Reply to review comments

We thank the reviewers for assessing this manuscript and for their time and effort. The constructive comments are much appreciated and have improved the manuscript. The original review comments are given below in black, our reply in blue, and quotes from the revised manuscript in gray.
Please find attached to this reply a revised manuscript where text changes are highlighted.

1 Sarah Schlunegger Referee #1

1.1 General

Hameau et. al. provides a multi-model (CMIP5) assessment of the relative timing of emergence of anthropogenic change in thermocline oxygen and thermocline temperature. To facilitate multi-model assessment of time of emergence (ToE), they provide a new metric, relative ToE. They find that for most of the global ocean changes in temperature emerge prior to changes in oxygen, however for some regions, changes in oxygen emerge prior to changes in temperature. Anthropogenic changes in oxygen emerge prior to temperature in regions where reduced solubility and ventilation work in tandem to reduce oxygen concentrations. In these locations, reduced ventilation also slows the propagation of anthropogenic warming signals from the upper ocean into the ocean interior, further contributing to the delay between emergence of thermocline oxygen and temperature. This paper represents an important contribution to the growing body of mechanistic interpretations of emergence timescales in the global ocean. Additionally, it confronts the known challenge of model-intercomparison with provision of a simple, yet powerful new metric, relative ToE. I recommend its publication in Biogeosciences, but only after minor-to-moderate revision of the text and inclusion of an additional figure to better visualize the regions for which thermocline oxygen robustly emerges prior to temperature.

Sincerely, Sarah Schlunegger, PhD Princeton University Program in Atmospheric and Oceanic Science
Thank you for your general support and your constructive comments.

1.2 Specific:

1. Improve Abstract In the abstract, I would suggest the following rewording to replace some of lines 7-10: Changes in thermocline oxygen emerge prior to changes in temperature because oxygen declines occur due to the confluence (or additivity) of both reduced solubility and ventilation. Otherwise, the abstract could be further streamlined and focused. Feel free to borrow from my summary above.
Thank you very much for the suggestion. The abstract has been revised and streamlined accordingly. Please see the revised manuscript draft at the end of this document.

2. Improve motivation/framing of the introduction Page 3 frames the question of "Which emerges first, physical or biogeochemical variables?" However, this is an ill-formed question, as the chronology of emergence has already been documented to not follow strict temporal separation of physical and biogeochemical variables. For example, Rodgers et al., 2015 evaluates ESM projection of 4 variables and find the following general emergence sequence: pH, SST, O2, NPP – biogeochemical, physical, physical, biogeochemical. Schlunegger et al., 2019, evaluates ESM projection 20 physical/biogeochemical variables, finding their emergence timescales are separated by their association to either chemical (gas-exchange) or physical (warming)

impacts of climate change on the ocean, and NOT to whether the variable itself is physical or biogeochemical. Variables that are first-order impacted by increasing atmospheric pCO2 and increased gas-exchange, like DIC, pH etc., will emerge the most rapidly (carbonate-chemistry related biogeochemical). Later, as atmospheric warming propagates into the ocean (physical) warming signals emerge, and quickly after solubility induces changes (like O2sat). Even later, circulation eventually adjust (physical), thereby altering nutrient supply (biogeochemical), and subsequently primary production (biogeochemical), export production (biogeochemical), etc. The broad order of emergence could then be described as 1. Carbonate-chemistry related biogeochemical and biology (hard tissue pump), 2. Surface temperatures and solubility-related variables (like O2), 3. Ocean dynamics, 4. Biogeochemical impacted by ocean dynamics (nutrients, production).

We have re-written the question. We now emphasize that we are concerned with the detection of the warming and oxygen signals in the ocean interior, as opposed to the surface, and in variables that are routinely observed versus less frequently measured. The text now reads:

Beyond the combined impact of physical and biogeochemical changes, an interesting question is whether anthropogenic changes in the ocean interior are first detectable in variables that are routinely and frequently measured such as temperature (T) or in variables with a relatively low observational coverage but potentially high impact for ecosystems such as O_2 (Joos et al., 2003). The answer may have implications for measurement strategies to detect anthropogenic changes in subsurface waters as well as for the impacts of physical and biogeochemical change on marine life. For the surface ocean, earlier studies (Keller et al., 2014; Rodgers et al., 2015; Frölicher et al., 2016; Schlunegger et al., 2019) showed that the anthropogenic signals of pH and pCO2 emerge earlier than sea surface temperature and O_2 change and earlier than productivity changes. Changes in surface O_2 are tightly coupled to temperature-driven solubility changes and O_2 varies hand in hand with sea surface temperature and the two signals emerge typically concomitantly. Regarding the ocean interior, the sequence of emergence for O_2 and T is less clear. Global warming increases surface ocean temperature, which tends to reduces O_2. On the other hand, O_2 is also influenced by non-thermal processes, such as respiration and the redistribution by ocean circulation and mixing. Respiration of organic matter in the ocean interior may have a larger influence on O_2 change than temperature-driven solubility change in a more stratified and less ventilated ocean. One could therefore expect that, under global warming, the combined effect of increased O_2 consumption and decreased O_2 solubility will accelerate the O_2 depletion in subsurface waters and that O_2 may be detectable before the warming reaches that layer.

3. Additional methodological explanation required Explanation of how S is computed, and its statistical properties relative to N, should be addressed more clearly. For instance, there are broadly two distinct usages of the term "time of emergence" – the first is to define the point in time at which the ocean state/variable is distinctly out of the range of the pre-industrial state, the second, is to define the point in time at which a forced trend is outside the range of how large natural trends are likely to be. I infer that the first meaning is used here, but that should be made clear.

We thank the reviewer for this pointer. Reviewer 2 raised a similar question. Therefore we have extended the paragraph related to the ToE methodology. In addition, a figure is now provided in the appendix to increase clarity (Fig. A1). The following sentence was also added:

Here, ToE represents the moment in time at which the ocean state becomes distinct from the preindustrial state.

4. Potentially reordering the figures to improve narrative flow. Another immediately odd thing that I see is that Figure 1a and 1b at first glance oppose the main hypothesis of the manuscript. This is just visually the case because O2 has higher model disagreement and T and therefore more area where at least 2 of the models think emergence does not occur this century. However, I know many (mostly senior) scientists who just read the abstract, skim the method and the look at the figures. From this alone, the figures do not tell your story until Figure 6. In fact, Figure 1 is somewhat irrelevant to the stated hypothesis in the introduction. The order of figures could therefore be revised to Figure 2, 3, 6, 7, 8, 1,4,5. With the text also reordered.

We thank the reviewer for this suggestion. We have We have now added a new figure in the discussion that is summarizing our main findings regarding early emergence of T versus O_2. We also formulated the abstract more concisely. We feel that these changes will convey our main message to all readers.

5. Additional figure to better visualize regions of ToE(O2) < ToE(T) These regions should be better vi-

sualized in the paper, potentially in a synthesis map showing locations of ToE(O2)<ToE(T) robust across models (could use the 7/9 threshold for example) and/or a figure/bargraph for which regions are aggregated and the average relTOE(O2) vs relTOE(T) is ploted for each model. So the figure would have regions on the X axis and ToE(O2)-ToE(T) on the Y axis. From there the reader could see the spread in the models as well as pick out regions where it is most likely for ToE(O2) to be significantly shorter than ToE(T).
Thank you for the suggestion. We have added a "binary" type map (Fig. 9) that shows the regions in blue/brown where seven out of nine models simulate emergence of oxygen/temperature before temperature/oxygen. We have added following text to the Discussion section:

On average across the nine models, an area covering 35±11 % (Fig. 6) of the global thermocline shows emergence in O_2 change before temperature change. Yet, the exact locations of relatively early emergence of O_2 differ across models. Hence, the regions where at least seven out of the nine models show consistently an earlier emergence of O_2 than T is smaller and amounts to 17 % of the global thermocline area. As shown in Figure 9 (blue areas) the O_2 signal emerges consistently in at least seven models before the T signal in parts of the Pacific subtropical gyres, the Southern Ocean and the southeast Indian Ocean.

1.3 Additional Technical/Editorial Corrections

Page 1

Line 3: remove word "as" .
Done and acknowledged.

Line 12-14: Simplify sentence with: To normalize across disparate trends and variability of the CMIP5 ensemble, we compute the local ToE relative to the global mean ToE within each model.
Modified as suggested

Page 2

Line 4-5: correct to remove "and". . . adversely affect marine organisms, ecosystems, and the services they provide.
Modified as suggested

Line 6: remove "a": . . .experienced significant warming
Modified as suggested

Line 9: Make "scale" plural. . .on regional to local scales
Modified as suggested

Line 13: Slightly awkward jump to discussion of ESMs.
The sentence has been moved. See comment on page 4, line 25)

Line 17: Remove first phrase, "Concomitant with ocean warming" and just start with "Observation-based studies. . ." The mechanisms of observed oxygen decline is subsequently explained, so it needn't be partially explained initially.
Modified as suggested

Lines 19-21: Break into 2 sentences and add context. "In subsurface waters, oxygen concentration is also affected by ventilation, remineralization of organic matter and air-sea disequilibrium. In the contemporary ocean, oxygen decreases are mostly dominated by a reduction in ventilation and increased consumption of oxygen during remineralization (references)."
We split the sentence as suggested. However, we find it confusing for a general reader to say that changes in air-sea disequilibrium affect O_2 in subsurface waters. Further, most model studies also suggest that changes in ventilation are, at least on larger scales as considered in this study, more important than changes in

export production of organic material. The text reads now:
Increased surface temperature reduces oxygen solubility, limiting atmospheric oxygen dissolution into the upper ocean. In subsurface waters, oxygen concentration is also affected by changes in ventilation and the remineralisation of organic matter. In the contemporary ocean, oxygen decreases in the interior are mostly dominated by a reduction in ventilation with a smaller role for changes related to the production of organic matter, O_2 solubility, and air-sea equilibration of O_2 in surface waters (Bopp et al., 2002; Plattner et al., 2002; Bopp et al., 2017; Tjiputra et al., 2018; Hameau et al., 2019).

Line 21: change to. . . "The largest oxygen declines are located"
Modified as suggested

Line 23: define "late" industrial period.
The sentence reads now:
The largest oxygen declines are located in the Pacific Ocean (equator and northern hemisphere) and the Southern Ocean. However, observations are relatively sparse and only start in the second half of the 20th century.

Line 23-24: change to. . . "Therefore, it is challenging to distinguish human-caused trends from natural variations in the observational record of ocean O2."
Modified as suggested

Line 24: remove "also". . . Modelling studies agree on the sign of . . .
Modified as suggested

Line 24-26 could be combined instead with 13-16 and form their own paragraph discussing the current state of modeling O2 trends.
The description of models studies for temperature and oxygen are now combined in one paragraph.
Global climate models, such as the Earth system models that participated in phase 5 of the Coupled Model Intercomparison Project (CMIP5) reproduce the long-term trend in global ocean heat content over the last 50 years when uncertainties of observation-based estimate and internally generated natural variability are taken into account (Frölicher and Paynter, 2015; Cheng et al., 2019). Modelling studies agree on the sign of oceanic O_2 changes, but likely underestimate the magnitude of loss (Cocco et al., 2013; Bopp et al., 2013; Oschlies et al., 2017). In particular in the tropical regions, models are not able to reproduce observed O_2 decrease in equatorial low-oxygen zones (Stramma et al., 2008; Cocco et al., 2013; Cabré et al., 2015).

Line 32: include reference, " (e.g. Deutch et. al., 2015)"
Reference included as suggested.

Page 3

Line 11-12: rewrite. . ."is critical to understanding contemporary O2 and temperature changes."
Modified as suggested

Line 17-29: rewrite "One study, (Hameau et al., 2019), uses a single model [which model?], to investigate ToE of temperature in the thermocline, finding that anthropogenic ocean warming emerges much earlier than the O2 signal in low and midlatitude regions. Delayed emergence of changes in O2 is due to the opposing effects of decreases in O2 solubility and O2 consumption. In the high latitudes and the Pacific subtropical gyres, deoxygenation emerges before ocean warming in [name model]. This occurs because decrease in oxygen solubility are reinforced by increased O2 consumption, leading to strong O2 depletion However, it is unknown if this single-model result is robust across a suite of different Earth system model simulations. Here, we conduct a multi-model study to more broadly test the hypothesis that anthropogenic deoxygenation in the thermocline emerges prior to anthropogenic warming. Since the primary objective is

to test the consistency across models of the order of emergence (deoxygenation prior to warming) within a single model, we introduce and use a relative ToE to conduct the intercomparison, rather than the absolute year of ToE. We define relative ToE as. . . ”
Modified as suggested

Page 5

Lines 1-10: Over what time frame is S estimated? The texts read “Enting, 1987”... Ending in 1987? At first I thought it was a typo but now I see it is a reference. To clarify, maybe state that S is a time series that extends from pre-industrial to 2100 (If I am understanding correctly). Secondly, with what “fitting?” –linear fit, polynomial, etc.? Perhaps the word “fit” should be excluded, if it is in fact meant to just say that the time-series is low-pass filtered.
According to your comment page 2, line 31, the corresponding paragraph has been improved and clarified. Moreover, it is now stated that the period of the signal S extends the entire period of the simulation (1860 – 2099).

Line 20/21: change request to require and is to are: “We therefore require that ToE values are defined for at least seven out of nine models to compute the multi-model statistics (median and spread).“
Modified as suggested

Line 24: New section, entitled “Separating mechanisms of oxygen change” or something like this.
Modified as suggested

Line 24-25: Potentially rewrite: To diagnose processes driving the simulated changes in ocean O2, the direct thermal/solubility component of change (O2sat) can be isolated from the total O2 change. The residual, Apparent Oxygen Utilization (AOU), represents the summation of all non-thermal changes, including those resulting from changes in ventilation and remineralization.
Modified as suggested

Page 9

The early emergence of T in the Eastern Equatorial pacific should be discussed, and if possible, mechanistically explained. At first glance, it is surprising that an area of such interannual variability at the surface can have relatively little variability at intermediate depths.
The strong variability in the Eastern Pacific upwelling system does not reach the 200 – 600m layer. As shown in Fig. 1, internal variability is confined between the surface and 200 m. Therefore, the temperature increase at 200 – 600 m depth range are detectable relatively rapidly.
The region in the Eastern Tropical Pacific is now explicitly discussed in the main text.
The combination of a strong signal and small variability results in early detection of the changes. This is the case in the Southern Ocean at 45° S (in the Atlantic and Indian regions; Fig. 1a), where the anthropogenic warming is strong (up to 4 ° C; Fig. 4b) but the internal variability is relatively small (0.1 ° C to 0.3 ° C; Fig. 4a). However, early emergence of anthropogenic changes can also occur when the signal is relatively small, if the variability is even smaller. This is the case in the tropical oceans such as in the Arabian Sea and the equatorial Atlantic, where water masses warm modestly (up to 1.5 ° C), but vary naturally between 0.1 ° C and 0.2 ° C only. It is also the case in the eastern equatorial Pacific, where the early emergence arise from the very weak internal variability in the thermocline, although, the temperature increase (∼0.80 ° C) is also relatively weak. In this region, the substantial variability in O_2 and T is largely confined to the top 200 m. No emergence by the end of the 21st century, such as simulated in the subtropical gyres of the Indian and Pacific oceans, results from a relatively weak signal combined with a relatively strong variability in these regions.

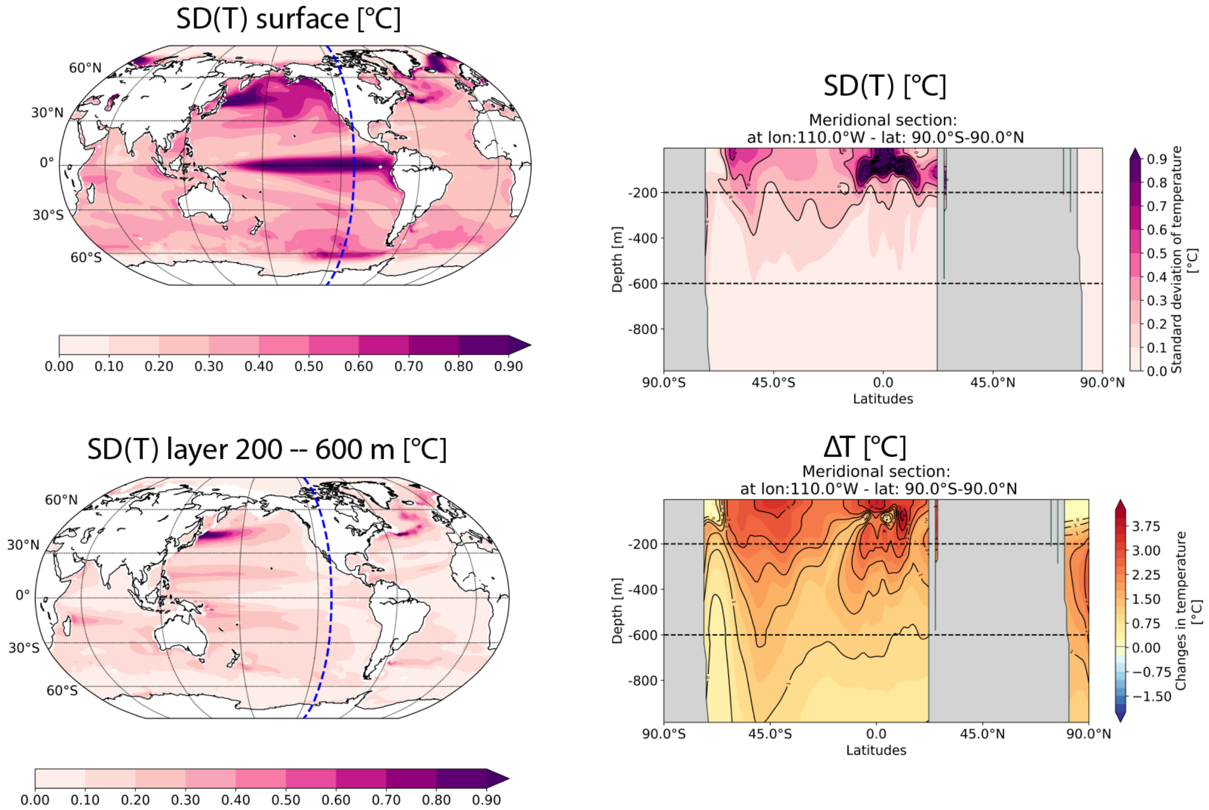

Figure 1: Internal temperature variability simulated by the CESM model under preindustrial conditions (top left) at the surface, (bottom left) for the averaged layer 200 – 600m, (top right) along the meridional 110 ° W (illustrated by the blue dashed line in the left panels). The bottom right panel shows the temperature increase by the end of the 21st century along the same longitude.

Line 20: remove word 'typically'
Modified as suggested

Page 10

Line 11: remove word 'surprisingly'
Modified as suggested

Line 12: what is the confidence interval referenced? Is this references figure 1e, the multi-model spread? If the medium relTOE is 10 and the spread is also 10, then what does "confidence interval" mean in this context? And in that case, a confidence interval value of 1 is not "high". Usually confidence interval terminology is used to describe the probability of rejecting of a null hypothesis. The implicit hypothesis of the statement would be that in the CMIP5 ensemble, this region's relTOE(O2) does not significantly differ from the global average TOE(O2). Potentially a better way to convey this is to state that the multi-model ensemble agrees that the North Pacific represents a region for which emergence timescales are representative of the globally-averaged emergence timescales?
We agree with your comment on the use of "confidence interval". The sentence reads now:
However, $ToE_{rel}(O_2)$ in this region is within ∼10 years (Fig. 1d), with a relatively low spread (±10 years) compared to $ToE_{abs}(O_2)$ (±30 years).

Line 15-16: remove sentence that begins "Using". If not, then replace with ... "allows for more equitable comparison of projections of CESM with those of lower sensitivity / higher variability models (Figs.3 and 2)." [because we cannot define high-sensitivity as a 'bias' since we do not know the sensitivity of the Earth's climate yet]
Modified as suggested and word biases replaced with model-model differences
The ToE_{rel} allows the comparison of ToE resulting from CESM output with the results from the 8 models in spite of these model-model differences (Figs. 2 and 3).

