Environment and Climate Change Canada | Environnement et Changement climatique Canada

The Big Leap: Replacing 4D-Var with 4D-EnVar and life ever since

Symposium: 20 years of 4D-Var at ECMWF

26 January 2018

Mark Buehner[1], Jean-Francois Caron[1] and Ping Du[2]

[1]Data Assimilation and Satellite Meteorology Research Section

[2]Data Assimilation and Quality Control Development Section

Contents

1. Why would someone consider replacing 4D-Var with anything else?
2. Brief overview of current/future operational NWP systems at ECCC
3. Scale-dependent ensemble covariance localization applied to global NWP (Caron and Buehner 2017, MWR submitted)
4. A new "hybrid" approach for estimating the impact of observations – Forecast Sensitivity Observation Impact (FSOI) (Buehner et al. 2017, MWR EOR)

Background

- Before November 2014, Environment Canada had 2 relatively independent state-of-the-art DA systems
- 4D-Var (Gauthier et al 2007) and EnKF (Houtekamer et al 2009):
  - both operational since 2005
  - both use GEM forecast model and assimilate obs
- 4D-Var used to initialize global deterministic forecasts
- EnKF is used to initialize global ensemble forecasts
- Can the EnKF be used to satisfy all assimilation needs?
- Intercomparison of approaches in carefully controlled context: similar forecast quality from EnKF and 4D-Var, 4D-Var with B_{ens} better (Buehner et al 2010, MWR)

Ensemble-Variational assimilation: EnVar

- 4D-EnVar uses a **variational assimilation approach** in combination with the already available **4D ensemble background-error covariances** from the EnKF
- By making use of the 4D ensembles, 4D-EnVar performs a 4D analysis without the need of the tangent-linear and adjoint of the forecast model
- Hybrid covariances are used in 4D-EnVar by averaging the ensemble covariances with the static climatological covariances
- Currently, our EnKF has 256 members, assimilates perturbed observations, and uses no recentering

4D-EnVar as an alternative to 4D-Var

The thinking around 2012

- Overall, 4D-EnVar analysis ~6X faster than 4D-Var on half as many cpus, and higher resolution increments
- Wall-clock time of 4D-Var was close to allowable time limit and model TL/AD did not scale well
- To progress with 4D-Var, significant work required to improve scalability at resolutions used in 4D-Var
- Decision made to try to replace 4D-Var with more efficient 4D-EnVar → if 4D-EnVar is at least as good as **current** 4D-Var
- Decided to take the risk of replacing 4D-Var and focus efforts on improving the ensemble and its use in the EnKF and 4D-EnVar

EnVar formulation

- In 4D-Var the 3D analysis increment is evolved in time using the TL/AD forecast model (here included in $\mathbf{H}_{4D}$):

$$J(\Delta\mathbf{x}) = \frac{1}{2}(H_{4D}[\mathbf{x}_b] + \mathbf{H}_{4D}\Delta\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1}(H_{4D}[\mathbf{x}_b] + \mathbf{H}_{4D}\Delta\mathbf{x} - \mathbf{y}) + \frac{1}{2}\Delta\mathbf{x}^T \mathbf{B}^{-1}\Delta\mathbf{x}$$

- In EnVar the background-error covariances and analysed state are explicitly 4-dimensional, resulting in cost function:

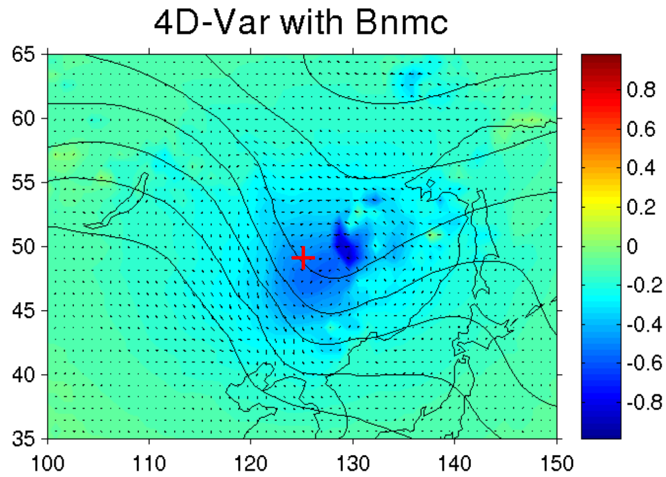
$$J(\Delta\mathbf{x}_{4D}) = \frac{1}{2}(H_{4D}[\mathbf{x}_b] + \mathbf{H}\Delta\mathbf{x}_{4D} - \mathbf{y})^T \mathbf{R}^{-1}(H_{4D}[\mathbf{x}_b] + \mathbf{H}\Delta\mathbf{x}_{4D} - \mathbf{y}) + \frac{1}{2}\Delta\mathbf{x}_{4D}{}^T \mathbf{B}_{4D}^{-1}\Delta\mathbf{x}_{4D}$$

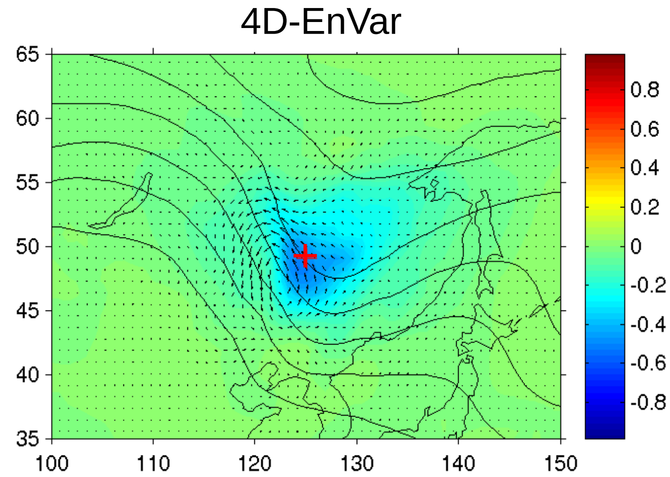
- Computations involving ensemble-based $\mathbf{B}_{4D}$ can be more expensive than with $\mathbf{B}_{nmc}$, but can be easily parallelized

Single observation experiments

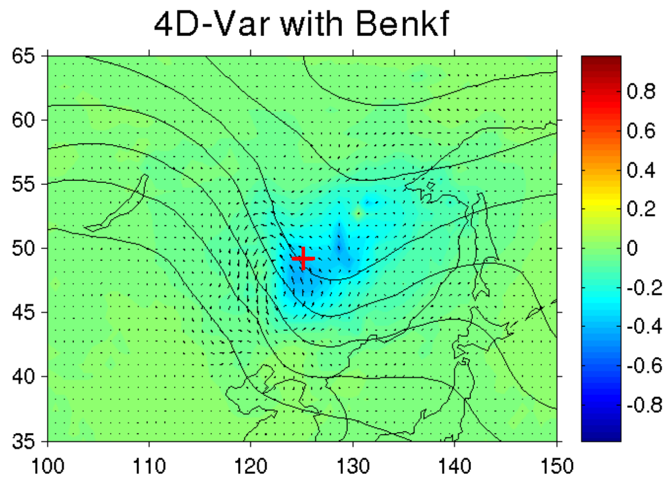
Difference in temporal covariance evolution

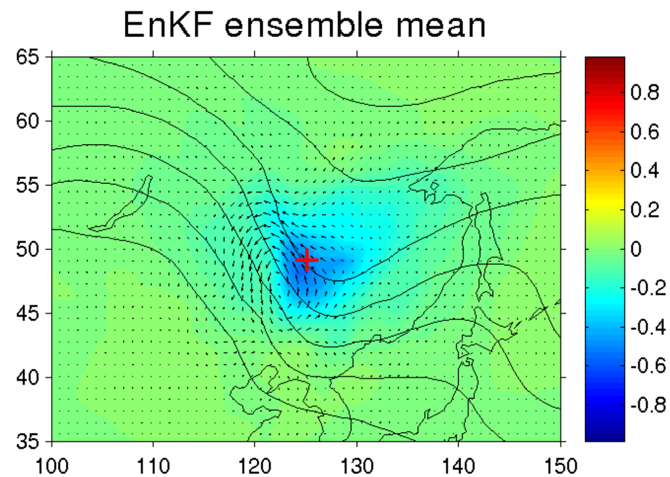
- radiosonde temperature observation at 500hPa
- observation at **beginning of assimilation window (-3h)**
- with same **B**, increments very similar from **4D-Var**, **EnKF**
- contours are 500hPa GZ background state at 0h (ci=10m)





contour plots at 500 hPa

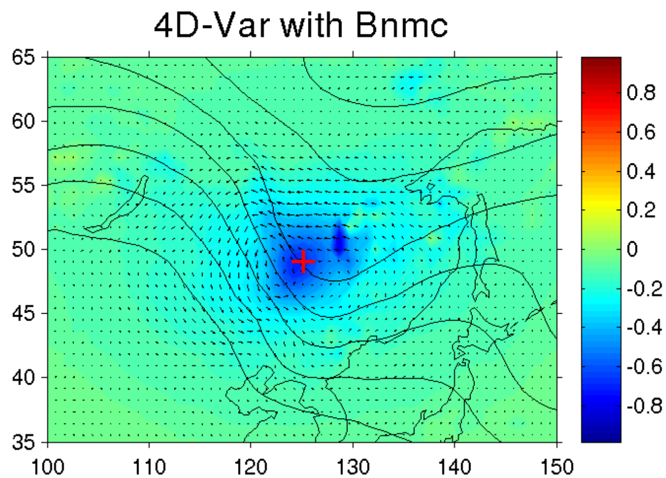


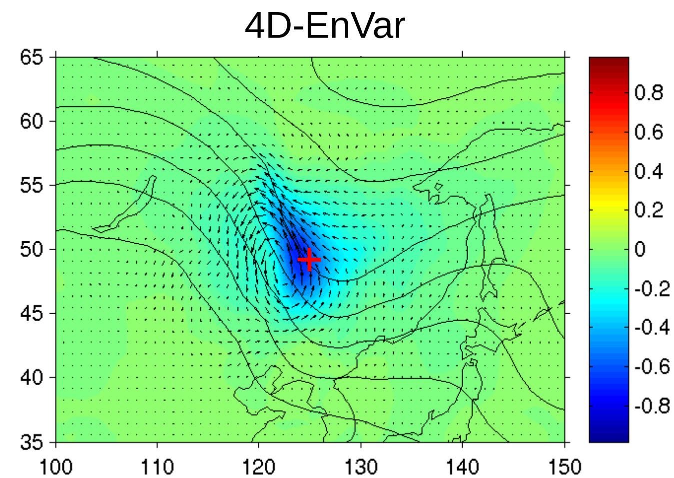


Single observation experiments

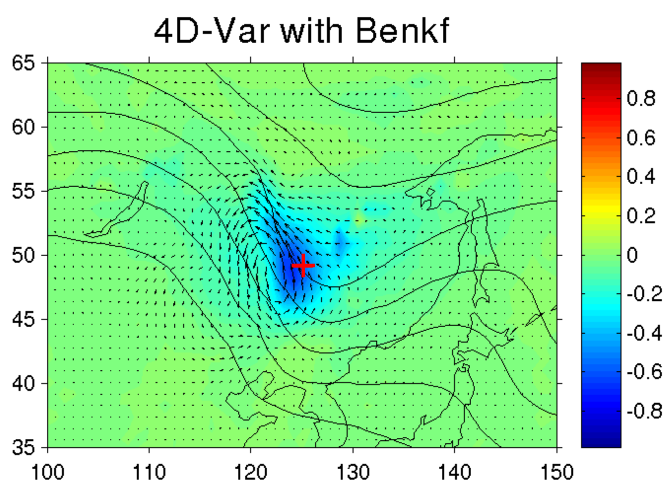
Difference in temporal covariance evolution

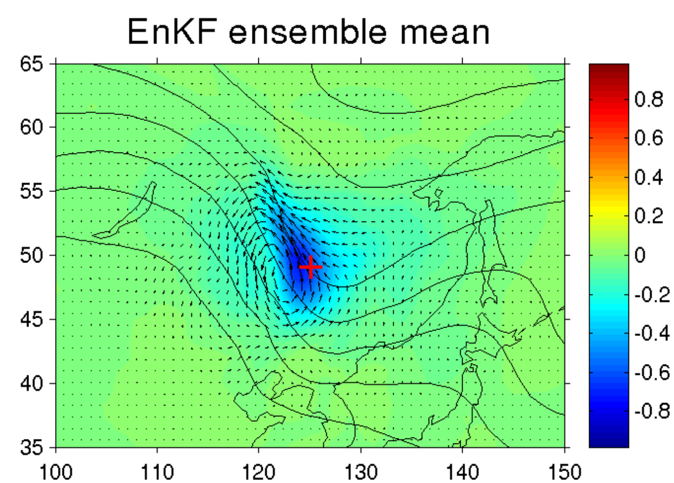
- radiosonde temperature observation at 500hPa
- observation at **middle of assimilation window (+0h)**
- with same **B**, increments very similar from **4D-Var**, **EnKF**
- contours are 500hPa GZ background state at 0h (ci=10m)





