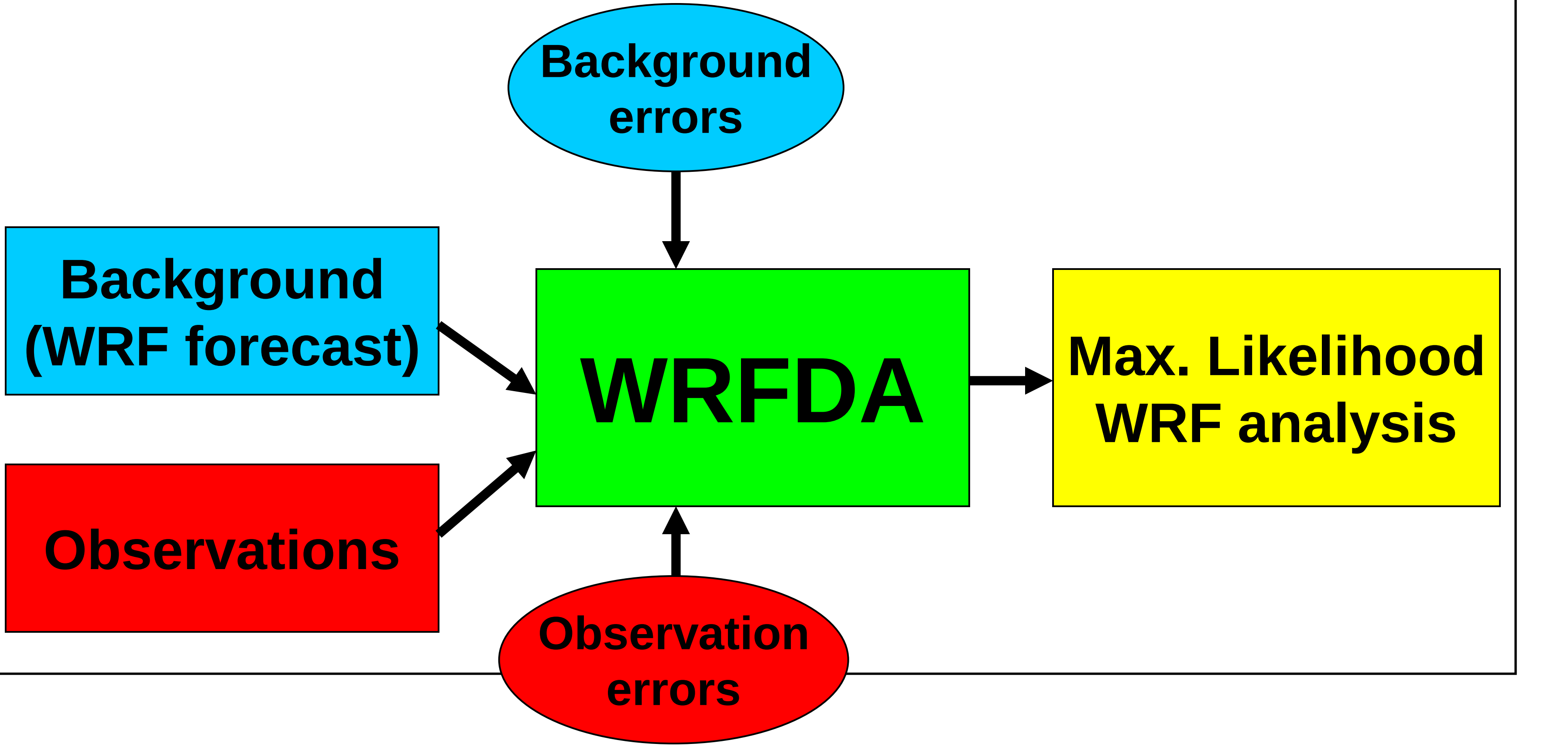
WRF HYBRID VARIATIONAL ENSEMBLE DATA ASSIMILATION WITH INITIAL CONDITION AND MODEL PHYSICS UNCERTAINTY

Dr. Luke Peffers
Florida State University

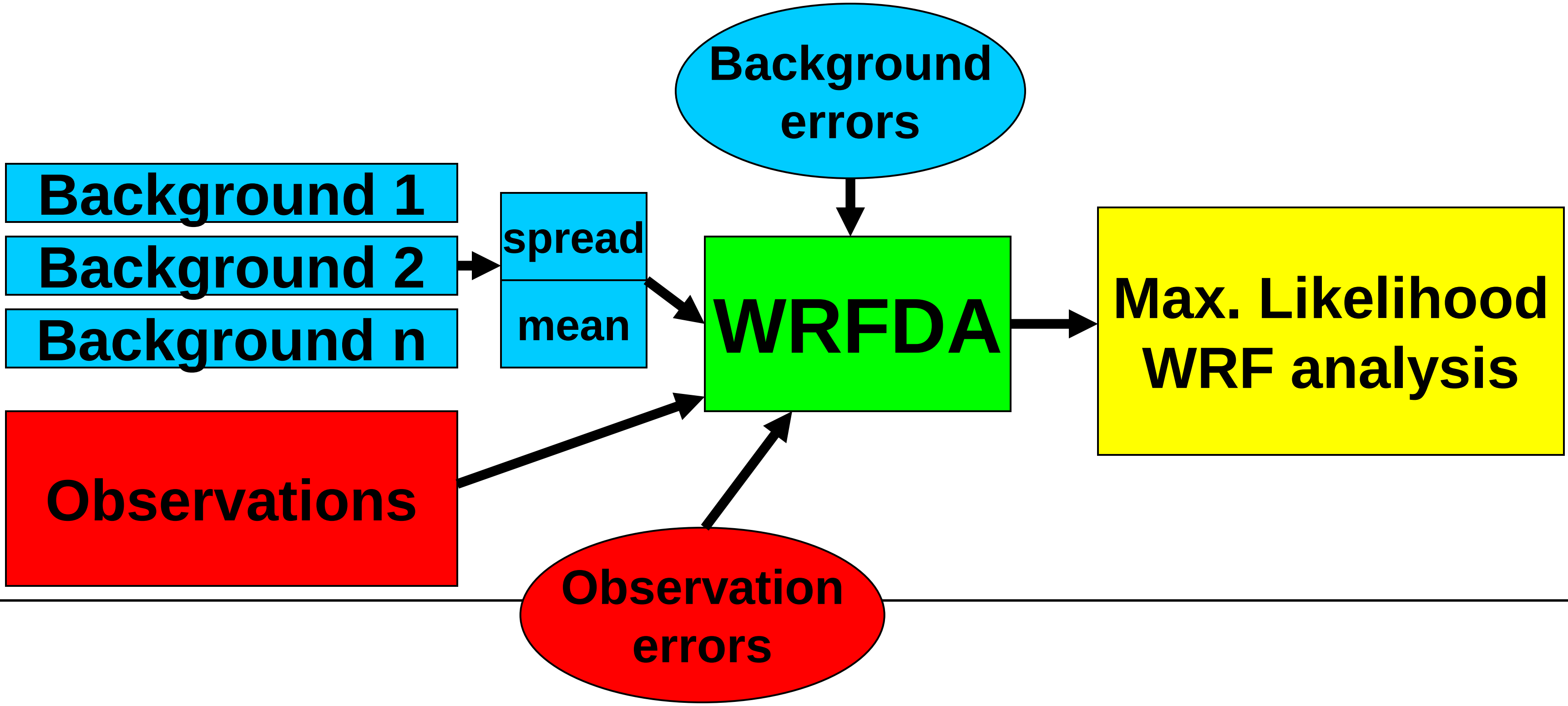
Outline

- What is that?
- How do you do it?
- How well does it work?
- How can it be improved?

WRF Variational Data Assimilation
Out-of-the-box WRF 3DVAR



WRF Hybrid Ensemble Variational Data Assimilation (Flow-dependent + 3DVAR errors)



The Hybrid Cost Function

- Wang et al. 2008ab

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

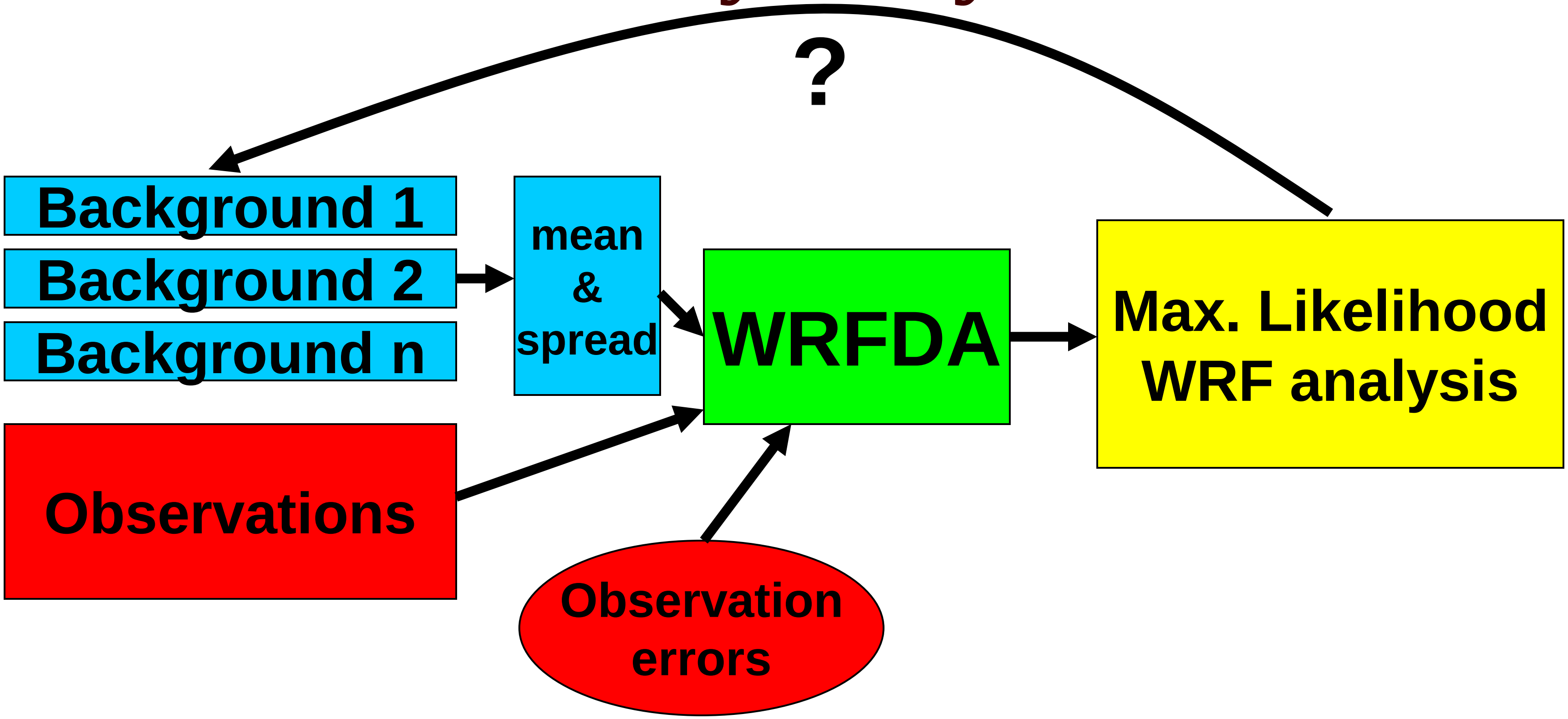
$$\mathbf{x}' = \mathbf{x}_1' + \sum(\alpha_k \circ \mathbf{x}_k^e)$$ Hybrid analysis increment

$\mathbf{x}_1'$ 3DVAR analysis increment

$\mathbf{x}^e$ Ensemble perturbations

α extended control variables

But how do you cycle this?



Ensemble generation technique

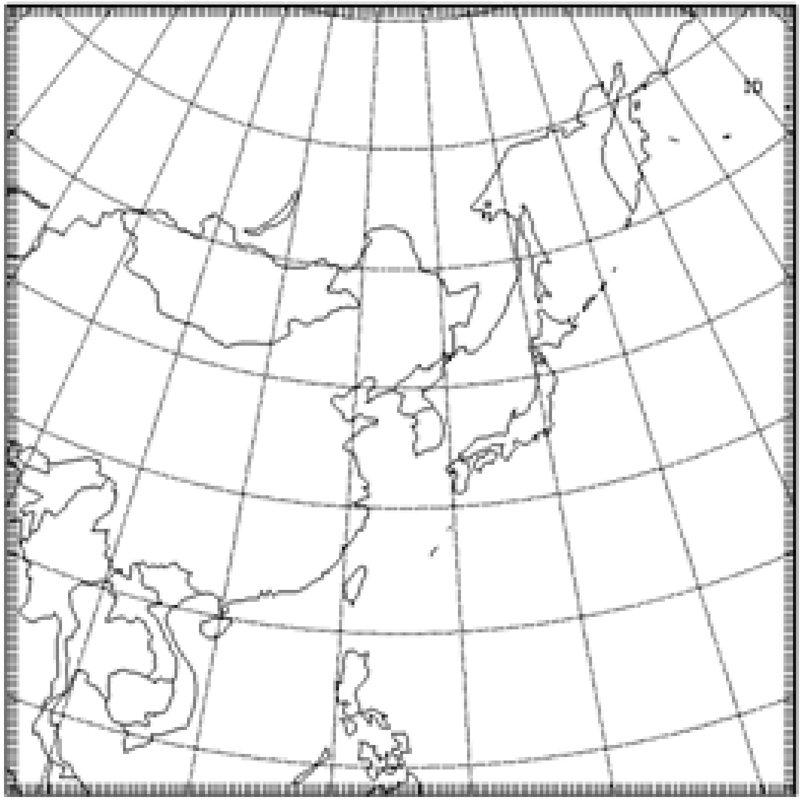
- Ensemble **transform** Kalman filter (E**T**KF)

$$\mathbf{P}^a = \Pi \mathbf{X}^f \mathbf{T}\mathbf{T}^{\mathrm{T}} \mathbf{X}^{f\mathrm{T}} \qquad \mathbf{P} \approx \mathbf{P}_e = \overline{(\mathbf{x}^f - \overline{\mathbf{x}^f})(\mathbf{x}^f - \overline{\mathbf{x}^f})^{\mathrm{T}}}$$

- $\mathbf{X}^f$ = Matrix w/columns of ensemble perturbations
- The scalar factor Π attempts to reduce the systematic underestimation of $\mathbf{P}^a$
- Transforms forecast perturbations into analysis perturbations
- Add these perturbations to ensemble mean

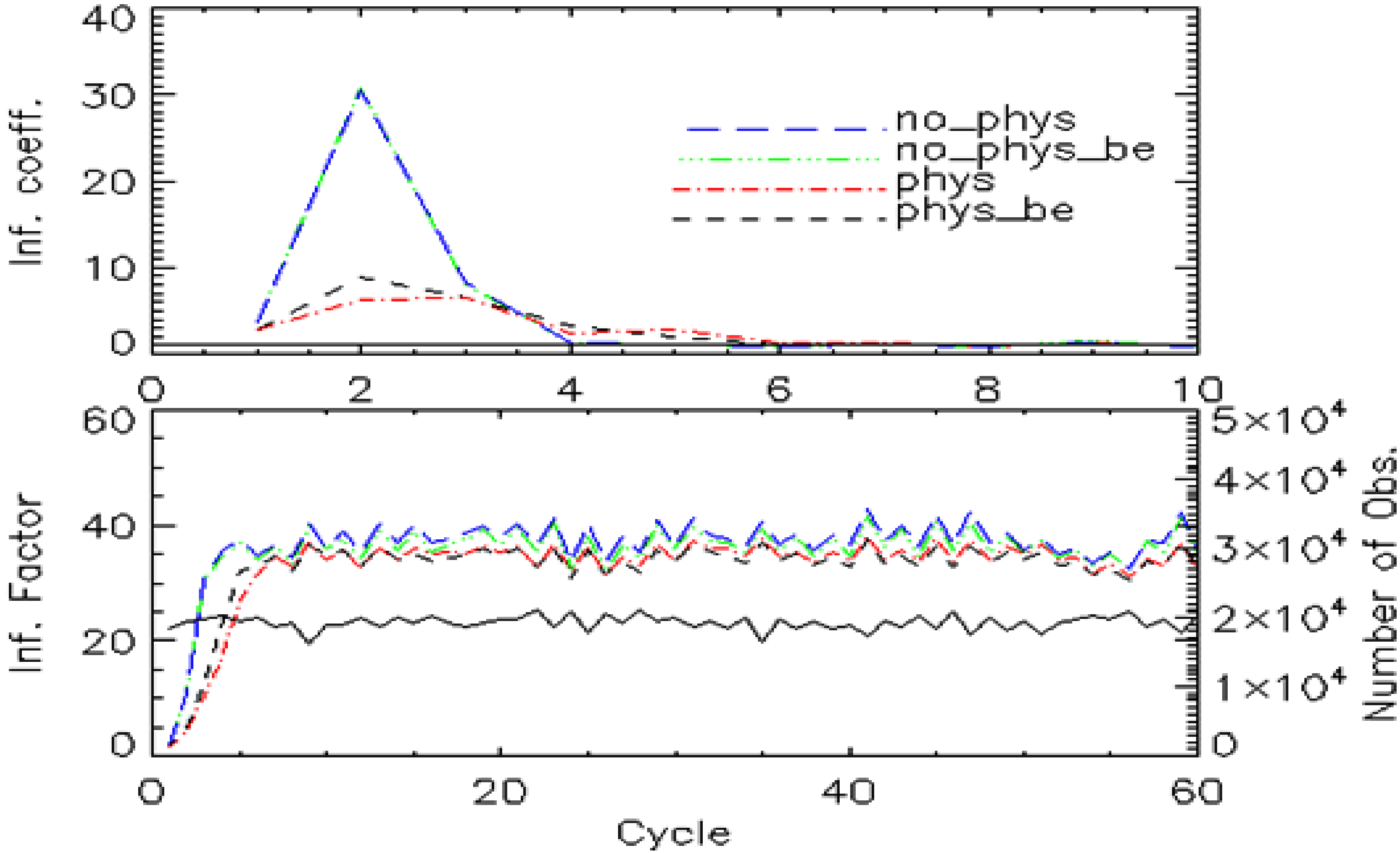
Methodology → Model configuration

- WRF V3.2
- 45-km resolution
- Single domain
- 160 × 160 grid points
- 35 vertical levels
- 20-mem ensemble
- 1 - 31 March 2010
- Assimilate all “standard” obs. + atmospheric motion vectors



