A hybrid ETKF-3DVAR data assimilation scheme for WRF

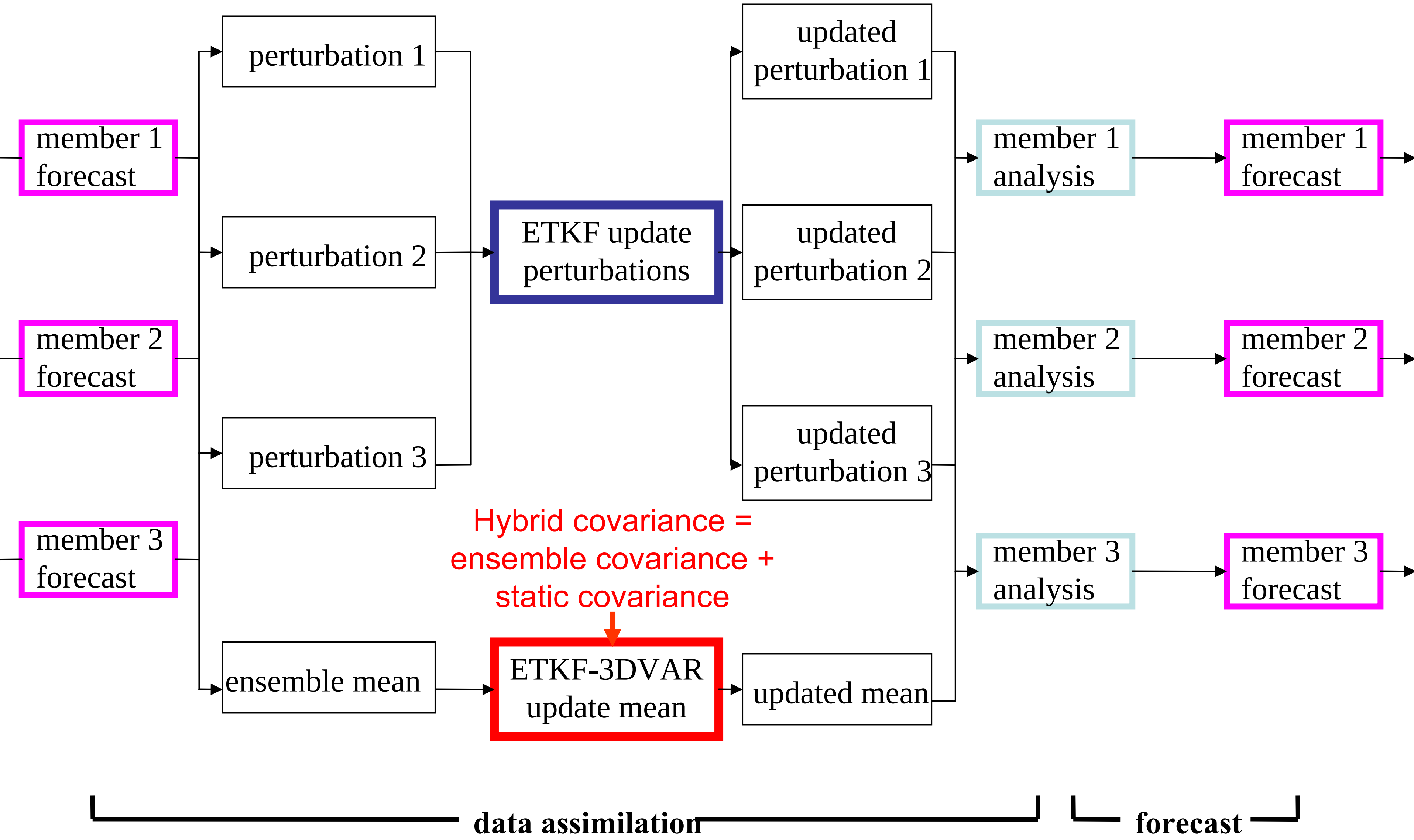
Xuguang Wang

NOAA Earth System Research Lab,
Physical Science Division, Boulder, CO

- Motivation
- Theory to understand namelist variables
- CONUS experiment results
- what next?
- flow chart to understand script
- namelist variables

What's Hybrid ETKF-3DVAR ?

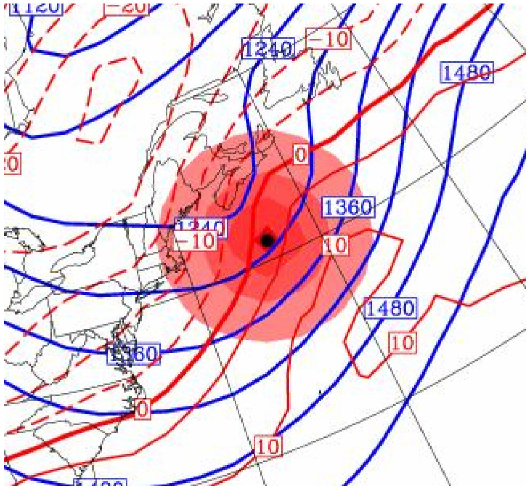
(Wang et al. 2007a, MWR)



Why Hybrid ETKF-3DVAR ?

