Earth System Research Laboratory
Physical Sciences Division

Developments in Ensemble DA

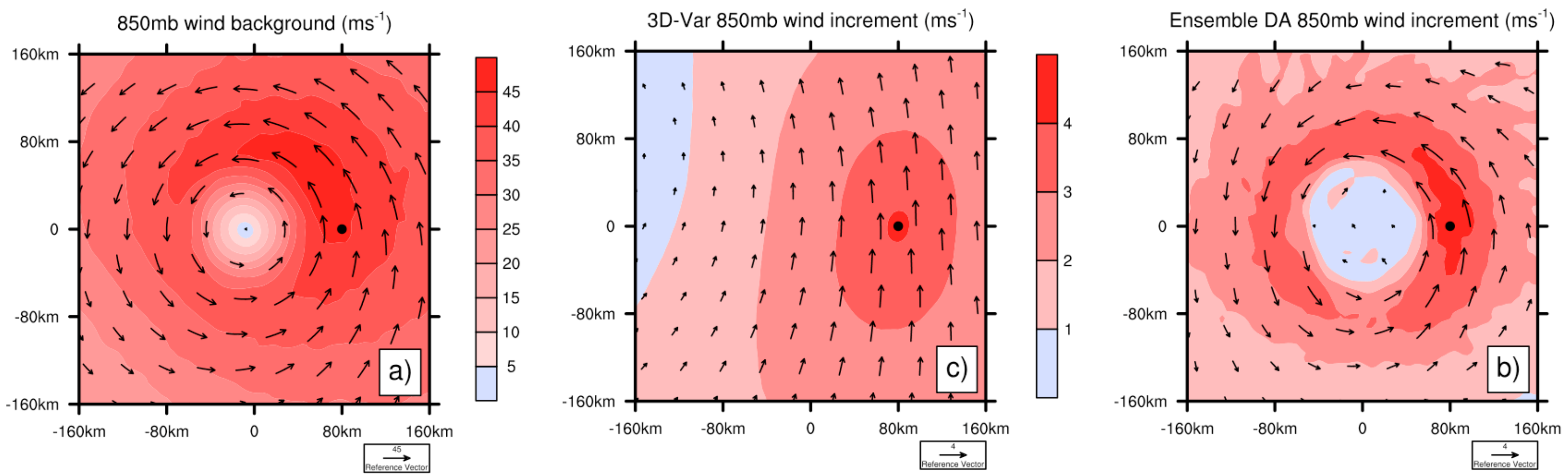
Jeff Whitaker
NOAA Earth System Research Lab, Boulder, CO USA
<jeffrey.s.whitaker@noaa.gov>

What makes the EnKF different?

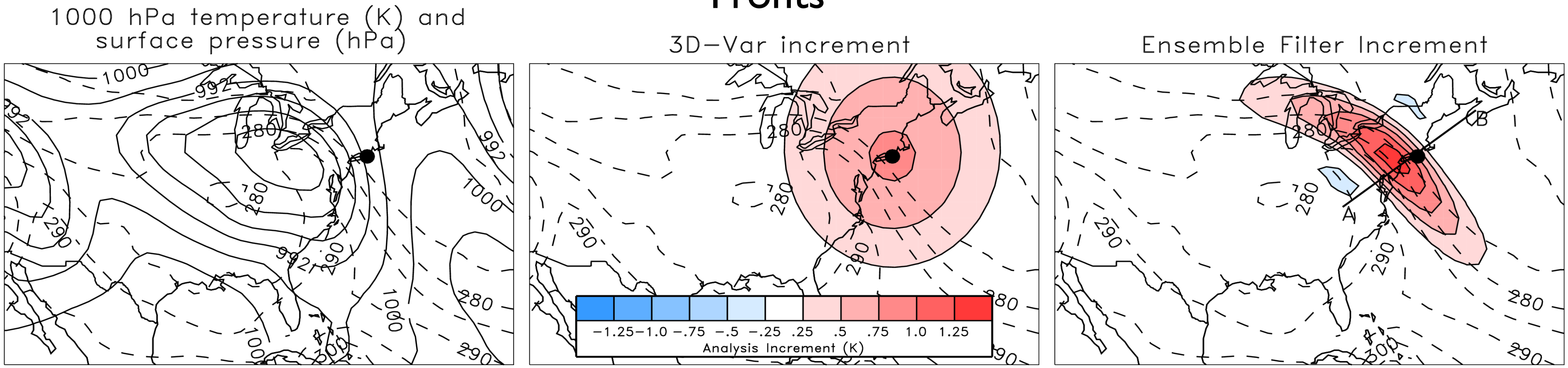
- Data assimilation requires "background-error covariances"
  - Describe error characteristics of first-guess forecast.
  - Determines how forecast and new observations are blended.
- In EnKF, these are *estimated from an ensemble*.
  - They can change with the dynamical situation.
- This leads to:
  - Improved quality analyses.
  - "Situation-dependent" estimates of analysis uncertainty are captured from ensemble of analysis states.

Benefits of Flow-Dependent Background Errors: Idealized Examples

Hurricanes



Fronts



Data assimilation terminology

- **y** : Observation vector (weather balloons, satellite radiances, etc.) with expected error ε.
- **x** : model state vector. Superscript *b* denotes prior (background), *a* posterior (analysis), *t* "truth".
- **H** : operator to convert model state to observation space, i.e. $\mathbf{y}=\mathbf{H}\mathbf{x}^t + \varepsilon$
- **R** : Observation-error covariance matrix, i.e. $\langle \varepsilon\varepsilon^T \rangle$
- $\mathbf{P}^b$: Background-error cov matrix, s.t. $\mathbf{x}^t = N(\bar{\mathbf{x}}^b, \mathbf{P}^b)$

The Kalman Filter (KF)

Assume:

Gaussian forecast errors $\mathbf{x}^t = N(\bar{\mathbf{x}}^b, \mathbf{P}^b)$

Gaussian observation errors $\epsilon = N(0, \mathbf{R})$

Bayes rule $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ implies:

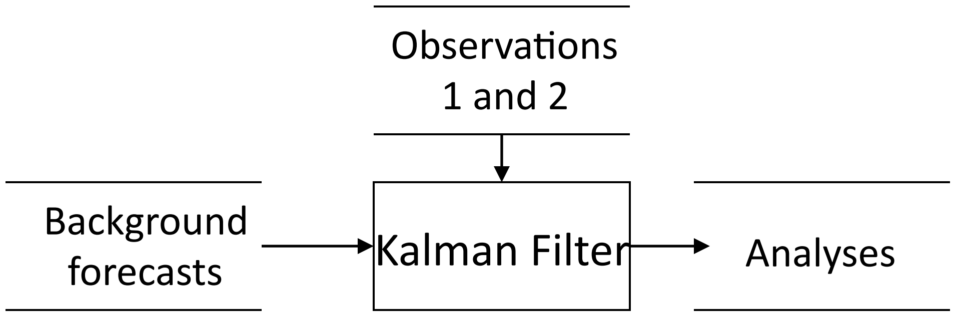
$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^b\right); \ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^b$$

