Biogeosciences
Discussions

*Interactive comment on* "An observation-based evaluation and ranking of historical Earth System Model simulations for regional downscaling in the northwest North Atlantic Ocean" *by* Arnaud Laurent et al.

**Anonymous Referee #2**

Received and published: 14 September 2020

In this paper, the authors compare the output of CMIP5 and CMIP6 Earth System Models (ESMs) to observations in order to determine which models are suitable to build boundary conditions for projections. A ranking analysis was performed on a large array of ESMs. However, they are only looking at surface values of 3 variables and far away from the regional model boundaries, even though they mention on lines 44-46 that it is important to look at the information imposed at the boundaries. I think the objective stated on line 67 "Our objective is to assess the performance of a number of available ESMs in reproducing present conditions on the NWA shelf in contrast to a

high-resolution regional model" is more in line with what is presented in the manuscript since there is no analysis at the boundaries. They are not discussing the processes that lead to the observed values in the region under study and they are not analysing if the models do represent these processes correctly. I believe salinity should be included, as surface temperature depends strongly on atmospheric forcing while salinity is more representative of the different water masses in that region. Moreover, it is very surprising that a similar study (Lavoie et al. 2019) in the exact same area, with the same purpose, and using some of the same ESMs is hardly mentioned at all. No comparison of the results of this study with the 2019 study is made. Also there is not enough details on the comparison with the data, they appear to be comparing different time periods (see detailed comments) or on how the ESMs were brought to a single grid. There is only a vague mention of what the improvements are between the CMIP5 and CMIP6 models. What was improved should be stated (not only biogeochemistry of physics) so that the reader can judge on the potential impact on the ranking. Increasing the model resolution in order to improve the representation of the circulation in the NWA has been mentioned by many authors (e.g. Loder, Brickman, Yool). Here it is stated that the resolution does not have an impact. This is a big statement, considering the general agreement, and it should be demonstrated. The authors could show the changes in circulation of a few models they are giving in example for this. All these points should be addressed in order for the conclusions to be more convincing (ranking based on analysis of shelf surface conditions representative of boundary conditions). Also, Lavoie et al. (2019) estimated that the boundary conditions obtained with the ESMs were not as reliable for the simulation of the conditions on the Scotian Shelf and in the Gulf of Maine. It would good to know if there was an improvement in this regard with the CMIP6 ESMs. With regards to the level of information lacking in the manuscript I recommend a major revision before it can be published.

Abstract

Line 11: Here you say that the coarse resolution is not appropriate to represent the

circulation and elemental flux but later on you say that increasing the resolution does not matter. Is it important or not?

Line 14: ability to reproduce surface observations . . .

Line 15: why is it particularly sensitive?

Line 16: The spatial mismatch in large-scale circulation was not demonstrated. There are references for CMIP5 but what about CMIP6. Changes, or not, in circulation should be shown/mentioned after an inspection of the ESMs results.

Line22: How can we say just by looking at the surface temperature, nitrate and chl a that the top three models are appropriate for boundary forcing? The model boundaries are hundreds of meters deep (and more) and are not located in the regions analysed. It should be mentioned what are the tracers that will be downscaled at the boundaries? Salinity is certainly one of them, why was is not included in the analysis?

Main text

Line 71: why look only at three variables? What about salinity?

Line 78: historical simulations are not used for projections. This should be rephrased.

Lines 115-116: The ESMs horizontal resolution in the region of interest should be given in Table 1. Some models have a variable resolution and it might not be that bad in the NWA .

Line 117: MR and HR mean medium resolution and high resolution respectively. If they share the same grid where does the change in resolution come from?

Line 123: From where were the satellite data obtained. Who did the averaging?

Line 128: Which data from the AZMP were used? Along the Halifax line only? Why were the data averaged seasonally and not monthly like the other data?

Line 132: So the model results are brought back onto 3 different grids, one for each

variable. Are the time period also adjusted? For example SST goes from 2006 to 2016. The CMIP5 historical period ends in 2005. How can the two be compared then? Also, there are probably models that have a higher resolution than 1° (see my comment for lines 115-116), what is the impact of decreasing the resolution (converting from higher to lower resolution) and having on the ranking analysis. And how is the conversion of one grid to the other done? Also, how the thickness of the first grid cell compares between the different models?

Line 175: What is the main difference between the CMIP5 and CMIP6 groups, why is it better? Improved BGC? Same question hold for nitrate.

Line 200: Figure 6 does not show the annual cycle.

Line 208: Could you explain why? From local atmospheric forcing or circulation change?

Line 209: What are the improvements in the CMIP6 models?

Line 215: So this means that the model ranked 2 (and others)might not be ranked as high? What is the impact on the final choice?

Line 218: Could the fact that you are using different time periods and different grid resolution for the three variables explain the lack of correlation?

Line 221: How do you explain that?

Line 230: Why? Does it relate to temperature-dependant phytoplankton growth?

Line 245: What are the years compared for the ACM and the glider data?

Line 260: Correlation coefficients are high for nitrate despite having a large bias and RMSD. This should be explained.

Line 270: See my previous comments about time-period and grid differences. I think that a statement about a misrepresentation of ocean physics as the cause should be

backed up since later on the cause for nitrate mismatch is stated as coming from the BGC behavior (line 279). There are refs for the CMIP5 models but was there an improvement in circulation with the CMIP6 group or not?

Line 288: So here again, the model-data comparison was made on a different grid than for the ESMs. Shouldn't it be done on the same grid for an appropriate comparison?

Line 295: Lavoie et al. (2019) also point at the misrepresentation of the remineralisation depth in those models as a likely cause. This also explain why some models having a coarse resolution still have good results with biogeochemistry. But the statement made below that improving the model resolution does not improve the representation of circulation and main features in the models, such as the representation of the Gulf Stream detachment point and flow around the Grand Banks should be demonstrated. There is a large consensus on that and it should not be stated lightly. The authors could actually show the mean currents between the two versions of a same model with improved resolution. Especially that you state that higher resolution is required to refine the projections on line 306. There is a contradiction here.

Line 310: Here again it appears to be contradictory as you previously mentioned that BGC improvements we the cause for improvements in the CMIP6 ranking. There are likely different versions of the 4 BGC models mentioned, which should be specified in the table and considered in the analysis. Also, it could relate to the processes that control nitrate in the regions under study, they are different for your 3 regions. And how well are these processes represented by the ESMs?

Line 335: What was updated in the ocean biogeochemistry?

Figure 4f: In the suppl. figures, there is more chl a in the model than in the obs. The opposite is shown here.

Figure 7 : Maybe use ACM instead of ROMS.

Figure 8: Could specify that ACM has the same rank for the three variables (only see

one point)