Answer to reviewer 2: Review of the manuscript Validation of the coupled physical-biogeochemical ocean model NEMO-SCOBI for the North Sea-Baltic Sea system by Ruvalcaba Baroni et al., (2023)

Note: Authors answers are given in bold

Reviewer #2

I have completed the review of the manuscript Validation of the coupled physical-biogeochemical ocean model NEMO-SCOBI for the North Sea-Baltic Sea system by Baroni et al., (2023). In this work the Authors describe a run made using the newly ocean-biogeochemical model NEMO-SCOBI that has been specifically developed and tuned for the North Sea-Baltic Sea System. I found the manuscript well written and the validation of the model properly carried out by the Authors. On the other hand, I think that the manuscript needs some major revisions whose reasons are listed below

We thank reviewer #2 for his time and useful comments, which we address in detail below.

Comments

1. While I was reading the manuscript, although it is clear that having a coupled model for the North Sea-Baltic Sea represents a clear advancement for the scientific community, it was not clear to me if its performances are better or worst that other 3D modeling tools available for the region that are cited by Authors at line 80 for example. Are the performances of NEMO-SCOBI better or worse than other modeling tools? Does NEMO-SCOBI improve the biases of other modeling tools or not? Im asking that since a potential user should have all this information to decide to use your modeling tool/simulated data than others provided by another modeling tool.

As mentioned in lines 450-455 and specified in line 585, our results show that NEMO-SCOBI is neither better nor worse than previous model results depending on the variable and area. This means that in some areas our model performs better than other models for a specific variable and in others not (details are in sections 3.2 and 3.3). Importantly, the aim of this paper is to evaluate NEMO-SCOBI, not to make a model intercomparison, which requires a total different methodology and extensive additional work (see for example Gröger et al., 2022). Because no model is perfect, the best way is to combine results from several models and use ensembles, when possible. Currently, there are only three biogeochemical-ocean models (including NEMO-SCOBI) that account for both the North Sea and the Baltic Sea, and therefore our work represents an important step to create representative ensembles from independent models that cover both the North Sea and the Baltic Sea. We have also discussed in detail (in section 3.4) the strong and weak points of NEMO-SCOBI model, as well as its possible application as a stand alone model. These points are also summarized in section 3.5. As concluded in line 663, NEMO-SCOBI can fully compete with previous models that account for the North Sea and the Baltic Sea. However, we will read through the text to see where we can make this more clear and extend section 3.5. One of the major benefits in a model covering both North Sea and Baltic Sea is that the open boundary is moved out from the transition zone between the seas (as mentioned in lines 580-581). This is a very important step to better study the Baltic Sea and the Skagerrak-Kattegat area as it allows to study these areas without interactions with the open boundary. We will highlight this more in the text.

2. Strictly correlated with that: please provide more quantitative information about these performances. Good, Comparable, Acceptable and so on are not informative from my point of view. Please quantify in the manuscript the biases and the values of the trends and their statistical significance.

As written in section 2.2.3, and mentioned in lines 265-270, the words "good" and "acceptable" have a statistical value attached to them that is also used in other studies - see for example Edman and Anderson (2014); van Leeuwen et al. (2023). They are specified according to the Pearson correlation coefficient (r) as 1-r and the cost function (CF), which are statistical methods broadly used for model validation against observations. In summary, the model performance is considered to be "good" when both 1-r and CF of an assessed parameter fall within an inner quarter circle (see figs. 10 and 11). If the CF evaluation is equal to or lower than 1, the model values are very similar to the observed values. A value of 1-r that is equal to or lower than 0.33 means that the variability in the model for the evaluated parameter matches well that of the observations. When combining both 1-r and CF, if values fall withing the inner circle, it means that both the absolute values and the variability of the model are in good agreement. The model performance is considered to be "acceptable" if the combined value of 1-r and CF fall within the inner and outer quarter circle (i.e. when 1-r and CF values are in between 0.33,1 and 0.66,2). If 1-r and CF values are larger than 1 and 2, respectively, the model data is far from observations, meaning that there are significant differences in both absolute values and variability between model and observations. We will add clarifications where relevant in the results section. Note that the value of the absolute bias is also given where relevant in the text (e.g., line 308 and line 357).

Regarding the trends, the regression analysis and the corresponding p-values (which give the statistical significance) are calculated and mentioned where relevant in the results (e.g, lines 319, 321, 326, 346 and 364). We do not see the need for more statistics in this case, as the goal of this paper is not to detect specific temporal trends, as mentioned in line 234.

