Oxygen paper comments Referee #2:

Cabre et al examine 11 Earth System Models with respect to their representation of present Oxygen Minimum Zones (OMZs), and their projection of OMZs and oxygen inventory. For the mean state, they find considerable discrepancies between models and observations in pattern and extent of OMZs. They identify potential reasons for the mismatch between simulated and observed oxygen and OMZs in the mean state, among them model physics (in particular: representation of the equatorial dynamics, ventilation pathways) as well as biogeochemistry (particularly sinking and remineralization of POC). Cabre et al. further examine projections of changes in OMZ, and compare the long-term changes to the effects of interannual variability on AOU, oxygen concentration and saturation in different domains.

The paper is a very comprehensive and thorough analysis of Earth System Models, and about their skill in representing oceanic oxygen dynamics. I am convinced it will contribute to future model analysis and development. I particularly appreciated the examination of the mean state, which covers so many different angles of model analysis: physical and biogeochemical diagnostics (including a very useful metric), as well a structural analysis of the different model components. I definitely recommend this paper for publication. I have just a few, minor comments, which perhaps could help to fine-tune and streamline the paper a bit.

Generally, I think the analysis of the mean state (section 3.1) could be a bit more connected to sections 3.2 (Oxygen changes in the Pacific from 1990-2090) and 3.3 (interannual variability vs long-term changes).

In our intentions the three section were naturally linked through the discussion of the mechanisms responsible for the observed behavior across models. However, we agree that the link might not appear evident to the reader so we introduced some changes to improve this aspect.

We added a sentence in the intro of Section 3.2:
"In this section we present changes in oxygen concentrations and the extent of OMZs in the Pacific throughout the 21st century, as well as the mechanisms responsible for those changes. We also explore how the biases found across CMIP5 models in the mean state (Section 3.1) propagate into 100-year timescale changes."

We added and modified some sentences in the intro of Section 3.3:

In this section we highlight the mechanisms controlling oxygen variations on both interannual and long-term timescales between 10ºS and 10ºN and east of 115ºW, at a depth of 100 to 200m. This domain is chosen to enclose the upper portion of the eastern tropical Pacific OMZ, found to have interesting oxygen dynamics due to the strong compensation between AOU and O_2sat through the 21st century (Section 3.2, Fig. 8c,f,i). We explore here the same compensation mechanisms discussed in Section 3.2 but on interannual timescales. This region is also of interest because tropical oxygen is underestimated in many CMIP5 models compared to observations (Fig. 3a), since models do not properly separate the northern and southern OMZs (Fig. 1) as explained in Section 3.1.

Further, a few words about the relevance of different criteria for OMZ definition and their relevance for organisms (e.g., Hofmann et al., 2013, Kinetic bottlenecks to respiratory exchange rates in the deep-sea – Part 1: Oxygen. Biogeosciences, 10, 5049–5060; Seibel, 2011, Critical oxygen levels and metabolic suppression in oceanic oxygen minimum zones, Jour. Exp. Biol., 214, 326- 336) may illustrate the effects of these assumptions onto organisms (the "real world").

We modified the following paragraphs to add a bit more background on OMZ. However, we acknowledge that this analysis focuses more on inter-model differences than on the real world.

"Most marine organisms suffer and might die in hypoxic conditions, i.e. when the oxygen concentration falls below 60-80 $mmol/m^3$ (Gray et al., 2002; Stramma et al., 2008). Note that this is a limited definition, since the specific survival and performance of organisms also depends on the species, temperature, oxygen partial pressure, and CO_2 levels (Seibel, 2011). However, a definition based on oxygen concentration facilitates the comparison across models. Oxygen minimum zones (OMZs) develop on the eastern outskirts of the subtropical gyres owing to poor ventilation combined with high biological consumption, a consequence of strong upwelling and biological productivity."

...

"With climate change, increased temperatures result in decreased O_2 solubility (thermal effect) while an increase in stratification is expected to reduce the ventilation of the subsurface and hence O_2 concentration (Keeling and Garcia, 2002). Significant global deoxygenation over the past 50 years has been observed (Helm et al., 2011) together with expansion of the OMZ in the North Pacific... with a small associated change in OMZ volume. The expansion of low-oxygen zones will result in the transition, adaptation, and/or extinction of different species in the real ocean. "

New reference:

Seibel, B. A.: Critical oxygen levels and metabloic suppression in oceanic oxygen minimum zones, The Journal of Experimental Biology, 214, 326-336, 10.1242/jeb.049171, 2011.

