REVIEWER 1

GENERAL COMMENTS:

REV1:

The manuscript investigates the processes driving near-bed oxygen changes on the Northwest European Continental Shelf under a high-emissions climate change scenario, with a focus on the intermodel uncertainties in these processes and their effects on oxygen. This work extends and qualifies the results of a previous study (Wakelin et al., 2020) by adding two additional sets of regionally downscaled model projections within the high-emissions forcing scenario (RCP 8.5).

Ocean deoxygenation and coastal hypoxia under climate change pose a serious threat to marine ecosystems. Robust understanding and projection of these processes is important for effective adaptation of ecosystem services. Given the lack of skill of coarse resolution global ESMs in coastal regions, regional downscaling of ESM projections will likely play a critical role in exploring this topic.

Although these additional model simulations provide valuable new insights into the fate of the oxygen in the region, some of the main conclusions reached by the authors are not well supported by the evidence presented. The scope of the study is not well defined and the manuscript overall lacks focus and rigor. While the scientific premise of the study is valuable, major revisions are required for this work to be fit for publication in Biogeosciences.

ANSWER:

We thank Reviewer 1 for the useful comments. We thoroughly revised the manuscript according to all reviewers' comments, we took special care in providing additional support to our conclusions, we better defined the scope of the study and its limitations.

SPECIFIC COMMENTS:

REV1:

(1)

A major result of the paper is the attribution of the deoxygenation hotspot in the Norwegian Trench to a relaxation reversal of the Norwegian Trench Current; but this interpretation is not well supported or well argued. The authors argue that (1) a relaxation of the advective current causes a freshening of the shelf region causing increased stratification, and (2) correlation suggests that the increased stratification is responsible for deoxygenation. Holt et al (2018) argue that changes in stratification are responsible for the relaxation of the current, opposite to the authors' explanation. In most cases, an increase in stratification would come from surface warming and precipitation changes; this null hypothesis should be disproved before seeking alternative explanations.

(2)

It is also not clear in the results whether vertical mixing or horizontal advective transport is dominating oxygen supply to the Norwegian Trench region, which should guide the conclusions made. Note that Wakelin et al (2020) do link reduced current to a recoupling of export with near-bed respiration; perhaps this is connected to the change in sign of correlation between SS and stratification (320).

(3)

Lastly, ‘tight coupling’ in Figure 11 is not necessarily convincing by eye. A stronger link has to be made.

ANSWER:

(1)

We appreciate how this may have not been entirely clear in the text but our conclusions about the causes of the relaxation of the WNT current are not at odds with Holt et al. 2018. In both works it is increased stratification at the northern entrance of the trench that reduces oceanic inflow into the North Sea, this in turn increases retention of fresh water from continental Europe and the Baltic within the North Sea, driving freshening (Holt et al. fig1e,f, this study, fig 3.) and a further increase in stratification in the North Sea. Then there certainly is a component of the increase in stratification due to the atmospheric temperature forcing, but this cannot explain the hotspot of increased stratification as surface warming is homogeneous across the domain (fig. 3.).

We revised the manuscript to make all of this clearer.

(2)

We acknowledge the reviewer’s comment, however in our model configuration, the lateral transport of oxygen from the Baltic open boundary does not change in time. We acknowledge that this was not clarified in the manuscript, but in GFDL and IPSL, the Baltic open boundary has fixed climatological values for all tracers, including oxygen. This choice was made because ESMs are scarcely reliable for an enclosed sea such as the Baltic. This ensures that the deoxygenation signal we detected in the Norwegian Trench does not originate from the Baltic boundary through lateral transport.

HADGEM also uses a fixed climatology at the boundary for both biogeochemical variables (including oxygen) and freshwater input, with the difference that the Baltic boundary is treated like a river, rather than an open boundary.

Both boundary treatment choices do have some limitations, however they also rule out lateral transport from the Baltic as the source of the deoxygenation signal in the Norwegian Trench.

