Hybrid Variational/Ensemble Data Assimilation

Zhiquan Liu (liuz@ucar.edu)

NCAR/MMM

Slides from D. M. Barker, M. Demirtas, Y. Chen, X. Wang

Outline

- Motivation of hybrid DA
- Elements of hybrid DA
- Preliminary results
- Alternative for updating perturbations
- Practice introduction

Why Hybrid?

- Background errors are flow-dependent
  - 3D-Var uses static BE
  - 4D-Var implicitly uses flow-dependent information, but still starts from static BE
  - Hybrid: using flow-dependent information from ensemble in WRF-Var
- Hybrid can be robust for small size ensembles.

Flow-Dependent Ensemble BE Covariances

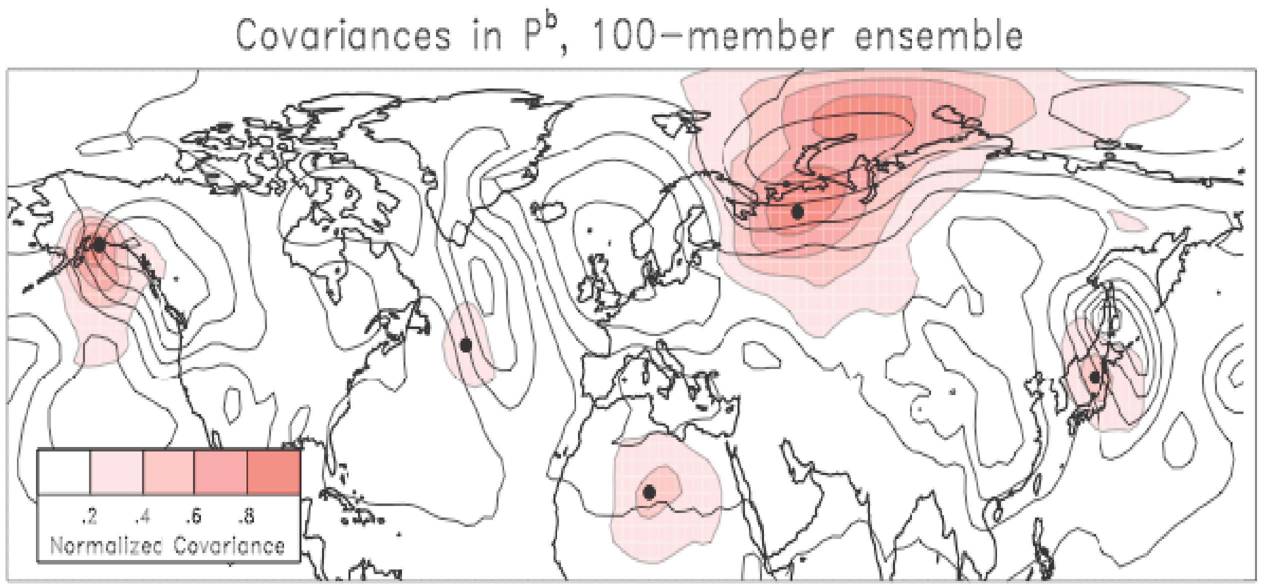


Figure 2. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

(Hamill 2006)

