An Ensemble Kalman Filter for NWP based on Variational Data Assimilation: VarEnKF

International Symposium on Data Assimilation
18-22 July 2016

Mark Buehner
Data Assimilation and Satellite Meteorology Research Section

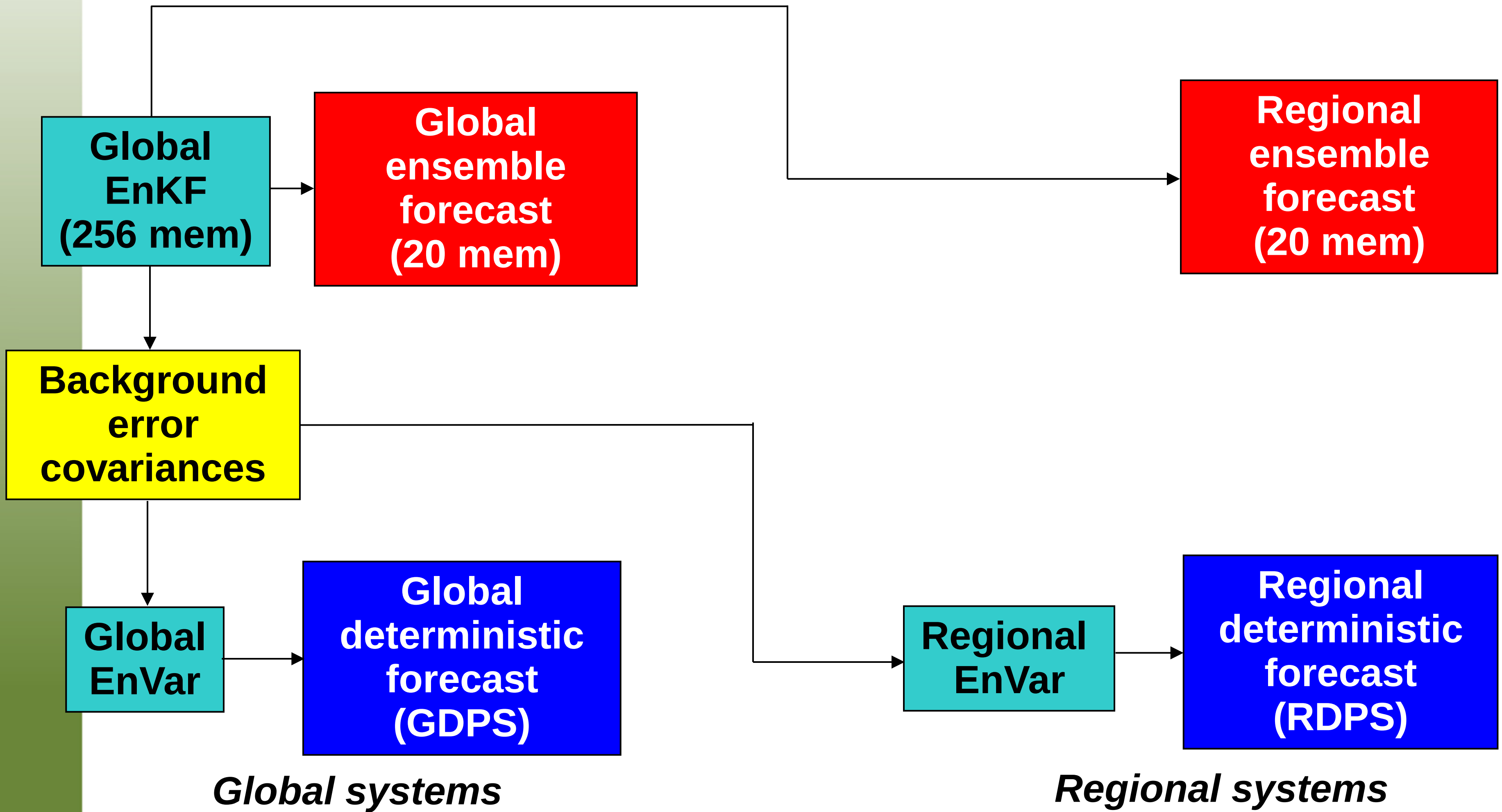
With contributions from Ron McTaggart-Cowan and Sylvain Heilliette

Context: EnVar Data Assimilation

- 4DEnVar replaced 4DVar for ECCC operational regional and global deterministic prediction systems Nov. 2014
- EnVar uses a **variational assimilation approach** with **4D ensemble covariances** from 256-member perturbed-obs EnKF
- Future improvements to the ensembles will benefit both ensemble and deterministic prediction systems
- EnKF and EnVar estimate covariances from ensembles in a similar way, but important differences are unavoidable (e.g. localization, use of hybrid covariances, VarQC)

Current Organization of the NWP Suites at Environment Canada

EnKF is an independent assimilation system, does not use the deterministic analysis state



Motivation to Explore Alternative Ensemble Data Assimilation Approaches

- Only small fraction of Fortran code shared between EnKF and EnVar – significant effort required to increase sharing
- Due to differences in EnKF and EnVar algorithms, changes to observations or covariances must be fully tested in both
- Due to computational cost, current EnKF algorithm limits the volume of observations (~40% of GDPS obs, no IR)
- EnVar uses model space **B** localization and can use hybrid covariances, variational QC, and scale-dependent localization (see talks by A. Shlyaeva tomorrow, J-F Caron on Friday)

