Review of the manuscript "An assessment of the variability in temperature and salinity of the Baltic Sea from a simulation with data assimilation for the period 1990 to 2020" written by Ye Liu, Lars Axell, and Jun She, submitted to Ocean Science journal.

The study presents a new dataset of the physical parameters of the Baltic Sea. The dataset is covering the period 1990-2020 (30 years). It is produced using a regional circulation model for the Baltic Sea (NEMO-Nordic setup) with a data assimilation (DA) technique called the local Singular Evolutive Interpolated Kalman filter (LSEIK), assimilating both in-situ T/S profiles and satellite products. To estimate the performance of the above-mentioned model setup, the study provides some validation of the results against in-situ T/S profiles, sea surface elevation from tide gauges, SST from a satellite product, and mixed layer depth (MLD) derived from in-situ observations. The validation is based on well-known quality metrics. In addition, the study investigates seasonal, inter-annual, and multidecadal dynamics of the water temperature and salinity produced by the authors' model setup at different vertical layers (surface, 60m, 100m, and bottom). By comparing model runs with and without DA, the study shows that DA improves the model quality in the Baltic Sea region. Despite some persisting uncertainties, the authors are convinced that the dataset demonstrates exemplary performance and can be utilized for further research. The study also

identifies a warming trend across the Baltic Sea, with the temperatures at intermediate depths increasing faster than at the surface. As for salinity, the study highlights opposite trends in SSS across the Baltic Sea, with slight freshening in the northern part of the sea and slight salinization in the southern part. At the same time, the authors observed a notable positive trend in salinity below the halocline in the Baltic Propper. Based on the comparison with other studies and products, the authors conclude that the temperature and salinity trends are comparable to the estimates provided earlier by the community.

We thank you for your valuable comments. We have implemented all requested changes. Please find our detailed response in blue below.

The study is well-structured and generally well-written. However, I have some comments on the general scientific questions it tackles. If the study's objective is solely to present and validate the new dataset, which can be inferred neither from the title nor the chosen journal, the other journal is probably the better choice (e.g., Geoscientific Model Development from the same publisher).

The study aims to provide reliable estimates for long-term and seasonal variations of temperature and salinity (T/S) in the Baltic sub-basins during the past three decades. By utilizing the reanalysis dataset, we demonstrated a reliable and gapless variability of T/S at various depths within the Baltic sub-basins for the past three decades. Therefore, validating the reanalysis dataset is a critical component of this study, ensuring its reliability and robustness.

To be published in the current journal, the authors need, in my opinion, to invest more work into the paper. First of all, I would appreciate a more detailed comparison with, e.g., the current publicly available CMEMS Baltic Sea physics reanalysis (https://doi.org/10.48670/moi-00013) covering the period from 1993 to 2023, which is the same duration as for the presented reanalysis, or possibly with other state-of-the-art Baltic Sea reanalyses (one could include it to the validation part). This analysis would demonstrate the added value of the new reanalysis described in the study.

We added a detail comparison of this study to the CMEMS Baltic physical reanalysis product:

” Although both this reanalysis and the current CMEMS Baltic Reanalysis product (CBRP, BALTICSEA_MULTIYEAR_PHY_003_011) covered the past three decades, they differ significantly in model setup, atmospheric forcing, DA method, and assimilated satellite observation sources. Unlike this study, CBRP is based on NEMO-Nordic 2.0 with a resolution of 1 nm (Kärnä et al. 2021), as well as a coarser atmospheric forcing (ERA5 dataset with 31 km resolution), and boundary conditions from the CMEMS North West Shelf multi-year product. This study uses a multivariate EOF method to generate a sample ensemble of background error covariances, which enhances the identification of key large-scale variability patterns while minimizing noise and reducing dimensionality. This method allows for sampling various combinations of patterns from the ensemble space, enhancing the diversity of the ensemble and improving DA accuracy, leading to a more reliable simulation. Furthermore, the study assimilated the satellite observations from OSISAF SST and IceMap, whereas the CBRP used CMEMS SST observations (SST_BAL_PHY_L3S_MY_010_040). When compared to the CBRP, the sea level anomaly in this study exhibits a stronger correlation with observations, except in Visby, Talinn, Grena, and Viken. A comparison of CFs revealed both datasets have comparable salinity quality at the same Baltic stations. However, there are notable discrepancies in SBS in both the southern Bothnian Sea and the southwestern Bothnian Bay. In the southwestern Bothnian Bay, the quality of SBS reported in this study was classified as good, whereas the CBRP was classified as poor. Conversely, in the southern Bothnian Sea, the CBRP data was of higher quality than the SBS in this study. Regarding temperature, the CBRP was of high quality across all stations, while this study reported slightly lower quality in the southern Bothnian Sea, though still within acceptable limits. Lastly, all reanalysis provide valuable insights into the Baltic Sea’s physical conditions. ”

