A hybrid ensemble transform Kalman filter (ETKF)-3DVAR data assimilation scheme for WRF

Xuguang Wang

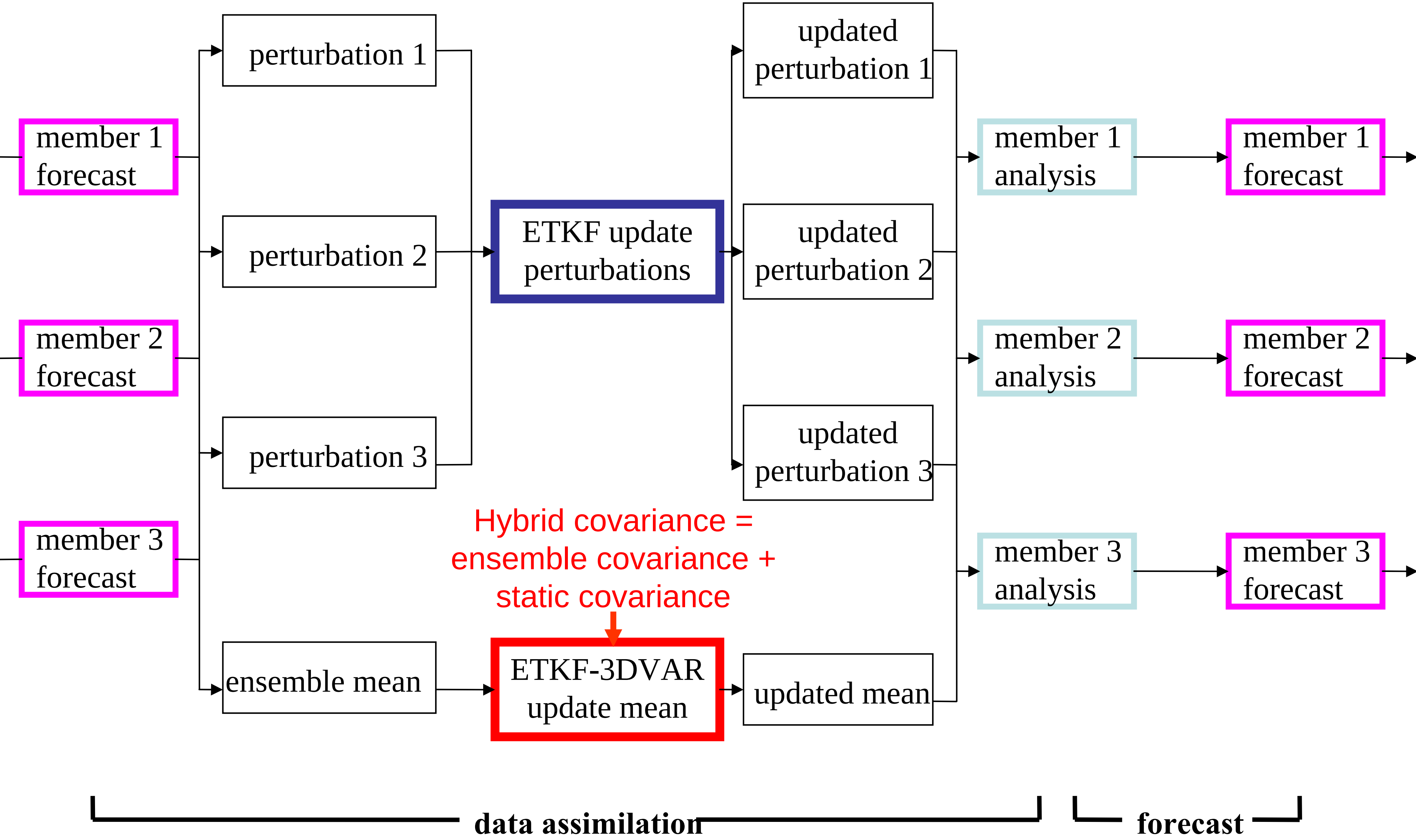
NOAA Earth System Research Lab,
Physical Science Division, Boulder, CO

Dale Barker, Chris Snyder (NCAR), Tom Hamill (NOAA)

8th WRF user's workshop, Boulder, CO, June, 2007

What's Hybrid ETKF-3DVAR ?

