General Response

We thank the three reviewers for their constructive and insightful comments. The main concerns among the reviewers fall into 3 categories. We give a brief overview of how we propose to address them and include the proposed new figures here in the general response, and then provide a detailed response below.

1. **Role of changes in stratification and productivity in dissolved oxygen projections.** The reviewers ask for an expanded investigation of potential changes in upper ocean stratification and the biological sink on oxygen changes. Specifically, increased stratification could contribute to deoxygenation in the northern Arabian Sea, and decreased productivity and respiration rates could contribute to oxygenation in the South Equatorial Current (SEC). We propose to add a subsection to our *Results* in which we examine trends in upper ocean stratification (Figure GAC1) and the export of organic carbon at 100 m (Figure GAC2). Both the northern Arabian Sea and SEC regions are projected to experience increases in stratification between 100 - 200 m, as well as decreases in export at 100 m, with opposing effects on oxygen. The spatial correlation of these two trend fields is expected, given the relationship between stratification and nutrient supply via vertical mixing. This suggests that the effects of stratification and export trends, at least, partially compensate each other in both of these regions. We propose to discuss this compensation and how it is consistent with the results of our sections 3.5 and 3.6 (using Optimum Multiparameter (OMP) analysis) which suggests that changes in circulation and water mass fractions related to lateral advective pathways can explain nearly all simulated oxygen changes (direct calculations of mixing-driven oxygen changes will be added in the revised manuscript; Figure GAC3). We have also performed additional sensitivity tests with the OMP method (see new Figure GAC3 and new Method details at end of general response) and propose to add a discussion of the robustness and limitation of the OMP method, and the implications for our results (to be added to *Discussion* section 4.4).
2. **Clarification of approach.** One reviewer is concerned with the definition of Oxygen Minimum Zone (OMZ) used in the manuscript which is a continuum covering a broad range of oxygen values extending beyond the commonly used hypoxic thresholds (e.g., 60 umol/kg). We propose to clarify in the *Methods* section 2.2.1 and throughout the text the rationale for using this broad definition: (1) The relevant thresholds for 'low oxygen' environments can vary significantly depending on the application, such as (in order of increasing threshold) denitrification, mortality of marine organisms, and sublethal stresses on marine organisms. Specifically, thresholds of ~150 umol/kg (and even as high as 200 umol/kg) are commonly used as habitat boundaries for large commercial fish species (eg Vaquer-Sunyer and Duarte, 2008); (2) Using a broad continuum of OMZ thresholds enables us to interpret changes in oxygen volume (including the hypoxic volume) within the context of larger-scale forced changes (contraction, redistribution, expansion).
3. **Robustness of findings across individual Earth System Models (ESM)**. The reviewers note that much of the analysis presented in the manuscript is based on the presentation of Multi-Model Means (MMM) of the ESM ensemble, which may mask

> compensating biases and different dynamical responses across models and may be sensitive to model selection. We propose to a) expand the discussion of the maps/sections of individual ESMs already included in the Supplementary Information (SI) of the original manuscript, b) add individual ESMs when presenting OMZ volume and OMZ volume trends in the main text (new Figures GAC4, GAC5, GAC6), and c) compare the effects of water mass fraction changes (from OMP analysis) across individual models (Figure GAC3). Overall, we find that most models exhibit the same regional pattern of oxygen trends and regimes of OMZ volume trends as the MMM, with varying amplitudes of features across models. This will also be discussed in Discussion section 4.4.

We are confident that these revisions will address the concerns of the reviewers. Below, we provide specific details and figures for these revisions, and we address additional reviewer comments.

