Hybrid methods: an introduction

Javier Amezcua

NCEO/DARC training course

Data Assimilation Research Centre

University of Reading

The set-up

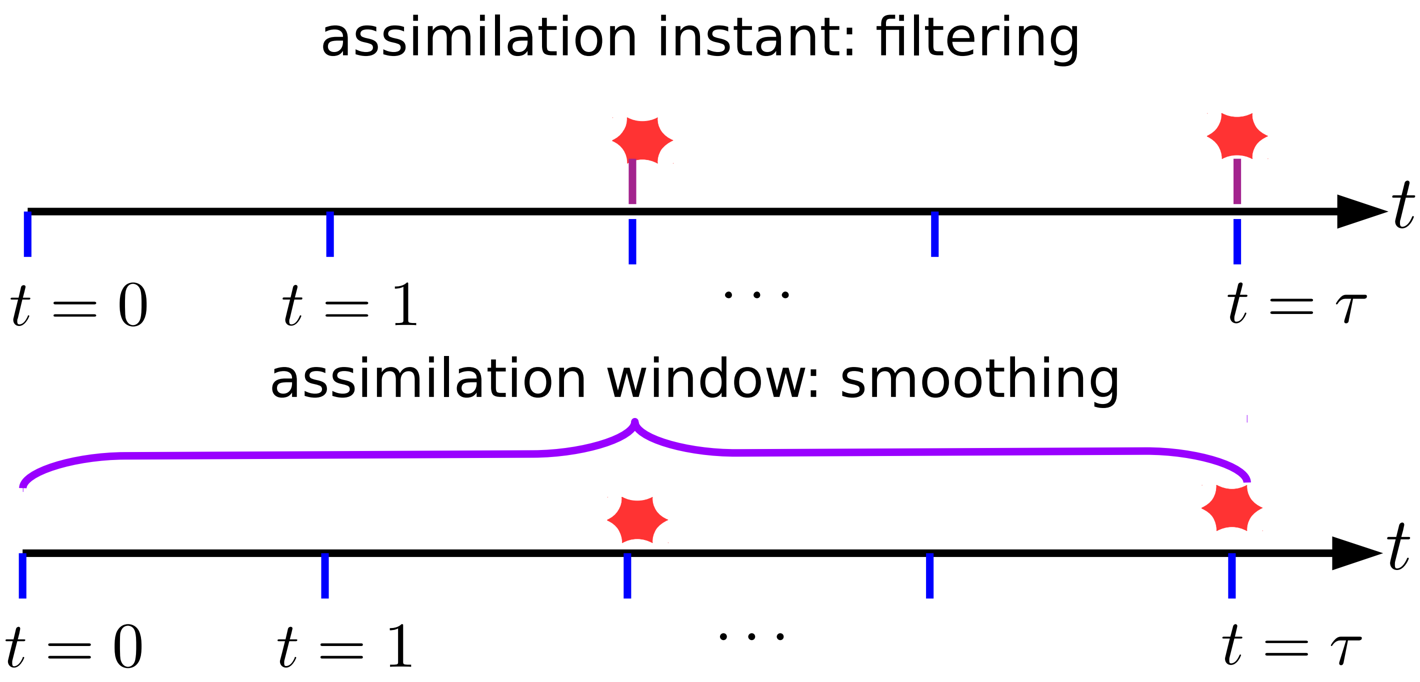
$\mathbf{x}^t \in \mathcal{R}^{N_x}$ **Model variables**

$\mathbf{y}^l \in \mathcal{R}^{N_y}$ **Observations**

$$\mathbf{x}^t = m^{(t-1)\to t}\left(\mathbf{x}^{t-1}\right) + \boldsymbol{\beta}^t$$

$$\mathbf{y}^l = h^l\left(\mathbf{x}^{t=l}\right) + \boldsymbol{\eta}^l$$

$$\mathbf{x}^0 \ r.v., \ \mathbf{x}^0 \perp \boldsymbol{\beta}^t \perp \boldsymbol{\eta}^l$$

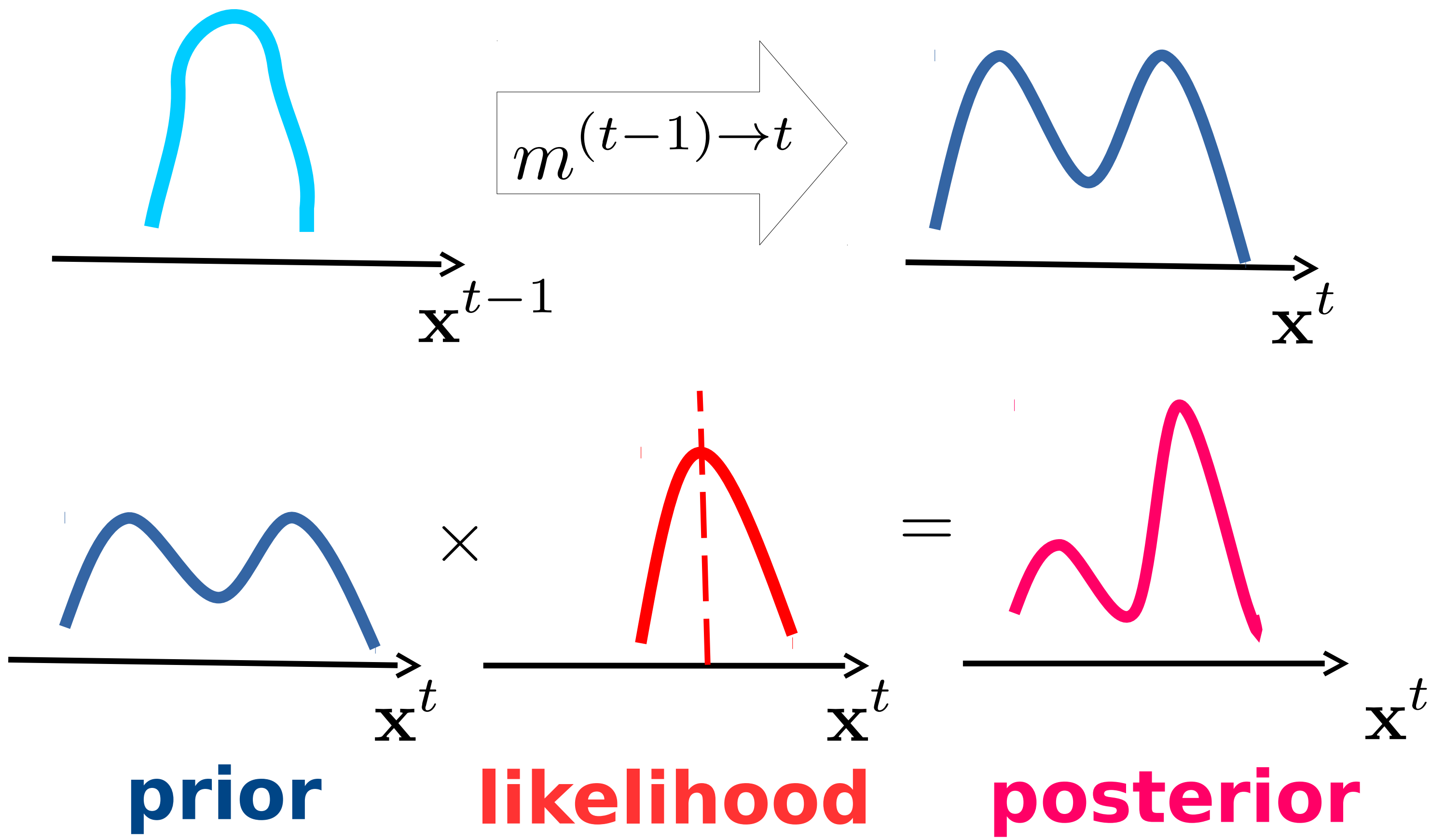


The ideal solution

Consider the following 1-step scenario:

Forecast:

Analysis:



The reality

Variational: Mode
Maximum a posteriori estimate

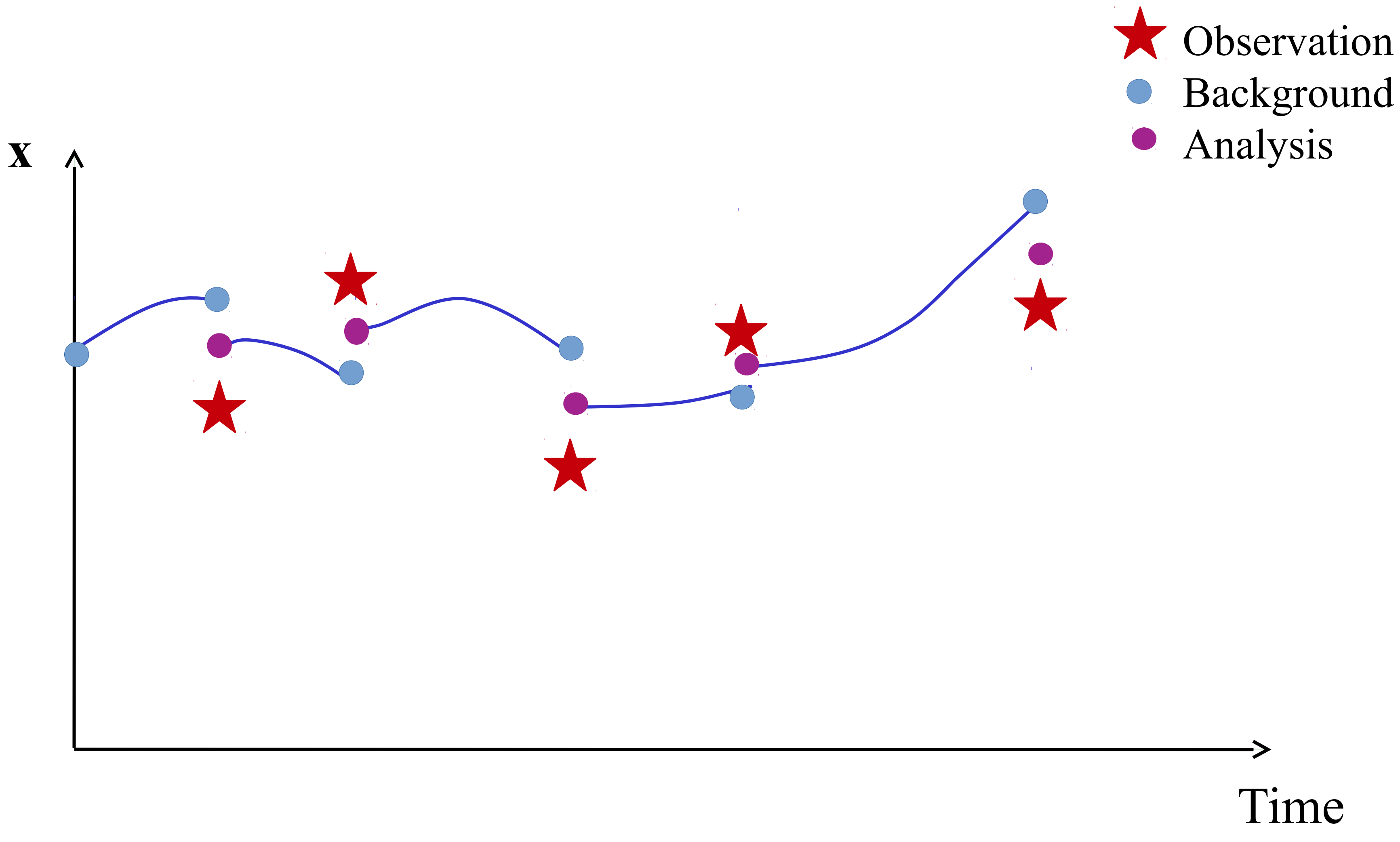
Kalman-based: (Ensemble) mean and covariance
Minimum variance estimation

Actually this is exact (and the same) when:

- Forecast model and observational operator are linear.
- Errors are Gaussian.

