

Data Assimilation with the Ensemble Kalman Filter and the SEIK Filter applied to a Finite Element Model of the North Atlantic

L. Nerger* S. Danilov, G. Kivman, W. Hiller, and J. Schröter

Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

Abstract

Currently there are different approaches to filter algorithms based on the Kalman filter. One of the most used filter algorithms is the Ensemble Kalman Filter (EnKF). It uses a Monte Carlo approach to the filtering problem. Another approach is given by the Singular Evolutive Extended Kalman (SEEK) and Singular Evolutive Interpolated Kalman (SEIK) filters. These filters operate explicitly on a low-dimensional error space which is represented by an ensemble of model states. The EnKF and the SEIK filter have been implemented within a parallel data assimilation framework in the Finite Element Ocean Model FEOM. In order to compare the filter performances of the algorithms, several data assimilation experiments are performed. The filter algorithms have been applied with a model configuration of FEOM for the North Atlantic to assimilate the sea surface height in twin experiments. The dependence of the filter estimates on the represented error subspace is discussed. In the experiments the SEIK algorithm provides better estimates than the EnKF. Furthermore, the SEIK filter is much cheaper in terms of computing time.

Key words:

Data assimilation, Kalman filter, Ensemble Kalman filter, SEIK filter, North Atlantic

1 Introduction

Filter algorithms based on the Kalman filter (KF) are widely used to perform data assimilation with atmospheric and oceanographic problems. The aim of these algorithms is to combine the information provided by observations

* Corresponding author.

Email address: lnerger@awi-bremerhaven.de (L. Nerger).

with the state of a numerical model in order to estimate the real state of the geophysical system. The filters assimilate the available observational data in a sequential manner. In a forecast phase, the model is integrated up to the time when observations are available. At this time, a new model state is computed in an analysis phase on the basis of the predicted model state and the observations with weights computed from error estimates of both the observations and the model state estimate. Subsequently a new forecast phase is performed. The algorithms based on the KF share the virtue that they not only integrate the state estimate but also an estimate of the error in the state which is prescribed by the state error covariance matrix. The propagation of the error estimate by the model dynamics has the potential to provide more accurate error estimates than methods like optimal estimation which rely on static error estimates or empiric propagation of the errors. KF algorithms are of multivariate nature. In the case that observations of only one type of physical field are available, like satellite altimetry which is commonly used, other fields of the numerical model are updated in the analysis phase via cross-correlations contained in the error covariance matrix.

There are several advanced filter algorithms based on the KF which have been developed for data assimilation with large-scale numerical models. The Ensemble Kalman Filter (EnKF) (Evensen, 1994; Burgers et al., 1998), including several of its variants, is one of the most used filter algorithms. A review on applications of the EnKF can be found in (Evensen, 2003). The Singular Evolutive Extended Kalman (SEEK) filter (Pham et al., 1998a) and the Singular Evolutive Interpolated Kalman (SEIK) filter (Pham et al., 1998b) provide alternative approaches. These filters have been used in several applications, e.g. in the Tropical Pacific (Hoteit et al., 2002) or in the North Atlantic (Brusdal et al., 2003). In the latter article, the SEEK filter has been compared with the EnKF in a realistic configuration of an OGCM. Only 8 model state evaluations were necessary in the SEEK filter to obtain comparable estimation errors to the EnKF which operated with an ensemble size of 150. This points to the possibility that there are alternatives to the EnKF which might allow for data assimilation with much smaller computing times. Since data assimilation has very high demands on computing time, it is of highest interest to find the algorithm which provides the best filter results - in terms of the estimated model state - with the lowest required computing time. The result by Brusdal et al. (2003) is, however, difficult to interpret, since this study used slightly different model configurations and different initial conditions for the compared filters and the dependence of the filter results on the ensemble size has not been studied. A first quantitative comparison of the EnKF with the SEEK and SEIK filters with a simple shallow-water model has been performed by Nerger et al. (2005a). In this study the SEIK filter was superior to both the EnKF and the SEEK filter. First, the SEIK filter behaved more stable than the SEEK filter. Second, while also the EnKF behaves rather stable, it requires significantly more computing time than the SEIK filter to obtain the

same filter performance. This difference was due to fact that, for equal filter performance, the EnKF required the ensemble size to be between 1.5 and 2 times larger than the ensemble size required for the SEIK filter.

To get further insight into the relative filter performances of the EnKF and the SEIK filter, we compare the filter performance of the SEIK filter with that of the EnKF using a realistic configuration of the Finite Element Ocean Model FEOM (Danilov et al., 2005) for the North Atlantic. Hence, this study extends the idealized experiments of Nerger et al. (2005a). To allow for a quantitative comparison of the filter behavior, both filters are applied to the same model in an identical configuration. Further, the filter initialization is performed in equal manner for both algorithms. The filters are applied in twin experiments, assimilating synthetic observations of the sea surface height. The influence of the data assimilation on the estimate of the sea surface height is examined and the relation of the assimilation updates of the state estimate to the true errors and estimated errors is discussed.

2 Filter algorithms

The filter algorithms under consideration, the EnKF and the SEIK filter, are shortly discussed here. For a detailed description including all equations of the algorithms we refer to Nerger et al. (2005a). Both algorithms rely on the representation of the error statistics, i.e. the state estimate and the corresponding state covariance matrix, by an ensemble of model states. In the theory of the Kalman filter, the state covariance matrix represents the uncertainty in the state estimate. In data assimilation (DA) this matrix is usually interpreted as an error covariance matrix of the state. For this reason, we use the expression error covariance matrix in the sequel. The major differences between the EnKF and SEIK algorithms stem from the initialization of the state ensemble and the analysis phase of the filters as will be discussed below.

