Ensemble methods in data assimilation

Summer school on data assimilation

IIRS, India, December 2012

Javier Amezcua
Ack: Ross Banister, Nancy Nichols, Stephano Migliorini, Michael Goodliff
Data Assimilation Research Center
University of Reading

University of Reading

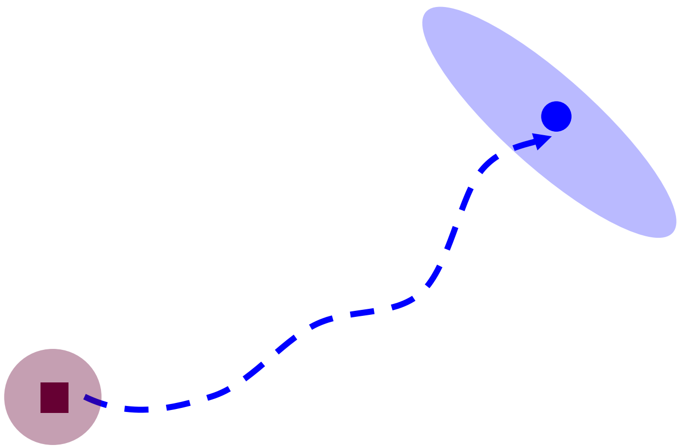
Content

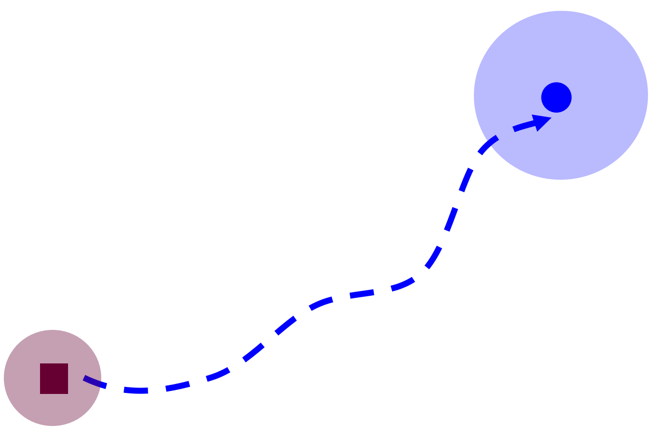
- Kalman filter and extended Kalman filter
- Ensemble Kalman filter
- Implementation aspects
- Parameter estimation
- Hybrid data assimilation

Kalman Filter and Extended Kalman Filter

Evolution of the uncertainty

In 3DVar, we considered the background covariance **B** to be fixed or static. In a way, it characterized the climatology of the system.





But from analysis to analysis, the background covariance can vary, it is **flow-dependent**.

How do we estimate this?

The Kalman Filter

Consider a system with linear evolution, linear observation operator, and Gaussian errors.

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t \qquad \mathbf{w} \sim N(\mathbf{0}, \mathbf{Q}) \qquad \mathbf{y}_t = h(\mathbf{x}_t) + \mathbf{v}_t \qquad \mathbf{v} \sim N(\mathbf{0}, \mathbf{R})$$

$$\mathbf{F} \in \Re^{N \times N}, \mathbf{x} \in \Re^{N} \qquad \mathbf{y} \in \Re^{L}$$

$$Cov[\mathbf{w}, \mathbf{v}] = \mathbf{0}$$

Independence of model and observational error.

In this case, it is enough to estimate the state of the mean and covariance of the system (closure at the 2nd moment).

$$E[\mathbf{x}] = \bar{\mathbf{x}} \qquad Cov[\mathbf{x}] = \mathbf{P}$$

The Kalman Filter

The filter has 2 steps.

Forecast
$$\begin{cases} \bar{\mathbf{x}}_t^b = \mathbf{F}\bar{\mathbf{x}}_{t-1}^a \\ \mathbf{P}_t^b = \mathbf{F}\mathbf{P}_{t-1}^a\mathbf{F}^T + \mathbf{Q} \end{cases}$$

Analysis
$$\begin{cases} \bar{\mathbf{x}}_t^a = \bar{\mathbf{x}}_t^b + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{H}\bar{\mathbf{x}}_t^b\right) \\ \mathbf{P}_t^a = \left(\mathbf{I} - \mathbf{K}_t\mathbf{H}\right)\mathbf{P}_t^b \end{cases}$$

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

The Extended Kalman Filter

When the dynamics is nonlinear, we use a 1st order Taylor expansion.

Forecast

$$\begin{cases} \overline{\mathbf{x}}_t^b = f\left(\overline{\mathbf{x}}_{t-1}^a\right) \\ \mathbf{P}_t^b = \mathbf{F}'\mathbf{P}_{t-1}^a\mathbf{F}'^T + \mathbf{Q} \end{cases}$$

Analysis

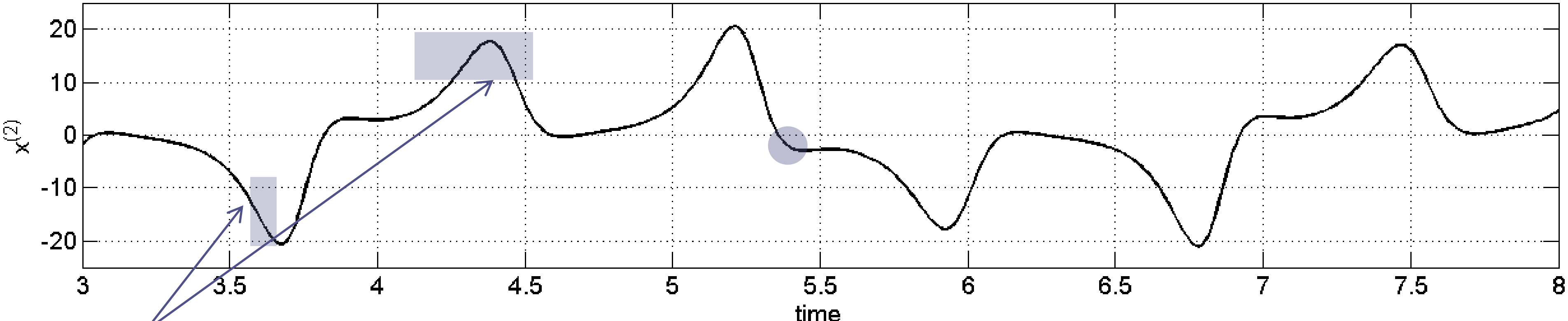
$$\begin{cases} \overline{\mathbf{x}}_t^a = \overline{\mathbf{x}}_t^b + \mathbf{K}'_t\left(\mathbf{y}_t - \mathbf{H}'\overline{\mathbf{x}}_t^b\right) \\ \mathbf{P}_t^a = \left(\mathbf{I} - \mathbf{K}'_t\,\mathbf{H}'\right)\mathbf{P}_t^b \end{cases}$$

Kalman gain

$$\mathbf{K}_t = \mathbf{P}_t^b\mathbf{H}'^T\left(\mathbf{H}'\mathbf{P}_t^b\mathbf{H}'^T + \mathbf{R}\right)^{-1}$$

$$= \mathbf{P}_t^a\mathbf{H}'^T\,\mathbf{R}^{-1}$$

$$\mathbf{F}' = \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}_{t-1}^a} \qquad \mathbf{H}' = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}_t^b}$$



The accuracy of the linear approximation depends on the length of the forecast window.