Compared to 3DVAR:

3DVAR problem:
static isotropic covariance



- Hybrid can benefit from ensemble-estimated flow-dependent error statistics (examples later).

Compared to conventional ENSDA:

- Hybrid may be more robust for small ensemble size and/or large model error (Wang et al. 2007a,b, MWR).

- Hybrid can be conveniently adapted to the existing operational variational framework; potentially less expensive.

Hybrid DA Theory

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

$$J(\mathbf{x}_1^{'}, \boldsymbol{\alpha}) = \beta_1 J_1 + \beta_2 J_e + J_o$$

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

Extra term associated with extended control variable

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

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.

Hybrid DA Theory

- Wang et al. 2007c show solution equivalent to

$$J(\mathbf{x}') = \frac{1}{2}\mathbf{x}'^T\left(\frac{1}{\beta_1}\mathbf{B} + \frac{1}{\beta_2}\mathbf{P}^e \circ \mathbf{S}\right)^{-1}\mathbf{x}' + \frac{1}{2}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}'\right)$$

Weighted average of static and ensemble covariance

- To preserve total variance

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

Hybrid DA Theory

- In current system, ensemble covariance localization applied through recursive filter.

$$J = \beta_1 \frac{1}{2} \mathbf{x}_1^{'T} \mathbf{B}^{-1} \mathbf{x}_1^{'} +$$

$$\beta_2 \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{C}^{-1} \boldsymbol{\alpha} +$$

$$\frac{1}{2} \left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}' \right)^T \mathbf{R}^{-1} \left(\mathbf{y}^{o'} - \mathbf{H}\mathbf{x}' \right)$$

Preconditioned by $\mathbf{x}_1^{'} = \mathbf{U}_1 \mathbf{v}_1$ and $\mathbf{U}_1 \approx \mathbf{B}^{1/2}$

Extended control variables constrained by correlation matrix C, which defines ensemble covariance localization. Only horizontal localization considered.

Preconditioned by $\boldsymbol{\alpha} = \mathbf{U}_2 \mathbf{v}_2$ and $\mathbf{U}_2$, $\mathbf{U}_2 \approx \mathbf{C}^{1/2}$, is modeled by recursive filter.

Ensemble generation by ETKF

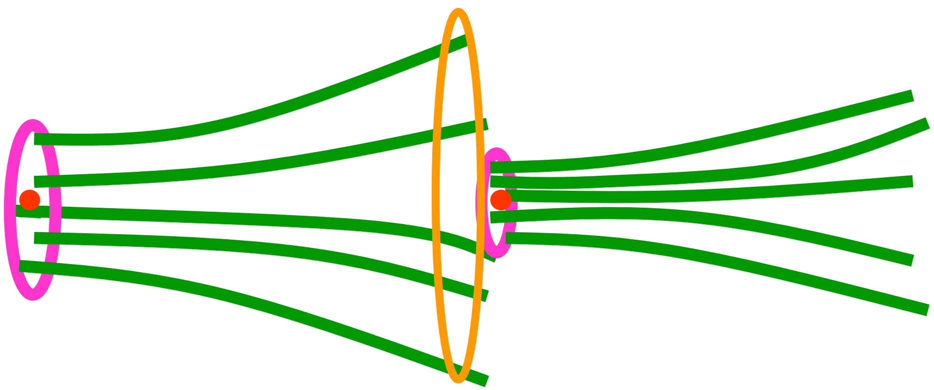
- ETKF generates ensembles by rescaling forecast perturbations with a transformation matrix (e.g., Wang and Bishop 2003, Wang et al. 2004, 2007a)