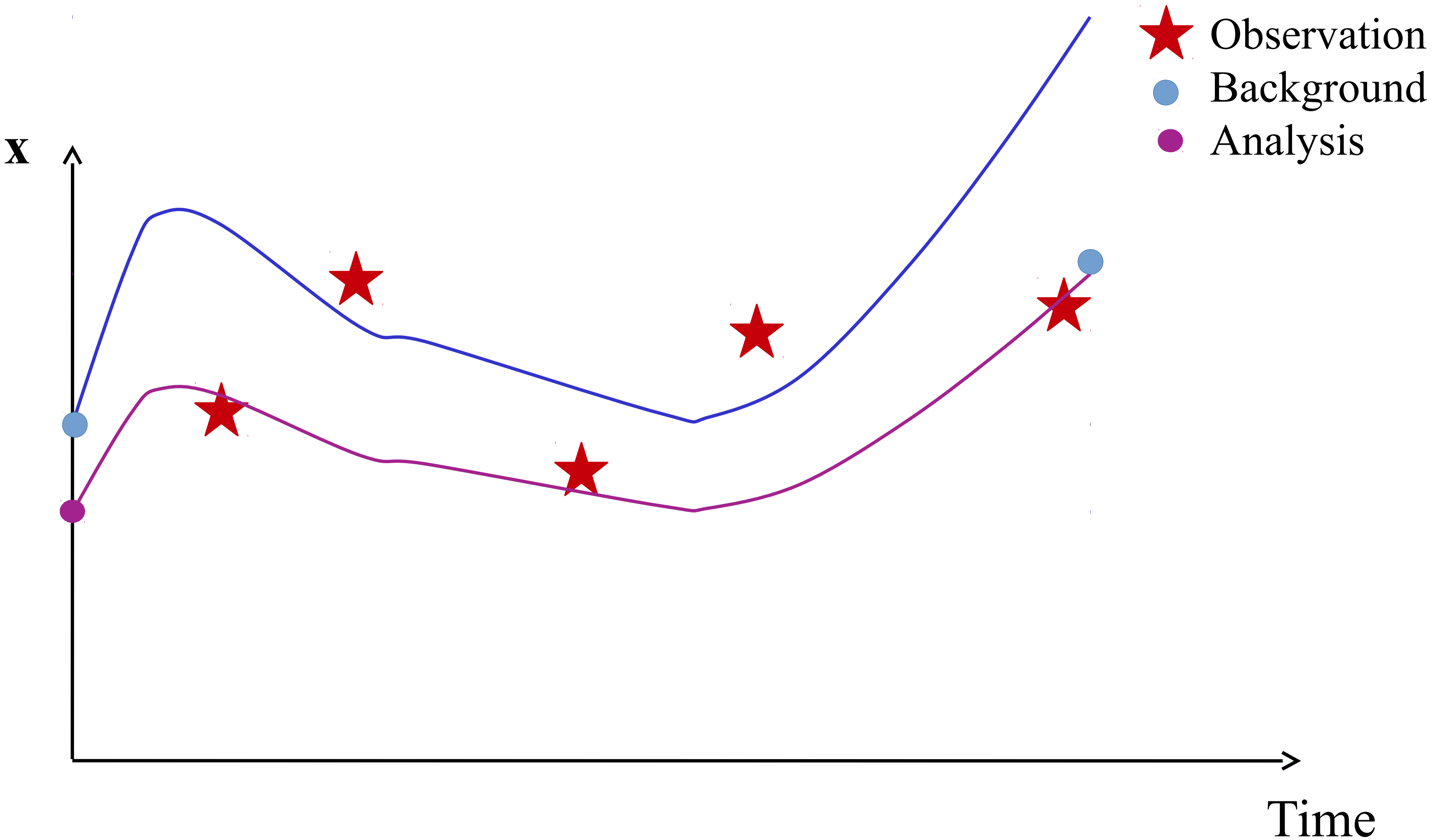
Filters

Assimilate every time observations are available.



Smoother

Assimilate observations over a time window.



Characteristics of traditional DA methods

	Method		Observations		Covariance	
	Variational	Kalman	Sequential	Smoother	Static	Dynamic
3DVar	✓		✓		✓	
4DVar	✓			✓	(✓)	✓
Optimal Interpolation		✓	✓		✓	
Kalman Filters		✓	✓			✓
Kalman Smoother		✓		✓		✓

Solution is got using (iterative) **minimisation** techniques.

Solution is got using explicit **linear algebra**.

Estimation is done for an **instant**.

Estimation is done within a **time window**.

Uncertainty is considered **fixed** in time.

Uncertainty evolves in time.

Why do we need hybrid methods?

The role of the covariance matrix. Filtering example

$$\mathbf{x} = \begin{bmatrix} u \\ v \end{bmatrix}$$

$$\mathbf{y}^o = \begin{bmatrix} u^o \end{bmatrix}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}$$

$$\mathbf{B} = \begin{bmatrix} \sigma^b_{uu} & \sigma^b_{uv} \\ \sigma^b_{vu} & \sigma^b_{vv} \end{bmatrix}$$

$$\mathbf{R} = r^2$$

Let’s write the explicit solution for this problem.

Analysis equations

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\mathbf{d}^{ob}$$

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

$$\mathbf{d}^{ob} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b$$ Innovation

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathbf{T}}\boldsymbol{\Gamma}^{-1}$$ Gain

$$\boldsymbol{\Gamma} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}} + \mathbf{R}$$ Total error covariance

Elements of the filtering solution

$$\mathbf{d}^{ob} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^b = \begin{bmatrix} u^o \end{bmatrix} - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} u^b \\ v^b \end{bmatrix} = \begin{bmatrix} u^o - u^b \end{bmatrix}$$

$$\mathbf{\Gamma} = \mathbf{HBH^T} + \mathbf{R} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma^b_{uu} & \sigma^b_{uv} \\ \sigma^b_{vu} & \sigma^b_{vv} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} r^2 \end{bmatrix} = \begin{bmatrix} \sigma^b_{uu} + r^2 \end{bmatrix}$$

$$\mathbf{K} = \mathbf{BH^T\Gamma}^{-1} = \begin{bmatrix} \sigma^b_{uu} & \sigma^b_{uv} \\ \sigma^b_{vu} & \sigma^b_{vv} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma^b_{uu} + r^2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\sigma^b_{uu}}{\sigma^b_{uu} + r^2} \\ \frac{\sigma^b_{vu}}{\sigma^b_{uu} + r^2} \end{bmatrix}$$

The solution

Mean

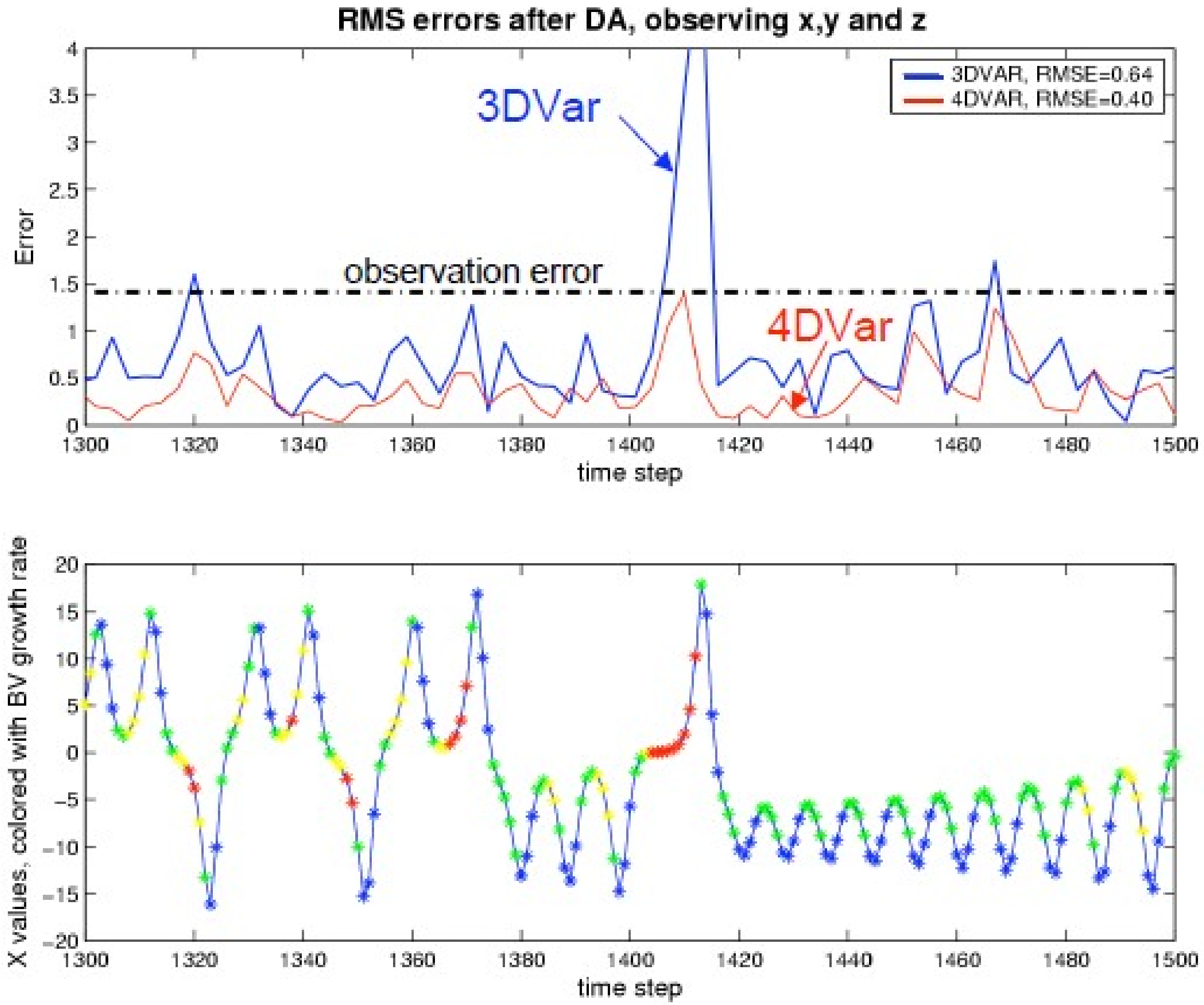
$$\begin{bmatrix} u^a \\ v^a \end{bmatrix} = \begin{bmatrix} u^b \\ v^b \end{bmatrix} + \begin{bmatrix} \frac{\sigma^b_{uu}}{\sigma^b_{uu}+r^2} \\ \frac{\sigma^b_{vu}}{\sigma^b_{uu}+r^2} \end{bmatrix} (u^o - u^b) = \begin{bmatrix} u^b + \frac{\sigma^b_{uu}}{\sigma^b_{uu}+r^2}(u^o - u^b) \\ v^b + \frac{\sigma^b_{vu}}{\sigma^b_{uu}+r^2}(u^o - u^b) \end{bmatrix}$$

Do the covariance as an exercise

3D vs 4DVar

4DVar has important information from the future (after all, it is a smoother), 3DVar does not.

