REVIEWER 2

GENERAL COMMENTS

REV2:

Giovanni Galli et al. have evaluated the trends and controls of O2 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 O2 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:

ANSWER:

We thank Reviewer 2 for the useful comments, we thoroughly revised the manuscript according to all reviewers' comments. Among other change we improved the ensemble description removing possible sources of ambiguity as well as the methods description and stated the study objectives more clearly.

SPECIFIC COMMENTS

REV2:

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?

ANSWER:

We acknowledge "intra-scenario variability" was misplaced here. We would refrain to use '21st century trends' because our ensemble is not large enough to produce a robust assessment of expected trends, or to quantify uncertainties. However, under Reviewer1's suggestion, we changed 'intra-scenario variability' to 'multi-model comparison' to reflect this concern.

REV2:

Abstract: Could you quantify some of your statements? What is new here?

ANSWER:

We added some quantification of results in the abstract, and some more in the results section.

We also improved the last paragraph of the abstract by stating the novelty and relevance of this study. In short, we expand on Wakelin et al (2020) showing that the projections of near-bed oxygen presented there are robust in a multi model context and that trends and drivers of change remain coherent at different warming intensities. Finally we want to assess the impact of the change in circulation presented by Holt et al. (2018) on oxygen change similarly across models. These are new findings that allow a better understanding of the projection of near-bed oxygen in the NWES and its drivers. We believe (and reviewer 1 seem to concur) that these are important questions to assess the impact of climate change on shelf seas and plan the needed adaptation, and it may suggest directions for future research.

REV2:

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.

ANSWER:

Our aim here was not really comparing the two results, which we believe may be equally valid, but showing that when looking at oxygen change results may vary also according to the spatial scales and domains being analysed (coastal vs shelf). Then global ESMs have indeed some known limitations in resolving coastal and shelf processes and we rephrased the part of text when we explain this to make it clearer and more specific.

REV2:

L116-122: which reference(s) is this all based on?

ANSWER:

It's all Wakelin et al. 2020, we are summarising the results of that paper and that took some text, we appreciate this may not have been entirely clear, so we repeated the citation in the text.

REV2:

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?

ANSWER:

We rephrased the sentence so that we refer to the components of variability (Frölicher et al. 2016): internal, model and scenario variability, and provided some examples of variability sources for internal and model variability. Since this particular sentence is general, it applies to both regional and global models, to atmosphere, ocean or land, it may not be advisable to mention all possible sources of variability.

REV2:

L131: Why not CMIP6?

ANSWER:

These simulations were implemented before outputs from CMIP6 were available. While we acknowledge the differences between the CMIP5 and the CMIP6 scenarios, CMIP5 still provide a useful set of climate scenarios that are still valuable and that can be associated to the CMIP6 ones.

REV2:

L 134: you did not investigate ecosystem responses?

ANSWER:

We rephrased the sentence, to make clear that we look at near-bed oxygen change.

REV2:

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?

ANSWER:

We have 3 downscaled ensemble members, each one is forced with one single CMIP5 ESM. We appreciate this may not have been clear in the text so we revised it accordingly.

REV2:

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 hdps://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015Gti005338). Also, your research questions are not so clear to me. Why did you focus on the near-bed O2 specifically, why not the whole water column and then near-bed as a separate focus?

ANSWER:

We acknowledge the claim of addressing [comprehensively] some aspect of variability may be misplaced here. We improved the final paragraphs of the introduction where we describe the aim of the study so that we don't refer to this. We also changed in the title and throughout the text the term "intra-scenario variability" to "multi-model comparison", which is perhaps more appropriate here. We also stated in the discussion which sources of variability our study fails to address.

We also improved the exposition of aim of the study that is further qualifying near-bed oxygen projections, their trends, and drivers in the NWES in a multi-model context, and assessing the impact of circulation changes on near-bed oxygen similarly in a multi-model context. All of which represent understudied questions.

We choose to focus on near-bed oxygen because the bottom layer is more vulnerable to deoxygenation than the surface, due to the seasonal decoupling from the atmosphere. Furthermore, near-bed oxygen dictates habitat suitability for benthic sessile or scarcely motile species and it is used as an indicator of eutrophication in the North Sea (Devlin et al. 2023).

We also added to the introduction some more explanation on why the focus on near-bed oxygen.

REV2:

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, drif should be subtracted from the data at the least (and a thorough discussion should be provided why you can still use the data).

