Practical Aspects of Ensemble-based Kalman Filters

Lars Nerger

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

and

Bremen Supercomputing Competence Center BremHLR
Bremen, Germany

lars.nerger@awi.de

BremHLR
Kompetenzzentrum für Höchstleistungsrechnen Bremen

AWI

Outline

Aspects

- Computing
- Analysis formulation
- Localization

Collaborations:

AWI: W. Hiller, J. Schröter, S. Loza, P. Kirchgessner, T. Janjic (now DWD)

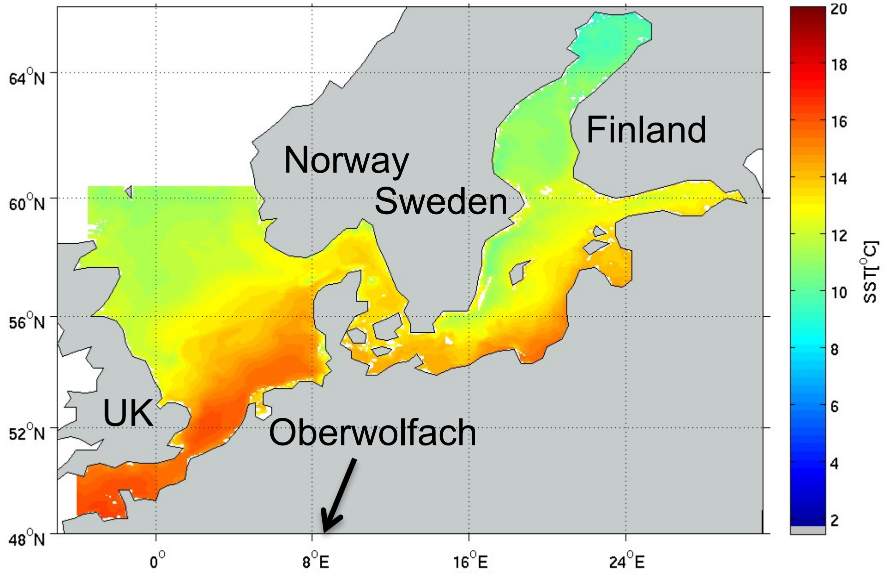
BSH: F. Janssen, S. Massmann

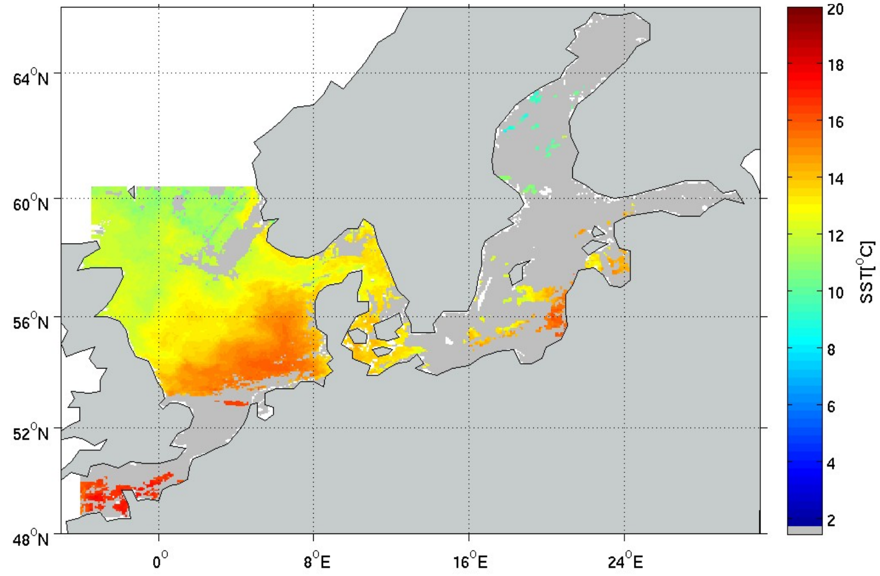
Bremen University: A. Bunse-Gerstner

The Problem

Application Example

DeMarine





- Forecasting in North & Baltic Seas
- Combine model and observations for optimal initial condition
- State vector size: $2.6 \cdot 10^6$ (4 fields 3D, 1 field 2D)
- Obervations: 10000 – 37000 (Surface temperature only)
- Ensemble size 8