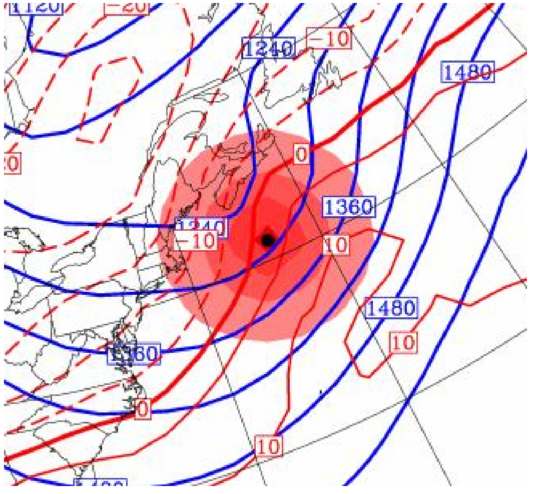
(Wang et al. 2007a, MWR)



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 Data Assimilation Theory

- Ensemble covariance is included in the 3DVAR cost function through augmentation of control variables (Lorenz 2003, QJ; Buehner 2005, QJ; Wang et al. 2007c, MWR)

Extra term associated with extended control variable

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

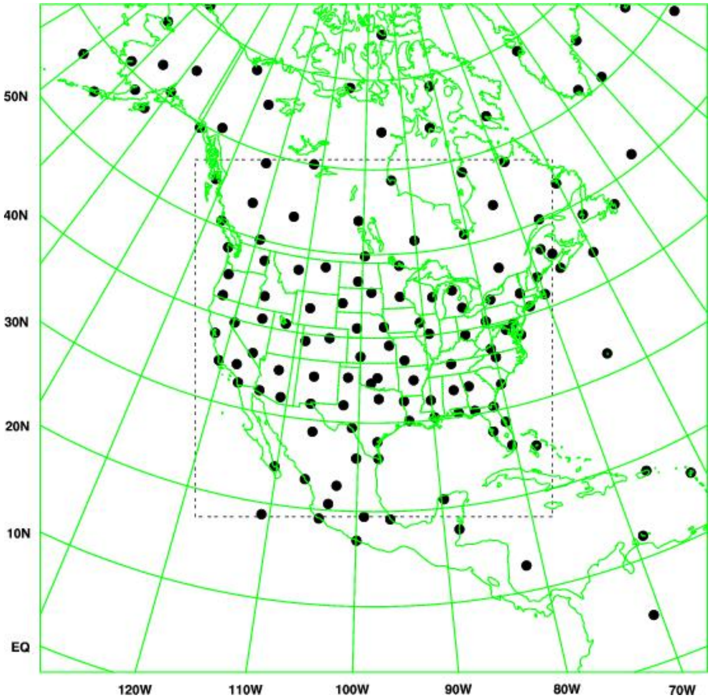
$$\mathbf{x}^{'} = \mathbf{x}_1^{'} + \underline{\sum_{k=1}^{K} \left(\boldsymbol{\alpha}_k \circ \mathbf{x}_k^e\right)}$$

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.

OSSE and real obs. experiment with WRF

WRF domain, observation locations and verification region



WRF domain: North America; 200km resolution; 45x45x28 grid

Observation: simulated (OSSE) and real (real obs. Exp) U,V,T radiosondes

Cycle: every 12h during Jan 2003.

ETKF ensemble: 50-member

LBC ensemble: AVN analysis + 50 perturbations drawn from default 3DVAR static covariance.