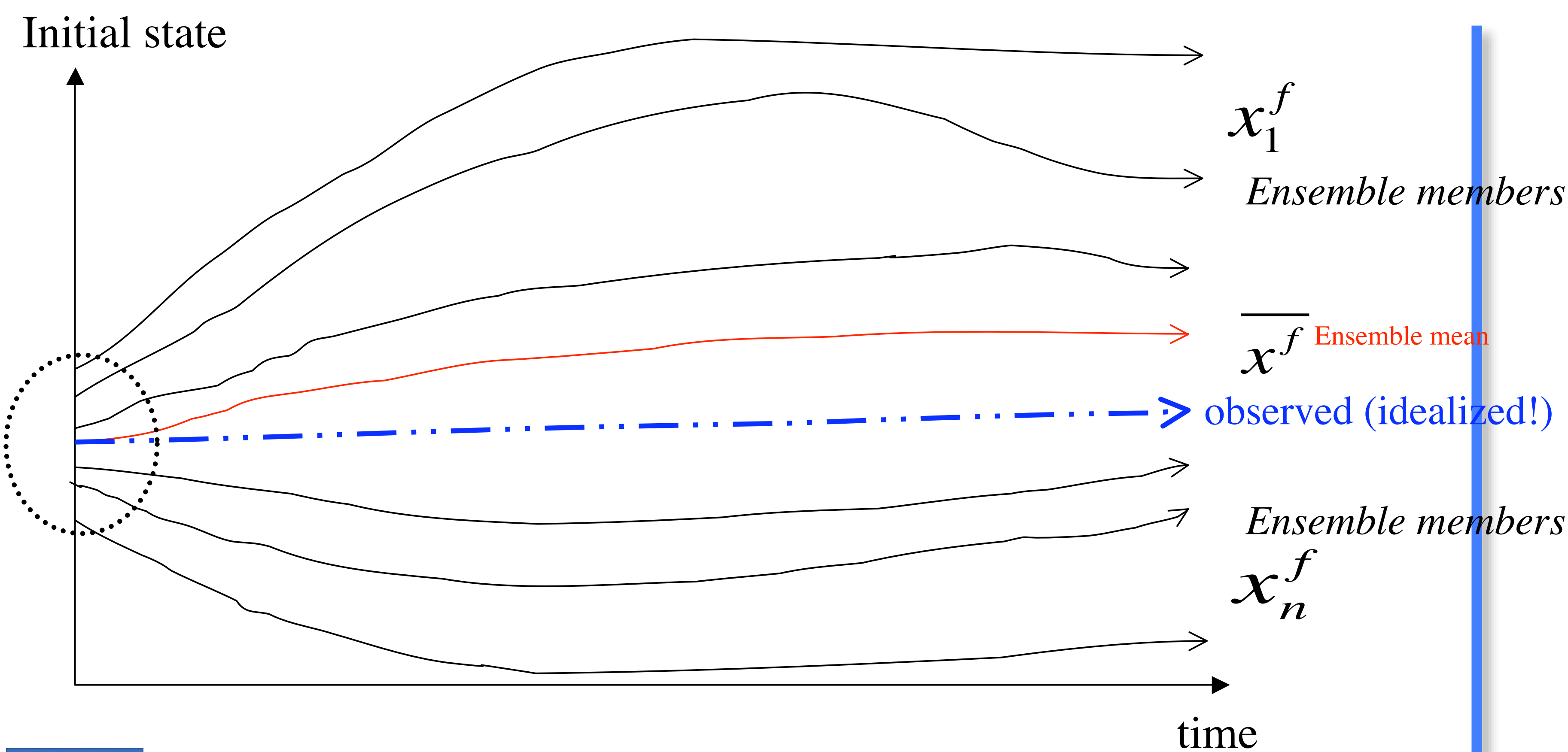
Elements of Hybrid DA

- 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.

Ensembles to address uncertainties in the initial state



Ensemble Mean and Perturbations

Assume the following ensemble forecasts:

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

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

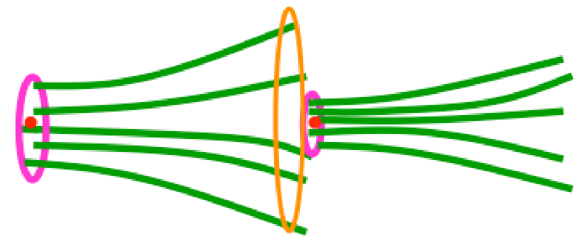
Ensemble perturbations: $\delta x_n^f = x_n^f - \overline{x^f}$

Forecast error covariance: $\mathbf{P}^f = \frac{1}{N-1}(\delta\mathbf{X}^f)(\delta\mathbf{X}^f)^{\mathrm{T}}$

ETKF formulation

- The ETKF (Bishop et al. 2001) finds the transformation matrix **T** to update forecast/background perturbations to analysis perturbations

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



- So that the analysis error covariance

$$\mathbf{P}^a = \frac{1}{N-1}(\delta\mathbf{X}^a)(\delta\mathbf{X}^a)^{\mathrm{T}}$$

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

- Where **C** contains eigenvectors and **Γ** eigenvalues of a NxN (N is ensemble size) matrix