S. Loza et al., Journal of Marine Systems 105 (2012) 152-162

AWI

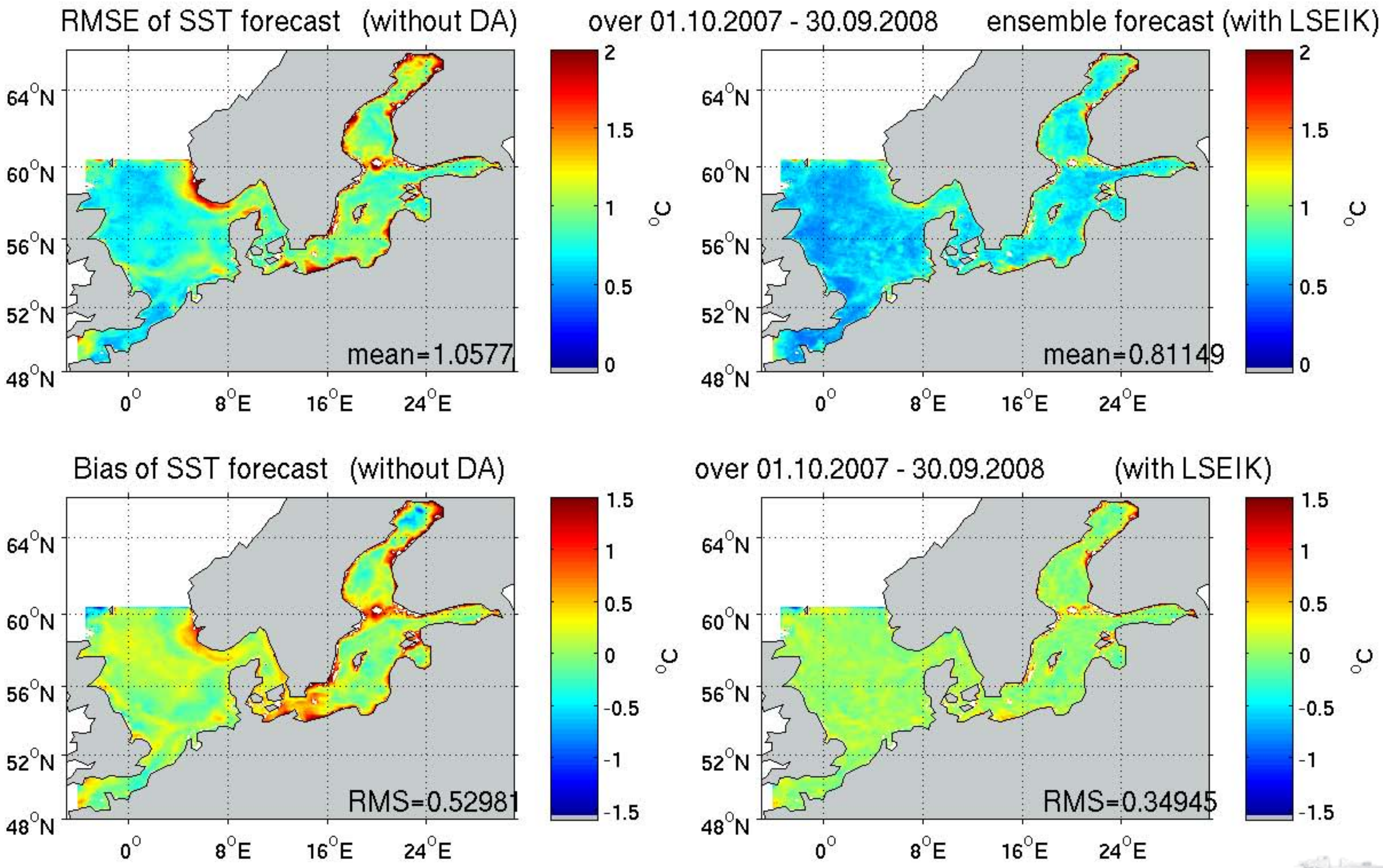
Forecast deviation from satellite data

DeMarine

No assimilation — Assimilation

RMS

bias



Improvements also sub-surface and in other fields

AWI

Data Assimilation

Problem: Estimate model state (trajectory) from

- guess at initial time
- model dynamics
- observational data

Characteristics of system:

- approximated by discretized differential equations
- high-dimension - $\mathcal{O}(10^7\text{-}10^9)$
- sparse observations
- non-linear

Current “standard” methods:

- Optimization algorithms (“4DVar”)
- Ensemble-based estimation algorithms

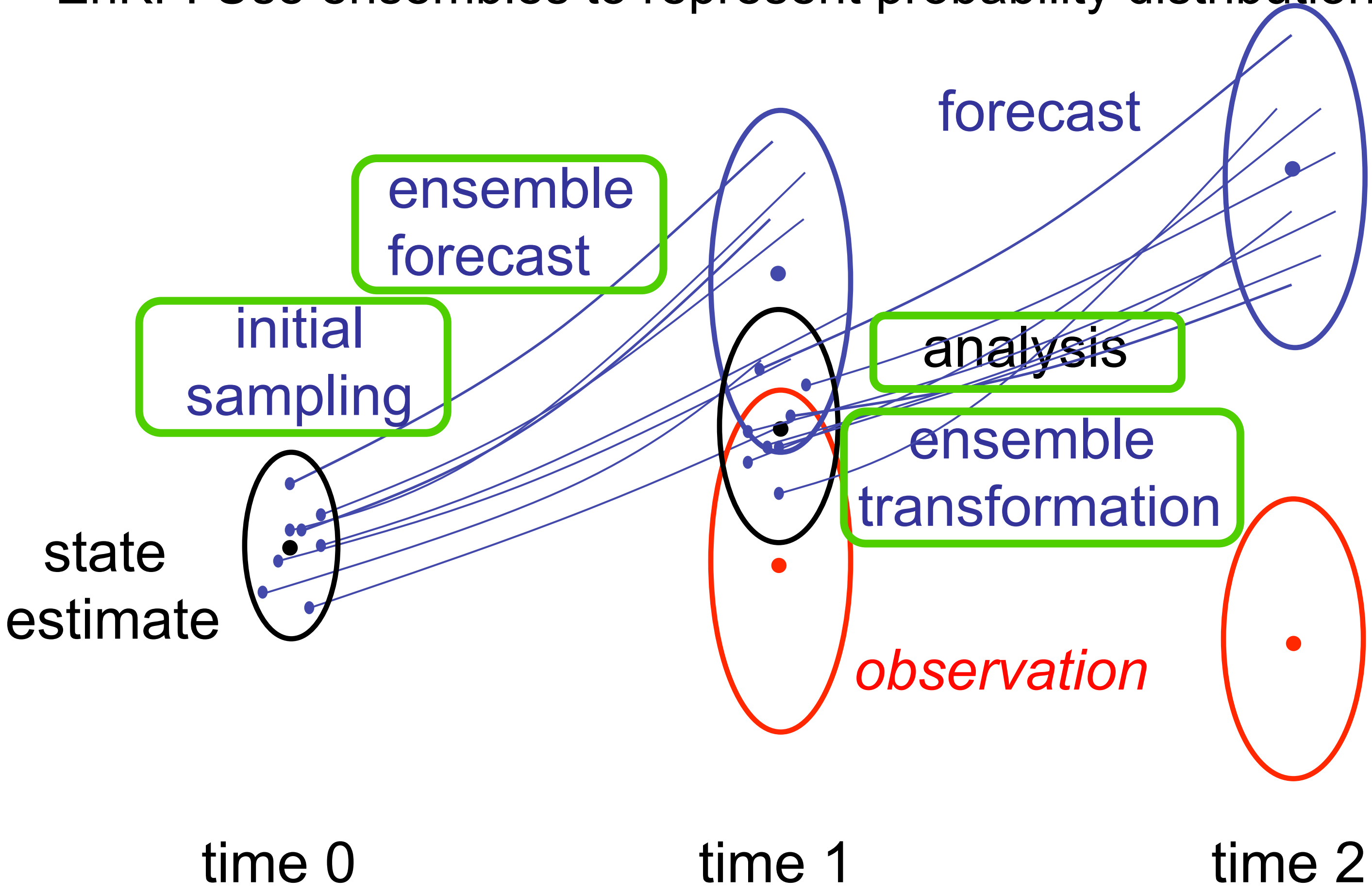
This talk!

Ensemble-based Kalman Filter

First formulated by G. Evensen (EnKF, 1994)

Kalman filter: express probability distributions by mean and covariance matrix

EnKF: Use ensembles to represent probability distributions



Looks trivial!

BUT:

There are many possible choices!

Computational and Practical Issues

Data assimilation with ensemble-based Kalman filters is costly!

Memory: Huge amount of memory required
(model fields and ensemble matrix)

Computing: Huge requirement of computing time
(ensemble integrations)

Parallelism: Natural parallelism of ensemble integration exists
(needs to be implemented)