Likewise, some few sentences about the effect of (model parameterizations of) DOM on simulated OMZs may complete the discussion about biogeochemistry and errors in OMZ.

We have added a discussion on DOM as suggested by the referee. A complete explanation can be found under point 9 in the answer provided to Referee #1.

Minor specific comments: —————————-

p 6527, lines 10 and 13: "Karstensen"

Changed

p 6527, line 17: "northern tropical Pacific" or "Northern Tropical Pacific"

Changed

p 6528, lines 8-10: The models by Kriest et al (2010, 2012) did not include explicit denitrification, but a more recent one (Kriest and Oschlies, 2015, Geosci. Model Dev. Dev. Discuss., 8, 1945–2010) does and examines this with focus on OMZ - skip "denitrification"
or refer to the more recent reference?

Agree. We chose to skip the word denitrification in this part of the Intro.

p. 6529, line 18 ff: What was the reason for choosing 1960-1999 - because there are observations available for this period? Would results look very different if 1860-1899 was chosen as reference period?

Yes, we chose a sufficiently long time span (40 years) to smooth out interdecadal differences and centered around 1980 to make the comparison to WOA results easier and direct. However, preindustrial results look very similar.
We added the following sentence to the text*:* The historical period is chosen around 1980 for a direct comparison to observations

p. 6530, line 5: what was the criterion for "equilibrium"?

We looked at time series of biogeochemical parameters in the historical period. Most models are stable but there is one model - CMCC-CESM - that shows significant trends during this period so we decided not to include it in our analysis for this reason.
We added the following text: "...and had reached equilibrium (no significant trends during the period 1850-1950)."

Section 3.1: It would help the reader if there were more specific references to panels in the figures. (E.g., Fig 1, panel a).

We added some specific references to Fig.1 and 2 in section 3.1 (in red in the new updated text) *p. 6532, line 5: point "(a)" is difficult to see from figure 2.*

We removed the reference to Fig. 2 as Fig.1 is enough here to explain point (a)

p. 6533, lines 15-18: As far as I can see from Fig 3, IPSL-CMA5 has approximately 1 uM PO4 around 100m, while the observations show around 2 uM. Could this be really limiting/the reason for the underestimation? IPSL-CMA5 also shows far too high oxygen at that depth - could that be another reason for the underestimation in subsurface OMZ?

We agree that too high oxygen above 100m, like in the IPSL models, could have some effect on the overestimation of oxygen occurring below this depth through downward diapycnal diffusion. However, from Figure 3 we can see that most models show a marked nutricline between 20 and 40m depth while the IPSL models show a smoother increase of PO4 with depth and, overall, the lowest concentrations. For oxygen, the inter-model differences are similar and opposite which leads us to think that the control on oxygen here is mainly driven by biology and that this control is also propagated deeper in the water column.

p. 6536, lines 14-26: Could another reason for the overestimate in deep oxygen by some models be due to errors in biogeochemistry / too little remineralization having taken place in these waters? (Instead of "only" errors in physics.)

Yes, we state that in line 27, " However, the anomalously deep OMZ found in some models in the tropics (Fig. 1c-d and Fig. 2) does not seem related to ventilation sources such as AABW, AAIW, or NPIW, as models having the same biogeochemical module (MPI-ESM and NorESM1-ME) show similar OMZ extent despite having significantly different representations of different water masses (Appendix Table A2). For example, NorESM1-ME shows one of the highest deep O_2 levels but still develops a deep OMZ (Fig. 2). This problem is discussed in Section 3.1.2."

p. 6537, line 1 "[...] having the same biogeochemical module (MPI-ESM and NorESM1-ME) [...]" - do they really have the SAME module, or just "similar" ones? (E.g., same equations, but different parameters?)