Handling **P**

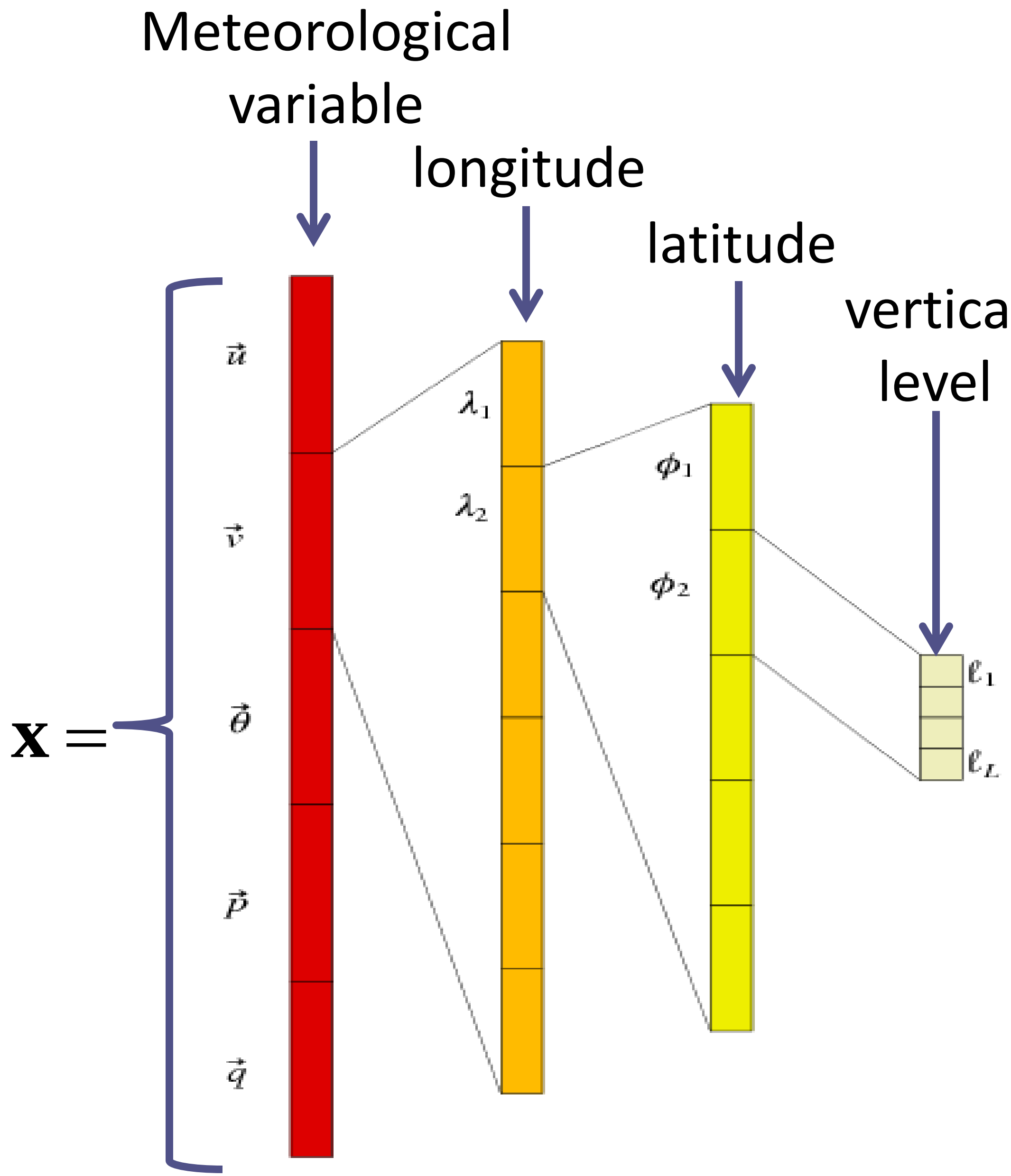
Both the KF and EKF require computations on the covariance matrix **P**.

In NWP applications:

$$\mathbf{x} \in \Re^N, N \sim 10^8$$

$$\mathbf{P} \in \Re^{N \times N}$$

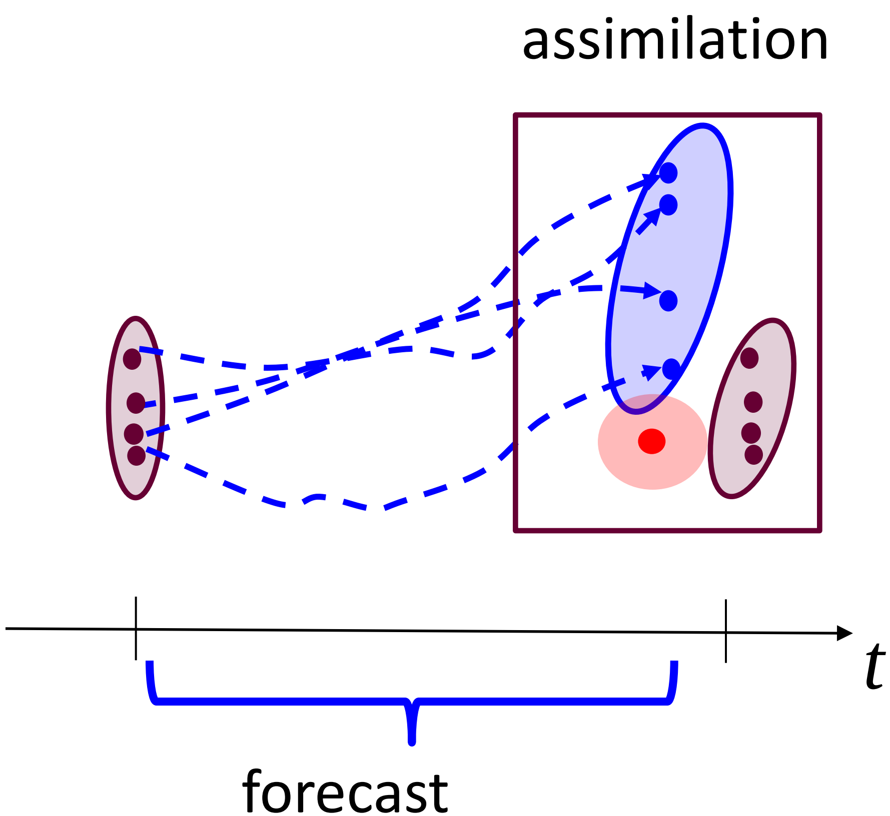
It is **not feasible to constantly compute it and store it**!



Ensemble Kalman Filter

Ensemble Kalman Filter

The Ensemble Kalman Filter is a Monte-Carlo implementation of the KF.
Study a family of *M* solutions, which we call ensemble (Evensen, 1994).



$$\mathbf{X} = \left[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_M \right] \in \Re^{N \times M}$$

Ensemble Kalman Filter

$$\mathbf{X} = \left[\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \cdots \,|\, \mathbf{x}_M \right] \in \Re^{N \times M}$$

We use the **sample estimators** of the mean and covariance for the KF analysis step.

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m \qquad \mathbf{P} = \frac{1}{M-1} \hat{\mathbf{X}} \hat{\mathbf{X}}^T$$

Where we define an ensemble of perturbations:

$$\hat{\mathbf{X}} = \left[\mathbf{x}_1 - \bar{\mathbf{x}} \,|\, \mathbf{x}_2 - \bar{\mathbf{x}} \,|\, \cdots \,|\, \mathbf{x}_M - \bar{\mathbf{x}} \right] \in \Re^{N \times M}$$

EnKF features

- Unlike the EKF, the EnKF **makes use of the full nonlinear model**: more accurate determination of error growth (and of error saturation).

- The Kalman gain **K** is computed **without the need to calculate $\mathbf{P}^b$**. This is particularly advantageous when observations are processed serially.

- No need to linearize the observation operator *h.*

EnKF features

- Each ensemble member can be propagated forward in time independently: **highly parallel algorithm**.

- Model error term can be included as a perturbation of the deterministic forecast.

So, how does it work?

Ensemble Kalman Filter

To assimilate the mean and covariance, we just follow the KF equations.

background analysis

$$\mathbf{P}^b \Rightarrow \mathbf{P}^a$$

$$\overline{\mathbf{x}}^b \Rightarrow \overline{\mathbf{x}}^a$$

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

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}\ \mathbf{H}^T\right)\mathbf{P}^b$$

However, how to assimilate the individual ensemble members is not trivial.

$$\mathbf{X}^b \Rightarrow \mathbf{X}^a$$

Not unique…, since the square root of a matrix is not uniquely defined.

Ensemble Kalman Filter: families

Stochastic (Burgers *et al.,* 1996; Houtekamer and Mitchell, 1998)

- Apply the KF equations to each ensemble member.
- Requires perturbed observations for each member.

Deterministic implementation: ensemble square root filters (Tippet *et al.*, 2003).

- Serial EnSRF (Whitaker and Hamill, 2002)
- Ensemble Adjustment Kalman Filter (Anderson, 2001)
- Ensemble Transform Kalman Filter (Bishop et al, 2001), LETKF (Hunt et al, 2007)

Stochastic EnKF

Having our ensemble: $\mathbf{X} = \left[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_M\right] \in \Re^{N\times M}$

$$\mathbf{x}_m^a = \mathbf{x}_m^b + \mathbf{K}\left(\mathbf{y}_m - \mathbf{H}\mathbf{x}_m^b\right)$$

$$\mathbf{y}_m = \mathbf{y} + \boldsymbol{\eta}_m, \quad \boldsymbol{\eta} \sim N\left(\mathbf{0}, \mathbf{R}\right)$$

Why do we need to perturb observations? To ensure correct analysis covariance.

Note that we fulfill the KF covariance equation statistically.

Deterministic EnKF

Having our ensemble: $\mathbf{X} = [\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\cdots|\, \mathbf{x}_M] \in \Re^{N\times M}$

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

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b$$

$$\left[\begin{array}{l} \hat{\mathbf{X}}^a = \mathbf{S}^a \hat{\mathbf{X}}^b \longrightarrow \text{EAKF} \\ \hat{\mathbf{X}}^a = \hat{\mathbf{X}}^b \mathbf{T}^a \longrightarrow \text{ETKF} \end{array}\right.$$

This transformation has to respect the KF covariance equation.

ETKF

Having our ensemble:

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

$$\mathbf{T}^a \in \Re^{M \times M}$$

The transform matrix is relatively small, since *M*~10-100.

