Assimilating NOAA SST data into BSH operational circulation model for the North and Baltic Seas: Part 2. Sensitivity of the forecast's skill to the prior model error statistics

Svetlana N. Losa[a,*], Sergey Danilov[a], Jens Schröter[a], Tijana Janjić[c], Lars Nerger[a], Frank Janssen[b]

[a]*Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany*
[b]*Federal Maritime and Hydrographic Agency (BSH), Hamburg, Germany*
[c]*Hans Ertel Center for Weather Research, Data Assimilation Branch, German Weather Service*

Abstract

A data assimilation (DA) system has been developed for the operational circulation model of the German Federal Maritime and Hydrographic Agency (BSH) in order to improve the forecast of hydrographic characteristics in the North and Baltic Seas. It is based on the local Singular Evolutive Interpolated Kalman (SEIK) filter algorithm and assimilation of the NOAA AVHRR-derived sea surface temperature (SST). The DA system allows one to improve the agreement of the SST forecast with the satellite observations by 27% on average over the period of October 2007 – September 2008. However, a sensitivity analysis of the forecasting system performance shows a significant impact of initial model error statistics on ice fields and bottom temperature. A reinitialisation of model error covariances in accordance with

[*]Corresponding author
Email address: `Svetlana.Losa@awi.de` (Svetlana N. Losa)

seasonality of the model error statistics was required in order to maintain the predictive skill with respect to these variables. The success of the DA system is quantified by the comparison with independent data from MARNET stations as well as sea ice concentration measurements. In addition, the Maximum Entropy approach is used to assess the system performance and the prior and posterior model error statistics.

1. Introduction

Any estimate of the reality based on various sources of information depends on the quality of these sources and the method used for extracting and combining the information. Dealing with estimation of the ocean state or hydrography of different water basins, we are often uncertain about numerical model deficiencies and errors of the data (especially if the data are from satellite remote sensing). This makes the task of data assimilation (DA) into ocean models rather challenging (Brusdal et al., 2003; Penduff et al., 2002; Testut et al., 2003; Bertino and Lisæter, 2008; Brasseur et al., 2005; Cummings et al., 2009; Storkey et al., 2010; Kurapov et al., 2011). It pertains not only the implementation of DA algorithms but also the approximation of the error statistics (Counillon et al., 2009; Janjić et al., 2011; Fu et al., 2011; Simon and Bertino, 2012; Lermusiaux, 2007), which in each case demands a study on its own.

This is in full measure related to the development of a DA system for the operational forecasting model of the North and Baltic Seas run by the

2

German Federal Maritime and Hydrographic Agency (BSH), which was described in Losa et al. (2012). The DA system is based on Singular Evolutive Interpolated Kalman filter (SEIK, Pham, 2001; Pham et al., 1998). Setting up the system required calibration efforts in order to properly account for errors in assimilated NOAA's satellite sea surface temperature (SST) data (Losa et al., 2012). Implementation of the developed and tuned ensemble-based DA system has allowed the authors to improve the SST forecast over the calibration period of October 2007 and real-time forecasting period in March 2011. It was emphasized that the performance of the DA system depends on how the prior probabilities of model and data errors are estimated and prescribed. Relatedly, the skill of DA system can be improved by careful selection of the estimates, always keeping in mind the conditional nature of the statistics.

The present paper can be considered as a second part of the study. While the first paper was assessing assumptions on satellite SST data errors, the current paper focuses on the sensitivity of the forecasting system to the estimates of statistics describing BSH circulation model uncertainties. The prior model uncertainty in Losa et al. (2012) has been substituted by model variability under different atmospheric, tidal and river-runoff conditions over the period of October–December 2007. Here we discuss the performance of the data assimilation system during the seasonal cycle with various priors.

The dynamical conditions in the North and Baltic Seas change considerably from storms in autumn to partial sea ice cover in winter followed by the formation of a shallow thermocline in spring and summer. It turns out that this seasonality should properly be reflected in the error covariances to

reach "optimal" performance. In order to demonstrate the need for plausible estimates or hypotheses on model errors and their seasonal variability we use independent data on sea ice concentration and bottom temperature and salinity data from MARNET stations. As an additional criterion of plausibility of our assumptions on error statistics and the system performance itself, we use the Principle of Maximum Entropy (PME, Kivman et al., 2001).

The next section briefly describes the BSH operational model, the data and the DA algorithm. Section 3 proposes sensitivity experiments. The results are presented in section 4 and discussed in section 5. The last section concludes.

2. System description

Here we briefly describe the forecasting system augmented by DA. The reader is referred to the study by Losa et al. (2012) (hereafter, Lea12) for more details.

BSH operational circulation model

A 5 km horizontal resolution version of the BSH operational circulation model is nested within a coarser resolution ($\sim$10 km horizontally) model for the North East Atlantic and coupled with a sea ice model (BSHcmod, see Kleine, 1994; Dick, 1997; Dick et al., 2001; Kleine, 2003). The model setup, including the model domain, boundary and initial conditions, forcing and time stepping is similar to that used in Lea12.

4

Data

The information about the hydrographic characteristics provided by the BSH circulation model is sequentially, every 12 hours, combined with sea surface temperature measurements obtained with the Advanced Very High Resolution Radiometer (AVHRR) aboard polar orbiting NOAA's satellites. 12-hourly composites (around midnight and noon) of the measurements are processed and gridded by the BSH satellite data service. The observations are cumulative over the 12 hours window and should be representative of the averaged half day, unless gaps due to meteorological conditions prevent that.

As independent information on temperature and salinity, we use time series from the Marine Environmental Monitoring Network (MARNET). MARNET is operated in the North and Baltic Seas by the BSH and includes several automated measuring stations (Table 1).

DA method

The ensemble based SEIK filter with second order exact resampling (Pham, 2001) is implemented for assimilating the AVHRR-derived SST into the BSHcmod. The algorithm has been coded within the Parallel Data Assimilation Framework (PDAF, Nerger et al., 2005; Nerger and Hiller, 2012) and applied locally (Nerger et al., 2006) for each water column of the model domain with an observation error of $\sigma_{sst} = 0.8^{o}$C and exponentially decreasing data influence within the radius of 100 km. These localisation conditions had been found to be the best among others tested in Lea12, based on experiments for October 2007. We would like to reiterate that the σ_{sst} in the DA algorithm is not just the standard deviation of the data errors including measurement and representativeness errors, but reflects the ratio between

model and data quality due to a link between model uncertainties and data errors relative to the reality. Lea12 pointed out on a need of careful calibration of the data assimilation system with respect to possible approximations of the data errors and model uncertainties. Focusing on model deficiencies, in this study, we extend the system validation period to a one-year period and analyse the forecasting system's performance under initializations with various ensembles in distinct seasons.

3. Experiment design

In order to explore the sensitivity to initial error statistics, we carried out a set of experiments differing in when the system has been initialised and how the initial ensembles have been calculated. The experiments consist of control run (Experiment 1) and Experiments 2 – 6 described below.

Control run (Experiment 1)

The BSH forecasting system augmented by the data assimilation is validated over the period from 1 October 2007 to 30 September 2008 with corresponding atmospheric forcing, river run-off data and satellite SSTs. The data are sequentially assimilated every 12 hours. Lea12 considered various timing and periods of the analysis and forecast. Here we assess the system performance only with respect to produced 12-hourly forecast.

On 1 October 2007, the system has been initialised with the same ensemble of initial model states (temperature, salinity, current velocities, sea surface elevation) as in Lea12. The ensemble has been generated based on covariance matrices computed using 12-hourly snapshots of BSHcmod integration over the following three months periods: October - November -

6

December (autumn-winter). The mean values over the corresponding integration period have been substracted from the model solution. The leading empirical orthogonal functions (EOFs) have been used for generating the ensemble of 8 members. As mentioned in Lea12, under the used localisation conditions, this quite rough approximation of the initial model error covariances works sufficiently well for the system.

The ensemble of the initial model states evolves in time and so do the forecast error covariances. These dynamically changing forecast error covariances are inflated via the so-called forgetting factor. The forgetting factor has been introduced in the work of Pham (2001) as a number less than 1 that multiples the observation error covariance in the calculation of the Kalman gain. Its effect is the same as inflating the forecast error covariance with the factor inverse the forgetting factor. This inflation in part accounts for model error, as it compensates for the net effect of all factors leading to an under-dispersive ensemble. However, it assumes that the model error is proportional to the forecast error covariance, i.e. model error is assumed to belong to the same error subspace as the forecast error. Therefore, when the forecast error covariance matrix tends to zero the model error also tends to zero. Based on some sensitivity experiments previously conducted for the system (not shown) it is proposed to use the forgetting factor of 0.95. There is, however, an evidence of possible further improvement of the forecast quality with the parameter variable in time. This issue is not discussed in the current paper.