These models have the same equations, and same remineralization parameters. We changed 'same biologeochemical module' to 'similar' to account for possible changes in other parameters.

p. 6541, lines 2-4: Does this mean that in this model (CESM1-BGC) decay of organic matter continues without using oxidants (oxygen, nitrate, sulfate, ...)?

Referring to: "The model CESM1-BGC does not develop a deep OMZ despite using an exponential curve (Figs. 6b and 2), similar remineralization parameters and ballast terms to GFDL-ESM2 models, partly because the denitrification length scale is much lower than in GFDL-ESM2 (260m compared to 1500 m) and also because the model adjusts the nitrate consumed to explicitly avoid running out of nitrate (Lindsay et al., 2014) and expanding the OMZ too deep."

Lindsay et al. (2014) state the following: " In order to avoid an unrealistic loss of nitrate, we artificially scale model predicted denitrification by NO3=110, where NO3 is the model predicted nitrate concentration in units of mol/l." To answer the referee, this means that the decay of organic matter continues at the same rate without using as much nitrate as it would usually be needed given the equations parameters. It also means that organic matter is consumed normally in low-nitrate zones, and hence not transferred to depth as it would happen if there was no oxygen nor nitrate. That alleviates the expansion of OMZ to depth.

We added a bit more explanation in the text: "The model CESM1-BGC does not develop a deep OMZ despite using an exponential curve (Figs. 6b and 2), similar remineralization parameters and ballast terms to GFDL-ESM2 models, partly because the denitrification length scale is much lower than in GFDL-ESM2 (260m compared to 1500 m). Additionally, the CESM1-BGC model adjusts the nitrate consumed during denitrification to lower values to avoid running out of nitrate (Lindsay et al., 2014), which might suppress the transfer of POC to depth, and hence the expansion of OMZ to depth.."

It may be too much for this paper, but: would it help / wouldn't it be interesting to also examine the models with respect to how they treat (a) "oxidant conservation", and (b) "DOM" dynamics (which may affect formation and persistence of OMZs particularly in the equatorial Pacific).

We agree that "oxidant conservation" is an interesting topic, however, except for CESM1-BGC (see previous comment) we haven't found any indication in the models' literature that oxidant is not conserved. We will consider, however, this aspect for future analysis. We have added some consideration on the role of DOM in determining models' behaviour in Section 3.1.2 (see also answer to comment 9 by referee #1).

p. 6543, lines 4ff: This seems to be a really informative and helpful metric!

Just a suggestion: In section 3.2 I first had some problems following the authors' train of thought. It helped a bit to see Fig 10 (afterwards), because I assume the distinction among the different domains in that figure and in section 3.2 seems to be roughly the same. If this is indeed the case, it might help to refer to Fig 10 earlier.

Good idea. We added " Fig. 10a shows a schematic of the regions of interest described in the following text" before describing the different regions.

p. 6545, lines 21-22 "decrease age (Fig 8l) and increase AOU (Fig 8i)": where (at what depth level) exactly is this happening?

Referring to:" We see that, over the 21st century, climate-driven decreases in deep ocean ventilation along the AABW in the Southern Ocean and along the NPIW in the North Pacific decrease age (Fig. 8l) and increase AOU (Fig. 8i), contributing to a decrease in O2 along these ventilation pathways (Fig. 8c)."

We changed it to increase age and increase AOU (typo!), which makes much more sense. These lines refer to high latitudes (100-1000m) and deep ocean (deeper than 1000m) in Fig. 8.

p. 6546, line 4 "Karstensen"

Changed.

p. 6547, lines 1-2 "The trends in both AOU and O2sat are consistent across models (Fig. 8f and i)" - For AOU (8i) there seem to be many unhatched areas in intermediate low latitude depths.

We changed it to: The decreasing trends in both AOU and O_2sat are consistent across models in the most central part of this region (Fig. 8f and 8i).

p. 6547, lines 15-28 and Fig 9: I find the spread of projected OMZ volume quite considerable, even when disregarding the two most extreme outliers (IPSL and HadGEM2).

Including the variability, for the very low thresholds (<5 uM) the remaining models seem to encompass -10% to +20%. Perhaps discuss this a bit more (if only briefly), as these are the ranges/thresholds only a few organisms will be able to cope with, and where most likely denitrification sets in. Further, in Fig 7 the authors have already used varying thresholds for model skill assessment - perhaps discuss these findings, and there implications for assessment of future development a bit more at this point?

