Hybrid Variational/Ensemble Data Assimilation

Dale Barker

Data Assimilation Course,

Buenos Aires, Argentina,

November 5th 2008

Outline of Talk

1) Introduction

2) Background Error Modeling

3) Examples of Flow-Dependent Errors in Variational DA.

4) Hybrid Variational/Ensemble Data Assimilation

1. Introduction

Motivation

- Intuitively expect forecast errors to be flow-dependent.
- Ensemble DA implicitly flow-dependent, but issues (e.g. sampling error).
- Studies (e.g. Barker 1999, Hamill and Snyder 2000, Etherton and Bishop 2004, Wang et al. 2007) show promising results of a hybrid variational/ensemble approach.
- Efficient leveraging of variational/ensemble DA resources.

Flow-Dependent EnKF Forecast Error Covariances

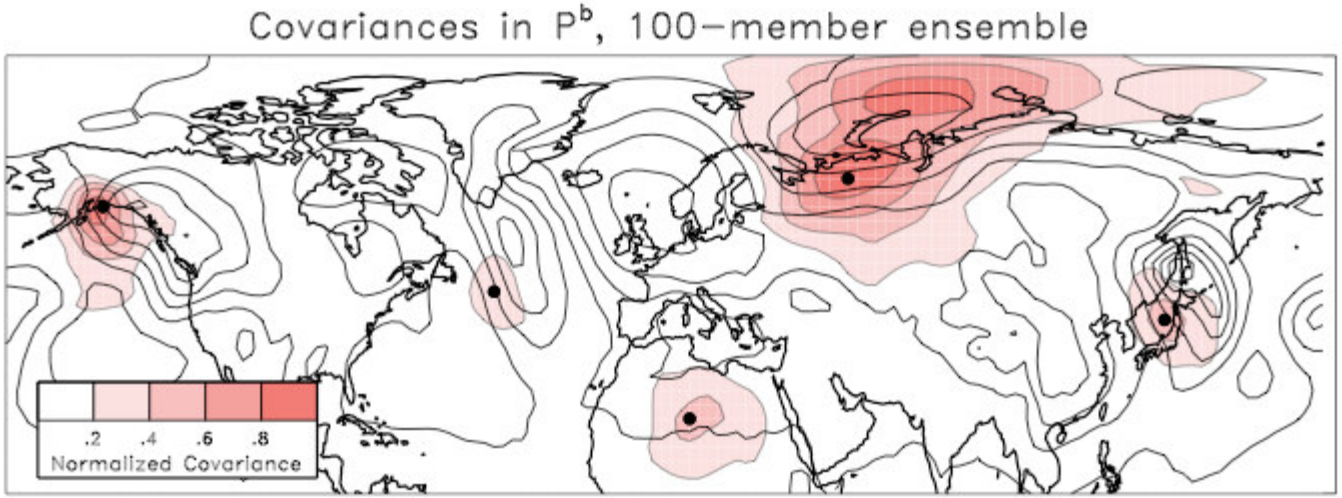


Figure 2. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured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

From Hamill (2006)

2. Background Error Modeling

Variational Data Assimilation

- The components J_b and J_o of the cost function are defined as

$$J_b\left[\mathbf{x}(t_0)\right] = \frac{1}{2}\left[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)\right]^T \mathbf{B}_o^{-1}\left[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)\right]$$

$$J_o\left[\mathbf{x}(t_0)\right] = \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{y}_i - \mathbf{y}_i^o\right]^T \mathbf{R}_i^{-1}\left[\mathbf{y}_i - \mathbf{y}_i^o\right]$$

- $\mathbf{B_0}$ is an *a priori* weight matrix estimating the error covariance of $\mathbf{x^b}$.

- $\mathbf{R_i}$ is the observation error covariance matrix at time *i*.

- Direct calculation of J_b and J_o impossible for NWP problems ($\mathbf{B_0}$, $\mathbf{R}$ are matrices of dimension 10^7). Therefore many practical simplifications required.

- *Incremental Var* produces analysis increments that are added back to a *first guess* field $\mathbf{x^g}$ to produce the analysis, i.e.

$$\mathbf{x}^a(t_0) \equiv \mathbf{x}^g(t_0) + \delta\mathbf{x}(t_0)$$

Model-Based Estimation of Climatological Background Errors

- Assume background error covariance estimated by model perturbations *x'* :

$$\mathbf{B}_0 = \overline{(\boldsymbol{x}^b - \boldsymbol{x}^t)(\boldsymbol{x}^b - \boldsymbol{x}^t)^T} \approx \overline{\boldsymbol{x}'\boldsymbol{x}'^T}$$

Two ways of defining ***x'***:

- The NMC-method (Parrish and Derber 1992):

$$\mathbf{B}_0 = \overline{\boldsymbol{x}'\boldsymbol{x}'^T} \approx A\overline{(\boldsymbol{x}^{t2} - \boldsymbol{x}^{t1})(\boldsymbol{x}^{t2} - \boldsymbol{x}^{t1})^T}$$

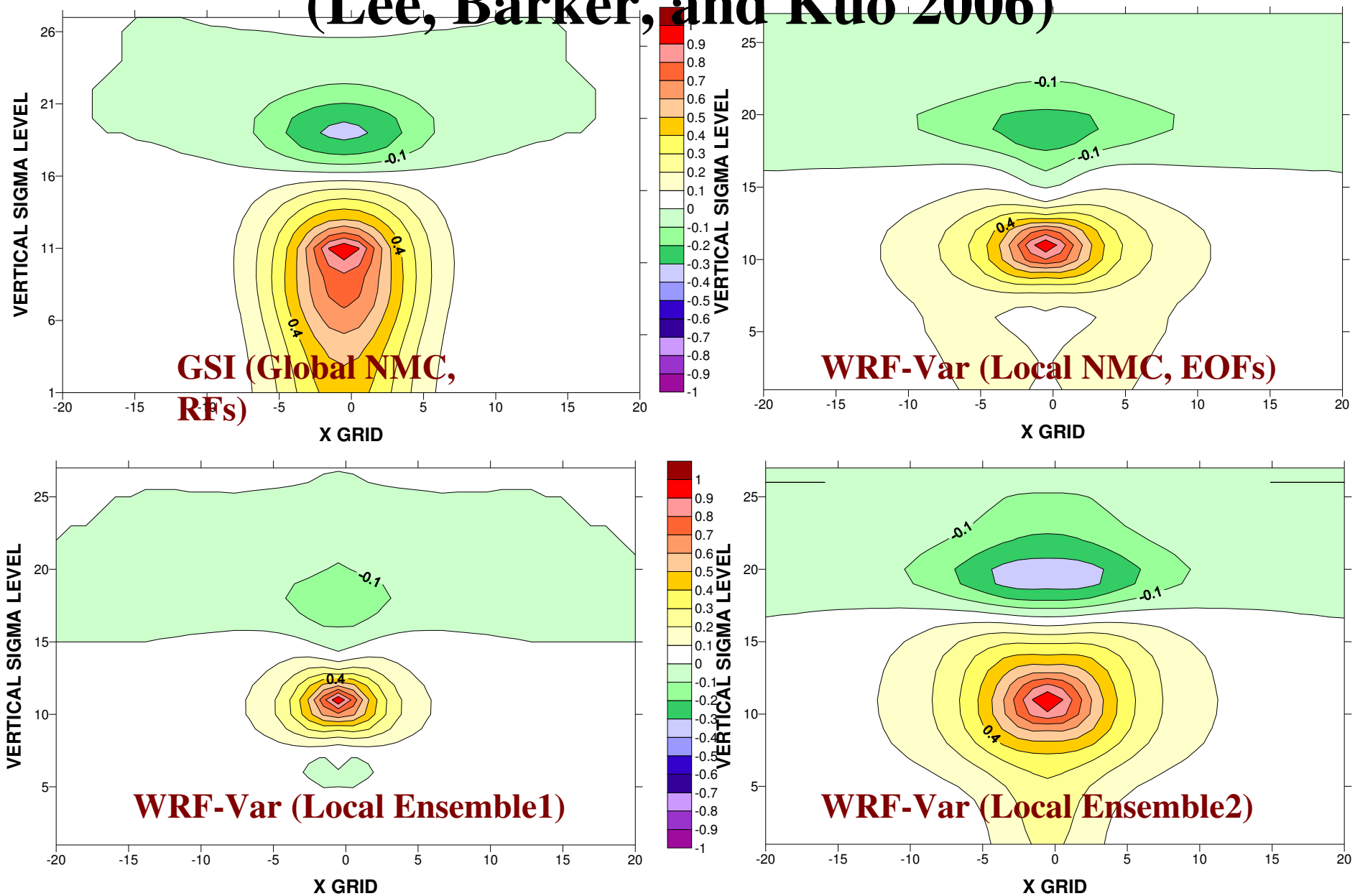
  where e.g. t2=24hr, t1=12hr forecasts…

- …or ensemble perturbations (Fisher 2003):

$$\mathbf{B}_0 = \overline{\boldsymbol{x}'\boldsymbol{x}'^T} \approx C\overline{(\boldsymbol{x}^k - \langle \boldsymbol{x} \rangle)(\boldsymbol{x}^k - \langle \boldsymbol{x} \rangle)^T}$$

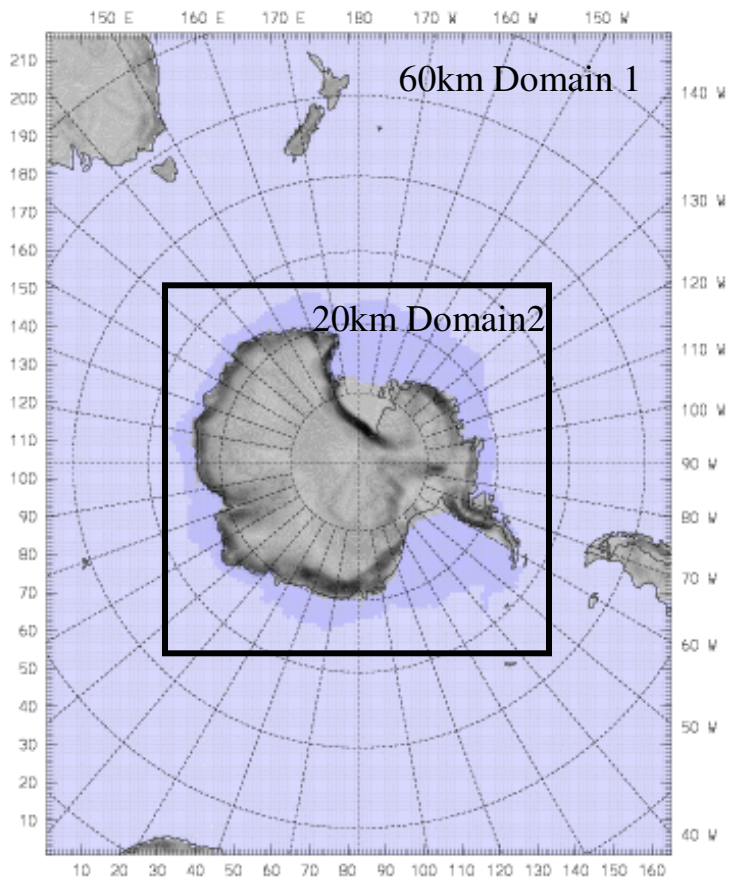
- Tuning via innovation vector statistics and/or variational methods.

Analysis Sensitivity to Background Error Model (Lee, Barker, and Kuo 2006)



T increments : T Observation (1 Deg , 0.001 error around 850 hPa)

Sensitivity to Forecast Error Covariances in Antarctica (Rizvi et al 2006)



60km Verification (vs. radiosondes)

T+24hr Temperature

GFS-based errors

WRF-based errors

Incremental WRF-Var J_b Preconditioning

$$J_b\left[\delta\mathbf{x}(t_0)\right] = \frac{1}{2}\left\{\delta\mathbf{x}(t_0) - \left[\mathbf{x}^b(t_0) - \mathbf{x}^g(t_0)\right]\right\}^T \mathbf{B}_o^{-1}\left\{\delta\mathbf{x}(t_0) - \left[\mathbf{x}^b(t_0) - \mathbf{x}^g(t_0)\right]\right\}$$

- *Define* ***preconditioned control variable*** $\boldsymbol{v}$ space transform

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v}$$

  where **U** transform CAREFULLY chosen to satisfy $\mathbf{B_o} = \mathbf{UU}^T$.

- Choose (at least assume) control variable components with uncorrelated errors:

$$J_b\left[\delta\mathbf{x}(t_0)\right] = \frac{1}{2}\sum_n v_n^2$$

- where n~number pieces of independent information.

Speedy Model Background Error Modeling

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v} \qquad \mathbf{B_o} = \mathbf{U}\mathbf{U}^{\mathbf{T}}$$

- *In Speedy:* $\mathbf{U} = \mathbf{VCA}$
- **A** = Background error standard deviation.
- **C** = Spatial correlation function (ignore vertical correlations)
- **V** = Impose multivariate correlations:

$$u = u_u + r_1 u_g(p_s, T) \qquad v = v_u - r_2 v_g(p_s, T)$$

WRF-Var Background Error Modeling

cv_options		2 (original MM5)	3(GSI)	4 (Global)	5(regional)
Analysis increments	$\mathbf{x}'$	$u', v', T', q', p_s'(i,j,k)$			
Change of Variable	U_p	$\psi', \chi', p_u', q'(i,j,k)$	$\psi', \chi_u', T_u', \tilde{r}', p_{su}'(i,j,k)$		
Vertical Covariances	U_v	$\mathbf{B} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}}$	RF	$\mathbf{B} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}}$	
Horizontal Correlations	U_h	RF		Spectral	RF
Control Variables	$\mathbf{v}$	$\mathbf{v}(i,j,m)$		$\mathbf{v}(l,n,m)$	$\mathbf{v}(i,j,m)$

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h\mathbf{v}$$

Define control variables:

$$\psi'$$

$$r' = q'/q_s(T_b, q_b, p_b)$$

$$\chi' = \chi_u' + \chi_b'(\psi')$$

$$T' = T_u' + T_b'(\psi')$$

$$p_s' = p_{su}' + p_{sb}'(\psi')$$

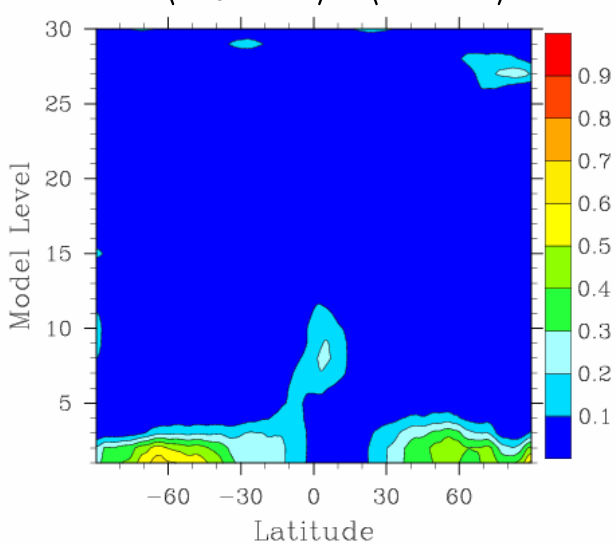
WRF-Var Statistical Balance Constraints

- Define statistical balance after Wu et al (2002):

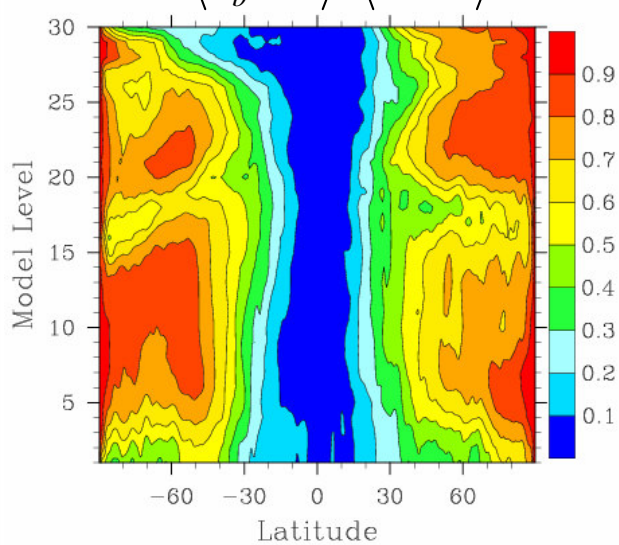
$$\chi_b^{'} = c\psi^{'} \quad T_b^{'}(k) = \sum_{k1} G(k,k1)\psi^{'}(k1) \quad p_{sb}^{'} = \sum_{k} W(k)\psi^{'}(k)$$

- How good are these balance constraints? Test on KMA global model data. Plot correlation between "Full" and balanced components of field:

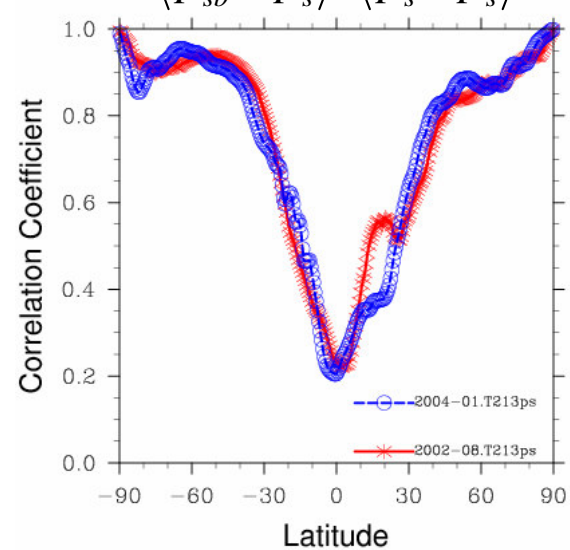
$\langle \chi_b \bullet \chi \rangle / \langle \chi \bullet \chi \rangle$



$\langle T_b \bullet T \rangle / \langle T \bullet T \rangle$

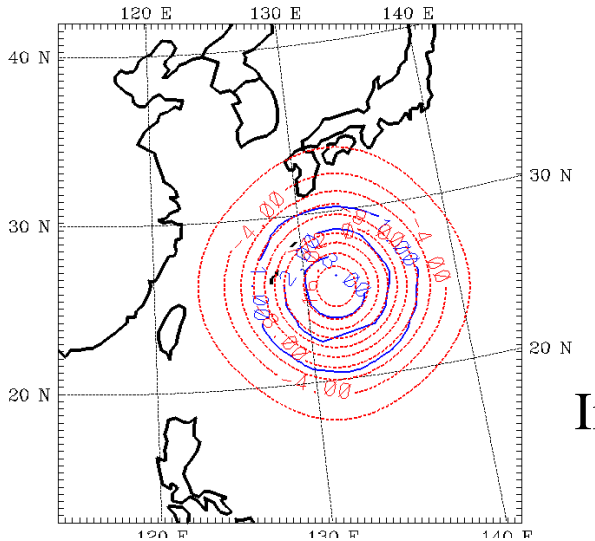


$\langle p_{sb} \bullet p_s \rangle / \langle p_s \bullet p_s \rangle$

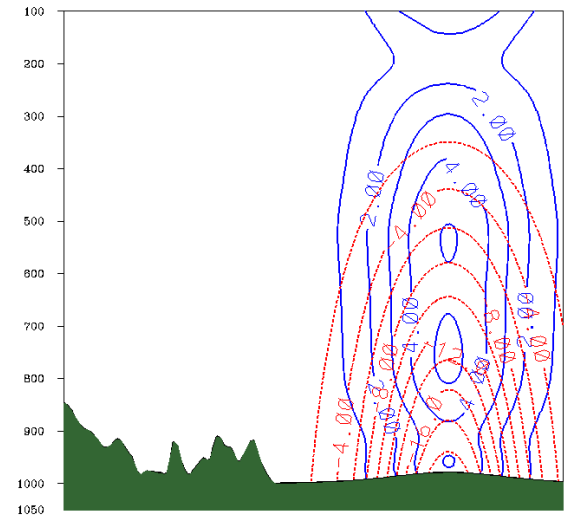


Flow-Dependence Via Nonlinear (Dynamical) Balance Equation

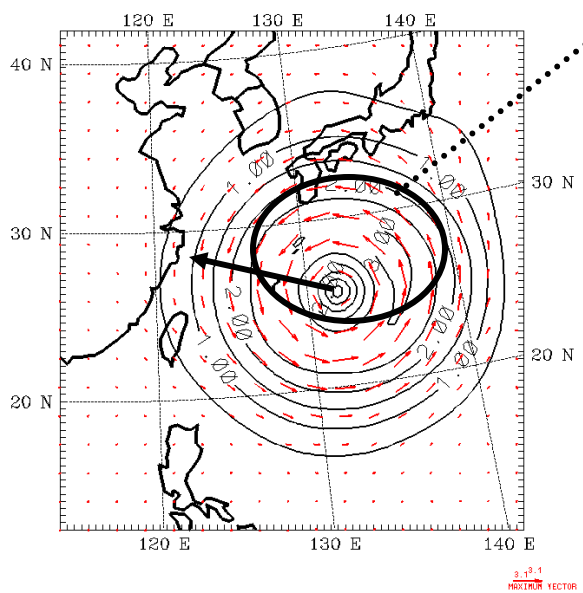
$$\nabla^2 p_b' = -\nabla \cdot \bar{\rho}[\bar{\mathbf{v}} \cdot \nabla \mathbf{v}' + \mathbf{v}' \cdot \nabla \bar{\mathbf{v}} + f\mathbf{k} \times \mathbf{v}']$$



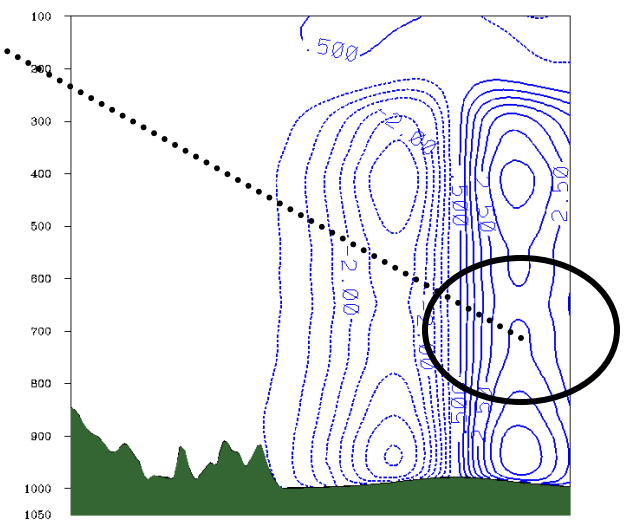
Pressure,
Temperature



Impact of cyclostrophic balance term



Wind Speed,
Vector,
v-wind component.