The figure shows a comparison of the performance of the two methods. Taken from Evans et al, 2005.



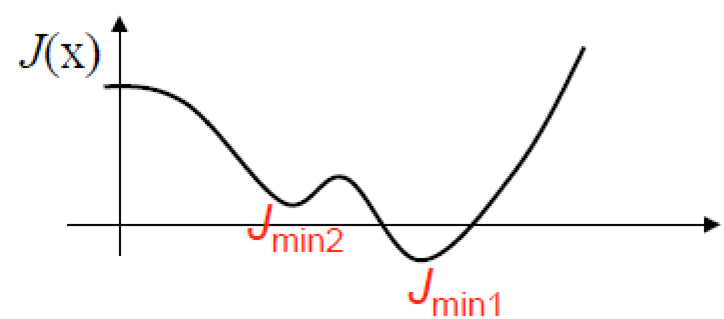
DA cycle and observations: $8\Delta t$, $\mathbf{R}=2*\mathbf{I}$
4D-Var assimilation window: $24\Delta t$

How long should the assimilation window be?

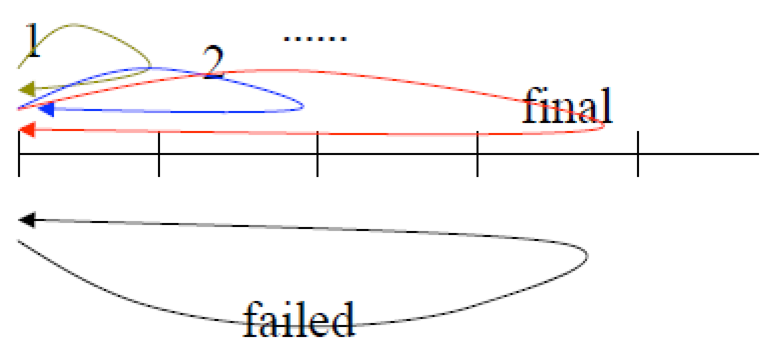
The longer the 4D assimilation window the more observations we'll have… but also the more nonlinear the forecast will be. The best should be somewhere in the middle.

	Win=8	16	24	32	40	**48**	56	64	72
Fixed window	0.59	0.59	0.47	0.43	0.62	**0.95**	0.96	0.91	0.98
Start with short window	0.59	0.51	0.47	0.43	0.42	**0.39**	0.44	0.38	0.43

Performance of 4DVar using the Lorenz 1963 and different lengths of assimilation window (Kalnay *et al.*, 2007).



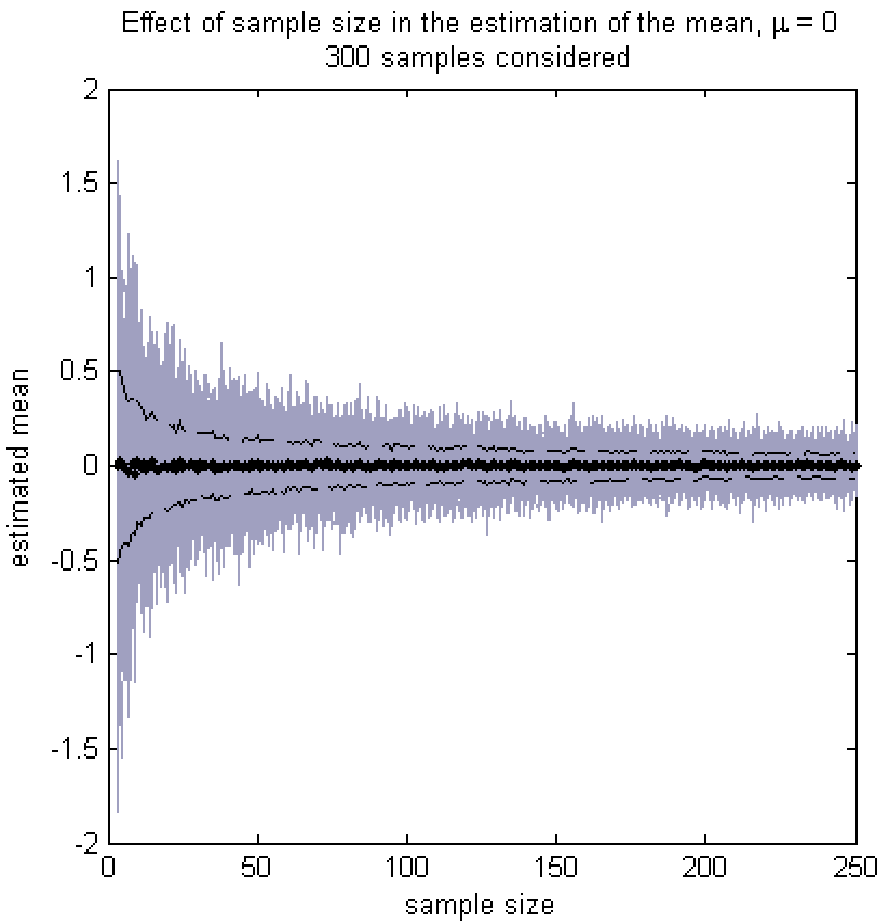
It is recommendable to do the minimization progressively while increasing the assimilation window (Pires *et al.*, 1996).

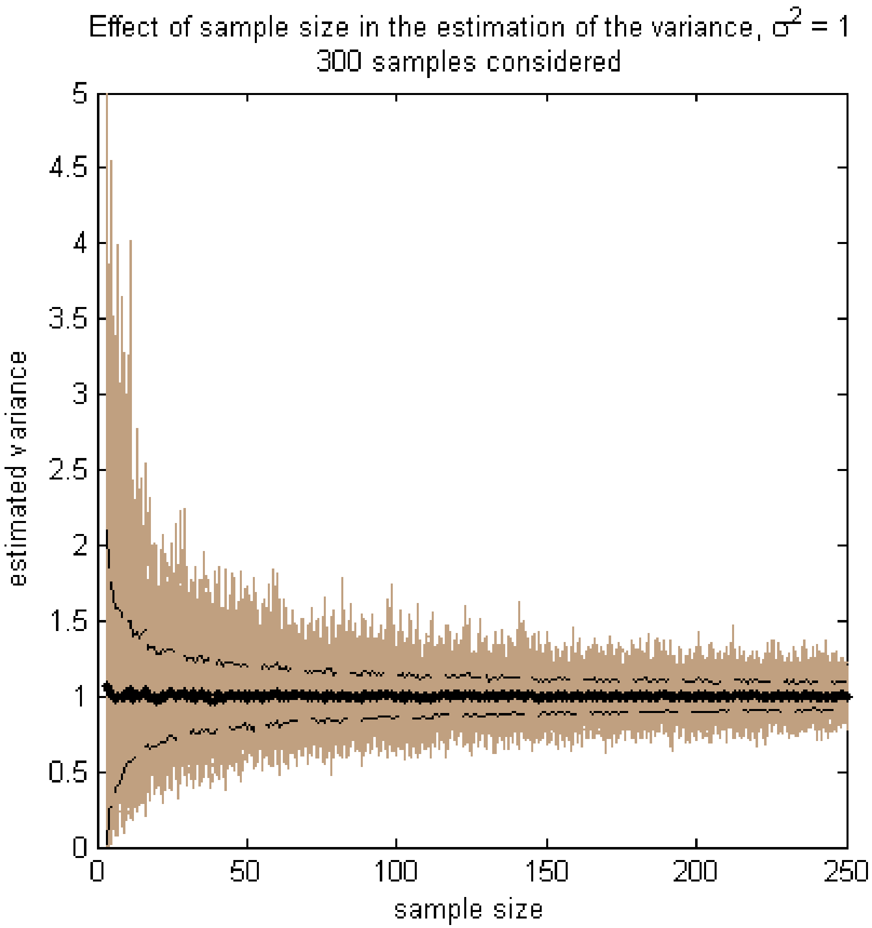


Sampling

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**.

