MFO, Oberwolfach, Germany, October 6, 2016

Localization-induced filter instability and a simple adaptive localization method

Lars Nerger

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

with

Paul Kirchgessner, Angelika Bunse-Gerstner

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR-
UND MEERESFORSCHUNG

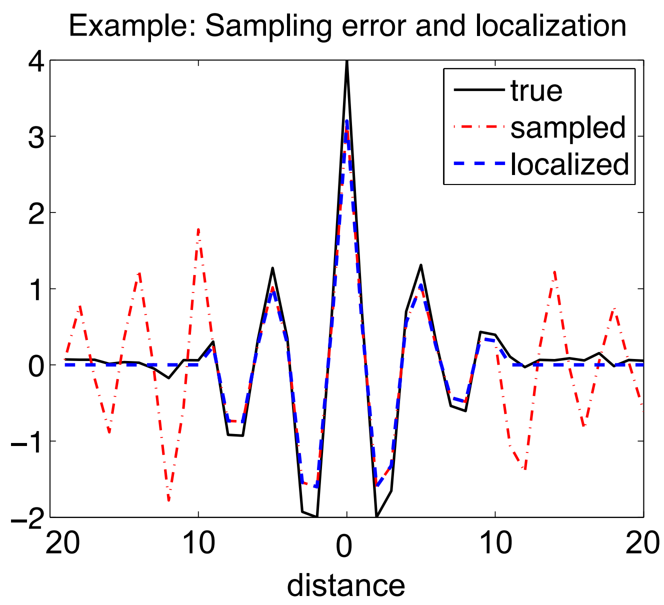
Outline

- Instability of serial observation processing filter in case of localization
- Adaptive localization following the degrees of freedom given by the ensemble

Localization

Localization: Why and how?

- Combination of observations and model state based on ensemble estimates of error covariance matrices
- Finite ensemble size leads to significant sampling errors
  - errors in variance estimates
    - usually too small
  - errors in correlation estimates
    - wrong size if correlation exists
    - spurious correlations when true correlation is zero



- Assume that long-distance correlations in reality are small
  - damp or remove estimated long-range correlations

Covariance localization

- Applied to forecast covariance matrix
- Element-wise product with correlation matrix C of compact support to reduce covariances

$$\mathbf{K}_{loc} = \left(\mathbf{C} \circ \mathbf{P}^f\right) \mathbf{H}^T \left(\mathbf{H} \left(\mathbf{C} \circ \mathbf{P}^f\right) \mathbf{H}^T + \mathbf{R}\right)^{-1}$$

- Only possible if forecast covariance matrix is computed (not in ETKF or SEIK)

E.g.: Houtekamer & Mitchell (1998, 2001), Whitaker & Hamill (2002)

Domain & Observation localization

Domain localization (local analysis)

- Perform local filter analysis with observations from surrounding domain

Observation localization (Hunt et al. 2007)

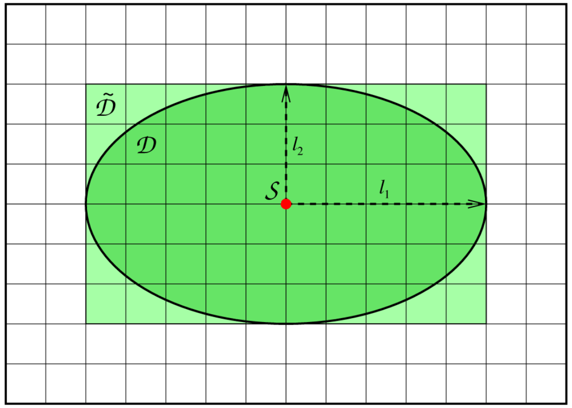
- Use non-unit weight for observations
- reduce weight for remote observations by increasing variance estimate

$$\mathbf{R}_\sigma^{-1} = \tilde{\mathbf{C}}_\sigma \circ \mathbf{R}^{-1}$$

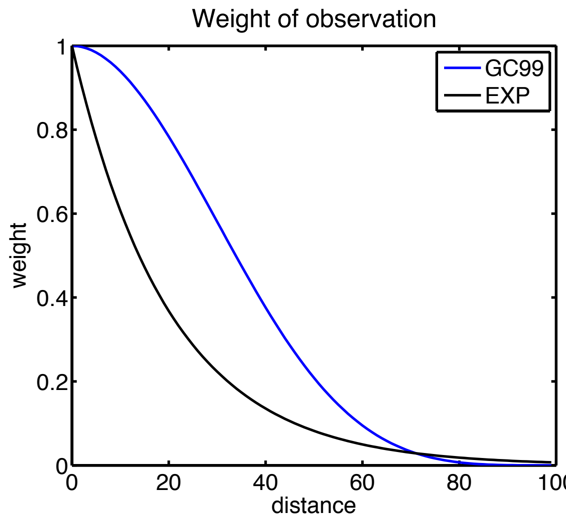
- Localization effect similar to covariance localization
- equivalence to *covariance localization* only shown for single observation (Nerger et al. QJRMS, 2012)

E.g.: Brankart et al. (2003), Evensen (2003), Ott et al. (2004), Hunt et al. (2007)

Domain Localization



S: Analysis region
D: Corresponding data region



Instability of serial observation processing filters in case of localization

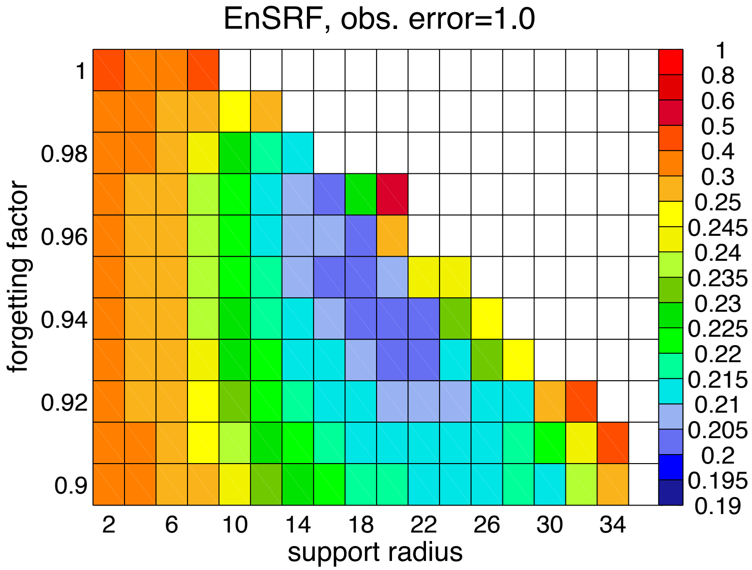
(EnSRF, EAKF)

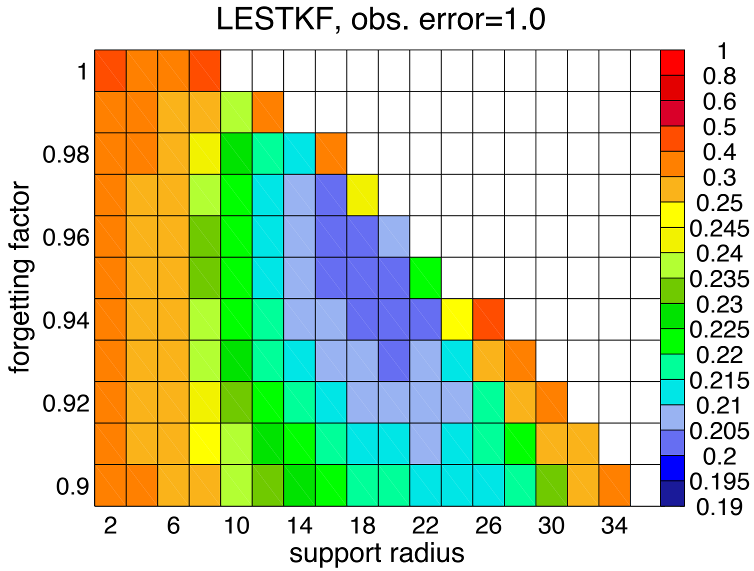
Serial observation processing

Serial observation processing	**Synchronous assimilation**
EnSRF, EAKF	***ETKF, SEIK, ESTKF, (EnKF)***
• Perform a loop assimilating each single observation • Efficient: Avoids matrix-matrix operations • Requires diagonal observation error covar. matrix	• Assimilation all observation at a given time at once • Usually using ensemble-space transformations • Possible for arbitrary observation error covar. matrices
Use **covariance localization**	Use **observation localization**

(EnSRF: Whitaker & Hamill, 2002; EAKF: Anderson, 2001)