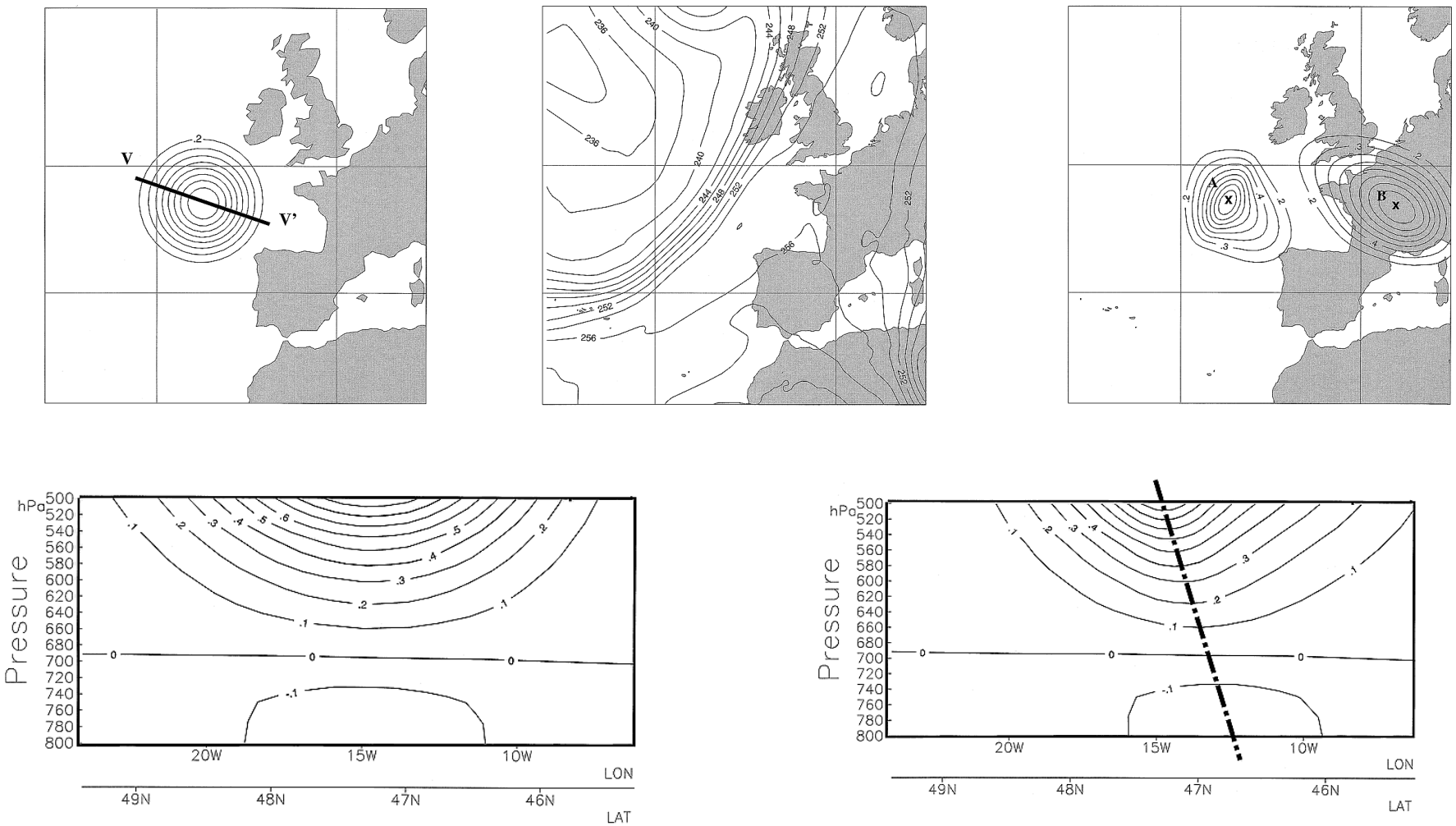
Barker et al (2004)

3. Examples Of Flow-Dependent Errors in Variational DA

Flow-Dependent Errors in VAR

- Commonly assumed in research community that 3D-Var error covariances are limited to static/isotropic.
- Many different techniques tried to implement flow-dependence in 3D-Var.
- Generally harder to implement flow-dependence in VAR DA schemes than ENS DA schemes.
- 4D-Var permits flow-dependence, but background errors typically static.
- Limited success to date (other than 4D-Var).

Semi-Geostrophic Transformation

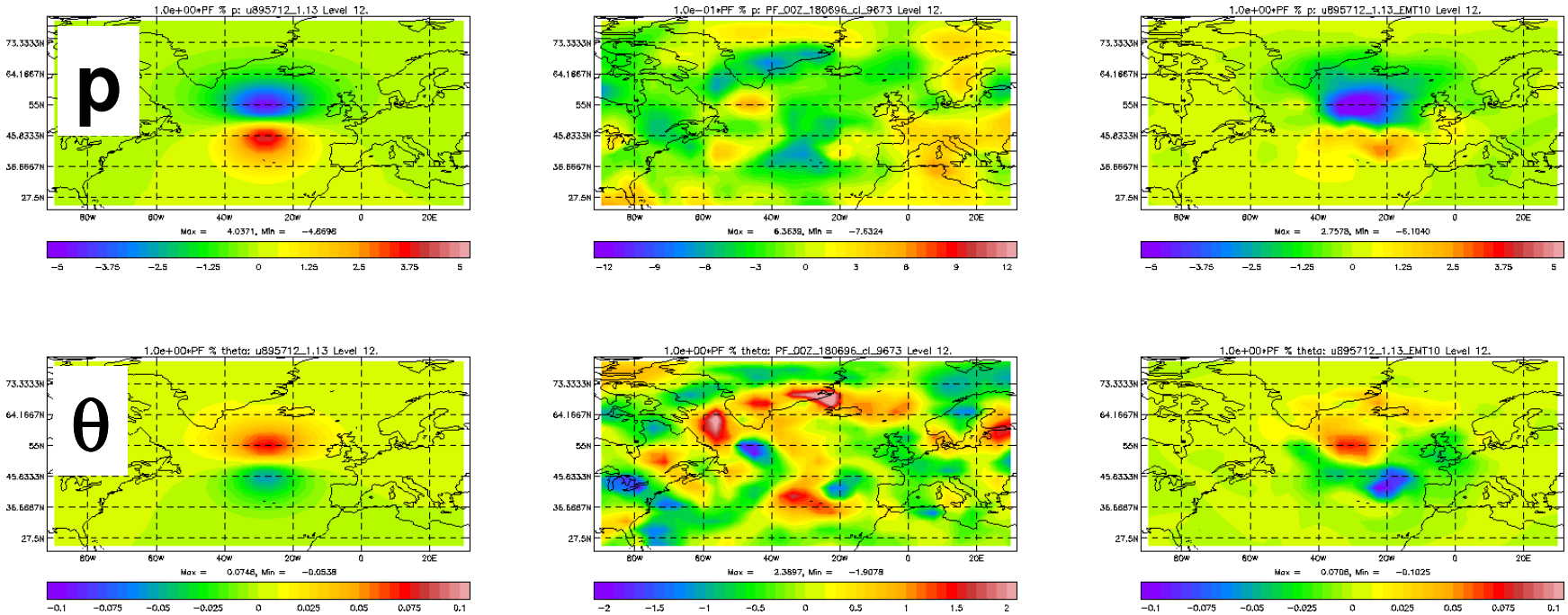


From Desroziers (1997)

Flow-Dependence Via Extra Control Variables 1

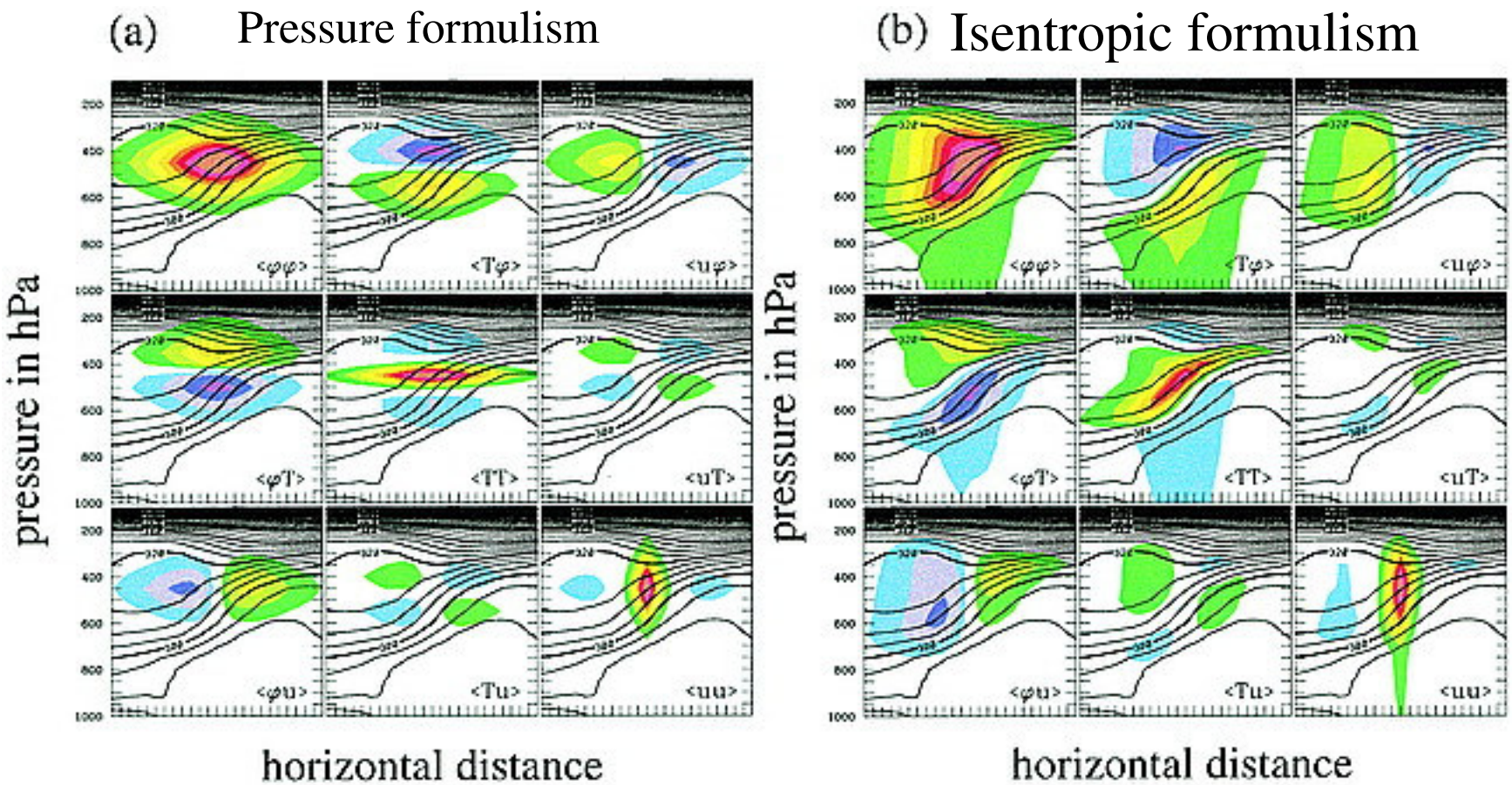
- Increments due to a u-wind observation at 50N, 30W (O-B=1m/s)

Static | Bred-Mode | Flow-Dependent



Barker (1999)

Observation-space 3D-Var (e.g. PSAS, NAVDAS)

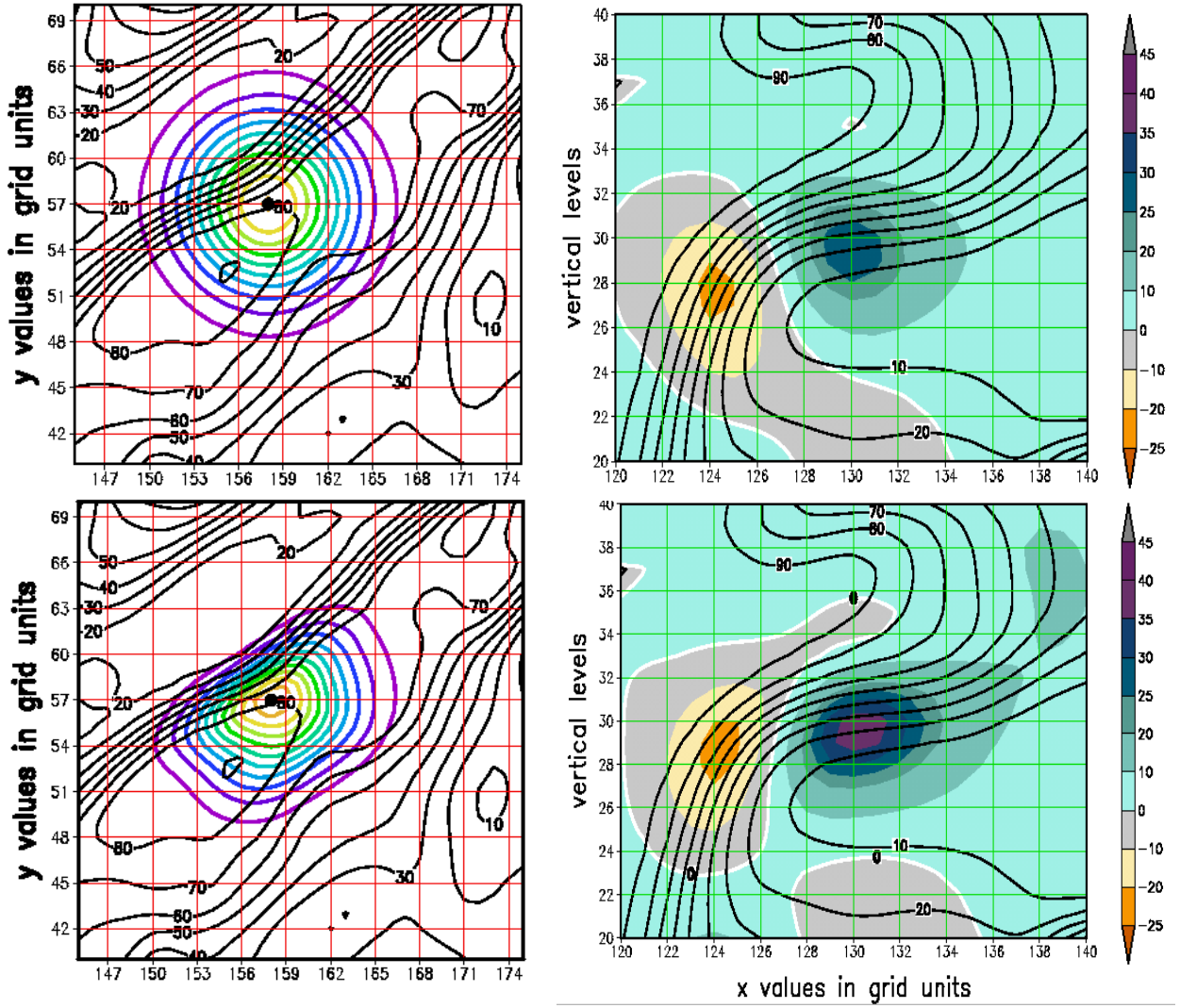


(from Daley and Barker 2000)

Flow-Dependence Via Anisotropic Recursive Filters (courtesy NCEP/EMC)

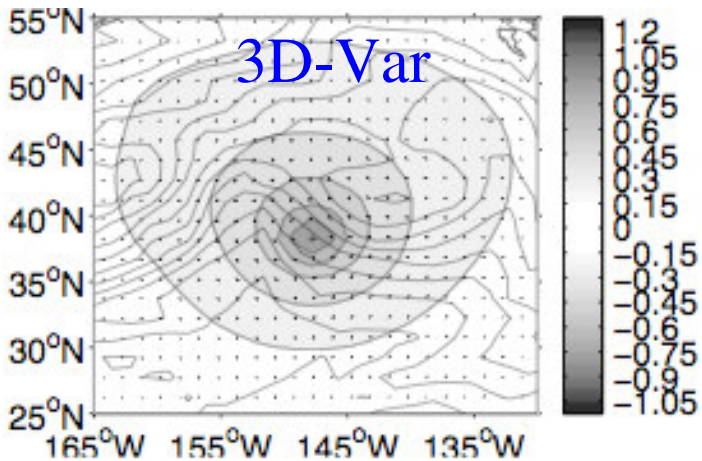
Isotropic:

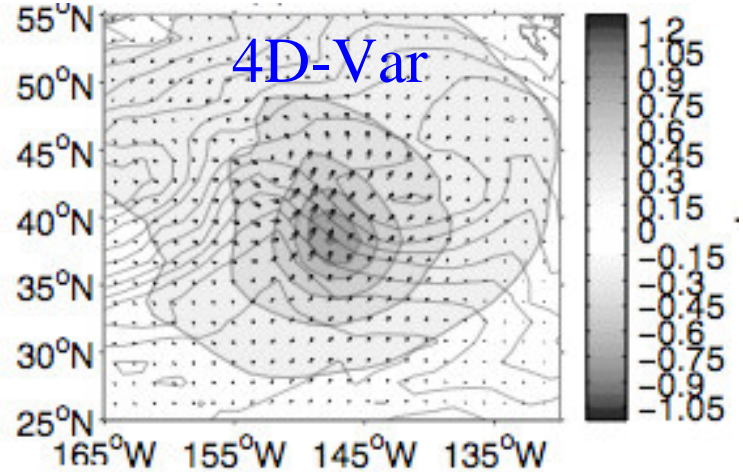
Anisotropic:

