Advanced data assimilation methods-EKF and EnKF

Alghero, Lecture 4

Hong Li and Eugenia Kalnay

University of Maryland

Optimal Interpolation for a scalar

Information available in OI

$T_b = T_t + \varepsilon_b$ background or forecast $\overline{\varepsilon_b} = 0; \quad \overline{\varepsilon_b \varepsilon_b} = \sigma_b^2$

$T_o = T_t + \varepsilon_o$ New observations $\overline{\varepsilon_o} = 0; \quad \overline{\varepsilon_o \varepsilon_o} = \sigma_o^2$

analysis=forecast+optimal weight * observational increment:

$$T_a = T_b + w(T_0 - T_b)$$

$$w = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$ Optimal weight

$$\sigma_a^2 = (\sigma_b^{-2} + \sigma_o^{-2})^{-1} = (1 - w)\sigma_b^2$$ Analysis error

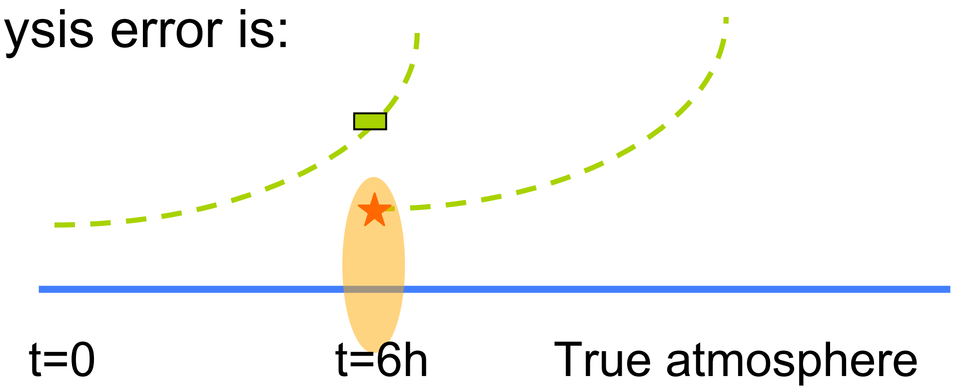
Recall the basic formulation in OI

- OI for a Scalar:

$$T_a = T_b + w(T_0 - T_b)$$

Optimal weight to minimize the analysis error is:

$$w = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$



- OI for a Vector:

$$\mathbf{x}_i^f = M\mathbf{x}_{i-1}^a$$

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{W}(\mathbf{y}_i^o - H\mathbf{x}_i^b)$$

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}$$

$$\mathbf{P}^a = [\mathbf{I} - \mathbf{W}\mathbf{H}]\mathbf{B}$$

$$\mathbf{B} = \overline{\delta\mathbf{x}_b \bullet \delta\mathbf{x}_b^T} \qquad \delta\mathbf{x}_b = \mathbf{x}_b - \mathbf{x}_\mathbf{t}$$

- B is statistically pre-estimated, and constant with time in its practical implementations. Is this a good approximation?

OI and Kalman Filter for a scalar

- OI for a scalar:

$$T_b(t_{i+1}) = M(T_a(t_i); \quad \text{assume} \quad \sigma_b^2 = (1+a)\sigma_a^2 = \frac{1}{1-w}\sigma_a^2 = const.$$

$$w = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}; \quad \sigma_a^2 = (1-w)\sigma_b^2$$

$$T_a = T_b + w(T_0 - T_b)$$

- Kalman Filter for a scalar:

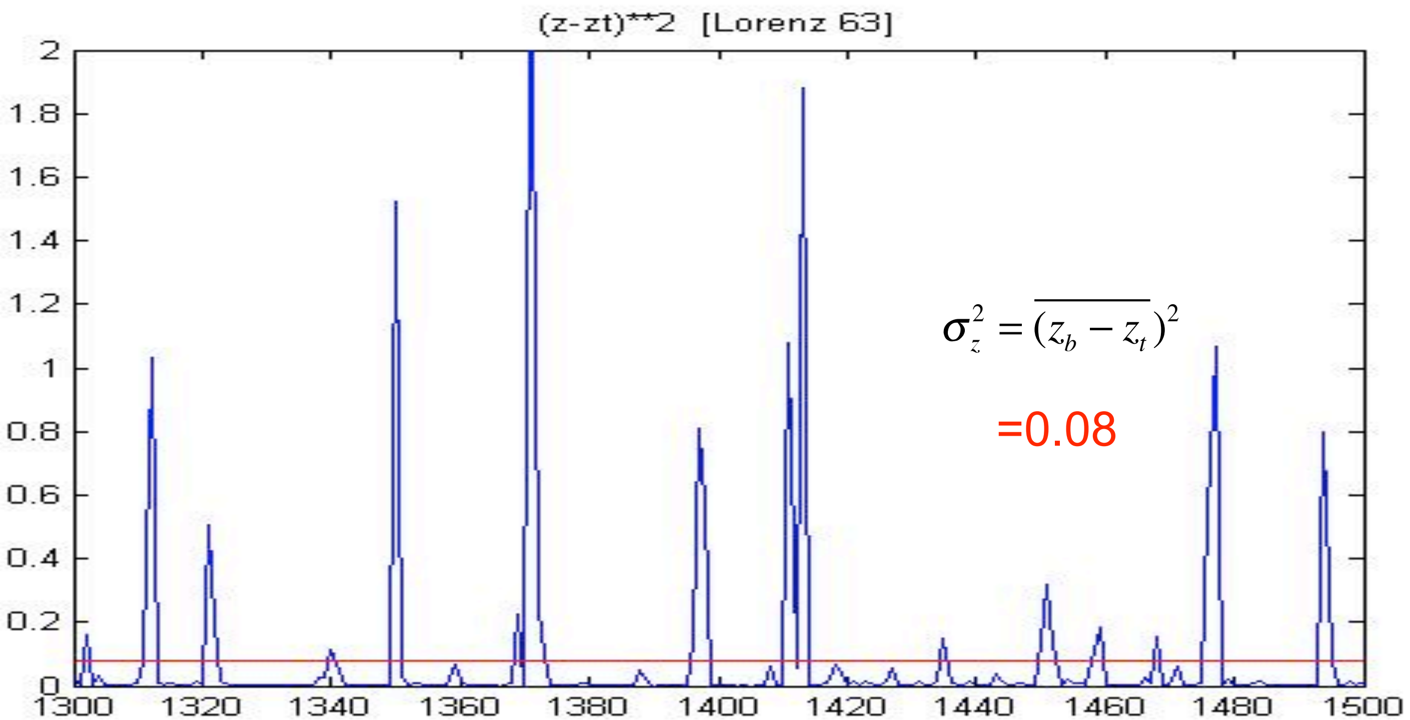
$$T_b(t_{i+1}) = M(T_a(t_i)); \quad \sigma_b^2(t) = (L\sigma_a)(L\sigma_a)^T; \quad L = dM / dT$$

$$w = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}; \quad \sigma_a^2 = (1-w)\sigma_b^2$$

$$T_a = T_b + w(T_0 - T_b)$$

- Now the background error variance is forecasted using the linear tangent model L and its adjoint L^T