Transformation matrix solved from Kalman filter theory

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{T}$$

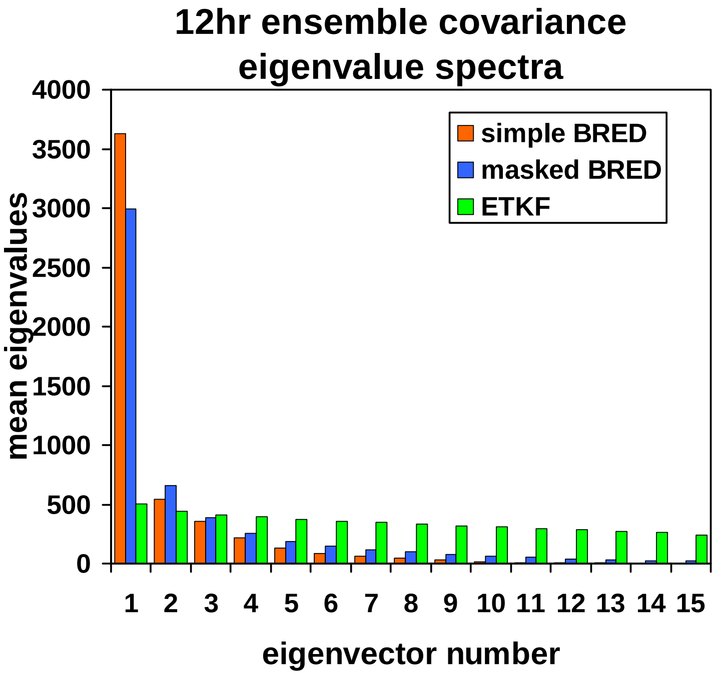
Initial ensemble perturbation

forecast ensemble perturbation



Ensemble generation by ETKF

- Previous studies indicated advantages of ETKF over breeding method (Wang and Bishop 2003).

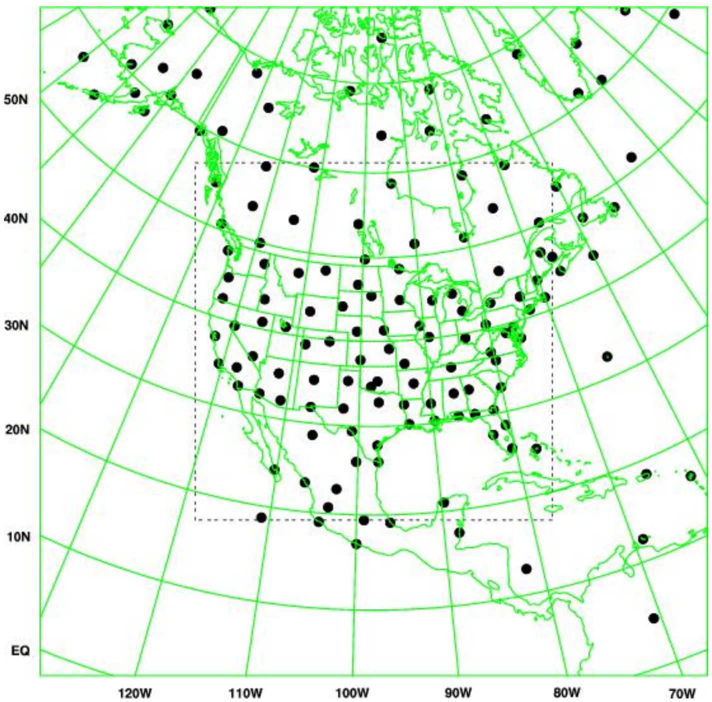


- Relatively inexpensive for ensemble size of o(100).

- Systematic underestimate of the analysis error variance due to sampling error. Ameliorated by two parameters.