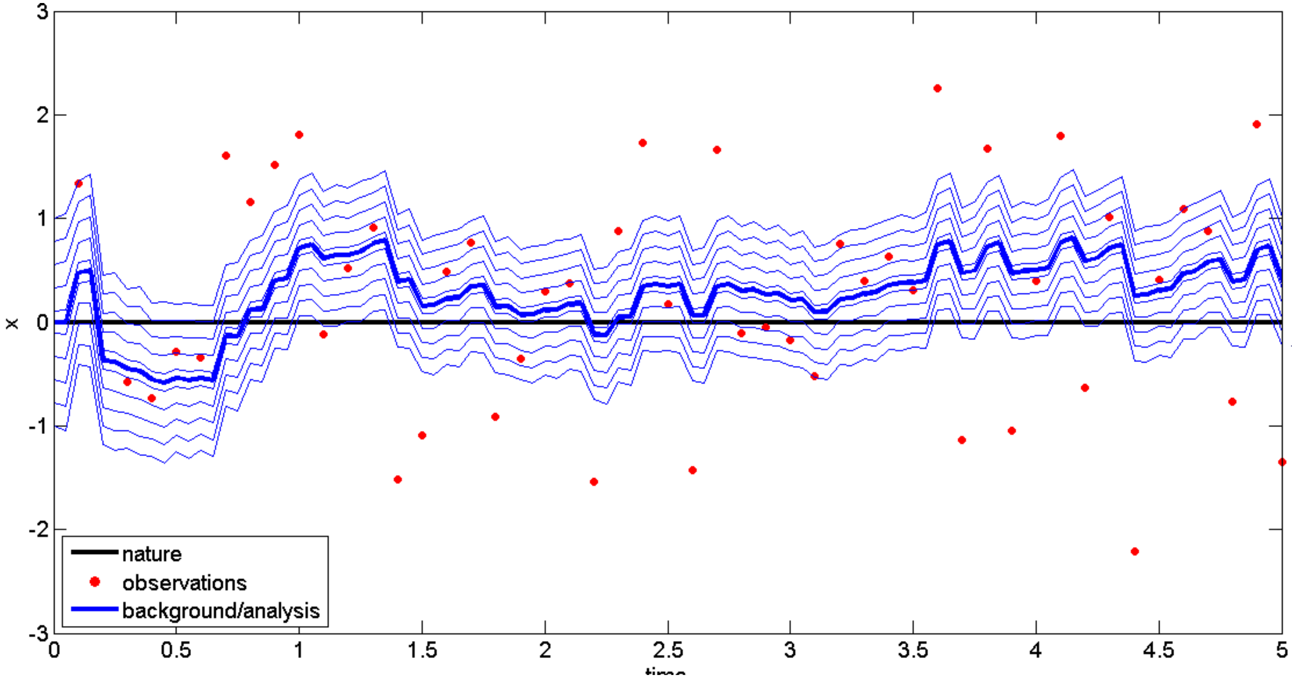
$$\mathbf{T}^a = \mathbf{C}\left(\mathbf{I} + \mathbf{\Gamma}\right)^{-1/2}\mathbf{C}^{\mathrm{T}} = \left((M-1)\tilde{\mathbf{P}}^a\right)^{1/2}$$

This is the analysis covariance computed in the ensemble space, which is low dimensional.

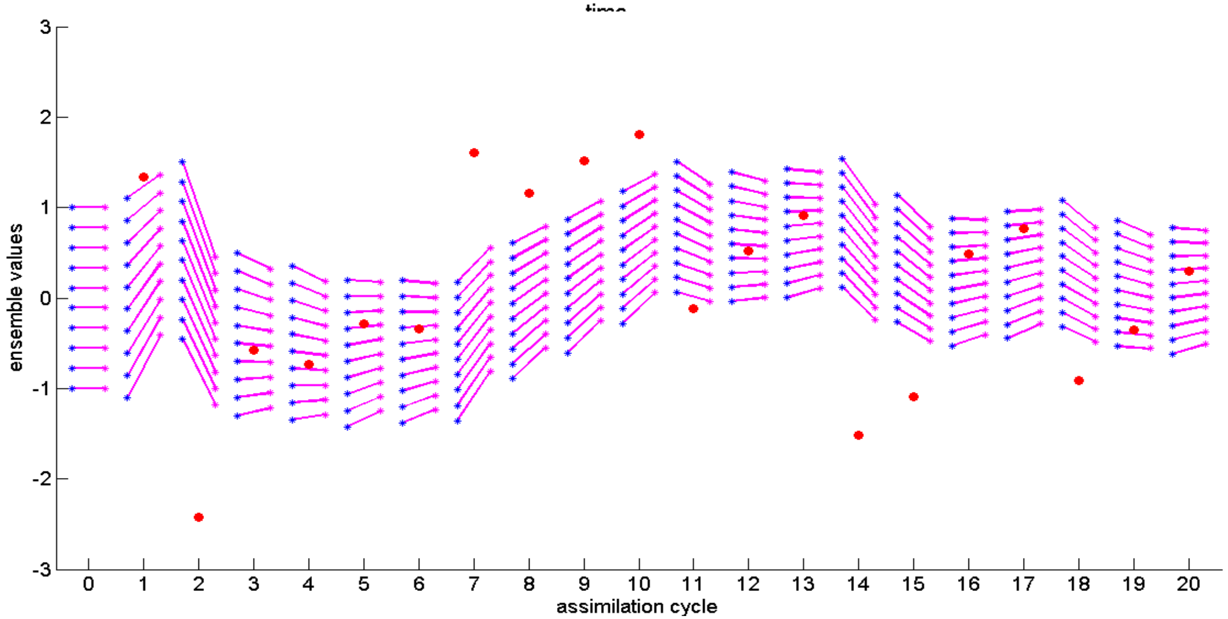
$$\mathbf{C}\,\mathbf{\Gamma}\,\mathbf{C}^{\mathrm{T}} = eig\left(\frac{\left(\mathbf{H}\mathbf{X}^{\mathbf{b}}\right)^{\mathbf{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{X}^{\mathbf{b}}\right)}{M-1}\right)$$