contour plots at 500 hPa

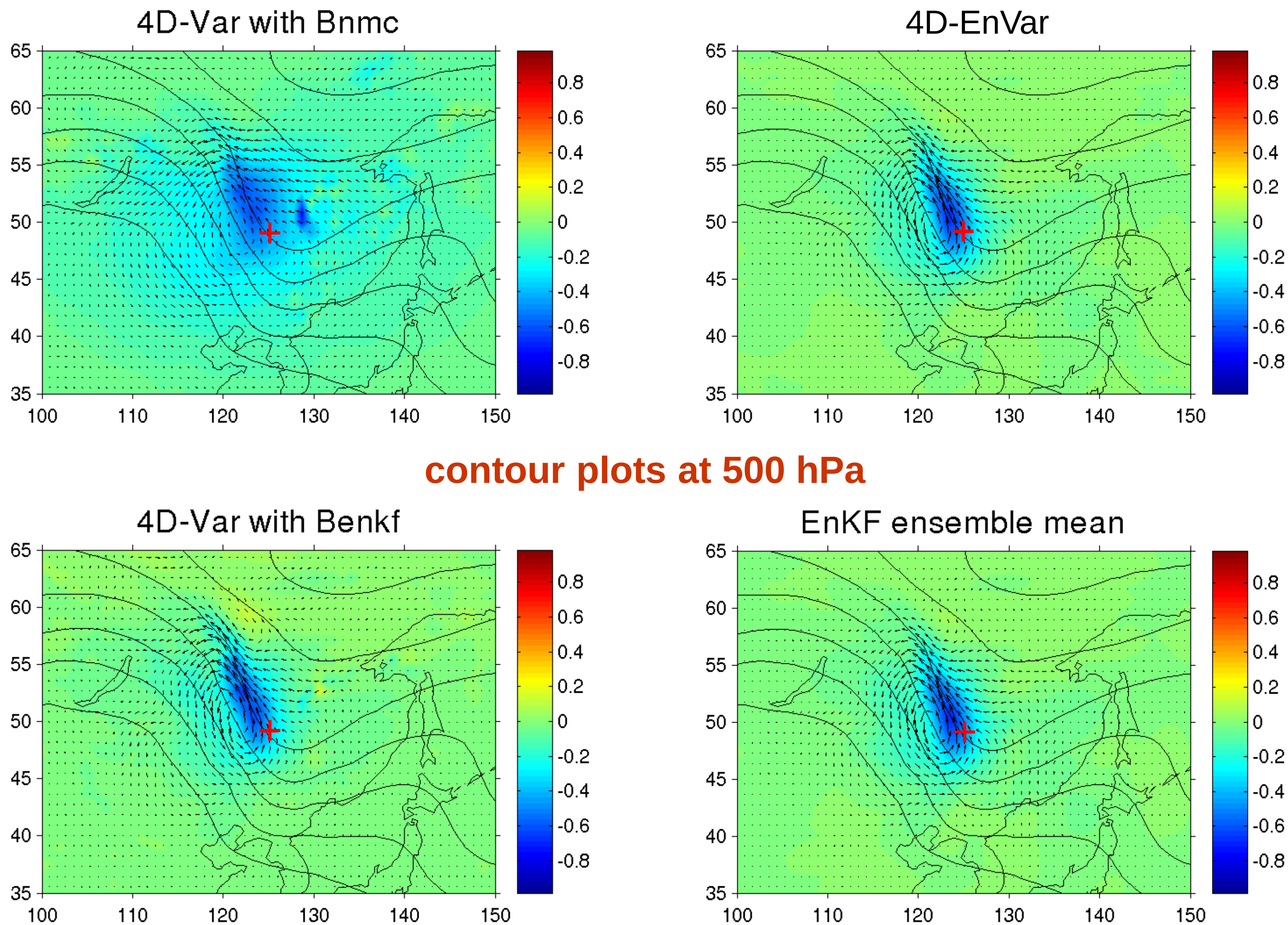




Single observation experiments

Difference in temporal covariance evolution

- radiosonde temperature observation at 500hPa
- observation at **end of assimilation window (+3h)**
- with same **B**, increments very similar from **4D-Var**, **EnKF**
- contours are 500hPa GZ background state at 0h (ci=10m)



Experimental results:
Configuration (Buehner et al. 2013, NPG)

4D-EnVar tested in comparison with version of forecast system implemented in operations in Feb, 2013:

- model top at 0.1hPa, 80 levels
- model has ~25km grid spacing
- 4D-Var analysis increments with ~100km grid spacing

4D-EnVar experiments use ensemble members from following configuration of EnKF:

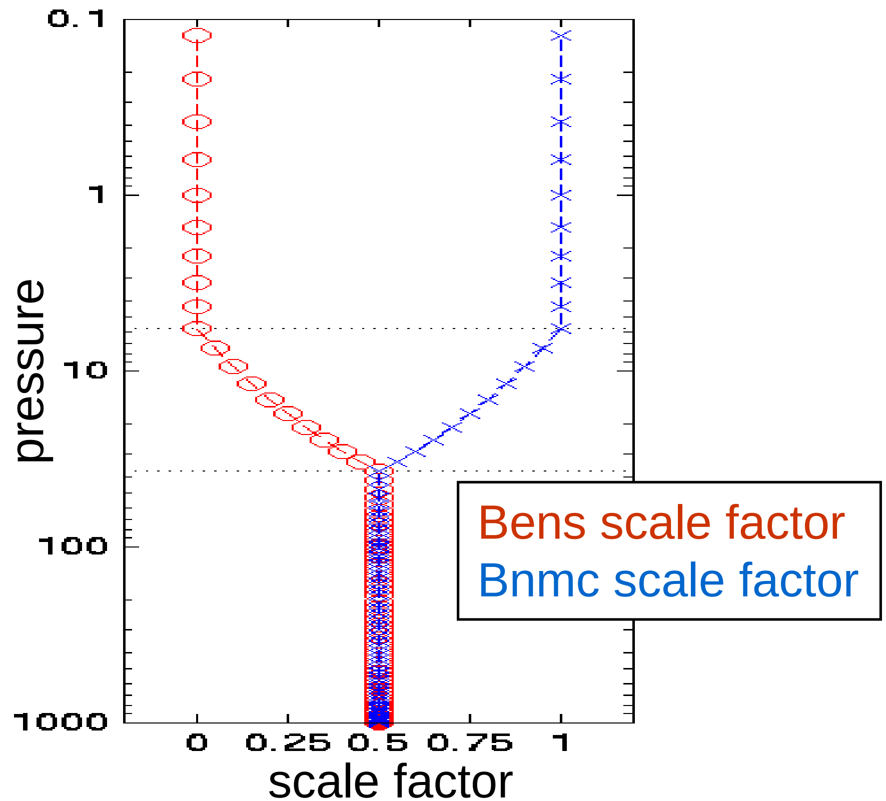
- 192 members every 60min in 6-hour window
- model top at 2hPa, 75 levels
- model ~66km grid spacing → EnVar increments ~66km

EnVar uses Hybrid Covariance Matrix

Model top of EnKF is lower than GDPS

Bens and Bnmc are averaged in troposphere ½ & ½, tapering to 100% Bnmc at and above 6hPa (EnKF model top at 2hPa)

