Ensemble Forecasting and Data Assimilation: Two Problems with the Same Solution?

Eugenia Kalnay
Department of Meteorology and Chaos Group
University of Maryland

References and thanks:

Ott, Hunt, Szunyogh, Zimin, Kostelich, Corazza, Kalnay, Patil, Yorke, 2003: Local Ensemble Kalman Filtering, MWR, under review.

Patil, Hunt, Kalnay, Yorke and Ott, 2001: Local low-dimensionality of atmospheric dynamics, PRL.

Corazza, Kalnay, Patil, Yang, Hunt, Szunyogh, Yorke, 2003: Relationship between bred vectors and the errors of the day. NPG.

Kalnay, 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341 pp.

Data assimilation

- Combination of a forecast (background) T_b with observations T_o to give a "best" estimate of the true state of the atmosphere (analysis) T_a.
- We need information about the errors: σ_b^2, σ_o^2

$$T_a = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2} T_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} T_o \qquad \frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

We can now use the analysis as initial conditions for the next forecast, get new a new observation, and repeat, this is called the "analysis cycle".

Ensemble Forecasting

- Normally a single control forecast is integrated from the analysis (initial conditions)
- In ensemble forecasting several forecasts are run from slightly perturbed initial conditions (or with different models)
- The spread among ensemble members gives information about the forecast errors

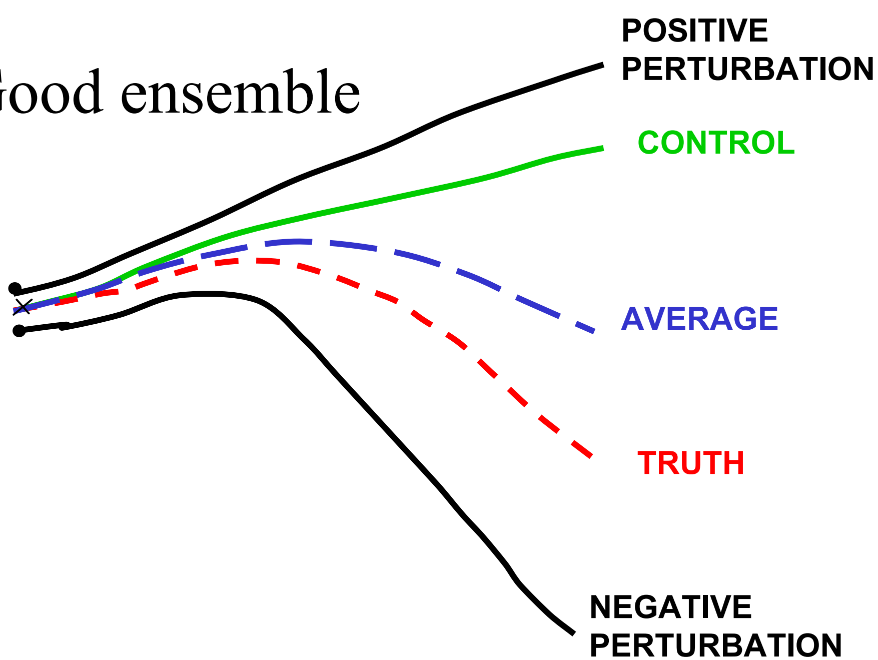
Ensemble forecasts

An ensemble forecast starts from initial perturbations to the analysis…

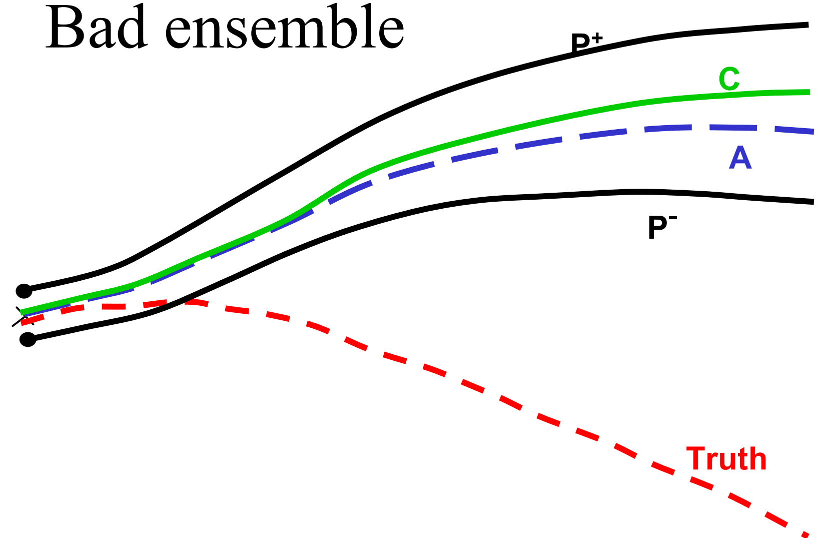
In a good ensemble "truth" looks like a member of the ensemble

The initial perturbations should reflect the analysis "errors of the day"

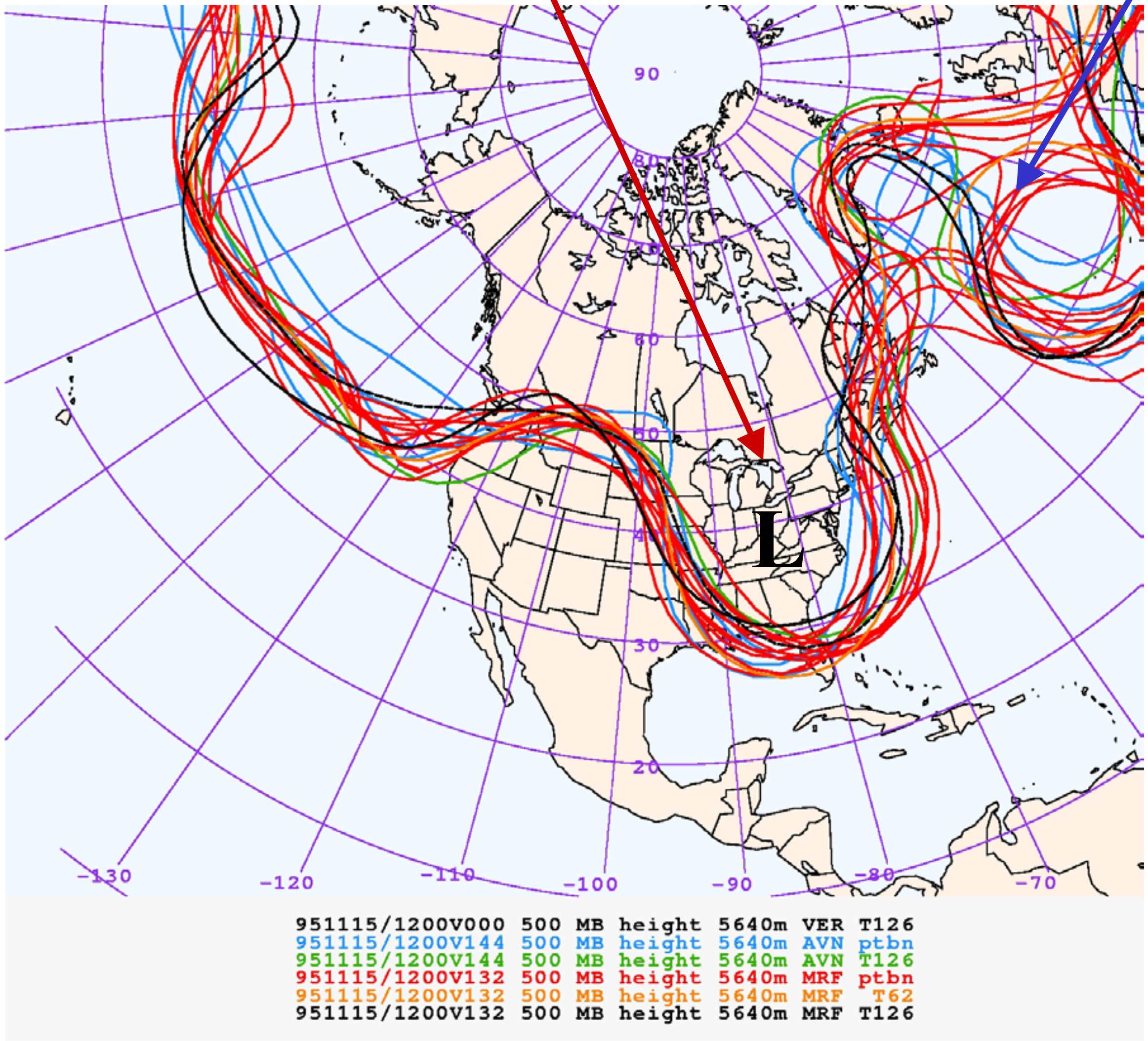
Good ensemble



Bad ensemble

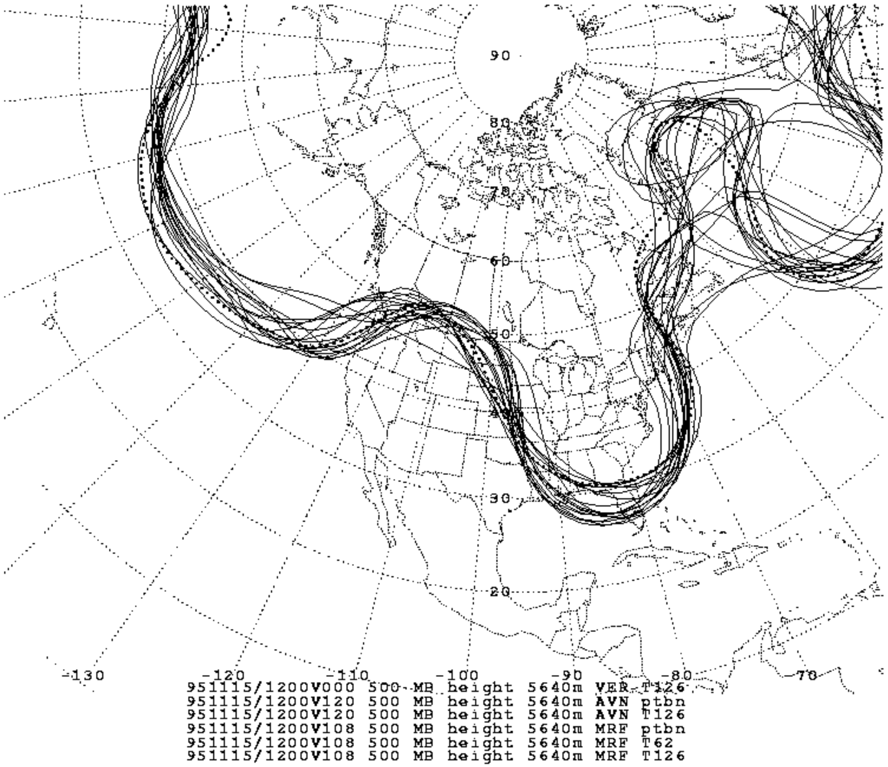


Example of a very predictable 6-day forecast, with "errors of the day"