Hybrid ETKF-3DVAR experiment

WRF domain, observation locations and verification region

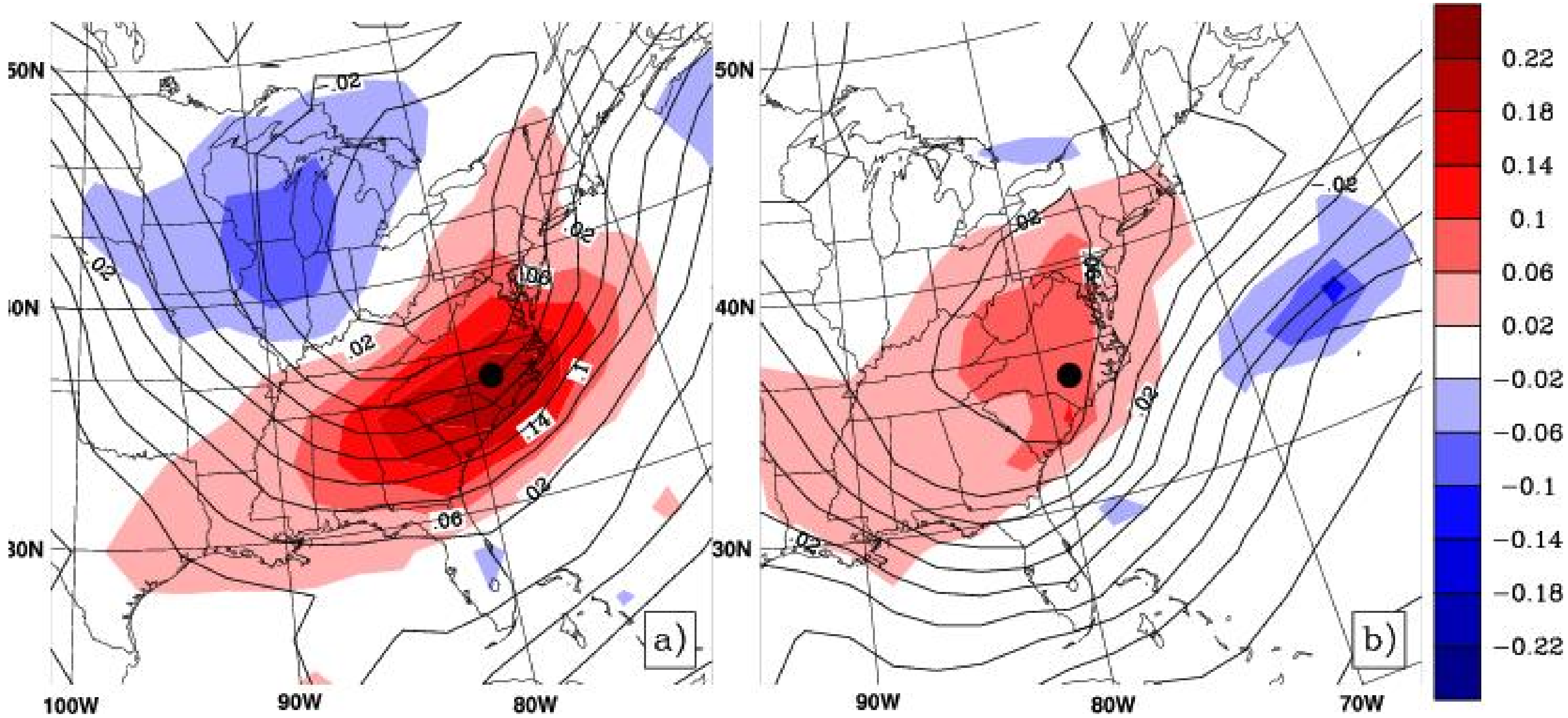


- WRF domain: North America; coarse resolution (Δx=200km; 28 levels)
- Observation: radiosonde wind and temperature
- Test period: Jan 2003
- Ensemble size: 50 members
- Verifications: Compare forecasts initialized by the hybrid and 3DVAR analyses.

Wang et al. 2008ab, MWR

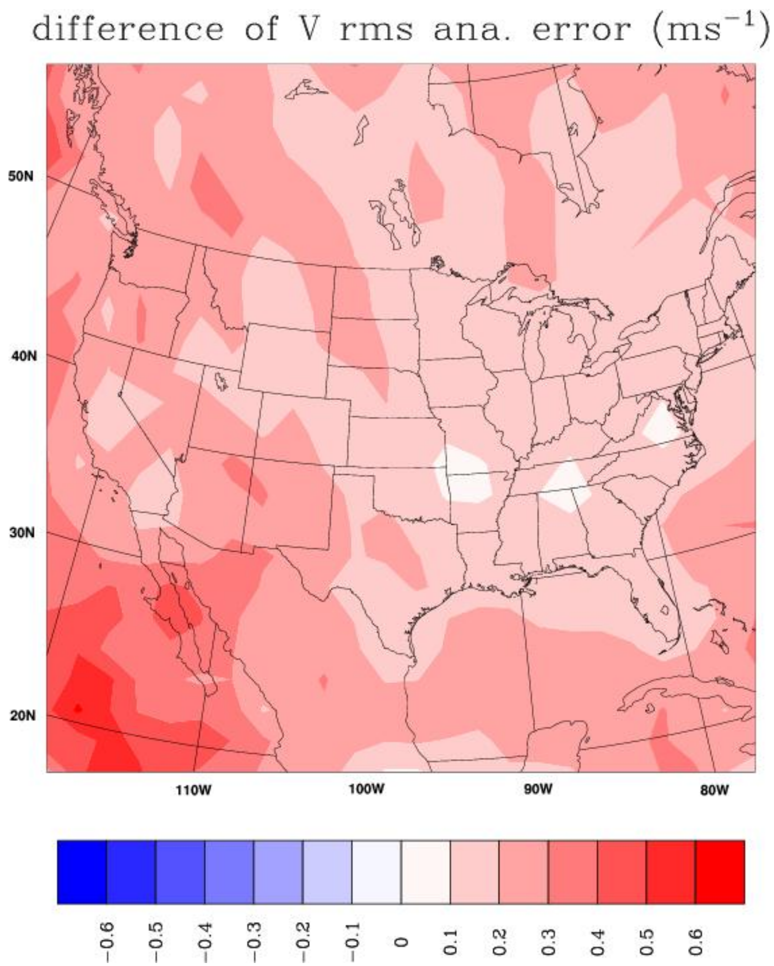
flow-dependent increments by the hybrid

850mb T increment (k)



- The hybrid system can provide flow-dependent increments.