Line 19-20: remove first 2 sentences, and begin with: "Broadly, temperature changes are. . ."
The two first sentences have been removed. The beginning of the paragraph reads now:
In general, temperature changes are detectable before O_2 changes in around 64±11 % of the thermocline

Line 27: "Model disagreement on the relative timing of TOE(o2) and TOE(T) is highest in the Atlantic. . ." A figure that bins the regions and computes regional SD across the models for the quantity ToE(T)-ToE(O2) would be useful. Then you could make statements about the probabilities of certain regions following your hypothesis. For example, the Atlantic certainty does not, but others may. It would give some better spatial insight into where the distributions of Figure 8 are occurring.
As suggested, we have added a "binary" type map (Fig. 9) that shows the regions where at least 7 models project earlier emergence in oxygen (blue) or temperature (brown)

Line 27-28: The discussion of the subtropics sounds like an extension of the discussion of the Atlantic. It should be made clear that you are now talking more broadly about the subtropics globally (I think) or still just discussing the Atlantic.
The sentence has been updated and reads now:
Model results for the Atlantic subtropical gyres are mixed. Some models suggest O_2 changes to be detectable earlier than T changes (HadGEM2 and the IPSL family), whereas in other models the O_2 signal does not even emerge.

Page 11

Line 1: remove "it is striking that" and later correct to "typically show a decrease"
Modified as suggested

Line 15: replace "noticeable" with "notable"
Modified as suggested

Lines 1-15 are understandable but could benefit from some heavier rewording and condensing."
The paragraph has been rewritten and reads now:
Regions with early emergence of anthropogenic O_2 compared to T show typically a decrease in [-AOU] (Fig. 6 versus Fig. 7), whereas regions with early emergence of T compared to O_2 show typically an increase in [-AOU]. For example, [-AOU] is decreasing in 77±8 % of the areas with early emergence of O_2, while only 22±8 % of these regions show an increase in [-AOU] (Fig. 8; blue). In most regions where T is emerging before O_2 (Fig. 8; green), [-AOU] is increasing (62±12 %). A decreasing trend in [-AOU] is indicative of a reduced ventilation induced by upper ocean warming and increased stratification (e.g. Capotondi et al., 2012). A more sluggish ventilation slows the supply of O_2 from the surface to the ocean interior. Consequently, thermocline [O_2] and [-AOU] are both decreasing. This leads to a strong and thus early detectable anthropogenic deoxygenation. In addition, a more sluggish ventilation slows the penetration of the anthropogenic warming signal from the surface to the interior, and similarly the penetration of the thermally driven O_2 signal ([$O_{2,sol}$]). The detection of the temperature changes is thus delayed compared to AOU and O_2. There are some exceptions to this relationship between [-AOU] and the earlier emergence of O_2 than T. For example, O_2 change emerges before warming in the GFDL model around 30° S and 120° W, although [-AOU] is increasing in this region. However, warming is emerging very late as the GFDL models simulate weak warming and even some cooling (Fig. S4) in this part of the thermocline. Thus, in this special case, the early emergence of O_2 relative to T is due to the absence of large warming in a region with notable temperature internal variability.

Line 17: "leading to relatively smaller changes in [O2]"
Modified as suggested

Line 21: rewrite to be specific: "anthropogenic change in temperature is detectable earlier than anthropogenic change in O2 most of the global ocean"
Modified as suggested

Page 12

Line 5: "trend" should actually be "change"
The sentence reads now:
Using ToE as a metric allows for the assessment of anthropogenic changes by comparing the magnitude of the human-induced changes with the magnitude of internal variability.

Line 5-6: Rewrite "Both the magnitude of anthropogenic change and internal variability are model dependent, rendering the absolute year of ToE (strongly*) model-dependent*. Evaluating differences in absolute year of ToE, however, can obscure important model agreement upon the spatial patterns and progression of emergence within a multi-variable framework. We therefore introduce a new metric. . ."
Modified as suggested

*However, I note a recent NCC paper (Nijsse et. al., 2019) that argues that there is a correlation (and mechanistic relation) between climate sensitivity and decadal variability. Presumably if there is compensation between the two, then ToE could be relatively more robust across models than S or N.
Thank you for pointing to this publication. We have added to following sentence in the discussion.
Nijsse et al. (2019) suggest that the magnitude of simulated decadal variability and climate sensitivity might be correlated. They suggest that models with a high climate sensitivity tend to simulate a high decadal variability. This may imply a compensation between the simulated signal and noise on the decadal scale.

Line 9: “within a model” is misleading – because this sounds as if the models are normalized by some external factor, say by taking “temperature of emergence” (the amount of global warming required for a signal in any chosen variable to emerge). I think you could add a qualifier, and way “signal emerges relatively early or late relative to the signals global average in the given model.”

We changed ”within a model” by ”for a given model”. The proposed sentence tends to repeat the definition of relative ToE. But here, we want to highlight the general benefit of using it. The sentence reads now:

Absolute years of emergence are thus not considered by this metric and it only illustrates whether a signal emerges relatively early or late for a given model.

Line 12: Replace short-coming with limitation or caveat

Modified as suggested

Line 14: Remove “Perhaps not surprising,”

Modified as suggested

Page 13

Line 16: Is it known why subtropical gyres in the Atlantic and the eastern equatorial Pacific have elevated noise? Please include.

The Eastern Equatorial Pacific shows a low variability in the thermocline. We corrected this wrong statement. We did not investigate in detail why there is high variability in the subtropical gyres of the Atlantic. We suspect that this is related to a large variability in the winds which force these gyres. The text now reads:

Exceptions are the subtropical gyres in the Atlantic, where it takes approximately two additional decades to detect the temperature changes, mainly because of the relatively large internal variability there.

Line 23: Add word “spatial” before pattern.

Modified as suggested

Lines 25-27: Rewrite: Even though the internal natural variability is low in the tropical regions, the O2 signal does not emerge from the noise, so is the signal.”

The proposed modification seems confusing to us. However, we revised the sentence for clarity:

Although internal natural variability is low in the tropical regions, the O_2 signal does not emerge by 2100. This is because the projected changes are also small.

Line 30: add “relatively” before “strong”.

Modified as suggested

Line 35: change “highly” to “significantly”

Below the updated sentence:

For example, the transient climate response of the individual models and therefore the ocean heat uptake, thermocline warming and deoxygenation can be substantially different (Bopp et al., 2013).

Page 14

Line 1: remove “largely”, if needed replace with “significantly”

Below the updated sentence:

In addition, the simulated internal variability considerably differs across models (e.g. Resplandy et al., 2015; Frölicher et al., 2016).

Line 2-3: Rewrite: “ToEs computed from CESM1.0 projections, for example, differ by many decades in absolute terms from other CMIP5 models, mostly due to a very weak internal natural variability.”

The sentence reads now: ToEs computed from CESM1.0 projections, for example, differ by many decades in absolute values from other CMIP5 models, mostly due to a very weak internal variability.

Line 3-4: Rewrite, something like "To extract valuable insights as to the relative spatial and temporal features of emergence across models and variables, we introduced. . ." However this notion is redundant with previous page so it could be excluded altogether or merged.
The sentence has been modified as suggested

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-339, 2019.

2 Anonymous Referee #2

2.1 General

This paper presents an analysis of the local time of emergence of an anthropogenic temperature and oxygen changes in the global oceans. In a recent study, the same authors (with the exception of Frolicher) used a single model to investigate the same topic. This paper went a step further by using an ensemble of Earth System models (ESMs) included in CMIP5. The idea of using a single metric, the time of emergence (ToE) to determine the pint in time when the anthropogenic signal becomes larger than natural variability, is simple and appealing. The authors applied ToE to temperature and O2. Because ToE varies a lot among ESMs, they introduced the concept of relative ToErel by subtracting the global area-averaged ToE from ToE at each model grid point.

Nevertheless, the results on ToE and ToErel would likely be sensitive to the threshold value (2) selected in Equation (1) as well as the way how S (anthropogenic signal) and N (internal natural variations or background noise are calculated).
We agree with the reviewer that methodological choices affect results for ToE in detail as shown earlier (Rodgers et al., 2015; Frölicher et al., 2016; Hameau et al., 2019). Our main conclusion - the signal of anthropogenic O_2 is emerging earlier than the anthropogenic warming signal in many ocean regions - was found to be robust regarding specific methodological choices applied to results from a single model (CESM) (Hameau et al. (2019)). Here, we find that our main conclusion is also robust across the different ESMs and thus for a range of different noise levels. The following text has been added in the introduction to clarify this:

The following paragraph has been added to the discussion:
Although the generic definition of ToE is under consensus, the methodologies applied to estimate ToE differ in the published literature (IPCC (2019)) as mentioned in the introduction. Depending on the spatial and temporal scale of a given variable, the threshold for which emergence is defined and the reference period applied, the absolute value of ToE can differ. In addition, the ToE also depends on the definition of the background variability, here acting as noise (Hameau et al., 2019). Estimating the background noise as the standard deviation (SD) of the internal chaotic variability from the control simulation (Frölicher et al., 2016), or as the SD of the variability from the industrial period (after removing anthropogenic trends; Keller et al., 2015; Henson et al., 2016) result in earlier ToE for both O_2 and T as when estimating the noise from the total (internal and externally-forced) natural variability over the last millennium. Yet, the finding that anthropogenic O_2 change emerges before anthropogenic warming in large ocean regions is robust across investigated choices. The anthropogenic signal is frequently computed as a linear trend over a few decades (Rodgers et al., 2015; Henson et al., 2017; Tjiputra et al., 2018). However, the resulting slope depends on the time window used to calculate the linear trend. Therefore, Hameau et al. (2019) use a low-pass filtered output to estimate the signal.

Although similar calculations were reported in previous papers, the authors need to describe how S and N were calculated and examine the sensitivity or robustness of the model results. There are also questions why the same methodology can be used for different regions of the global oceans? Can you use the same methodology for the tropics and mid-latitudes?
Thank you for pointing out that we do not provide sufficient explanations for those readers not familiar with the ToE concept. The method section has been extended and an additional figure has been added to the appendix (Fig. A1). The sensitivity of results to methodological choices is explored and quantified in our previous publication (Hameau et al., 2019) as well as in IPCC (2019) - Chap5 - Box5.1 entitled: "Time for Emergence and Exposure to Climate Hazards"; please see our answer to the previous comment (page 11, line 14). In this manuscript, the robustness of relative and absolute ToE estimations across the model range is discussed in Sect. 3.1 and illustrated by Fig. 1 (middle and lower panels).

Yes, the same ToE method can be applied to any time series and to different ocean regions as is it computed in each grid point. However, the definition of ToE could differ among regions because of specific behaviour of the considered variability with depth and time for example. Nevertheless, our objective here is to investigate a specific methodology across models and thus we leave this latter point for future studies.

The following technical explanation is now given in the method section:
We define the absolute ToE as the first year when the anthropogenic signal S becomes equal or larger than twice the noise of internal variability N (Eq. 1; following Hameau et al., 2019; Fig. A1). The threshold is set to two in order to distinguish the signal from the noise at 95 % confidence level. Annual O_2 and T data are first averaged over the thermocline (200 – 600 m) at each grid point of the horizontal grid and local S and N are computed from these depth-averaged values for each model, variable and (horizontal) grid-point. Annual anomalies are calculated relative to the preindustrial period (1860 – 1959)....
The background noise, N, is computed as one standard deviation (SD) of O_2 and of T from the annual preindustrial control output. Although, the length of the control simulations differ between models, the entire duration of the control simulation is considered for each model to estimate the background noise....
The annual output of the forced, transient simulation (1860 – 2099) is smoothed by a low pass spline filter (Enting, 1987) to estimate S for each (horizontal) grid point in the thermocline. The cut-off period of the spline is set to 80 years to remove decadal to multi-decadal variations (e.g. associated with internal variability). The signal S is then the value of the spline at each point in time.

The simple concept of ToE or ToErel also has its drawback, making it hard to interpret the model results. The authors provided little or no interpretations of the major models results (Figures 1-7). After reading the manuscript, I was left with an impression that it was a purely numerical exercise.
We disagree with the reviewer regarding the importance of the ToE concept and of this study, and refer to the general comments of reviewer #1 (see page 1, line 11) and #3 (see page 17, line 7). We agree that it is difficult to interpret ToE and ToE_{rel} in a mechanistic way even within a single model. ToE reflects the ratio between the magnitude of change and the magnitude of variability - two quantities that are, at least partly, influenced by different processes. Providing mechanistic explanations in this multi-model study is beyond the scope of this work. We note that we quantify both magnitude of anthropogenic change and variability individually and that we distinguish between solubility-driven and remineralization-driven O_2 changes. We show that anthropogenic deoxygenation emerges before anthropogenic warming in about a third of the global thermocline. This, together with the finding that human-caused O_2 changes leave the bounds of natural variability in many regions is likely relevant to assess the risks associated with anthropogenic greenhouse gas emissions. We conclude that monitoring biogeochemical variables such as O_2 will help to better identify environmental risks.

2.2 Some detailed comments:

(1)

First paragraph in Section 3.1.1 on page 7. Why does ToErel (T) show early emergence in low latitudes and between 30° and 60° S and late emergence in the western tropical Pacific?
The driver of early/late emergence are discussed in details in Sect 3.3. The corresponding sentences have been updated to clarify this:
The combination of a strong signal and small variability results in early detection of the changes. This is the case in the Southern Ocean at 45° S (in the Atlantic and Indian regions; Fig. 1a), where the anthropogenic warming is strong (up to 4 ° C; Fig. 4b) but the internal variability is relatively small (0.1 ° C to 0.3 ° C; Fig. 4a)....
...However, early emergence of anthropogenic changes can also occur when the signal is relatively small, if the variability is even smaller. This is the case in the tropical oceans such as in the Arabian Sea, the equatorial Atlantic and the western equatorial Pacific, where water masses warm modestly (up to 1.5 ° C), but vary naturally between 0.1 ° C and 0.2 ° C only....

In order to guide the reader, the beginning of the result section reads now:
We start by discussing the multi-model median and spread of relative ToE estimates for potential temperature (Fig. 1a, b) and dissolved oxygen (Fig. 1d, e) changes in the thermocline (200 – 600m). An analysis of the roles of internal variability and anthropogenic change ToE and why anthropogenic change is detectable early or late is presented in Sect 3.3.

Before jumping to ToErel, tell us the global mean ToE first.
The following text is added in section 3.1.1. and 3.1.2:
ToE values for individual horizontal grid cells are globally averaged to obtain an area-weighted global mean ToE for the thermocline and each model. These global mean values range between year 1963 and year 2033 for the nine models (see subtitles in Fig. 2). These global differences between models are removed by definition in ToE $_{\text{rel}}$. The multi-model median of ToE $_{\text{rel}}$(T) shows early emergence in low latitudes and between 30^o S and 60^o S, ...

As for temperature, a large range in absolute ToE is found with globally-averaged ToE(O_2) ranging between the year 1991 and 2046 for the nine models (see subtitles in Fig. 3).

Why is there no emergence in the subtropical gyres of the Indian and the Pacific oceans?
No emergence of warming until the end of the 21st century result from the combination of relatively low anthropogenic warming and relatively strong internal temperature variability. As discussed in Sect 3.3, high temperature variability is simulated in the subtropical thermocline. Subtropical gyres are regions of vertical subduction and thus active wind ventilation (Pedlosky, 1996). This induces relatively large variability in these regions and may explain late emergence. Differences in vertical profiles of temperature and salinity and in effective thermocline depths among the different basins may explain the specific behavior of the Pacific and Indian oceans as compared to the Atlantic in our diagnostics. We added the following sentence:
No emergence by the end of the 21st century, such as simulated in the subtropical gyres of the Indian and Pacific oceans, results from a relatively weak signal combined with a relatively strong variability in these regions.

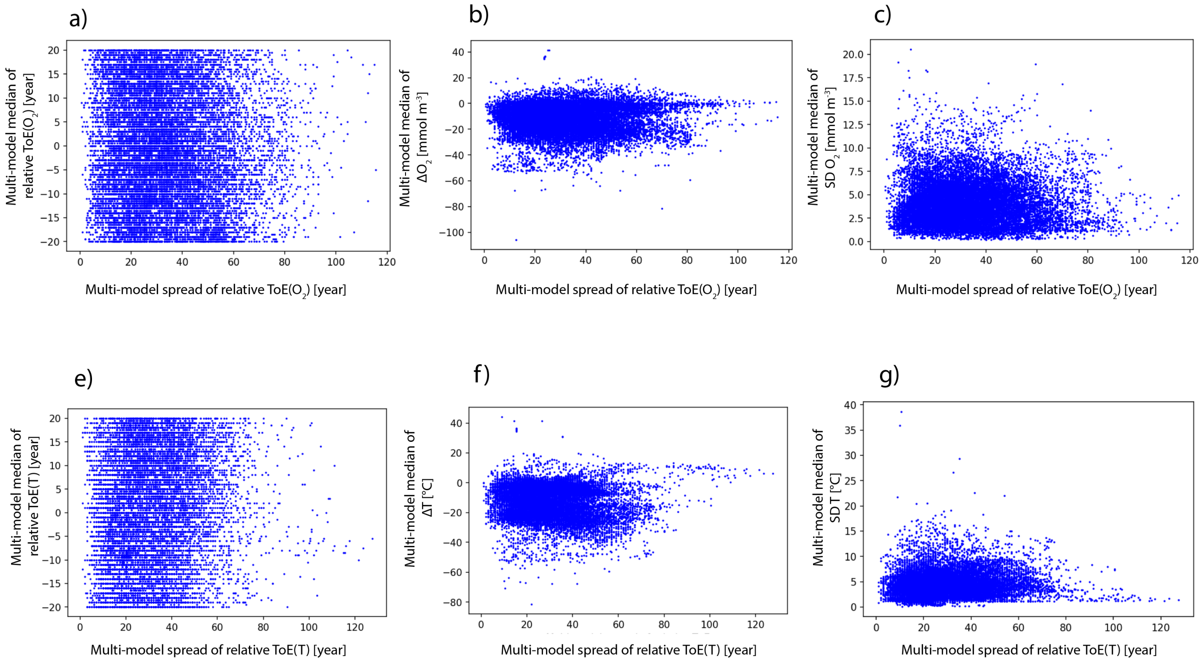

Figure 2: Scatter plots between the multi-model spread in relative (top panel) ToE(O_2) and (lower panel) ToE(T) and the multi-model median of (a) relative ToE, (b) anthropogenic change (1860-1959 – 2070-2099) and (c) internal chaotic variability. Each point represents a grid cell in the thermocline.

(2)

Second paragraph in Section 3.1.1 on page 7. Why is the spread among ESMs small in some regions but large in other regions?

This is an interesting question. We are not able to answer this question. We do not find a clear relationship between the multi-model spread in ToE and the multi-model median in ToE, or the multi-model median of the anthropogenic signal, or the multi-model median of the variability for both T and O_2 as evidenced by the scatter plots in Fig. 2 of this reply.

(3)

Third paragraph. Can you use individual model projections to obtain a quantitative estimate on the robust/uncertainty in estimating the mean ToE from the ESM ensemble?
Yes, please see Fig. 1 and related discussion in the manuscript. As usual in multi-model analyses, individual variables, here ToE and ToE_{rel}, are computed first for each individual model. Then, additional metrics are computed including the multi-model median and the multi-model spread. The interquartile ranges for ToE and ToE_{rel} for T and O_2 are shown in Fig. 1b and e. The interquartile range represent a measure of model uncertainty. The following text is added to the method section to clarify this:
S, N, ToE and ToE_{rel} are first computed from the annual output for each model and at each (horizontal) grid cell. Then, multi-model median and spread (interquartile range) of the multi-model estimations are computed from the model ensemble. The median represents a "best" estimate and the interquartile range a measure of model uncertainty. Uniform weights are applied to each model configuration to compute these statistics. ..

(4)

First paragraph in Section 3.1.2 on page 7. Why is ToErel (O2) relative homogeneous?
We thank the reviewer for pointing out this misleading wording. Text replaced by:
In contrast to ToE $_{rel}$(T), most of the thermocline shows no emergence of the anthropogenic O_2 change by the end of the 21st century (Fig. 1d). In the remaining regions ToE $_{rel}$(O_2) varies by about ±40 years.