System's sensitivity to the initial ensemble (Experiments 2 – 6)

The initial ensemble samples the probability distribution of the initial model states. In this respect, the generation of the ensemble based on co-variances computed using model solution over a certain integration period is quite a standard and widely used approximation. With such an approach, however, the real initial model error statistics are replaced by the variability of the prescribed model dynamics under variable atmospheric forcing. The degree of its closeness to the real set of probable uncertainties in initial states determines the quality of sampling and, thus, the time evolution of the forecast error statistics.

By setting up the DA system we found sensitivity of the success to a dynamical change following the seasonal cycle. To evaluate the robustness of the forecasting system performance with the dynamically evolving error covariances, additionally to the control run (Experiment 1), a number of experiments (from 2 to 6) have been conducted by initialising the system in spring (on 1 March 2008) and summer (on 1 June 2008) with specially designed ensembles of the initial model states. For convenience, information on the experiments and conditions is summarized in Table 2.

The first set of sensitivity experiments (Experiments 2, 3, 4) focuses on spring season. In Experiment 2, the ensemble generated for the control run have been used to reinitialise the DA system on 1 March 2008. Two additional initial ensembles have been generated based on covariance matrices computed using 12-hourly model snapshots over spring season (March - April - May). In Experiment 3, the initial ensemble spreads around the seasonal mean, which was initially, as described above, subtracted from the model

solution before computing the covariances. Such an ensemble consistently prescribes the simulated variability. The ensemble mean, nevertheless, could be biased with respect to actual system state at the time of initialisation. The obtained covariances can, however, be considered as an approximation of uncertainties in specification of the initial model states. The clouds of the samples, then, can be spread around the model state at the particular date of the system initialisation. In this case, the mean of the ensemble is also dynamically consistent. As such, our second spring ensemble is spread around the model state on 1 March 2008 following 11 days of model integration started from the climatology (Experiment 4). As in Experiment 2, in Experiments 3 and 4 the DA system is initialised on 1 March 2008.

Two more ensembles of the initial model states have been generated to reinitialise the system on 1 June 2008 (Experiments 5 and 6). One initial ensemble is based on summer covariances, obtained by model integration over June - July - August. The second one uses the already described spring covariances. Both ensembles have been spread around the model state on 1 June 2008 and used for the sensitivity experiments (Experiments 5 and 6, respectively) with the focus on the summer season.

4. Results

To evaluate the performance of the forecasting system augmented by data assimilation, at first, we consider one year SST evolution based on 12-hourly local analysis. Then, we assess the system performance with respect to forecasting sea ice concentrations (a variable not included into the analysis state vector) and bottom temperature (one of the "unobserved", in terms of DA,

9

variables, but updated through the state covariances).

Control run: SST validation over a one year period

Figure 1 illustrates the ability of the DA system to reduce the deviation of the predicted SST from the satellite observations over a one year period. Spatial distribution of the root mean squared (RMS) deviation averaged over the period 1 October 2007 – 30 September 2008 obtained with DA (top right panel) is compared against that obtained without DA (top left panel). On average, over the model domain, the RMSE has been reduced from $\sim 1.1^{o}C$ to $\sim 0.8^{o}C$, which implies $\sim 27\%$ improvement. A similar comparison for the annual mean deviations (bottom right panel against bottom left panel) also shows the reduction from $\sim 0.5^{o}C$ to $\sim 0.3^{o}C$. For instance, augmenting the forecasting system by DA has enabled us to reduce bias in the Norwegian Trench and in the Gulf of Bothnia. Systematic errors in reproducing SST in the coastal region have been also reduced.

Comparison of the simulated sea surface temperature with independent SST time series at MARNET stations (Fig. 2) also shows significant improvement of the SST forecast quality when assimilating satellite data. Figure 2 shows the temporal evolution of the SST at Darss Sill and Arkona Basin MARNET stations (see Table 1 for the locations). As seen from Figure 2, the BSHcmod forecast without DA agrees quite well with the observations all year round except for May and June, when the model deviates from the observations up to $4^{o}C$. The deviations of the SST forecast from observations over this time interval are reduced by DA very substantially. Figure 2 also illustrates that the model deviation from the data has some systematic component (bias) which changes in time.

Table 3 summarizes the RMS and mean deviations of the forecast from MARNET SST observations with and without data assimilation. The statistics have been improved with the DA. It is worth noting that Figure 2 illustrates the DA impact at the most problematic stations with respect to the improved statistics. Table 3 also includes estimates of deviation of the NOAA's satellite temperature from the SST observed at the six MARNET stations. One can see that, except for the Darss Sill station, the SST prediction based on DA is more accurate than the AVHRR-derived temperature. Errors of the satellite data were the focus of our previous study. In particular, we addressed the problem of the data bias due to the algorithm used to convert the signal detected by satellite sensors to the bulk temperature at 1 m depth. In the current paper we do not discuss the quality of the observations and refer the reader to Lea12.

Ice simulation

In order to demonstrate skills of the implemented DA system we analyse its predictions with respect to a system variable that has not been included in the statistically updated state vector. As such, we compare predicted sea ice concentrations to observations. Since the sea ice model is coupled to the circulation model, the hydrography forecast serves as a factor influencing the ice dynamics. Since we do not assimilate sea ice data, it is natural to exclude the ice variable from the state vector, but evaluate the influence of the improved forcing on the sea ice forecast. To illustrate sensitivity of the predicted sea ice to the prior error statistics of the circulation model, we compare Experiment 1 and Experiment 4. Figures 3 and 4 depict the ice concentration on 27 March 2008 and 14 April 2008 respectively. Shown are the

observed ice concentration (top left panel) and the model ice concentration forecast without (top right panel) and with SST DA for the control run and experiment 4 (bottom panels). One can see that, generally, the BSHcmod underestimates the sea ice concentrations in the Gulf of Bothnia on 27 March 2008. The predicted ice extent, however, agrees quite well with observations, except for the central part of the gulf. Here the model forecasts a polynia, which is not observed. Assimilation of the satellite SST (note that there is no SST data in areas covered by ice), in Experiment 1, allowed ice of still small concentration to close the polynia. On 14 April 2008 (Fig. 4), the ice obtained with SST DA tends to cover the eastern part of the Gulf of Bothnia towards the north from 65ºN following the observations. But, at the same time, the simulated ice extends too much towards the south from65ºN. These results seem to indicate that the forecast error covariances evolving through time from 1 October 2007 are not able to sufficiently well approximate true error covariances for March and April. The forecast of the sea ice concentrations for the control run is compared against those obtained after the system reinitialisation on 1 March 2008 with the ensemble of initial model states based on spring error covariances (Experiment 4). The bottom left panels of Figs. 3 and 4 show the predicted concentrations of sea ice produced by the DA forecasting system reinitialised on 1 March 2008. Obviously, the new covariances have enabled the forecasting system to improve the quality of the predicted sea ice edge. Nevertheless, the ice concentration is still underestimated in most of the regions. In this respect, it should be noted that the ice model has been calibrated based on regular BSH circulation model, which is biased (see Fig. 1). While Figure 1 depicts the forecast deviation from obser-

vations averaged over a one year period, the patterns of the cold bias in the Gulf of Bothnia is similar to those estimated over March - April 2008. The magnitude of the spring bias is, however, smaller ($\sim 0.6^oC$) in comparison with the annual mean. Simultaneous SST and sea ice data assimilation (see Bertino and Lisæter, 2008) would be a way to optimize the BSH operational forecasting system.

Bottom temperature

The impact of initial model error covariances is even more pronounced for bottom temperature simulations. The bottom temperature is one of the model state variables statistically updated after the filter analysis. Figure 5 shows the bottom temperature forecast at the MARNET stations Darss Sill and Arkona Basin against observations (green). The forecast has been obtained for the period 1 March 2008 – 15 September 2008 without DA (black curve) and with LSEIK filtering for Experiments 2 (blue curve), 3 (black dashed) and 4 (red). One can see that, at the location of the stations, only the ensemble mean forecast based on experiment 4 has been able to follow the observed temperature during the warming and shallowing of the upper mixed layer (UML) over the period of 15th of May 2008 – 21st of June 2008 till the summer shallow UML has formed. Then the model trajectories lost the proper covariances. A comparison of the predicted bottom temperature without DA against the observations indicates some systematic model errors. These model uncertainties result in $\sim 5^oC$ colder bottom water relative to the observed one during June – July 2008. Modifying the initial model error correlations based on model variability under spring forcing and taking the ensemble spread around the model state on 1 March (Experiment 4) has

13

allowed the forecasting system to compensate for such a model deficiency, yet until the dynamical regime shift due to well established summer thermocline. After this shift, the system could not correctly approximate the true model error statistics with respect to the bottom temperature forecast. Figure 5 also demonstrates that constructing the initial ensemble around the model state on the initialisation date (Experiment 4) leads to certain advantages in comparison to the case where the ensemble is spread around the seasonal mean (Experiment 3) even if the ensemble is double in size.