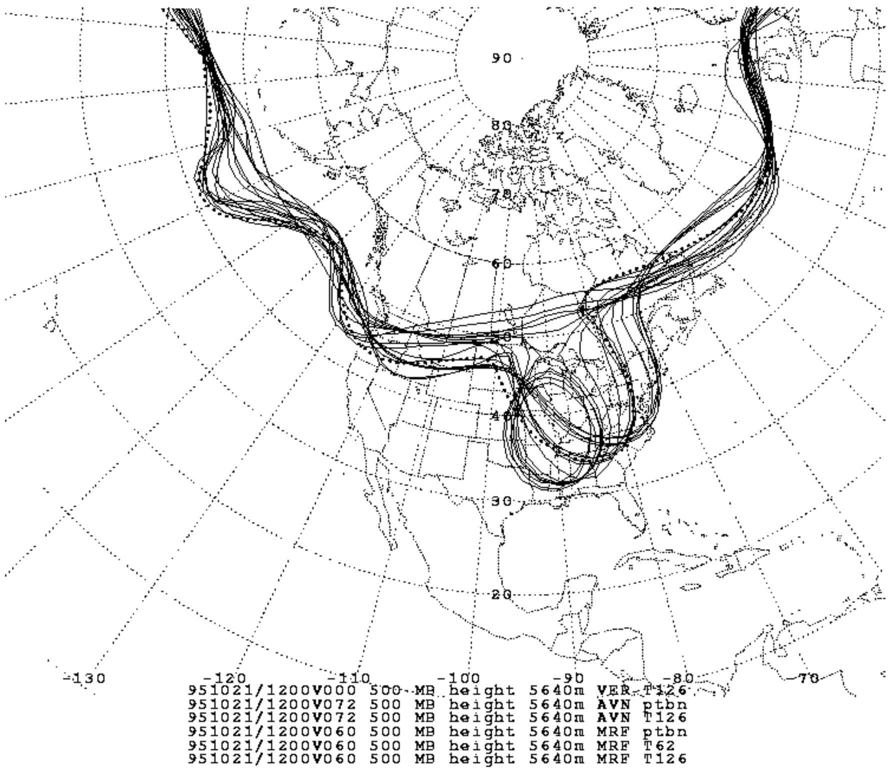
Errors of the day tend to be localized and have simple shapes
(Patil et al, 2001)

The errors of the day are instabilities of the background flow. At the same verification time, the forecast uncertainties have *the same shape*



4-day forecast verifying on the same day

Strong instabilities of the background tend to have simple shapes (perturbations lie in a low-dimensional subspace)



2.5 day forecast verifying on 95/10/21.

Note that the bred vectors (difference between the forecasts) lie on a 1-D space

It makes sense to assume that the errors in the analysis (initial conditions) have the same shape as well:

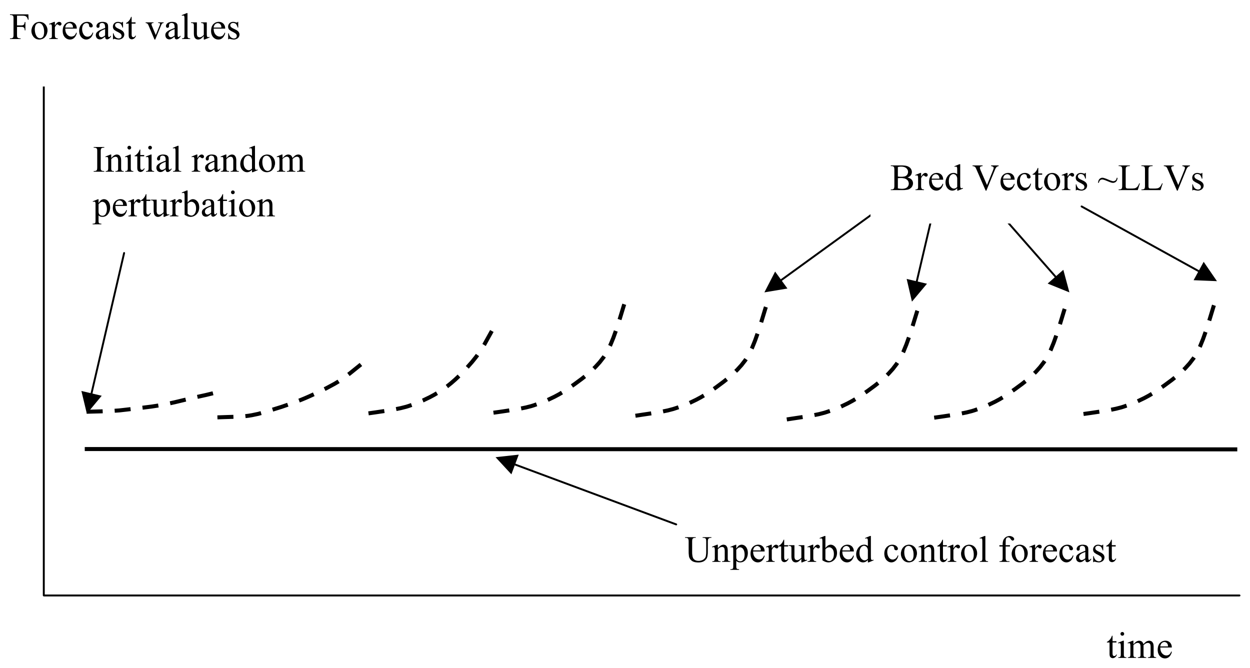
the errors lie in the subspace of the bred vectors

Errors of the day

- They are instabilities of the background flow
- They dominate the analysis and forecast errors
- They are not taken into account in data assimilation except for 4D-Var and Kalman Filtering (very expensive methods)
- Their shape can be estimated with breeding
- Their shape is frequently simple (low dimensionality, Patil et al, 2001)

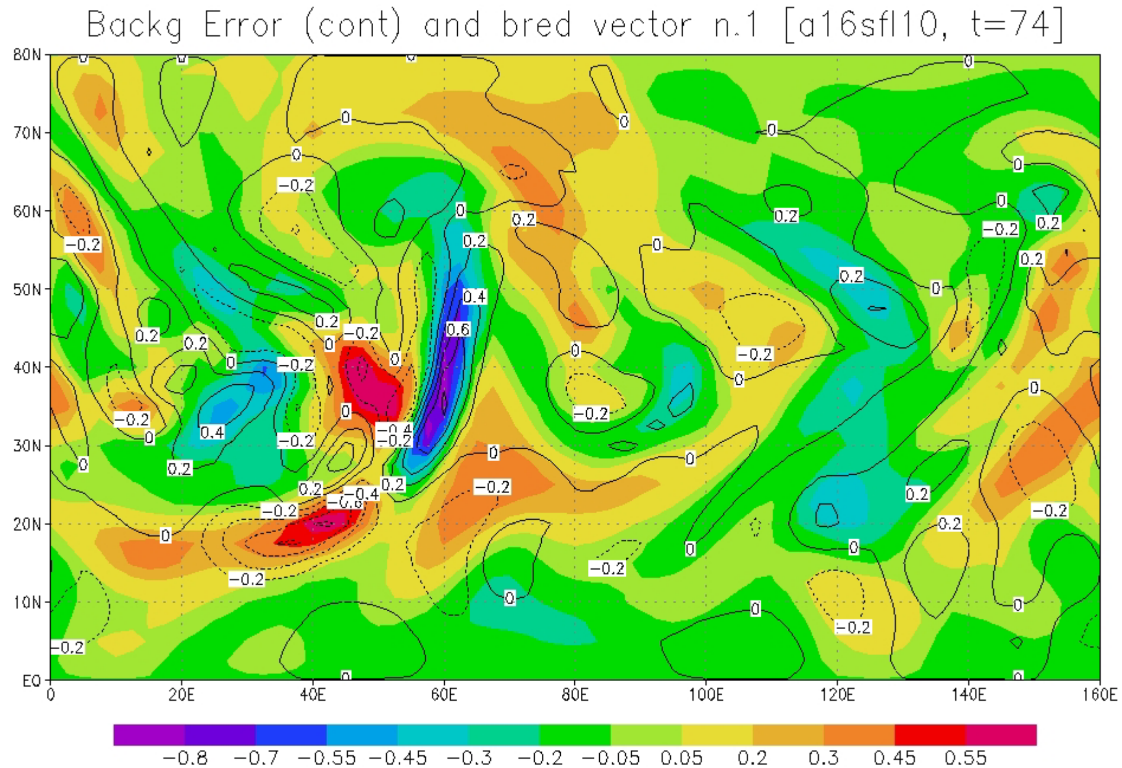
One approach to create initial perturbations for ensemble forecasting with errors of the day: breeding

- Breeding is simply running the nonlinear model a second time, from perturbed initial conditions



QG simulation of data assimilation (Corazza et al, 2003)

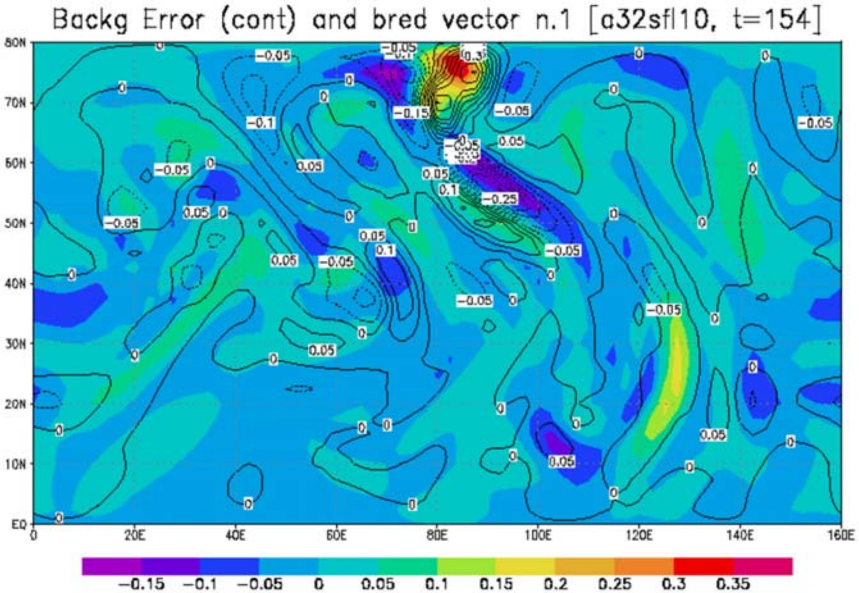
Bred vectors (color) have shapes similar to forecast error (contours).