(5)

Line 3 in the third paragraph in Section 3.3 on page 9. Large warming of more than ∼4 Co is projected in the northern North Atlantic and round the subantarctic water. Can this projection be trusted?
Text modified to read:
Large warming of more than 4.0±0.7 ° C is projected in the northern North Atlantic and around the subantarctic waters in the Indian and Atlantic Oceans (Fig. 4b and Fig. S4). We note that these projections are also characterised with the largest inter-model spread (±1.5 ° C; Fig. 4d and Fig. S4) and uncertainties in these regional warming projections are large.

Many ESMs showed biases when simulating historical periods. Were these biases removed before the ToE analysis was applied?
No, we did not apply any bias corrections. Removing the bias in the climatological mean state would not affect the magnitude of the anthropogenic change (S) nor the variability (N). Therefore, all our results for ToE, S, and N would be the same. Another potential approach might be to apply scaling factors to modelled variability and anthropogenic change. However, the observational information on industrial period change and on variability in T and O_2 in the thermocline is limited on the grid cell scale. This limitation regarding observational data may also apply to other more sophisticated correction methods.

(6)

Section 3.4. It was a good idea to check changes in AOU in order to better distinguish the O2 and temperature signals. Can you check if ventilation of the thermocline indeed decreases in regions with decrease in [-AOU] rather than relying on cited references?
The ideal age tracer was only provided for only one model family in the CMIP5 dataset (only GFDL-ESM2G and GFDL-ESM2M provide this tracer). This is stated in the manuscript and the corresponding text in section 2 reads: "Output of an ideal age tracer is not available for most models".

The authors were on the right track here to get at the mechanisms but did not go far enough. Similar mechanistic analysis should be done to explain the other results.
In the context of a multi-model study, it is challenging and difficult to develop robust mechanistic analyses for each single model as some of them do not provide all the variables necessary (for example in this case, the ideal age tracer).

(7)

Second paragraph in Section 4 (page 12). Most ESMs do not have fine resolutions to simulate the oxygen minimum zones (OMZs) well. As the authors indicated, the ESMs diverge in their projections for the physical and biogeochemical changes in OMZs. Some models even showed an opposite trend to the observations in recent decades.

This raised an important concern about the merit of even using such models to investigate ToE because they will lead to misleading results. Why didn't you remove those ESMs that did not capture the past changes?

We thank the reviewer for bringing up this issue. Nowadays, there is no generally accepted metric to weight the individual models regarding the simulated oxygen concentrations in the thermocline. This is due to the relative lack of observation. Cabré et al. (2015) show that all CMIP5 models present biases in O_2 concentration and in the extent of the O_2 minimum zone (e.g. see Fig.1 from Cabré et al., 2015). This is related to limited process understanding and coarse model resolution.

Moreover, Knutti and Sedláček (2013) underline the relatively high uncertainties in CMIP5 projections, and suggest the most reliable climate projection is given by a multi-model averaging. The limited number of models used in this study (only 4 different model family) is a caveat of our study that we discuss now in Sect 4 paragraph 2, which reads:

This multi-model study relies on results from only four different model families (GFDL-ESM, HadGEM2-CC, IPSL, MPI-ESM and CESM) applied in nine model configurations. All model configurations available from CMIP5 that provide 3-dimensional fields for O_2 and T for the control, historical and future-RCP8.5 scenario simulations have been incorporate into the analysis. Nevertheless, using a larger model ensemble would increase confidence in our results (Knutti and Sedláček, 2013).

3 Anonymous Referee #3

This paper focus on detecting the anthropogenic signals of both thermocline temperature and o2 (i.e. physical and biogeochemical properties) from a suite of CMIP5 models under the future projections. The study is (to some extent) based on the previous study of Hameau et al., (2019) extending to multi-model perspectives detecting the ToE of the thermocline temperature and o2 to assess the robustness of the results. The authors also introduce the relative ToE concept, results in reducing the inter-model spread compared to the traditional ToE and allows them to conduct more robust comparison. In general I think it is important to aim on understanding changes in both physical and biogeochemical tracers together to better understand the resulting changes in marine ecosystems and combining multi-tracers could provide additional insights. I think the topic and contents of this study fits into the scope of the special issue in Biogeosciences. However, I have comments on the current manuscript.

We thank the referee for the positive and constructive review.

3.1 General comments

1. Abstract:

I suggest to include some discussion (possibly in the section 4) on following up the statement "... the detection of anthropogenic impacts become more likely when using multi-tracer observations" in the abstract. Combination of two tracers will definitely provide additional information on further implications from both physical and biogeochemical perspectives. Despite the fact that two of these properties emerge on different timescales, what would authors expect to see from (or should be aware of for monitoring) future multi-tracer observations?

Multi-tracer observations would inform on the potential earlier impacts and consequences of the changing climate on marine ecosystems. Moreover, an important part of the biogeochemical processes simulated by climate models are still based on empirical values. Extended multi-tracer observations would contribute to better constrain climate models. The associated paragraph in the discussion has been extended and reads now:

Published studies addressing the detection of anthropogenic ocean warming focus on temperature at sea surface. To our knowledge, only a single study Hameau et al. (2019) using output from a single model is assessing ToE(T) in the thermocline. Yet, the thermocline is habitat for many fish and other species. Warming in combination with changes in other stressors, such as deoxygenation, ocean acidification and hypocapnia, may reduce the habitat suitability of marine life in a future climate (e.g. Deutsch et al., 2015; Gattuso et al., 2015; Breitburg et al., 2018; Cheung et al., 2018). Multi-tracer analyses contribute to a better understanding of the potential impact on marine ecosystems in a changing ocean.

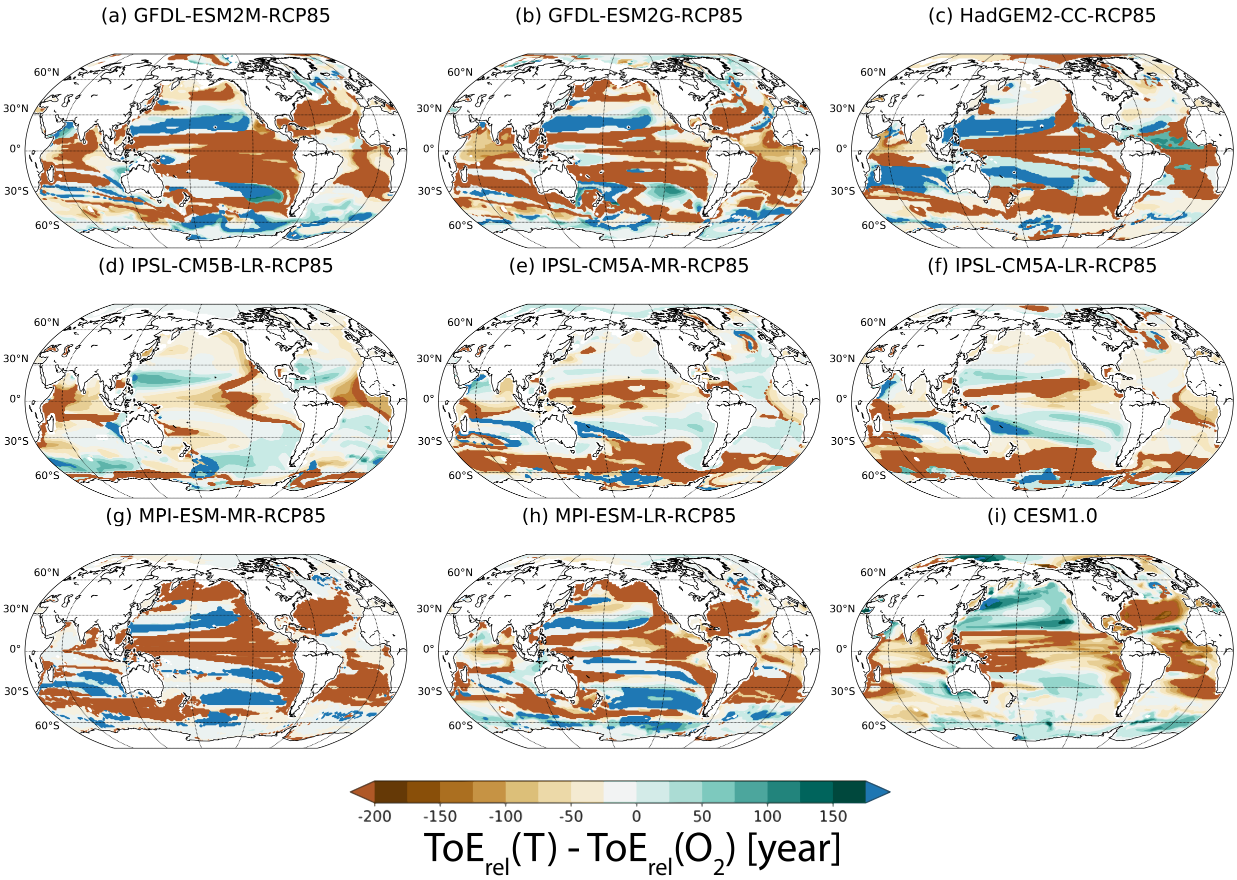

Figure 3: Difference in ToE_{rel} of T and O_2 for nine CMIP5 model configurations. The saturated colours indicate that one of the variables has not emerged by 2099.

2.

I agree that the ToE comparison among the models are not straightforward and the advantage of relative ToE is to "reduce the inclusion model uncertainty in the metric" (as stated in section 3.2 in details). However, in section 3.4 (results on ToE comparison between the two variables), the author calculate the difference between the "absolute ToE" for each models. I thought this will still include more model bias (from global ToE, which is subtracted in relative ToE). Since the author introduced an improved ToE metric, it might be better to come up with a metric comparing "relative ToE" from the two variables. This might not be straightforward but can you think of further metrics based on comparing two relative ToEs? If authors think the this will not make a difference, please explain in more details.

The goal of the metric relative ToE is to allow for a better comparison of ToE across models by removing potential biases (such as too high/low sensitivity to external forcing). We expect that these biases affect ToE(T) and ToE(O_2) in a similar direction in an individual model. Therefore, the effect of biases may be reduced for the difference ΔToE=ToE(T)-ToE(O_2), which is evaluated for each model and location individually. Further, and perhaps more important, the interpretation of the relative metric $\Delta ToE_{rel}=ToE_{rel}(T)-ToE_{rel}(O_2)$ is not very clear. ΔToE indicates whether the O_2 signal or the T signal emerges first from the background variability, but this information is not provided anymore by ΔToE_{rel}. We evaluated ΔToE_{rel} for each model for comparison with ΔToE. The results for ΔToE (Fig. 6 of main manuscript) and ΔToE_{rel} (Fig. 3 of this reply) are similar for individual models. In addition, the multi-model median of ΔToE=ToE(T)-ToE(O_2) and of ΔToE_{rel} are also very similar (not shown). Therefore, we do not discuss ΔToE_{rel} in the manuscript.

3.

Regarding to the terminology used in the manuscript, the "internal natural variability" and "natural variability" are mixed used in the manuscript. It is not always clear what exactly the terminology defines in this context. From what I understand, the internal natural variability meant here is the variability stemming from internal climate system (specific example will be ENSO, PDO etc.) and the natural variability includes the natural "external" forcing (such as volcanic eruptions) correct? It was mixture of terminology (particularly in the discussion, which I saw the two terms were inter-changeable in some sentences) and I suggest to clearly define the terminology in the beginning and use the term in a consistent manner.

The reviewer points out an important potential source of confusion. The term "internal natural variability" has been replaced by the term "internal variability".

4.

I suggest the authors to explain some of the statistics in more details in the method. The noise (N) is defined as the standard deviations from the pre-industrial control simulations from each models but did the author defined standard deviations based on the temporal standard deviation using full control simulation period? This may not be a huge difference but I assume periods differs among the model. Also, the CESM1 in this study uses the last millennium spinup but is this different from the preindustrial control simulations (I assumed yes)? I think introducing a schematic based on for example Figure A1 c) d) (or similar figure) will help explaining the N and S, and at which point you define the ToE used in this study in a more visualized way.
We have modified the method section and added a Figure to the Appendix to clarify the points raised by the reviewers. Please see also our answers to reviewer # 1 and # 2 who raised similar issues. First, as the reviewer correctly points out, the preindustrial control simulations differ in their length: from 300 years (IPSL-CM5A-LR) to 1000 years (MPI-ESM-LR). We clarified this in the method section:
Although, the length of the control simulations differ between models, the entire duration of the control simulation is considered for each model to estimate the background noise.
Secondly, the reviewer correctly points out that the control simulation performed with the CESM1 was forced under 850 CE conditions, whereas CMIP5 control simulations are run under 1850 CE conditions. The following sentence was updated for more clarity:
The CMIP5 model simulations are branched off from preindustrial control simulations, whereas the CESM1.0 simulation is an extension of a last millennium simulation run under 850 CE conditions (Lehner et al., 2015)
Finally, we added a new figure to better illustrate the ToE definition (see Fig. A1)

3.2 Specific Comments

- Page 2, L24-25: I suggest to cite one of the Oschlies review paper in addition to Cocco and Bopp's papers (underestimating the trend and variability of o2 in the model simulations).
Reference to Oschlies et al. (2017) included as suggested. Acknowledged.

- Page 4, Method, Earth system models section: I am guessing this will not affect much on the overall results but why did you use your own CESM1 (with different spin-up procedures) for multi-model comparison instead of using CMIP5 CESM? In addition, is the CESM1 used in this study the same as the one in the early Hameau et al., (2019)?
The study considers all the available CMIP5 ESMs that provide oxygen, salinity and temperature 3-d fields for the piControl, historical and rcp85 simulations. Furthermore, the "in-house" CESM simulations that has already been described in Hameau et al., 2019 is used. The CESM simulation also follows the CMIP5 protocol (with the exception of the spin-up).
We have modified the corresponding text to:
In order to extent the multi-model ensemble from four to five family-models, we also included the output from simulations performed with the Community Earth System Model (CESM1.0) conducted at the Swiss Supercomputing Centre.

- Page 9, L1: What do you exactly mean by "combining climate sensitivity to anthropogenic forcing and natural variability in one metric"? I understand combining the anthropogenic forcing and natural variability part but I was not fully sure about the climate sensitivity statement.
We agree with the reviewer that this sentence might be confusing. We rewrote the sentence to:
The ToE allows for a comparison across climate models, by combining the amplitude of the climate response to anthropogenic forcing and the amplitude of natural variability in one metric.

- Page 12, L18: "an increasing ventilation" (following Gnanadesikan et al., 2007): Strictly speaking I would not state "an increasing ventilation" but it is more of a consequence of reduced upwelling as discussed in Gnanadesikan et al., 2007.
Thank you very much for this pointer. We agree with the reviewer and have modified the text as suggested.

- Figure 1. I understand from the Figure 1 that the SD reduces for the relative ToE but I also have some impression that two metrics could still give similar information. It might help to show additional map of ToE SD difference between Figure 1 (b) and (c) for example to show the bias (spread) reduction using this metric.
Thank you for this comment. We have replaced the lower panel (c, f) by the difference of the spread as suggested.

- Figure 6. For consistency, I suggest to used the same hatching as the previous figures to show the regions that one of the variables has not emerged by 2099 rather than saturated colors.
As suggested, hatching have been added in areas where both the T and O_2 signals do not emerge by the end of the 21st century. The caption of the figure has been updated with the following sentence:
No emergence in both T and O_2 are shown by the hatched areas.

- Figure 8. I like this summary figure aiming on incorporating emergent signals of both thermocline temperature and o2, along with AOU information (mainly indicating the water mass age information). Minor things on this, I think the x-axis is supposed to be -AOU (it puzzled me for a moment) and x-axis label should be corrected.
Many thanks for the positive comment. The label of the x-axis has been corrected as suggested.

References

Bopp, L., Le Quéré, C., Heimann, M., Manning, A. C., and Monfray, P.: Climate-induced oceanic oxygen fluxes: Implications for the contemporary carbon budget, Global Biogeochemical Cycles, 16, doi:10.1029/2001GB001445, URL http://onlinelibrary.wiley.com/doi/10.1029/2001GB001445/abstract, 2002.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, Biogeosciences, 10, 6225–6245, doi:10.5194/bg-10-6225-2013, URL http://www.biogeosciences.net/10/6225/2013/, 2013.

Bopp, L., Resplandy, L., Untersee, A., Mezo, P. L., and Kageyama, M.: Ocean (de)oxygenation from the Last Glacial Maximum to the twenty-first century: insights from Earth System models, Phil. Trans. R. Soc. A, 375, doi:10.1098/rsta.2016.0323, URL http://rsta.royalsocietypublishing.org/content/375/2102/20160323, 2017.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science, 359, doi:10.1126/science.aam7240, URL http://science.sciencemag.org/content/359/6371/eaam7240, 2018.

Cabré, A., Marinov, I., Bernardello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, doi:10.5194/bg-12-5429-2015, URL http://www.biogeosciences.net/12/5429/2015/, 2015.

Capotondi, A., Alexander, M. A., Bond, N. A., Curchitser, E. N., and Scott, J. D.: Enhanced upper ocean stratification with climate change in the CMIP3 models, Journal of Geophysical Research: Oceans, 117, doi:10.1029/2011JC007409, URL http://onlinelibrary.wiley.com/doi/10.1029/2011JC007409/abstract, citation Key Alias: capotondi_enhanced_2012-1, 2012.

Cheng, L., Abraham, J., Hausfather, Z., and Trenberth, K. E.: How fast are the oceans warming?, Science, 363, 128–129, doi:10.1126/science.aav7619, URL https://science.sciencemag.org/content/363/6423/128, 2019.

Cheung, W. W., Jones, M. C., Reygondeau, G., and Frölicher, T. L.: Opportunities for climate-risk reduction through effective fisheries management, Global change biology, 24, 5149–5163, 2018.

Cocco, V., Joos, F., Steinacher, M., Froelicher, T. L., Bopp, L., Dunne, J., Gehlen, M., Heinze, C., Orr, J., Oschlies, A., Schneider, B., Segschneider, J., and Tjiputra, J.: Oxygen and indicators of stress for marine life in multi-model global warming projections, Biogeosciences, 10, 1849–1868, doi:10.5194/bg-10-1849-2013, wOS:000317010600040, 2013.

Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H.-O., and Huey, R. B.: Climate change tightens a metabolic constraint on marine habitats, Science, 348, 1132–1135, doi:10.1126/science.aaa1605, URL http://science.sciencemag.org/content/348/6239/1132, 2015.

Enting, I. G.: On the use of smoothing splines to filter CO2 data, Journal of Geophysical Research: Atmospheres, 92, 10 977–10 984, doi:10.1029/JD092iD09p10977, URL http://onlinelibrary.wiley.com/doi/10.1029/JD092iD09p10977/abstract, 1987.

Frölicher, T. L. and Paynter, D. J.: Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales, Environmental Research Letters, 10, doi:10.1088/1748-9326/10/7/075002, URL http://stacks.iop.org/1748-9326/10/i=7/a=075002, 2015.

Frölicher, T. L., Rodgers, K. B., Stock, C. A., and Cheung, W. W. L.: Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors, Global Biogeochemical Cycles, 30, 1224–1243, doi:10.1002/2015GB005338, URL http://onlinelibrary.wiley.com/doi/10.1002/2015GB005338/full, 2016.

Gattuso, J.-P., Magnan, A., Billé, R., Cheung, W. W. L., Howes, E. L., Joos, F., Allemand, D., Bopp, L., Cooley, S. R., Eakin, C. M., Hoegh-Guldberg, O., Kelly, R. P., Pörtner, H.-O., Rogers, A. D., Baxter, J. M., Laffoley, D., Osborn, D., Rankovic, A., Rochette, J., Sumaila, U. R., Treyer, S., and Turley, C.: Contrasting futures for ocean and society from different anthropogenic CO2 emissions scenarios, Science, 349, doi:10.1126/science.aac4722, URL http://science.sciencemag.org/content/349/6243/aac4722, 2015.

Hameau, A., Mignot, J., and Joos, F.: Assessment of time of emergence of anthropogenic deoxygenation and warming: insights from a CESM simulation from 850 to 2100 CE, Biogeosciences, 16, 1755–1780, doi:10.5194/bg-16-1755-2019, URL https://www.biogeosciences.net/16/1755/2019/, 2019.