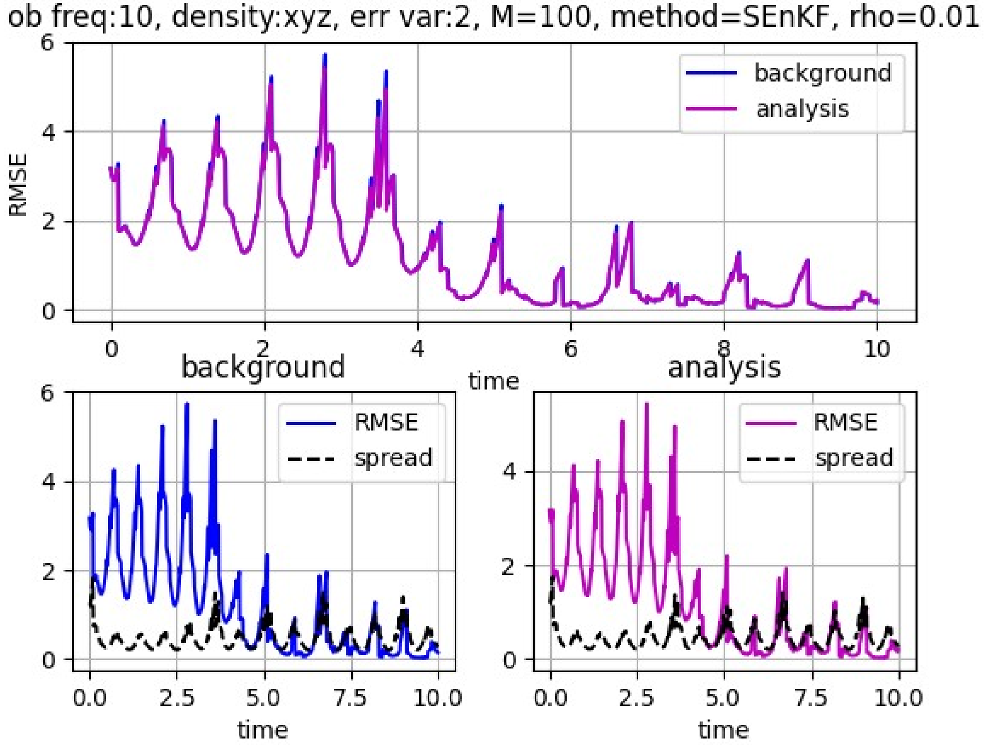
Fixes:

- Covariance inflation
- Covariance localization

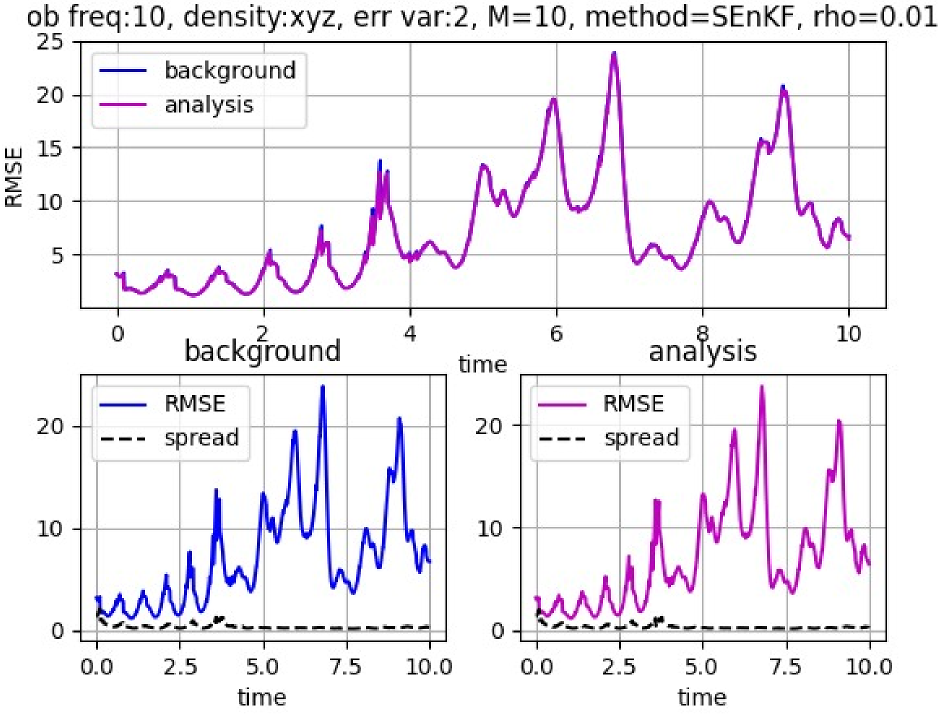
Also, the sample covariance matrix is singular for $N>M$…

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

Sampling



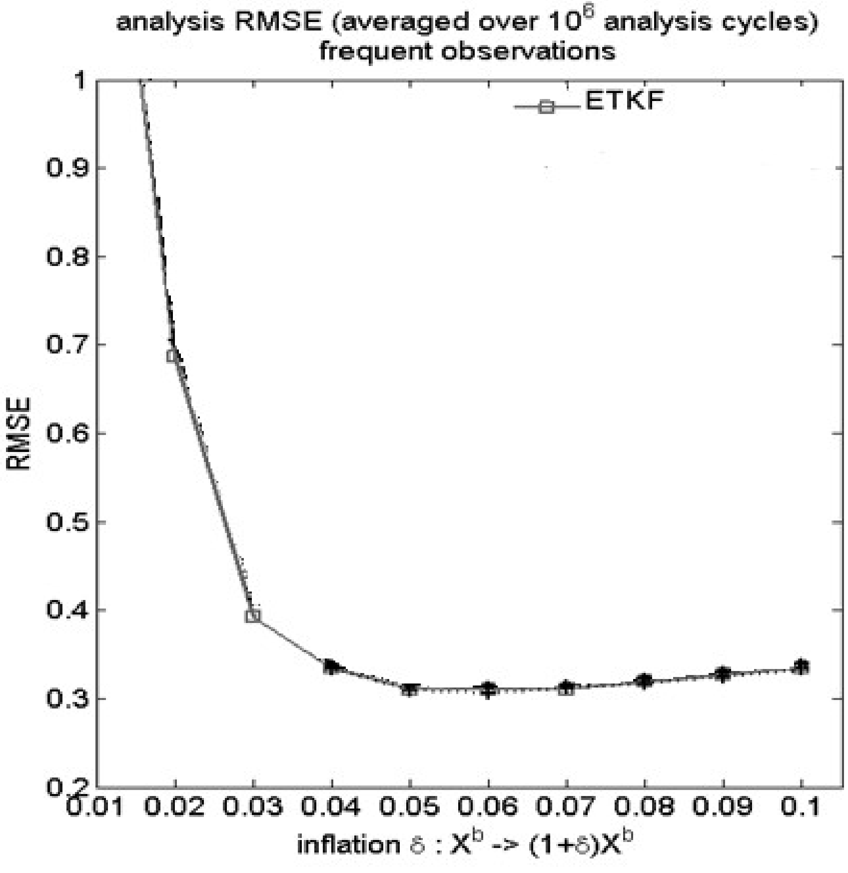
ob freq:10, density:xyz, err var:2, M=100, method=SEnKF, rho=0.01
RMSE
time
background
analysis
RMSE
spread

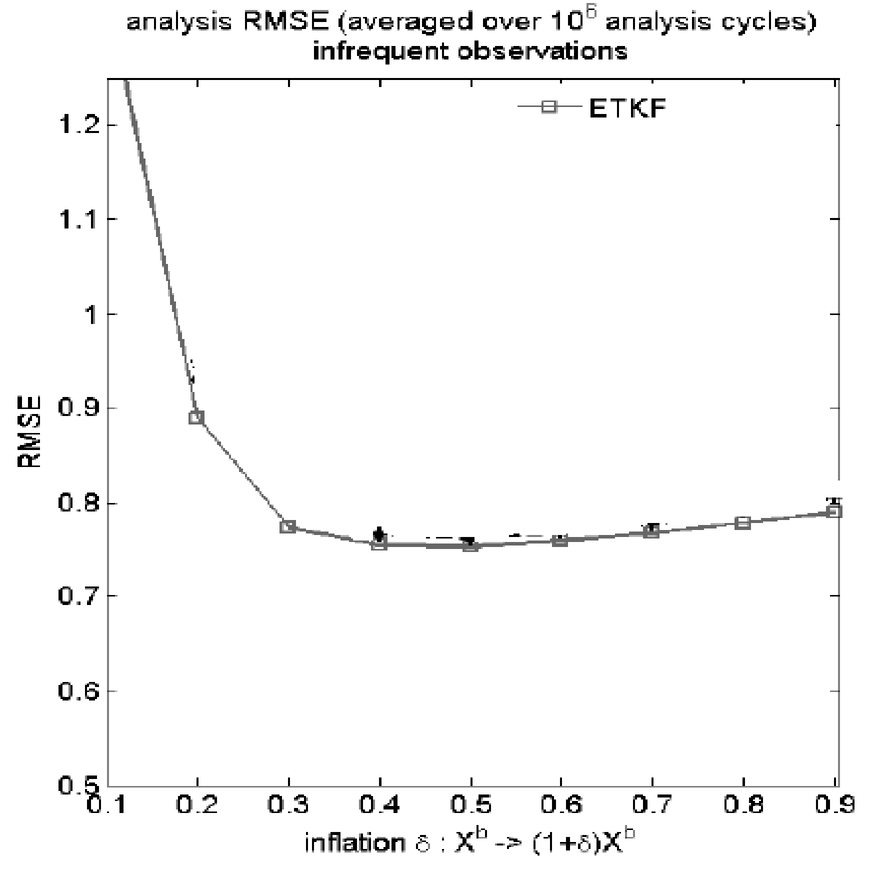


ob freq:10, density:xyz, err var:2, M=10, method=SEnKF, rho=0.01
RMSE
time
background
analysis
RMSE
spread

Covariance inflation and performance.

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





Amezcua et al, 2012.

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.

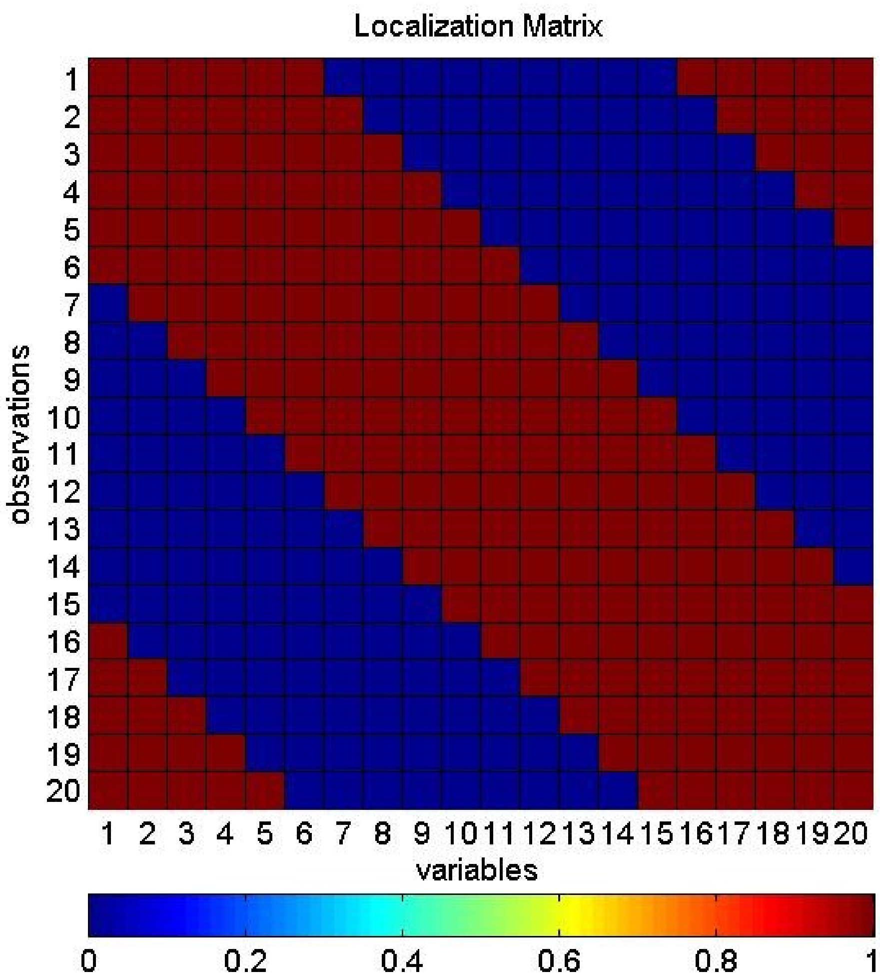
- This ensures that only the components of $\mathbf{P}^b$ believed to represent the corresponding components of $\mathbf{P}^b$ accurately are retained.