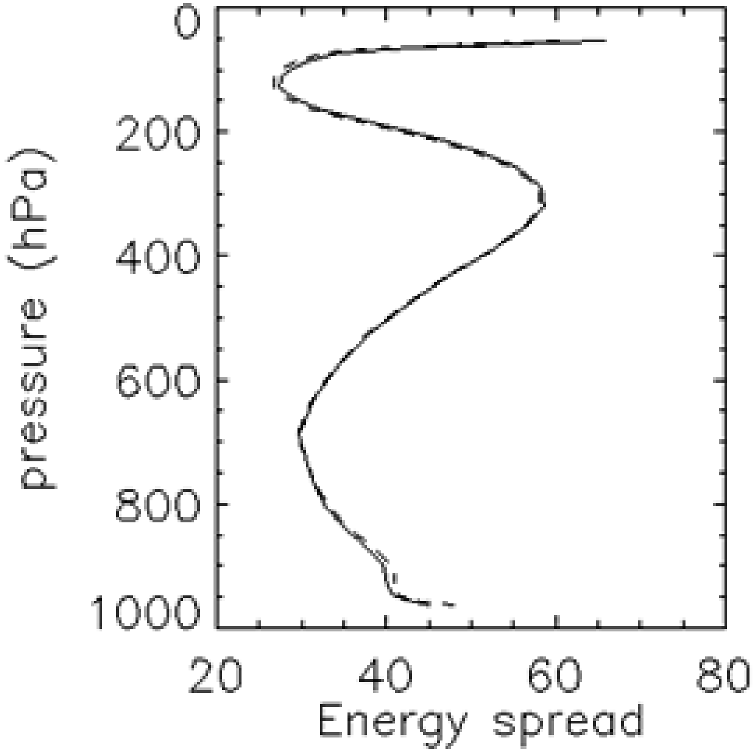
RESULTS: The ETKF ensemble (100% B)

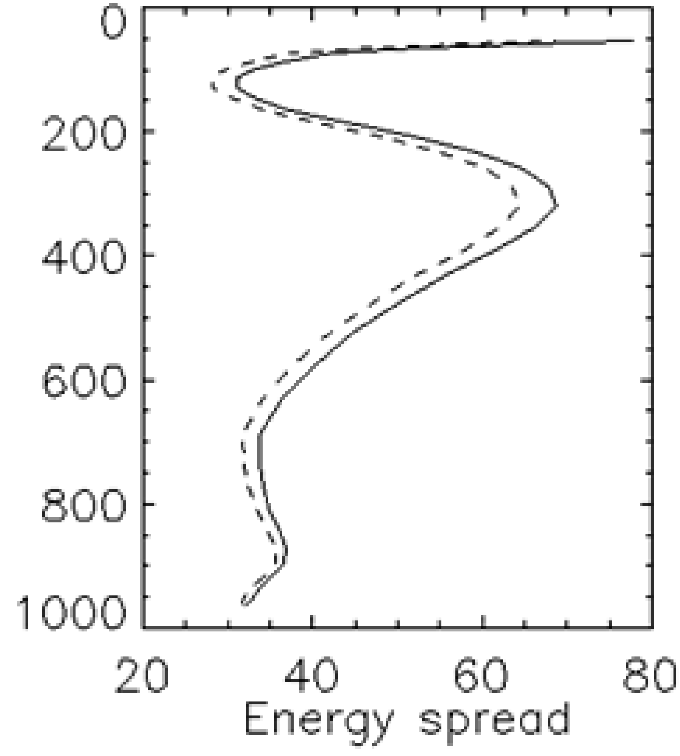
- Ensemble spread → inflation



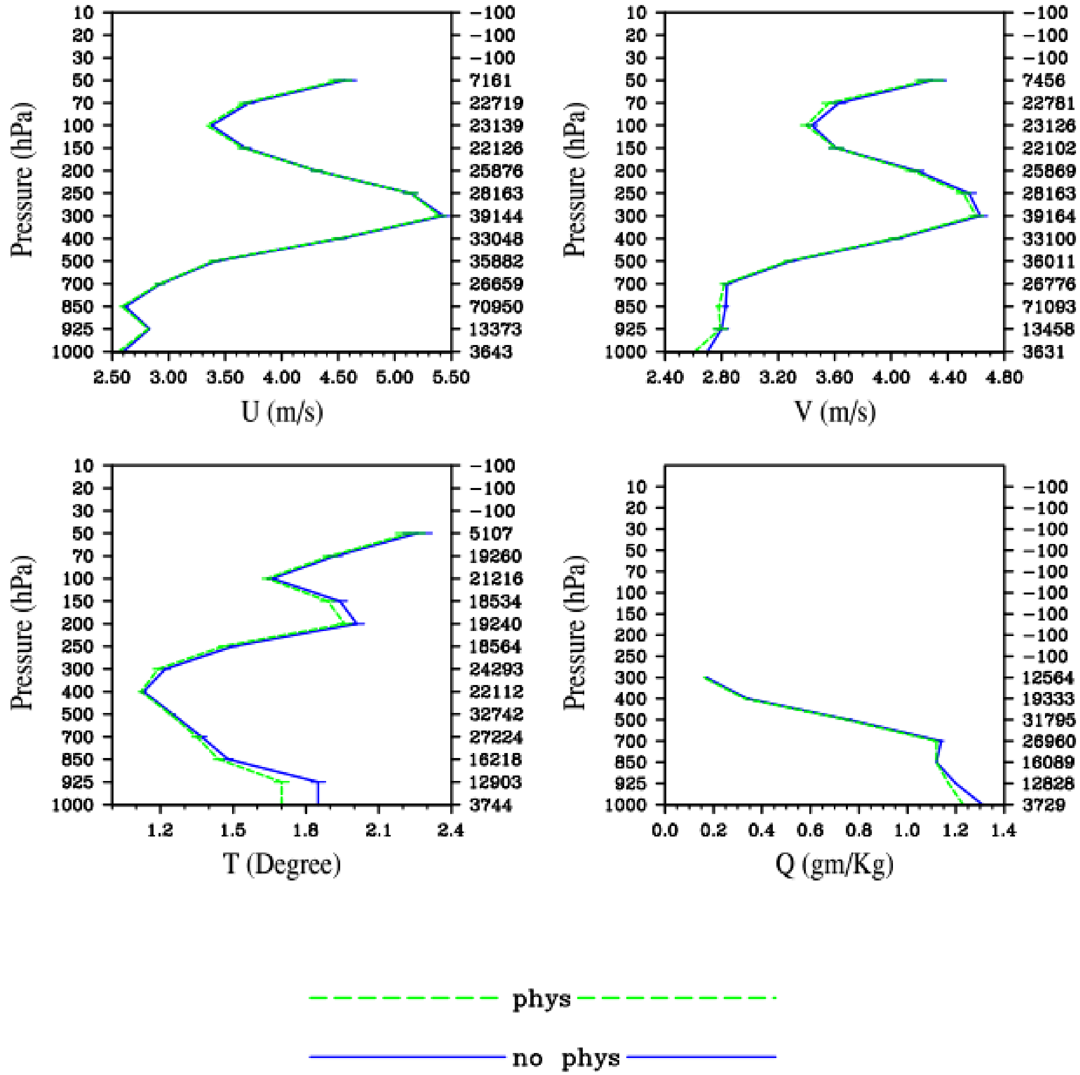
ETKF transformation of ensemble forecast perturbations

Month average total energy profiles for ***phys*** (left) and ***no_phys*** (right). Dashed = *a priori*, solid = *a posteriori*





First Guess (Ensemble Mean) RMSE

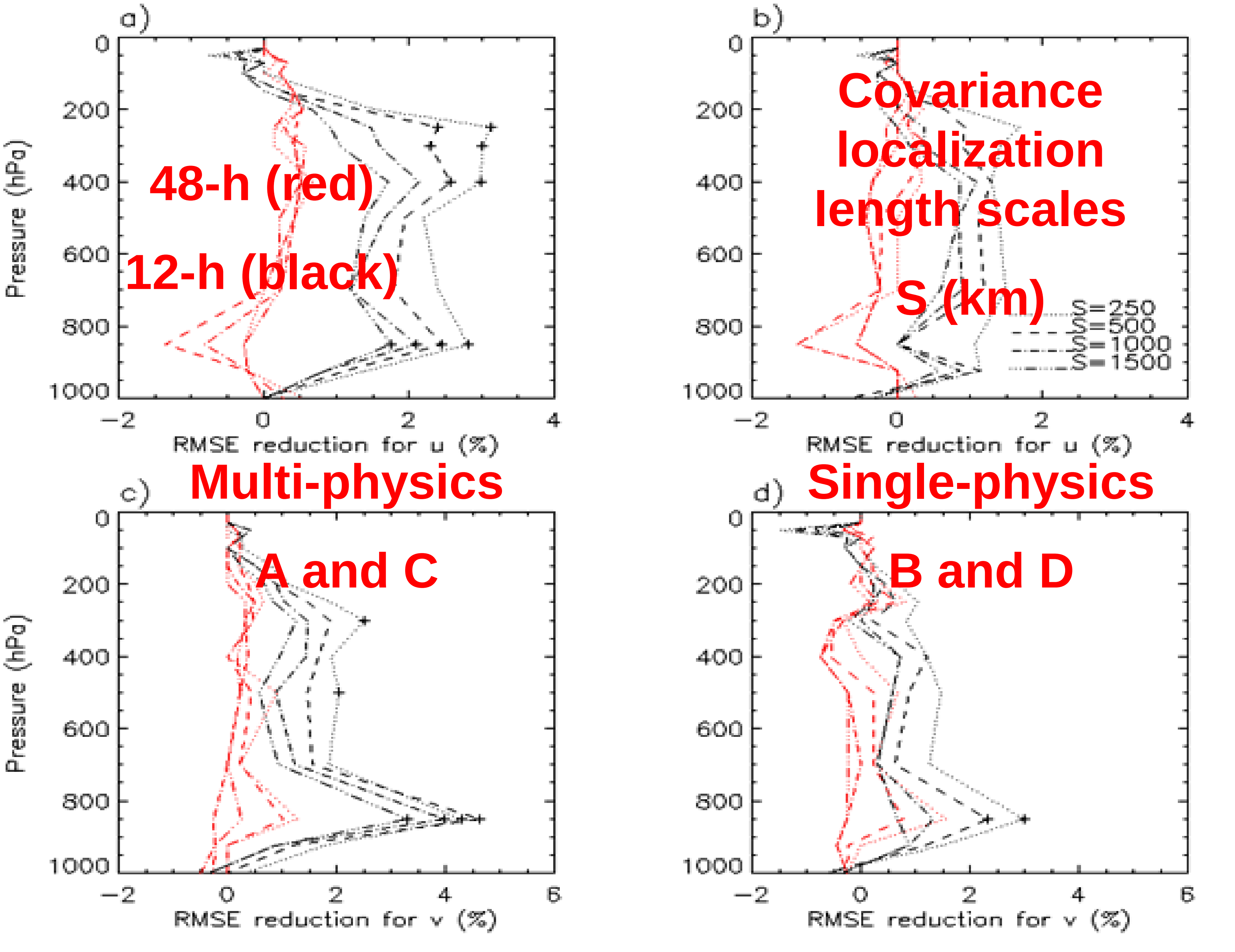


ETKF ensemble:

- Requires large inflation factor to maintain ensemble spread
- Provides stable month-long ensemble
- Ensemble 3DVAR more skillful than standard 3DVAR! (6 or more % RMSE reduction!)
  - Multi-physics ensemble mean most skillful
  - Single-physics cycling DA and deterministic forecasts most skillful

RESULTS: Hybrid 3DVAR-ETKF data assimilation

- 12- and 48-h deterministic forecasts from each analysis during the 1-month long cycling expts.
- T-test used to confirm differences are statistically significant (indicated by +)
- RMSE reduction is relative to ensemble 3DVAR (i,e., 100% **B**)
- Only showing $1/\beta_2 = 0.2$ (20% $\mathbf{P}^f$, 80% **B**)


a)
b)
c)
d)
Pressure (hPa)
RMSE reduction for u (%)
RMSE reduction for v (%)
48-h (red)
12-h (black)
Covariance localization length scales
S (km)
S=250
S=500
S=1000
S=1500
Multi-physics
Single-physics
A and C
B and D

Summary

- ETKF provides *a method* to maintain an ensemble for the hybrid system
- Use of the ensemble mean as the first guess in the 3DVAR cost function significantly improves analysis skill
- Incorporating ensemble-based error covariances into the hybrid cost function added skill to the analysis
  - This was in addition to the skill added by using the ensemble mean as the first guess
  - Multi-physics ensemble covariances proved to be most beneficial

Future Work

- Hybrid system is fairly robust but needs a skillful ensemble
- ETKF ensemble is not optimal
  - It doesn't localize (requires large inflation factor)
  - It's not consistent with the hybrid cost function (simply estimates what EnKF would provide)
- WRFDA system can produce its own ensemble!
  - Lanczos minimization algorithm provides estimate of inverse Hessian of hybrid variational cost function (this is the analysis error covariance) (Collaboration with Tom Auligne: NCAR)

Thanks!

Additional slides

Outline

- Introduction
- Purpose of study
- Description of the WRF hybrid scheme
- Methodology
- RESULTS: Optimization of the ETKF ensemble
- RESULTS: Hybrid 3DVAR-ETKF data assimilation
- RESULTS: HLEF as an alternative to ETKF
- Concluding remarks

Introduction → 3DVAR

- Minimize cost function (maximize likelihood)

$$J = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathbf{b}}\right)^{T}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathbf{b}}\right)+\frac{1}{2}\left(\mathbf{y}-H\left(\mathbf{x}\right)\right)^{T}\mathbf{R}^{-1}\left(\mathbf{y}-H\left(\mathbf{x}\right)\right)$$

- Incremental formulation → WRF 3DVAR

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^{\mathbf{g}}$$

Outer loop $\mathbf{x}^{\mathbf{g}} = \mathbf{x}^{\mathbf{b}}$

$$J = \frac{1}{2}\delta\mathbf{x}^{T}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\left(\mathbf{d}-\mathbf{H}\delta\mathbf{x}\right)^{T}\mathbf{R}^{-1}\left(\mathbf{d}-\mathbf{H}\delta\mathbf{x}\right)$$

Introduction → EnKF

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{P}_i^f \mathbf{H}_i^T + \mathbf{R} \right)^{-1}$$

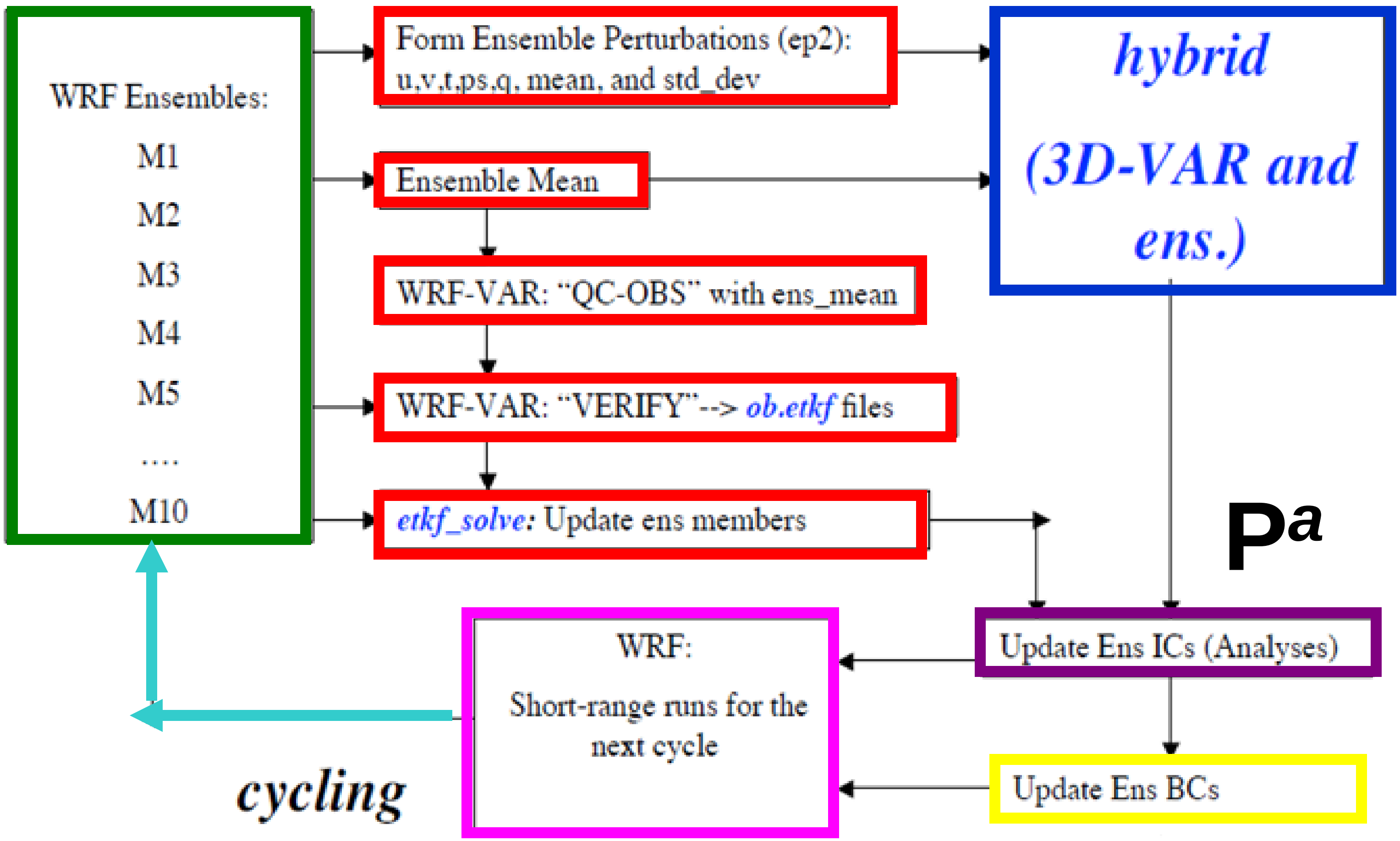
$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i \left[\mathbf{y}^o - H\left(\mathbf{x}_i^f \right) \right]$$

$$\mathbf{P}_i^a = \left(\mathrm{I} - \mathbf{K}_i \mathbf{H}_i \right) \mathbf{P}_i^f$$

$$\mathbf{x}_{i+1}^f = M\left(\mathbf{x}_i^a \right)$$

$$\mathbf{P}_{i+1}^f = \mathbf{M}_i \mathbf{P}_i^a \mathbf{M}_i^T + \mathbf{Q}_i$$

