The HBM-PDAF assimilation system for operational forecasts in the North and Baltic Seas

L. Nerger (1), S. N. Losa (1), T. Brüning (2), and F. Janssen (2)

(1) Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany. lars.nerger@awi.de

(2) Federal Maritime and Hydrographic Agency (BSH), Hamburg, Germany

Abstract: The HIROMB-BOOS Model (HBM) has been coupled with the Parallel Data Assimilation Framework PDAF (http://pdaf.awi.de) in order to improve hydrographic forecasts in the North and Baltic Seas. The coupled forecast system assimilates satellite sea surface temperature as well as in situ data of temperature and salinity profiles to initialize forecasts up to 5 days. The assimilation uses an ensemble Kalman filter, which dynamically estimates the uncertainty of the state estimate with an ensemble of model states and applies spatially localized updates to improve the ocean state. The structure of the assimilation system, which can analogously be used to extend other forecast models for data assimilation, is discussed. Applying the assimilation reduces errors of the surface temperature by about 0.2°C.

Keywords: data assimilation, ensemble Kalman filter, North Sea, Baltic Sea

1. INTRODUCTION

The German Federal Maritime and Hydrographic Agency (BSH) has a large need for ocean forecasting data to run its internal operational services like the sea level prediction and storm surge warning service and the ice service, and to support external customers like the national search-and-rescue centers, the Central Command for Maritime Emergencies or the German Navy. In order to fulfill all these operational obligations, the BSH runs and maintains a comprehensive numerical ocean forecasting system that is under permanent revision. A fruitful cooperation in the Baltic area led to a spread of the original circulation model code developed at the BSH – called BSHcmod (Dick, 1997, Dick *et al.*, 2001, Kleine, 1994) – in the Baltic Sea community. Branches were installed at the Swedish Meteorological and Hydrological Institute and at the Danish Meteorological Institute (DMI). The three model lines have been actively developed over several years and somehow diverged over time. During recent years and with support of the MyOcean projects an effort has been made to merge the three development lines into one. The outcome of this effort is the HIROMB-BOOS Model (HBM, see Berg and Poulsen, 2012) nowadays jointly developed and operationally used by BSH, DMI, the Finnish Meteorological Institute and the Marine Systems Institute of Tallinn University. At the BSH the complete transition from the current operational model code BSHcmod towards HBM should be finished in 2015.

To improve the forecast skill of the HBM, information from observations can be taken into account by means of data assimilation. In data assimilation the information from a numerical model and from observations are combined to generate an improved estimate of the modeled state. By now the operational forecasts at the BSH are computed without the application of data assimilation. At the DMI, both optimal interpolation (Fu et al., 2011) and 3D-Var methods (Fu et al., 2012) have been applied.

Methods like optimal interpolation and 3D-Var use a single state realization and parameterized covariance matrices or weight matrices to prescribe the weight of deviations of the model state from observations. Current state-of-the-art data assimilation algorithms use an ensemble of model forecasts to estimate the uncertainty of the model state. These methods are typically called sequential data assimilation algorithms and base on the original ensemble Kalman filter (EnKF, Evensen, 1994). Several new and computationally more efficient methods have been developed, that are classified as ensemble square-root Kalman filters (Tippett *et al.*, 2003, Nerger *et al.*, 2012) or error-subspace Kalman filters (Nerger *et al.*, 2005). Here, the aim is to extend HBM for data assimilation with state-of-the-art ensemble filters. This goal is achieved by coupling HBM to the parallel data assimilation framework (PDAF, *Nerger et al.*, 2005). For the validation of the resulting assimilation system, the SEIK filter (Pham, 2001) is used in combination with a local analysis that builds the LSEIK filter (Nerger *et al.*, 2006).

2. HIROMB-BOOS MODEL

The HBM is a three-dimensional hydrostatic circulation model using the primitive equations. It uses spherical horizontal and generalized vertical coordinates (Kleine, 2003). The model domain extends from 4°W to 30.5°E and from 48.5°N to 60.5°N in the North Sea and to 66°N in the Baltic Sea. The horizontal grid spacing is ~5 km (5′ in longitude and 3′ in latitude). In the vertical, the model is discretized using 36 vertical layers. The layer thickness is about 2 m at the surface, up to 3 m in the upper 50 m of the water column. Below 50 m

it increases up to a thickness of 100m, whereas the bottom layer thickness is always about 3m.
In the North Sea, the model configuration has a northern open boundary, which is closed with a sponge layer. Within this layer, the temperature and salinity are restored towards monthly mean climatological values. A similar sponge region is included at the entrance to the English Channel.
A two-dimensional model for the North East Atlantic, which is run separately by the BSH, provides information on external surges at the open boundaries. Tidal forcing is implemented using 14 tidal constituents and flooding and drying of tidal flats is applied. The atmospheric forcing at the surface is based on meteorological forecast data provided by the German Weather Service (DWD). River runoff is prescribed as freshwater fluxes at the boundaries opened in the regions of main rivers. HBM includes a sea-ice model component, which describes sea ice thermodynamics and incorporates Hibler-type dynamics (Hibler, 1979).