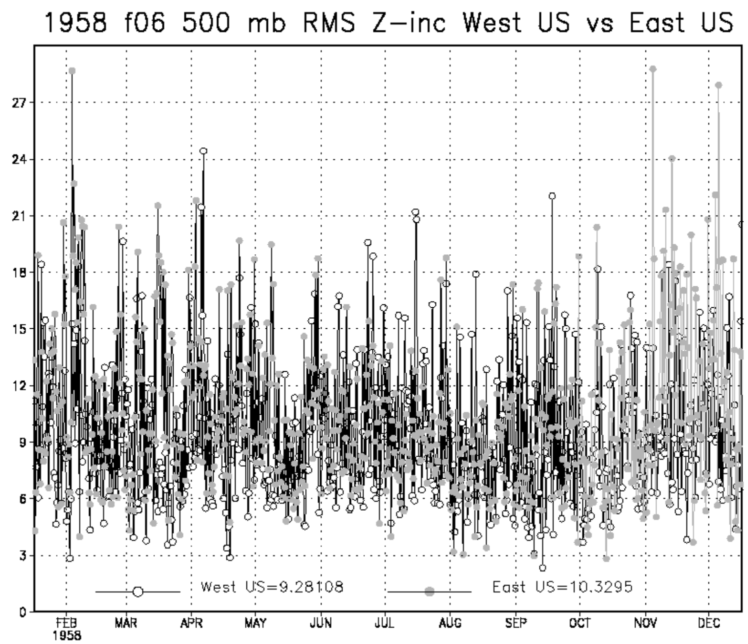
“Errors of the day” computed with the Lorenz 3-variable model: compare with rms (constant) error

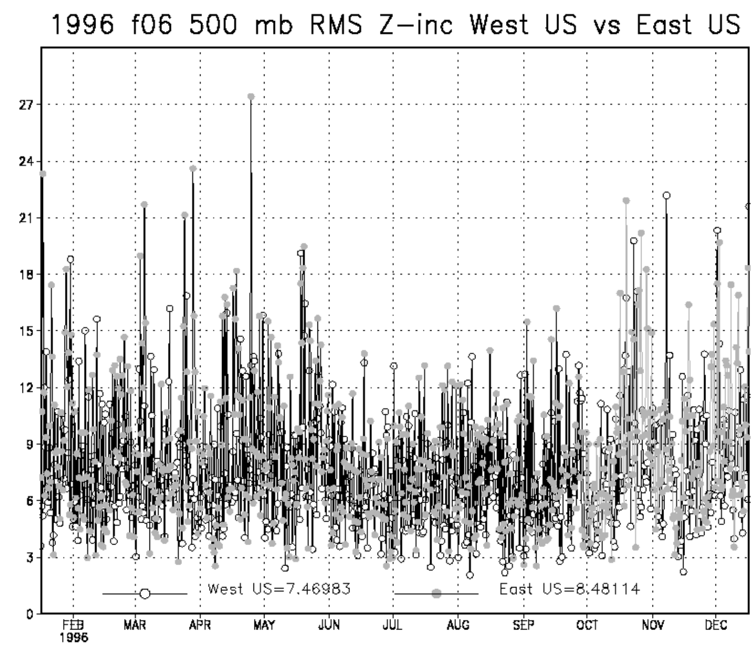


➢Not only the amplitude, but also the structure of B is constant in 3D-Var/OI

➢This is important because analysis increments occur only within the subspace spanned by B

“Errors of the day” obtained in the NCEP reanalysis (figs 5.6.1 and 5.6.2 in the book)





- Note that the mean error went down from 10m to 8m because of improved observing system, but the “errors of the day” (on a synoptic time scale) are still large.
- In 3D-Var/OI not only the amplitude, but also the structure of B is fixed with time

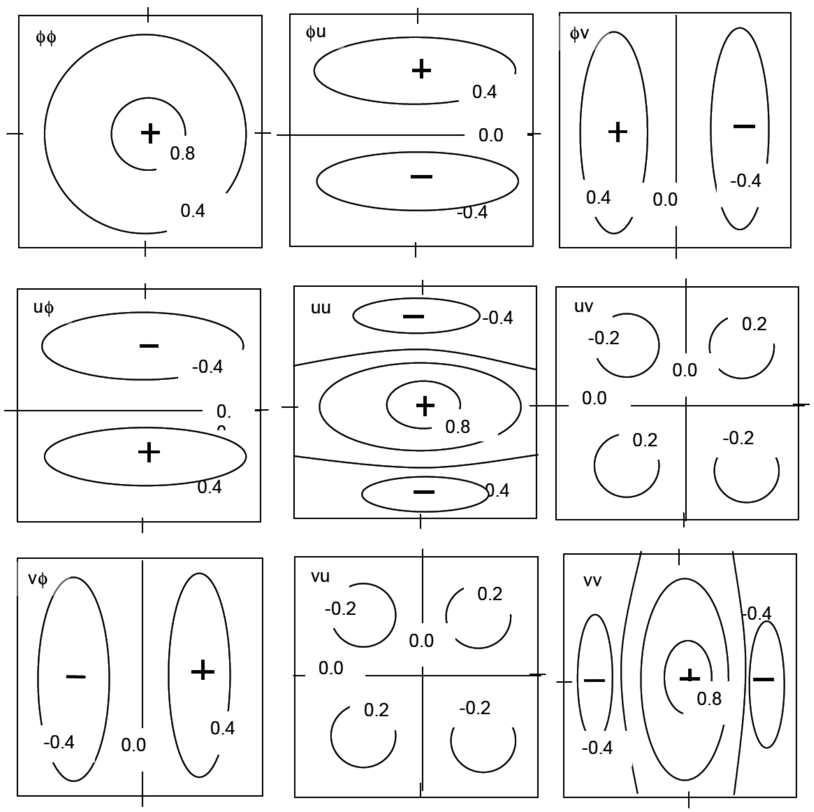
Flow independent error covariance

➢If we observe only Washington, D.C, we can get estimate for Richmond and Philadelphia corrections through the error correlation (*off-diagonal term* in B).

$$B = \begin{bmatrix} \sigma_1^2 & \mathrm{cov}_{1,2} & \cdots & \mathrm{cov}_{1,n} \\ \mathrm{cov}_{1,2} & \sigma_2^2 & & \mathrm{cov}_{2,n} \\ \cdots & \cdots & & \cdots \\ \mathrm{cov}_{1,n} & \mathrm{cov}_{2,n} & \cdots & \sigma_n^2 \end{bmatrix}$$