EnKF based on Variational Approach: Benefits

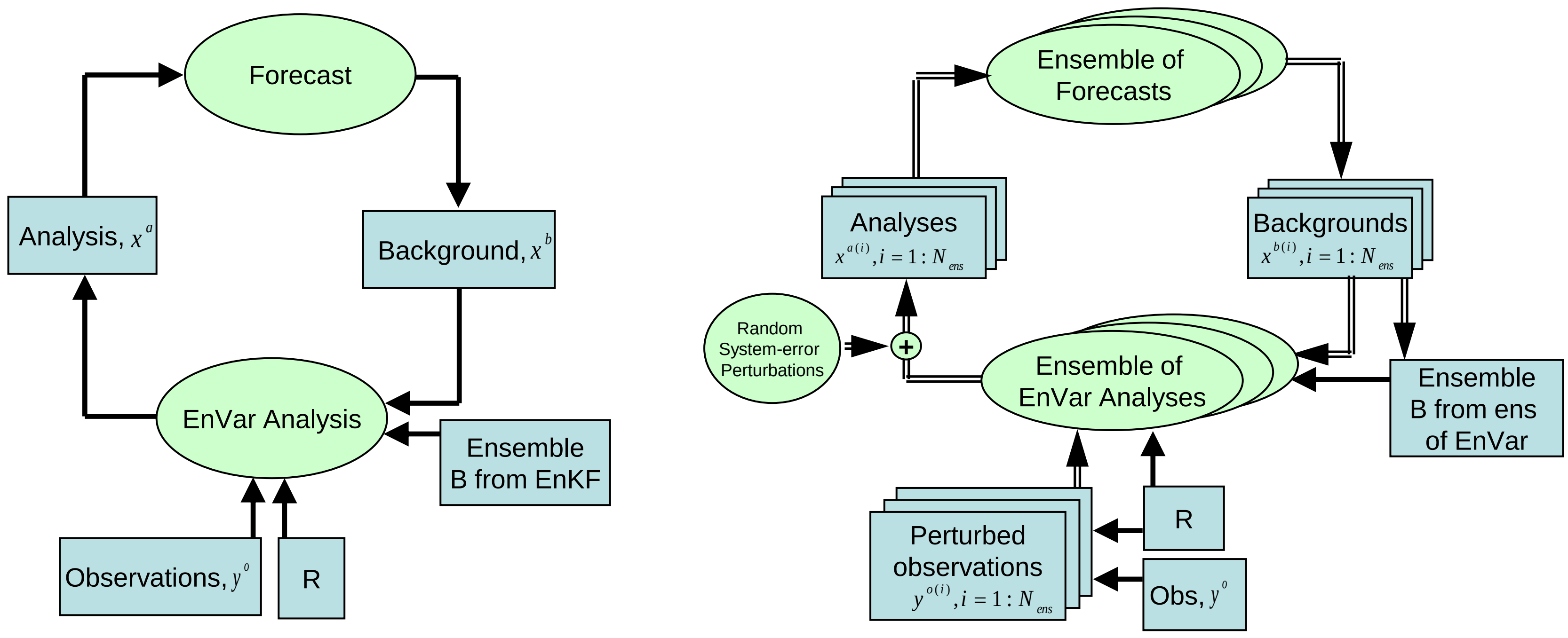
1. Reduce effort to maintain and improve systems (same unified code/algorithm for all systems)
2. Reduce amount of required testing (same assimilation algorithm and obs, therefore impact of changes more consistent for all systems)
3. Possibly improve quality of ensemble forecast (increased volume of assimilated obs and improved treatment of covariances)

Especially interesting for centers without existing EnKF

Ensemble of Variational Analyses (EDA)

Some centers using this (MetOffice, ECMWF, M-F)

- Most direct approach: ensemble of EnVar (or 4DVar) data assimilation cycles, each assimilating independently perturbed observations – **Very costly vs. EnKF**!



Deterministic EnVar

Ensemble of EnVars

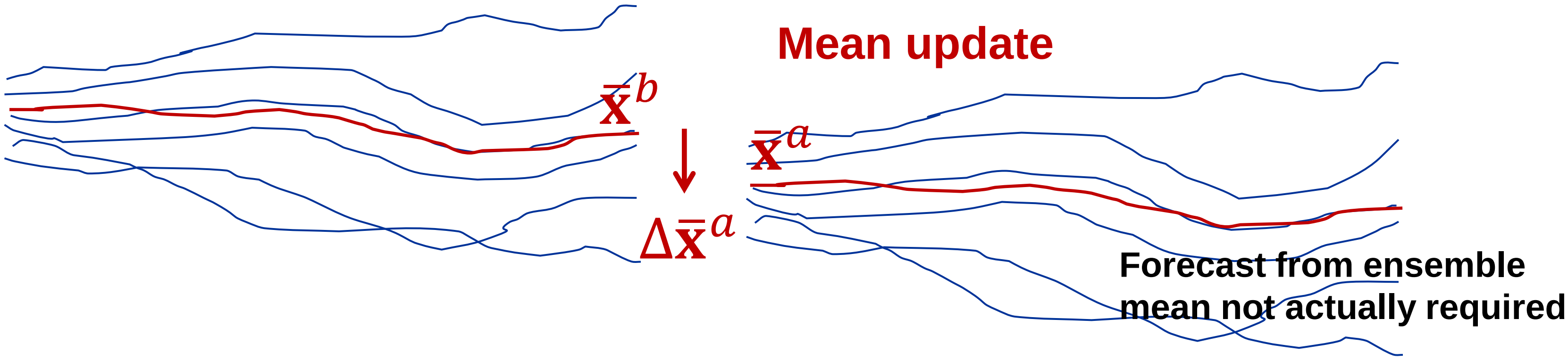
Separating Ensemble Mean and Perturbations

As suggested by Lorenc et al. (2016, submitted)

- To reduce computational cost, perform the analysis step separately for the ensemble mean and ensemble perturbations and simplify the problem for perturbations:

$$\mathbf{x}_k^a = \bar{\mathbf{x}}^a + \mathbf{x}_k^{a\prime}, \qquad \mathbf{x}_k^b = \bar{\mathbf{x}}^b + \mathbf{x}_k^{b\prime}$$

$$\Delta\mathbf{x}_k^a = \Delta\bar{\mathbf{x}}^a + \Delta\mathbf{x}_k^{a\prime}$$



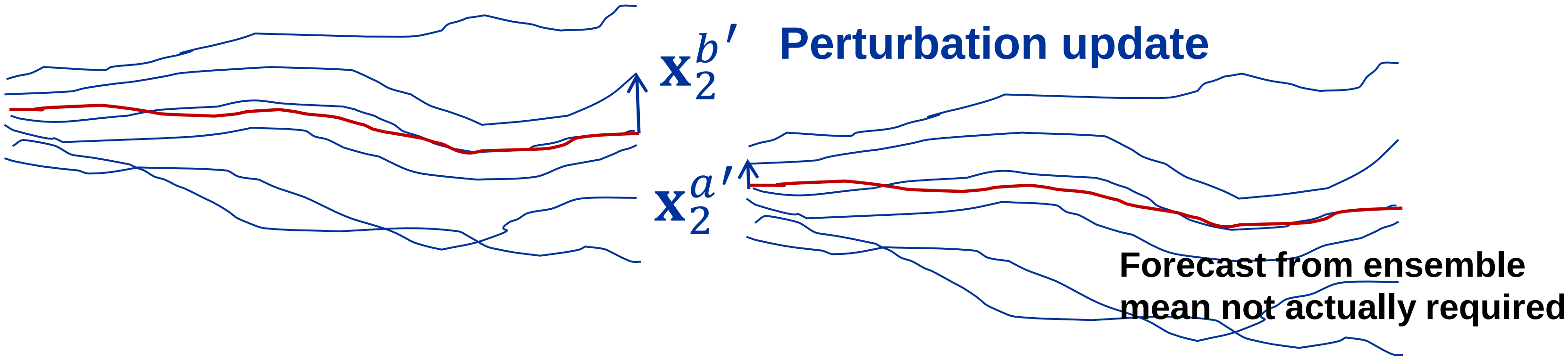
Separating Ensemble Mean and Perturbations