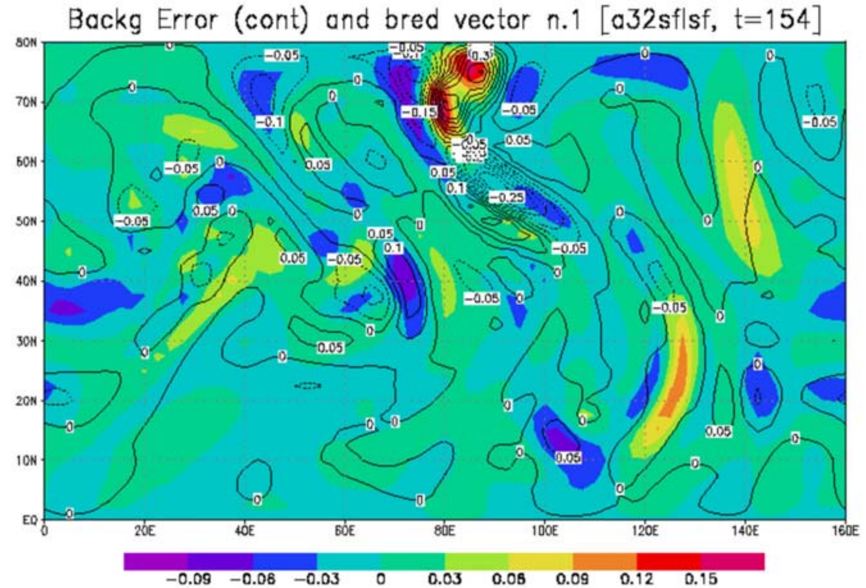
The bred vector clearly knows about the "errors of the day"

Bred vectors (like forecast errors) are independent of the norm

Bred vector normalized using **enstrophy** norm



Bred vector normalized using **streamfunction** norm



Data assimilation: combine a forecast with observations. We make a temperature forecast T_b and then take an observation T_o. A popular way to optimally estimate the truth (analysis) is to minimize the "3D-Var" cost function:

$$J(T) = \frac{1}{2}\left[\frac{(T - T_b)^2}{\sigma_b^2} + \frac{(T - T_o)^2}{\sigma_o^2}\right]$$

The analysis is given by $T_a = T_b + K(T_o - T_b)$

where $K = \sigma_b^2 / (\sigma_b^2 + \sigma_o^2)$

and the analysis error variance is smaller than the forecast or obs. $\sigma_a^2 = (1 - K)\sigma_b^2$

3D-Var used in operational forecasting centers

$$J = \min \frac{1}{2}[(\mathbf{x}_b - \mathbf{x}_a)^{\mathbf{T}} \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}_a) + (\mathbf{y}_o - H\mathbf{x}_a)^{\mathbf{T}} \mathbf{R}^{-1} (y_o - Hx_a)]$$

Distance to forecast　　　　　　　　Distance to observations

It's the same as the scalar formula for T, but now **x** is a model state vector, with $10^{6\text{-}8}$d.o.f., and $\mathbf{y}_{\mathrm{o}}$ is the set of observations, with $10^{5\text{-}9}$ d.o.f.

- **R** is the observational error covariance, **B** the forecast error covariance.
- In 3D-Var **B** is constant: it does not include "errors of the day"

As in the scalar case, the 3D-Var analysis is given by

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}_o - H\mathbf{x}_b)$$

where the weight matrix is

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

and the analysis error covariance is given by

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

- In 3D-Var **B** is *assumed* to be constant: it does not include "errors of the day"
- 4D-Var is very expensive and does not provide the analysis error covariance.
- In Kalman Filtering **B** is forecasted**.** It is like running the model N times, where N~$10^{6\text{-}8}$, so that it is impractical without simplifications

The solution: Ensemble Kalman Filtering

1) Perturbed observations and ensembles of data assimilation

- Evensen, 1994
- Houtekamer and Mitchell, 1998

2) Square root filter, no need for perturbed observations:

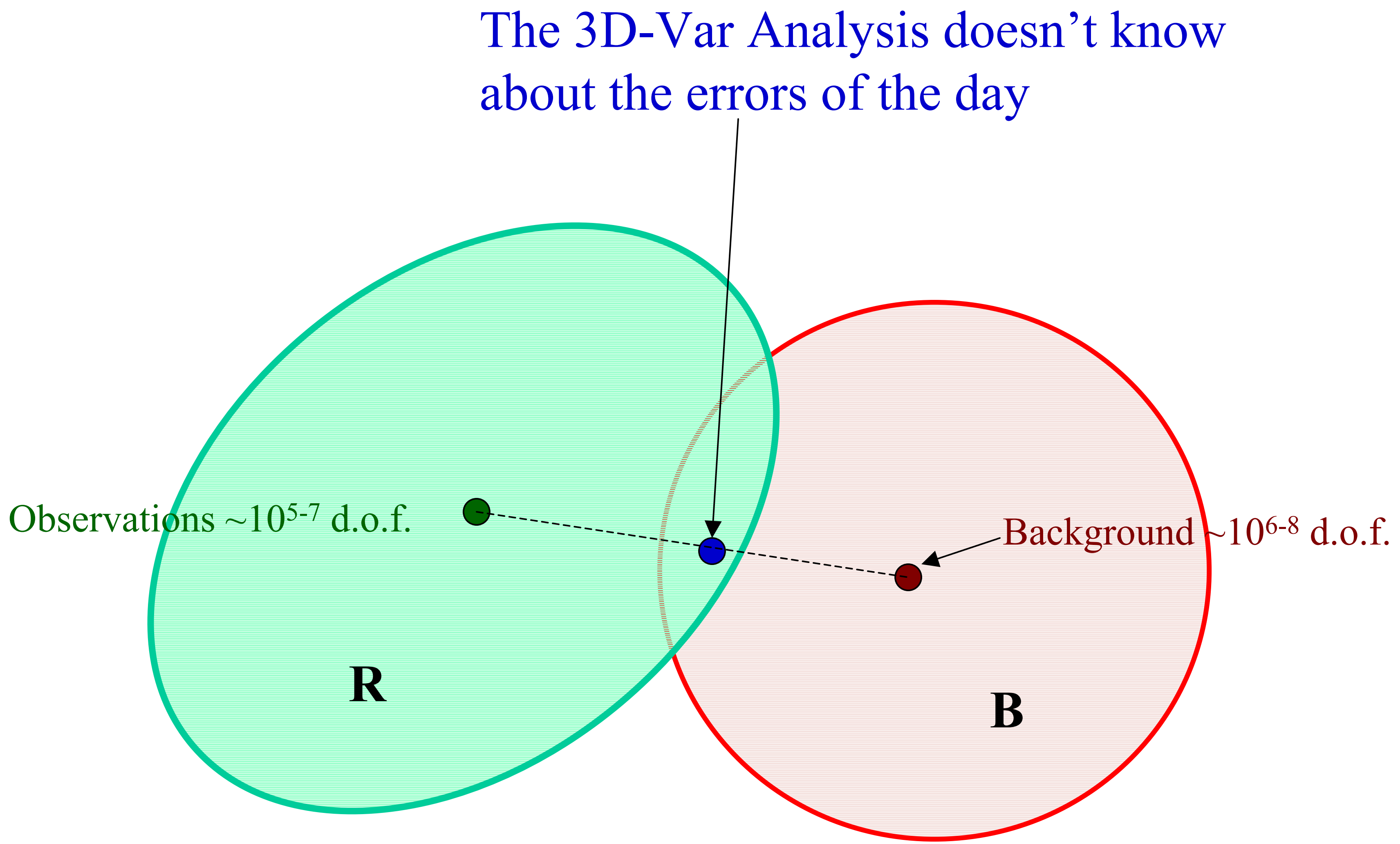
- Tippett, Anderson, Bishop, Hamill, Whitaker, 2003
- Anderson, 2001
- Whitaker and Hamill, 2002
- Bishop, Etherton and Majumdar, 2001

In these, the obs are assimilated one at a time

3) Local Ensemble Kalman Filtering, also a square root filter, but done in local patches

