Authors' Comments to Reviewer 2

Reviewers comments formatted in italic dark red font.

Our response is formatted in thin black font.

Summary: The presented study analyses interannual variations of ocean heat content (OHC) and freshwater content (FWC) of the Baltic Sea based on modern data analysis techniques (random forests) and puts these into perspective to other state variables of the ocean and atmospheric forcing. The underlying data consist of reanalysis products, covering the period 1993-2003. The authors highlight that interannual basin mean FWC variations refer mainly to FWC variations in the halocline, while OHC variations refer mainly to both, variations in the seasonal thermocline and upper halocline.

In a second step, the study elaborates on the potential impact of atmospheric forcing and some oceanographic factors. The authors suggest that similar analysis techniques might be applicable also to other regions.

Major Comments:

The subject of the study is interesting and I appreciate the use of modern data analysis techniques. Also, the manuscript is well written. Still, I have some points which need in my eyes to be addressed:

(1) My major concern is that I find the fitting procedure for the random forest models not well explained. Generally, I would expect distinct data for the fitting procedure and for testing the fit - to evaluate how well the RF-model generalizes to unseen observations. I am not sure if this has been done. Also, I am not sure how the hyperparameters were chosen and would appreciate some more details (e.g., the number of trees, tree depth, minimum samples for splits).

We will have a separate subsection explaining implemented RF models in detail in the Methods section.

To optimize the performance of the Random Forest (RF) model while ensuring robustness and generalizability, a set of hyperparameters was selected based on sensitivity analysis. The minimum leaf size (MinLS) was set to 1, allowing the trees to fully grow and capture complex data patterns. The number of predictors to sample at each split (Pred2Samp) was dynamically determined as one-third of the total number of predictors, tackling a balance between feature randomness and predictive strength. This approach promotes diversity among trees while preventing excessive correlation. The number of trees (NumTrees) in each RF model was set to 100 (10^2), providing sufficient ensemble stability while maintaining computational efficiency. Additionally, Out-of-Bag (OOB) error estimation was enabled to assess model performance without requiring a separate validation set. Feature importance was evaluated using OOB permuted predictor importance, identifying the most influential features in the prediction process.
To further enhance predictive reliability, assess uncertainty, and evaluate the stability of both predictions and feature importances, an ensemble of 150 independently trained RF regression models was constructed. The final prediction was obtained by computing the mean output across all ensemble

members, reducing the impact of individual model variations. Similarly, the feature importance scores were averaged across the ensemble, ensuring a more robust and reliable interpretation of feature contributions.

Additionally to number of trees we have conducted additional sensitivity on leaf size hyperparameter, which have improved the fit-of the predicted timeseries capturing the variability extremes better. The RMSD's have changed from ~0.35 to ~0.2 .

(2) I must admit that I find it sometimes difficult to keep overview over all the RF-models. I would find it helpful if the many random forests would be described in an extra sub-section in the Methods and/or a table of all RF-models mentioned in the text might be nice (including some respective measures of the goodness of fit).

We will have a separate subsection explaining implemented RF models in detail in the Methods section.

In this study we use four different RF models: RF1 for Ocean Heat Content (OHC) with meteorological variables, RF2 for Freshwater Content (FWC) with meteorological variables, RF3 for OHC with temperature profiles T(z), and RF4 for FWC with salinity profiles S(z).

Specific Comments:

Introduction, Ln.38ff: Winsor et al. (2001) and Rodhe and Winsor (2002) might be interesting to mention here for FWC. For OHC Dutheil et al. (2023) might be interesting.

We are adding a sentence referring to the paper by Winsor et al. (2001) and Rodhe and Winsor (2002). Also we include reference to Dutheil et al. (2023).

Ln. 55/56: I would suggest rather to talk about Granger-causality here (https://en.wikipedia.org/wiki/Granger_causality) because statistical relationships cannot identify "real" cause-and-effect relationships.

We modify the sentence "*The third stage is analyzing the forcing functions and integral state characteristics together, which enables identifying potential causalities between them.*" We would like not to use the term Granger-causality here, because we do not perform the Granger causality tests. Granger causality and random forest are different in their approaches to understanding relationships between variables. Traditional Granger causality assumes linear relationships between variables. Random forest can model non-linear dependency. Indeed, Granger causality and RF can be used in combination.

Line 79ff: It might also be interesting to add that many studies reported a strong warming of the Baltic during the past decades when compared to the global oceans (e.g. Kniebusch et al, 2019).

We will refer to previous studies about strong warming of the Baltic Sea.

Line 99ff: I would find it nice to get a bit more background information on FWC and OHC – why it is considered so important and how it's calculated? (e.g. I guess the reference salinity from Raudsepp et al. 2023 has been updated?).

We will explain the importance of FWC and OHC in the revised manuscript. We will include a more detailed calculation of FWC and OHC. In general, OHC is a quantitative measure of energy and provides a robust indicator of changes in the Earth's climate system. Mean water temperature lacks the depth and precision needed for understanding the full impacts of energy changes in the ocean. Ocean FWC provides a more detailed, physically relevant measure of the addition or removal of freshwater, making it a better choice for studying climate change, hydrological cycles, and their impacts on ocean dynamics. Mean salinity gives a broad idea of salt concentrations.

Methods, Random Forest: Line 109ff: As outlined in the major comments, I find the fitting procedure for the random forest models not well explained. Generally, I would expect distinct data for the fitting procedure and for testing the fit - to evaluate how well the RF-model generalizes to unseen observations. I am not sure if this has been done and how the hyper parameters were chosen.

