The SEEK Filter Revisited

L. Nerger, W. Hiller, J. Schröter

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

AWI
Alfred Wegener Institute
for Polar and Marine Research

Introduction

The Singular Evolutive Extended Kalman (SEEK) filter (Pham et al, 1998) is a low-rank approximation of the Extended Kalman Filter (EKF). Several successful applications of it have been reported in the literature.

This work reconsiders the SEEK algorithm with respect to its application to large-scale non-linear numerical models. The mathematical formulation and numerical requirements are compared with the widely used Ensemble Kalman Filter (EnKF, Evensen, 1994) and the less common Singular Evolutive Interpolated Kalman (SEIK) filter (Pham et al, 1998). The SEIK filter has been invented as an interpolated variant of the SEEK filter, but one can also interpret it as an ensemble filter using a preconditioned ensemble. The application of the three algorithms to a numerical model using the shallow water equations with non-linear evolution demonstrates the different abilities and the similarities of the filters.

The SEEK filter approximates the state covariance matrix used in the EKF by a matrix of low rank which is stored in decomposed form. The equations of the EKF are re-formulated to respect the decomposed form of the covariance matrix. A re-orthonormalization phase improves the numerical stability of the algorithm by constraining the modes of the covariance matrix.

The SEIK and EnKF filters not just approximate the EKF. They apply nonlinear ensemble forecasts which have the ability to better represent the prediction of the state covariance matrix and state estimate than the SEEK filter.

The major differences between the SEIK and the EnKF rely in the proposed initialization of the ensemble and in the analysis phase. Both filters apply the EKF analysis which assumes Gaussian error statistics, but the EnKF updates each single ensemble member while the SEIK updates the ensemble mean followed by a resampling of the ensemble.

Filter Algorithms

SEEK 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 decomposed form.

Forecast: Evolve the state estimate with the non-linear model and the modes of the covariance matrix with the tangent-linear model or a gradient approximation.

Analysis: Apply the update step of the Extended Kalman Filter (EKF) to the state forecast. The covariance matrix is approximated by the forecasted modes. It is updated by a relation derived from the Riccati equation.

Re-orthonormalization: Ooccasionally perform a re-orthonormalization of the modes of the covariance matrix to avoid successive alignment of these vectors.

EnKF The Ensemble Kalman Filter applies a Markov-Chain Monte-Carlo method to forecast the error statistics. A random ensemble is forecasted. In the analysis each single ensemble member is updated.

Initialization: Sample the initial error statistics given by the prescribed state estimate and error covariance matrix approximately by a stochstic ensemble of model states.

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

Analysis: Apply the EKF update step to each single ensemble member with an observation vector from an observation ensemble which has to be generated. The covariance matrix is approximated by the ensemble statistics. The error statistics are updated implicitely with the ensemble update. The state estimate is given by the ensemble mean.

SEIK The Singular Evolutive Interpolated Kalman Filter can be interpreted as a reduced-rank preconditioned ensemble Kalman filter. It is formulated to use a particular ensemble and applies an analysis analogous to the SEEK filter.

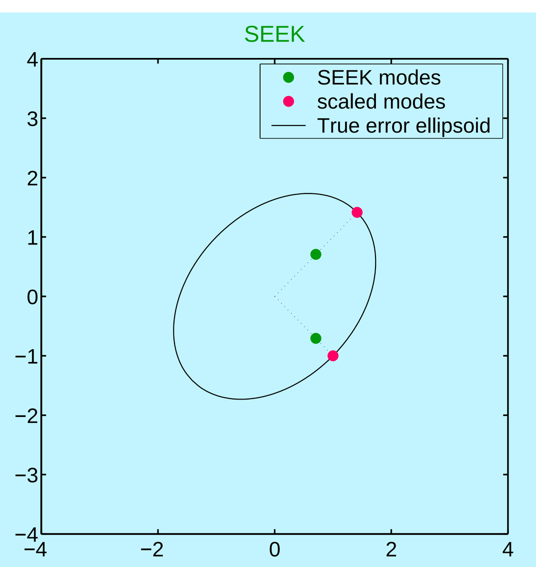
Initialization: Initialize as in the SEEK filter. Then, by a transformation of the modes, generate an ensemble of model states of minimum size which exactly represents the low rank covariance matrix.

