A Comparison of Data Assimilation with the Ensemble Kalman Filter and the SEEK Filter applied to non-linear Shallow Water Equations

Alfred Wegener Institute
for Polar an Marine Research

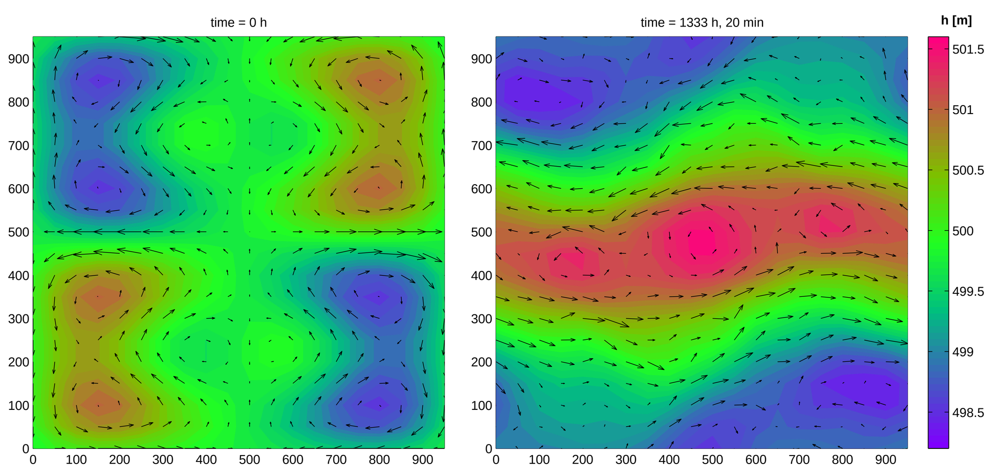
L. Nerger, W. Hiller, J. Schröter

Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
Contact: lnerger@awi-bremerhaven.de

The Problem

During the last years there has been an extensive development of stochastic filtering algorithms based on the Kalman filter intended for application to high-dimensional numerical models. Of those filters, we directly compare two widely used algorithms: The Ensemble Kalman filter (EnKF) and the Singular Evolutive Extended Kalman filter (SEEK). In addition we consider the Singular Evolutive Interpolated Kalman filter (SEIK), which can be regarded as an interpolated version of the SEEK algorithm or as an ensemble filter using a preconditioned ensemble.

The comparison focuses on the mathematical foundations of the algorithms and their numerical requirements as well as on their application to a model ocean. In twin experiments with synthetic observations of the surface elevation the assimilation behavior of the algorithms is assessed. The computational burden and filter performance depend strongly on the ensemble size and rank of the state covariance matrix. Hence the ensemble size and the rank are used as a parameter in the experiments.



Model state at time step 0 and after 40000 time steps.

The numerical models used for the filter experiments are shallow water equations with non-linear evolution.

The model domain is chosen as a box measuring 950 km per side, discretized by a grid of 20x20 points. Periodic boundary conditions are applied, and a constant Coriolis parameter and a flat bottom are assumed. The model is initialized in geostrophic balance and evolved with a time step of 2 minutes.

The assimilation experiments assume an exact model, thus model errors vanish.

Filter Algorithms

EnKF[a] The Ensemble Kalman Filter applies a Monte-Carlo method to solve the Fokker-Plank equation governing the evolution of the statistics of a stochastical model.

Initialization: Generate an ensemble of model states whose ensemble statistics approximate the prescribed initial state estimate and error covariance matrix.

Forecast: Evolve each of the ensemble member states with the full numerical model.

Analysis: When observations are available apply the update step of the Extended Kalman Filter with a state covariance matrix approximated by the ensemble statistics. Each of the forecasted ensemble states is analyzed using an observation from an observation ensemble which is generated. The error statistics are updated implicitely with the ensemble update.