Figure 6 illustrates possible improvement of the bottom temperature forecast for the period of June – September 2008 after reinitialization the system on 1 of June 2008 (Experiments 5 and 6). At this time, the system is started first with one ensemble based on model state covariances over June – July – August (Experiment 5). The bottom temperature obtained in this experiment at the Darss Sill and in the Arkona Basin (red curve) are, then, compared against the results of Experiments 3 (black dashed), 4 (red dashed) and Experiment 6 (blue, with an initial ensemble constructed from spring covariances and spread around the model state on 1 June). Among all the experiments, Experiment 5 simulates the observed summer bottom temperature at the MARNET stations best. However, by mid of August 2008, at the Arkona Basin station, the ensemble mean bottom temperature once again drops to the regular model bottom temperature (without DA, black) which is 5^oC colder than observed (green). After that the system no longer keeps the right bottom error covariances at this deepest MARNET station.

These results demonstrate that it is possible to create a DA system which would allow one to improve forecast even of the state variables that are not

14

observed. This is nevertheless conditional to the prescribed model error statistics (including uncertainties in the forcing). In the presence of uncertainties in specification of such statistics and lack of independent observational information there is a need for a criterion for evaluating the assumptions on model or data errors and, therefore, for assessing the DA system itself. Following Kivman et al. (2001) we propose to apply the principle of Maximum Entropy as such a criterion.

Maximum Entropy as an additional criterion to validate the system

In practical applications, uncertainties in the information coming from the model and measurements are among the reasons of sub-optimality of any DA analysis (see Lea12). The error statistics of the model and data are, however, often poorly known. As proposed by Kivman et al. (2001), in this situation, the most plausible assumption to make is that the information we are dealing with is most uncertain, and the best framework to think about it is the entropy. Kivman et al. (2001) in their study show how to apply an entropy approach to selecting priors in data assimilation problem (see Appendix A) so that the state estimates would be less biased with respect to the priors. In the study by Losa et al. (2004), the implementation of such an approach for a problem of state and parameter estimation in biogeochemical modeling allowed the authors to obtain reliable estimates of physiological parameters and, moreover, to infer about data quality and model uncertainties. Here, we apply the PME for calibrating our ensemble-based forecasting DA system (see Appendix A) and look at the summer results described above from the entropy point of view. In other words, hypothesising on the prior model error statistics, we use the PME as an additional criterion of plausi-

15

bility of our assumptions and forecasting system performance. The entropy estimates are given in Table 4. From this Table one can see that, over the period of 25 June – 8 August, the system has higher entropy under conditions of Experiment 5 than in case of Experiment 2, 3, 4 and 6. This circumstance leads us to conclude that initialization of the DA system with the 'summer' ensemble results in the least biased forecast among considered. This conclusion is in addition supported by comparison to the independent MARNET temperature data, as shown above.

5. Discussions

Arkona bottom temperature

Let us discuss in more details the fact that, initialised with 'summer' model error covariances, the DA system was able to recover the BSHcmod uncertainties in simulating bottom temperature at the deepest MARNET station "Arkona Basin" in summer period. The following questions naturally arise in this respect.

1. What could be the reason for the model discrepancy in reproducing bottom temperature at the Arkona Basin? Is it the resolution, errors in atmospheric forcing (or dare we say forcing errors relative to the resolution) or parameterisation of the horizontal and vertical mixing?

2. Is it possible to diagnose and distinguish the model uncertainties via DA?

3. Why does LSEIK filtering allow the forecasting system to compensate the model deficiencies in reproducing the Arkona bottom temperature in case of summer covariances but not in other cases?

16

Answering question 1, it is worth noting that, at first glance, the model cold bias of 5^oC near the bottom (40m) could indicate that model vertical mixing in the Arkona basin is too weak. For instance, the comparison of the predicted temperature profiles without DA against the observations (Fig. 7, top and middle panels) on 18 June and 1 July 2008 shows that the model underestimates the upper mixed layer (UML). However salinity measurements (Fig 8) indicate that the Arkona bottom water (much saltier than subsurface one) is formed somewhere in the Danish Straits under the North Sea water influence. The bottom temperature increase in the Arkona Basin during June – July 2008 is accompanied by a strong salinity increase (Fig. 8), which means that the bottom temperature changes due to lateral advection. This advective transport is missing in the free run (Fig. 7, middle and bottom panels). The resolution of 5 km used by us does not allow the model to simulate the North Sea and Baltic Sea waters interactions with necessary degree of realism. According to Figure 7, the forecasting system with DA better reproduces the salty water intrusion as concerns the bottom temperature and salinity. To what an extent the overall circulation pattern is improved remains an open question as we do not have any observational data to validate the velocity field.

The system with DA also better reconstructs the observed event of temperature minimum at depths close to 30m (Fig. 8). Indeed, as seen from Figure 8, the observed summer hydrography at the "Arkona Basin" is explained by interactions of at least 3 watermasses: the local (warm and relatively fresh) surface watermass; the bottom one originated from the saline North Sea waters; and the intermediate cold water initially formed close to

the Gulf of Bothnia, but absent in the model solution. The DA does a better job in simulating the watermass interactions than pure BSHcmod (Fig. 7, middle and bottom panels). The result is, however, imperfect and shifts with time and depth. Further improvement could be expected by calibrating the analysis localisation radius (see Lea12).

Concerning question 2 on the model uncertainties diagnostics we note that calculating covariances of the LSEIK analysis increment ("forecast-minus-analysis") over the period 25 June – 8 August 2008 sheds a light on some model error structures. It should be borne in mind that such diagnostics depends on how close the LSEIK analysis is itself to the reality. The increments presenting the model uncertainties are different in Experiments 4, 5 and 6. From the information entropy point of view, the increments obtained as a result of Experiment 5 (with higher entropy) should better correspond to the reality. The vertical temperature increment covariances at the "Arkona Basin" indicate the highest model uncertainties in bottom sea levels (see Fig. 9). This is in agreement with the aforementioned independent observations. The model error structure is, however, distinct from that in spring (Exp. 4; integration period: 9 March – 25 June 2008), with higher variances in the upper water levels (Fig. 9, middle left panel).

For the moment, it is not clear how to distinguish which factor is responsible for the model uncertainties. It can be errors in the atmospheric forcing, errors due to model resolution or parameterisations of horisontal and vertical mixing. It can also be errors in the bathymetry used by the model. All they are mixed together, so improvement with respect to a particular source of uncertainty, for example in forcing, is conditional to other acting error

18

sources.

To answer question 3, we compare our a posteriori estimates of the forecast error covariances with the prior initial model error statistics (Fig. 9). Given the initial model uncertainties approximated relying on model variability under spring forcing (Fig. 9, bottom left panel), the ensemble based forecasting system was not able to well mend the model deficiencies in deep water levels even with the dynamically changing forecast error covariances. Doubling the ensemble size (not shown) has not improved the situation and kept the bottom error correlations practically close to initial, which is wrong for the summer.

The initial forecast ensemble obtained with summer forcing, nevertheless, already possessed some significant features (compare top and bottom panels of Figure 9), allowing the system to account for and to catch the true forecast error covariances. This means that the model on its own (as expected) is able to simulate seasonal dynamics, but due to a bias– for whatever reason,– some processes do not occur at right place and time. Forcing errors could be the reason of such biases (Skandrani, 2009). In this case, stochastically perturbing the forcing would probably help (Brusdal et al., 2003; Sakov et al., 2012). However, we stress that, in the particular example of the dynamics at the "Arkona Basin" station, it could be just a compensation for uncertainties in model resolution. Thus, once again, we are dealing with forcing errors conditionally to resolution and/or the parameterisation of horizontal and vertical mixing.

One more reason why the forecasting system fails in the summer if it is initialised in spring based on the spring covariances could be the localisation

19

used for the flow dependent forecast error covariances at the analysis steps (see Lea12). As seen from Figure 10, which depicts spatial correlation of the forecast errors around the Arkona Basin, the error correlation length drastically increased in summer compared to spring. It could happen that, the localisation radius of 20 grid points given the exponential data weighting (see Lea12) is not big enough for that region to sufficiently approximate the changing summer error covariances. It points out to a need in spatially and temporally variable localisation in accordance to the observed dynamics.

General discussion on possible ways of accounting for model errors

Certainly, the story of the temperature at the Arkona basin is just an example. And yet it illustrates and highlights general points discussed in the DA community.

One major contributor to the forecast uncertainty is the model error. The reasons behind the model error include unrepresented subgrid scale processes, inaccurate forcing and boundary conditions, errors in representation of orography as well as parameterisation uncertainty. Statistical representations of possible model errors substantially effect the data assimilation results. Several approaches are available in the literature for the representation of model error statistics. These include use of multiple physics packages for the same model, inclusion of stochastic kinetic energy backscatter scheme to represent the unresolved scales at every time step, parameter variations, as well as use of deterministic stochastic dynamical models (Berner et al., 2011). However, as Berner et al. (2011) stated "there is no unique method the scientific community has agreed upon".