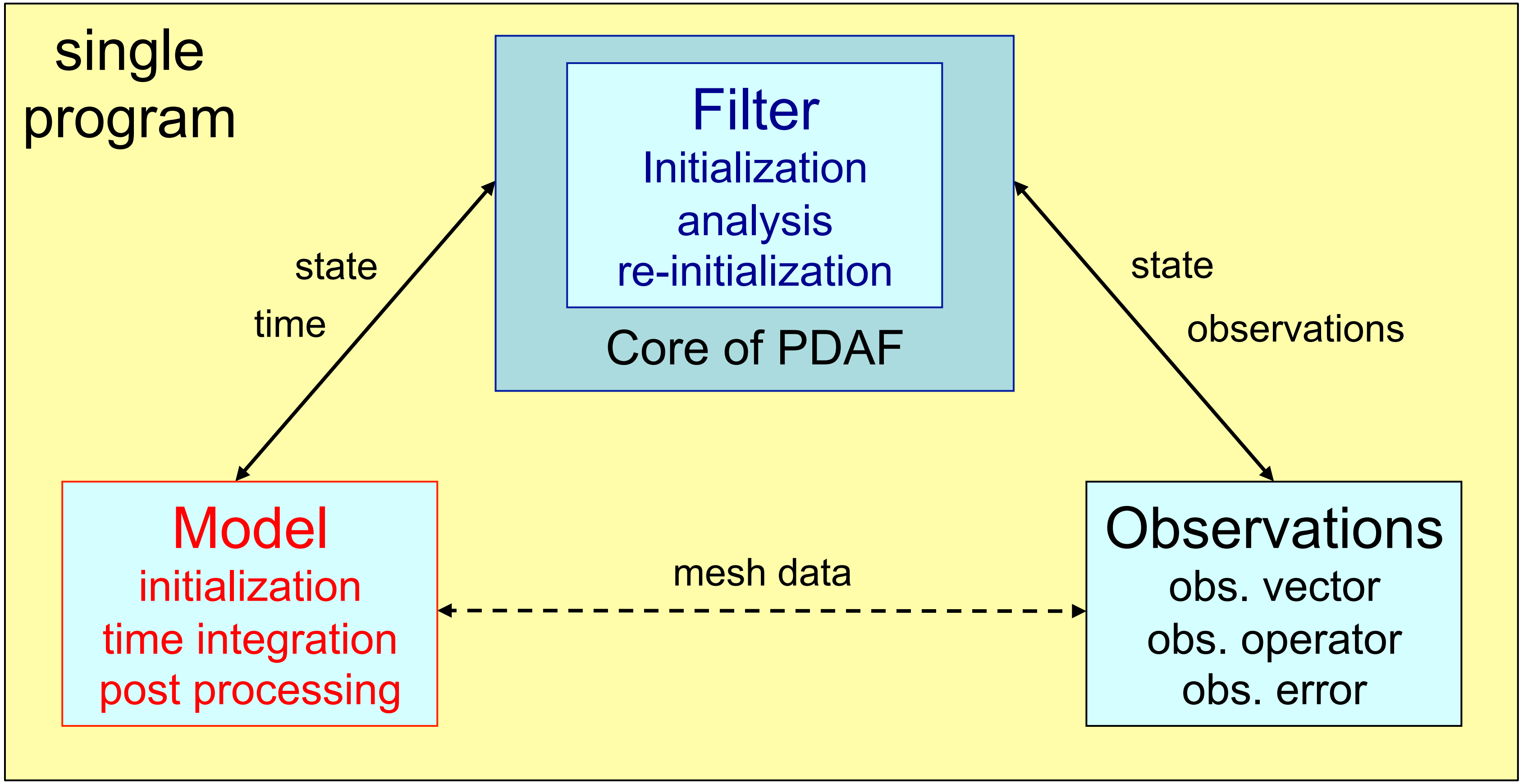
„Fixes": Filter algorithms do not work in their pure form
(„fixes" and tuning are needed)
because Kalman filter optimal only in linear case

What we are looking for…

- Goal: Find the assimilation method with
  - smallest estimation error
  - most accurate error estimate
  - least computational cost
  - least tuning
- Want to understand and improve performance
- Difficulty:
  - Optimality of Kalman filter well known for linear systems
  - No optimality for non-linear systems
  - ➜ limited analytical possibilities
  - ➜ apply methods to test problems

Computing

Logical separation of assimilation system



Nerger, L., Hiller, W. (2012). Software for Ensemble-based DA Systems – Implementation and Scalability. Computers and Geosciences. In press. doi:10.1016/j.cageo.2012.03.026

PDAF: A tool for data assimilation

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 4800 processors)

More information and source code available at

http://pdaf.awi.de

Analysis Formulations

Ensemble-based/error-subspace Kalman filters

A little "zoo" (not complete):

EnKF(94/98)

Studied in Nerger et al. (2005)

RRSQRT

ROEK

SEEK

SEIK

New study (Nerger et al. 2012)

EnKF(2003)

EnKF(2004)

EAKF

EnSRF

ETKF

ESTKF

MLEF

SPKF

ESSE

RHF

anamorphosis

New filter formulation

ESTKF: L. Nerger et al., Monthly Weather Review 140 (2012) 2335-2345

Model Equations

Stochastic dynamic model:

$$\mathbf{x}_i^t = M_{i,i-1}[\mathbf{x}_{i-1}^t] + \boldsymbol{\eta}_i, \quad \mathbf{x}_i^t, \boldsymbol{\eta}_i \in \mathbb{R}^n$$

Stochastic observation model:

$$\mathbf{y}_k = H_k[\mathbf{x}_k^t] + \boldsymbol{\epsilon}_k, \quad \mathbf{y}_k, \boldsymbol{\epsilon}_k \in \mathbb{R}^m$$

Assumptions:

$$\boldsymbol{\eta}_i \propto \mathcal{N}(\mathbf{0}, \mathbf{Q}_i); \quad \left\langle \boldsymbol{\eta}_i \boldsymbol{\eta}_j^T \right\rangle = \mathbf{Q}_i \delta_{ij}$$ Model error

$$\boldsymbol{\epsilon}_k \propto \mathcal{N}(\mathbf{0}, \mathbf{R}_k); \quad \left\langle \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_l^T \right\rangle = \mathbf{R}_k \delta_{kl}$$ Observation error

$$\mathbf{x}_i^t \propto \mathcal{N}(\bar{\mathbf{x}}_i^t, \mathbf{P}_i)$$

$$\left\langle \boldsymbol{\eta}_k \boldsymbol{\epsilon}_k^T \right\rangle = 0; \quad \left\langle \boldsymbol{\eta}_i (\mathbf{x}_i^t)^T \right\rangle = 0; \quad \left\langle \boldsymbol{\epsilon}_k (\mathbf{x}_k^t)^T \right\rangle = 0$$

The Ensemble Kalman Filter (EnKF, Evensen 94)

Initialization:

Generate random ensemble $\{\mathbf{x}_0^{a(l)}, l = 1, \ldots, N\}$

Ensemble statistics approximate $\mathbf{x}_0^{\mathrm{a}}$ and covariance $\mathbf{P}_0^{\mathrm{a}}$

Forecast:

$$\mathbf{x}_i^{a(l)} = M_{i,i-1}[\mathbf{x}_{i-1}^{a(l)}] + \boldsymbol{\eta}_i^{(l)}$$

Analysis:

$$\mathbf{x}_k^{a(l)} = \mathbf{x}_k^{f(l)} + \mathbf{K}_k\left(\mathbf{y}_k^{(l)} - \mathbf{H}_k\mathbf{x}_k^{f(l)}\right)$$

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

Kalman filter

$$\mathbf{P}_k^f := \frac{1}{N-1}\sum_{l=1}^{N}\left(\mathbf{x}_k^{f(l)} - \overline{\mathbf{x}_k^f}\right)\left(\mathbf{x}_k^{f(l)} - \overline{\mathbf{x}_k^f}\right)^T$$

$$\mathbf{x}_k^a := \frac{1}{N}\sum_{l=1}^{N}\mathbf{x}_k^{a(l)}$$

Issues of the EnKF94

Monte Carlo Method

- ensemble of observations required
  (samples matrix **R**; introduces sampling error)

Inversion of large matrix $\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \in \mathbb{R}^{m \times m}$

(can be singular, possibly large differences in eigenvalues >0)

Alternative:

- Compute analysis in space spanned by ensemble

Methods: *Ensemble Square-Root Kalman Filters*, e.g.