Therefore, EnVar not expected to be better than 3D-Var above ~10-20hPa

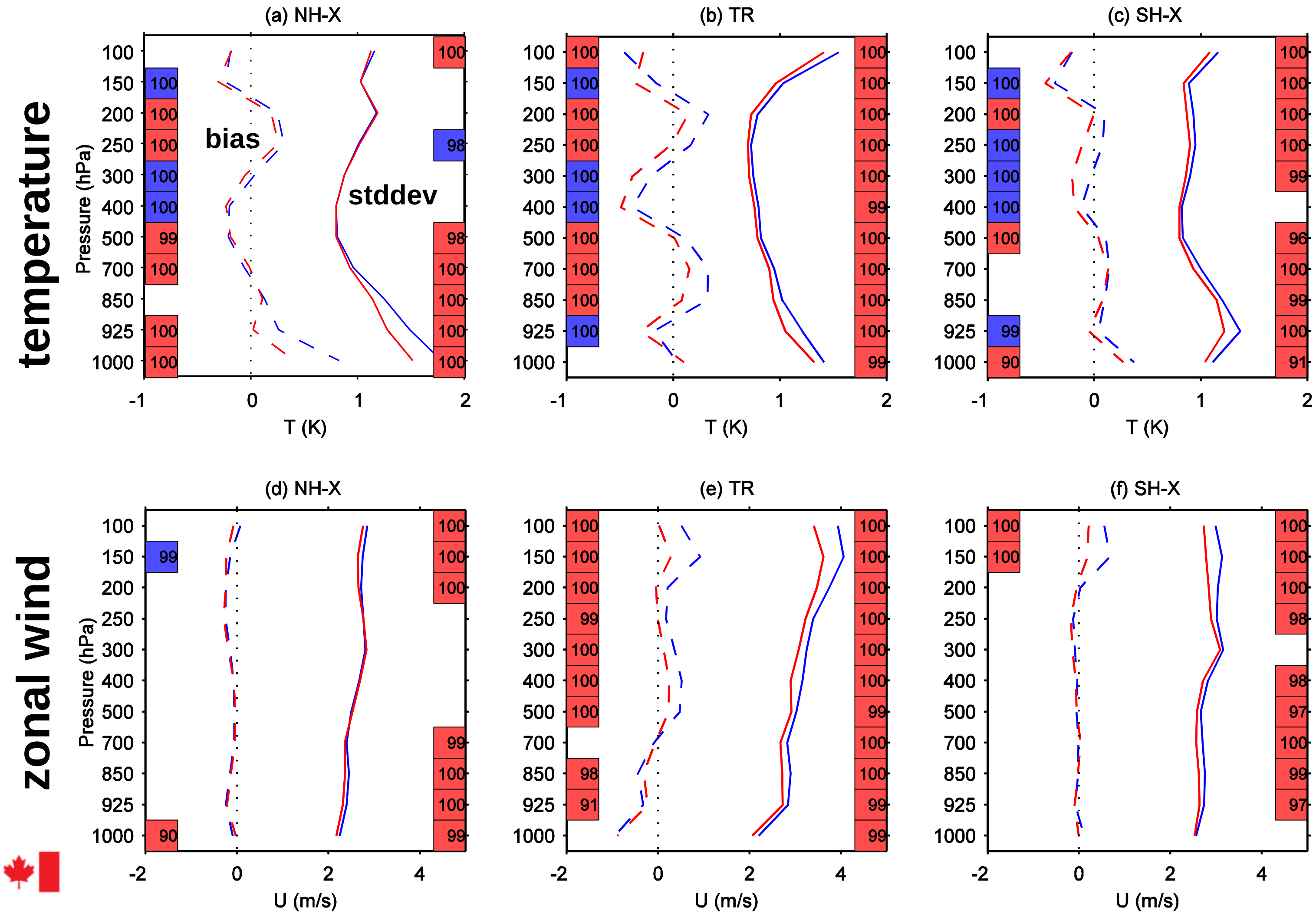


Forecast Results: 4D-EnVar vs. 4D-Var

Radiosonde verification scores – 6 weeks, Feb/Mar 2011

6h forecast

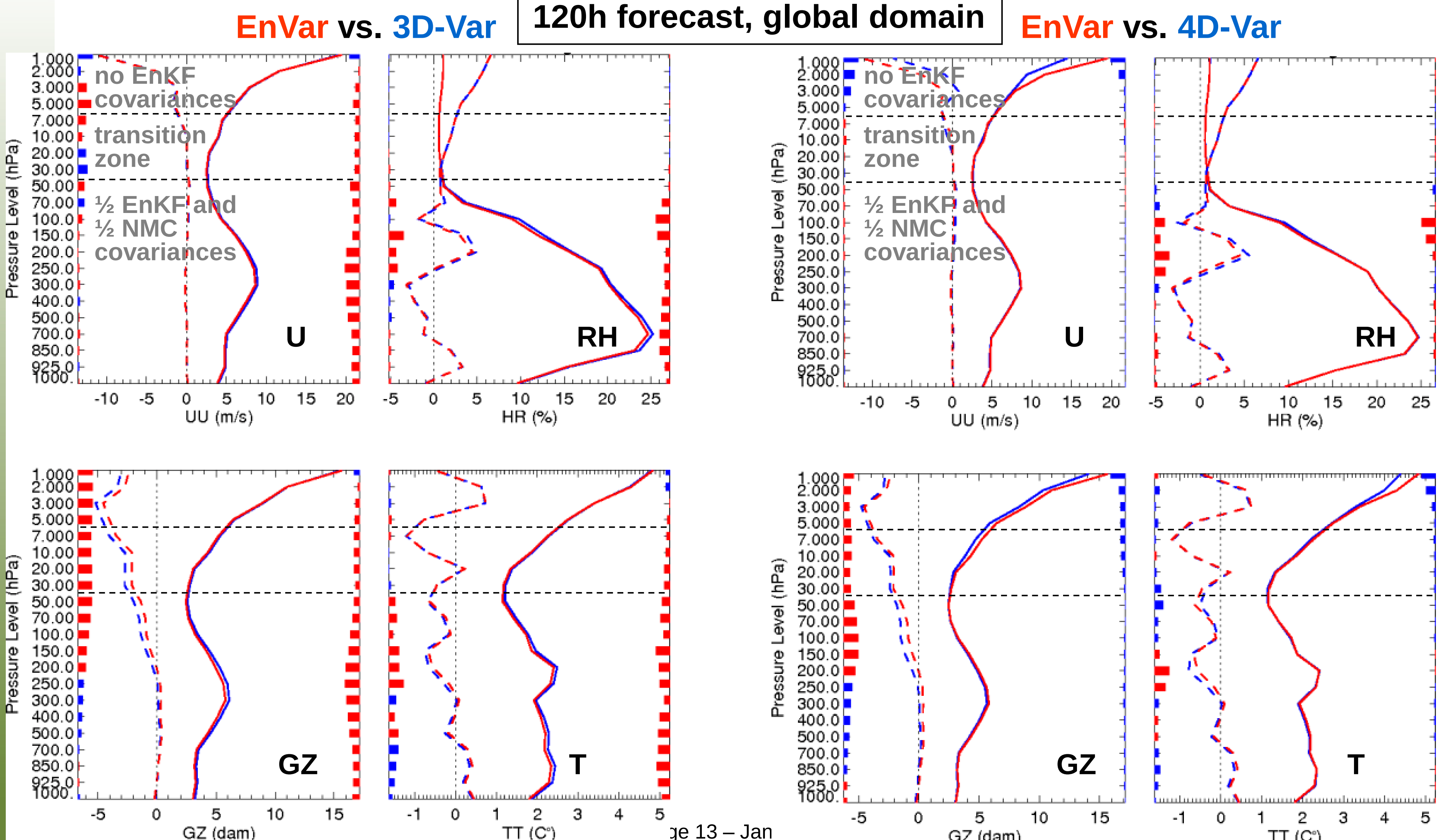
North | Tropics | South



Forecast Results: EnVar vs. 3D-Var and 4D-Var

Verification against ERA-Interim analyses – 6 weeks, Feb/Mar 2011

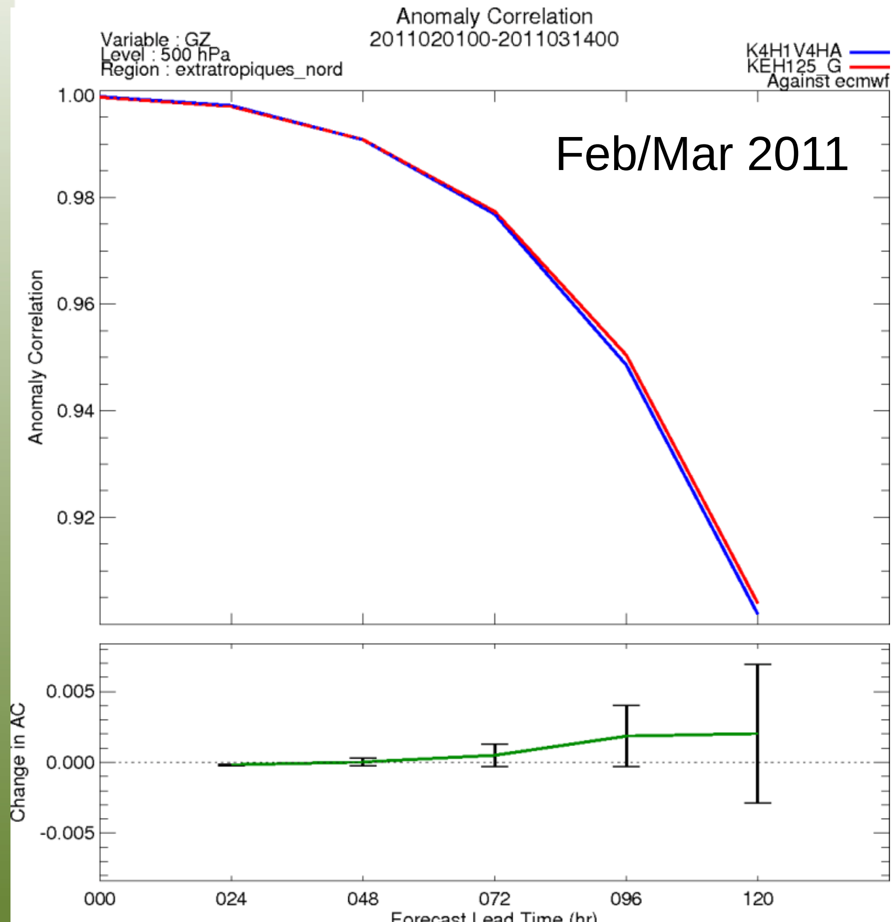
120h forecast, global domain

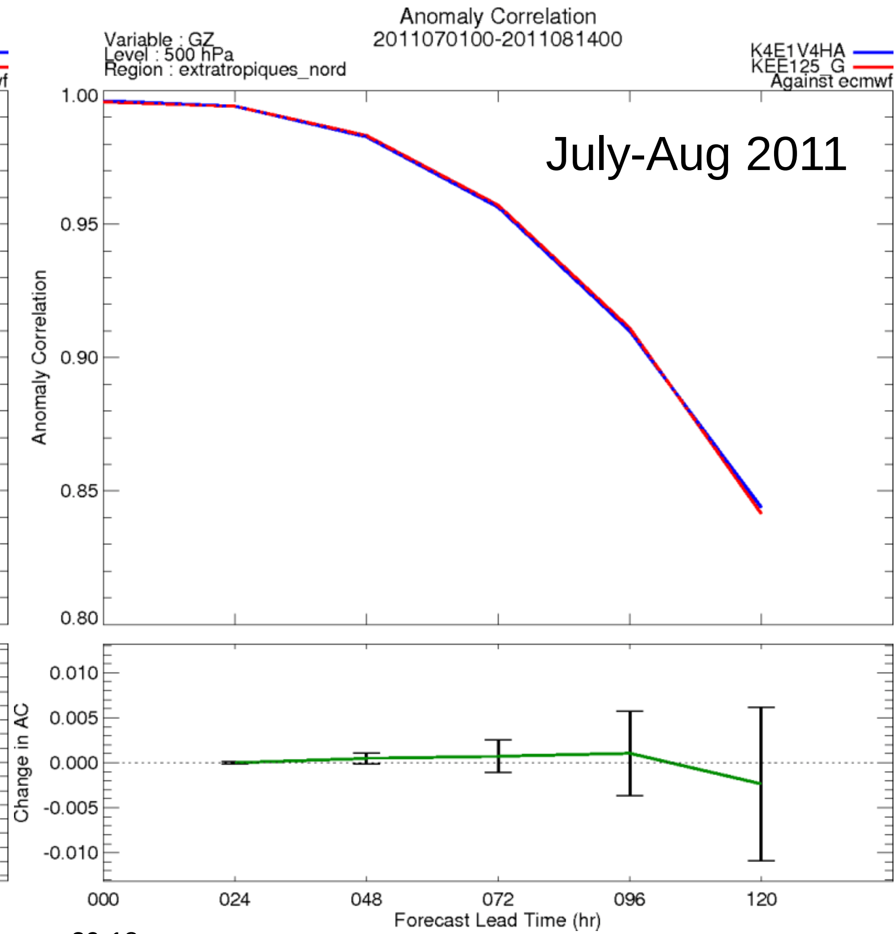


Forecast Results: 4D-EnVar vs. 4D-Var

Verification against ERA-Interim analyses – 6 weeks

North extra-tropics
500hPa GZ correlation anomaly

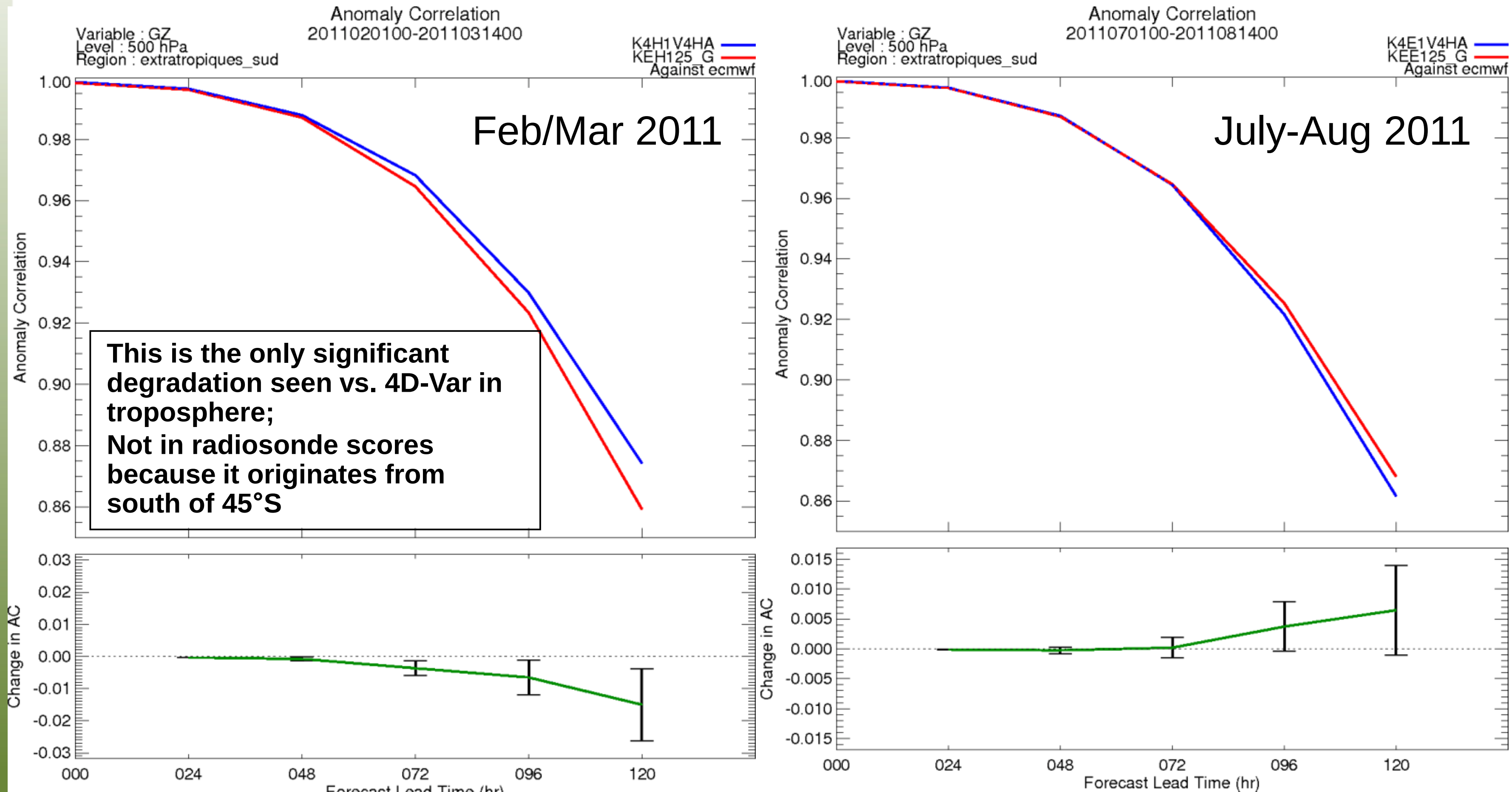




Forecast Results: 4D-EnVar vs. 4D-Var

Verification against ERA-Interim analyses – 6 weeks, Feb/Mar 2011

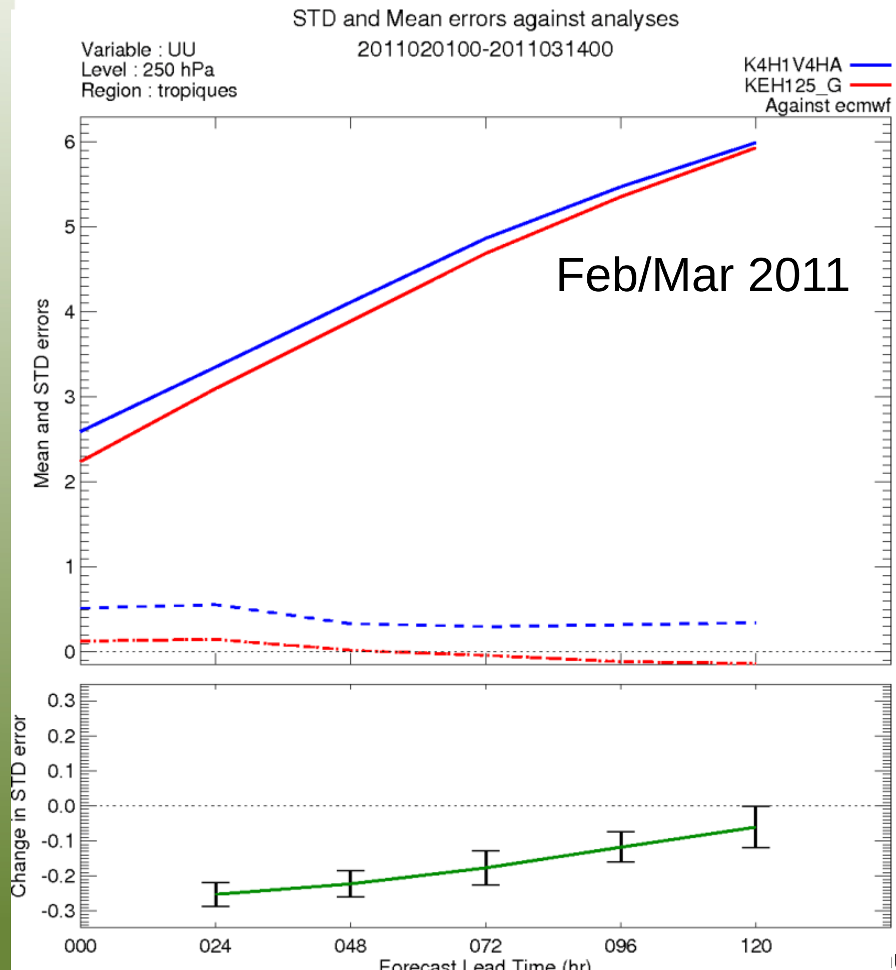
South extra-tropics
500hPa GZ correlation anomaly