ETKF in a simple univariate model

$$x_{t+1} = (1+\Delta)x_t \,;\, x_{true} = 0$$

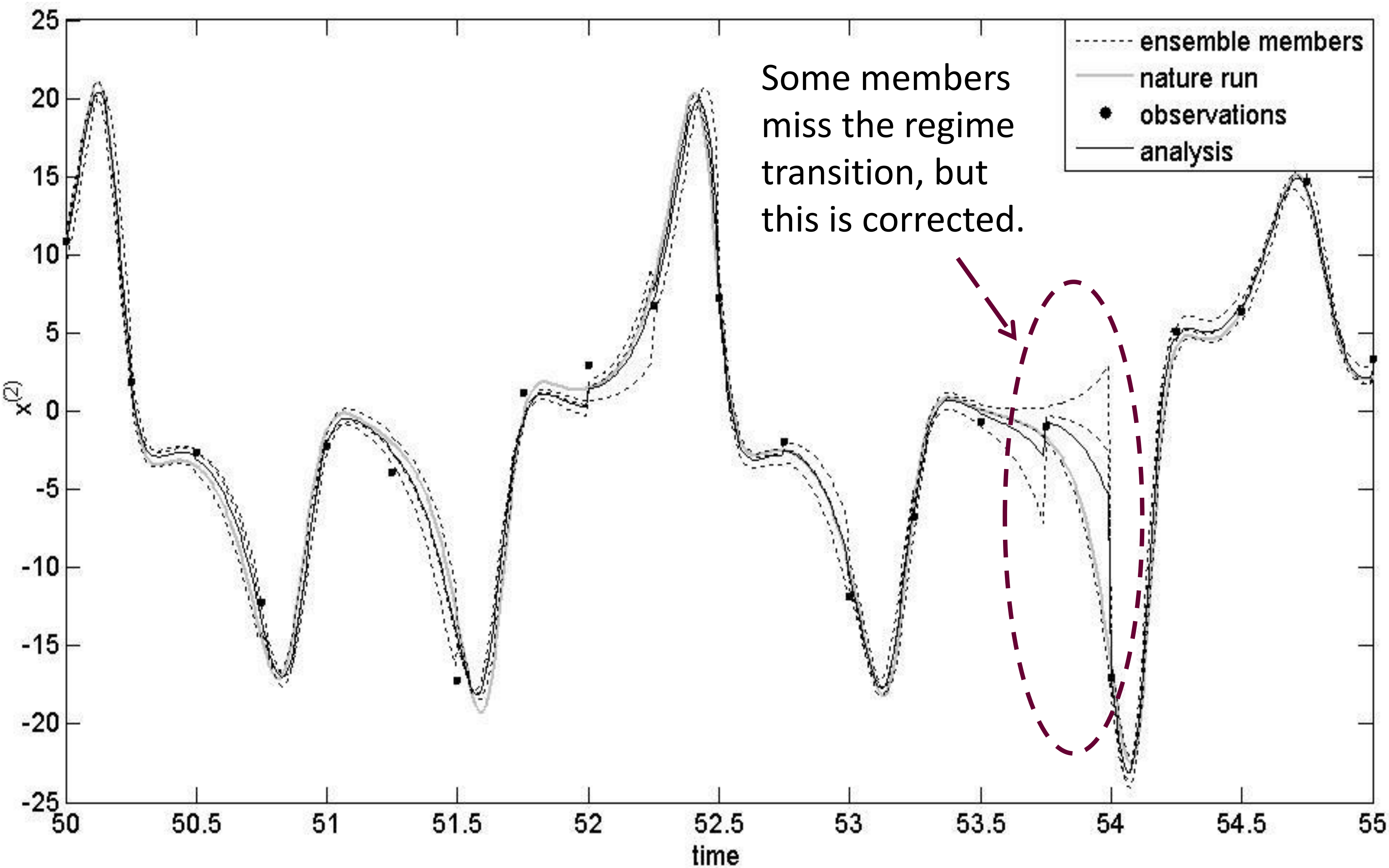


Time evolution



Only analysis

ETKF in Lorenz 1963 with 3 members

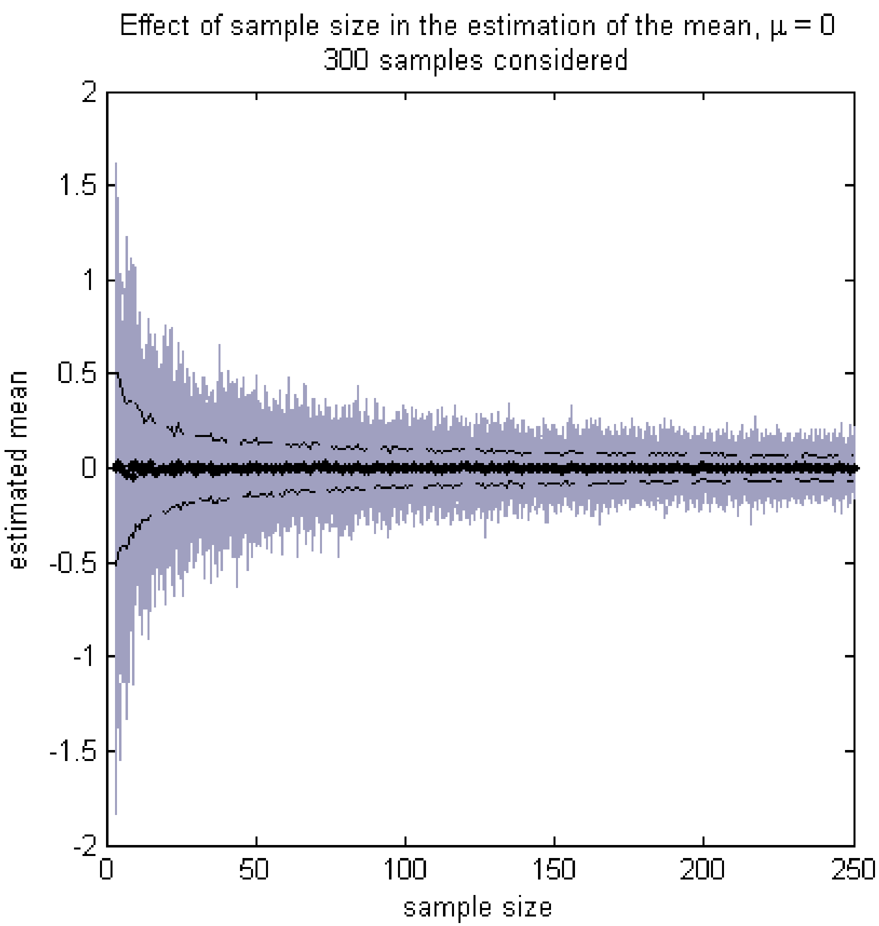


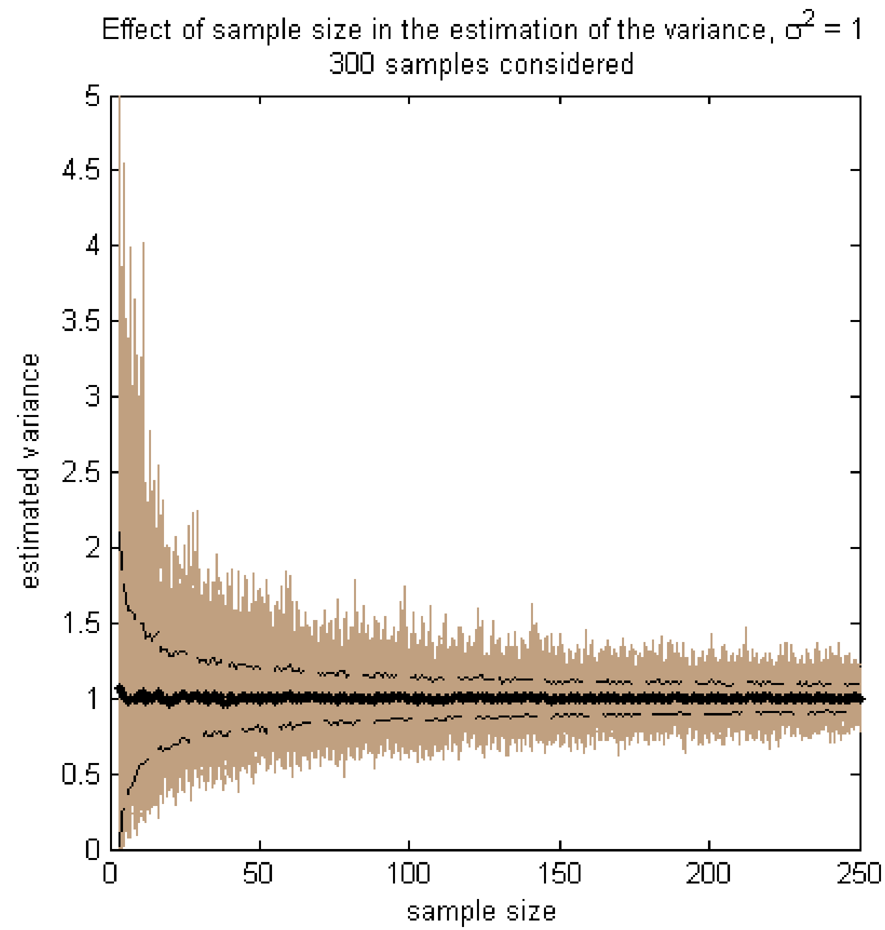
Implementation aspects

Sampling

There is always sampling noise in the estimators, this reduces as the ensemble size increases.

Example with a univariate Gaussian distribution.





Sampling

Two effects of **finite sample size**:

- **Underestimation** of sample **covariance**.
- **Spurious** long-range **correlations**.

Fixes:

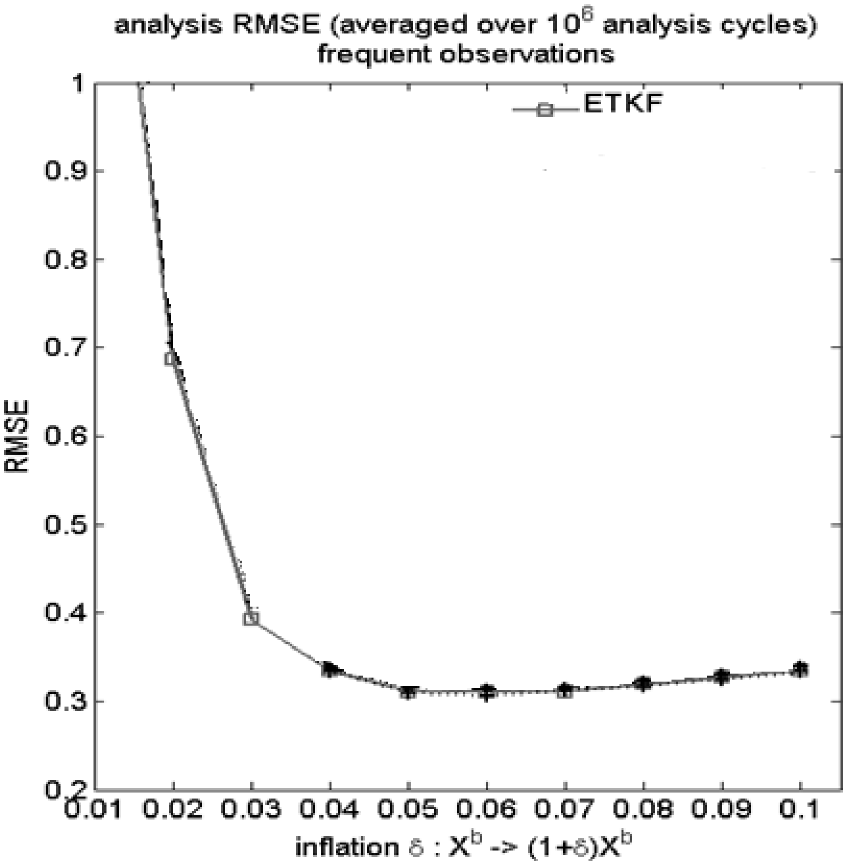
- Covariance inflation
- Covariance localization