The Ensemble Kalman Filter (EnKF) (Evensen, 1994) applies a Monte-Carlo approach to the KF. The state estimate \mathbf{x}_0 of dimension n and the corresponding error covariance matrix \mathbf{P}_0 are sampled by an ensemble of N model state realizations $\{\mathbf{x}^{(i)}, i = 1, \dots, N\}$. Thus, the ensemble approximates the matrix \mathbf{P}_0 . Using the ensemble representation, the forecast phase is computed by integrating each ensemble state by the numerical model. The statistics of the forecasted ensemble, i.e. the ensemble mean and covariance matrix, represent the estimate of both the state and the covariance matrix at the observation time. In the analysis phase, the available observations are assimilated by updating each ensemble state according to the analysis equations of the KF. For consistency of the error estimate after the analysis, an ensemble of observations has to be generated for the analysis which represents the covariance

matrix of the observation errors (Burgers et al., 1998). The initial state ensemble and the observation ensemble can be generated by a transformation of independent random numbers. Let the state error covariance matrix of rank q be given in the form $\mathbf{P} = \mathbf{L}\mathbf{L}^T$ where the $N \times q$ matrix \mathbf{L} can be obtained, e.g., by a singular value decomposition of \mathbf{P} . Now an ensemble state with index i is given by $\mathbf{x}^{(i)} = \mathbf{x}_0 + \mathbf{L}\mathbf{b}^{(i)}$ where $\mathbf{b}^{(i)}$ is a random vector of dimension q filled by numbers from a normal distribution of zero mean and unit variance. Due to the finite ensemble size, the mean of the generated ensemble usually deviates from \mathbf{x}_0 . This deviation can be corrected, e.g., by shifting the ensemble to the correct mean. Note that this initialization technique is not a pure Monte-Carlo sampling, but represents the covariance matrix following the relative weights of its singular vectors.

The SEIK filter (Pham et al., 1998b) is equivalent with the EnKF in the forecast phase. Both filters integrate an ensemble of model states to propagate the state estimate as well as the error covariance matrix of the state. However, the initialization and analysis phases of the SEIK filter are distinct from those of the EnKF. The SEIK filter is based on a low-rank approximation of the initial error covariance matrix \mathbf{P}_0 . This is typically done by a truncated eigenvalue decomposition (or singular value decomposition) of \mathbf{P}_0 which only retains a small number r of leading eigen-values and corresponding eigen-vectors (or eigen-modes). With this \mathbf{P}_0 is approximated in a decomposed form as $\mathbf{V}\mathbf{U}\mathbf{V}^T \approx \mathbf{P}_0$ where \mathbf{U} is a diagonal $r \times r$ matrix which holds the leading eigenvalues. The matrix \mathbf{V} holds in its r columns the corresponding eigen-modes. For the forecast phase a random ensemble $\{\mathbf{x}^{(i)}, i = 1, \dots, N\}$ of minimum size $N = r + 1$ is generated which has the properties that it exactly represents the state estimate \mathbf{x}_0 and the approximated covariance matrix. This ensemble can be obtained by second-order exact sampling, see e.g. Pham (2001). Based on the fact that the EnKF and SEIK algorithms are in fact independent from the method used to generate the ensemble, see (Nerger, 2003), Evensen (2004) proposed to use a sampling scheme for the EnKF which is analogous to the scheme just described. If this scheme is used, the difference of the filters remains in the analysis phase. In the case of the SEIK filter, the analysis equations of the KF are applied to update the ensemble mean state and the matrix \mathbf{U} . The equations are formulated to treat the covariance matrix in the decomposed form $\mathbf{V}\mathbf{U}\mathbf{V}^T$. Subsequently to the analysis phase, the state ensemble is transformed in a re-initialization phase to represent the updated state estimate and the corresponding error covariance matrix.

Both filter algorithms are commonly used in conjunction with a so-called forgetting factor ρ (Pham et al., 1998a), ($0 < \rho \leq 1$). In the context of the EnKF, this technique is typically applied as an inflation factor, see e.g. (Anderson, 2001). The forgetting factor increases the error estimates in the error covariance matrix. This can stabilize the filter process as both the EnKF and SEIK algorithms are known to underestimate variances. For the EnKF, the

forgetting factor is used to inflate the ensemble around its mean by the inverse of the square-root of the given value ρ .

3 Experimental Setup

The data assimilation experiments are performed with an eddy-permitting configuration of the 3D finite-element primitive-equation ocean model FEOM (Danilov et al., 2004) for the North Atlantic. FEOM is based on the primitive equations using a splitting of the dynamic equations into barotropic and baroclinic subproblems. The discretization uses tetrahedrons which rely on a horizontally refined mesh in regions of steep topography. In the vertical a z-layer discretization is used. A sloping bottom is represented similarly to the shaved-cells approach used in finite-volume models. In the configuration used (Danilov et al., 2005), the mesh extends over the North Atlantic from 7°N to 80°N and from 18°E into the Caribbean. At the northern and southern boundaries closed boundaries for the velocities are considered. For temperature and salinity, relaxation to seasonal Levitus climatology is performed in buffer zones. The southern zone, in which relaxation is performed, occupies a 5° belt. The north-eastern buffer zone extends to 60°N. The mesh has a variable resolution which ranges from 1/15° to 2°. The strong variability of the resolution is visible in figure 1 which shows a part of the mesh at the surface. The high resolution near the coast lines allows for an accurate representation of the coasts. In total, the mesh contains approximately 16000 surface nodes and 220000 3D nodes discretized in 23 layers. The state dimension amounts to 925000. Results of the model integration for 1990-1998 under the realistic NCEP wind forcing and relaxation to the climatological temperature and salinity at the surface were presented and thoroughly discussed by Danilov et al. (2005). Here this model trajectory is used as the ocean “true” state for the twin experiments.

To assess the filter performance of both filter algorithms, twin experiments are performed starting in December 1992. Synthetic observations of the full sea surface height are assimilated at the initial time and in monthly intervals for 3 months. The observations are generated by adding uncorrelated Gaussian noise to the true model trajectory. For this, a standard deviation of 5cm is assumed for the observations as well as that the errors are uncorrelated. To initialize the filter algorithms, the initial state estimate has been chosen from a perpetual 1990 model spin-up run. The error covariance matrix is chosen to be represented implicitly by the variability of the 9-year “true” trajectory. This covariance matrix is dominated by a small number of large-scale modes. Accordingly, it can be very well approximated by a matrix of significantly lower rank. The dominance of the leading eigen-modes of the covariance matrix is visible in figure 2 which shows the relative truncation

error of the total variance of a low-rank approximation of the matrix as a function of the number of eigen-modes taken into account. This unexplained variance is given by one minus the sum of eigenvalues up to the number of the used eigen-modes divided by the total sum of eigenvalues of the covariance matrix. The first 7 eigen-modes explain already 86 % of the variance while 31 eigen-modes are sufficient to explain 97 %. To obtain consistent initializations of the EnKF and SEIK algorithms, the state ensembles are generated on the basis of the chosen initial state estimate and covariance matrix either by the transformation of independent random numbers (EnKF) or by second-order exact sampling (SEIK), as described in section 2.

