Interactive comment on “Assimilating High-resolution Sea Surface Temperature Data Improves the Ocean Forecast in the Baltic Sea” by Ye Liu and Weiwei Fu

Anonymous Referee #1

Received and published: 27 March 2018

The marine status of the Baltic Sea is highly variable and influenced by the forcing from atmosphere and freshwater influx due to shallow topography and semi-enclosed restriction. As a stable observation source, the high-resolution SST from satellite is rather important to improve the ocean operational forecast to serve the Baltic industry needs. The article of “Assimilating High-resolution Sea Surface Temperature Data Improves the Ocean Forecast in the Baltic Sea” use a localized Singular Evolutive Interpolation Kalman (SEIK) filter to assimilate the OSISAF SST during one year of 2010. Compared with dependent and independent observations, the evaluation of the model runs with and without assimilating the SST shows the SST modeling has been improved clearly. This study is suitable for publication in OS, but there are still some obvious defects like the experimental illustration is not clear, lack of conclusions or analysis methods to inspire the readers.

We appreciate the referee for the good comments, which definitely contributes to the improvement of this study. Our responses are in blue.

The main comments are listed as follow:

1) In this study, only to assimilate the OSISAF SST in the Baltic sea. In fact, there are more SST candidates with equivalent high-resolution like OSITA (CMEMS) and RTG_SST_HR (http://polar.ncep.noaa.gov/sst/rtg_high_res/). So if assimilating one or two additional SST products, the related results will be more help the reader to well understand about them. On the other word, the special features about the OSISAF SST in the Baltic Sea have not been highlighted at current, which looks not to support the study focused on it.

We thank the reviewer for this comment. We used the OSISAF in this study for a couple of reasons. First, it is level 2 product and is retrieved directly from the satellite, which means there is no hind-analysis information included; Second, the OSISAF has a high resolution in the Baltic Sea, which makes it more suitable for our operational forecast system.

In the revised manuscript, we added a few sentences to clarify:

"For operational forecast, the SST from OSISAF is the most important dataset in the Baltic Sea because it differs from hindcast analyzed product like OSTIA (Operational SST and Sea Ice Analysis) data. As a level 2 product, the OSISAF SST has both good temporal and spatial coverage in the Baltic Sea. As there is no hindcast information included in the OSISAF SST, we are able to assess direct impacts of assimilating SST observations"

2) Lines 163-165, this SST product from AVHRR is available twice daily. It is not clear how to assimilate in the experiment. The **a**ssimilation time window is daily? How to calculate the innovation, is it asynchronous?

In section 3.1, we mention "only the subskin SST at night, which is comparable to in situ (buoy) measurement, is used …".

In the revision, we clarified how to calculate the innovation: "Further, we define a two-day assimilation window in the assimilation experiment. As a result, the observations in the two days before the assimilation time were used to calculate the innovation with observation operator. When we calculated the innovation we also changed the observation error according to the observation time by

$$\varepsilon = 0.4 \times \exp(-0.15\Delta t) \qquad (9),$$

here Δt is the absolute time difference between observation time and DA time."

3) In the first paragraph of 3.1, the assimilated SST has been filtered by the quality. But it is not clear how to consider the sea ice. Do you use the sea ice concentration of OSISAF to mask the SST product, and how to do?

We didn't use the sea ice concentration of OSISAF to mask the SST product. By the quality filter, we checked observation position, innovation relative to model result and the quality flag provided by OSISAF. If the model is covered by sea ice, the SST observation will be excluded.

4) The observation error for the OSISAF SST is important for this study, is it a constant of 0.5 degree used? As a good consistence check, some diagnostic about the assimilation stability like Rodwell et al. (2016) is beneficial to understand the system reliability and the observation error.

Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., Richardson, D. S., and Yamaguchi, M.: Reliability in ensemble data assimilation, Q. J. Roy. Meteor. Soc., 142, 443–454, doi:10.1002/qj.2663, 2016.

