Answers to referees for the submitted manuscript

Evolution of oxygen and stratification in the North Pacific Ocean in CMIP6 Earth System Models

L. Novi, A. Bracco, T. Ito and Y. Takano

Referee #3

Clarifications to the methods for non-specialists

- I am not familiar with the methods regarding data-mining tools (∂-Maps) nor Information Entropy (IE). I found the explanation of how IE was calculated very well put, and I was able to follow without much difficulty. The exception here, however, was when the authors state "the explicit dependence of the entropy quantifier one is removed using the maximum entropy formulation". At this point I was not sure of what the authors were doing, since the way an explicit e is removed is explained in *Ikuyajolu et al (2021)* that the authors point to and I am not familiar with.
  We added a sentence to explain this point. More in detail, the recent work of Prado et al. (2020) proposes a method to free the entropy calculation from the selection a distance threshold (epsilon). They analyzed the dependency of Entropy on epsilon, and found that the Entropy-epsilon relationship has a well-defined maximum (Smax in Fig 4 of Prado et al. (2020)), which is robust and relatively stable within a range of epsilon values, and that this maximum entropy is strongly correlated with the Lyapunov exponent. In our work, we applied the heuristic explained in detail in Ikuyajolu et al (2021) for the calculation of Smax through an iterative procedure that calculates the recurrence entropy for varying epsilon until a maximum is found and retained. This algorithm requires three input parameters: the microstate dimension (that we set at m=4 but explored other values as well in the revision), the number of random samples to compute the microstates distribution in the RP (here 10000) and the number of random sub-samples used to determine the epsilon for which entropy is max (here 1000).

- Another thing, am I interpreting things correctly if the choice to use 4 microstates means that their IE calculation uses 4 probabilities of occurrence (k=4 in equation 2)? Doesn't this mean that the authors are choosing four different e values to create these 4 microstates from the same timeseries of IPV* Euclidean Distances (Eqn. 1)? So here I am confused about how e and Eqn. 1 is actually being done.

  We thank the reviewer for having pointed this out, and we acknowledge that the explanation was not clear, therefore we corrected it in the revised version. In particular, the sentence "using 4 microstates" at line 134 of the initially submitted manuscript should have been instead "using microstates of dimension 4". Indeed, the dimension of a microstate is the size of the NxN matrix introduced at line 124 (of the initially submitted version), i.e. N=4 in this case. Microstates are calculated sampling matrices of size NxN inside the recurrence plot (RP), and the total number of possible configurations of a microstate of dimension N is N_{tot} = 2^(N^2). Therefore, the Probability of occurrence of the generic k^{th} microstate is P_k is used in eq (2) of the initially submitted version and detailed at line 124 (initially submitted version). Therefore, that equation uses N_{tot} different values in the summation. We thank very much the referee for this useful comment, as it helped clarify the method in the revised version.
  We also note that in response to Rev 2, we will be adding in the supplementary a sensitivity analysis based on the number of microstates.

- It is also not clear to me on reading the methods how you calculate the 95th percentile of mean, variability and extremes in Eqn. 3. For the mean (indicator 1), as an example, are you computing the differences across all years in Period 1, then calculating the 95th percentile of these differences? But actually, this doesn't seem to be the case, because you state that *ind1j*is equal to $yseasm2 - yseasm$*), where yseasm1 and yseasm2 are multi-*

year seasonal means in Period 1 and 2, respectively. Multi-year is arbitrary, and on first reading I am inclined to believe that its averaging across the whole length of the period. This suggests to me that *ind1j* is one number, and so it is not clear to me how you then retrieve a 95th percentile. Could you please make this explanation clearer?

The indicators and SEDs are computed point by point, i.e. each grid point has one value. The percentile is therefore computed spatially over all the grid points. We added a sentence in the text to clarify this point.

Other recommendations:

I very much agree with Reviewer 2 in that the paper would benefit from more signposting throughout, and that a reader only really appreciates what they have learned from the results in the final sentences of the conclusions.

I also agree that there is probably additional studies to point to so that the work can be placed amongst the wider literature.

We have addressed all reviewer 2' comments and restructured greatly the presentation of the results to help the readers following the presentation of all the analyses and how they are relevant to the stated hypotheses.