In the experiments discussed below, ensembles with 8 and 32 members have been used. To stabilize the assimilation process, a forgetting factor of $\rho = 0.8$ was applied in both filter algorithms. Apart from this, no model error was simulated. The filter algorithms are implemented in the parallel data assimilation framework PDAF (Nerger et al., 2005b). The data assimilation system is configured to perform 8 model integrations in parallel. Each of the 8 model tasks was executed by 4 processors, thus 32 processors were used in total on an IBM pSeries 690 computer system.

4 Data Assimilation Experiments

To discuss the results of the DA experiments we focus on the behavior of the filter algorithms on the estimate of the SSH field at the initial time and at the end of the experiments, i.e. after three one-month integrations with analysis updates after each month. The discussion focuses on the general comparison rather than physical details of the filter results.

4.1 First Analysis Phase

The first analysis phase is performed at the initial model time, before computing any model integration. Hence, the update of the state estimate by the filter algorithms is determined by the initial state estimate and the estimated covariance matrix. This permits to compare the influence of the different methods used to initialize the state ensemble. The initial estimate of the sea surface height (SSH) for the SEIK filter with an ensemble size of 8 members is shown together with the true initial SSH field in figure 3. In small regions in the scale of a few Rossby radii, the estimated SSH deviates up to 40 cm from the true SSH field. The initial estimate of the SSH by the EnKF is very similar to that of the SEIK filter and hence not shown here. The differences due to the distinct sampling schemes used to initialize the EnKF and SEIK algorithms

do not exceed 9 cm.

For the first analysis phase with ensembles of 8 members, figure 4 compares the behavior of the SEIK (left hand side) and EnKF (right hand side) algorithms with respect to the SSH. In the upper panels the improvement of the estimate by the analysis update is shown. It is given by the difference of the absolute values of the true estimation error for the SSH from the estimated state before and after the analysis. Thus, an improvement of the estimate is indicated by positive values while negative values represent a deterioration of the estimate. In the middle and lower panels of figure 4, the true and estimated variance fields for the SSH are shown, respectively. In the first analysis phase, the SEIK algorithm improves the estimate of the sea surface height already with a state ensemble of size 8. The estimate of the SSH field is improved in wide regions as is visible from the upper-left panel of figure 4. The largest improvement is obtained near 40°W and 52°N . The estimate in the Norwegian and Greenland Seas is improved, too. At several locations, the state estimate is worse after the analysis compared to the initial estimate. This is due to the averaging character of the global filter analysis performed here. Dependent on the overall weights, the estimates do not need to be improved at every single grid point. For the EnKF, the maximal amplitude of the improvements of the state estimate is bigger. However, also the amplitude in the regions in which the state estimate deteriorates is bigger for the EnKF compared to the SEIK filter. Overall, the improvement-field appears to be noisier for the EnKF than for the SEIK filter.

The overall root mean square (RMS) error after the analysis is bigger for the EnKF than for the SEIK filter, despite the larger improvements by the EnKF. This is caused by the fact that the errors in the initial state estimate are bigger for the EnKF than for the SEIK filter. This is visible in the true variance fields shown in the middle panels of Fig. 4. As mentioned in section 2, the EnKF can also be initialized by second-order exact sampling as the SEIK filter. In this case, both filters provide the same state estimate after the analysis update. However, also for the distinct sampling schemes, the true variance fields exhibit a very similar spatial structure for both filters. For both algorithms, the highest deviations from the true SSH are of local scale of a few Rossby radii. They are located in the regions with the largest nonlinearities, i.e., in the Gulf Stream region and in the Gulf of Mexico. The estimated variance fields, which are displayed in the lower panels, are also very similar for both filter algorithms. While these fields show maximum values which are of the same order as the true variance fields, the estimated and true variance fields do not correspond in all regions. This inconsistency between the estimated and true variances is also typical for practical situations with real observations. It results in suboptimal analysis updates since a part of the true error space lies in the null-space of the low-dimensional error subspace estimated by a filter algorithm. An example of this effect can be seen for the SEIK filter in the Gulf of Mexico. Here the

filter algorithm estimates minimal errors. Thus, the existing deviation of the state estimate from the true state, which is caused by eddies in this region, cannot be reduced in the filter analysis phase.

If the ensemble size is increased to 32 members, the filtering performance improves. The behavior of the EnKF and SEIK filters for $N=32$ is shown in figure 5 analogous to the case of $N=8$. The improvement of the SSH field by the analysis update of the filters, shown in the upper panels, is now much more similar than for $N=8$. This corresponds to the increased similarity of the state estimates, visible from the true variance fields (middle panels) as well as the estimated variances (lower panels). As in the case of $N=8$, the state estimate of the SEIK filter is slightly better than for the EnKF. The RMS estimation error for the SSH after the analysis update is reduced to about 85% for the EnKF and 82% for the SEIK filter compared to a free evolution of the initial state estimate. Comparing the results from the two different ensemble sizes, shown in figures 4 and 5, it is visible that the character of the changes in dependence on the ensemble size is different for the two filters. For the SEIK filter the amplitude of the corrections has increased. In addition, the larger ensemble size results in additional improvements on the local scale. While for $N=8$ no significant corrections are obtained in the Gulf Stream region and in the Gulf of Mexico, the largest improvements are found in these regions for $N=32$. For $N=32$ this is also the case for the EnKF. However, for the EnKF it is visible that the increase of the ensemble size results in a reduction of the noisiness of the improvement. Even for $N=8$ the SSH estimate is corrected in the Gulf Stream and the Gulf of Mexico, but for the larger ensemble the areas in which the estimation error is increased by the analysis are reduced. This distinct behavior of the two filters in dependence on the ensemble size is due to the different sampling schemes used in the EnKF and SEIK algorithms. For the SEIK filter the behavior is caused by the exact sampling of a low-rank approximated covariance matrix. As described in section 2, the approximation is done by a truncated eigenvalue-decomposition of the prescribed covariance matrix. Thus only the leading eigen-modes are represented by a small ensemble. The leading eigen-modes, which have a rather coarse structure, represent the major part of the model variance. The higher modes add more details to the covariance structure, while their influence on the total variance estimate is small. Thus, a small ensemble in the SEIK filter represents the coarse modes of high variance while an increasing ensemble size provides a successive representation of the variability of finer scales. In contrast to this, the sampling scheme of the EnKF takes into account all modes for any ensemble size. The modes are, however, weighted by the variance they represent. Thus a small ensemble will represent the coarse structures while showing significant sampling errors on all spatial scales. For larger ensembles, the sampling errors are reduced which also improves the representation of fine-scale structures. The reduction of the sampling errors results in diminishing noise in the estimates of the state and covariance matrix.