- SEIK (Pham et al., 1998)
- ETKF (Bishop et al., 2001)

Ensemble Transform Kalman Filter - ETKF

Ensemble perturbation matrix

size

$$\mathbf{X}'_k := \mathbf{X}_k - \overline{\mathbf{X}_k}$$

(n x N)

Analysis covariance matrix

$$\mathbf{P}^a = \mathbf{X}'^f \mathbf{A} (\mathbf{X}'^f)^T$$

(n x n)

“Transform matrix” (in ensemble space)

$$\mathbf{A}^{-1} := (N-1)\mathbf{I} + (\mathbf{H}\mathbf{X}'^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{X}'^f$$

(N x N)

Ensemble transformation

$$\mathbf{X}'^a = \mathbf{X}'^f \; \mathbf{W}^{ETKF}.$$

(n x N)

Ensemble weight matrix

$$\mathbf{W}^{ETKF} := \sqrt{N-1}\mathbf{C}\mathbf{\Lambda}$$

(N x N)

- $\mathbf{C}\mathbf{C}^T = \mathbf{A}$ (symmetric square root)
- $\mathbf{\Lambda}$ is identity or random orthogonal matrix with EV $(1, \ldots, 1)^T$)

SEIK Filter

Error-subspace basis matrix

size

$$\mathbf{L} \quad := \quad \mathbf{X}^f \ \mathbf{T}$$

(n x N-1)

(T subtracts ensemble mean and removes last column)

Analysis covariance matrix

$$\tilde{\mathbf{P}}^a = \mathbf{L}\tilde{\mathbf{A}}\mathbf{L}^T$$

(n x n)

"Transform matrix" (in error subspace)

$$\tilde{\mathbf{A}}^{-1} := (N-1)\mathbf{T}^T\mathbf{T} + (\mathbf{HL})^T\mathbf{R}^{-1}\mathbf{HL}$$

(N-1 x N-1)

Ensemble transformation

$$\mathbf{X}'^a = \mathbf{L}\ \mathbf{W}^{SEIK}$$

(n x N)

Ensemble weight matrix

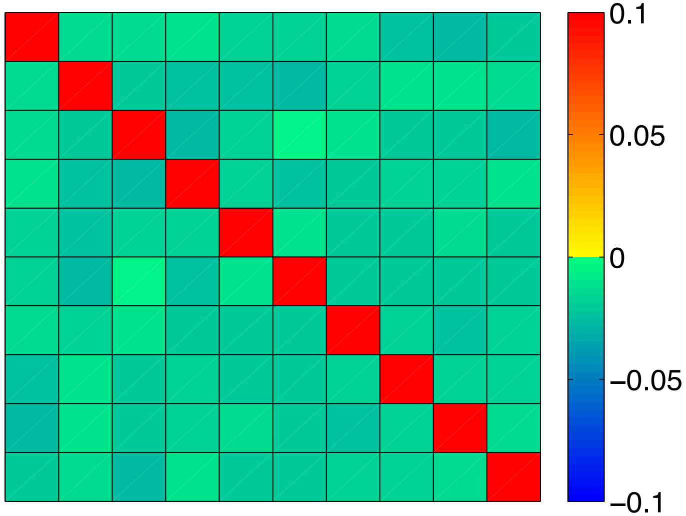
$$\mathbf{W}^{SEIK} := \sqrt{N-1}\tilde{\mathbf{C}}\mathbf{\Omega}^T$$

(N-1 x N)

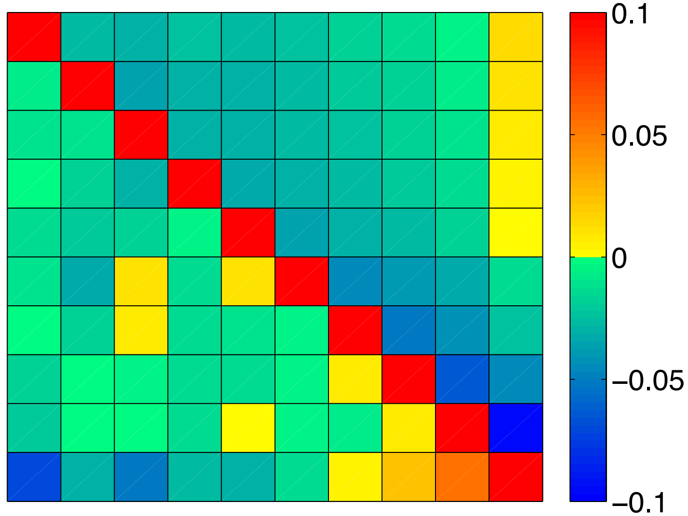
- $\tilde{\mathrm{C}}$ is square root of $\tilde{\mathrm{A}}$ (originally Cholesky decomposition)
- $\mathbf{\Omega}^T$ is transformation from N-1 to N (random or deterministic)

Weight Matrices (W in $X^{a'} = X^f W$)

ETKF



SEIK-Cholesky sqrt



ETKF

main contribution from diagonal (minimum transformation)