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

Inflation factor in ETKF

- Analysis error variance is usually underestimated from ETKF due to sampling error, i.e., analysis perturbations' spead is too small.
  - So perturbations need to be inflated by a factor ∏

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

  - Inflation factor is estimated from innovation vector (Wang and Bishop, 2003).

Hybrid DA formulation

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables.

Extra term associated with extended control variable

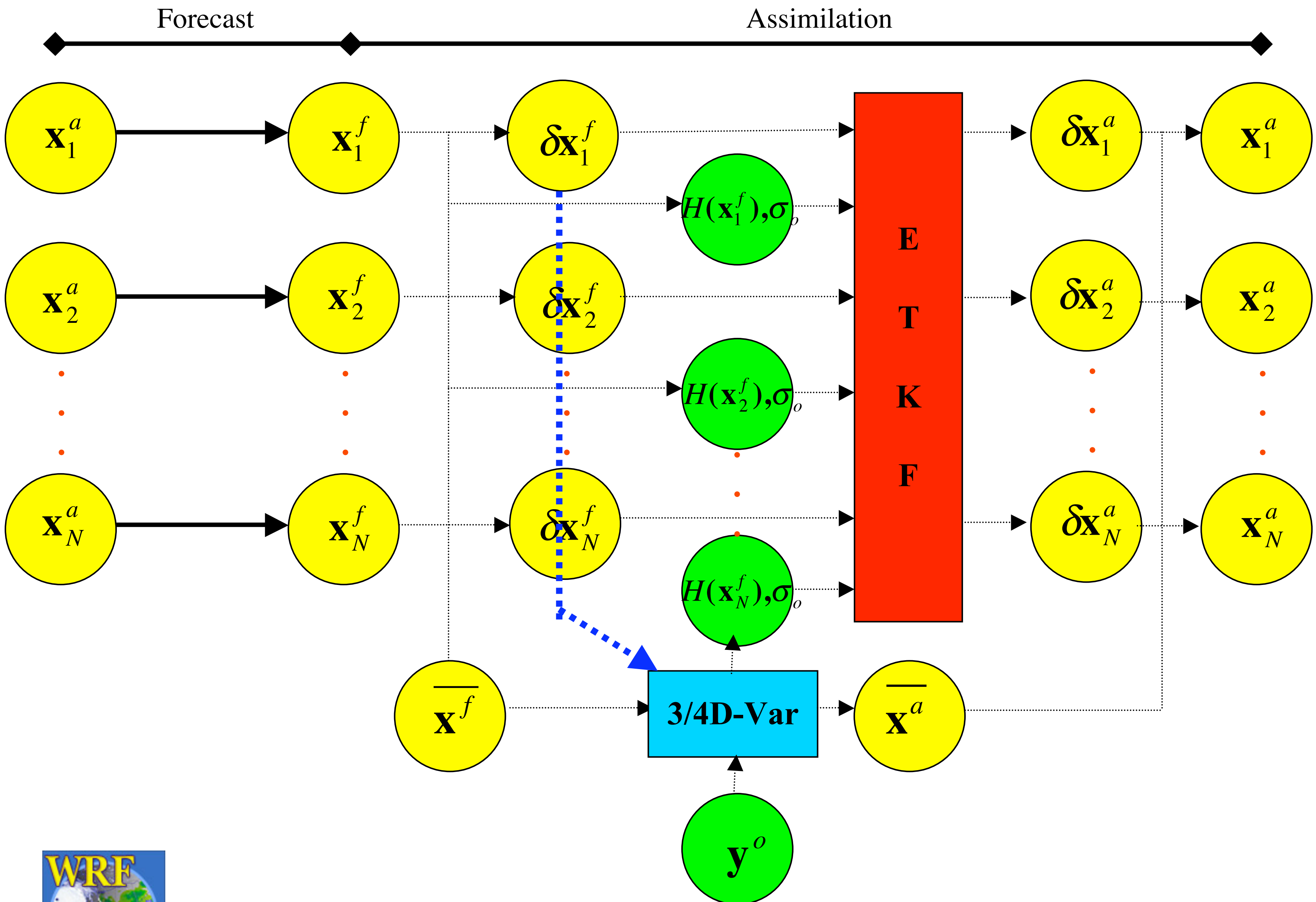
$$J(\mathbf{x}_1', \boldsymbol{\alpha}) = \beta_1 J_1 + \beta_2 J_e + J_o$$
$$= \beta_1 \frac{1}{2} \mathbf{x}_1'^{T} \mathbf{B}^{-1} \mathbf{x}_1' + \underline{\beta_2 \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{C}^{-1} \boldsymbol{\alpha}} + \frac{1}{2} \left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)^T \mathbf{R}^{-1} \left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)$$