We agree that consistency check and assimilation stability are important for operational forecast systems with DA. We used a constant observation error similar to Rodwell et al. (2016) in this study, but our DA design is different from that paper. The major difference between these two studies is that we estimate the background error covariance from stationary ensembles and avoid the perturbation of observation error. Therefore, the diagnostic of the assimilation stability can be directly obtained from the forecast error, like the RMSE, in Fig.4, which shows comparable bias and RMSE in the assimilation and free forecast.

In the revision, we cited the Rodwell et al. (2016) and discussed the assimilation stability in section 6.

The corresponding text are added:

"Further, the reliability of the DA system is worth being assessed. In Rodwell et al. (2006), a perfect reliable system error variance for ensemble assimilation was calculated by the sum of the variance of the sample ensemble, the square of innovation (misfit between observation and model), the variance of observation at assimilation time. In this study, we used a constant observation error similar to Rodwell et al. (2016) because our DA design is different from that paper. The major difference between these two studies is that we estimate the background error covariance from stationary ensembles and avoid the perturbation of observation error. Therefore, the variance of the sampled ensemble and observation is univariate and the diagnostic of the assimilation stability can be directly obtained from the forecast error like the RMSE in Fig.4."

5) The IceMap has been used for evaluation as one independent SST observation. It is not objective and only twice for one week. In fact, another surface water temperature data set from SMHI collected by Ferry (http://www.smhi.se/hfa_coord/BOOS/Ferrybox/BSNI/BSNI-Wtemp.png) is more useful and independent for this study.

We agree that Ferrybox data is a very good source for model evaluation. In this study, we aim to evaluate the overall impact of OASIF SST product on the model forecast in the Baltic Sea. In this sense, the IceMap data is more preferable due to its spatial coverage and quality while the Ferrybox data has limited spatial coverage. At the same time, we have also used the independent in situ SHARK observations to verify the experiment results. The Ferrybox data may corroborate our conclusions but we think it is not a critical factor for our evaluation and conclusions.

6) The two in situ observations at Arkona and BY15 is super case to show the impact of assimilating SST only. It is valuable to do more specific analysis by diagnosing dynamic variables. Firstly, **i**nvestigating the mixed layer depth in the two runs can clearly show the mixing strength for Fig.5 and Fig.6. Secondly, the temp/salinity misfits in vertical can be shown and mutual authentication with the SHARK results.

We thank the reviewer for this important comment. To address the reviewer's comment, we compared the mixed layer depth in the two runs (Fig. 7) in the revised manuscript. We also used the SHARK data to examine the misfits of temperature and salinity at both inside and outside of the Baltic Sea(Fig.9).

In section 5.2, we added

"The mixed layer depth (MLD) was calculated at the Arkona and BY15 station and compared with the SHARK observation in Fig. 7. We used the temperature criterion to define the MLD, i.e., the depth at which the temperature deviated from the surface value by 0.5 ^{o}C (Fu et al., 2012). Figure 7 shows that the MLD at Arkona had larger variability relative to the MLD at BY15. The reason contributed to this feature is that the deeper water at Arkona is easy affected by wind forcing because of the shallow bathymetry and well mixing, whereas the temperature variation in upper water at BY15 difficulty influences the deeper water because of the strong stratification. Both runs had reproduced the MLD variability feature similar as the observations. For example, the minimum MLD appeared in summer, which was about several meters. The assimilation of satellite SST caused strong changes in the MLD at both stations, especially in winter. One explanation was that the Baltic Sea was largely affected by wind forcing and the winter wind was much stronger than the summer wind. Further, strong heating in summer promoted stratification in summer and shoaled the MLD."

7) Based on the current results, it indicates the salinity looks no remarkable improvement. However, the salinity peak in Sep 2010 at 7 m can be reduced by assimilation even this model run has an underestimation before. This event is a nice case to explore which factor contributes that positive correction.

We appreciate the reviewer's comment, but it is hard to attribute the improvement in September 2010 to a specific factor. There are a couple of reasons for this: firstly, at the depth of 7 m, the model salinity was strongly affected by the simulation of advection, mixing and E-P flux. Bias in any of these factors could contribute to the large bias especially after mid-September. In other words, any improvement of these factors also helped to correct the salinity bias. Secondly, the salinity at 7 m is generally decreased irrespective of the model bias, suggesting that the method is stable. Therefore, it is very likely that the improvement is a cumulative effect of our data assimilation, including the effect of the changes of circulation and mixing (shown in the mixed layer depth in Fig. 7).