Rather then suggesting the periodical reinitialisation of the forecasting

system by seasonally based error statistics we would like to point at the possible directions in dealing with the model error statistics. A solution would be to introduce the model error statistics implicitly in the data assimilation algorithm, with additive model error covariance varying with the season. This would allow for the deficiencies in the time evolving error covariances to be compensated through the seasonally dependent model error. The approach would be similar to a hybrid discussed by Simon and Bertino (2012) when a constant ("static"), centered within 3 months period, covariances are used for observed variables, and dynamical covariances are used for unobserved variables. The authors pointed out on the possible sensitivity of the dynamically changed error covariances to model biases. The inclusion of different seasonal error structures in the data assimilation algorithm would allow us to keep the correct spread of the ensemble going through the different seasons.

A second approach that can partially alleviate the problem is a seasonally and temporally changing localisation radius. This would allow for the forecast error covariance depending on the season to search for the analysis in the different space. Continuous transformation between the localisation lengths would need to be ensured.

Additional work is required in these directions.

6. Conclusions

The BSH operational forecasting system extended by satellite sea surface temperature (SST) data assimilation (DA) based on a local version of the ensemble Kalman-type filter algorithm (SEIK) has been validated over the period 1 October 2007 – 30 September 2008. On average, during that period

21

the quality of the SST forecast has been improved by $\sim 27\%$ relatively to the forecast produced by regular BSHcmod without any DA. The improved performance of the forecasting system with DA has been also confirmed by comparison of the predicted SST against independent time series data at the six MARNET stations. The forecast estimates appeared to be even more accurate there than the satellite observations.

The changes in SST by DA affected sea ice forecast, while sea ice variables were not included in the model state vector updated during local SEIK analysis step. However, evolving through different seasons the forecast ensemble was not able to account for true model error covariances. As a consequence, the predicted sea ice as well as bottom temperature at MARNET stations were not optimal. The forecasting system had to be reinitialised on 1 March and 1 June with new ensembles. Several experiments with different initial ensembles of model states on 1 March and on 1 June 2008 have demonstrated substantial sensitivity of the sea ice and bottom temperature forecast quality to the assumptions on the model error statistics. These initial ensembles have been generated based on the model variability under particular seasonal (autumn-winter, spring, summer) forcing, which, as assumed, approximated the model error correlations. It has been found that the best forecasts for the periods of March – June 2008 and June – August 2008 have been obtained with spring and summer model error covariances, respectively. As a criterion of plausibility of the prior error covariances and the DA system performance, the principle of maximum entropy is used. The estimates of the entropy confirmed that the June initialisation with the summer ensemble has enabled the system to produce the least biased forecast with respect to

the used priors. The statistics of model errors is seasonally variable, because when simulating different dynamical regimes, the model has different systematic errors. The forecasting system has to account for this seasonality of the model deficiencies.

Our study is, in fact, just a validation of a newly developed ensemble based DA forecasting system for the North and Baltic Seas and an illustration of the sensitivity of the forecast performance to the information about model uncertainties, which is a priori unknown. The paper does not provide best solution to the problem, leaving space for future research. But, obviously, a posteriori analysis of model errors would help to educate the forecasting system and optimise it by including its statistics into the DA algorithm.

Appendix A. Principle of Maximum Entropy

General formulation by Kivman et al. (2001)

From a probabilistic point of view, the problem of data assimilation into dynamical models is formulated as estimating $\rho(x|y)$, the probability density function (PDF) of probable model trajectories realisations x given the data

y (van Leeuwen and Evensen, 1996). This conditional a posteriori PDF, also called the analysis PDF, expresses the state of our knowledge about the model state when data are observed.

Following Kivman et al. (2001), this analysis PDF should maximize the entropy

$$S(\rho) = -\int_X \rho(x|y) \ln \frac{\rho(x|y)}{\mu(x)} \prod dx, \tag{1}$$

where $\mu(x)$ is the lowest information about the system state x. The most probable analysis x^a or mean with respect to such $\rho(x|y)$ is

$$x^a = M_m x_m + M_d x_d, \tag{2}$$

where x_m and x_d are any system states satisfying the model equations $L(x) = f$ and data $H(x) = y$, respectively. Here, L is the model operator describing internal processes, f is external forcing, H is an observational operator. As shown in the study by Kivman et al. (2001), the operators M_m and M_d depend on both the operators L and H and also on our assumptions on the prior error statistics. Kivman et al. (2001) prove that M_m and M_d are nonnegative, self-adjoint, and that

$$M_m + M_d = I. \tag{3}$$

In our applications, we are certain neither in model error covariances nor in data errors. In this case, by virtue the properties of M_m and M_d, the so-called operator-valued measure M generated by the operators can be used for assessing the assumed prior error statistics if the entropy of M is defined following Kivman et al. (2001)

24

$$S(M) = -\mathrm{tr}(M_d \ln M_d + M_m \ln M_m) = -\sum_{i=1}^{N}[\lambda_i \ln \lambda_i + (1-\lambda_i)\ln(1-\lambda_i)]. \quad (4)$$

Given eigenvalues λ_i of M_d or M_m matrixes, one can calculate the entropy. Assessing the assumptions on the model and data error statistics, we search for the prior which generates M with the highest entropy value.

S in terms of Kalman-type filtering

In terms of Kalman-type filtering, the maximum probable x or so called state vector analysis x^a is expressed as the following

$$x(t_n)^a = x(t_n)^f + K_n(y_n - Hx(t_n)^f), \quad (5)$$

where $x(t_n)^a$ and $x(t_n)^f$ denote the analysis and forecast of the model state at certain time t_n, y_n are observations available at t_n, and K_n is the Kalman gain

$$K_n = P_n^f H (H P_n^f H^T + R)^{-1}. \quad (6)$$

Here, following Pham (2001), P_n^f is the forecast error covariance matrix, H is the observation operator and R is the observational error covariance matrix.

Comparing Eq. 5 to Eq. 2, one can see that the operator-valued measure M is now determined by Kalman gains. To calculate the entropy $S(M)$ (4), one just need to know λ_i of the Kalman gain matrix. Such a matrix could be constructed by collecting and considering $K_n H$, for instance, globally over a certain period of time or locally. The local variant is valuable for validation of localisation conditions. λ_i of the matrix are obtained by SVD decomposition.

References

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A., Snyder, C., 2011. Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. Mon. Wea. Rev 139, 1972–1995.

Bertino, L., Lisæter, K. A., 2008. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. Journal of operational oceanography 1(2), 15–18.

Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J. M., Ferry, N., Losa, S., Remy, E., Schröter, J., Skachko, S., Testut, C. E., Tranchant, B., van Leeuwen, P. J., Verron, J., 2005. Data assimilation for marine monitoring and prediction: the MERCATOR operational assimilation systems and the MERSEA developments. Quarterly Journal of the Royal Meteorological Society 131(613), 3561–3582.

Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. Journal of Marine Systems 40–41, 253–289.

Counillon, F., Sakov, P., Bertino, L., 2009. Application of a hybrid EnKF-IO to ocean forecasting. Ocean Sci. 5, 389–401.

Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M., Mogensen, K., Oke, P., Testut, C., Verron, J., Weaver, A., 2009. Ocean Data Assimilation Systems for GODAE. Oceanography 22(3), 96–109.

Dick, S., 1997. Operationelles Modellsystem für Nord- und Ostsee. In: FORUM, Proc. 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.

Fu, W., She, J., 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.

Janjić, T., Nerger, L., Albertella, A., Schröter, J., Skachko, S., 2011. On domain localization in ensemble based Kalman filter algorithms. Monthly Weather Review 136 (7), 2046–2060.

Kivman, G. A., Kurapov, A. L., Guessen, A., 2001. An entropy approach to tuning weights and smoothing in the generalized inversion. Journal of atmospheric and oceanic technology 18, 266–276.

Kleine, E., 1994. Das operationalle Modell des BSH für Nord- und Ostsee: Konzeption und übersicht. Berichte des BSH, Hamburg.

Kleine, E., 2003. A Class of Hybrid Vertical Co-ordinates for Ocean Circulation Modelling. In: Proceedings of the 6th HIROMB Scientific Workshop. Morzaschita, St. Petersburg, pp. 7–15, [ISBN 5- 86813-132-0].

Kurapov, A. L., Foley, D., Strub, P. T., Egbert, G. D., Allen, J. S., 2011. Variational assimilation of satellite observations in a coastal ocean model off Oregon. J. Geophys. Res. 116, 1–19.

Lermusiaux, P. F. J., 2007. Adaptive modeling, adaptive data assimilation and adaptive sampling. Physica D 230, 172–196.

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., Kivman, G. A., Ryabchenko, V. A., 2004. Weak constrant parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data. Journal of Marine Systems 45, 1–20.

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

Nerger, L., Hiller, W., 2012. Software for Ensemble-based Data Assimilation Systems — Implementation Strategies and Scalability. Computers and Geosciences doi:10.1016/j.cageo.2012.03.026, (In press.).