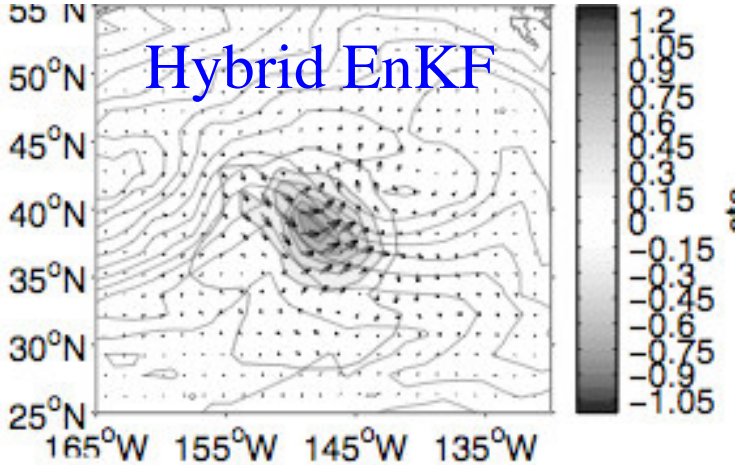


Flow-Dependence Via Extra Control Variables 2

- Temperature increments of single temperature observation

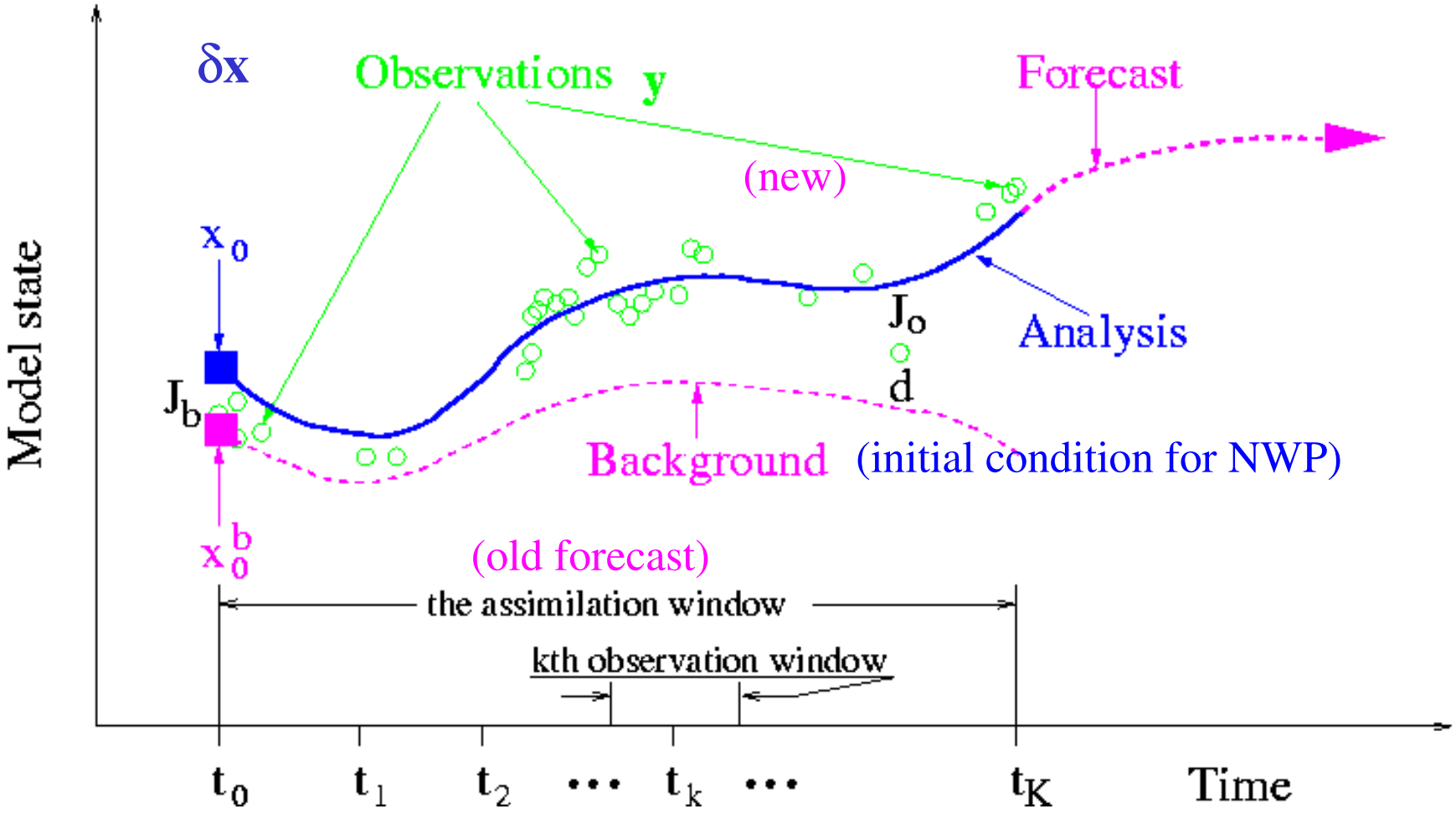




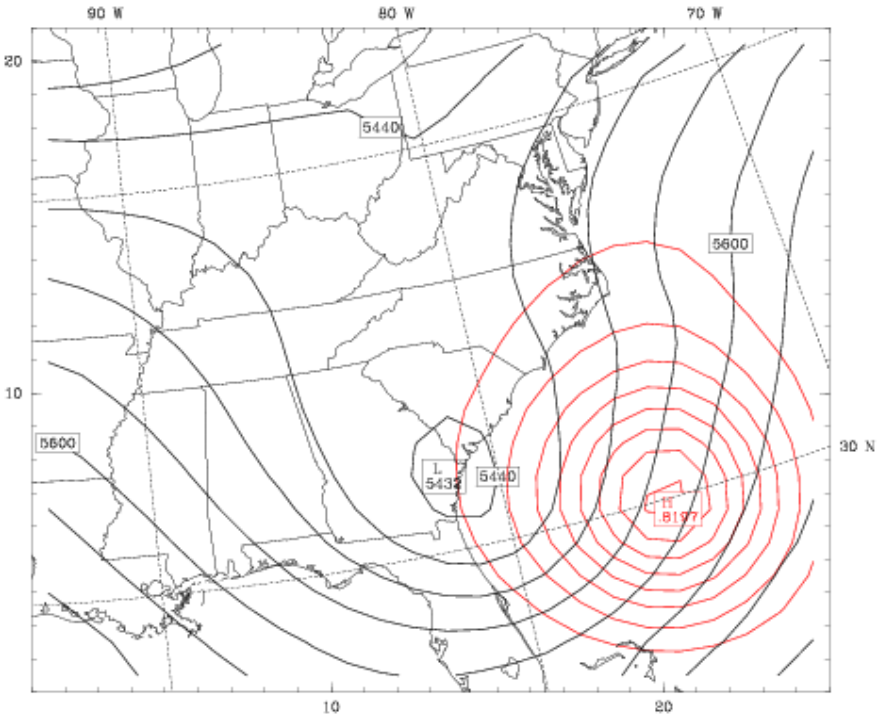


Buehner (2005)

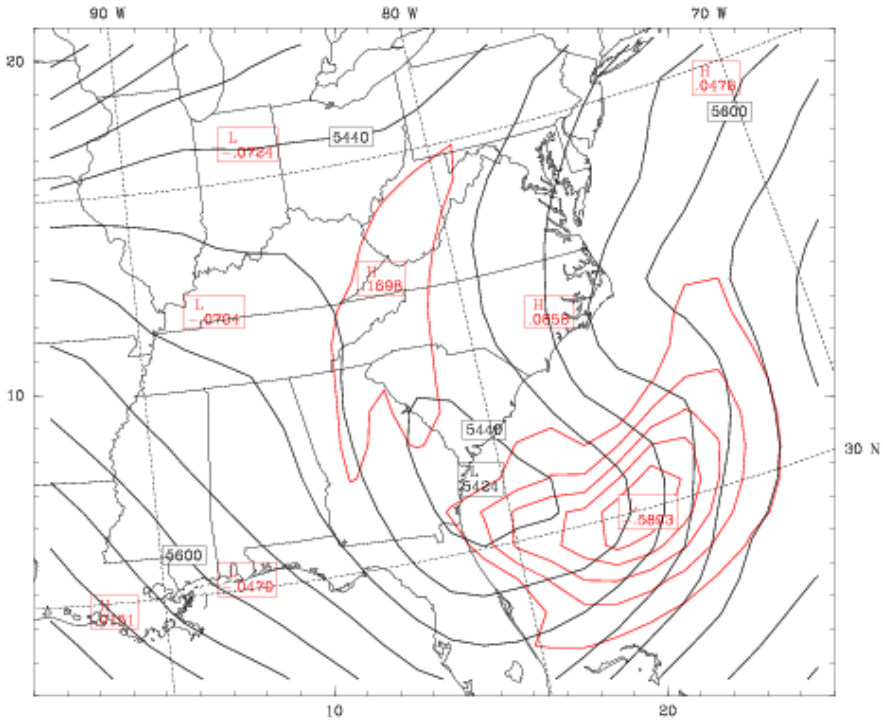
4D Variational Data Assimilation



500mb θ increments from 3D-Var at 00h and from 4D-Var at 06h due to a 500mb T observation at 06h

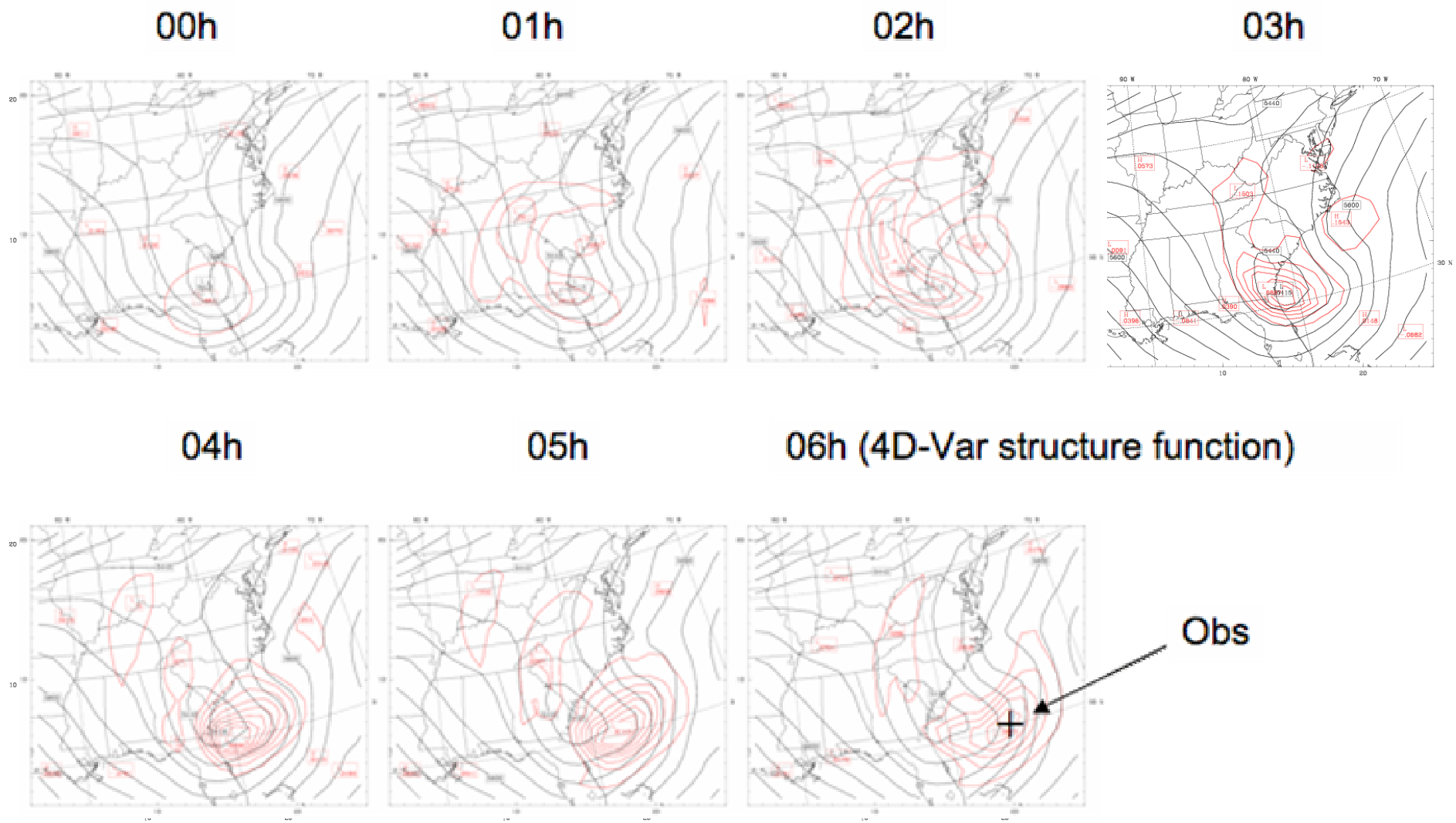


3D-Var (NoFGAT)

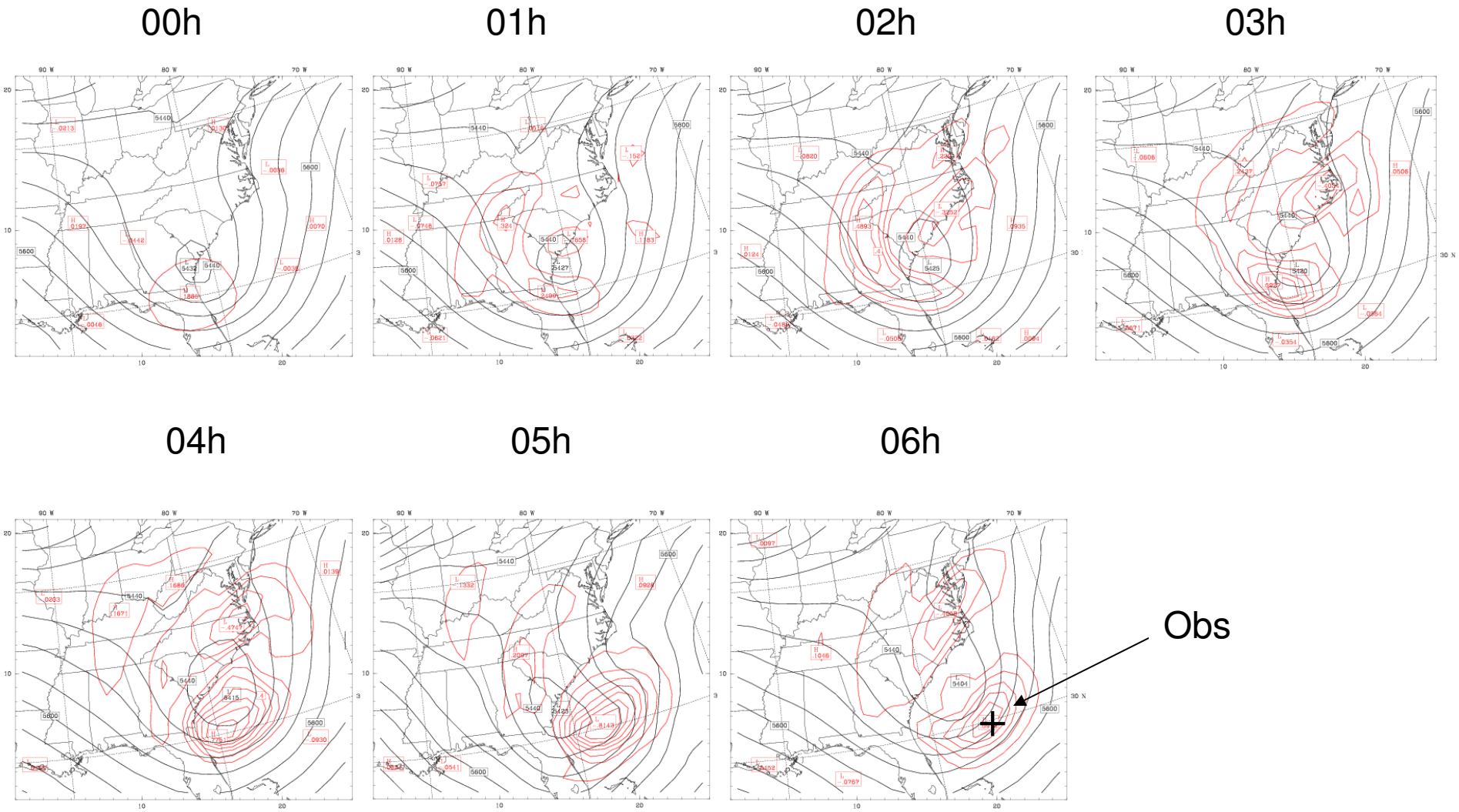


4D-Var

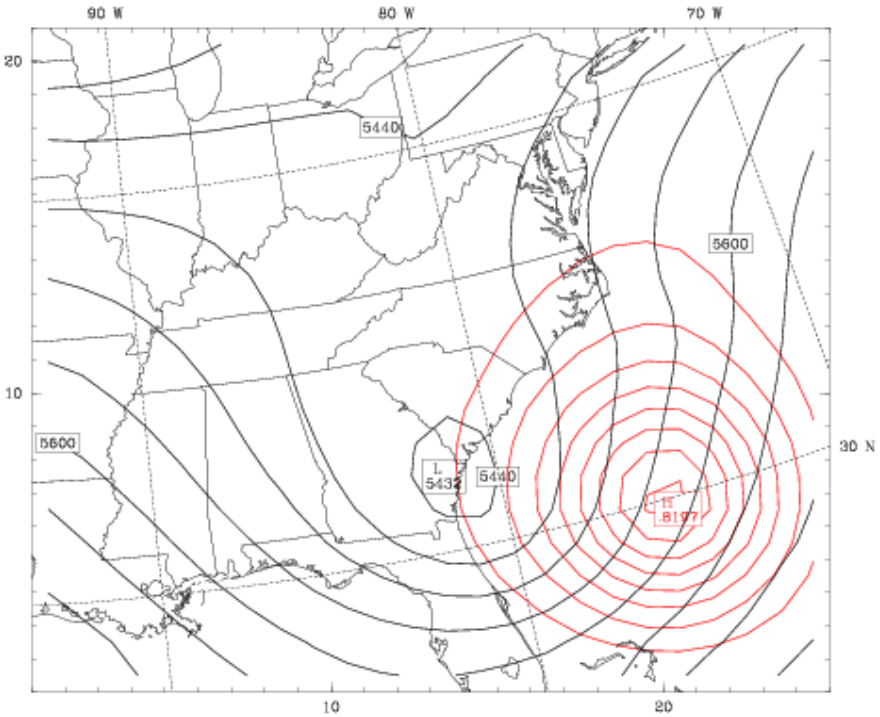
500mb θ increments at 00,01,02,03,04,05,06h to a 500mb T ob at 06h



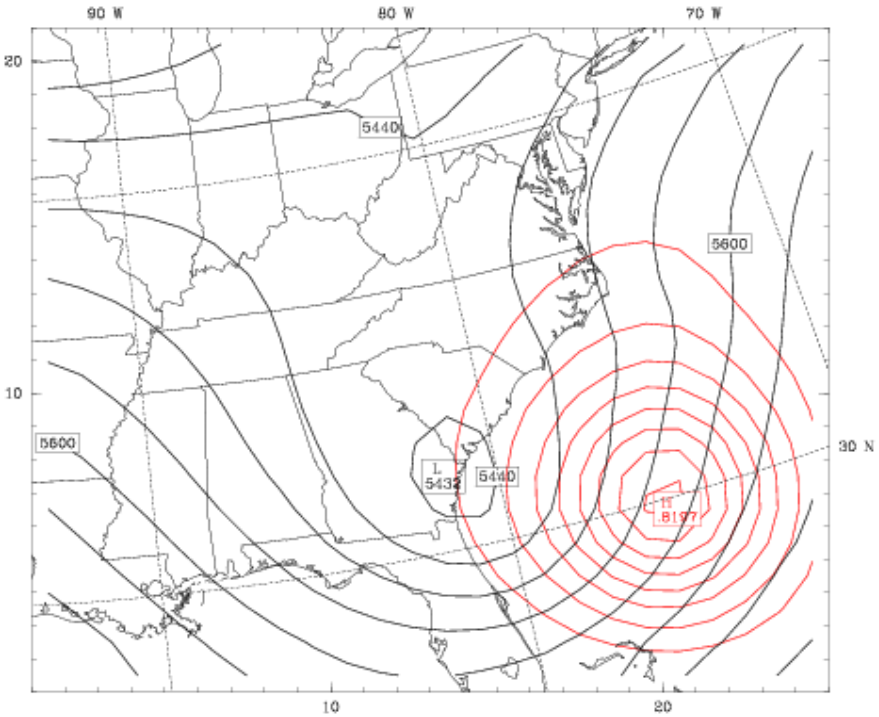
500mb θ difference at 00,01,02,03,04,05,06h between two nonlinear runs (one from background; one from 4D-Var)