The current analysis of temperature and salinity variability and its discussion seem to be a continuation of the reanalysis validation. In my opinion, the study, in its current form, does not deliver any new insights into T and S variability in the Baltic Sea. I think the interesting part is an investigation of T and S variability at different depths, not just at the surface, but this part needs to be improved, e.g., authors could elaborate more on the causes of the strong trends in T and S at the intermediate depths in the Baltic Propper (was it just the inflow activity or something else). In addition, some analyses and results are also currently questionable (for a detailed explanation, see the comments below). Therefore, I suggest the authors undertake a major revision and resubmit the revised manuscript. Below, I provide a few detailed comments on the manuscript:

We greatly appreciate the referee’s thoughtful feedback. While we understand that the current analysis of temperature and salinity variability may appear as an extension of the reanalysis validation, the primary goal of this study is to provide reliable estimates for long-term and seasonal variations of T/S across the Baltic Sea, particularly in the sub-basins and at various depths.

We agree that focusing on the variability of T/S at different depths—beyond just the surface—offers important insights. This is a key strength of our study, and we value the referee’s suggestion to explore this aspect further. We have revised the manuscript to provide more detailed discussion and context regarding the causes of the strong trends in T/S, especially at intermediate depths in the Baltic Proper. These trends may result from a combination of factors such as inflow activity, changes in stratification, and other regional dynamics. We have added more focus on these causes in the revised manuscript.

We also acknowledge that some aspects of our analysis and results may require further clarification. We are committed to improving the manuscript in response to your suggestions. Below we highlight the key findings of our study that directly address the study's objectives and make a significant contribution to understanding Baltic Sea dynamics:

- **Filling the knowledge gap**: Most previous Baltic Sea variability studies have focused on surface or bottom trends, with only a few addressing intermediate temperature trends (e.g., Liblik and Lips, 2019; Dutheil et al., 2023). Our study fills this gap by providing a **comprehensive, gapless analysis** of temperature and salinity across **four depths** in the Baltic Sea, with a particular focus on the **intermediate waters**. This is especially crucial for data-sparse regions where observations are limited by coverage and model biases.

- **Trends at intermediate depths**: Our analysis revealed that **temperature and salinity fluctuations have been more pronounced in recent years**, particularly at intermediate depths (e.g., at 100m). These strong trends highlight the critical role of mid-depth dynamics in shaping the Baltic Sea's response to climate and oceanographic changes, which has not been fully explored in previous studies.
- **Impact of inflows and other factors**: Our findings underscore that the **strong trends** observed at intermediate depths in **the Baltic proper** are likely driven by a combination of inflow activity, changes in water column stratification, and other local factors. We provide more detailed discussion on how these factors have contributed to the observed trends in the revised manuscript.

- **Dataset's significance**: The dataset presented in this study provides a **comprehensive reference** for analyzing temperature and salinity at **any depth** in the Baltic Sea. By extending the time period and offering gapless coverage, it enables researchers to **analyze long-term and seasonal variations** beyond short-term observations, which can help improve predictions and environmental monitoring for the Baltic Sea.

- **Environmental implications**: Our study highlights a combination of **warming and increased salinity stratification**, suggesting that the Baltic Sea is experiencing more pronounced **layering** between surface and deep waters. This enhanced stratification could have significant ecological implications, such as reducing oxygen transport to deeper layers and **exacerbating hypoxia** in deep basins. These changes could negatively impact marine **ecosystems and fish populations**, disrupt **nutrient mixing**, and alter **primary productivity** and **biogeochemical cycling**.

We hope these changes address the referee's concerns and provide a clearer emphasis on the study's contributions and findings. We are confident that the revised manuscript, with an expanded discussion of the causes of T/S trends, will enhance the understanding of variability in the Baltic Sea and make an important contribution to the field.

L21: "its significant impact on the Nordic climate change": I would reformulate the sentence since the Baltic Sea is considered to be significantly driven by external forcing, not the other way round (see, e.g., Stigebrandt & Gustafsson, 2003).

We appreciate the comment and have revised the sentence to "The variability of the Baltic Sea has received considerable attention from both the scientific and political-economic communities due to its significant response to external forcings, which play a substantial role in influencing the Nordic climate change."

L23-24: “The Baltic Sea exchanges with the North Sea through the Danish Strait in the transition zone.”: Danish StraitS. I would also reformulate this sentence, e.g., The water exchange between the Baltic and North Seas happens through the shallow and narrow Danish Straits.

Thanks, we revised it.

L27: “(Matthäus et al. 2008).”: I think a reference to newer research could be added, e.g.,

(Mohrholz, 2018).

We updated the reference and added Mohrholz, 2018 and Lehmann et al. 2022.

L31: “The Baltic Sea exhibits various features over time due to local changes in climate and forcing”: How do the changes in climate differ from the changes in forcing? The sentence does not sound right to me; please reformulate it.

Climate changes refer to long-term atmospheric and oceanic shifts, while forcing includes external factors, both natural and human-driven, that can occur on different timescales, ranging from seasonal to decadal variations. Therefore, we revised the sentence to “The Baltic Sea exhibits various features over time due to both gradual shifts in climate and changes in external forcings.”