P^f = B → OI & 3DVAR



WRF Ensembles:
M1
M2
M3
M4
M5
....
M10
Form Ensemble Perturbations (ep2): u,v,t,ps,q, mean, and std_dev
Ensemble Mean
WRF-VAR: "QC-OBS" with ens_mean
WRF-VAR: "VERIFY"--> ob.etkf files
etkf_solve: Update ens members
hybrid (3D-VAR and ens.)
Pa
Update Ens ICs (Analyses)
Update Ens BCs
WRF: Short-range runs for the next cycle
cycling

Ensemble generation techniques

- Inflation factor

The squared innovations should equal the trace of the sample background forecast-error covariance estimate – That is, the ensemble spread should approximate the ensemble mean error

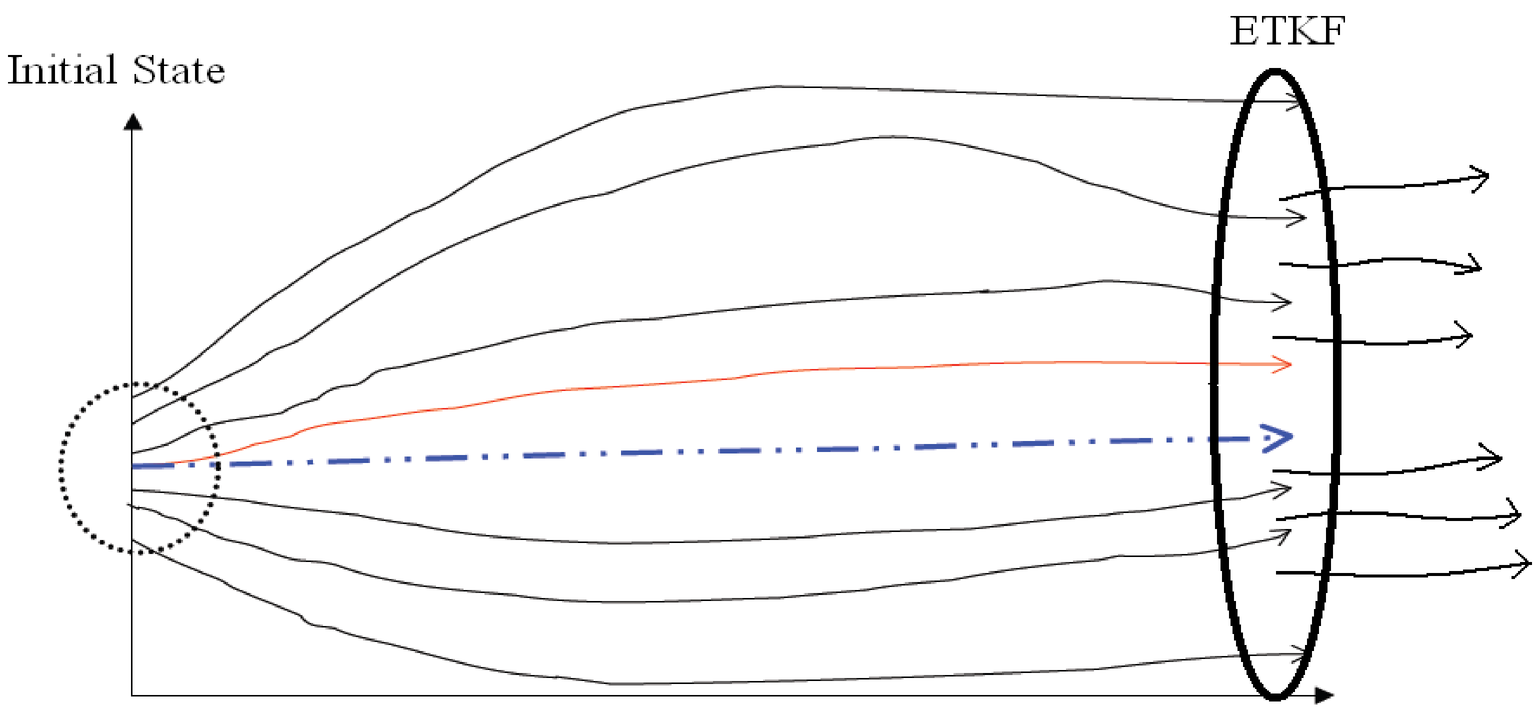
$$\tilde{\mathbf{d}}_i^{\mathbf{T}}\tilde{\mathbf{d}}_i = Tr(\tilde{\mathbf{H}}\alpha_i\mathbf{P}^e\tilde{\mathbf{H}}^{\mathbf{T}} + \mathbf{I})$$

$$\tilde{\mathbf{d}} = \mathbf{R}^{-1/2}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}^f}\right)$$

$$\prod_i = \prod_{i-1}\sqrt{\alpha_i}$$

Ensemble generation techniques

- ETKF rescaling of ensemble perturbations



Theory → HLEF

- MLEF

  State space analysis covariance matrix is

  $$\mathbf{P_a^{1/2} = P_f^{1/2} [I + (X)^T X]^{-1/2}}$$

- HLEF

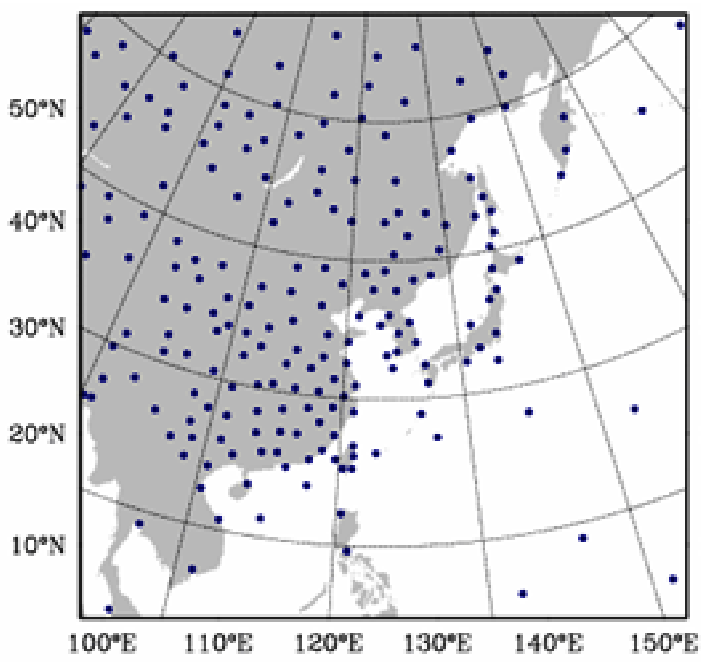
  Replace $\mathbf{[I + (X)^T X]}$ (generalized Hessian at the optimal point) with an equivalent estimate of the Hessian at the optimal point

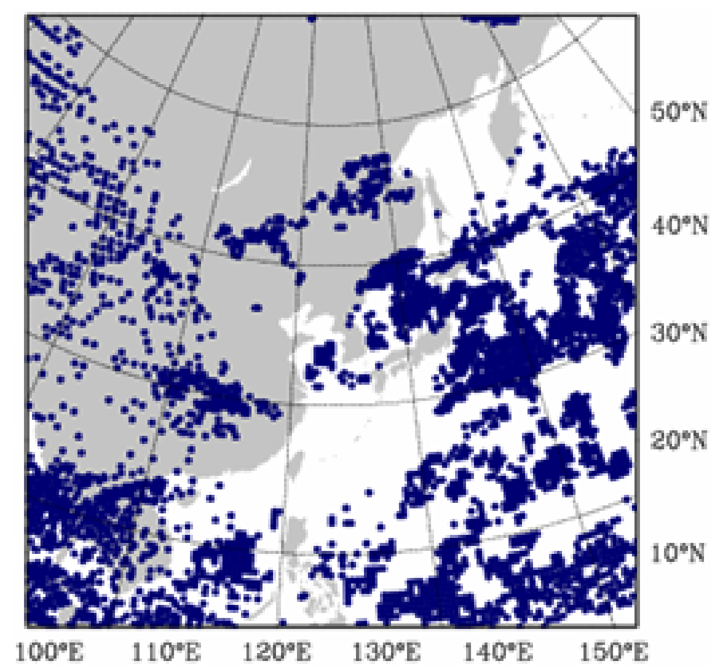
Ensemble generation techniques

- Lanczos minimization algorithm
  - A conjugate gradient minimization algorithm
  - Gives us an equivalent estimate of the inverse Hessian of the variational cost function
- Find solution to $\boldsymbol{Ax=b}$
  - Define $\Phi(x) = \frac{1}{2}x^T\boldsymbol{A}x - x^Tb$
  - Minimize Φ over a set of Lanczos vectors
  - $\|\boldsymbol{A}x - b\|_2$ = residual (of optimal solution)!

Methodology → Observations

- Assimilate "standard NCEP observations"
- And satellite atmospheric motion vectors (AMVs)





Methodology → **ETKF experiments**

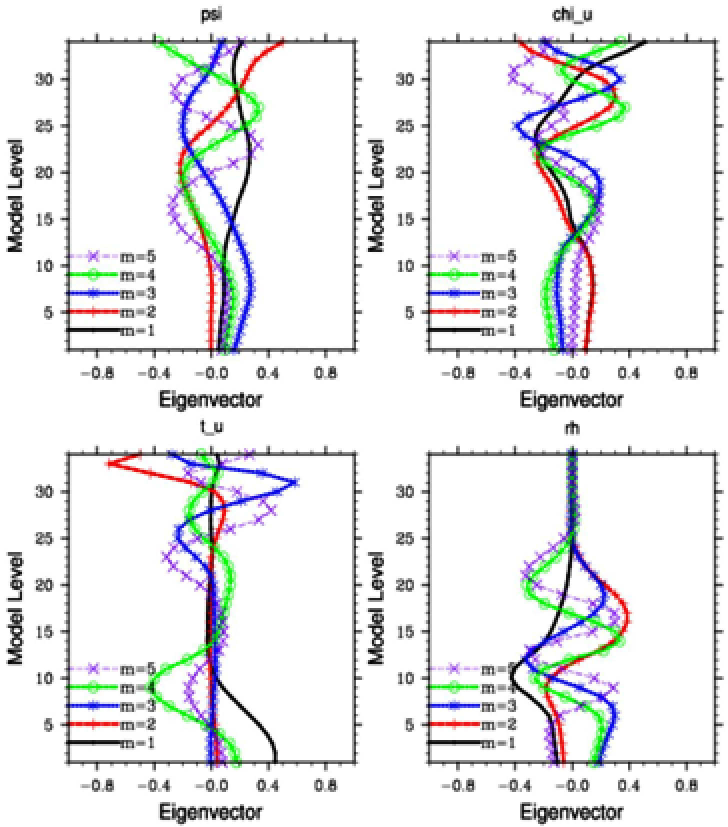
- Vertical covariance localization (secondary consideration)
  - Propose that vertical covariance localization should be consistent with horizontal covariance localization
  - Rossby radius of deformation relation

$$L_R \equiv \frac{NH}{f_0}$$

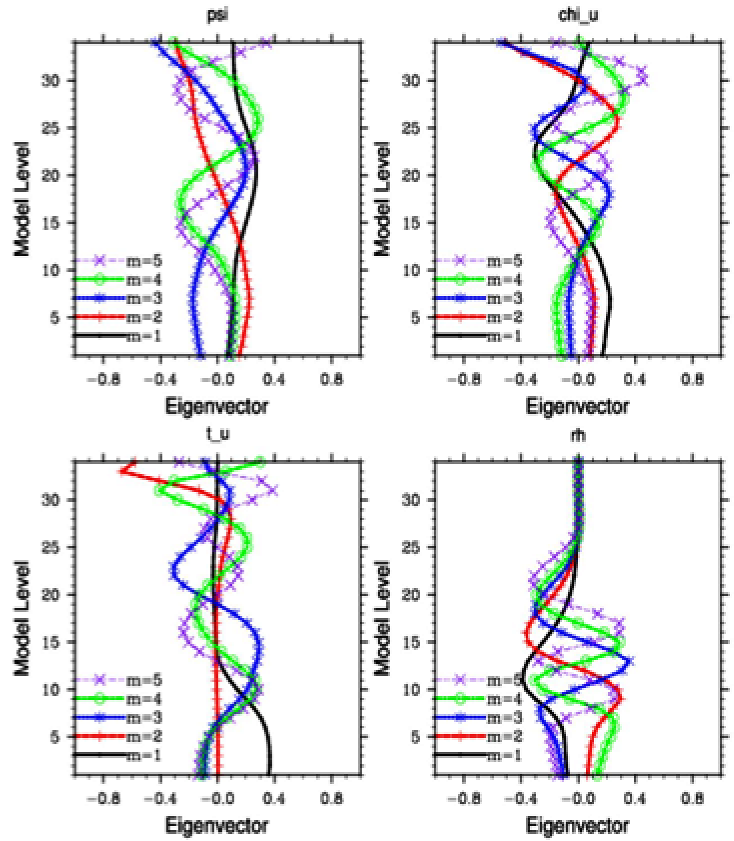
Methodology → Evaluation method

- Diagnose ETKF ensemble maintenance
  - Ensemble spread
  - Ensemble perturbation space directions
  - Dependence on covariance inflation
- Performance of analyses
  - 48-h deterministic forecasts verified at 12 and 48 hours for 30-days (60 analyses)

Tuned 3DVAR background error covariances

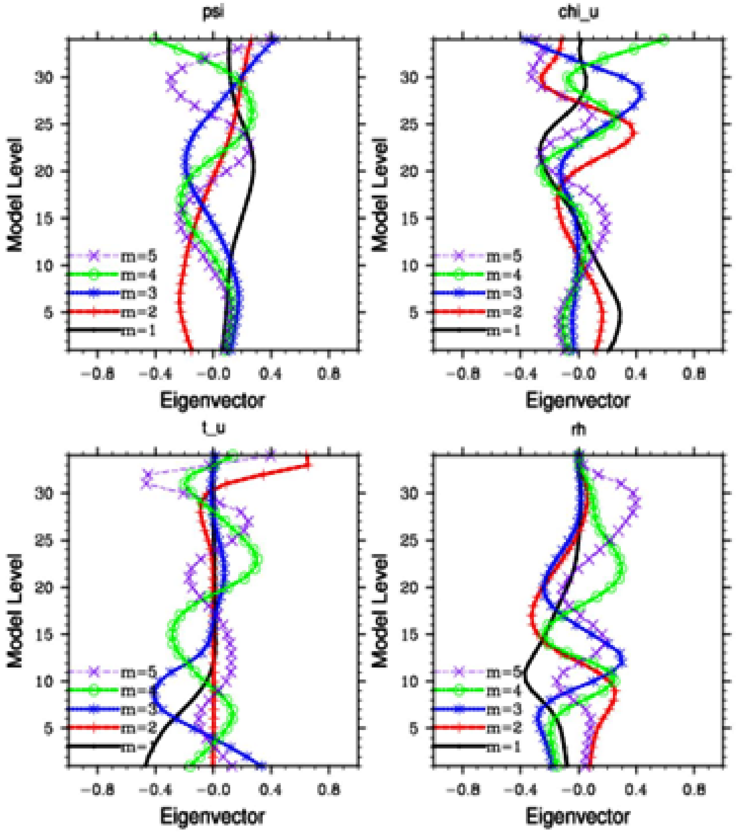


NMC method



single-physics ensemble method

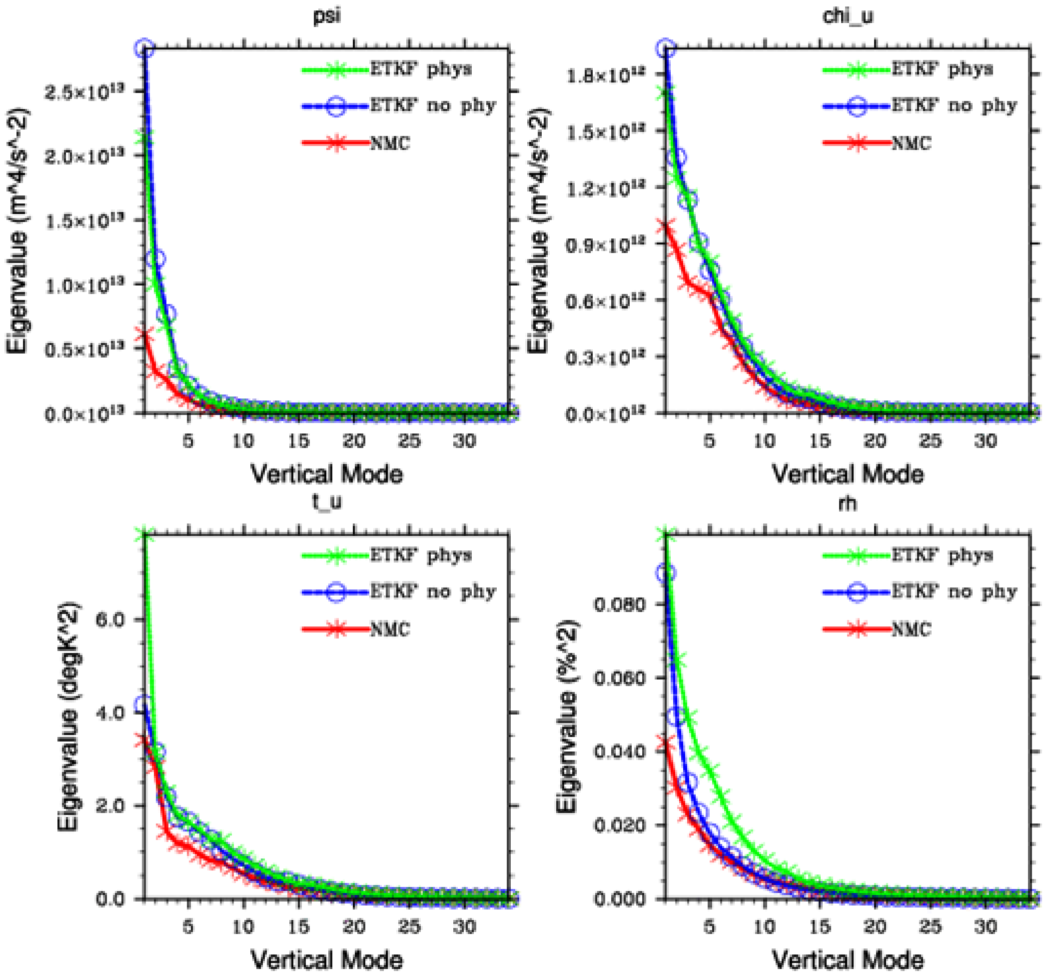
Tuned 3DVAR background error covariances