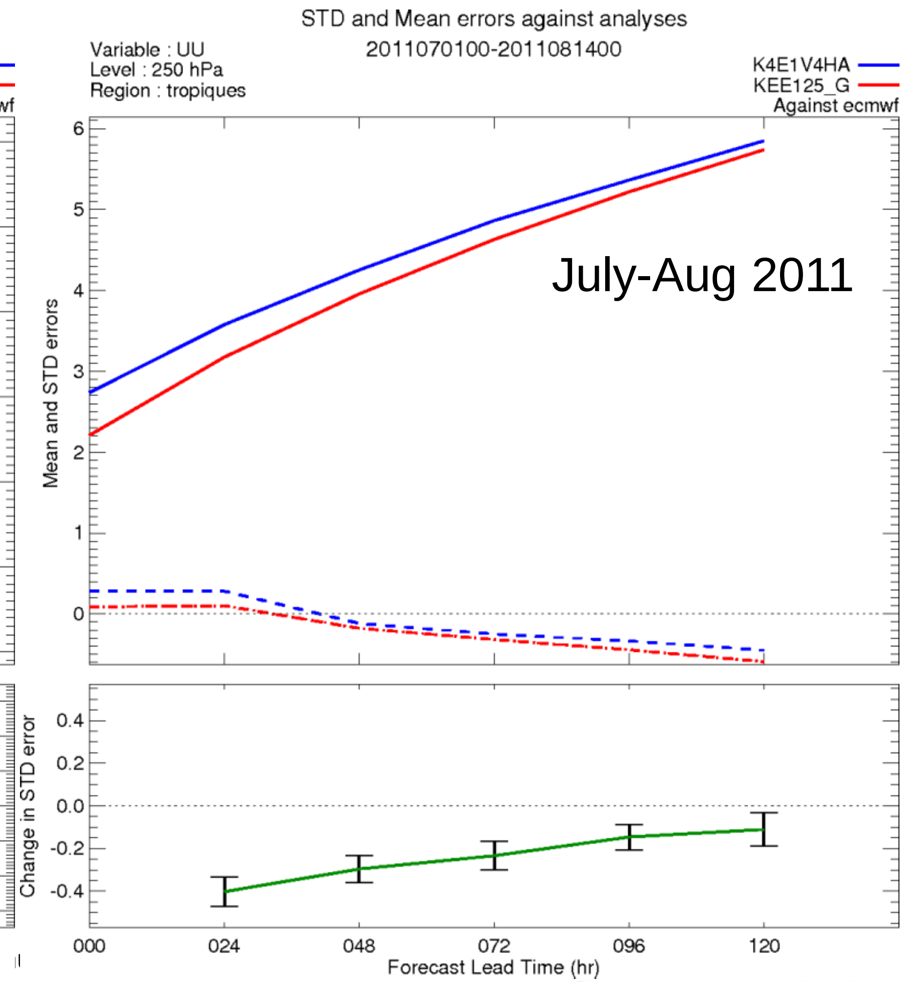


Forecast Results: 4D-EnVar vs. 4D-Var

Verification against ERA-Interim analyses – 6 weeks, Feb/Mar 2011

Tropics
250hPa U-wind STDDEV

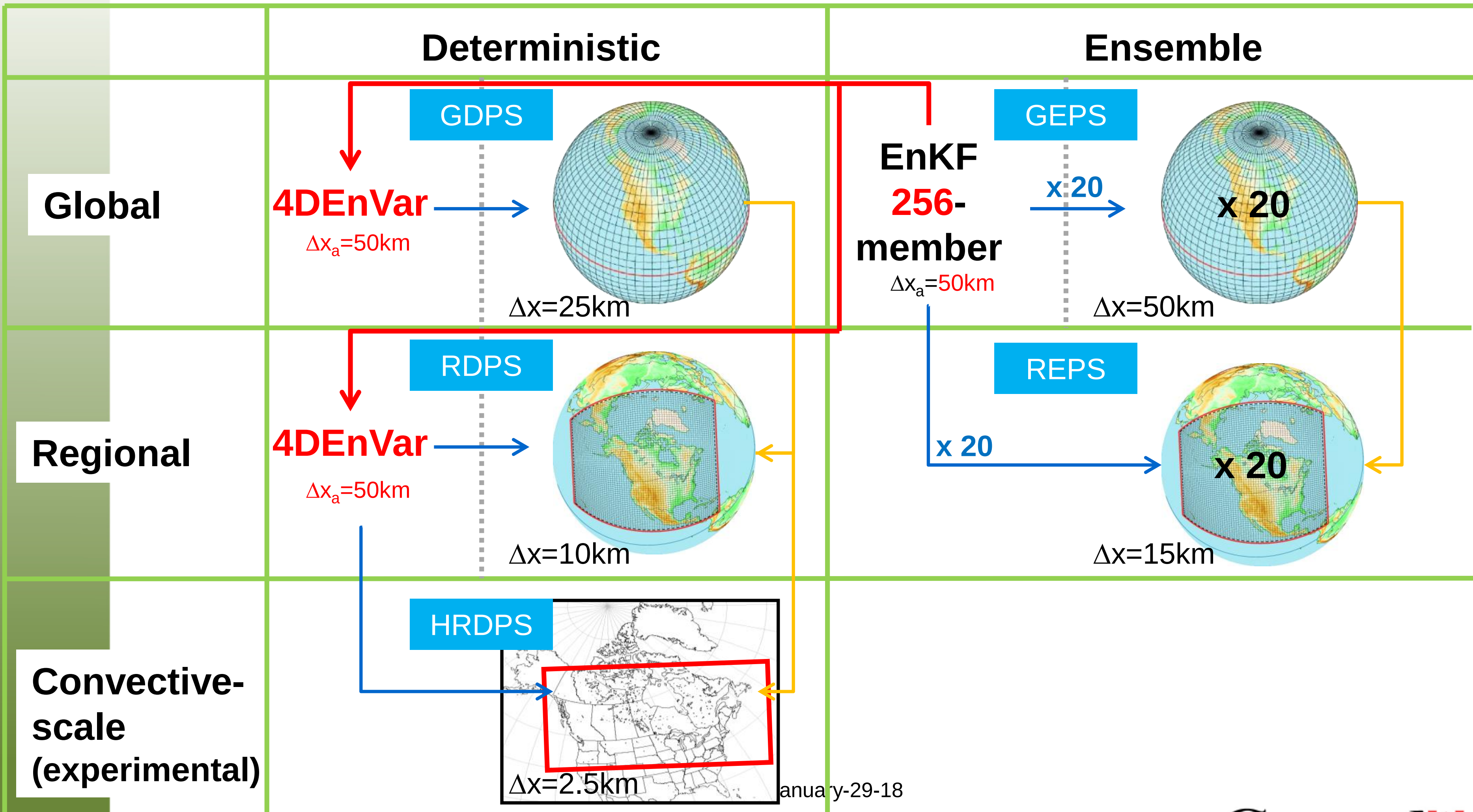




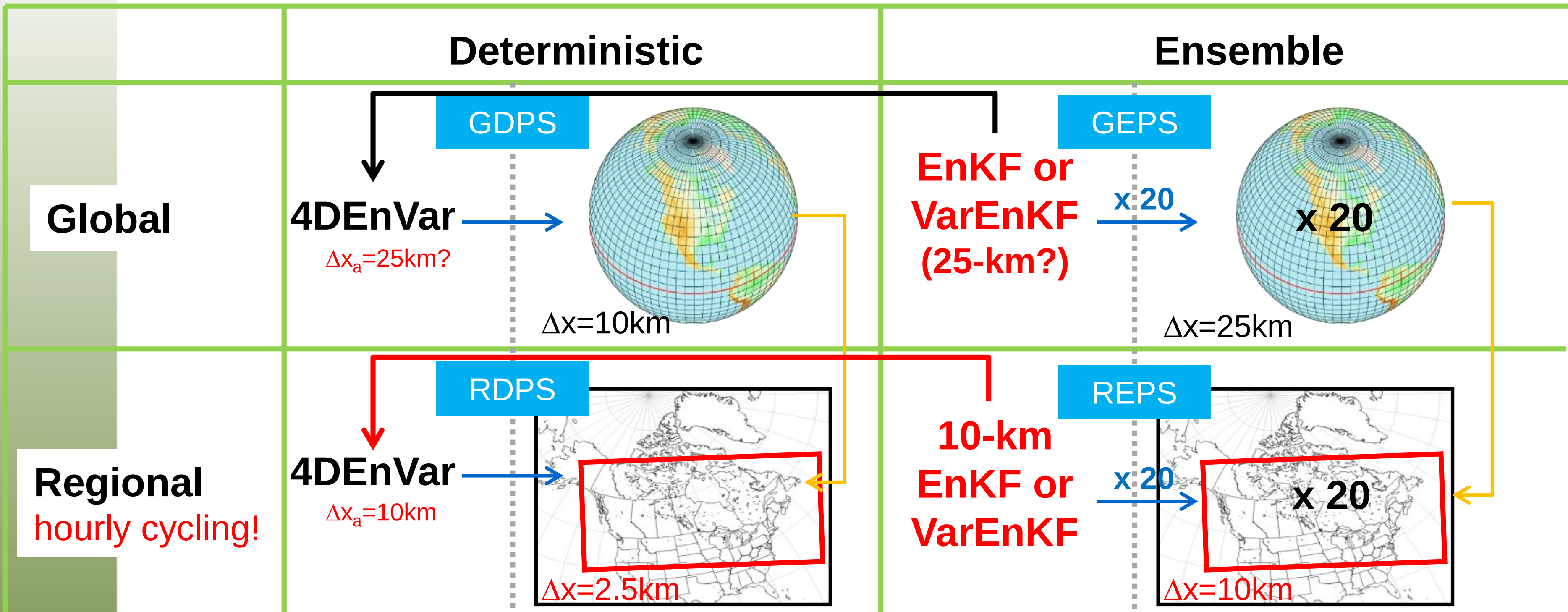
Conclusions

- Comparison of 4D-EnVar with 4D-Var (and 3D-Var):
  - EnVar produces similar quality forecasts as 4D-Var below ~20hPa in extra-tropics (except southern extra-tropical summer), significantly improved in tropics
  - above ~20hPa, scores similar to 3D-Var, worse than 4D-Var; potential benefit from raising EnKF model top to 0.1hPa
- 4D-EnVar is an attractive alternative to 4D-Var:
  - like EnKF, uses full nonlinear model dynamics/physics to evolve covariances; no need to maintain TL/AD version of model
  - computational saving allows increase in analysis resolution and more computational resources for EnKF and forecasts

ECCC's NWP systems since 2016



ECCC's NWP systems in ~2020



The range of analysed scales will increase with time <u>in both</u> global and limited-area NWP. DA methods that can cope with this challenge are needed: scale-dependent localization

Scale-dependent covariance localization

Motivation

- Currently, EnVar uses single horizontal and vertical localization length scales, very similar to our EnKF
- Comparing various studies, seems it is best to use different amount of localization depending on application:
  - convective-scale assimilation: ~10km
  - mesoscale assimilation: ~100km
  - global-scale assimilation: ~1000km – 3000km
- In the future, global systems will resolve convective scales
- Therefore, need a general approach for applying appropriate localization to wide range of scales in a single analysis procedure: **Scale-dependent localization**

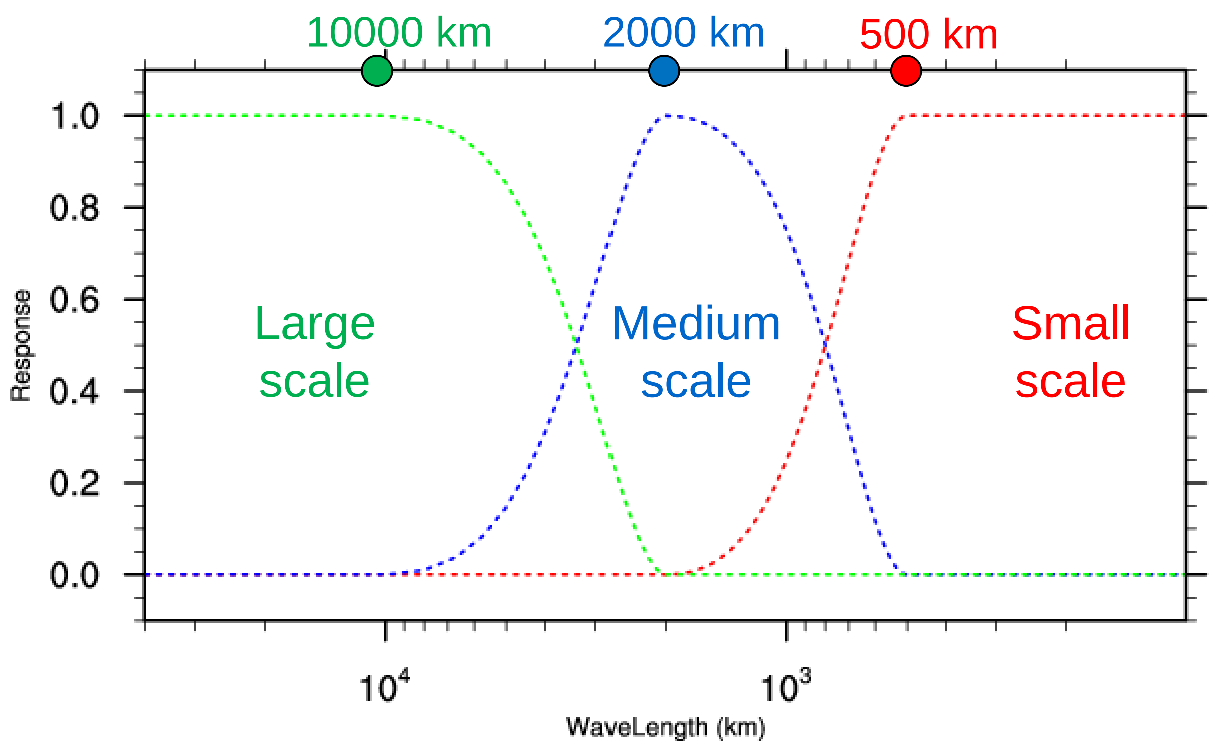
Scale-dependent covariance localization

General Approach

- Ensemble perturbations decomposed with respect to a series of overlapping spectral wavebands
- Apply scale-dependent spatial localization to the scale-decomposed covariances, both **within-scale** and **between-scale** covariances (Buehner and Shlyaeva 2015)
- Keeping the between-scale covariances is necessary to maintain heterogeneity of ensemble covariances

Horizontal Scale Decomposition

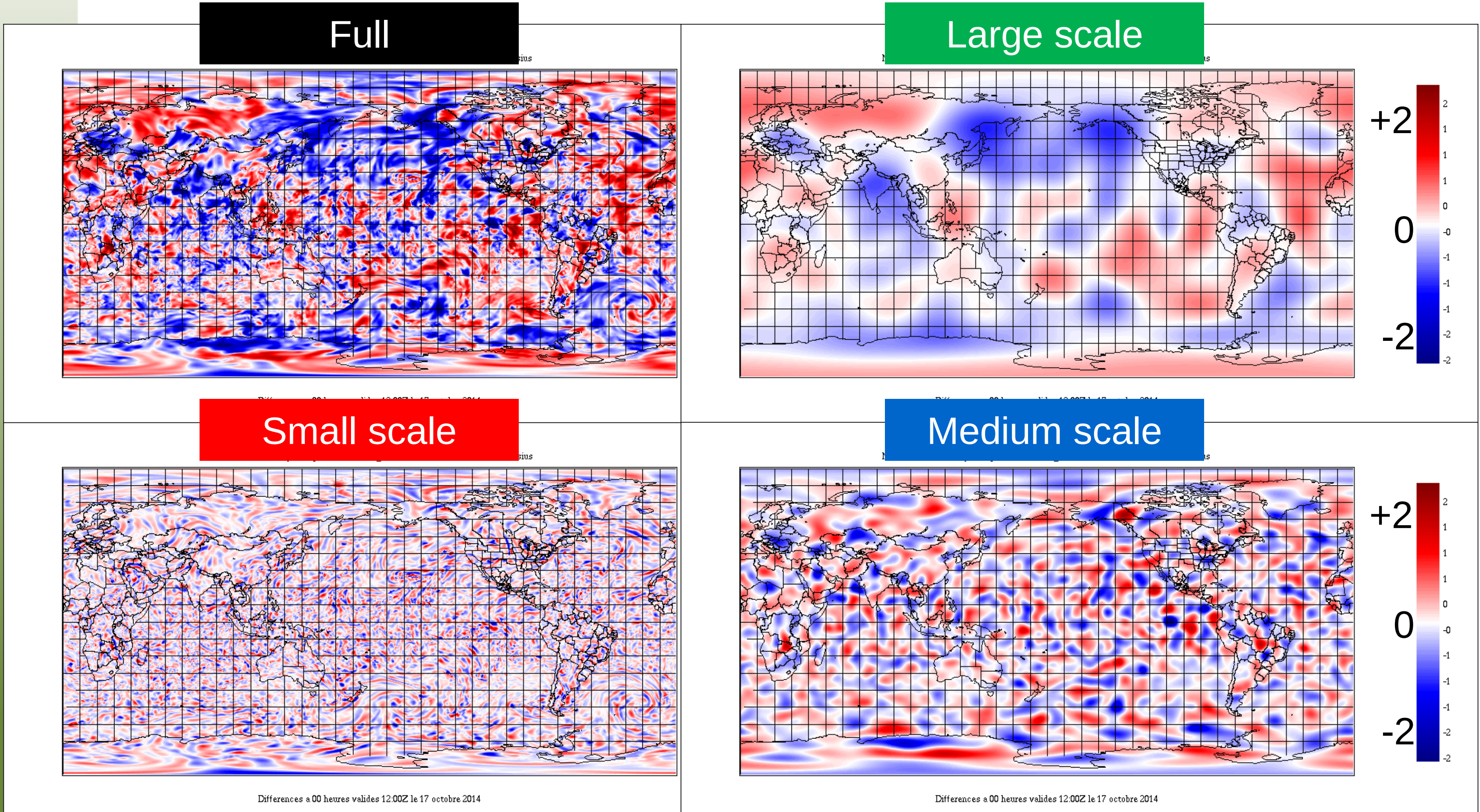
Filter response functions for decomposing with respect to 3 horizontal scale ranges