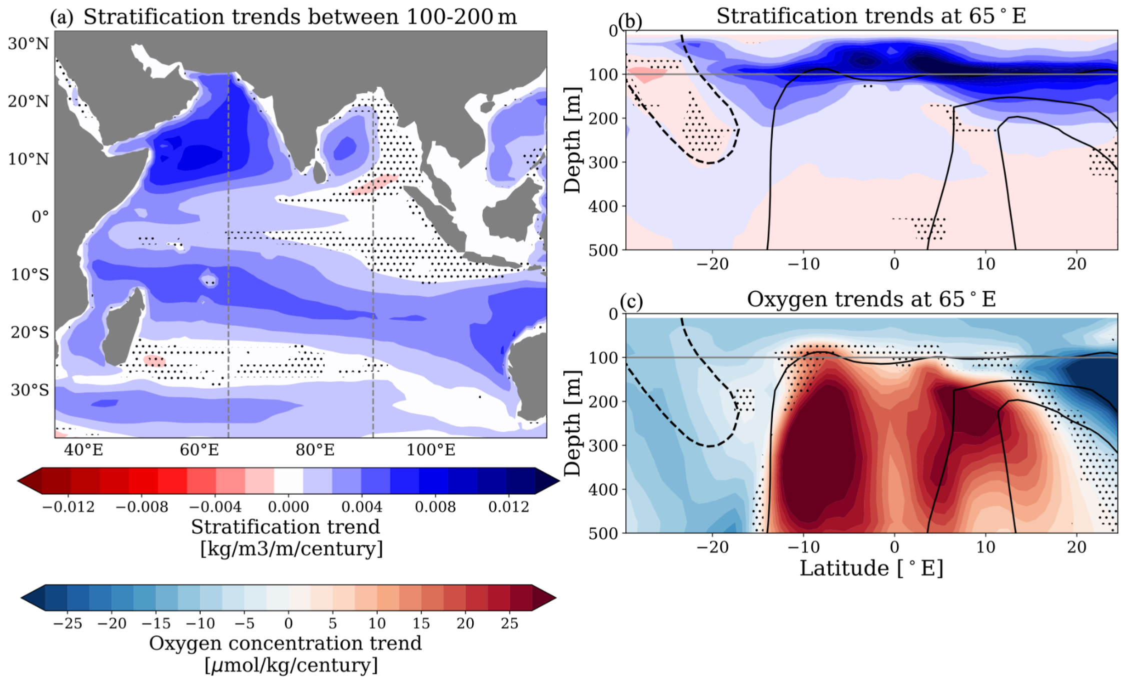

Figure GAC (General Author Comment) 1: Multi-model mean (MMM) stratification trends under SSP5-8.5 scenario forcing (1915-2100). Stratification trends (a) between 100 and 200m, and (b) at 65E. (c) dissolved oxygen trends at 65E. (b,c) Solid black contours represent 20, 60 and 150 µmol/kg oxygen. Dashed black contours highlight salinity signature of Subtropical Underwater (STUW). Results are stippled where less than 75% (6/8) of models agree on sign of trend.

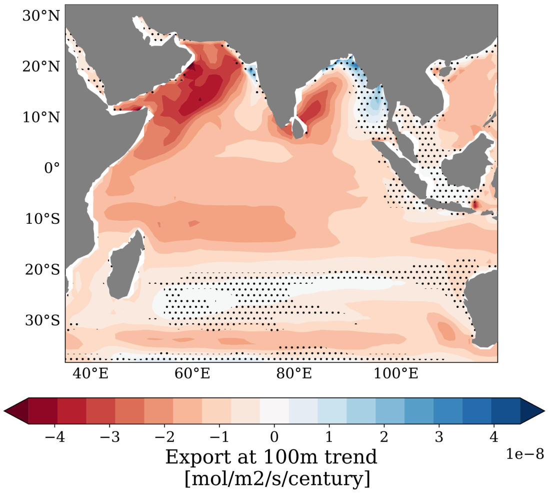

Figure GAC2: Multi-model mean (MMM) export of organic carbon at 100 m trends under SSP5-8.5 scenario forcing (1915-2100). Results are stippled where less than 75% (6/7) of models agree on sign of trend.

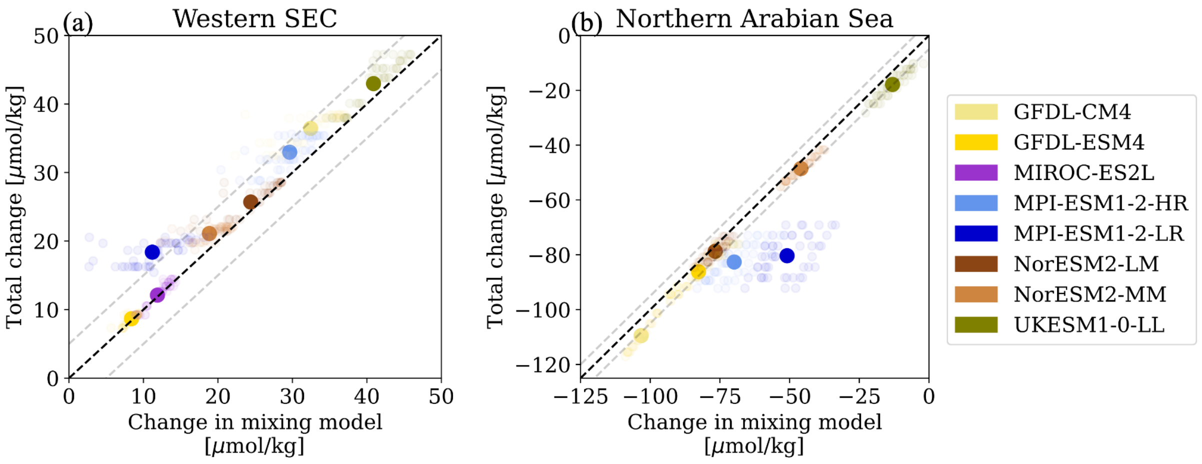

Figure GAC3: Changes in dissolved oxygen accounted for by mixing model analysis. Total oxygen changes (simulated by ESMs) versus oxygen changes attributed to shifts in water mass composition in the OMP mixing mode for the (a) western South Equatorial Current (SEC) and (b) northern Arabian Sea. Results in the western SEC averaged between 40-80E, 5-15S. Results in the northern Arabian Sea averaged between 62-70E, 20-26N. Solid markers are the mean of 50 runs (light markers) of the OMP analysis with perturbed source water properties (See end of general response for details on new methods). Black dashed line marks 1-to-1 relationship, gray dashed lines mark 5 umol/kg deviations from 1-to-1.

