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
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

- 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)
- Observations: 10000 – 37000 (Surface temperature only)
- Ensemble size 8

S. Loza et al., Journal of Marine Systems 105 (2012) 152-162
Forecast deviation from satellite data

No assimilation

RMSE of SST forecast (without DA)

Assimilation

over 01.10.2007 - 30.09.2008

ensemble forecast (with LSEIK)

RMS

bias

Improvements also sub-surface and in other fields

Lars Nerger – Practical Aspects of Ensemble-based KFs
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 - $\mathcal{O}(10^7-10^9)$
• sparse observations
• non-linear

Current “standard” methods:
• Optimization algorithms (“4DVar”)
• Ensemble-based estimation algorithms

Lars Nerger – Practical Aspects of Ensemble-based KFs
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

Lars Nerger – Ensemble square-root KFs
Logical separation of assimilation system

- **Single program**
- **Filter**
  - Initialization
  - Analysis
  - Re-initialization
- **Model**
  - Initialization
  - Time integration
  - Post processing
- **Observations**
  - Obs. vector
  - Obs. operator
  - Obs. error

- **State**
- **Time**
- **Observations**
- **Mesh data**

Explicit interface
Indirect exchange (module/common)

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)
- RRSQRT
- ROEK
- SEEK
- SEIK

Studied in Nerger et al. (2005)

- EnKF(2003)
- EnKF(2004)
- EAKF
- EnSRF
- ETKF
- ESTKF

New study (Nerger et al. 2012)

- MLEF
- SPKF
- ESSE
- RHF

New filter formulation

**Model Equations**

**Stochastic dynamic model:**
\[
x_i^t = M_{i,i-1} x_{i-1}^t + \eta_i, \quad x_i^t, \eta_i \in \mathbb{R}^n
\]

**Stochastic observation model:**
\[
y_k = H_k x_k^t + \epsilon_k, \quad y_k, \epsilon_k \in \mathbb{R}^m
\]

**Assumptions:**
\[
\eta_i \sim \mathcal{N}(0, Q_i); \quad \mathbb{E}[\eta_i \eta_j^T] = Q_i \delta_{ij} \quad \text{Model error}
\]
\[
\epsilon_k \sim \mathcal{N}(0, R_k); \quad \mathbb{E}[\epsilon_k \epsilon_l^T] = R_k \delta_{kl} \quad \text{Observation error}
\]
\[
x_i^t \sim \mathcal{N}(\bar{x}^t_i, P_i)
\]
\[
\mathbb{E}[^T \eta_k \epsilon_k^T] = 0; \quad \mathbb{E}[\eta_i (x_i^t)^T] = 0; \quad \mathbb{E}[\epsilon_k (x_k^t)^T] = 0
\]
The Ensemble Kalman Filter (EnKF, Evensen 94)

Initialization:
Generate random ensemble \( \{ x_0^{a(l)}, l = 1, \ldots, N \} \)
Ensemble statistics approximate \( x_0^a \) and covariance \( P_0^a \)

Forecast:
\[ x_i^{a(l)} = M_{i,i-1} [ x_{i-1}^{a(l)} ] + \eta_i^{(l)} \]

Analysis:
\[ x_k^{a(l)} = x_k^{f(l)} + K_k \left( y_k^{(l)} - H_k x_k^{f(l)} \right) \]
\[ K_k = P_k^f H_k^T \left( H_k P_k^f H_k^T + R_k \right)^{-1} \]
\[ P_k^f := \frac{1}{N-1} \sum_{l=1}^{N} \left( x_k^{f(l)} - \bar{x}_k^f \right) \left( x_k^{f(l)} - \bar{x}_k^f \right)^T \]
\[ x_k^a := \frac{1}{N} \sum_{l=1}^{N} x_k^{a(l)} \]
Issues of the EnKF94

Monte Carlo Method

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

Inversion of large matrix $H_k P_k^f H_k^T + R_k \in \mathbb{R}^{m \times m}$
(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)

Lars Nerger – Practical Aspects of Ensemble-based KFs
Ensemble Transform Kalman Filter - ETKF

Ensemble perturbation matrix

\[ \mathbf{X}'_k := \mathbf{X}_k - \overline{\mathbf{X}}_k \]  

(size \( n \times N \))

Analysis covariance matrix

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

(size \( n \times n \))

“Transform matrix” (in ensemble space)

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

(size \( N \times N \))

Ensemble transformation

\[ \mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}^{ETKF} \]  

(size \( n \times N \))

Ensemble weight matrix

\[ \mathbf{W}^{ETKF} := \sqrt{N - 1} \mathbf{C} \mathbf{A} \]  

(size \( N \times N \))

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

Error-subspace basis matrix

\[ L := X^f T, \]

\( (T \text{ subtracts ensemble mean and removes last column}) \)

Analysis covariance matrix

\[ \tilde{P}^a = \tilde{L}\tilde{\Lambda}L^T \]

\( (n \times n) \)

“Transform matrix” (in error subspace)

\[ \tilde{A}^{-1} := (N - 1)T^TT + (HL)^TR^{-1}HL \]

\( (N-1 \times N-1) \)

Ensemble transformation

\[ X'^a = L \tilde{W}^{SEIK} \]

\( (n \times N) \)

Ensemble weight matrix

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

\( (N-1 \times N) \)

- \( \tilde{C} \) is square root of \( \tilde{A} \) (originally Cholesky decomposition)
- \( \Omega^T \) is transformation from \( N-1 \) to \( N \) (random or deterministic)
Weight Matrices ($W$ in $X^{a'} = X^f W$)

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

Lars Nerger – Practical Aspects of Ensemble-based KFs
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

SEIK–sym: Difference of transformation matrices

![Matrix diagram showing the difference in transformation matrices with a color scale from -4 to 4 x 10^{-3}](image)

(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!

Lars Nerger – Practical Aspects of Ensemble-based KFs
Revised $T$ matrix

Identical transformations require different projection matrix for SEIK:

$$L := X^f T.$$ 

For SEIK:

- $T$ subtracts ensemble mean and drops last column

→ Dependence on order of ensemble members!
→ Solution:
  → Redefine $T$: Distribute last member over first N-1 columns
  → Also replace $\Omega$ by new $\hat{T}$

New filter formulation:

Error Subspace Transform Kalman Filter (ESTKF)
T-matrix in SEIK and ESTKF

**SEIK:**

\[ T_{i,j} = \begin{cases} 
1 - \frac{1}{N} & \text{for } i = j, i < N \\
-\frac{1}{N} & \text{for } i \neq j, i < N \\
-\frac{1}{N} & \text{for } i = N 
\end{cases} \]

**ESTKF:**

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

- 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 $\mathbf{A}$
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

Localization Types

Covariance localization
- Modify covariances in forecast covariance matrix $P_f$
- Element-wise product with correlation matrix of compact support

Requires that $P_f$ is computed (not in ETKF or SEIK)


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

Possible in all filter formulations

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

Simplified analysis equation:

$$x^a = x^f + \frac{P_f}{P_f + R} (y - x^f)$$
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

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

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

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

Figure 1. Effective weight

Figure 2. Regulated Localization

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!

Lars Nerger – Practical Aspects of Ensemble-based KFs
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!