Off-diagonals of similar weight

➜ Minimum change in distribution of ensemble variance

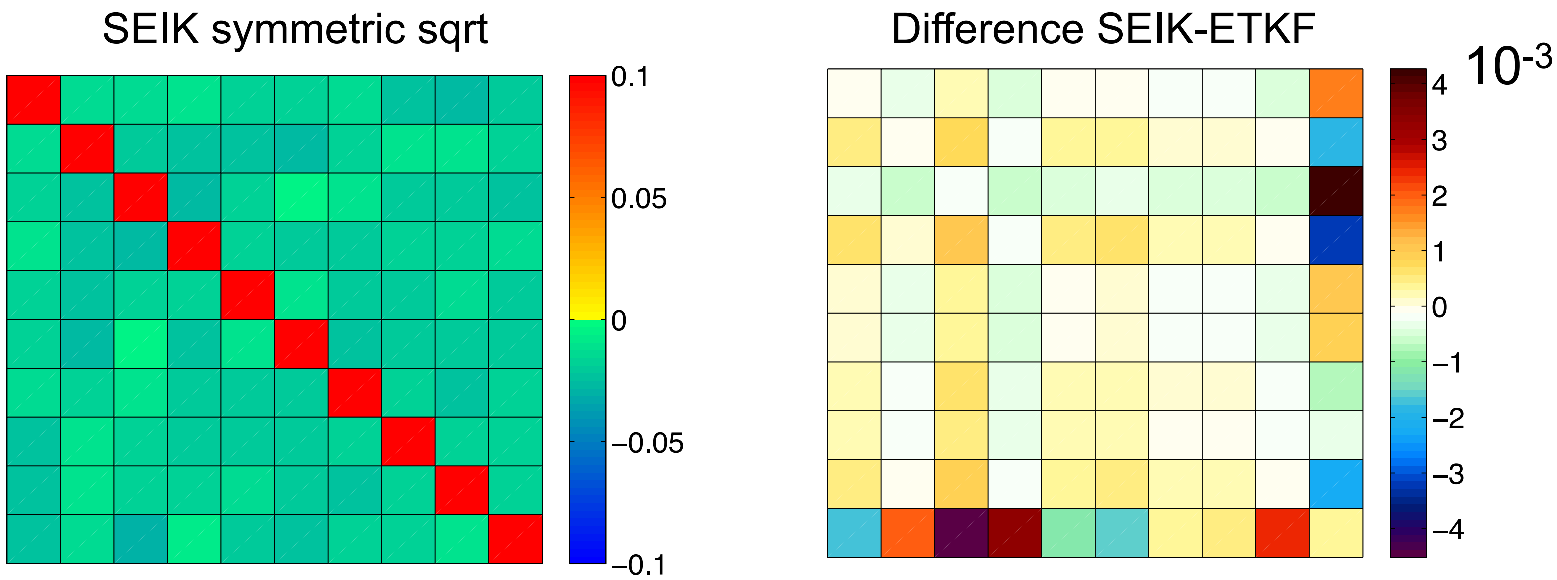
SEIK with Cholesky sqrt

main contribution from diagonal

Off-diagonals with strongly varying weights

➜ Changes distribution of variance in ensemble

Transformation Matrix of SEIK/symmetric sqrt



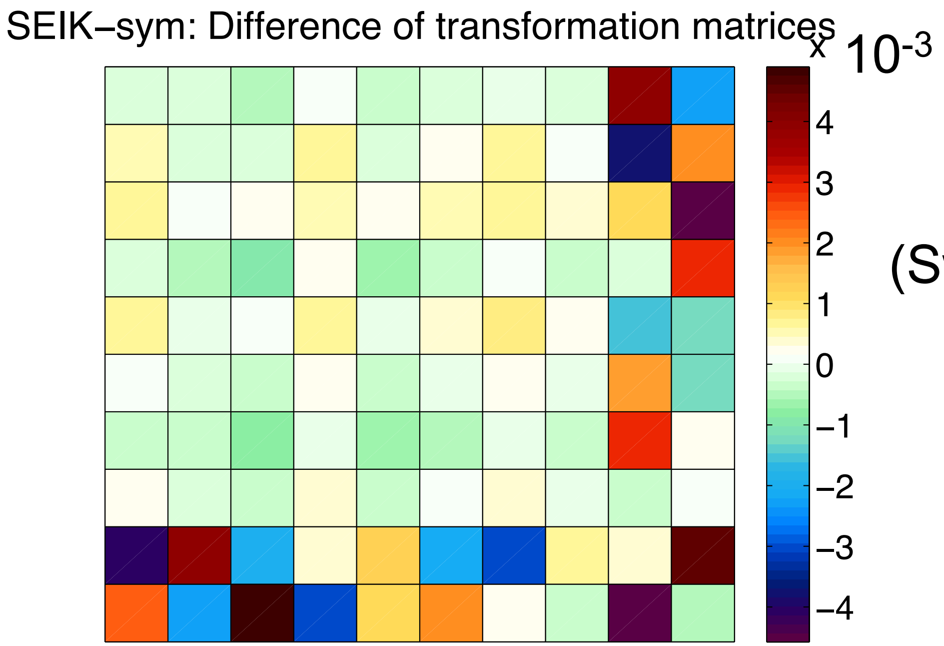
Transformation matrices of ETKF and SEIK-sym very similar

Largest difference for last ensemble member

(Experiments with Lorenz96 model: This can lead to smaller ensemble variance of this member)

SEIK depends on ensemble order

Switch last two ensemble members



(Switched back last two columns & rows for comparison)

Ensemble transformation depends on order of ensemble members (For ETKF the difference is 10^{-15})

Statistically fine, but not desirable!

Revised T matrix

Identical transformations require different projection matrix for SEIK:

$$\mathbf{L} \;:=\; \mathbf{X}^f\,\mathbf{T}$$

For SEIK:

$\mathbf{T}$ subtracts ensemble mean and drops last column

➜ Dependence on order of ensemble members!

➜ Solution:

- ➜ Redefine $\mathbf{T}$: Distribute last member over first N-1 columns
- ➜ Also replace $\mathbf{\Omega}$ by new $\hat{\mathbf{T}}$

New filter formulation:

Error Subspace Transform Kalman Filter (ESTKF)

T-matrix in SEIK and ESTKF

SEIK:

$$\mathbf{T}_{i,j} = \begin{cases} 1 - \frac{1}{N} & \text{for } i = j, i < N \\ -\frac{1}{N} & \text{for } i \neq j, i < N \\ -\frac{1}{N} & \text{for } i = N \end{cases}$$

ESTKF:

$$\hat{\mathbf{T}}_{i,j} = \begin{cases} 1 - \frac{1}{N}\frac{1}{\frac{1}{\sqrt{N}}+1} & \text{for } i = j, i < N \\ -\frac{1}{N}\frac{1}{\frac{1}{\sqrt{N}}+1} & \text{for } i \neq j, i < N \\ -\frac{1}{\sqrt{N}} & \text{for } i = N \end{cases}$$

- Efficient implementation as subtraction of means & last column
- ETKF: improve compute performance using a matrix **T**

ESTKF: New filter with identical transformation as ETKF

New filter ESTKF – properties like ETKF:

- Minimum transformation
- Transformation independent of ensemble order

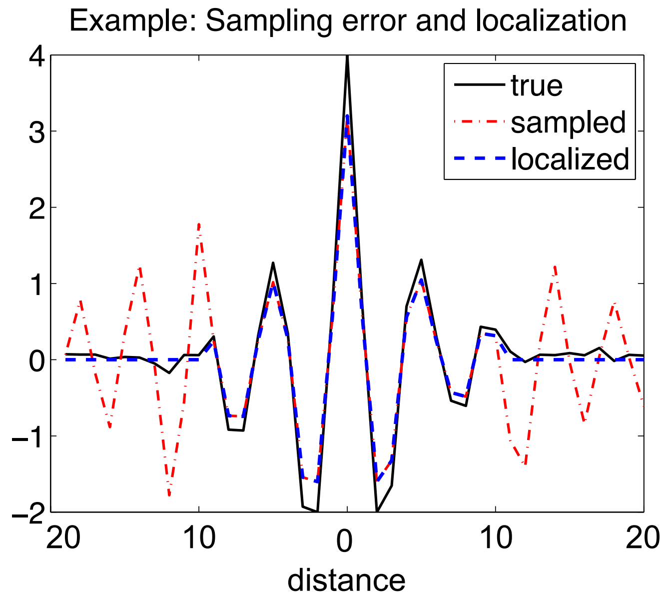
But:
- analysis computed in dimension N-1
- direct access to error subspace
- smaller condition number of **A**

L. Nerger et al., Monthly Weather Review 140 (2012) 2335-2345

Localization

Localization: Why and how?

- Combination of observations and model state based on estimated error covariance matrices
- Finite ensemble size leads to significant sampling errors
  - particularly for small covariances!