➢In OI(or 3D-Var), the scalar error correlation between two points in the same horizontal surface is assumed homogeneous and isotropic. (p162 in the book)

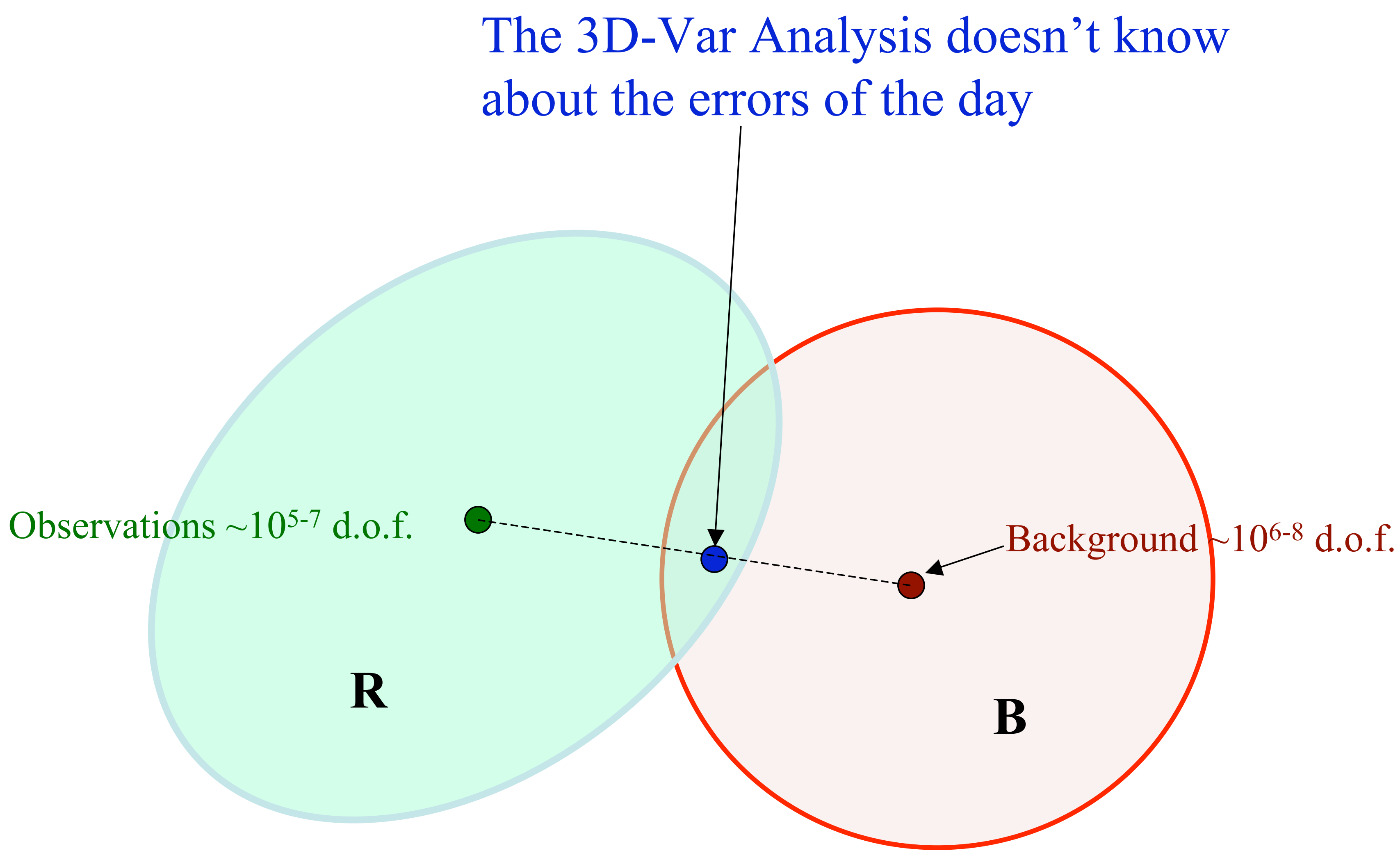
Typical 3D-Var error covariance



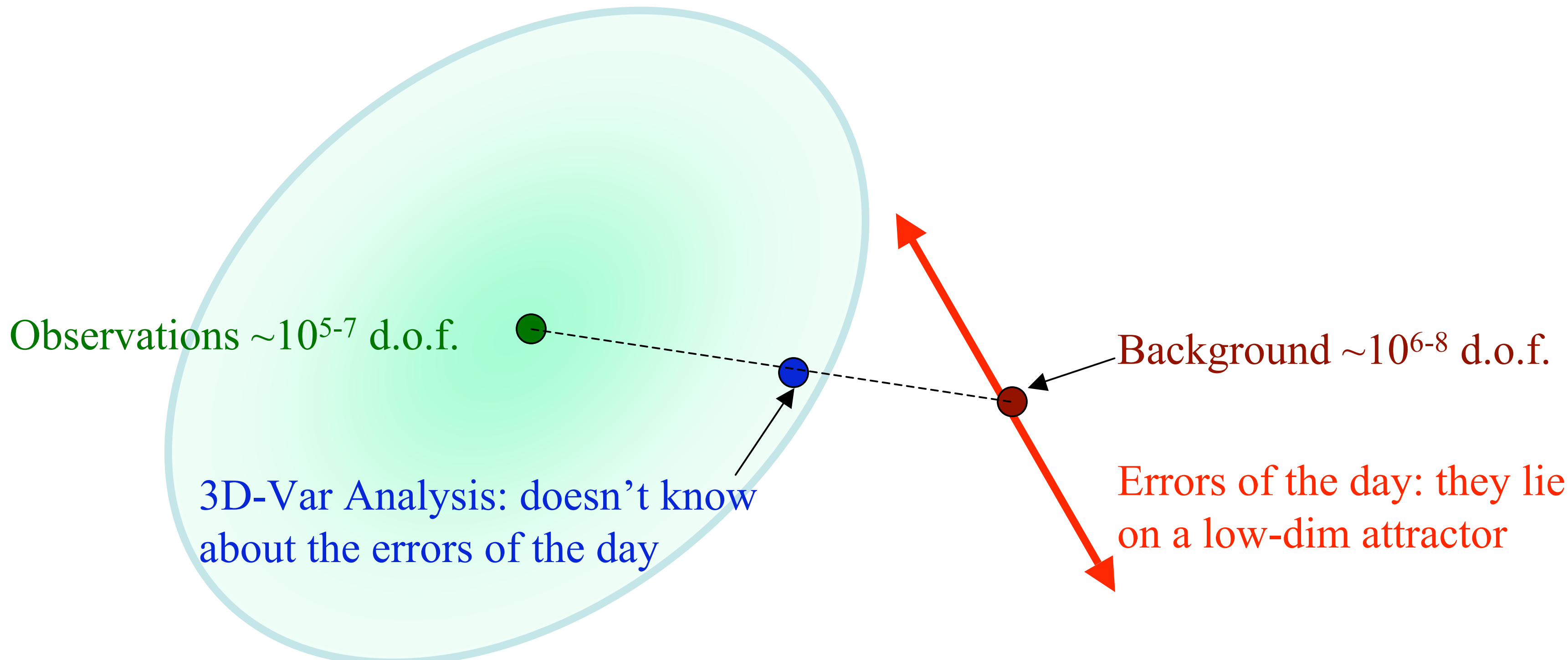
$$B = \begin{bmatrix} \sigma_1^2 & \mathrm{cov}_{1,2} & \cdots & \mathrm{cov}_{1,n} \\ \mathrm{cov}_{1,2} & \sigma_2^2 & & \mathrm{cov}_{2,n} \\ \cdots & \cdots & & \cdots \\ \mathrm{cov}_{1,n} & \mathrm{cov}_{2,n} & \cdots & \sigma_n^2 \end{bmatrix}$$