Single-physics ensemble method

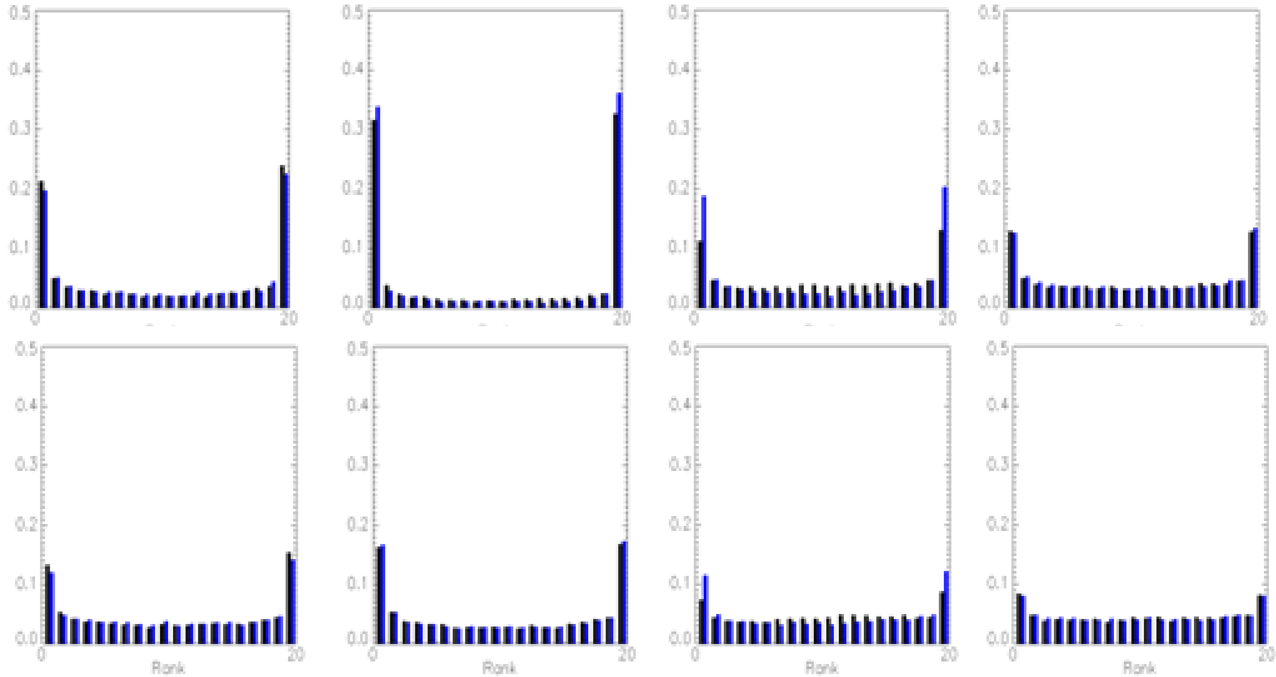
Tuned 3DVAR background error covariances

Eigenvalues corresponding to eigenvectors



RESULTS: Optimization of the ETKF ensemble

- Ensemble spread → Rank histograms



U and V winds for ***phys*** and ***no_phys*** (top = original, bottom = added observations noise to ensemble members). Cycles 2,4,6,8

RESULTS: Optimization of the ETKF ensemble

- ETKF transformation of ensemble forecast perturbations
- Define Total dry energy spread

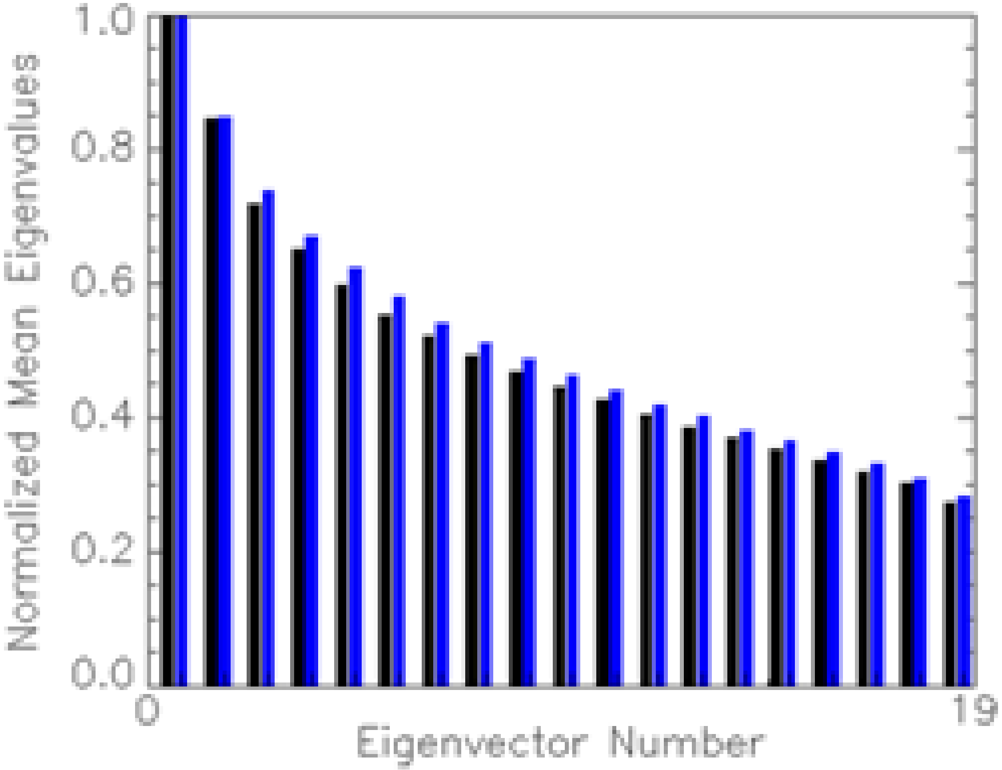
$$E(i,j,k)=\frac{1}{2}\left[u^2(i,j,k)+v^2(i,j,k)+\frac{C_p}{T_r}T^2(i,j,k)\right]$$

RESULTS: Optimization of the ETKF ensemble

- Maintenance of variance in orthogonal directions

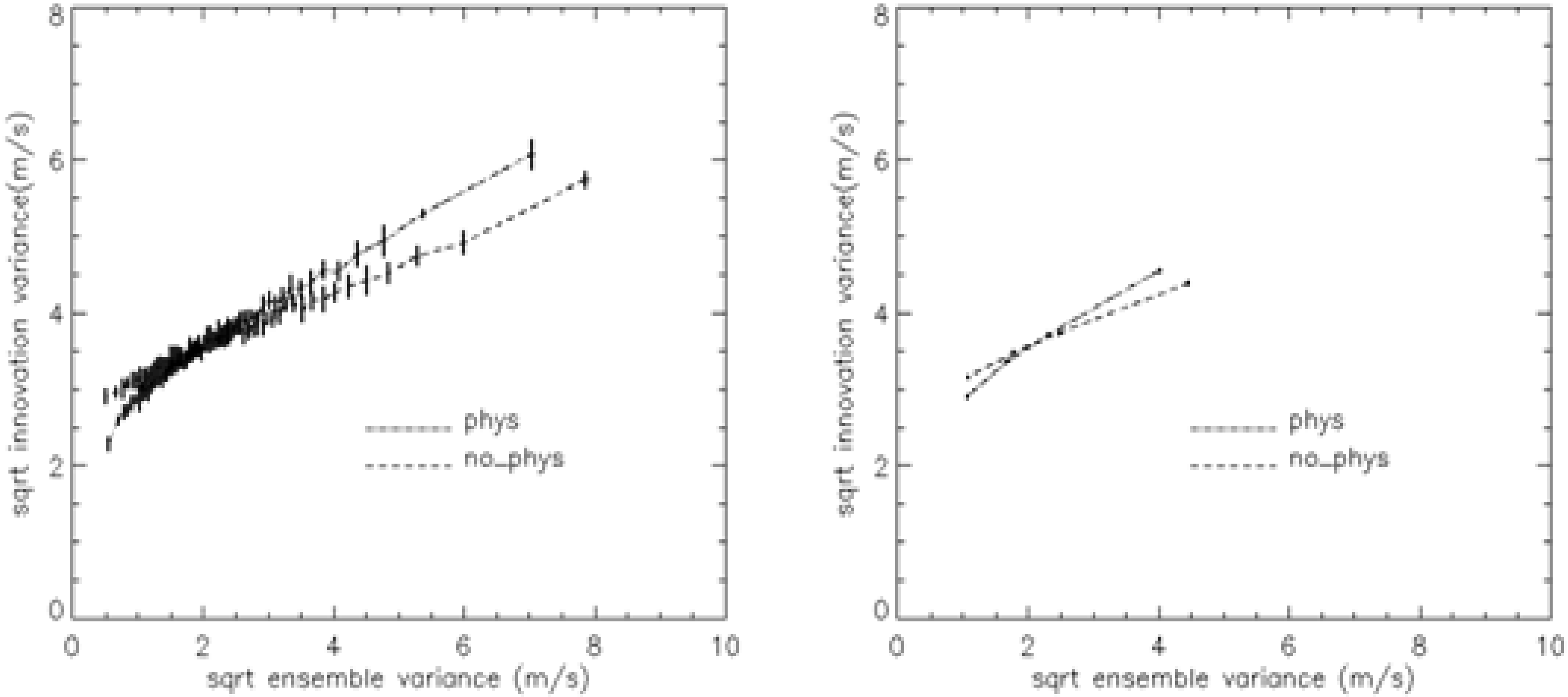
Eigenvalues of ensemble forecast spread in obs. space

Phys
No_phys
1-month average



RESULTS: Optimization of the ETKF ensemble

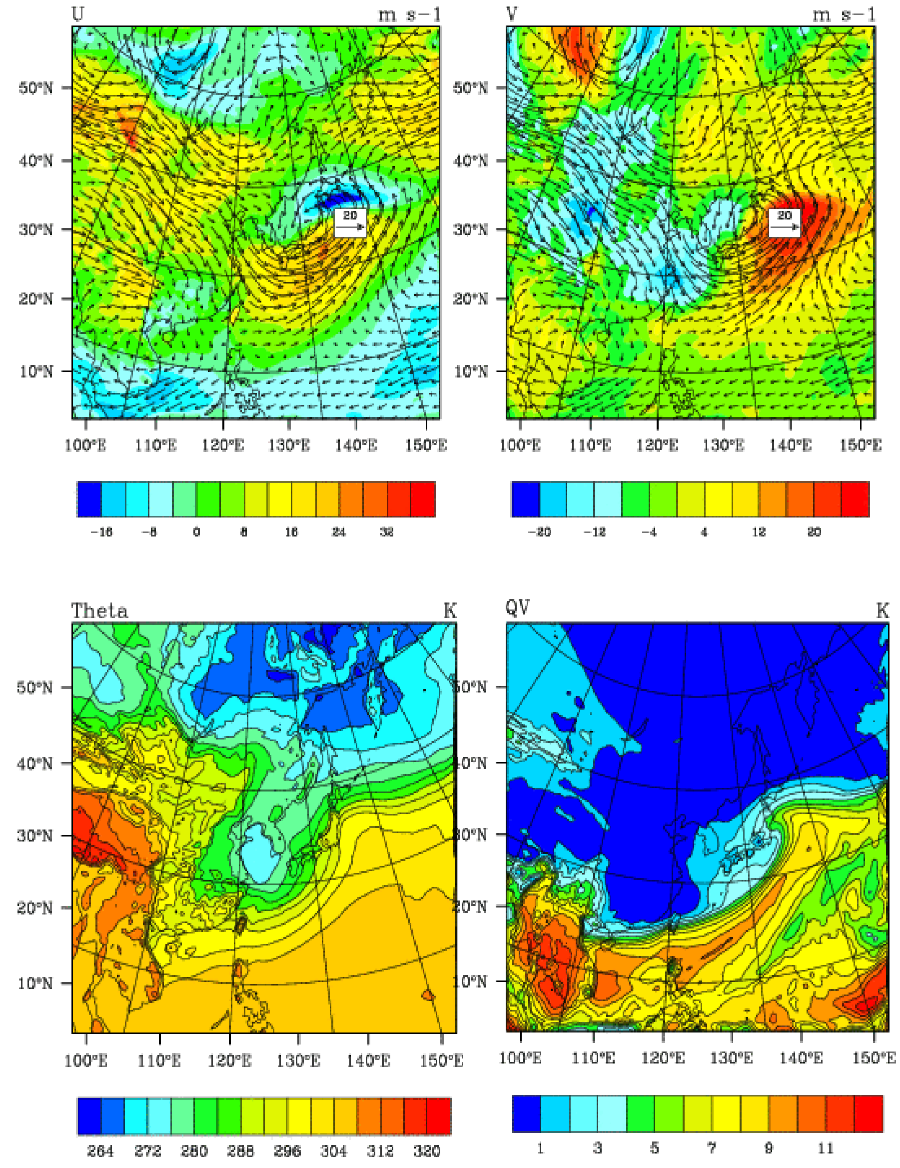
- Ensemble precision (range of resolved innovation variance)



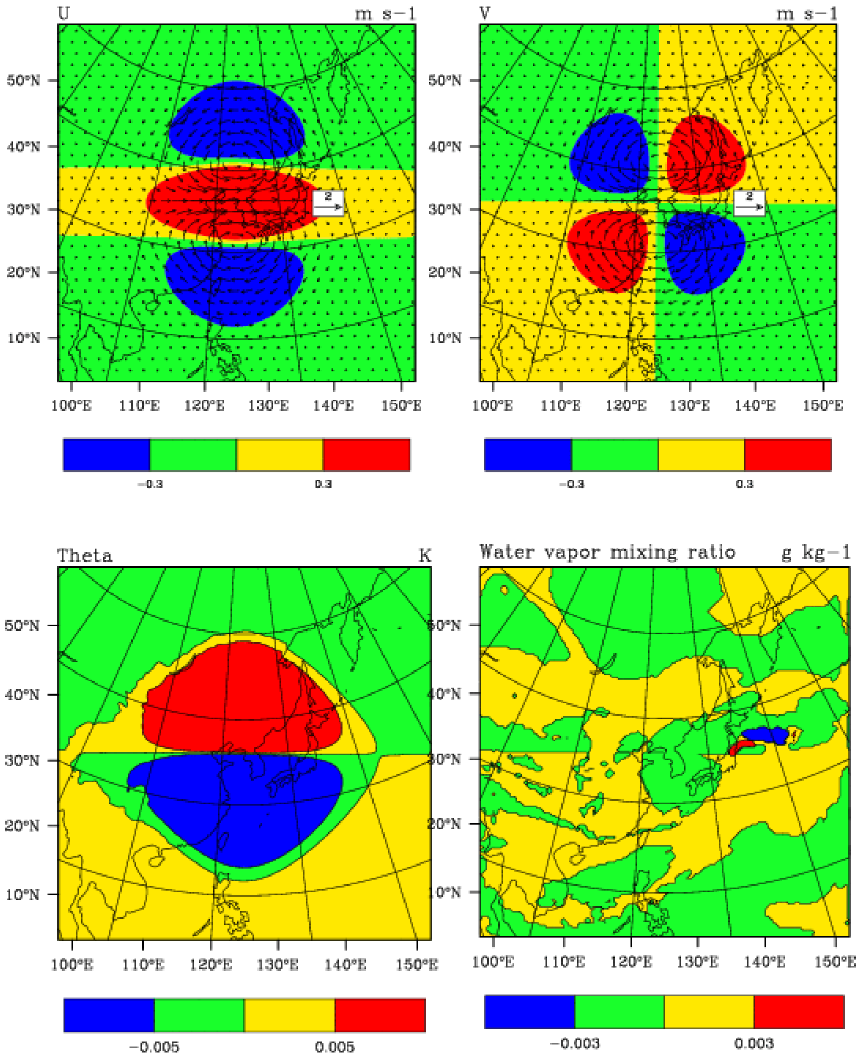
RESULTS: Hybrid 3DVAR-ETKF data assimilation

- Hybrid analysis increments
  - Pseudo observation tests for "tuned" **B**
  - Then, Pseudo observation tests for $\mathbf{P^f + B}$
    - Innovation = 5 m/s
    - Obs error = 1 m/s
    - Center of domain at ~ 800 hPa level
    - Valid 10 March 2010
      - Mature ensemble perturbations