$$\text{where } \mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

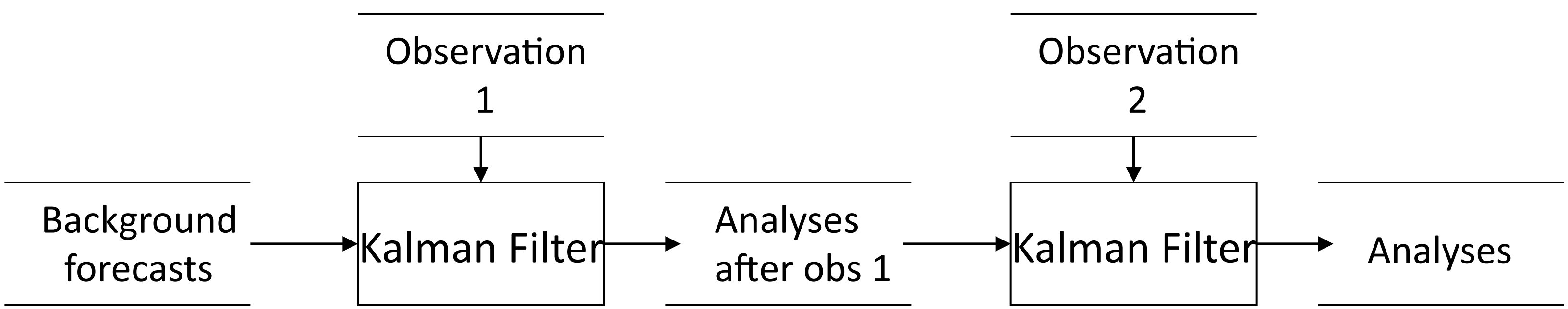
- Computationally hard since $\mathbf{P}^b$ is N_x x N_x (N_x = dim $\mathbf{x}$).
- **EnKF** uses sample of **P** of size N_e , converges to **KF** as N_e approaches N_x (with linearity, Gaussianity, …).

Computational shortcuts in EnKF:
(1) serial processing of observations (requires observation error covariance **R** to be diagonal)

Method 1



Method 2



Computational shortcuts in EnKF:
(2) Simplifying Kalman gain calculation

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

$$\textit{define} \quad \overline{H\mathbf{x}^{\mathrm{b}}} = \frac{1}{m} \sum_{i=1}^{m} H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}$$

$$\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}} \right) \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right)^{\mathrm{T}}$$

$$H\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right) \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right)^{\mathrm{T}}$$

The key here is that the huge matrix $\mathbf{P}^{\mathrm{b}}$ is never explicitly formed

Computational shortcuts in EnKF:
(3) Covariance localization

- Calculate covariances only between "nearby" model priors and observation priors.
  - Assumes large scale separation means small covariance.
- Since $N_x >> N_e$ covariance estimate is rank deficient anyway.
  - Noisy covariance estimates will cause $\mathbf{P}^a$ to be underestimated.
  - To reduce sampling noise, taper covariance estimate as a function of separation (using Gaussian-ish function).
  - Increases effective rank of sample covariance matrix.

This (and covariance inflation) is the key to making the whole thing work!

Algorithmic details

Basically two types of EnKF codes are being used:

- ✓ '**stochastic**' EnKF (original formulation by Houtekamer and Mitchell, 1998 *MWR*) treats obs as ensemble by adding N(0,**R**) noise. This is needed to prevent underestimation of $\mathbf{P}^a$ when every member updated with the same KF update equations.

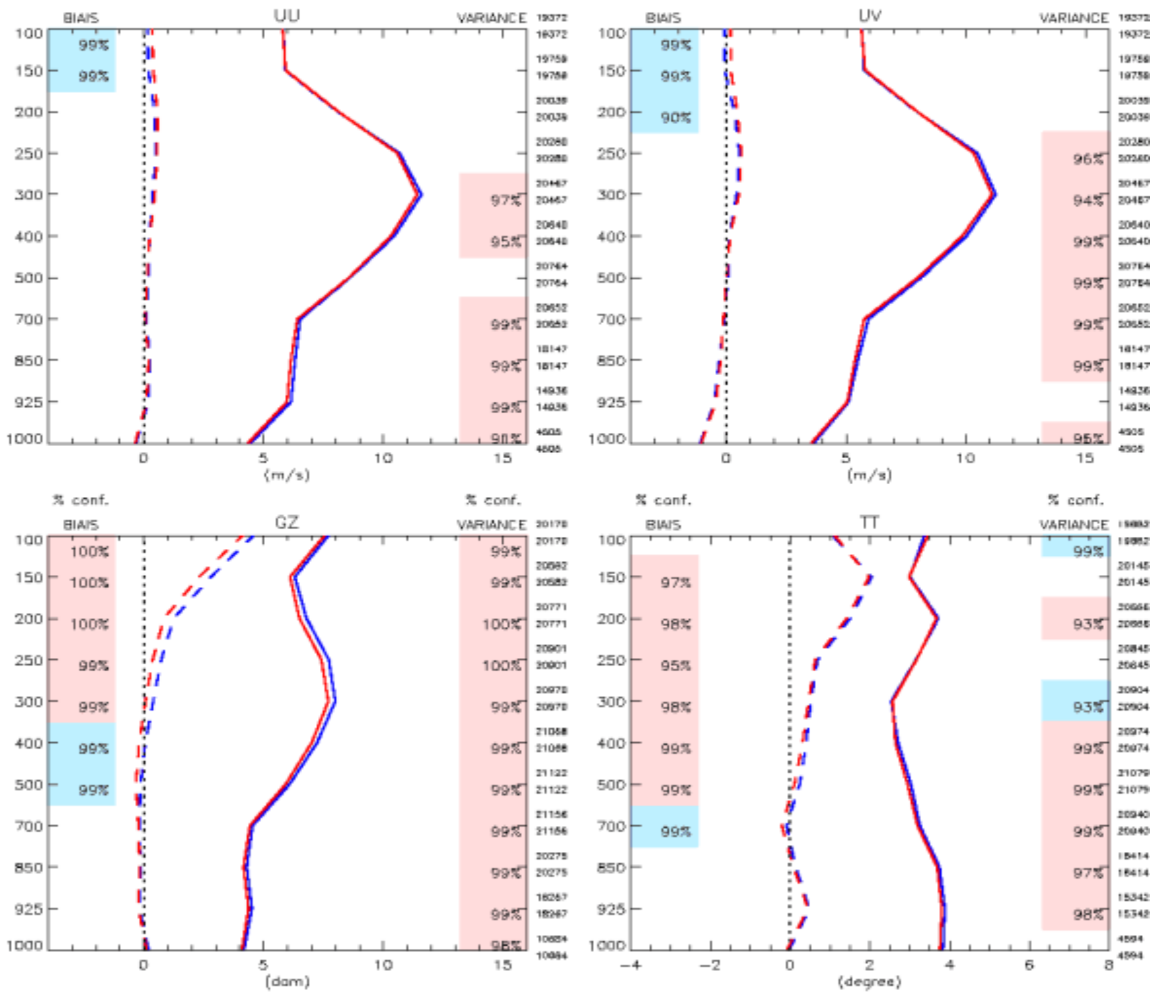
- ✓'**deterministic**' EnKF (LETKF, Hunt et al 2007, Physica D; serial EnSRF, Whitaker and Hamill 2002 MWR) avoids this by updating ensemble perturbations separately from mean in such a way that $\mathbf{P}^a$ consistent with KF is obtained.