3. THE DATA ASSIMILATION FRAMEWORK PDAF

The parallel data assimilation framework PDAF (Nerger et al, 2005, Nerger and Hiller, 2013, http://pdaf.awi.de) is a software framework that allows extending a numerical model with data assimilation functionality for ensemble-based Kalman filters. The framework provides support for ensemble forecasts within a single model program and several different ensemble square-root filters to compute the actual data assimilation (called 'analysis step') in which the forecast ensemble is combined with observations. The extension of the HBM for data assimilation is described in the following section. The assimilation framework uses a parallelization with the Message Passing Interface (MPI, Gropp et al., 1994) to allow for the ensemble integrations. Thus, while running a single program it is logically split into as many model tasks as there are ensemble states. The filter algorithms are parallelized using both MPI and OpenMP (see, e.g. Chandra et al., 2000).
For the setup of a data assimilation system, PDAF follows a strict logical separation of the program into three parts. These are the numerical model, the observations, and the data assimilation method. PDAF implements the data assimilation method such that it is completely independent from the model and the observations. In particular, the assimilation methods only consider so-called state vectors, i.e. the different model fields are combined in a single vector. With these state vectors, the analysis step can be performed by means of linear algebra without considering the different variables of the model. The data assimilation methods are part of the core of PDAF, which can be compiled as a program library as a user does not need to modify these functions. The information exchange between the model and PDAF is conducted through two subroutines that write the model fields into the state vector and vice versa. These routines are model-specific and coded upon the implementation of the assimilation program. The information about observations is handled in a set of subroutines that are used by PDAF as 'call-back' routines. Thus, these routines called by the core routines of PDAF through specified interfaces. These routines are implemented when the data assimilation program is created and have to be consistent with the model grid and the particular observations to be assimilated.

4. COUPLING HBM AND PDAF

The data assimilation system is implemented by coupling HBM to PDAF through the insertion of subroutine calls to functions of PDAF into the source code of HBM. The program flow is shown in Fig. 1. The blue fields show the parts of HBM that are executed. Without the data assimilation the model is first initialized, e.g. by reading the configuration of the model mesh from files and by reading initial model fields. Subsequently, the time stepping of the model is computed in which a prediction of the model state is computed. After the time stepping, the model can perform post-processing before the program ends.
For the data assimilation extension three different subroutine calls are inserted into the source code of HBM. The first call 'init_parallel_pdaf', executes a routine that sets up the parallelization of the data assimilation program. HBM has options to use both MPI and OpenMP for its own parallelization. In this study, we only used the OpenMP parallelization of HBM and used MPI to distribute the ensemble members. The parallelization is configured such the number of generated model tasks is equal to the ensemble size.
The second call 'init_pdaf' executes a subroutine that reads the configuration for the data assimilation. Further the subroutine itself executes the PDAF core routine 'PDAF_init', which initializes the assimilation framework internally and calls a call-back routine to initialize the ensemble of model states at the initial model time.
The third subroutine that is called 'assimilate_pdaf' executes the PDAF core routines for the analysis step. The routine can check if an analysis step should be performed at the time when it is called, if not, the time stepping of HBM continues.
During the analysis step, the PDAF core routine for the analysis step calls several call-back routines that perform the observation handling. As the filter algorithms needs to compute the difference between an observation and its model estimate, one of these operations is the so-called observation operator. This is a model- and observation-specific operator to extract the observed part of a state vector. Other