500mb θ increments from 3D-Var at 00h and from 4D-Var at 00h due to a 500mb T observation at 00h



3D-Var (NoFGAT)

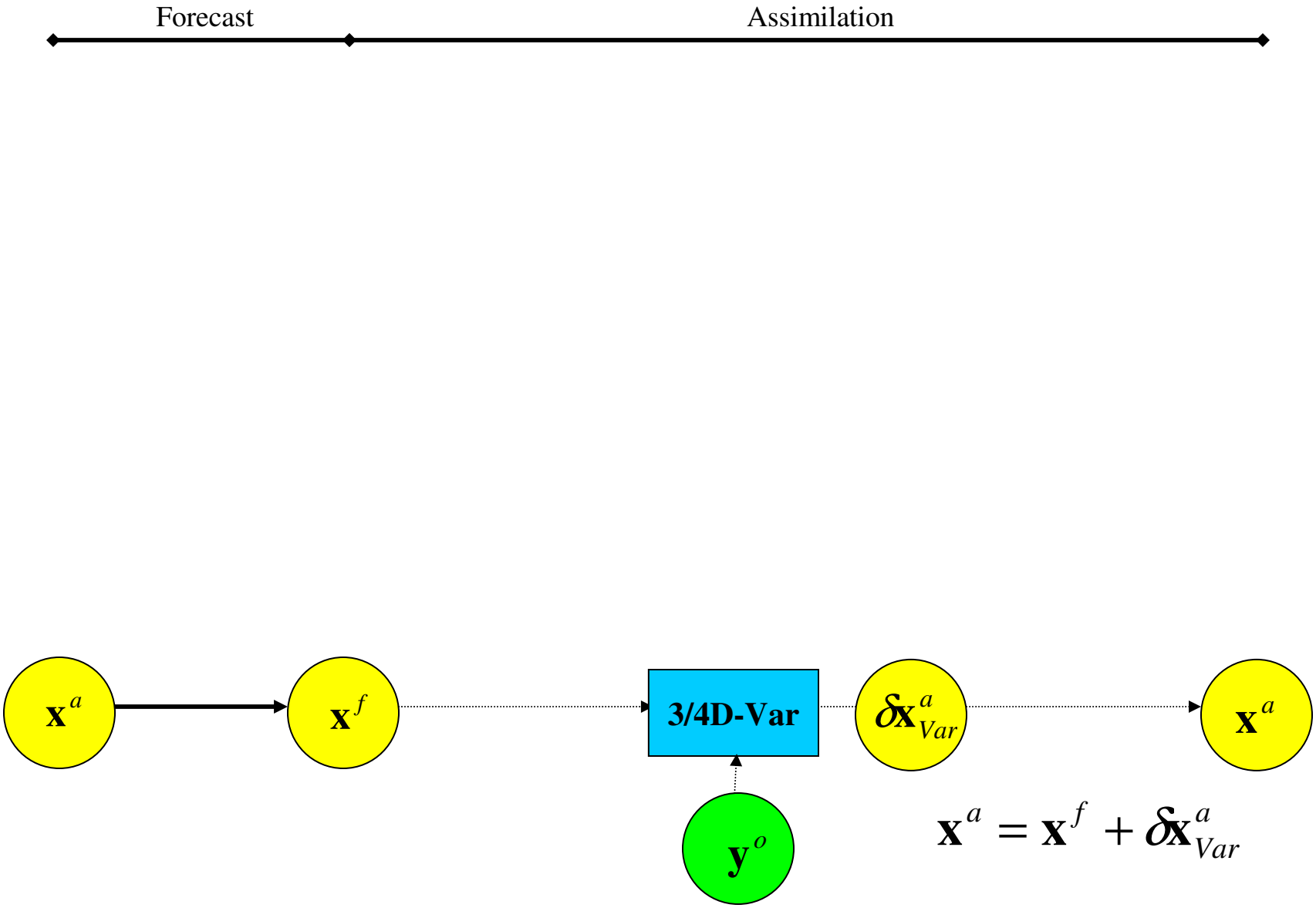


4D-Var

4D-Var introduces no flow-dependence for observations at analysis time!

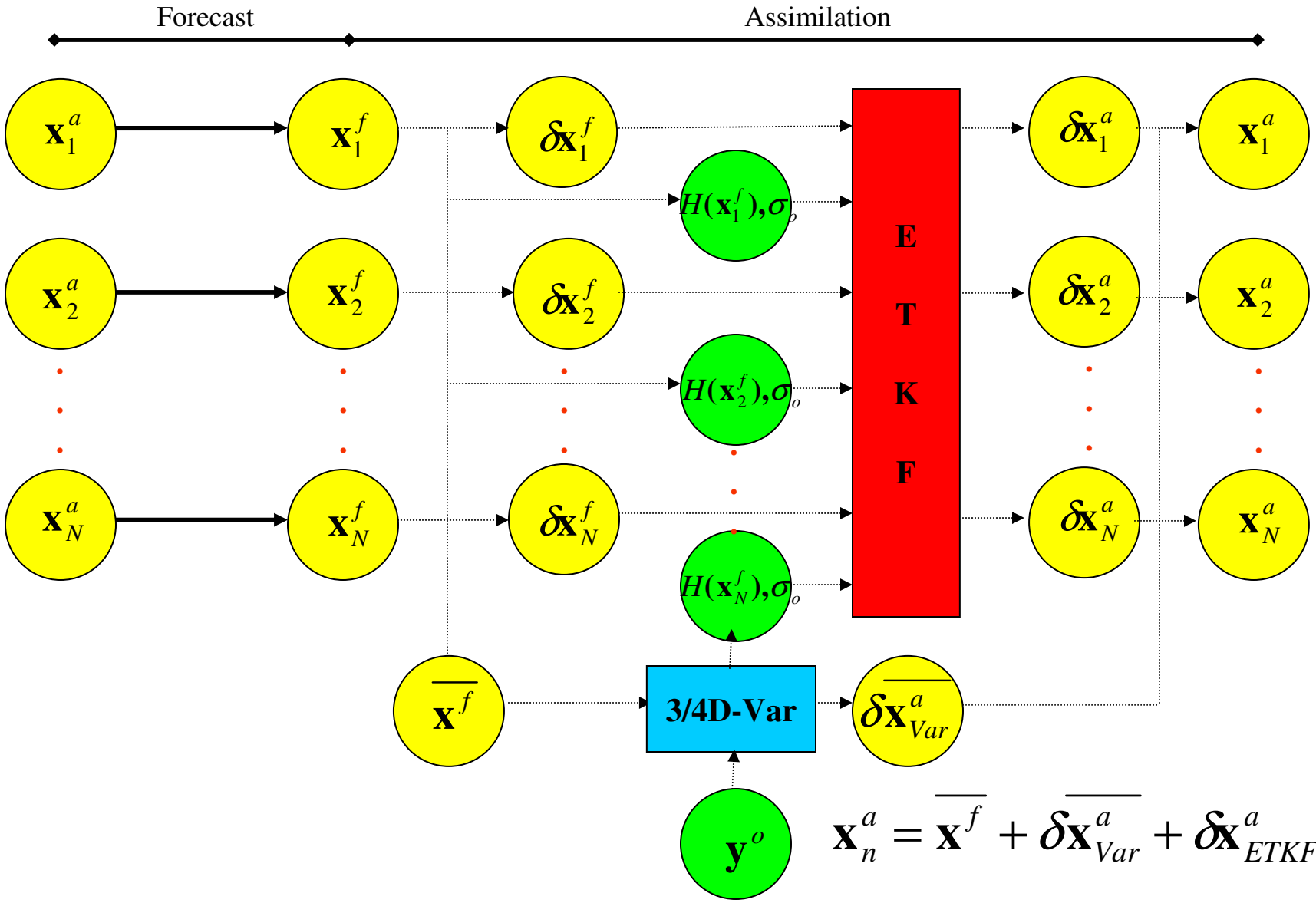
4. Hybrid Variational/Ensemble Data Assimilation

1. Deterministic Cycling NWP System



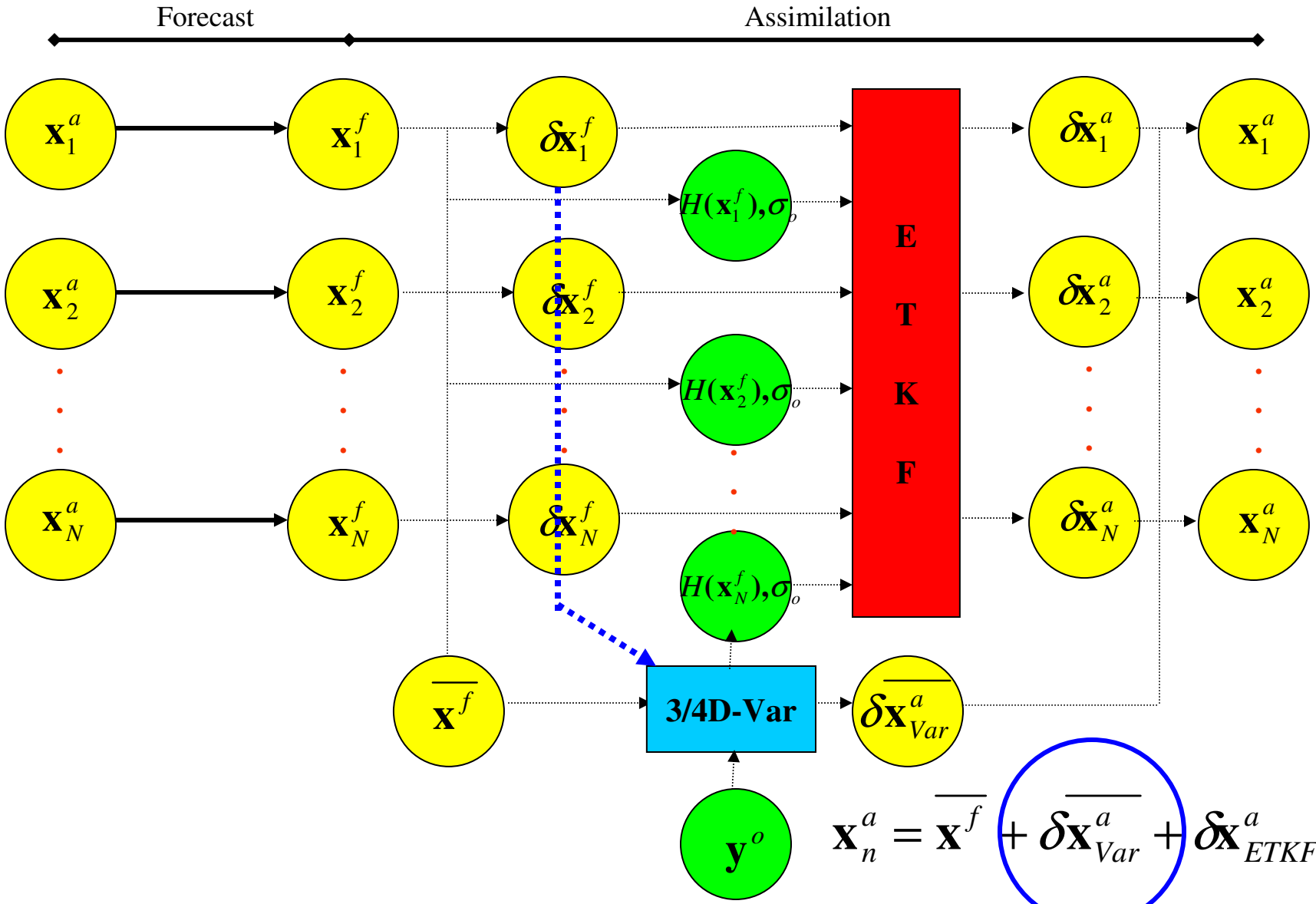
$$\mathbf{x}^a = \mathbf{x}^f + \delta\mathbf{x}^a_{Var}$$

Cycling WRF/WRF-Var/ETKF System



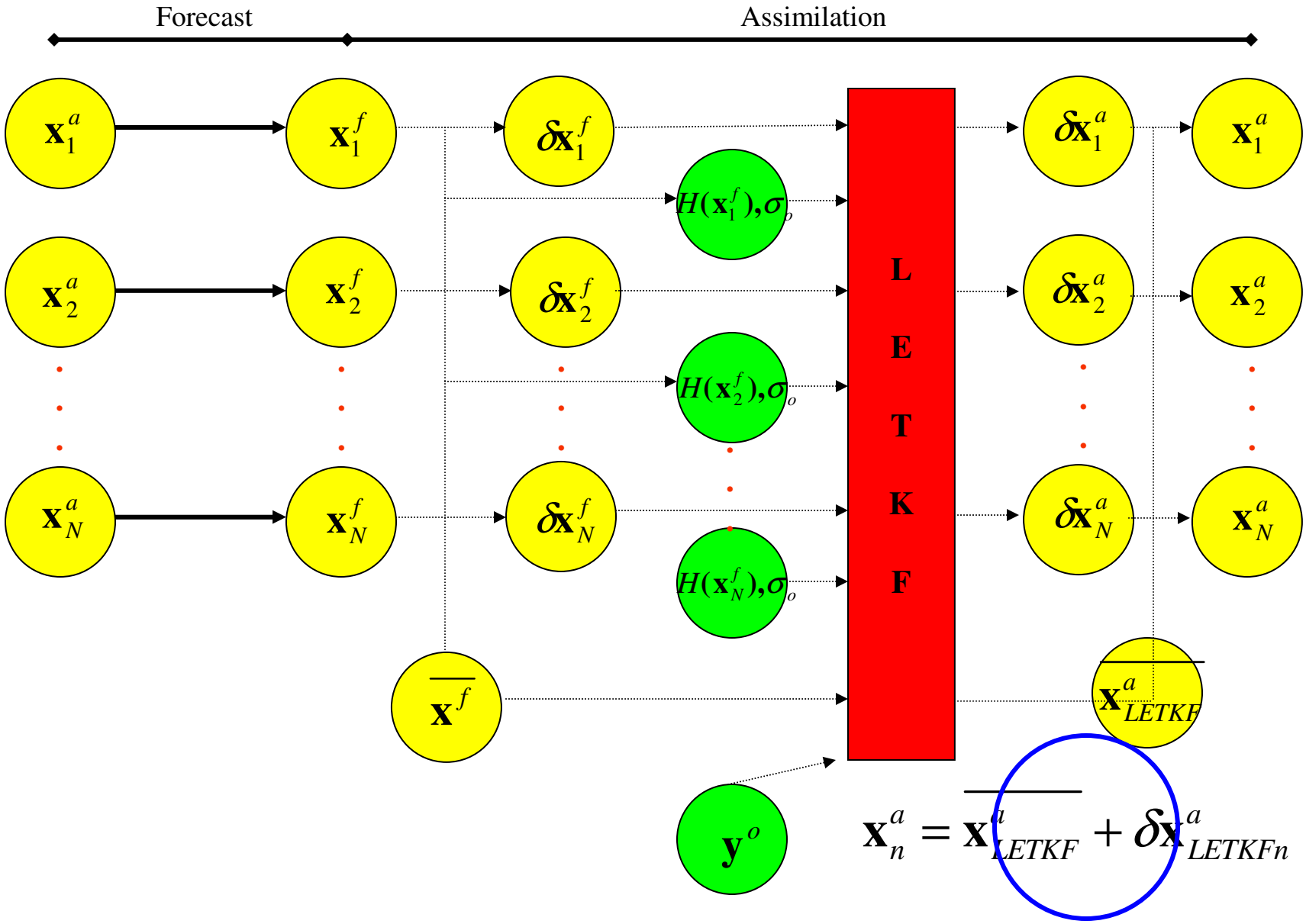
$$\mathbf{x}_n^a = \overline{\mathbf{x}^f} + \overline{\delta\mathbf{x}_{Var}^a} + \delta\mathbf{x}_{ETKFn}^a$$

Cycling WRF/WRF-Var/ETKF System (Hybrid DA)



$$\mathbf{x}_n^a = \overline{\mathbf{x}^f} + \overline{\delta\mathbf{x}_{Var}^a} + \delta\mathbf{x}_{ETKFn}^a$$

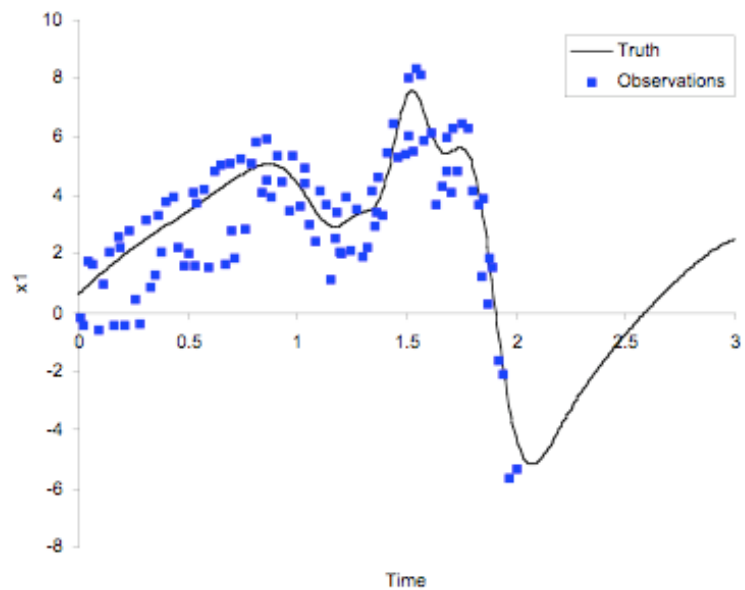
Cycling Ensemble (e.g. LETKF) Data Assimilation System



$$\mathbf{x}_n^a = \overline{\mathbf{x}_{LETKF}^a} + \delta\mathbf{x}_{LETKFn}^a$$

Hybrid Testing With Lorenz 1996 model (K. Y. Chung: KMA Visitor to NCAR)

$$\frac{d\mathbf{x}_i}{dt} = (\mathbf{x}_{i+1} - \mathbf{x}_{i-2})\mathbf{x}_{i-1} - \mathbf{x}_i + \mathbf{F}$$



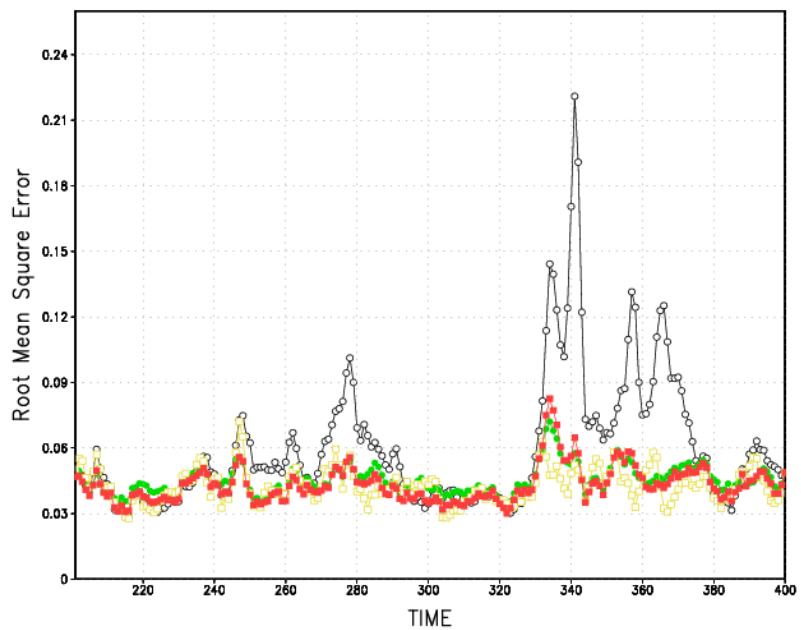
- 40-variable model (i=1, 40). **F**=8.
- Periodic boundary conditions $\mathbf{x}_1 = \mathbf{x}_I$
- OSSE: dt=6h. Simulate obs every 12h. 400d run, verify last 200d.
- Hybrid 3D-Var/EnKF (Hamill and Snyder 2000): $\mathbf{P}^f = (1-\beta)\mathbf{P}_e^f + \beta\mathbf{B}_0$
- Flow-dependence $\mathbf{P}_e^f$ via a) Lagged forecast diff., b) EnKF perturbations.