EnKF - Current state of the art

Env. Canada EnKF vs 4DVar (Buehner et al MWR, 2010)

- Fit of 120-h control forecasts to radiosondes (NH) EnKF red, 4DVar blue

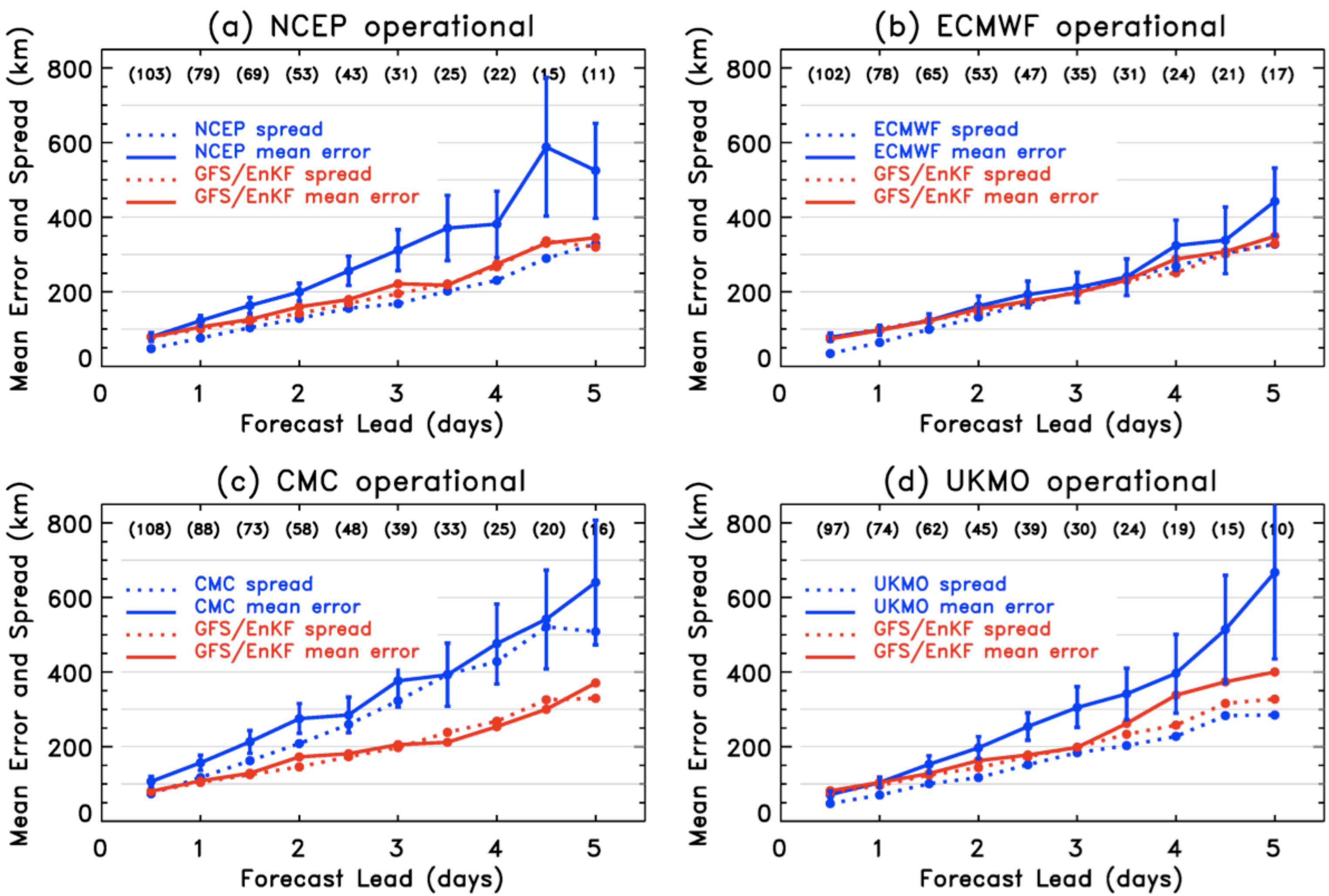
- EnKF run at 100 km resolution, 4DVar 35 km (outer loop), 150 km (inner loop).



EnKF performance nearly identical to operational 4DVar

EnKF - Current state of the art

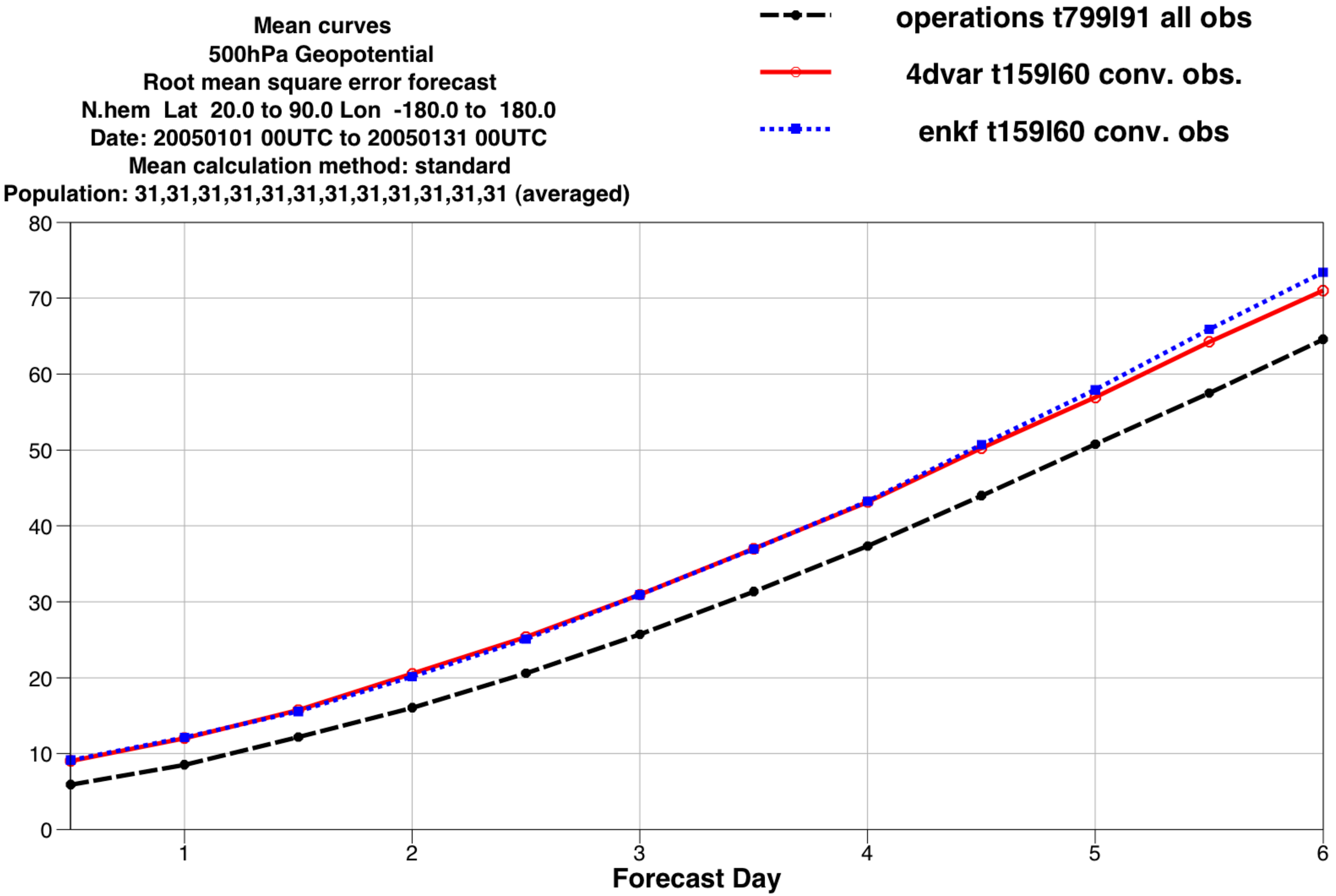
Global ensemble hurricane track forecasts (Hamill et al MWR, 2010)



GFS/EnKF ensemble better than UKMO, Canada, NCEP, close to EC.