4.2 Fourth Analysis phase

The fourth analysis phase takes place at the end of the third cycle of alternating one-month forecasts and analysis updates. Figure 6 shows the estimated variance for the SEIK filter with $N=32$ at this time. The forecasting of the ensemble, which represents this variance field, leads to a strong change of the spatial structure compared to the initial estimate. The ensemble spread is damped near 40°W and 52°N resulting in diminishing variance estimates in the course of the forecast. On the other hand, the nonlinear ensemble forecast leads to a larger ensemble spread in small regions with strong nonlinearity. Accordingly, the major variance estimates are focused to small regions where the ensemble spread has increased strongly after each single month of integration. The error estimate considers the model state to be quite accurate in all places except the North Sea as well as the Gulf Stream region and small areas in the Gulf of Mexico where the variance estimate is large. These high-variance regions dominate the analysis update and correspond to the regions in which the filter analysis step changes - and in most cases improves - the estimated sea surface height.

To compare the behavior of the EnKF and SEIK filters at the 4th analysis update, we focus on the region with the highest estimated variances which is the same region as shown in figure 1. Figure 7 compares the behavior of the EnKF and SEIK filter for $N=32$ at the 4th analysis update. The behavior for the case $N=8$ is similar and hence not shown here. In general, the improvement of the state estimate at the 4th analysis update is very similar for both filter algorithms. The magnitude of the corrections is of the same order as for the first analysis phase. The correspondence between the estimated and true variance is better for the fourth analysis than for the first one. Thus, the nonlinear evolution of the ensemble states provides a rather realistic estimate of the spatial structure of the true errors. However, the estimated variances are much smaller than the true ones. This is caused by the rather small ensemble size as well as the inconsistency between the estimated and true variances at the first analysis update. The inconsistency leads to an underestimation of the variances directly after the first analysis phase. The forgetting factor applied in the experiments, which amplifies the estimated variances, reduces the underestimation only to a small degree. Compared to a free evolution of the initial state estimate, the RMS estimation errors of the SSH after the fourth analysis phase are 15% smaller for the EnKF and 16.4% smaller for the SEIK filter. The experiments show that both filters are able to improve the estimate of the sea surface height on a fine spatial scale already for an ensemble of 32 members.

4.3 Computing Times

Overall, the experiments resulted in state estimates of similar quality for the EnKF and the SEIK filters. However, the results of the SEIK filter still appear to be slightly better, but not to that extent which was found by Nerger et al. (2005a) for a shallow water model. A major difference between the EnKF and SEIK algorithms remains in the computing time required to evaluate the analysis step. Table 1 shows the computing times of a single ensemble integration of one month length and the times required for one filter update phase with the EnKF and SEIK filters. The SEIK filter took only about 2 seconds for a single analysis update with ensemble size 8. In contrast, the EnKF algorithm lasted on average 1640 seconds. This time should be related to the computing time of about 7000 seconds for the ensemble integration over one month, i.e., the analysis phase of the EnKF for $N=8$ took almost one fourth of the time required for the model integration between two successive analysis phase in the case of eight concurrent model integrations. For the ensemble of size 32, the time for the filter update of the EnKF is reduced to about 8% of the integration time, since the number of concurrent model integrations is not increased. For the SEIK filter, the time for the analysis and re-initialization phase is negligible to the integration time.

The strong difference in the computation time between the two filter algorithms is due to the fact that the SEIK filter explicitly takes into account the low-dimensional error sub-space which is spanned by the state ensemble. Since the correction of the state estimate can only lie within the error sub-space, it is most economical to account explicitly for it by limiting the expected number of degrees of freedom to the dimension of the error-subspace which is ensemble size minus one. The EnKF algorithm, in its original form as used here, is formulated for a number of degrees of freedom given by the dimension of the assimilated observation vector. Due to the large difference between the ensem-

Table 1

Average computing time of different parts of the assimilation program in seconds for ensembles of size $N=8$ and $N=32$. The experiments are performed using 32 processors of an IBM pSeries 690 computer. 8 model tasks are executed concurrently. The integration time is to compute the full ensemble forecast of one month. For the SEIK filter, the filter update time includes the time for the analysis and re-initialization phases. Shown is the average time for a single update phase.

Filter	N	Integration (s)	Filter update (s)
SEIK	8	7000	2
EnKF	8	7000	1640
SEIK	32	28000	5
EnKF	32	28000	2360

ble size of 8 or 32 members to about 16000 observations, the EnKF algorithm can be expected to take much more time than the SEIK filter. Despite this formulation of the EnKF, the correction to the state estimate can lie only in the error sub-space represented by the state ensemble also for the EnKF. Based on this fact, Evensen (2004), developed a new variant of the ensemble Kalman filter, which hasn't been published at the time our experiments have been performed. The new formulation explicitly accounts for the error sub-space by projecting the observation error covariance matrix onto this space. Accordingly, the computing time can be expected to be significantly reduced. To achieve this smaller computing time in the EnKF, it is necessary to approximate the observation error covariance matrix by a low-rank ensemble. To which extent this approximation might influence the filter estimates is yet unknown.

5 Conclusion

The EnKF and SEIK filter algorithms have been applied to an eddy-permitting configuration of the finite element ocean model FEOM for the North Atlantic. Under identical conditions, the data assimilation results are found to be very similar for both filter algorithms for a moderate ensemble size of 32 members. The filtering performance is slightly better for the SEIK filter than for the EnKF algorithm. In addition, the EnKF requires significantly more computing time to evaluate the analysis update than the SEIK filter. This difference is caused by the distinct algorithmic formulations of the filters. Since the SEIK filter directly accounts for the low-dimensional space spanned by the state ensemble, it is superior from the theoretical point of view.