Henson, S. A., Beaulieu, C., and Lampitt, R.: Observing climate change trends in ocean biogeochemistry: when and where, Global Change Biology, 22, 1561–1571, doi:10.1111/gcb.13152, URL http://onlinelibrary.wiley.com/doi/10.1111/gcb.13152/abstract, 2016.

Henson, S. A., Beaulieu, C., Ilyina, T., John, J. G., Long, M., Séférian, R., Tjiputra, J., and Sarmiento, J. L.: Rapid emergence of climate change in environmental drivers of marine ecosystems, Nature Communications, 8, doi:10.1038/ncomms14682, URL http://www.nature.com/ncomms/2017/170307/ncomms14682/full/ncomms14682.html, 2017.

IPCC: Summary for policymakers, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Pörtner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N., URL https://www.ipcc.ch/srocc/home/, in Press, 2019.

Joos, F., Plattner, G.-K., Stocker, T. F., Körtzinger, A., and Wallace, D. W. R.: Trends in marine dissolved oxygen: Implications for ocean circulation changes and the carbon budget, Eos, Transactions American Geophysical Union, 84, 197–201, doi:10.1029/2003EO210001, URL http://onlinelibrary.wiley.com/doi/10.1029/2003EO210001/abstract, 2003.

Keller, K. M., Joos, F., and Raible, C. C.: Time of emergence of trends in ocean biogeochemistry, Biogeosciences, 11, 3647–3659, doi:10.5194/bg-11-3647-2014, 2014.

Keller, K. M., Joos, F., Lehner, F., and Raible, C. C.: Detecting changes in marine responses to ENSO from 850 to 2100 C.E.: Insights from the ocean carbon cycle, Geophysical Research Letters, 42, doi:10.1002/2014GL062398, URL http://onlinelibrary.wiley.com/doi/10.1002/2014GL062398/abstract, 2015.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369–373, doi:10.1038/nclimate1716, URL https://www.nature.com/articles/nclimate1716, 2013.

Lehner, F., Joos, F., Raible, C. C., Mignot, J., Born, A., Keller, K. M., and Frölicher, T. L.: Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, Earth System Dynamics, 6, 411–434, doi:10.5194/esd-6-411-2015, URL http://www.earth-syst-dynam.net/6/411/2015/, 2015.

Nijsse, F. J. M. M., Cox, P. M., Huntingford, C., and Williamson, M. S.: Decadal global temperature variability increases strongly with climate sensitivity, Nature Climate Change, 9, 598–601, doi:10.1038/s41558-019-0527-4, URL https://www.nature.com/articles/s41558-019-0527-4, 2019.

Oschlies, A., Duteil, O., Getzlaff, J., Koeve, W., Landolfi, A., and Schmidtko, S.: Patterns of deoxygenation: sensitivity to natural and anthropogenic drivers, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375, doi:10.1098/rsta.2016.0325, URL http://rsta.royalsocietypublishing.org/content/375/2102/20160325, 2017.

Pedlosky, J.: Ocean Circulation Theory, Springer-Verlag, Berlin Heidelberg, doi:10.1007/978-3-662-03204-6, URL https://www.springer.com/gp/book/9783540604891, 1996.

Plattner, G.-K., Joos, F., and Stocker, T. F.: Revision of the global carbon budget due to changing air-sea oxygen fluxes, Global Biogeochemical Cycles, 16, 1096, doi:10.1029/2001GB001746, URL http://onlinelibrary.wiley.com/doi/10.1029/2001GB001746/abstract, 2002.

Resplandy, L., Séférian, R., and Bopp, L.: Natural variability of CO2 and O2 fluxes: What can we learn from centuries-long climate models simulations?, Journal of Geophysical Research: Oceans, 120, 384–404, doi:10.1002/2014JC010463, URL http://onlinelibrary.wiley.com/doi/10.1002/2014JC010463/abstract, 2015.

Rodgers, K. B., Lin, J., and Frölicher, T. L.: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model, Biogeosciences, 12, 3301–3320, doi:10.5194/bg-12-3301-2015, URL https://www.biogeosciences.net/12/3301/2015/, 2015.

Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Frölicher, T. L., Dunne, J. P., Ishii, M., and Slater, R.: Emergence of anthropogenic signals in the ocean carbon cycle, Nature Climate Change, pp. 1–7, doi:10.1038/s41558-019-0553-2, URL https://www.nature.com/articles/s41558-019-0553-2, citation Key Alias: schlunegger_emergence_2019, 2019.

Stramma, L., Johnson, G. C., Sprintall, J., and Mohrholz, V.: Expanding Oxygen-Minimum Zones in the Tropical Oceans, Science, 320, 655–658, doi:10.1126/science.1153847, URL http://science.sciencemag.org/content/320/5876/655, 2008.

Tjiputra, J. F., Goris, N., Lauvset, S. K., Heinze, C., Olsen, A., Schwinger, J., and Steinfeldt, R.: Mechanisms and Early Detections of Multidecadal Oxygen Changes in the Interior Subpolar North Atlantic, Geophysical Research Letters, 45, 4218–4229, doi:10.1029/2018GL077096, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077096, 2018.

Is deoxygenation detectable before warming in the thermocline?

Angélique Hameau[1, 2], Thomas L. Frölicher[1, 2], Juliette Mignot[3], and Fortunat Joos[1, 2]

[1]Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
[2]Oeschger Centre for Climate Change Research, Bern, Switzerland
[3]LOCEAN/IPSL, Sorbonne Université (SU)-CNRS-IRD-MNHN, Paris, France

Correspondence: Angélique Hameau (hameau@climate.unibe.ch)

Abstract. Multiple lines of evidence from observation- and model-based studies show that anthropogenic Anthropogenic greenhouse gas emissions cause ocean warming and oxygen depletion, with adverse impacts on marine organisms and ecosystems. Temperature is considered as Warming is one of the main indicators of anthropogenic climate change, but, in the thermocline, anthropogenic changes in biogeochemical tracers such as oxygen changes in oxygen and other biogeochemical tracers may emerge from the bounds of natural variability before changes in temperatureprior to warming. Here, we compare the local time of emergence assess the Time of Emergence (ToE) of anthropogenic temperature and oxygen changes in the thermocline change in thermocline temperature and thermocline oxygen within an ensemble of Earth system model simulations from the fifth phase of the Coupled Model Intercomparison Project(CMIP5). Anthropogenic deoxygenation emerges from natural internal variability before warming in . Changes in temperature typically emerge from internal variability prior to changes in oxygen. However, in about a third (35 ±11 %) of the global thermocline . Earlier emergence of oxygen than temperature change is simulated by all models in parts of the subtropical gyres of the Pacific and the Southern Ocean. Earlier detectable changes in oxygen than temperature are typically related to decreasing trends in ventilation. The supply of oxygen-rich surface waters to the thermocline is reduced as evidenced by an increase in apparent oxygen utilisation over the simulations. Concomitantly, deoxygenation emerges prior to warming. In these regions, both reduced ventilation and reduced solubility add to the oxygen decline. In addition, reduced ventilation slows the propagation of the warming signal is hindered by slowing ventilation, which delays the warming in the thermoclineanthropogenic warming from the surface into the ocean interior, further contributing to the delayed emergence of warming compared to deoxygenation. As the magnitudes Magnitudes of internal variability and simulated temperature and oxygen changesof anthropogenic change, which determine ToE, vary considerably among models , we compute the local ToErelative to the global mean ToEwithin each modelleading to model-model differences in ToE. We introduce a new metric, relative ToE, to facilitate the multi-model assessment of ToE. This reduces the inter-model spread in the relative ToE compared to the traditionally evaluated absolute ToE. Our results underline the importance of an ocean biogeochemical observing system and that the detection of anthropogenic impacts becomes more likely when using multi-tracer observations.

1 Introduction

Carbon emissions from human activities are causing ocean warming (Rhein et al., 2013) and ocean deoxygenation, i.e. a decrease in the oceanic oxygen (O_2) concentration (Sarmiento et al., 1998; Bopp et al., 2002; Matear and Hirst, 2003; Battaglia and Joos, 2018). Both warming and deoxygenation adversely affect marine organismsand , ecosystems and the services they provide (e.g. Pörtner et al., 2014; Deutsch et al., 2015; Gattuso et al., 2015; Magnan et al., 2016).

All major ocean basins have experienced a significant warming over the last few decades. Warming is generally strongest at the surface and weaker at deeper layers, indicative of heat penetrating from the surface towards the deep ocean as expected from atmospheric greenhouse gas forcing. The strongest warming in the top 2000 m has been observed in the Southern Ocean (Roemmich et al., 2015) and the tropical/subtropical Pacific and Atlantic Ocean (Cheng et al., 2017). On regional to local scalescales, the anthropogenic warming signal may be masked by natural interannual to multi-decadal variability. For example, decadal-scale cooling trends in the tropical Pacific and Indian oceans may arise from natural El Nino-Southern Niño-Southern Oscillation and/or Indian Ocean Dipole variability (Han et al., 2014). Similarly, decadal variability in the Atlantic Meridional Overturning is observed to modulate temperature and heat content change in the North Atlantic (Chen and Tung, 2018). Global climate models, such as the Earth system models that participated in phase 5 of the Coupled Model Intercomparison Project (CMIP5) reproduce the long-term trend in global ocean heat content over the last 50 years when uncertainties of observation-based estimate and internally generated natural variability are taken into account (Frölicher and Paynter, 2015, Cheng et al., 2019).

Concomitant with ocean warming, observation-based Observation-based studies indicate that the global ocean oxygen content has decreased since 1960 (e.g. Schmidtko et al., 2017). Increased ocean surface temperature reduces oxygen solubility, limiting atmospheric oxygen dissolution into the upper ocean. Increased surface temperature reduces oxygen solubility, limiting atmospheric oxygen dissolution into the upper ocean. In subsurface waters, oxygen concentration is also affected by ventilation, changes in ventilation and the remineralisation of organic matterand air-sea disequilibrium but the oxygen decrease is . In the contemporary ocean, oxygen decreases in the interior are mostly dominated by a reduction in ventilation and increased consumption (Bopp et al., 2002, Bopp et al., 2017,Hameau et al., 2019with a smaller role for changes related to the production of organic matter, O_2 solubility, and air-sea equilibration of O_2 in surface waters (Bopp et al., 2002,Plattner et al., 2002, Bopp et al., 2017,Tjiputra et al., 2018, Hameau et al., 2019). The largest losses oxygen declines are located in the Pacific Ocean (equator and northern hemisphere) and the Southern Ocean. However, observations are relatively sparse and only start during the late industrial period. It remains therefore still difficult to precisely distinguish human caused in the second half of the 20th century. Therefore, it is challenging to distinguish human-caused trends from natural variations in the observational record of ocean O_2.

Global climate models, such as the Earth system models that participated in phase 5 of the Coupled Model Intercomparison Project (CMIP5) reproduce the long-term trend in global ocean heat content over the last 50 years when uncertainties of observation-based estimate and internally generated natural variability are taken into account (Frölicher and Paynter, 2015, Cheng et al., 20 Modelling studies also agree on the sign of oceanic O_2 changes, but likely underestimate the magnitude of loss (Cocco et al., 2013, Bopp et a

In particular in the tropical regions, models are not able to reproduce observed O_2 decrease in equatorial low-oxygen zones (Stramma et al., 2008; Cocco et al., 2013; Cabré et al., 2015).

It is expected that ocean warming and deoxygenation, and the combination thereof, increases the risk of adverse impacts on marine organisms and ecosystem services (Pörtner et al., 2014). Warming of the ocean influences the physiology and ecology of almost all marine organisms. Reduced oceanic O_2 concentrations can disrupt marine ecosystems by pushing organisms to their species-specific limits of hypoxic tolerance, below which the species are no longer able to meet their metabolic O_2 demand. The species-specific metabolic demand of O_2 is also a function of temperature, as warmer temperatures increase metabolic rates and oxygen requirements (Deutsch et al., 2015). At the same time, higher ocean temperatures also decrease oxygen supply through reduced ventilation, enlarging the regions with limited O_2 concentrations and thus shifting ecosystem distribution (Cheung et al., 2011).

Beyond the combined impact of physical and biogeochemical changes, an interesting question is whether anthropogenic changes in the ocean interior are first detectable in ~~physical variables~~ variables that are routinely and frequently measured such as temperature (T) or in ~~biogeochemical variables such as , pH, or DIC (Joos et al., 2003; Keller et al., 2015)~~variables with a relatively low observational coverage but potentially high impact for ecosystems such as O_2 (Joos et al., 2003). The answer may have implications for measurement strategies to detect anthropogenic changes in subsurface waters as well as for the impacts of physical and biogeochemical change on marine life. ~~On the one hand, physical processes generally influence the biogeochemistry of the ocean. For example, global~~ For the surface ocean, earlier studies (Keller et al., 2014),Rodgers et al., 2015,Frölicher e Schlunegger et al., 2019 showed that the anthropogenic signals of pH and pCO2 emerge earlier than sea surface temperature and O_2 change and earlier than productivity changes. Changes in surface O_2 are tightly coupled to temperature-driven solubility changes and O_2 varies hand in hand with sea surface temperature and the two signals emerge typically concomitantly. Regarding the ocean interior, the sequence of emergence for O_2 and T is less clear. Global warming increases surface ocean temperature, which ~~reduces solubility and decreases air-sea gas exchange of~~ tends to reduces O_2. On the other hand, O_2 is also influenced by non-thermal processes, such as respiration and the redistribution by ocean circulation and mixing. Respiration of organic matter in the ocean interior ~~will~~ may have a larger influence on O_2 change than temperature-driven solubility change in a more stratified and less ventilated ocean. One could therefore expect that, under global warming, the combined effect of increased O_2 consumption and decreased O_2 solubility will accelerate the O_2 depletion in subsurface waters and that O_2 may be detectable before the warming reaches that layer.

~~In the context of climate change, the distinction between anthropogenic induced changes and natural variability is pivotal to gain understanding on and temperature changes.~~ The concept of Time of Emergence (ToE; Christensen et al., 2007; Hawkins and Sutton, 2012) is often used to determine the point in time when the anthropogenic signal becomes larger than the range of natural variability. ToE has been broadly used in climate change detection for physical climate variables (e.g. surface temperature: Hawkins and Sutton, 2012; Frame et al., 2017), land carbon fluxes (Lombardozzi et al., 2014) or marine biogeochemical variables (e.g. pH, alkalinity, DIC, pCO_2: Hauri et al., 2013; Keller et al., 2014; marine biological productivity: Henson et al., 2016). A limited number of studies addressed anthropogenic deoxygenation detection in the subsurface layers (Rodgers et al., 2015; Frölicher et al., 2016; Henson et al., 2016; Long et al., 2016; Henson et al., 2017; Hameau et al., 2019). ~~Only one~~

study(Hameau et al., 2019), using One study, Hameau et al. (2019), uses a single model , investigated (CESM), to investigate ToE of temperature and oxygen in the thermocline. One main finding of their study is that anthropogenic driven , finding that anthropogenic ocean warming emerges much earlier than the O_2 signal in low and midlatitude regions. This Delayed emergence of changes in O_2 is due to the opposite effect of decreases in solubility and consumption, delaying the changesopposing effects of O_2 solubility and O_2 consumption. In the high latitudes and the Pacific subtropical gyres, deoxygenation emerges before ocean warming in their model. This is because the CESM. This occurs because decrease in oxygen solubility is reinforced by an increase in are reinforced by increased O_2 consumption, leading to strong depletion.

Even though this earlier study indeed identifies regions with earlier emergence of in comparison with temperature, consistent with our outlined hypothesis aboveO_2 depletion. However, it is currently unclear unknown if this single-model result is robust across a suite of different Earth system model simulations. A Here, we conduct a multi-model study that addresses and compares the emergence of anthropogenic warming and of to more broadly test the hypothesis that anthropogenic deoxygenation in the thermocline is currently missing. However, such a comparison across models is delicate, as the absolute years emerges prior to anthropogenic warming. Since the primary objective is to test the consistency across models of the order of emergence (Keller et al., 2014; Henson et al., 2017) is highly dependent of the ToE methodology. We therefore introduce deoxygenation prior to warming) within a single model. We introduce and use a relative ToE , considered to conduct the intercomparison, rather than the absolute year of ToE. We define relative ToE as a deviation relative to the model mean ToE for improved model intercomparison.

In this study, we analyse and compare the relative ToE(T) and ToE(O_2) in the thermocline (200 – 600 m) using nine different CMIP5 Earth system models. We also assess the impact of using the relative ToE in comparison to the classical approach using absolute ToE. In addition, we discuss the magnitude of background internal variability and anthropogenic signal, and their translation into ToE. Finally, we analyse the role of solubility, ventilation and respiration for the emergence of anthropogenic changes in oxygen and temperature.

2 Method

2.1 Earth system models

We use output from eight different configurations of four Earth system models (ESMs) that participated in the Coupled Model Intercomparison Project 5 (CMIP5; Taylor et al., 2012): GFDL-ESM2M, GFDL-ESM2G, HadGEM2-CC, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MPI-ESM-LR and MPI-ESM-MR (Table 1). In ~~addition, output from simulations~~ order to extent the multi-model ensemble from four to five family-models, we also included the output from simulations performed with the Community Earth System Model (CESM1.0) conducted at the Swiss Supercomputing Centre~~are included in the analysis.~~ . The horizontal ocean model resolution is ~~generally~~ about 1° ~~(both~~ in both the GFDL models and CESM1.0~~). Exceptions are the~~ . HadGEM2-CC and IPSL models ~~, which~~ have a horizontal resolution of about 2° and the MPI models ~~, which~~ have a horizontal resolution of about 0.4° (MR) and 1.5° (LR). Of the nine models, all but one (GFDL-ESM2G, isopycnal vertical coordinate) use a pressure-based vertical coordinate. For additional information on the individual model setups, the reader is referred to the references listed in Table 1.

Both the CMIP5 ESMs and the CESM1.0 were run under prescribed anthropogenic and natural greenhouse gas and aerosol forcing. All simulations span the historical 1861-2005 period and the 2006-2100 period following the Representative Concentration Pathway 8.5 (RCP8.5) scenario. The RCP8.5 represents a high emission scenario with a radiative forcing of 8.5 W m^{-2} in year 2100 (Riahi et al., 2011). These simulations are complemented with output from corresponding control runs with constant preindustrial forcing. The CESM1.0 simulations differ from the CMIP5 simulations only with regard to the spin-up procedure: The CMIP5 model simulations are branched off from preindustrial control simulations, whereas the CESM1.0 simulation is an extension of a last millennium simulation run under 850 CE conditions (Lehner et al., 2015). For this study, all CMIP5 models are used for which the ~~3-d~~ 3 dimensional output of oxygen, temperature and salinity for all simulations were available on the Earth System Grid. We regridded all model output onto a regular 1° x 1° grid. Even though the model drift in the control simulations is relatively small in the thermocline (3.6 ± 2.4 x 10^{-3} mmol m^{-3} $year^{-1}$ for trend in global mean oxygen concentration and 7.2 ± 6.6 x 10^{-5} °C $year^{-1}$ for trend in global mean temperature averaged over 200 – 600 meters), we detrended all model output with a linear trend obtained from the preindustrial control simulation in each grid cell. The CESM1.0 simulation also shows some model drift.Therefore, an exponential curve was fitted to the annual output of its associated control simulation at each grid cell. The detrending procedure is described in detail in Hameau et al. (2019).

2.2 Multi-model analysis methods

We use the concept of Time of Emergence (ToE; e.g. Hawkins and Sutton 2012) to compare anthropogenic changes in O_2 and temperature (signal; S) with internal ~~natural~~ variations (background noise; N). Here, ToE represents the moment in time at which the ocean state becomes distinct from the preindustrial state. Appendix Figure A1 provides a graphical illustration of the method used to compute ToE.