EnKF - Current state of the art

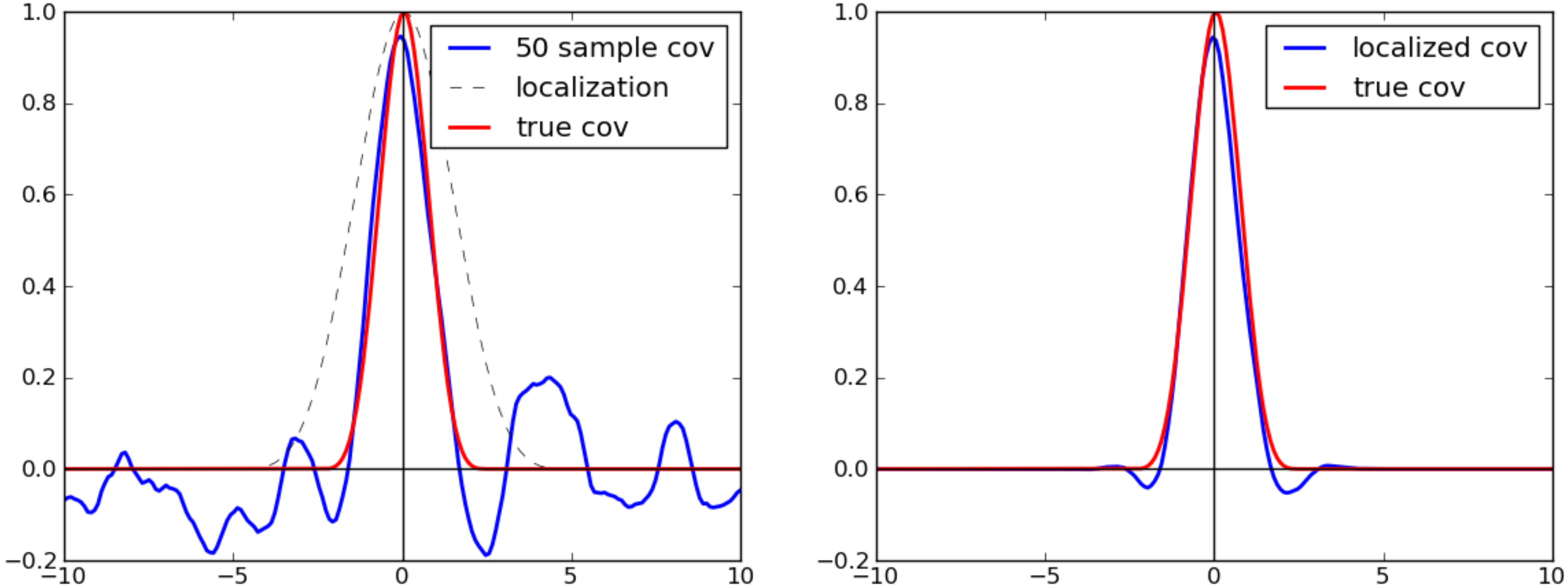
ECMWF EnKF vs 12-h 4DVar (T159), conv obs only



What makes the EnKF suboptimal?

- Var and ensemble methods both attempt to solve the KF eqns, but take different shortcuts!
- EnKF is optimal IFF:
  - Observation and forecast errors Gaussian
  - Ensemble size large enough so that sampling errors are small ($N_x \sim N_e$) ***covariance localization***
  - All sources of error sampled by ensemble, including model errors! ***covariance inflation***
- EnKF development is focused on better ways to deal with sampling and model errors, and other sources of un(der)represented errors.

Covariance localization

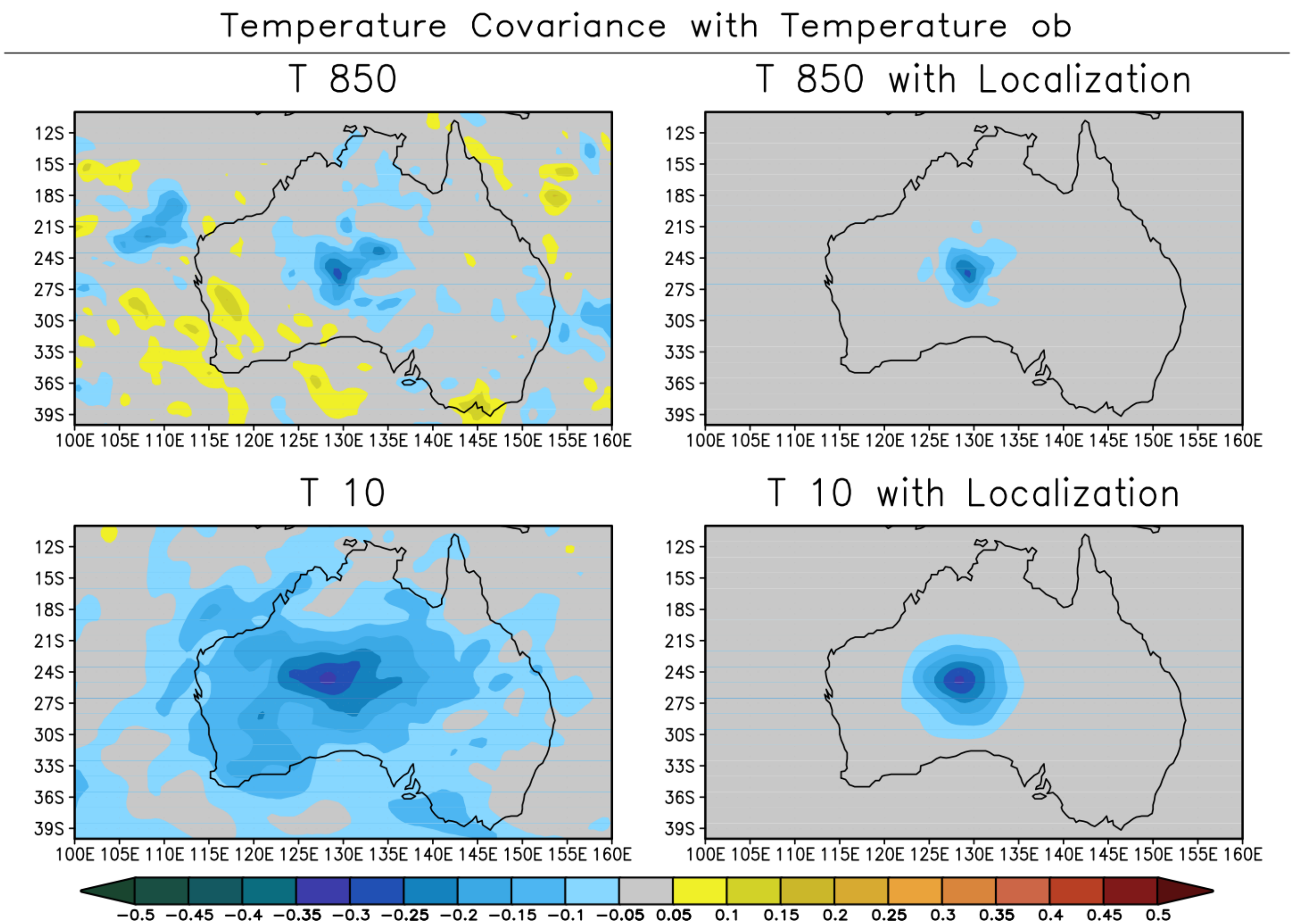


- *statistical noise degrades the spread of information from observation locations to model variables.*
- *signal-to-noise small when covariance is small.*
- *Methods used now are not flow-dependent.*

Localization: is flow-dependence needed?