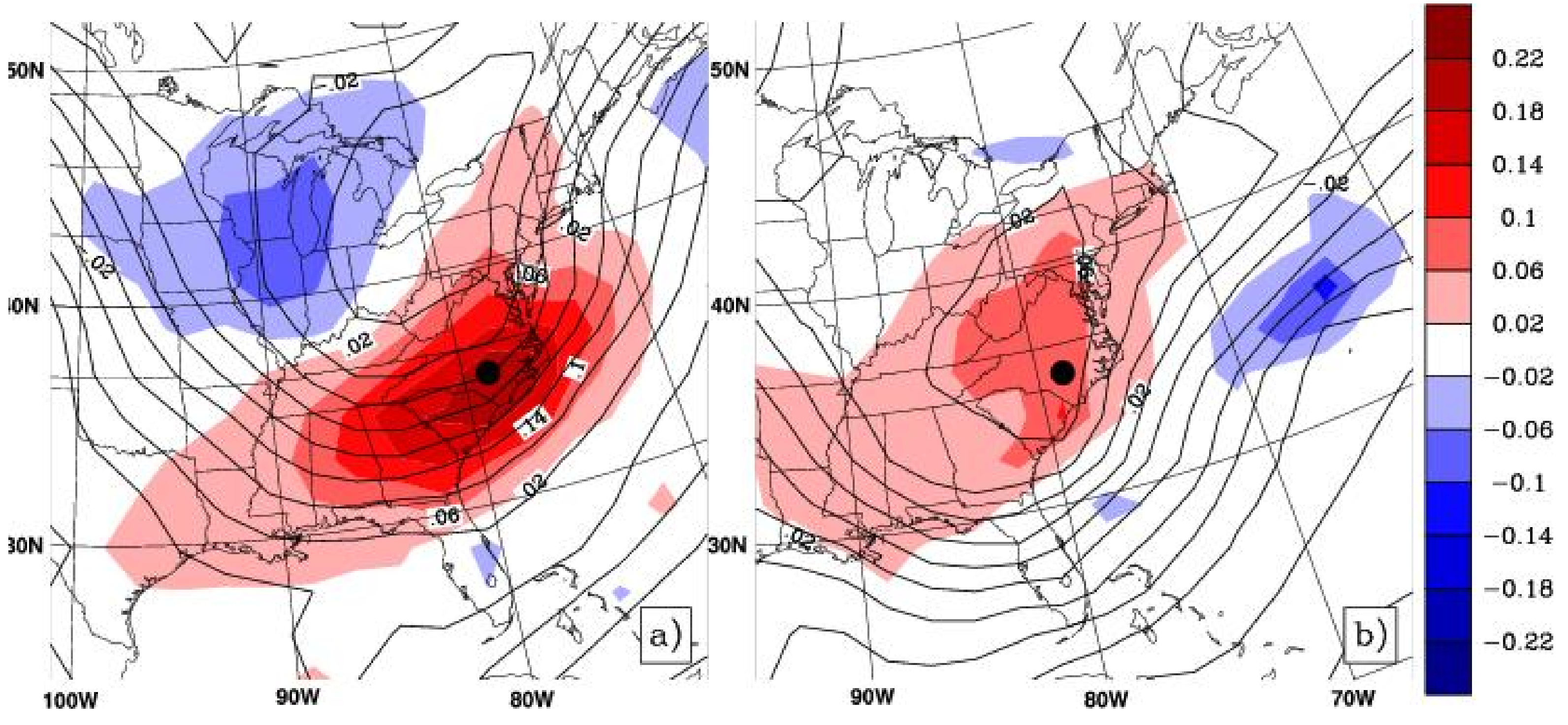
Verifications:

OSSE: against “truth” at all grids and variables within the verification domain

REAL: fitting 12h forecasts to obs. within the verification domain.