$$\mathbf{x}' = \mathbf{x}_1' + \underline{\sum_{k=1}^{K} \left(\boldsymbol{\alpha}_k \circ \mathbf{x}_k^e\right)}$$

Extra increment associated with ensemble

$\mathbf{B}$ 3DVAR static covariance; $\mathbf{R}$ observation error covariance; K ensemble size;
$\mathbf{C}$ correlation matrix for ensemble covariance localization; $\mathbf{x}_k^e$ kth ensemble perturbation;
$\mathbf{x}_1'$ 3DVAR increment; $\mathbf{x}'$ total (hybrid) increment; $\mathbf{y}^{o'}$ innovation vector;
$\mathbf{H}$ linearized observation operator; β_1 weighting coefficient for static covariance;
β_2 weighting coefficient for ensemble covariance; $\boldsymbol{\alpha}$ extended control variable.

Cycling WRF/WRFDA/ETKF System (Hybrid DA)

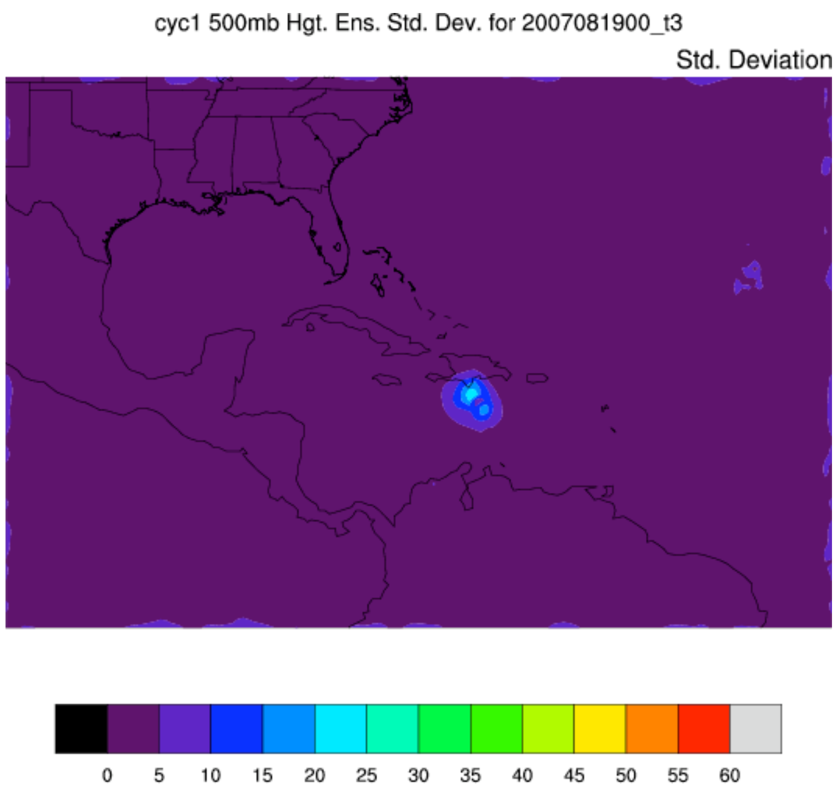