New method to be added to Optimum Multiparameter (OMP, section 2.2.4) analysis for sensitivity analysis and used to produce Figure GAC3 above:

“From the mixing model results, we quantify the change in oxygen accounted for by shifts in water mass composition for key regions: $\Delta O_{mixing} = \sum_{i} (f_{i,1} O_{i,1} - f_{i,0} O_{i,0})$, where f_i and O_i are the fraction and oxygen concentration of source water type i, for future (1) and historical (0) states. Oxygen is allowed to be remineralized, so ΔO_{mixing} need not approximate the total simulated change in oxygen. To test the sensitivity of the mixing model results to small changes in the definition of source water types and potential density layer, we average 50 realizations of the experiment applying random perturbations to source water locations and density values. Source water locations are perturbed by up to 5 degrees in latitude and longitude for the South Equatorial Current region and 2 degrees for the Arabian Sea. For both locations, the value of the potential density layer is perturbed by up to 0.1.”

We note regarding the OMP analysis that we lowered the weight on oxygen in the optimization compared to the submitted manuscript (lowered from 20% the weight of temperature and salinity to 2%) to more appropriately account for the range of oxygen values in the analysis compared to the range and temperature and salinities (see Appendix A of submitted manuscript for technical details of OMP).

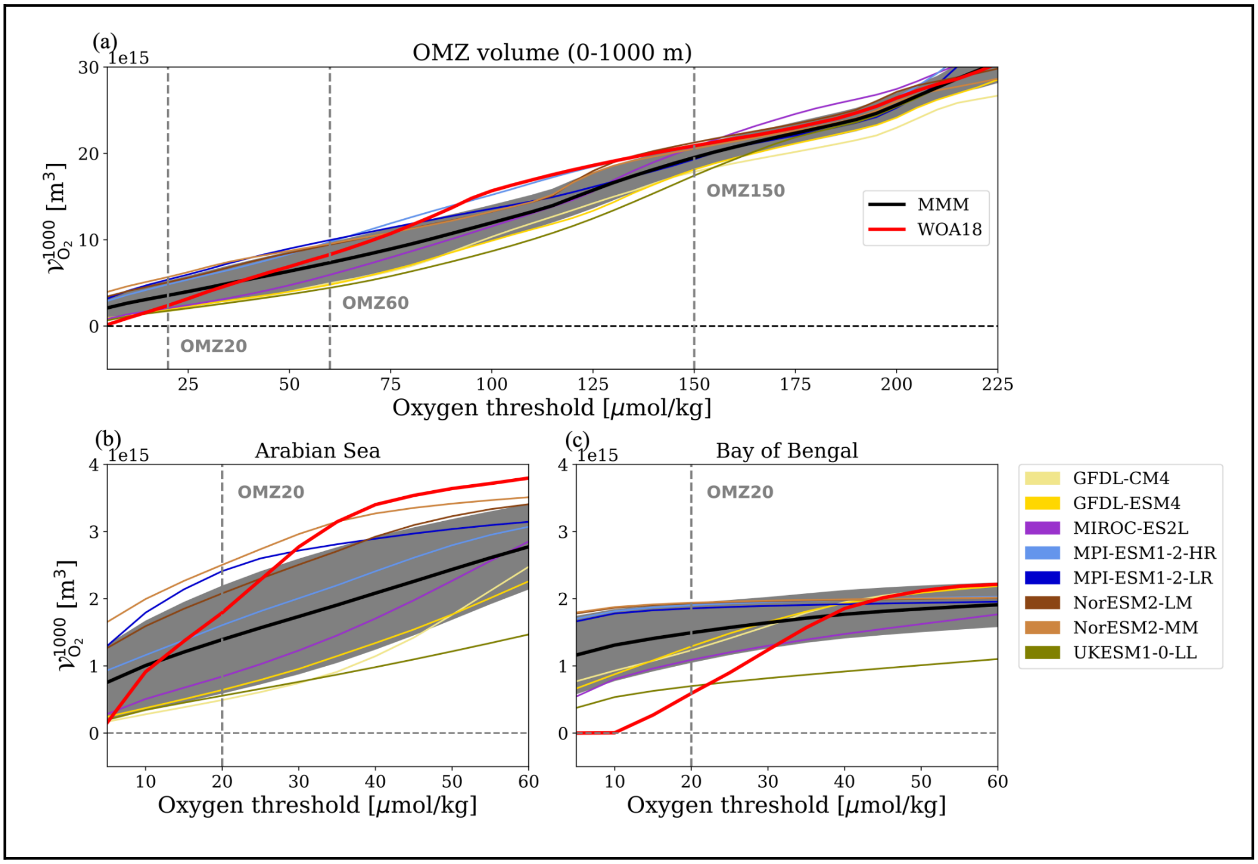

Figure GAC4: (**To replace Figure 2 in main text**) OMZ volume taken between 0-1000 m in the (a) Indian Ocean and sub-basins (b) Arabian Sea and (c) Bay of Bengal for the multi-model mean averaged over 1950-2015 (MMM; black) and observed climatology (WOA18; red). Shading represents one standard deviation of the model spread. Individual ESMs shown in colored curves. Southern boundary of Arabian Sea and Bay of Bengal are at 5N.