We define the absolute ToE as the first year when the anthropogenic signal S becomes equal or larger than twice the noise of internal natural variability N (Eq. 1; following Hameau et al., 2019).

$$\mathrm{ToE}: \frac{S}{N} \geqslant 2$$

We estimate N ; Fig. A1). The threshold is set to two in order to distinguish the signal from the noise at 95 % confidence level. Annual O_2 and T data are first averaged over the thermocline (200 – 600 m) at each grid cell by calculating point of the horizontal grid and local S and N are computed from these depth-averaged values for each model, variable and (horizontal) grid-point. Annual anomalies are calculated relative to the preindustrial period (1860 – 1959).

$$\mathrm{ToE}: \frac{S}{N} \geqslant 2 \tag{1}$$

The background noise, N, is computed as one standard deviation (SD) of the annual means in O_2 and of T from the preindustrial control simulation. Thus, annual preindustrial control output. The entire duration of the control simulation is considered for each model to estimate the background noise. N represents the noise due to the internal chaotic variability of the climate system. Note that this definition of the noise differs from Hameau et al. (2019), who used internal plus externally-forced natural variability from a last millennium simulation to assess the standard background noise.

The annual output of the forced, transient simulation (1860 – 2099) is smoothed by a low pass spline filter (Enting, 1987) to estimate S is estimated at each grid cell from the forcedsimulation by fitting the annual evolution of the considered variable with a low-pass filter (for each (horizontal) grid point in the thermocline. The cut-off period of the spline is set to 80 years ; Enting, 1987) in order to remove short term variations , to remove decadal to multi-decadal variations (e.g. associated with internal natural variability. variability). The signal S is then defined as the value of the spline at each point in time. To ensure that S indeed detects anthropogenic trend, we also apply a criterion for the sign of S to define ToE: S needs to have the same sign as the difference between the last 30 years of the future simulation and the preindustrial average for the corresponding variable and grid cell. Annual O_2 and T data are first averaged over the thermocline (200 – 600 m) at each horizontal grid cell and local S and N are computed from these depth-averaged values for each model, variable and grid-pointpoint.

In order to minimise inter-model differences and to highlight the common spatial patterns of ToE, we introduce a new metric, the relative ToE ($\mathrm{ToE}_{\mathrm{rel}}$). It is defined as the absolute ToE ($\mathrm{ToE}_{\mathrm{abs}}$) minus the global area-averaged ToE ($\mathrm{ToE}_{\mathrm{glob}}$; Eq. 2).

$$\mathrm{ToE}_{\mathrm{rel}} = \mathrm{ToE}_{\mathrm{abs}} - \mathrm{ToE}_{\mathrm{glob}} \tag{2}$$

Median S, N, ToE and $\mathrm{ToE}_{\mathrm{rel}}$ are first computed from the annual output for each model and at each (horizontal) grid cell. Then, multi-model median and spread (interquartile range) of the multi-model estimations are computed from the annual outputs of the model ensembleand uniform model ensemble. The median represents a "best" estimate and the interquartile range a measure of model uncertainty. Uniform weights are applied to each model configuration to compute those statistics.

Tests have been performed using a weighted median as several simulations stem from the same model family (CESM x 1; GFDL x 0.5; HadGEM2 x 1; IPSL x 0.3; MPI x 0.5). However, median and interquartile range of the multi-model ensemble are not sensitive to the weighting scheme applied (not shown). Because an anthropogenic signal may not emerge before the end of the simulation in year 2100, ToE can be undefined. We therefore ~~request~~ require that ToE values ~~is~~ are defined for at least seven out of nine models to compute the multi-model statistics (median and spread). If more than two models have an undefined ToE, we mask the grid points in maps of the multi-model median and of the multi-model the spread.

~~To understand the processes behind~~

2.3 Separating mechanisms of oxygen change

To diagnose processes driving the simulated changes in ocean ~~, we decompose the changes into solubility () or thermal components and~~ O_2, the direct thermal/solubility component of change ($O_{2,sol}$) can be isolated from the total O_2 change. The residual, Apparent Oxygen Utilisation (AOU)~~or~~ , represents the summation of all non-thermal ~~components:~~ changes, including those resulting from changes in ventilation and remineralisation.

$$[O_2] = [O_{2,\,sol}] + [\text{-AOU}] \quad (3)$$

The solubility component for each model is computed following Garcia and Gordon (1992), which requires local salinity and temperature output. The solubility depends mostly on temperature with a small contribution of salinity. The non-thermal component ([-AOU]) is deduced from the difference between ~~and~~ $O_{2,sol}$ and O_2 following Eq. 3. In Sect. 3.4, we will use changes in [-AOU] as a proxy for changes in water mass age and ventilation. Output of an ideal age tracer is not available for most models. A decrease in water exchange between the surface ocean and the thermocline typically leads to an increase in water mass age in the thermocline. Therefore, changes in ventilation affect the balance between the rate of supply of O_2-rich waters from the surface and the rate of O_2 consumption by remineralisation of organic matter. It has been demonstrated in earlier studies (e.g. Gnanadesikan et al., 2012; Bopp et al., 2017; Hameau et al., 2019) that a decrease in [-AOU] typically corresponds to a decrease in ventilation and an increase in water mass age, as simulated changes in the remineralisation rates of organic material and in associated O_2 consumption are relatively small over the 21st century.

3 Results

3.1 Relative Time of Emergence

We start by discussing the multi-model median and spread of relative ToE estimates for potential temperature (Fig. 1a, b) and dissolved oxygen (Fig. 1d, e) changes in the thermocline (200 – 600m). An analysis of the roles of internal variability and anthropogenic change ToE and why anthropogenic change is detectable early or late is presented in Sect. 3.3.

3.1.1 Anthropogenic warming

$ToE_{rel}(T)$ shows early emergence in low latitudes and between 30° S and 60° S, and late emergence in the western tropical Pacific, in the Atlantic subpolar gyre and the subtropical gyres of the Indian and Pacific Ocean (Fig. 1a). The northern Indian Ocean and the eastern equatorial Atlantic stand out as the regions with earliest emergence in anthropogenic warming, i.e. 70 years (median of nine $ToE_{rel}(T)$) before the global average ToE. No emergence of warming by the end of the 21st century (for at least 3 models; cf. Sect. 2.2) is simulated in the subtropical gyres of the Indian and the Pacific oceans, south of Greenland and locally south of 60° S.

The multi-model spread in $ToE_{rel}(T)$ is generally small in regions with early emergence (Fig. 1b). This is the case in many regions of the Pacific and the Southern Ocean (±15 years). However, in the Atlantic subtropical gyres and in the Arabian Sea, the early $ToE_{rel}(T)$ estimates are associated with a wider spread across models (±25 to ±45 years). Large inter-model spread is also found in the Kuroshio extension and in the Indian and Atlantic region of the Southern Ocean (±50 years). In the global On average, the multi-model spread for $ToE_{rel}(T)$ is about 25 years.

ToE values for individual horizontal grid cells are globally averaged to obtain an area-weighted global mean ToE for the thermocline and each model. These global mean values range between year 1963 and year 2033 for the nine models (see subtitles in Fig. 2). The patterns of $ToE_{rel}(T)$ for each individual model are shown in Fig. 2. As described previously, low latitude regions and parts of the Southern Ocean show earlier emergence compared to mid- and other high-latitude regions. The HadGEM2-CC model (Fig. 2c) is an exception in that respect as temperature emerges later (+30 to +50 years) than the global average in the tropical Atlantic and Pacific. In the Pacific and Indian subtropical gyre regions, the models show late (IPSL family) or no emergence. And finally, CESM and the IPSL family models are the only models that show emergence before the end of the 21st century in the subtropical gyres of the Pacific.

3.1.2 Anthropogenic deoxygenation

In contrast to $ToE_{rel}(T)$, the pattern of $ToE_{rel}(O_2)$ is relatively homogeneous most of the thermocline shows no emergence of the anthropogenic O_2 change by the end of the 21st century (Fig. 1d) and only . In the remaining regions $ToE_{rel}(O_2)$ varies by about ±40 years~~between regions.~~. Early emergence is found in the subtropical gyre of the North Pacific, the northern North Atlantic, the Atlantic sector of the Southern Ocean, and generally south of 60° S. No emergence is simulated in 47 % of the ocean area by the end of the 21st century including large parts of the tropics and the subtropical gyres of the Atlantic Ocean and the Indian Ocean.

The multi-model spread for $ToE_{rel}(O_2)$ is 20 years in the global average and thus somewhat smaller than for $ToE_{rel}(T)$. The models show a high spread for $ToE_{rel}(O_2)$ (±50 years) at low latitudes, such as in the southern Arabian Sea or in the equatorial Atlantic, whereas high model agreement is found in parts of the central North Pacific and the northern Indian Ocean (spread of ±15 years) (Fig. 1e). In the eastern tropical Atlantic, the spread for $ToE_{rel}(O_2)$ is, despite a smaller global mean spread, larger than for $ToE_{rel}(T)$. In summary, even though the median pattern of $ToE_{rel}(O_2)$ is relatively uniform in comparison to $ToE_{rel}(T)$, the spread for $ToE_{rel}(O_2)$ varies between regions as for $ToE_{rel}(T)$.

The multi-model median O_2 signal does not emerge in 47 % of the global thermocline as noted above. Mid and low latitudes show no emergence by the end of the 21st century in most of the models (Fig. 3). However, the exact regions of no emergence differ between models. This regional mismatch, in combination with the requirement that at least seven out of nine models need to show an emerging signal (Sect. 2.2), explains why in the multi-model analysis many grid cells are masked, indicating no emergence in the median (Fig. 1d-f). The area fraction with no emerging O_2 signal is smaller in individual models than in the multi-model median and ranges between 10 and 30 %.

As for temperature, a large range in absolute ToE is found with globally-averaged $ToE(O_2)$ ranging between the year 1991 and 2046 for the nine models (see subtitles in Fig. 3). The analysis of $ToE_{rel}(O_2)$ for individual models reveals some additional notable differences (Fig. 3). GFDL-ESM2M, GFDL-ESM2G, HadGEM2-CC and CESM1.0 simulate early emergence in the Southern Ocean, but the IPSL models project no emergence of deoxygenation in this region by the end of the 21st century. In addition, the IPSL models and the CESM1.0 model show relatively early emergence in many grid cells of the western tropical Pacific, a region with no emergence in other models. $ToE_{rel}(O_2)$ also diverges across the models in the Atlantic subtropical gyres: in the HadGEM2 and IPSL simulations, oxygen changes are simulated to emerge relatively early ($ToE_{rel}(O_2) \sim 40$ to 60 years), whereas in the GFDL, MPI and CESM simulations, the changes are not yet detectable by the end of the 21st century.

3.2 Relative versus absolute ToE

Mapping ToE_{rel} for different models is intended to emphasise common patterns across models by removing the global mean bias between models, while model-model differences in ToE_{abs} are indicative of an overall model uncertainty.
The multi-model spread for ToE_{abs} is ~~by design~~ in average larger than the multi-model spread for ToE_{rel} for temperature (Fig. 1~~b,~~ c) and oxygen (Fig. 1~~e,~~ f), while spatial patterns are similar for ToE_{rel} and ToE_{abs}. On global average, the spread is reduced from ±30 years for $ToE_{abs}(T)$ to ± 23 years for $ToE_{rel}(T)$ and from ±20 years for $ToE_{abs}(O_2)$ to ±17 years for $ToE_{rel}(O_2)$. Regionally, the reduction can be larger. For example, in the equatorial regions, the Atlantic and the Southern Ocean, the spread

is reduced by 20 to 50 years when computed for $ToE_{rel}(T)$ instead for $ToE_{abs}(T)$. Similarly, the spread in ToE(O_2) is reduced from ±35 to ±5 years in parts the North Pacific.

3.3 Internal ~~natural~~ variability and anthropogenic signals

The ToE allows for a comparison across climate models, by combining ~~climate sensitivity~~ the amplitude of the climate response to anthropogenic forcing and the amplitude of natural variability in one metric. The magnitude and the spatial patterns of the internal ~~natural variability (SD)~~ variability and of the anthropogenic signal for both thermocline temperature and oxygen are discussed next.

The multi-model median of internal ~~natural~~ variability for thermocline temperature fluctuates with an amplitude typically ranging between ±0.1 °C in the tropics and the Arctic Ocean, and ±0.5 °C in mid-to-high latitudes (Fig. 4a). SD(T) is the largest (up to ±0.9 °C) in the Western Boundary Currents such as the Kuroshio Current and the Gulf Stream. The internally generated variability is also relatively large along the equatorward flanks of the subtropical gyres. It is also in these regions where SD(T) differs most among models (up to ±0.5 °C along the North Atlantic Current; Fig. 4c).

In the multi-model median, temperature in the thermocline is projected to increase on global average by 1.2±0.7 °C (Fig. 4b) by the end of the 21st century under the RCP8.5 scenario relative to the period 1861-1959, in accordance with (Levitus et al., 2009, 2012; Bilbao et al., 2019). Large warming of more than 4.0±0.7 °C is projected in the northern North Atlantic and around the subantarctic water in the Indian and Atlantic Oceans (Fig. 4b and Fig. S4). We note that these ~~regions~~ projections are also characterised with the largest inter-model spread (±1.5 °C; Fig. 4d and Fig. S4) and uncertainties in these regional warming projections are large . Finally, disagreement among models in simulating changes in thermocline temperature is also large in the Arctic Ocean, possibly related to different simulated changes in sea ice cover (Stroeve et al., 2012; Wang and Overland, 2012).

The combination of a strong signal and small variability ~~typically~~ results in early detection of the changes. This is the case in the Southern Ocean at 45° S (in the Atlantic and Indian regions; Fig. 1a), where the anthropogenic warming is strong (up to 4 °C; Fig. 4b) but the internal variability is relatively small (0.1 °C to 0.3 °C; Fig. 4a). However, early emergence of anthropogenic changes can also occur when the signal is relatively small, if the variability is even smaller. This is the case in the tropical oceans such as in the Arabian Sea~~and~~ , the equatorial Atlantic and the western equatorial Pacific, where water masses warm modestly (up to 1.5 °C), but vary naturally between 0.1 °C and 0.2 °C only. It is also the case in the eastern equatorial Pacific, where the early emergence arise from the very weak internal variability in the thermocline, although, the temperature increase (∼0.80 °C) is also relatively weak. In this region, the substantial variability in O_2 and T is largely confined to the top 200 m. No emergence by the end of the 21st century, such as simulated in the subtropical gyres of the Indian and Pacific oceans, results from a relatively weak signal combined with a relatively strong variability in these regions.

Natural Internal variability of dissolved oxygen concentrations is particularly large in the northern North Pacific and North Atlantic, the Southern Ocean and along the equatorward boundaries of the subtropical gyres with SD(O_2) of up to 10 mmol m^{-3} (Fig. 5a). The multi-model spread of SD(O_2) (Fig. 5c) is about equally large as the median of SD(O_2) (Fig. 5a) along the equatorward boundaries of the subtropical gyres. Looking at the individual model responses, the natural O_2 internal variability shows a wide range of different patterns (Fig. S5). The GFDL and MPI models simulate high natural internal variability of oxygen in the entire thermocline, whereas CESM, HadGEM2 and IPSL models show high variability regionally.

The O_2 concentration in the thermocline (Fig. 5b) is projected to decrease under global warming, in accordance with previous model studies (e.g. Sarmiento et al., 1998; Cocco et al., 2013; Bopp et al., 2017). The anthropogenic decrease in O_2 is large in the Southern Ocean, in the North Pacific subtropical gyre and in the North Atlantic subpolar gyre. In tropical regions, the changes are projected to be small, except for the western Indian ocean, where more than 70 % of the models project an increase of O_2 concentration. The simulated O_2 changes differ most across models in high latitudes and in the subopolar gyres, as well as in the equatorial Indian ocean (Fig. 5d).

Despite differences in the simulated magnitude of changes and natural O_2 changes and internal variability patterns of O_2 between the different models, the resulting ToE$_{\text{rel}}$() are surprisingly O_2) are robust across models. For example, the decrease in O_2 spans from -12 to -40 mmol m^{-3} (Fig. S6) and SD(O_2) spans from $\pm$5 to $\pm$15 mmol m^{-3} (Fig. S5) in the central North Pacific. Moreover, the spatial locations of the maximum O_2 depletion differ across the models. However, ToE$_{\text{rel}}$(O_2) in this region is within $\pm$ $\sim$10 years (Fig. 1d), with a relatively high confidence interval low spread ($\pm$10 years) compared to ToE$_{\text{abs}}$(O_2) ($\pm$30 years). Another example is the CESM model. The very early detection of anthropogenic changes (for temperature and oxygen) in the CESM model described in Sect. 3.1, results from a particularly weak of internal variability (Figs. S3i and S5i; see also Hameau et al., 2019) combined with a high climate sensitivity of the model (Figs. S4i and S6i). Using the The ToE$_{\text{rel}}$ allows the comparison of ToE resulting from CESM output with the results from the 8 models in spite of these biases model-model differences (Figs. 3 and 2 2 and 3).

3.4 Comparison of ToE(O_2) with ToE(T)

Are changes in detectable earlier than warming in the thermocline? We examine this question with the help of Fig. 7, which shows ToE(T) minus ToE() for the individual models. In general, temperature changes are detectable before O_2 changes in around 64$\pm$11 % of the thermocline (yellow to brown colours in Fig. 7). As discussed in section 3.1, the anthropogenic O_2 signal emerges late or not at all in many low latitude regions, while the anthropogenic warming signal is emerging in most regions and typically early around the equator. However, there are also areas where anthropogenic deoxygenation is detectable earlier than anthropogenic warming in all models (green to blue colours in Fig. 7). These cover 35$\pm$11 % of the global thermocline in the nine models. They are mainly located in the mid latitudes, especially between ∼15° N and 30° N in the North Pacific, around Antarctica (including the Ross and Weddell Sea), along the Western Australian Current and the Pacific southern subtropical gyre region. Model results for the Atlantic subtropical gyres are mixed. Some models suggest O_2 changes to be detectable earlier than T changes in the subtropical gyres (HadGEM2 and the IPSL family), whereas in other models the O_2 signal does not even emerge.

The exact locations of relatively early emergence of O_2 differ across models. Hence, the regions where at least seven out of the nine models show consistently an earlier emergence of O_2 than T is smaller and amounts to 17 % of the global thermocline area. As shown in Fig. 6 (blue areas) the O_2 signal emerges consistently in at least seven models before the T signal in parts of the Pacific subtropical gyres, the Southern Ocean and the southeast Indian Ocean.

A mechanistic explanation of early or late emergence of the O_2 signal relative to the temperature signal is not straightforward as two ratios (S/N) are involved. Nevertheless, changes in apparent oxygen utilisation (Δ[-AOU]; Fig. 8) provide some insight into underlying mechanisms. We use Δ[-AOU] as a proxy for changes in water mass age and ventilation as noted in Sect 2.2.

It is striking that regions Regions with early emergence of anthropogenic O_2 compared to T show typically a decrease in [-AOU] in the future (Fig. 7 versus Fig. 8). , whereas regions with early emergence of T compared to O_2 show typically an increase in [-AOU]. For example, [-AOU] is decreasing in 77±8 % of the areas with early emergence of O_2, while only 22±8 % 22±8 % of these regions show an increase in [-AOU] (Fig. 9; blue). On the other hand, -AOUis increasing in most of the regions (62±12 %) In most regions where T is emerging before O_2 (Fig. 9; green). We interpret these results as follows. A decrease brown), [-AOU] is increasing (62±12 %). A decreasing trend in [-AOU] suggests that the ventilation of the thermocline is decreasing. Under global warming , this can be due to an increase in surface stratification (not shown; see also Gnanadesikan et al., 2007). In turn, the supply rate is indicative of a reduced ventilation induced by upper ocean warming and increased stratification (e.g. Capotondi et al., 2012). A more sluggish ventilation slows the supply of O_2 from the surface is decreasing, and consequently to the ocean interior. Consequently, thermocline [O_2] and [-AOU] are both decreasinginducing . This leads to a strong and thus early detectable anthropogenic deoxygenation. At the same time, a decrease in ventilation tends to slow down In addition, a more sluggish ventilation slows the penetration of the anthropogenic warming signal from the surface into the thermoclineto the interior, and similarly the penetration of the thermally driven O_2 signal ([$O_{2,sol}$]). The detection of these signals the temperature changes is thus delayed compared to AOU and this may partly counteract the effect of ventilation on the early detection of O_2. There are some exceptions to this mechanismrelationship between [-AOU] and the earlier emergence of O_2than T. For example, the GFDL models simulate an increase in -AOUO_2 change emerges before warming in the GFDL model around 30° S - and 120° W, while emerges before T. The although [-AOU] is increasing in this region. However, warming is emerging very late as the GFDL models simulate weak warming and even some cooling in this part of the thermocline (Fig. S4) , moderate T variability (Fig. S3) and, therefore, no or late emergence of the warming signal (Fig. 2)in this part of the thermocline. Thus, in this special case, the early emergence of O_2 relative to T is due to the absence of large warming in a region with ~~noticeable temperature~~ notable temperature internal variability.