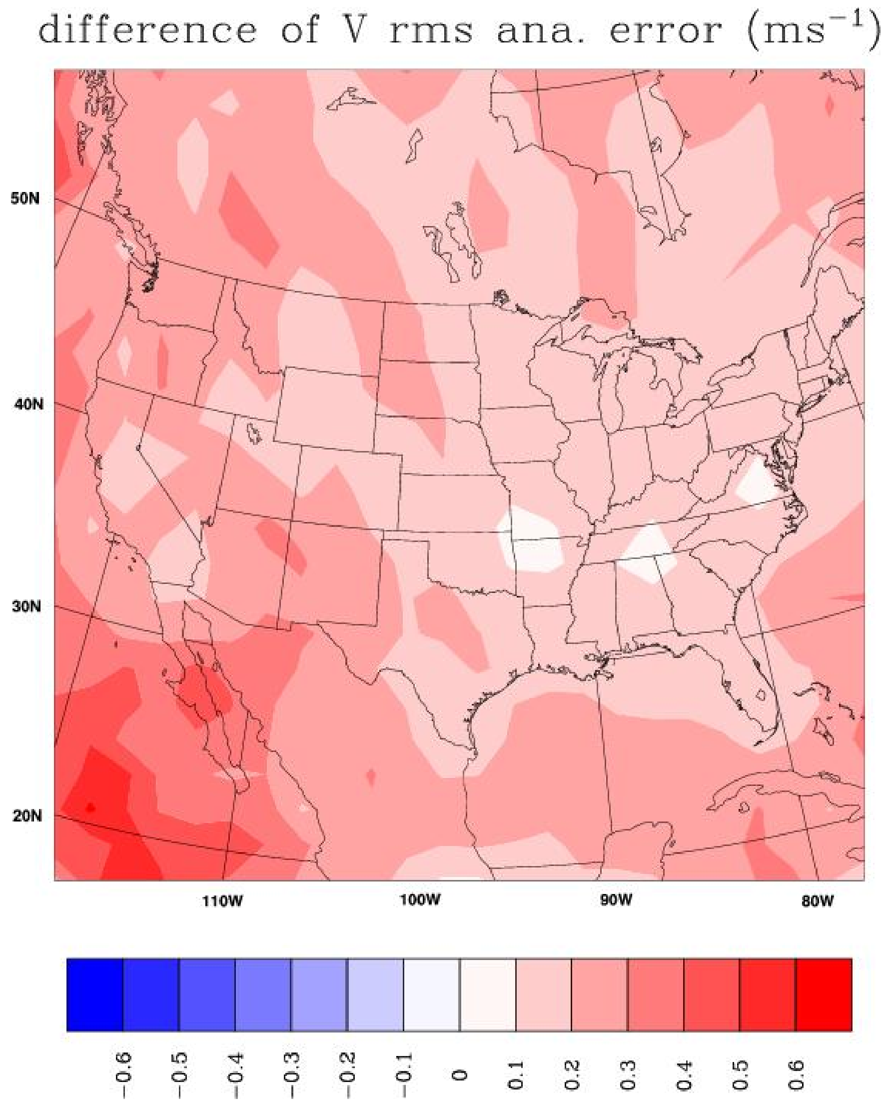
OSSE: single observation test

850mb T increment (k)



- The hybrid system (using extended control variable) can incorporate flow-dependent ETKF ensemble covariance.