Scales of covariances can dependent on flow, localization should too.

Bishop & Hodyss, 2009, Tellus present a strategy for doing this



Flow-Adaptive Localization based on sample correlations (Bishop + Hodyss 2011)

Localization function based on sample correlations computed using smoothed, normalized perturbations.

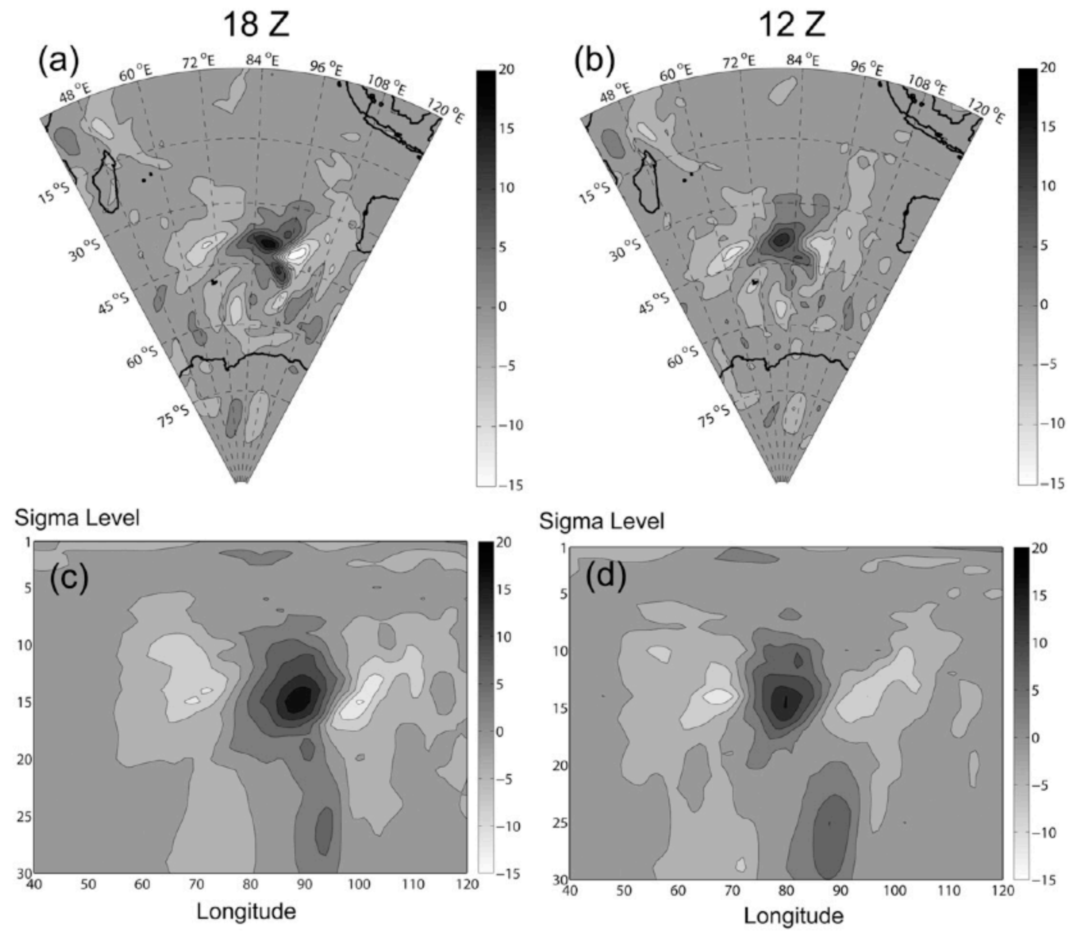


FIG. 3. Unlocalized ensemble covariance function of meridional wind at 1800 and 1200 UTC with 1800 UTC meridional wind variables at 40°S, 90°E and σ-level 15 (about 400 hPa). The ensemble has 128 members. The horizontal cross sections at σ-level 15 of the covariance function at (a) 1800 and (b) 1200 UTC. The zonally oriented vertical cross sections at 40°S of the covariance function at (c) 1800 and (d) 1200 UTC.

Flow-Adaptive Localization based on sample correlations (Bishop + Hodyss papers)

Localization function based on sample correlations² computed using smoothed, normalized perturbations.