➢In OI(or 3D-Var), the error correlation between two mass points in the same horizontal surface is assumed homogeneous and isotropic.(p162 in the book)

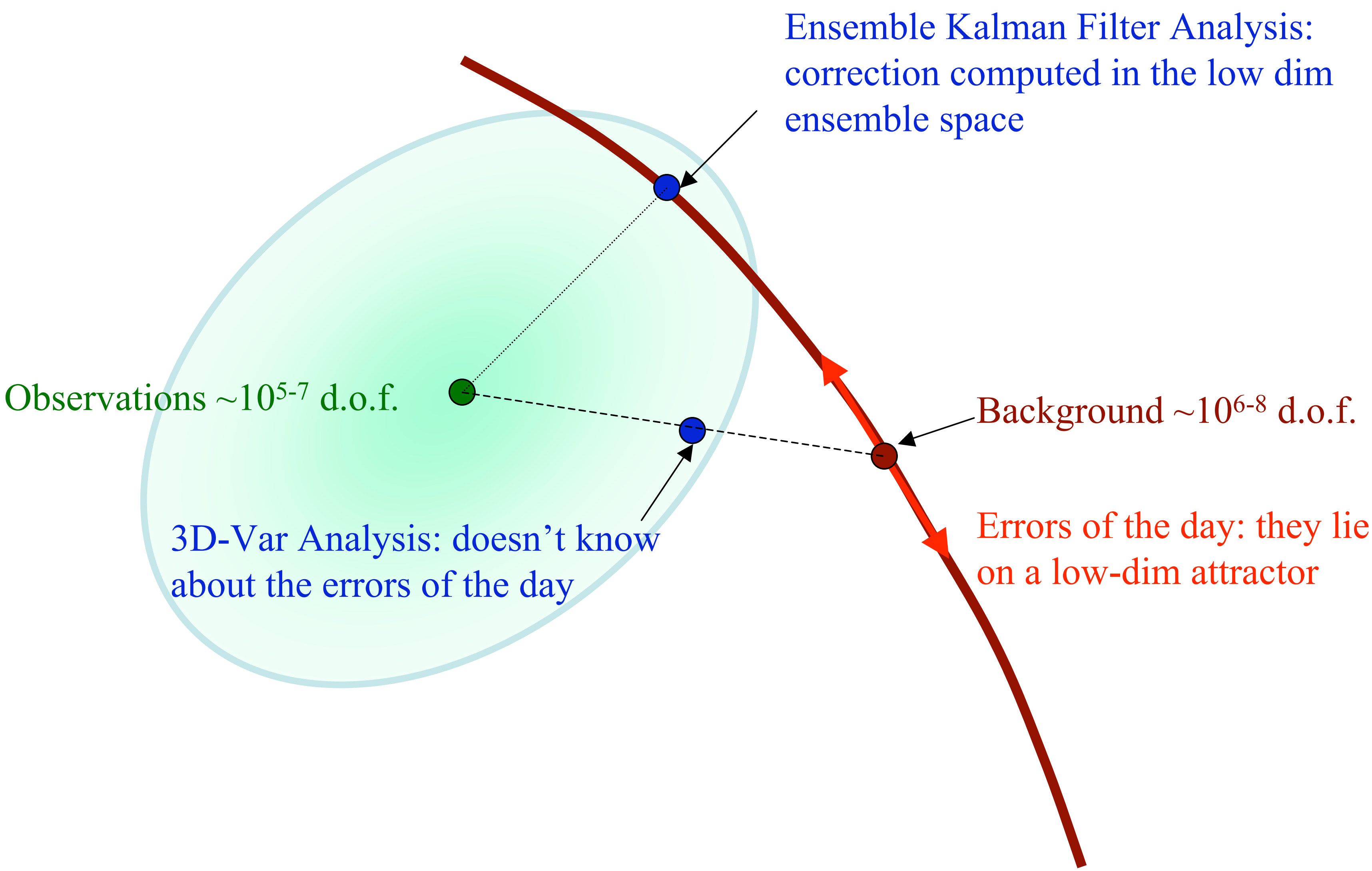
Suppose we have a 6hr forecast (background) and new observations