Localization

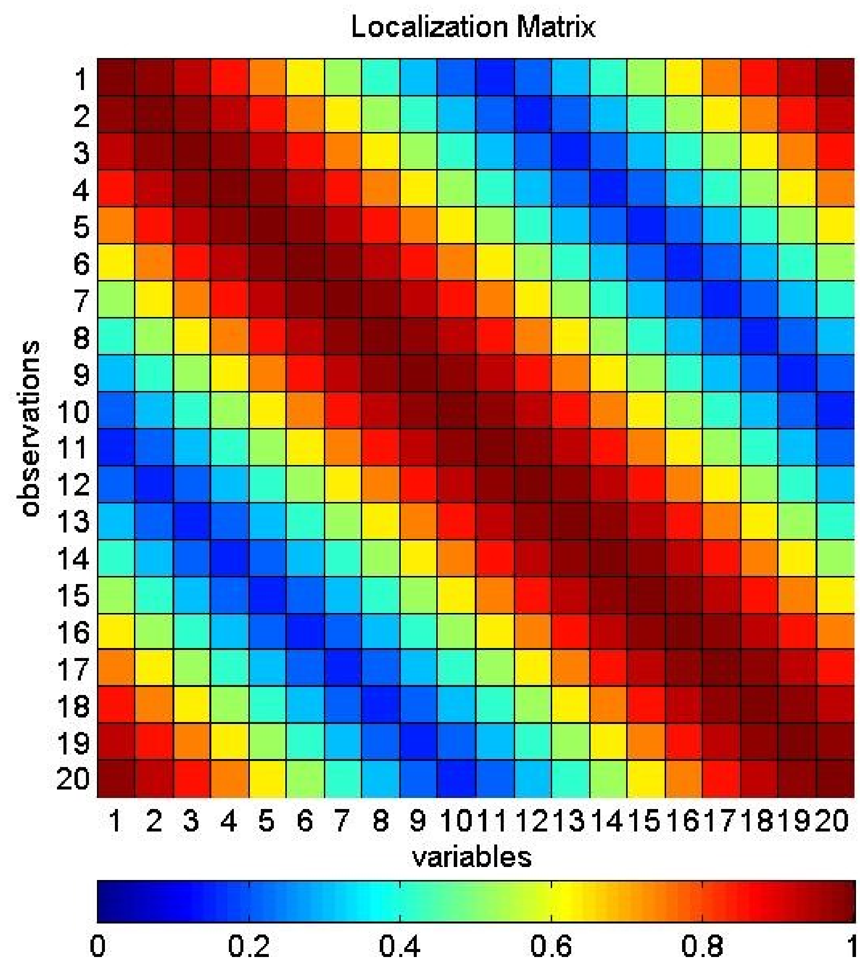
Example using Lorenz 1996

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

Cut-off



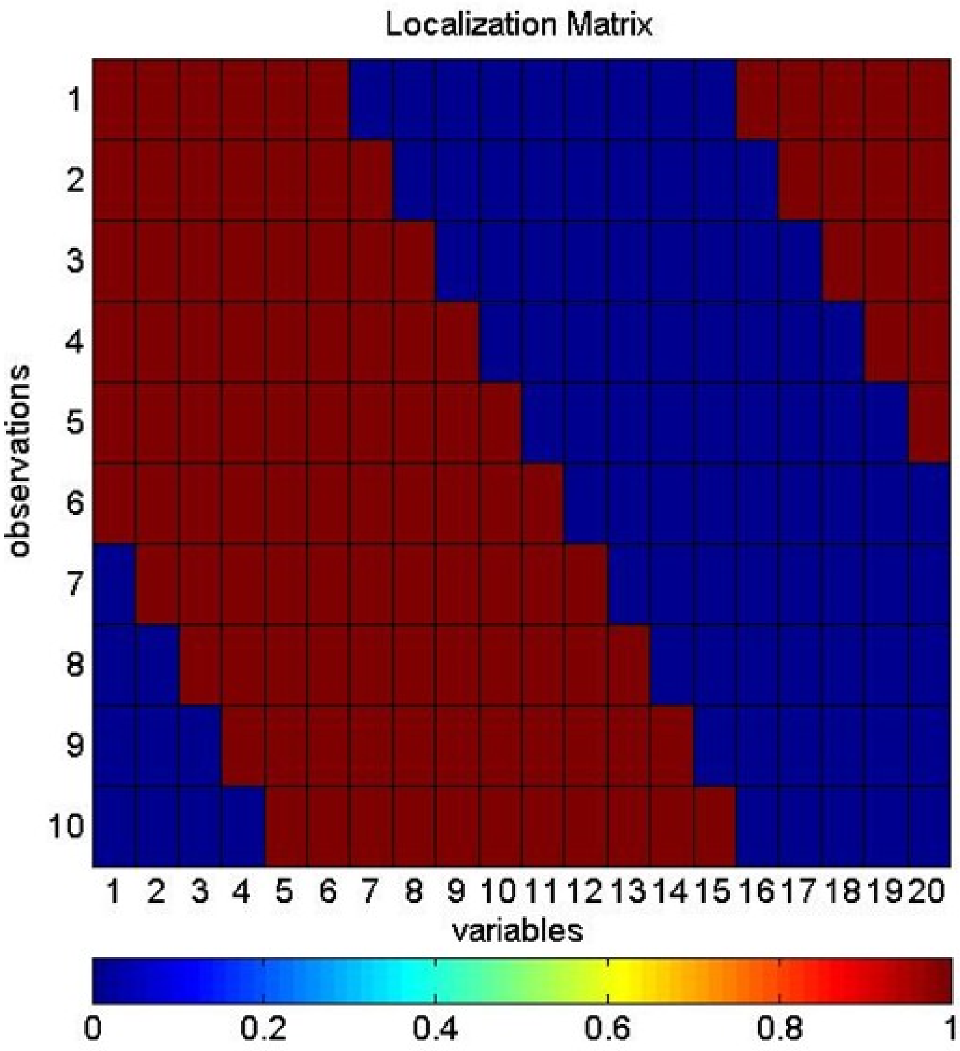
Gaspari-Cohn



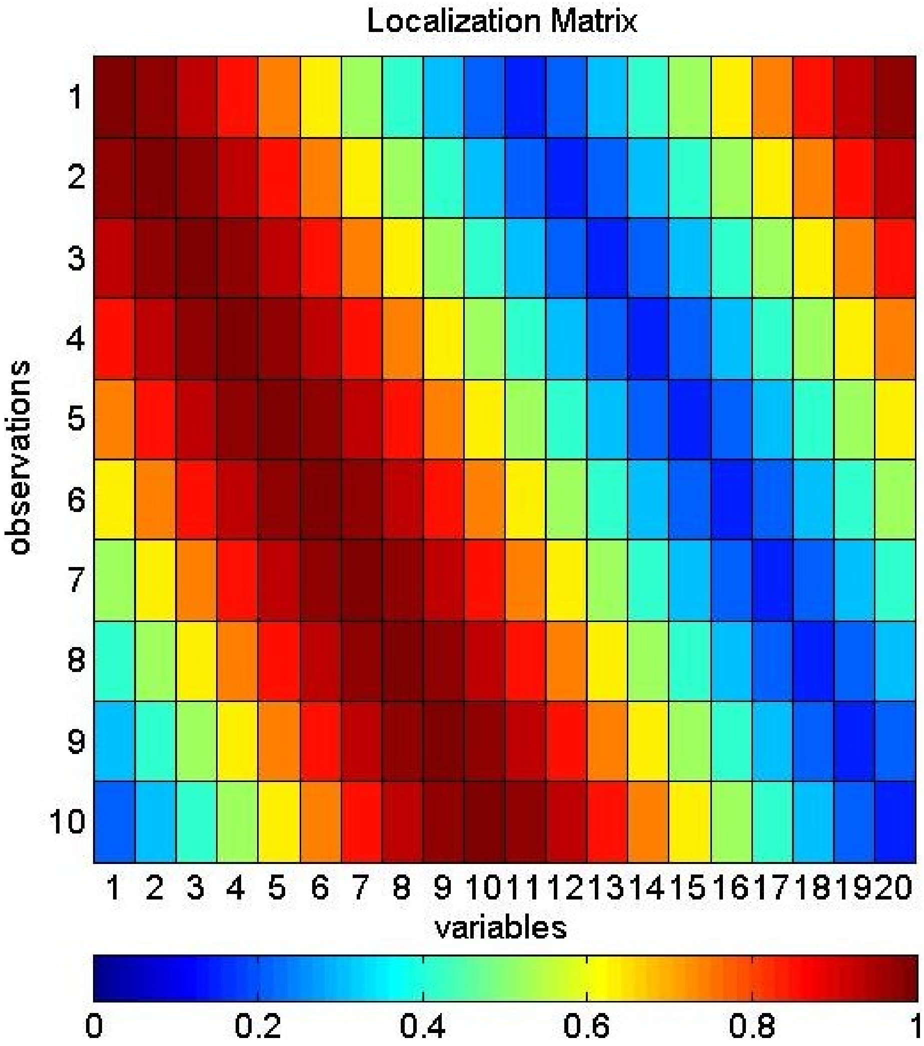
Localization $\mathbf{C} \circ \left(\mathbf{P}^b \mathbf{H}^T \right)$

Example using Lorenz 1996, observing every other variable.