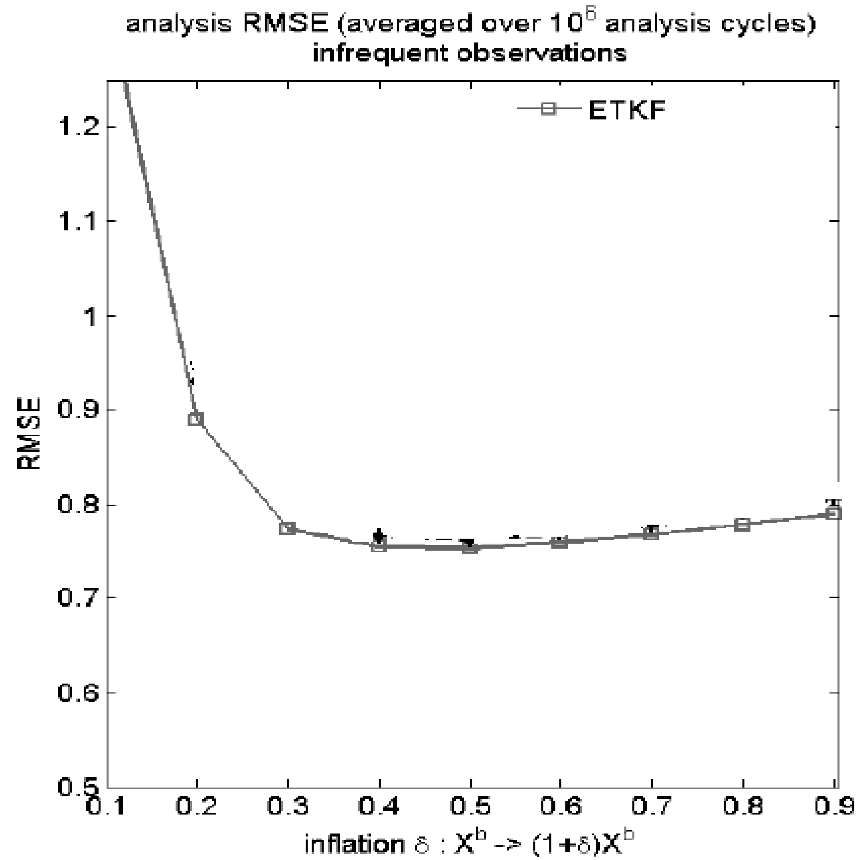
Also, the sample covariance matrix is singular for $N>M$…

How many members would we need? At least as many as the **number of unstable directions of error growth**?

Covariance inflation and performance.

Lorenz 1963 $\mathbf{H} = \mathbf{I}, \mathbf{R} = 2\mathbf{I}, M = 3$





Covariance localization

- When forecast error covariance is mispecified (e.g., due to neglecting model error, or when $M << N$), it may include spurious correlations between very distant grid points.

- A common solution is to multiply each $\mathbf{P}^b$ element by an appropriate weight that reduces long-distance correlations.

- This ensures that only the components of $\mathbf{P}^b$ believed to represent the corresponding components of $\mathbf{P}^b$ accurately are retained.

Covariance localization: an example

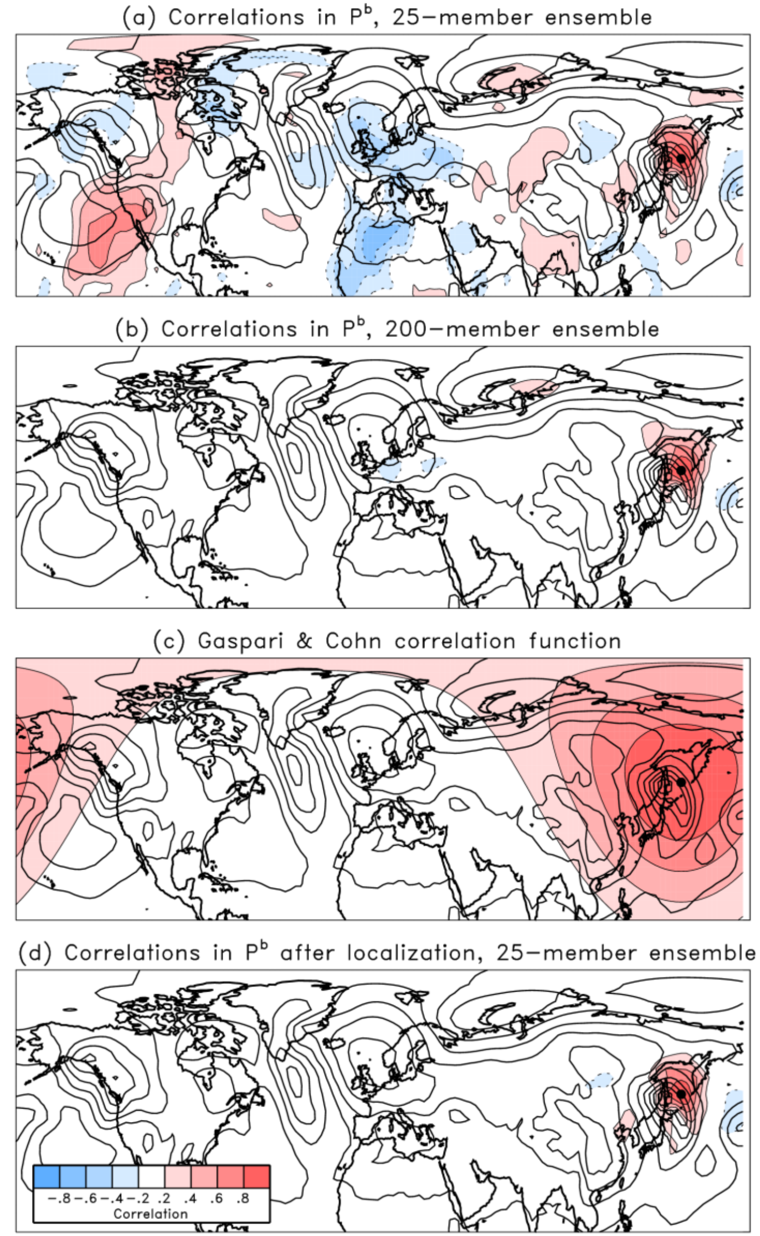
(a) $\mathbf{P}^b_e$ (N=25)

(b) $\mathbf{P}^b_e$ (N=200)

(c) Correlation function with compact support

(d) localized $\mathbf{P}^f_e$ (N=25)