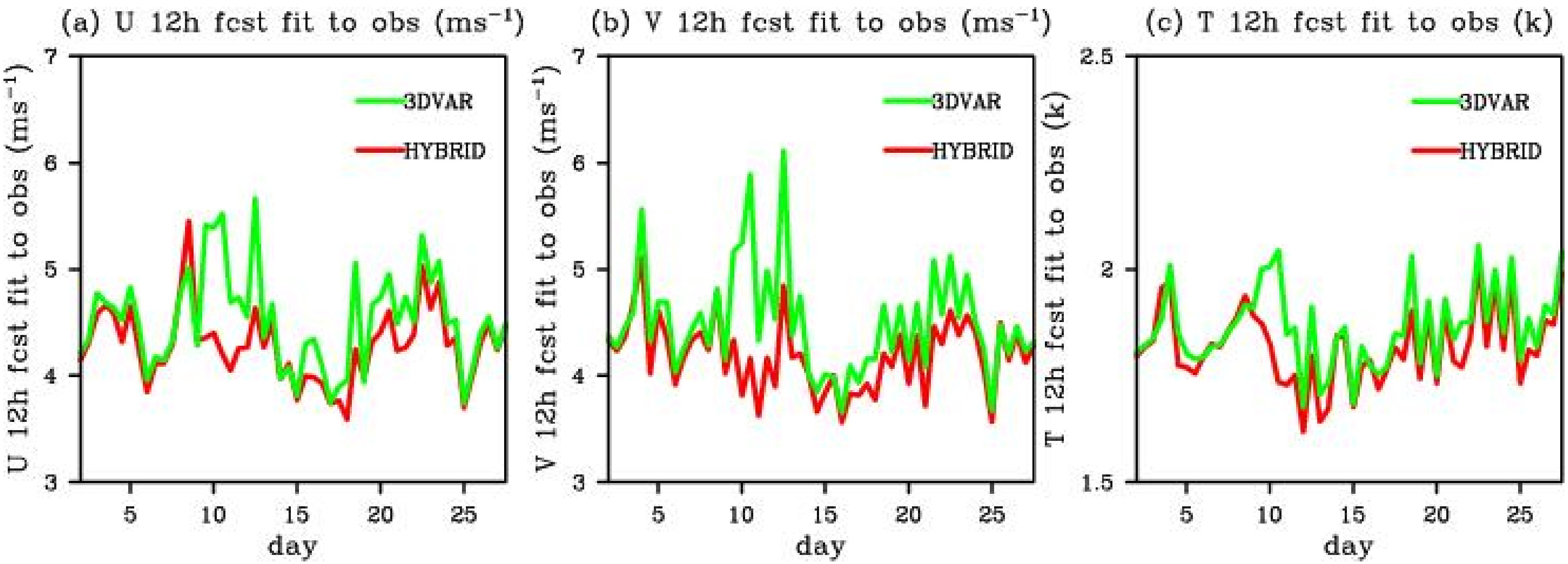
OSSE: root mean square analysis error



- Hybrid has larger improvement over data sparse regions, e.g., ocean; Flow-dependent ensemble covariance has the largest impact where observation is sparse.

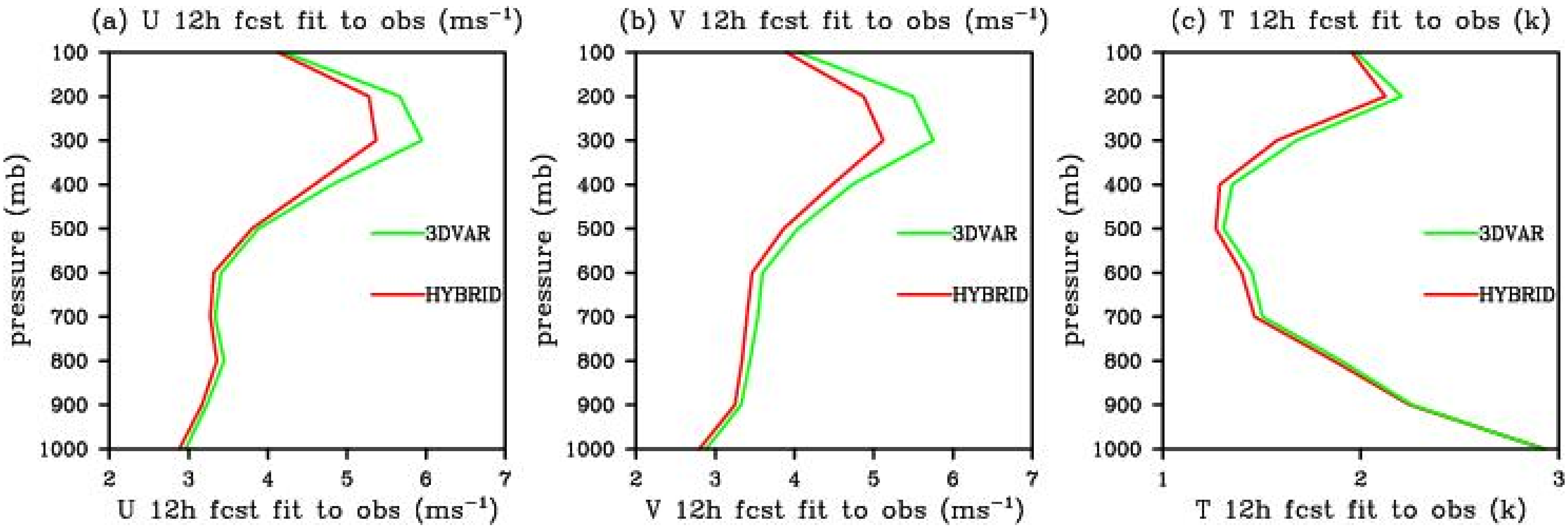
- Hybrid has larger improvement over western continent than the east; hybrid extrapolates observations info properly (flow dependently) to the upstream data void region.

Real obs. experiment: 12h forecast fit to obs.