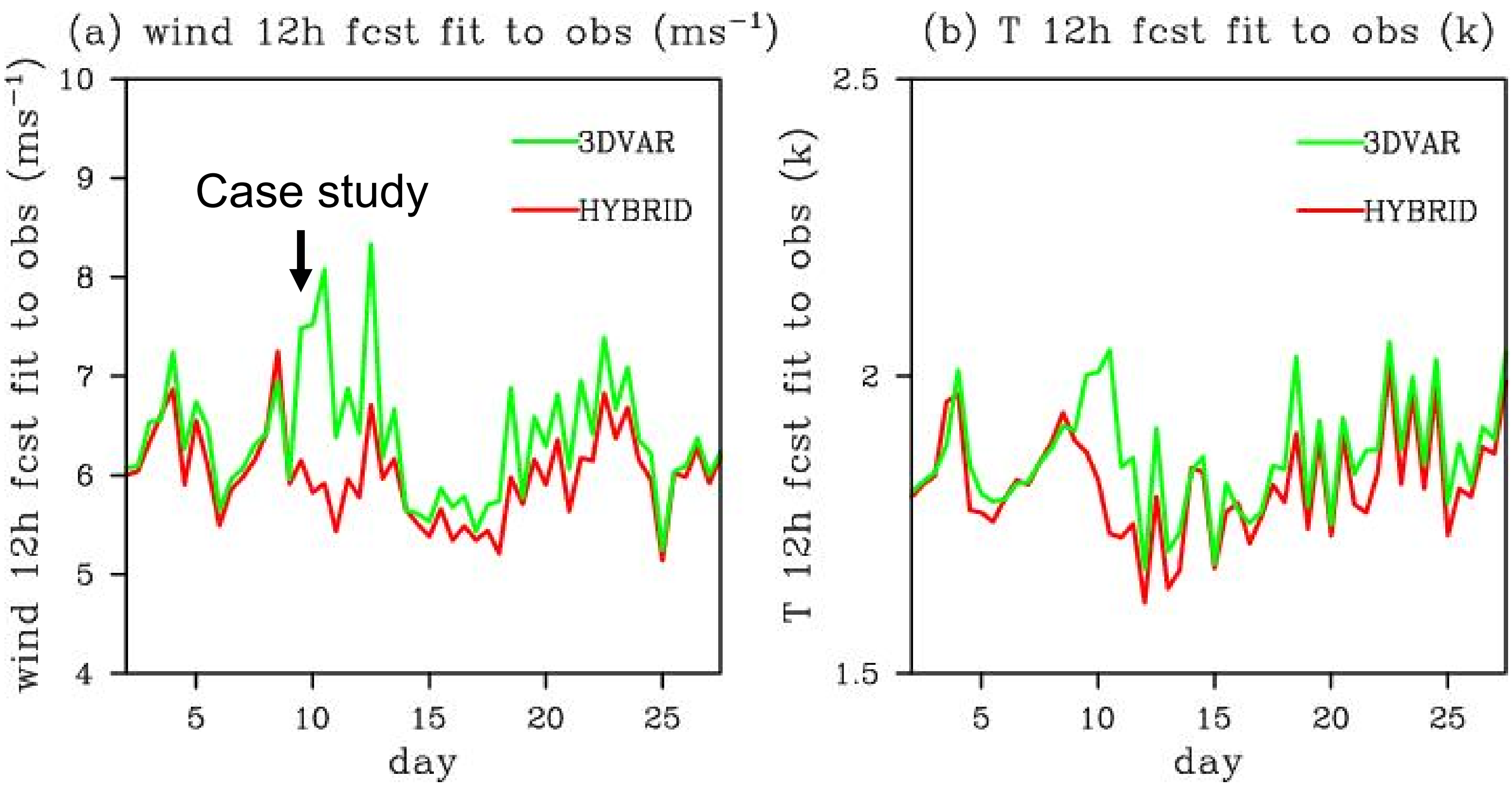
Improvement relative to obs. density



The simulated observation experiment shows that

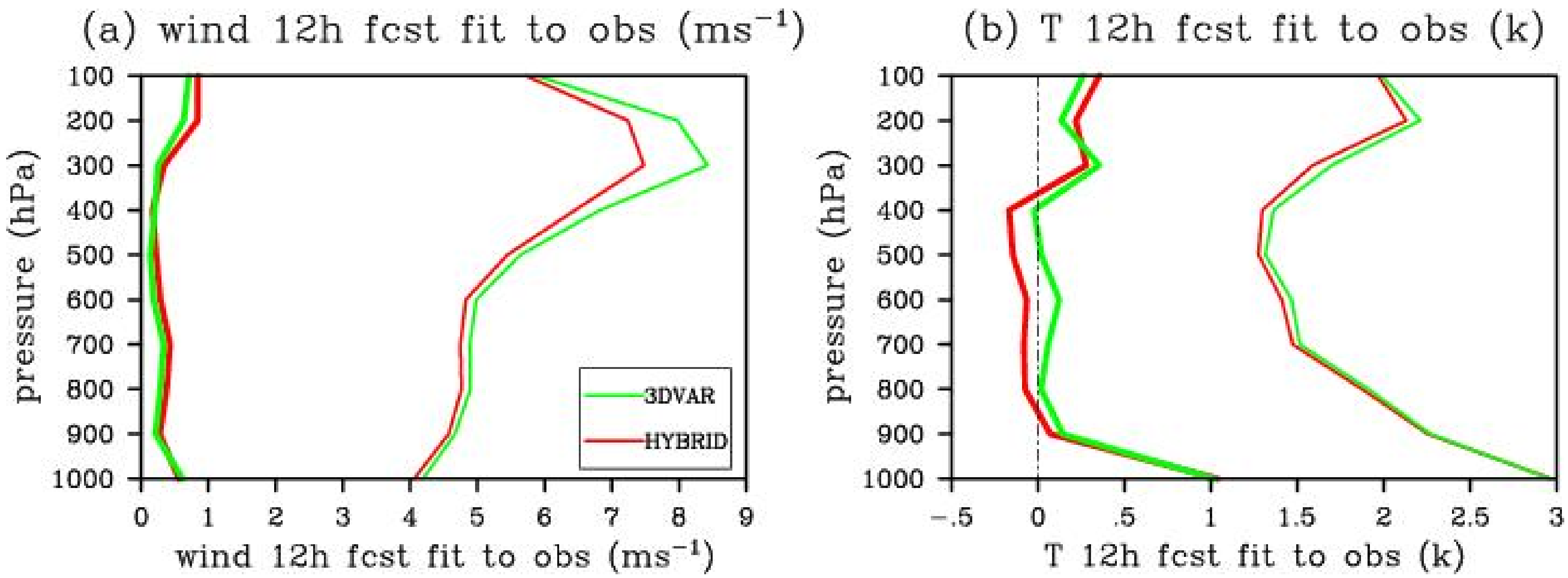
- Hybrid has larger improvement over and downstream of data sparse regions.