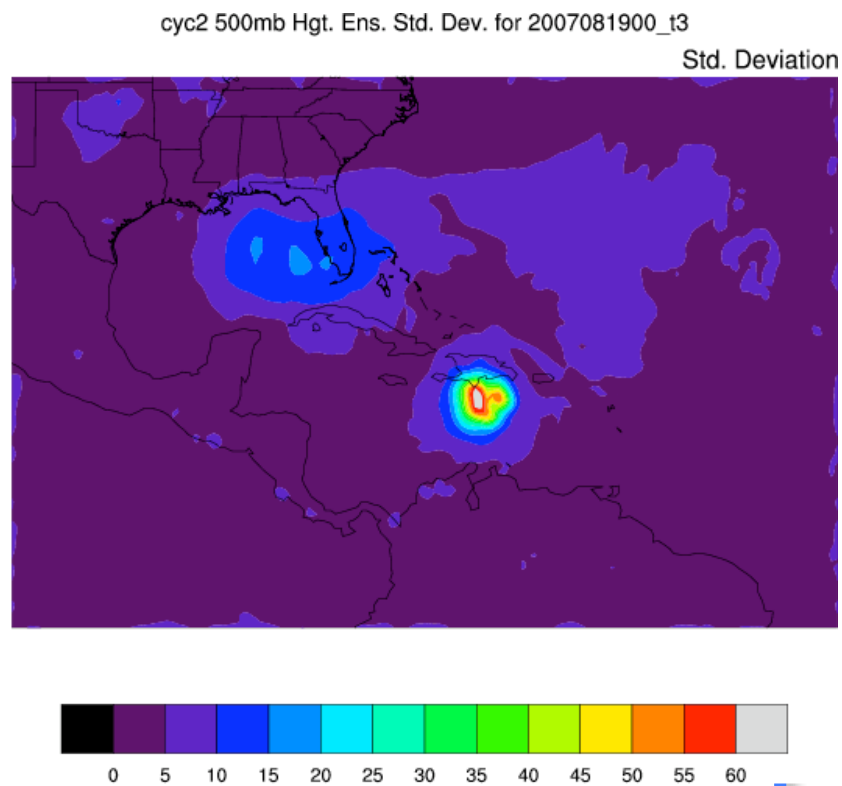
Preliminary results

- 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

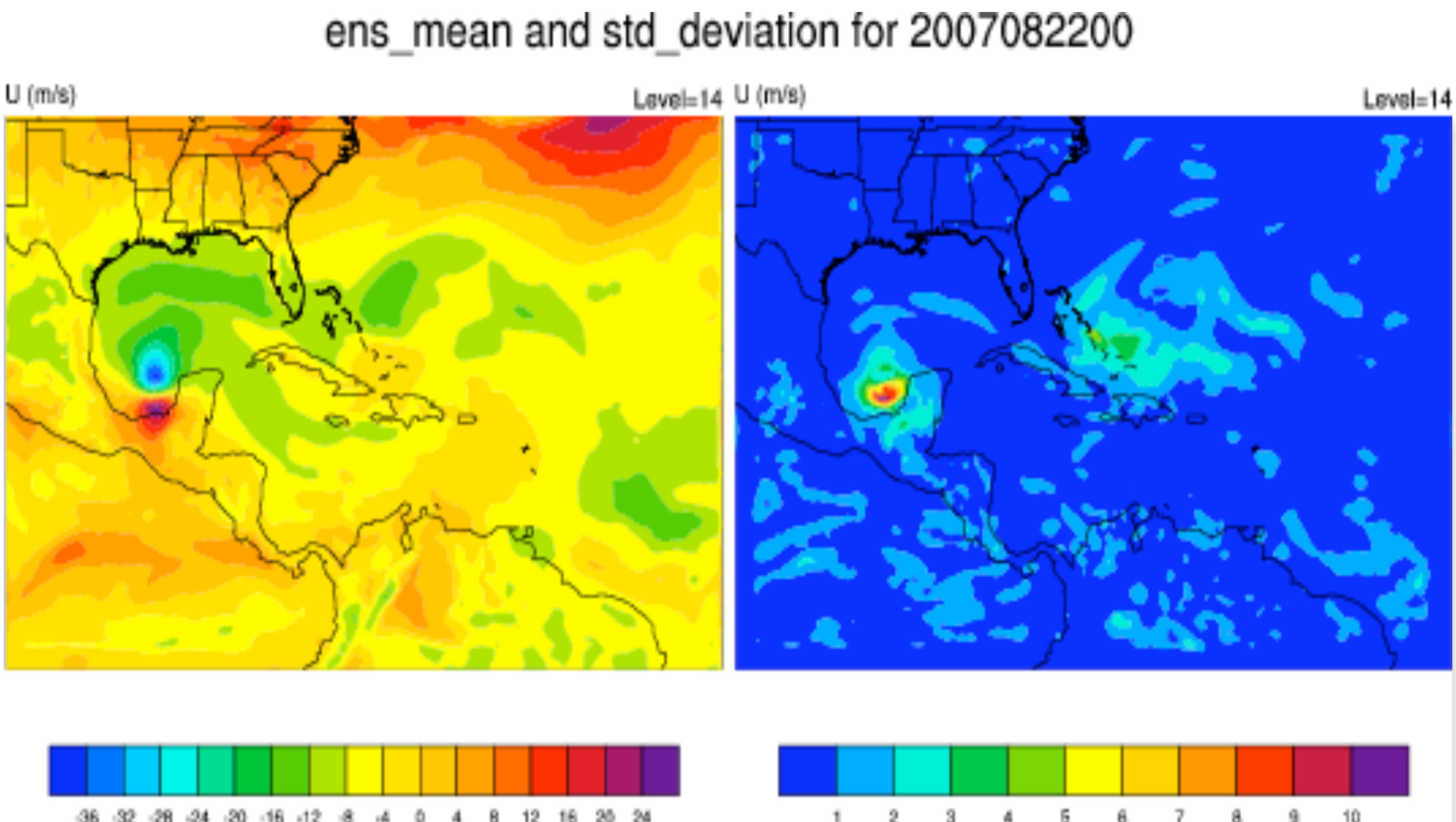




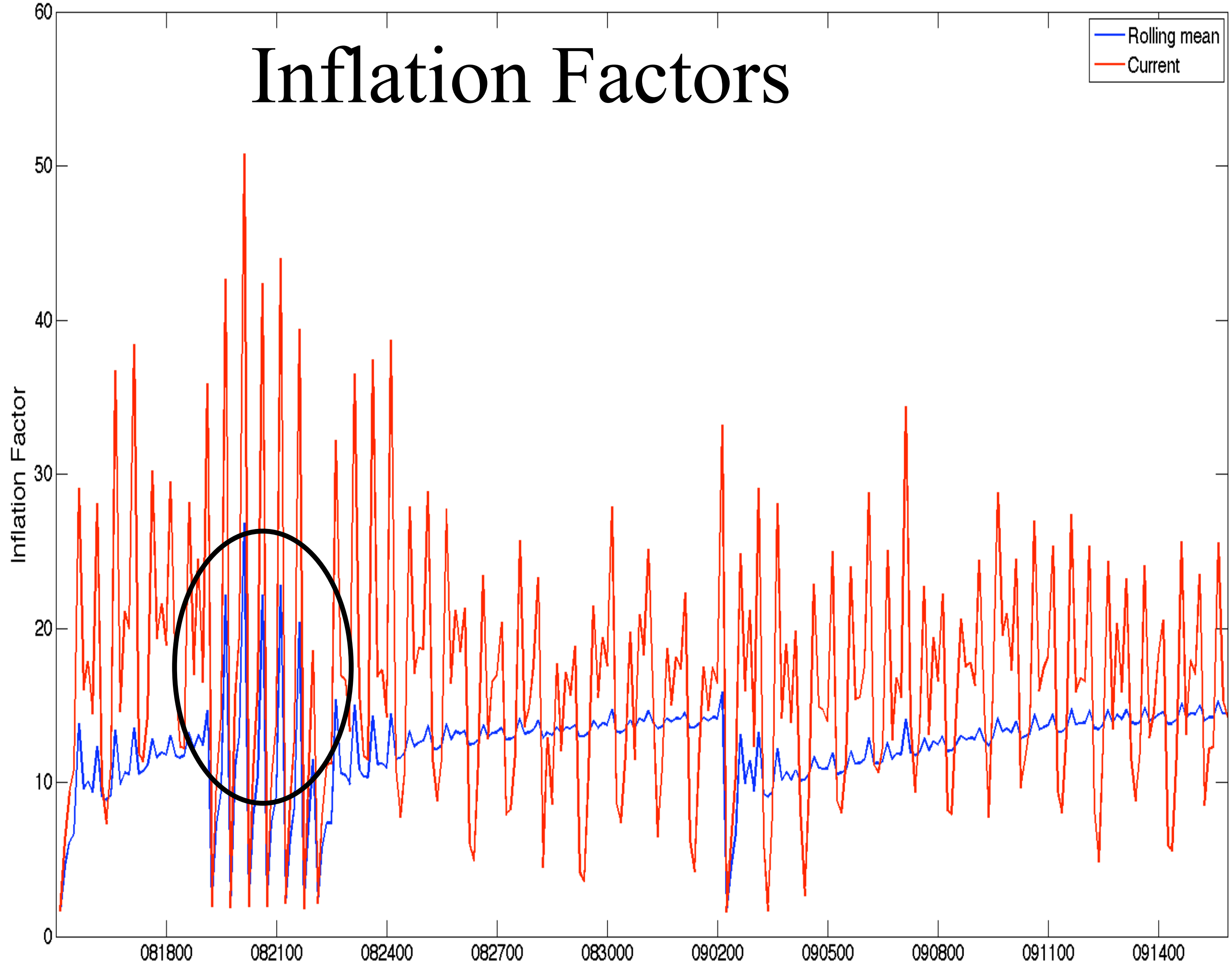
Modest inflations factors used

Higher inflations factors used

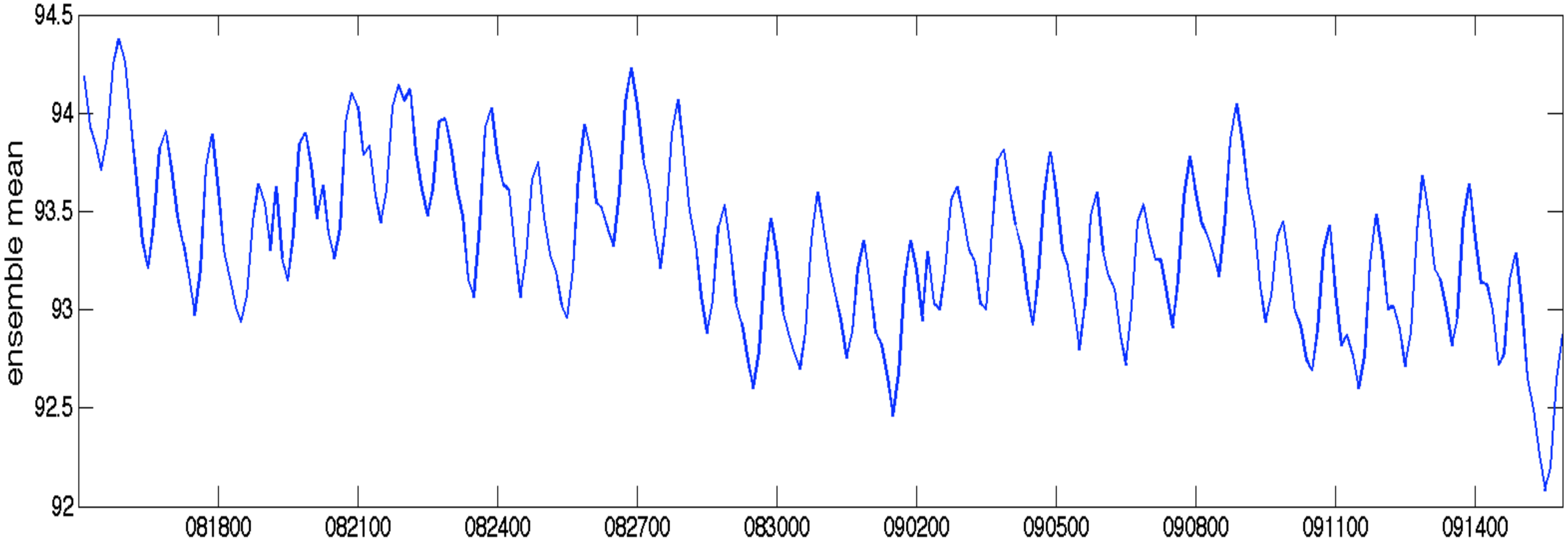
Ensemble Mean and Std. Deviation (spread)