SEEK[b] The Singular Evolutive Extended Kalman Filter is derived from the Extended Kalman Filter by approximating the state error covariance matrix by a matrix of reduced rank and evolving this matrix in decomposed form.

Initialization: Choose the initial estimate for the model state and an approximate state covariance matrix of low rank in the decomposed form $\mathbf{LUL}^\mathbf{T}$

Forecast: Evolve the guessed state with the full non-linear model and the column vectors $\mathbf{L_i}$ with the tangent-linear model.

Analysis: For analysis compute the updated state covariance matrix by an equation for the matrix $\mathbf{U}$ which relates the model state error to the observation error in the spirit of the Riccati equation. With this updated covariance matrix the state update is given by the analysis step of the Extended Kalman Filter.

Re-orthogonalization: To avoid successive alignment of the vectors $\mathbf{L_i}$, occasionally perform a re-orthogonalization of these vectors.

SEIK[c] The Singular Evolutive Interpolated Kalman Filter was originally derived from the SEEK algorithm. Alternatively it can be interpreted as a reduced-rank-preconditioned ensemble Kalman filter.

Initialization: In a process called minimum second order exact sampling, generate a state ensemble of minimum size whose ensemble statistics yield exactly the low-rank covariance matrix used in SEEK. For rank r the minimum ensemble size is $r+1$.

Forecast: Evolve each of the ensemble member states with the full numerical model.

Analysis: Perform the analysis analogous to the SEEK filter. As the matrix $\mathbf{U}$ has been used for the ensemble generation during the initialization step, a new $\mathbf{U}$ is computed which only implicitely relates the model error to the observation error.

Resampling: Resample the state ensemble to represent the updated error statistics of the model state.

[a] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res, **99 (C5)** (1994) 10143

[b] D. T. Pham, J. Verron, M. C. Roubaud, A singular evolutive extended Kalamn filter for data assimilation in oceanography, J. Mar. Sys., **16** (1998) 323

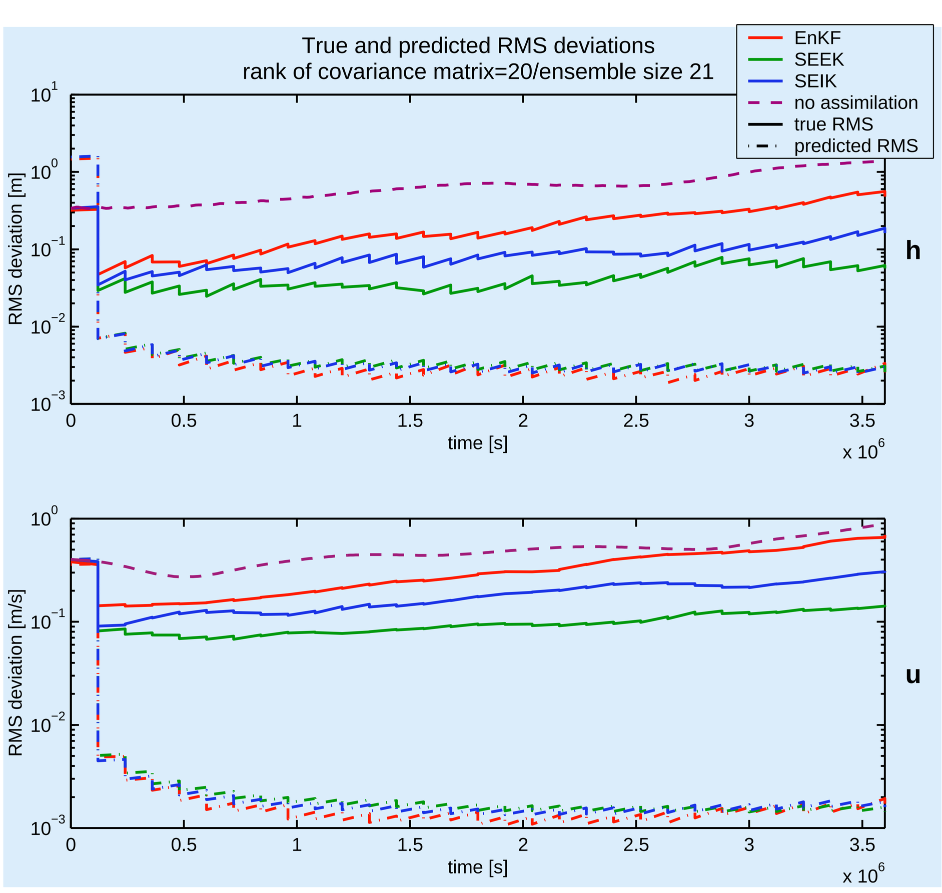
[c] D. T. Pham, J. Verron, L. Gourdeau, Filtres de Kalman singuliers évolutif pour l'assimilation de données en océanographie, C. R. Acad. Sci Terre Planètres, **326** (1998) 255