The data assimilation process leads to an improvement of the estimate of the sea surface height field. Due to the inconsistency of the initial estimate of the error covariance matrix with the true estimation errors, which will also be the typical situation in realistic applications, the variances are underestimated by the filter algorithms already after the first analysis phase. The forgetting factor applied in the experiments can only partially compensate for this effect. In addition, there remains a null-space of errors which are not recognized by the filter algorithms. For these reasons, a more realistic representation of model errors, which further reduces the underestimation of the variances and leads to error estimates also in the null-space of the covariance matrix, can be expected to enhance the filter performance.

The experiments also show that the error in the estimate of the sea surface height can be increased by the analysis update at some locations. This is due to the globally averaging character of the analysis equations of the EnKF and SEIK algorithms. Here, algorithmic developments beyond the global analysis

and the tuning of model errors are required to obtain better filter estimates. One possibility is the localization of the analysis. If the analysis update of the state estimate at some location is computed using only observations within some influence radius around this location, the analysis equations would be no longer global, see e.g. (Houtekamer and Mitchell, 2001). This approximation increases the number of degrees of freedom for the analysis update, see e.g. (Evensen, 2003), and will provide a smaller estimation error during the analysis step.

For a better general ability of the filters to provide improved state estimates, an adaptive modification of the error subspace would be useful. An approach to this was discussed by Brasseur et al. (1999) where the remaining residual between the state estimate and the observations was inverted and subsequently used as a new direction of the error subspace. In principle, this adaptivity can account for the null-space of the estimate error which is not represented by the initial state ensemble. However, the inversion of the state-observation residual is not well defined.

Acknowledgments

We thank three anonymous reviewers for their comments which helped to improve the manuscript. The computations have been performed on the IBM pSeries 690 computer of the HLRN network under the project hbk00010.

References

- Brasseur, P., Ballabrera-Poy, J., and Verron, J. 1999. Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. *J. Mar. Syst.* **22**, 269-294.
- Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., and Verron, J. 2003. A demonstration of ensemble based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* **40-41**, 253-289.
- Burgers, G., van Leeuwen, P. J., and Evensen, G. 1998. On the analysis scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.* **126**, 1719-1724.
- Danilov, S., Kivman, G. and Schröter, J. 2004. A finite-element ocean model: Principles and evaluation. *Ocean Mod.*, **6**, 125-150
- Danilov, S., Kivman, G. and Schröter, J. 2005. Evaluation of an eddy-

- permitting finite-element ocean model in the North Atlantic. *Ocean Mod.*, **10**, 35-49.
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99(C5)**, 10143–10162.
- Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.
- Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* **54**, 539–560.
- Anderson, J. L. 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.* **129**, 2884-2903.
- Houtekamer, P. L. and Mitchell H. L. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123-137.
- Hoteit, I., Pham D.-T. and Blum, J. 2002. A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *J. Mar. Syst.*, **36**, 101-127.
- Nerger, L. 2003. Parallel filter algorithms for data assimilation in oceanography. Ph.D. thesis, University of Bremen, Germany, 189pp.
- Nerger, L., Hiller, W., and Schröter, J. 2005a. A comparison of error subspace Kalman filters. *Tellus*, in press.
- Nerger, L., Hiller, W., and Schröter, J. 2005b. PDAF - The Parallel Data Assimilation Framework: Experiences with Kalman filtering. In *Proceedings of the 11. ECMWF Workshop on Use of High Performance Computing in Meteorology*, Reading, UK, October 25-29, submitted.
- Pham, D. T., Verron, J., and Roubaud, M. C. 1998a. A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* **16**, 323–340.
- Pham, D. T., Verron, J., and Gourdeau, L. 1998b. Singular evolutive Kalman filters for data assimilation in oceanography. *C. R. Acad. Sci., Ser. II* **326(4)**, 255–260.
- Pham, D. T. 2001. Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Mon. Wea. Rev.* **129**, 1194-1207.

List of Figures

- 1 Part of the surface mesh of FEOM. Visible is the unstructured character of the mesh in combination with the strongly varying resolution. 15
- 2 Relative truncation error of the total variance of a low-rank approximation of the of the state covariance matrix. The first 7 eigen-modes already explain 86 % of the variance. 97 % of the variance are explained by the first 31 eigenvalues. 16
- 3 Initial estimate of sea surface height (SSH) for the SEIK filter with ensemble size $N = 8$ (left) and true initial SSH field (right). 17
- 4 Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the first analysis time for an ensemble size 8. Shown is the improvement or error reduction (top), the true variance field (center), and the estimate variance field (bottom). 18
- 5 Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the first analysis time for an ensemble size 32. Shown is the improvement or error reduction (top), the true variance field (center), and the estimate variance field (bottom). 19
- 6 SEIK-estimated variance of the sea surface height before the fourth analysis update with $N=32$. 20
- 7 Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the fourth analysis time for an ensemble size 32 in the region with the largest variances. Shown is the improvement or error reduction (top), the true variance field (center), and the estimate variance field (bottom). 21