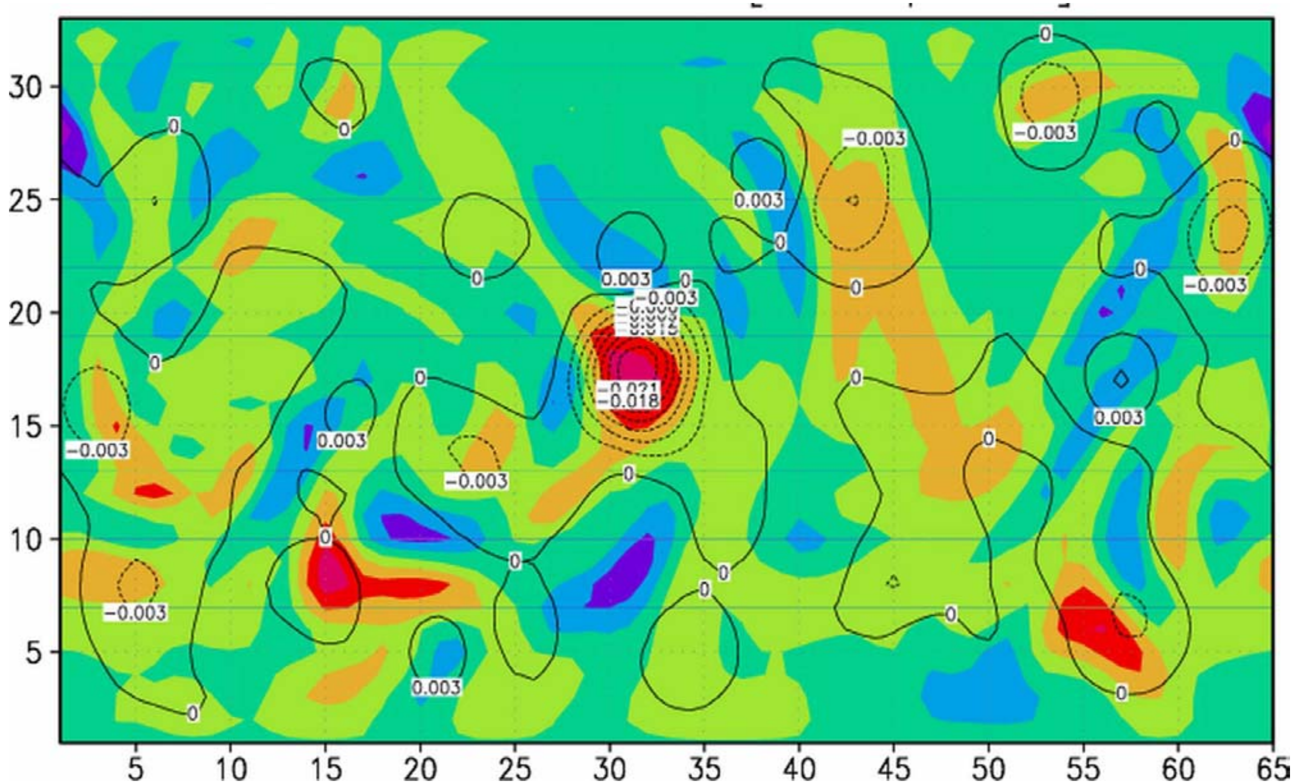
- Ott et al, 2003, MWR under review

Suppose we have a 6hr forecast (background) and new observations



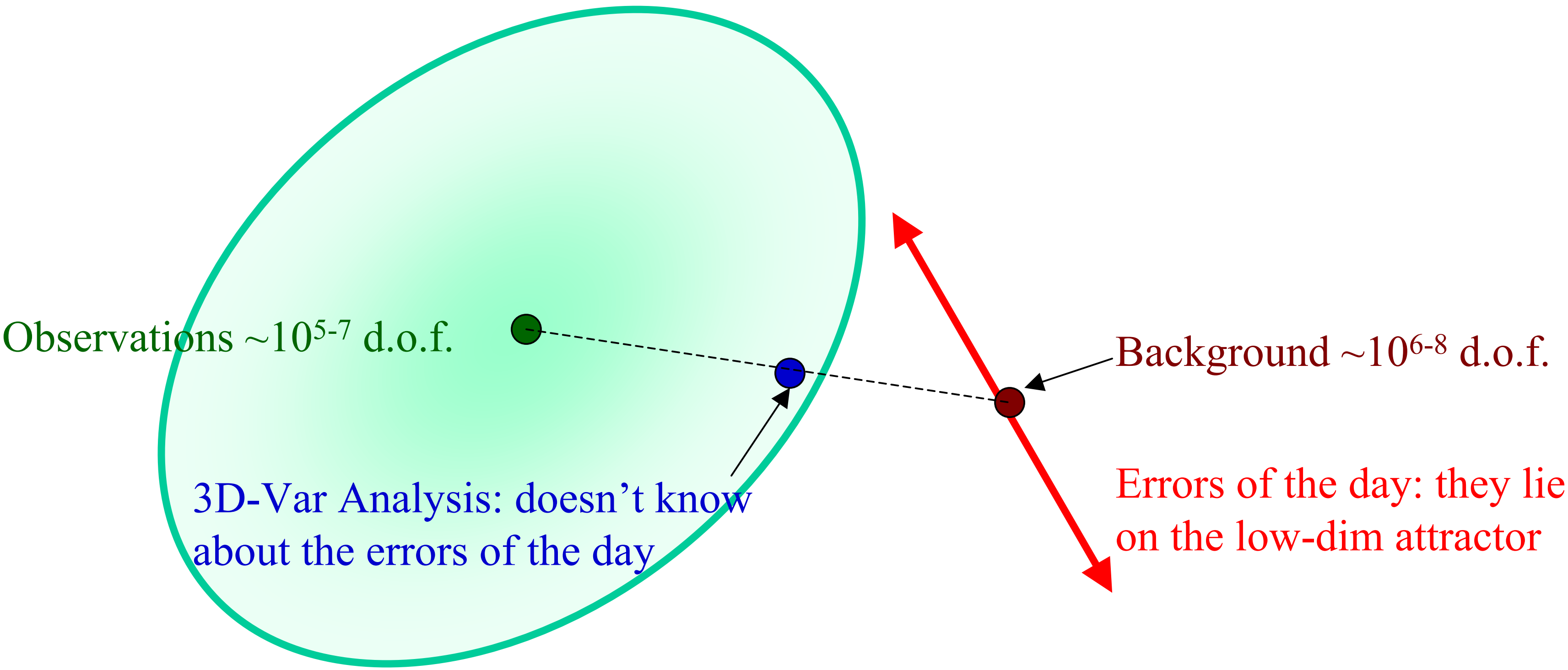
An example with the QG system (Corazza et al, 2003)

Background error (color) and 3D-Var analysis correction (contours)

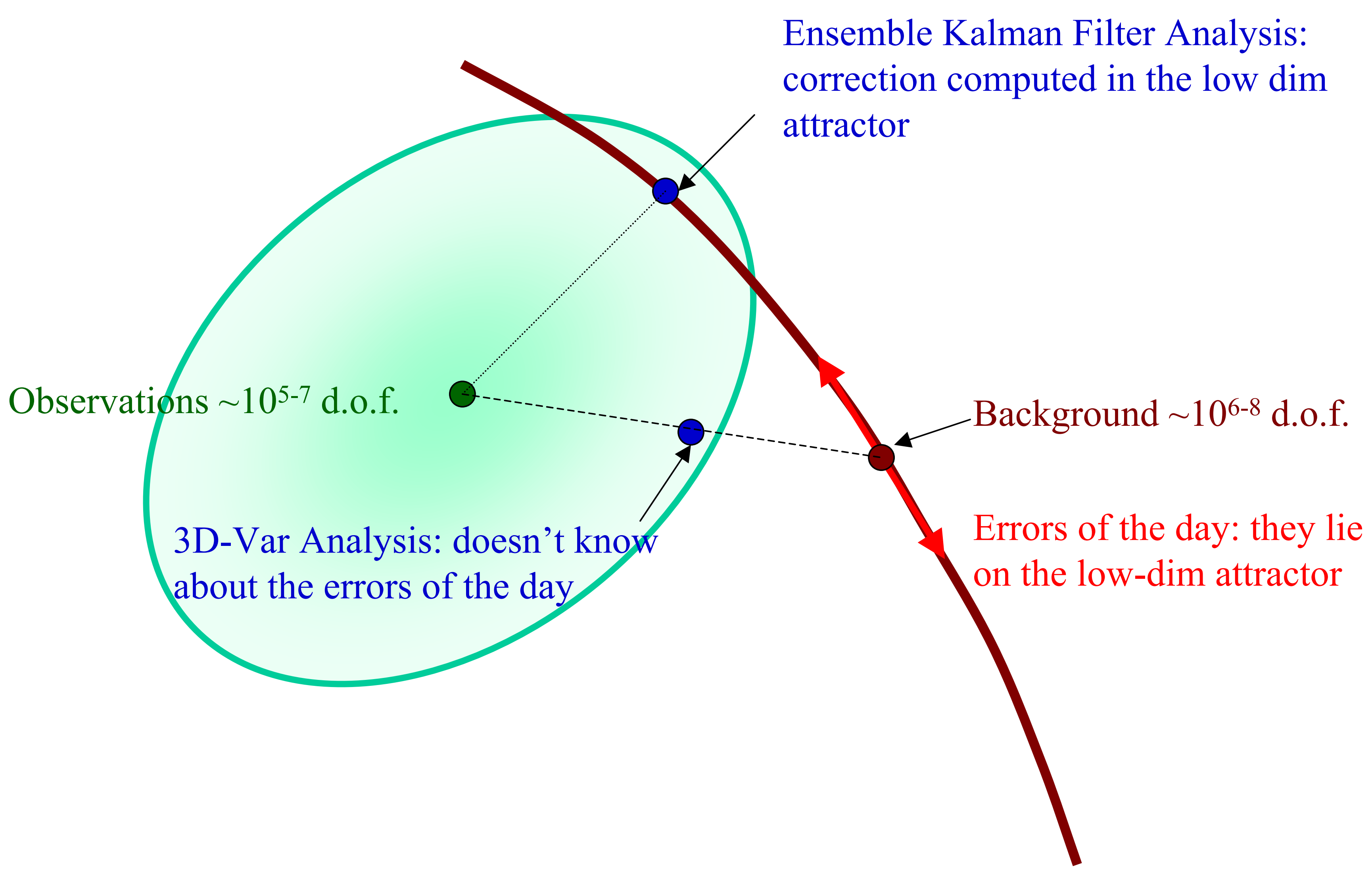


The analysis corrections due to the observations are isotropic because they don’t know about the errors of the day

With Ensemble Kalman Filtering we get perturbations pointing to the directions of the "errors of the day"