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

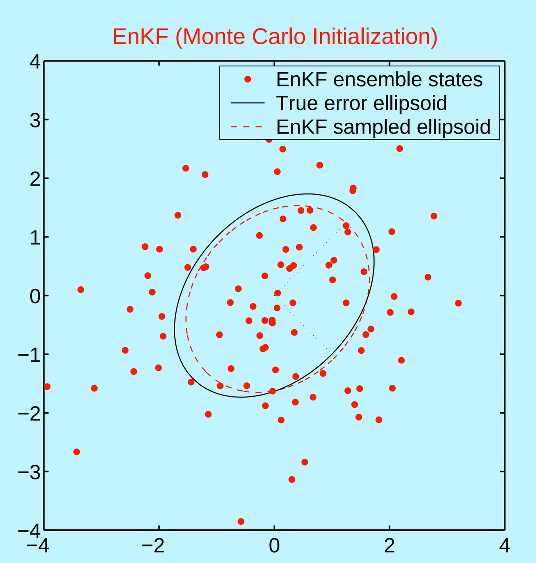
Analysis: Perform the analysis analogous to the SEEK filter. But here, apply the EKF update step to the ensemble mean. The covariance matrix is approximated by the forecasted ensemble. It is updated analogous to the SEEK analysis.

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

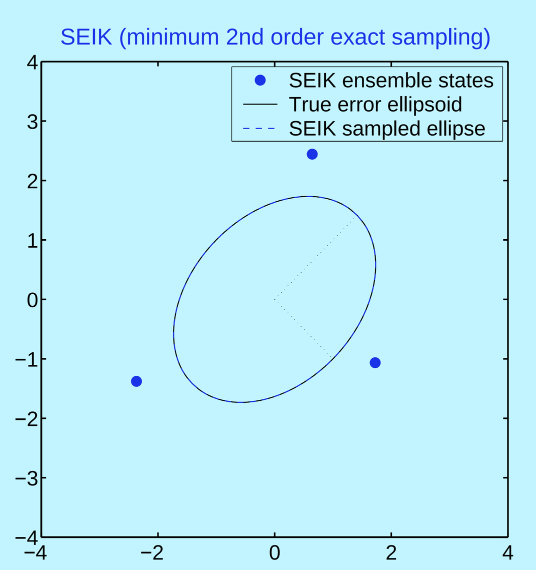
Different Initializations



Initialization of SEEK by modes of the low-rank approximated covariance matrix.



Initialization of EnKF with Monte Carlo sampling (ensemble of 100 states).



Initialization of SEIK with second order exact sampling of the low-rank approx. matrix.

All three filters differ in their initialization and approximation of the error statistics prescribed by the state estimate and state covariance matrix. We examplify here the initialization with a simple 3-dimensional example.

Consider a Gaussian probability density which is fully prescribed by the covariance matrix $\mathbf{P}$ and the mean state $\mathbf{x}$ given by

$$\mathbf{P} = \begin{pmatrix} 3.0 & 1.0 & 0 \\ 1.0 & 3.0 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This density can be visualized by an error ellipsoid prescribed by the eigenvectors and eigenvalues of $\mathbf{P}$. A low-rank approximation of rank 2 ($\mathbf{P}_2$) can be performed introducing only a small error due to the small third eigenvalue of $\mathbf{P}$. It is used by the SEEK and the second order exact sampling applied in SEIK.

The SEEK filter uses directly the modes of unit length of $\mathbf{P}_2$ to represent it. (Alternatively it is also possible to formulate SEEK to use modes scaled by the square roots of the eigenvalues, thus resembling the RRSQRT algorithm by Verlaan and Heemink (1995).) The modes are forecasted under the assumption that, without model error, they still represent the principal axes of the error ellipsoid.

The EnKF algorithm uses Monte Carlo sampling to generate an ensemble of random states which represents approximately the density given by $\{\mathbf{P}, \mathbf{x}\}$. No rank-reduction has to be performed, but the sampling converges rather slow. The forecast of the ensemble does not assume particular directions.

The SEIK filter applies a second order exact sampling to generate an ensemble of random states which exactly represents the low-rank approximation $\mathbf{P}_2$. Dependent on the eigenvalue spectrum of $\mathbf{P}$ a much smaller ensemble than in the EnKF is required to reach the same sampling error. The forecast is equivalent to that of the EnKF. (Acting on an error subspace SEIK is analogous to the ESSE concept introduced by Lermusiaux and Robinson (1999).)