Regions where the warming signal is detectable before the deoxygenation are typically associated with an increase in [-AOU]. Such increase counteracts the decrease in [$O_{2,sol}$], leading to ~~small~~ relatively smaller changes in [O_2], which are thus often not detectable. There are again a few exceptions. For example, the IPSL models simulate a decrease in [-AOU] in the northern North Pacific, but an earlier ToE for T than for O_2 in this region.

In summary, ~~temperature is in general~~ anthropogenic change in temperature is detectable earlier than anthropogenic ~~.~~ change in O_2 in most of the global ocean. However, there are large ocean regions where anthropogenic O_2 changes are detectable earlier in the thermocline in all models. Early emergence of deoxygenation relative to warming is typically detected in regions where thermocline ventilation and [-AOU] are decreasing over the simulation and late emergence of O_2 changes where ventilation and [-AOU] are increasing.

4 Discussion and conclusions

We analysed the time of emergence (ToE) of human-induced changes in oxygen (O_2) concentrations and temperature (T) in the thermocline (200 – 600 m) using nine Earth system model simulations of the climate over the historical and the future period. Using ToE as a metric allows for the assessment of anthropogenic changes by comparing the magnitude of the anthropogenic trend human-induced changes with the magnitude of natural variability. Both these magnitudes vary among models, e.g. , due to different climate sensitivities, and this metric was thus found to be relatively strongly internal variability. Both the magnitude of anthropogenic change and internal variability are model dependent, rendering the absolute year of ToE strongly model-dependent. Here, we introduced Evaluating differences in absolute year of ToE, however, can obscure important model agreement upon the spatial patterns and progression of emergence within a multi-variable framework. We therefore introduce a new metric, the relative ToE (ToE_{rel}), to better compare ToE across different models and variables. ToE_{rel} is computed by subtracting the global mean ToE from the ToE field. Absolute years of emergence are thus not considered by this metric and it only illustrates whether a signal emerges relatively early or late within a for a given model. We investigated whether anthropogenic T or O_2 changes emerge first and link patterns in of ToE(T)-ToE(O_2) to changes in apparent oxygen utilisation (Δ[-AOU]) and ventilation of the thermocline. In addition, we also identified the processes for earlier/later detection in O_2 changes compared to temperature changes.

This multi-model study relies only on results from only four different model families (GFDL-ESM, HadGEM2-CC, IPSL, MPI-ESM and CESM), applied in nine model configurations. All model configurations available from CMIP5 that provide 3-dimensional fields for O_2 and T for the control, historical and future-RCP8.5 scenario simulations have been incorporated into the analysis. Nevertheless, using a larger model ensemble would increase confidence in our results (Knutti and Sedláček, 2013).

A short-coming limitation of our study is that all the Earth system models included have a relatively coarse resolution for simulating the complex processes in the O_2 minimum zones (Margolskee et al., 2019). Perhaps not surprising, Earth System models diverge in projecting physical and biogeochemical changes in these regions (Brandt et al. 2015; Cabré et al. 2015). Some models used in this study project a large increase in [-AOU] (Fig. 8) and considerable warming (Fig. S6) in the eastern tropical Atlantic, likely indicative of an increasing ventilation (Gnanadesikan et al., 2007a reduced upwelling (Gnanadesikan et al., 2007). Observations show a decrease in O_2 and an expansion of hypoxia in the tropics (Stramma et al., 2008, 2012) over recent decades, contradicting the long-term projections from some models. However, these observed trends in the tropics may also be a result of natural variability acting on multi-decadal timescales associated with the Pacific Decadal Oscillation.

Comparing ToE estimates from different studies is delicate due to the model and method dependencies of ToE. Hameau et al. (2019) Alth the generic definition of ToE is under consensus, the methodologies applied to estimate ToE differ in the published literature as mentioned in the introduction (e.g. IPCC (2019)). Depending on the spatial and temporal scale of a given variable, the threshold for which emergence is defined and the reference period applied, the absolute value of ToE can differ. In addition, the ToE also depends on the definition of the background variability, here acting as noise (Hameau et al., 2019). Estimating the background noise as the standard deviation (SD) of the internal chaotic variability from the control simulation (Frölicher et al., 2016), or as the SD of the variability from the industrial period (after removing anthropogenic trends; Keller et al., 2015; Henson et al., 2016) result in earlier ToE for both O_2 and T as when estimating the noise from the total (internal and externally-forced) natural variability over the last millennium. Yet, the finding that anthropogenic O_2 change emerges before anthropogenic warming in large ocean regions is robust across investigated choices. The anthropogenic signal is frequently computed as a linear trend over a few decades (Rodgers et al., 2015; Henson et al., 2017; Tjiputra et al., 2018). However, the resulting slope depends on the time window used to calculate the linear trend. Hameau et al. (2019) use a low-pass filtered output to estimate the signal. They showed that ideally the noise (N) component of ToE should be estimated from simulations that include natural variability forced by explosive volcanic eruptions and changes in total solar irradiance, especially when assessing regional to global scale ToE estimates. However, these authors also find that on a grid cell scale, internal natural variability is typically the dominant contribution to overall natural variability during the last millennium. Therefore, estimating the noise from control simulations that include internal natural variability only, as done in this study, appears justified.

Another limitation of our study lies in the assumption that the anthropogenic signal emerges from interannual to multi-decadal internal variability. The anthropogenic signal S and the noise N is estimated by smoothing the model output with a multi-decadal spline filter. Any potential natural centennial variations are retained in the signal S and removed from the noise N. Results from a forced simulation over the past millennium with CESM1.0 show that potential biases in ToE arising from the neglect of long-term natural variability are small for this model (Hameau et al., 2019). However, our multi-model analysis reveals centennial variations in some grid cells and models causing multiple emergence of the signal from the noise (Fig. A2). This may bias the detection of the anthropogenic signal towards early emergence. Here, we constrained detection to partly circumvent problems with re-emerging signals; we require that the trend of the signal at the time of emergence must have the same sign as the change between the last and first 30 model years. Re-emerging signals are found in only a few grid cells, except in HadGEM2, and centennial natural variability appears to play a minor role in these simulations. We expect therefore that our estimates of ToE are reliable for the model ensemble.

Published studies addressing the detection of anthropogenic ocean warming focus on the sea surfacetemperaturetemperature at sea surface. To our knowledge, only a single study Hameau et al. (2019) using output from a single model is assessing ToE(T) in the thermocline. Yet, the thermocline is habitat for many fish and other species. Warming in combination with other stressors, such as deoxygenation, ocean acidification and hypocapnia, may reduce marine life habitat suitability and extent the habitat suitability of marine ecosystems in a future climate (e.g. Deutsch et al., 2015; Gattuso et al., 2015; Breitburg et al., 2018; Cheung et al., 2018). Multi-tracer analyses contribute to a better understanding of the potential impact on marine ecosystems in a changing ocean.

We find that thermocline anthropogenic warming emerges first in low latitudes, followed by the Southern Ocean and the high northern latitudes. No emergence is detected in parts of the subtropical gyres of the Pacific and Indian Ocean. The rapid emergence at low latitudes is explained by the small internal natural variability, but moderate to strong warming signals. Exceptions are the subtropical gyres in the Atlanticand the eastern equatorial Pacific, where it takes approximately two additional decades to detect the temperature changes, mainly because of the relatively large internal natural variability there. The warming in mid- to high latitude thermocline emerges approximately 60 to 80 years later than in low latitudes. No emergence is simulated for the Pacific and Indian subtropical gyres, because the changes in temperature are relatively small and the internal natural variability relatively high there (in accordance with Hameau et al., 2019). For comparison, surface temperature changes emerge at first in low latitudes and then in midlatitudes (Henson et al., 2017).

The time of emergence spatial pattern of thermocline oxygen changes is almost opposite to the one of temperature. Rapid emergence for O_2 is simulated at midlatitudes, whereas low latitudes generally do not experience emergence of the O_2 signal by the end of the 21st century (Rodgers et al., 2015; Frölicher et al., 2016; Long et al., 2016). Even though the internal natural Although internal variability is low in the tropical regions, the O_2 signal does not emerge from the noise, because the changes projected by the models are even smaller. by 2100. This is because the projected changes are also small. This is due to the opposite responses of O_2 components. The thermal component is simulated to decrease (due to temperature increase), but [-AOU] is on average projected to increase, counteracting the $O_{2,sol}$ trend (Frölicher et al., 2009; Cocco et al., 2013; Bopp et al., 2017). Some regions show similar relative ToE but for different reasons. For example, in the North Pacific subtropical gyre and the Southern Ocean, both the oxygen depletion and the natural variability are relatively strong. In the Arabian Sea, natural internal variability and anthropogenic response are both rather weak. Nevertheless, the S/N ratio results in very similar relative ToE for all these regions.

Comparing ToE across models is not straightforward. For example, the The transient climate response of the individual models and therefore the ocean heat uptake, thermocline warming and deoxygenation can be highly different substantially differ among models (Bopp et al., 2013). In addition, the simulated internal natural variability largely differs The simulated internal variability also differs considerably across models (e.g. Resplandy et al., 2015; Frölicher et al., 2016). The ToEs computed from CESM1.0 projections, for example, shows very different absolute ToE values for oxygen and temperature compared to other differ by many decades in absolute values from other CMIP5 models, mostly due to a very weak internal natural variability(Hameau et al., 2019). We partly resolved these inter-model discrepancies by introducing variability. Nijsse et al. (2019) suggest that the magnitude of simulated decadal variability and climate sensitivity might be correlated. They suggest that models with a high climate sensitivity tend to simulate a high decadal variability. This may imply a compensation between the simulated signal and noise on the decadal scale. To extract valuable insights as to the relative spatial and temporal features of emergence across models and variables, we introduced a new metric, the relative time of emergence. By normalising the ToE using the global globally averaged ToE as reference allows for a more direct comparison with the other models. As a result, the patterns and time of emergence of anthropogenic changes in O_2 and warming in CESM1.0

are more coherent with the other models for ToE$_{\mathrm{rel}}$ than for the ~~raw (absolute)~~ absolute ToE.

Following Hameau et al. (2019), we compared the ToE(T) with the ToE(~~)~~ O_2) in nine models. We find that ~~in most of the thermocline, the anthropogenic increase in temperature is expected to emerge before anthropogenic changes. However, in 35±11 %~~ the anthropogenic decline in O_2 emerges before anthropogenic warming in a significant part of the thermocline. On average across the nine models, an area covering 35±11 % of the global ~~ocean the signal emerges before the temperature signal. In the Pacific subtropical gyres, the Southern Ocean and the West Australian Current, the~~ thermocline shows emergence in O_2 change before temperature change. Yet, the exact locations of these patterns differ across models. Only 17 % of the global thermocline show agreement (seven out of the nine models) on earlier emergence of O_2 ~~signal emerges before the temperature signal in all 9 models.~~ changes prior to T changes, Thus, our multi-model analysis confirms earlier findings using output from a single model only (Hameau et al., 2019). The early emergence of O_2 suggests that the monitoring of biogeochemical variables would be particularly useful to detect early signals of anthropogenic change in the ocean interior (Joos et al., 2003). Multi-tracer observations of both physical and biogeochemical variables may enable an earlier detection of potential changes than temperature-only data (Keller et al., 2015) in specific regions and for specific processes.

Hameau et al. (2019) established a direct link between the early emergence in O_2 with a slow down of ventilation. A weaker ventilation leads to a decrease in [-AOU], and therefore to a reduction in O_2, with a minor role for organic matter export changes in their simulation. We used [-AOU] as a ventilation age proxy for our model ensemble and concluded that the slow down of the ventilation induces O_2 changes to be detectable before T changes in many regions. A slower ventilation seems to shift the balance between O_2 supply from the surface and O_2 consumption by organic matter remineralisation. Moreover, a more stratified upper ocean delays the propagation of the temperature signal from the surface into the subsurface waters. Note that the exact locations of early O_2 emergence and reductions in [-AOU] and ventilation diverge ~~between~~ among the models. This is partly due to model biases in terms of ocean dynamics. In addition, the use of depth coordinates to define a thermocline layer from 200 – 600 m may lead in our analysis to the inclusion of different water masses for different models. Another approach would be to perform the analysis on isopycnal levels instead on depth levels.

To conclude, normalising ToE across models (relative ToE) or estimating ToE in relation to another variable (ToE(T) - ToE(O_2)), reduces the multi-model spread arising from method and model dependencies. We find that in about 35 % of the thermocline anthropogenic O_2 depletion emerges before anthropogenic warming. This relative early emergence of O_2 is linked to a more sluggish ventilation of these subsurface waters under global warming. Our study also suggests that temperatures in the thermocline have already left the bounds of ~~natural~~ internal variability in much of the tropical ocean and that temperatures will have left these bounds in most of the thermocline by 2100 under unabated global warming.

Data availability. The CMIP5 simulations are available on https://esgf-node.ipsl.upmc.fr. The CESM1.0 simulations are available upon request.

5 Figures

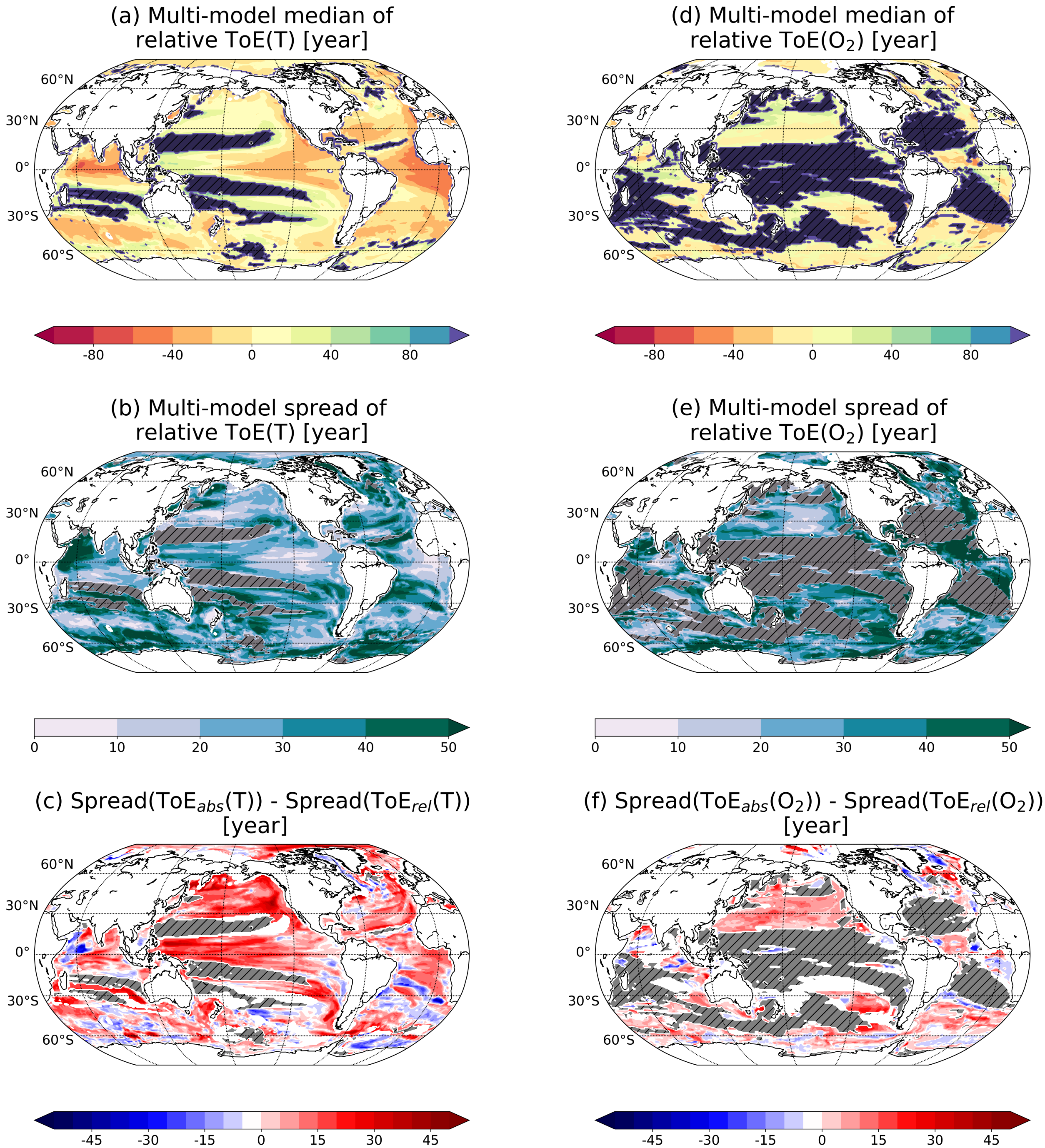

Figure 1. Multi-model median (top panel) and spread (middle panel) of relative ToE for temperature (left column) and dissolved oxygen (right column) for the thermocline (200 – 600 m). The spread is computed as the interquartile range. ~~Multi-model spread of~~ Difference (lower panel)between the multi-model spread of absolute ToE estimates with the multi-model spread of relative ToE estimates for (c) temperature and (f) dissolved oxygen. The hatched areas show regions with no emergence for at least 3 models. For temperature (oxygen), the relative ToE estimates are shown for each model in Fig. 2 (3) and the absolute estimates in Fig. S1 (S2).

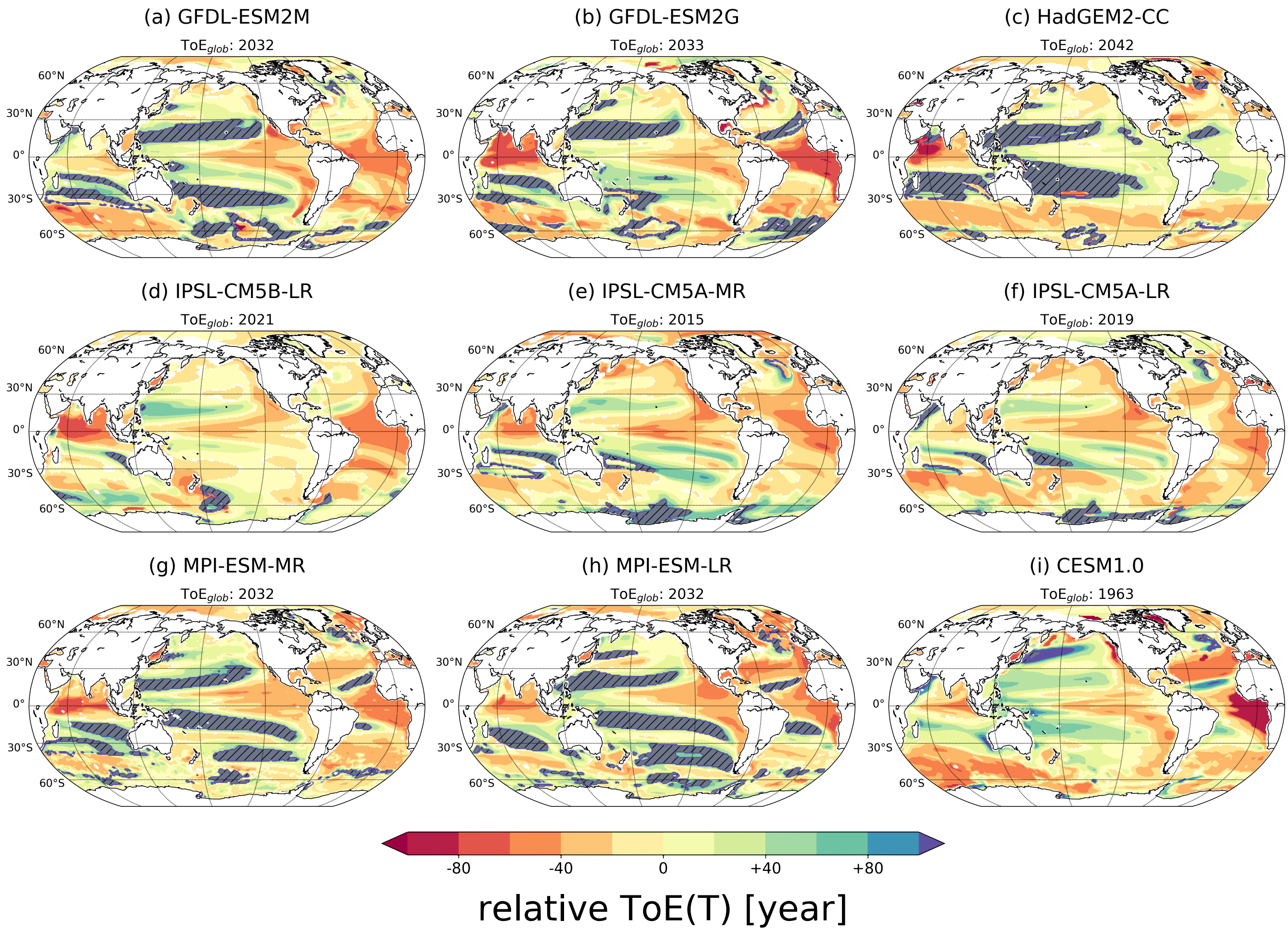

Figure 2. Time of Emergence (ToE) of T in the thermocline (200 – 600 m) relative to the averaged ToE in that layer for each simulation. The hatched areas show regions with no emergence by the end of the 21st century.The values of the global average ToE, ToE$_{glob}$, are given above each panel. The absolute ToE estimates are shown in Fig. S1.