We clarified this in the methods, and added some discussion about the limitation from having a climatological boundary at the Baltic while being able to rule out lateral transport as a contributing factor for deoxygenation hotspots.

(3)

We complemented section 3.9 (now 3.7) with correlation coefficients for the analysed timeseries and revised the text according to the results.

REV1:

The title of the manuscript suggests that the focus of the paper is on ‘intra-scenario variability’; however, it is unclear what the scope of this is and how effectively it can actually be investigated with available tools. Uncertainty in ESM climate projections (and by extension, downscaled projections) fall broadly into three categories, regarding (1) internal model variability, (2) intermodel uncertainty, (3) and scenario uncertainty. The term ‘intra-scenario’ would suggest that you look at both internal variability and intermodel uncertainty, which is not really the case. Due to the small sample size (three models) and inconsistencies in the model and methods used for downscaling in the older HADGEM run versus the IPSL and GFDL simulations, neither internal variability nor intermodel uncertainty is well sampled nor well isolated. Perhaps the term 'multi-model comparison' used in the abstract is more appropriate here. This is already addressed somewhat in the introduction (125-135), but should be clarified and given more thought. Claims like "we added an intra-scenario variability dimension (375)" are unclear and misleading, and should be changed.

ANSWER:

We appreciate the focus on "Intra-scenario variability" may be misleading, and we concur with the reviewer that the scope here is to compare the projections of oxygen from the small "multi-model" ensemble. Therefore, we changed "Intra-scenario variability" in the title and throughout the text with "multi-model comparison", or deleted it, revised the introduction by mentioning the categories of uncertainty in projections and by more clearly stating the aim of the study, shifting the focus away from variability estimation. We revised the discussion by stating which sources of variability were not addressed in this study. We revised the conclusions highlighting the importance of sampling different sources of variability while building regional climate model ensembles.

REV1:

Throughout the study, the authors claim that oxygen changes in the study region across the three simulations scale with the climate sensitivity of the parent ESMs. If quantifications of these sensitivities are available, they should be presented here. Additionally, an issue with this claim is that the differences in downscaling methods for HADGEM vs the IPSL and GFDL simulations provide uncontrolled degrees of freedom. The authors should provide an argument whether the differences in downscaling techniques should significantly impact the magnitude of oxygen changes. If possible, the authors could run some short sensitivity experiments using the new (used for IPSL, GFDL) setup to test sensitivity to e.g. vertical resolution.

ANSWER:

We added the estimates for the global equilibrium climate sensitivity of the three parent ESMs (these are 4.59, 4.12 and 2.39K for HADGEM2-ES, IPSL-CM5A-MR and GFDL-ESM2G respectively).

We also added in section 4 a more detailed discussion on how the different downscaling methods from HADGEM may influence the results. Unfortunately producing conclusive evidence requires ad-hoc experiments, and as the reviewer suggest, this can be quite an expensive task that is not always feasible for multiple reasons, including availability of resources. While we agree that the differences in the model set-up may play a role in the dynamic, these will not be the driving cause of the patterns projected by the model.

Nonetheless, despite some noticeable different responses, the bulk of the behaviour of our ensemble members is still coherent with the tested climate change intensities. This we think shows that our results are still robust with respect to the [limited] model variability represented in our ensemble.

REV1:

In the model used by Wakelin et al (2020), oxygen is not included in open boundary conditions of the regional model so that changes in open ocean oxygen is not included. Is this the case here? This is very important for how the results may be interpreted and should be documented carefully.

ANSWER:

We appreciate this was not explicitly mentioned in the text. In our IPSL and GFDL members oxygen is indeed included in the open ocean and Baltic boundaries, whilst Wakelin et al. (2020), that is our HADGEM, uses a zero gradient-scheme (i.e. boundary concentration equals concentration inside the domain) for most biogeochemical tracers, including oxygen, at the open ocean boundary. The only tracers that are forced with external data at the open ocean boundary are nutrients and inorganic carbon. At the Baltic boundary HADGEM uses climatological values for all tracers, including oxygen.