Work of Jean-Francois Caron

Horizontal Scale Decomposition

Perturbations for ensemble member #001 – Temperature at ~700hPa

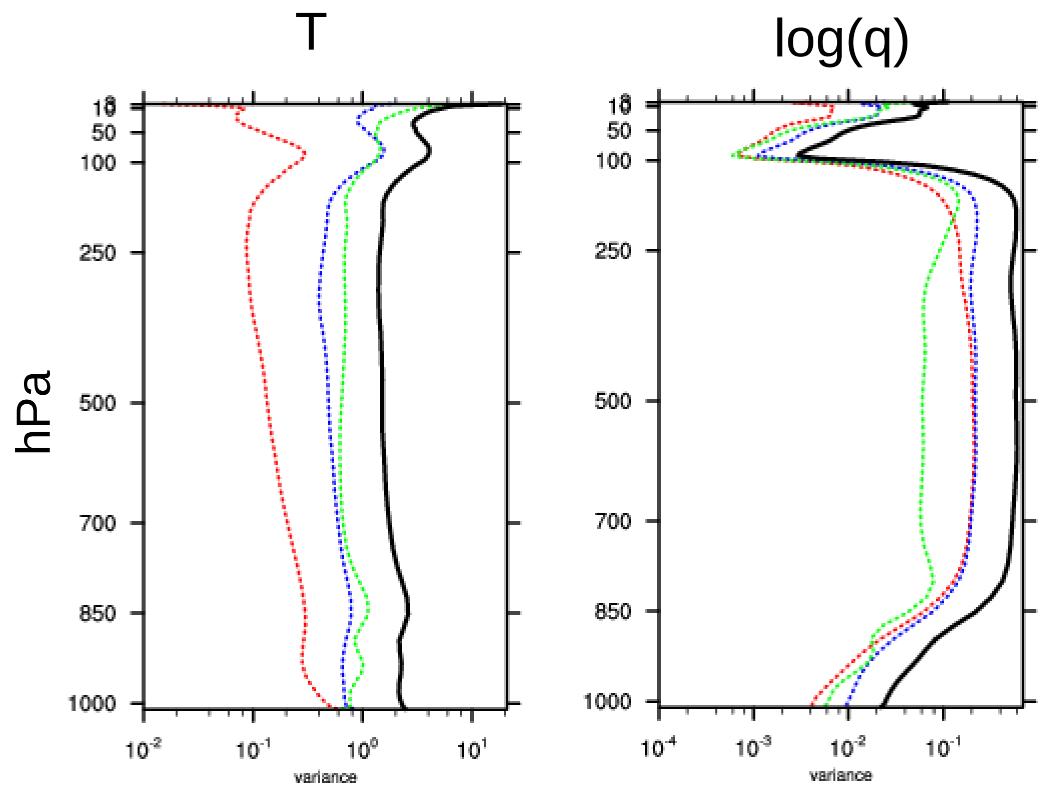


Horizontal Scale Decomposition

Waveband integrated variances

Large scale
Medium scale
Small scale
All the scales

Horizontal scale-dependent localization leads to (implicit)…
<u>variable-dependent</u>
and
<u>level-dependent</u>
horizontal localization



6-h perturbation from 256-member EnKF

Work of Jean-Francois Caron

Scale-dependent covariance localization

Implementation in EnVar

Current (one-size-fits-all) Approach

- **Analysis increment computed from control vector ($B^{1/2}$ preconditioning) using:**

$$\Delta \mathbf{x} = \sum_k \mathbf{e}_k \circ \left(\mathbf{L}^{1/2} \mathbf{v}_k \right)$$

k: member index

Scale-dependent Approach (Buehner and Shlyaeva, 2015, *Tellus*)

- **Varying amounts of smoothing applied to same set of amplitudes for a given member**

$$\Delta \mathbf{x} = \sum_k \sum_j \mathbf{e}_{k,j} \circ \left(\mathbf{L}_j^{1/2} \mathbf{v}_k \right)$$

k: member index
j: scale index

where $e_{k,j}$ is scale j of normalized member k perturbation

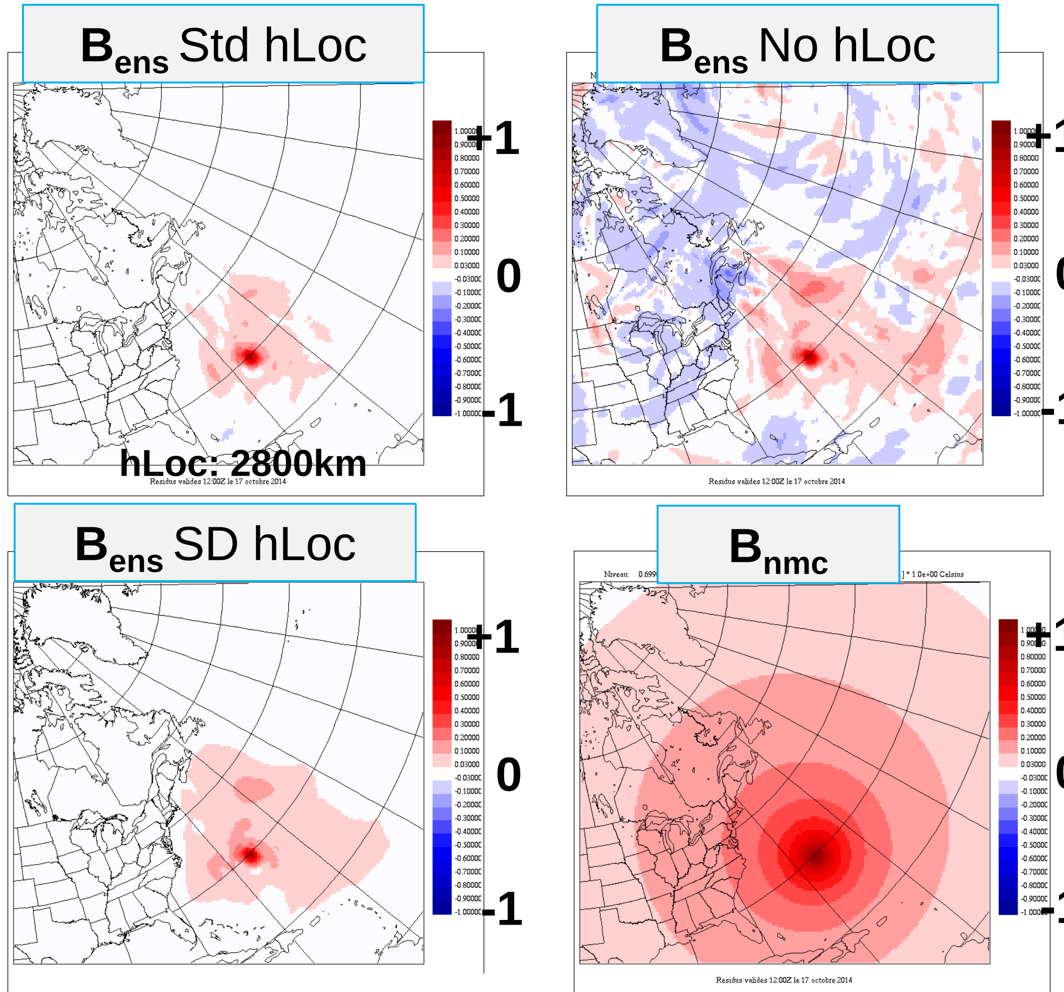
Work of Jean-Francois Caron

Scale-dependent covariance localization

Impact in single observation DA experiments

700 hPa T observation at the center of **Hurricane Gonzalo** (October 2014)

Normalized temperature increments (correlation-like) at 700 hPa resulting from various B matrices.



hLoc: 1500km / 4000km / 10000km

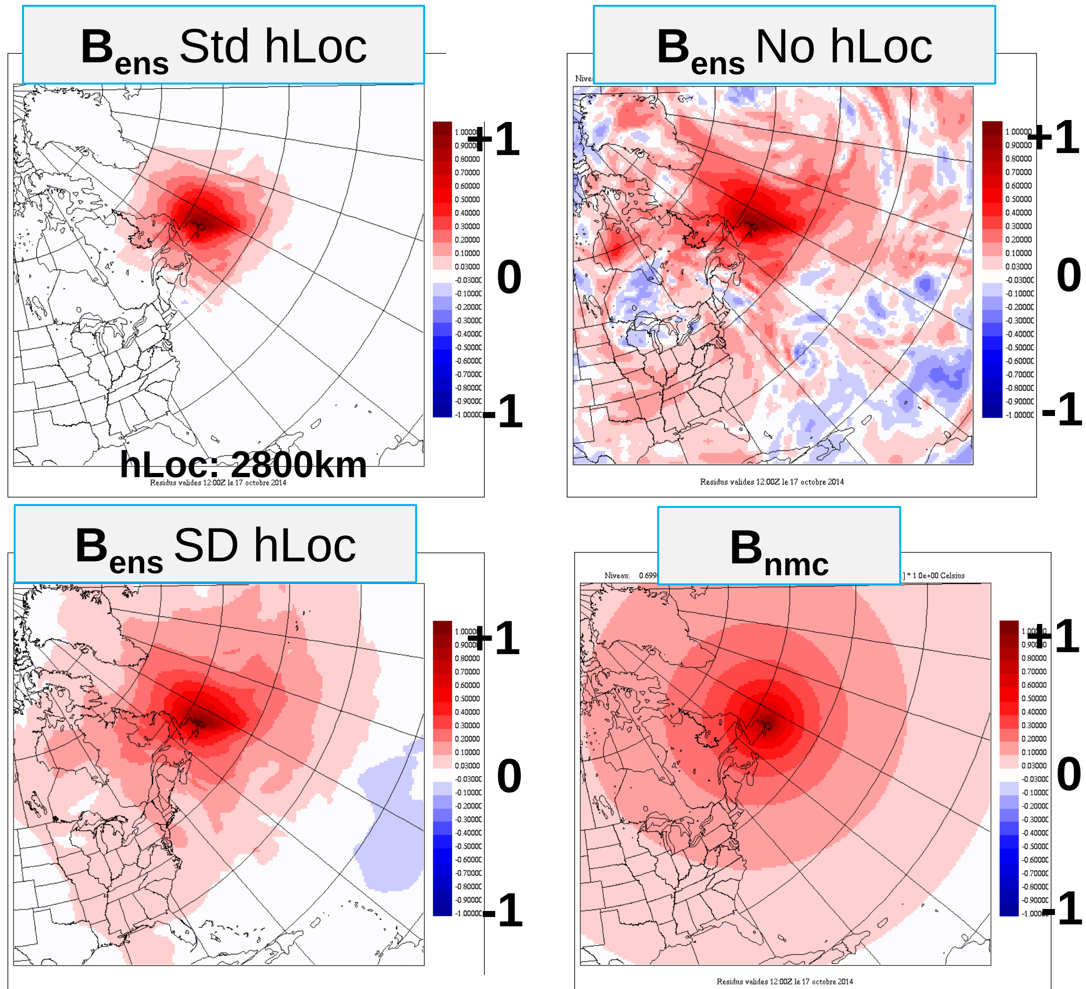
Work of Jean-Francois Caron

Scale-dependent covariance localization

Impact in single observation DA experiments

700 hPa T observation at the center of a **High Pressure**

Normalized temperature increments (correlation-like) at 700 hPa resulting from various B matrices.



hLoc: 1500km / 4000km / 10000km

Work of Jean-Francois Caron

Scale-dependent covariance localization

Forecast impact

- 2.5-month trialling (June-August 2014) in our global NWP system.
- 3D-EnVar with 100% B_{ens} used in both experiments

1) **Control experiment:** hLoc = 2800 km, vLoc = 2 units ln(p)
2) **Scale-Dependent experiment** with a 3 horizontal-scale decomposition

   I. Small scale uses hLoc = 1500 km
   II. Medium scale uses hLoc = 2400 km
   III. Large scale with uses = 3300 km

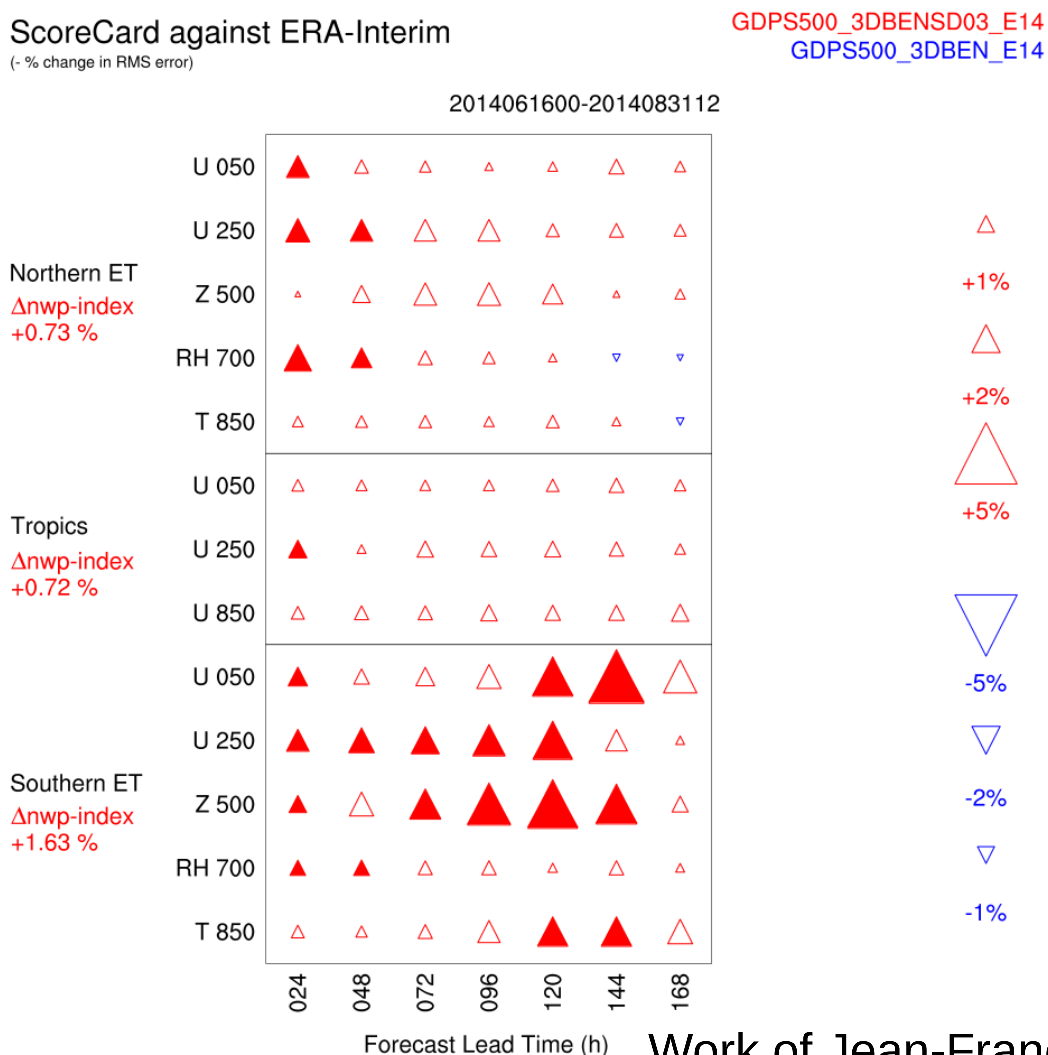
   Ad hoc values!