Lorenz Model OSSE Analysis Error, N=10

Root Mean Square Analysis Error (x100)

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3D-Var	5.70								
Hybrid	4.21	4.33	**4.19**	4.46	4.32	4.37	4.86	5.00	5.20
EnKF	4.59								

Optimal β=0.3 consistent with Etherton and Bishop (2004)



3D-Var

β=0.3

β=0.7

EnKF

Example Hybrid Result (Wang et al 2007)

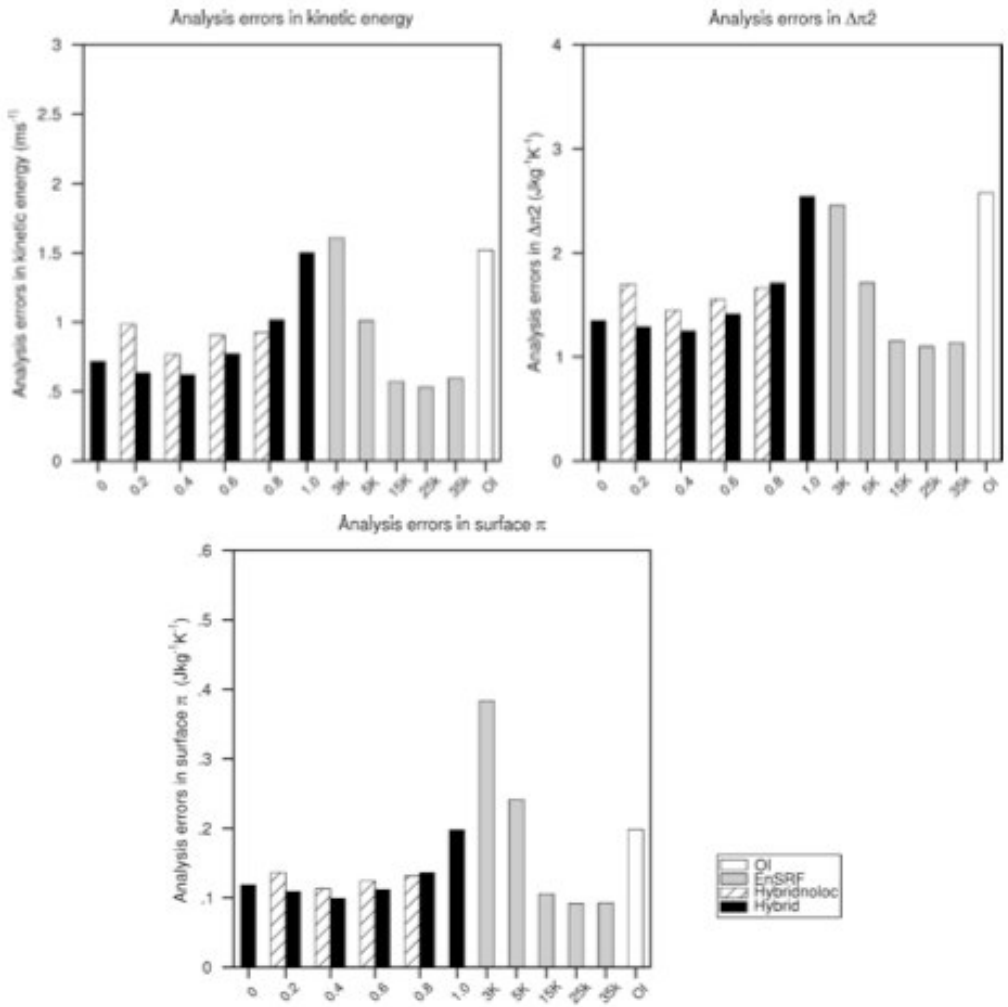


FIG. 5. Root-mean-square analysis error for kinetic-energy norm, upper-layer thickness $\Delta\pi_2$ norm, and surface π norm for 50-member ensembles. The black bars are results for the hybrid ETKF–OI scheme with localization. The hatched bars are results for the hybrid ETKF-OI with no localization. The gray bars are results for the EnSRF. The white bar is for the OI experiment.

Hybrid DA Via Additional Control Variables

- Define the matrix of ensemble perturbations as

$$\delta\mathbf{X}_f = \left(\delta\mathbf{x}_{f1}, \delta\mathbf{x}_{f2}, \ldots, \delta\mathbf{x}_{fN}\right) \tag{1}$$

- Hybrid 3/4D-Var analysis increments give by

$$\delta\mathbf{x}_0 = \delta\mathbf{x}_{0d} + \delta\mathbf{X}_f \bullet \mathbf{a} = \mathbf{U}\mathbf{v} + \delta\mathbf{X}_f \bullet \mathbf{a} \tag{2}$$

- Note flow-dependence $\delta\mathbf{X}_f$ constrained by a new set of control variables

$$\mathbf{a}^{\mathrm{T}} = \left(\alpha_1, \alpha_2, \ldots, \alpha_N\right) \tag{3}$$

- Could alternatively define the hybrid in control variable space, e.g.

$$\delta\psi_0 = \delta\psi_{0d} + \delta\psi_f \bullet \mathbf{a} \qquad \delta\chi_{u0} = \delta\chi_{u0d} + \delta\chi_{uf} \bullet \mathbf{a} \tag{4}$$

- (4) better than (2) more when balance well known.

Cost Functions and Variance Conservation

- Flow-dependence is constrained by an additional cost-function $J_{\alpha,}$ i.e.

$$J = \frac{W_b}{2}\delta\mathbf{x}_0{}^T\mathbf{B}_o^{-1}\delta\mathbf{x}_0 + \boxed{\frac{W_\alpha}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a}} + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]^T\mathbf{R}_i^{-1}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]$$

- Define empirical alpha covariance matrix $\mathbf{A} = \sigma_\alpha^2\mathbf{A}_c$

- W_b and W_α are weights defined to conserve forecast error.

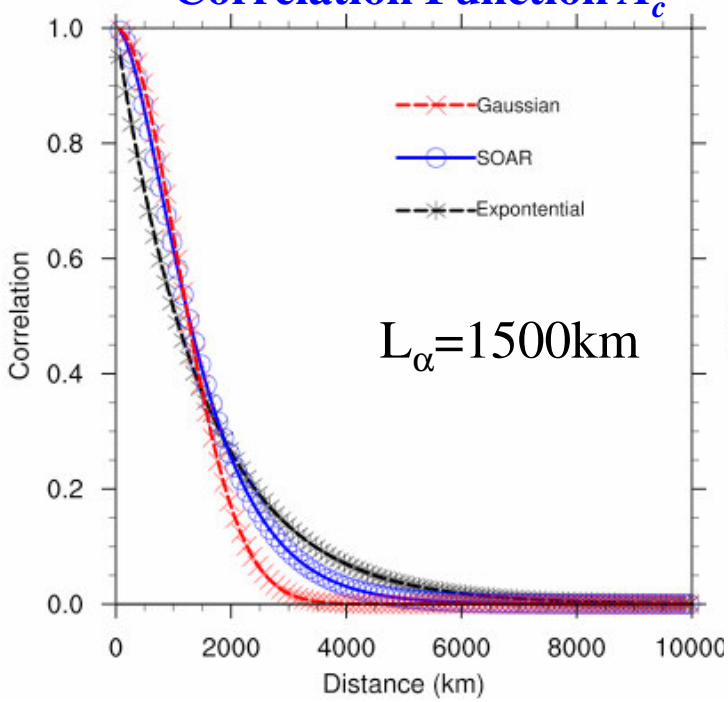
- Lorenc (2003)-type hybrid conservation: $$\frac{1}{\sqrt{W_b}} + \frac{1}{\sqrt{\left(W_\alpha / \sigma_\alpha^2\right)}} = 1$$

- Forecast error variance conservation: $$\frac{1}{W_b} + \frac{1}{W_\alpha / \sigma_\alpha^2} = 1$$

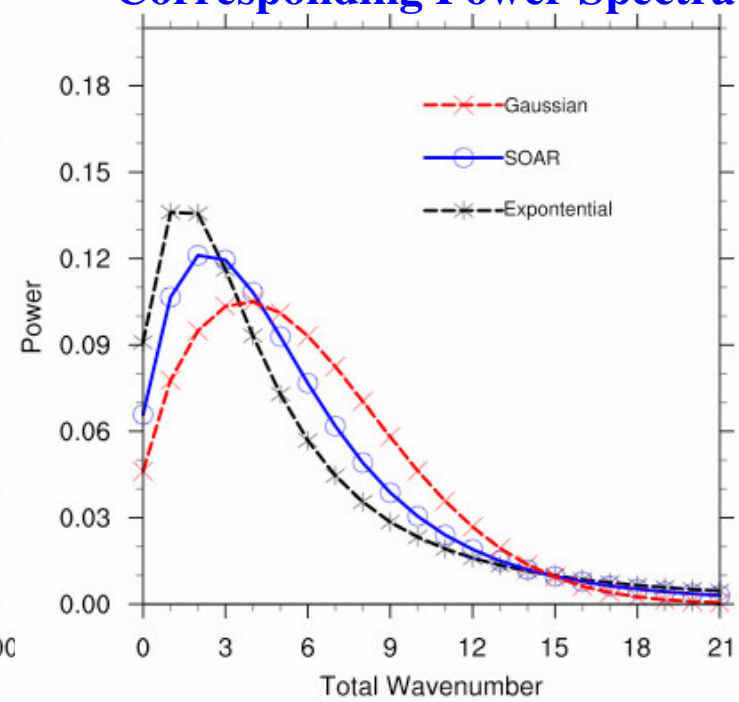
Example Application of ACV in Global WRF-Var

- Alpha correlation A_c is empirical function (e.g. Gaussian) with prescribed scale L_α.
- Lorenc (2003) suggest equivalence between A_c and covariance localization.
- Test ACV approach in global WRF-Var (horiz. correlations in spectral space)

Correlation Function A_c



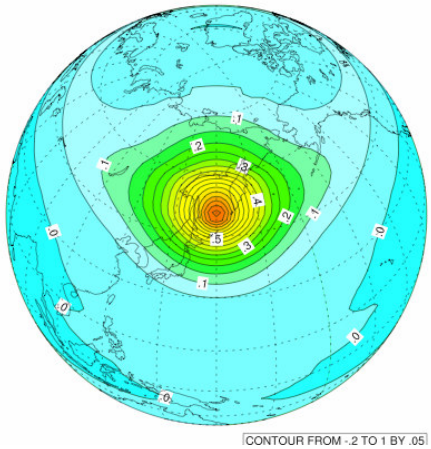
Corresponding Power Spectra



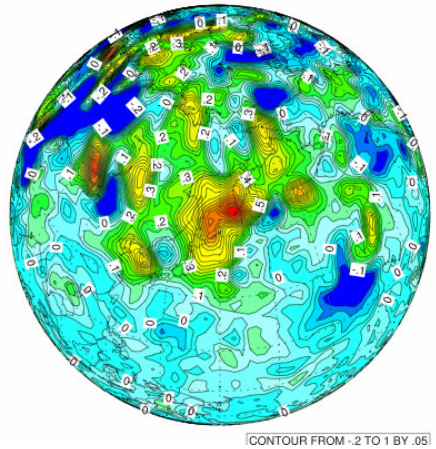
Single Observation Test - Alpha CV

- Specify single T observation (O-B, σ_o=1K) at 50N, 150E, 500hPa.
- Example: Flow-Dependence given by 1 member of KMA's EPS.

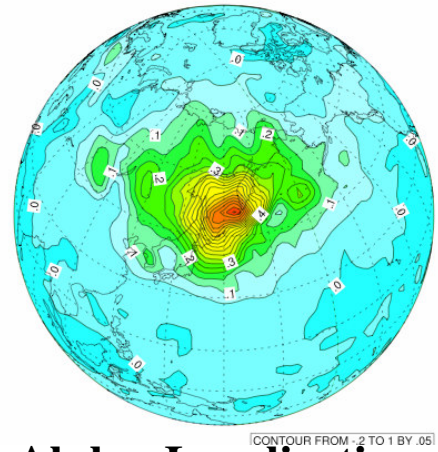
T' increment



No Alpha

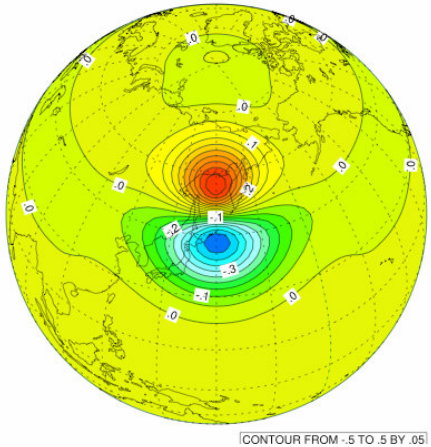


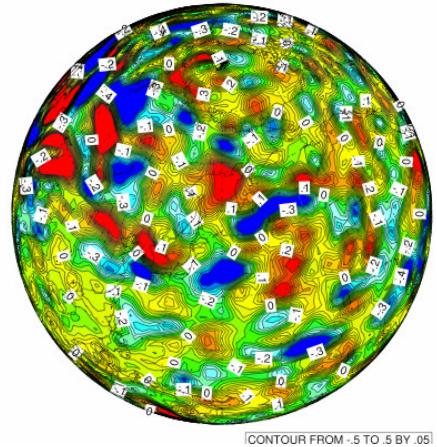
Alpha, No Localization

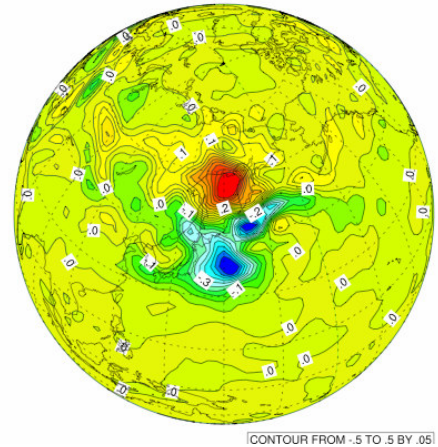


Alpha, Localization (1500km Gaussian)

u' increment



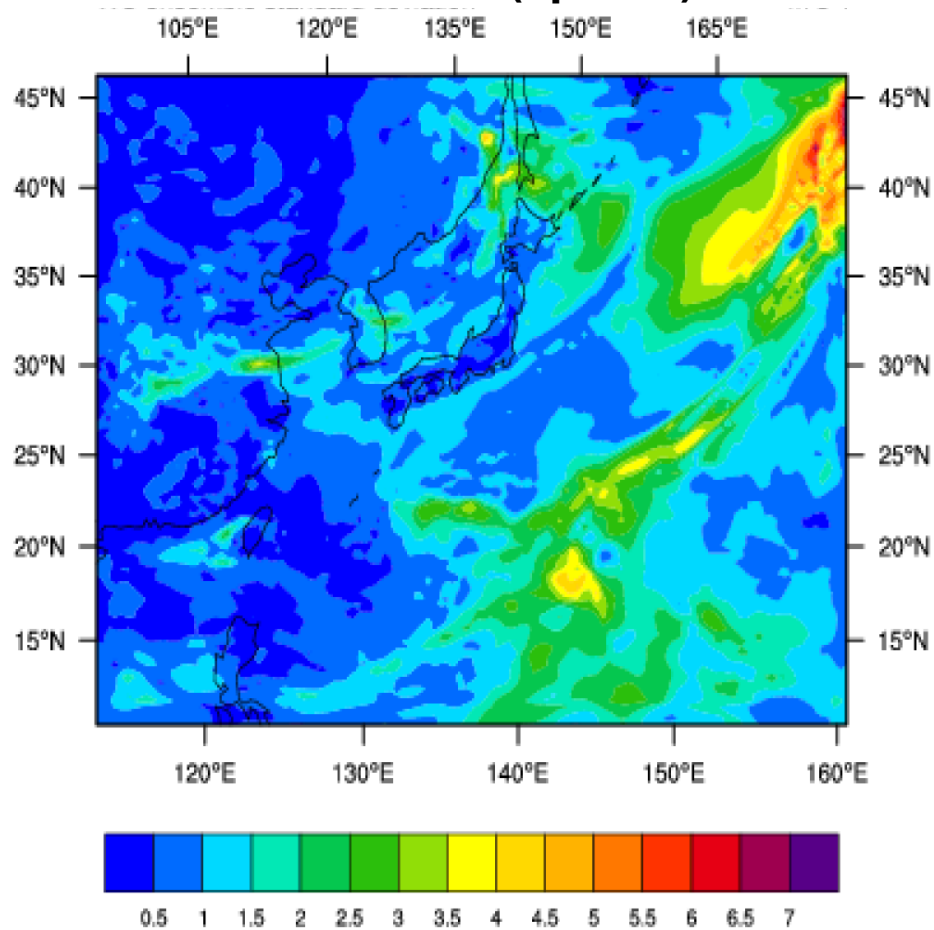




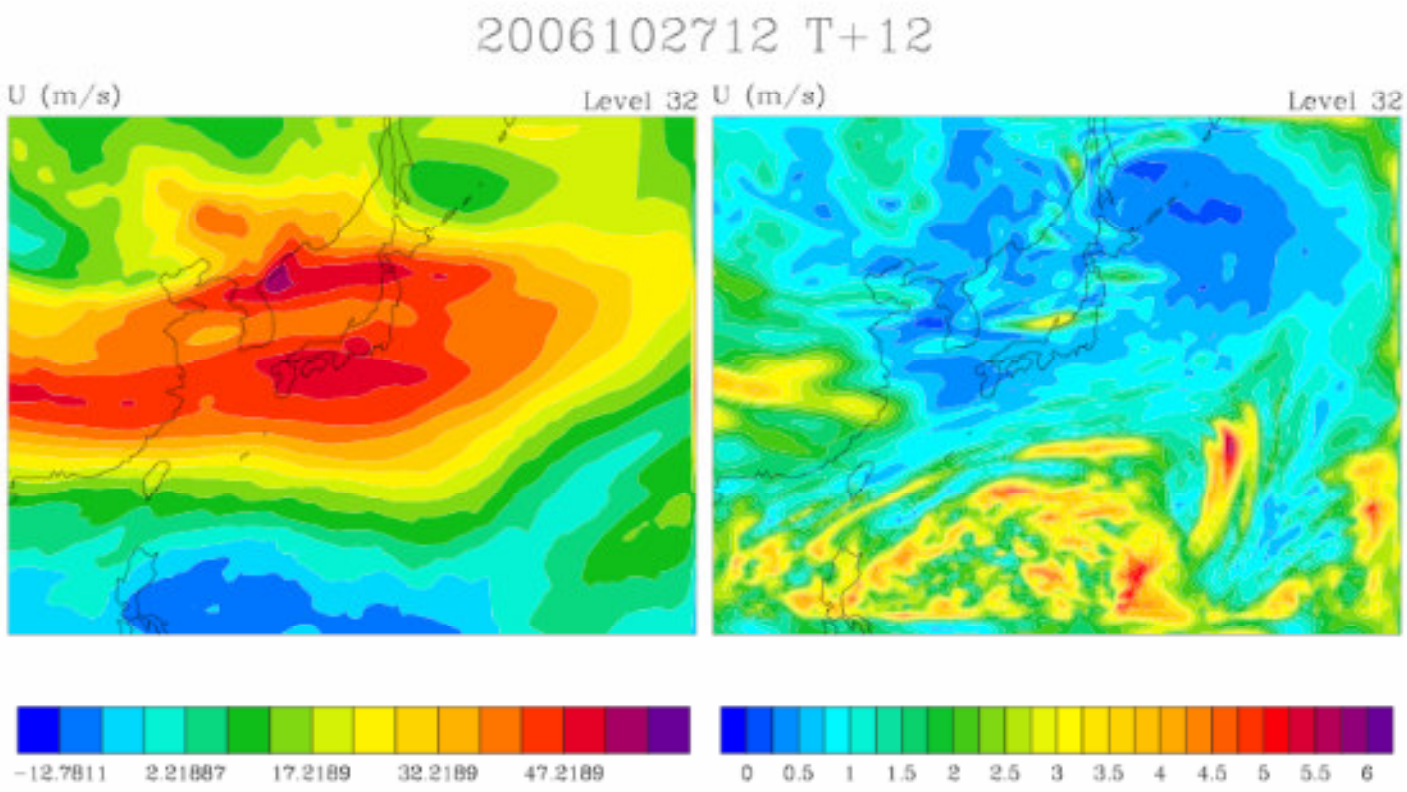
Joint Mesoscale Ensemble (US AFWA)