Test with Lorenz9[568] Model



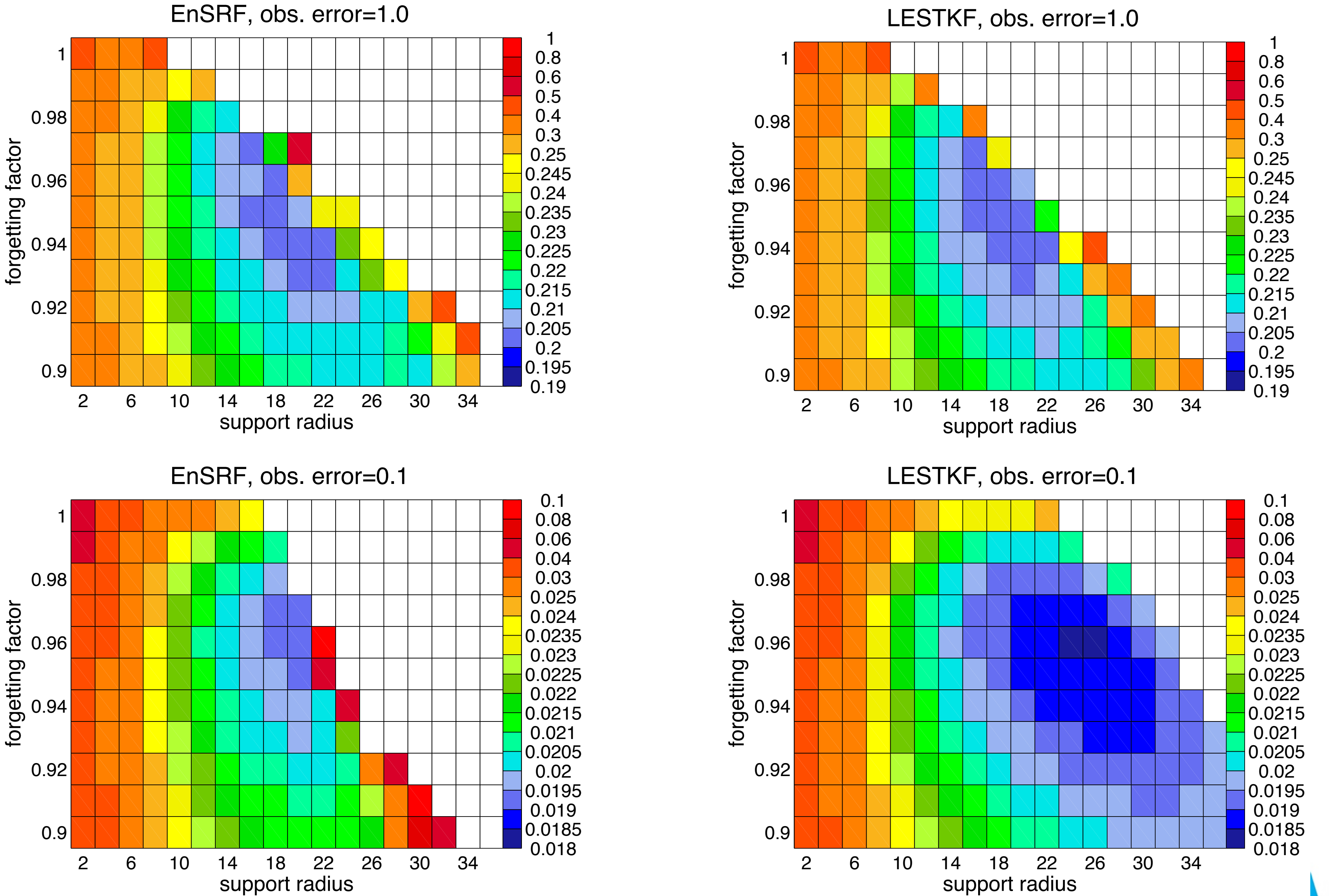


EnSRF (Whitaker & Hamill 2002)

For obs. error=1.0:

- EnSRF and LESTKF almost identical

Test with Lorenz9[568] Model

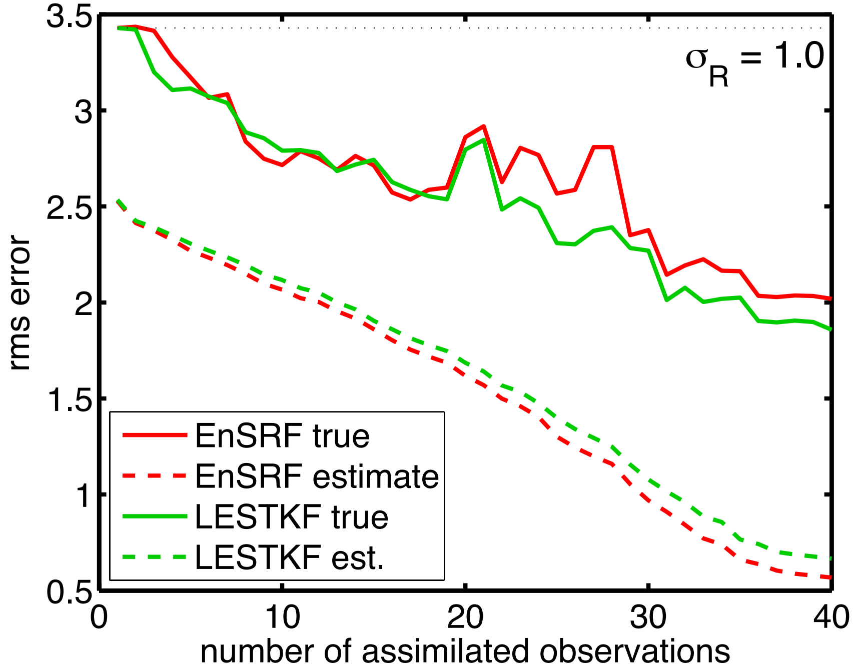


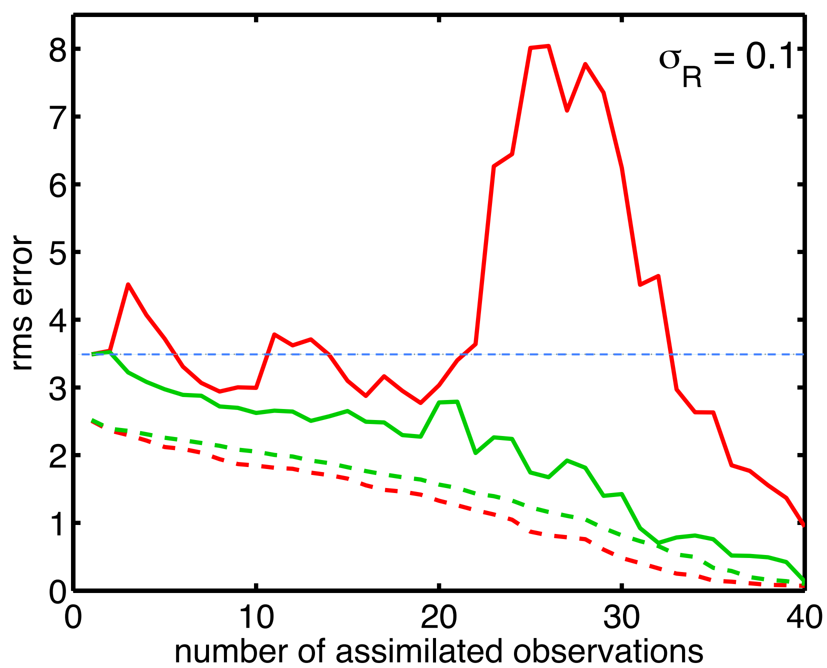
RMS error over number of observations

How does the RMS error develop during the loop over all observations?

At first analysis step:

- EnSRF: Compute RMS errors at each iteration
- LESTKF: Do 40 experiments with increasing number of obs.

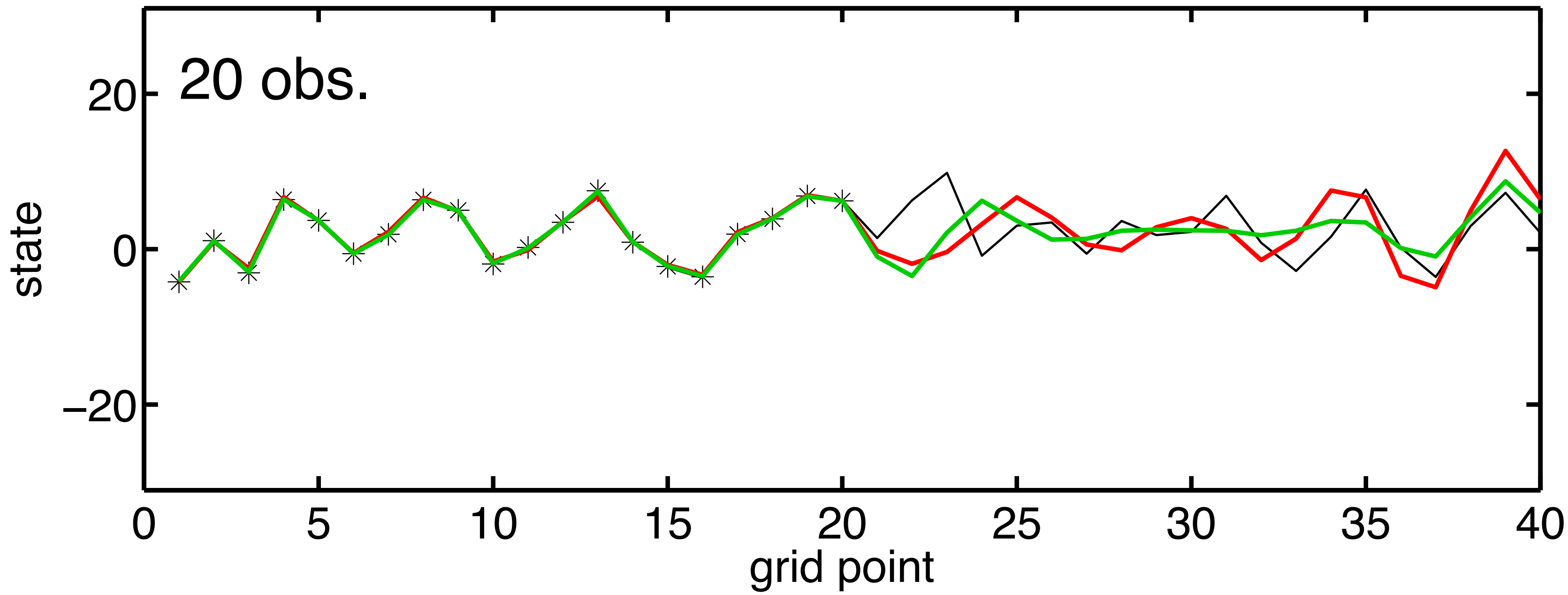




Instability of serial obs. processing with localization

More detailed view:

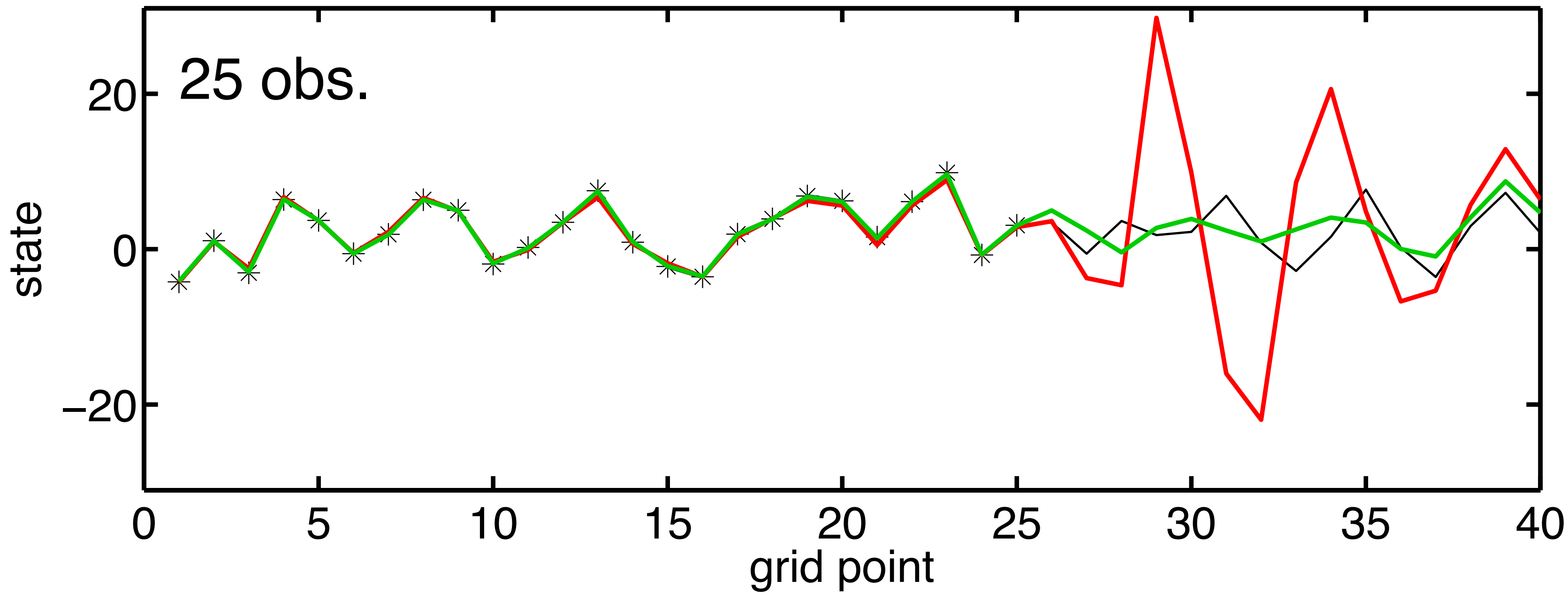
- State estimate for different numbers of observations



Instability of serial obs. processing with localization

More detailed view:

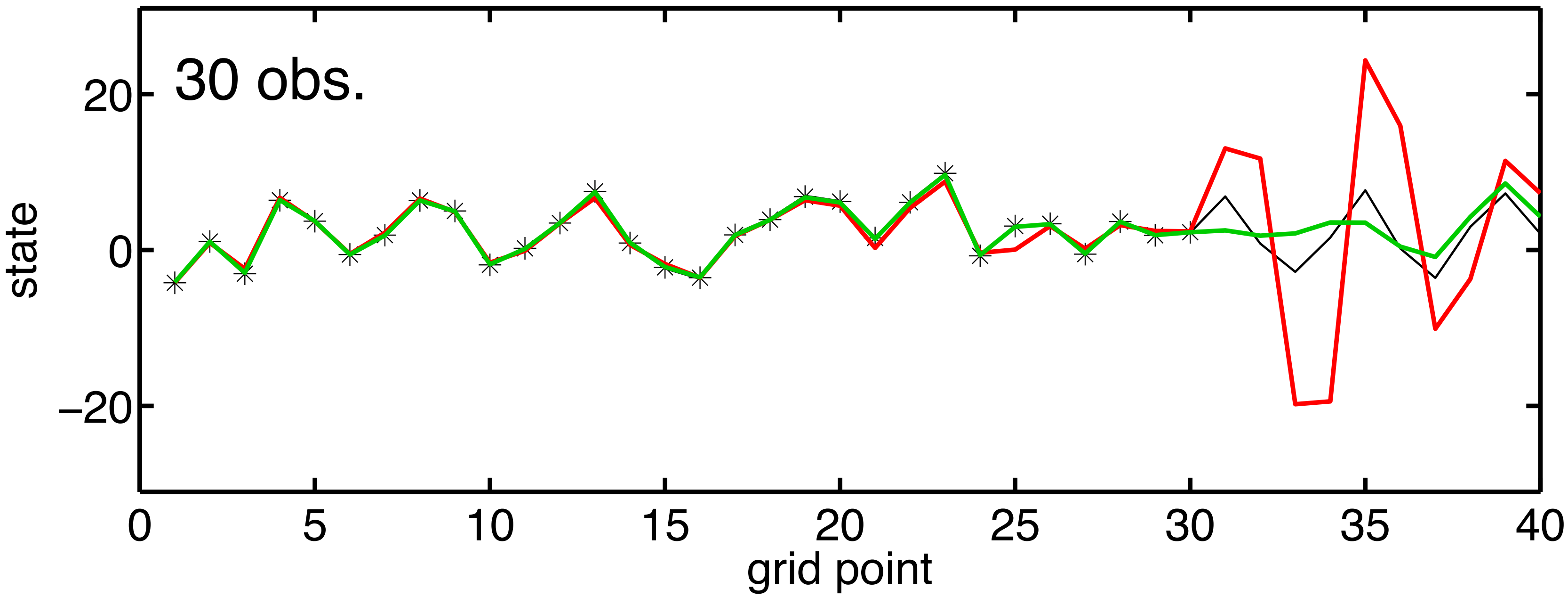
- State estimate for different numbers of observations



Instability of serial obs. processing with localization

More detailed view:

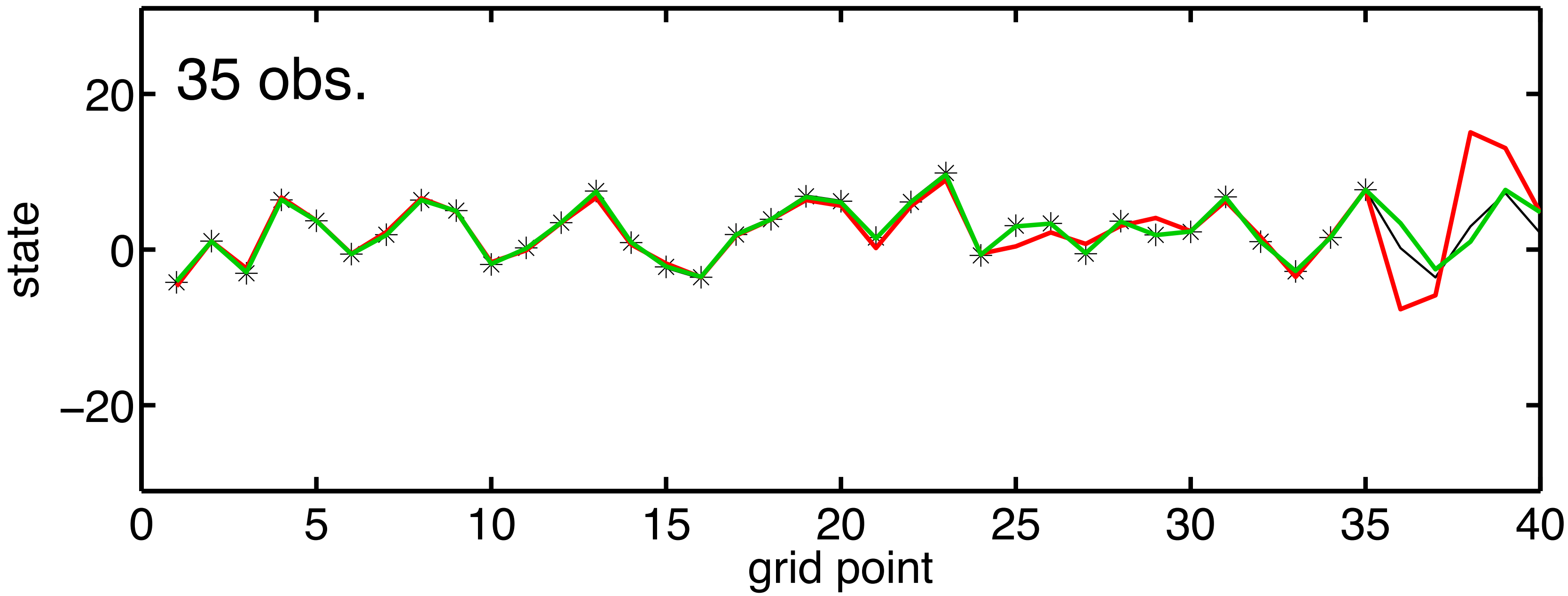
- State estimate for different numbers of observations



Instability of serial obs. processing with localization

More detailed view:

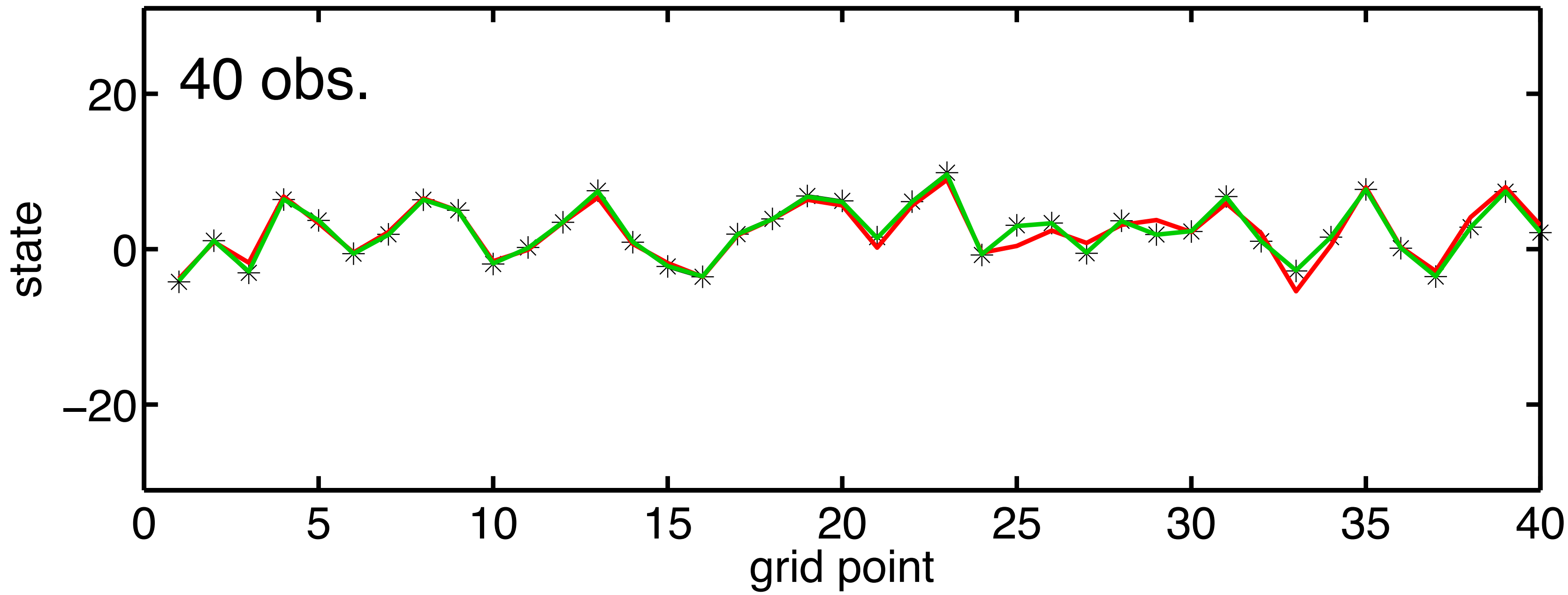
- State estimate for different numbers of observations



Instability of serial obs. processing with localization

More detailed view:

- State estimate for different numbers of observations



Inconsistent matrix updates

The Kalman filter updates the covariance matrix according to

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f\,(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (1)$$

With the Kalman gain

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \quad (2)$$

this simplifies to

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f \quad (3)$$

(1) and (3) yield same result **only** with gain (2)!

Not fulfilled with localization:

$$\mathbf{K}_{loc} = \left(\mathbf{C} \circ \mathbf{P}^f\right)\mathbf{H}^T\left(\mathbf{H}\left(\mathbf{C} \circ \mathbf{P}^f\right)\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

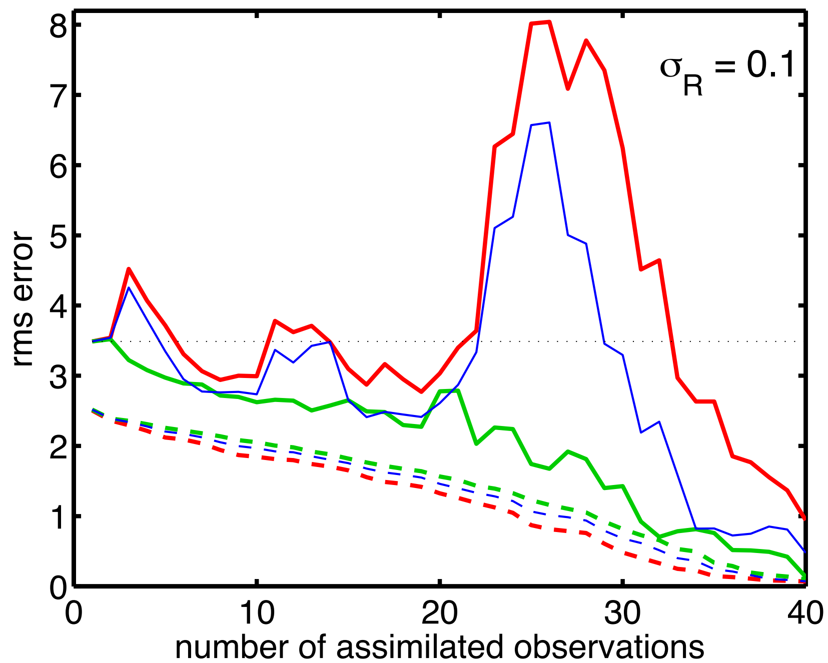
- Update of **P** is inconsistent in localized EnSRF (already noted by Whitaker & Hamill (2002), but never further examined)

Inconsistent matrix updates (2)

The inconsistency also occurs in LETKF, LESTKF, LSEIK ...

- But here: update is only done once followed by ensemble forecast

- LETKF with serial observation processing also shows instability



Blue: LETKF with serial observation processing

Simple Example

State estimate & covariance matrix

$$\mathbf{x}^f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \quad \mathbf{P}^f = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

Observation

$$\mathbf{y} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad \mathbf{R} = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}; \quad \mathbf{H} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Localization matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$$

Simple Example

Bulk update (all observations at once)

$$\mathbf{P}^a_{(sym.)} = \begin{pmatrix} 0.089 & 0.007 \\ 0.007 & 0.089 \end{pmatrix}; \quad \mathbf{P}^a_{(1sided)} = \begin{pmatrix} 0.080 & 0.058 \\ 0.058 & 0.080 \end{pmatrix}$$

$$\mathbf{x}^a_{(bulk)} = \begin{pmatrix} 0.077 \\ 0.077 \end{pmatrix}$$

Serial update

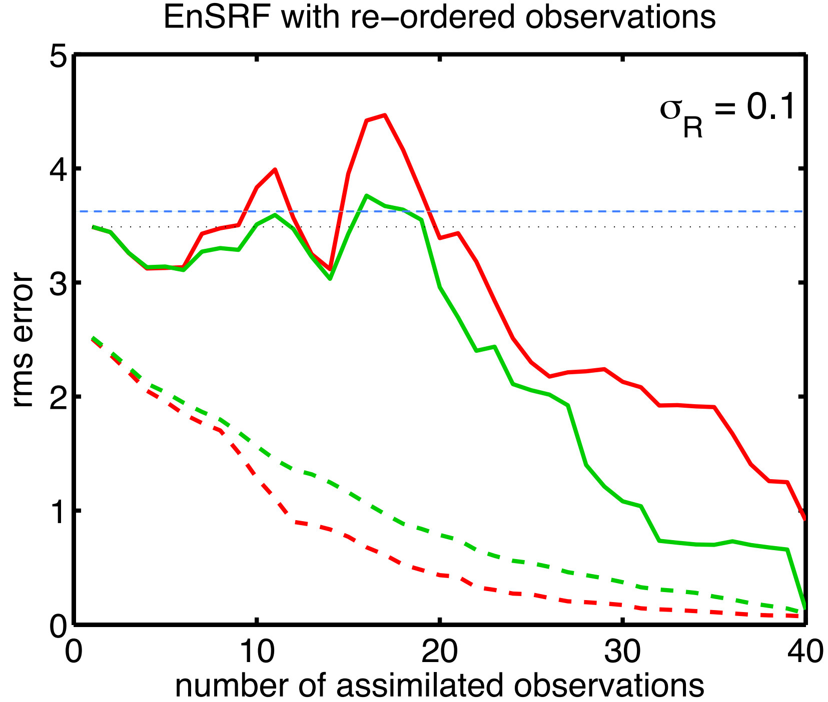
$$\mathbf{P}^a_{(sym.)} = \begin{pmatrix} 0.088 & 0.009 \\ 0.009 & 0.088 \end{pmatrix}; \quad \mathbf{P}^a_{(1sided)} = \begin{pmatrix} 0.089 & 0.055 \\ 0.055 & 0.076 \end{pmatrix}$$

$$\mathbf{x}^a_{(sym.)} = \begin{pmatrix} 0.097 \\ 0.073 \end{pmatrix}; \quad \mathbf{x}^a_{(1sided)} = \begin{pmatrix} 0.091 \\ 0.046 \end{pmatrix}$$

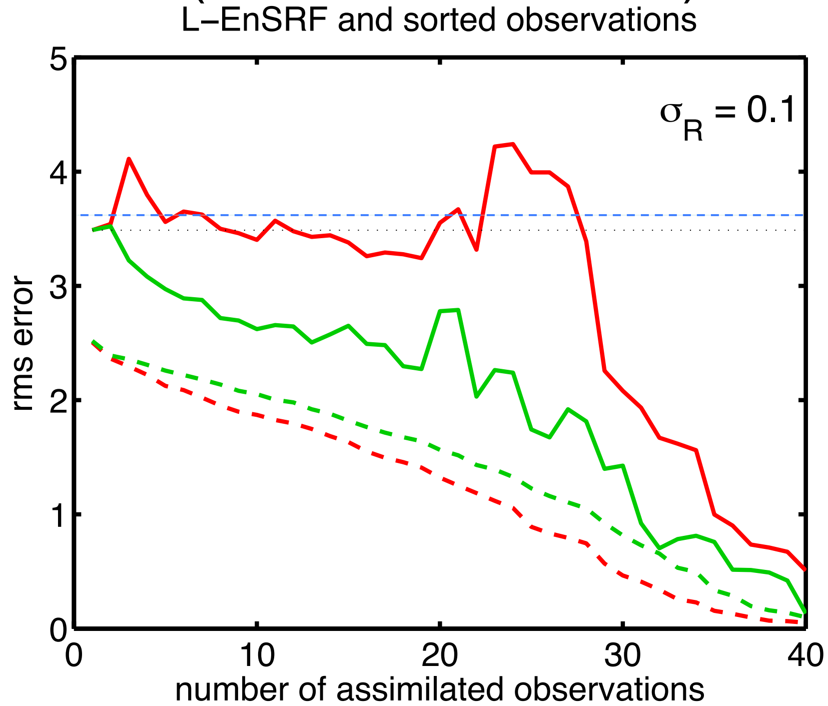
Effect of observation reordering

- Before: Assimilated observation from grid point 1 to 40 with increasing index
- What is the effect when we re-order the observations?