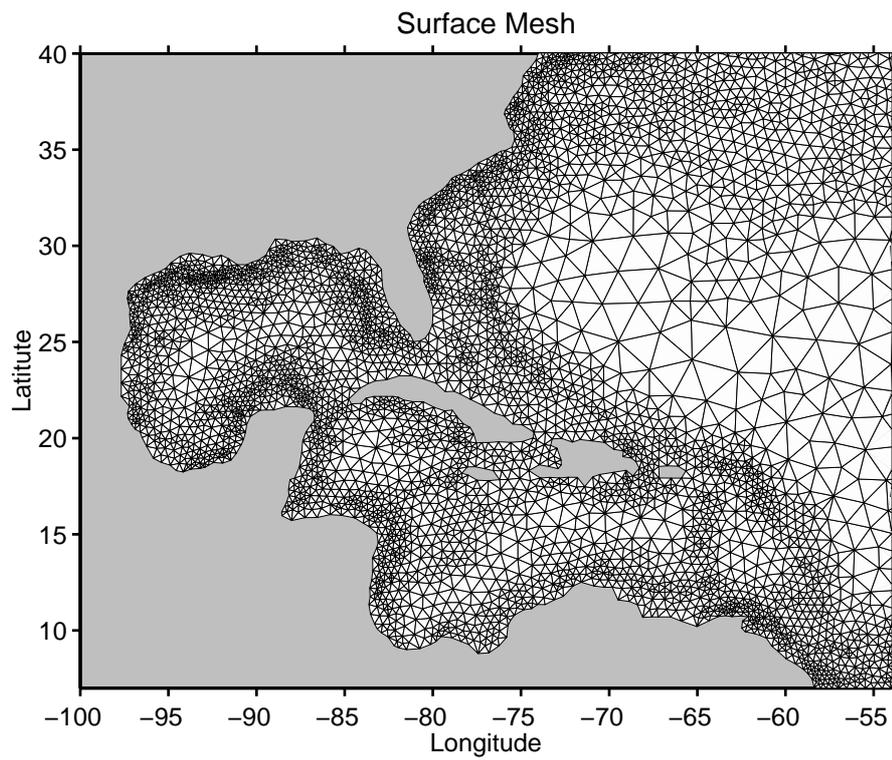


Fig. 1. Part of the surface mesh of FEOM. Visible is the unstructured character of the mesh in combination with the strongly varying resolution.

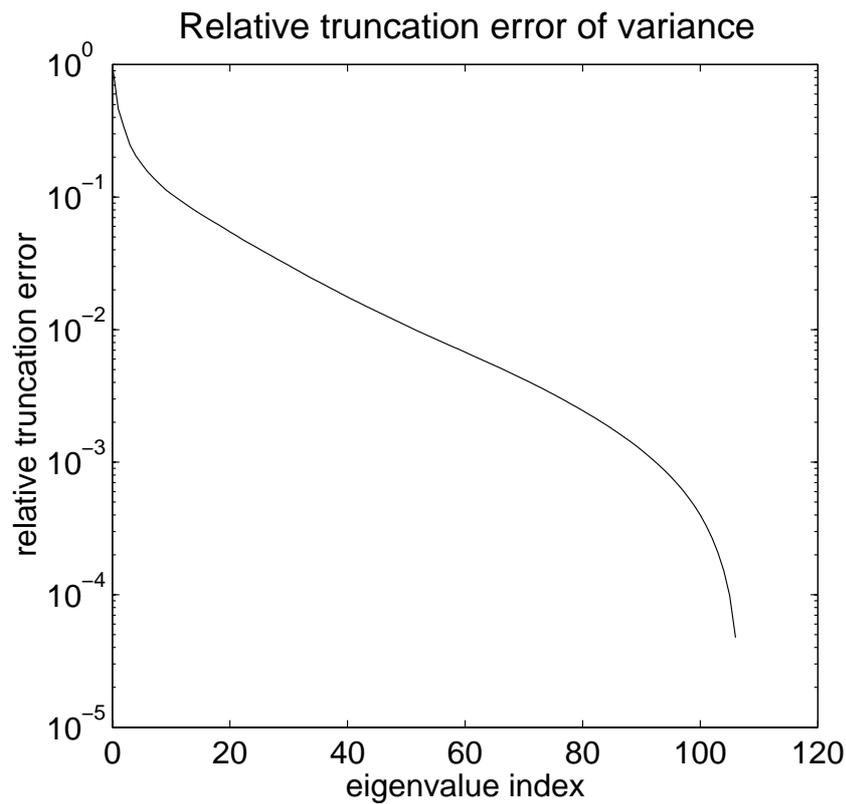


Fig. 2. Relative truncation error of the total variance of a low-rank approximation of the of the state covariance matrix. The first 7 eigen-modes already explain 86 % of the variance. 97 % of the variance are explained by the first 31 eigenvalues.

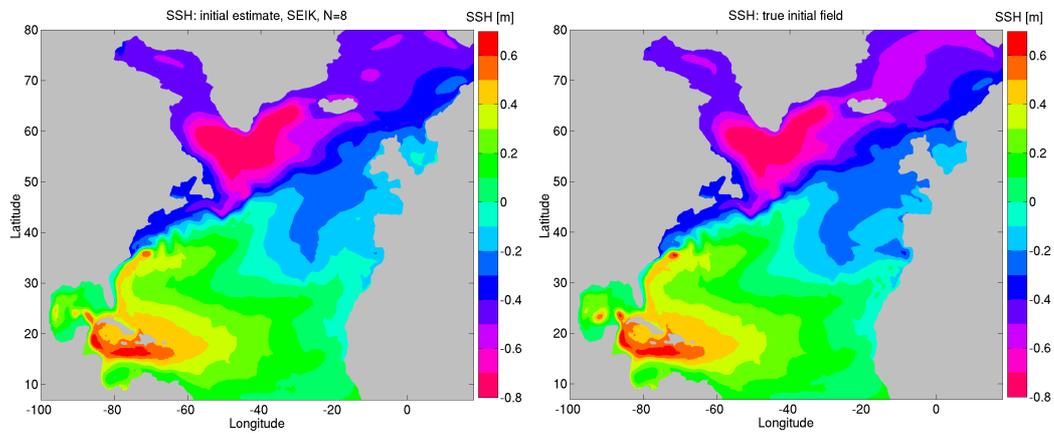


Fig. 3. Initial estimate of sea surface height (SSH) for the SEIK filter with ensemble size $N = 8$ (left) and true initial SSH field (right).