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 Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology. Singapore: World Scientific, Reading, UK, 25 - 29 October 2004, pp. 63–83.

Penduff, T., Brasseur, P., Testut, C.-E., Barnier, B., Verron, J., 2002. A four-year eddy-permitting assimilation of sea-surface temperature and al-

timetric data in the South Atlantic Ocean. Journal of Marine Research 60(6), 805–833.

Pham, D. T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Mon. Weather Rev. 129, 1194–1207.

Pham, D. T., Verron, J., Gourdeau, L., 1998. Singular evolutive Kalman filters for data assimilation in oceanography. C. R. Acad. Sci. Paris, Earth and Planetary Sciences 326, 255–260.

Sakov, P., Counillon, F., Bertino, L., K, A. L., Oke, P. R., Korablev, A., 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. Ocean Sci. 8(4), 633–656.

Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: application to a 1D ocean ecosystem model. J. Mar. Sys. 89, 1–18.

Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P., Siddorn, J. R., 2010. Forrecasting the ocean state using NEMO: The new FOAM system. Journal of operational oceanography 3(1), 11–15.

Testut, C.-E., Brasseur, P., Brankart, J.-M., Verron, J., 2003. Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean. Journal of Marine Systems 40-41, 291–316.

van Leeuwen, P. J., Evensen, G., 1996. Data assimilation and inverse methods in terms of probabilistic formulation. Month. Weath. Rev. 124, 2898–2913.
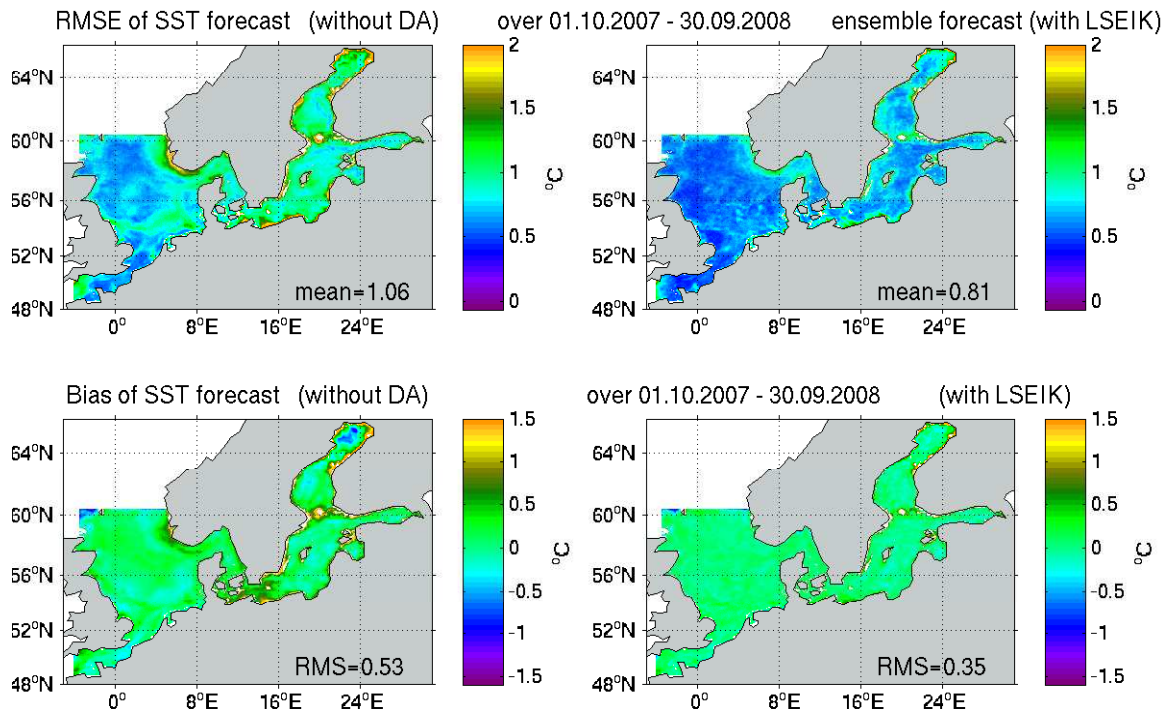
Figure 1: Deviation of the SST forecast from the satellite observations averaged over the period 1st of October 2007 – 30th of September 2008.: root mean squared (top panels) and mean (bottom panels) estimates obtained without DA (left panels) and when locally SEIK filter implementing (right panels).
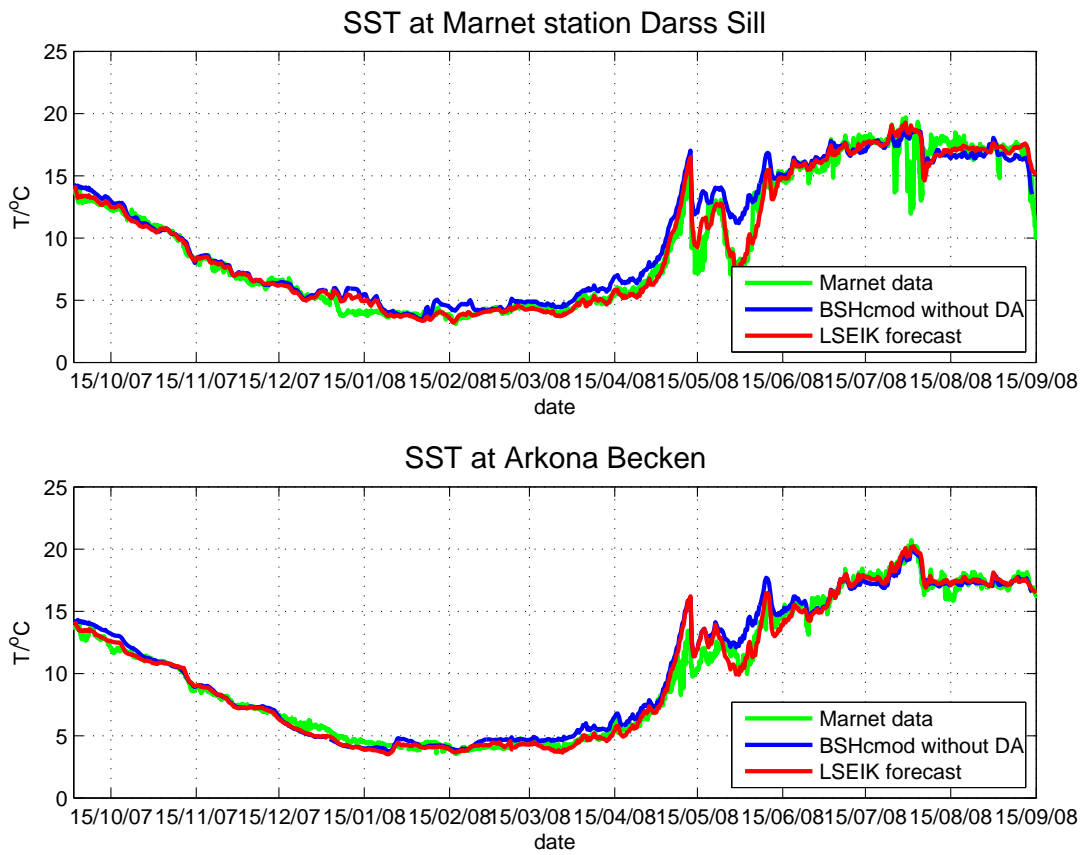
30

Figure 2: Temporal evolution of the sea surface temperature at MARNET stations "Darss Sill"(54°42$'$ N, 12°42$'$ E,top panel) and "Arkona basin" (54°53$'$ N, 13°52$'$ E, bottom panel) over the period 1st of October 2007 – 30th of September 2008: observation (green), model solution without data assimilation (blue), forecast based on local SEIK analysis (red).