   Same vLoc (2 units of ln(p)) for every horizontal-scale

Work of Jean-Francois Caron

Scale-dependent covariance localization

Forecast impact – Comparison against ERA-Interim



SDL is better

Control is better

Solid triangles indicate statistically significant differences

Work of Jean-Francois Caron

Summary – Scale-dependent localization

- Scale-dependent localization is feasible, but more expensive than single-scale localization (like having a larger ensemble)
- Preliminary results using a <u>horizontal</u>-scale-dependent <u>horizontal</u> localization results in modest forecast improvements in our global NWP system
- Expect larger improvements in a system with larger range of scales assimilating dense high-resolution observations and/or with fewer ensemble members
- Finding the optimal SDL setup is **not** straightforward

Work of Jean-Francois Caron

FSOI adapted for 4D-EnVar: Motivation

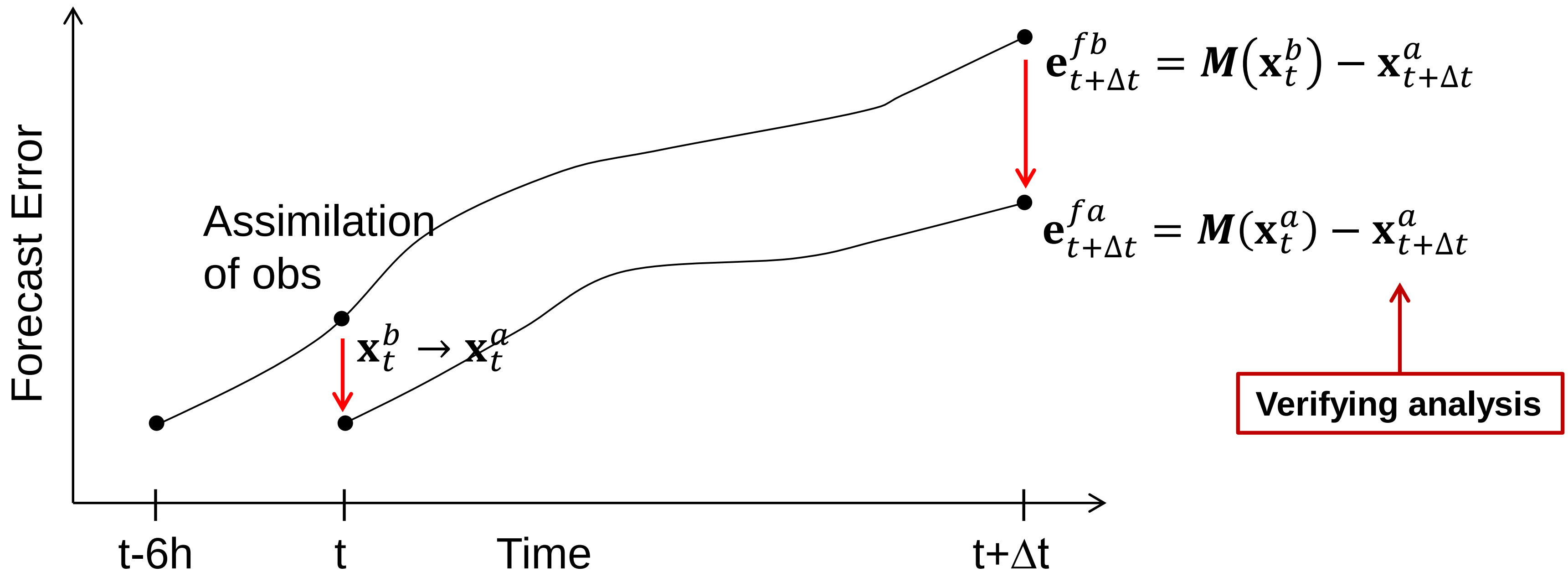
- Since retirement of 4D-Var, development discontinued of tangent linear and adjoint of forecast model
- Therefore, to perform FSOI in context of 4D-EnVar, requires adapting approach to avoid use of adjoint of forecast model
- Pure ensemble approach exists (e.g. as used at NCEP), but can only give impact of observations assimilated in EnKF
- At ECCC, numerous observation types assimilated in 4D-EnVar **not** assimilated in EnKF (AIRS, IASI, CrIS, SSMIS, Geo-rad, GB-GPS)

Basic idea of FSOI

- Goal is to partition, with respect to arbitrary subsets of observations, the forecast error reduction from assimilating these observations:

$$\Delta e^2 = \left(\mathbf{e}^{fa}_{t+\Delta t}\right)^T \mathbf{C}\left(\mathbf{e}^{fa}_{t+\Delta t}\right) - \left(\mathbf{e}^{fb}_{t+\Delta t}\right)^T \mathbf{C}\left(\mathbf{e}^{fb}_{t+\Delta t}\right)$$

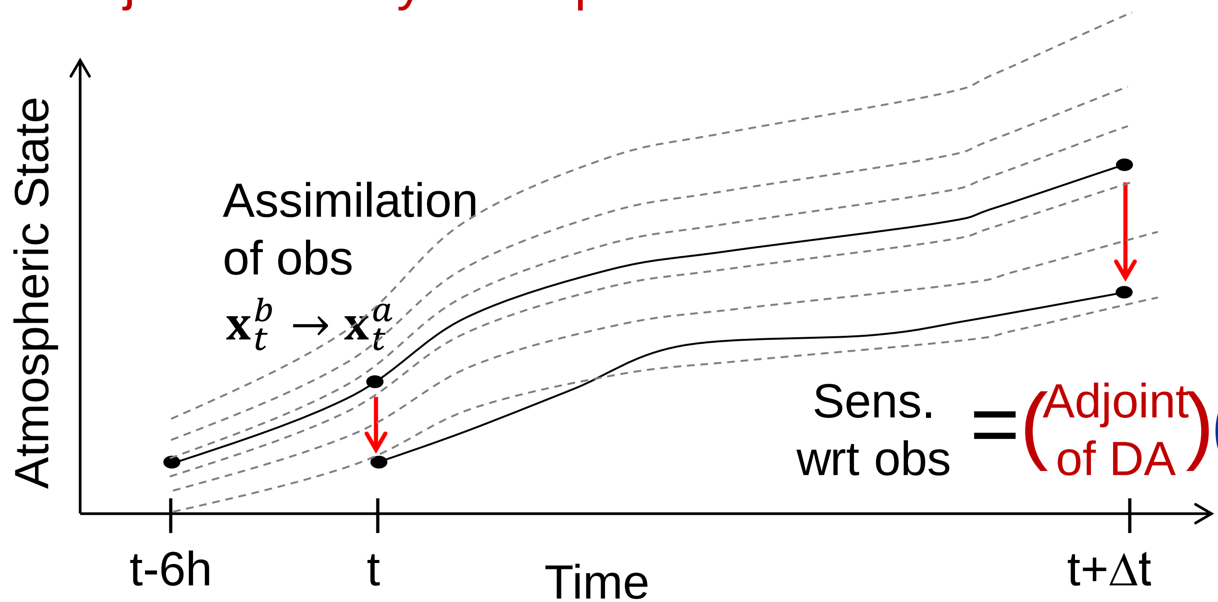
Scalar measure of forecast error



New FSOI approach <u>(Idea from Lorenc working paper)</u>

Forecast step uses ensemble, DA step like variational approach

- Instead of using adjoint of forecast model, sensitivities propagated to analysis time using extended background ensemble forecasts → <u>requires use of 100% ensemble **B** in analysis step</u>
- The analysis increment is a (spatially varying) linear combination of the background ensemble*****, the propagated increment is assumed to be the same linear combination of the ensemble***** at the forecast time
- Adjoint of analysis step uses standard variational approach



*****actually the deviations of the ensemble members from the ensemble mean state

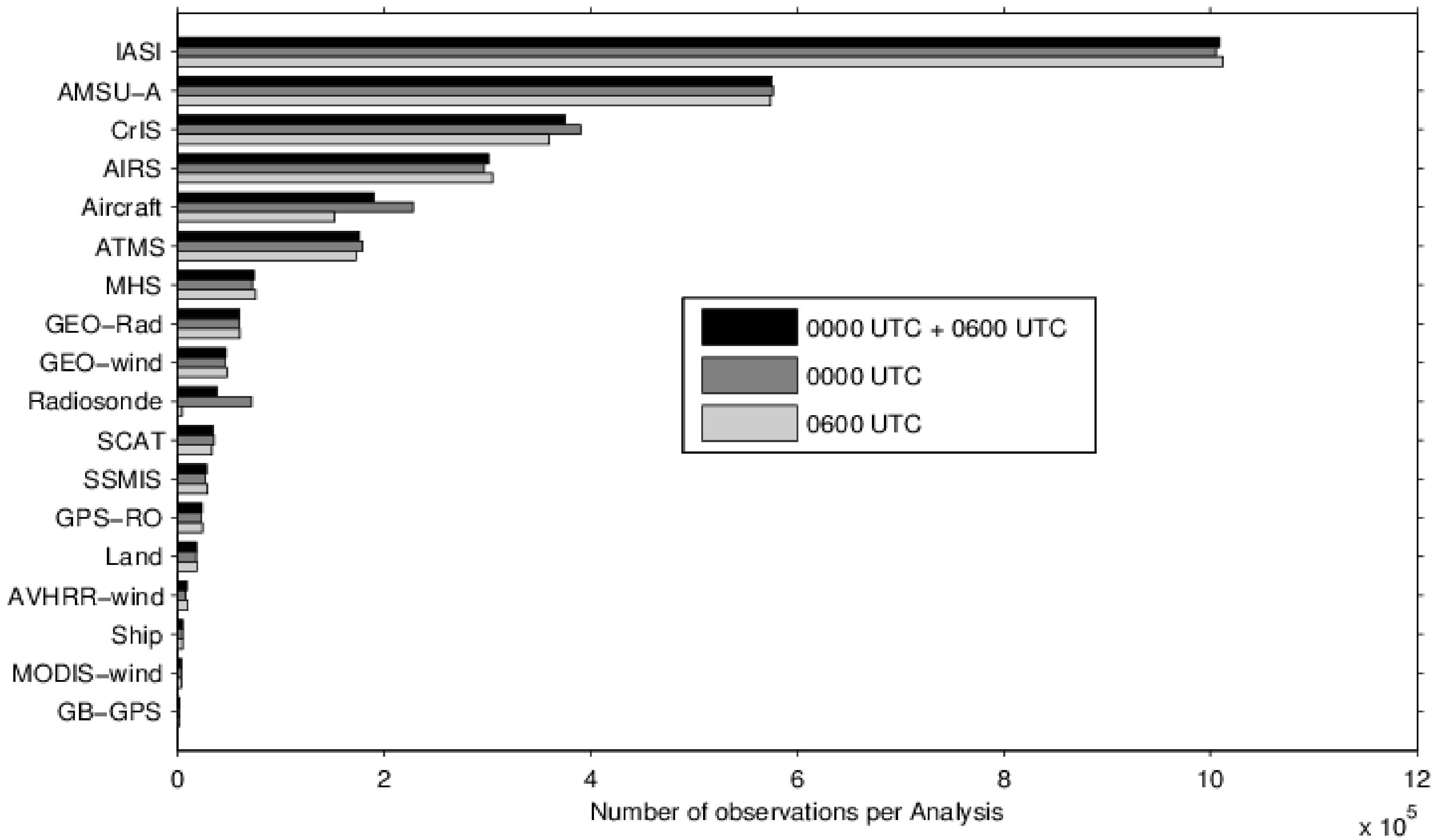
Sens. wrt obs = (Adjoint of DA)(Adjoint of Fcst)(Sens. wrt fcst.)

FSOI experiments with new approach

- Performed 4D-EnVar data assimilation experiment similar to operational configuration, but with 100% ensemble **B**
- Forecast error measured with **dry global energy norm up to 100hPa** relative to operational GDPS analyses
- For new ensemble-variational approach, computed FSOI both with and without horizontal advection of the localization (0.75 × wind)
- Compared results with using adjoint of forecast model to propagate sensitivities from forecast→analysis time