Cut off

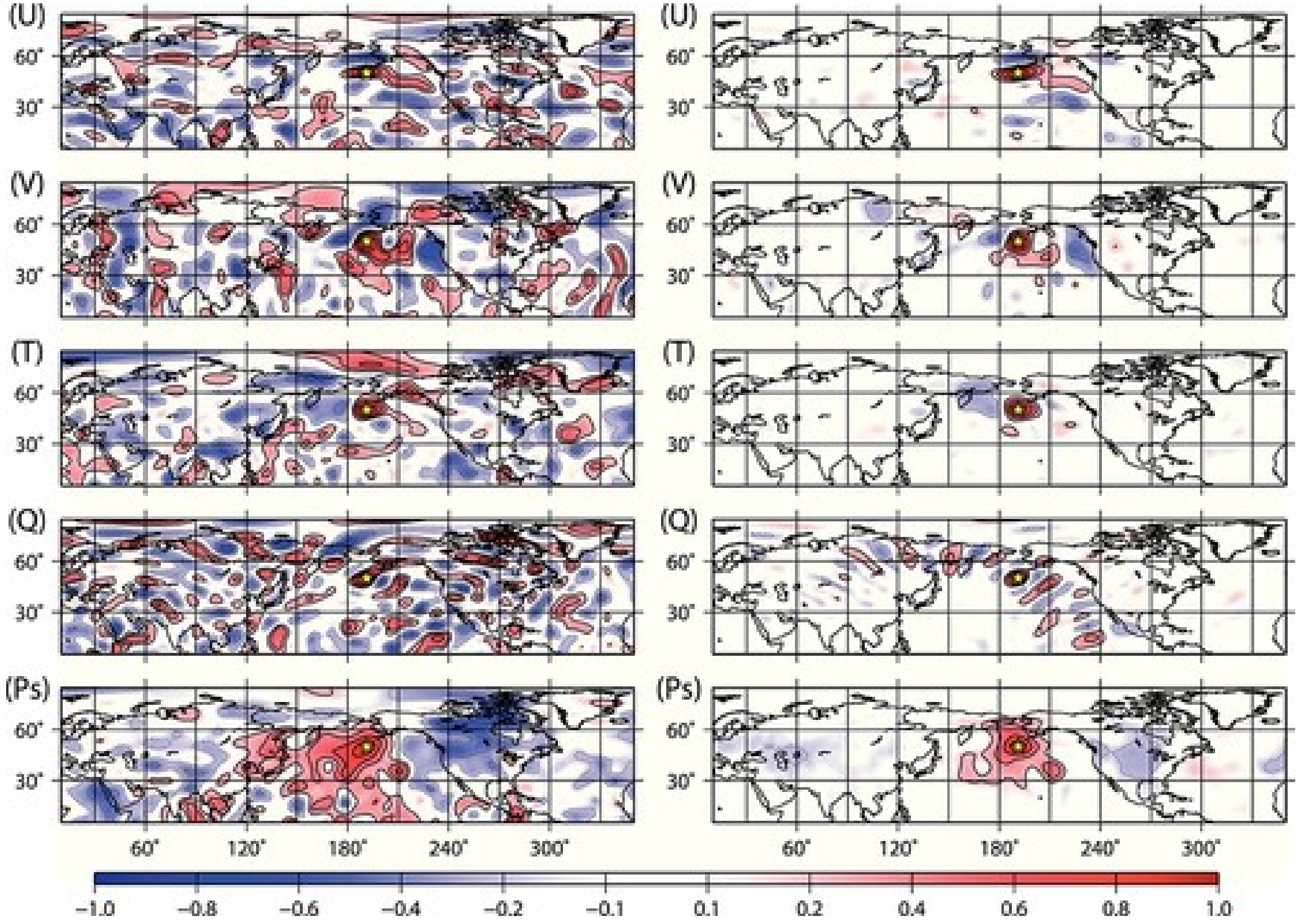


Gaspari-Cohn



Is the correlation real of an artifact?

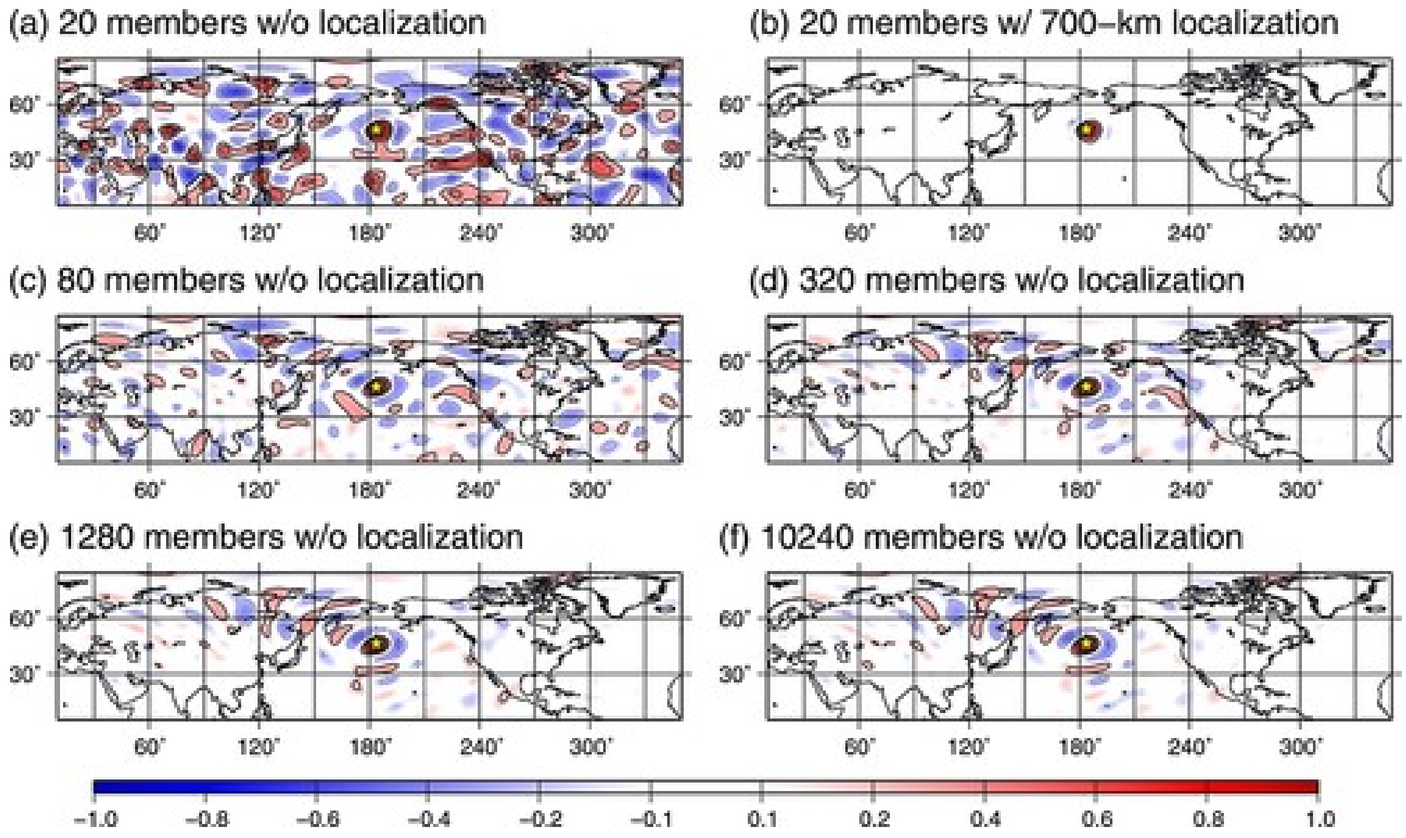
Miyoshi (2014)



Ensemble-based autocorrelations from the yellow star point (50.099 N, 168.75 W) at 00:00 UTC 17 January with (left column) 20 members and (right column) 10,240 members for (row 1) zonal wind component (U), (row 2) meridional wind component (V), (row 3) temperature (T), (row 4) specific humidity (Q), and (row 5) surface pressure (Ps). Except for Ps, the fourth model level (sigma = 0.51, or approximately 500 hPa level) is shown.

Is the correlation real of an artifact?

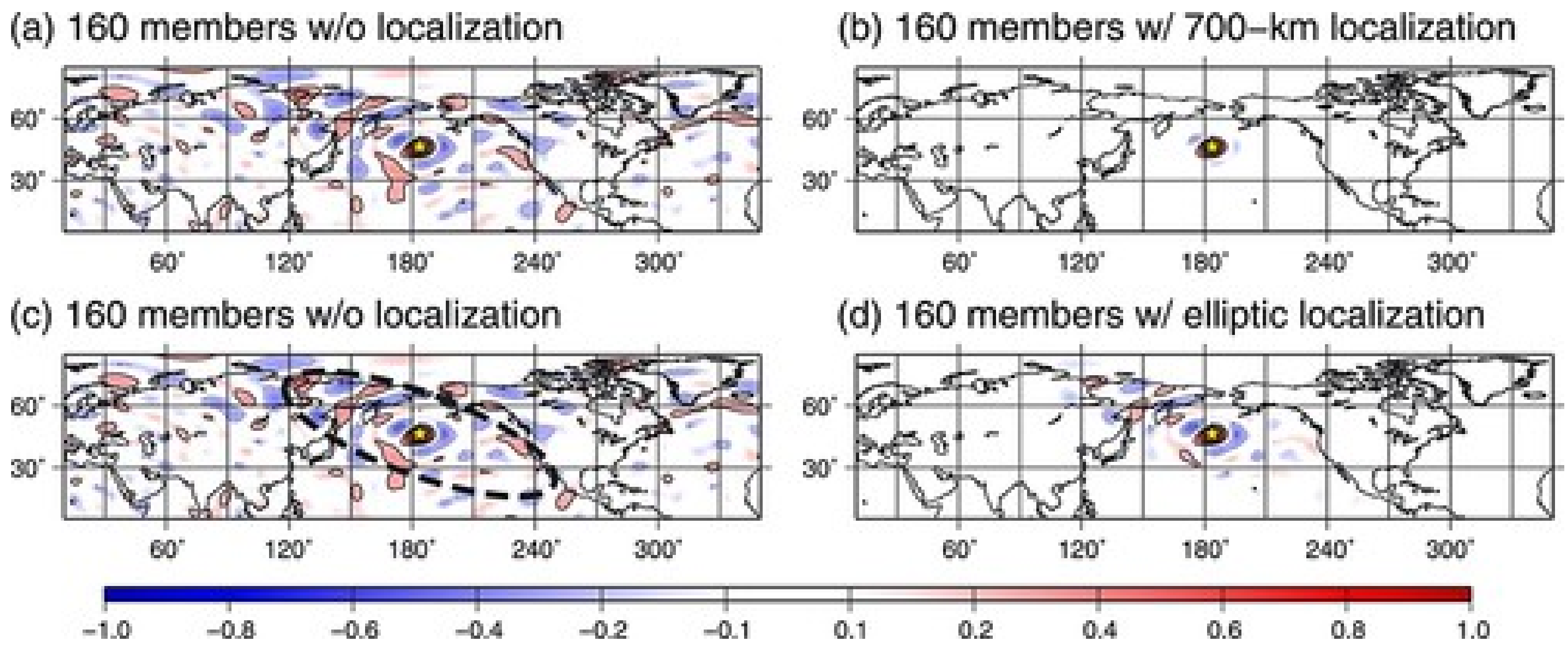
Miyoshi (2014)



Similar to Figure 1 but at 00:00 UTC 18 January with the yellow star point at 46.389°N, 176.25°W and for different ensemble sizes ((a) 20, (c) 80, (d) 320, (e) 1280, and (f) 10,240 members) and (b) with localization for 20 members.