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



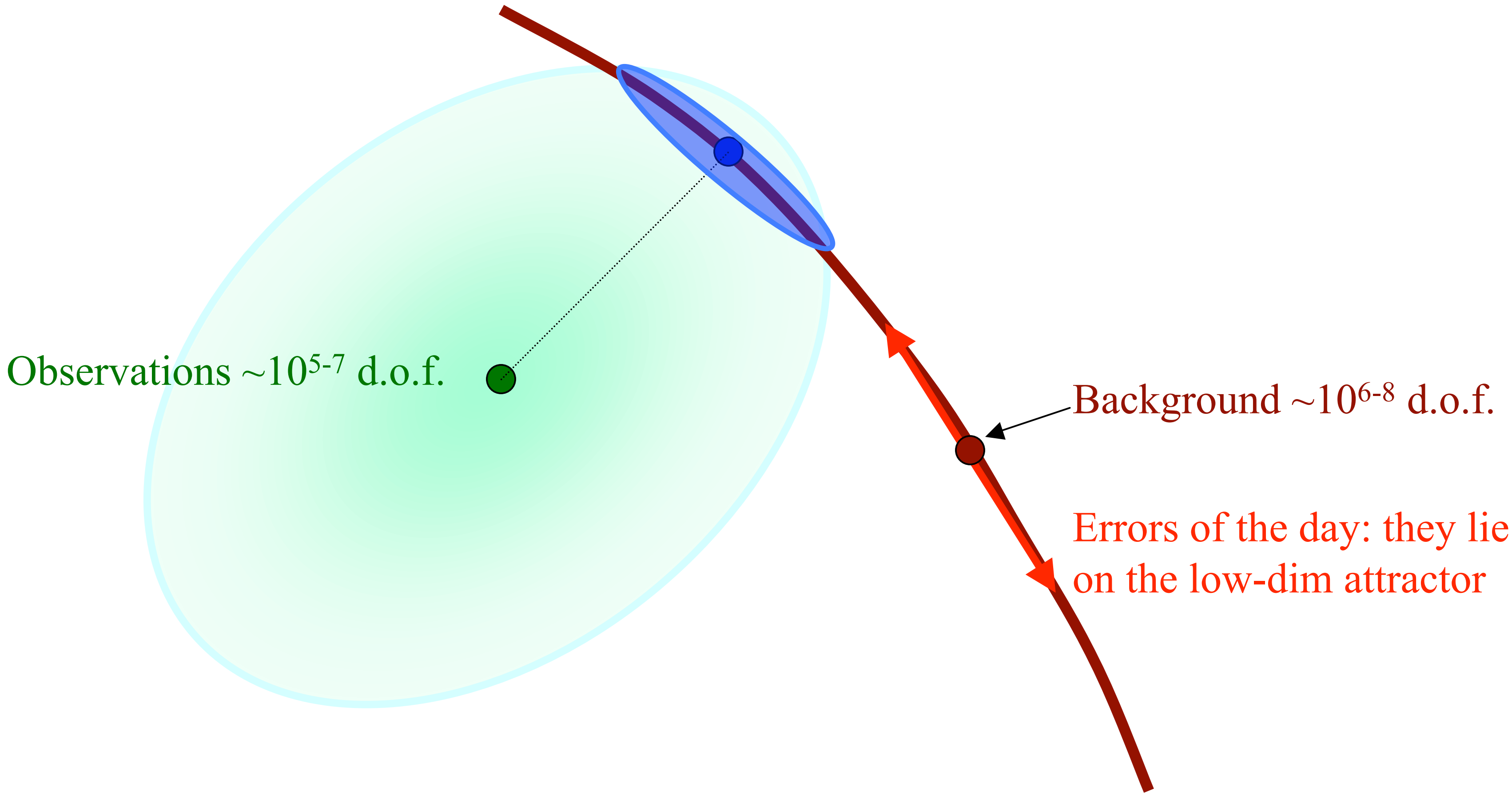
Ensemble Kalman Filtering is efficient because matrix operations are performed in the low-dimensional space of the ensemble perturbations



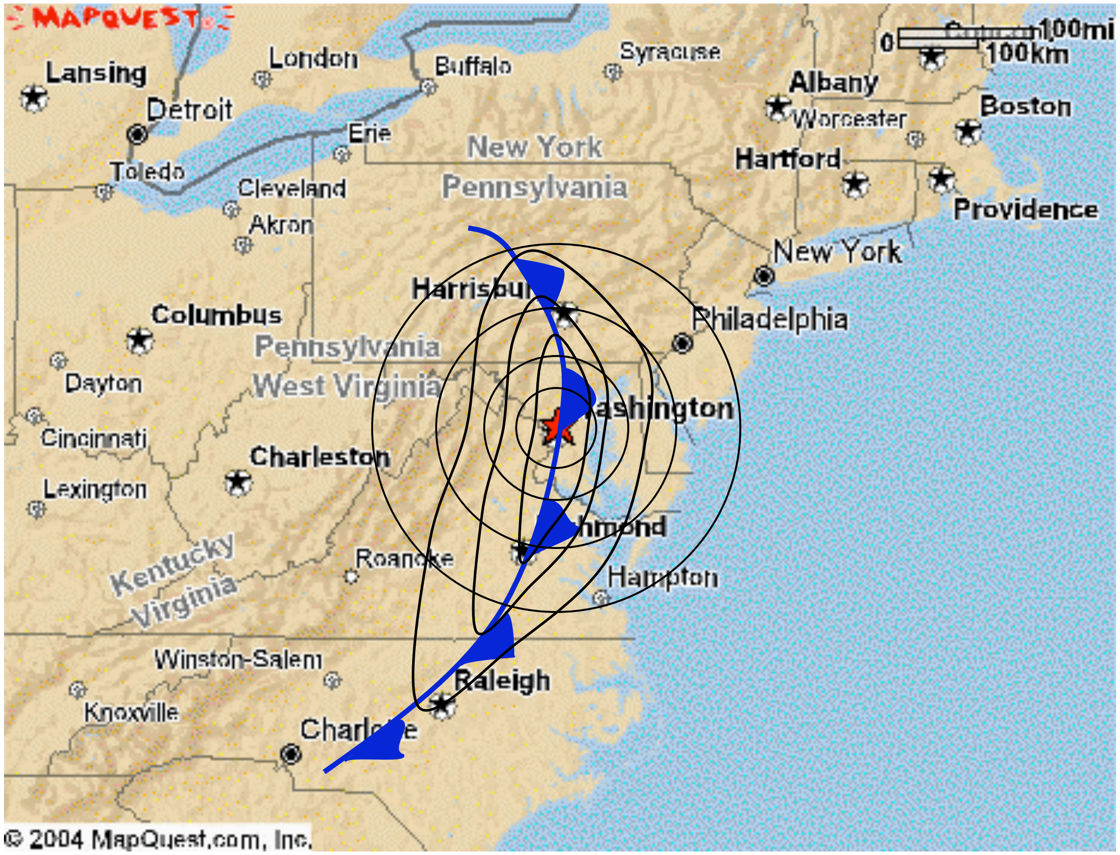
After the EnKF computes the analysis and the analysis error covariance $\mathbf{A}$, the new ensemble initial perturbations $\delta \mathbf{a}_i$ are computed:

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

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



Flow-dependence – a simple example (Miyoshi,2004)



There is a cold front in our area…

What happens in this case?

This is not appropriate

This does reflects the flow-dependence.