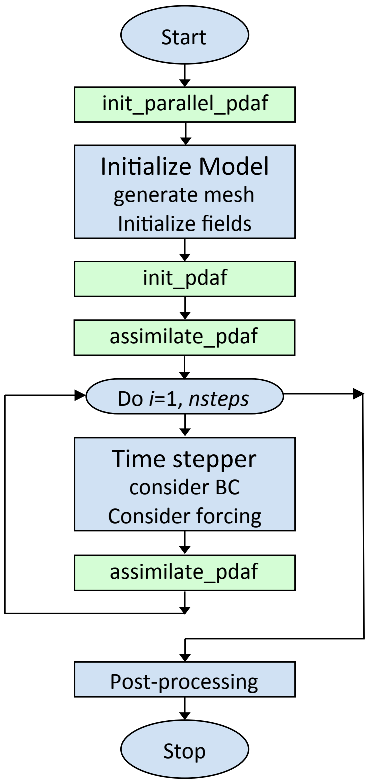

Fig. 1. Program flow of the HBM model (blue) with extension by subroutine calls to couple to the data assimilation framework PDAF.

operations are, for example to fill the vector of observations by reading them from a file and the computation of the distance of an observation from a grid point of the HBM model grid. For the analysis step, two options are implemented. In an operational setting, the analysis step is usually computed at the initial model time, and subsequently a long forecast of up to 72 hours is computed. In contrast, for hindcast experiments that are used to optimize the parameter settings for the data assimilation, the analysis step is usually not computed at the initial time, but after (fixed) intervals of time steps (e.g. after each 12 h). The two options are realized by placing the call to the routine 'assimilate_pdaf' both just before the time stepping loop and within it. By choosing a parameter at run time, one can then choose the mode in which the data assimilation is performed.

5. VALIDATION OF THE ASSIMILATION SYSTEM

5.1. Configuration of the assimilation system

The behavior of the assimilation system obtained by coupling HBM and PDAF is assessed with a realistic configuration that follows the operational forecast configuration at the BSH. However, while the operational forecast is computed without data assimilation, we here apply data assimilation to assess its impact on the forecast skill in a hindcast experiment. Following Losa *et al.* (2012, 2014), observations of the sea surface temperature (SST) from NOAA satellites are assimilated. This type of data is received by the BSH as level-2 data and further processed by the BSH. Thus, it is usable for operational purposes. An example of a 12-hour composite of SST data is shown in the upper panel of Fig. 3.

The assimilation is performed as a cycling experiment with an analysis step after each forecast of 12h over the full month of October 2007, which is chosen for consistency with Losa *et al.* (2012). Analogous to this study the LSEIK filter is applied. For the local analysis an observational influence radius of 100 km is used within which the observation influence decreases exponentially. Further it is assumed that the errors in the SST observations are uncorrelated with an error of 0.8°C. This rather high error takes into account the representation error, i.e. the inability of the model to exactly represent the observation, e.g. due to the model resolution of 5km. The initial model state estimate at October 1, 2007 is taken from the model forecast without data assimilation. The initial uncertainty that is represented through the ensemble spread is generated from the model variability by means of second-order exact sampling (Pham, 2001).

5.2. Assimilation results

Figure 2 shows the root-mean square (RMS) error between the modeled SST and the satellite observations. The RMS errors for the assimilation show the typical shape with error increases during the forecasts and an error reduction at each analysis time. For the 12h-forecasts from the analysis states, the RMS error is about 0.2°C lower than for the model without assimilation. Exceptions are at October 11, 22 and during October 24-26. Here, the model dynamics drive the error in the state estimate, computed as the mean of the ensemble, very close to the error from the free running model. On October

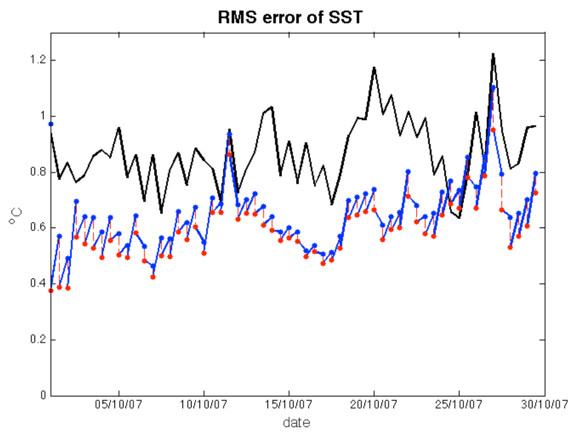

Fig. 2. RMS deviation of the modeled SST from the assimilated observations: (black) model without data assimilation, (blue) error of each 12h forecast, (red) error directly after the analysis step.

24 and 25, the forecast errors are slightly larger than the error from the free-running model without data assimilation. Over the full month, the data assimilation reduces the RMS error from 0.87°C for HBM without data assimilation to 0.67°C for the 12-hour forecasts. For the state estimates directly after the analysis step the RMS error was reduced to 0.59°C.