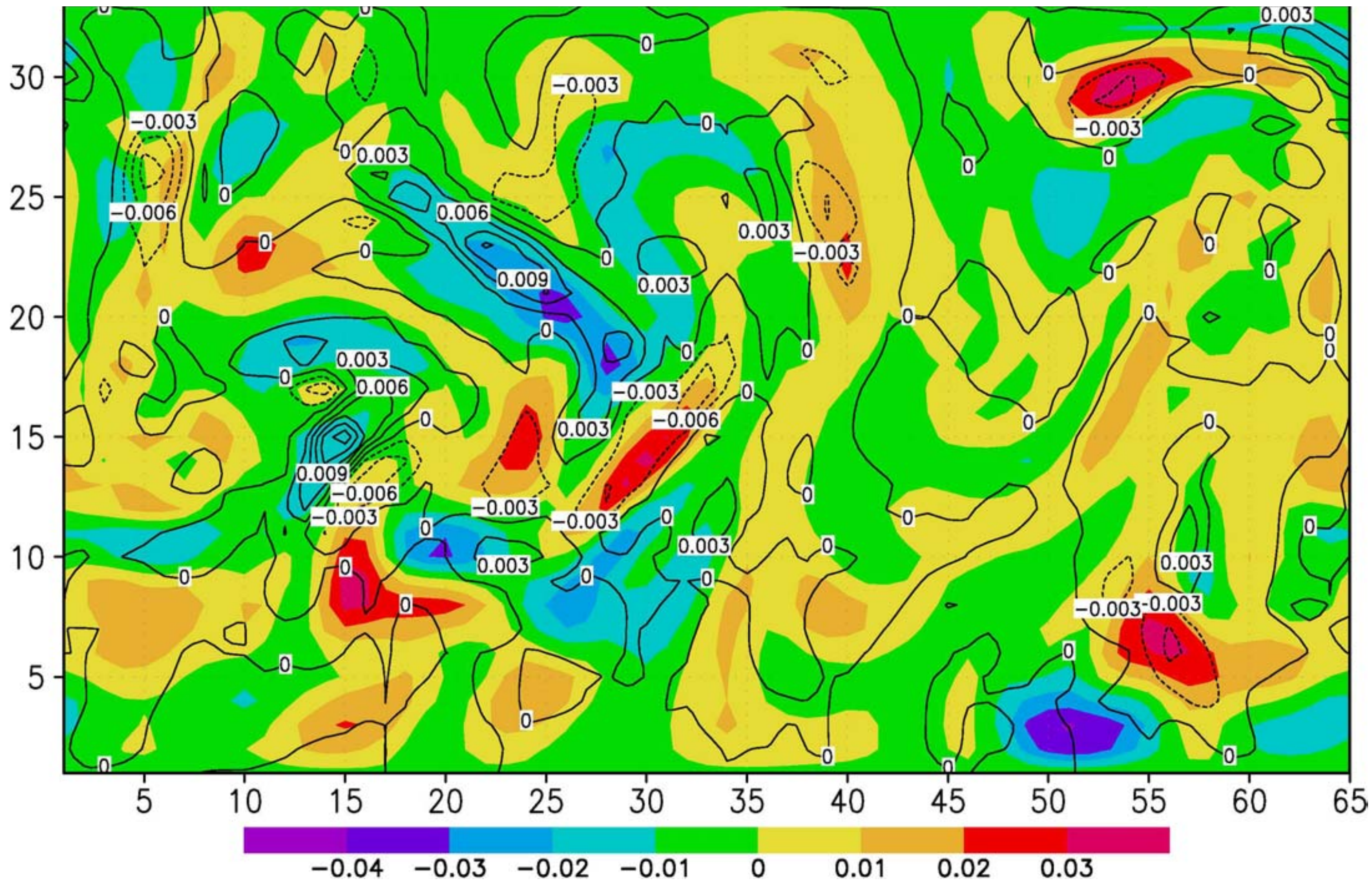


Ensemble Kalman Filtering is efficient because matrix operations are performed in the low-dimensional space of the ensemble perturbations



QG model example of Local Ensemble KF (Corazza et al)

Background error (color) and LEKF analysis correction

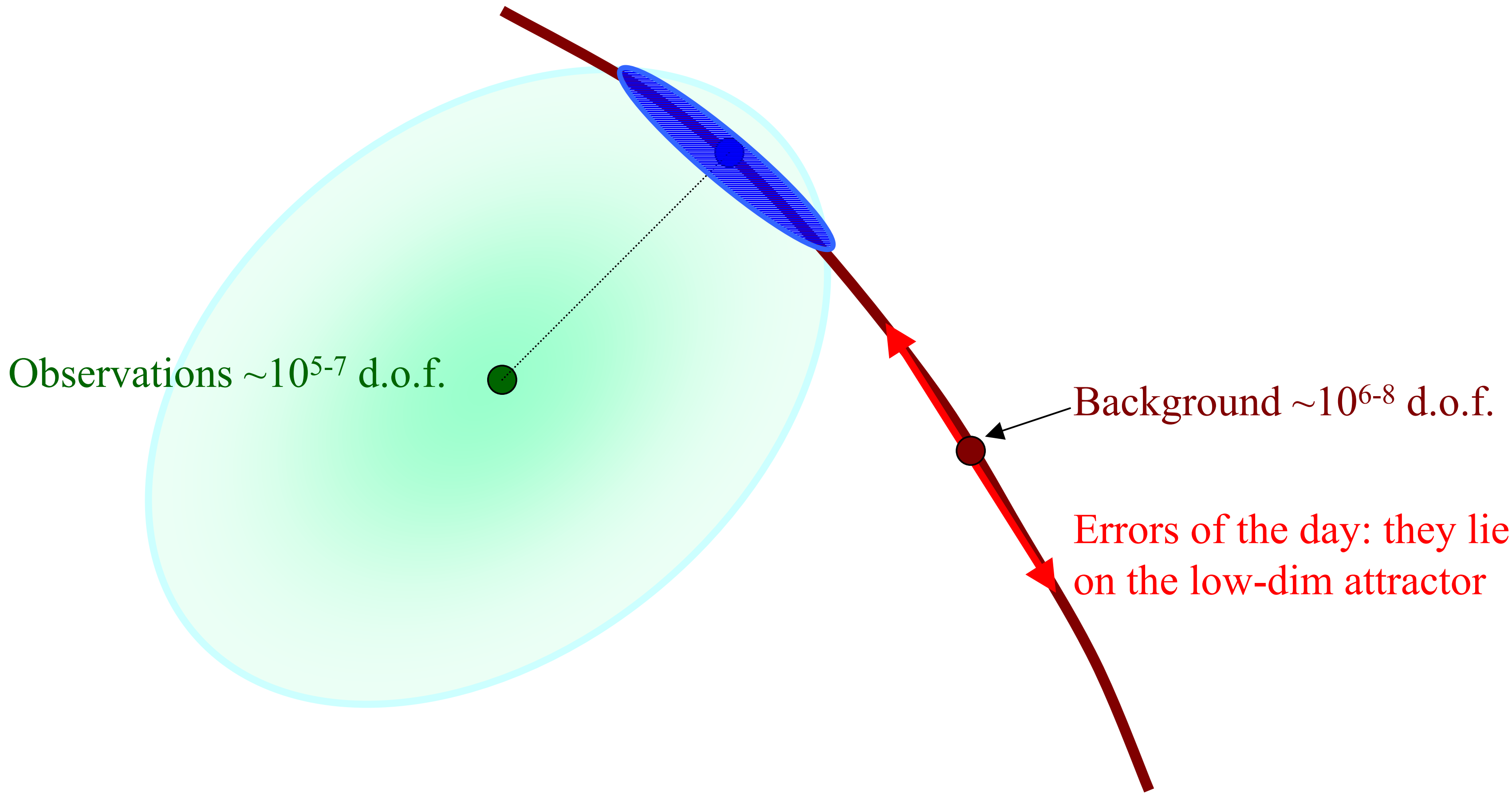


The LEKF does better because it captures the errors of the day

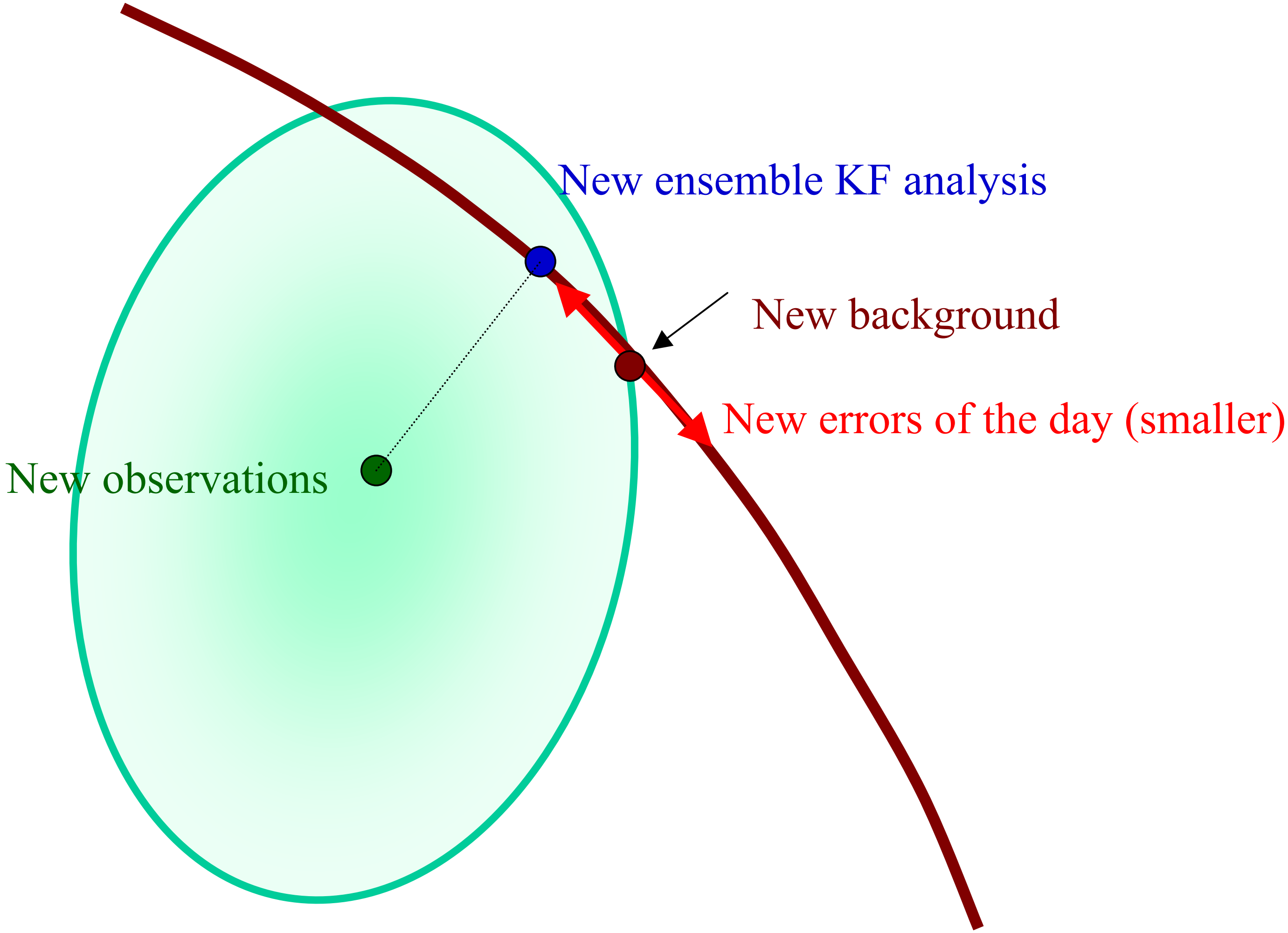
After the EnKF computes the analysis and the analysis error covariance **A**, the new ensemble initial perturbations $\delta \mathbf{a}_i$ are computed:

$$\sum_{i=1}^{k+1} \delta \mathbf{a}_i \delta \mathbf{a}_i^T = \mathbf{A}$$

These perturbations represent the analysis error covariance and are used as **initial perturbations** for the next ensemble forecast