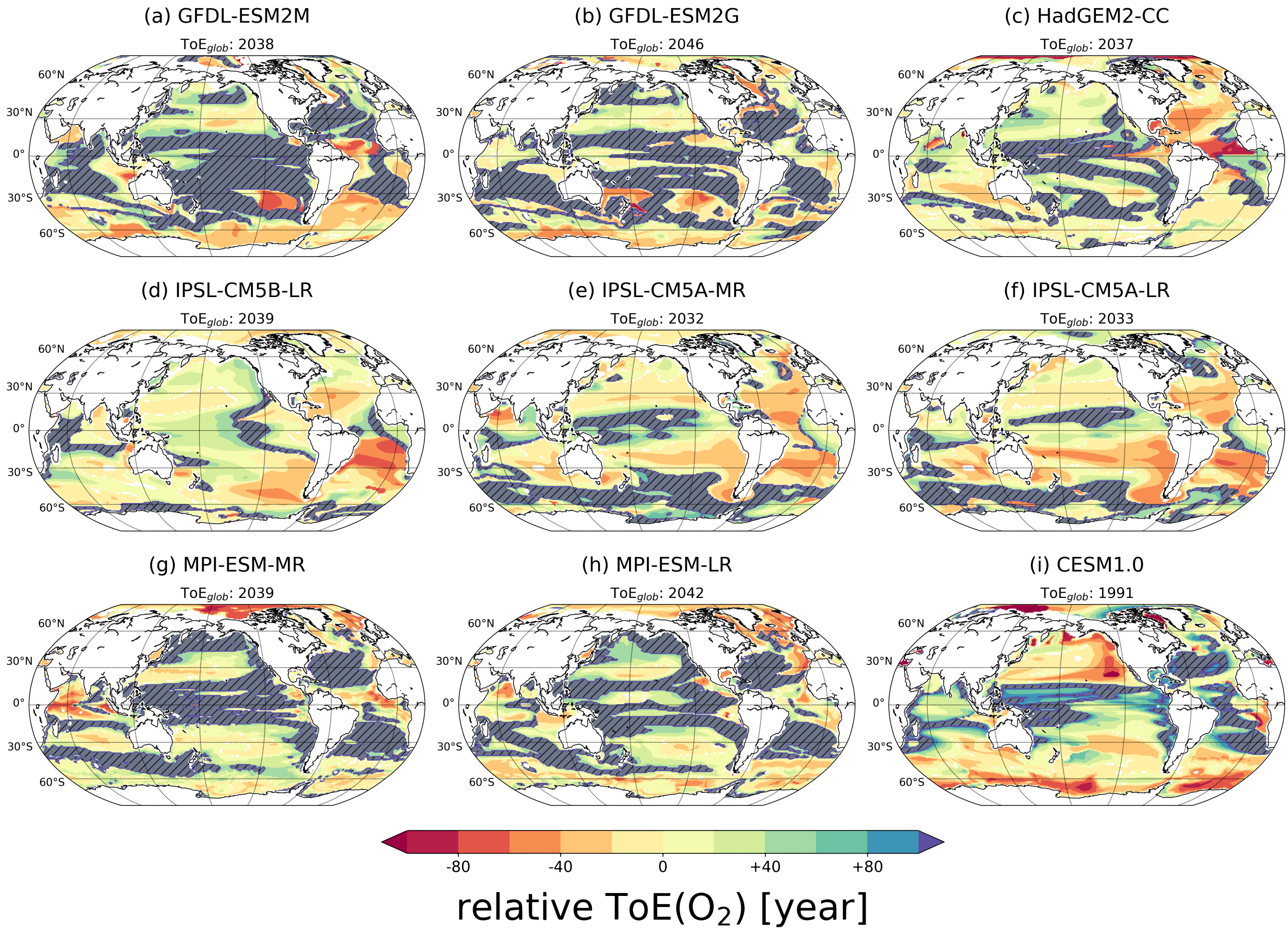

Figure 3. Time of Emergence (ToE) of O_2 in the thermocline (200 – 600 m) relative to the averaged ToE in that layer for each simulation. The hatched areas show regions with no emergence by the end of the 21st century. The absolute ToE estimates are shown in Fig. S2. The global average ToE, ToE$_{glob}$, is shown for each model

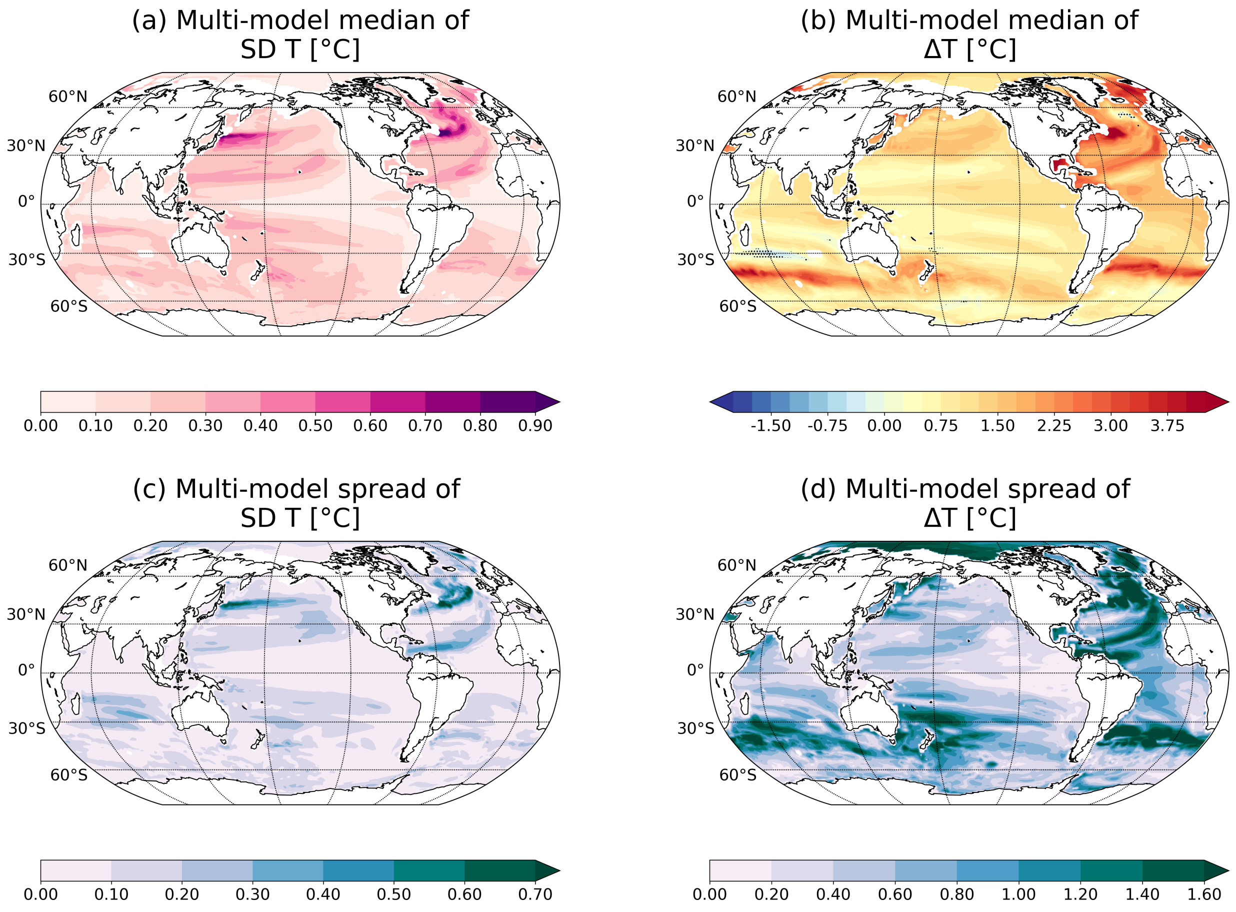

Figure 4. Median (top panels) and spread (bottom panels) of multi-model natural variability (standard deviation of control simulation; left panels) and changes by the end of the 21st century (right panels) of ocean temperature between 200 and 600 m. The individual responses for each model are shown in Figs. S3 and S4.

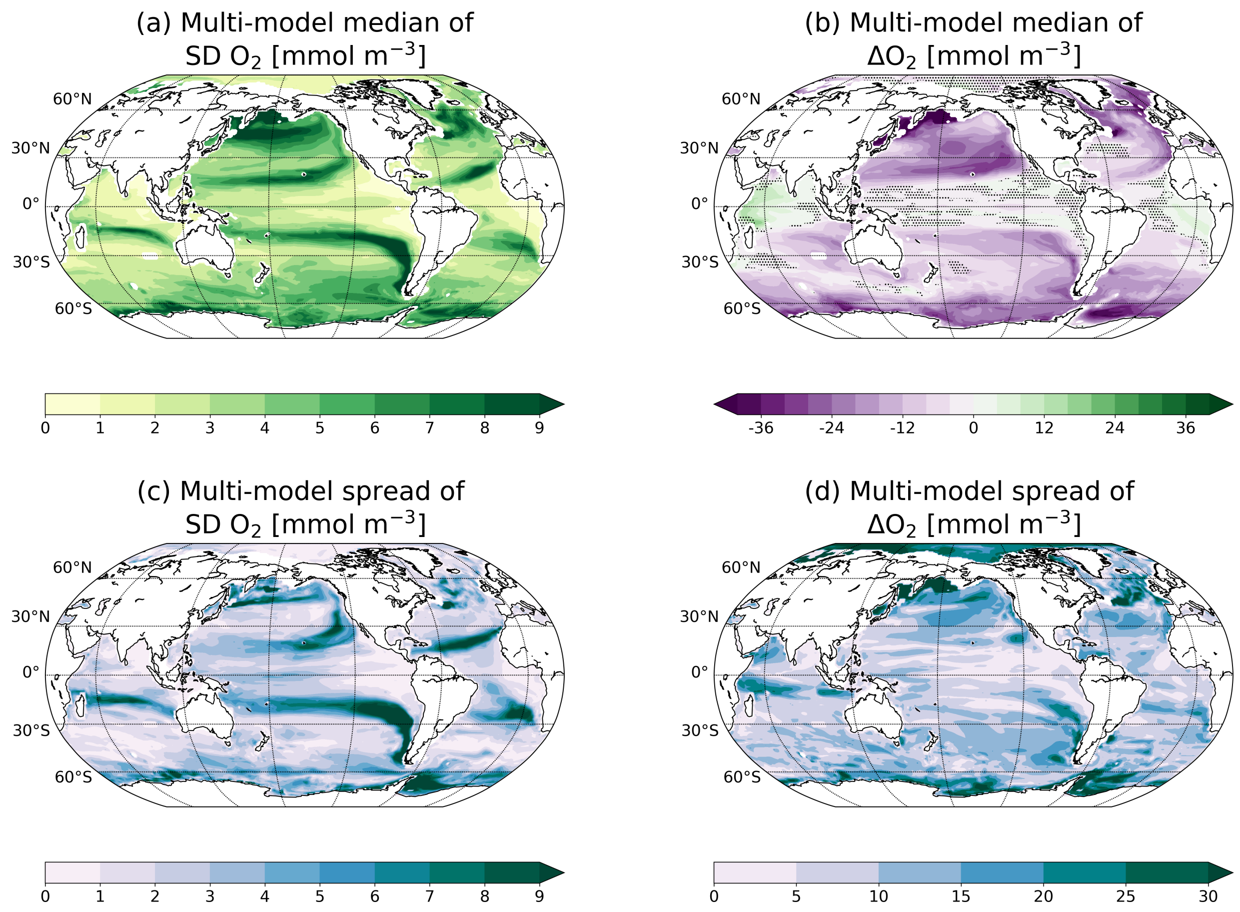

Figure 5. Median (top panels) and spread (bottom panels) of multi-model natural variability (standard deviation of control simulations; left panels) and changes by the end of the 21st century (right panels) of O_2 between 200 and 600 m. The hatched areas in panel b show regions where at least 70 % of the models do not agree on ΔO_2 sign. The individual responses for each model are shown in Figs. S5 and S6.

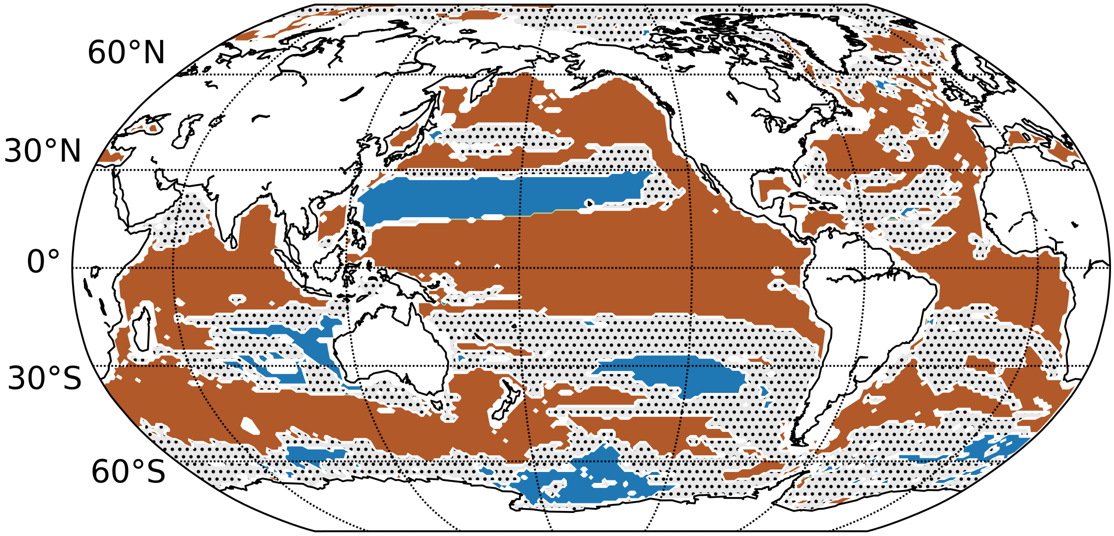

Figure 6. Summary map showing the regions where oxygen changes emerge before temperature changes (blue areas; ∆ToE>0) and where temperature changes emerge before oxygen changes (brown areas; ∆ToE<0) for at least seven out of nine models. The dashed areas show the regions where more than three models differ in the sign of ∆ToE.

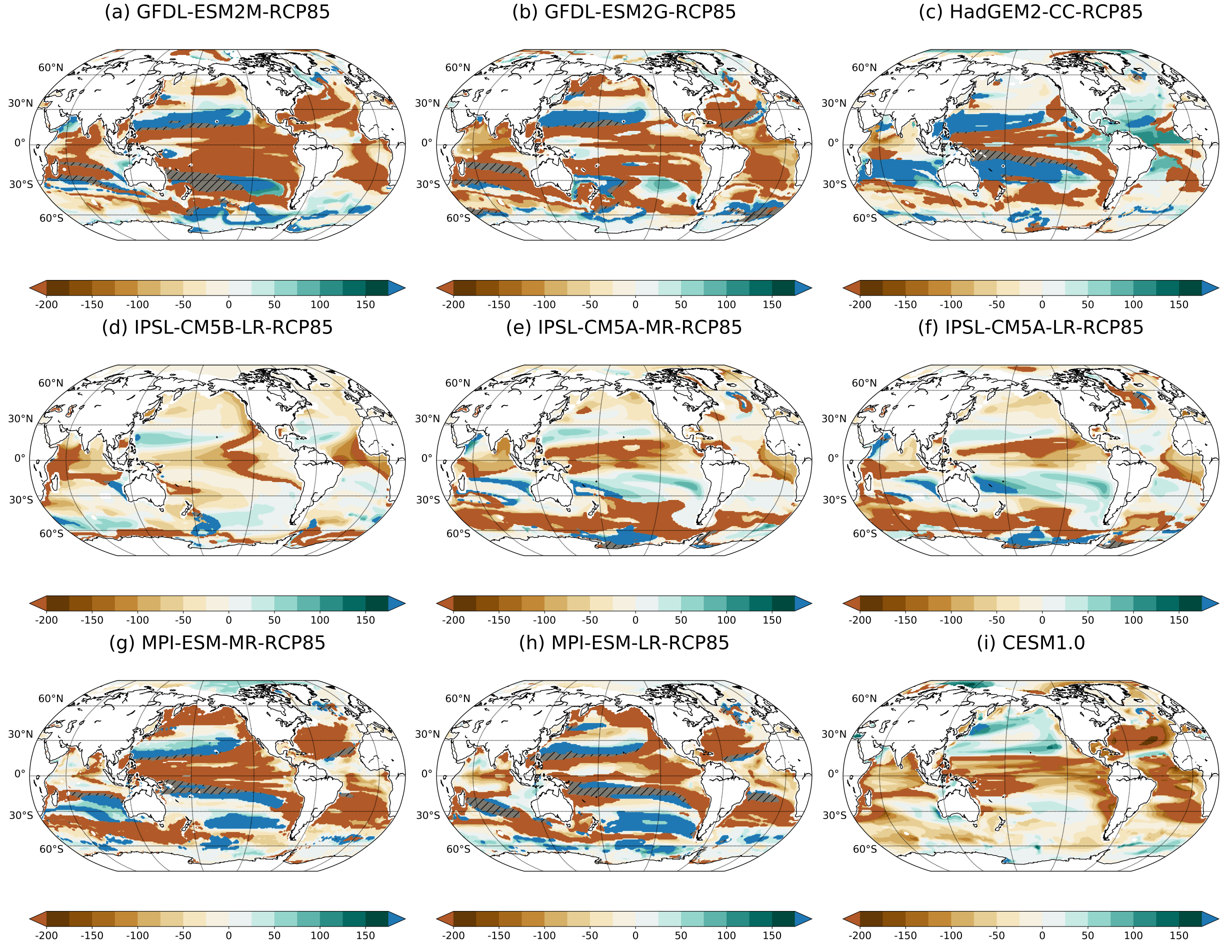

Figure 7. ToE(T) minus ToE(O_2) for each simulation in the thermocline. Blueish colours indicate earlier emergence of oxygen. Brownish colours indicate earlier emergence of temperature. The saturated colours mean that one of the variables has not emerged by 2099. No emergence in both T and O_2 are shown by the hatched areas.