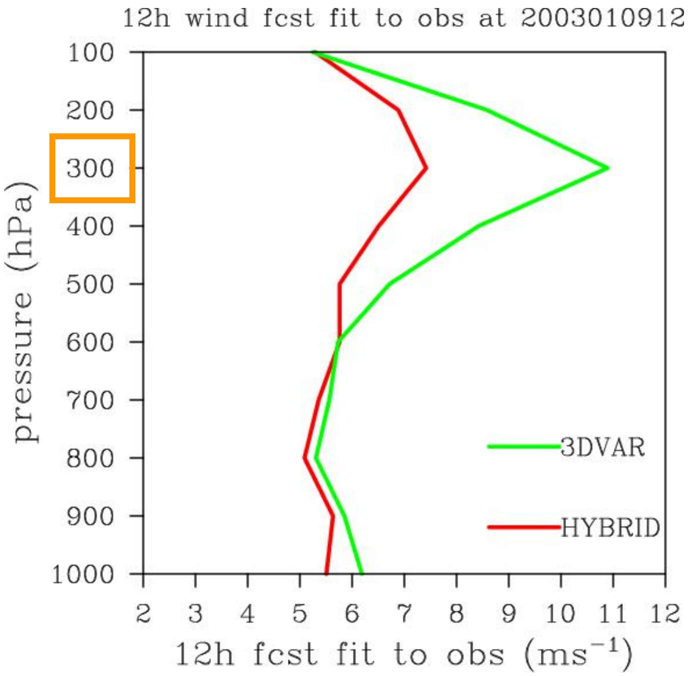
- Hybrid 12h forecast is more accurate than the 3DVAR for most time.

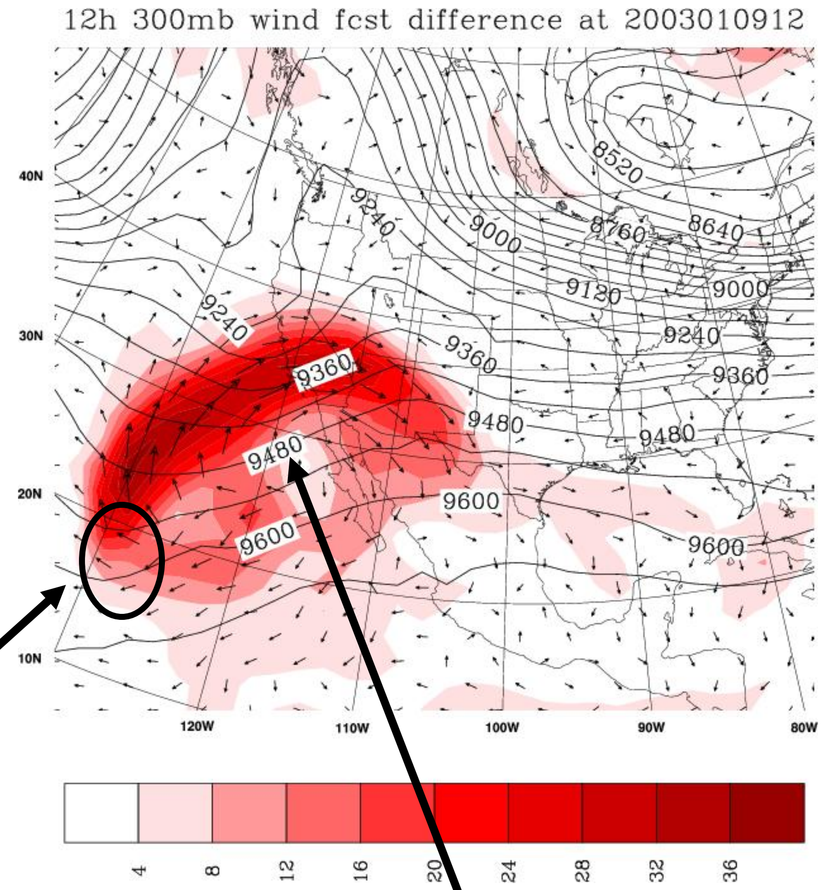
Real obs. experiment: 12h forecast fit to obs.