„maximum distance“
(1, 21, 11, 31, 6, 26,...)

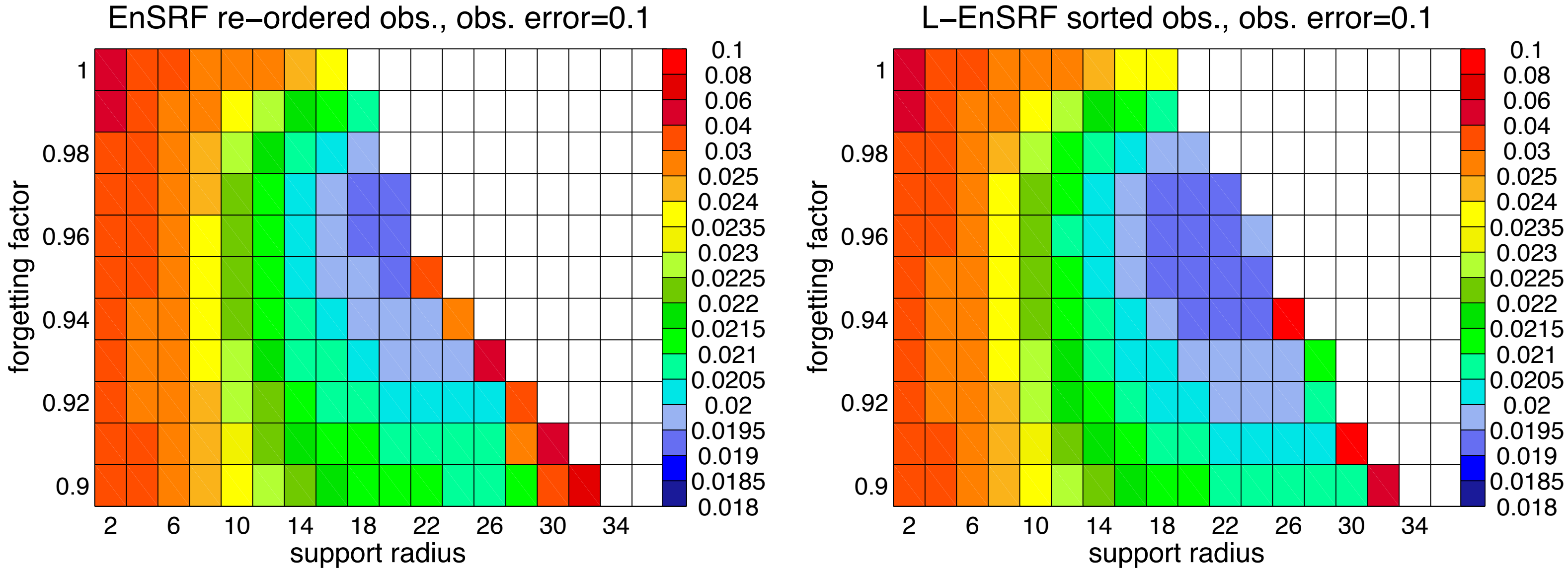


local observation sorting
(Whitaker et al. 2008)



Observation reordering

Full experiment over 50000 analysis steps, N=10



Maximum-distance reordering	EnSRF with local observation sorting
• practically no effect on final results	• improves stability • But not minimum error

L. Nerger, Mon. Wea. Rev. 143 (2015) 1554-1567

Optimal Localization Radius

Domain & Observation localization

Localization radius can depend on

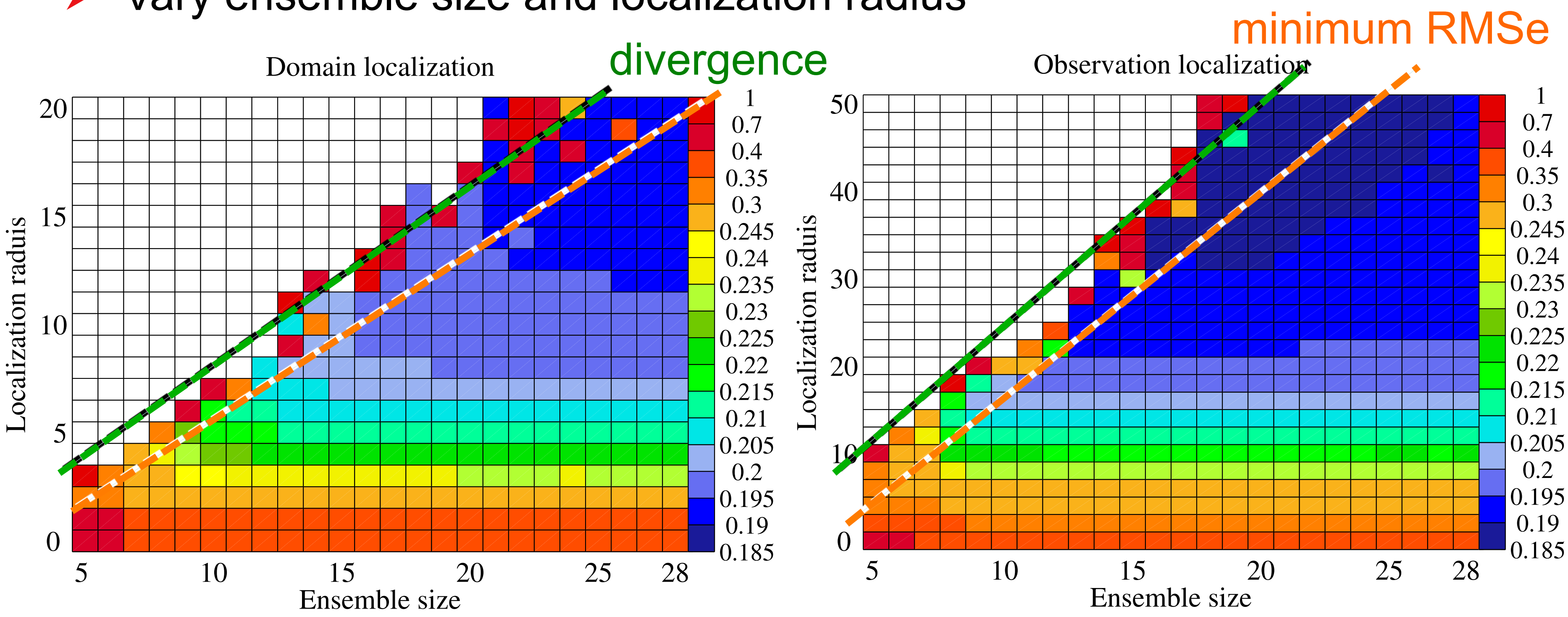
- Ensemble size
- Model dynamics & resolution
- Field

Optimal localization radius

- Typically determined experimentally (very costly)
- Some authors proposed adaptive methods (e.g. Anderson, Bishop/Hodyss, *Harlim*)
    - still with tunable parameters

Relation between ensemble size and localization radius

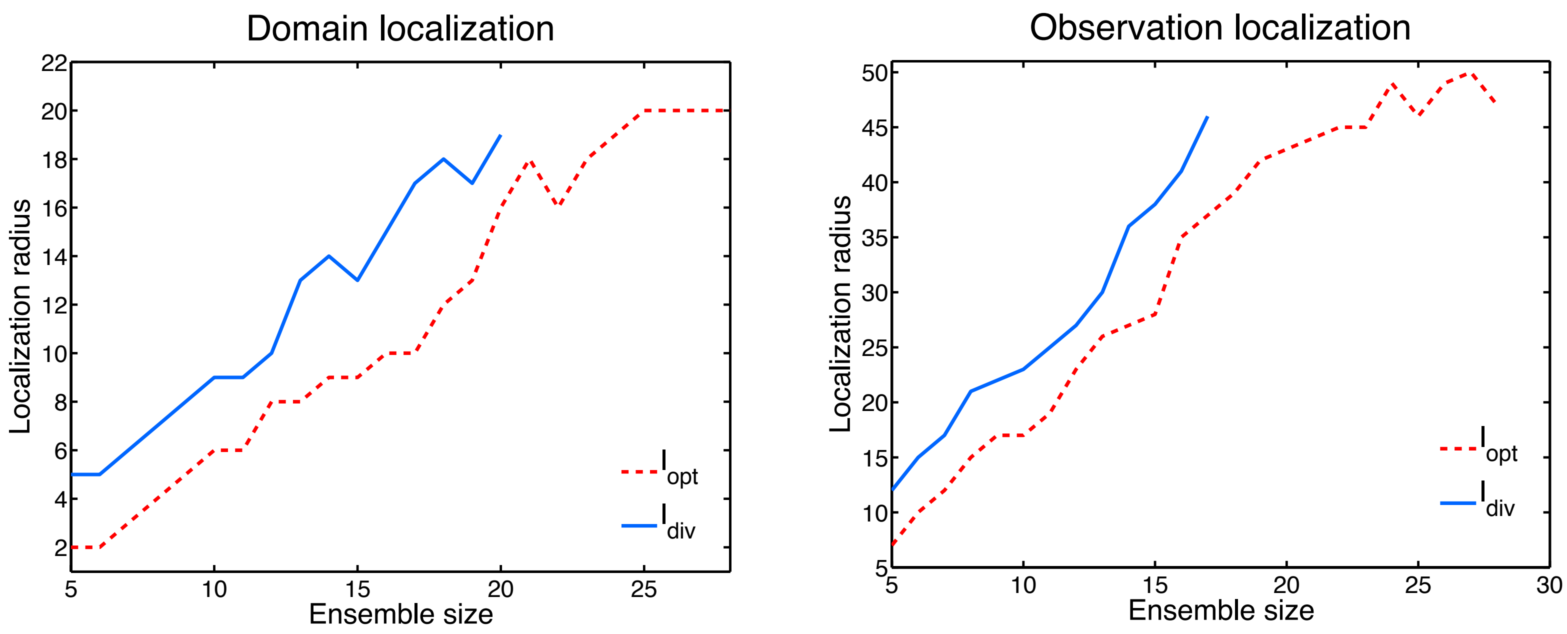
- Test runs with Lorenz-96 model
- Vary ensemble size and localization radius



