The SEIK Filter - an Alternative to the Ensemble Kalman Filter!?

L. Nerger, W. Hiller, and 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 Ensemble Kalman Filter (EnKF), recently reformulated by its inventor G. Evensen, is one of the most used filter algorithms for data assimilation in meteorology and oceanography. The EnKF algorithm promises to provide good data assimilation results while being relatively simple to implement and to apply. On the other hand, the algorithm exhibits problems related to its computational cost for large-scale problems and approximations made by the EnKF. The comparison with the SEIK filter, introduced by D.T. Pham, shows that this alternative formulation of an ensemble based Kalman filter exhibits better properties with regard to computational costs and required approximations than the typical formulation of the EnKF. We discuss the differences between the two filter algorithms and advantages of each filter. The practical consequences of the different algorithmic formulations are shown using results from an application of both filter algorithms to the finite element model FEOM in a configuration for the North Atlantic.

Assimilation Experiments

The Finite-Element Ocean circulation Model (FEOM) developed at AWI is based on the primitive equations and serves to model ocean circulation on a temporal scale from years to decades. FEOM uses tetrahedral spatial discretization, backward Euler time stepping and approximates the model fields by linear functions on elements. The model is applied to simulate the North Atlantic circulation at eddy-permitting resolution (0.2^0–2^0). It relies on a horizontally refined mesh in regions of steep topography and allows the sloping bottom to be represented within the z-coordinate vertical discretization, similar to the so called shaved cell approach. The mesh consists of about 16000 surface nodes and 23 z-levels. In total there are 220000 3D nodes. The state dimension amounts to 925000.

We performed data assimilation experiments with the SEIK filter and the classical EnKF94 algorithm. In twin experiments synthetic observations of the sea surface height ζ are assimilated which have been generated by adding Gaussian noise of constant variance (σ = 5cm) to the sea surface height ζ of a model trajectory. The size of the observation vector amounts to about 16000.
To assess the filtering performance of the algorithms, data assimilation results are compared for the same ensemble size for both filters. Thus the same number of model integrations is computed for both filters.
The state covariance matrix is estimated by the covariance matrix of a 9-year model state state trajectory starting from January 1991 initialized from climatology. State ensembles are generated according to this covariance matrix either by Monte Carlo sampling (EnKF) or by second-order exact sampling (SEIK).

The initial state estimate was taken from a perpetual 1990 model spin-up. Analysis phases are performed at the initial time and in monthly intervals for three months. No explicit model error was simulated during the model forecasts. The analysis phase was performed with a forgetting factor of 0.8 for both filters.
The figures show filter results for the sea surface height ζ for the EnKF and SEIK filters for the initial time and after three months. The experiments are performed with an ensemble size of 32 members.
The experiments are configured such that 8 model integrations are performed concurrently. The filter update is computed in parallel with 8 processes each holding 4 members of the state ensemble.

Conclusions

⇒ If the classical EnKF94 is properly initialized, both the SEIK and EnKF94 algorithms can provide almost identical estimates.

⇒ The analysis schemes of the new EnKF04 and the SEIK filter are equivalent, apart from the observation ensemble used in EnKF04. In addition, the forecast phases are equivalent. Both algorithms can be initialized with the same ensemble.

⇒ SEIK is significantly faster than the classical EnKF94. Comparing the computational complexity of SEIK and the new EnKF04 algorithm, it is visible that the EnKF04 will require a computing time similar to SEIK. A remaining overhead of the EnKF04 will be the generation of the ensemble of observations.

EnKF and SEIK

EnKF

Flow of the classical Ensemble Kalman Filter EnKF94 (Evensen, 1994):

Initialization: Sample the initial error statistics given by the prescribed state estimate and error covariance matrix approximately by a stochastic 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 ensemble member with an observation vector from an observation ensemble. The covariance matrix is approximated by the ensemble statistics; the state estimate by the ensemble mean. The error statistics are updated implicitly with the ensemble update.

• The new EnKF algorithm (EnKF04; Evensen, 2004) uses a square root formulation. The flow of the EnKF04 is analogous to that of SEIK: The ensemble mean is updated in the analysis and the ensemble is subsequently transformed. For efficiency, EnKF04 still requires an ensemble-approximation of the observation error covariance matrix **R**. The influence of this approximation is unknown.

• Other ensemble square-root formulations (e.g. Anderson (2001), Bishop et al (2001), Whitaker and Hamill (2002)) are also analogous to the flow of SEIK. These filter use the inverse matrix $\mathbf{R}^{-1}$ in their equations.