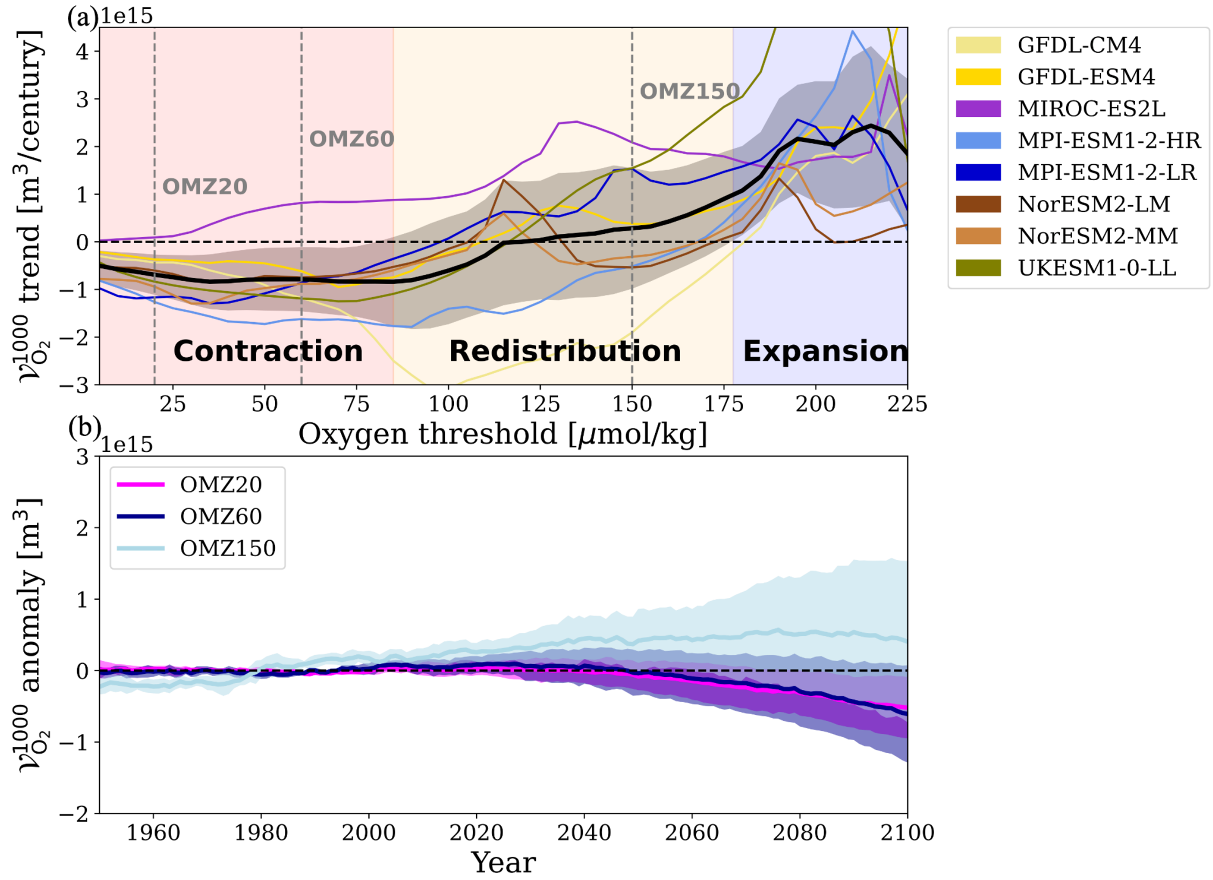

Figure GAC5: (**To replace Figure 3 in main text**) Multi-model mean thermocline OMZ volume changes (between 0 - 1000 m) under SSP5-8.5 scenario forcing (2015-2100). (a) V_{o2}^{1000} trends as a function of oxygen threshold for the Indian Ocean. The 20, 60 and 150 umol/kg thresholds bounding OMZ20, OMZ60 and OMZ150 are indicated with grey dashed lines. (b) Time series of V_{o2}^{1000} anomaly from 1950-2100 (anomaly referenced to 1950-2015 mean) for OMZ60 (dark blue) and OMZ150 (teal). Shading represents one standard deviation of model spread.