As suggested by Lorenc et al. (2016, submitted)

- To reduce computational cost, perform the analysis step separately for the ensemble mean and ensemble perturbations and simplify the problem for perturbations:

$$\mathbf{x}_k^a = \bar{\mathbf{x}}^a + \mathbf{x}_k^{a\prime}, \qquad \mathbf{x}_k^b = \bar{\mathbf{x}}^b + \mathbf{x}_k^{b\prime}$$

$$\Delta\mathbf{x}_k^a = \Delta\bar{\mathbf{x}}^a + \Delta\mathbf{x}_k^{a\prime}$$



EnVar for the Mean Analysis Update

- **<u>For ensemble mean</u>**: full 4DEnVar analysis using **ensemble mean background state** and **unperturbed observations** to compute increment to ensemble mean:

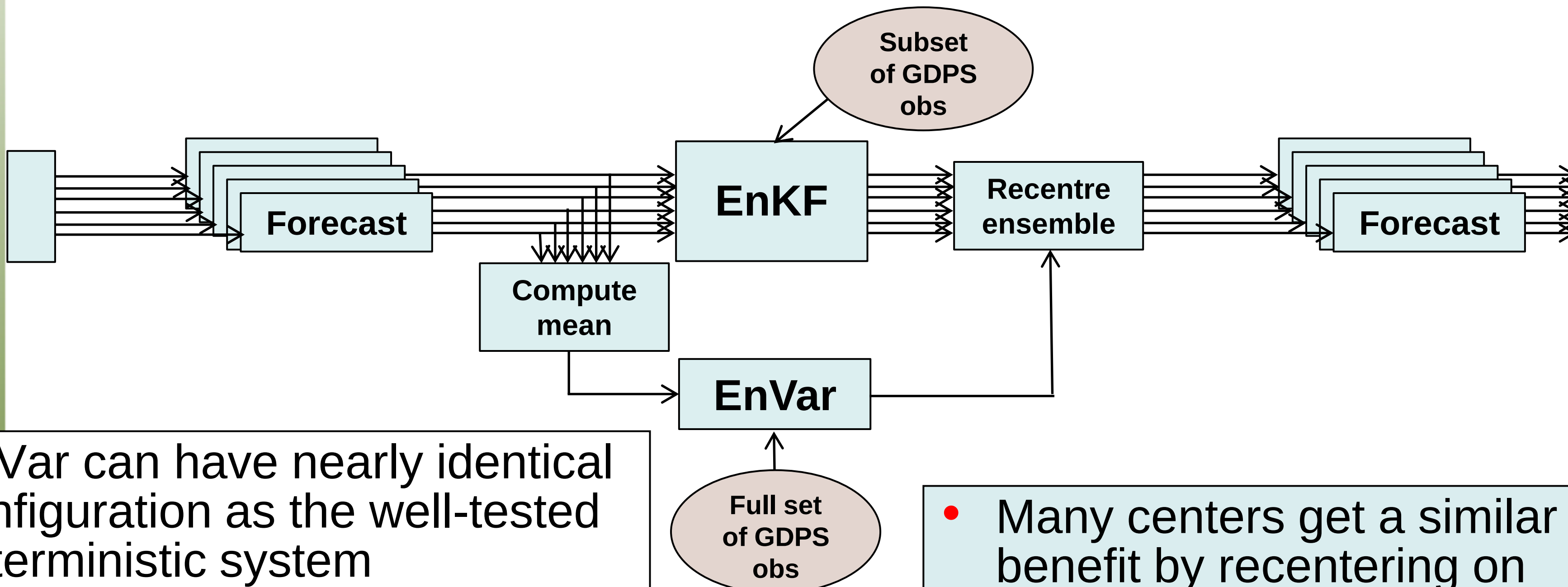
$$J(\Delta\bar{\mathbf{x}}^a) = \frac{1}{2}(\Delta\bar{\mathbf{x}}^a)^{\mathrm{T}}\mathbf{B}^{-1}(\Delta\bar{\mathbf{x}}^a) + \frac{1}{2}\left(\mathbf{y}^o - \boldsymbol{H}(\bar{\mathbf{x}}^b) - \mathbf{H}\Delta\bar{\mathbf{x}}^a\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}^o - \boldsymbol{H}(\bar{\mathbf{x}}^b)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

- Variational approach highly efficient for single analysis

EnVar for <u>only</u> the Mean Analysis Update

- Little added cost to use **EnVar to update the ensemble mean** and the **EnKF to update the perturbations** (*k* is member index)*:*

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \Delta\bar{\mathbf{x}}^a{}_{\text{envar}} + \Delta\mathbf{x}_k^{a\prime}{}_{\text{enkf}}$$



- EnVar can have nearly identical configuration as the well-tested deterministic system

- Many centers get a similar benefit by recentering on deterministic analysis

Control member forecast results showing impact of using EnVar for only ensemble mean