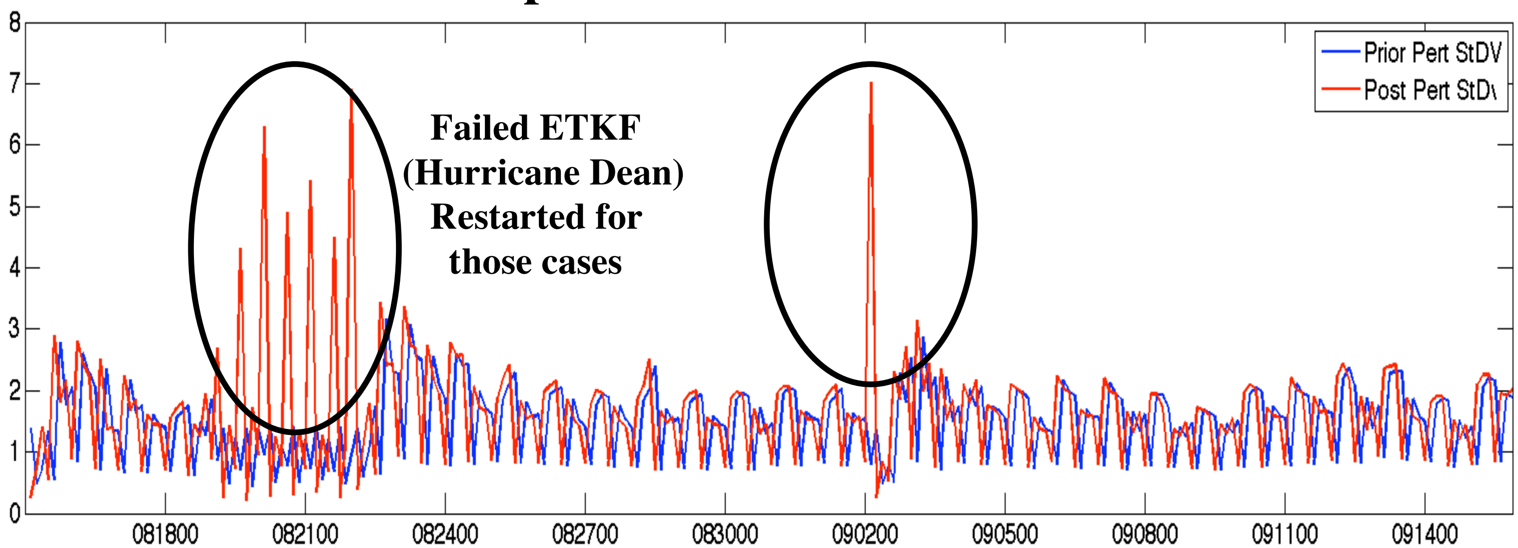
Inflation Factors

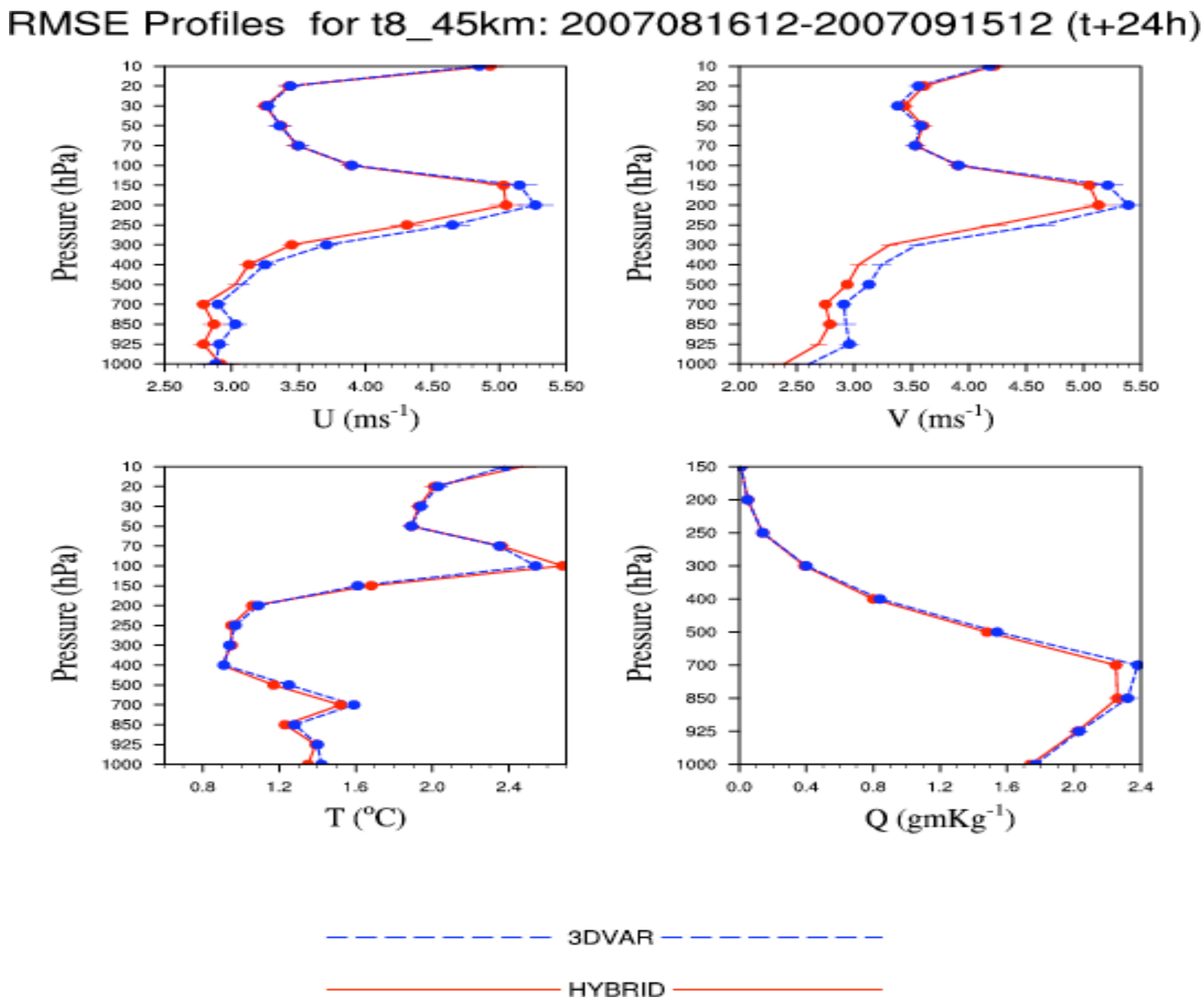


Prior Ensemble Mean

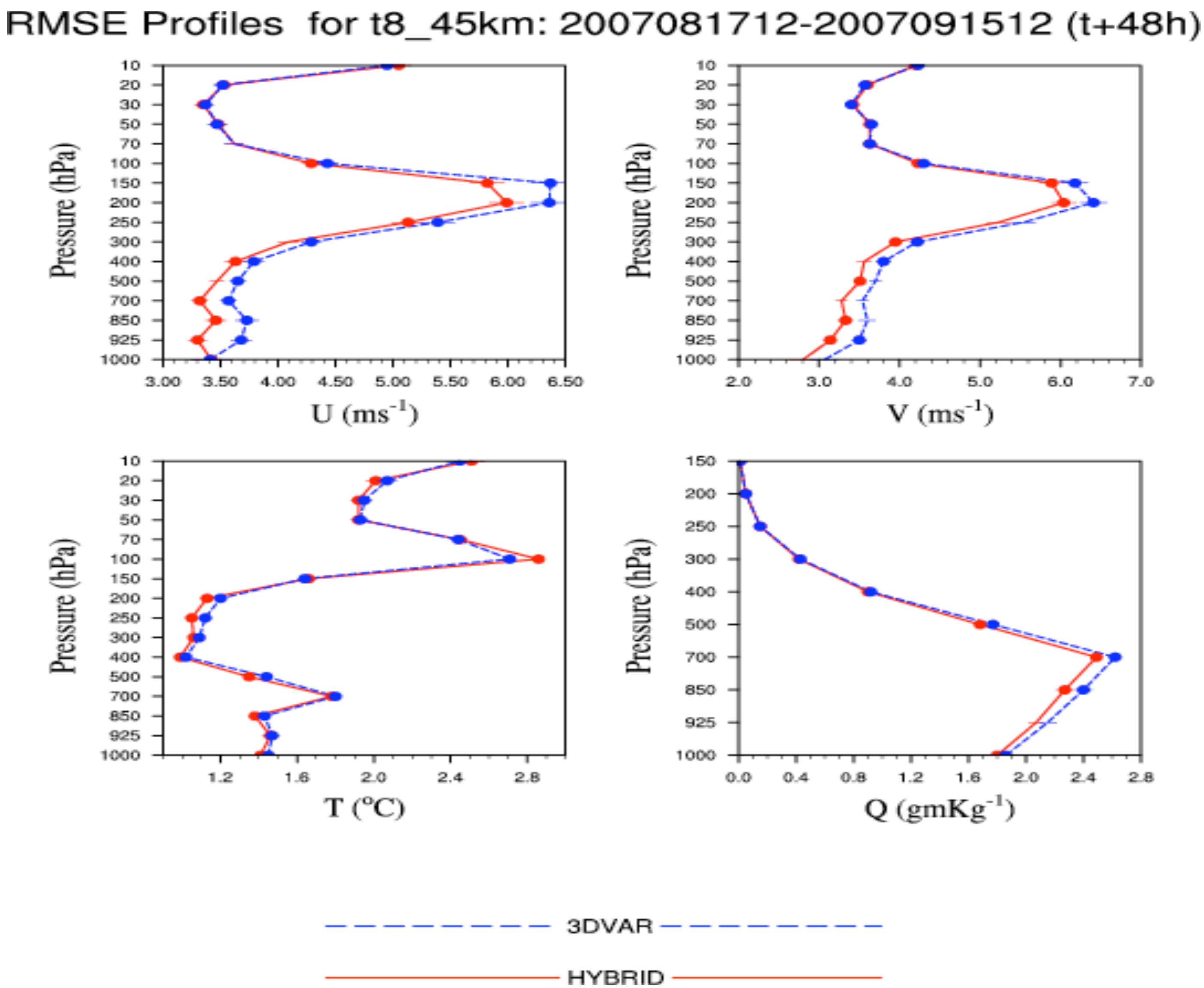


Ensemble Spread





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



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

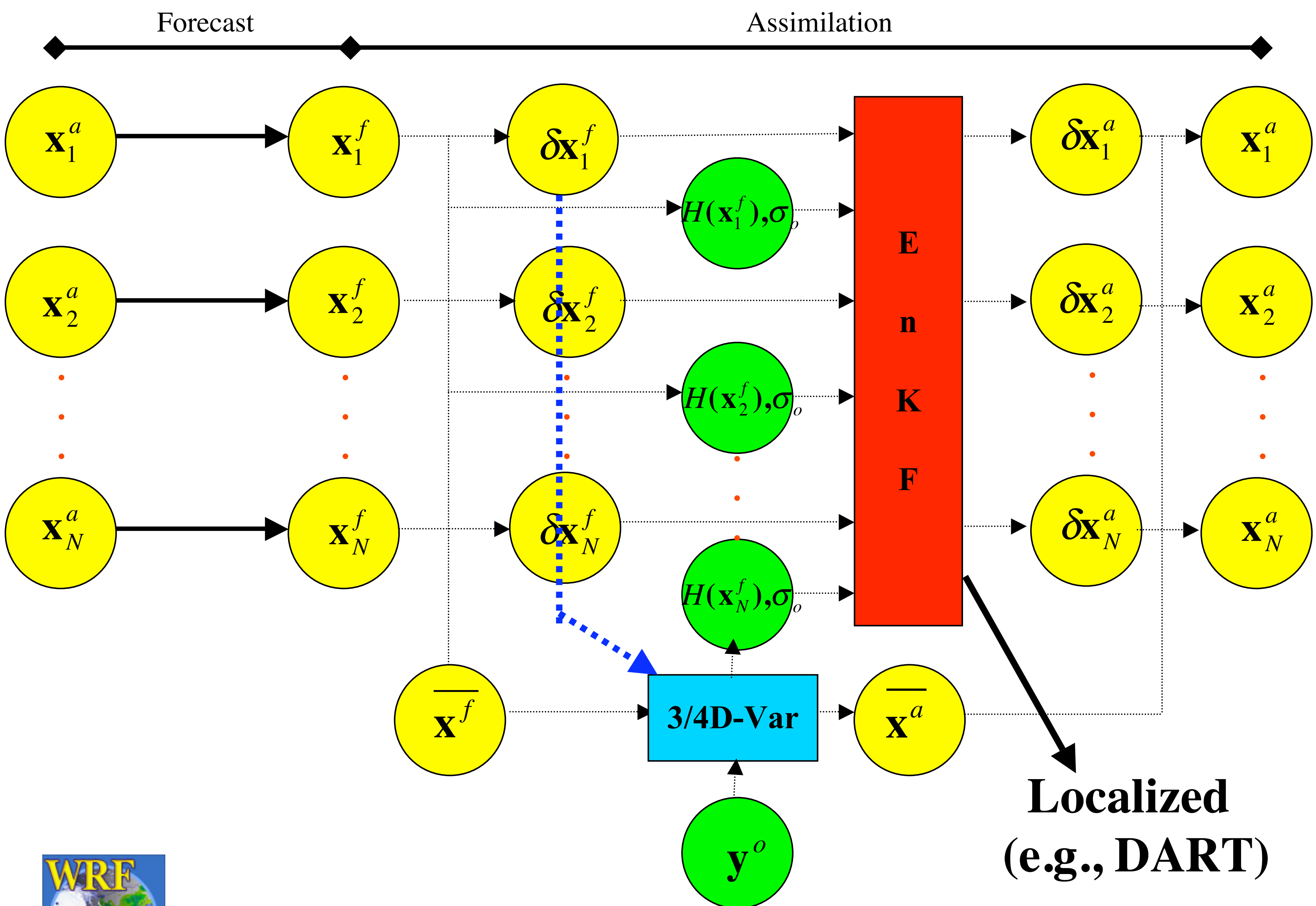
Pros and Cons of ETKF

- 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. It can even lead failure for Hurricane applications!

- Exist some localized versions of ETKF, expensive than non-localized ETKF.

Alternative: replace ETKF by localized EnKF



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.
- **ETKF part not included yet 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

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.

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.