• The classical EnKF94 and the EnKF04 directly use the matrix **R**. This can be advantageous, if **R** is changing very frequently.

SEIK

Flow of the Singular "Evolutive" Interpolated Kalman filter (Pham et al., 1998):

Initialization: Approximate covariance matrix by a low-rank matrix in the form $\mathbf{P} = \mathbf{V}\mathbf{U}\mathbf{V}^T$. By a transformation of the columns of **V**, 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: Apply the EKF update step to the ensemble mean and the "eigenvalue matrix" **U**. The covariance matrix is approximated by the forecasted ensemble.

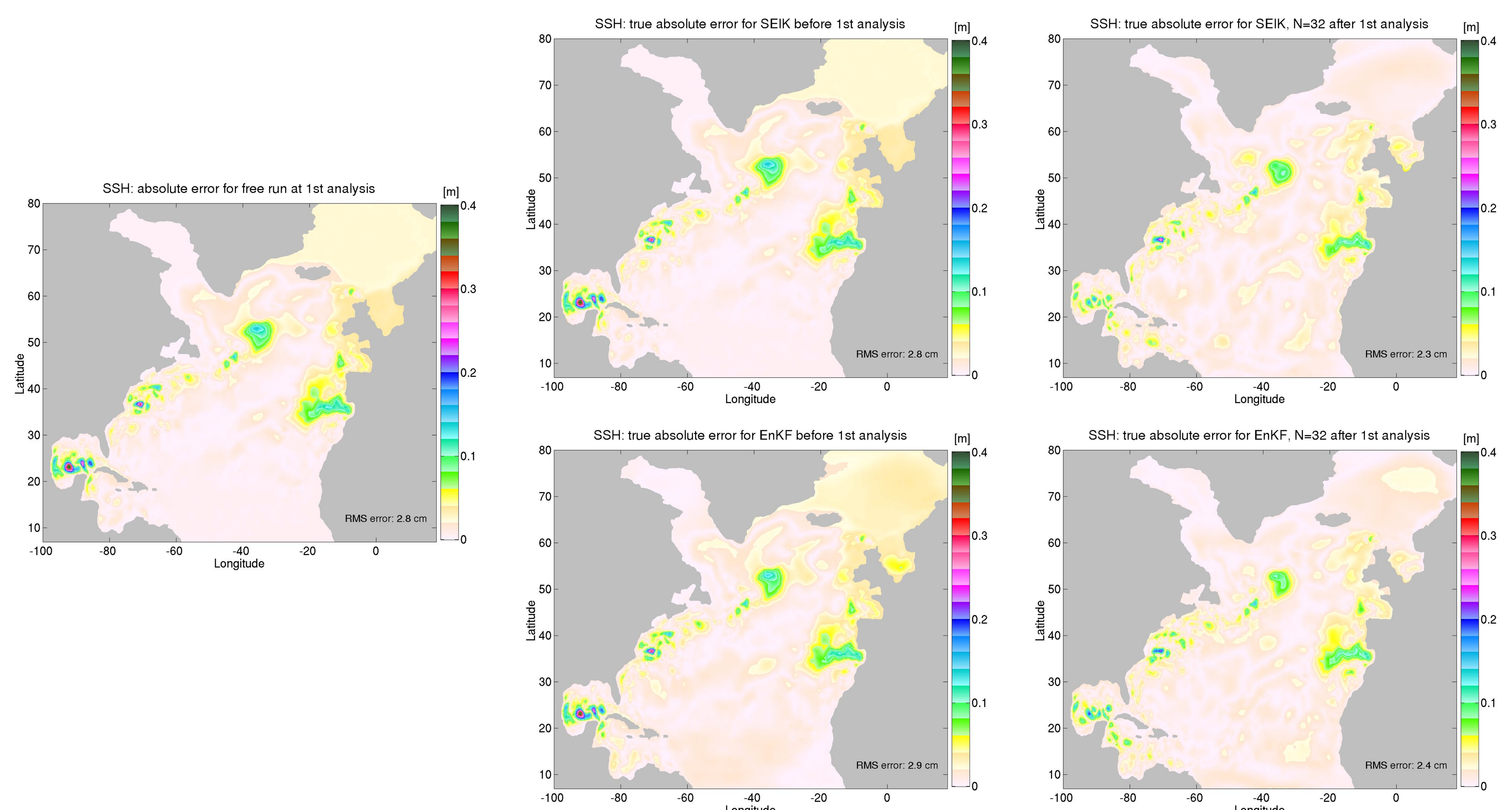
↓

Re-Initialization: Transform the forecasted state ensemble to represent exactly the updated error statistics of the model state.