- White: Filter divergence

Optimal localization radius

- Optimal localization radius as function of ensemble size

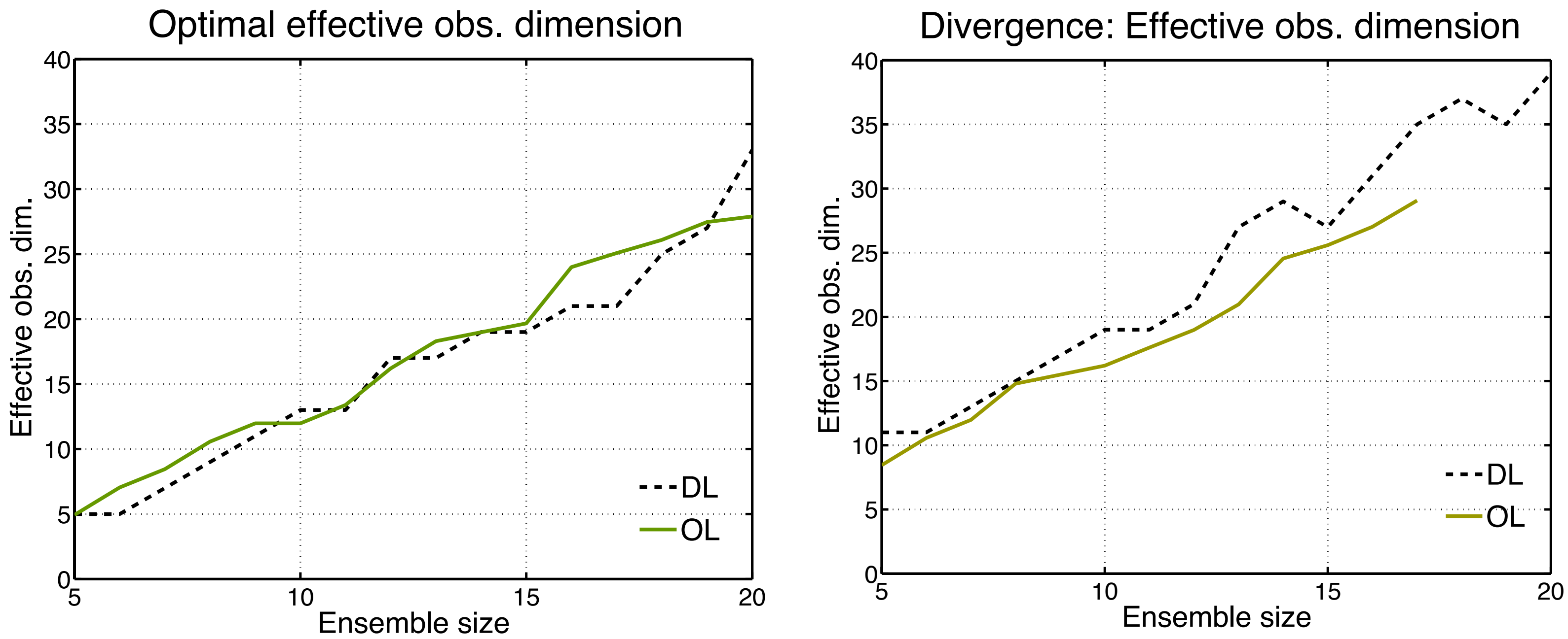


- Linear dependence for domain and observation localization
- Radius larger for OL than DL

Relate domain and observation localizations

- Define:

  Effective observation dimension d_W = sum of observation weights



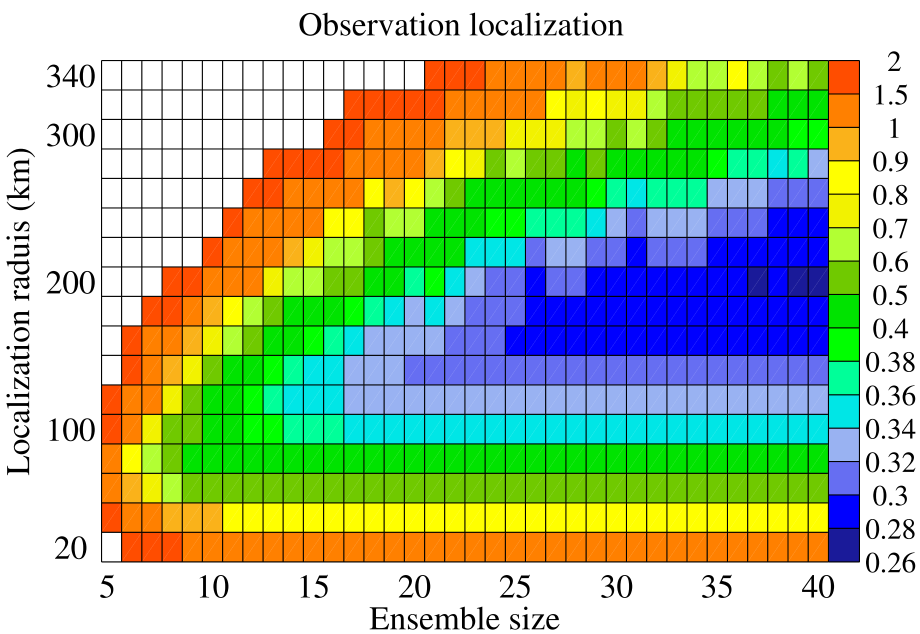
- Minimum RMS errors when effective obs. dimension slightly larger than ensemble size

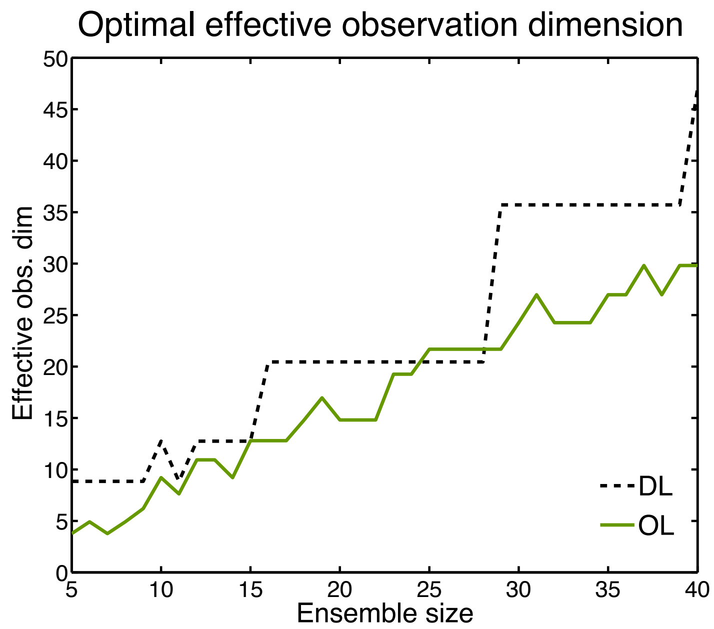
- When $\mathbf{d_W=N}$, errors are almost as small
  (optimal use of degrees of freedom from ensemble?)