31
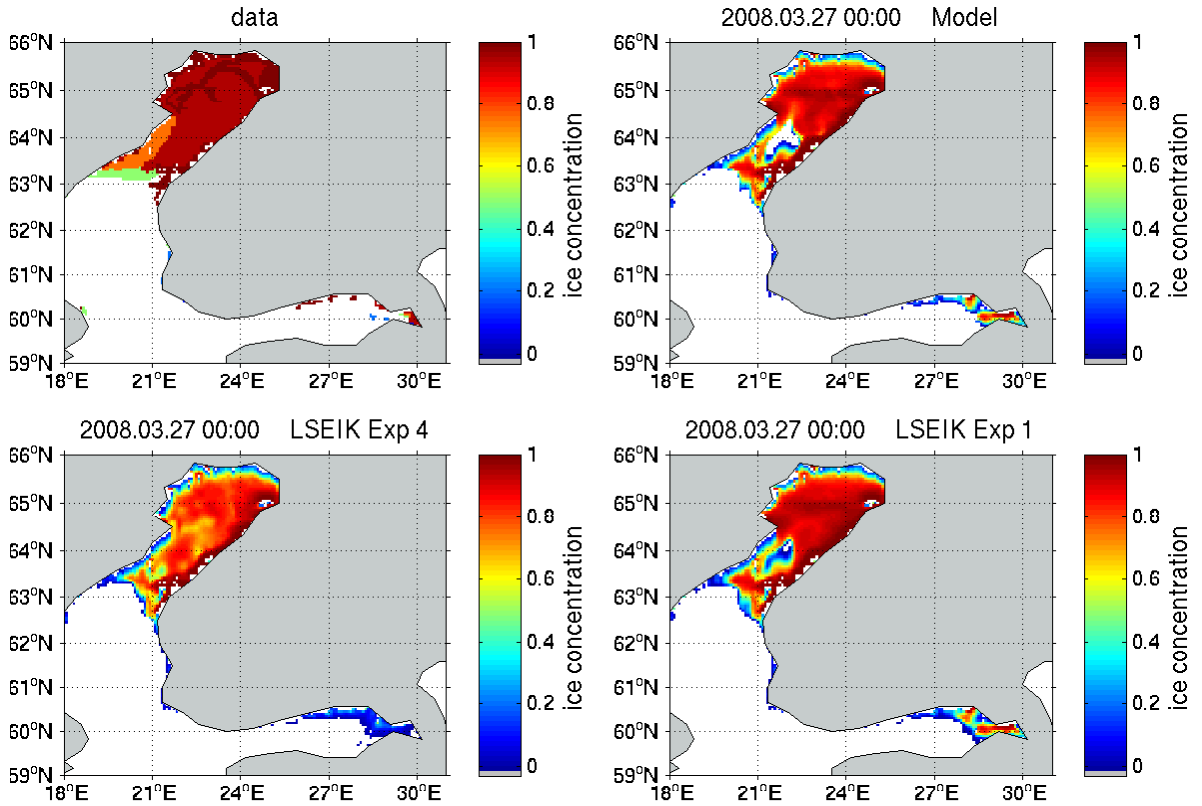
Figure 3: Ice concentrations on 27th of March 2008: data (top left panel); BSHcmod forecast without DA (top right panel); LSEIK ensemble mean forecast for Experiment 4, based on the ensemble initialised on 1st of March 2008 with the given spring covariance matrix (bottom left panel); LSEIK ensemble mean forecast based on the control ensemble initialised on 1st of October 2007 with the given winter covariance matrix (bottom right panel).
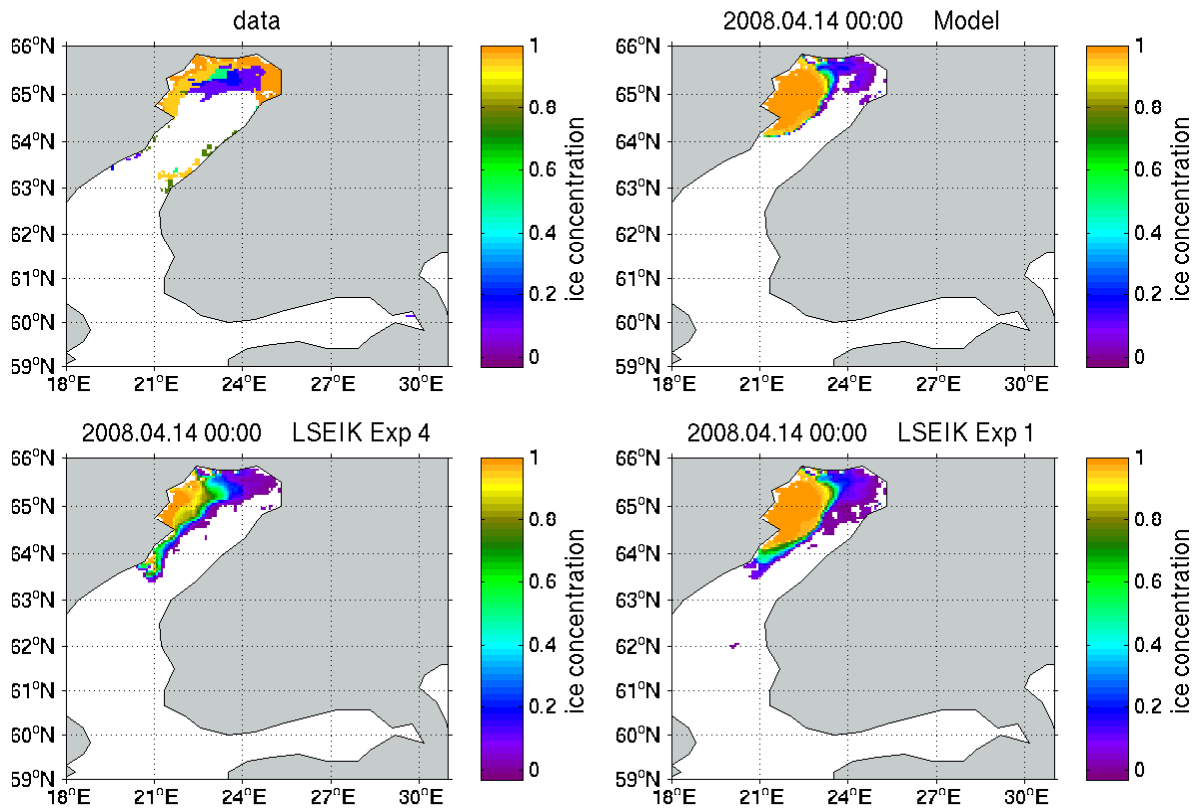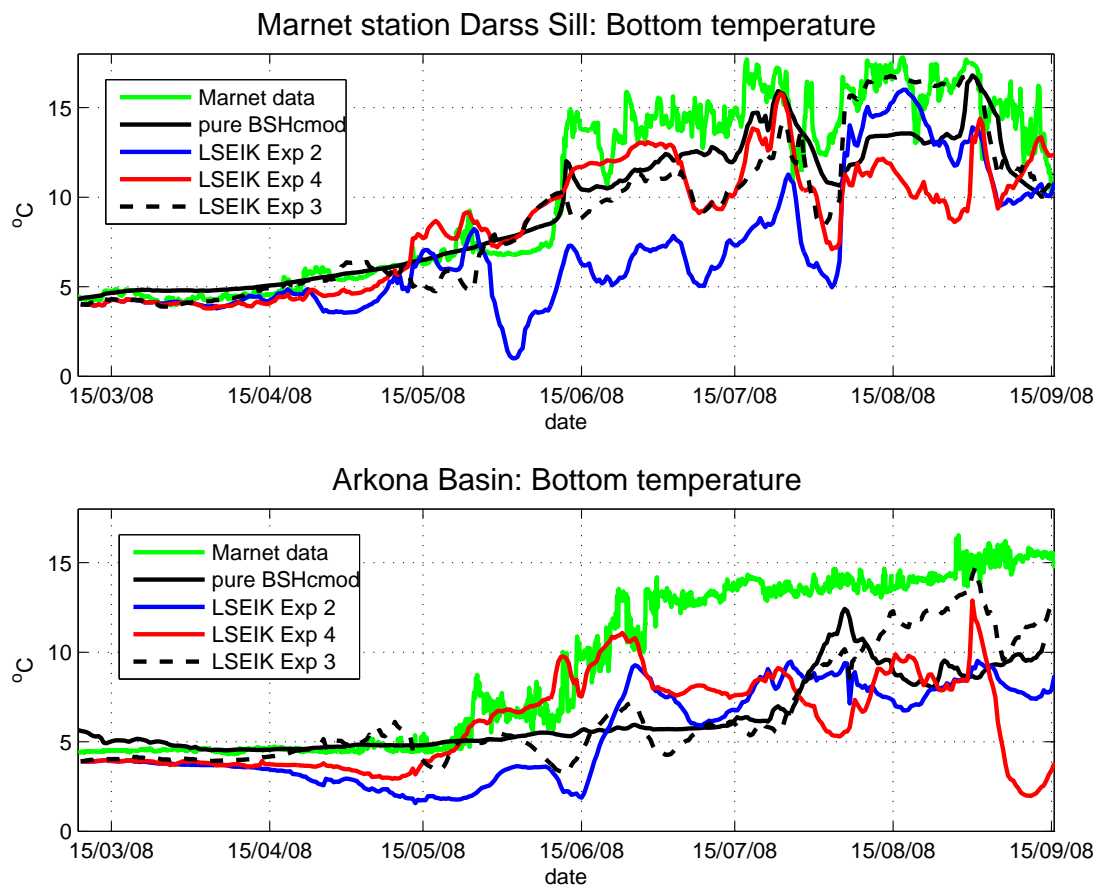
Figure 4: Same as in figure 3, but on 14th of April 2008.