L36-37: “a comprehensive climate assessment directly from the observations is still lacking due to the spatiotemporal coverage limitations, especially in the deep Baltic Sea.”: I think most observations are focused on the deep Baltic Sea, where most monitoring stations are located. Therefore, I would assume assessing the climate impact on the coastal areas is more challenging.

While most observations are concentrated in the deeper regions of the Baltic Sea, where the majority of monitoring stations are located, these stations do not provide full spatial coverage of the entire Baltic Sea. As a result, using these observations alone to assess climate impacts in the subsurface Baltic Sea can only offer a general understanding. This approach does not capture the detailed variability in areas with sparse or no observations, limiting the comprehensiveness of the climate assessment. We revised the text:

”a comprehensive climate assessment based on the observations alone remains limited due to gaps in the spatiotemporal coverage , especially in subsurface waters of the Baltic Sea.”

L92: “The spin-up run (1 January 1975 – 31 December 1989)”: Could you please say a few words about the spin-up duration? I would say it is rather short.

We revised the text to

“The spin-up run (1 January 1975 – 31 December 1989) was initialized using the stable restart file from Hordoir et al., (2019), with the same model version and configuration. Although the restart file represents a stable state, the simulation was continued to ensure the system adjusts to any differences in boundary conditions or forcing, allowing the model to fully stabilize within our specific setup.”

L119: “It should be note”: It should be noteD.

Corrected.

L138-139: “of 3 oC or 3 PSU, respectively”: Please comment on the thresholds. I would assume the one for salinity should be lower since its variability is lower than that of temperature.

These values are used to identify the differences between the model and observations, based on a rough comparison between the model (without data assimilation) and observational data. If the difference between the model and observation exceeds the thresholds, we do not assimilate that observation into the model, as large adjustments could lead to instability. It is important to note that the values for salinity and temperature may differ from one another depending on the comparison between the model and observations.

Eq.5: Are you sure it should be -0.005 under exponent, not +0.005? Now, it looks like the error variances are decreasing after 70m, which contradicts the text.

The value should be -0.005 because we assume that observation error decreases with increasing depth. This assumption allows us to assign more weight to observations relative to the model values in the data assimilation process. By doing so, we aim to adjust the model more significantly in deeper waters, where observation errors are considered smaller.

L157: “oN”: Typographical error.

Corrected.

L170: “low-resolution”: In the ICES dataset, they are called “high-resolution CTD data.”

Most of the ICES station observations have relatively low temporal and spatial resolutions. For instance, the vertical resolution of ICES observations typically ranges from 5 to 25 meters, and the temporal resolution averages about 3 to 7 profiles per month. While certain stations in the ICES database offer higher resolution, the majority of stations still operate at lower resolutions.

L173: “the SHARK and ICES data center”. Why only those two? There is also, e.g., the IOW dataset for the Baltic Sea (https://odin2.io-warnemuende.de/).

our research is constrained by the exclusion of certain available observations for assimilation. We used the SHARK and ICES data, as these two data centers contain most of the important stations in the Baltic Sea.

We point out this limitation of this study by revised the text:

“it is important to acknowledge that our research is constrained by the exclusion of certain available observations for assimilation. For example, we haven’t assimilated the Baltic Sea- In Situ Near Real Time Observations (INSITU_BAL_PHYBGCWAV_DISCRETE_MYNRT_013_032), consist of level 2 data sourced from various platforms such as ferry boxes, gliders, and Argo floats. However, the presented study has already assimilated the T/S profile observations from important Baltic mooring stations in the Baltic Sea- In Situ Near Real Time Observations. To the surface satellite data, we utilize a coarse level 2 satellite product that aligns more closely with the horizontal resolution of the model employed in this study. Further, the satellite SST assimilated in current Baltic CMEMS reanalysis serves as a benchmark for validating the reanalyzed SST of this study. “

L177: “increases significantly with time.” This conclusion contradicts Fig.2, which shows a substantial gap in 2012.

We revised it to “There is a significant trend towards an increase in the number of observed profiles over time.”

Eq.7: Add “=”

Added.

Eq.12: You use the standard deviation of the observations to calculate the cost function. It might affect your estimates if the observations are, for example, seasonally biased or if there are only a few profiles available. It may be better to use the model standard deviation instead. I assume the model should capture the variability more or less correctly.

We used a modified cost function by Eilola et al. (2011) to calcuated the assessement values. The equation has been corrected: $C = \frac{\overline{|m-O|}}{\sigma_O}$. Therefore, the standard deviation of the observations is suitable for this calculations. We agree with you that the assessments are influenced by the observation's distribution. We have discussed this point.

L240-241: "The salinity shows a stronger stratification at the eastern Gotland Basin (BY15) compared to the Bornholm Basin (BY5)." Based on Fig.4, I would say the opposite is the case since the vertical salinity gradient seems stronger at BY5 compared with BY15. Please check.