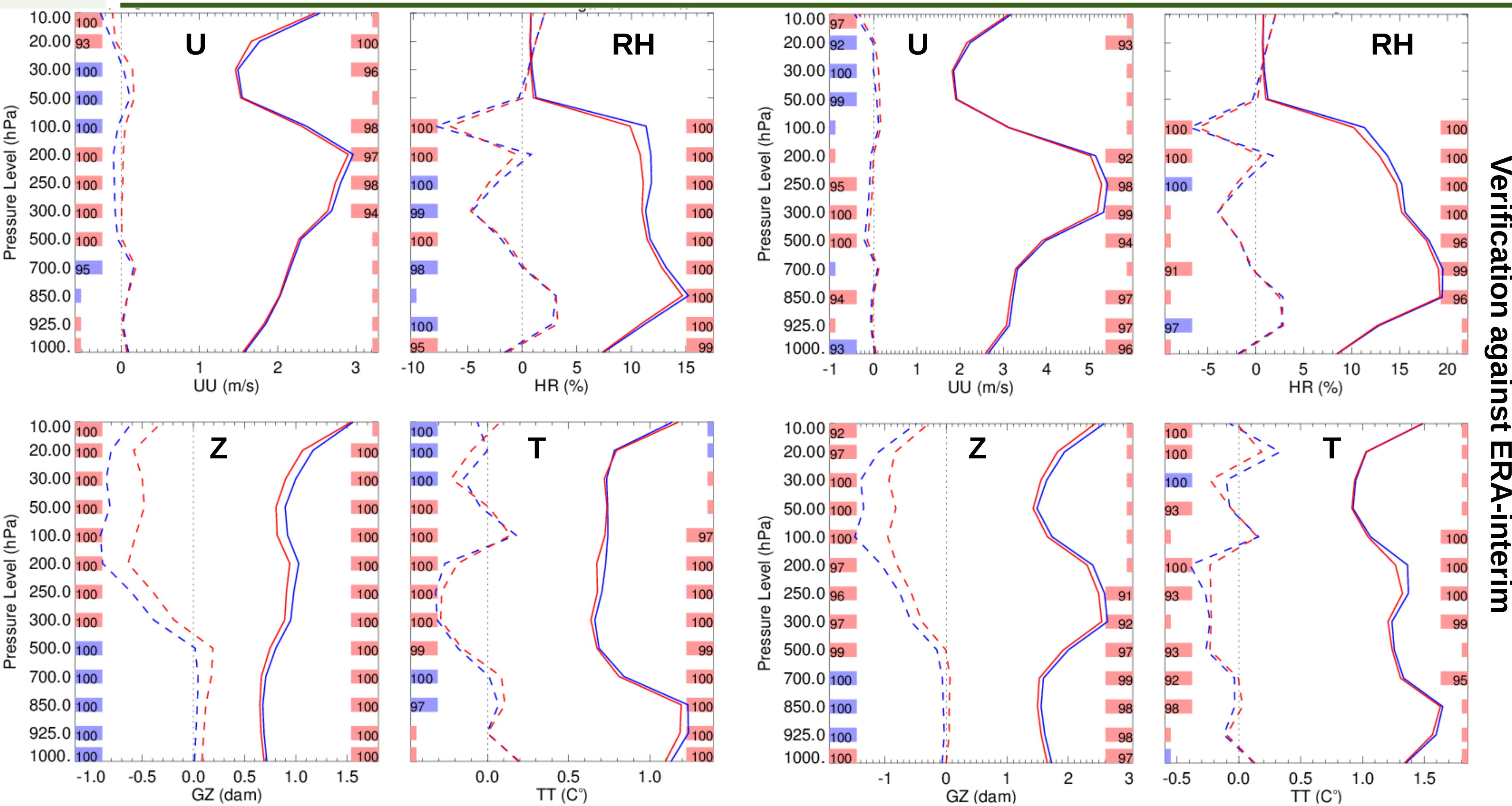
- **Ensemble mean: EnVar with full set of GDPS obs* vs. Current EnKF**
- **Ensemble perturbations: Both use current EnKF**
- **Experiments cover 3 January – 15 January 2015 (26 forecasts)**

***Ensemble mean update also includes non-zero inter-channel obs-error correlations, hybrid background-error covariances (10% B_{nmc} + 90% B_{ens}), and variational QC**

Results: EnVar vs. EnKF for ensemble mean

Control member forecasts (deterministic forecast from mean analysis)

24h global forecasts **72h global forecasts**



Using EnVar with all GDPS obs to only update the ensemble mean gives significant improvements for control member vs. Current EnKF

Using the Var System to Update Perturbations

Var efficient for single analysis, not really suited to ensembles, so…

- **<u>For perturbations</u>**: use much cheaper, less precise approach to compute increment to all 256 ensemble perturbations $\Delta\mathbf{x}_k^{a\prime} = \mathbf{x}_k^{a\prime} - \mathbf{x}_k^{b\prime}$

$$\Delta\mathbf{x}_k^{a\prime} = \mathbf{K}\left(\mathbf{r}_k - \mathbf{H}\mathbf{x}_k^{b\prime}\right), \text{ where } \mathbf{r}_k \sim N(\mathbf{0}, \mathbf{R})$$

- Perform minimization with simplifications to efficiently solve equivalent variational problem for each perturbation analysis:

$$J(\Delta\mathbf{x}_k{}') = \frac{1}{2}(\Delta\mathbf{x}_k{}')^{\mathrm{T}}\mathbf{B}^{-1}(\Delta\mathbf{x}_k{}') + \frac{1}{2}\left(\mathbf{r}_k - \mathbf{H}\mathbf{x}_k^{b\prime} - \mathbf{H}\Delta\mathbf{x}_k{}'\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{r}_k - \mathbf{H}\mathbf{x}_k^{b\prime} - \mathbf{H}\Delta\mathbf{x}_k{}'\right)$$

Variational Minimization to Update Perturbations

- Need way to speed up minimization for each perturbation:
  - Reduce number of iterations
  - Simplify **B** matrix used (3D instead of 4D, fewer ensemble members, or use only climatological **B** matrix)
  - Reduce spatial resolution of analysis increment
  - Reduce quantity of observations assimilated
- In the context of ensemble prediction, evidence that computing full perturbations with simple approaches works surprisingly well vs. sophisticated approaches:
  - Magnusson et al. (2009): ECMWF system
  - Raynaud and Bouttier (2015): Météo-France AROME
- Simplification to variational approach for perturbation increments should be less extreme than these

Simplified Configuration for Perturbations

- Initial test uses the following (extreme) simplifications:
  - Only climatological **B** matrix with reduced resolution (3DVar)
  - Reduced quantity of observations: no AIRS, IASI, CRIS, SSMIS, GeoRad (also not used in current EnKF)
  - **Same number of iterations** as deterministic system (70)

- The simplified **B** matrix and reduced volume of observations decrease the memory requirements and execution time

- Reduction in size of problem allows many jobs to be run in parallel:
  - **1 Perturbation update has 2.5% the cost of full 4DEnVar!**
  - 256 members takes ~24min wall clock on 2048 processors
  - Trivial to parallelize further (up to 256 jobs, each taking ~1min)

Impact of using 3DVar minimization vs. current EnKF for perturbation updates

- Ensemble mean: Both use EnVar with full set of GDPS observations and same configuration
- Ensemble perturbations: 3DVar vs. current EnKF each assimilating similar subset of observations
- **Experiments cover 3 January – 15 January 2015 (26 forecasts)**
- Using simplified 3DVar certainly NOT expected to be better than EnKF, but is it significantly worse?