Assimilation Experiments

Configuration

For the experiments we generated synthetic observations from a model run by disturbing the surface elevation by normally distributed noise.

The initial state estimate was chosen as the mean state of a model run over 40000 time steps. The state covariance matrix was computed as the variation of this state sequence about the mean. By an incomplete eigendecomposition of this matrix, retaining only the largest eigenmodes, we generated the low-rank approximation for use with the SEEK and SEIK algorithms. The analysis step was performed after each 1000 time steps.

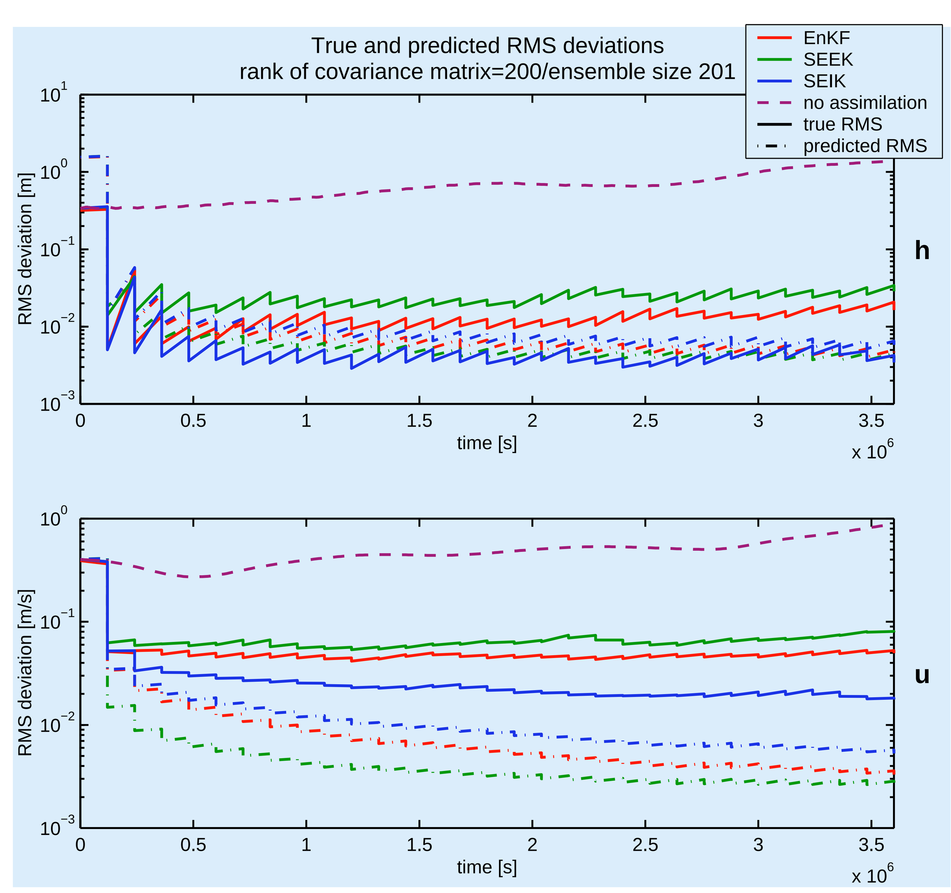


RMS deviations from the true state for surface elevation (h) and velocity x-component (u) for rank r=20. Shown are the true RMS deviation, that predicted by the filter, and the deviation without assimilation.