- 10 Ensemble members.
- Model error via WRF physics perturbations.
- 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 for verification

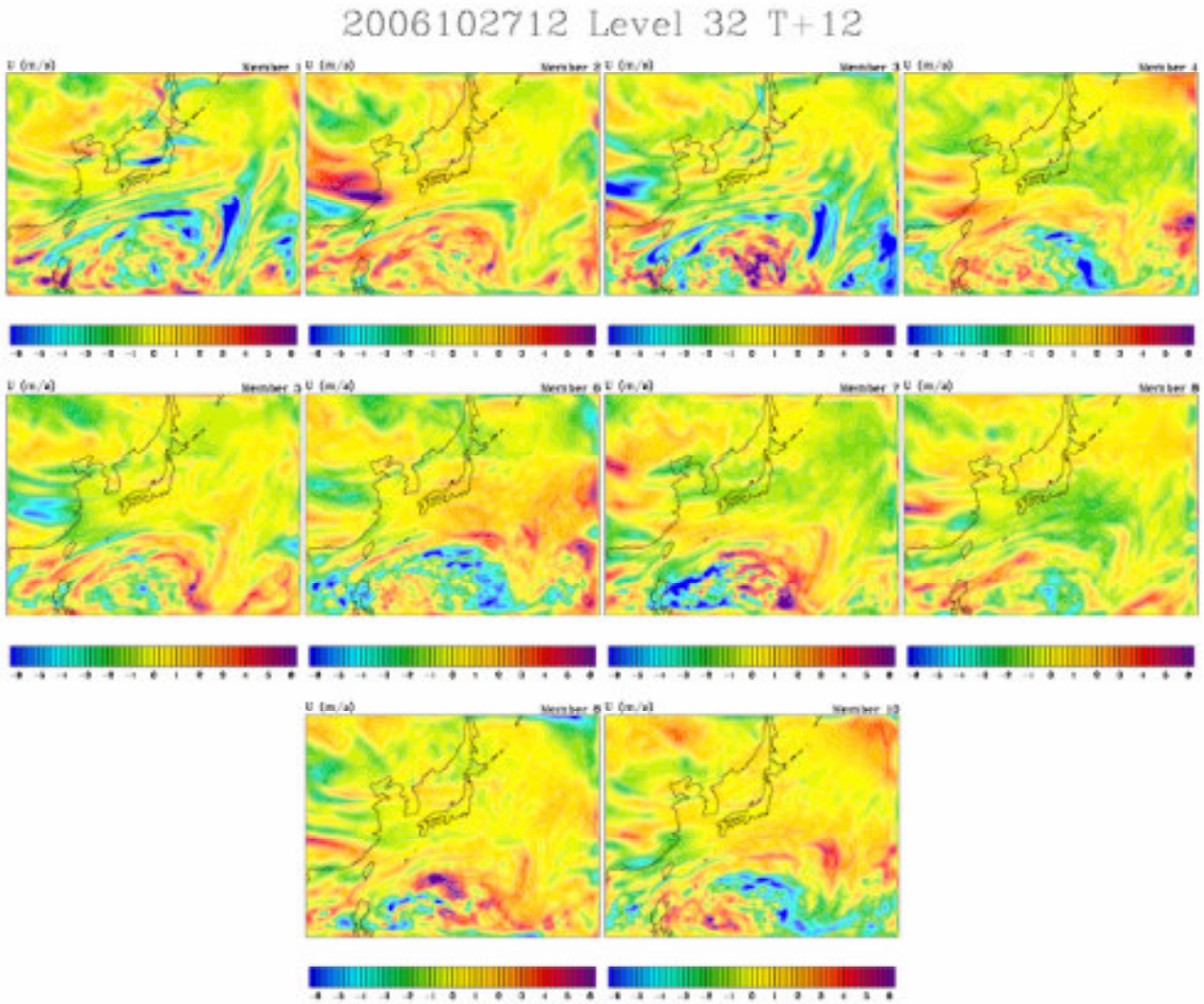
10-m wind speed ensemble standard deviation (spread)



JME Test Case: u (level 32) mean and std. dev. (da_run_ensmean.ksh)

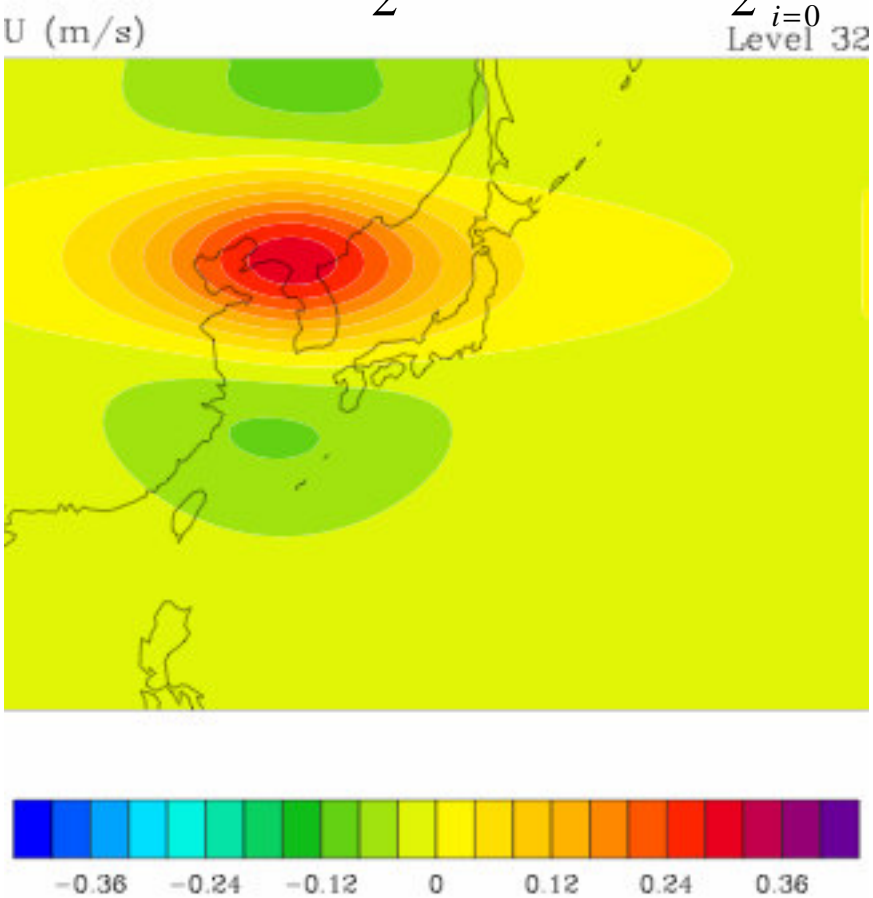


JME Test Case: u (level 32) perturbations (da_run_ens_ep.ksh)



WRF-Var No-Hybrid Single ob test (250hPa u, O-B=1m/s, sigma_o=3.3m/s)

$$J = \frac{W_b}{2}\delta\mathbf{x}_0{}^T\mathbf{B}_o^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]^T\mathbf{R}_i^{-1}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]$$

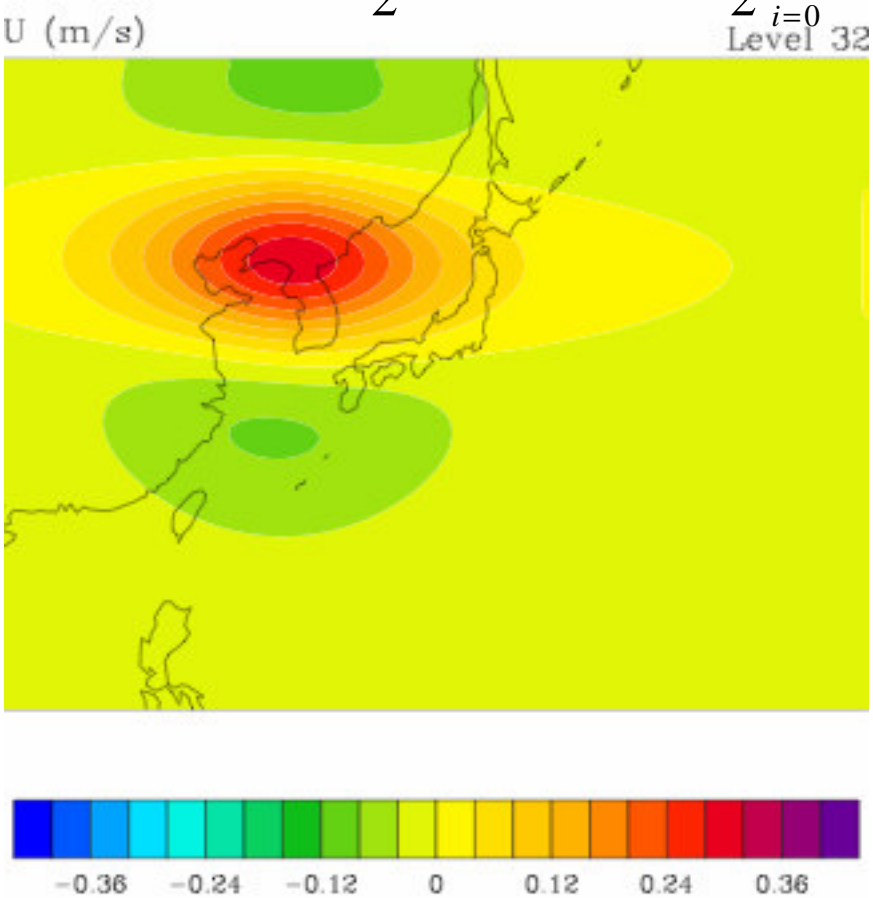


W_b=1

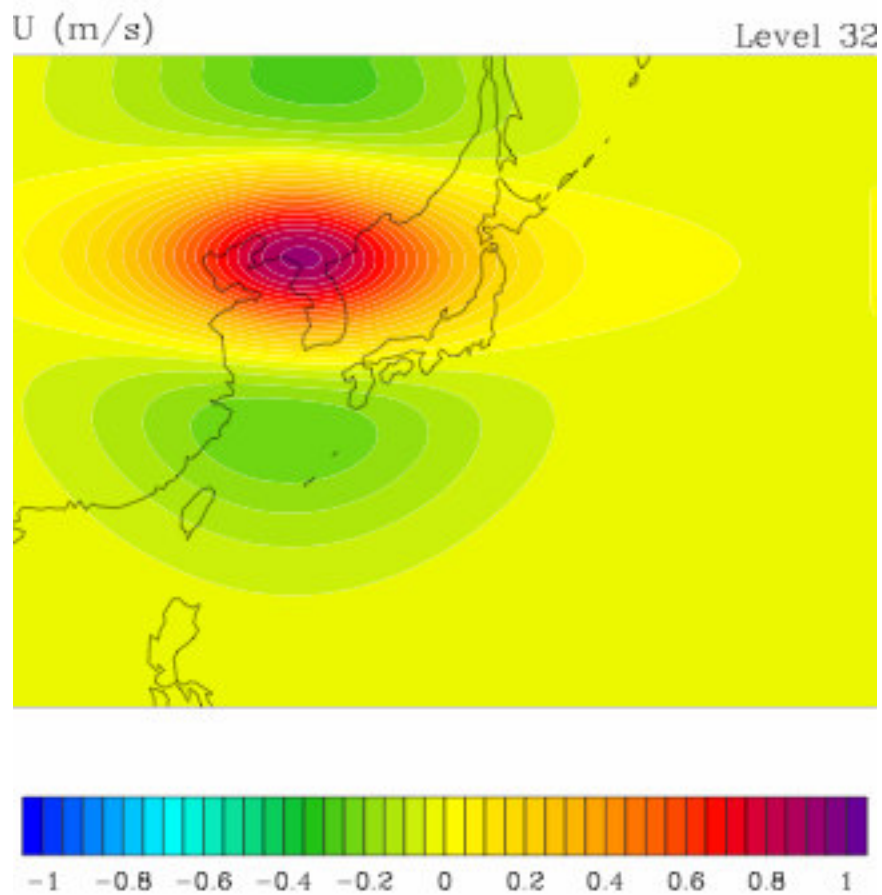
W_b=0?

WRF-Var No-Hybrid Single ob test (250hPa u, O-B=1m/s, sigma_o=3.3m/s)

$$J = \frac{W_b}{2}\delta\mathbf{x}_0{}^T\mathbf{B}_o^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]^T\mathbf{R}_i^{-1}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]$$



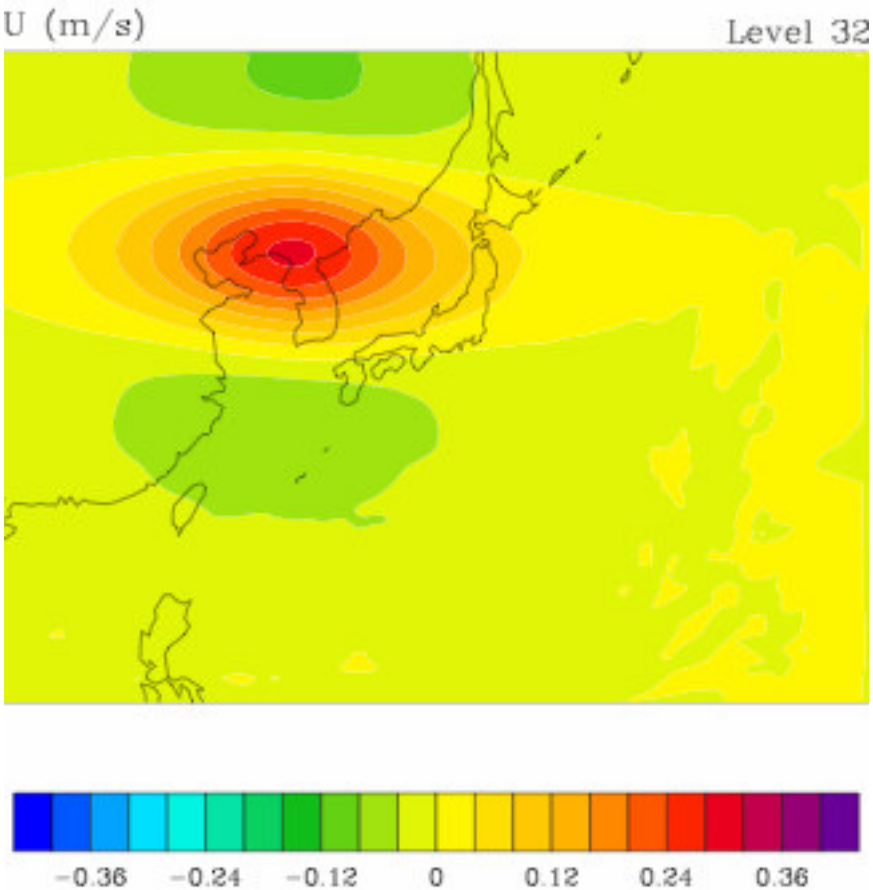
W_b=1



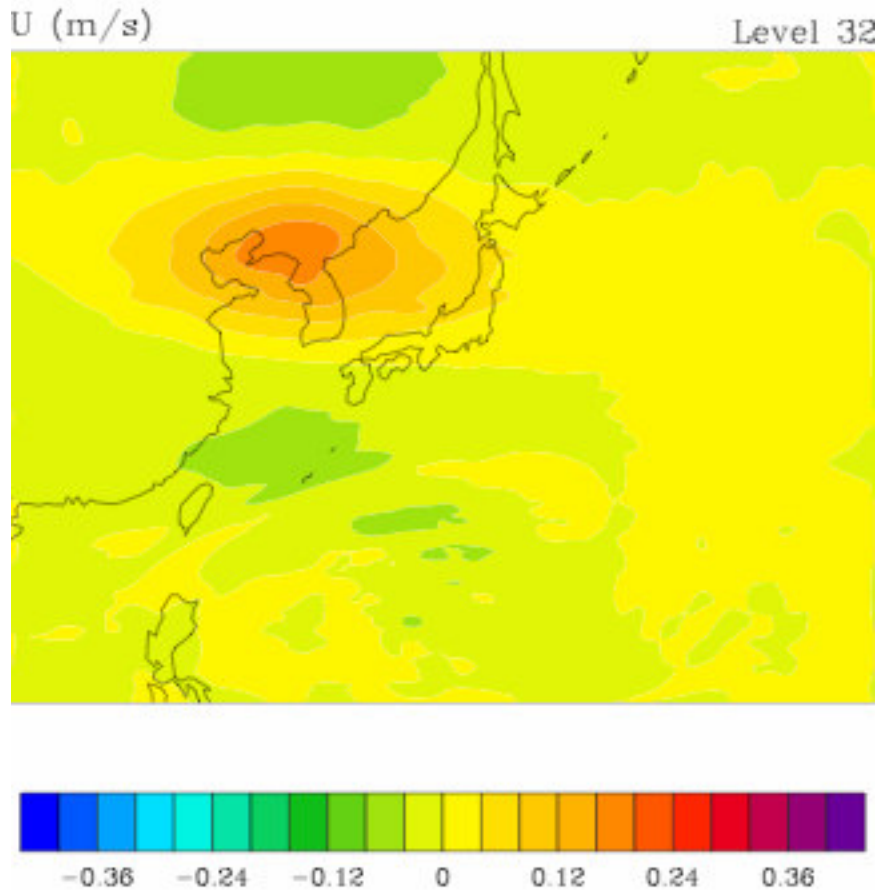
W_b=0

WRF-Var Hybrid Single ob test (250hPa u, O-B=1m/s, sigma_o=3.3m/s)

$$J = \frac{W_b}{2}\delta\mathbf{x}_0{}^T\mathbf{B}_o^{-1}\delta\mathbf{x}_0 + \frac{W_\alpha}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]^T\mathbf{R}_i^{-1}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]$$

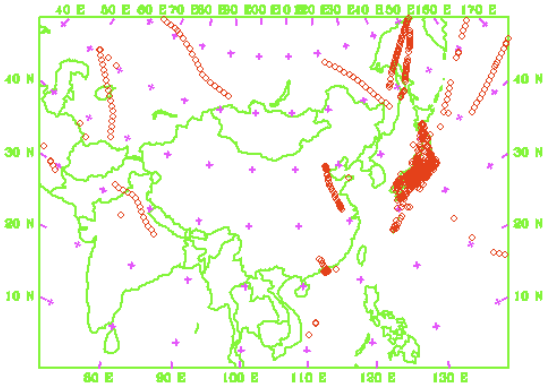


W_e=10 , W_b=1.11

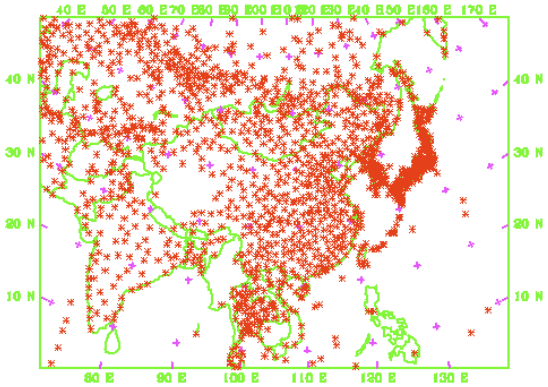


W_e=2 , W_b=2

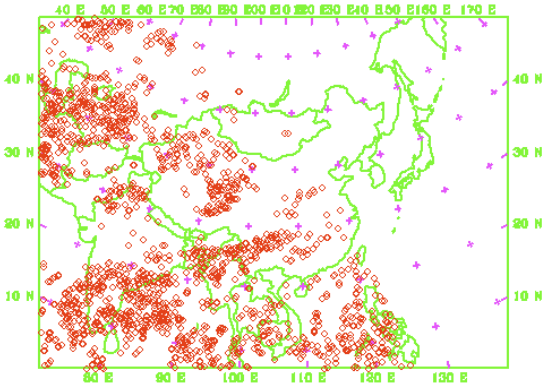
JME Test Case: 2006102800



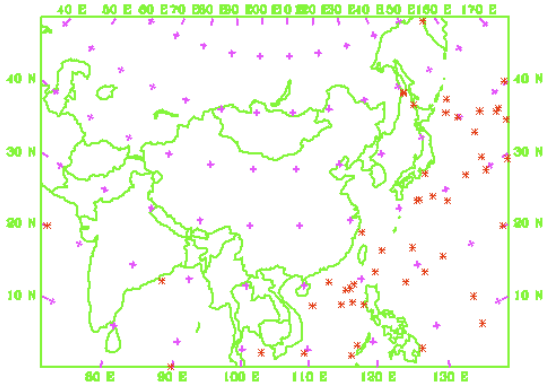
◇ 930 AIREP 2006-10 [27_21:00,28_03:00]