The upper panel of Fig. 3 shows the 12-hour composite of SST data on October 30, 2007 centered at midnight. Data is only available in regions without clouds. The weather conditions lead to a situation in which most of the Baltic Sea is unobserved. As the assimilation can only utilize the observational information in regions close to observations, most parts of the Baltic Sea will be not influence by the assimilation on October 30.

The middle panel of Fig. 3 shows the prediction of SST from HBM without data assimilation. Compared to the SST observations, it is visible that the SST in the northern part of the North Sea is too low. Further, the SST in the English Channel is underestimated. By applying the LSEIK filter throughout the month, the estimates of SST are improved with data assimilation (bottom panel of Fig. 3). In particular the SST estimates are higher in the northern North Sea and the English Channel. Also, the SST in the Baltic Sea around the Bay of Gdansk at about 55°N appears to be better represented by the assimilation estimate. However, along the Norwegian Coast in the Skagerrak, the assimilation underestimates the SST.

6. SUMMARY

A data assimilation system has been generated by coupling the HIROMB-BOOS Model (HBM) with the Parallel Data Assimilation Framework (PDAF). The HBM will be used as the next operational model code at the German Federal Maritime and Hydrographic Agency (BSH) to compute forecasts of the North and Baltic Seas. PDAF is a software framework to build ensemble-based data assimilation frameworks. It provides support for the ensemble integrations and provides fully implemented filter methods for data assimilation. Coupling PDAF and the numerical model into a single program results in the most efficient combination as the different parts of the data assimilation program can utilize the parallelization of both the ensemble integrations, the model, and the filter methods.

The coupled data assimilation program was validated with the assimilation of surface temperature observations from NOAA satellites over the month of October 2007. Using the same assimilation parameters as in Losa *et al*. (2012), where the assimilation was applied with the older model version BSHcmod, the RMS deviations of the model forecasts from the observations are reduced

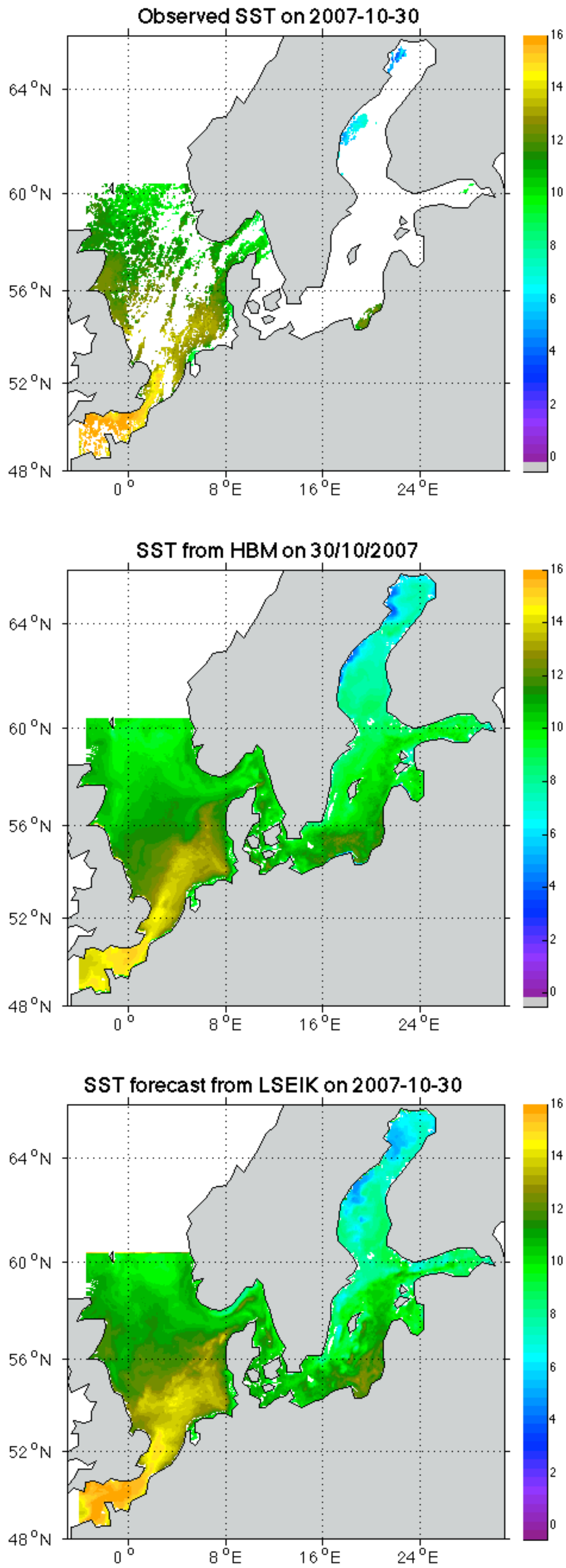

Fig. 3. SST on October 30, 2007, 00:00h: (top) Composite of satellite observations. (middle) SST estimate from HBM without assimilation. (bottom) Improved SST estimate from LSEIK assimilation.