Results: 3DVar vs. current EnKF for perturbations

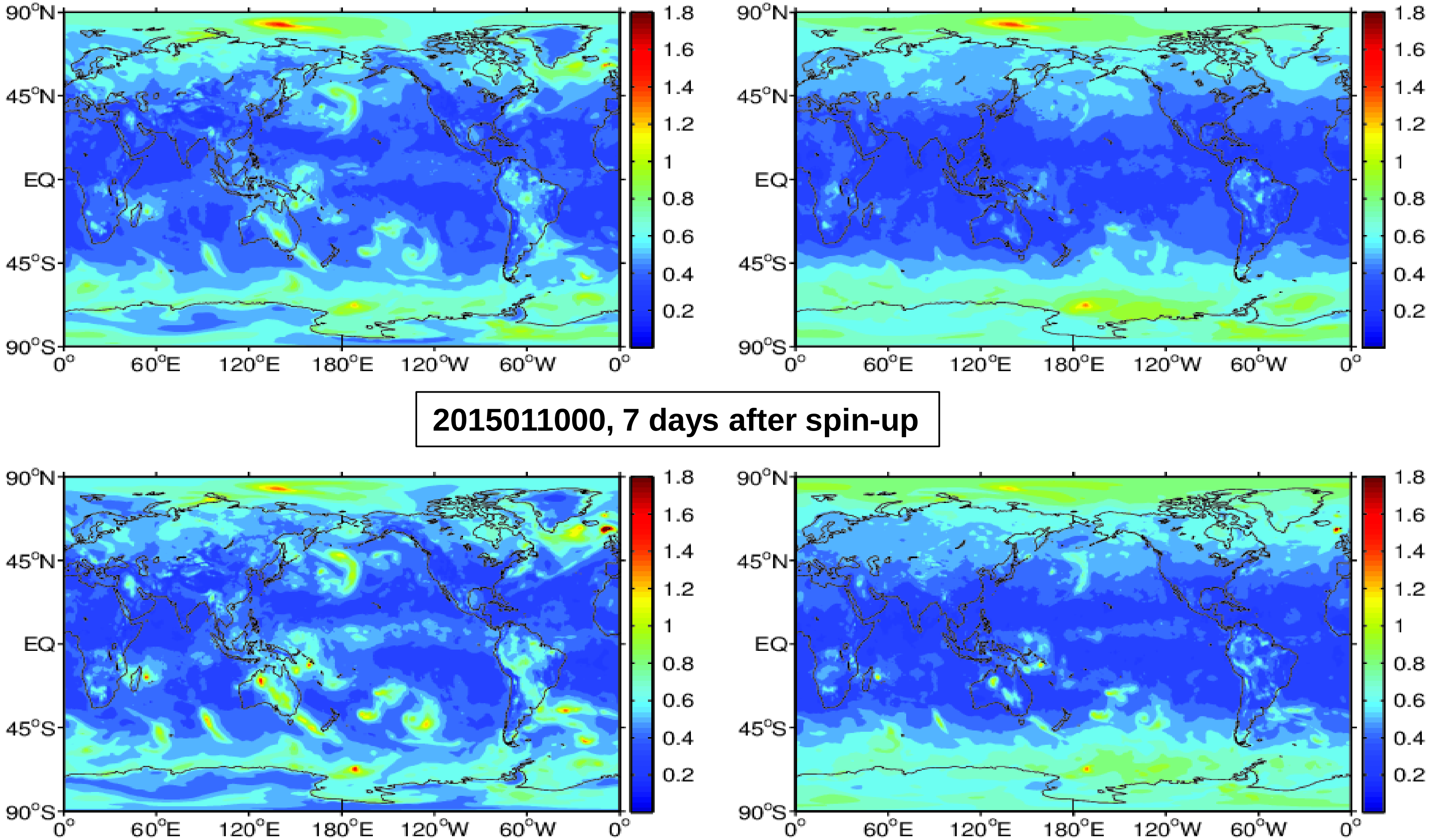
Perturbation increments computed with:

Background ensemble

Analysis ensemble
(after adding system error perturbations)