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

A case study: investigating 12h forecast improvement by hybrid at 12Z Jan 9 2003



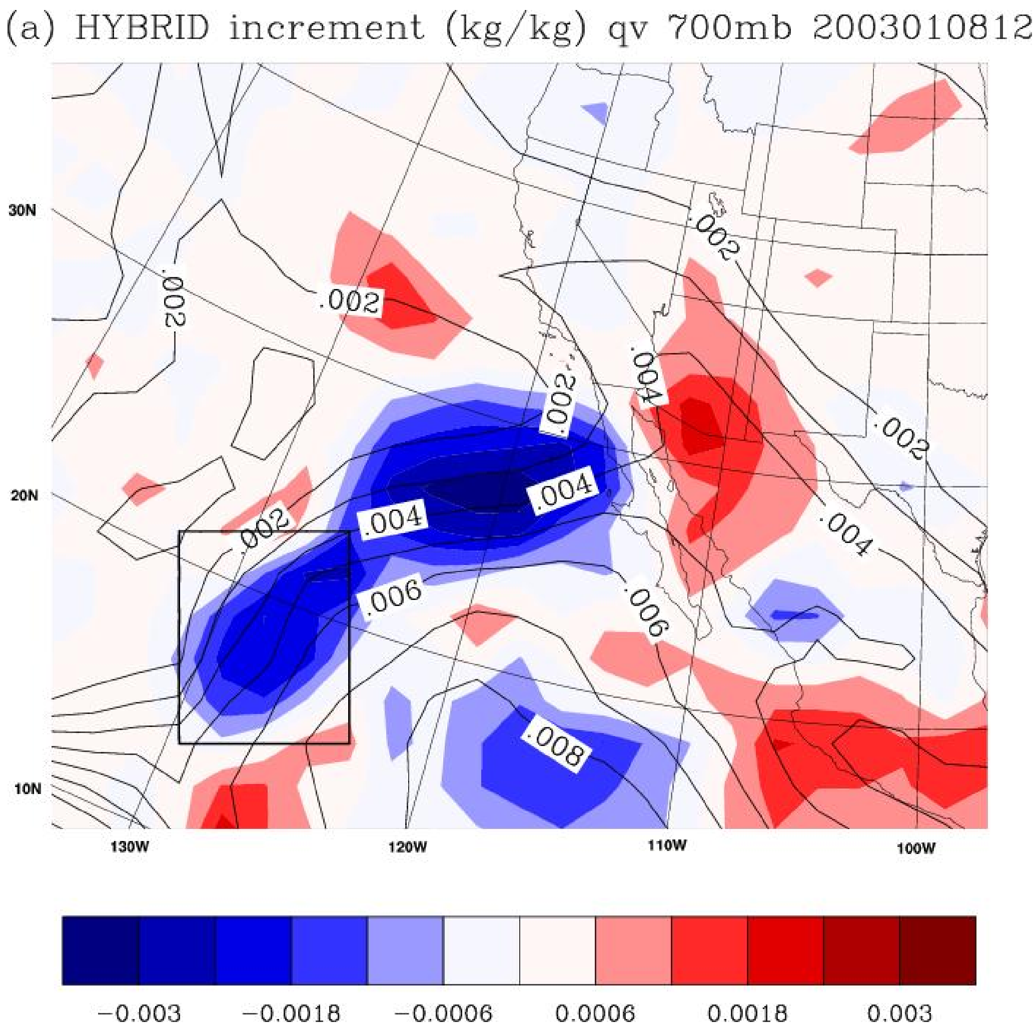


Spurious ridge anomaly in 3DVAR

- Spurious ridge anomaly started to appear two cycles earlier. It seemed to be associated with precipitation that was in 3DVAR, but not in the hybrid.

- Hybrid properly updated the upstream data void region to suppress the rain?

A case study: flow-dependent update of the data void upstream region by the hybrid



- Hybrid corrected the moisture field over the data void upstream region using observations far in land. It dried the lower troposphere along the warm front, reaching the region where 3DVAR was raining.

Summary and Discussion

- The hybrid ETKF-3DVAR scheme for WRF provided more accurate analyses and forecasts than WRF 3DVAR.

- The hybrid scheme can properly extrapolate observations to the data void region according to the flow and thus improved the forecasts downstream.

- For the real data experiment, expect more improvement with improved LBC ensembles and better treatment of model errors.

- Future work, test with more complete obs. network; applications on mesoscales; ETKF-4DVAR; etc.

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

$$\underbrace{\mathbf{X}^a}_{\text{Initial ensemble perturbation}} = \underbrace{\mathbf{X}^b}_{\text{forecast ensemble perturbation}} \underbrace{\mathbf{T}}_{}$$

Initial ensemble perturbation

forecast ensemble perturbation

- Computationally inexpensive for ensemble size of o(100), since transformation fully in perturbation subspace.

Fitting forecasts at 2003010912 to soundings at Vandenberg and El Paso