Results

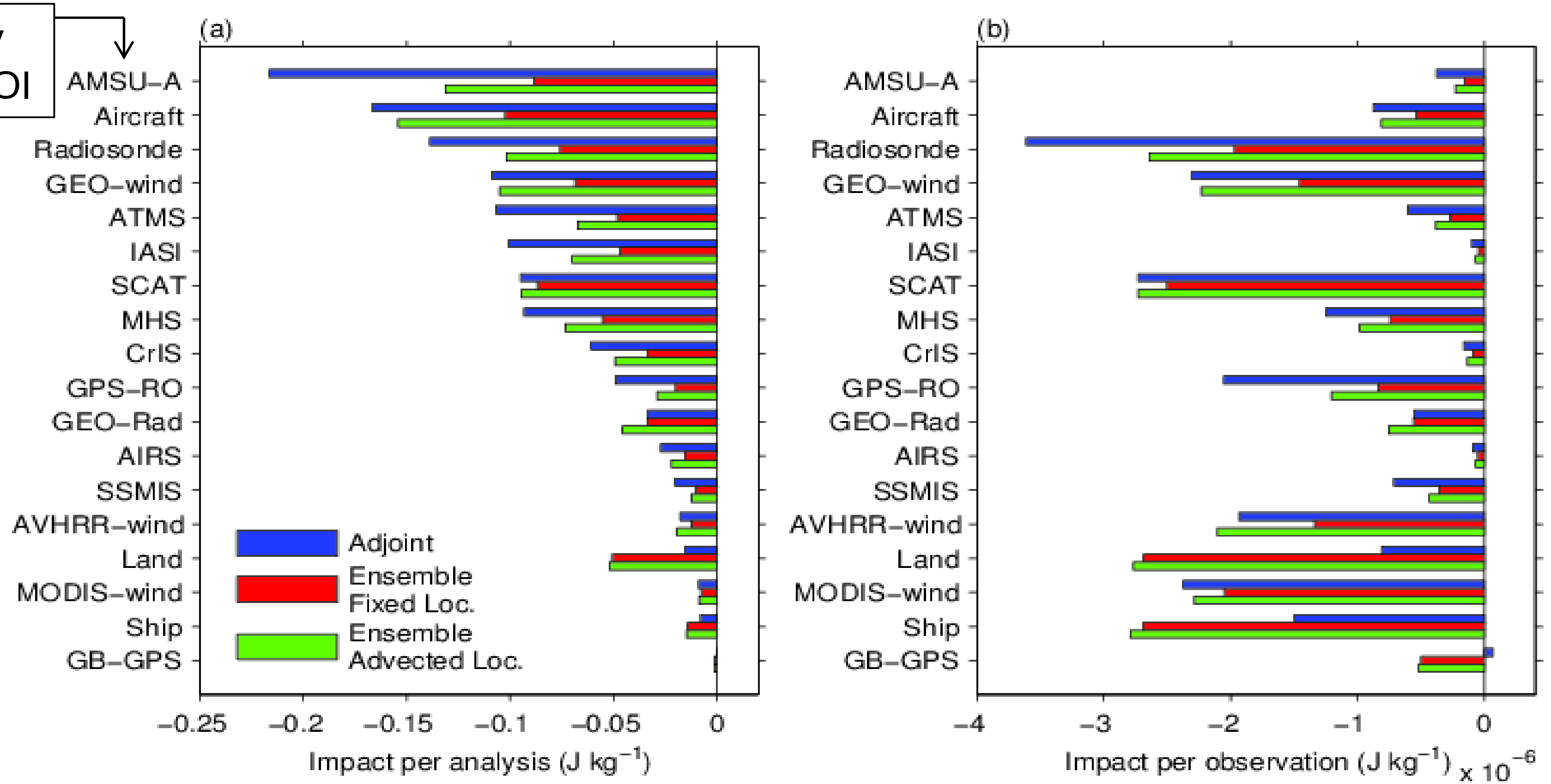
Number of assimilated observations



Results

Average impact per analysis on 0Z+6Z 24h forecasts

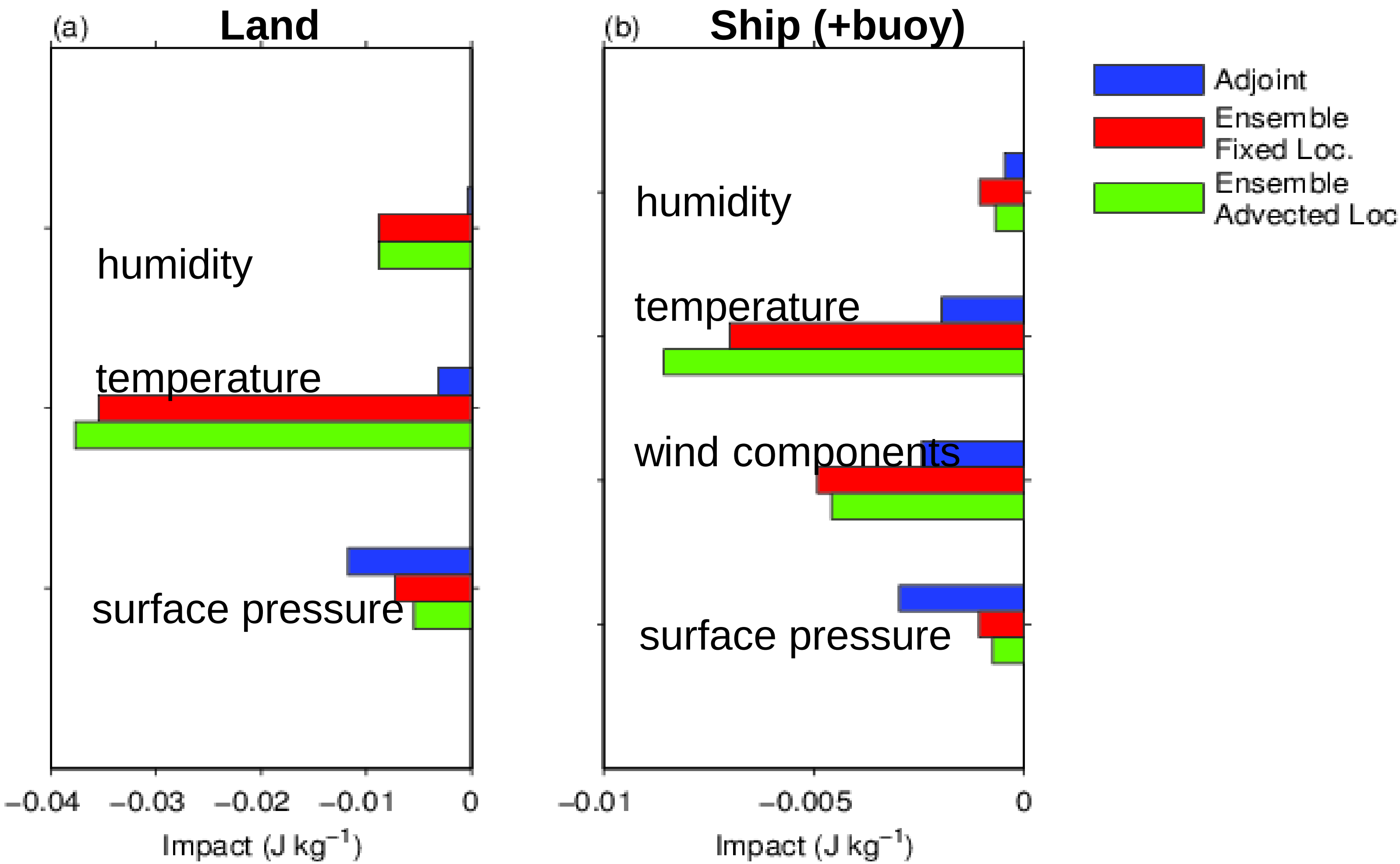
Ordered by Adjoint-FSOI



- Overall, similar results between using ensemble or adjoint model
- Advection increases apparent impact when using ensemble
- In-situ surface obs have larger apparent impact when using ensemble
- Radiances, Raob and GPS-RO have lower impact with ensemble

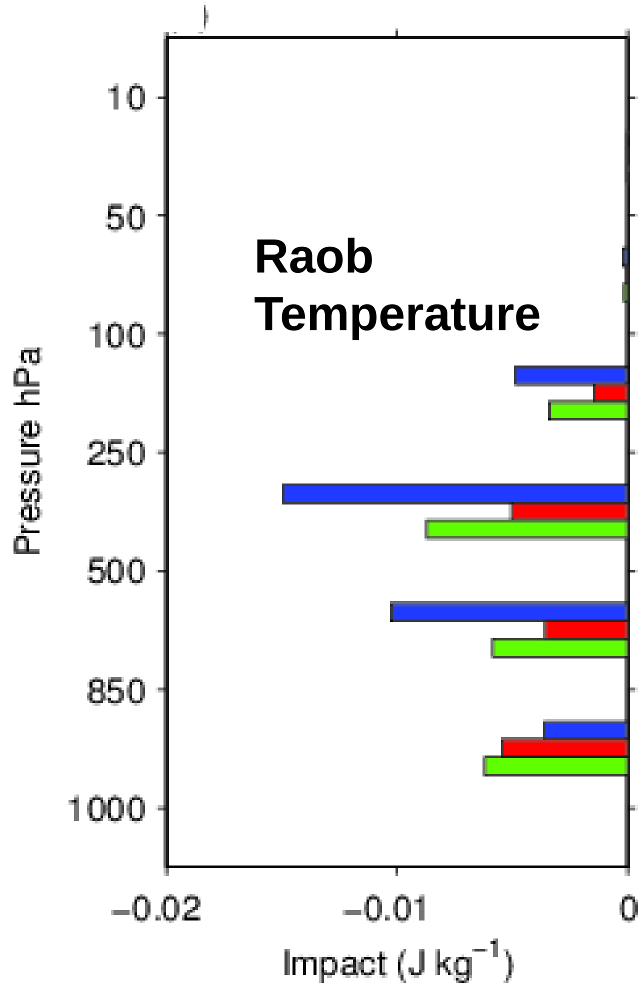
Results: 24h forecasts

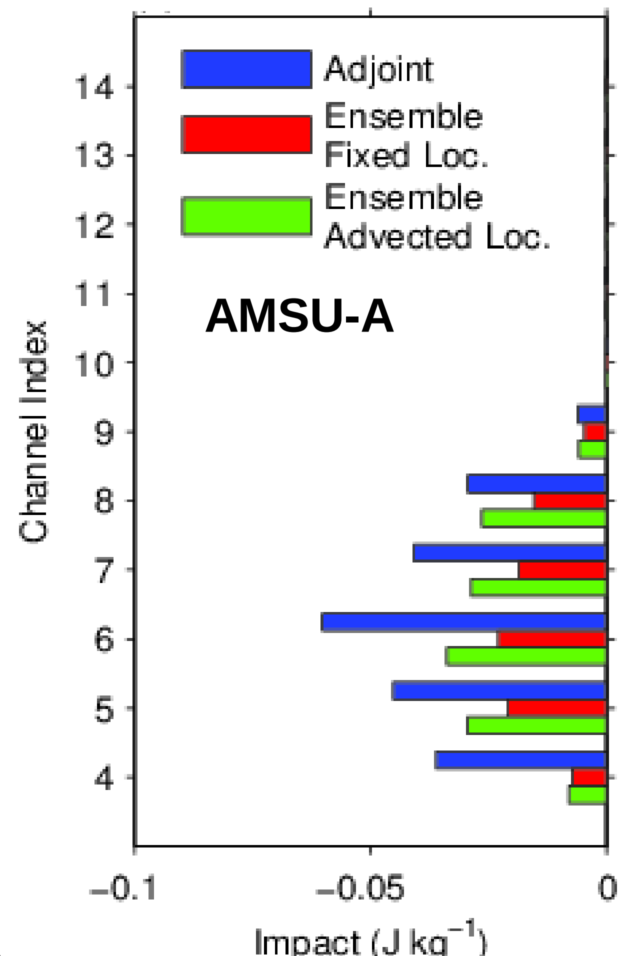
Average impact of Land and Ship (+buoy)



Results: 24h forecasts

Vertical distribution: Impact of Raob and AMSU-A

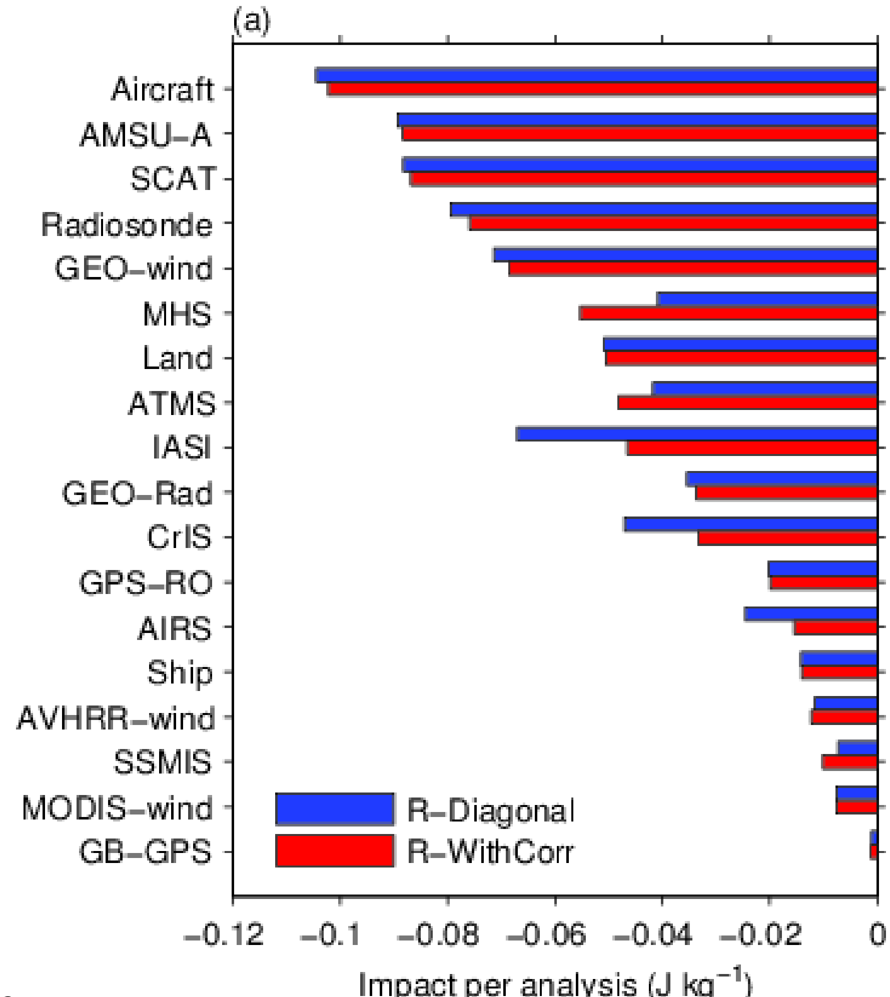




Results: 24h forecasts

Impact when using diagonal versus non-diagonal R

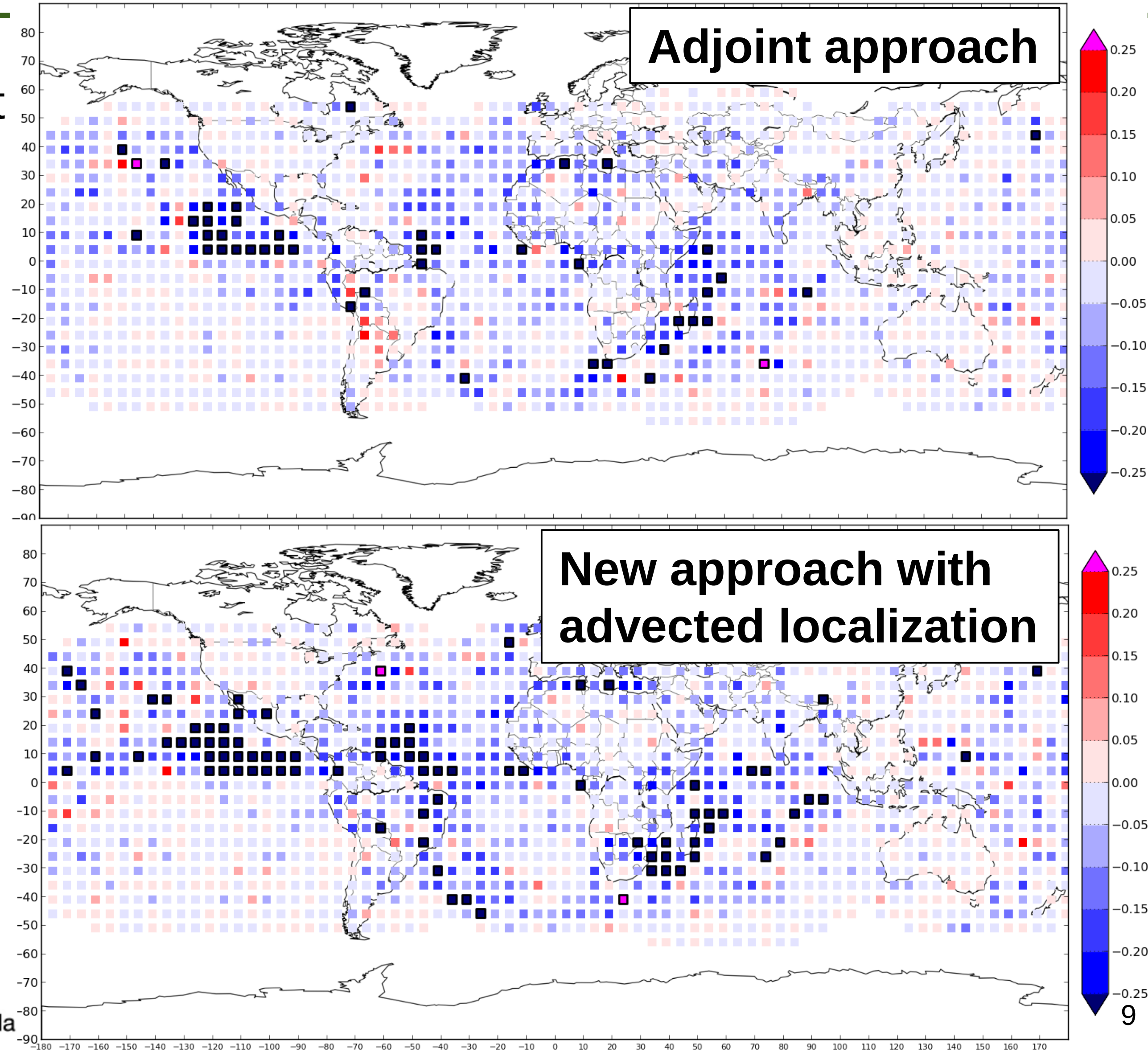
- Inclusion of inter-channel error correlations combined with reduction of obs error variances for highly correlated humidity channels
- Use of correlations decreases impact of hyper spectral IR sensors
- Cannot separate impact of inter-correlated channels