Added some sentences to: "As oxygen remains approximately constant with climate warming in mid-depth tropics, 21st century changes in tropical OMZ volume are small in most CMIP5 models (Fig. 9). In anoxic regions (O_2 < 5 mmol/m^3) CMIP5 models show an increase or decrease in volume (by about 10%) depending on the balance between O_2sat and AOU contributions and changes are indistinguishable from natural variability for GFDL models. Changes in the volume of anoxic regions are of critical importance for the survival of organisms and for denitrification, which occurs at low oxygen values. However, the skill of models in representing anoxic regions is poor (Fig. 7a), so this result should be taken with caution. The volumes of hypoxic regions (O_2 < 80 mmol/m^3) mostly increase, although slightly, due to decreased overall ventilation and increased AOU. Hypoxic volumes encapsulate wider regions than anoxic regions, so these are not dominated anymore by local consumption but by overall ventilation. As an exception, IPSL-CM5A predicts an increase of both anoxia and hypoxia, as changes in oxygen levels are dominated by decreased ventilation even within anoxic regions. CESM1-BGC similarly predicts an increase in OMZ volume everywhere driven by changes in ventilation at intermediate depths (500-1000m). The anomalously high interannual variability in HadGEM2 in anoxic regions is due to a small OMZ size, which artificially boosts the relative changes."

p. 6548, line 27 "expected response" - what is this?

Referring to: " It is worth noting that we found no correlation between biases in the mean state and the expected response to climate change, suggesting that the mechanisms underlying the change are similar and robust in all models."

We changed it to:
" It is worth noting that we found no correlation between biases in the mean state and responses to climate change, suggesting that the mechanisms underlying the change are similar and robust in all models."

Section 3.3: This section/analysis seems to focus on a region (10S-10N, east of 115W) that has not been examined before, in the very careful analysis of model mean state. The closest analysis would be Fig 3 (5S-5N, 80-100W). Would it make sense to harmonize this a bit more? This also takes up on the very important finding from that first part, namely that most models have problems representing the bimodal structure of the eastern tropical Pacific OMZ. I'd suggest to refer to that section a bit more.

We added some lines to relate to section 3.1 results
The most interesting dynamics in the oxygen system is the strong compensation between AOU and O_2sat through the 21st century below the thermocline in the eastern tropical Pacific (Fig. 8c,f,i). In this section we highlight the mechanisms controlling oxygen variations on both interannual and long-term timescales between 10ºS and 10ºN and east of 115ºW, at a depth of 100 to 200m. This domain is chosen to enclose the upper portion of the eastern tropical Pacific OMZ, found to have interesting oxygen dynamics due to the strong compensation between AOU and O_2sat through the 21st century (Section 3.2, Fig. 8c,f,i). We explore here the same compensation mechanisms discussed in Section 3.2 but on interannual timescales. This region is also of interest because tropical oxygen is underestimated in many CMIP5 models compared to observations (Fig. 3a), since models do not properly separate the northern and southern OMZs (Fig. 1) as explained in Section 3.1.

p. 6552, line 20 "would not trigger denitrification so easily" - do you really mean "denitrification",
or rather "remineralization"?

Referring to: "With realistic equatorial ventilation, the depth of the OMZ would probably be less sensitive to changes in POC profiles because it would not trigger denitrification so easily"

We meant denitrification.
We changed it to: ... would not switch from aerobic remineralization to denitrification so easily.

p. 6552, line 23-25: "We recommend ..." - Sediment models, and their exchange with the water column, would have to be calibrated and their skill be examined, too. Be a bit more cautious here?

Referring to: " We recommend the use of sediment models, since the lack of sediments burial at the sea floor might contribute to anomalously deep OMZ if all the remaining POC is remineralized instantaneously at the sea floor."

We changed it to: " We recommend the calibration and examination of sediment models, since the lack of sediments burial at the sea floor might contribute to anomalously deep OMZ if all the remaining POC is remineralized instantaneously at the sea floor."

p. 6553, line 20: what is "oxygen behavior"?