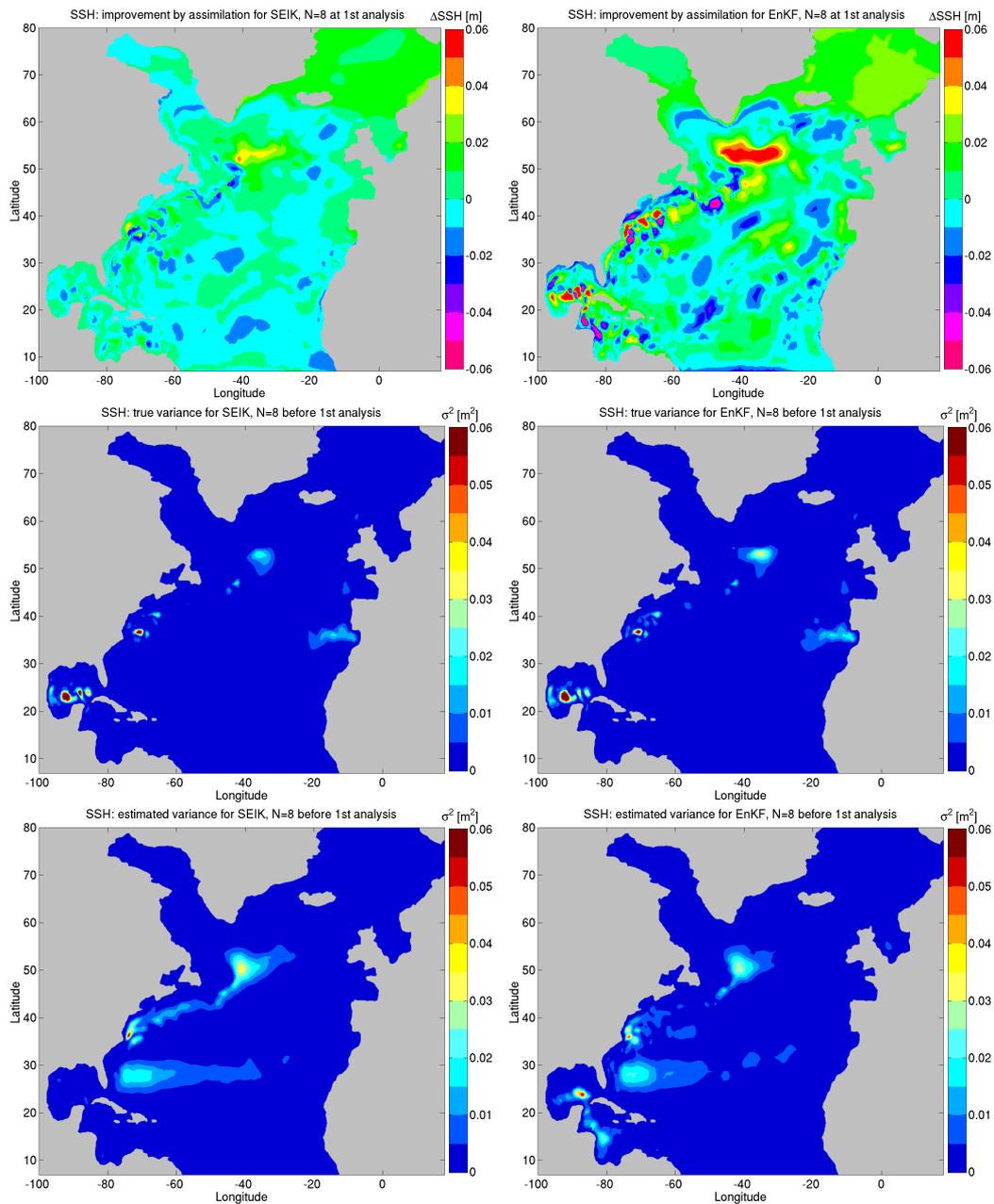


Fig. 4. Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the first analysis time for an ensemble size 8. Shown is the improvement or error reduction (top), the true variance field (center), and the estimated variance field (bottom).

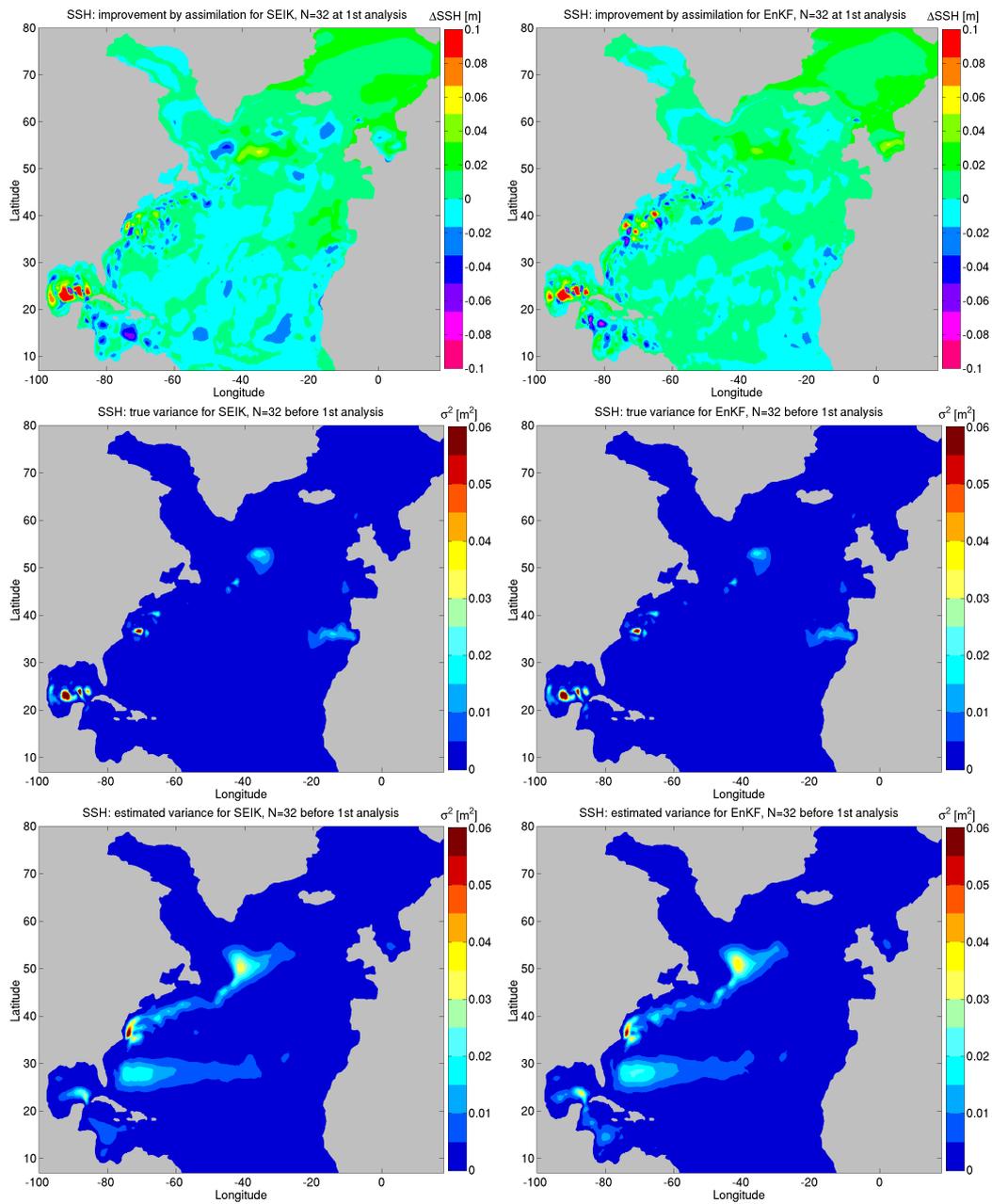


Fig. 5. Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the first analysis time for an ensemble size 32. Shown is the improvement or error reduction (top), the true variance field (center), and the estimated variance field (bottom).