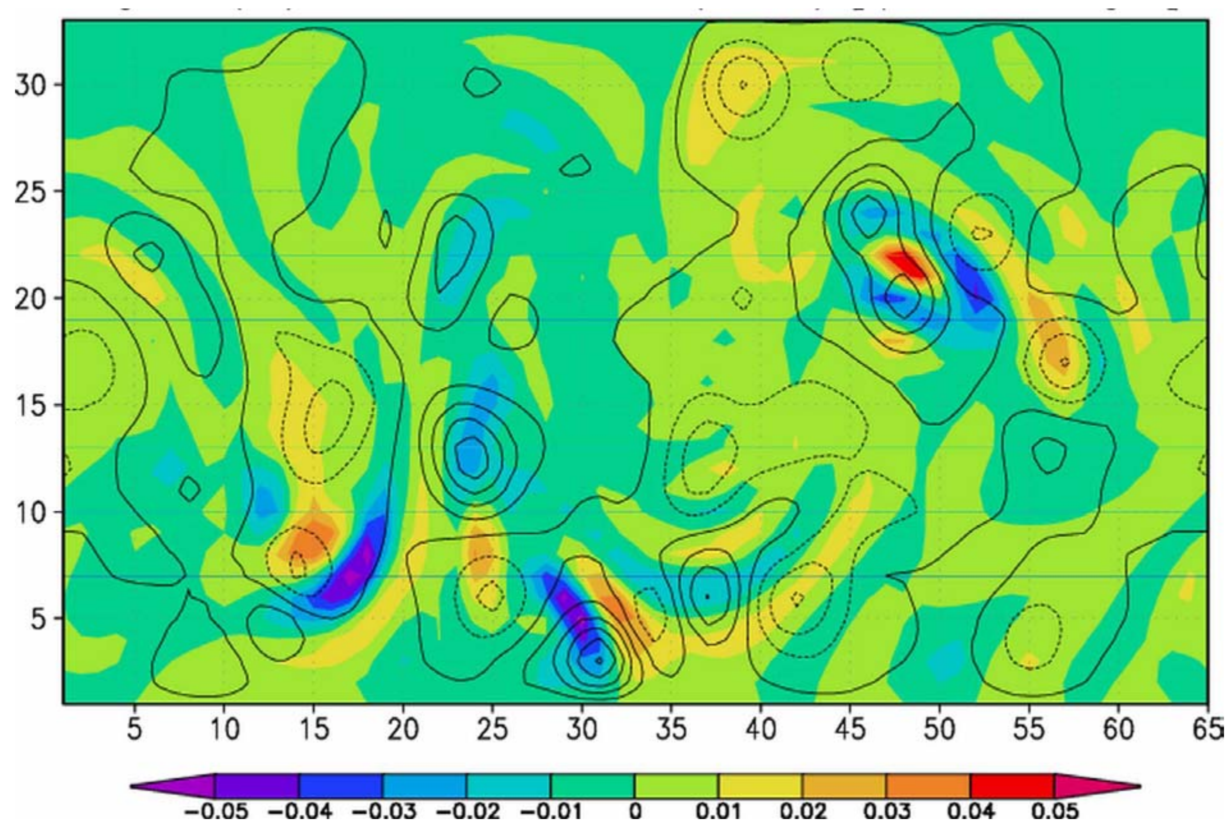


The process is repeated: an ensemble of forecasts is started from each of the initial perturbed analyses and integrated for 6 hours. The new background is the average of the forecasts, and the new low-dimensional attractor is given by the forecast perturbations.



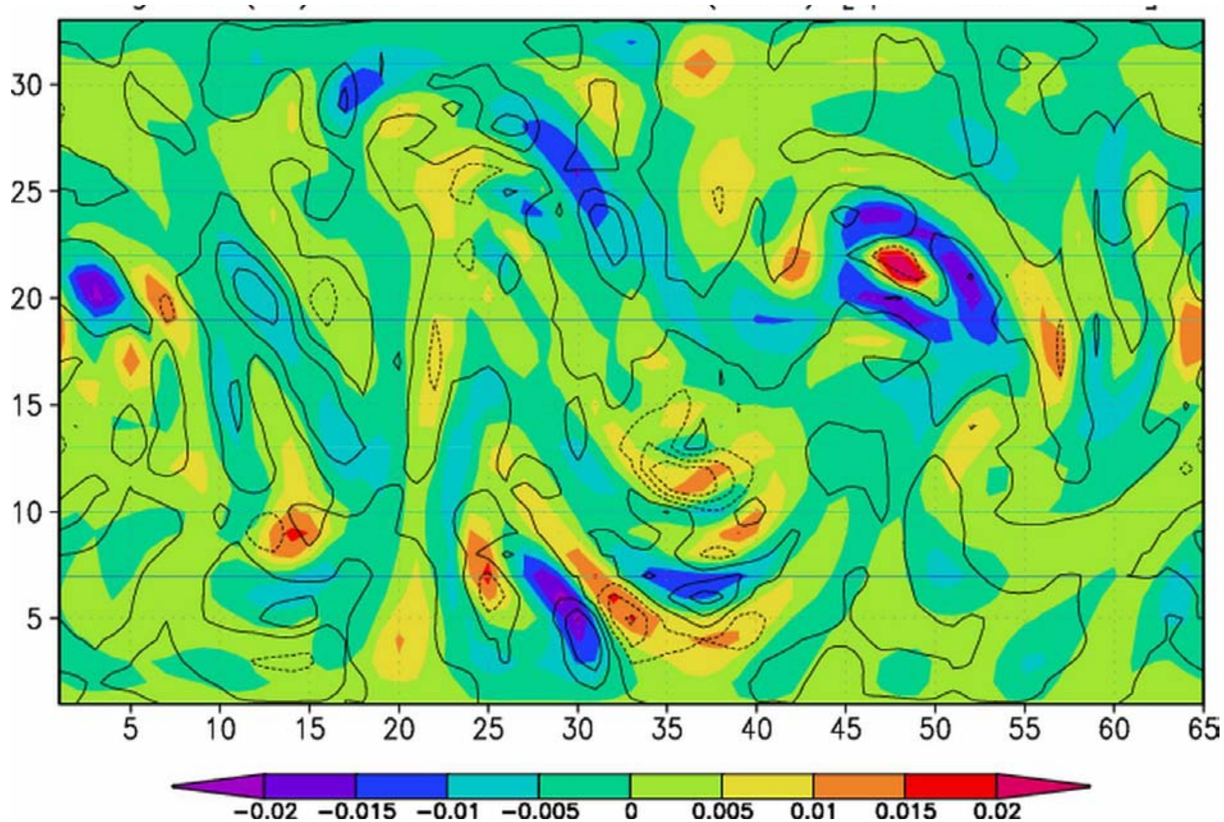
Again, from the QG simulation (Corazza et al, 2003)

Background error and 3D-Var analysis increment, June 15



The 3D-Var does not capture the errors of the day

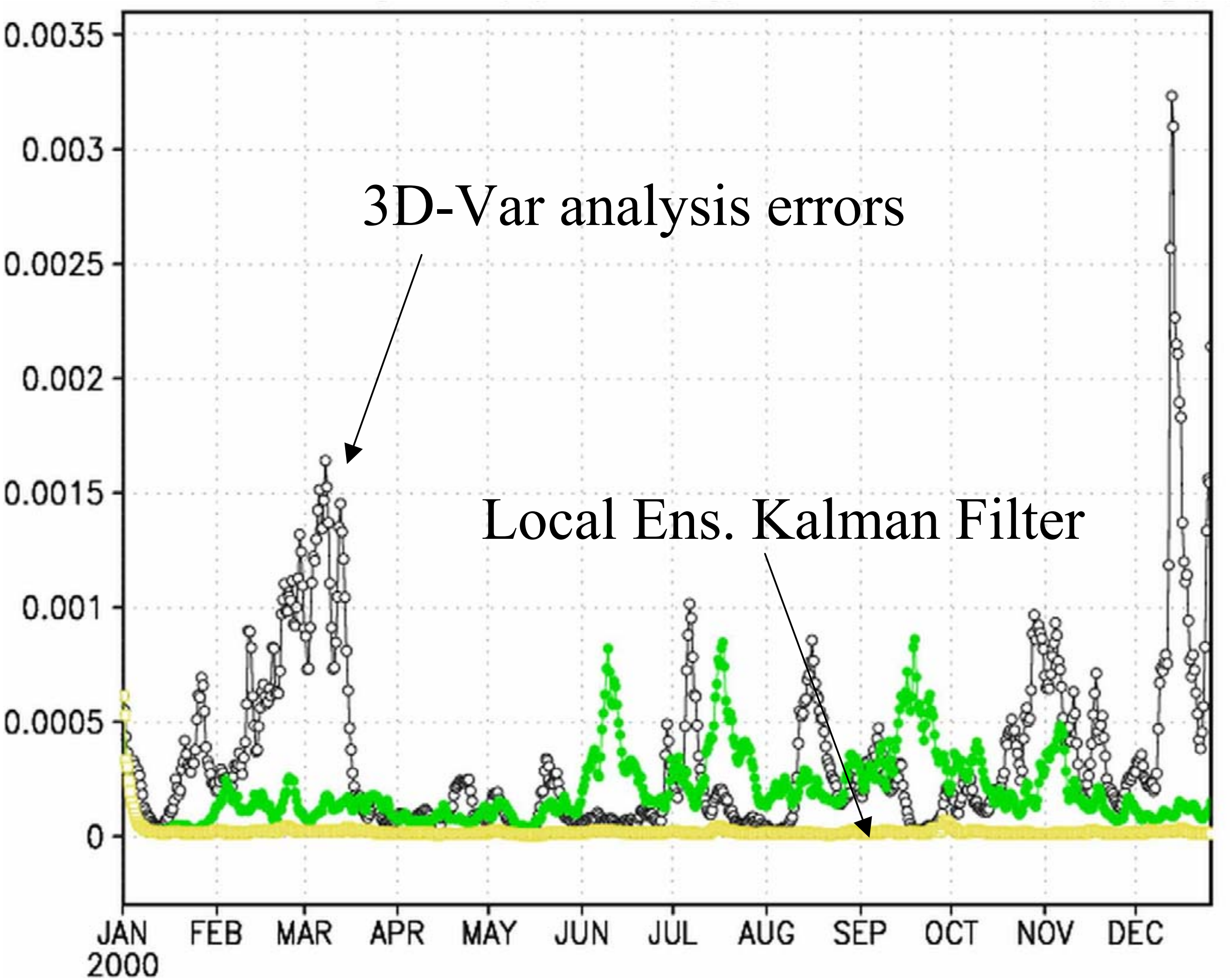
Background error (color) and LEKF analysis increments (contours), June 15