Referring to:" We also find coherent patterns of oxygen behavior with climate change in the Pacific surface (Sect. 3.2, Fig. 10a), where oxygen decreases due to warming (O2sat decrease)."

We changed 'behavior' for 'trends'.

Fig 1 (as well as other figures below): I found it very difficult to distinguish some line colours. If possible, could the black and grey lines (observations and model means) be a bit thicker? Further, instead of distinguishing the different variants of models by different shades (e.g., purple and pink for IPLS; light and dark green for MPI; light and dark blue for GFDL), maybe one colour, but with straight and dashed lines could be easier to see?

We have updated Fig. 1 as suggested. We previously tried other options to represent all the models at once (including the one suggested by the referee) and decided to stick with the one that we show because it was the clearest option.

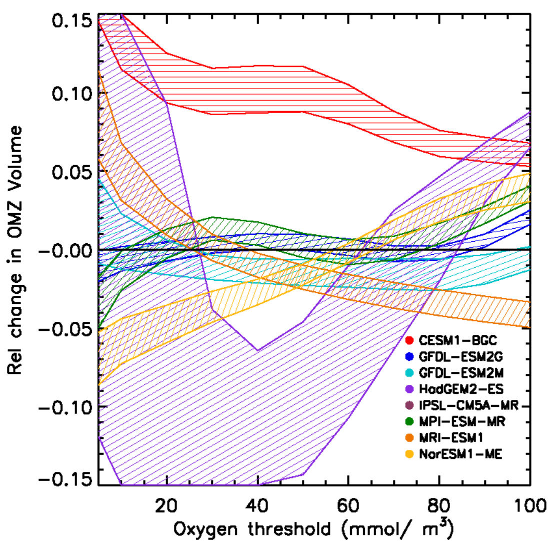
Fig 2 (as well as other figures below): I am a bit undecided about the vertical log-scale: although it helps a lot to see the fine structures at the surface, it is somehow difficult to translate these plots into something that is of significance on a global scale. E.g., an underestimate of oxygen from 1000 m to the bottom would have a huge impact on global model inventory, but not show up strongly in these plots. I don't want to actually recommend changing the scale to a linear, but just note, that these may be somehow difficult to interpret.

We agree that the log-scale is not standard, but in this case it is helpful to focus on intermediate scales where OMZ form, so we find it adequate and would like to keep it as it is.

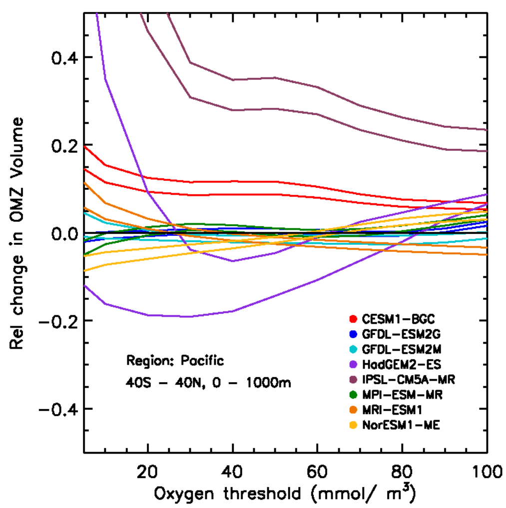
Fig 9: This is an interesting plot, but the two lines for each model are a bit confusing: Is there any way to plot this with mean plus/minus 1 SD, e.g. as horizontal bars, or with transparent shades? Alternatively, the mean with bold lines, and plus/minus one SD as thin dashed lines (if it does not become too crowded).

We suggest changing the figure to the following figure (which excludes IPSL model), so we added the old figure into supp figures (and updated the text accordingly).

NEW FIGURE 9



NEW FIGURE S8 (old Fig. 9)



Rel change in OMZ Volume
0.4
0.2
0.0
−0.2
−0.4
20
40
60
80
100
Oxygen threshold (mmol/ m3)
Region: Pacific
40S - 40N, 0 - 1000m
CESM1-BGC
GFDL-ESM2G
GFDL-ESM2M
HadGEM2-ES
IPSL-CM5A-MR
MPI-ESM-MR
MRI-ESM1
NorESM1-ME