From Fig. 6.4
of Hamill, 2006

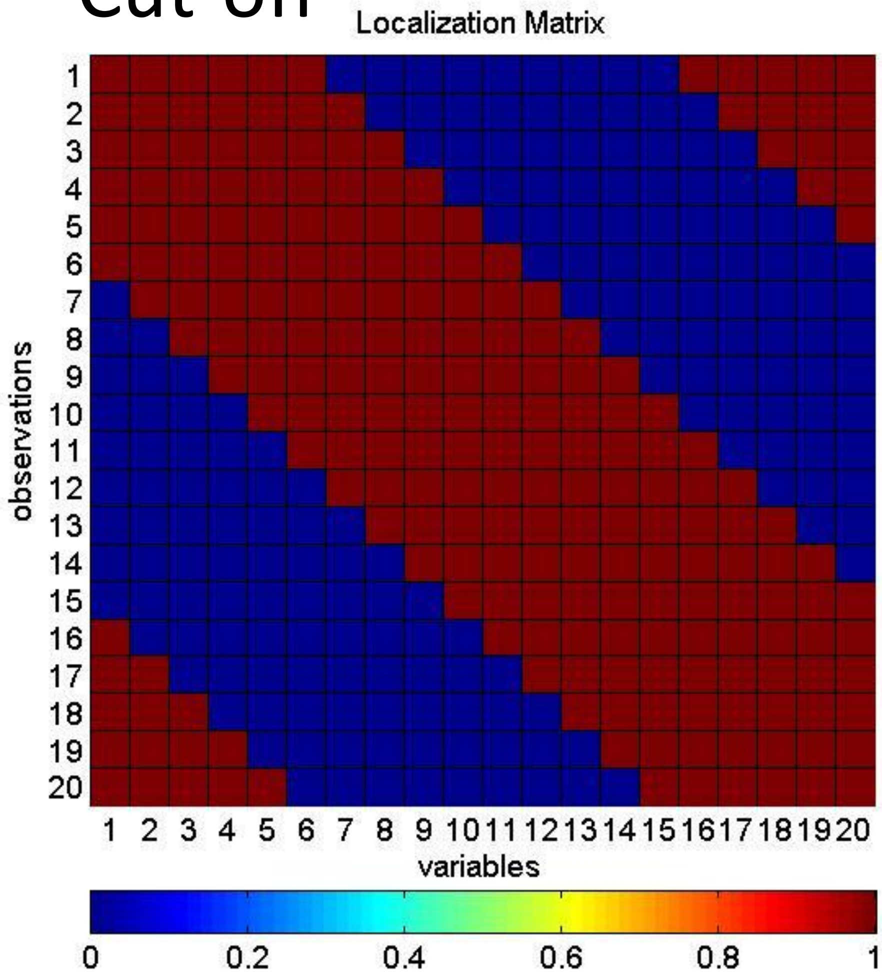


Localization

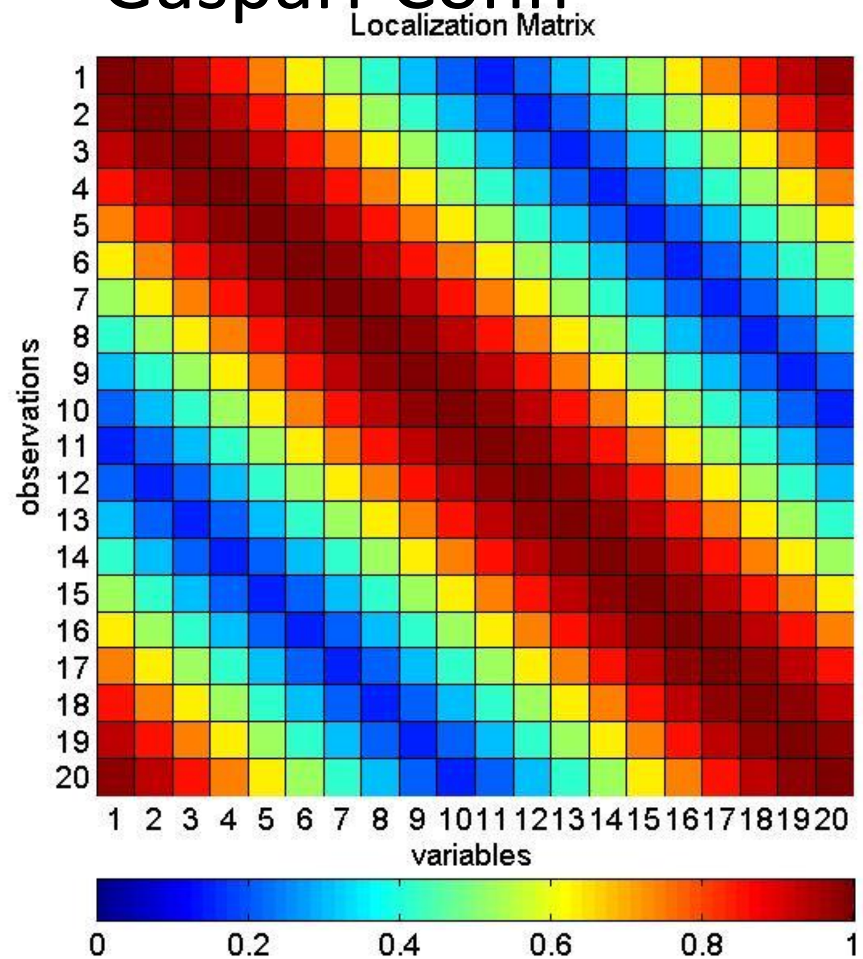
Example using Lorenz 1996

$$\mathbf{C} \circ \mathbf{P}^b$$

Cut-off



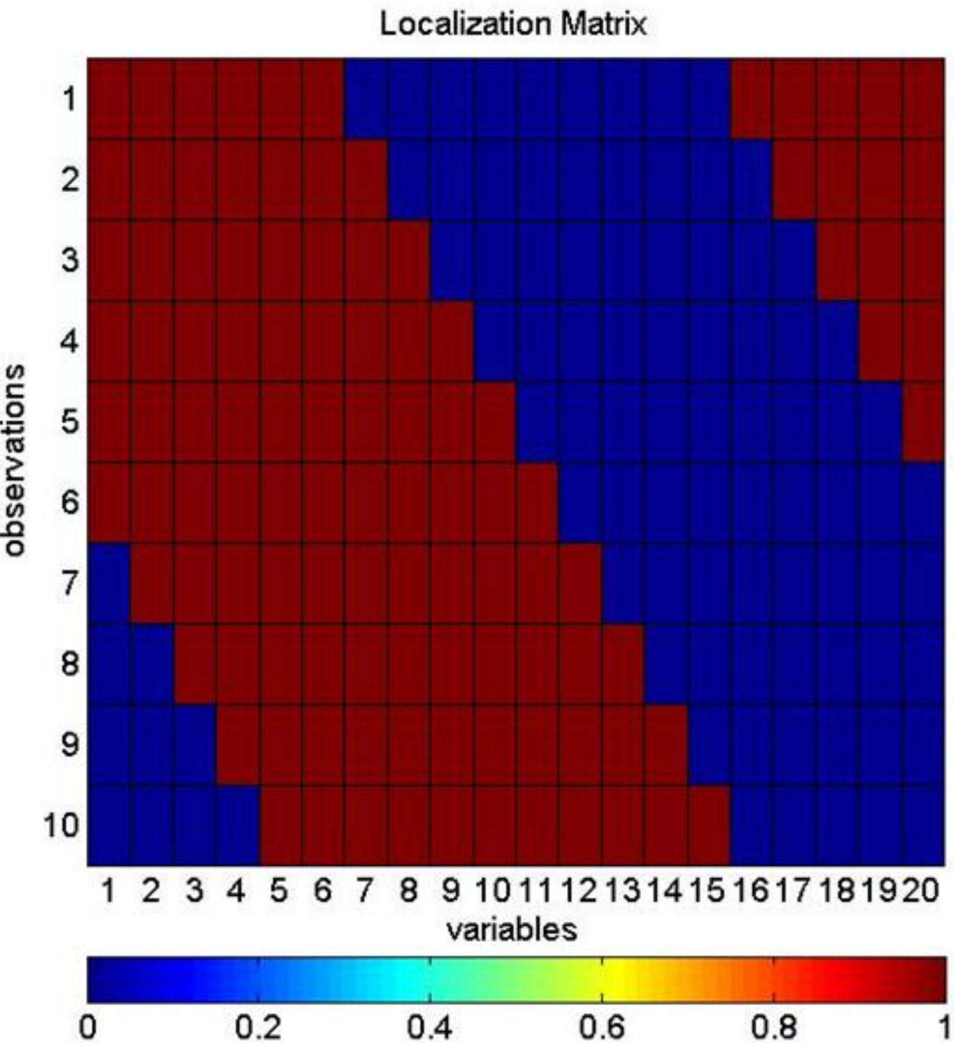
Gaspari-Cohn



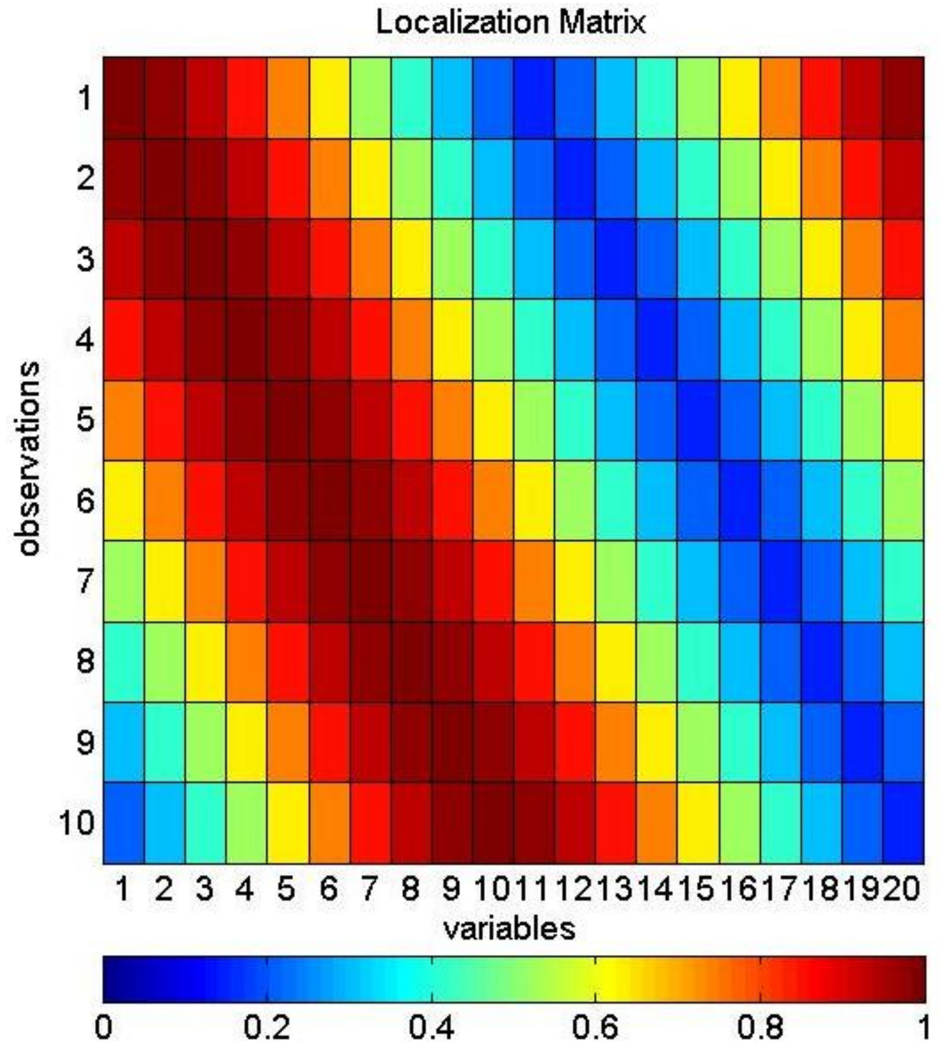
Localization $\mathbf{C} \circ \left(\mathbf{P}^b \mathbf{H}^T\right)$

Example using Lorenz 1996, observing every other variable.

Cut off



Gaspari-Cohn



Gridpoint R-localization: LETKF

We go gridpoint by gridpoint and perform the update using observations within a radius of influence.