Due to equivalent forecasts of EnKF and SEIK it is possible to use Monte Carlo initialization and second order exact sampling in both algorithms. Then their differences rely in the analysis and resampling stages.

Summary

⇒ The SEEK filter is a low-rank approximation of the EKF. It is numerically better suited for large scale problems, but it does not improve the abilities of the EKF to handle non-linearities.

⇒ Both, the EnKF and the SEIK filters show better abilities than the SEEK to treat non-linearities and are able to handle large scale problems.

⇒ Using preconditioned ensembles in EnKF (like generated with 2nd order exact sampling) can improve also the filter performance of EnKF.

⇒ The higher numerical complexity of the SEIK allows for data assimilation with smaller ensembles than the EnKF - at least for moderately nonlinear problems.

Filter Experiments

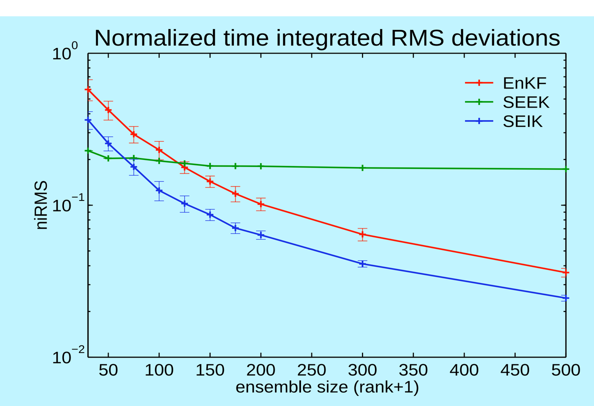
We performed data assimilation experiments with all three filter algorithms using shallow water equations with nonlinear evolution.

We initialized the state estimate with the mean state of a simulation over 8000 time steps (denoted the truth). The covariance matrix $\mathbf{P}$ was computed as the variation about the mean state. Further we generated synthetic observations of the surface elevation by adding Gaussian noise to the true states. These observations were assimilated each 200 time steps.

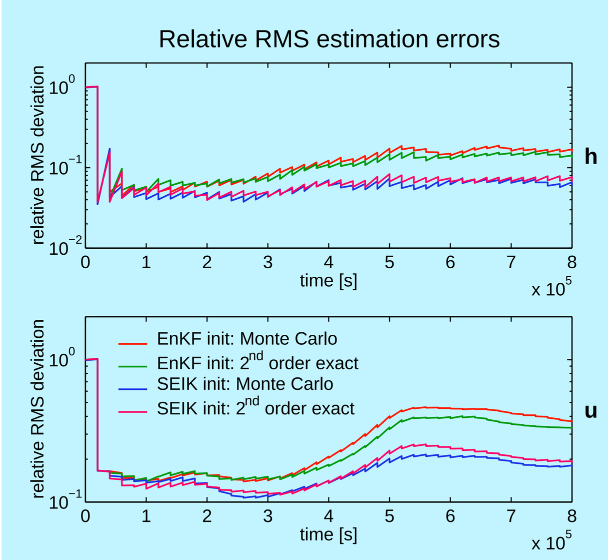
Under these equal conditions the three filters show quite different performances in estimating the true state depending on the ensemble size. The SEEK behaves distinct from the EnKF and SEIK filters which is due to the different forecast schemes. EnKF and SEIK converge quite similar with increasing ensemble size with the SEIK showing the better performance.

Comparing directly the EnKF and SEIK, it is evident that both filters initially yield almost the same estimation error but subsequently the performance of the EnKF deteriorates. This is due to noise introduced by the observation ensemble required for the analysis in EnKF.

Using second order exact sampling for the EnKF improves the filter performance slightly for an ensemble of size 100.



Mean time intergrated RMS deviations for the three filters normalized by the time integrated RMS deviation of a free run from the true state sequence. The means and standard deviations are computed over each 20 simulations with different random numbers in the initialization.



Relative RMS estimation errors for surface elevation (h) and zonal velocity (u) over time for ensembles of size 100 for EnKF and SEIK with different initializations.