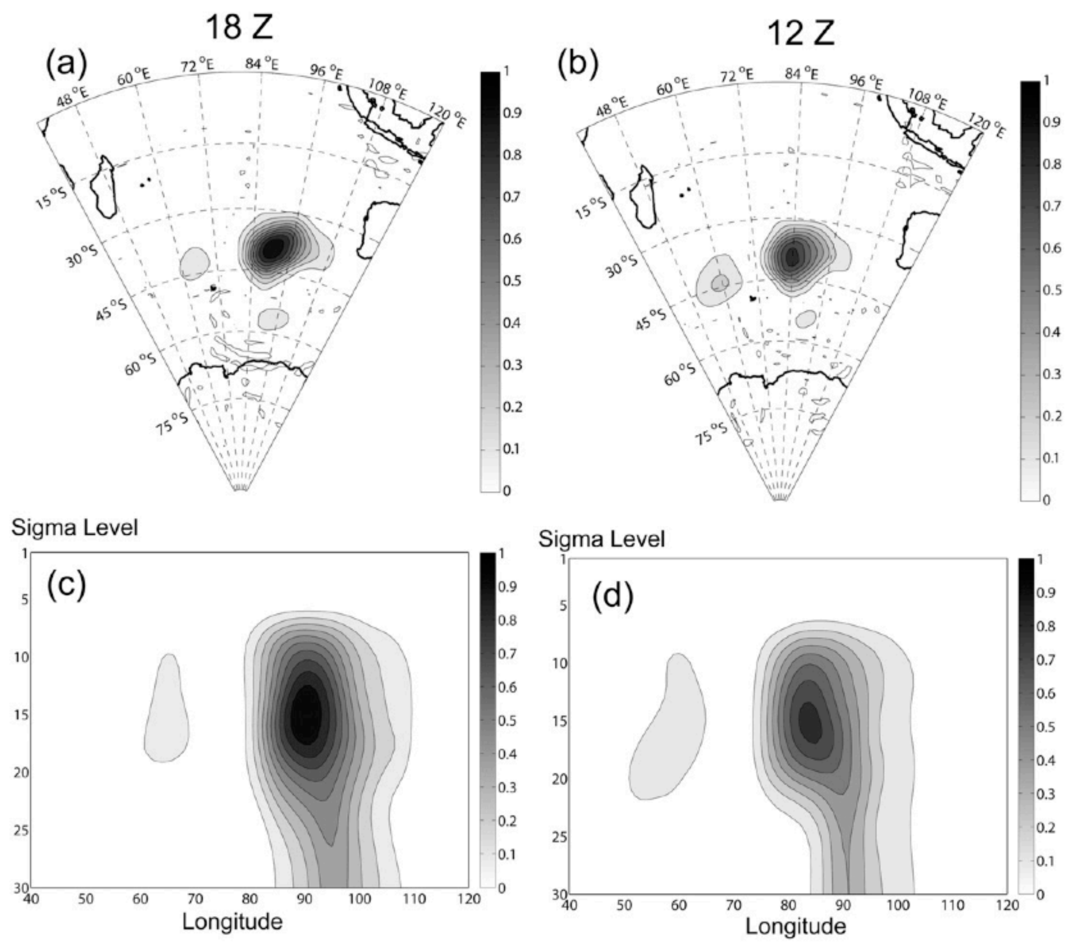
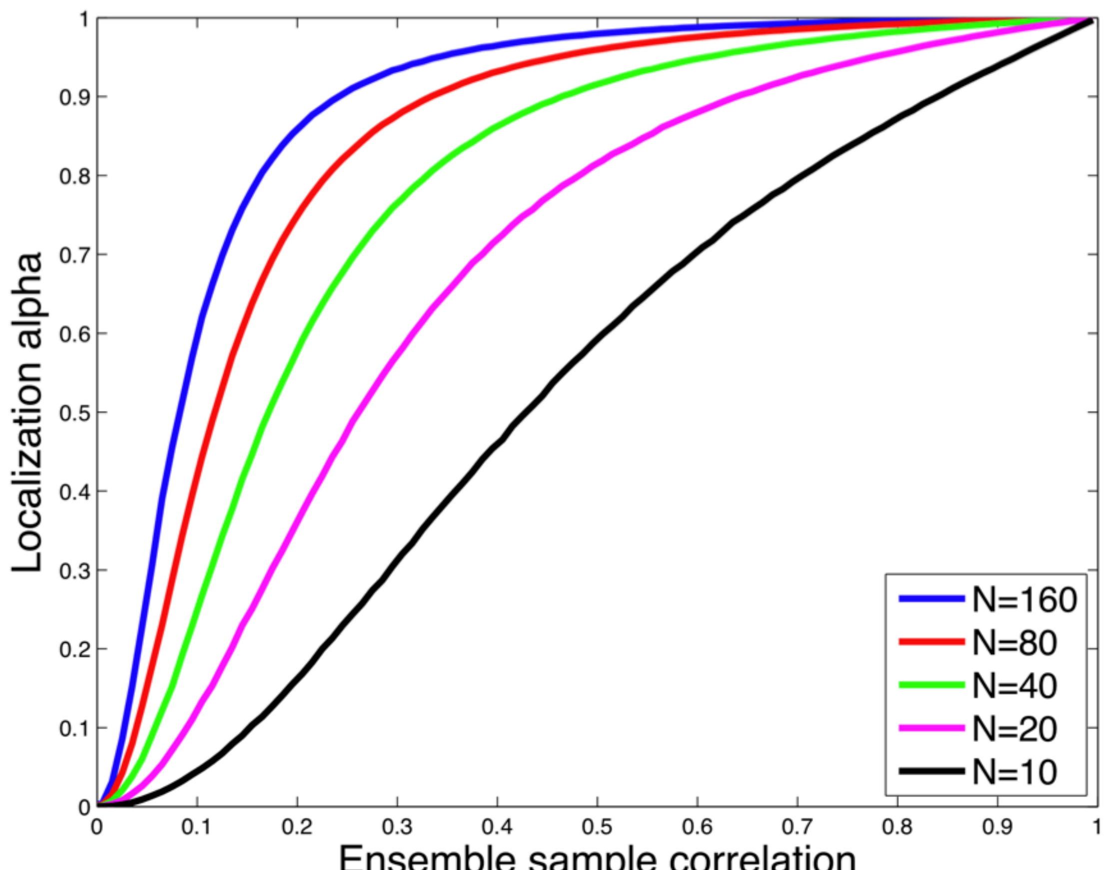


FIG. 4. The AECL function for the raw covariance function shown in Fig. 3 is shown. This localization function is the element-wise square of the correlation function of a 128-member ensemble of smoothed and normalized streamfunction fields.

Simpler version proposed by Jeff Anderson

(2011 AMS talk *Localization and Correlation in Ens. Kalman Filters*)

Localization α as function of ensemble size N and sample correlation $\hat{r}$.



Other localization issues

- Localization should really be done in model space, not localization space (Campbell et al MWR, 2009)
  - May be important for accurate obs with complicated forward operators (radiances?)
  - Ensemble Var systems localize in model space, EnKF localizes in ob space (because of the way covariances are calculated, see slide 10).
- What about localization in 'variable' space? (Kang et al, JGR 2011)
  - Covariance between observation priors and model priors can be essentially all sampling noise even if physical separation is zero (e.g. tracer observation, temp variable)
  - Need a more general concept of "distance", or a method like Bishop+Hodyss that uses sample correlations.
- What to do in LETKF (when obs. prior/model prior covariance not explicitly computed)?
  - Local analyses already deals with rank deficiency, like 'box method' in OI. Abrupt transitions can lead to noisy increments.
  - To get smoother increments, can also apply 'observation error localization' - similar to covariance localization, but instead of modulating covariances increase obs. error as a function of distance from analysis point (Greybush et al, MWR 2011)

Un(der)-represented error sources in an EnKF ensemble

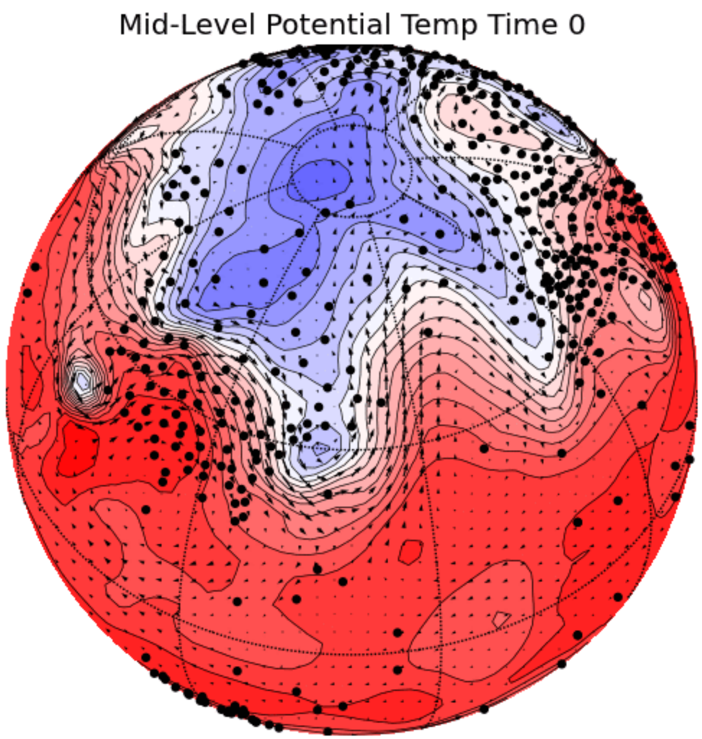
Model error	$\mathbf{M}\mathbf{x}_\mathrm{a}$
Sampling error	$\frac{1}{N}\sum_{j=1}^{N}(N << \infty)$
Observation error	$\mathbf{R}$
Boundary condition error	$T(z=0) \Rightarrow T_s$
Forward operator error	$\mathbf{H}\mathbf{x}_\mathrm{b}$