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

2D Shallow Water Model

- Shallow water model simulating a double gyre in a box
- Assimilate sea surface height at each grid point





- For DL: steps due to addition of observations
- d_W optimal if about or slightly lower than ensemble size
- relation holds for different weight functions

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

PDAF: A tool for data assimilation

PDAF Parallel Data Assimilation Framework

PDAF - Parallel Data Assimilation Framework

- a software to provide assimilation methods
- an environment for ensemble assimilation
- for testing algorithms and real applications
- useable with virtually any numerical model
- also:
  - apply identical methods to different models
  - test influence of different observations
- makes good use of supercomputers
  (Fortran and MPI; tested on up to 17000 processors)
- first public release in 2004; continued development

More information and source code available at

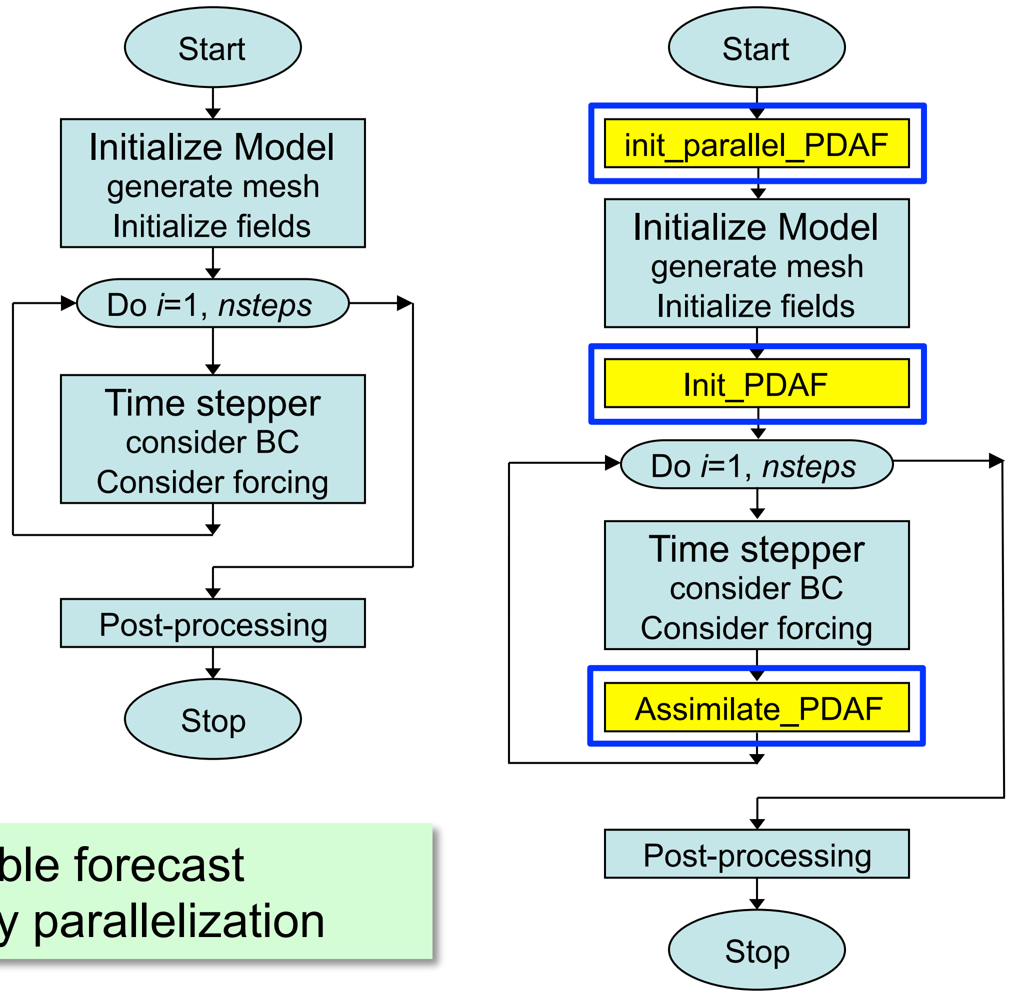
http://pdaf.awi.de

L. Nerger & W. Hiller, Computers & Geosciences 55 (2013) 110-118

AWI

Extending a Model for Data Assimilation

Model



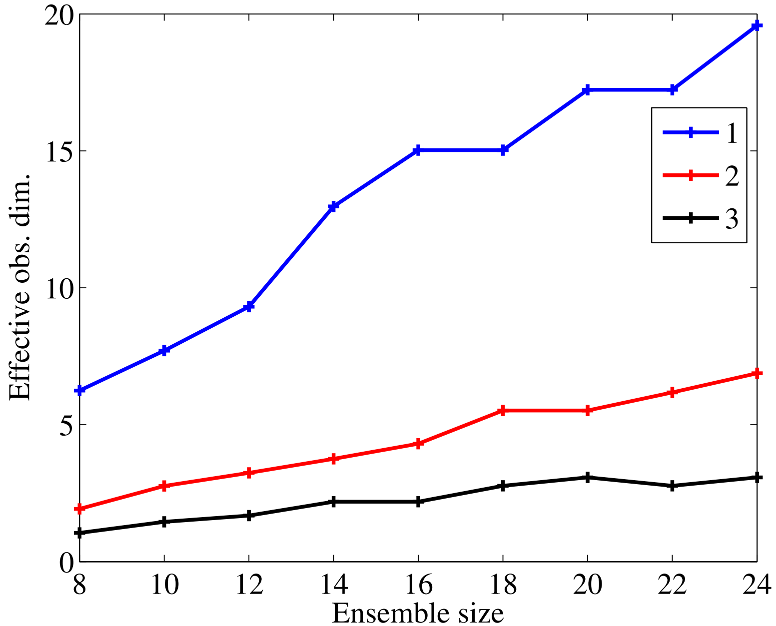
Extension for data assimilation

plus:
Possible model-specific adaption.

ensemble forecast
enabled by parallelization

2D Shallow Water Model

- Sparser observations
  - 1/4 and 1/9 of observations

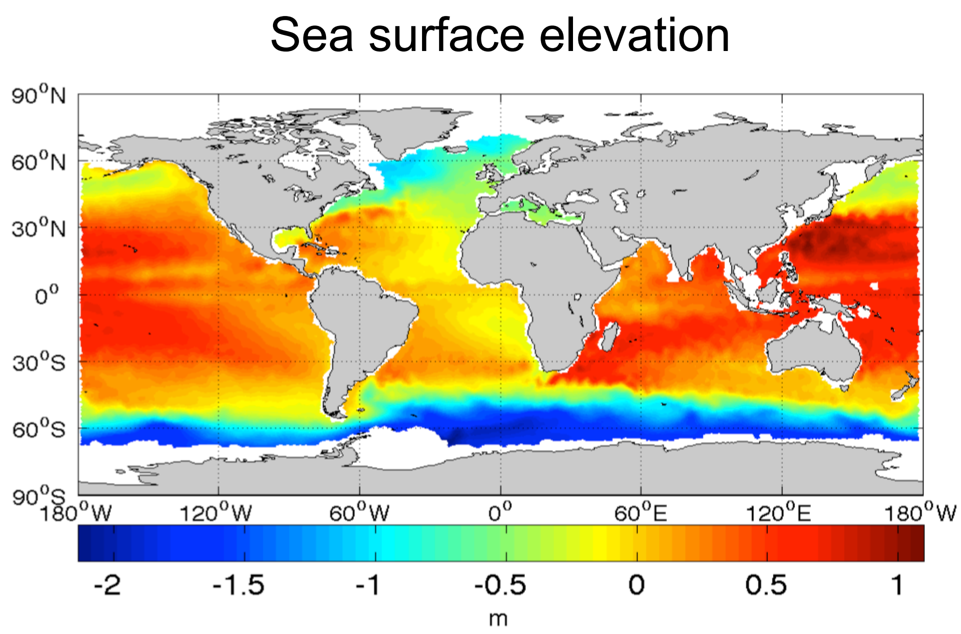


- Still linear dependence between effective obs. dimension and N
- Effective obs. dimension has to be scaled by obs. density

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

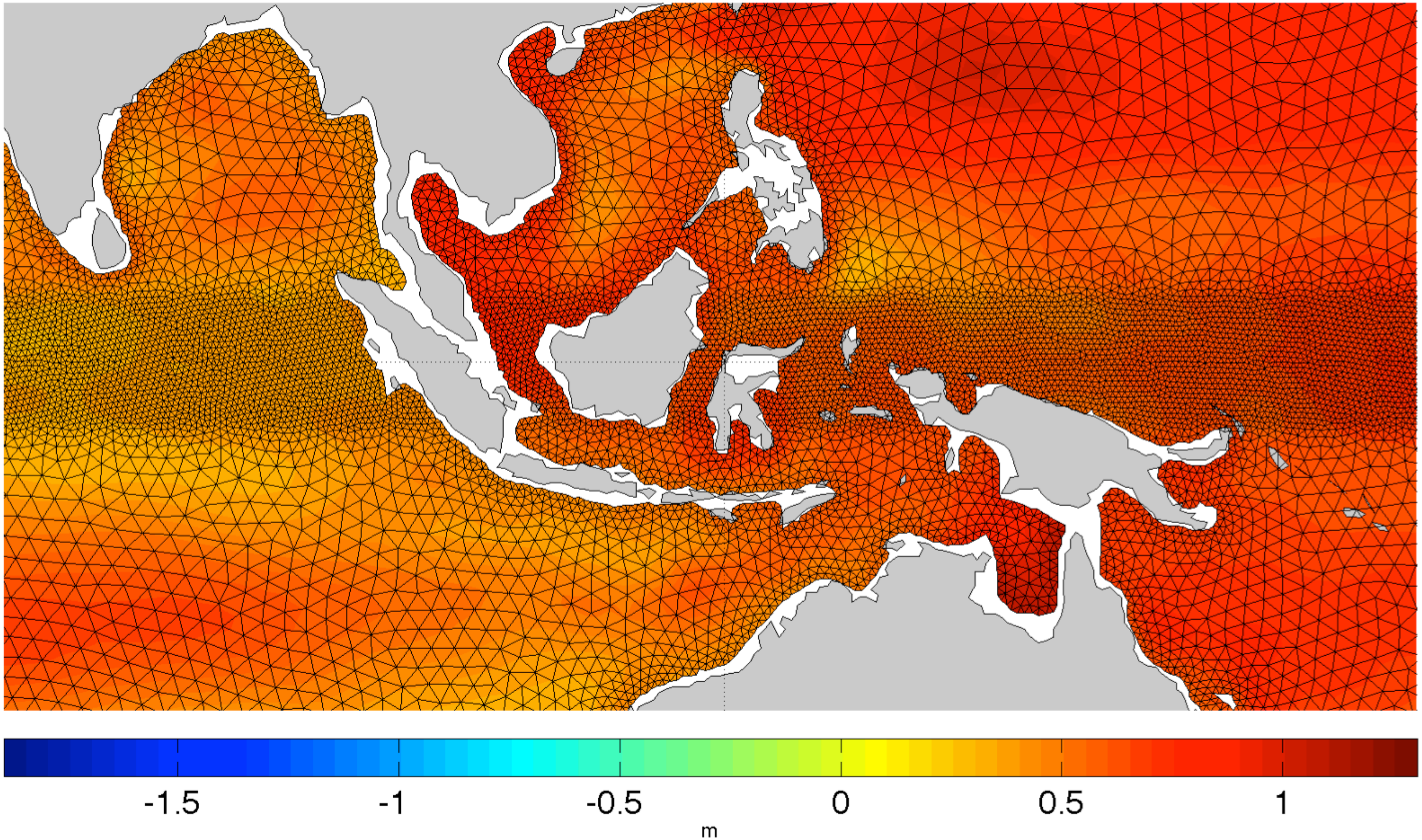
Large scale data assimilation: Global ocean model