Thanks to point out this inaccurate statement. The Baltic Sea is strongly stratified, both by salinity and temperature. Although the salinity gradient at BY5 is larger compared to that at BY15, the stratification at BY15 is still stronger than that at BY5.

To clear, we removed this sentence: "The salinity shows a stronger stratification at the eastern Gotland Basin (BY15) compared to the Bornholm Basin (BY5)."

L242: "Compare": CompareD.

Corrected.

L246: "As shown in Fig. 3, error of DA simulation is very small": Shouldn't you refer to Fig.4 instead?

Thanks, we corrected it.

L315-316: "using a seawater density criterion to define the MLD as the depth at which the seawater density deviates by 0.03 kg m^{-3} from the surface value (Chrysagi et al., 2021)." I suggest adding the original publication on this method: https://doi.org/10.1029/2004JC002378. In addition, please clarify whether you used in-situ or potential density. To remove the effects of pressure, the density should be potential. But I think the difference should be very minor in the Baltic Sea.

We added the reference:

Montégut, D. B., Madec, C., G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the globalocean: An examination of profile data and a profile-based climatology, J. Geophys. Res., 109, C12003, doi:10.1029/2004JC002378., 2004.

We made it clear that the density for calculation is the in-situ density.

L322-323: "The MLD was shallower in the far ends of the elongated Baltic Sea, including the Bothnian Bay, the Arkona Basin, and the Gulf of Finland, compared to the MLD in the central Baltic Sea": Is it only due to the different depths?

The MLD in the far ends of the elongated Baltic Sea is smaller than that in central Baltic Sea has been caused by various factors, such as reduced wind forcing, freshwater input from rivers, and the influence of regional atmospheric conditions, which can limit vertical mixing in these areas.

L345: “intermedia”: intermediaTE.

Corrected.

L349: “well mixing”: wrong use of English language.

We revised the text and removed this sentence.

L352: “This pattern indicates a clear temporal variation of the T/S trends in the Baltic Sea.”: I would add a reference here.

We revised the text and added a reference:

“This study showed that the temperature and SSS of the Baltic Sea decreased between 1990 and 1995, , followed by a increase between 2010 and 2015, reflecting clear temporal variations in the T/S trends in the Baltic Sea between 1990 and 2015 (Kniebusch et al, 2019).”

L378-379: “the tendency for increased salinity was more pronounced at deeper waters, as evidenced in both the Baltic Proper and the Bornholm Basin.”: Please add additional analysis on that topic.

We have revised the text and re-analysis the trends distribution.

- The warming was least pronounced at the surface, while the most significant warming occurred at mid-depths (100 m), suggesting warming is not solely driven by surface heat exchange but also by internal processes, such as stratification and reduced vertical mixing.
- The increased salinity in deeper waters of the Baltic proper is primarily due to the stratification of the water column, the limited vertical mixing between surface and deep waters, and the inflow of saltier water masses from the North Sea, which tend to sink and accumulate at greater depths.

L381: “notable, significant”: a general comment. I would estimate the statistical significance of the trends presented in the paper and discuss them accordingly.

Thanks, we revised the figure and text by adding the statistical significance of the trends.

L427: “On data assimilation scheme, please write”: please write.

We removed these words.

L431: “Fig. 3”: maybe Fig. 4?

Corrected.

L436: “Fig. 4”: maybe Fig. 3?

Corrected.

L465: “Fig. 3”: maybe Fig. 2?

Corrected.

L474: “which well match with observed values,”: English language

We revised it to “adding observational information through DA improves the simulation accuracy and provides a reliable distribution of T/S.”

L479: “to accurate representation”: English language

we revised it to “as smaller or unevenly distributed datasets might limit the model's ability to accurately represent T/S dynamics across the studied region”

L480: “data is”: data are

corrected

L493: “which is helpful to constrain the bottom simulation”: I didn’t understand what the bottom simulation is.

We revised it to “the study introduces a more advanced approach by allowing the observation error to vary with water depth for a given profile, improving the accuracy of the bottom T/S simulation of the Baltic Sea using an ensemble DA method, such as LSEIK.”

Table 1: The decrease of salinity at intermediate depths looks pretty suspicious. This means that, generally, the stratification is unstable, which cannot be accurate. The same problem is visible in Fig.9. One possible explanation could be that the shallow transition region was included in the averaging at the surface but not at the intermediate depth. To avoid that and account for the different conditions in different Baltic Sea regions, I recommend providing the analysis for different basins of the Baltic Sea separately (the HELCOM definition can be applied here). In addition, I would calculate trends at all available depths and provide the vertical profiles of trends rather than focusing on specific vertical layers. If the authors still want to focus on particular depths, they should, in my opinion, account for area differences when averaging data. Are trends statistically significant?