8) Fig8 shows the vertical impact for temp/saln. It is better to separate into two parts internal and out of Baltic sea.

We separate the Bias and RMSE calculation in the two regions now. The figure caption of Fig. 8 was changes as Fig. 9.

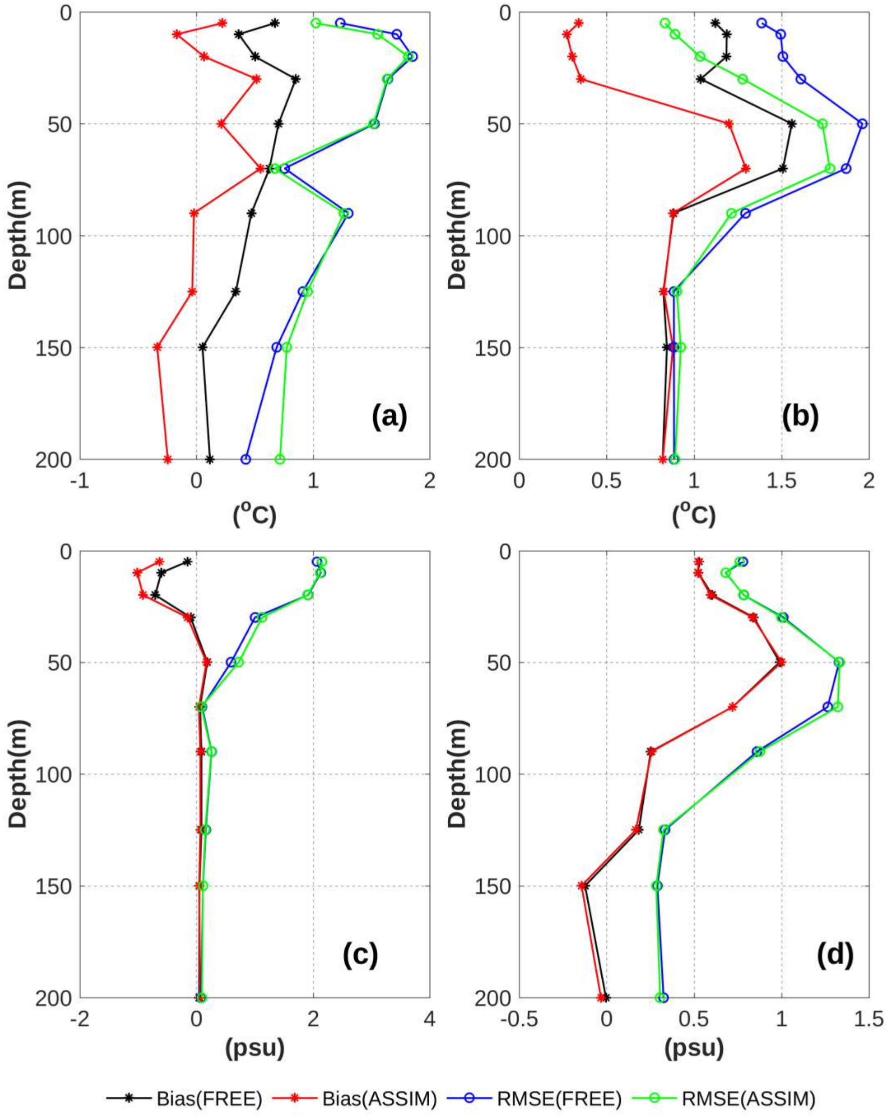

Figure 9. The overall RMSE and bias of temperature (up panel) and salinity (down panel) from FREE and ASSIM relative to observations as a function of water depth inside (b,d) and outside (a,c) of the Baltic Sea.