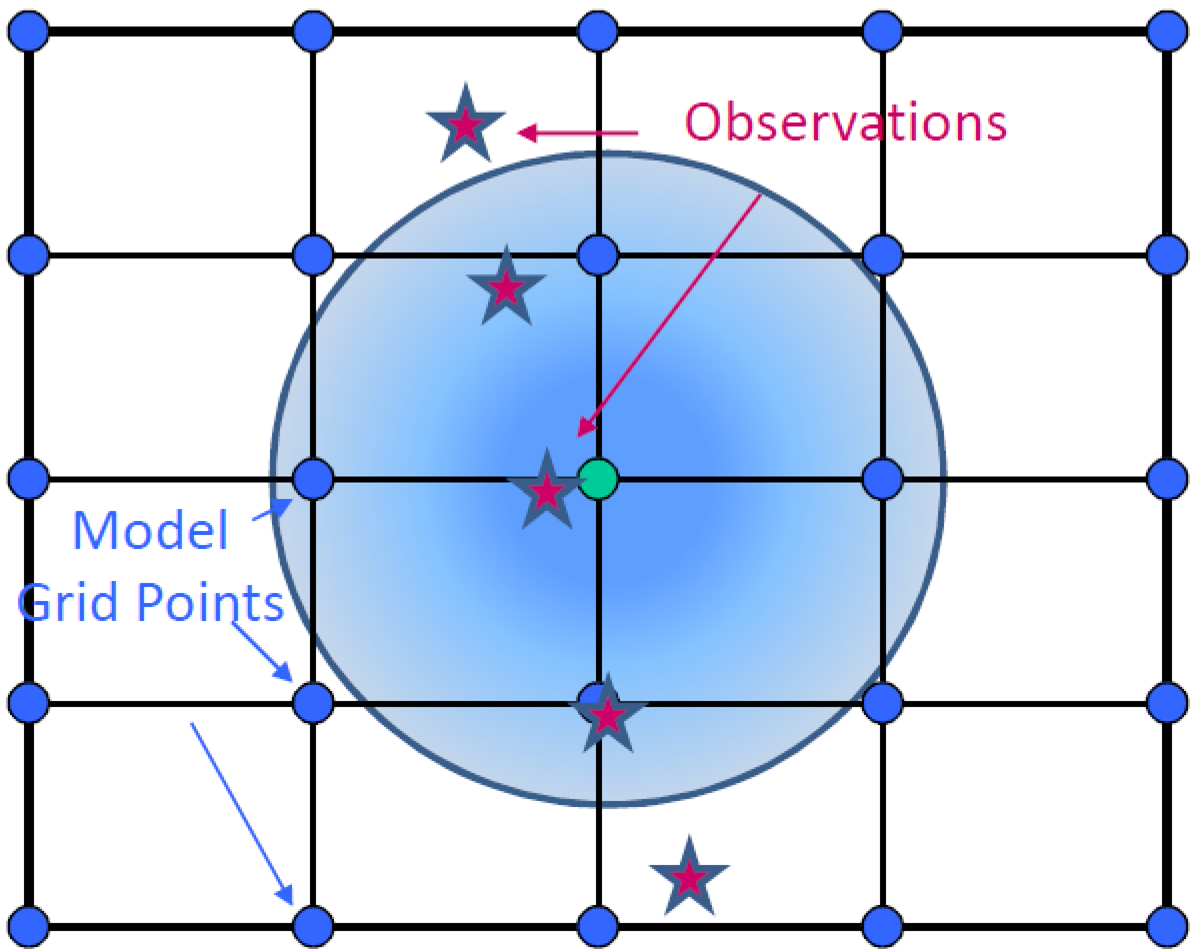
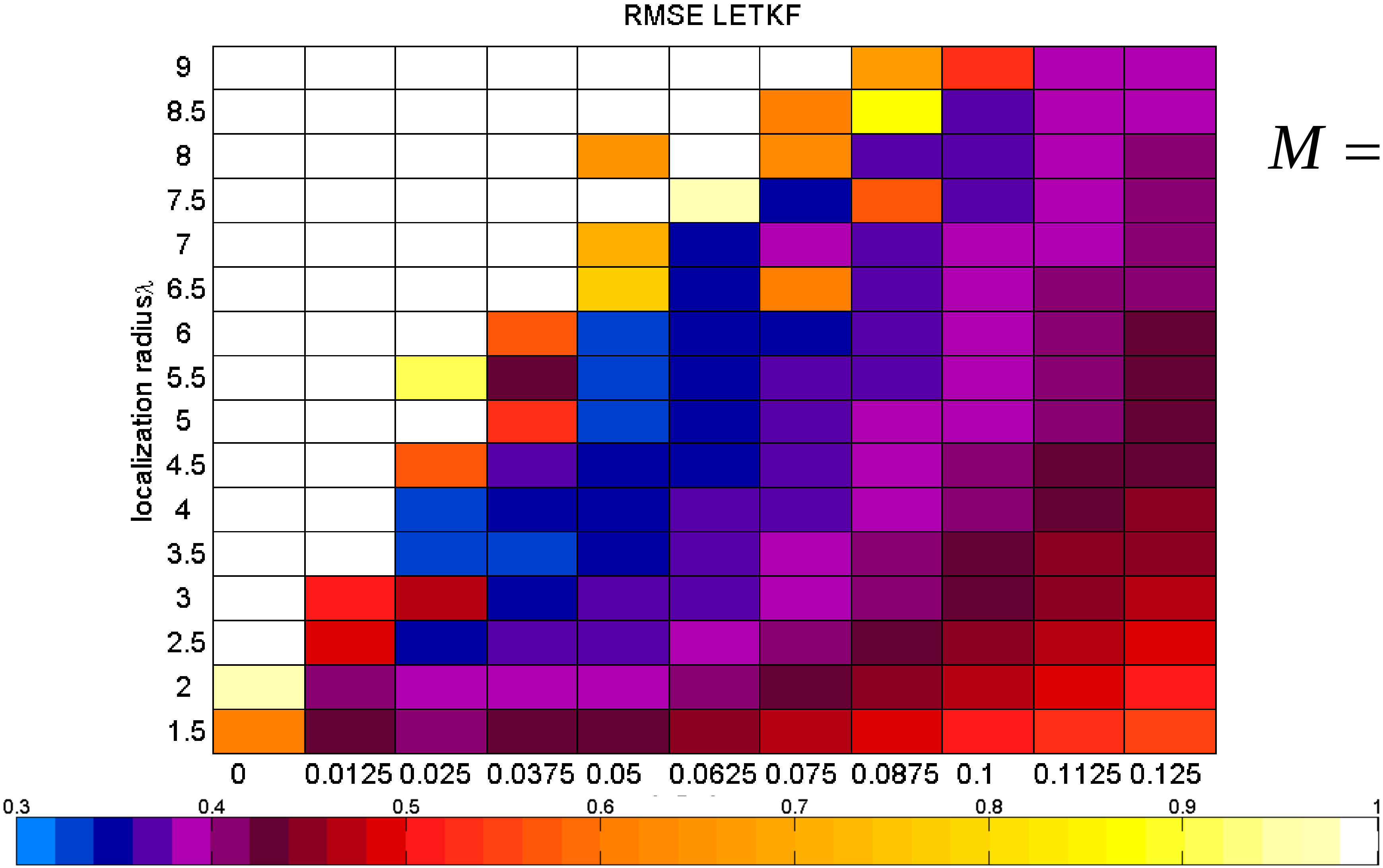


Image courtesy of Steven Greybush.

Combined effects of inflation and localization

Experiments with Lorenz 1996 and 40 variables, observing every 2 time steps and every other variable.



$$M = 10$$

Parameter estimation

Extending the state vector

The state vector can be extended with the parameters of the model. We can use the covariance to update values of poorly known parameters.