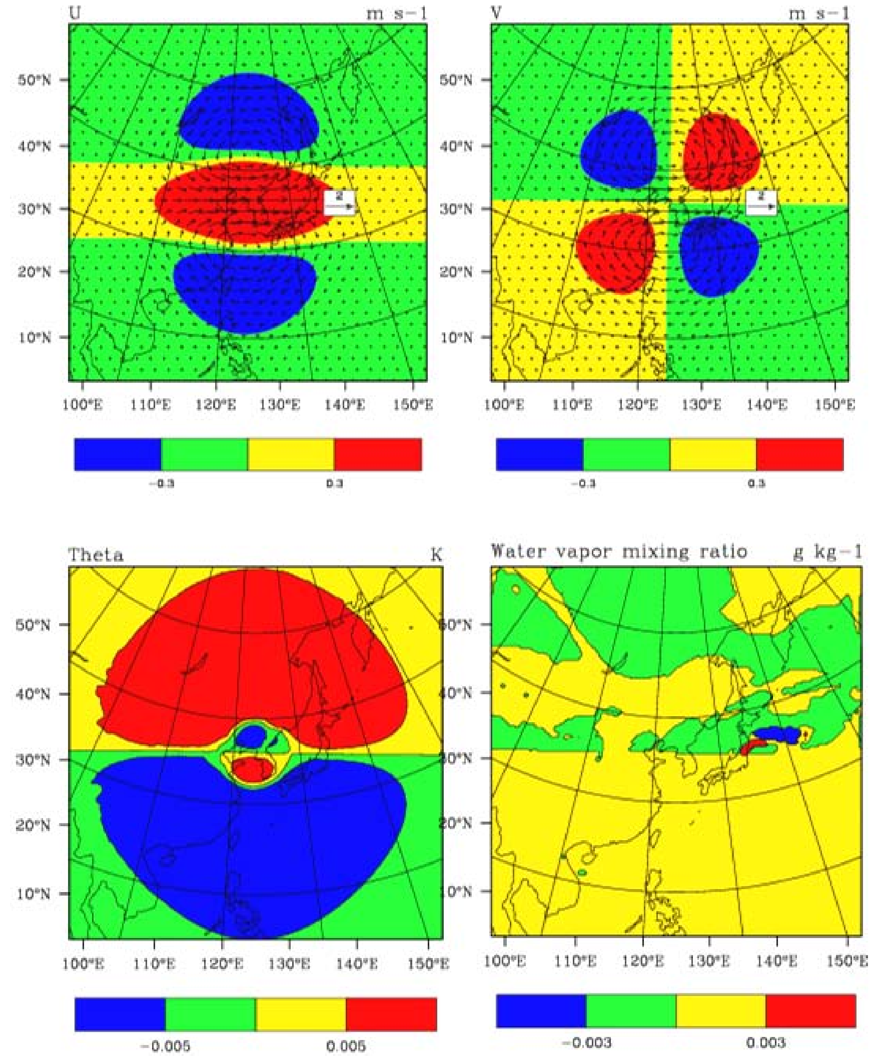
Forecast Valid 10 Mar ‘10



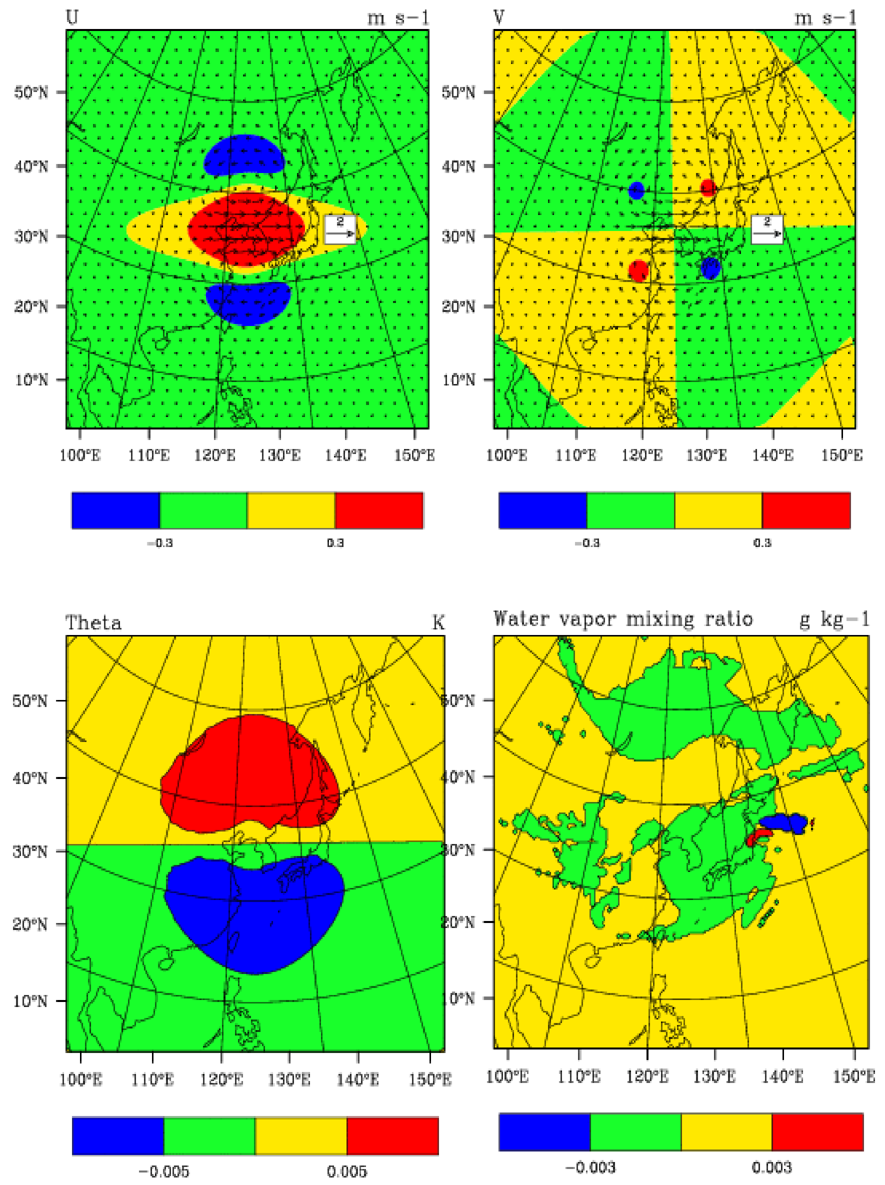
3DVAR **B** Analysis Increments



no_phys-tuned



phys-tuned



U
m s−1
V
m s−1
Theta
K
Water vapor mixing ratio
g kg−1
50°N
40°N
30°N
20°N
10°N
100°E
110°E
120°E
130°E
140°E
150°E
−0.3
0.3
−0.005
0.005
−0.003
0.003

NMC method

Hybrid Analysis Increments

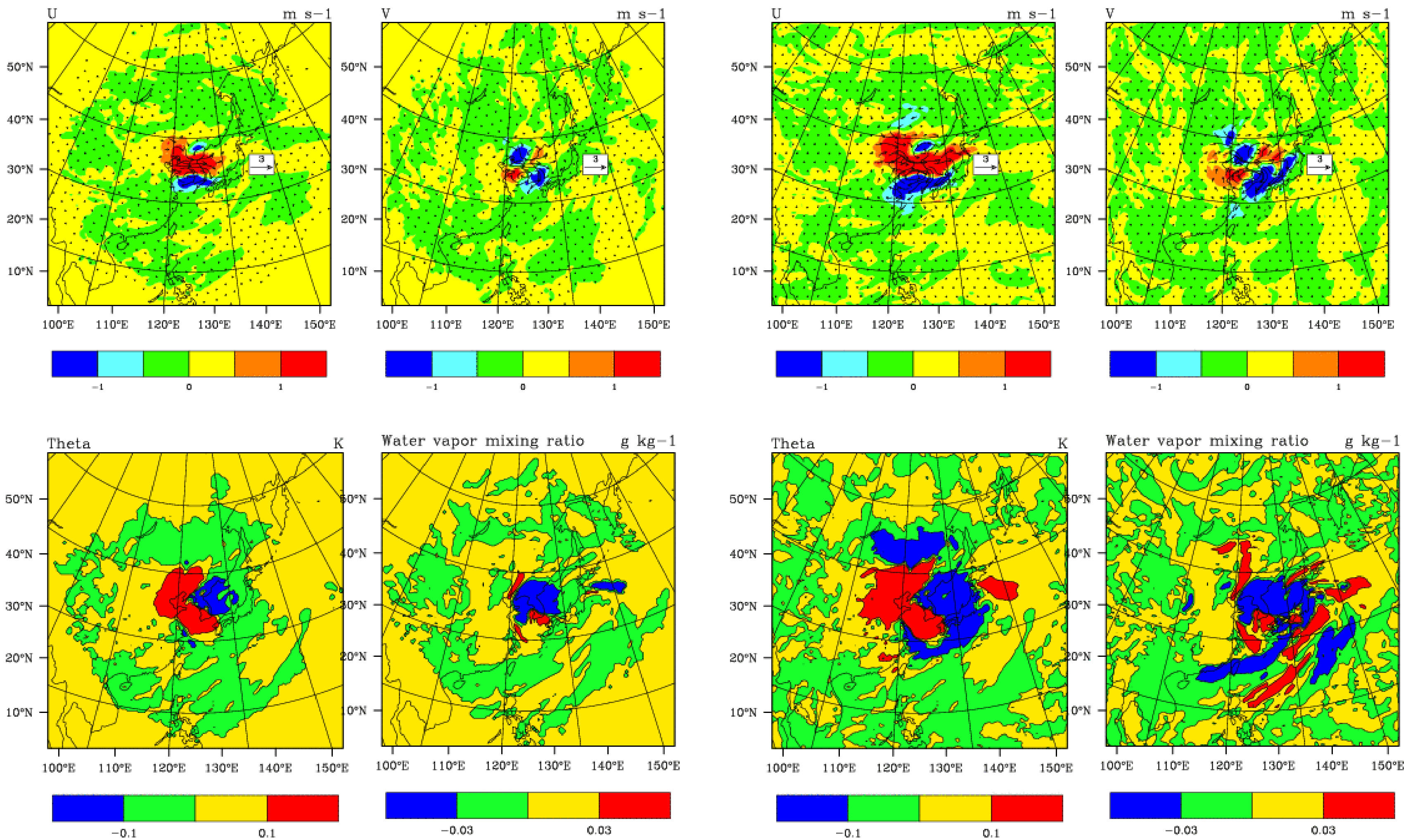
Pure ensemble

$1/\beta_2 = 1.0$

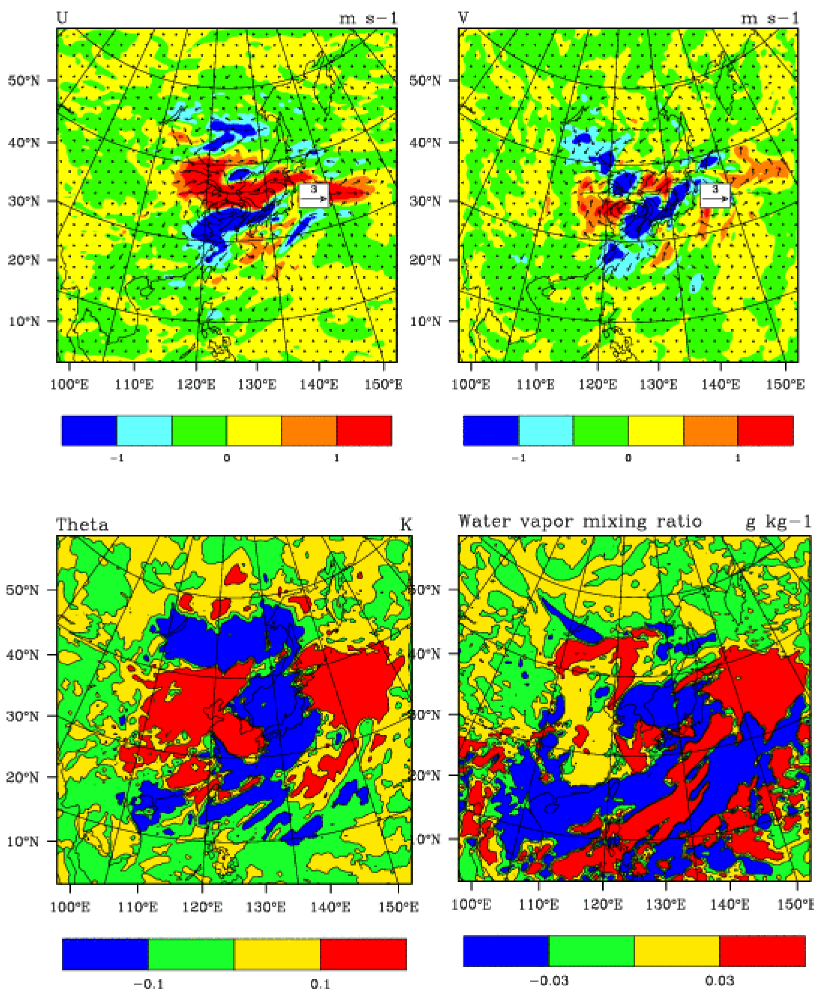
100% $\mathbf{P}^f$, 0% $\mathbf{B}$

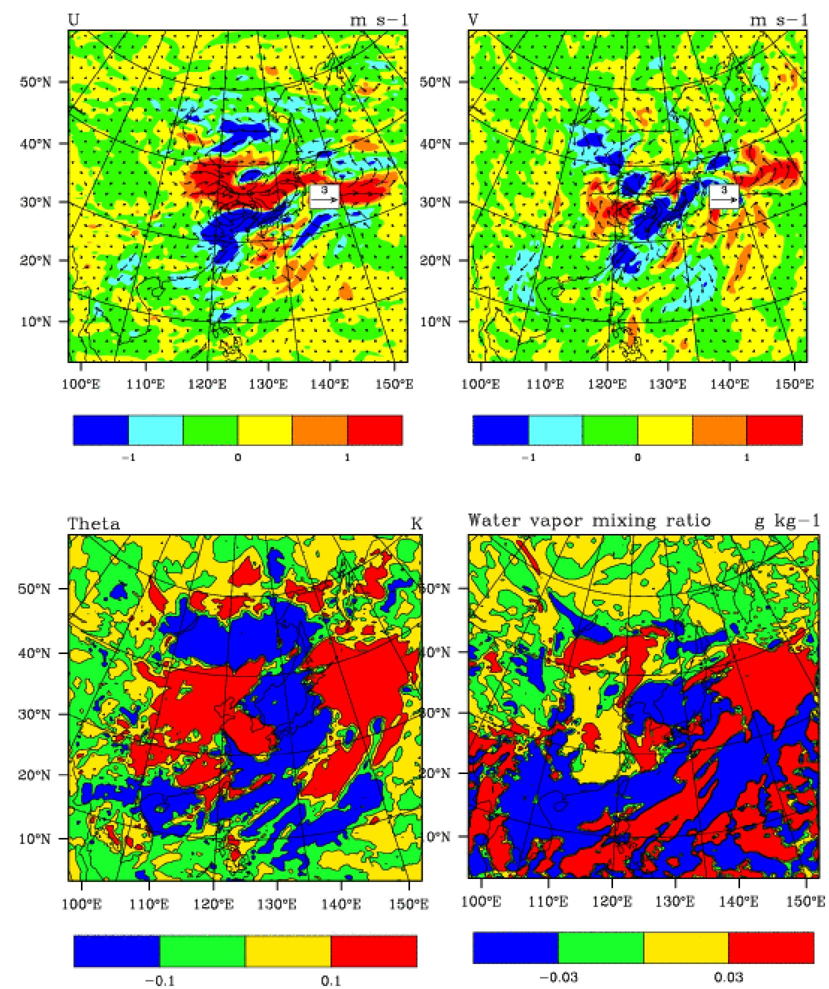
Covariance localization

1500, 1000, 500, and 250 km



U
m s−1
V
m s−1
U
m s−1
V
m s−1
50°N
40°N
30°N
20°N
10°N
100°E
110°E
120°E
130°E
140°E
150°E
3
−1
0
1
Theta
K
Water vapor mixing ratio
g kg−1
Theta
K
Water vapor mixing ratio
g kg−1
−0.1
0.1
−0.03
0.03


U
m s−1
V
m s−1
50°N
40°N
30°N
20°N
10°N
100°E
110°E
120°E
130°E
140°E
150°E
−1
0
1
Theta
K
Water vapor mixing ratio
g kg−1
−0.1
0.1
−0.03
0.03


U
m s−1
V
m s−1
50°N
40°N
30°N
20°N
10°N
100°E
110°E
120°E
130°E
140°E
150°E
−1
0
1
Theta
K
Water vapor mixing ratio
g kg−1
−0.1
0.1
−0.03
0.03