We improved the Ensemble description in the manuscript, so that all boundary schemes are clearly described.

The impact of the different treatment of the boundary will largely impact the open ocean part of the domain (that is excluded from the analysis), while the shallow depth and intense winter mixing of the NWES makes so that ocean-atmosphere exchange will reset the oxygen to saturation every winter, or more frequently, throughout the water column, so that oxygen on shelf is scarcely coupled with oceanic oxygen.

REV1:

The authors need to be careful when interpreting correlation as causation. Correlations are only meaningful when there is a process that can explain the relationship. Please be thorough about when a physical/ biogeochemical mechanism can explain a correlation and when a correlation cannot be explained. For example, why would you have a positive correlation between SS and stratification in some regions (Fig. 7)? If strong but erroneous correlations are prevalent, why can we still trust the results? The authors should also provide a discussion of any covariances that may influence the results (e.g. between temperature, stratification, respiration, NPP)

ANSWER:

We agree with the reviewer that correlation does not automatically imply causation, and we can support our interpretation by improving the presentation of the results and the discussion of the attribution of correlations. In particular we added detail about:

[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 see later, 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).

REV1:

In calculating correlations, the long-term trend is removed. I see how this avoids false positives, but how can you assess the drivers of forced changes after removing the long term trends? In this case, it seems that correlations just classify the drivers of short-term variability, which is not what you purport to be investigating. Please explain/ clarify.

ANSWER:

We appreciate this may be of concern regarding our methodology. While analysing the data we did conduct exploratory analyses where trends were not removed, which only resulted in slight improvements of some detected correlations, with no relevant changes in sign. We concluded that trend removal was in this case the most conservative practice.

This perhaps could be justified in systems with short turnover rates where drivers of short- and long-term trend overlap. For example, warming reduces oxygen solubility both on the long-term, through increasing mean temperatures, and on short-term, e.g. during summer months. Or increasing NPP produces oxygen both on the short-term, during a bloom, and the long-term (if coupled with enough mixing) if the productivity of a region increases over time.

Nonetheless, we see a solid point can indeed be made in favour of retaining the trend when calculating correlations, if the aim is explaining the trend, and taking care that, when interpreting results, some patterns will be explained by covariances rather than causal links (false positives). This is what we did. As for the revised results the only relevant changes are:

1) Negative correlation between O2sat and atmospheric temperature in the Norwegian Trench (all members, instead of non-significant).

2) Negative correlation between SSO2 and atmospheric temperature in the Norwegian Trench and eastern part of the North Sea (HADGEM and IPSL, instead of non-significant).

3) Negative correlation between SSO2 and PEA along the Norwegian Trench (HADGEM and IPSL, instead of non-significant)

for 1) and 3) a case can be made for a causal link, for 2) the pattern is more easily explained by covariance with PEA.

None of these change our conclusions substantially but 3) removes the need for Fig 8. “running correlations between SSO2 and PEA mediated over the Norwegian Trench”.

The correlations involving biogeochemical variables (NPP, BResp) didn't show any significant change.

We attach a revised version of the part of results that changed for a more complete exposition.

REV1:

What is gained by decomposition in section 2.3 as opposed to a traditional O2sat, AOU decomposition? Why is O2phys-ch (O2sat scaled by the initial saturation state) a more meaningful metric than O2sat? The authors end up using O2sat and SS (a.k.a 1-AOU) anyway, so this section can be removed entirely.

ANSWER:

Please note that we re-worked the methods section according to Reviewer3's comments, as a result the definition of ΔO2_other changed slightly and O2,phy-ch,t is no longer present, the comment still applies though.

The ΔO2_phys-ch and ΔO2_other metrics we presented in the methods are indeed related to the traditional O2sat, AOU decomposition, with the difference that they describe the partitioning of oxygen change relative to a reference period, rather than the distance from equilibrium at any specific moment.

This renders them interesting as metrics because they are directly comparable, being both Δ concentrations (unlike Osat (concentration), AOU (Δ)) and they sum up to the total ΔO_2. This allows to quantify how much of the observed change can be attributed to each component.