"There is also very little discussion or mention of the other processes affecting O2. Apart from ventilation, there is also the solubility effect of warming and biogeochemical processes affecting oxygen demand. I think the paper would benefit from a discussion of how these two other factors come into play. For the solubility effect, it is not obvious how one would separate it from the ventilation component captured by IPV, since both are driven by warming and the IPV-O2 relationship should actually encompass both the ventilation and solubility effects. I leave it to the authors to think about how the solubility effect could be separated from the ventilation effect. However, for the biogeochemical processes, it would not be so difficult to calculate preformed O2 from the T and S of each model and redo some of your analysis. An analysis of the IPV* - preformed O2 relationship would eliminate any impact of the biogeochemical process, and then allow you to focus on predictability of physical O2 injection. I would expect substantially more predictability and over a much greater area. Similarly, you could take preformed O2 away from your O2 tracer to get AOU, and you could look at the predictability of AOU, which is likely not predictable at all from IPV? Maybe give this a try, and see if some interesting results jump out. This would at least allow you diagnose why IPV-O2 predictability falls over in some regions?"
We thank the referee for this insightful comment. We re-ran all calculations to evaluate the IPV*-linked predictability potential for AOU, i.e. the areas where IPV* and AOU time series are positively correlated with correlation coefficients > 0.5. These areas are very similar to the ones obtained by studying the relationship O2-IPV* proposed in the submitted manuscript. We also separated the solubility part O2sol, which is a very good approximation for preformed O2 at the considered depths, and computed the anticorrelations areas (i.e. where c.c. < -0.5, as for the IPV*-O2 relationship). Interestingly, these areas are not too small (especially in the hindcast) but are mostly superimposed to high-entropy/low predictability areas for IPV*. These results are reported in the figures below. We added a discussion of the two contributions in the revision and the difference in spatial patterns.

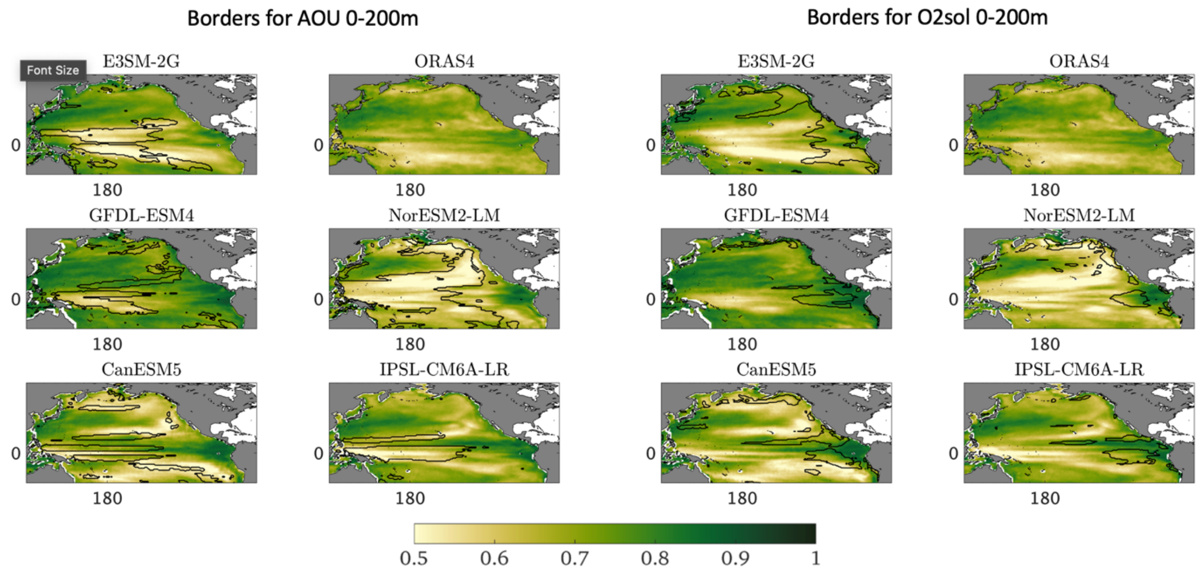

Fig.1: IPV* entropy field in the historical interval for the ESMs, and in the historical 1960-2014 period for the hindcast and ORAS4 with superposed the contours of the areas where IPV* and AOU time series are correlated with correlation coefficients > 0.5 (left), and where IPV* and O_{2sol} time series are anticorrelated with correlation coefficients < -0.5 (right)

Specific comments:

- Line 182: More accurate to say "We obtain three indicators grouped into four seasons for each variable"?

  We thank the referee for the proposed rewording. This is correct and clearer that what stated, so we updated the text as suggested.

- Line 214: I know gridded Argo doesn't provide T and S as far back as your period 1, but couldn't you compare the ORAS4 against gridded Argo in Period 2? This would then provide some measure of how much bias there is in ORAS4, with which you are then using to assess bias in the models. Case in point is that the IPSL IPV fields looks (at least by eye) the least biased. It is not a coincidence because the IPSL and ORAS4 both use NEMO as their ocean models.

  Thank you for this comment, we attach below the comparison between the ORAS4 IPV* climatology over the "period 2" (1988-2014 for this reanalysis) with the corresponding climatology computed using SODA3.4.2, which is suitable as it uses a different ocean component compared to ORAS4 and overcomes the limitations of T and S data from ARGO before 2002. As shown in the figure below, the difference across reanalysis that use different models but assimilate the same observations is much smaller (about one order of magnitude) that the signal, and smaller than any model bias. We will include this information in the Supplementary Material.