Filtering

To relate the filter performance to the computational burden, all three algorithms were configured in such a way that each algorithm required the same number of model evaluations. In addition, we implemented the algorithms to achieve minimum computing times.

For large ocean models the number of model evaluations is usually quite restricted due to limited computing power and time. Therefore we tested the filter performances for a small ensemble of size 21. Additionally, we performed experiments with a large ensemble of size 201, which is expected to provide a much better representation of the error statistics.



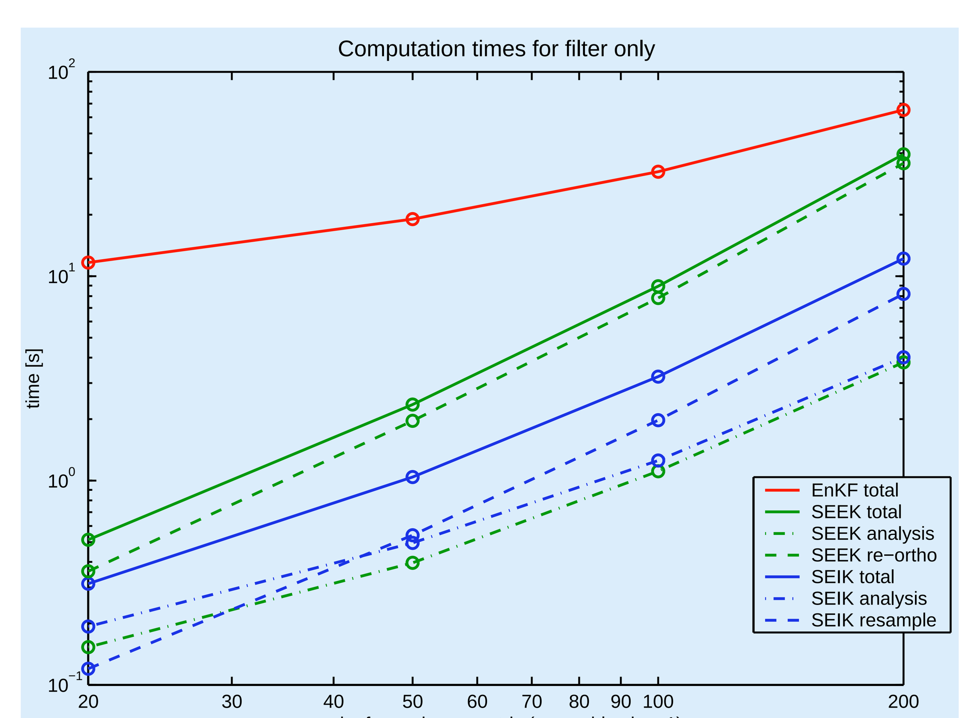
RMS deviations from the true state for surface elevation (h) and velocity x-component (u) for rank r=200. Shown are the true RMS deviation, that predicted by the filter and the deviation without assimilation.

Summary

For the experiments presented here the SEEK algorithm shows superior filter performance for the small ensemble and is comparatively fast. The SEIK filter is faster than the SEEK but yields better performance only for larger ensembles. The EnKF is expected to be fastest for very large ensembles. It shows a filter performance similar to that of the SEIK.

Computation Time

The model evaluations take more than 95% of the computing time. Since the number of model evaluations is equal for all three filters, we consider only the computing time of the filter.



Computation times for the filter algorithms. For the SEEK and SEIK algorithms, timings for analysis and re-orthogonalization/resampling step are also shown.