1) Line 11: What do you mean with differences? please explain

We mean that the model values and the values given by measurements for chlorophyll-a and nitrate differ considerably in the mentioned areas both in time and space. We will rephrase this sentence as follows:

"However, there are important differences for chlorophyll-a and nitrate between model results and observations in coastal areas of the southeastern ..."

2) Line 26 : Modelled or Observational studies? Please explain

Previous biogeochemical studies in the Kattegat-Skagerrak; both modelling and observations. This will be added into the text.

3) Line 31-40 : What about the intermediate layers of the basin?

This paragraph will be adjusted (also in response to reviewer #1, comment 1. More detail information on the water column structure will be added.

4) Line 54: What do you mean with point sources. Please explain

A point source is any single identifiable source of pollution from which pollutants/nutrients are discharged, such as sewage. This clarification will be added in the text.

5) Line 70-76: Please provide more information about these differences since it is the starting point for showing that this model is a step forward for the scientific community

We do not fully understand what differences the reviewer is referring to. If the reviewer refers to the difference between previous studies and our study, this is mentioned in lines 74-84 - i.e., our model covers both seas and only 2 other 3D models have this similar domain. In models that cover either the Baltic Sea or the North Sea, the Skagerrak or the Kattegat represent the boundaries of the models, which implies having simplified dynamics in these areas (this is written in lines 579-585 and also mentioned in our conclusions, line 659). Having several models covering the same area is a big advantage, as each model comes with their respective pros and cons. We have a dedicated section where we discuss the advantages that NEMO-SCOBI provides. Therefore, we will not add information on this in the introduction, but we will better highlight this in section 3.5. Please also see our reply 0.1 to reviewer #3.

6) Fig.1: It would be great to have super imposed here the bathymetry of the domain (eventually using contours)

Adding the bathymetry superimposed to this map it will be too messy and confusing. However, we will consider to add the bathymetry next to this figure or in appendix.

7) Line 107: Did you assess the drifts in the tracers to assess whether 14 years are (lets say) enough for the spin up or not? In many ocean/biogeochemical models even 30 years are not enough to stabilize the numerical solutions.

Yes, the years that are actually considered for validation correspond to those with less drift for biogeochemical parameters. The exact year is difficult to pick due to lack of observations before the 1980s. Therefore, we also performed several other runs also starting from 1961 and compared the drift. The initial drift generally decreased around the year 1975 in all runs. In addition, the biogeochemistry was not initialized from scratch - i.e. with homogeneous 3D tracer fields. The spinup was started from earlier SCOBI runs that represent already the physical-biogeochemical conditions as imprint in the initialization followed by the actual spinup. We will better clarify this in the text, line 168.

For the physical initial conditions we used restart files for the year 1973 from the simulation in Hordoir et al. (2019) that were the closest to the observations for physical properties at the start of the simulation (adding another 12 years of spin up).

8) Line 113: what kind of grid are you using? Regular, structured/unstructured? Please explain

We use a regular grid. This information will be added

9) Line 119: Is Iron not important in your domain of study? (since I remember that iron is important in the global ocean).

In case of dissolved iron in the water column as a limiting substance for the phytoplankton growth, the answer is no. Iron is not important in the domain of our study. However, for the

phosphorus cycling iron may be an important factor for the ability of the sediment to adsorb dissolved phosphate. One of the greatest challenges in implementing iron in this domain is the lack of observational data of iron in both the deep water and in the sediment as well as in the supply from atmosphere and land. Until enough observations are available, iron is assumed not to be a limiting factor.

10) Section 2.1.2 I would include in the paragraph the treatment of light and PAR (that I see located in the SM). How is primary production parametrized? Did you use Q10 formulation? Please explain here.

The full equations for primary production in NEMO-SCOBI are detailed in appendix A1 (equations A1 to A14) as written in lines 152-153. The specific equations where PAR is used is shown in equations A1 to A5. These equations shows that the modelled growth of phytoplankton is depending on temperature and can be limited by nutrients or light (i.e. we do not use Q10).