- Finite-element sea-ice ocean model (FESOM, Danilov et al.)
- Global configuration (~1.3 degree resolution with refinement at equator)
- State vector size: 10^7
- Scales well up to 256 processor cores



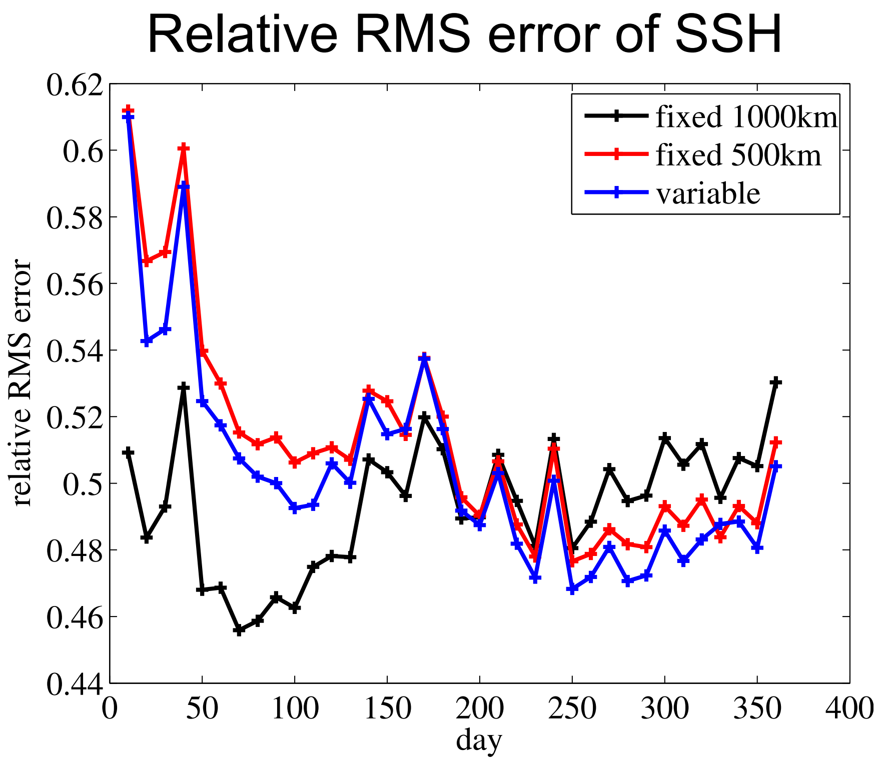
- Assimilate synthetic sea surface height (SSH) data for ocean state estimation
- Costly due to large model size (using up to 2048 processor cores)

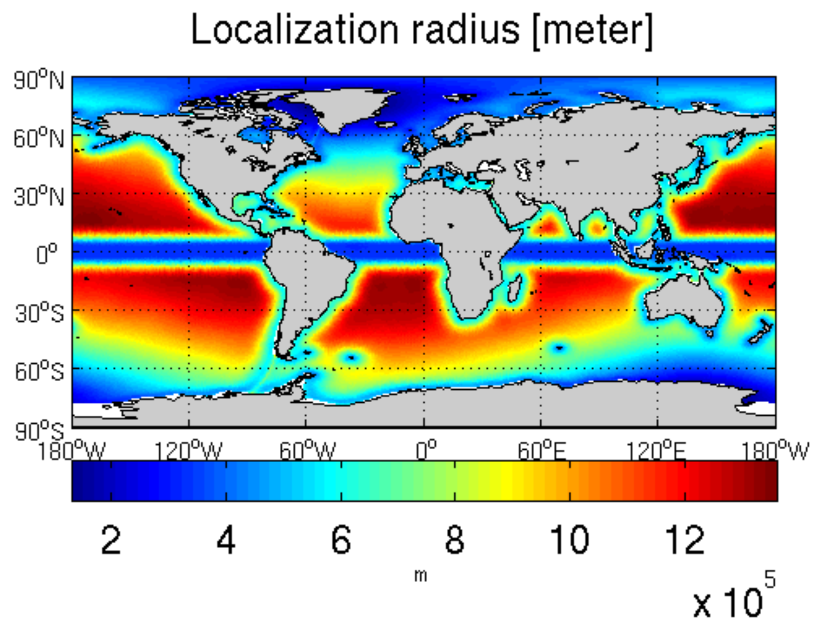
Model mesh at the equator



Adaptive localization radius in global ocean model

- Localization radius follows mesh resolution
- Fixed 1000km radius leads to increasing errors in 2nd half of year
- Lower RMS error in SSH than fixed 500km radius





Discussion on localization radius

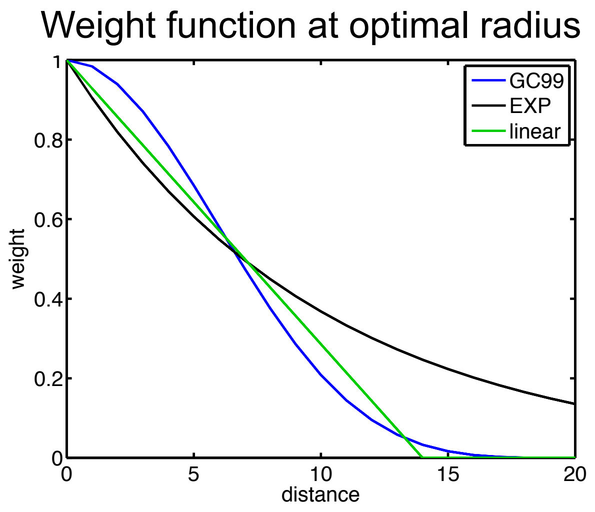
- Findings:
  - Effective observation dimension d_W relates to degrees of freedom
  - d_W close to ensemble size a good choice
  - No dependence on model dynamics
- Limitations
  - Observations at each grid point (optimal d_W smaller for incomplete observations)
  - Uniform observation error
  - Ignoring information content of observations (e.g. Migliorini, QJRMS 2013)

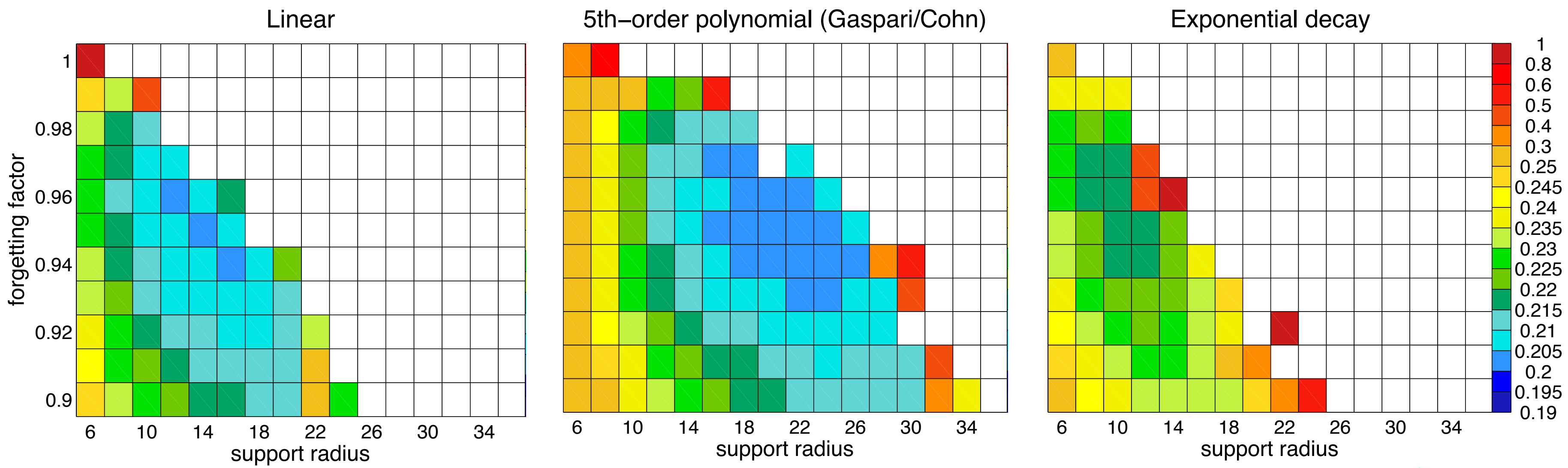
P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

Weight Functions

Weight function

- Why 5th-order Gaspari/Cohn polynomial?
- Covariance function not required for OL
- Furrer/Bengtsson (2007) indicate best sampling error reduction in $\mathbf{P}^f$ for exponential covariances
- For Lorenz96, some other functions give similar errors – but not significantly lower ones





Summary

- Serial observation processing filters can be unstable when used with localization
- Update of state error covariance matrix P inconsistent when localization is applied (all filters except classical EnKF)
- Estimation of adaptive localization radius dependent on ensemble size possible for “dense” observations
  - luckily a usual situation for ocean models assimilating satellite data

Thank you!