Extended Kalman Filter (EKF)

- Forecast step

$$\mathbf{x}_i^b = M\mathbf{x}_{i-1}^a$$

$$\mathbf{P}_i^b = \mathbf{L}_{i-1}\mathbf{P}_{i-1}^a\mathbf{L}_{i-1}^T + \mathbf{Q} \quad * \qquad \varepsilon_i^b = \mathbf{L}_{i-1}\varepsilon_{i-1}^a + \varepsilon_{i-1}^m$$

- Analysis step

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}_i(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f)$$

$$\mathbf{K}_i = \mathbf{P}_i^b\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R}]^{-1} \qquad * \; *$$

$$\mathbf{P}_{i\ n\times n}^a = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^b = [(\mathbf{P}_i^b)^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}$$

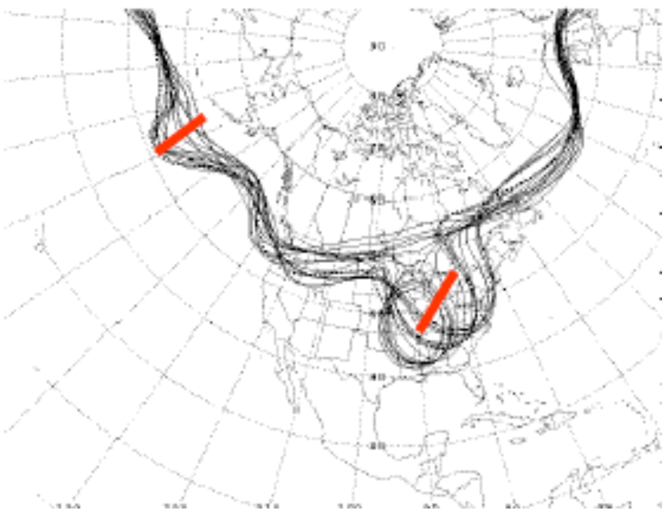
- Using the flow-dependent $\mathbf{P}_i^b$, analysis is expected to be improved significantly

However, it is computational hugely expensive. $\mathbf{P}_i^b$, $\mathbf{L}_i$ n*n matrix, $n \sim 10^7$

computing equation * directly is impossible

Ensemble Kalman Filter (EnKF)

$$\mathbf{P}_i^b = \mathbf{L}_{i-1}\mathbf{P}_{i-1}^a\mathbf{L}_{i-1}{}^T + \mathbf{Q} \qquad *$$



Physically,

▪"errors of day" are the instabilities of the background flow. Strong instabilities have a few dominant shapes (perturbations lie in a low-dimensional subspace).

▪ It makes sense to assume that large errors are in similarly low-dimensional spaces that can be represented by a low order EnKF.

❖ Although the dimension of $\mathbf{P}_i^f$ is huge, the rank ($\mathbf{P}_i^f$) << n (dominated by the errors of the day)

$$\mathbf{P}_i^b \approx \frac{1}{m}\sum_{k=1}^{m}(x_k^f - x^t)(x_k^f - x^t)^T$$

Ideally $m \to \infty$

❖ Using ensemble method to estimate *

$$\mathbf{P}_i^b \approx \frac{1}{K-1}\sum_{k=1}^{K}(x_k^f - \overline{x^f})(x_k^f - \overline{x^f})^T$$

$$= \frac{1}{K-1}\mathbf{X}^b \bullet \mathbf{X}^{bT}$$

K ensemble members, K<<n

❖ Problem left: How to update ensemble ? i.e.: How to get $\mathbf{x}_i^a$ for each ensemble member?

Ensemble Update: two approaches

1. Perturbed Observations method:

An "ensemble of data assimilations"

- It has been proven that an observational ensemble is required (e.g., Burgers et al. 1998). Otherwise $\mathbf{P}^a_{i\ n\times n} = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}^b_i$ is not satisfied.

- Random perturbations are added to the observations to obtain observations for each independent cycle

$$\mathbf{y}^o_{i\,(k)} = \mathbf{y}^o_i + noise$$

- However, perturbing observations introduces a source of sampling errors (Whitaker and Hamill, 2002).

$$\mathbf{x}^b_{i\,(k)} = M\mathbf{x}^a_{i-1(k)}$$

$$\mathbf{P}^b_i \approx \frac{1}{K-1}\sum_{k=1}^{K}(x^b_k - \overline{x^b})(x^b_k - \overline{x^b})^T$$

$$\mathbf{K}_i = \mathbf{P}^b_i\mathbf{H}^T[\mathbf{H}\mathbf{P}^b_i\mathbf{H}^T + \mathbf{R}]^{-1}$$

$$\mathbf{x}^a_{i\,(k)} = \mathbf{x}^b_{i\,(k)} + \mathbf{K}_i(\mathbf{y}^o_{i\,(k)} - H\mathbf{x}^b_{i\,(k)})$$

Ensemble Update: two approaches

2. Ensemble square root filter (EnSRF)

- Observations are assimilated to update only the ensemble mean.

$$\overline{\mathbf{x}_i^a} = \overline{\mathbf{x}_i^b} + \mathbf{K}_i(\mathbf{y}_i^o - H\overline{\mathbf{x}_i^b})$$

- Assume analysis ensemble perturbations can be formed by transforming the forecast ensemble perturbations through a transform matrix

$$\mathbf{x}_i^b = M\mathbf{x}_{i-1}^a$$

$$\mathbf{P}_i^b \approx \frac{1}{K-1}\sum_{k=1}^{K}(x_k^b - \overline{x^b})(x_k^b - \overline{x^b})^T$$

$$\mathbf{K}_i = \mathbf{P}_i^b\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R}]^{-1}$$

$$\overline{\mathbf{x}_i^a} = \overline{\mathbf{x}_i^b} + \mathbf{K}_i(\mathbf{y}_i^o - H\overline{\mathbf{x}_i^b})$$

$$\mathbf{X}_i^a = \mathbf{T}_i\mathbf{X}_i^b$$

$$\mathbf{x}_i^a = \overline{\mathbf{x}_i^a} + \mathbf{X}^a$$

$$\frac{1}{k-1}\mathbf{X}^a\mathbf{X}^{a^T} = \mathbf{P}_{i\ n\times n}^a = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^b = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\frac{1}{k-1}\mathbf{X}^b\mathbf{X}^{bT} \Rightarrow \mathbf{X}_i^a = \mathbf{T}_i\mathbf{X}_i^b$$

Several choices of the transform matrix

- EnSRF, Andrews 1968, Whitaker and Hamill, 2002)

$$\mathbf{X}^a = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{X}^b, \tilde{\mathbf{K}} = \alpha \mathbf{K}$$

- EAKF (Anderson 2001)

$$\mathbf{X}^a = \mathbf{A}\mathbf{X}^b$$

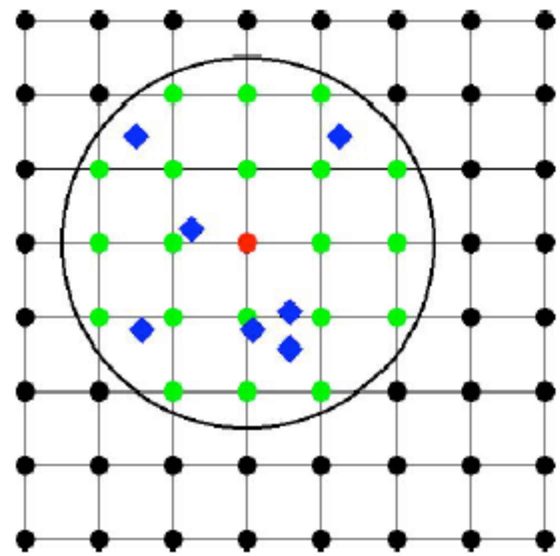
- ETKF (Bishop et al. 2001)