Figure 5: Temporal evolution of the bottom temperature at MARNET stations "Darss Sill"($54^o42'$ N, $12^o42'$ E,top panel) and "Arkona basin" ($54^o53'$ N, $13^o52'$ E, bottom panel) over the period 1st of March 2008 – 30th of September 2008: observation (green), model solution without data assimilation (black), forecast based on local SEIK analysis of Experiment 2 (blue), 3 (black dashed) and 4 (red).
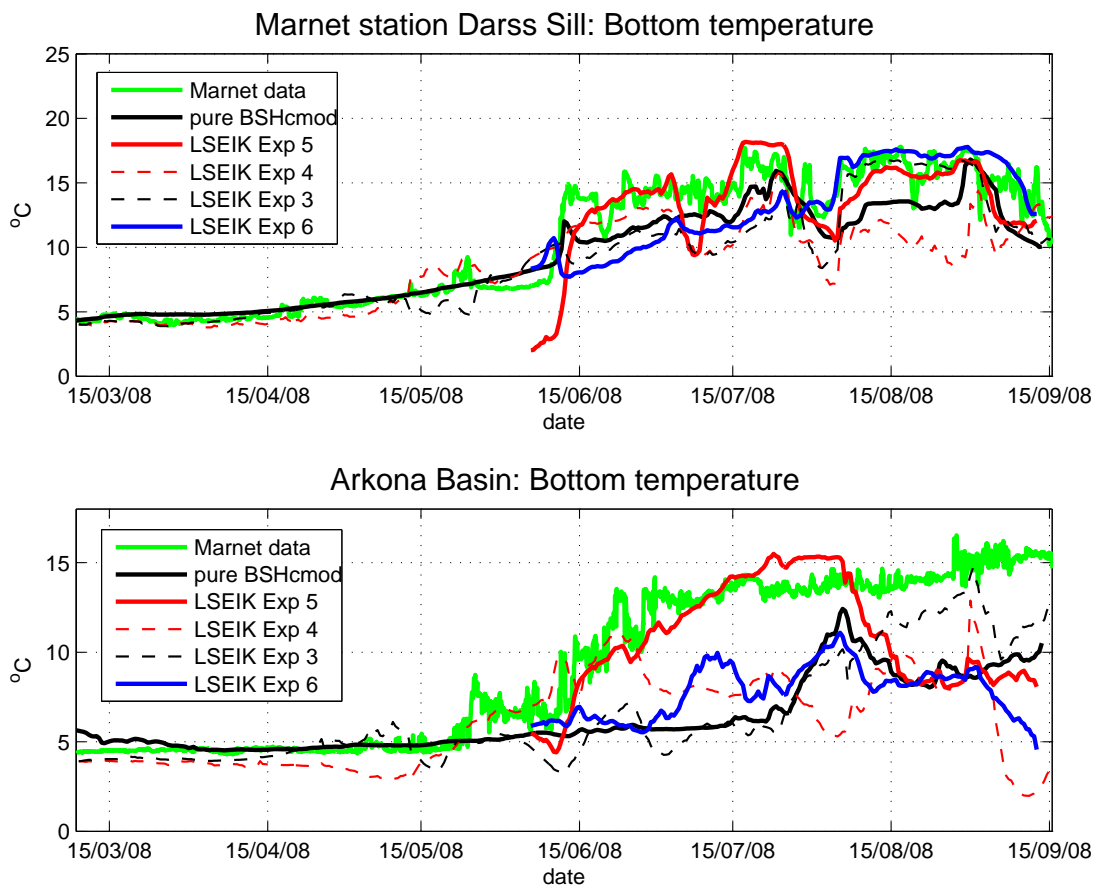
Figure 6: Temporal evolution of the bottom temperature at MARNET stations "Darss Sill"(54°42$'$ N, 12°42$'$ E,top panel) and "Arkona basin" (54°53$'$ N, 13°52$'$ E, bottom panel) over the period 1st of June 2008 – 30th of September 2008: observation (green), model solution without data assimilation (black), forecast based on local SEIK analysis of Experiment 4 (red dashed), 3 (black dashed), 5 (red) and 6 (blue).
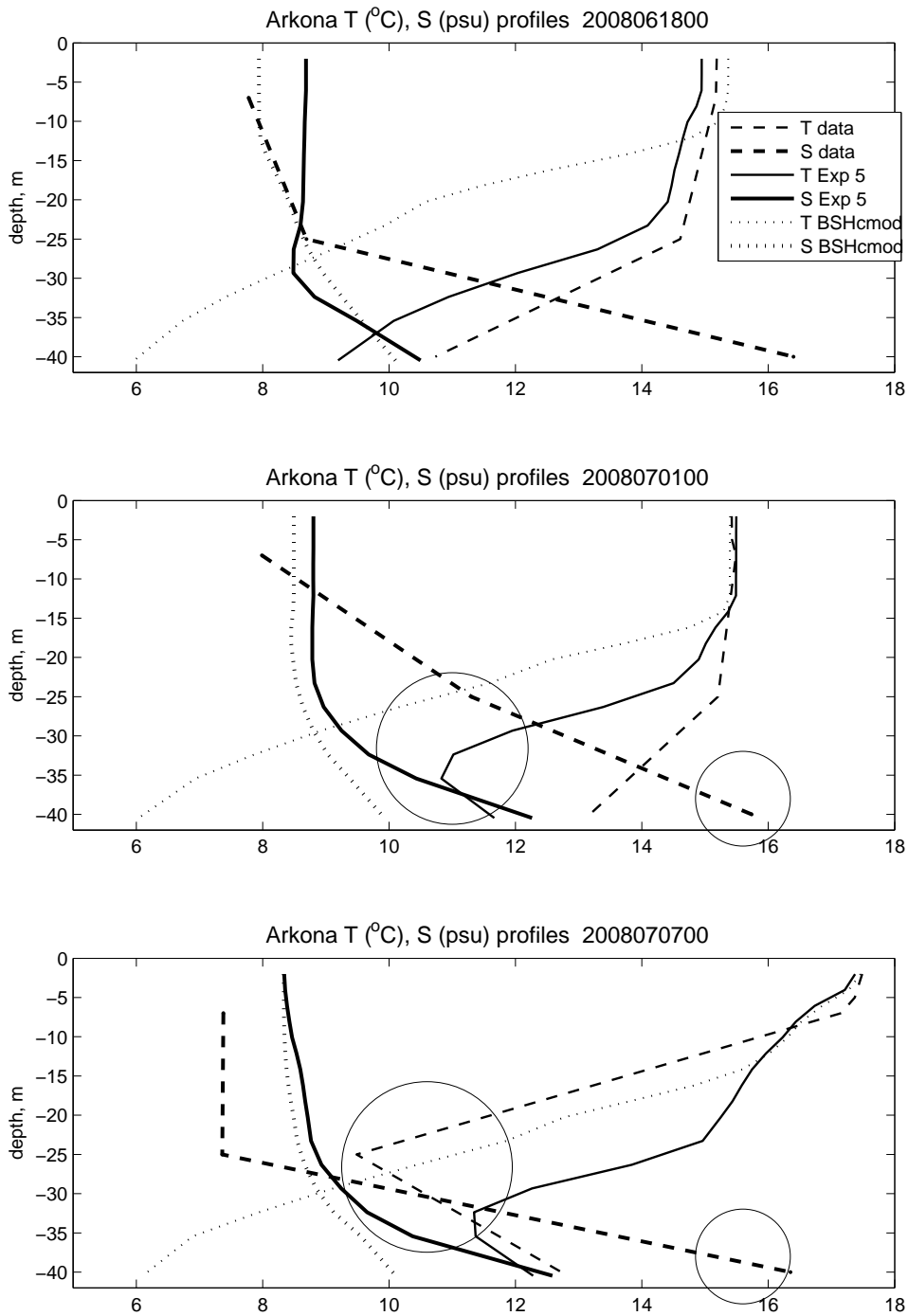
Figure 7: Temperature, salinity profiles at the "Arkona basin" station on 18 Juni 2008 (top panel), 1 July 2008 (middle panel) and on 7 July 2008 (bottom panel): forecast without DA (solid lines); forecast with DA (dotted lines); observations (dashed lines). In the circles are events of intrusion of bottom warm and salty and intermediate cold and less saline waters.
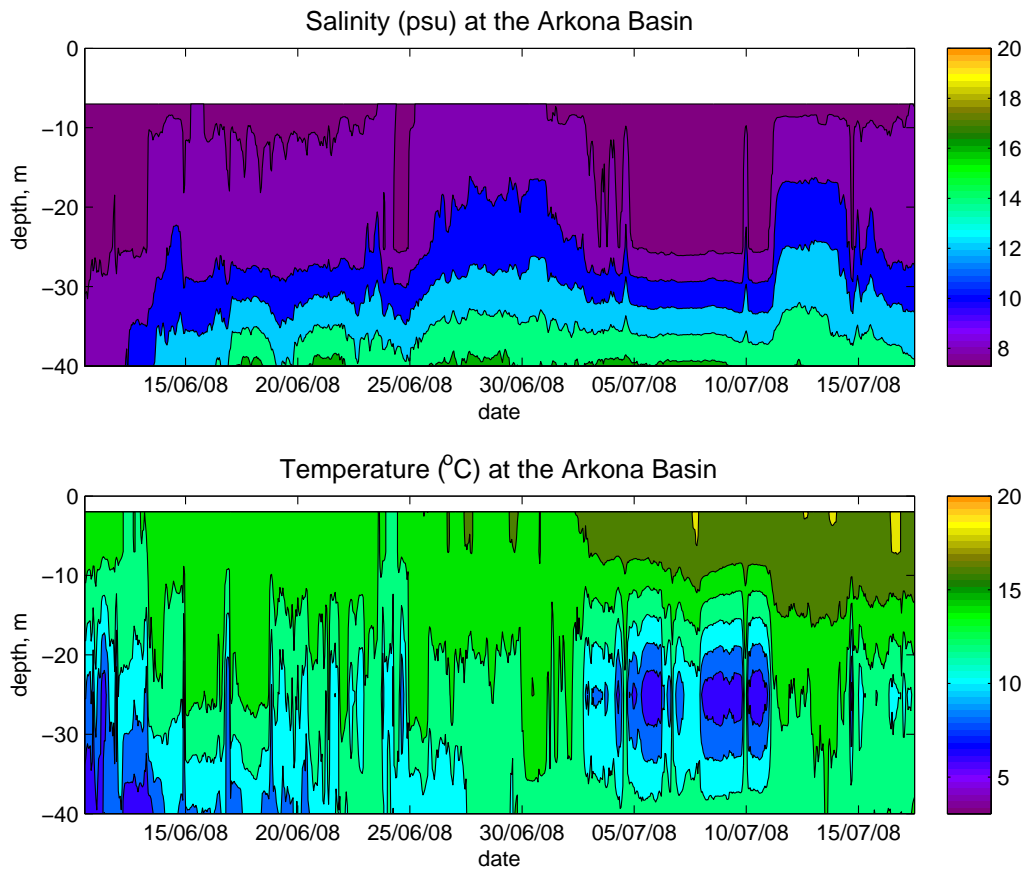
Figure 8: Temperature, salinity profiles temporal evolution observed at the "Arkona basin" station over the period 10 June 2008 - 17 July 2008.
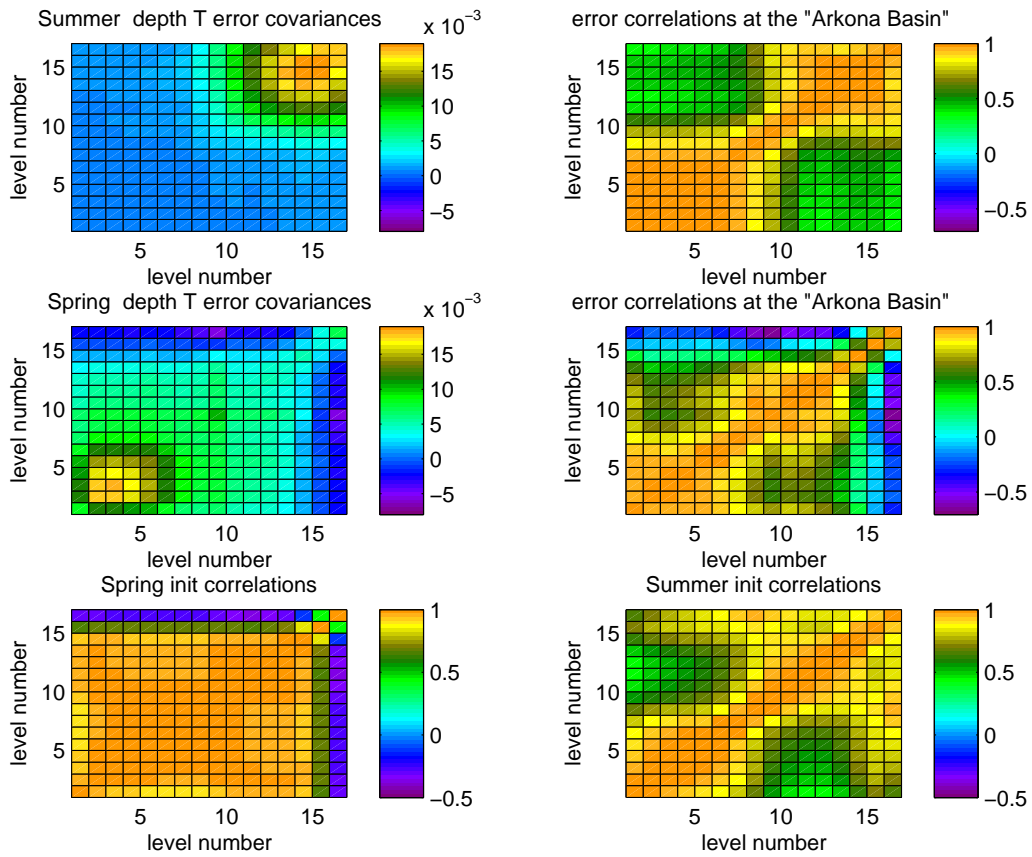
Figure 9: Temperature error vertical covariances and correlations at the "Arkona Basin": summer flow dependent forecast error covariances (top left panel) and correlations (top right panel); spring flow dependent forecast error covariances (middle left panel) and correlations (middle right panels); initial spring model error correlations (bottom left panel); initial summer model error correlations (bottom right panel).
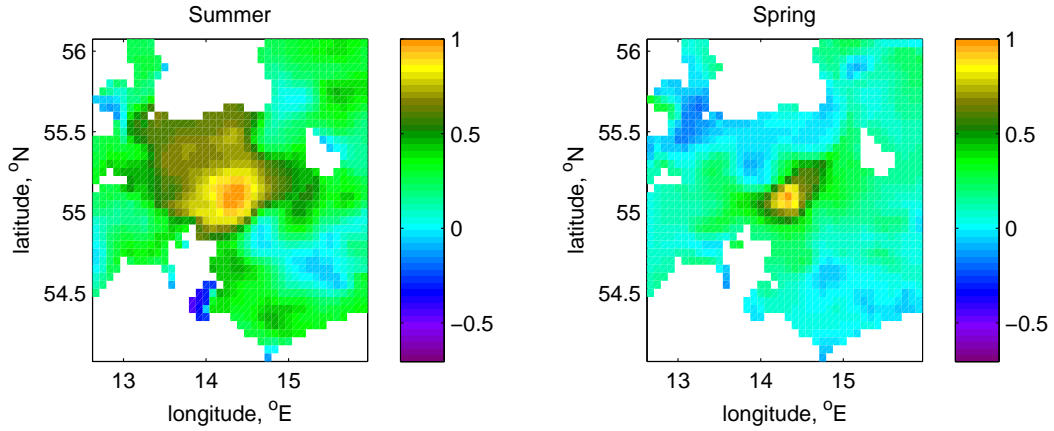
Figure 10: BSHcmod salinity spatial error correlations around the "Arkona Basin" : in summer (left panel) and spring (right panel).