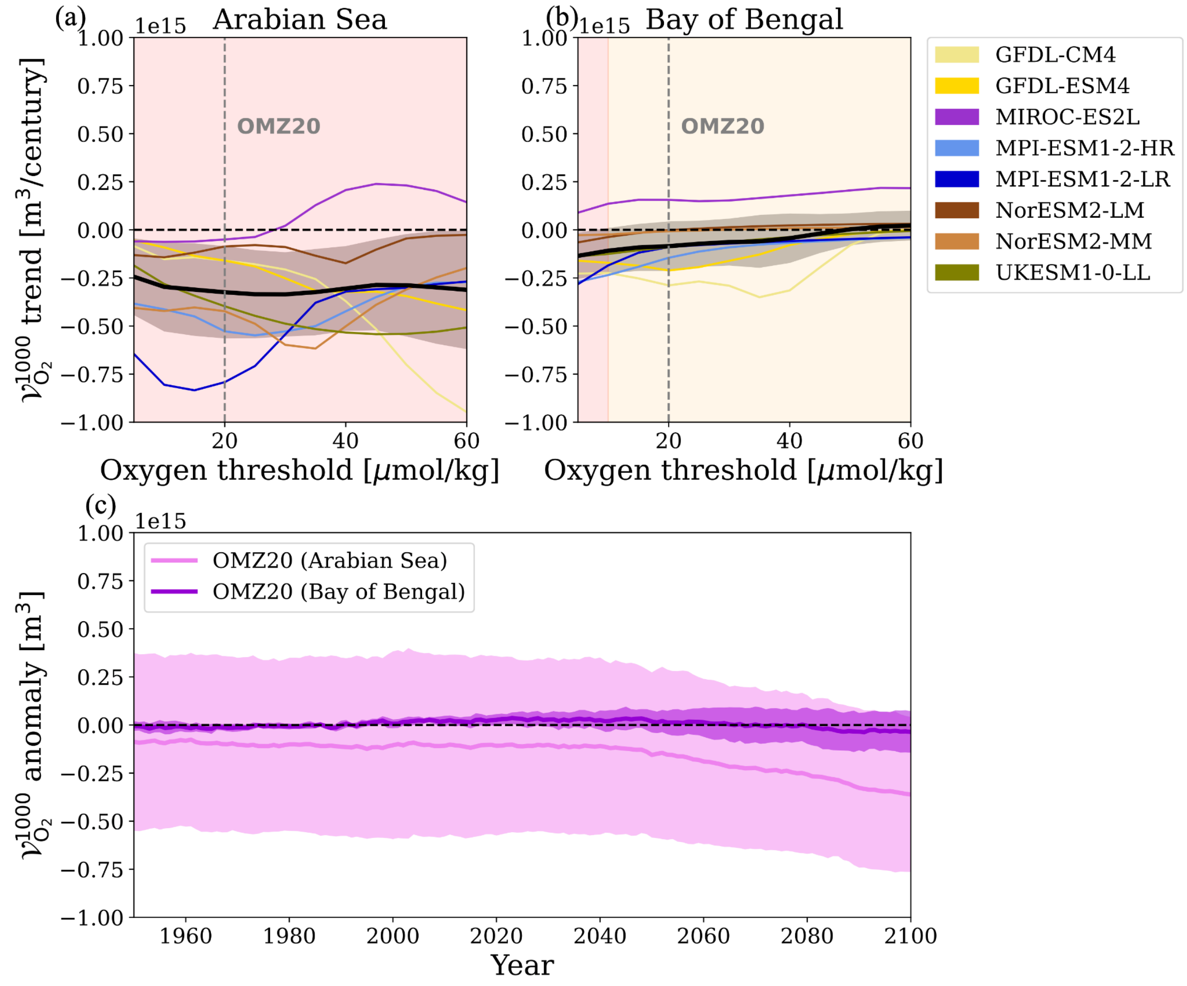

Figure GAC6: (**To be added to supplementary information**) Multi-model mean thermocline OMZ volume changes (between 0 - 1000 m) under SSP5-8.5 scenario forcing (2015-2100). V_{o2}^{1000} trends as a function of oxygen threshold for the (a) Arabian Sea and (b) Bay of Bengal with OMZ20 volumes indicated for each basin. (c) Time series of V_{o2}^{1000} anomaly from 1950-2100 for OMZ20 volumes in the Arabian Sea and Bay of Bengal. Shading represents one standard deviation of model spread.

Reviewer 2:

Review of Ditkovsky et al.
Anand Gnanadesikan
This paper considers drivers of changes in hypoxia in the Indian Ocean, a critical region for artisanal fisheries and one whose behavior under global warming has not been well characterized. The paper finds three regimes of oxygen change, which correspond to three different driving mechanisms for that change. The authors kindly credit me with distinguishing "single-pipe" from a "mixing network" models and describe how this can be used to distinguish the changes in the northern, central, and southern Indian Oceans and how it needs to be updated to do so. Frankly, I think their mapping onto the different regimes is clearer than what I wrote about- this is an elegantly written paper. The argument that there are three separate regimes is generally well made.

I have three comments, that in some ways parallel those made by the other reviewer. Normally, I would give this a "major" revision since they will require some new analysis, not merely clarification of existing analysis. But I think the basic analysis is sound and don't want to suggest "reconsideration" of the paper, so I'll call this minor.

1. The first comment regards the role of the overflows. There is a lot more to overflows than the volume that they deliver to the ocean. It's been a problem for a long time in models to get the depths of injection of overflow water correct. If I look at the salinity along 70E (Figure R1), it's clear that there's a signal from high-salinity shelf waters that penetrates for hundreds of m. It is unlikely that models, which generally have problems with numerical entrainment, correctly capture either this process or its sensitivity to changes in climate. This is even true for relatively high-resolution models (see Seddigh-Marvasti et al., 2015 for some discussion of this). It would be good to at least evaluate how much of a problem this is, rather than simply accepting the results of the MMM (cross-sections of the watermass fractions might be useful to look at this and discern whether there are any systematic errors here). This wouldn't need to be an extra figure in the main text but would make a good one in the Supplemental material and would be useful for evaluating whether there are any systematic errors here.