11) Line 155-170: How tick is your sponge layer in the ocean domain? Why do not you use ORAS5 that is more recent than ORAS4? Did you use bulk formulas for latent and sensible heat fluxes? What formulation of albedo did you use in your ocean model (since it influences the quantity of shortwave radiation reflected by the surface)

Because NEMO was originally intended for global setups with periodic boundary conditions, it has developed differently than other ocean model. Except for the nesting option called AGRIF, there is no sponge layer option in NEMO. Here, we have used the same settings as in Hordoir et al. (2019) for open boundary conditions (key _bdy). This means that we do not consider a sponge layer per say, instead we use a default thickness of one grid cell along the open boundaries. However, to make the model stable near the boundary, the viscosity/diffusivity coefficient has been significantly increased (by a factor of $\sim$ 10) above the halocline near the boundary, as explained in detail by Hordoir et al. (2019).

The reason we use ORAS4 instead of ORAS5 is simply that we have not yet prepared ORAS5 OBCs, but we may do it in the future. We used the CORE bulk formulation as mentioned in Hordoir et al. (2019).

Regarding the albedo, we used the original formulation and constants for the ocean, ice and snow as given in LIM3, which is the ice model to which NEMO is coupled with (Vancoppenolle et al., 2008). The LIM3 has been broadly used for climate simulations and operational oceanography. Its code is open access and therefore detailed formulation can be seen here
In summary, the feedback of the albedo of ice promotes high absorption of shortwave radiation. The constants used could of course be tested and better tuned for the Baltic Sea and the North Sea, but this has not yet been done.

12) Fig3-4 I would put the runoff as first panel in both figures.

We prefer having the top panels with nutrients (TP and TN) as these are the most relevant for this work.

13) Line 289: What do you mean with similar? please explain

we will add more detail as follows:

"This is because the model response at these two stations is similar, both in magnitude and trend behaviour, to that at stations within their corresponding region ..."

14) Figure 6 and after: please move the small panel Model Obs .. outside the first panel since it covers partially the lines beneath.

We will improve this figure.

15) Line 388 What do you mean with good period

This will be rephrased as follows:

All these stations show good model skill for both the seasonal and the entire period evaluation for temperature and salinity (Fig. 10a and Fig. 10b).

16) Paragraph 3.4. I think that this paragraph should go before the comparison between models and observations at the single stations. This could provide a general overview of the model performances better that the comparison with single point.

We have taken this feed back into consideration, but came to the conclusion that we do not agree with the reviewer suggestion. The comparisons per station are given with time (as times series and interannual averages). This gives a better overview of when the model has large bias or not. In addition, figures for only 4 stations are shown, but the analysis is done for 27 stations (as written in line 210), which are regrouped by regions and therefore, also give a spatial overview. The analysis is then summarized with the model evaluation plots (Figs 10 and 11) in sections 3.2 and 3.3.

17) Line 520-541, Line 551-560: These parts should go in the introduction since they provide an interesting description of the Baltic-North Sea system.

We will add information on this in the introduction

18) Figure 13-14: Put in first row the observations.

This will be changed according to reviewer's suggestions

References

Edman, M. K. and Anderson, L. G. (2014). Effect on pCO2 by phytoplankton uptake of dissolved organic nutrients in the Central and Northern Baltic Sea, a model study. *Journal of Marine Systems*, 139:166–182.

Gröger, M., Placke, M., Meier, H. E. M., Börgel, F., Brunnabend, S.-E., Dutheil, C., Gräwe, U., Hieronymus, M., Neumann, T., Radtke, H., Schimanke, S., Su, J., and Väli, G. (2022). The baltic sea model intercomparison project (bmip) – a platform for model development, evaluation, and uncertainty assessment. *Geoscientific Model Development*, 15(22):8613–8638.

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., Liu, Y., Pemberton, P., Schimanke, S., Andersson, H., et al. (2019). Nemo-Nordic 1.0: a NEMO-based ocean model for the Baltic and North seas–research and operational applications. *Geoscientific Model Development*, 12(1):363–386.

van Leeuwen, S. M., Lenhart, H.-J., Prins, T. C., Blauw, A., Desmit, X., Fernand, L., Friedland, R., Kerimoglu, O., Lacroix, G., van der Linden, A., Lefebvre, A., van der Molen, J., Plus, M., Ruvalcaba Baroni, I., Silva, T., Stegert, C., Troost, T. A., and Vilmin, L. (2023). Deriving pre-eutrophic conditions from an ensemble model approach for the North-West European seas. *Frontiers in Marine Science*, 10.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Beatty, C. K., and Maqueda, M. M. (2008). Lim3, an advanced sea-ice model for climate simulation and operational oceanography. *Mercator Newsletter*, 28:16–21.