The corresponding text are changed:

"Figure 9 shows the change of overall bias and RMSE of T/S with depth against the SHARK dataset. In the Baltic Sea, DA had large impact on the temperature forecast in the water above 100 m. The RMSE showed that the forecast of temperature was obviously improved from surface to thermocline in the ASSIM and the improvements generally decreased with depth. Above 100 m, the overall RMSE of temperature in ASSIM was decreased by 21.38% (from 1.59 to 1.25 ^{o}C). It was also found the temperature error had similar variability as the warm biases in two runs. In the transition zone, the RMSE in the ASSIM was reduced by 5.59% and -

20.31% above and below 100 m relative to the FREE, respectively. Below 90 m, the temperature was also over-adjusted, which changed the warm bias to cold bias. It is worth noting that the number of the deeper water observation in the transition zone is substantially less than that in the Baltic Sea. For the salinity, both RMSE and bias of the ASSIM showed very minor changes relative to the FREE inside the Baltic Sea. For the water above 100 m, the total RMSE of salinity was increased by 3.48% (from 1.15 psu in the FREE to 1.19 psu in the ASSIM) in the transition zone and 1.04% (from 0.96 psu in the FREE to 0.97 psu in the ASSIM) in the Baltic Sea."

9) The impact on SLA looks very small so I suggest replacing the related figure and table by a short paragraph.

We thank your good comment. We removed the Table 1 and added a Figure to show the variation by DA.

"We calculated the RMSE and correlation coefficients for both the FREE and ASSIM against the observations from tide gauges (Fig. 10). The overall RMSE was reduced by 1.8% and the correlation coefficients were slightly increased. Among the stations, RMSE at the Oskarshamn was decreased by 5.6%, which is larger than that at other station. The minimum RMSE change of SLA was seen at the Klagshamn. For the correlation coefficient, improvement on the SLA by the DA is very small. Simrishamn station showed the biggest change of correlation coefficient, which is 1.1%. The RMSE and correlation comparison demonstrated that the SST DA has generally positive effects on the forecast of the SLA."

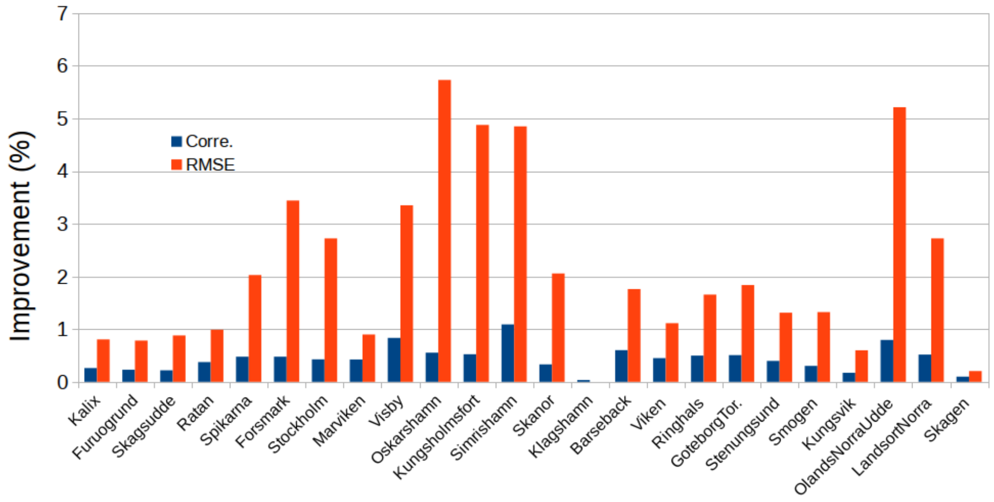

Figure 10. The improvement (%) of correlation coefficient and RMSE for the SLA at 10 tide gauges stations. The positions are shown in Fig. 8b.

Further, we also replaced the old figure 9 by Figure 11 to show the bias variation after data assimilation.