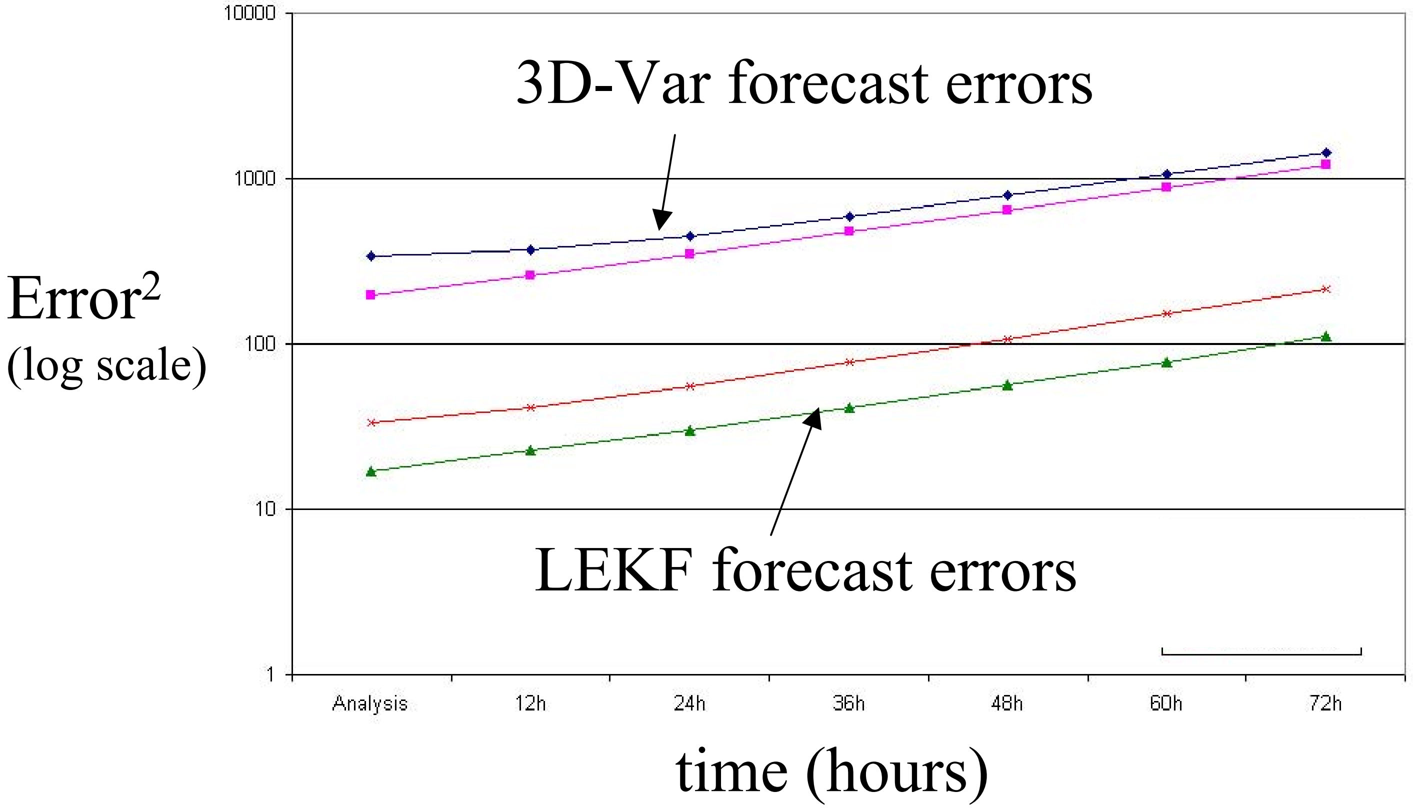
Contour interval: 0.005

The LEKF makes better use of the obs. because it includes the errors of the day

Area averaged Analysis Error: 3d-Var (black), LEKF (green), LEKF with covariance inflation (yellow)



This advantage continues into the 3-day forecasts



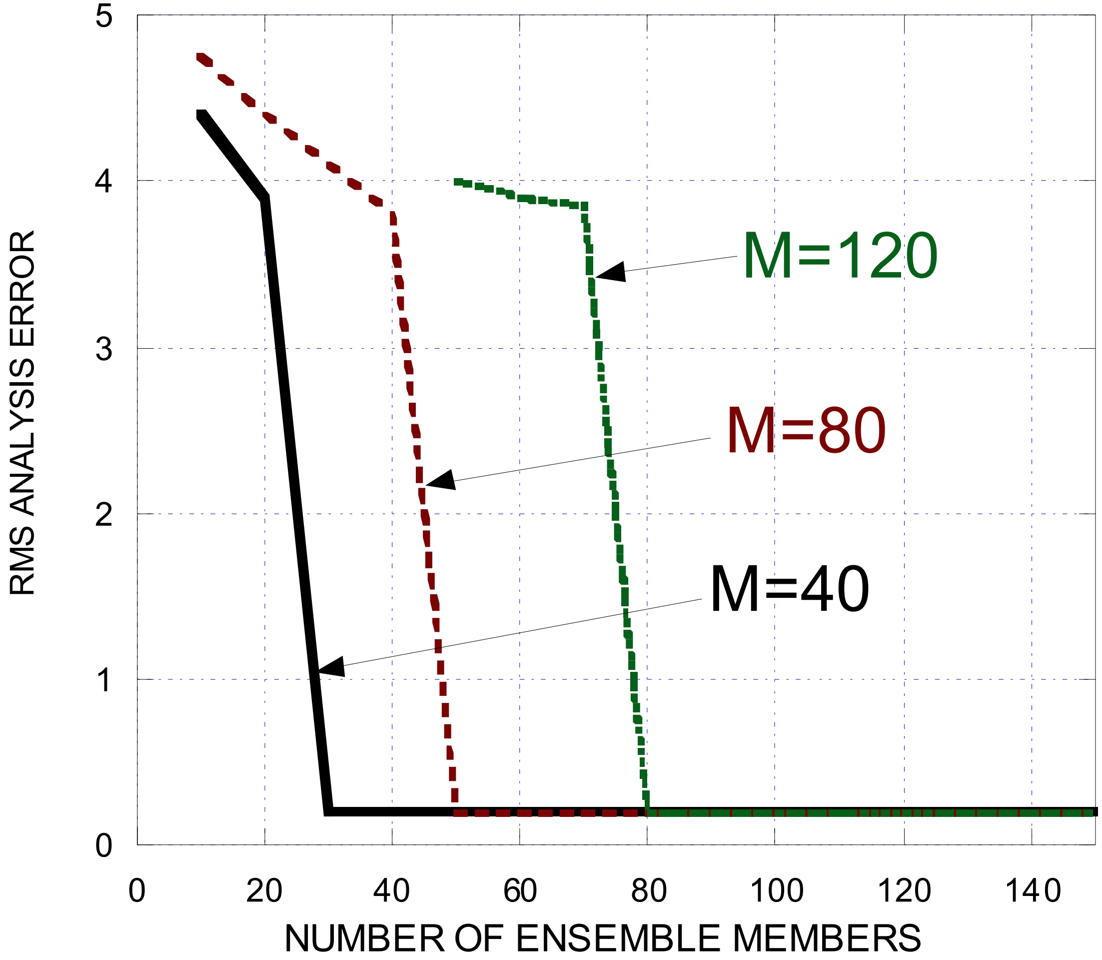
Why use a "local" ensemble approach?

- In the Local Ensemble Kalman Filter we compute the generalized "bred vectors" globally but use them locally (3D patches around each grid point of ~1000km x 1000km).
- These local columns provide the local shape of the "errors of the day".
- At the end of the local analysis we create a new global analysis and initial perturbations from the solutions obtained at each grid point.
- **This reduces the number of ensemble members needed.**
- **It also allows to compute the KF analysis independently at each grid point ("embarrassingly parallel").**