Hybrid Analysis Increments

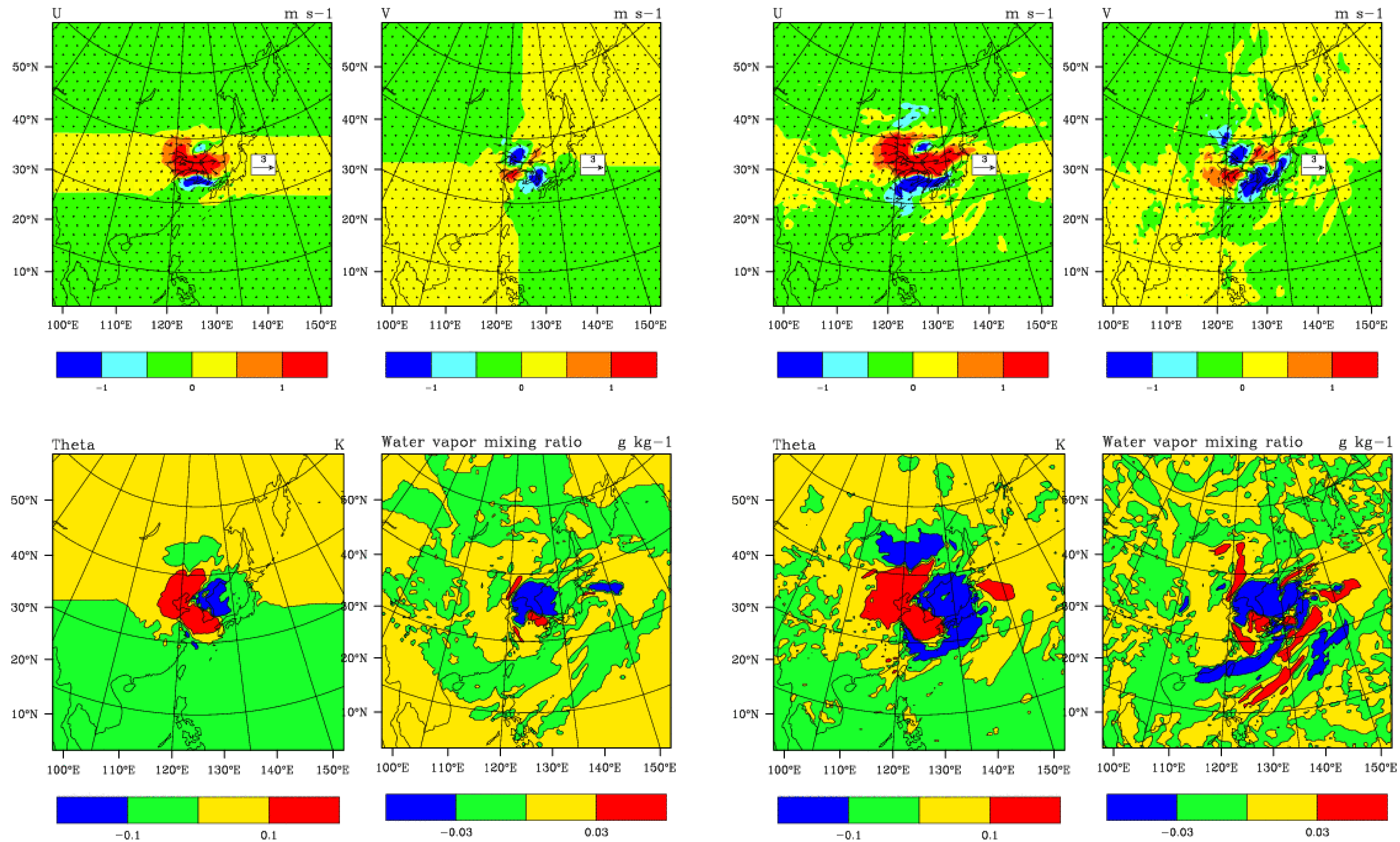
Hybrid

$1/\beta_2 = 0.5$

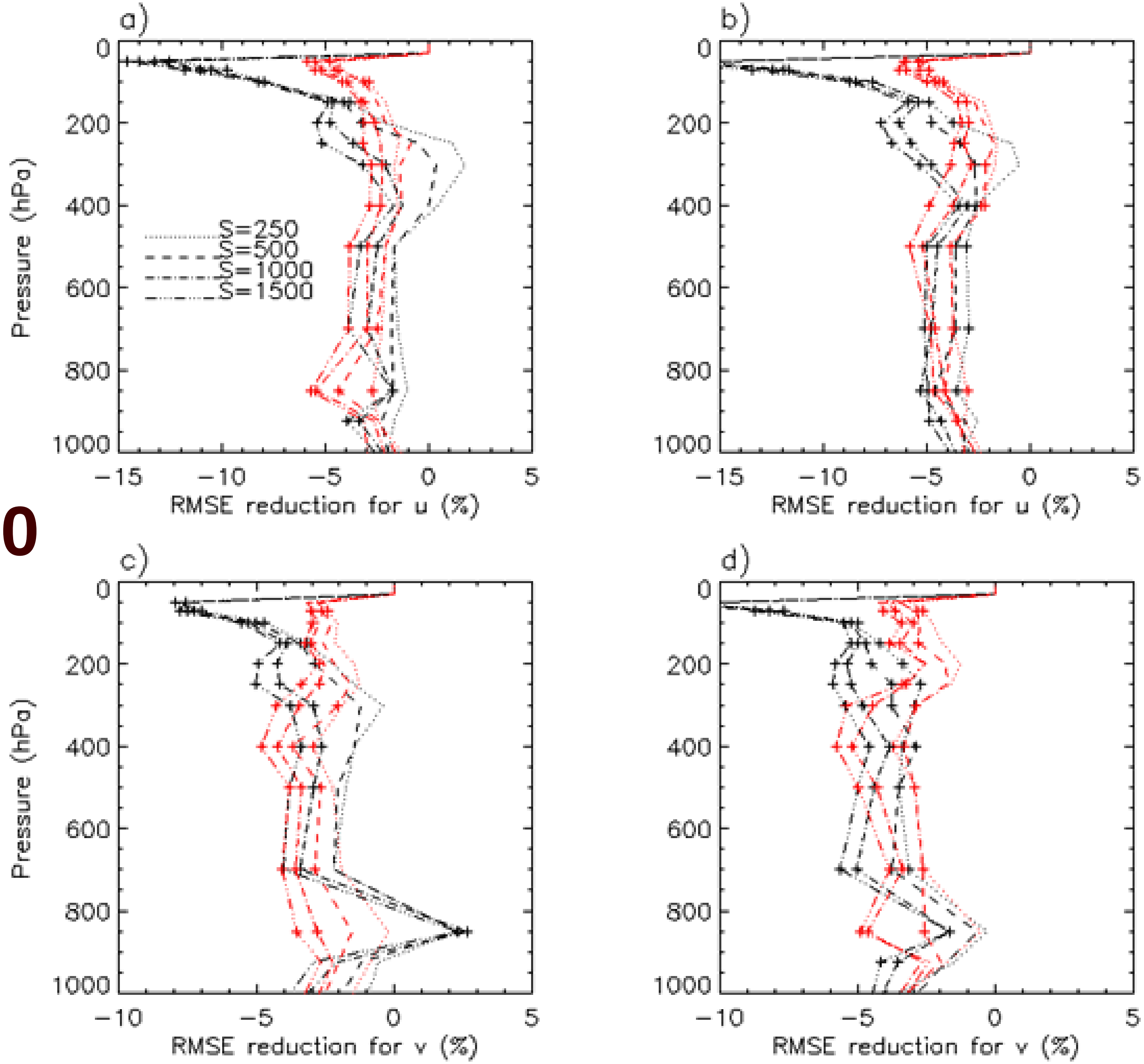
50% $\mathbf{P}^f$, 50% $\mathbf{B}$

Covariance localization

1500, 1000, 500, and 250 km

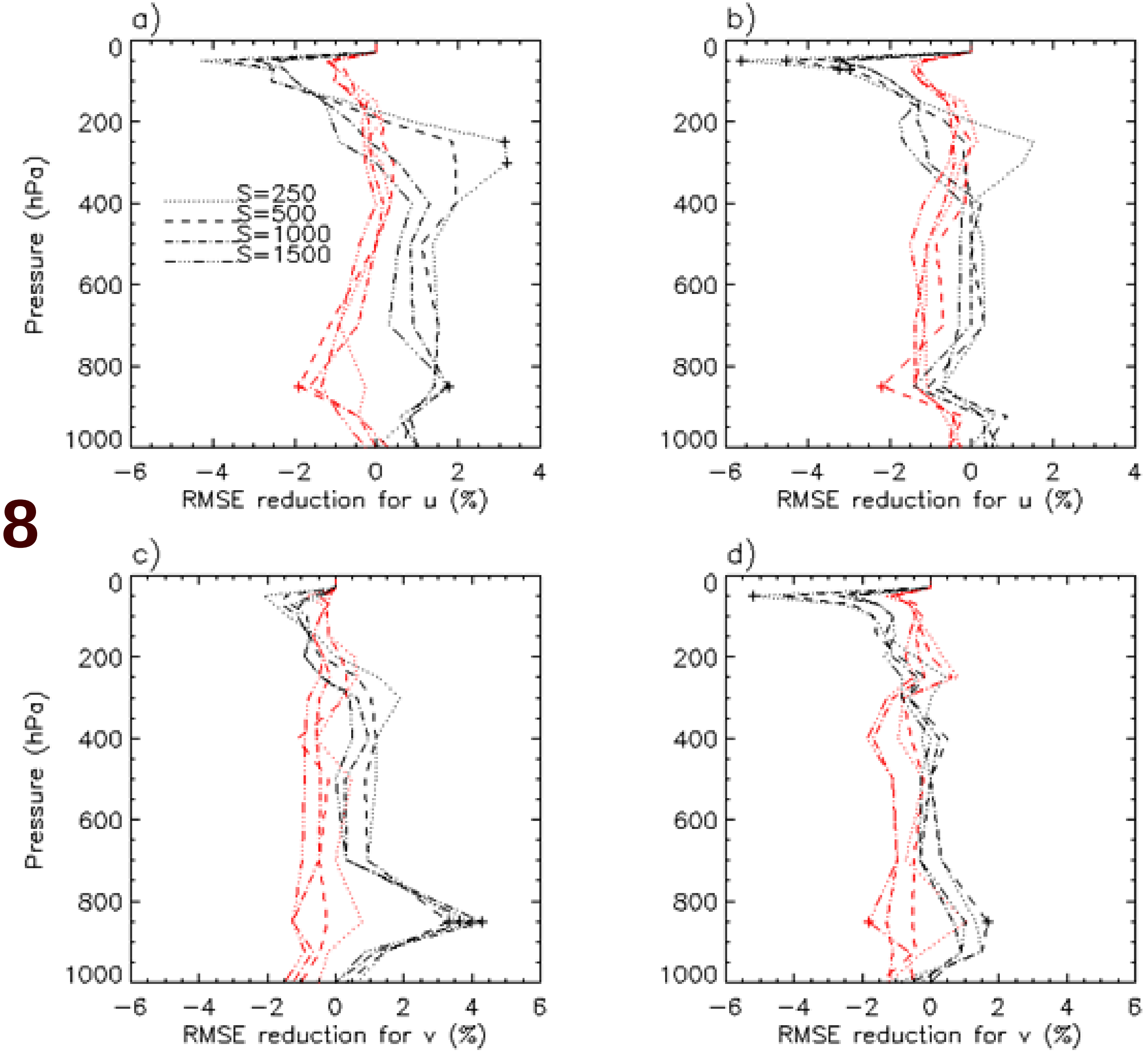

U
m s−1
V
m s−1
−1
0
1
Theta
K
Water vapor mixing ratio
g kg−1
−0.1
0.1
−0.03
0.03
100°E
110°E
120°E
130°E
140°E
150°E
10°N
20°N
30°N
40°N
50°N

1/β2 = 1.0

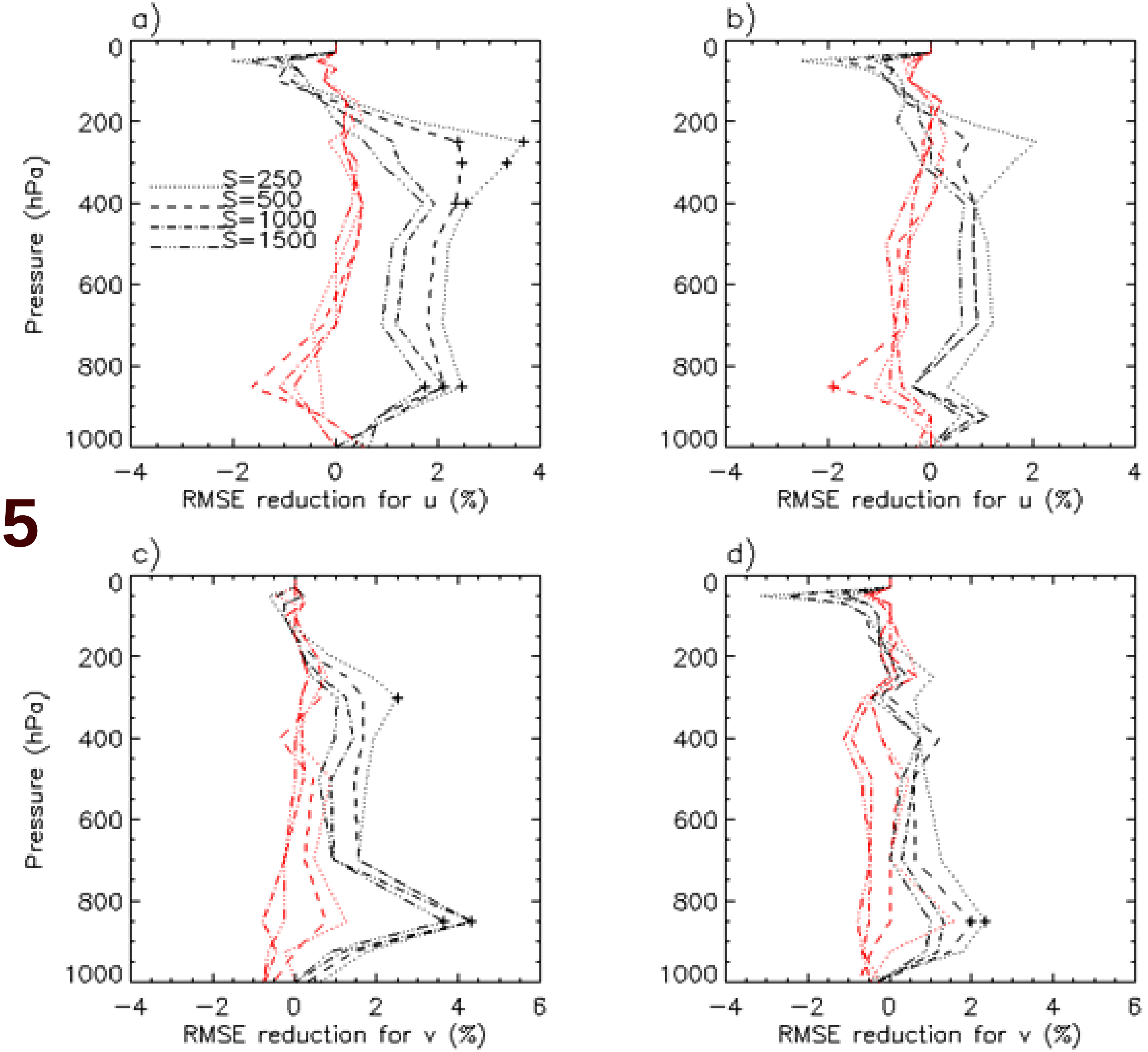


a)
b)
c)
d)
Pressure (hPa)
RMSE reduction for u (%)
RMSE reduction for v (%)
S=250
S=500
S=1000
S=1500

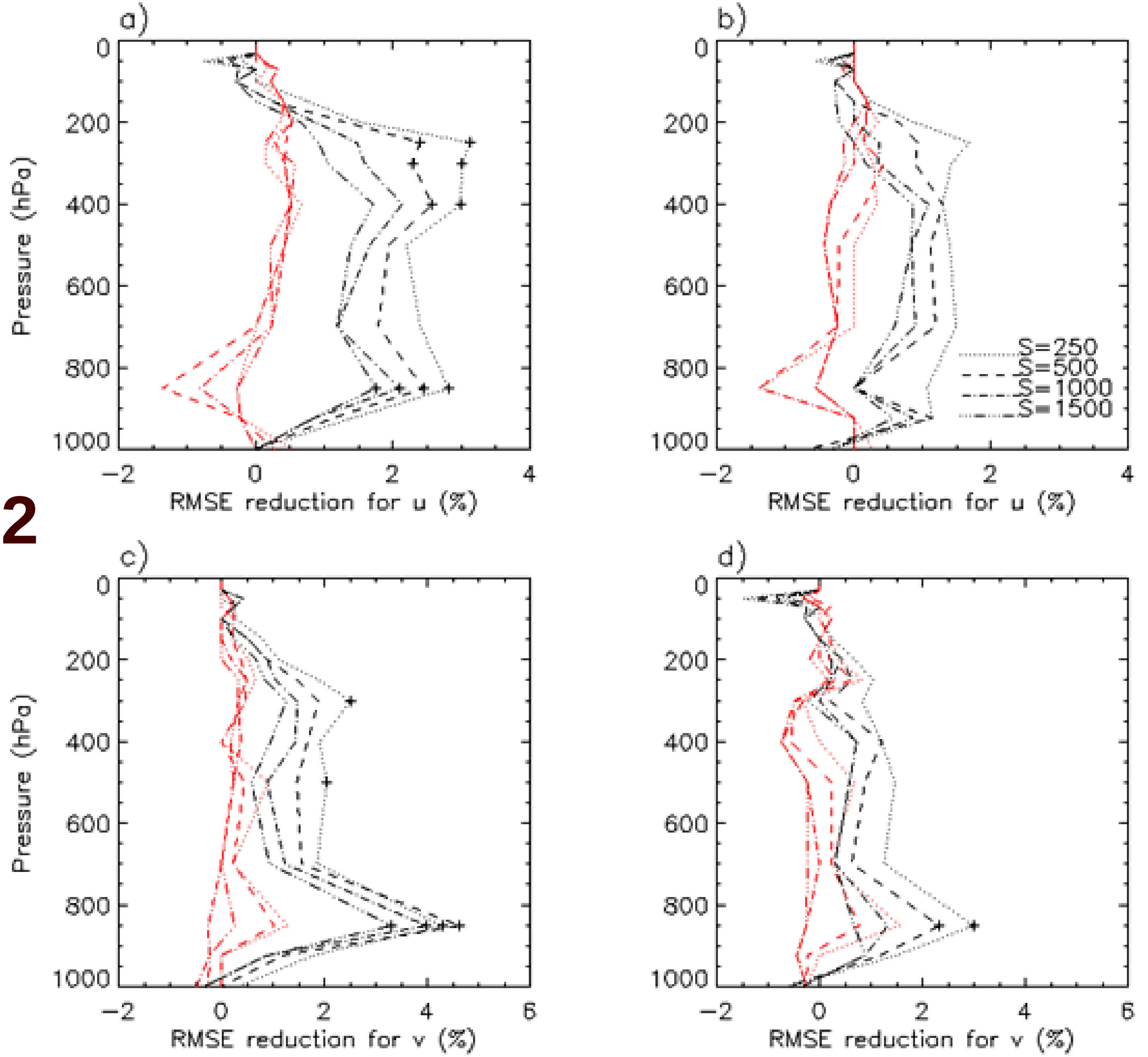
$1/\beta 2 = 0.8$


a)
b)
c)
d)
Pressure (hPa)
RMSE reduction for u (%)
RMSE reduction for v (%)
S=250
S=500
S=1000
S=1500

1/β2 = 0.5


a)
b)
c)
d)
Pressure (hPa)
S=250
S=500
S=1000
S=1500
RMSE reduction for u (%)
RMSE reduction for u (%)
RMSE reduction for v (%)
RMSE reduction for v (%)