Current EnKF

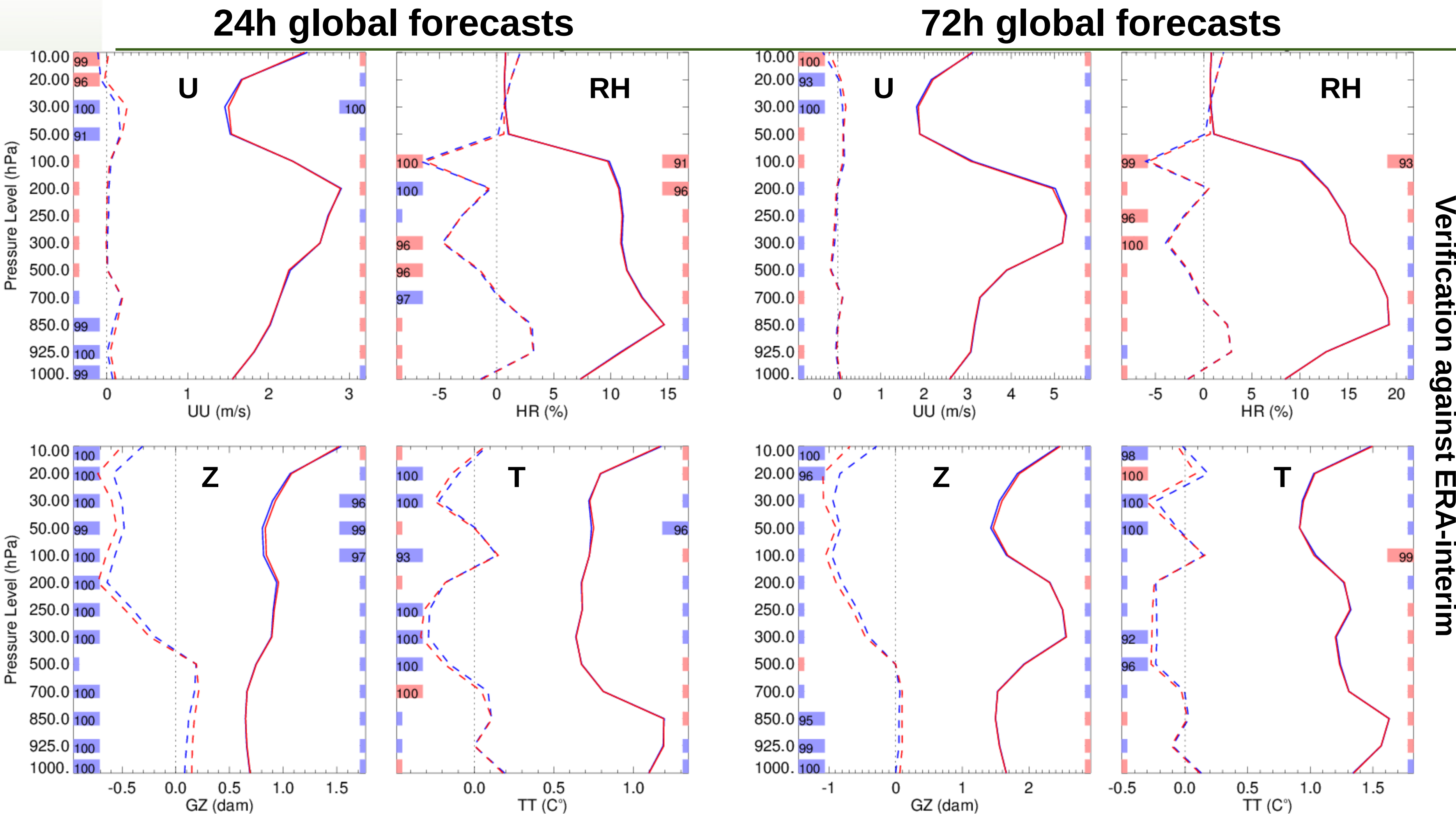
3DVar



Ensemble spread for Psfc (hPa) – both experiments use EnVar for mean

Results: 3DVar vs. current EnKF for perturbations

Control member forecasts (deterministic forecast from mean analysis)



Using 3DVar with reduced set of obs for perturbations nearly equivalent to Current EnKF for perturbations (both use EnVar for ens. mean)

Impact of VarEnKF* vs. current EnKF

- Ensemble mean: EnVar with full set of GDPS obs* vs. current EnKF

- Ensemble perturbations: 3DVar vs. current EnKF

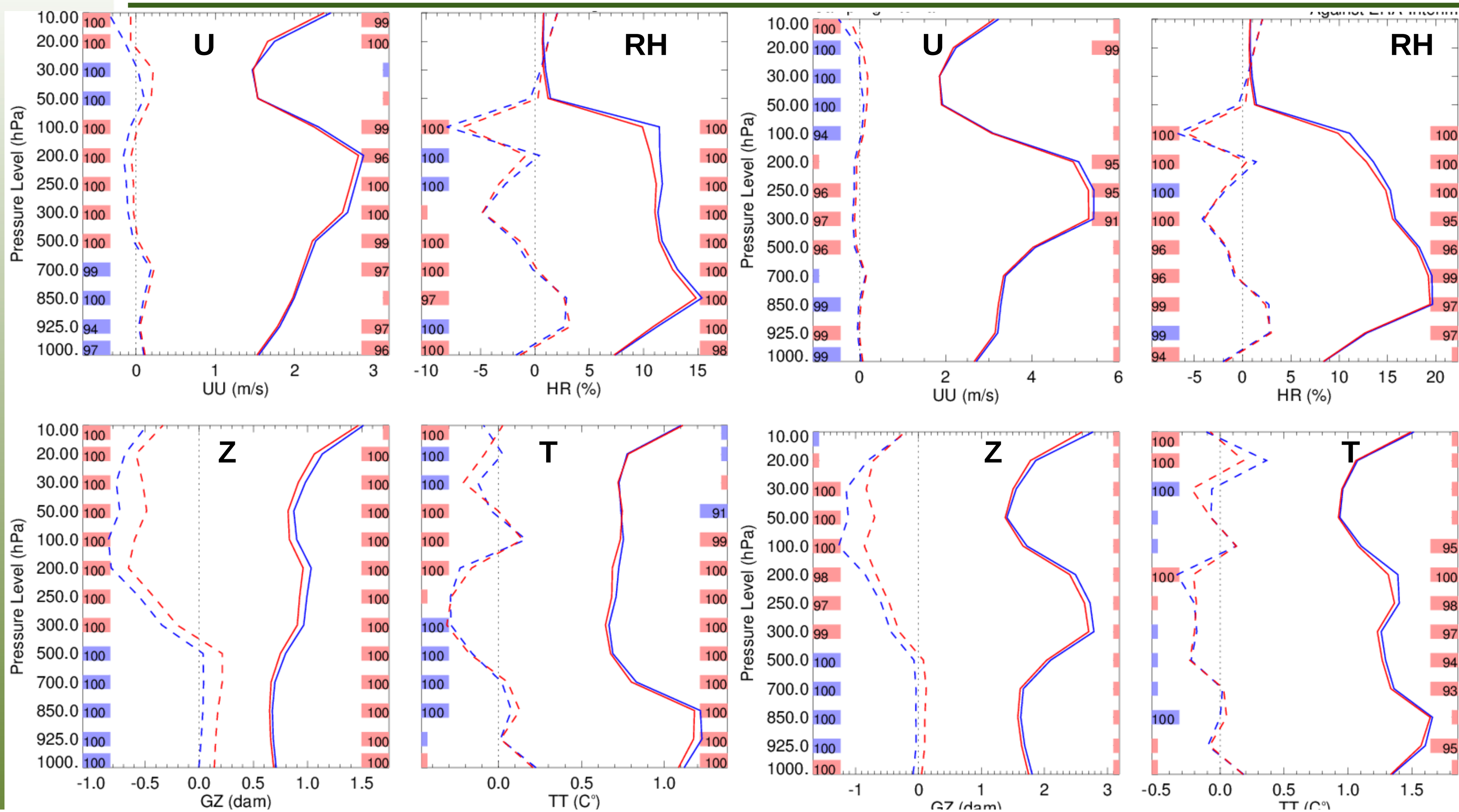
- Experiments cover 3 January – 28 January 2015 (52 forecasts)

*Ensemble mean update also includes non-zero inter-channel obs-error correlations and hybrid background-error covariances (10% B_{nmc} + 90% B_{ens})

Results: VarEnKF vs. current EnKF

Control member forecasts (deterministic forecast from mean analysis)