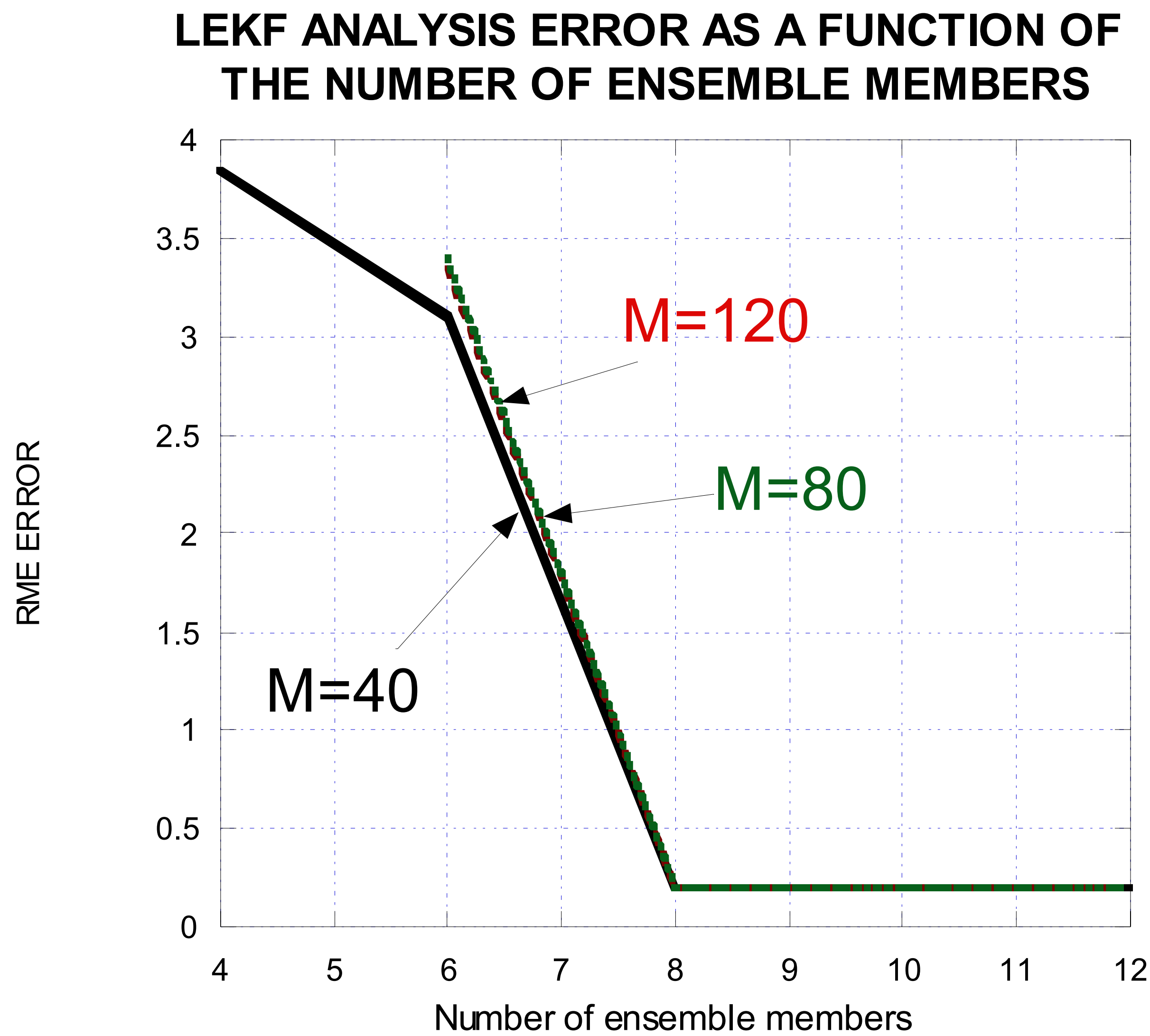
Results with Lorenz 40 variable model

- Used by Whitaker and Hamill (2002) to validate their ensemble square root filter (EnSRF)
- A very large global ensemble Kalman Filter converges to an “optimal” analysis rms error=0.20
- This “optimal” rms error is achieved by the LEKF for a range of small ensemble members
- We performed experiments for different size models: M=40 (original), M=80 and M=120, and compared a global KF with the LEKF

FULL ENSEMBLE KALMAN FILTER ANALYSIS ERROR AS A FUNCTION OF THE NUMBER OF ENSEMBLE MEMBERS



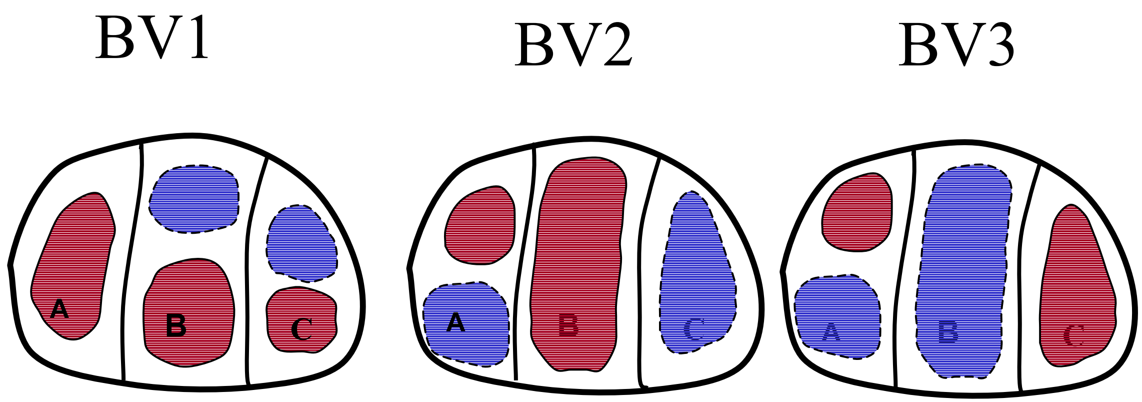
With the global EnKF approach, the number of ensemble members needed for convergence increases with the size of the domain M



With the local approach the number of ensemble members remains small

Why is the local analysis more efficient?

Schematic of a system with 3 independent regions of instability, A, B and C. Each region can have either wave #1 or #2 instability



From a local point of view, BV1 and BV2 are enough to represent all possible states.

From a global point of view, BV2 and BV3 are independent, and there are 63 possible different states…

Time mean error: optimal=0.20, eps=0.012

k = Rank of **B**

2l+1 =
Size of the box

	3	4	5	6	7	8	9
5	.24	.23					
7	.22	.22	.21	.22			
9	.22	.21	.21	.21	.21		
11	D	D	.20	.20	.20	.20	.20
13	D	D	.20	.20	.20	.20	.20
15	D	D	.22	.20	.20	.20	.20

Time mean error: optimal=0.20, 2l+1=13

K = Rank of **B**

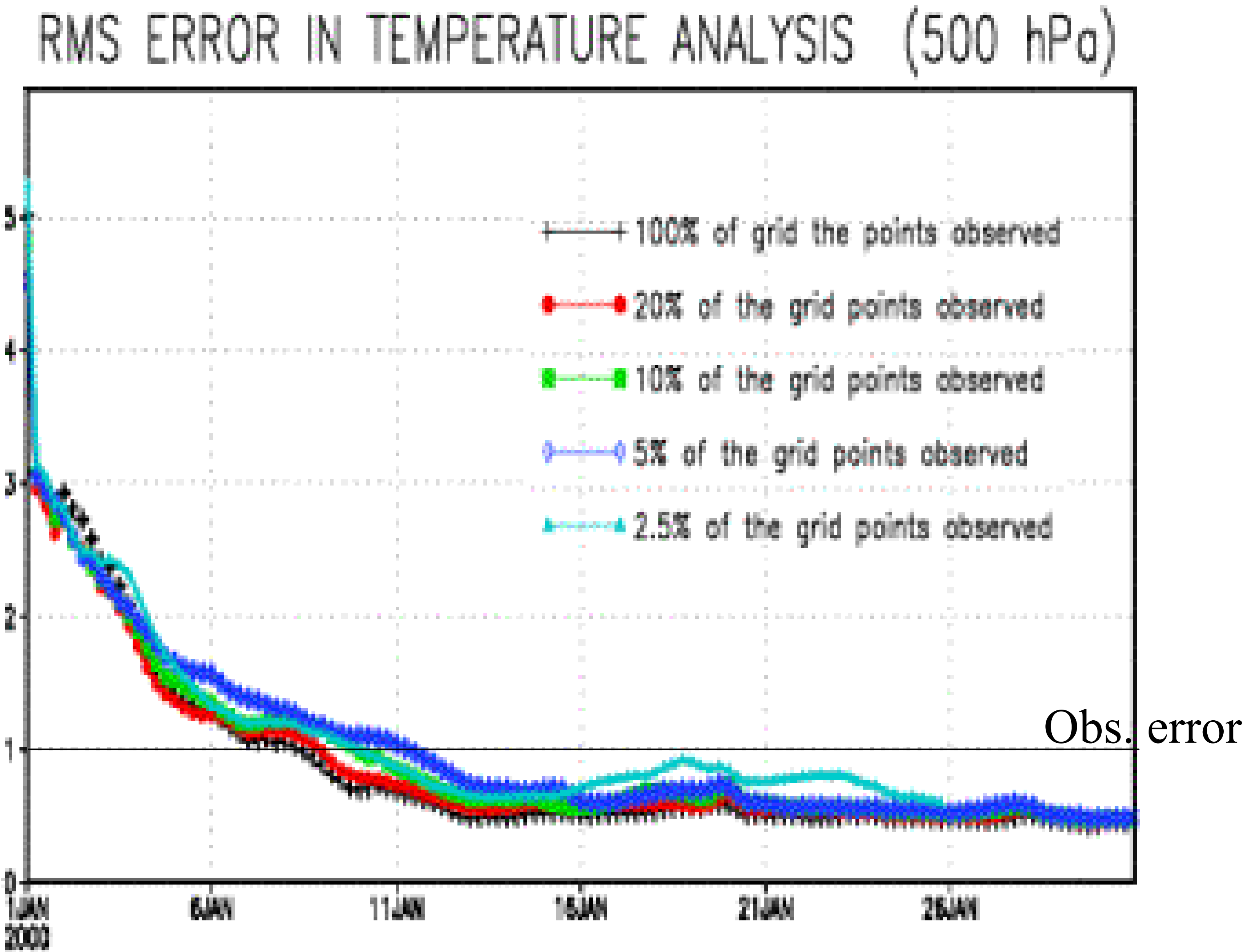
eps = enhanced inflation

	3	4	5	6	7	8	9
.008	D	D	.44	.20	.20	.20	.20
.010	D	D	.20	.20	.20	.20	.20
.012	D	D	.20	.20	.20	.20	.20
.014	D	D	.20	.20	.20	.20	.20
.016	D	D	.20	.20	.20	.20	.20
.018	D	D	.20	.20	.20	.20	.20

From Szunyogh, Kostelich et al

Preliminary LEKF results with NCEP's global model

- T62, 28 levels (1.5 d.o.f.)
- The method is model independent: the same code was used for the L40 model as for the NCEP global spectral model
- Simulation with observations at every grid point (1.5 million obs)
- Very parallel! Each grid point analysis done independently
- Very fast! 20 minutes in a single 1GHz Intel processor with 10 ensemble members



RMS ERROR IN TEMPERATURE ANALYSIS (500 hPa)
100% of grid the points observed
20% of the grid points observed
10% of the grid points observed
5% of the grid points observed
2.5% of the grid points observed
Obs. error
0
1
2
3
4
5
1JAN
2000
6JAN
11JAN
16JAN
21JAN
26JAN

From Szunyogh, Kostelich et al

Preliminary results with NCEP's global model

A) observations at every grid point

- With 40 members and no tuning, the rms error was half of the observations rms error

B) observations were thinned until only 2.5% of the grid points had observations

- The solution of LEKF converged to the same level of errors!!

Advantages of Ensemble KF

- It knows about the "errors of the day" through **B.**
- Matrix computations are done in a low-dimensional space.
- In LEKF computations for each grid point are independent from the neighbors (very parallel).
- It can handle many observations
- Both accurate and efficient: can be done frequently (e.g., once every hour)
- EnKF generates perfect initial perturbations for ensemble forecasting (bred vectors are now both scaled and rotated to represent the analysis error covariance).

In summary

- New ensemble Kalman Filtering methods have become feasible
- They provide optimal analysis and initial ensemble perturbations

However,…

- **The most important remaining problem is how to handle model deficiencies**
- EnKF may also be the most efficient way to tune models and reduce errors…