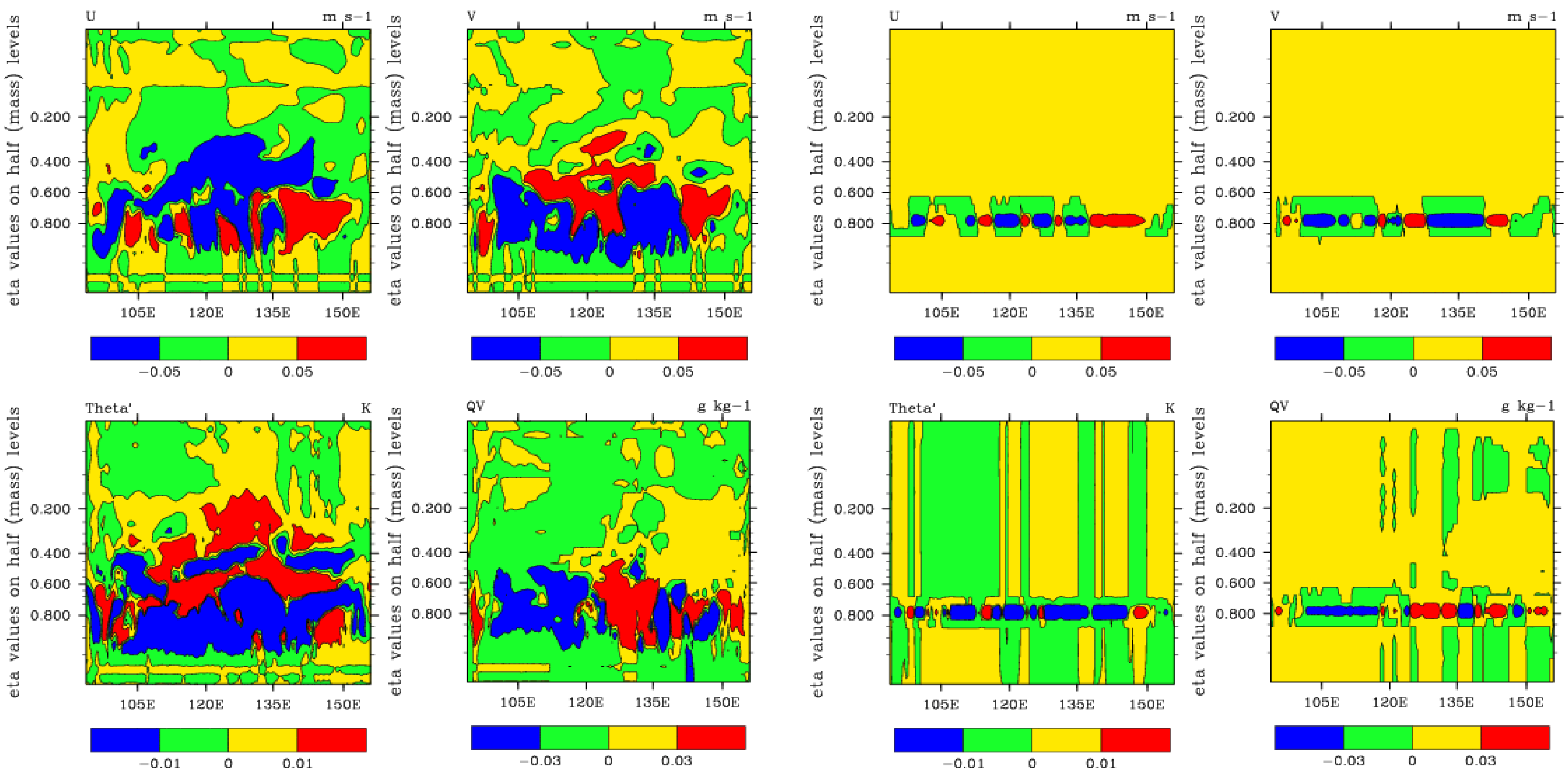
1/β2 = 0.2


a)
b)
c)
d)
Pressure (hPa)
RMSE reduction for u (%)
RMSE reduction for v (%)
S=250
S=500
S=1000
S=1500

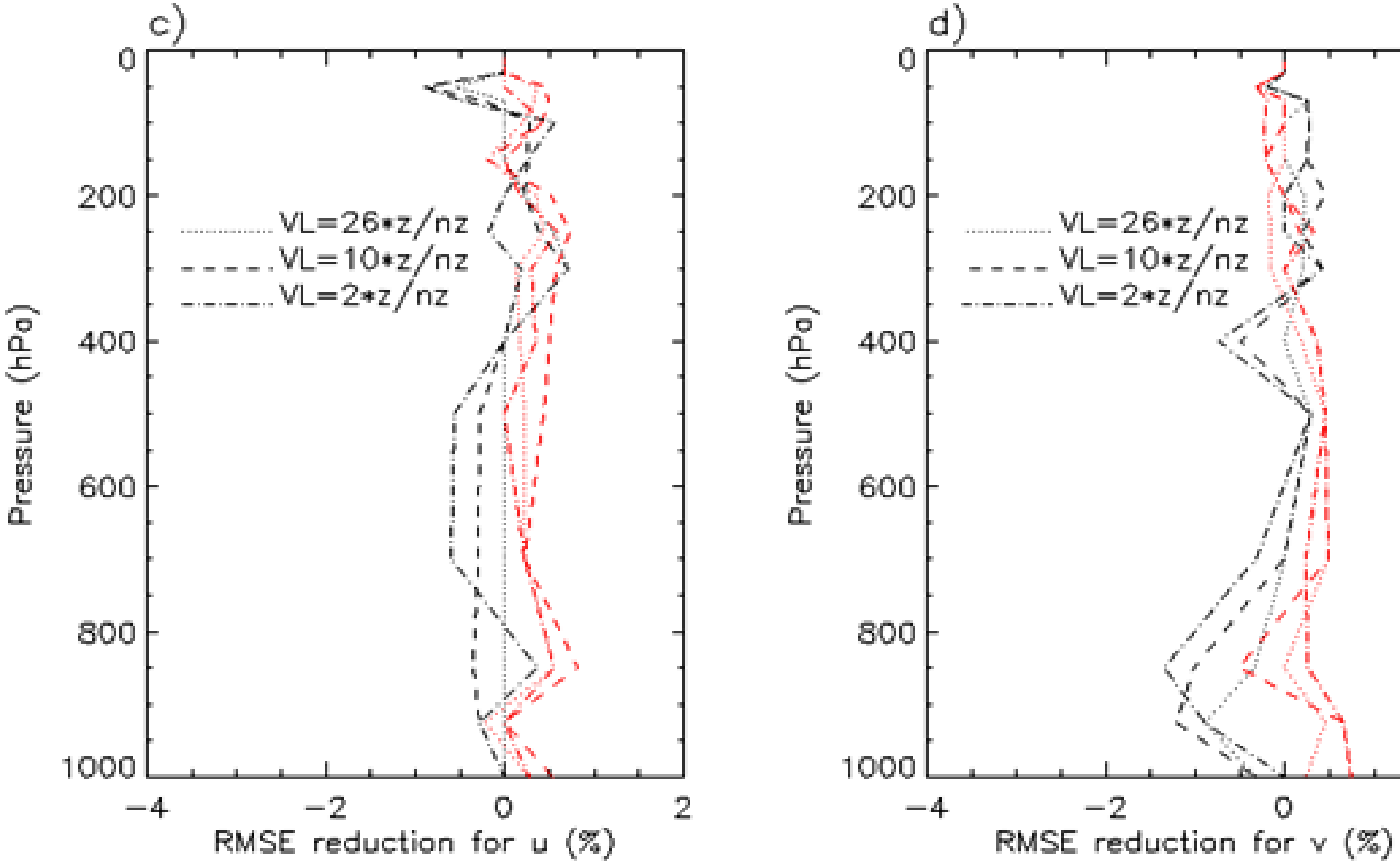
RESULTS: Hybrid 3DVAR-ETKF data assimilation

- Altitude-dependent vert. cov. localization

26 gp 2 gp



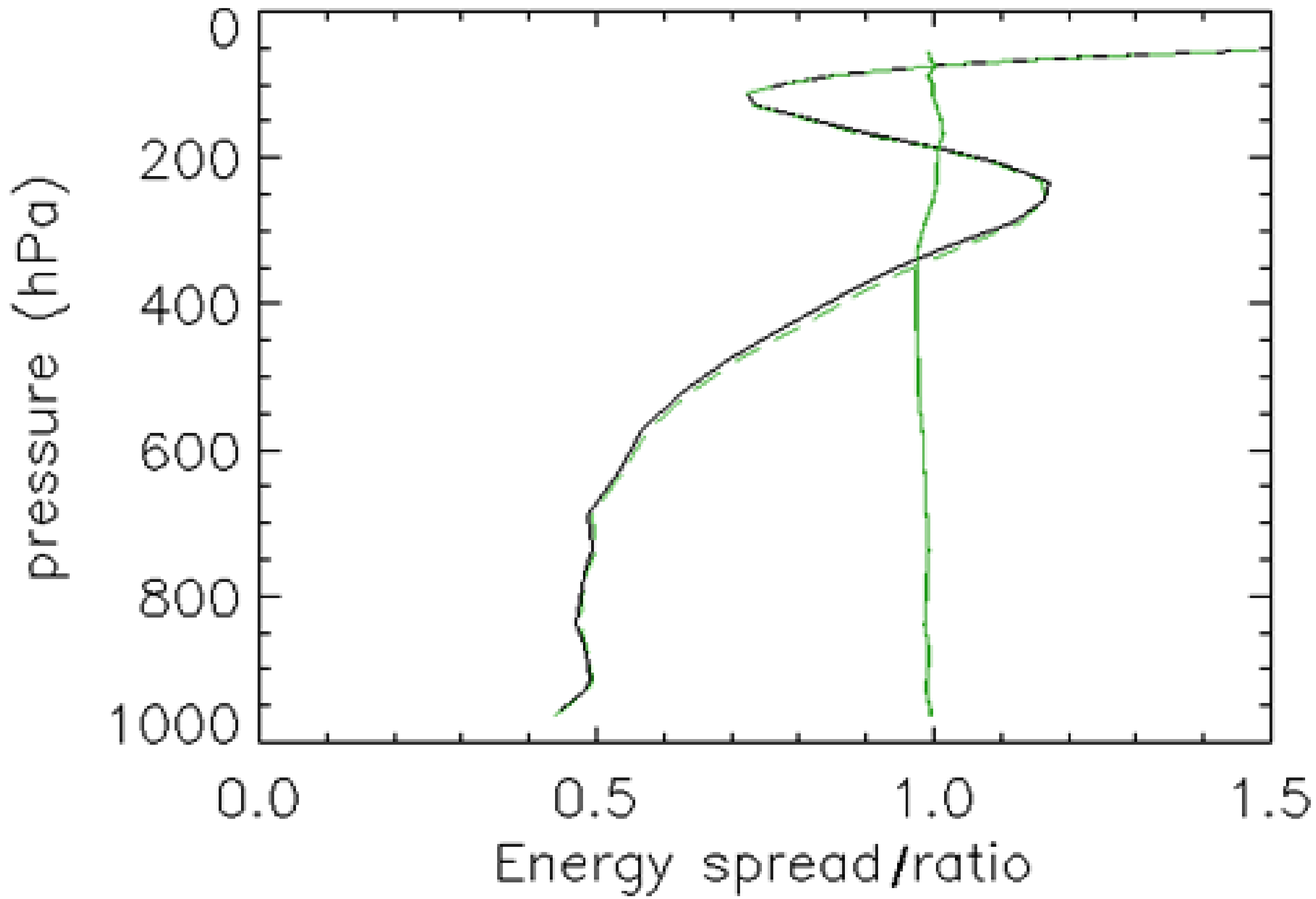
RESULTS: Hybrid 3DVAR-ETKF data assimilation



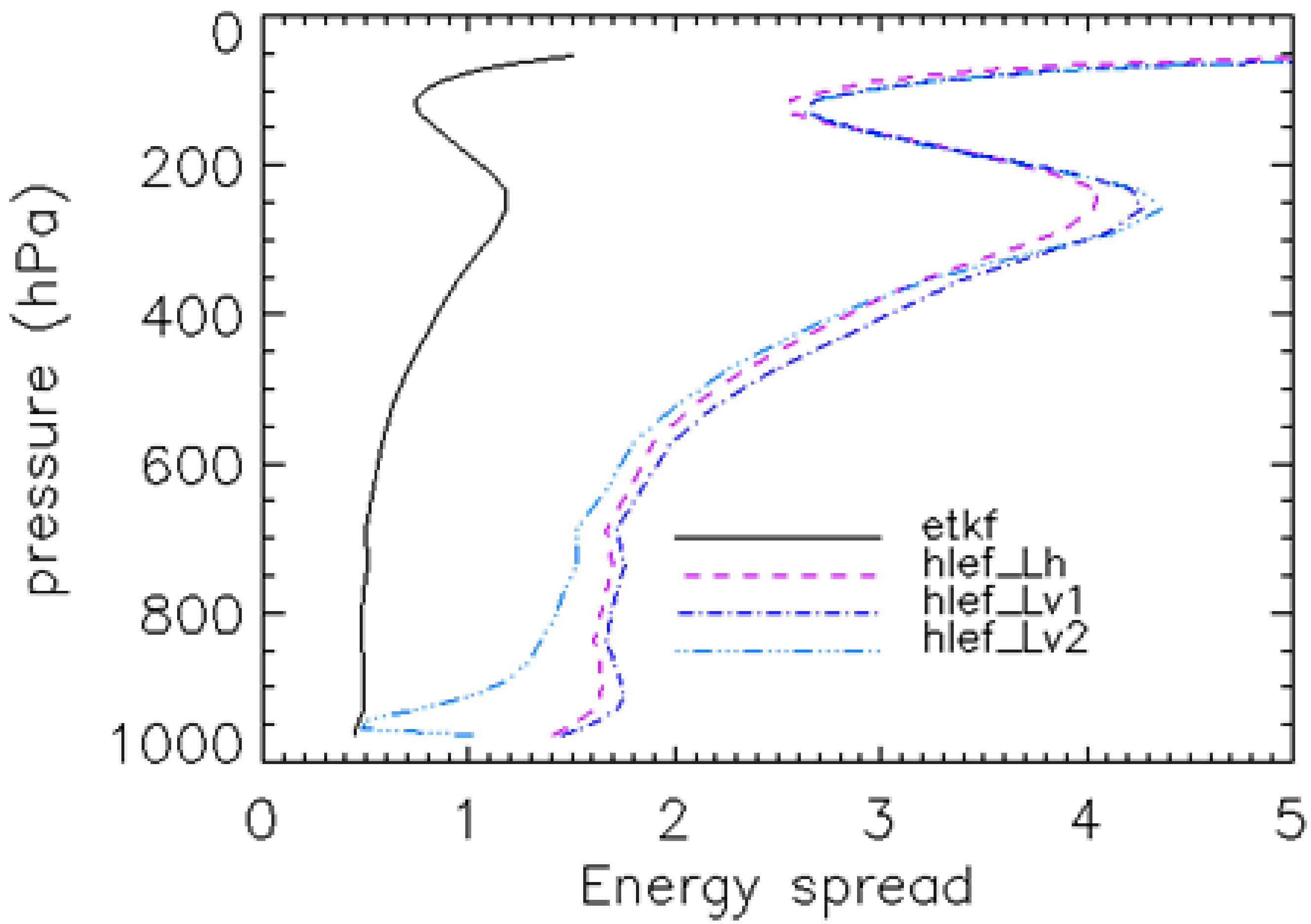
RESULTS: HLEF as an alternative to ETKF

- Theoretical equivalence
- Localization in HLEF perturbations
- Hybridizations in HLEF perturbations