AOU estimates oxygen consumption (and production) since a water parcel was last in contact with the atmosphere, assuming Osat doesn't change. Our metrics, by explicitly considering changes in Osat, allow to partition oxygen change into the two separate components.

We included a section in the methods explaining this, the relation between our metrics and AOU, and the hypotheses and limitation of both methods.

Results about $\Delta O_{2,phys\text{-}ch}$ and $\Delta O_{2,other}$ are presented in section 3.4 Contributions to near-bed oxygen change.

REV1:

How are there negative values in the root-mean-square distance calculation (Fig. 2)? Need to provide formulae here for nbias and nurmsd.

ANSWER:

we appreciate the metrics from Jollif et al. may not be as widely known as others, we addressed this by adding their definition in the methods (although we merely described the equations, rather than writing them down, as they are indeed trivial). As for the negative values of nurmsd, they arise by multiplying rmsd by the sign of the difference of model and data stds, so that a negative value indicates that the model's std is lower than that of the observations, and vice-versa for positive values. We explained also this in the text.

REV1:

Bias-correction for hypoxia measurements should be included in methods

ANSWER:

we included bias correction procedure in the methods.

TECHNICAL COMMENTS:

REV1:

Grammatical errors and inconsistent capitalization throughout. Please proofread carefully.

ANSWER:

We carefully proof-read the manuscript and corrected errors and capitalization.

REV1:

In all figures, panel labels need to be included.

ANSWER:

Panel labels have been added to all figures.

REV1:

Use consistent terminology for region names. Is the Danish strait the same as Skagerrak? Eastern North Sea is referenced throughout but not delineated on the map in Fig 1.

ANSWER:

We replaced Danish strait with Skagerrak, we also replaced 'Eastern North Sea' with 'eastern part of the North Sea', or similar, throughout the text.

REV1:

Nearly all instances of ‘in fact’ can be removed

ANSWER:

all instances of 'in fact have been removed or replaced'

3.5 Physical controls of oxygen change: temperature and stratification

Changes in $\Delta O_{2,phy\text{-}ch}$ and $O_{2,sat}$ are, for the greatest part, explained by warming (correlation between $O_{2,sat}$ and near-bed T ~-1 everywhere in all models, not shown). The driver of this is the temperature atmospheric forcing (Fig. 6) that in all models displays strong negative correlation with near-bed $O_{2,sat}$.

Conversely atmospheric temperature correlates positively with SS_{O2} in coastal regions around the British Isles and continental Europe (including the Southern North Sea Channel and Irish Sea) in all models. This appears to be mediated by a covariation with increasing NPP in these well mixed areas fuelling oxygen production (see section 3.6). Positive correlation between SS_{O2} and atmospheric temperature in the Central and Northern North Sea, which stratify seasonally, in IPSL and GFDL may instead be mediated by covariation with decreasing respiration in these areas, which is due to decreasing NPP (see section 3.6).

Atmospheric temperature and SS_{O2} instead are negatively correlated in IPSL and HADGEM in the regions of the deoxygenation hotspots, Norwegian Trench and eastern part of the North Sea. This is mediated, for both models, by covariation with increasing stratification in these regions (see below).