ANSWER:

While a 10y spinup is not enough for Earth System Models, it is for the regional model used here and it is in fact common practice (e.g. Tinker et al. 2014, Holt et al. 2018, Ciavatta et al. 2018). The North Sea has a flushing time of 2-4 years, the Norwegian Trench about 100d (Blaas et al. 2001), depths are relatively shallow and seasonal mixing is intense throughout the water column in most locations; all of this concurs in making these relatively short spinup times acceptable for this domain. Indeed we did not observe appreciable drifts in the model during the spinup period. In addition to that the biogeochemical initial conditions, which include the slowest components of the system (i.e. the benthos) are initialised with the reanalysis from Ciavatta et al. 2018, which ran for an additional 10y, including its own spin-up. We added a couple of lines explaining this.

REV2:

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?)

ANSWER:

We have 3 downscaled ensemble members that all use the NEMO-ERSEM suite, each one is forced with boundary conditions from one of three fully coupled CMIP5 ESMs, the "parent ESMs", which are GFDL-ESM2G, IPSL-CM5A-MR and HADGEM2-ES. So we have 3 downscaled ensemble members in total, which, for brevity, we call GFDL, IPSL and HADGEM (instead of "the downscaled ensemble member forced with boundary conditions from GFDL-ESM2G, IPSL-CM5A-MR or HADGEM2-ES"). We appreciate that our presentation might have been confusing, we revised it in order to make it clearer, and we took care of being consistent in the use of ensemble member instead of model.

As for the forcings in the 21st century they are as described in the text: those available from the CMIP5 (e.g. wind, lateral boundary conditions, etc.) are from CMIP5 (both historical, up to 2005, and climate runs), those not available from CMIP5 are from other sources (e.g. river nutrient loads are from a reanalysis, Ciavatta et al. (2018), multiplied by river discharge). We also revised the Ensemble description in the text in order to provide a clearer and more comprehensive list of all the forcing fields.

REV2:

L 150 and what are the ECS then of these models?

ANSWER:

We added to the text the estimates of the ECSs of the ESMs, thes are 4.59, 4.12 and 2.39K for HADGEM2-ES, IPSL-CM5A-MR and GFDL-ESM2G respectively (Andrews et al., 2012, Dufresne et al., 2013).

REV2:

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 O2 change have 3 different components, but then you can only separate into 2, right? Namely the temperature effect through its effect on O2 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 O2 sat and the contribution of O2 from circulation+bio is a simple calculation I would say, and I think the analysis should go beyond this and the correlations.

ANSWER:

Please note we reworked the methods section extensively, following Reviewer3's suggestions. The comment still applies though.

Our approach is indeed related to the classic AOU / O2sat decomposition, with the difference that our approach quantifies components of change at a point in space relative to a reference time period, whereas AOU measures the time-integrated amount of oxygen consumed since a water parcel has left the surface.

Whereas AOU implicitly assumes that O2sat of a water parcel doesn't change since its last contact with the atmosphere, our method explicitly accommodates changing O2sat (due to e.g. ocean warming). This way ΔO2 can be explicitly partitioned in two components, one related to changing O2sat and one related to changing SSO2.

ΔO2phy-ch and ΔO2other are directly comparable measures, being both components of the total ΔO2, as opposed to AOU (Δ concentration) and O2sat (concentration).

As for other metrics such as True oxygen utilisation (TOU Ito et al. 2004) and Evaluated Oxygen Utilisation (EOU, Duteil et al. 2013), the first must be evaluated explicitly at runtime to define the preformed oxygen, which unfortunately we haven’t implemented, and both address some known issues with AOU when O2 concentration and solubility are decoupled, which may happen e.g. when undersaturated water is subducted or when a water mass changes temperature away from the surface, or in the presence of sea ice. This is not quite the case in the NWES that is relatively shallow and well mixed, with intense ocean-atmosphere exchange. Which makes the assumptions behind AOU (and our method as well) a fair approximation. Overall, we felt the use of different metrics wouldn't have brought much improvement to the results. We added some lines in the discussion addressing the limitation of our method.

While we agree our analysis method is fairly simple, we still believe that this is still informative, and indeed we successfully exploited it to diagnose oxygen dynamics and controls in our ensemble members.

REV2:

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…).

ANSWER:

We don't bias correct extensively (with the exception of the hypoxia estimation in fig. 4) because we almost exclusively look at delta concentrations and correlations and these are not influenced by bias. Instead we use bias correction when calculating hypoxia because when fixed low oxygen thresholds are considered, absolute values are relevant. We added some lines to the manuscript to make this clearer.

REV2:

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?

ANSWER:

We don't indeed use normalised rmsd but normalised unbiased rmsd, which is normalised rmsd multiplied by the sign of the difference of model and data's stds, so that it can also assume negative values, which indicate that the model's std is smaller than that of the observations, and vice-versa for positive values. We explained this in the text, we didn't add in the full equation because that is indeed quite trivial.

REV2:

L 259: here you for the first time use the word downscaling, this should be introduced in the methods section.

ANSWER:

We mentioned "downscaling" several times throughout the text.

REV2:

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?

