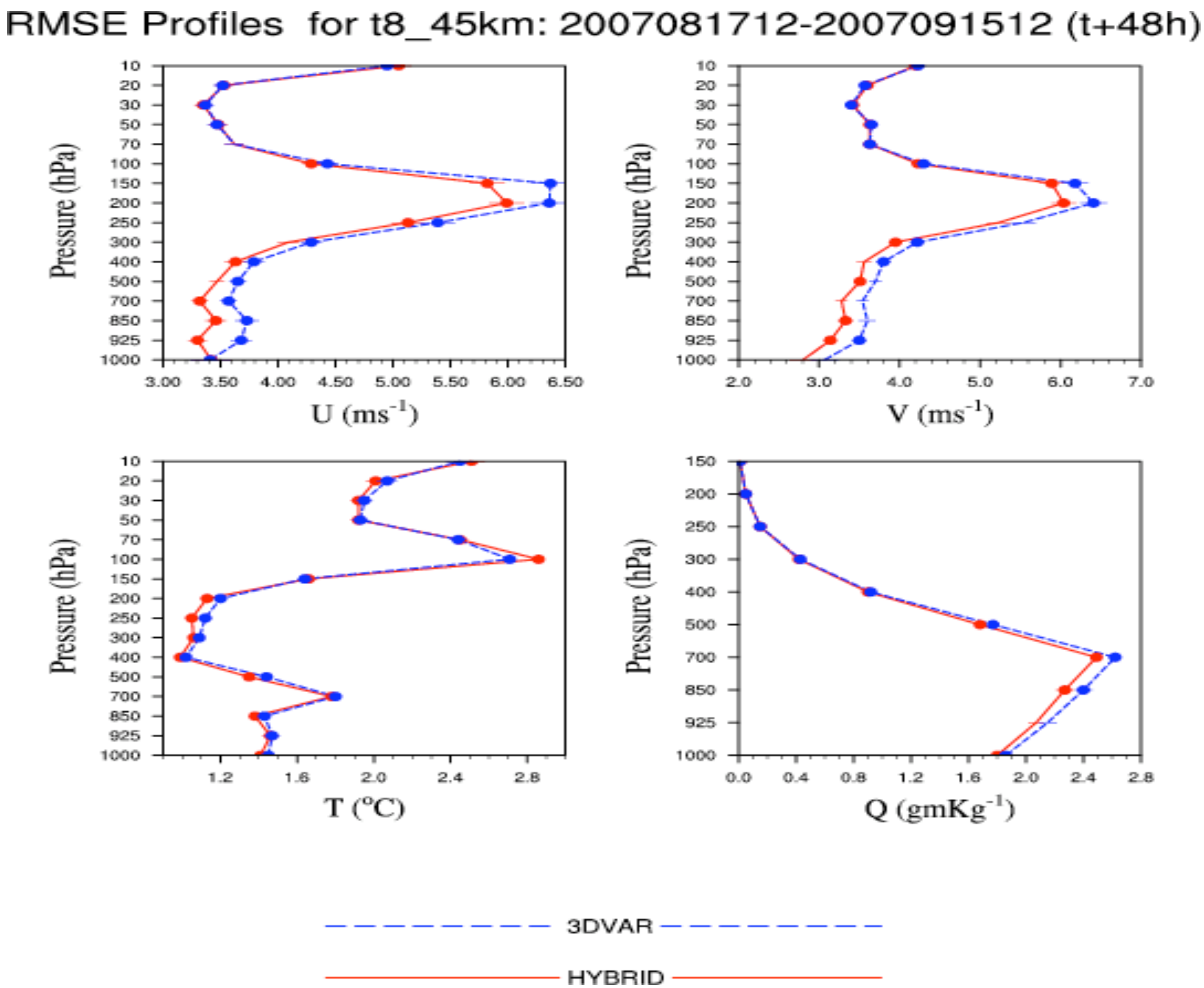
Prior Ensemble Mean
T
ensemble mean
94.5
94
93.5
93
92.5
92
081800
082100
082400
082700
083000
090200
090500
090800
091100
091400



Ensemble Spread
T
Prior Pert StDV
Post Pert StD\
Failed ETKF
(Hurricane Dean)
Restarted for
those cases
8
7
6
5
4
3
2
1
0
081800
082100
082400
082700
083000
090200
090500
090800
091100
091400



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



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

Introduction of Hybrid practice session

- **Computation:**
  - Computing ensemble mean.
  - Extracting ensemble perturbations (EP).
  - Running WRFDA 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.
- **Ensemble generation part not included in current practice**

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

References

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.

T. M. Hamill and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. Mon. Wea. Rev., 128, 2905–2919.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. Quart. J. Roy. Me- teor. Soc., 129, 3183–3203.

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.