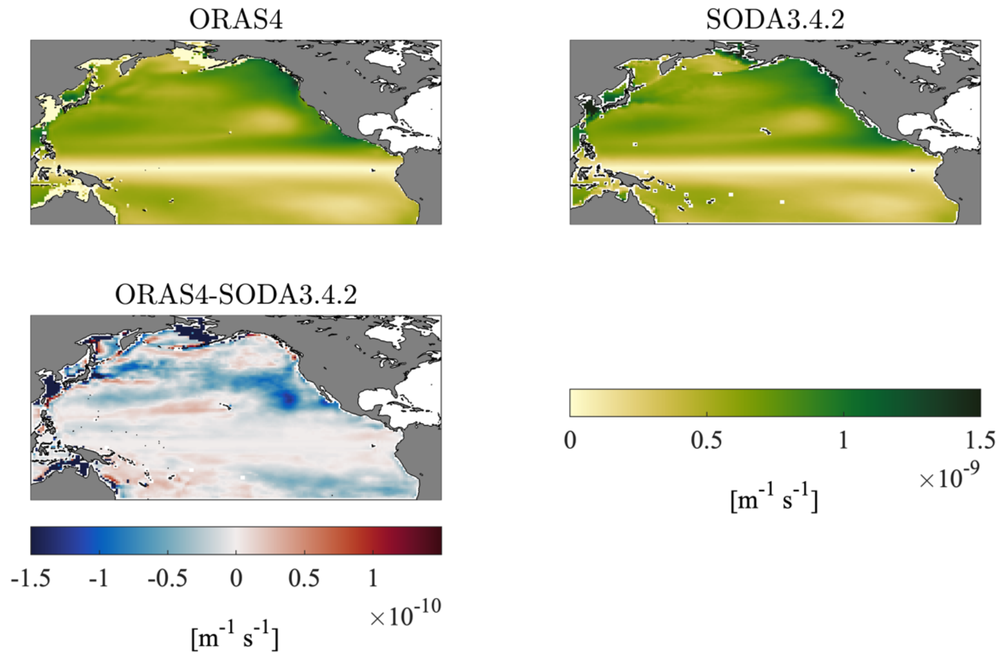

Fig.2: Comparison between IPV* 1988-2014 climatology computed with ORAS4 and SODA.3.4.2.

- Line 221: How is it that the RMSE of the NorESM2-LM is the lowest among the models? Are you sure this is calculated correctly?
  Thank you for this comment. We verified the rmse calculation with three different scripts, two of them in Matlab and one of them via cdo ("climate data operators"). In cdo, for example, we used the following: cdo -L -sqrt -fldmean -sqr -sub model_nor.nc rean.nc rmse_nor.nc
  where model_nor.nc and rean.nc are the IPV* climatology over 1950-2014 (multiyear seasonal means computed with cdo) of NorESM2-LM and ORAS4 respectively.
  We always obtained the same result, $4.92 * 10^{-9}$. We also verified that the climatologies were computed correctly. The two extratropical areas where the bias is higher (which we agree make the visual estimation of rmse hard) are likely more than counter-balanced by the tropical areas where the bias is small. In the revised version, we updated both the plots and the RMSE calculations to comply with what requested by Referee#2, i.e. to re-do all the calculations involving E3SM-2G over 1960-2014 (instead of 1958-2014 for these members) with no significant differences.

- Line 307: Maybe remind the reader here was Ind1 is.
  Thank you, we added a sentence in the text. We also followed the advice of referee#2 are used a more readable notation throughout the revised manuscript.

- Figures 4 and 5: Add text to the legend stating a more intuitive way of interpreting the plots. For *bIPV** you are presenting the change in IPV* (m-1 s-1) per change in SST (ºC-1), right?
  Yes, this is correct. We updated the figures as requested.

- Is the MPAS-O ocean model based on some version of MOM? The correspondence between the two models is striking.
  No, the MPAS-O ocean model is not based on MOM. See Ringler et al (2013).

Supplementary Material comments:

Figure S4: you've stated 1950-2014 twice?
Thank you for catching this typo. It is now corrected.

Thank you for considering my input to your research,
Pearse J. Buchanan.

References used in the answer to Referee # 3

Ikuyajolu, O. J., Falasca, F. and Bracco, A.: Information Entropy as Quantifier of Potential Predictability in the Tropical Indo-Pacific Basin. Front. Clim. 3:675840. doi: 10.3389/fclim.2021.675840, 2021.

Prado, T., Corso, G., Santos Lima, G., Budzinski, R., Boaretto, B., Ferrari, F., Macau, E.E.N. and Lopes, S.R.: Maximum entropy principle in recurrence plot analysis on stochastic and chaotic systems. Chaos 30:043123. doi: 10.1063/1.5125921, 2020

Ringler, T., Petersen, M., Higdon, R. L., Jacobsen, D., Jones, P. W., and Maltrud, M.: A multi-resolution approach to global ocean modeling, Ocean Model., 69, 211–232, doi:10.1016/j.ocemod.2013.04.010, 2013