Thank you for your insightful comments. In the original version, we inadvertently included the shallow transition zones between the North Sea and the Baltic Sea in our calculations, which may have influenced the results. Based on your suggestion, we have now recalculated the values, restricting our analysis to the Baltic Sea region east of 13°E longitude. This adjustment ensures a more accurate representation of the Baltic Sea’s conditions.

To address your concern about stratification and to improve the clarity of our analysis, we have now provided separate trend analyses for different sub-basins of the Baltic Sea, using the HELCOM definition. The trend maps are now presented at four distinct depths: the surface, two intermediate depths (shallow and deeper than the permeant thermocline/halocline), and the bottom layer. We believe these four depths offer a comprehensive representation of the Baltic Sea's vertical variability.

Additionally, we have included statistical significance checks for the trends. The updated table and figure reflecting these changes are provided below.

Table 1. The temperature and salinity variability in the Baltic Sea over the period 1990-2020.

Parameter	Trend [year $^{-1}$]	Mean	Maximum	Minimum
SST [^{o}C]	0.037±0.010	8.16±0.60	20.08±1.54	0.67±0.54
T60 [^{o}C]	0.044±0.008	4.16±0.55	6.76±0.52	2.31±0.62
T100 [^{o}C]	0.051±0.006	4.77±0.54	5.53±0.57	4.09±0.51
SBT [^{o}C]	0.041 ±0.008	5.34±0.53	10.76±0.51	1.96±0.49

SSS [PSU]	-0.004±0.002	5.83±0.10	6.50±0.10	5.01±0.20
S60 [PSU]	0.026±0.003	7.48±0.28	8.48±0.42	6.54±0.13
S100 [PSU]	0.049±0.005	9.16±0.51	9.63±0.55	8.61±0.51
SBS [PSU]	0.015 ±0.003	7.39±0.19	8.34±0.26	6.54±0.15

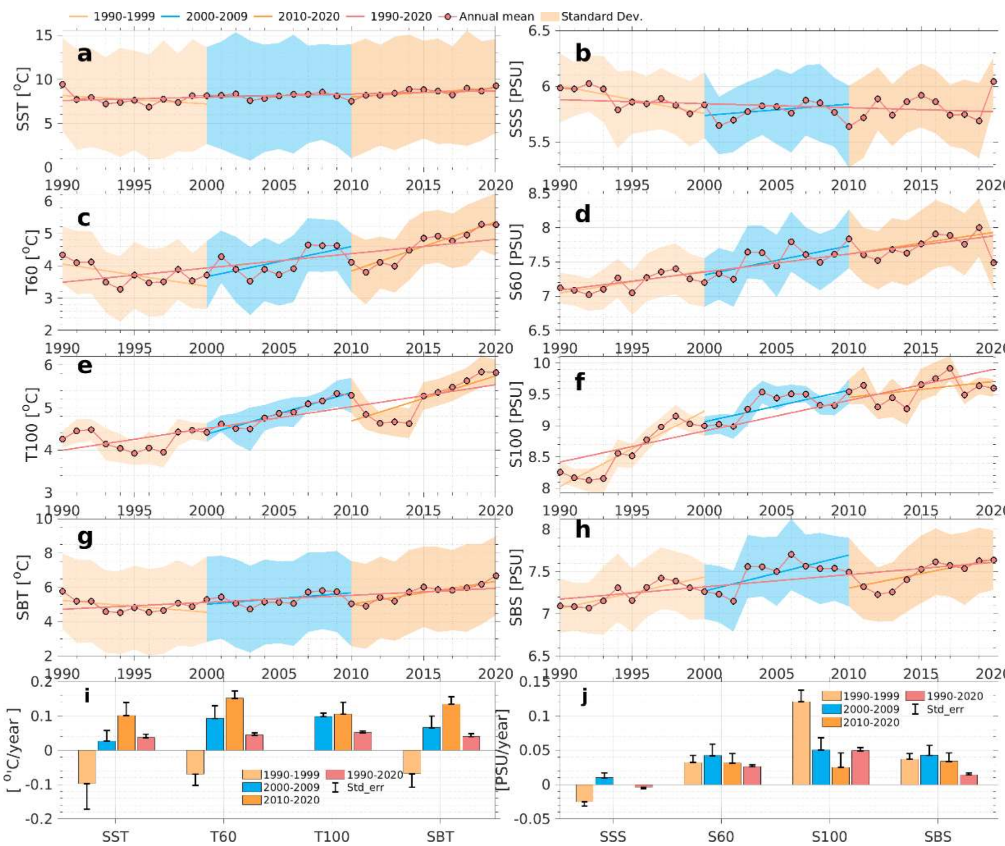
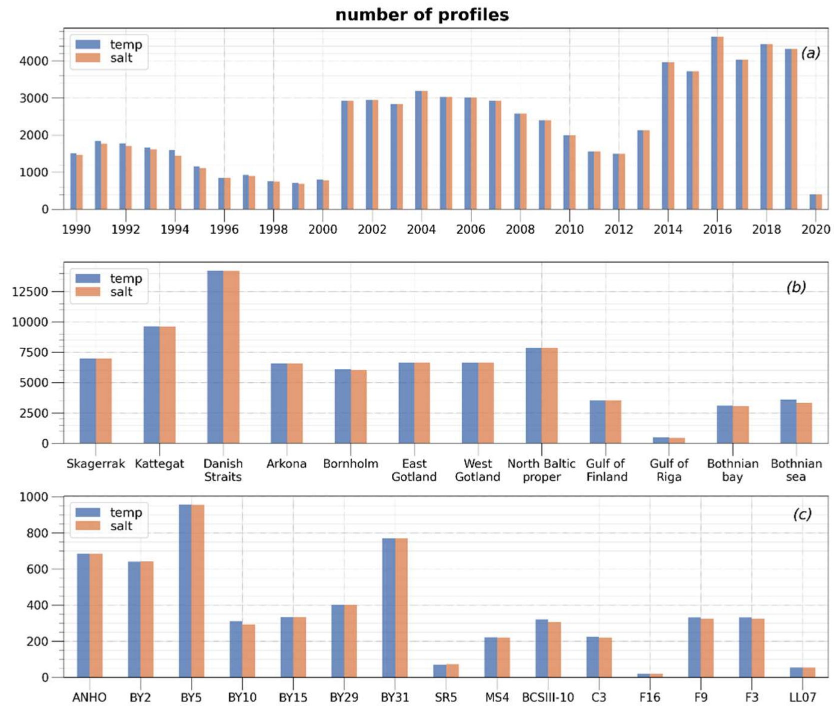