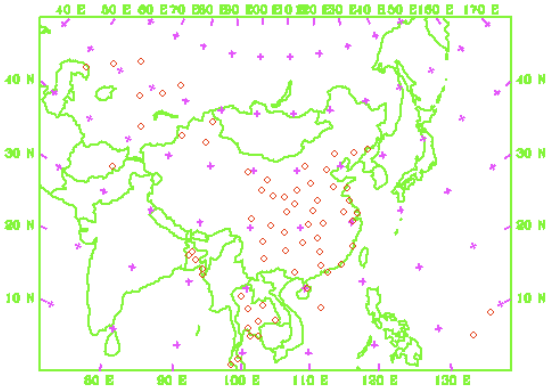
* 1634 METAR 2006-10 [27_21:00,28_03:00]



◇ 1788 SATOB 2006-10 [27_21:00,28_03:00]

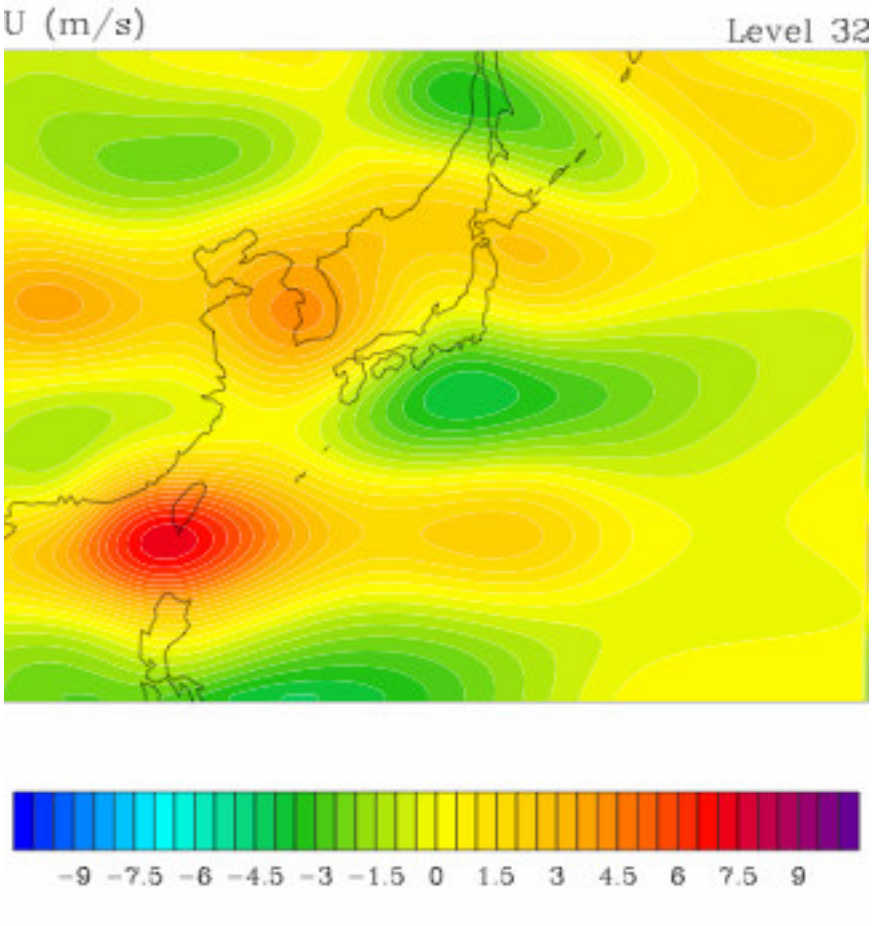


* 48 SHIPS 2006-10 [27_21:00,28_03:00]

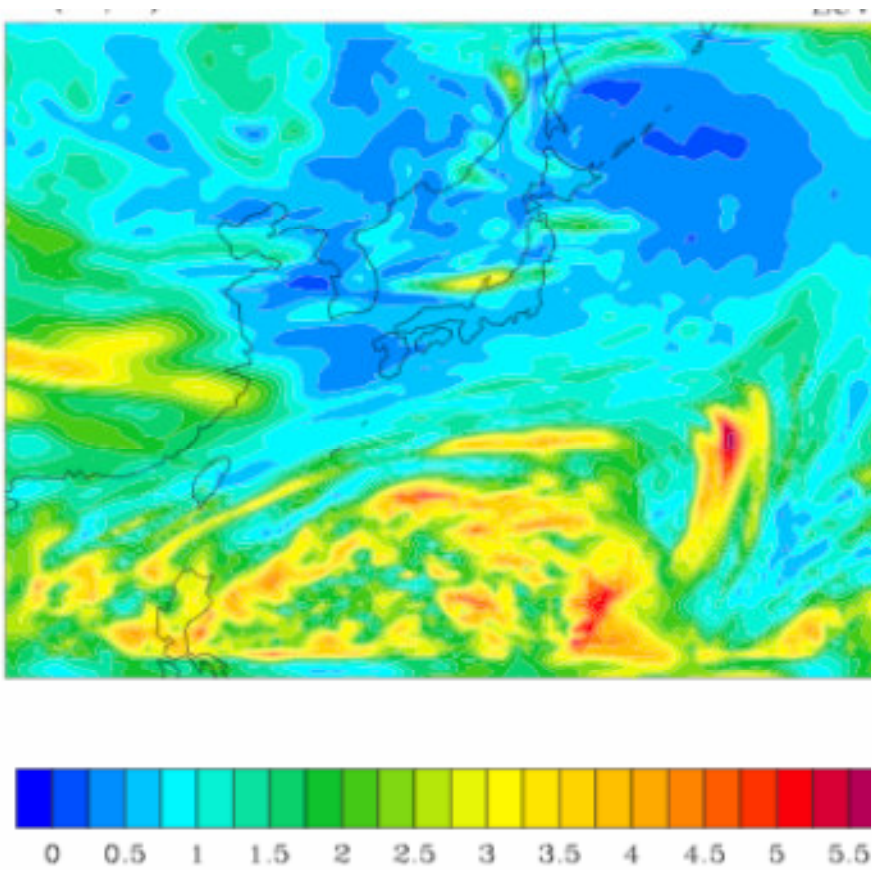


◇ 68 SOUND 2006-10 [27_21:00,28_03:00]

WRF-Var Full Assimilation



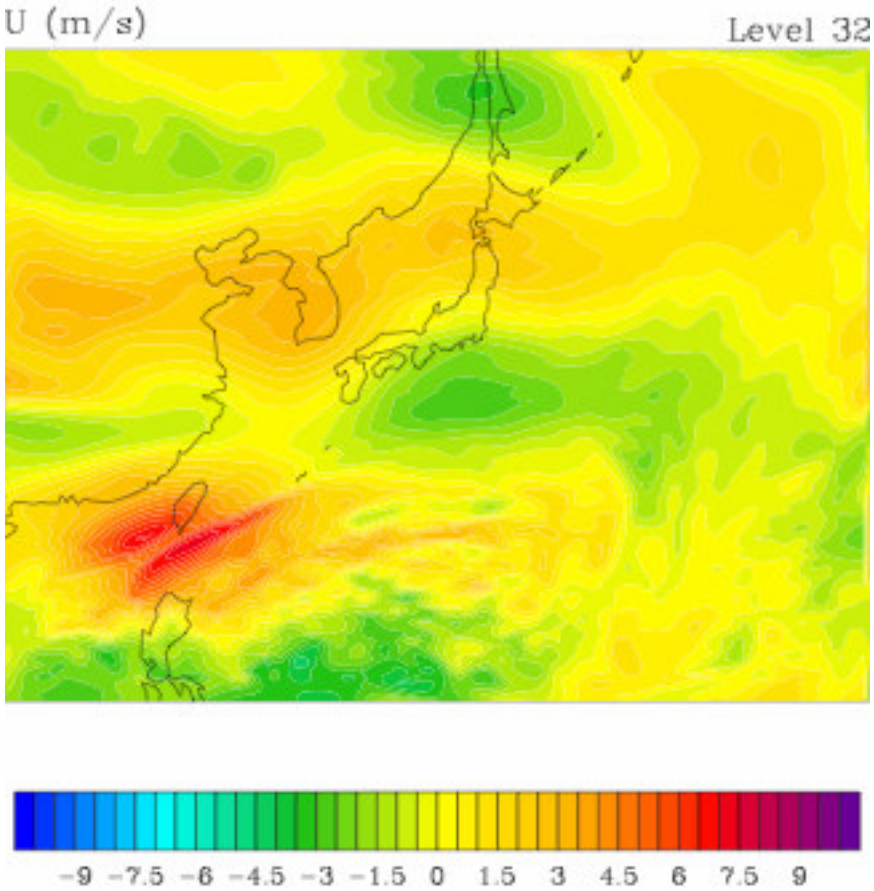
U250 Analysis Increment



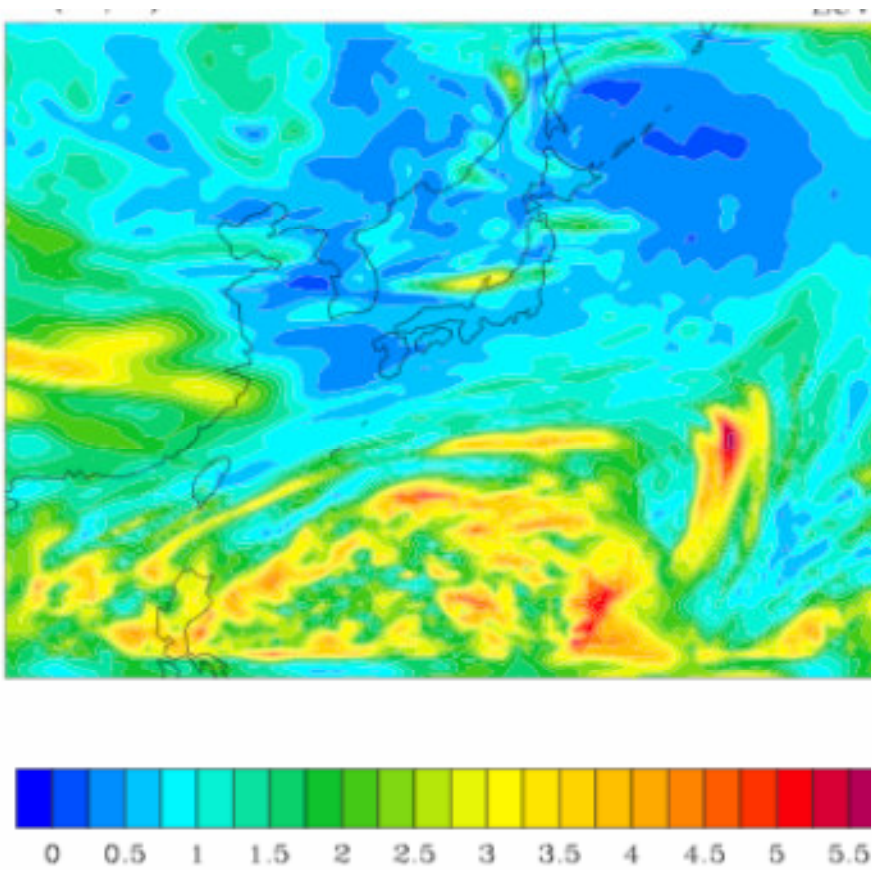
U250 Ensemble Spread

Control (no hybrid)