Indeed, the Arabian Sea is a region of strong salinity and oxygen bias in ESMs in large part due to overflows (eg Schmidt et al., 2021). A comparison of salinity sections at 65E (Figure R2AC1 below) suggests that there is significant inter-model variability in the strength of the overflow signals, with most models simulating too-saline subsurface waters compared to WOA. However, our model selection process, based on an oxygen criterion does mitigate the salinity bias in the Arabian Sea, as well as the oxygen bias. We propose to extend our explanation of model selection in section 2.1 as follows:

“Out of the 14 CMIP6 ESMs that provided monthly dissolved oxygen data for the pre-industrial control, historical, and SSP5-8.5 experiments, we exclude 6 models that simulate virtually no suboxic (<10 umol/kg) volume in the Arabian Sea (ACCESS-ESM1-5, CanESM5, CanESM5-CanOE, CNRM-ESM2-1, IPSL-CM6A-LR; Table R1AC1). We keep the 8 remaining ESMs (GFDL-CM4, GFDL-ESM4, MIROC-ES2L, MPI-ESM1-2-HR, MPI-ESM1-2-LR, NorESM2-LM, NorESM2-MM, UKESM1-0-LL). All six ESMs excluded from the multi-model mean exhibit above average salinity biases in the Arabian Sea from outflows (Fig. R2AC1) and four of six models exhibit Red Sea outflow rates over twice the observed rate (Fig. S6 *from original submission – Red sea outflow rates*}). Thus, the representation of marginal sea outflows may be improved significantly in our ensemble by excluding these ESMs.”

While salinity biases in our ensemble are improved by model selection, they are certainly not eliminated. We note that there are compensating salinity biases between models in our selected ensemble that may mask issues when examining a multi-model mean, namely the fresh MIROC-ES2L (which does not have overflows) and the deep salinity biases of the NorESM2-LM and NorESM2-MM. A discussion of these biases will be included in section 3.1 on model evaluation.

Consideration of these biases does not change our analysis, but it will factor into the discussion section 4.4 on interpreting our results for the Arabian Sea. The extended revision of this section is attached in response to your third comment.

2. The neglect of changes in productivity is understandable, but I note that it was also picked up by the other reviewer. One way of addressing this is to look at how much of the change in the oxygen can be accounted for by changes in the O2:age slope and how much can be accounted for by changes in the age itself (i.e. to the extent that $AOU=JO_2*age \rightarrow \Delta AOU=JO_2*\Delta age+age*\Delta JO_2$) Showing not just the correlation coefficients but the regression coefficients might help with this.

We propose to address changes in productivity by including an analysis of trends in export of organic carbon at 100 m (see General Response 1). ΔJO2 may be a difficult quantity to interpret, as it is an average rate over the lifetime of a water parcel, and thus will be strongly influenced by remote changes outside of the Indian Ocean. The relationship between AOU and JO2 can also be distorted by mixing giving spurious results (e.g. Guo et al., GRL 2023), as well as issues with ideal age in CMIP6 ESMs (requirement of long spin-ups). Instead, we propose to add a subsection to our *Results* that will present the following points regarding changes in the biological sink and stratification:

The mixing model experiments suggest that nearly all of the simulated oxygen changes in the SEC and northern Arabian Sea can be accounted for by shifts in water mass compositions driven by changes in advective ventilation pathways (Fig. GAC3). Oxygen changes captured by the mixing model generally agree within about 5 umol/kg of the total simulated oxygen change for most models (with the exception of the MPI models), and this result is robust to small changes in the definitions of source water mass types (Fig. GAC3 presents a composite of 50 runs with perturbed source water properties; see new methods for OMP). The residual oxygen changes that the mixing model does not capture, while small, tend to be systematically biased positive for the western SEC and negative for the northern Arabian Sea. This suggests the influence of additional processes not accounted for in the mixing model. These processes include increases in subsurface stratification which can influence the vertical mixing between different water masses and changes in biological oxygen consumption after a water parcel leaves its source water region. We examine trends in stratification under SSP5-8.5 forcing (Figure GAC1). Strong stratification increases are generally confined to the upper 100 m of the water column. However, in both the Arabian Sea and SEC regions, the multi-model mean simulates significant stratification increases between 100-200 m. These stratification increases are simulated consistently across the ESM ensemble for both regions (*Individual model stratification trends in supplementary*). Across the ensemble, most of the oxygenation in the SEC region occurs below 200 m, where stratification does not show strong changes, suggesting that changes in vertical mixing likely play a relatively minor role in this region. In contrast, most of the deoxygenation in the northern Arabian Sea is collocated with a stratification increase between 100-200 m, suggesting that reduced vertical mixing of oxygenated surface waters downward contributes to the deoxygenation simulated in this region.