Figure 10. The annual mean temperature (left panel) and salinity (right panel) at surface (a,b), 60 m(c,d), 100 m (e, f), and the bottom (g, h) of the Baltic Sea for the period 1990-2020 and their linear decadal trends (i, j). Linear trends are showed by solid lines and the standard deviations are showed by the shaded area.

Figure 3: A typo in the figure caption. There is no panel C in this figure. I suggest adding colorbar labels as well.

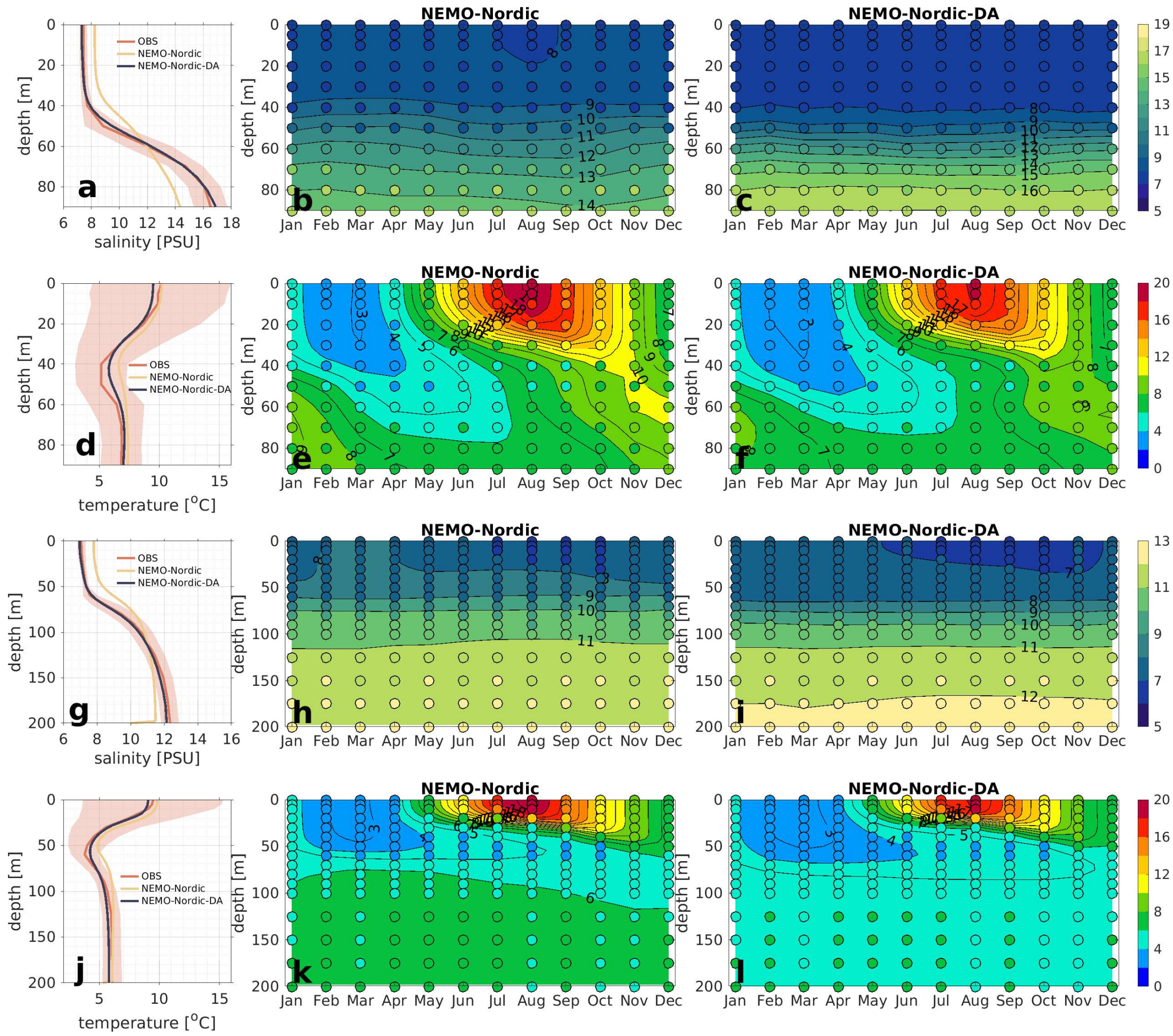
We added labels for Panel C and the yaxis in Figure 3 shows the size of the bars. Thanks for your suggestion. Since the yaxis shows the size of the observations and the grid lines make it easy to identify their size. The color bar is not very useful for identifying the size of the observations. the figure is revised to