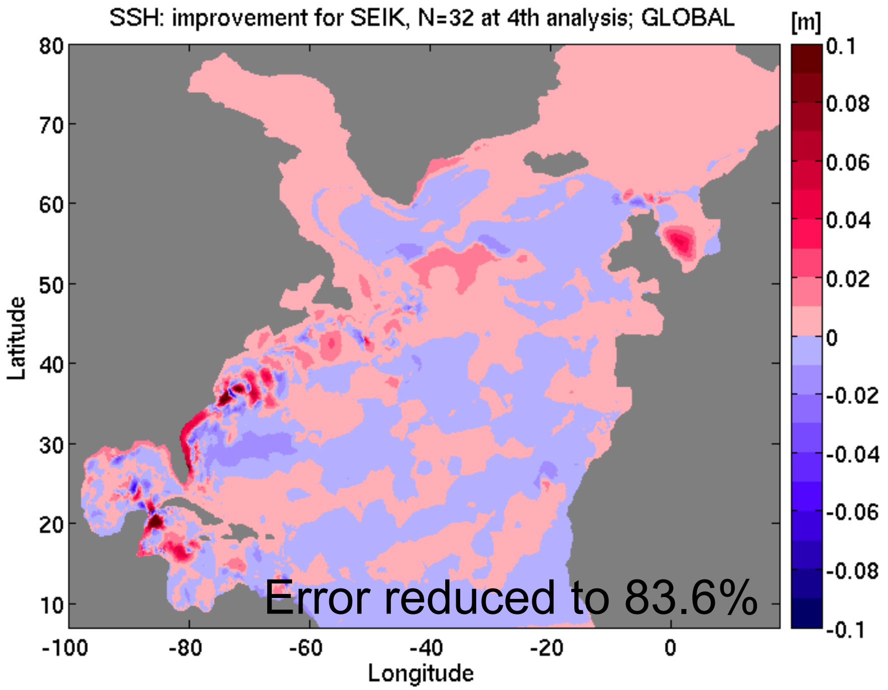


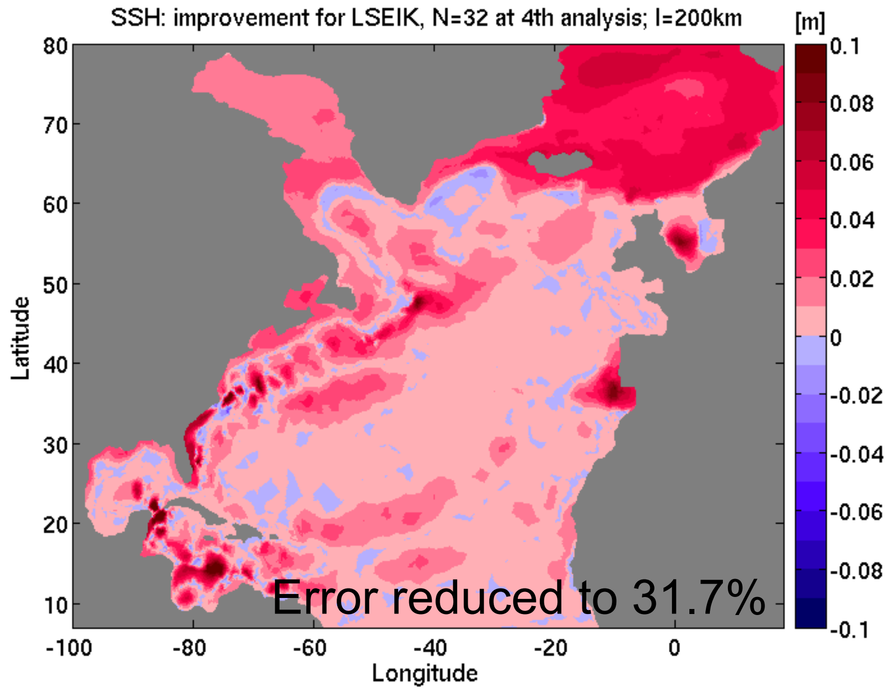
- Remove estimated long-range correlations

  ➜ Increases degrees of freedom for analysis (globally not locally!)

  ➜ Increases size of analysis correction

Global vs. local SEIK, N=32 (March 1993)





- Improvement is error reduction by assimilation
- Localization extents improvements into regions not improved by global SEIK
- Regions with error increase diminished for local SEIK
- Underestimation of errors reduced by localization

L. Nerger et al. Ocean Dynamics 56 (2006) 634

Localization Types

Simplified analysis equation:

$$\mathbf{x}^a = \mathbf{x}^f + \frac{\mathbf{P}^f}{\mathbf{P}^f + \mathbf{R}} \left(\mathbf{y} - \mathbf{x}^f\right)$$

Covariance localization

- Modify covariances in forecast covariance matrix $\mathbf{P}^f$
- Element-wise product with correlation matrix of compact support

Requires that $\mathbf{P}^f$ is computed (not in ETKF or SEIK)

Observation localization

- Modify observation error covariance matrix $\mathbf{R}$
- Needs distance of observation (achieved by local analysis or domain localization)

Possible in all filter formulations

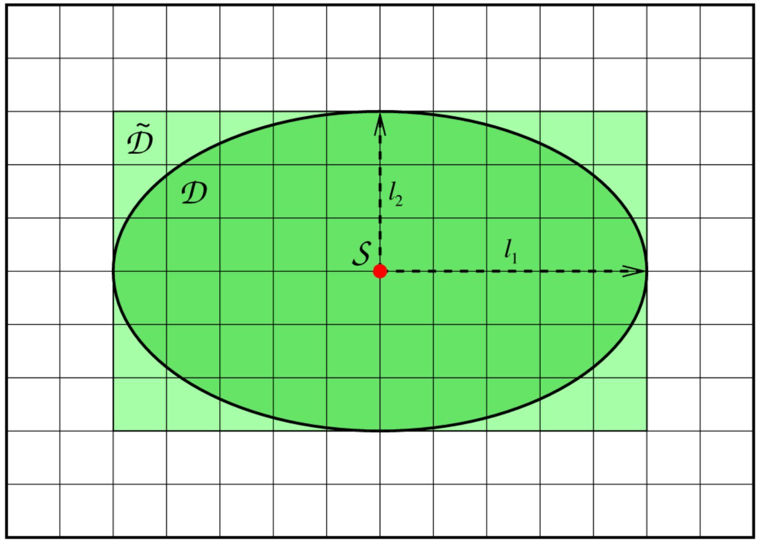
E.g.: Houtekamer/Mitchell (1998, 2001), Whitaker/Hamill (2002), Keppenne/Rienecker (2002)

E.g.: Evensen (2003), Ott et al. (2004), Nerger/Gregg (2007), Hunt et al. (2007)

Local SEIK filter – domain & observation localization

Local Analysis:

- Update small regions (like single vertical columns)
- Observation localizations: Observations weighted according to distance
- Consider only observations with weight >0
- State update and ensemble transformation fully local



S: Analysis region
D: Corresponding data region

Similar to localization in LETKF (e.g. Hunt et al, 2007)