We will have a separate subsection explaining implemented RF models in detail in the Methods section.

Methods, Line 136: I think it would help the readers if another subsection was introduced, explaining which data were used to generate the random forests, including the thoughts behind the design-choices (e.g. for using annual means and the selected predictors). Given the very limited amount of data I would try to keep the number of predictors as low as possible. Currently, many design information of the RF-models appear in the Result-Section and I found it sometimes hard to keep overview.

We will have a separate subsection explaining implemented RF models in detail in the Methods section.

Line 172ff: As said, I would find it easier to follow if the RF-design information would be moved to a respective sub-section in the Methods.

We will have a separate subsection explaining implemented RF models in detail in the Methods section.

Line 199: Could you provide some measure on the quality of this reconstruction? – for the readers to rate it against the fit when using all predictors. Also, I am not sure if this intermediate step is needed (couldn't deep salinity (or Mohrholz, 2018?) be included directly for FWC based on Fig.3?).

First we conduct RF model analysis without detrending the data (Lines 198-216). These results were not provided in the original manuscript. We will add appropriate figures to the Appendix of the revised manuscript.

Line 199ff: How do these results and design decisions fit to the foregoing findings that the overall interannual FWC variations are mainly due to changes of FWC in the halocline? (as impacting factors I would thus expect mainly P-E, runoff and inflows form the North Sea?). Ultimately, Fig 3d is a nice result and I think it might help the readability to have it a bit more included in the subsequent argumentation.

We will elaborate it more in the revised manuscript.

Line 212 Which criteria were used to identify MLD and UML? Also, I am not sure how to conclude line 216 from this. Maybe this is because I am not sure what is meant by "of the sea". Can you help me out?

We will include an explanation of how mixed layer depth is calculated (Bal REAN, Panteleit et al., 2023)

Indeed, the conclusion on line 216 is speculative. We will reformulate the text.

Line 224: So, these were not assimilated?

The model system assimilates satellite observations of SST (EU Copernicus Marine Service Product, 2023) and in situ temperature and salinity profile observations from the ICES database (ICES Bottle and low-resolution CTD dataset, 2022). Even so, the salinity values at the entrance region have notable errors (Lindenthal et al., 2024). Due to data assimilation, the salinity downstream from the Danish straits is acceptable (Lindenthal et al., 2024), but salt fluxes through the cross-section at the entrance to the Baltic Sea are not accurate enough.

Line 232: I would call this inflow instead of influx – but both formulations might be right.

We have changed the term influx to inflow.

Line 235. You lost me a bit here – maybe it helps to mark the strong inflow events in the time series plot and shorten the explanation so it's easier to follow?

We will mark the Major Baltic Inflows in the figure.

Discussion & Conclusion, Line 250: I miss some more arguments why OHC and FWC are considered so important that they define the Baltic Sea state (as opposed to common measures such as temperature and salinity).

We will expand Discussion and Conclusions to address the issue.

Line 251: I find "the forcing functions" a bit vague here.

We will be more specific in the revised manuscript

Line 270: replace "correlation" by "relation"? (could be non-linear)

We have changed "correlation" to "relation".

Figures:

Fig.2: I appreciate that the authors aim for as many information as possible in this plot but I am struggling to understand the colored dots in relation to the solid lines.

We will explain Fig. 2 in more detail.

Fig.4: Could you also describe the symbols in Fig.4 c and d in the caption? Line 244 FWC

We will correct Fig. 4 caption.

<u>References:</u>

Baltic REAN, Baltic Sea Physics Reanalysis, DOI: https://doi.org/10.48670/moi-00013;

Panteleit, T., Verjovkina, S., Jandt-Scheelke, S., Spruch, L., and Huess, V.: EU Copernicus Marine Service Quality Information Document for the Baltic Sea Physics Reanalysis, BALTICSEA_MULTIYEAR_PHY_003_011, Issue 4.0, Mercator Ocean International, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-BAL-QUID-003-011.pdf, 2023

EU Copernicus Marine Service Product: Baltic Sea – L3S Sea Surface Temperature Reprocessed, Mercator Ocean International [data set], https://doi.org/10.48670/moi-00312, 2023.

ICES Bottle and low-resolution CTD dataset: Extractions 22 DEC 2013 (for years 1990–2012), 25 FEB 2015 (for year 2013), 13 OCT 2016 (for year 2015), 15 JAN 2019 (for years 2016–2017), 22 SEP 2020 (for year 2018), 10 MAR 2021 (for years 2019–2020), 28 FEB 2022 (for year 2021), ICES, Copenhagen [data set], https://data.ices.dk (last access: 30 April 2024), 2022.

Lindenthal, A., Hinrichs, C., Jandt-Scheelke, S., Kruschke, T., Lagemaa, P., van der Lee, E. M., Maljutenko, I., Morrison, H. E., Panteleit, T. R., and Raudsepp, U.: Baltic Sea surface temperature analysis 2022: a study of marine heatwaves and overall high seasonal temperatures, in: 8th edition of the Copernicus Ocean State Report (OSR8), edited by: von Schuckmann, K., Moreira, L., Grégoire, M., Marcos, M., Staneva, J., Brasseur, P., Garric, G., Lionello, P., Karstensen, J., and Neukermans, G., Copernicus Publications, State Planet, 4-osr8, 16, https://doi.org/10.5194/sp-4-osr8-16-2024, 2024.