WRF-Var Full Assimilation



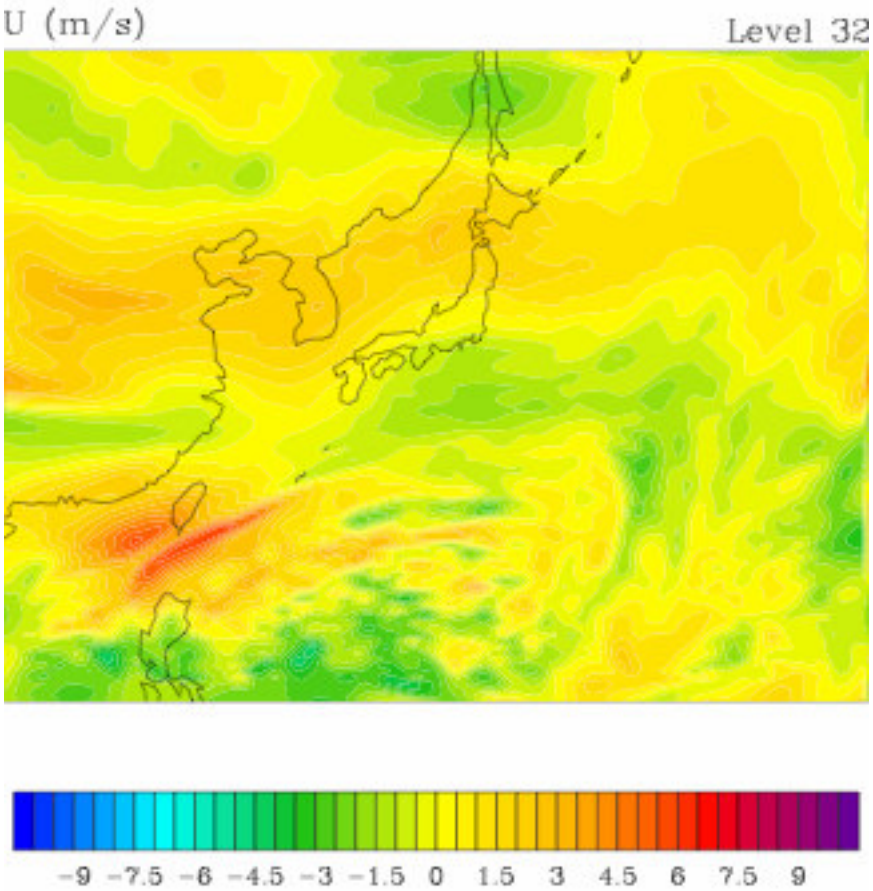
U250 Analysis Increment



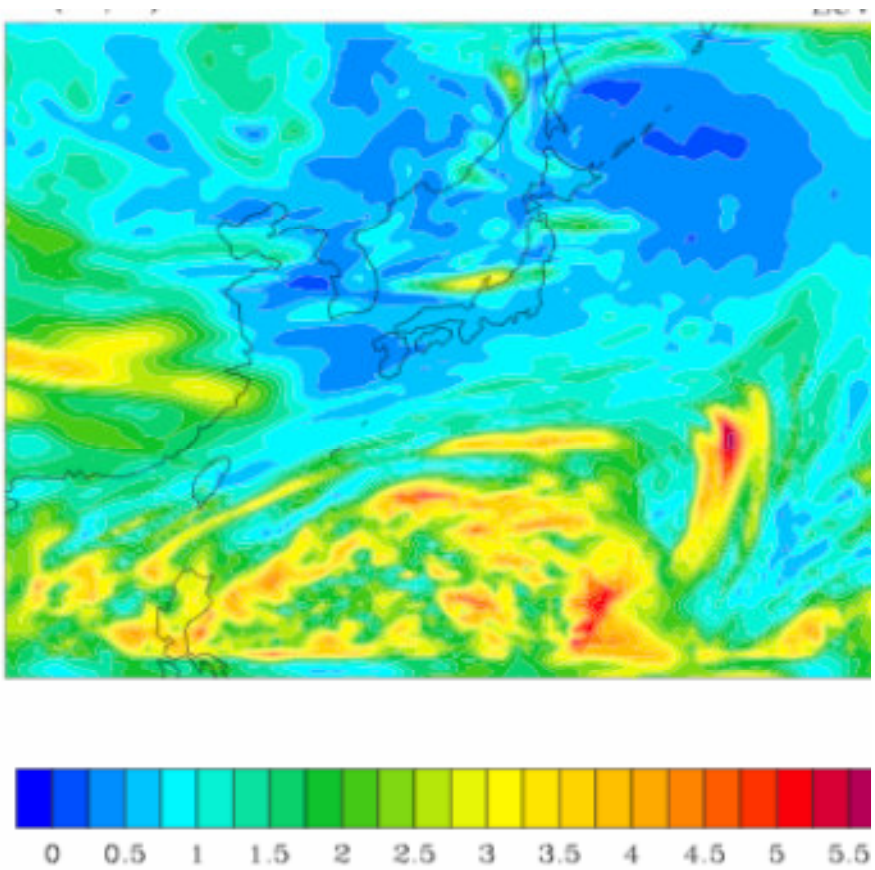
U250 Ensemble Spread

$W_e=2$, $W_b=2$

WRF-Var Full Assimilation



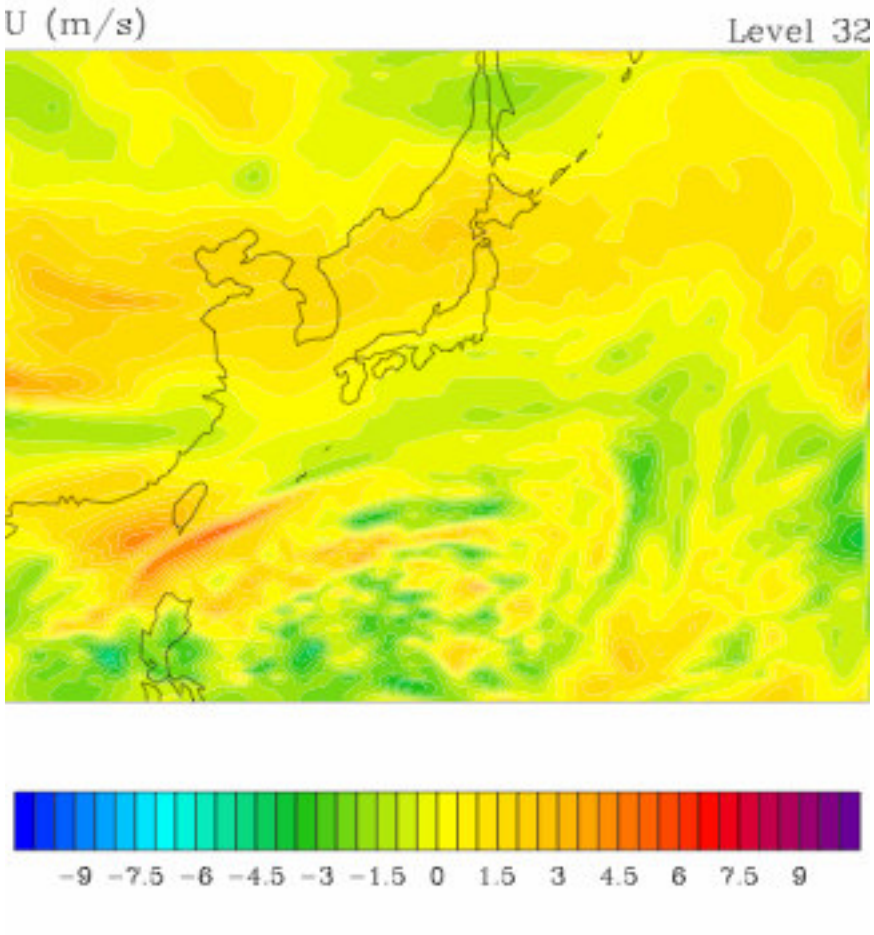
U250 Analysis Increment



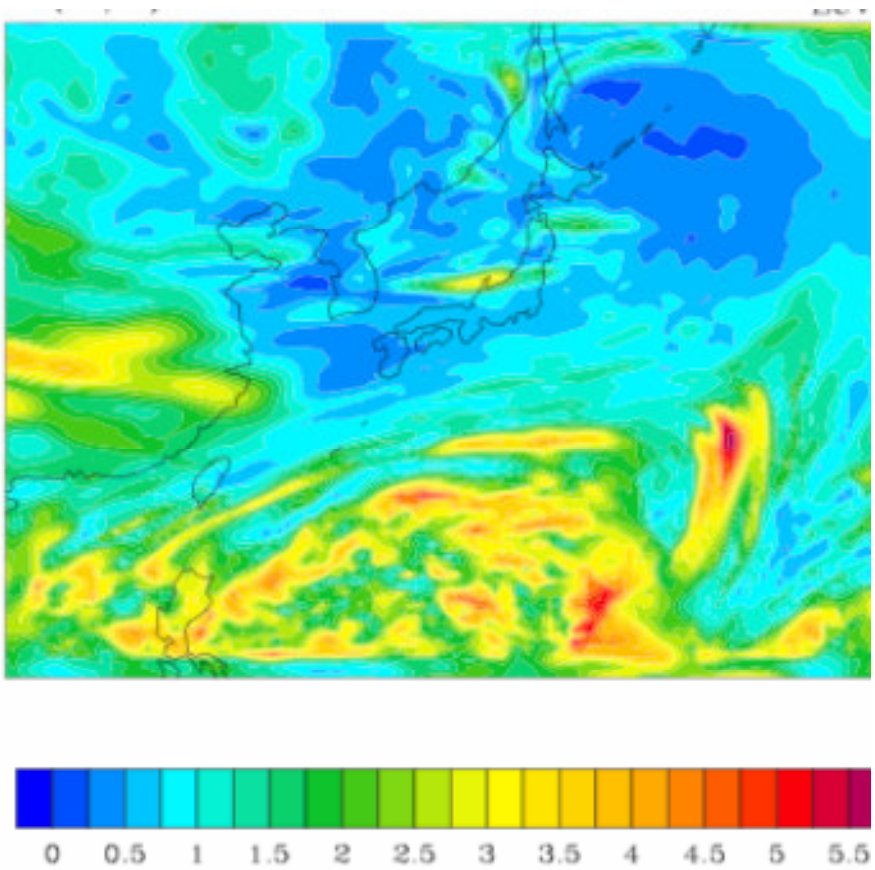
U250 Ensemble Spread

W_e=1.25 , W_b=5

WRF-Var Full Assimilation



U250 Analysis Increment



U250 Ensemble Spread

W_e=1.1, W_b=11

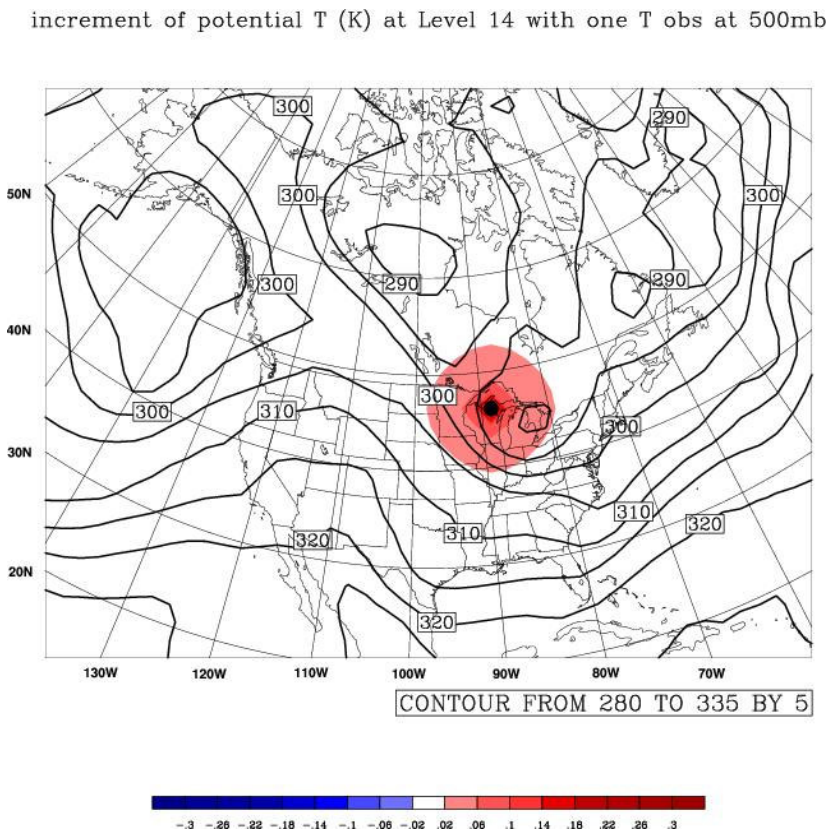
Domain For Initial WRF/ETKF Tests

- Low-resolution (200km) CONUS domain.
- Same domain used for initial 3D-Var/EnKF comparison.
- January 2003 test period.
- Assimilate sondes at 12hr intervals.
- 30/50 member ETKF.
- Sampled 3D-Var covariances used to provide initial/lateral boundary perturbations.

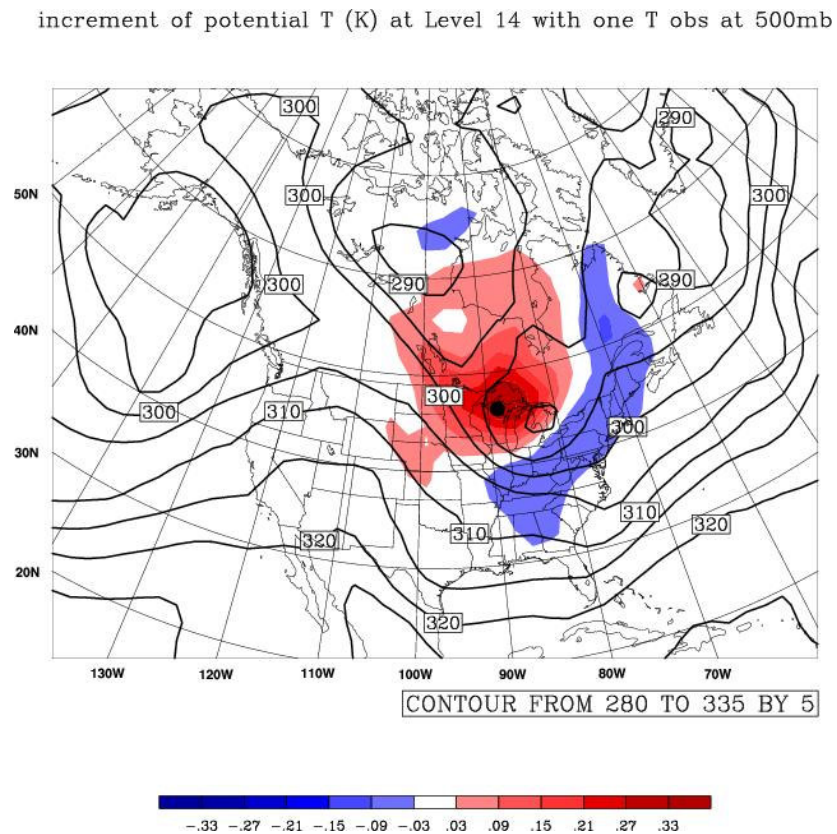
WRF Test with single observation (X. Wang)

Analysis increment

Static covariance

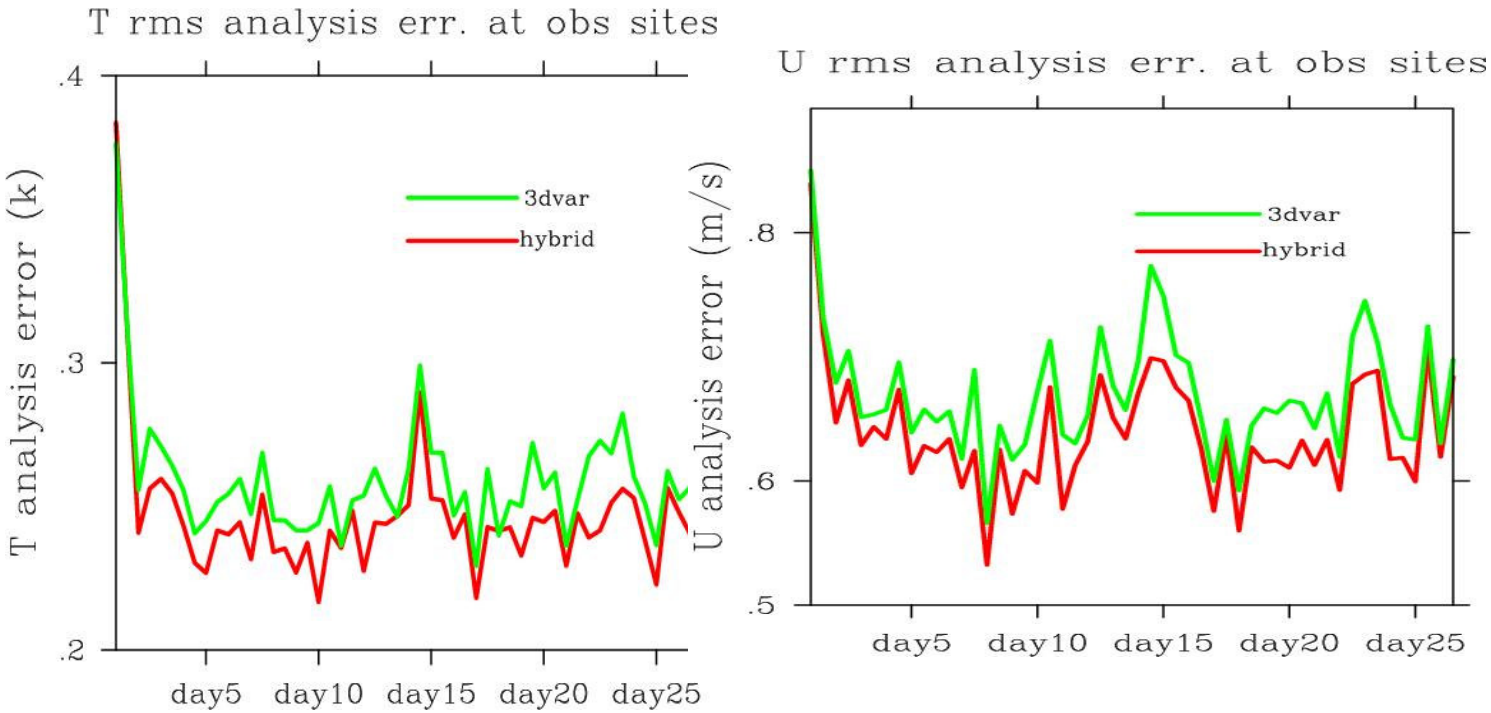


Ensemble covariance with localization



Flow-dependent ETKF ensemble covariance is successfully incorporated in WRF-Var

Hybrid 1-Month OSSE Trial: Analysis Error (Wang, Barker, Hamill and Snyder 2008a)



- Test hybrid with equal weight (0.5) on static/ETKF error covariances
- Hybrid analyses significantly better than the pure 3DVAR.
- Note yet cycling, nor tuned. Expect further improvements?

Hybrid 1-Month Real-Data (Sonde) Trial (Wang, Barker, Hamill and Snyder 2008b)

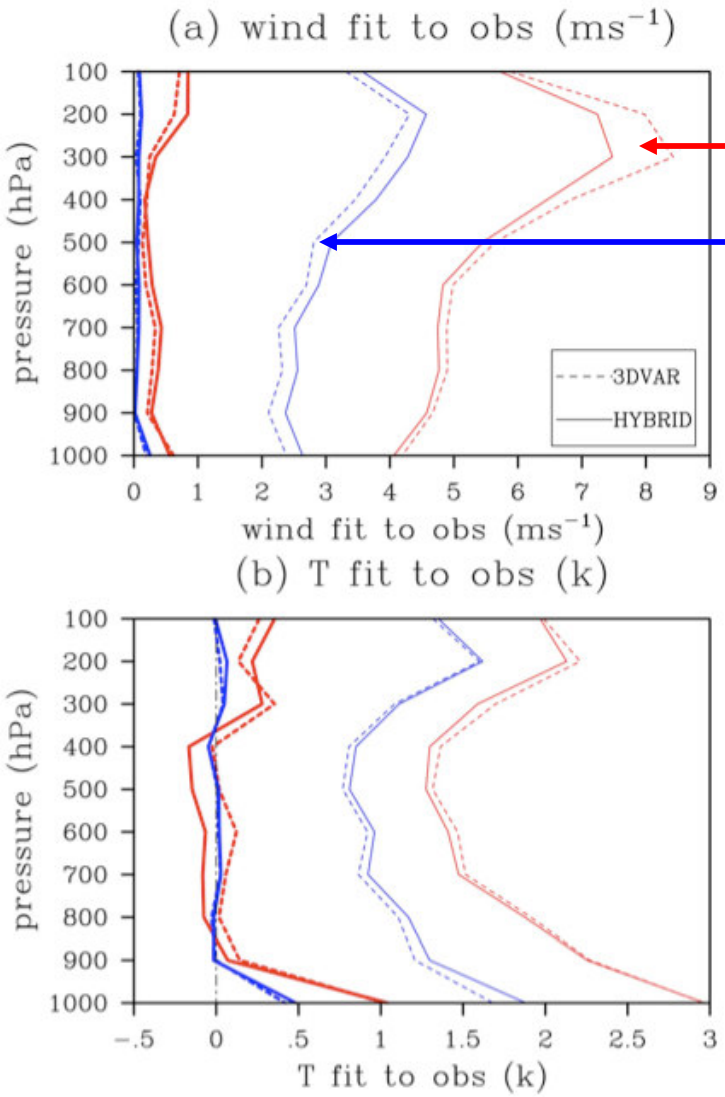
TABLE 2. Root-mean-square (rms) fit of the 12-hour wind and temperature forecasts to the radiosonde observations for the hybrid with various combinations of the weighing coefficients $1/\beta_1$ and the covariance localization scales S_e. Please see text for the definition of $1/\beta_1$ and S_e. Numbers in the parentheses indicate the percentage improvement relative to the 3DVAR with tuned static covariance. The smallest rms fits are highlighted. For $1/\beta_1 = 1.0$, experiments do not depend on S_e.

Wind (ms^{-1})	$S_e = 4242\ km$	2828 km	1414 km	707 km
$1/\beta_1 = 1.0$	6.255 (2.2)	—	—	—
0.8	5.991 (6.3)	6.001 (6.1)	6.046 (5.4)	6.186 (3.2)
0.5	5.997 (6.2)	**5.960 (6.8)**	5.998 (6.2)	6.146 (3.9)
0.2	5.964 (6.7)	6.010 (6.0)	6.045 (5.4)	6.160 (3.6)
0.0	6.457 (-1.0)	6.327 (1.0)	6.241 (2.4)	6.201 (3.0)

T (K)	$S_e = 4242\ km$	2828 km	1414 km	707 km
$1/\beta_1 = 1.0$	1.858 (0.6)	—	—	—
0.8	1.818 (2.7)	**1.813 (3.0)**	**1.813 (3.0)**	1.827 (2.2)
0.5	1.821 (2.6)	1.818 (2.7)	1.816 (2.8)	1.823 (2.5)
0.2	1.851 (1.0)	1.845 (1.3)	1.829 (2.1)	1.841 (1.5)
0.0	2.034 (-8.8)	1.979 (-5.9)	1.921 (-2.8)	1.886 (-0.9)

- Hybrid has larger impact on wind than temperature field.
- Optimal localization radius larger than OSSE (W2008a).
- Flow-dependence effect larger than skill of ensemble mean over control (2.2% for wind, 0.6% for temperature).

Hybrid 1-Month Real-Data (Sonde) Trial (Wang, Barker, Hamill and Snyder 2008b)

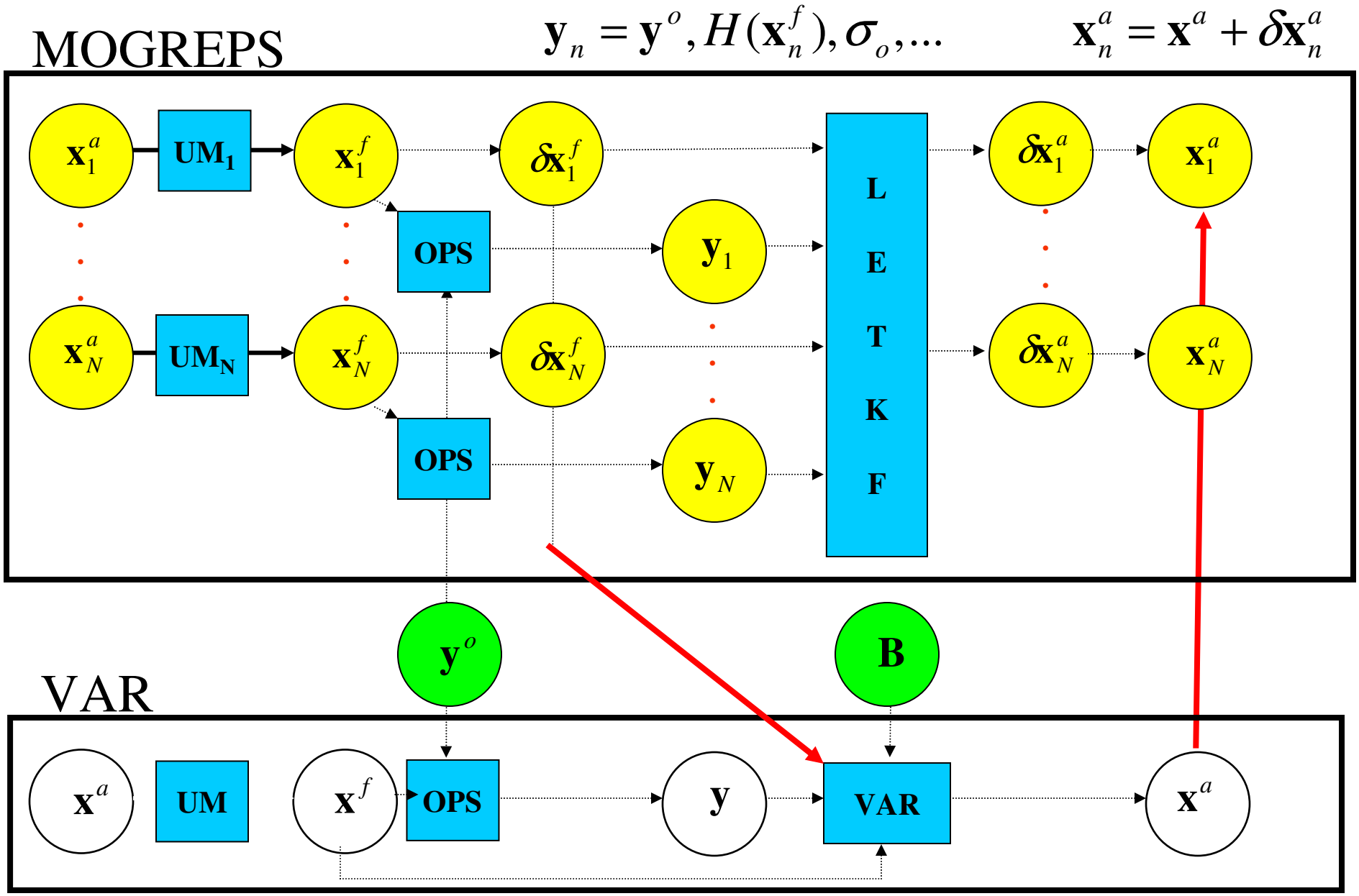


RMS 12hr Fcst Fit to Sonde

RMS Analysis Fit to Sonde

- 3D-Var analysis fits observations more closely.
- Hybrid forecast fits observations more closely.

Met. Office 4D-Var/LETKF Hybrid System



Benefits Of Hybrid Var/Ensemble DA

Benefits for Var:

- Introduces flow-dependent initial PDF in 3/4D-Var.
- Explicit coupling between moisture/temp/wind fields (tropics, high-res).
- Easily incorporated in Var framework.
- Relatively cheap (if properly preconditioned and compressed).
- Can't do worse – can switch off ensemble covariances if detrimental.

Benefits over ensemble filters:

- Hybrids more robust for small ensemble sizes and large model error.
- Cost does not scale with observations.
- Can couple with nonlinear QC (serial filter can't do that by itself).
- Ensemble used for covariance modeling only - can still run high-res mean.

Hybrid Var/Ensemble DA References

- Barker (1999): Demonstration with 1 bred-mode + UKMO 3D-Var.
- Hamill and Snyder (2000): Hybrid in Ensemble Framework.
- Etherton and Bishop (2004): QG model.
- Buehner (2005): Additional control variable/3D-Var: Small impact.
- Wang, Hamill, Whitaker and Bishop (2007): Compare EnSRF/Hybrid/OI.
- Wang, Barker, Hamill and Snyder (2008a, b): WRF AlphaCV+3D-Var.
- Barker, Clayton, and Lorenc (2009): Global 4D-Var/LETKF