Is the correlation real of an artifact?

Miyoshi (2014)



Similar to Figure 4 but for 160 members (a) without localization, (b) with 700 km circular localization, (c) flow-adaptive localization ellipse identified subjectively (black dashed), and (d) with elliptic localization based in Figure 5c.

Lesson: the dynamical situation may be important for localisation!

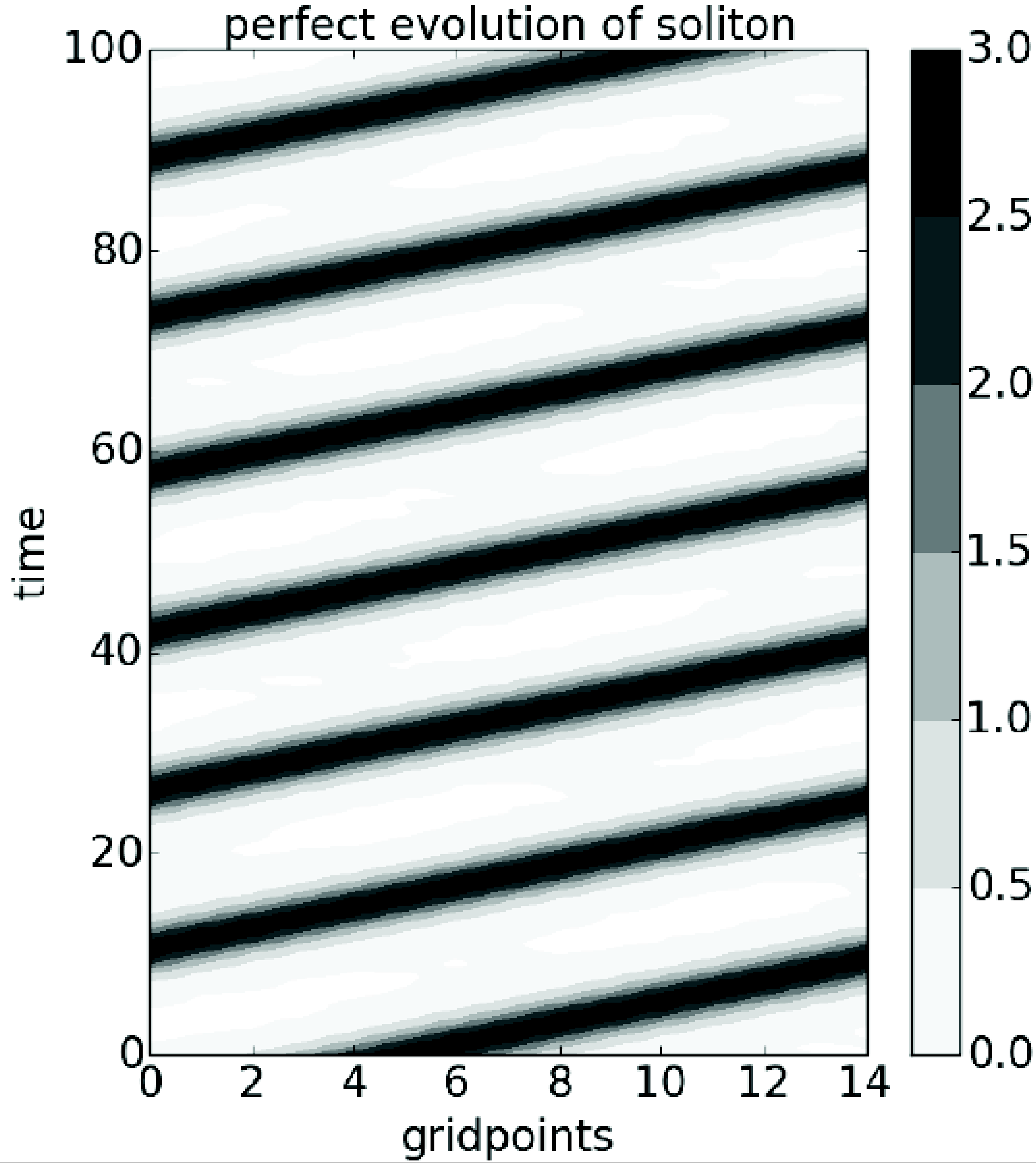
Example: Korteweg-DeVries model

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial s} + \frac{\partial^3 u}{\partial s^3} = 0$$

- Has soliton solutions, so solutions that don't change shape over time.
- We run it on a periodic domain and study how a covariance matrix is evolved by this system.

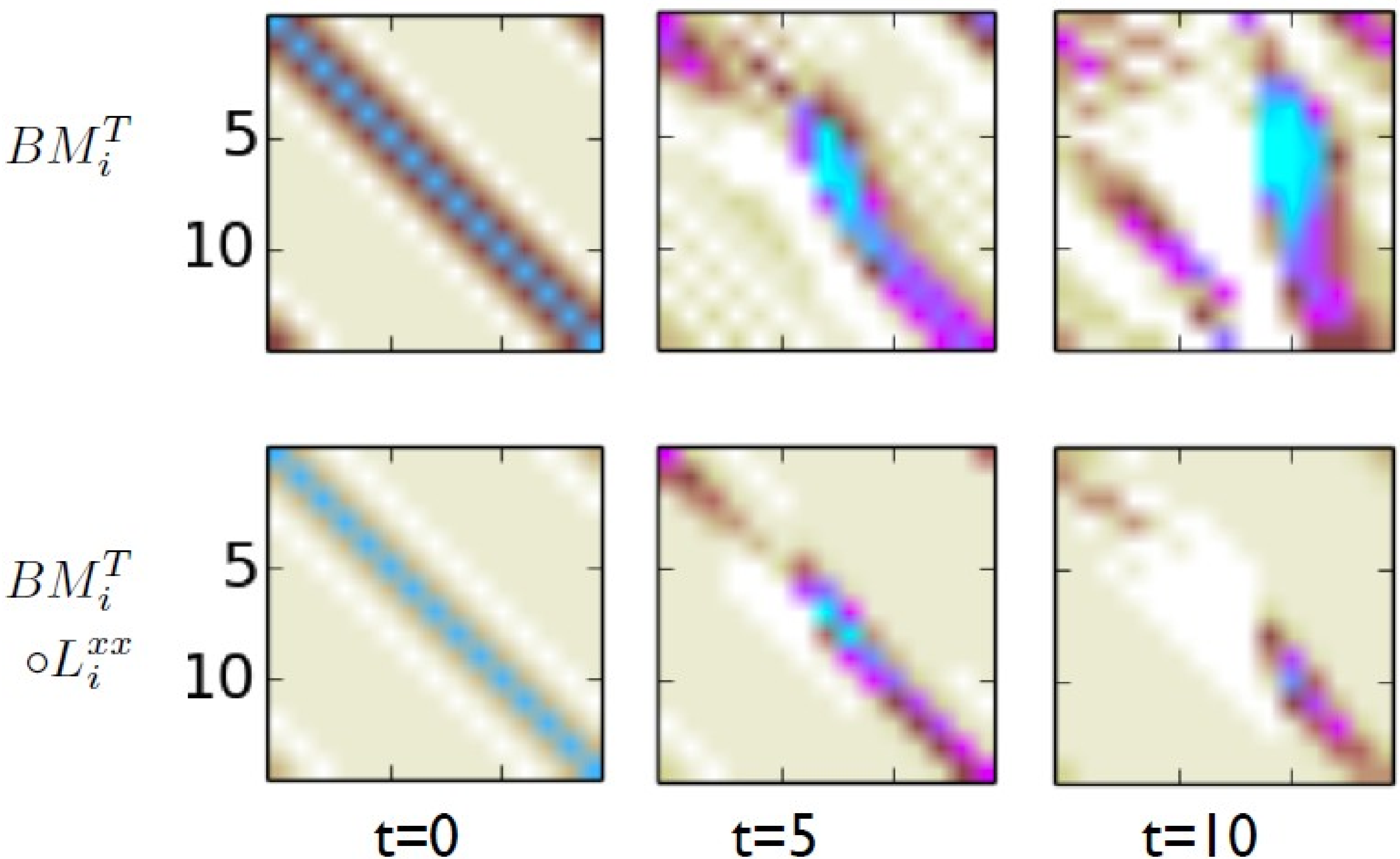
Amezcua et al, 2017.

Example propagation of soliton with KdV equation.



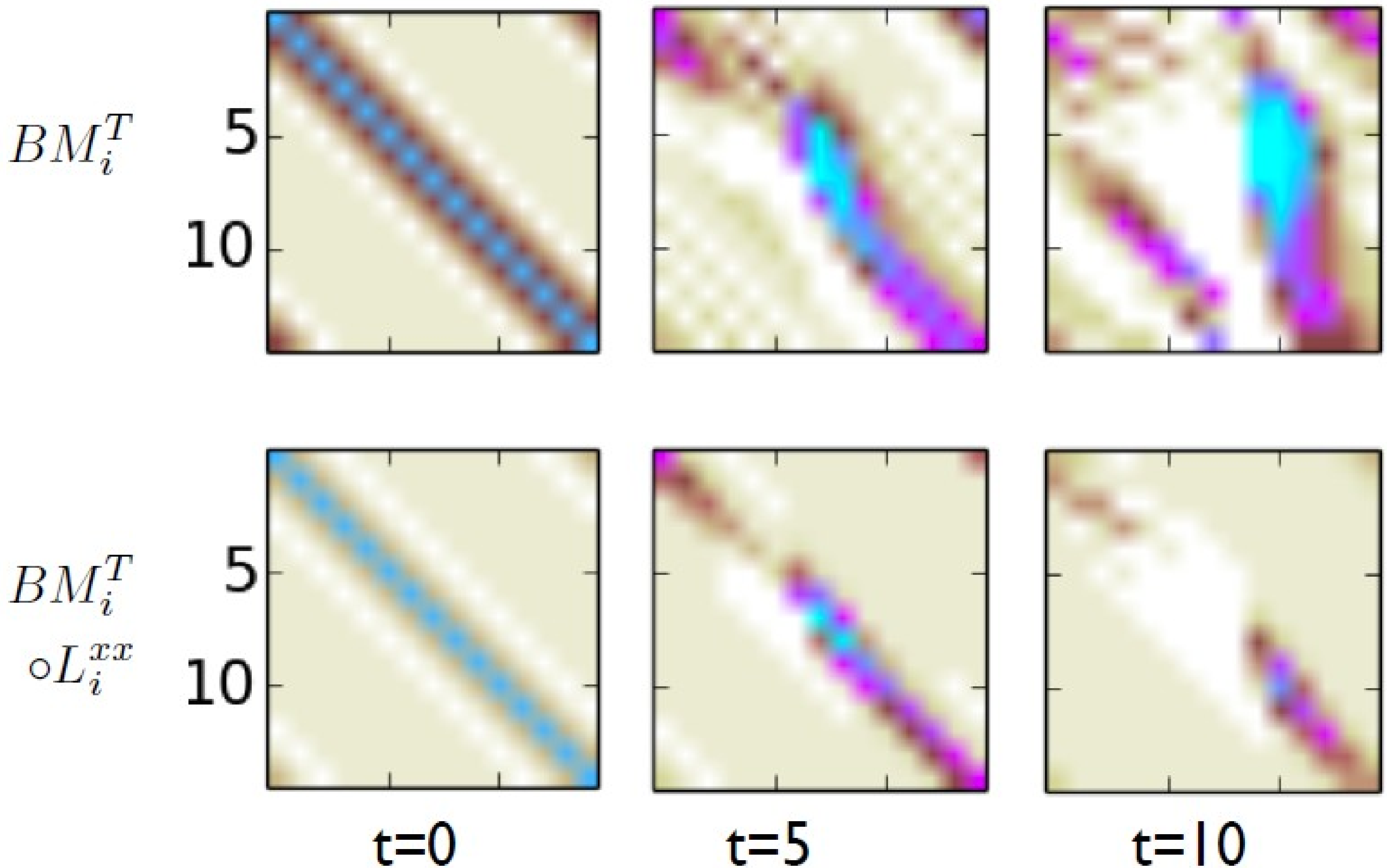
Amezcua et al, 2017.