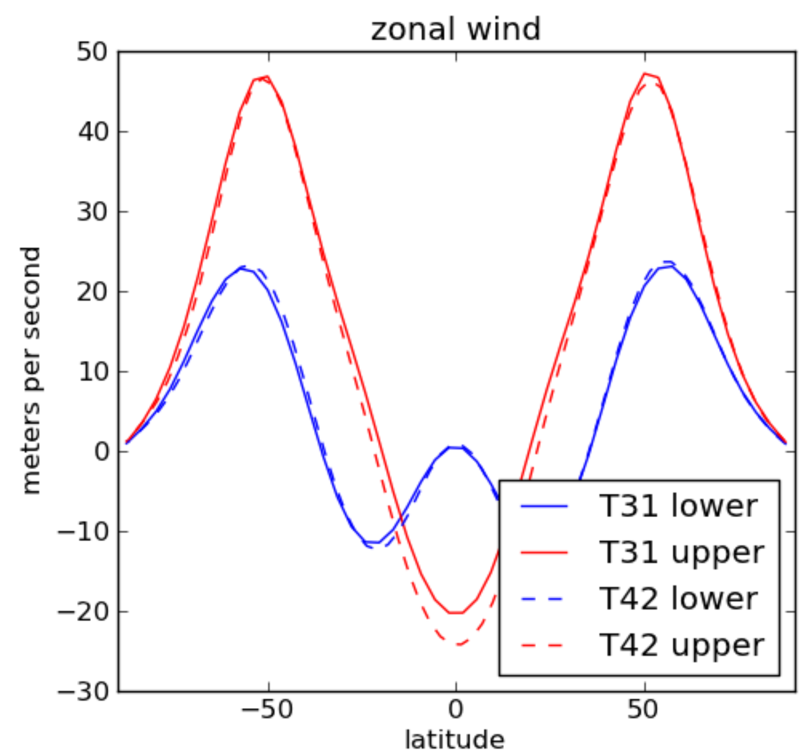
Neglecting or under-representing any of these will cause assimilation to give too little weight to observations

Idealized expts with 2-level PE model

(from WGNE model uncert. workshop)

- 2-level PE model on a sphere (Lee and Held, 1993 with parameters as in Hamill and Whitaker, 2010).
- 511 12-hourly obs of geopotential height at sonde locations (error = 10 m)
  - 20 member ensemble, serial determinstic (i.e. square-root) EnKF.
  - 1000 assimilation cycles, 3500 km localization (none in vertical)
- Truth from T42 nature run, assimilation with T31 model. Only sources of DA error are model error and sampling error.





Multiplicative inflation

- Simple constant inflation not suitable when observing network and dynamics vary in space and/or time.
- Both sampling error and model error are expected to be a larger fraction of the total background error where observations have a larger impact (where σ_b/σ_a is large).
- We use “relaxation to prior spread“ (RTPS)

$$\sigma^a \leftarrow (1-\alpha)\sigma^a + \alpha\sigma^b$$

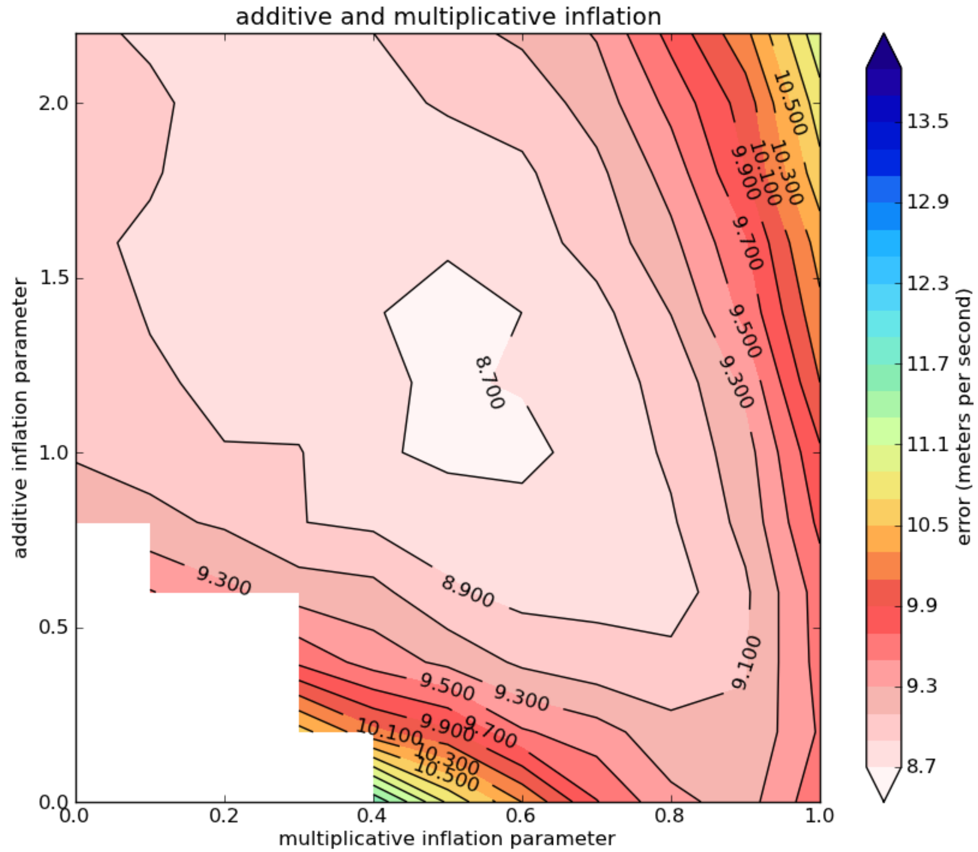
which implies $\mathbf{x}_i'^a \leftarrow \mathbf{x}_i'^a \sqrt{\alpha \frac{\sigma^b - \sigma^a}{\sigma^a} + 1}$

Additive inflation

- Add random samples from a specified distribution to each ensemble member after the analysis step.
- Env. Canada uses random samples of isotropic 3DVar covariance matrix.
- Here we use a dataset of 12-h forecast errors with the T31 model in which the initial conditions are perfect (T31 truncated states from the T42 nature run).