24h global forecasts **72h global forecasts**

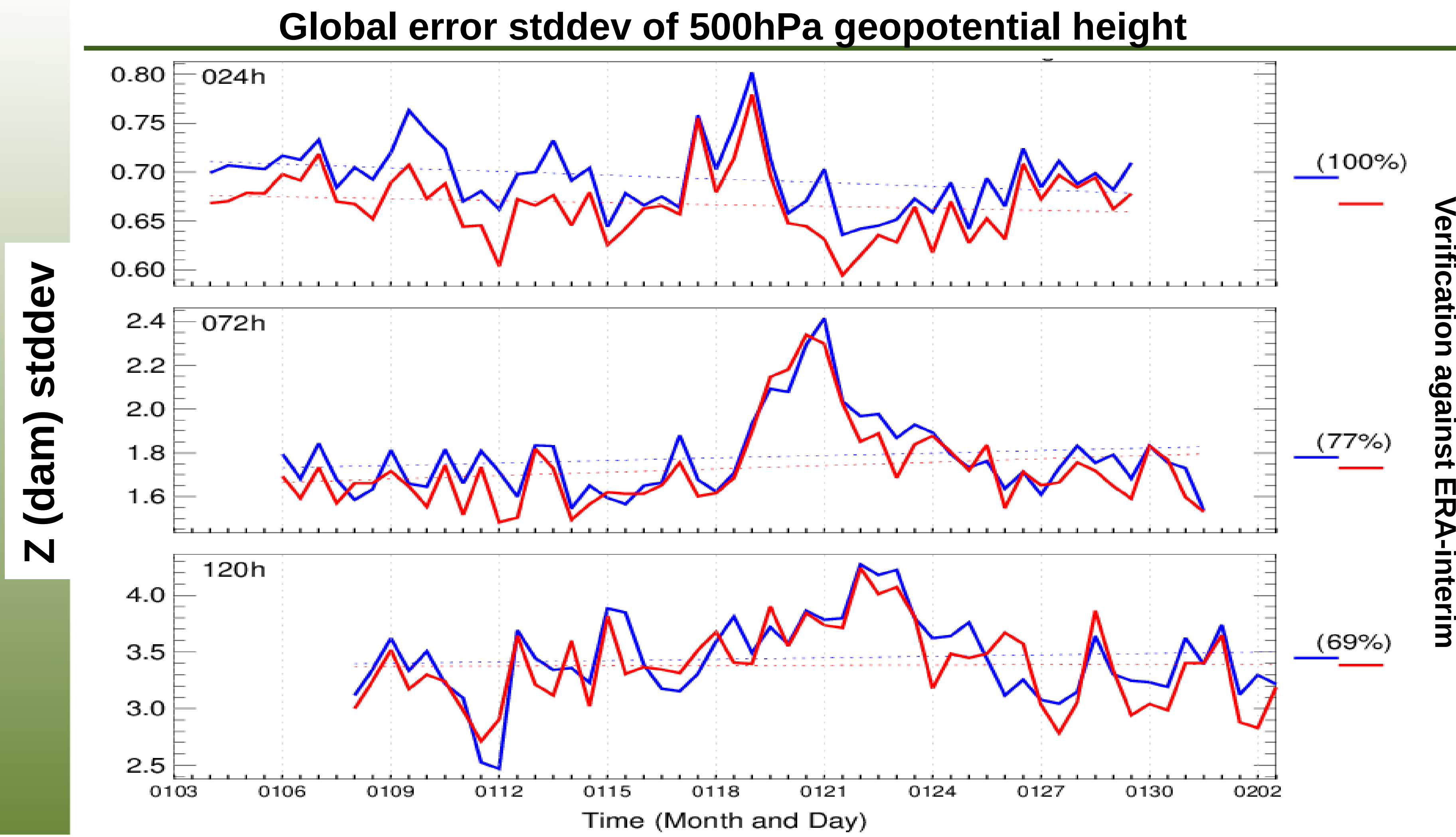


Verification against ERA-interim

Using 4DEnVar and full set of obs for mean and 3DVar for perturbations (VarEnKF) gives significant improvement vs. using current EnKF

Results: VarEnKF vs. current EnKF

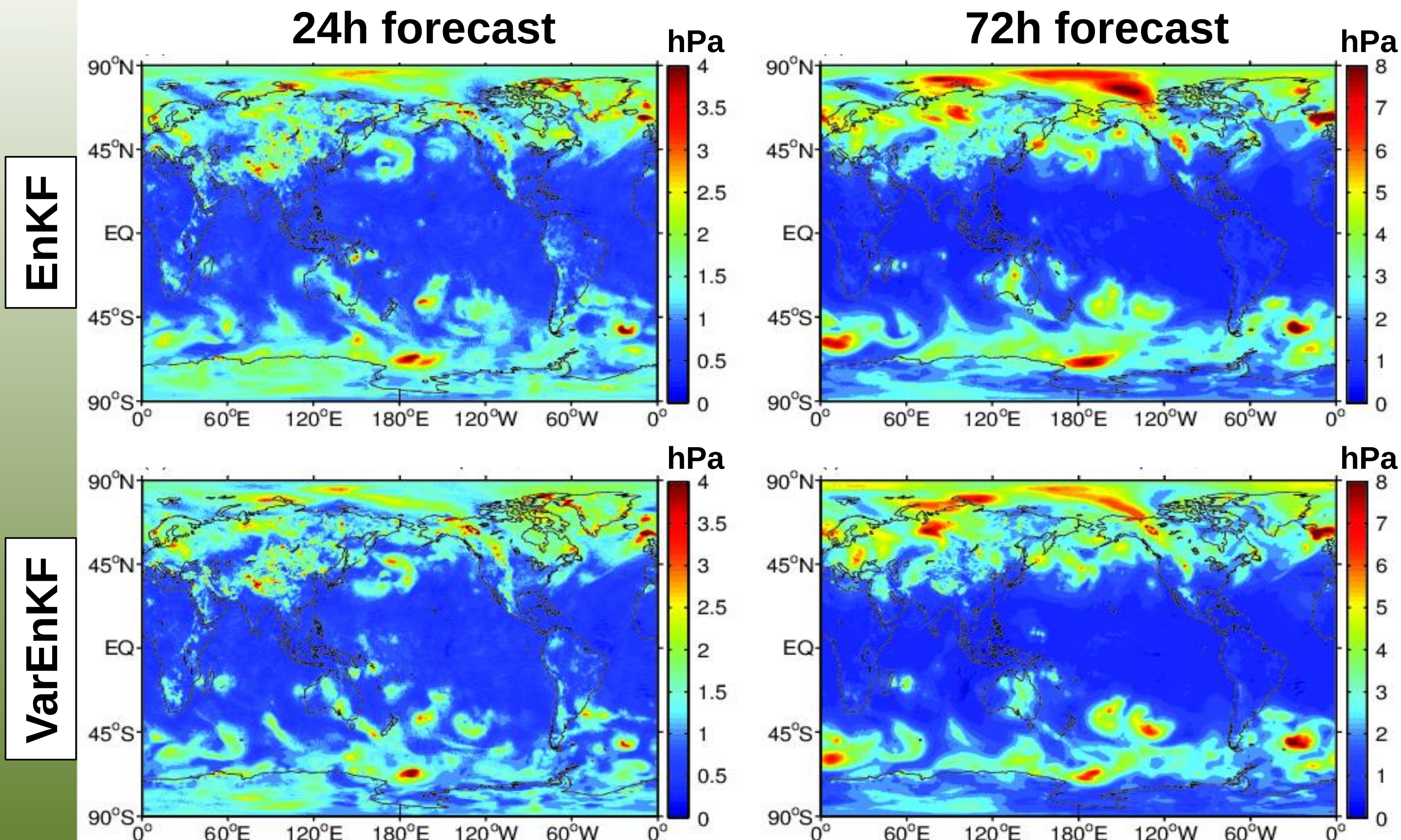
Control member forecasts (deterministic forecast from mean analysis)