shows that further tuning of the assimilation parameters is required with the HBM model to obtain optimal assimilation results.

Acknowledgements

The authors are grateful to NOAA and the BSH SST satellite data service for providing the data. The work was funded by the project DeMarine-2 by the German Federal Ministry of Economics and Technology (BMWi) through the German Aerospace Center (DLR) under grant number 50EE1113.

REFERENCES

Berg, P. and Poulsen, J. W., (2012) Implementation details for HBM. *DMI Technical Report*, No. 12-11, Copenhagen

Chandra, R., et al. (2000). *Parallel Programming with OpenMP.* Morgan Kaufmann, 250 pp.

Dick, S., 1997. Operationelles Modellsystem für Nord- und Ostsee. *FORUM, Proceedings der Fachtagung 'EDV im Seeverkehr und maritimen Umweltschutz'*. GAUSS, Bremen, pp. 22-25.

Dick, S., Kleine, E., Müller-Navarra, S.H., Klein, H., Komo, H., 2001. The operational circulation model of BSH (BSHcmod) model description and validation. *Berichte des BSH*. Hamburg, Germany

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 (C5), 10143-10162.

Fu, W., She, J., and Zhuang S. (2011). Application of an Ensemble Optimal Interpolation in a North/Baltic Sea model: Assimilating temperature and salinity profiles. *Ocean Modelling,* 40, 227-245

Fu, W., She, J., and Dobrynin, M., (2012). A 20-yr reanalysis Experiment in the Baltic Sea Using three Dimensional Variational (3DVAR) method. *Ocean Science*, 8, 827-844

Gropp, W., Lusk, E., Skjellum, A., (1994). *Using MPI - Portable Parallel Programming with the Message-Passing Interface*. The MIT Press, Cambridge. 307 pp.

Hibler, W.D. (1979). A dynamic/thermodynamic sea ice model. *Journal of Physical Oceanography*, 9, 815-846.

Kleine, E., 1994. Das operationelle Modell des BSH für Nord- und Ostsee: Konzeption und Übersicht. *Berichte des BSH*, Bremen, Germany

Kleine, E., (2003). A class of hybrid vertical coordinates for ocean circulation modelling. *Proceedings of the 6th HIROMB Scientific Workshop*. Morzaschita, St. Petersburg, 7-15.

Losa, S.N., Danilov, S., Schröter, J., Nerger, L., Massmann, S., Janssen, F., (2012.) Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas: Inference about the data. *Journal of Marine Systems*, 105-108, 152-162.

Losa, S.N., Danilov, S., Schröter, J., Janjic, T., Nerger, L., Janssen, F. (2014). Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas: Part 2. Sensitivity of the forecast's skill to the prior model error statistics. *Journal of Marine Systems*, 129, 259-270

Nerger, L., Hiller, W., Schröter, J. (2005). PDAF - the parallel data assimilation framework: experiences with Kalman filtering. In: Zwieflhofer,W.,Mozdzynski, G. (eds.). *Use of High Performance Computing in Meteorology - Proceedings of the 11. ECMWF Workshop*. WorldScientific, 63-83.

Nerger, L., Danilov, S., Hiller, W., Schröter, J. (2006). Using sea level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics*, 56, 634-649

Nerger, L., Janjic, T., Schröter, J, Hiller, W. (2012). A unification of ensemble square root Kalman filters. Monthly Weather Review, 140, 2335-2345

Nerger, L, Hiller W. 2013. Software for ensemble-based data assimilation systems - implementation strategies and scalability. *Computers & Geosciences,* 55, 110-118

Pham, D.T. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review,* 129, 1194-1207.

Tippett, M. K., Anderson, J. L., Bishop, C.H., Hamill, T. M., and Whitaker, J. S., 2003: Ensemble square root filters. *Monthly Weather Review*, 131, 1485-1490