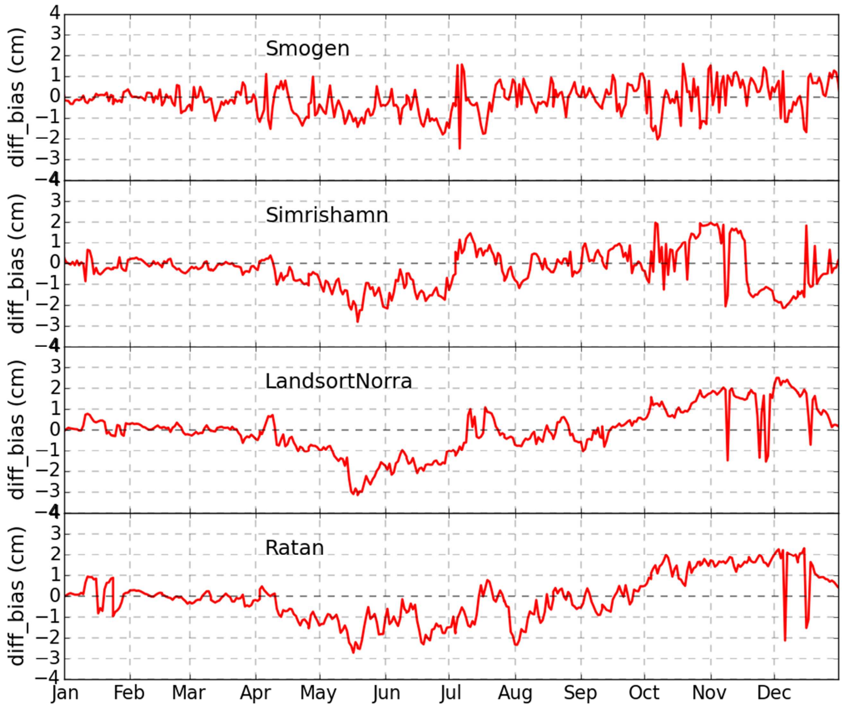

Figure 11. The difference of SLA biases between ASSIM and FREE against observations as a function of time at four observing stations.

10) Fig. 10 shows an improvement by assimilating SST. But the quantitatively comparison with the OSISAF concentration in the time series is helpful to know the impact in different sea ice seasons.

We added a new figure showing the comparison of monthly mean sea ice concentration in March and April and we also added the time series of the sea ice extent (SIE).

In the manuscript, we revised the text as:

"In March, compared to observation, the FREE produced low SIC in the western coast of the Bothnian Sea, Gulf of Finland, Gulf of Riga and the connect zone between the Bothnian Sea and Gulf of Finland. However, the model SIC in the FREE was higher than IceMap in the interior the Bothnian Bay. For instance, the SIC from FREE in the western Bothnian Sea was 40% higher than observation. In the south coast of the Arkona basin and Baltic proper, the FREE failed to reproduce the sea ice as in observation. After the DA, the high SIC was decreased in western Bothnian Sea and closer to that in IceMap in Bothnian Sea. In the Gulf of Finland and Gulf of Riga, the SIC error was increased in the ASSIM. In April, the large SIC error in the FREE was shown in the Bothnian Sea, the Bothnian Bay, Gulf of Rig and Gulf of Finland, where no clear improvements were seen in the ASSIM."

"The daily SIE from the FREE and ASSIM was compared with observations in Fig.13. The observed SIE was generally increased from January to February and reached the maximum in mid-February. During the period of March-May, SIE was decreased as temperature was increasing. SIEs in both the FREE and ASSIM experiments were generally underestimated by comparison with the observation in 2010, especially in the period from Mid-March to early April. The SIE bias in both runs was increased from January to early April. In early April, the maximum negative bias of SIE was found to be 105000 km^2 for the ASSIM and 10000 km^2 for the FREE. The impact of SST assimilation on the SIE was positive during the phase of sea ice formation. For example, the SIE bias was reduced 25000 km^2 at the end of February and in the Mid-December. However, during the phase of sea ice melting (March to April), SIE error was increased in the ASSIM even with the error of SST decreased. For example, the SIE bias in the

ASSIM was increased by 42000 km^2 relative to FREE in the early March. These increased SIE error in March mainly happened in the Gulf of Riga and Gulf of Finland (Fig.11)."