ANSWER:

We appreciate this could be an interesting angle to look at our projections, and we did run some additional analyses to look into it (see attached document, WNT_and_Warming_Level.pdf).

However, warming level, either global or regional (over the downscaled domain), doesn't seem to be relevant for the development of the change in circulation in the North Sea, whose effects on near-bed oxygen are an important focus of our manuscript. In particular, both IPSL and GFDL, which are exactly the same model with different forcings, show regional atmospheric warming in excess of 2K, but whilst IPSL develops the circulation change, GFDL doesn't.

In the manuscript we improved the statement of the aim of the study, shifting the focus away from the evaluation of model variability and differences. In short, we aim at testing the system's response (with a special focus on effects of circulation changes) at different climate change intensities, as this is at present poorly constrained for many biogeochemical variables (oxygen included) in the NWES. That is why we chose models covering a wide range of ECSs.

REV2:

Can Sect. 3.4-6 be merged?

ANSWER:

we merged 3.5-6 and also 3.7-8, the new titles are as follows:

3.5 Physical controls of oxygen change: temperature and stratification

3.6 Biogeochemical controls of oxygen change: primary production and respiration

REV2:

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.

ANSWER:

We improved the methods section in order to introduce more detail about the analyses we performed and whose results are presented in section 3.

REV2:

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?)?

ANSWER:

We thoroughly re-wrote the discussion focussing on the implication of our results and on the limitation of our methods, and on the evidence available from the literature. We also discussed the combined impacts with other stressors in the conclusions section.

REV2:

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.

ANSWER:

As we don't indeed provide a direct comparison with the oxygen field in the parent ESMs, we see how these statements were problematic. Providing such comparison would be, we believe, certainly interesting but also out of the scope of this paper. We rephrased the two sentences referring to literature rather than this study.

REV2:

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 O2 changes.

ANSWER:

We rephrased all instances of 'ecosystem impacts' or similar in the text to make it clear that our assessment is limited to near-bed oxygen change.

REV2:

L450-451: reference?

ANSWER:

we rephrased the sentence and added a reference (Devlin et al. 2023, https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/indicator-assessments/seafloor-dissolved-oxygen).

REV2:

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.

ANSWER:

We re-wrote the conclusions sections focussing on the answers to our research questions, we explicitly linked to what is stated in the abstract, and we improved the final recommendations. We also improved our discussion on how circulation change in the North Sea affects the development of deoxygenation hotspots.

REV2:

correlation is not causation

ANSWER:

We improved our results section by discussing in more detail the mechanisms that can observed correlations, including some that we failed to discuss earlier. This includes the discussion of covariances that determine correlations between variables also in the absence of a direct causal link. Here some examples:

[1] corr(SSO2, Tatm)>0 in southern coastal regions, all members, covariance explained by increasing NPP,

[2] corr(SSO2, PEA)>0 in coastal regions, covariance mediated by the seasonality in NPP.

[3] corr(SSO2, Tatm)<0 in the Trench and Eastern North Sea, all members, (new results without detrending under reviewer1’s suggestion, covariation with increasing PEA),

[4] corr(BResp, SSO2)>0 in the Norwegian Trench, IPSL, covariance explained by decreasing BResp, due to decreasing NPP, together with decreasing SSO2 due to increased stratification (no strong direct causal link).

MINOR REMARKS

REV2:

Some spelling errors that can be captured by any spellchecker are still in the text

ANSWER:

We thoroughly revised the manuscript and corrected typos.

REV2:

L76: possibly? Sometimes? Regularly?

ANSWER:

frequently

REV2:

L185: limit validation?

ANSWER:

limit validation!

Western Norwegian Trench current flux and warming levels.

We compared (Fig. 1) the time evolution of the Western Norwegian Trench current flux (WNT) and warming levels in an ensemble of three downscaled ocean climate projections. The three members of the ensemble are forced with BDYs from one of three rcp8.5 CMIP5 ESMs: HadGEM2-ES, IPSL-CM5A-MR and GFDL-ESM2G. BDYs extracted from the ESMs and used for downscaling include atmospheric surface temperature. The three downscaled ensemble members are termed (HADGEM, IPSL and GFDL for brevity).

WNT is calculated as per Holt et al. (2018) and smoothed with a gaussian filter. Global Warming Levels (GWL) are calculated as the year at which the average atmospheric surface temperature in the three ESMs reaches the thresholds of +1, +2.5, +2.0, +2.5, …, +5°C relative to pre-industrial average. ESM global mean temperature time-series are smoothed with a 21y running mean prior to determination of the warming levels.
The Regional Warming Level (RWL) is calculated as the average atmospheric surface temperature change over the downscaled domain, smoothed with a 21y running mean, relative to the period 1990-2011.
Note that, since the reference periods change, RWL starts from 0, while GWL starts from 1.