Increased subsurface stratification tends to limit the vertical mixing of oxygenated surface waters downward, but also limits the mixing of subsurface nutrients into the surface. This in turn limits primary productivity and the export of organic matter to the subsurface, setting a control on oxygen consumption rates at depth. We examine trends in the export of organic carbon at 100 m (available for 7 out of 8 ESMs; Figure GAC2). The export of organic carbon consistently declines under SSP5-8.5 forcing over the Indian Ocean in the multi-model mean, with local maximum declines in the western SEC and Arabian Sea regions. Declines in export are particularly strong in the Arabian Sea in 4 of 7 ESMs (MIROC-ES2L, MPI-ESM1-2-HR, MPI-ESM1-2-LR, UKESM1-0-LL) and in the western SEC in 2 of 7 models (MIROC-ES2L, UKESM1-0-LL; *Individual model export trends in supplementary*). The Arabian Sea and SEC experience both a significant decline in export rates and increased stratification at depth with opposing effects on oxygen changes, and thus are perturbed by processes not captured by the mixing model. The residual (non source water-driven) oxygen changes from the mixing model suggest that decline in export dominates over stratification increases in the SEC, and vice versa in the northern Arabian Sea. Though compensation between the effects of export and stratification changes may be key to allowing the mixing model to perform well in both regions. While the mixing model results plausibly show that the pattern of oxygen trends simulated by our ESM ensemble can be driven by changes in advective ventilation pathways, a quantitative decomposition of contributions from advective ventilation, biological sink and stratification changes is not possible with available data.

3. Finally, I wanted to see whether this picture seemed to work in my own suite of coarse models reported in Bahl et al., 2019. These models don't have a Red Sea or Persian Gulf, and also fail to generate an oxygen minimum zone in the Northern Arabian Sea (this, incidentally, supports the point of this paper and others that resolving the impacts of such water is important).I show results for two cases with low and high lateral mixing in the figure below. Interestingly, it does seem that the same 3 regimes show up.

Since I have a remineralized phosphate tracer in this model I can also directly attribute the changes to changes in accumulated phosphate, and again, this dominates the pattern at low mixing, though somewhat less at high mixing. Ultimately I think this highlights an interesting question of whether the real world Indonesian throughflow acts as a barrier to or enhancer of tracer mixing. It also raises the question of how much of the intermodel variability is due to how this subgridscale mixing is handled. That the basic framework seems to work in this model suite as well is encouraging and supports the publication of the manuscript. Note,however, that despite not having marginal seas, we still get a drop in the North. This seems to be driven by a shallowing of mixed layers (echoing something noted by the other reviewer), which reach over 100m on average in the winter in the Northern Arabian Sea in this model, but shallow substantially under global warming. This may be difficult to capture with the watermass analysis alone, as it is not clear (at least to me) that the MMM will necessarily capture the differences between overflow and surface watermasses in this region. It would be worth examining changes in mixed layer depth to see to what extent this plays a role in the more realistic models.

Thank you for providing these examples. It is interesting to see the same pattern of oxygen trends emerge, and the deoxygenation simulated in the Arabian Sea without overflows in your suite of models (this is also the case in MIROC-ES2L). We investigated forced changes in winter mixed layer depths (MLD) in the Arabian Sea. We found that while the winter MLD shoals by up to 30 m in some models, the MLD was still too shallow (sitting around 100 m) to explain the volumes of deoxygenation. So instead, we examined 3D fields of stratification trends more generally, and find that much of the deoxygenation is collocated with increased stratification in most models (see new results section above). We propose to interpret the relative influence of stratification changes versus changes in outflow ventilation – as well as other challenges in interpreting projections for the Arabian Sea – as an expanded discussion in section 4.4:

“Deoxygenation in the northern Arabian Sea has been detected in observations (reviewed by Lachkar et al., 2023). The rapid heating of marginal seas compared to the open ocean and the subsequent increase in stratification and buoyancy of the marginal sea outflows, particularly for the Persian Gulf is also supported by observations (Al-Yamani et al., 2017; Naqvi, 2021). The projected changes in outflow ventilation are consistent with the findings of ocean model simulation studies, which showed that the warming and shoaling of the Persian Gulf was a major driver of the deoxygenation in the northern Arabian Sea (Lachkar et al., 2019, 2021). Yet, ESMs tend to overestimate the oxygen transport by these marginal sea outflows to the Arabian Sea (e.g. Schmidt et al., 2021). This partially explains the systematic high oxygen bias in the Arabian Sea in the ensemble of eight CMIP6 ESMs used in this study (Fig. 1), the absence of a suboxia in the five ESMs that were excluded from the ensemble (Table R1AC1), and could influence the magnitude of the projected deoxygenation associated with the vertical displacement of these outflows. In fact, a study by Vallivattathillam et al. (2023), which applies downscaling methods to CMIP5 projections of Arabian Sea oxygen under RCP8.5 forcing, finds that deoxygenation in the region simulated in CMIP5 models is not preserved after a bias-corrected downscaling method is applied (sampling between 200 and 700 m). This suggests that ESM simulated mean state biases in the Arabian Sea can have a profound impact on projected changes.