$$\mathbf{x}_t = f\left(\mathbf{x}_{t-1};\boldsymbol{\theta}\right)$$

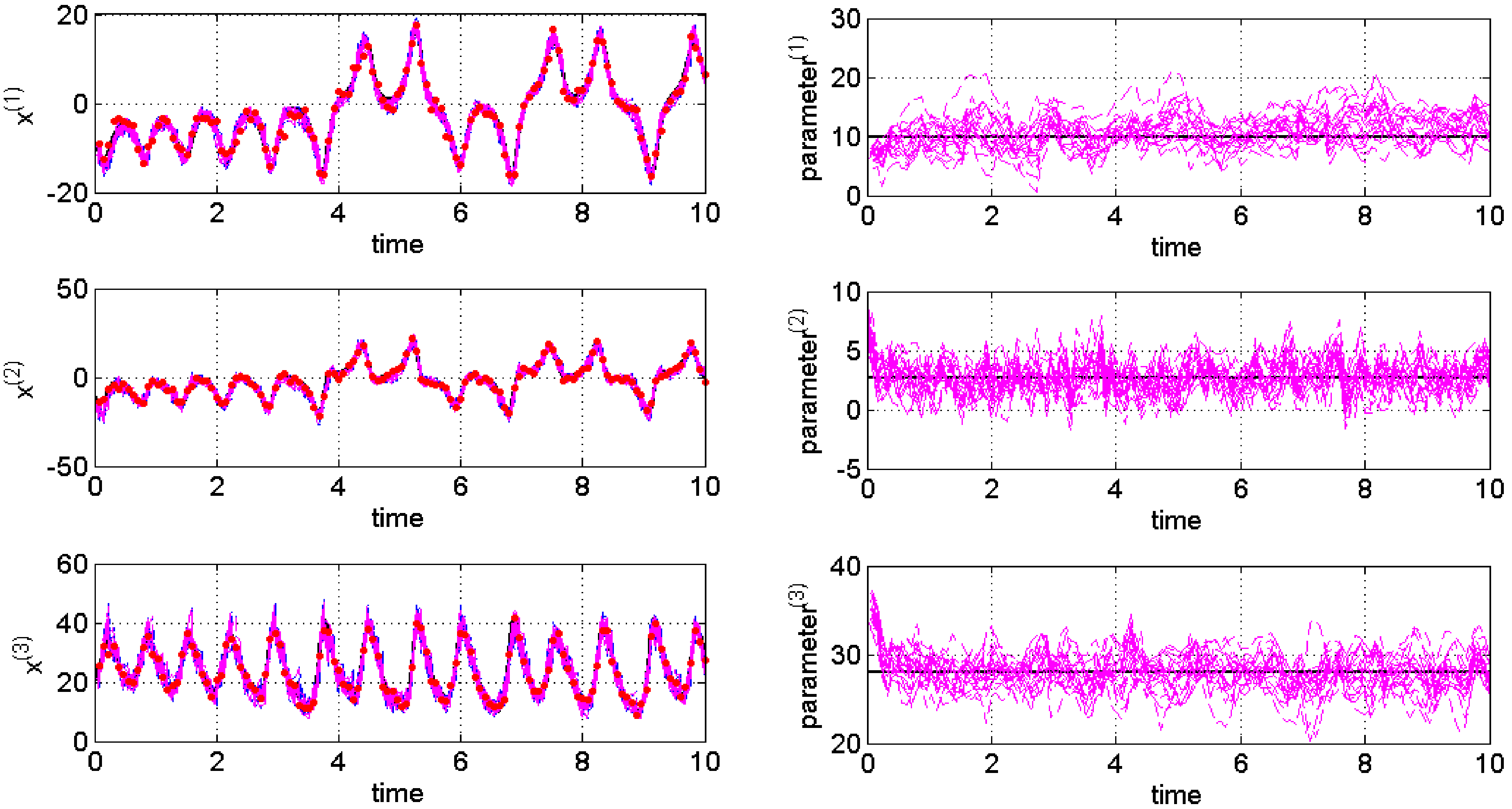
$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \cdots \\ \boldsymbol{\theta} \end{bmatrix} \qquad P = \begin{bmatrix} Cov\left(\mathbf{x},\mathbf{x}\right) & Cov\left(\mathbf{x},\boldsymbol{\theta}\right) \\ Cov\left(\mathbf{x},\boldsymbol{\theta}\right) & Cov\left(\boldsymbol{\theta},\boldsymbol{\theta}\right) \end{bmatrix}$$

This ‘cross-covariance’ carries information from state variables to parameters. Remember: **parameters are not observables.**

There are no dynamics for the parameters, one can perturb them during the forecast.

Example

Using Lorenz 1963, estimate the values of the state variables and the parameters.



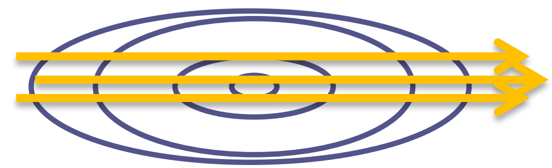
Hybrid data assimilation

Combining the best of 2 worlds

A static covariance is full rank, it is invertible, it gives idea of the climatology.

An ensemble covariance has information of the flow, but it can be singular and contains sampling errors.

Climatology



Flow/State Dependence

$$\mathbf{B} = \alpha \mathbf{B}_{static} + (1-\alpha)\mathbf{B}_{ensemble}$$

Compromise?

There are several ways to implement this.

Extended control variable (NCEP)

- Incorporate ensemble perturbations directly into variational cost function through extended control variable

$$J(\mathbf{x}_\mathrm{f}', \boldsymbol{\alpha}) = \boxed{\beta_\mathrm{f} \frac{1}{2} (\mathbf{x}_\mathrm{f}')^\mathrm{T} \mathbf{B}_\mathrm{f}^{-1} (\mathbf{x}_\mathrm{f}')} + \boxed{\beta_\mathrm{e} \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{\alpha}^n)^\mathrm{T} \mathbf{L}^{-1} (\boldsymbol{\alpha}^n)} + \frac{1}{2} (\mathbf{H}\mathbf{x}_\mathrm{t}' - \mathbf{y}')^\mathrm{T} \mathbf{R}^{-1} (\mathbf{H}\mathbf{x}_\mathrm{t}' - \mathbf{y}')$$

$$\mathbf{x}_\mathrm{t}' = \mathbf{x}_\mathrm{f}' + \sum_{n=1}^{N} (\boldsymbol{\alpha}^n \circ \mathbf{x}_\mathrm{e}{}^n)$$

β_f & β_e: weighting coefficients for fixed and ensemble covariance respectively

$\mathbf{x}_\mathrm{t}'$: (total increment) sum of increment from fixed/static **B** ($\mathbf{x}_\mathrm{f}'$) and ensemble **B**

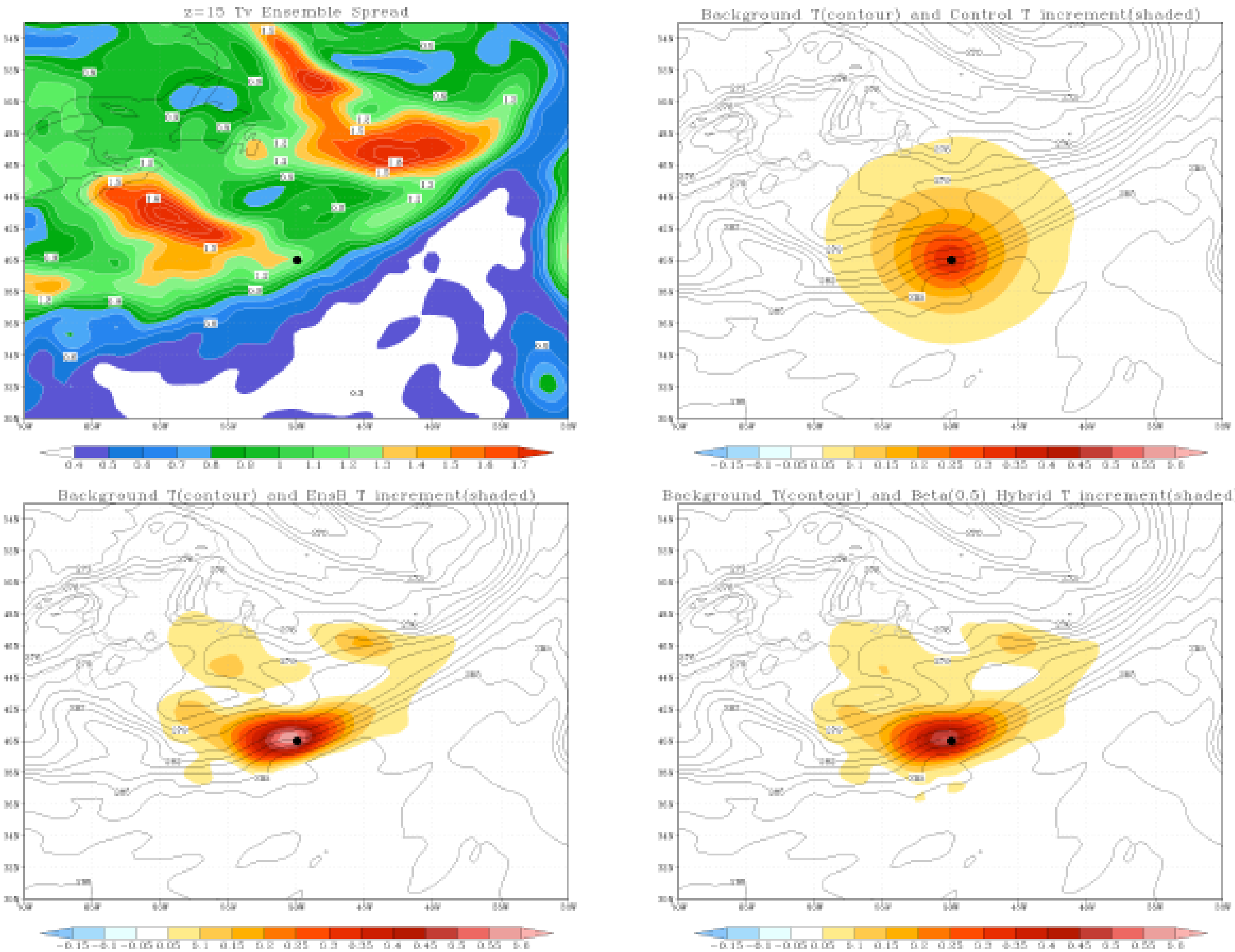
α_k: extended control variable; $\mathbf{x}_k^\mathrm{e}$:ensemble perturbations

- analogous to the weights in the LETKF formulation

L: correlation matrix [effectively the localization of ensemble perturbations]

Single observation experiment

3DVar, EnKF and Hybrid techniques using GFS.



Single 850mb Tv observation (1K O-F, 1K error)

Image courtesy of Daryl Kleist, NOAA.

Some words on ensembles

Nowadays, NWP does ensemble forecasts to quantify uncertainty. They are readily available for DA.

Sample covariance has information about the errors of the day, it ‘knows’ about the flow. Nonetheless, it has sampling error and can be singular.

Parameter estimation can be implemented in a straightforward fashion.

EnKFs do not require adjoints.