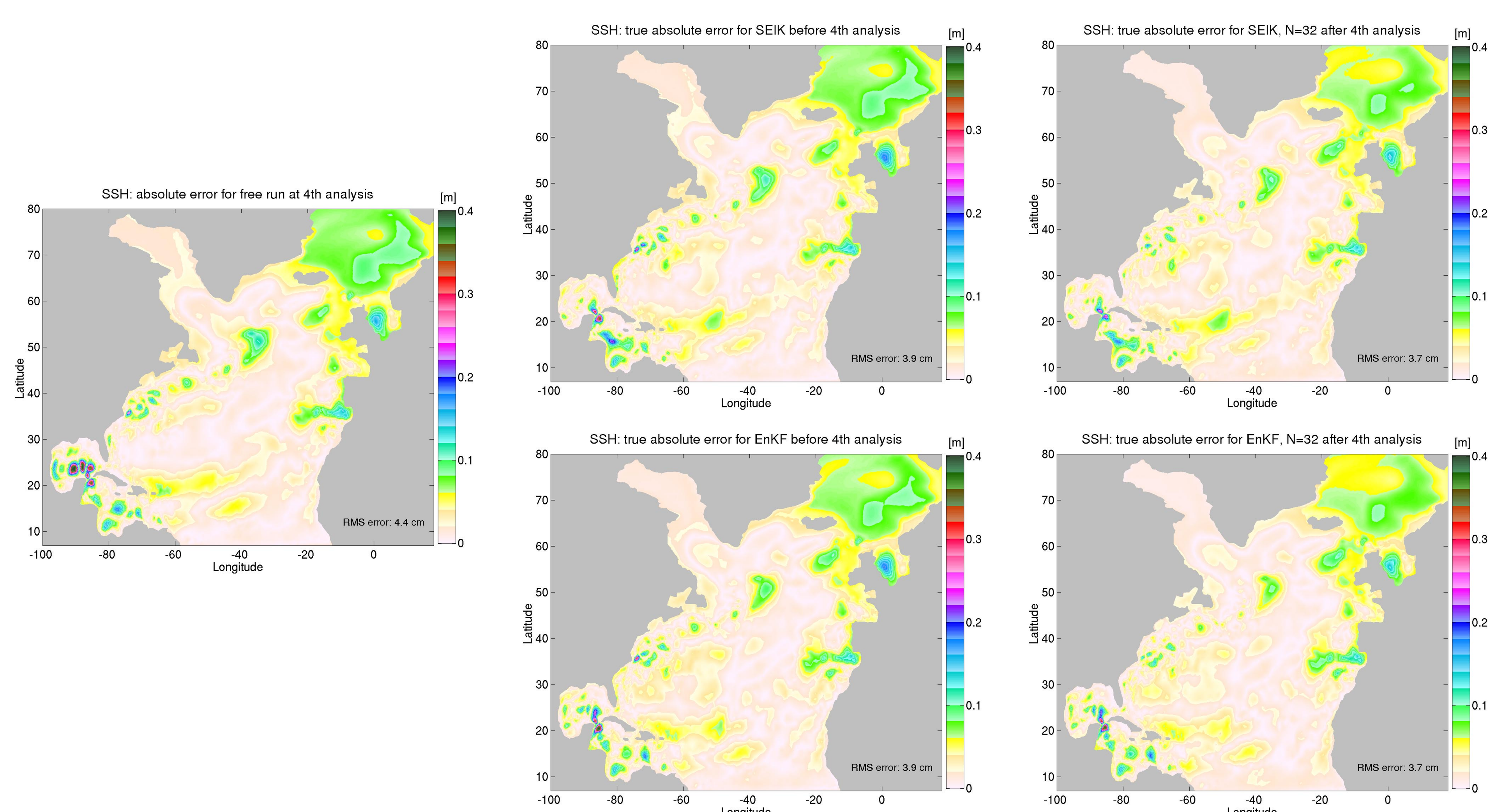
• SEIK does not assume that the observation error covariance matrix **R** is diagonal. It is efficient even with a full matrix **R**.

• Ensemble square-root formulations (e.g. Anderson (2001), Bishop et al (2001), Whitaker and Hamill (2002)) require a serial processing of observations or a diagonal observation error covariance matrix to be efficient.

• SEIK uses the inverse matrix $\mathbf{R}^{-1}$ in its equations. This can usually be computed off-line.



True absolute errors for the sea surface height at the first (initial) analysis time for ensembles of 32 members.



True absolute errors for the sea surface height at the end of three forecast/analysis cycles with ensembles of 32 members.