There are also aspects of the Arabian Sea which may limit the accuracy of the mixing model (Optimum Multiparameter; OMP) analysis employed in this study. Specifically, it may be difficult to distinguish the contribution of Persian Gulf Water compared to Arabian Sea surface water because both tend to be anomalously warm, saline and well oxygenated. Thus, it is possible that weakened vertical exchange with the surface is aliased as a decrease in Persian Gulf Water fraction. The presence of Northern Arabian Sea deoxygenation in MIROC-ES2L, an ESM which does not resolve the marginal seas and was thus excluded from the OMP analysis for the region, also suggests that stratification increases may play a first-order role in Northern Arabian Sea deoxygenation in the ESM ensemble. A previous study by Lachkar et al. (2021), which investigates deoxygenation in the Arabian Sea from 1982 to 2010 using a high-resolution regional ocean model, attributes about 75% of ventilation decreases over this period to vertical mixing changes and only 25% to advective changes. In contrast, Vallivattathillam et al. (2023) demonstrate that remote physical forcing (including marginal seas) drives deoxygenation below 200 m in the northern Arabian Sea while local forcing (including stratification from local warming) does not in downscaled projections under RCP8.5 forcing.

Further work, including for example a full oxygen budget analysis, would be required to better disambiguate the effects of marginal sea shoaling and increased stratification in the projections of Arabian Sea oxygen for the ESMs used in this study; however, data for such a study are not currently available. Based on these considerations, the relative effects of marginal sea shoaling and increased stratification in ESM projections of the northern Arabian Sea remain largely uncertain. Ultimately though, the present study supports the conclusions of the recent review by Lachkar et al. (2023): the Arabian Sea OMZ will be reshaped by changes in stratification and the Persian Gulf outflow driving deoxygenation in the northern subsurface layer, while increased ventilation from the south drives oxygenation of the deeper layers."

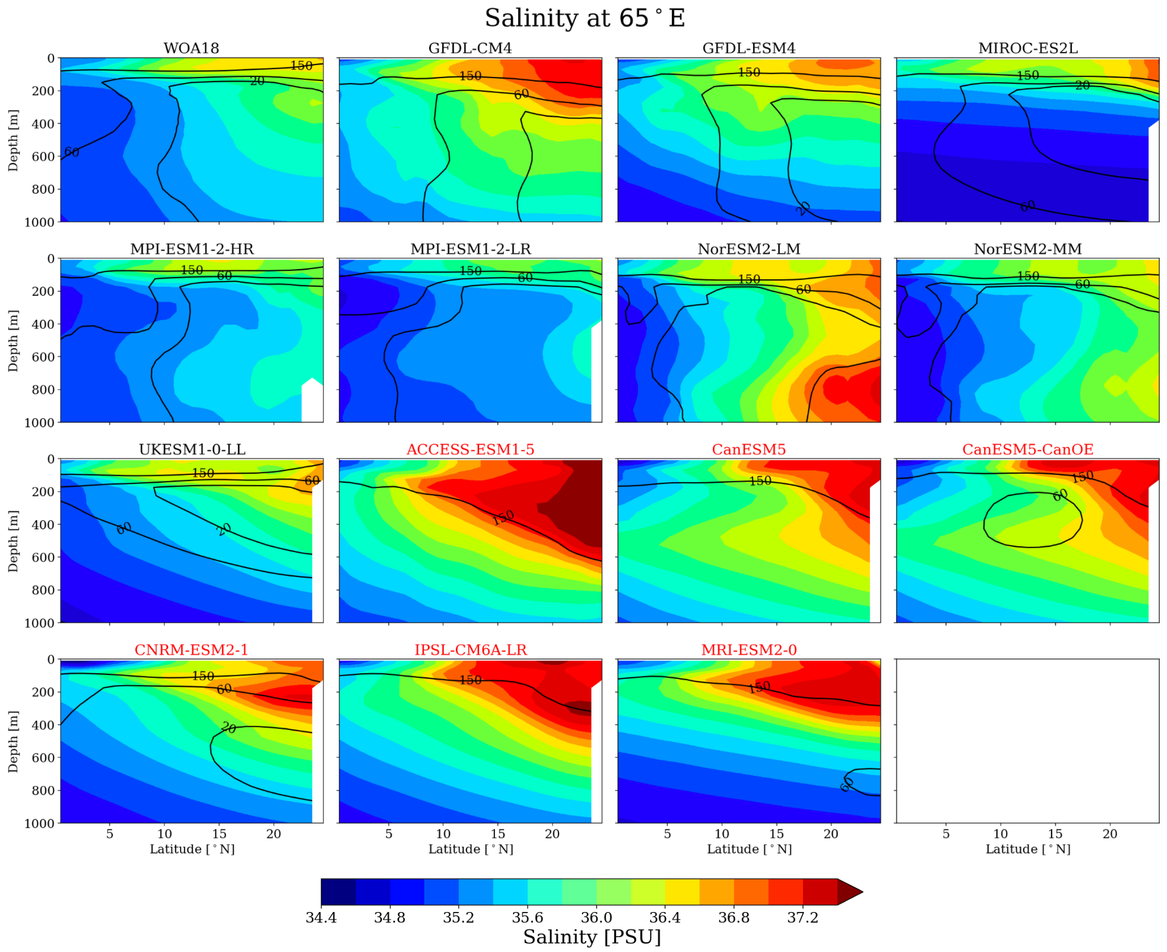

Figure R2AC1: Section of historical mean salinity in the Arabian Sea at 65E in World Ocean Atlas (WOA) and CMIP6 models. Historical mean in CMIP6 models taken from 1950-2015.