Table 1: Coordinates of the MARNET stations.

Station	*Location*
Arkona	54º53′ N, 13º52′ E
Darss	54º42′ N, 12º42′ E
Kiel	54º30′ N, 10º16′ E
Fehmarn	54º36′ N, 11º09′ E
Ems	54º10′ N, 6º21′ E
Germ. Bight	54º10′ N, 7º27′ E

Table 2: Experiment conditions

Experiment number	Initialization date (ID)	Covariance over period	Ensemble mean	Ensemble size
1	Oct 1st 2007	Oct - Dec	seasonal*	8
2	Mar 1st 2007	Oct - Dec	seasonal	8
3	Mar 1st 2008	Mar - May	seasonal	16
4	Mar 1st 2008	Mar - May	state on ID**	8
5	Jun 1st 2008	Jun - Aug	state on ID	8
6	Jun 1st 2008	Mar - May	state on ID	8

* – averaged over the covariance period; ** – model state on the ensemble initialization date spun up for 11 days of model integration started from climatology.

Table 3: Deviation of the SST predicted with and without DA and AVHRR-derived temperature from the MARNET observations.

	RMS (°C)			Mean (°C)		
Station	BSHcmod	LSEIK	NOAA	BSHcmod	LSEIK	NOAA
Arkona	0.88	0.58	0.61	0.29	0.0	-0.04
Darss	1.27	0.81	0.69	0.55	0.17	-0.01
Kiel	0.79	0.49	0.61	0.13	-0.07	-0.08
Fehmarn	0.63	0.43	0.56	0.16	-0.03	-0.16
Ems	0.67	0.45	0.49	-0.33	-0.20	-0.17
Germ. Bight	0.97	0.53	0.57	0.34	0.03	-0.27

Table 4: Entropy (an example for the integration period of 25 June – 8 August 2008).

Experiment number	Entropy value
2	3.97
3	4.24
4	4.44
5	4.89
6	2.71