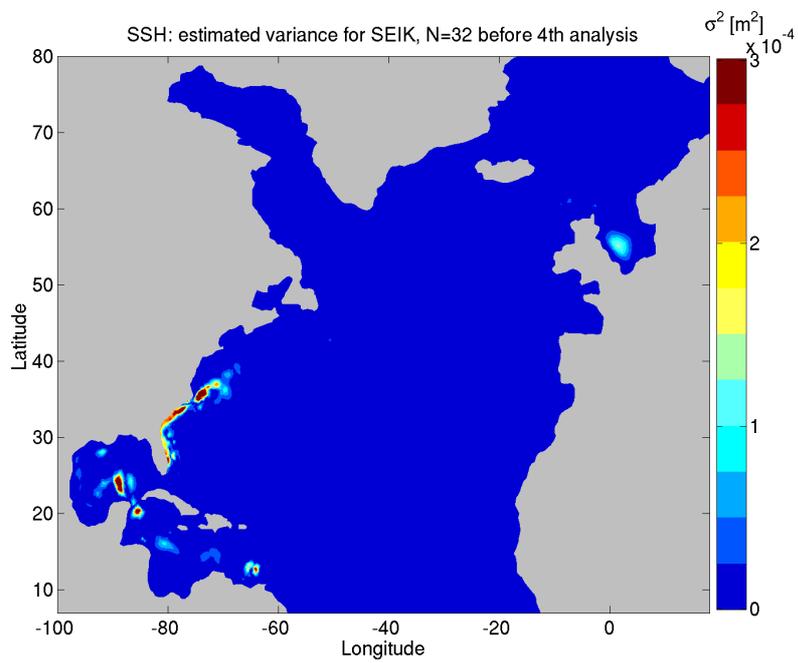


Fig. 6. SEIK-estimated variance of the sea surface height before the fourth analysis update with N=32.

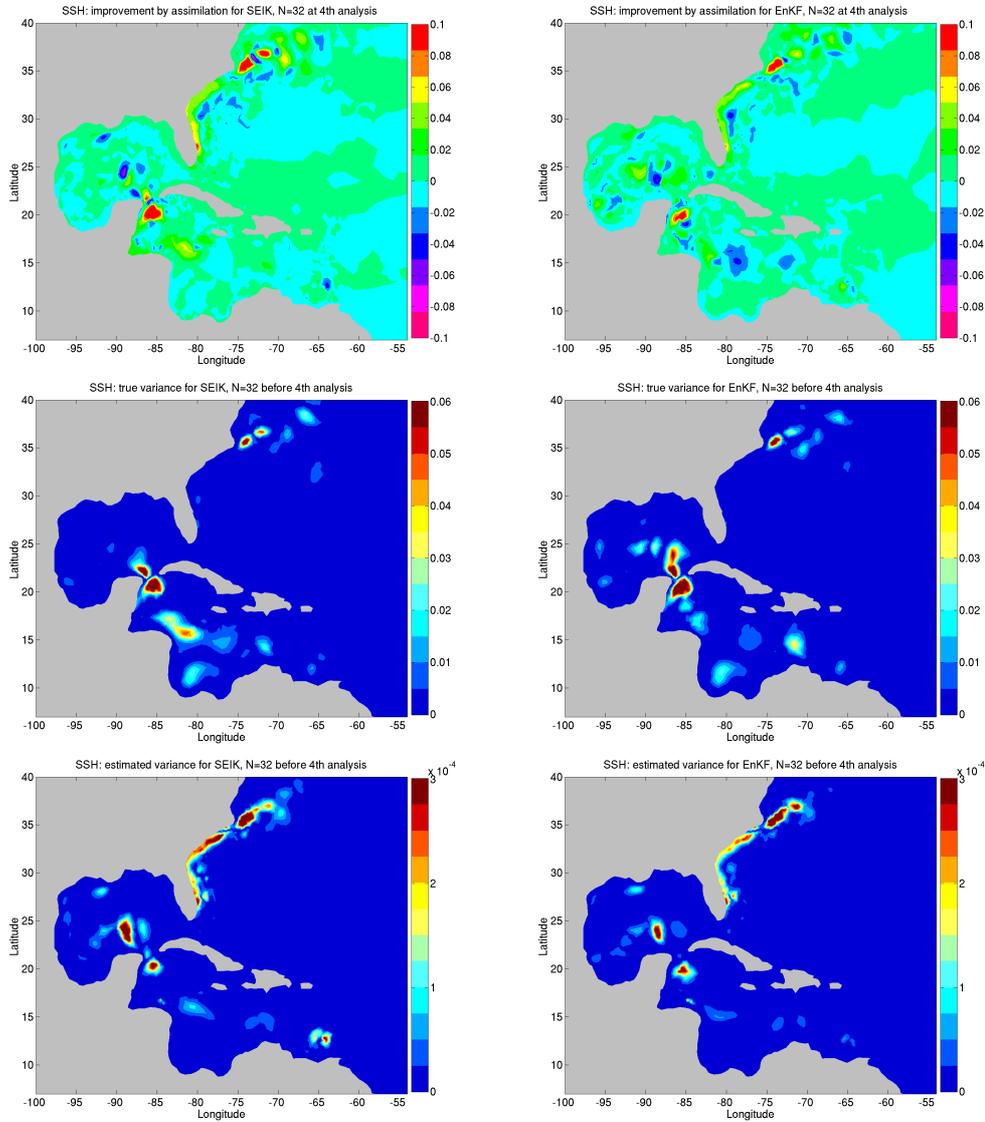


Fig. 7. Comparison of the behavior of the SEIK filter (left) and the EnKF (right) at the fourth analysis time for an ensemble size 32 in the region with the largest estimated variances. Shown is the improvement or error reduction (top), the true variance field (center), and the estimated variance field (bottom).