Results: 24h forecasts

Daily average impact for Geo-Radiances in 5˚x5˚ boxes

- Detailed spatial impact of geostationary radiances (1 channel per instrument) is generally similar between approaches



Conclusions – New FSOI approach

- Results with new FSOI approach adapted for use with EnVar qualitatively similar to using adjoint model
- Significant differences for some obs types (e.g. sfc):
  - At least partially due to vertical ensemble localization
  - Also due to nonlinear ensemble vs. linear adjoint propagation:
    - incomplete simplified physics and no surface sensitivities in adjoint model
    - use of multi-physics approach and independently evolving surface fields in ensemble
- Current approach (formally) limited to EnVar with 100% ensemble **B**

Other ongoing projects related to EnVar

- Next delivery will use ensembles with top at 0.1hPa, 39km resolution → ~10 min. for 4D-EnVar, 70 iter on 27 nodes
- Developing high-resolution 4D-EnVar for regional analysis
- Testing different strategies for recentering global EnKF members on a 4D-EnVar ensemble mean analysis
- Working towards atmosphere-ocean-ice strongly coupled DA (global coupled forecasts already operational)
- Many projects (e.g. FSOI, SDL, coupled DA) facilitated by work on increasing the modularization and generalization of DA Fortran code (continuous refactoring as needed)
- Hope to explore new ideas: treatment of horizontally correlated obs error, non-Gaussian errors (LPF), …

Extra slides

Scale-dependent covariance localization

Forecast impact

- Is it possible to do as well as SDL with a single localization approach?
- After all, perhaps our one-size-fits-all horizontal localization radius of 2800 km is not optimal
- Tried increasing and decreasing amount of localization and compare with using standard amount…

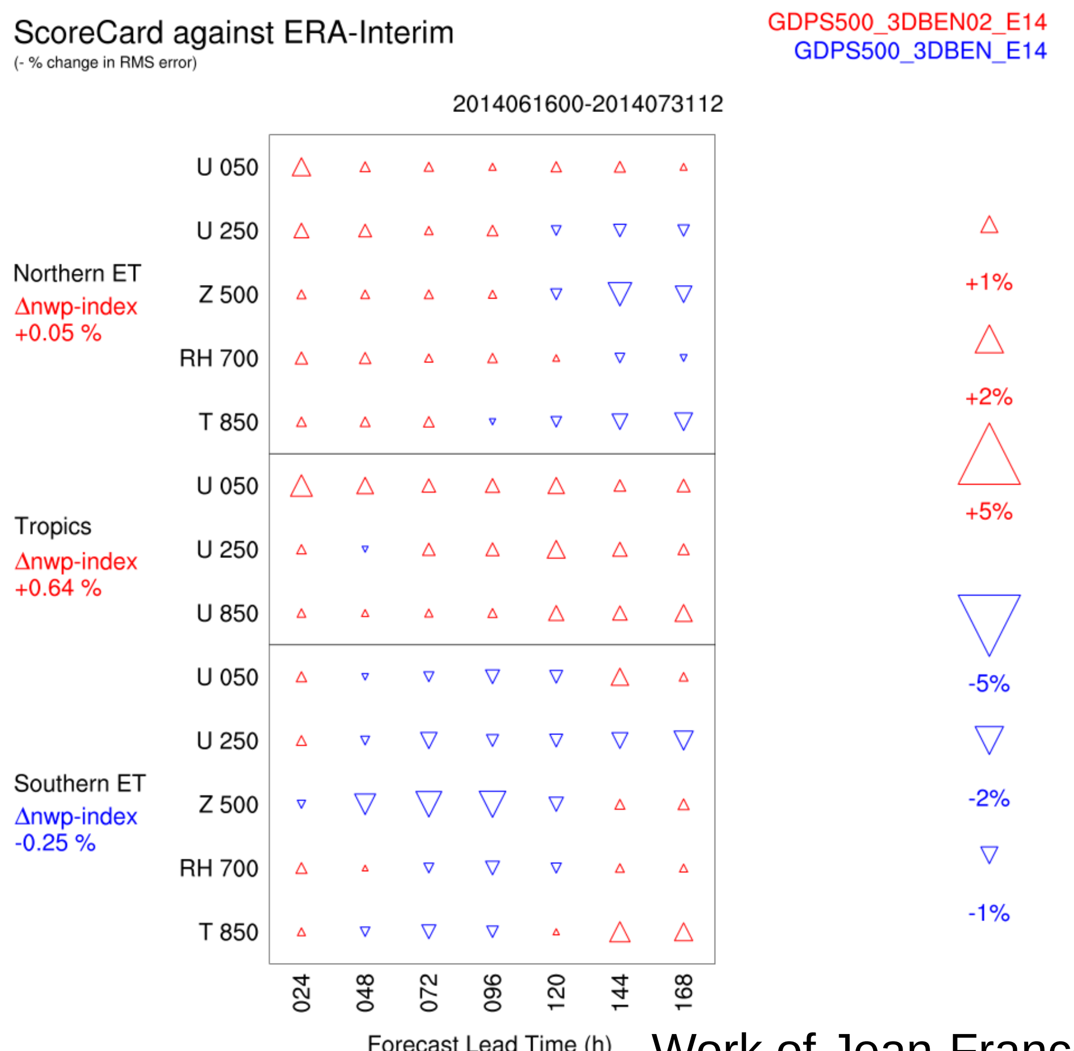
Work of Jean-Francois Caron

Scale-dependent covariance localization

Forecast impact – Comparison against ERA-Interim

2800 km is better

2400 km is better



Work of Jean-Francois Caron

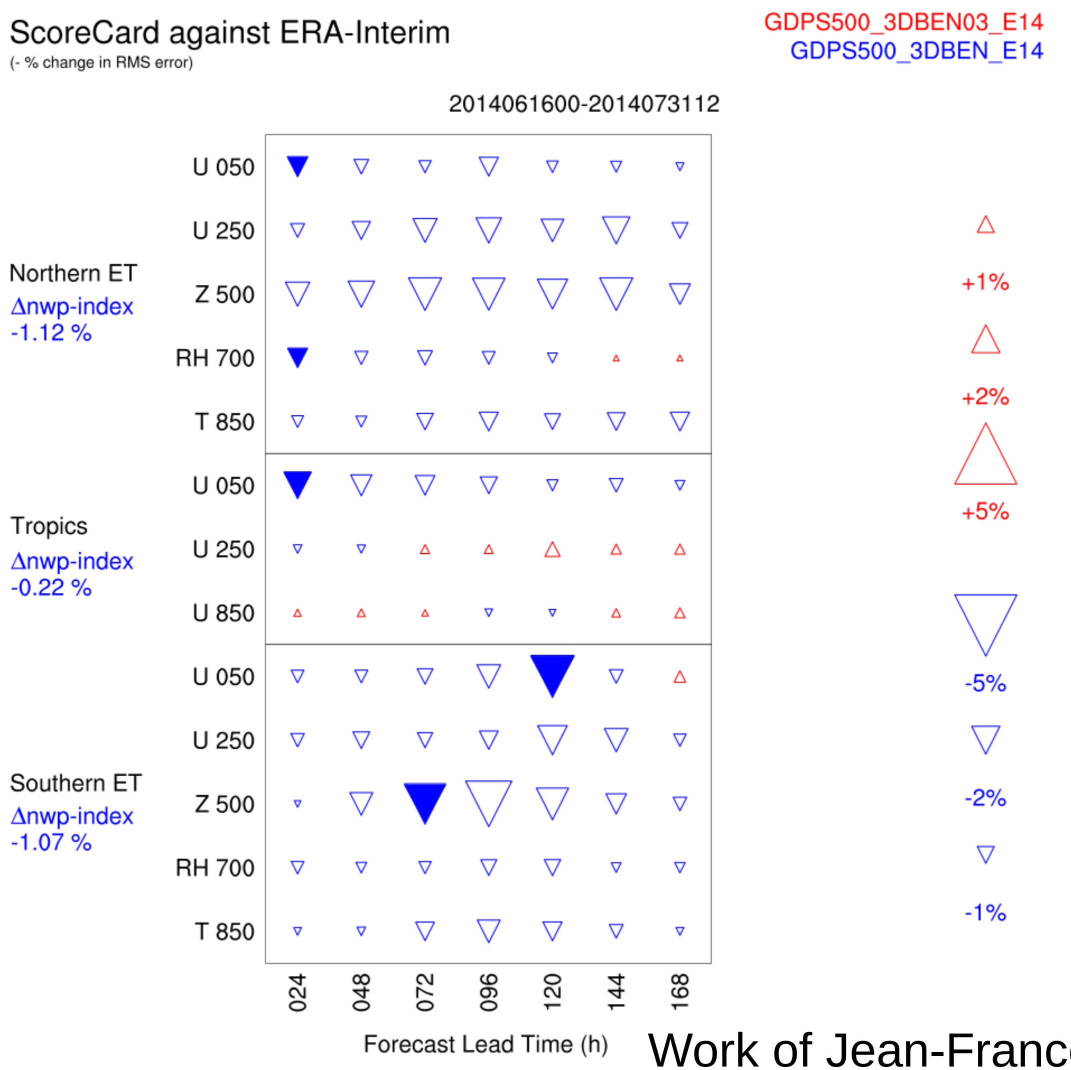
Environment and
Climate Change Ca

Scale-dependent covariance localization

Forecast impact – Comparison against ERA-Interim

2800 km is better

3300 km is better



Work of Jean-Francois Caron

Environment and Climate Change Ca

FSOI general approach

- Forecast error reduction from assimilating all observations:

$$\Delta e^2 = \left(\mathbf{e}_{t+\Delta t}^{fa}\right)^T \mathbf{C}\left(\mathbf{e}_{t+\Delta t}^{fa}\right) - \left(\mathbf{e}_{t+\Delta t}^{fb}\right)^T \mathbf{C}\left(\mathbf{e}_{t+\Delta t}^{fb}\right)$$

- This can be rewritten as a sum of contributions from each observation, allowing the calculation of contribution from any subset of obs:

$$\Delta e^2 \approx \sum_i \left(\mathbf{y}^o - H\left(\mathbf{x}^b\right)\right)_i \;\; \partial \Delta e^2 / \partial \mathbf{y}_i{}^o$$

- Where the sensitivity of the change in forecast error to each observation can be written as (using the chain rule):

$$\frac{\partial \Delta e^2}{\partial \mathbf{y}^o} = \left(\frac{\partial \mathbf{x}_t}{\partial \mathbf{y}^o}\right)\left(\frac{\partial \mathbf{x}_{t+\Delta t}}{\partial \mathbf{x}_t}\right)\left(\frac{\partial \Delta e^2}{\partial \mathbf{x}_{t+\Delta t}}\right)$$

$$\begin{matrix}\text{Sensitivity}\\ \text{wrt obs}\end{matrix} = \begin{pmatrix}\text{Adjoint}\\ \text{of DA}\end{pmatrix}\begin{pmatrix}\text{Adjoint}\\ \text{of Fcst}\end{pmatrix}\begin{pmatrix}\text{Sensitivity}\\ \text{wrt fcst.}\end{pmatrix}$$

Formulation (Idea from A. Lorenc working paper)

Change in forecast error at time t (with respect to norm $\mathbf{C}$) is given by:

$$\Delta e^2 = \left(\mathbf{e}_t^{fa}\right)^T \mathbf{C}\left(\mathbf{e}_t^{fa}\right) - \left(\mathbf{e}_t^{fb}\right)^T \mathbf{C}\left(\mathbf{e}_t^{fb}\right)$$

Denote gradient of this with respect to any quantity as: $\widehat{(\cdot)} = \partial \Delta e^2 \,/\, \partial(\cdot)$

Write Δe^2 as a sum of contributions from each obs:

$$\Delta e^2 \approx \left(\mathbf{y}^o - H\left(\mathbf{x}^b\right)\right)^T (\hat{\mathbf{y}}^o)$$

Based on extended ensemble at forecast time, replaces $\mathbf{B}_0^{T/2}\mathbf{M}^T$

Where:

$$\hat{\mathbf{y}}^o = \mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}\hat{\mathbf{a}}$$ →sensitivity wrt observations

$$\hat{\mathbf{v}} = \mathbf{B}_t^{T/2}\widehat{\boldsymbol{\delta}\mathbf{x}_t}$$ →sensitivity wrt control vector

$$\widehat{\boldsymbol{\delta}\mathbf{x}_t} = \mathbf{C}\left(\mathbf{e}_t^{fa} + \mathbf{e}_t^{fb}\right)$$ →sensitivity wrt forecast

And $\hat{\mathbf{a}}$ is obtained by minimizing the cost function (very similar to EnVar):

$$J(\hat{\mathbf{a}}) = \frac{1}{2}(\hat{\mathbf{a}} - \hat{\mathbf{v}})^T(\hat{\mathbf{a}} - \hat{\mathbf{v}}) + \frac{1}{2}\left(\mathbf{H}\mathbf{B}_0^{1/2}\hat{\mathbf{a}}\right)^T \mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}_0^{1/2}\hat{\mathbf{a}}\right)$$

Results

Actual and estimated change in <u>12h and 24h forecast error</u> from assimilating observations at 0Z and 6Z