- Flow-dependent ensemble covariance has the largest impact over and downstream of where observation is sparse.

Real obs. experiment: 12h forecast error



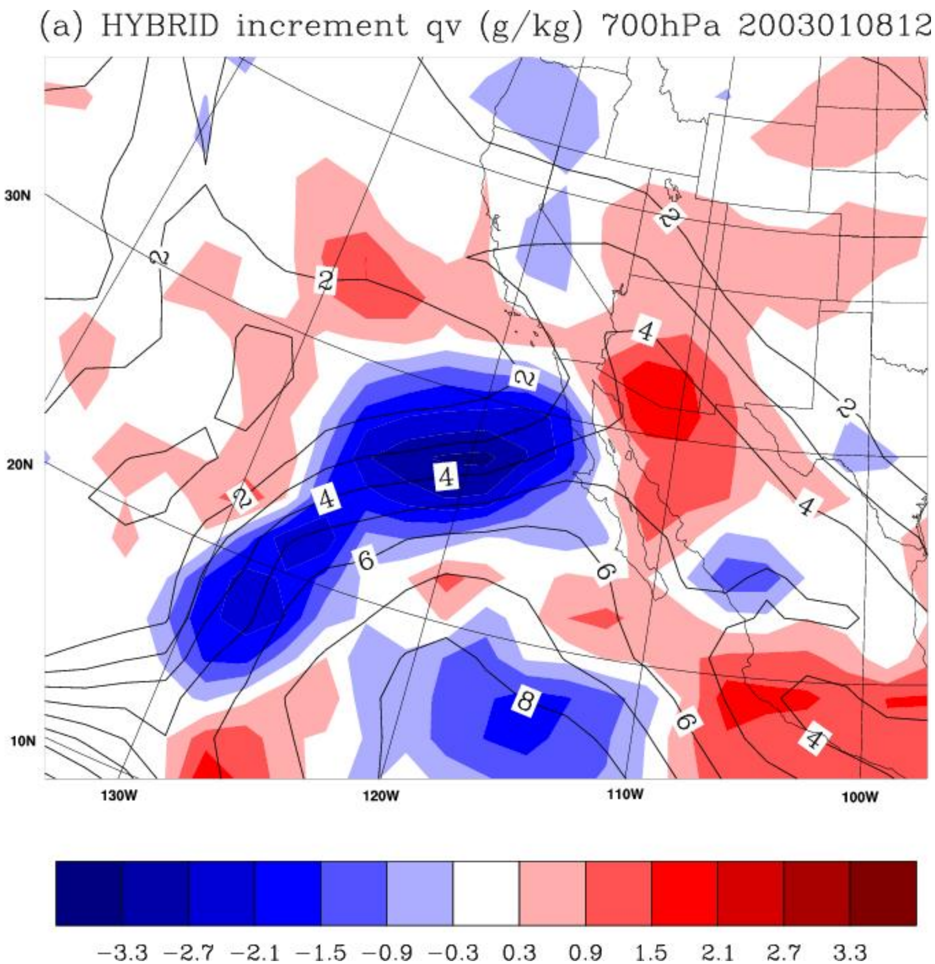
- Hybrid 12h forecast is more accurate than the 3DVAR for most time.