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{T}$$

one observation at a time

- LETKF (Hunt, 2005)

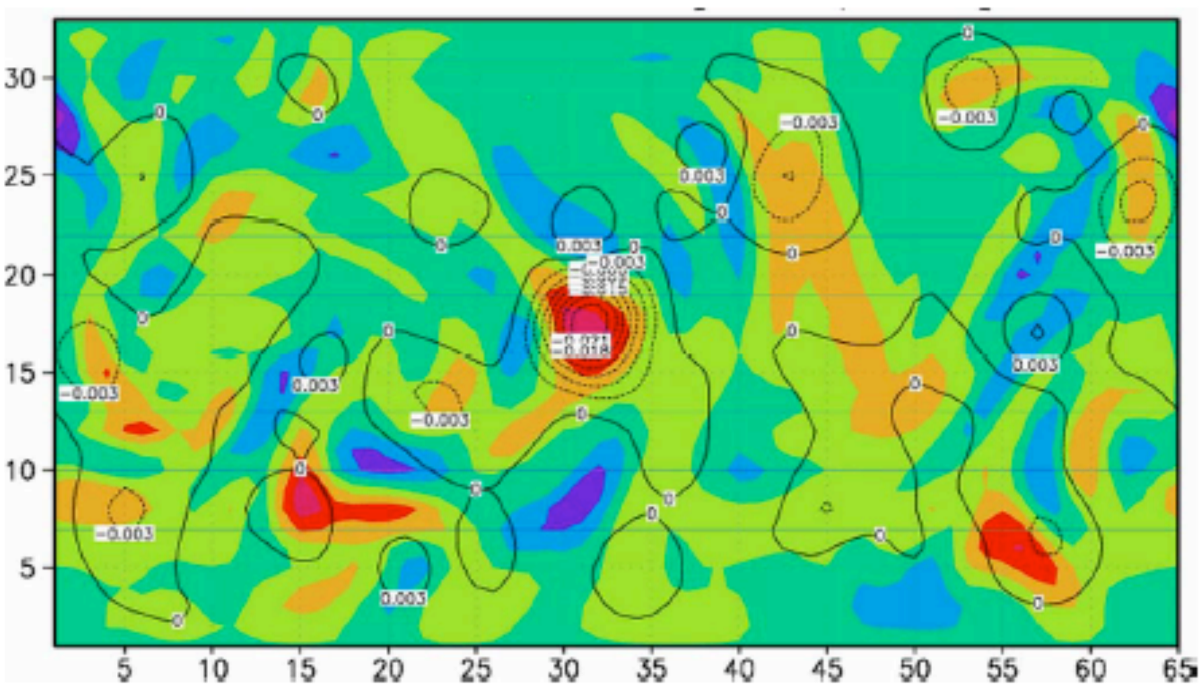
  Based on ETKF but perform analysis simultaneously in a local volume surrounding each grid point



An Example of the analysis corrections from 3D-Var (Kalnay, 2004)

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

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

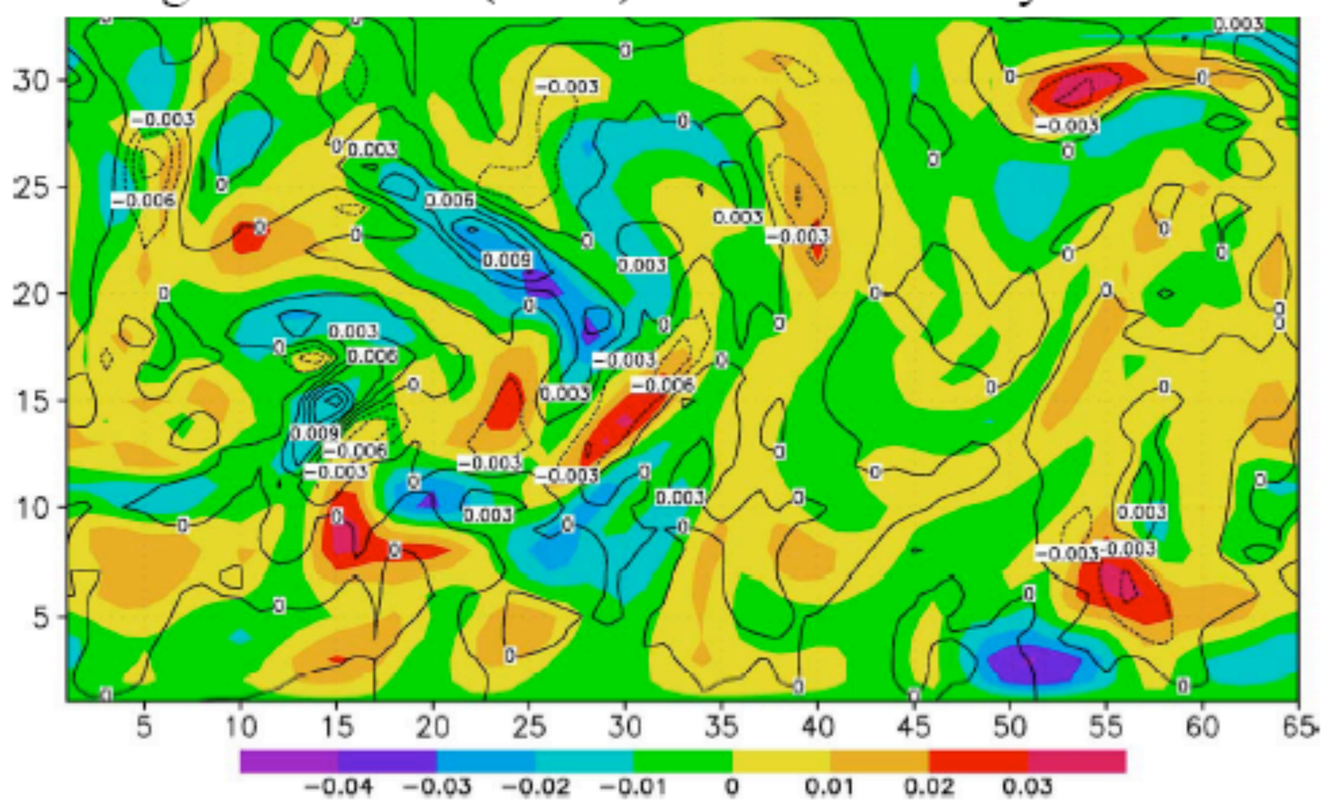


The analysis corrections due to the observations are isotropic because they don't know about the errors of the day