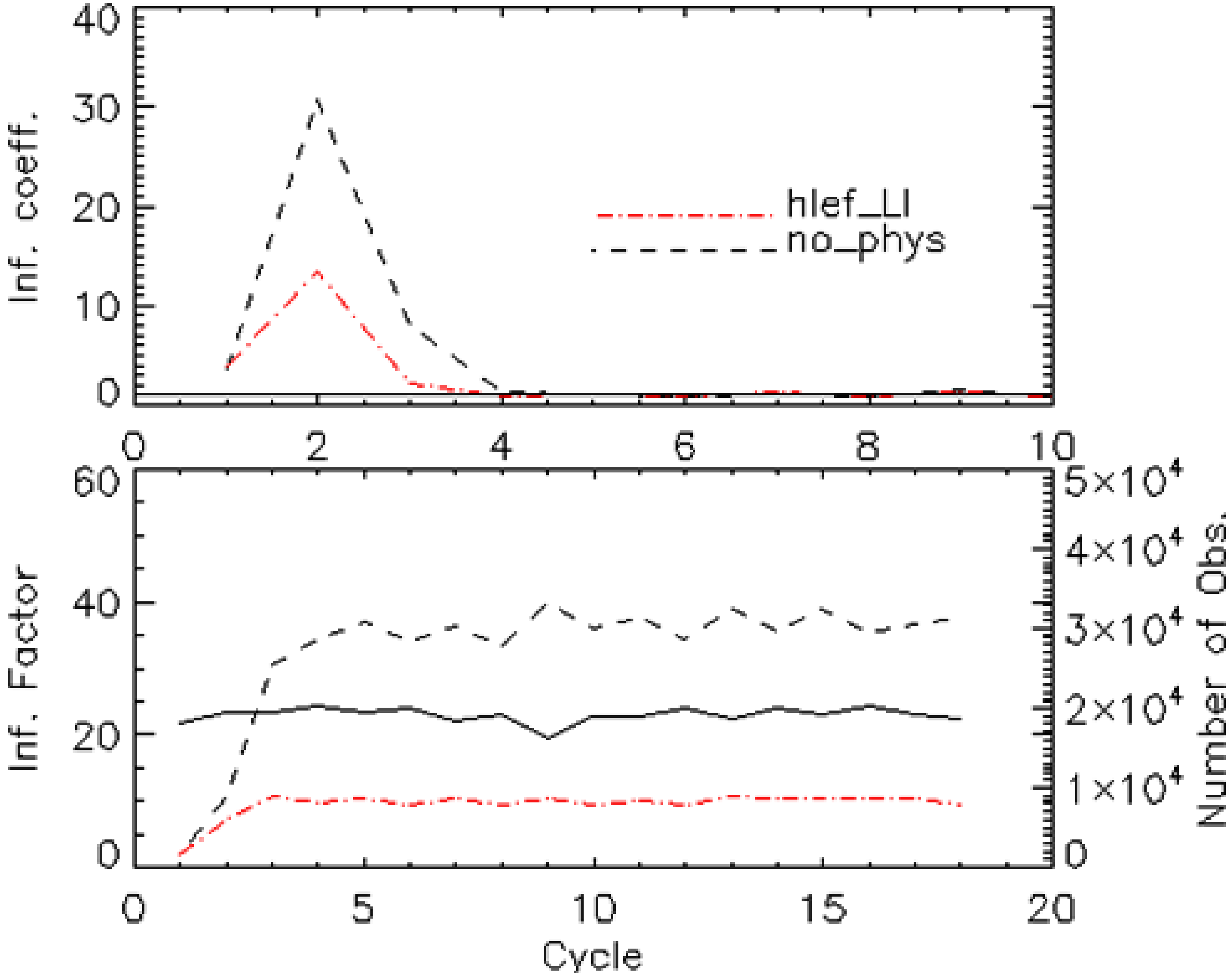
Theoretical equivalence



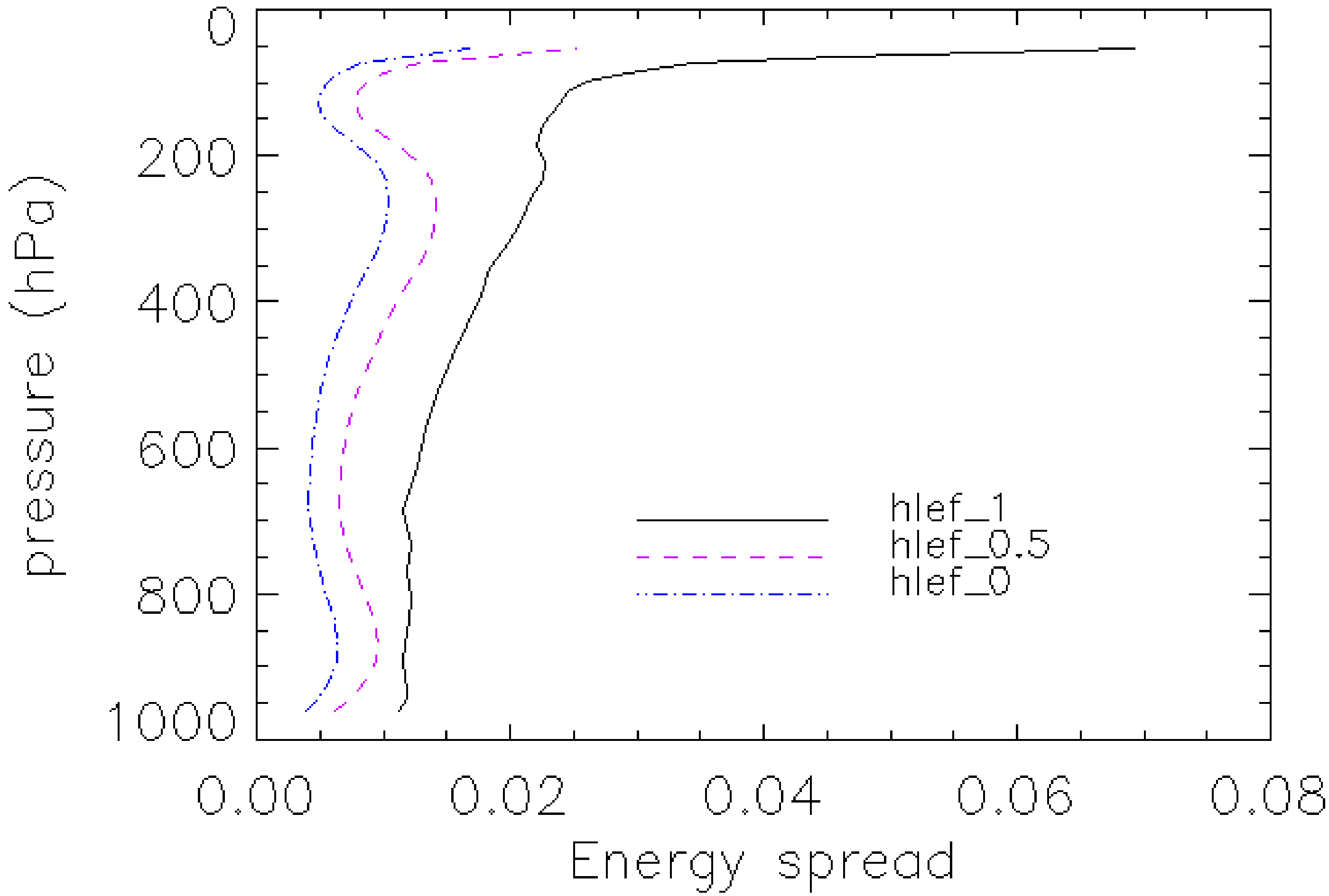
Localization in HLEF perturbations



Localization in HLEF perturbations



Hybridizations in HLEF perturbations



Summary and discussion

- The ETKF provides a sufficient (albeit slightly under-dispersive) ensemble
- Including multi-physics diversity:
  - reduces the ensemble's dependence on covariance inflation
  - resolves more innovation variance than the single-physics ensemble
  - Maintains variance more evenly in the available sub-space directions
- The multi-physics ensemble was characterized by larger error growth (at lower altitudes) than the single-physics ensemble
- Use of the ensemble mean as the first guess in the 3DVAR cost function significantly improved analysis skill

Summary and discussion

- Tuning **B** with the ETKF ensemble improved the skill of deterministic and ensemble 3DVAR
- Incorporating ensemble-based error covariances into the hybrid cost function added skill to the analysis
  - This was in addition to the skill added by using the ensemble mean as the first guess
  - Multi-physics ensemble covariances proved to be most beneficial
- Vertical covariance localization adds some skill to the analysis although more work is needed to determine an optimal localization criterion
  - The Rossby radius of deformation provides a baseline for future work

Summary and discussion

- HLEF was shown to be equivalent to ETKF
- HLEF improved upon ETKF through the effect of covariance localization
- The possibility of producing hybridized HLEF analysis perturbations and the hybrid analysis in a single step was shown

Thank you

Methodology

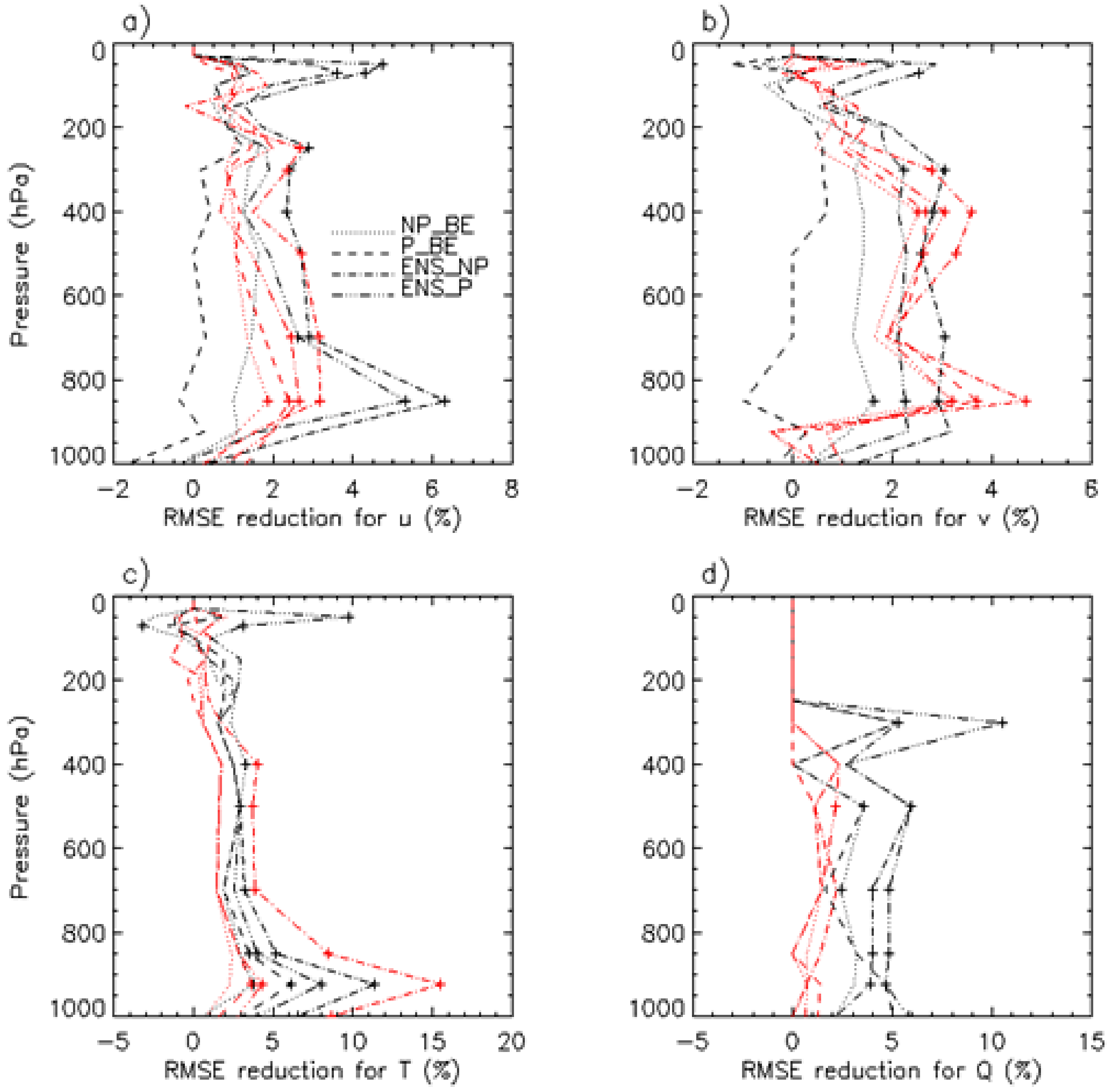
- Balance in the analysis
  - Ensemble mean is used as first guess (background)
  - Ensemble mean is not physically balanced
  - Analysis increments can cause imbalance
  - Covariance localization introduces imbalance
- Digital filter initialization used to balance each initial ensemble forecast

RESULTS: Ensemble mean as the first guess

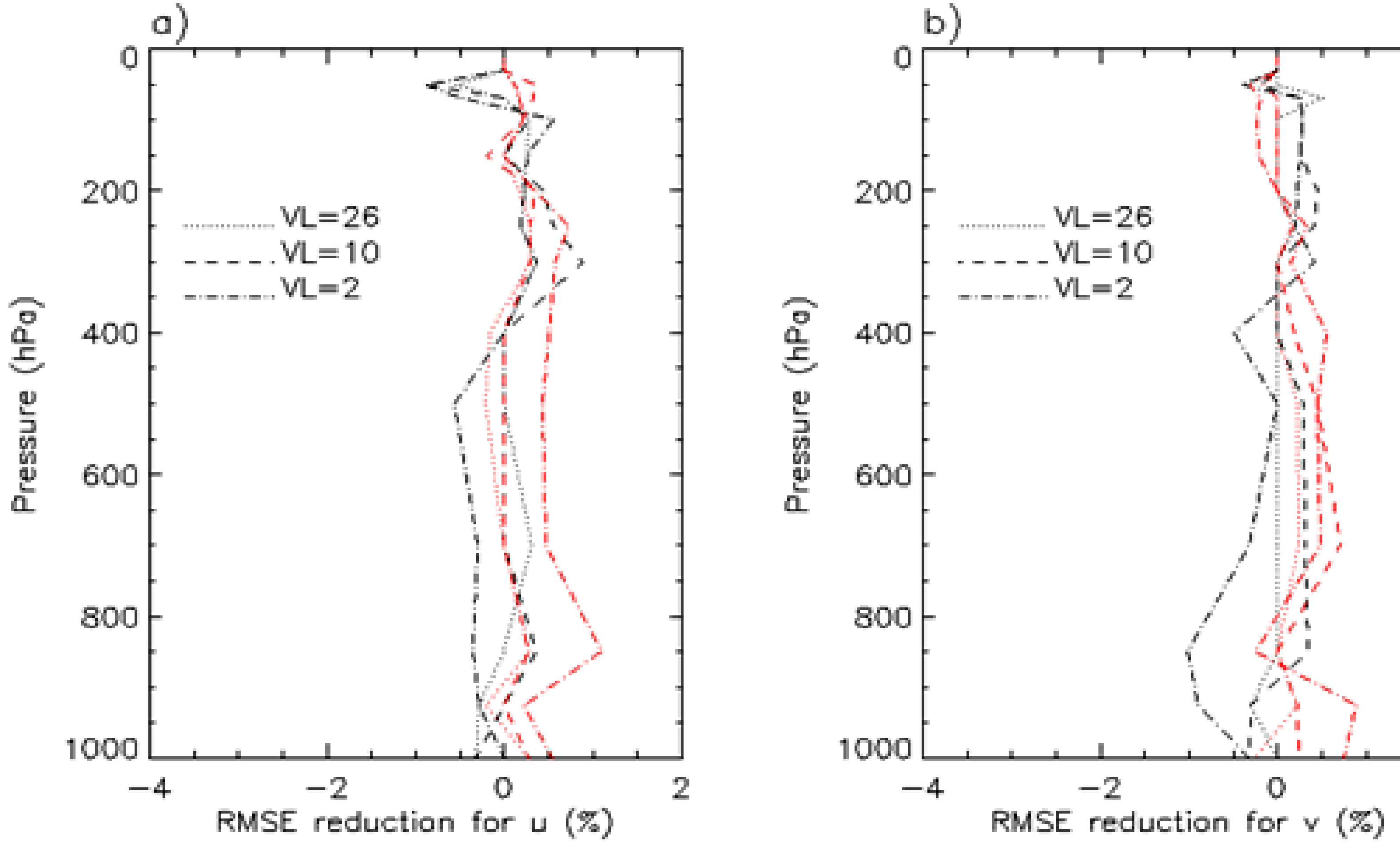
- Ensemble 3DVAR → Non-hybrid!!!
- 12- and 48-h deterministic forecasts from each analysis during the 1-month long cycling expts.
- T-test used to confirm differences are statistically significant (indicated by +)
- Rmse profiles of “rmse reduction” are shown (<0 indicates reduced skill and >0 indicated imporved skill)
- Reduction is based on 3DVAR with NMC-method rmse

48-h rmse

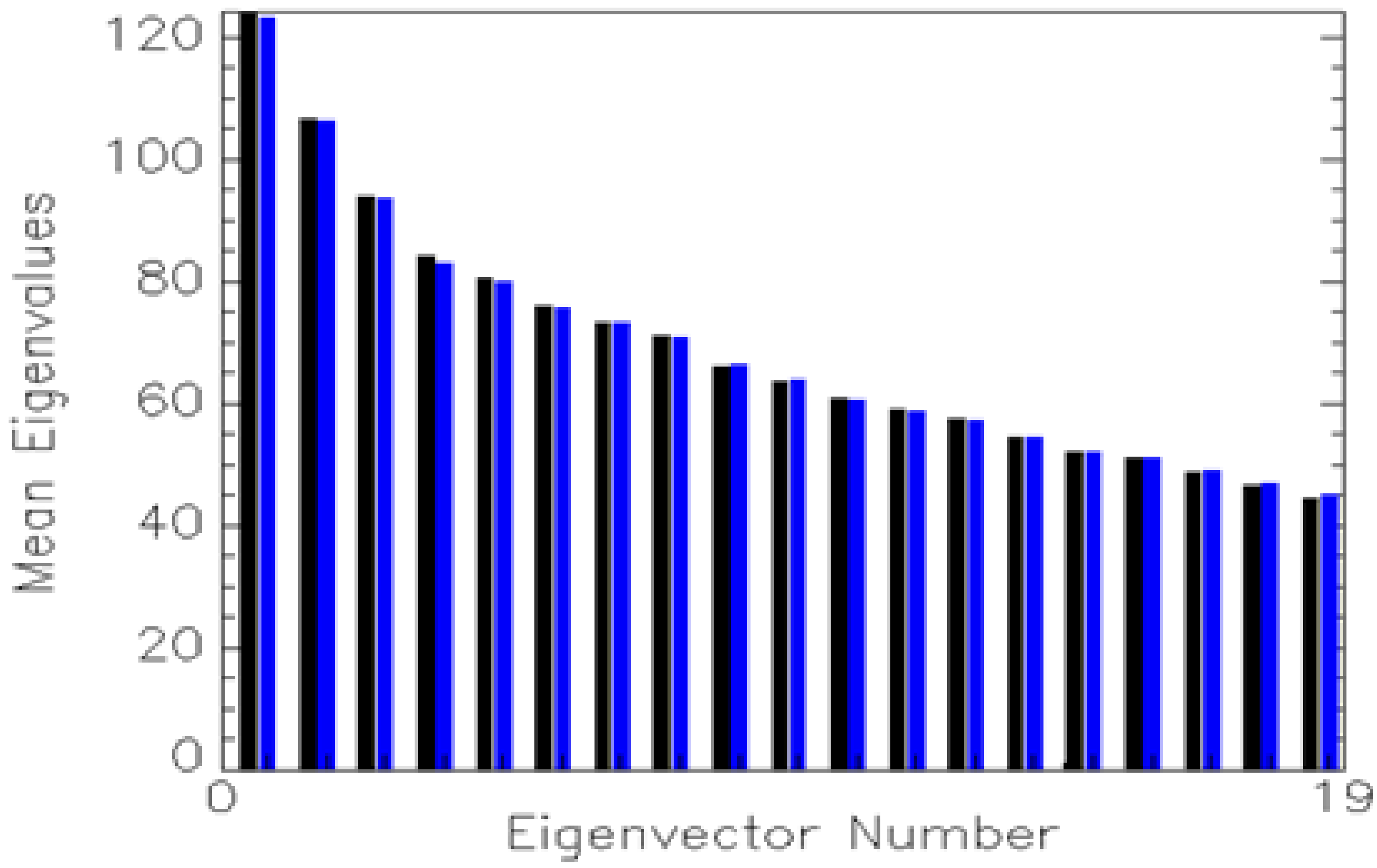
12-h rmse



RESULTS: Hybrid 3DVAR-ETKF data assimilation



Theoretical equivalence



Lanczos Algorithm

1) Supply first guess, x^g

2) Minimize cost function to find x^a

   a) generate sequence of orthogonal Lanczos vectors

   b) Lanczos iterations

   c) Determine eigenvalues and eigenvectors of the Lanczos tri-diagonal matrix

   d) check convergence criterion

   e) orthonormalize new gradient and continue Lanczos iteration if necessary.

3) Construct inverse Hessian by randomizing Lanczos tri-diagonal matrix

4) Add inverse Hessian elements (analysis perturbations) to analysis