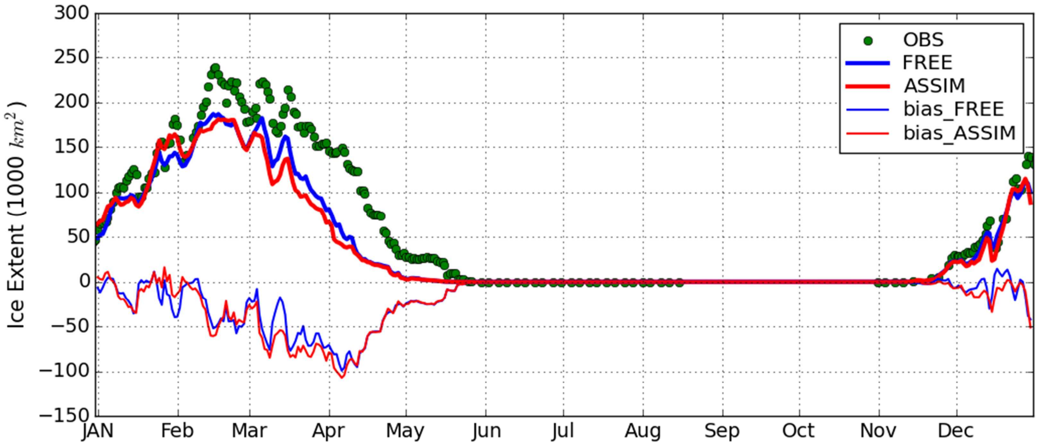

Figure 13. The daily sea ice extent from FREE, ASSIM and IceMap and the sea ice extent bias (modelled minus observed field), respectively.

Other small issues:

1) Line 137, the operator of Li in Eq. 3 has no illustration.

We added the line for the operator **L** illustration.

2) Line 159, "OSISAF product" is it means more general products or only SST?

To clarify, we delete " products are using in priority the European Meteorological satellites METEOSAT and MetOp and also several American satellites operated by NOAA, DMSP and NASA. Its"

3) Line 229, "model layer" replaced by "model level" because the model is not a layered model.

It was corrected.

4) Line 233, the forgotten factor is constant, or how to be defined?

We add a sentence " To define the forgetting factor, a one-month simulation experiment with varying the factor ρ was done in January 2010. At last, a factor ρ = 0.3 resulted in the best assimilation performance." At the end of Section 4.

5) Line 257, the evolution of SST based on 48-hourly local analysis. Does it mean all the SST comparison afterward use the 48 hourly forecast from the model?

Yes, we use the 48-hour forecast SST in the all comparison with observation.

6) Fig 1, the text is hard to identify. It is better to show the rivers involved in the model.

The two stations of Arkona and BY15 can be shown in Fig. 1 (or Fig. 7).

We add the Neva River and the position of Arkona and BY15 in Fig.1

7) Fig 6, the observed temperature at 70 m looks missing at Nov 2014, especially compared with other two depths or the salinity.

This temperature at 70 at BY15 station hasn't observation value at Nov 2010 in SHARK database.

8) Line 289, the obvious improvement in the Gulf of Finland. However, based on the snapshot of the observed SST distribution in Fig. 2 there are no observations.

The OSISAF observation at a specific basin may be missing like the Figure 2. Our Figure 3 is based on annual averaged IceMap SST comparison in 2010.

9) Line 278, "The model SST forecasts in both winter and summer (Fig2)". It is not corrected to say SST forecast in Fig. 2 because they only show the analyzed fields and the related increments, which not supports this conclusion.

Thank you. We changed it to "the SST DA has improved the simulated SST in both cases (Fig.2)"

10) Line 311, "The temperatures differ by about 15-22C between summer and winter" is confused. Does it mean the seasonal variability in observation?

We intended to show the seasonal variability. Since we only done one-year simulation, we delete the sentence "The temperatures differ by about 15–22 ^{o}C between summer and winter." to avoid confusion.

11) Line 314, "The reason perhaps ..." this kind of illustrations in this study require some proofs like MLD diagnosing or others.

We add the mixed layer depth analysis, which will support our conclusion.

12) Section 5.1, which mean ssh fields are used for tide gauges and the model simulations?

Since the mean SSH fields may be different from each other in the model and observation we do the comparison of SLA in this study. We calculated the mean SSH by directly averaging the tide gauges or model fields.