Vertical profile of 12h forecast error



- **Wind:** Hybrid has the largest improvement at 200mb-300mb;
- **Temperature:** Improvement smaller than wind. No improvement at lower troposphere (significant bias).

How was the moisture field updated?

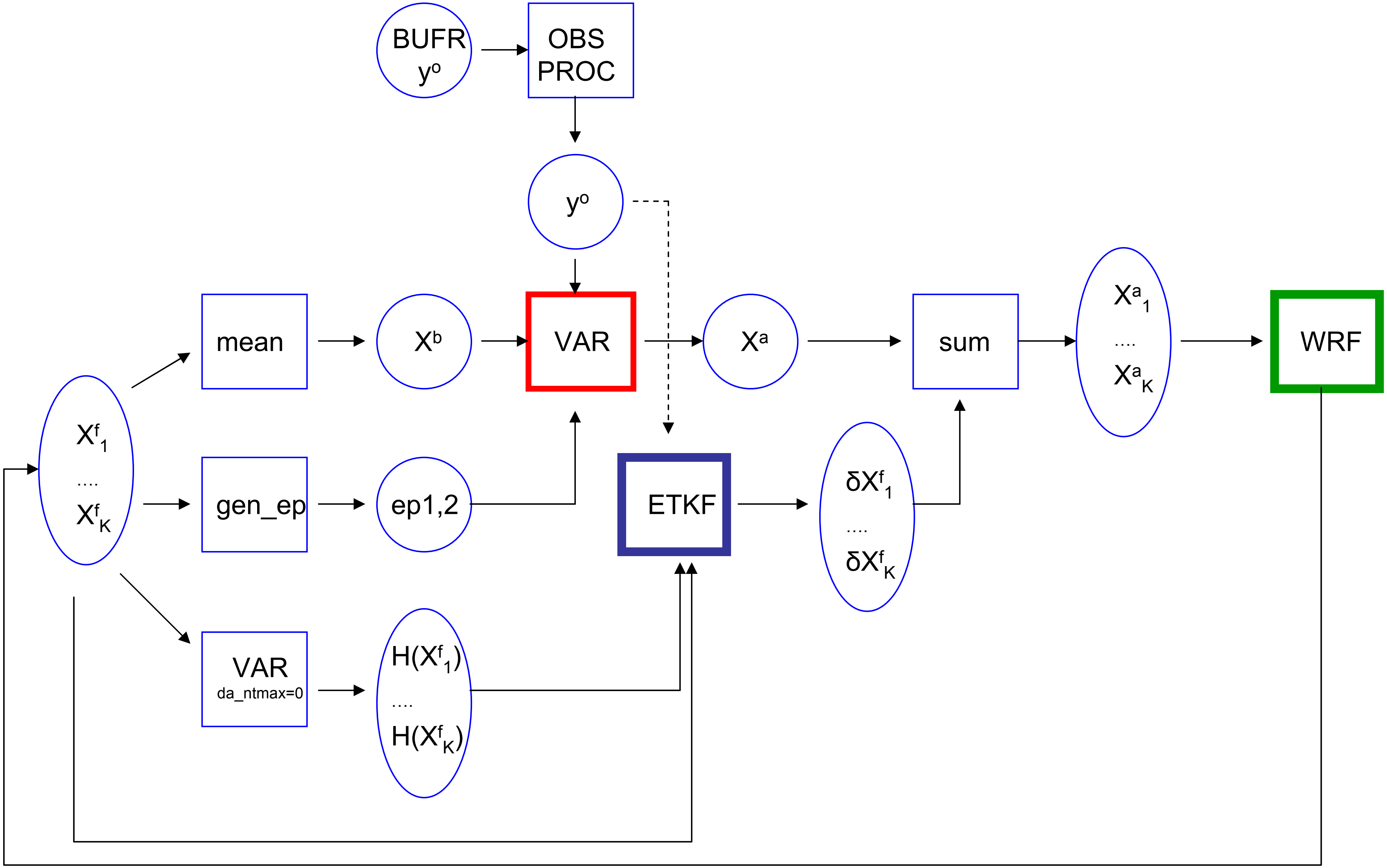


- flow-dependent adjustments produced by the hybrid extended a large distance into eastern Pacific data void region. It dried the lower troposphere along the front.