Propagation of B with KdV model



Amezcua et al, 2017.

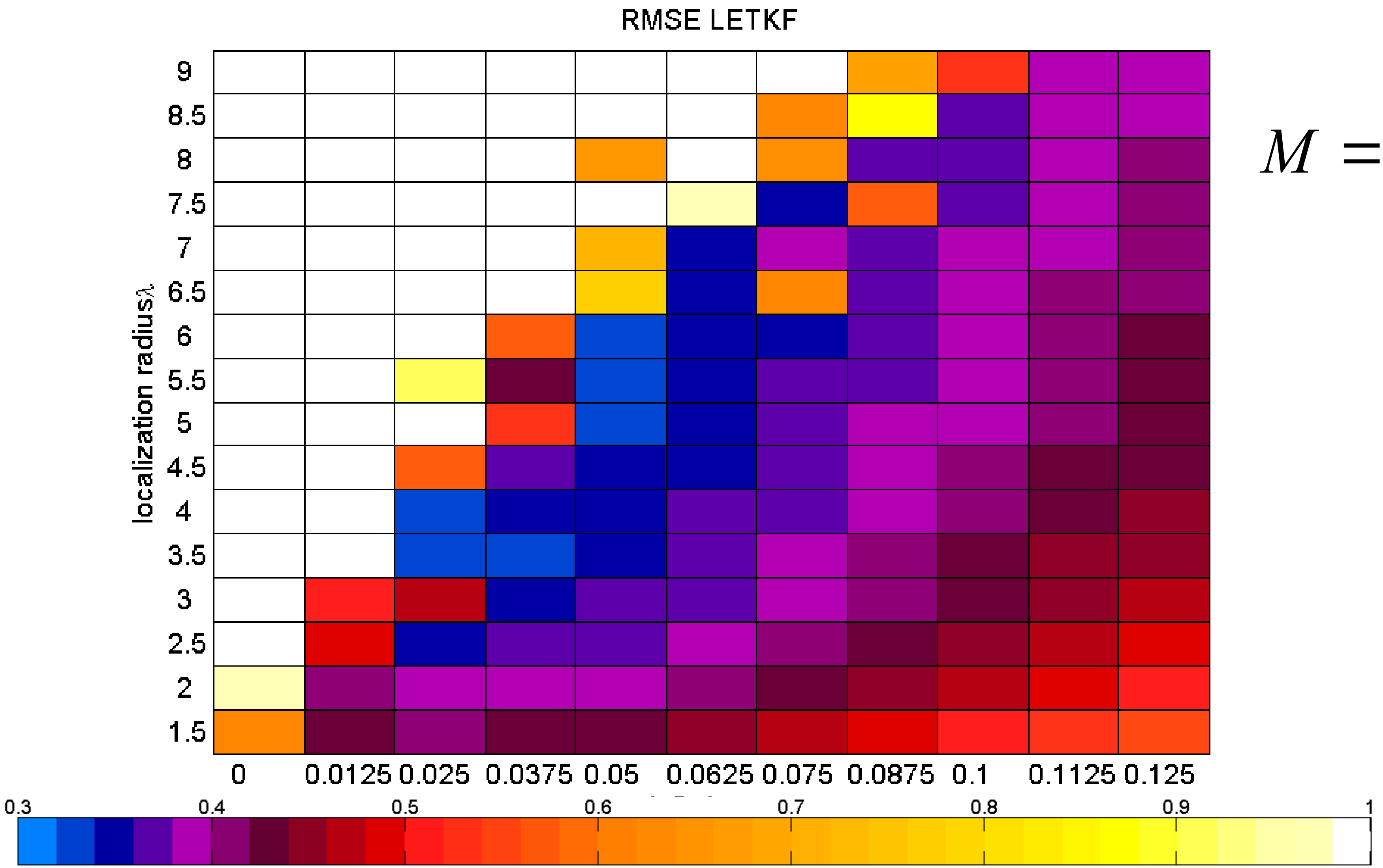
Propagation of B with KdV model



Amezcua et al, 2017.

Combined effects of inflation and localization

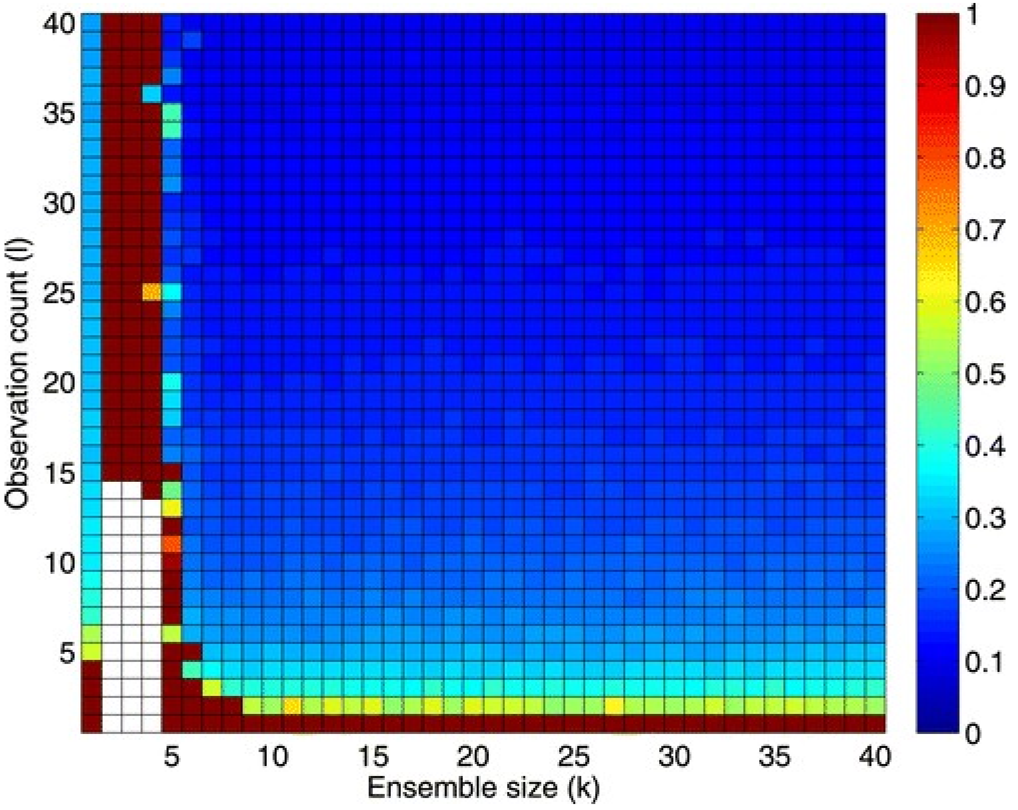
Experiments with Lorenz 1996 and 40 variables, observing every 2 time steps and every other variable.

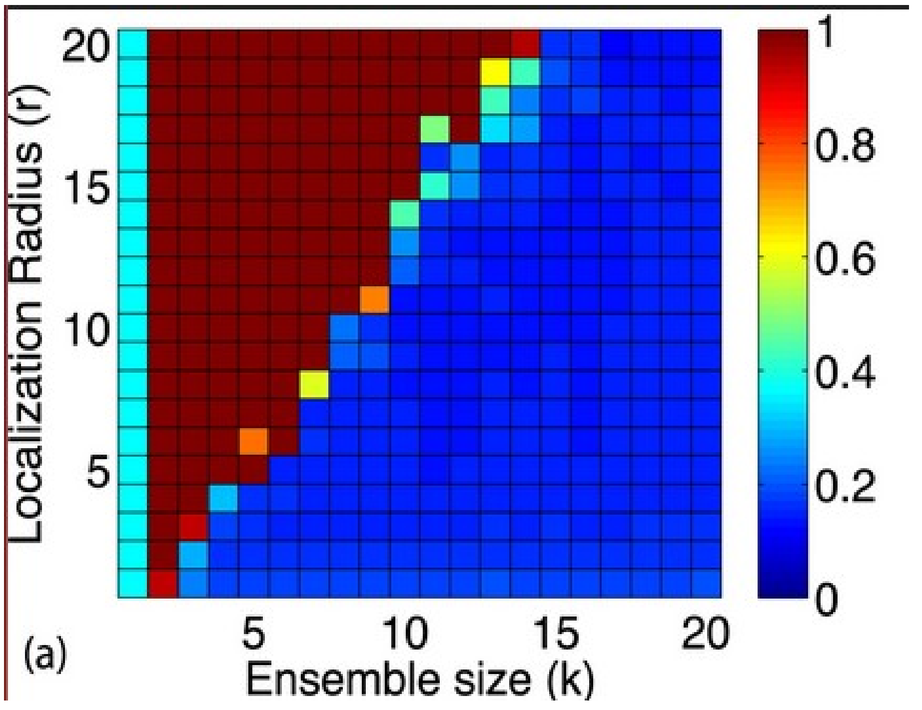


$M = 10$

Amezcua et al, 2012.

Interactions of different parameters in the EnKF





Penny, 2014

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

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

There are several ways to implement this.