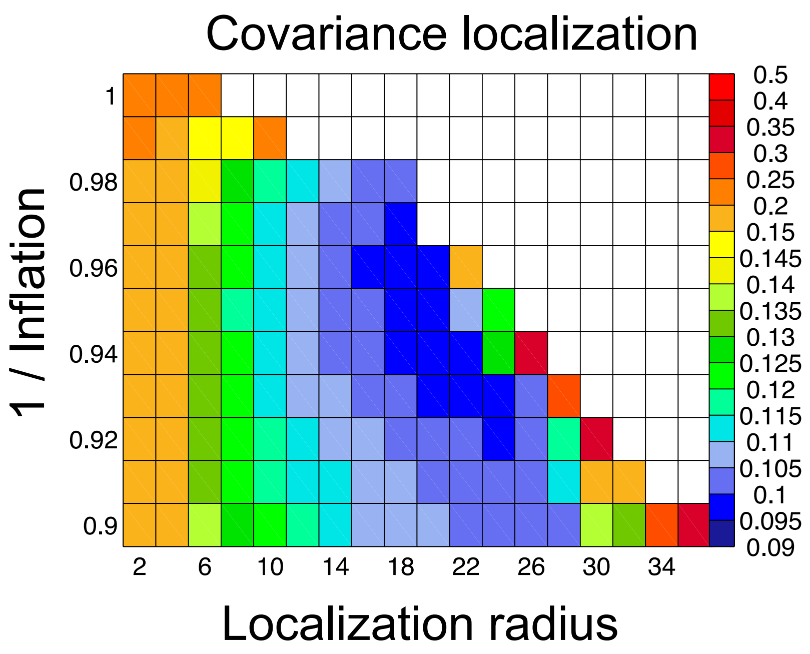
L. Nerger et al., Ocean Dynamics 56 (2006) 634
L. Nerger & W.W. Gregg, J. Mar. Syst. 68 (2007) 237

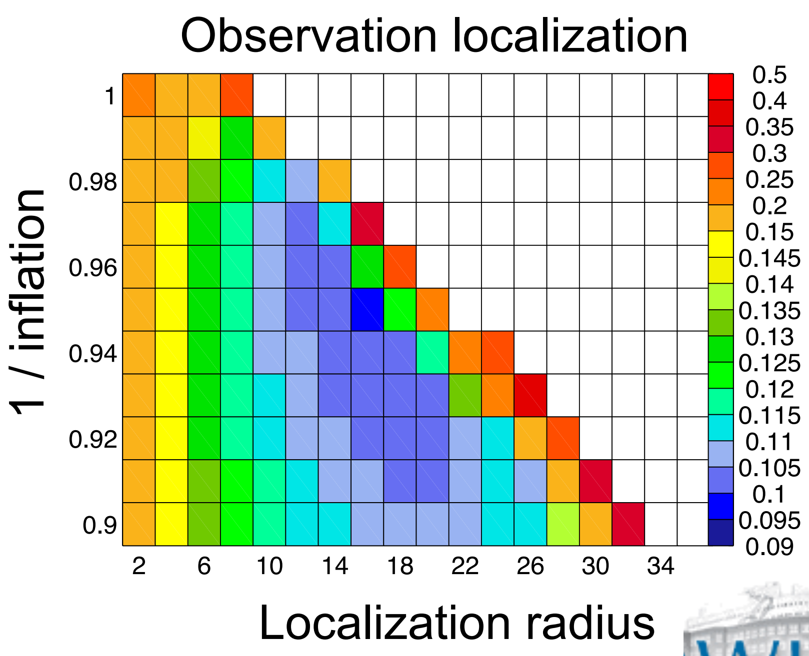
Different effect of localization methods

Experimental result:

- Twin experiment with simple Lorenz96 model
- Covariance localization better than observation localization
  (Also reported by Greybush et al. (2011) with other model)

Time-mean RMS errors

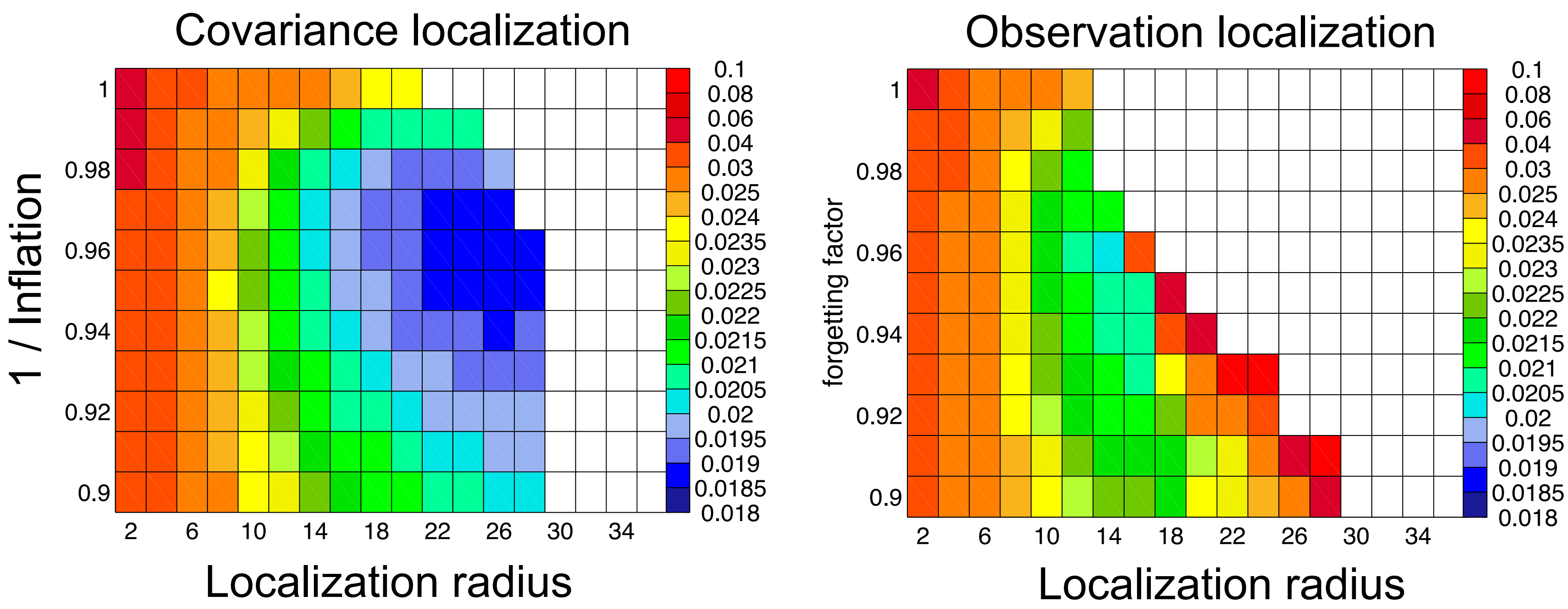




T. Janjic et al., Mon. Wea. Rev. 139 (2011) 2046-2060

Different effect of localization methods (cont.)