Annual number of profiles (a), total number of profiles in the Baltic sub-basins (b), and number of profiles at the selected stations (c) for the period 1990-2020.

Figure 4: I suggest adding captions to the plots (station name for the profiles and an indication of whether the simulation was with or without DA for 2D plots). In addition, I would make colorbars equally spaced to facilitate the discussion of stratification, etc., in the text.

Thanks for your suggestion, we revised the figure and added the captions:



Monthly, seasonal and period averages of salinity and temperature at BY5 (a-f) and BY15 (g-l) for the period 1990-2020. Time averages (a, d, g, j) are shown over the entire water column and the standard deviation of observations over the period are shown as a grey shaded area. The seasonal variable simulated from model with (c,f,i,l) and without (b,e,h,k) data assimilation are compared with the observational monthly average values (the filled circles).

Figure 5: Please add labels.

Added.

Figure 7: Labels as well.

Labeled.

Figure 8: Also, colorbar labels. In addition, instead of "size of the mixed layer depth," I would say "mixed layer thickness."

We revised this figure.

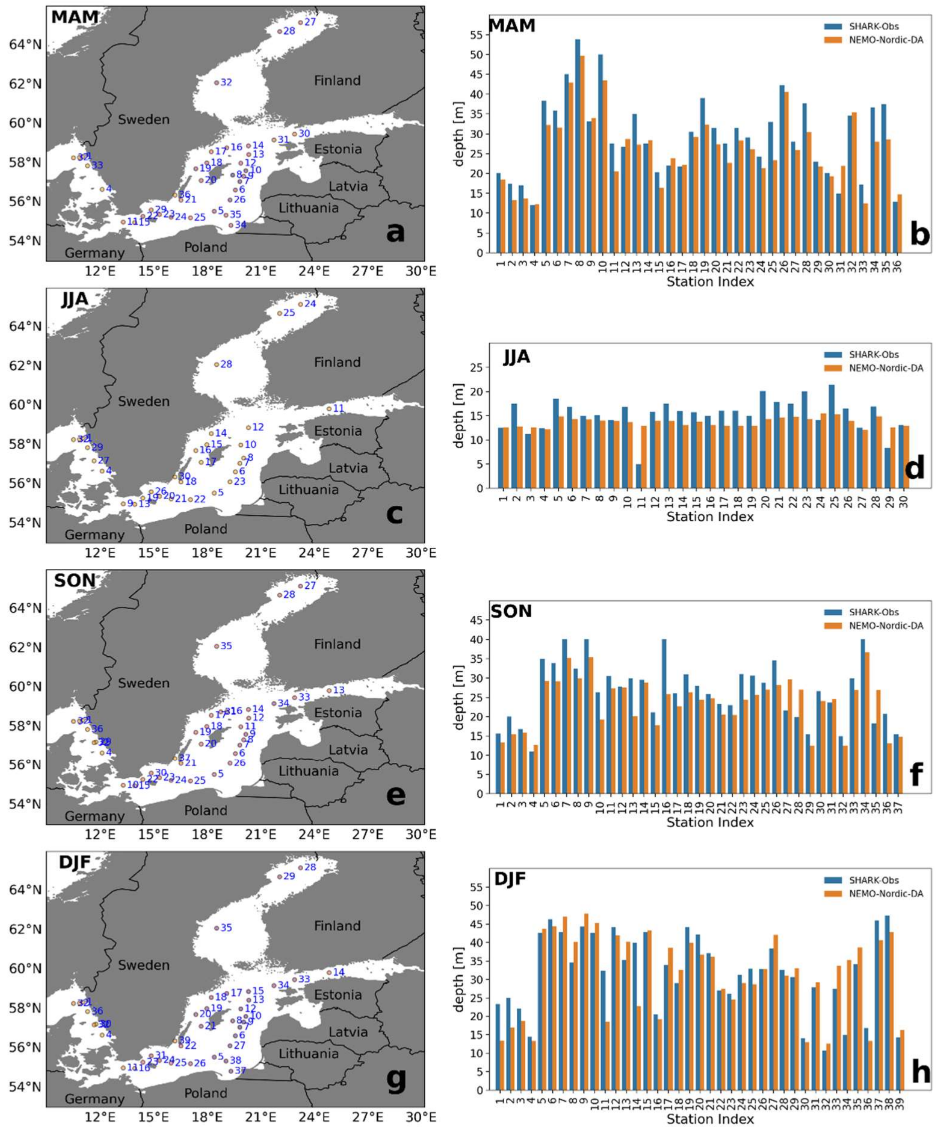


Fig 9. Seasonal mixed layer depth for the period 1990-2020: MAM: March-April-May; JJA: June-July-August; SON: September-October-November; DJF: December-January-Feburary. The left panel shows the station positions and the right panel shows the size of the mixed layer thickness.

Figure 9: Same as in Table 1. In addition, I would add an indication of seasonal variability on that kind of plot (e.g., +- sigma). Why is the salinity in [g kg-1] in this figure? I suggest replacing the label with PSU since, in the model, the EOS-80 equation of state was used.

We replaced the label with PSU and added the variability.

Figure 10: Trends significance and some discussion on observed results.

We revised it to a significant marked figure.

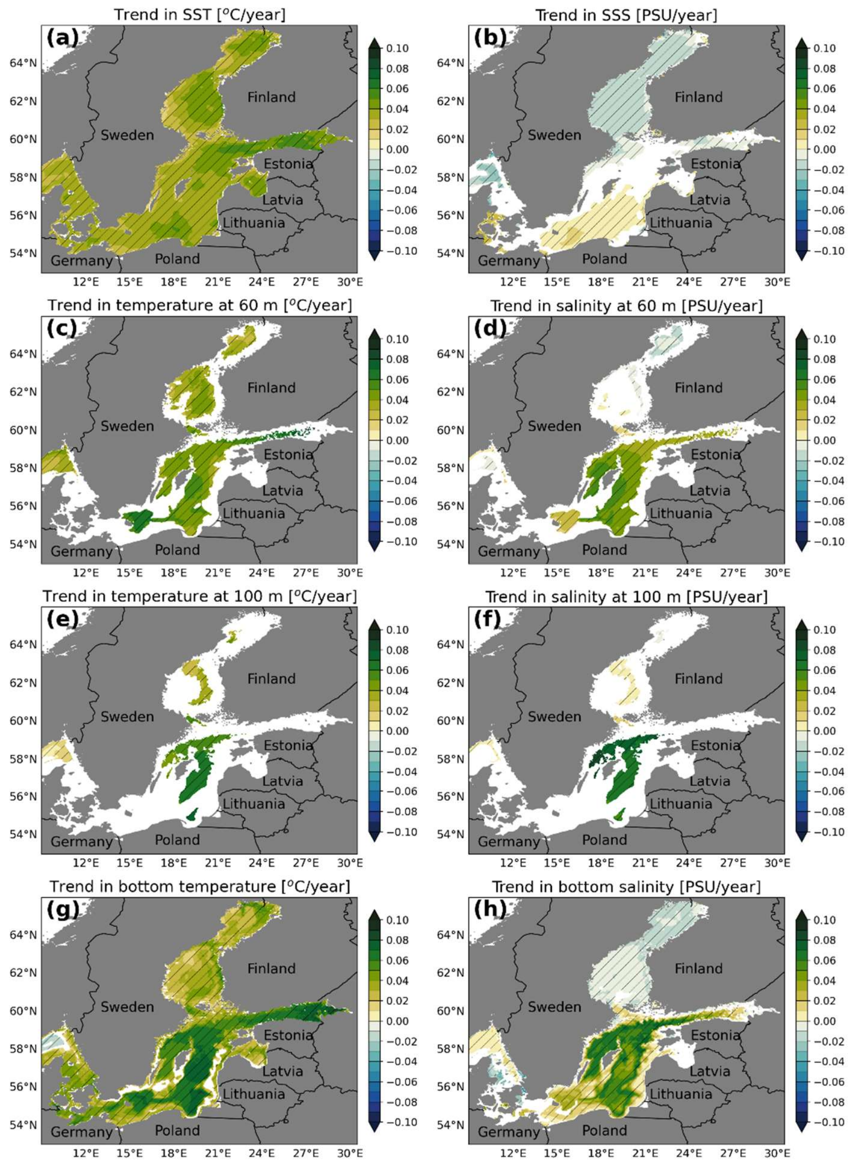


Figure 11. Linear trends in the water temperature and salinity in the Baltic Sea during the period 1990-2020, derived from the reanalysis. The Areas with a significance level are hatched (p values lower than 0.05), and areas with no significance (p values larger than 0.33) are excluded (white).