An Example of the analysis corrections from EnKF (Kalnay, 2004)

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

Summary steps of LETKF

1) Global 6 hr ensemble forecast starting from the analysis ensemble

2) Choose the observations used for each grid point

3) Compute the matrices of forecast perturbations in ensemble space X^b

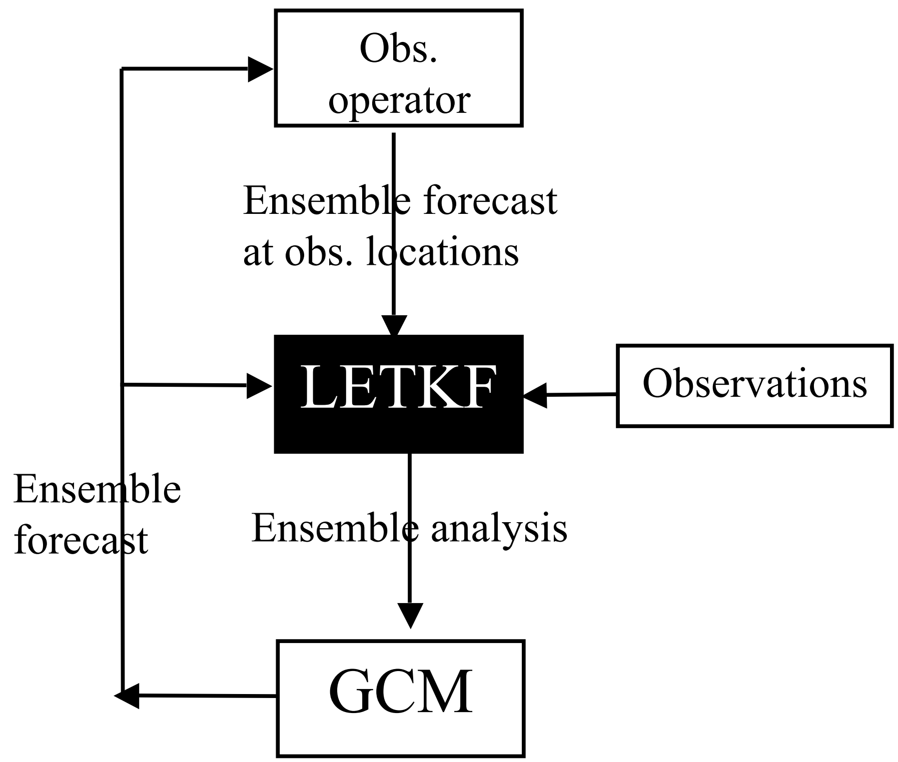
4) Compute the matrices of forecast perturbations in observation space Y^b

5) Compute P^b in ensemble space space and its symmetric square root

6) Compute w^a, the k vector of perturbation weights

7) Compute the local grid point analysis and analysis perturbations.

8) Gather the new global analysis ensemble. Go to 1



LETKF algorithm summary, Hunt 2005

Forecast step (done globally): advance the ensemble 6 hours, global model size n

$$\mathbf{x}_n^{b(i)} = M(\mathbf{x}_{n-1}^{a(i)}) \quad i = 1,\ldots,k$$

Analysis step (done locally). Local model dimension m, locally used obs s

$$\mathbf{X}^b = \mathbf{X} - \overline{\mathbf{x}}^b \text{ (mxk)} \quad \mathbf{Y}^b = H(\mathbf{X}) - H(\overline{\mathbf{x}}^b) \approx \mathbf{H}\mathbf{X}^b \text{ (sxk)}$$

$$\tilde{\mathbf{P}}^a = \left[(k-1)\mathbf{I} + \left(\mathbf{H}\mathbf{X}^b\right)^{\mathbf{T}} \mathbf{R}^{-1}\mathbf{H}\mathbf{X}^b \right]^{-1}$$ in ensemble space (kxk)

$$\mathbf{P}^a = \mathbf{X}^{aT}\mathbf{X}^a = \mathbf{X}^{bT}\tilde{\mathbf{P}}^a\mathbf{X}^b$$ in model space (mxm)

$$\mathbf{X}^a = \mathbf{X}^b(\tilde{\mathbf{P}}^a)^{1/2}$$ Ensemble analysis perturbations in model space (mxk)

$$\mathbf{w}_n^a = \tilde{\mathbf{P}}^a\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}_n^o - \overline{\mathbf{y}}_n^b)$$ Analysis increments in ensemble space (kx1)

$$\overline{\mathbf{x}}_n^a = \overline{\mathbf{x}}_n^a + \mathbf{X}^b\mathbf{w}_n^a$$ Analysis (mean) in model space (mx1)

$$\mathbf{x}_n^a = \overline{\mathbf{x}}_n^a + \mathbf{X}^a$$ Analysis ensemble in model space (mxk)

We finally gather all the analysis and analysis perturbations from each grid point and construct the new global analysis ensemble (nxk) and go to next forecast step

References posted: www.atmos.umd.edu/~ekalnay
Google: "chaos weather umd", publications

Ott, Hunt, Szunyogh, Zimin, Kostelich, Corazza, Kalnay, Patil, Yorke, 2004: Local Ensemble Kalman Filtering, Tellus, 56A,415–428.

Kalnay, 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341 pp. (3rd printing)

Hunt, Kalnay, Kostelich, Ott, Szunyogh, Patil, Yorke, Zimin, 2004: Four-dimensional ensemble Kalman filtering. Tellus 56A, 273–277.

Szunyogh, Kostelich, Gyarmati, Hunt, Ott, Zimin, Kalnay, Patil, Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. Tellus, 57A. 528-545.

Hunt, Kostelich and Szunyogh 2007: Efficient Data Assimilation for Spatiotemporal Chaos: a Local Ensemble Transform Kalman Filter. Physica D.

Whitaker and Hamill, 2002: Ensemble Data Assimilation Without Perturbed Observations. Monthly Weather Review, 130, 1913-1924.

EnKF vs 4D-Var

- EnKF is simple and model independent, while 4D-Var requires the development and maintenance of the adjoint model (model dependent)
- 4D-Var can assimilate asynchronous observations, while EnKF assimilate observations at the synoptic time.
- Using the weights w^a at any time 4D-LETKF can assimilate asynchronous observations and move them forward or backward to the analysis time

<u>Disadvantage of EnKF:</u>

- Low dimensionality of the ensemble in EnKF introduces sampling errors in the estimation of $\mathbf{P}^b$. 'Covariance localization' can solve this problem.