A change-point in WNT is detected in both HADGEM and IPSL, but not in GFDL, starting approximately in the late 2020s, with a progressive weakening and reversal of the current flux. This happens at approximately +1.5-2.0°C GWL and +0.5-1.0°C RWL in both IPSL and HADGEM. However in GFDL no change in WNT is detected despite GWL up to +3.0°C and RWL up to +2.0°C.

We conclude that warming level, either regional or global is not a determining factor explaining the onset of the circulation change.

The circulation change is triggered by an increase in oceanic stratification at the northern entrance of the Norwegian Trench (Holt et al. 2018). It may hence be triggered by other factors not necessarily linearly related to warming level, e.g. ice melting. The representation of such factors in the ESMs has likely a crucial role here.

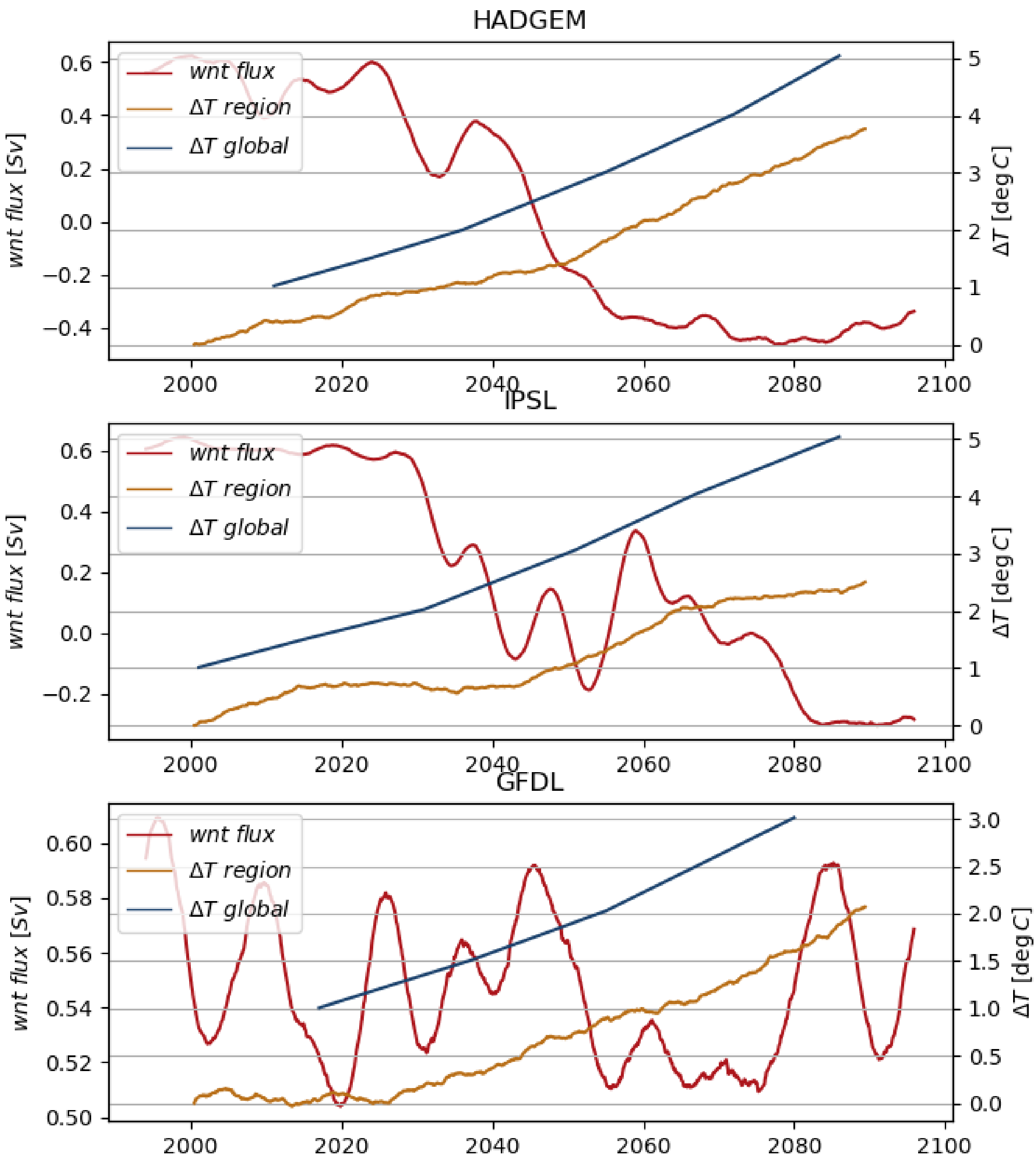

Figure 1 WNT current flux (red), Global Warming Level (ΔT global, blue) and Regional Warming Level (ΔT region, yellow).