- Although no moisture observations were assimilated, hybrid through cross-variable covariance estimated by the ensemble can update moisture field whereas 3DVAR can not.

What next?

- 3D extended control variables with vertical covariance localization
- Include microphysical state variables for e.g. radar DA
- Adaptive spatially varying weighting factors?
- Better methods to ameliorate ETKF sampling errors: spatially varying inflation, Local ETKF
- Better LBC ensembles, better representation of model errors in the ensemble
- ETKF-4DVAR; inter-comparison of different DA schemes
- Meso/convective scale application


BUFR
y^o
OBS
PROC
y^o
mean
X^b
VAR
X^a
sum
X^a_1
....
X^a_K
WRF
X^f_1
....
X^f_K
gen_ep
ep1,2
ETKF
δX^f_1
....
δX^f_K
VAR
da_ntmax=0
$H(X^f_1)$
....
$H(X^f_K)$

Hybrid DA namelist variables

- **alpha_corr_scale**

Correlation length scale (km) of recursive filter

- **jb_factor** ~ $\beta 1$
- **je_factor** ~ $\beta 2$

$1/\beta 1 + 1/\beta 2 = 1$

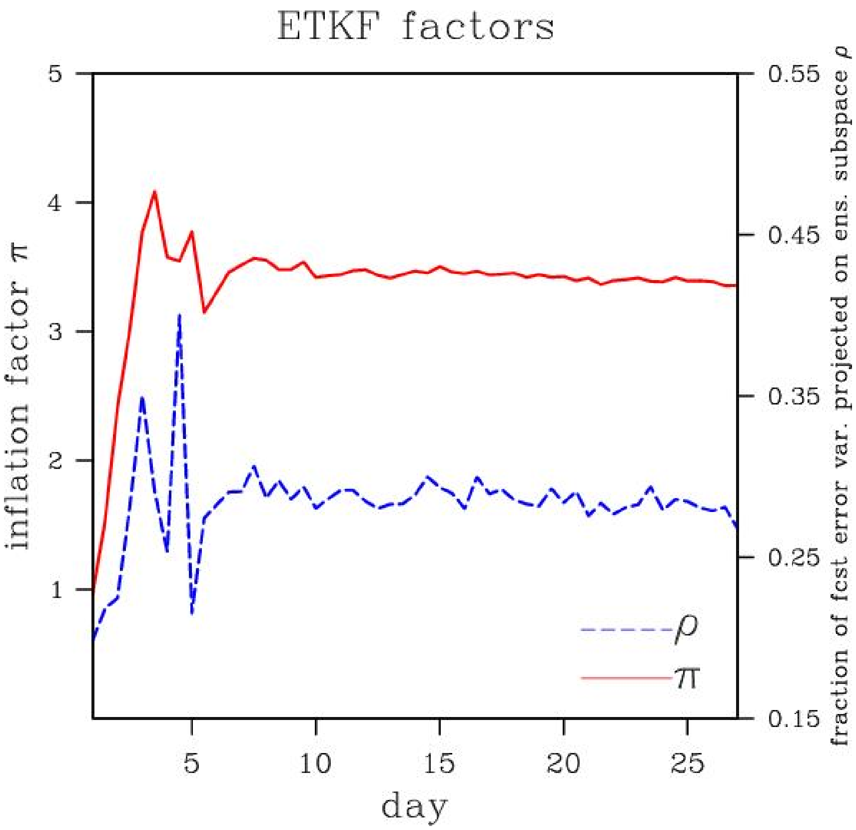
- **alphacv_method**

1 control-variable space ensemble perturbations

2 model space ensemble perturbations

ETKF namelist variables

- **tainflatinput**
  prescribed inflation factor
- **rhoinput**
  prescribed fraction of forecast error variance projected onto ensemble subspace
- **naccumt1/naccumt2**
  For online estimation of the parameters. Denote period over which running means of the parameters are calculated.



References

- 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., C. H. Bishop, and Simon J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590-1605.
- Wang, X., T. M. Hamill, J. S. Whitaker and C. H. Bishop, 2007a: A comparison of hybrid ensemble transform Kalman filter-OI and ensemble square-root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055-1076.
- Wang, X., C. Snyder, and T. M. Hamill, 2007c: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222-227.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, in press.
- Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008b: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.*, in press.