Using 4DEnVar and full set of obs for mean and 3DVar for perturbations (VarEnKF) gives significant improvement vs. using current EnKF

Results: Spread for VarEnKF vs. current EnKF

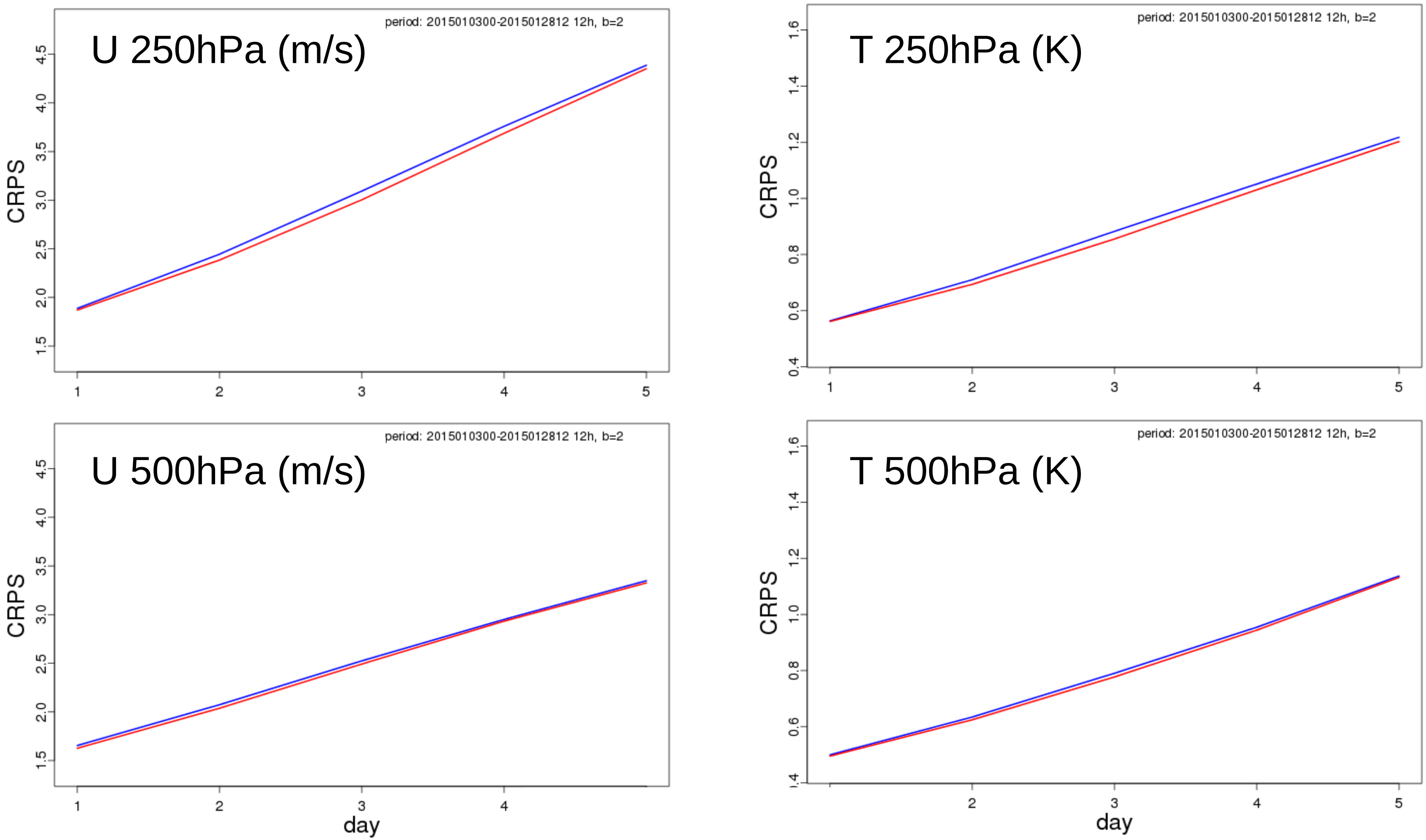
Standard deviation of surface pressure ensemble spread from 20-member ensemble forecasts on 0UTC, 10 January 2015



VarEnKF gives similar ensemble spread vs. current EnKF

Results: CRPS for VarEnKF vs. current EnKF

Continuous Ranked Probability Score measures accuracy of ensemble pdf relative to observations (radiosonde)



VarEnKF gives improved CRPS vs. current EnKF

Conclusions

- Ensemble mean: using 4DEnVar, with nearly identical configuration as deterministic system, leads to improved ensemble forecasts
  - EnVar assimilates higher volume of observations than EnKF
  - Uses non-zero inter-channel observation-error correlations, hybrid background-error covariances (10% / 90%) and variational QC
- Ensemble perturbations: severe simplifications to variational assimilation resulting in cost ~2.5% of deterministic system
  - Results in similar forecast quality relative to using current EnKF!
  - Are there better simplifications? Should test using ensemble σ_b
  - Main motivation is to eliminate need to maintain two DA algorithms/codes
- **Therefore, more efficient to dedicate development and computing resources to ensemble mean update than perturbation update**
  - How general is this? Still holds for more rapid cycling, higher resolution?
  - What other data assimilation algorithms can be adapted to take advantage of this? e.g. EDA: 4DVar for mean, 3DVar for perturbations

Future plans

- Evaluate impact of simply using existing deterministic analysis to recenter ensemble – consistent with other centres and probably get most of the improvement
- Evaluate VarEnKF approach in context of new higher-resolution regional DA system over only Canada (currently no operational deterministic or ensemble DA for this resolution/domain)