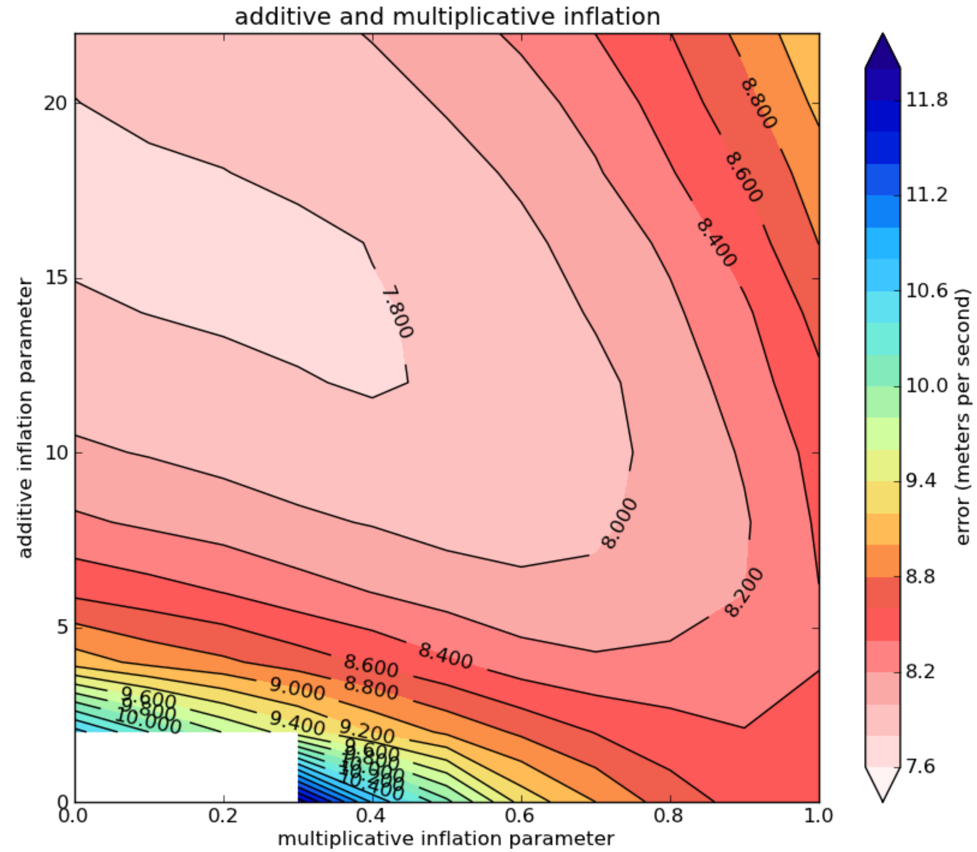
Multiplicative + Additive inflation

- Additive inflation alone outperforms multiplicative inflation alone (compare values y-axis to values along x-axis)
- A combination of both is better than either alone.
- Multiplicative and additive inflation representing different error sources in the DA cycle?



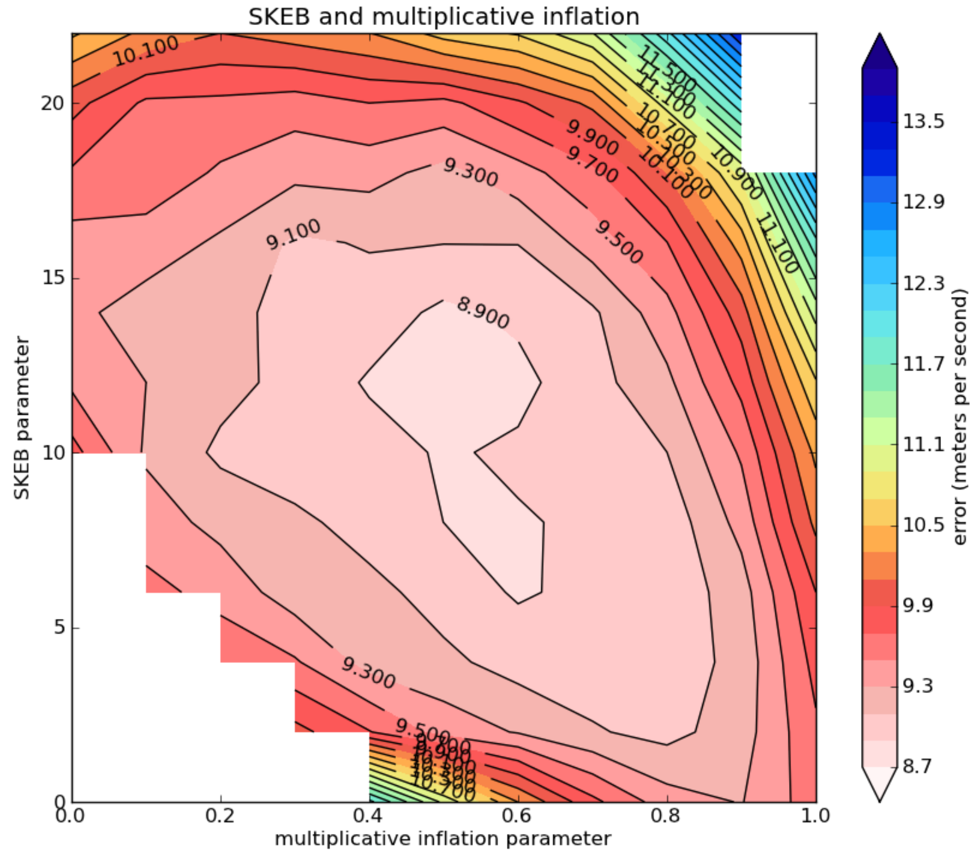
Large ensemble results (Additive + Multiplicative Inflation)

- 200 instead of 20 members, with model error. Min error reduced from 8.7 to 7.7.
- When sampling error is reduced, additive inflation alone outperforms combination of add +mult inflation.
- Suggests that additive inflation is better at capturing model-related errors.



Multiplicative inflation + *Stochastic Kinetic Energy Backscatter* (SKEB)

- A combination of SKEB and multiplicative inflation works better than either alone.
- SKEB alone comparable to multiplicative inflation alone (compare values along x and y axes).
- Results are slightly inferior to those obtained using additive + multiplicative inflation.
- y-axis is amplitude of random pattern (σ) – results do not change much if p (power law) or time-scale (τ) are varied.



Experiences with Env. Canada system

(Houtekamer, Mitchell and Deng, MWR July 2009)

- Operational EnKF tested with
  - Multiple parameterizations
  - SKEB (stochastic kinetic energy backscatter)
  - SPPT (stochastically perturbed physics tend)
  - Additive inflation (isotropic covariance structure)
  - Multi-physics plus additive inflation
- Most of these designed to represent specific model errors, additive inflation is ‘catch-all’ to represent what’s left.
- Multiplicative inflation not tested.

Experiences with Env. Canada system

(Houtekamer, Mitchell and Deng, MWR July 2009)

configuration	O-F (energy norm)	Energy spread in ob space
Additive inflation	3.1388	2.0622
Multi-physics	3.2978	1.2773
SKEB	3.4348	1.2671
SPPT	3.3899	1.1670
Multi-physics + add. Infln.	3.0846	2.1335
SKEB + SPPT	3.3352	1.3608
SKEB+SPPT+Mult-physics +rescaled additive infln.	3.0940	2.1092

- Biggest impact from ad-hoc additive inflation.
- Addition of multi-physics improves assimilation slightly.
- SPPT and SKEB have less impact (tuned for EPS?, model error not dominant?)

Summary

- EnKF algorithms now fairly mature, are highly scalable.
- Research now focused on treatment of sampling and model error (and other un(der) represented sources of error in the background ensemble).
  - Flow-adaptive localization has not yet been shown to out-perform non-adaptive localization in NWP systems.
  - Multiplicative and additive inflation are a tough baseline to beat.
- Now implemented in operations at Env Canada. Hybrid Var/EnKF system implemented at UKMO, NCEP in 2012. ECMWF has an experimental EnKF system.

Hybrid Var/EnKF - best of both worlds?

Features from EnKF	Features from VAR
Extra flow-dependence in $\mathbf{P}^b$	Localization done correctly (in model space)
More flexible treatment of model error (can be treated in ensemble)	Reduction in sampling error in time-lagged covariances (full rank evolution of $\mathbf{P}^b$ in assimilation window in 4DVar).
Automatic initialization of ensemble forecasts, propagation of covariance info from one cycle to the next.	Ease of adding extra constraints to cost function