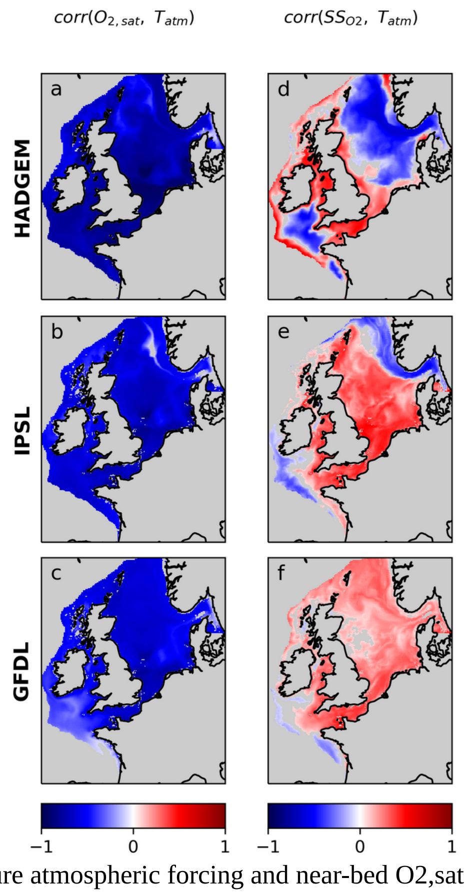

Fig. 6. correlation between Temperature atmospheric forcing and near-bed O2,sat and SS_{O2}.

The North Sea hotspots of oxygen decline in HADGEM and IPSL coincide with enhanced stratification hotspots and indeed SS_{O2} and potential energy anomaly (PEA - an indicator of stratification de Boer et al. 2008) are, in both ensemble members, strongly negatively correlated in this area (Fig. 7);

GFDL on the other hand only shows a moderate increase in stratification and no significant hotspots, with a weaker correlation between SS_{O2} and PEA than in the other two models. The main driver of stratification along the Norwegian Trench and in the eastern part of the North Sea is, for all models, surface salinity, that is strongly negatively correlated with PEA there and over much of the domain, especially in HADGEM and IPSL.

The positive correlation between PEA and SS_{O2} in coastal areas in the southern North Sea and around the British Isles (observed in all ensemble members) appears to be mediated by the seasonality of primary productivity. These shallow regions experience strong tides and remain well mixed year-round (PEA barely changes in the long term). Here stratification is not a meaningful indicator of vertical oxygen transport. However. The highest PEA values do happen in the summer months, when also NPP peaks, producing oxygen that contributes to high SS_{O2} values, while the opposite is true during winter; hence the positive correlation.

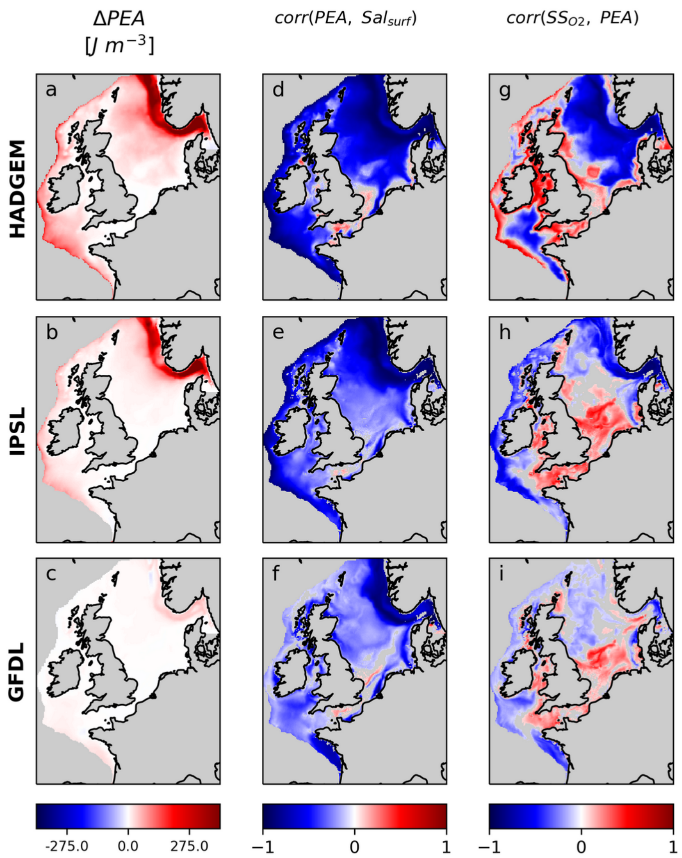

Fig. 7. Change in potential energy anomaly (PEA) and correlation between PEA and surface salinity and PEA and SS_{O2}.