Giovanni Galli et al. have evaluated the trends and controls of O_2 changes due to biogeochemical and physical changes in the NWES using model data for the 21st century. Unfortunately, the methods are unclear to me (ensemble description, analysis of O_2 change) as well as the research questions/novelty of the study. I agree it is important to assess biogeochemical changes and drivers (and their uncertainty) in such a heavily exploited and strongly changing region. As I could not follow the methods everywhere, I can only give an incomplete review of the manuscript at this stage. Here are some comments that may help to improve the manuscript:

Comments

- The title does not really seem to cover the results (How about ' 21st century trends and controls of near-bed oxygen change on the Northwest European Continental Shelf' or so?
- Abstract: Could you quantify some of your statements? What is new here?
- l85-92: You make an elaborate comparison here, but then also underlines the limited usability of Kwiatkowski et al. (2020). I would just highlight the limitation of ESMs to quantify this region if you like, but not compare.
- L116-122: which reference(s) is this all based on?
- L128: for regional models boundary conditions are also highly relevant. Spinup-times could also be mentioned here. Do you wish to provide an exhaustive list here?
- L131: Why not CMIP6?
- L 134: you did not investigate ecosystem reponses?
- L130-132: I get the feeling here you used three models and then ran 3 ensembles within each model (namely by using slightly different CMIP5 forcings), can you clarify this already here?
- Introduction: your final paragraph (lines 126-136) describes what your new contribution is. However, it is unclear at this stage what extra model variability you argue to have covered (and important to note that there are several sources of uncertainty, scenario, model variability, model uncertainty, see for example Fig. 3 in https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015GB005338). Also, your research questions are not so clear to me. Why did you focus on the near-bed O_2 specifically, why not the whole water column and then near-bed as a separate focus?
- Line 140: is a 10-year spin-up enough? In ESMs a few hundred years is more common. What drift do you have in your variables during this spinup? If significant, drift should be subtracted from the data at the least (and a thorough discussion should be provided why you can still use the data).
- Sect. 2.1: I do not follow. There are 3 models which all are part of the NEMO-ERSEM model suite (so 3 times almost the same model?). Are these the 3 members then? Which you then forced with ESM data (from GFDL, IPSL and HADGEM)? What are the parent ESMs then (as its says that the boundary conditions of these 3 ESMS are taken from parent ESMs (line 152), these CMIP5 data are fully coupled ESMs without boundary conditions except towards space)? So, do you have 9 model runs in total (3*3)? When you write about 'all models' in line 155 you seem to be discussing the ESMs as if you have been running the ESMs, but you used this NEMO-ERSEM setup, right? Anyway, I do not follow. Alternating between the word member and model might be inconsistently done? Maybe a table? What happens in the forcing in the 21st century (e.g., wind/freshwater forcing changes?)

- L 150 and what are the ECS then of these models?
- Sect. 2.3: SS_t0 is not defined here? How is this approach different from AOU (Apparent Oxygen Utilization) or even better TOU (True Oxygen Utilization)? You open with that O_2 change have 3 different components, but then you can only separate into 2, right? Namely the temperature effect through its effect on O_2 saturation and then biology+circulation as the 'other' term (which is like in AOU and I am not aware of a method that can distuingish all 3). Calculating O_2^{sat} and the contribution of O_2 from circulation+bio is a simple calculation I would say, and I think the analysis should go beyond this and the correlations.
- Sect. 3.1: Why don't you bias-correct and only use the model/ensemble trends (like you actually do in e.g. Fig. 3), considering the significant biases? Then the absolute errors are less important and can go into an appendix or so. You seem to have done so anyway for (part of?) your analysis (mentioned in line 276 and caption Fig. 4 only…).
- Fig. 2: based on the text units here are standard deviations? Maybe just use the full names instead of nurmsd and nbias? Or just call them Root mean square and bias and say that they are normalized? I think it would be good to get the equations from Jolliff et al. (2009) or to use more commonly used metrics like RSS?
- L 259: here you for the first time use the word downscaling, this should be introduced in the methods section.
- Fig. 3: If you would plot instead of a change over time a change at a certain global warming level (countering the differences in ECS, see Hausfather et al. (2022); 10.1038/d41586-022-01192-2), your model differences will likely be smaller? Would that be a more meaningful way of assessing model differences as showing differences in warming is inherent to choosing models with different ECS?
- Can Sect. 3.4-6 be merged?
- Sect. 3: I was actually a bit surprised about the section titles here, and it would be good to introduce the reader earlier what you will exactly cover in your results section to answer your research questions.
- Sect. 4: this mostly sounds like a conclusions/summarizing section except for the last paragraph. Please try to discuss limitations of your methods, implications, compare to other studies that may show something else? You find many confirmations/consistencies which is fine but makes your work sound less novel or complementary. What other stressors does the near-bed ecosystem experience (trawling/pollution?)?
- L 430: how does your study highlight this? Could you show your regional/downscaled model runs are superior to the ESM output? Same in line 441.
- You mention that you asses 'ecosystem impacts' throughout the manuscript, but I would say you mostly assessed a range of physical and biogeochemical changes and the possible drivers of the O_2 changes.
- L450-451: reference?
- Sect. 5: I do not see so well how this section connects to your results. Please quantify your results and focus your conclusions on the answers to your research question(s). E.g., your conclusions and abstract text are quite different while one would expect them to cover very similar statements. Sect. 3.9 is not discussed or concluded upon.
- correlation is not causation

Minor remarks:

- Some spelling errors that can be captured by any spellchecker are still in the text
- L76: possibly? Sometimes? Regularly?
- L185: limit validation?