Larger differences for smaller observation errors



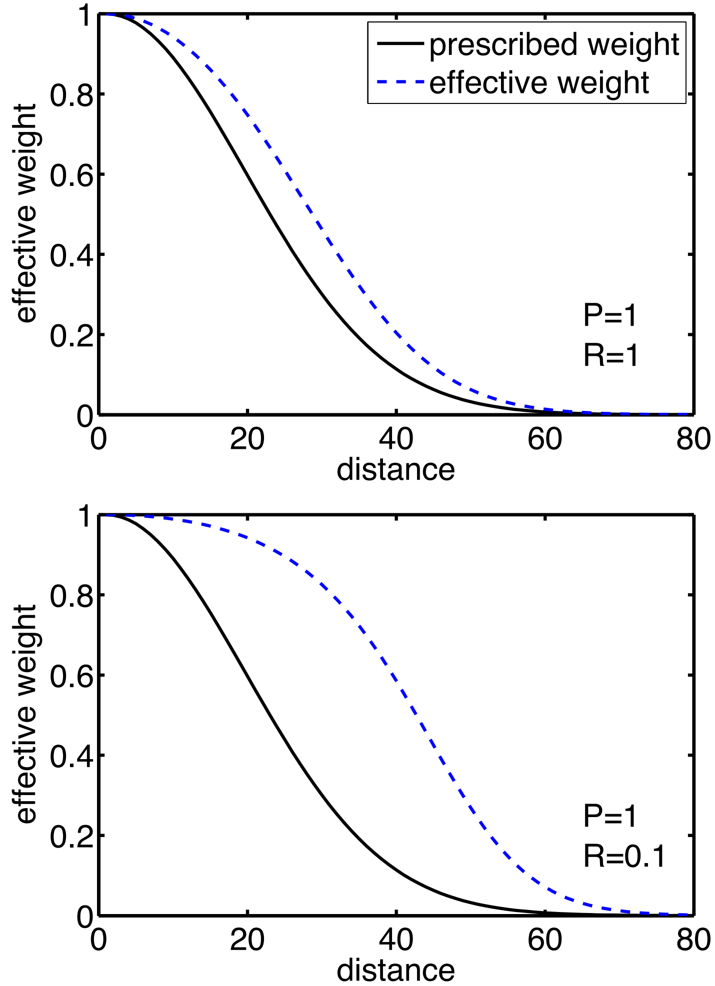
Covariance vs. Observation Localization

Some published findings:

- Both methods are "similar"
- Slightly smaller width required for observation localization

But note for observation localization:

- Effective localization length depends on errors of state and observations
  - ➢ Small observation error ➜ wide localization
  - ➢ Possibly problematic:
    - in initial transient phase of assimilation
    - if large state errors are estimated locally

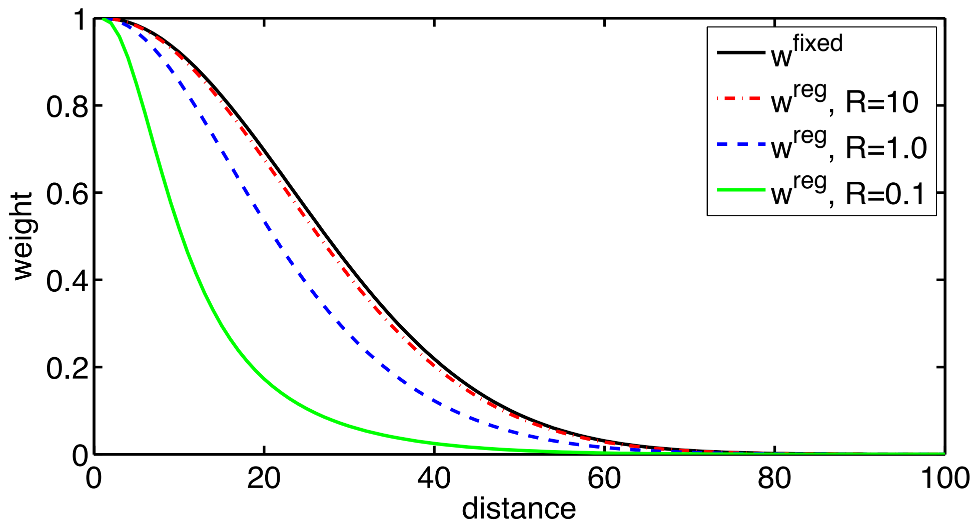


P: state error variance
R: observation error variance

Regulated Localization

➜ New localization function for observation localization

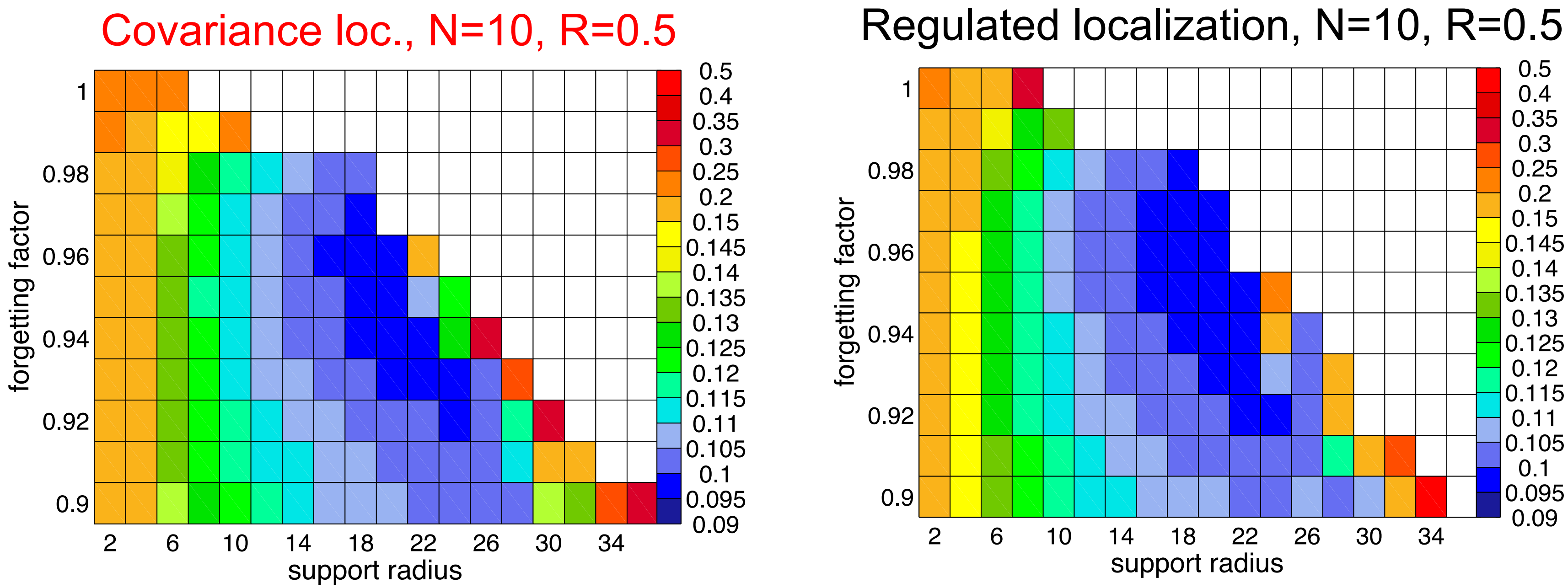
- formulated to keep effective length constant (exact for single observation)
- depends on state and observation errors
- depends on fixed localization function
- cheap to compute for each observation
- Only exact for single observation – works for multiple



P=1

L. Nerger et al. QJ Royal. Meterol. Soc. 138 (2012) 802-812

Lorenz96 Experiment: Regulated Localization



- Reduced minimum rms errors
- Increased stability region
- Still need to test in real application
- Description of effective localization length explains the findings of other studies!

Summary

- Ensemble-based KFs not exact
  - ➜ But they “work”!
- Improve methods
  - ➜ Least cost; least tuning; best state and error estimates
- Study relations for improvements
  - ➜ Efficient analysis formulations
  - ➜ Efficient localization

Thank you!