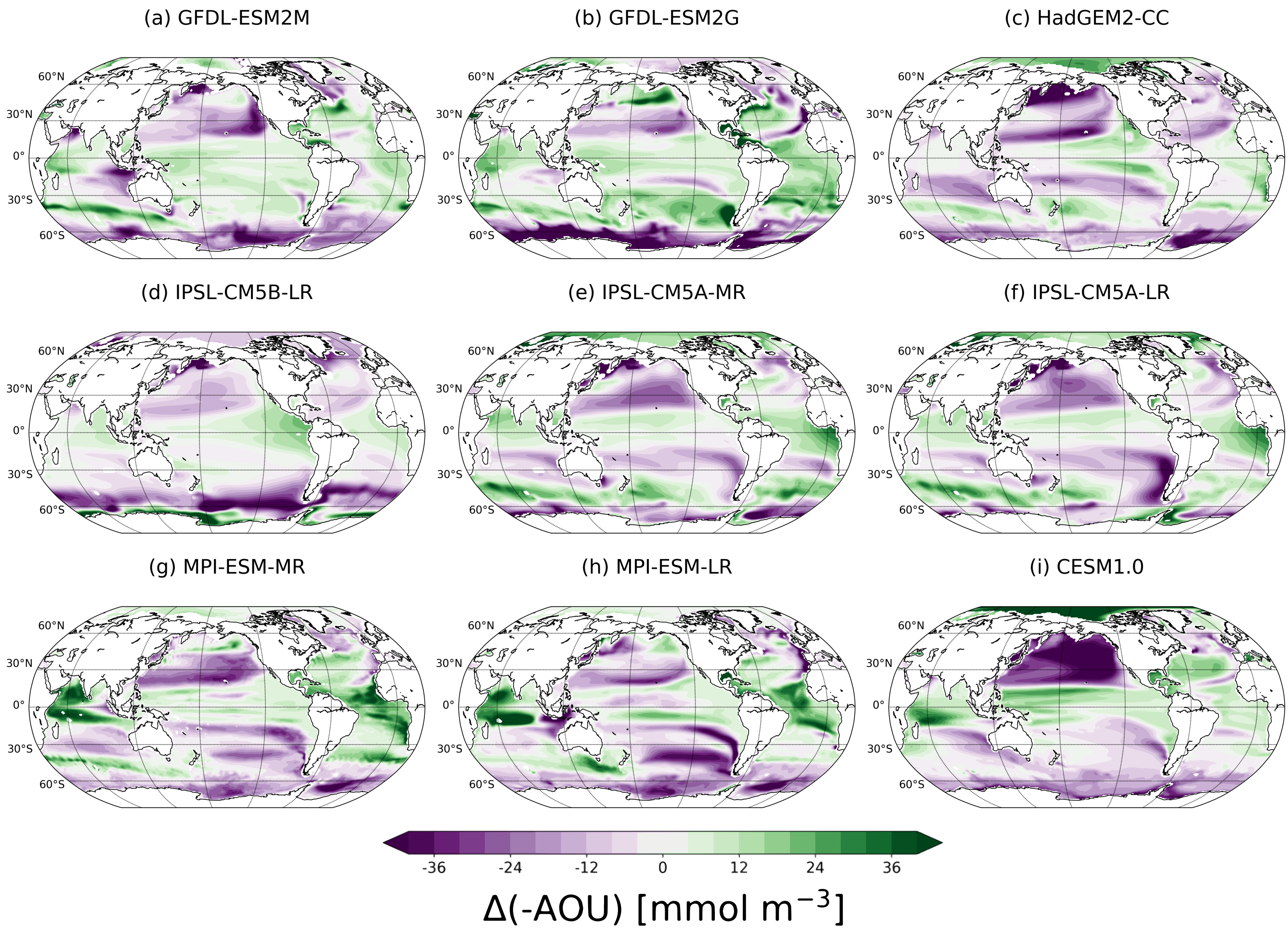

Figure 8. Anthropogenic changes ((2070-2099 CE) minus (1861-1959 CE)) in [-AOU] in the thermocline for each model.

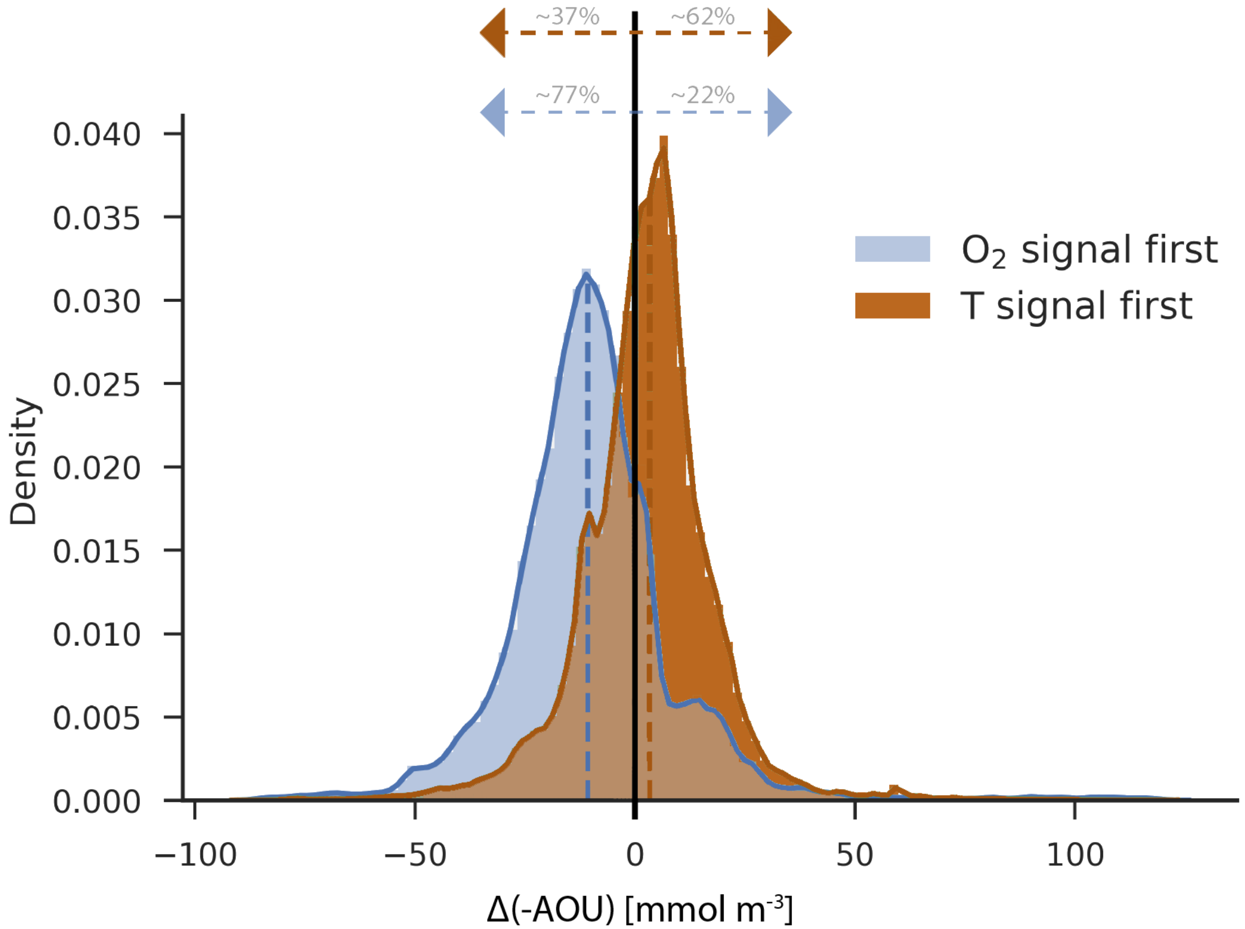

Figure 9. Density distribution of [-AOU] changes by 2099 for the grid points where the O_2 signal emerges first (blue) and where the temperature signal emerges first (~~green~~brown) in the thermocline for the ensemble of 9 models. Each distribution is centred around the median (dashed blue: -10.8 mmol m^{-3}; dashed ~~green~~brown: 3.3 mmol m^{-3}).

Table 1. Overview of the Earth system models used in this study, their configurations and vertical and approximated horizontal resolutions.

Earth system model	Physical ocean model	Biogeochemical ocean model	Vertical and horizontal ocean resolution
CESM1.0 Hurrell et al. (2013)	POP2 (Smith et al., 2010; Danabasoglu et al., 2011)	BEC (Moore et al., 2002, 2004)	60 levels ~1° x 1°
GFDL-ESM2M GFDL-ESM2G Dunne et al. (2012, 2013)	MOM4p1 (Griffies et al., 2011) GOLD (Hallberg, 1997)	TOPAZ2 (Dunne et al., 2013)	50 levels ~1° x 1°
HadGEM2-CC Collins et al. (2011)	HadGEM2 (Collins et al., 2011)	HadOCC (Palmer and Totterdell, 2001)	40 levels ~2° x 2°
IPSL-CM5A-LR IPSL-CM5A-MR IPSL-CM5B-LR Dufresne et al. (2013)	OPA (Madec et al., 2017)	PISCES (Aumont and Bopp, 2006)	31 levels ~2° x 2°
MPI-ESM-LR MPI-ESM-MR Giorgetta et al. (2013)	MPIOM (Jungclaus et al., 2013)	HAMOCC5.2 (Ilyina et al., 2013)	40 levels ~1.5° x 1.5° 40 levels ~0.4° x 0.4°

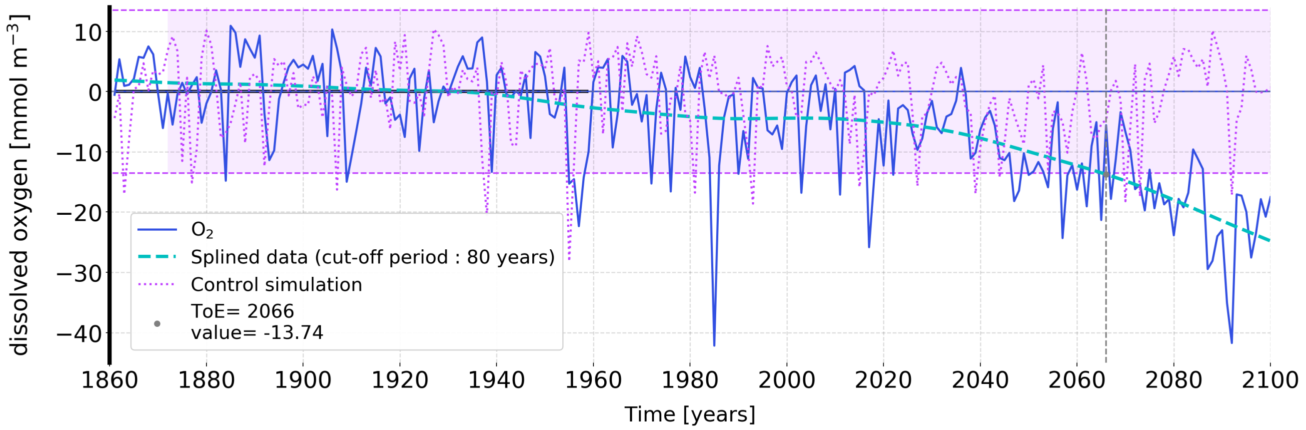

Figure A1. Illustration of the ToE method by using simulated oxygen concentration of the MPI-ESM-LR model averaged over 200 – 600 m in one grid cell in the Western North Pacific. The dark blue line shows that temporal evolution of oxygen anomalies from 1860 to 2099 relative to pre-industrial concentrations (1860 – 1959). The two pink lines indicate the magnitude of the noise, N, which is computed as two standard deviations from annual output of the control simulation. The anthropogenic signal, S, is the simulated oxygen concentration over 1860 to 2090 from the forced simulation splined with a 80 year cut-off period (dashed cyan). Here, the resulting ToE is the intersection between the lower limit of the noise and and the splined signal (vertical dashed line).

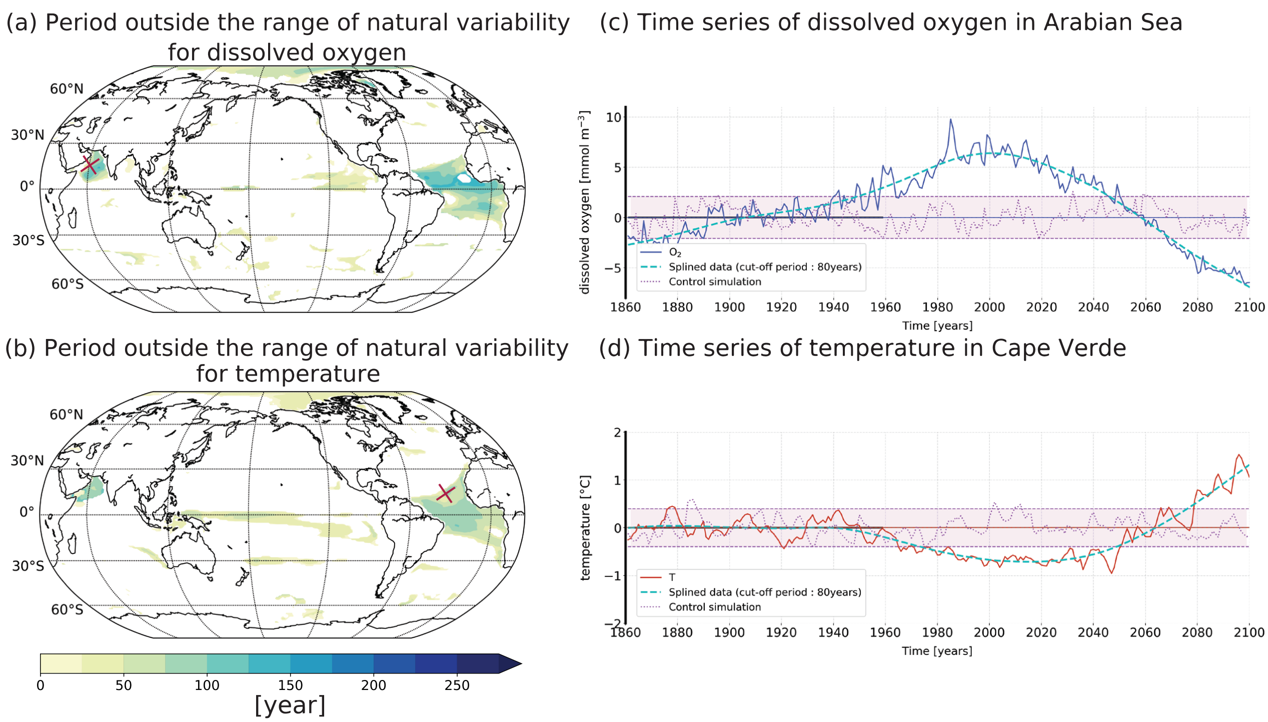

Figure A2. Period outside the range of natural variability for (a) oxygen concentration and (b) temperature in the thermocline for the model HadGEM2. The time series show two example of the temporal evolution of the oxygen concentration (c) and temperature (d) for a single grid point (red crosses in the left panels: Arabian Sea (c) and equatorial Atlantic (d)).

Author contributions. All authors contributed to the discussion and the writing of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. A. Hameau, F. Joos and T. Frölicher thank the Oeschger Center for Climate Change Research, the Swiss National Science Foundation (#200020_172476 and PP00P2_ 170687)~~and the H2020 project COMFORT~~ , and the European Union for financial support and the CSCS Swiss National Supercomputing Center for computing resources. This publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820989 (project COMFORT, Our common future ocean in the Earth system – quantifying coupled cycles of carbon, oxygen, and nutrients for determining and achieving safe operating spaces with respect to tipping points). The work reflects only the authors' view; the European Commission and their executive agency are not responsible for any use that may be made of the information the work contains. We thank C. Raible and F. Lehner for providing CESM output. We also thank the World Climate Research Programme's Working Group on coupled Modelling, which is responsible for CMIP5, and the climate modelling groups for producing and making available their model output. We thank the two anonymous referees and Sarah Schlunegger for their effort and time and their comments.

References

Aumont, O. and Bopp, L.: Globalizing results from ocean in situ iron fertilization studies, Global Biogeochemical Cycles, 20, https://doi.org/10.1029/2005GB002591, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GB002591, 2006.

Battaglia, G. and Joos, F.: Hazards of decreasing marine oxygen: the near-term and millennial-scale benefits of meeting the Paris climate targets, Earth System Dynamics, 9, 797, 2018.

Bilbao, R. A. F., Gregory, J. M., Bouttes, N., Palmer, M. D., and Stott, P.: Attribution of ocean temperature change to anthropogenic and natural forcings using the temporal, vertical and geographical structure, Climate Dynamics, https://doi.org/10.1007/s00382-019-04910-1, https://doi.org/10.1007/s00382-019-04910-1, 2019.

Bopp, L., Le Quéré, C., Heimann, M., Manning, A. C., and Monfray, P.: Climate-induced oceanic oxygen fluxes: Implications for the contemporary carbon budget, Global Biogeochemical Cycles, 16, https://doi.org/10.1029/2001GB001445, http://onlinelibrary.wiley.com/doi/10.1029/2001GB001445/abstract, 2002.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, Biogeosciences, 10, 6225–6245, https://doi.org/10.5194/bg-10-6225-2013, http://www.biogeosciences.net/10/6225/2013/, 2013.

Bopp, L., Resplandy, L., Untersee, A., Mezo, P. L., and Kageyama, M.: Ocean (de)oxygenation from the Last Glacial Maximum to the twenty-first century: insights from Earth System models, Phil. Trans. R. Soc. A, 375, https://doi.org/10.1098/rsta.2016.0323, http://rsta.royalsocietypublishing.org/content/375/2102/20160323, 2017.

Brandt, P., Bange, H. W., Banyte, D., Dengler, M., Didwischus, S.-H., Fischer, T., Greatbatch, R. J., Hahn, J., Kanzow, T., Karstensen, J., Körtzinger, A., Krahmann, G., Schmidtko, S., Stramma, L., Tanhua, T., and Visbeck, M.: On the role of circulation and mixing in the ventilation of oxygen minimum zones with a focus on the eastern tropical North Atlantic, Biogeosciences, 12, 489–512, https://doi.org/10.5194/bg-12-489-2015, http://www.biogeosciences.net/12/489/2015/, 2015.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science, 359, https://doi.org/10.1126/science.aam7240, http://science.sciencemag.org/content/359/6371/eaam7240, 2018.

Cabré, A., Marinov, I., Bernardello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, https://doi.org/10.5194/bg-12-5429-2015, http://www.biogeosciences.net/12/5429/2015/, 2015.

Capotondi, A., Alexander, M. A., Bond, N. A., Curchitser, E. N., and Scott, J. D.: Enhanced upper ocean stratification with climate change in the CMIP3 models, Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2011JC007409, http://onlinelibrary.wiley.com/doi/10.1029/2011JC007409/abstract, citation Key Alias: capotondi_enhanced_2012-1, 2012.

Chen, X. and Tung, K.-K.: Global surface warming enhanced by weak Atlantic overturning circulation, Nature, 559, 387–391, https://doi.org/10.1038/s41586-018-0320-y, https://www.nature.com/articles/s41586-018-0320-y, 2018.

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., and Zhu, J.: Improved estimates of ocean heat content from 1960 to 2015, Science Advances, 3, https://doi.org/10.1126/sciadv.1601545, https://advances.sciencemag.org/content/3/3/e1601545, 2017.

Cheng, L., Abraham, J., Hausfather, Z., and Trenberth, K. E.: How fast are the oceans warming?, Science, 363, 128–129, https://doi.org/10.1126/science.aav7619, https://science.sciencemag.org/content/363/6423/128, 2019.

Cheung, W. W., Jones, M. C., Reygondeau, G., and Frölicher, T. L.: Opportunities for climate-risk reduction through effective fisheries management, Global change biology, 24, 5149–5163, 2018.

Cheung, W. W. L., Dunne, J., Sarmiento, J. L., and Pauly, D.: Integrating ecophysiology and plankton dynamics into projected maximum fisheries catch potential under climate change in the Northeast Atlantic, ICES Journal of Marine Science, 68, 1008–1018, https://doi.org/10.1093/icesjms/fsr012, https://academic.oup.com/icesjms/article/68/6/1008/699298, 2011.

Christensen, J., Hewitson, B., A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R.K. Kolli, W.-T. Kwon, R. Laprise, V. Magaña Rueda, L. Mearns, C.G. Menéndez, J. Räisänen, A. Rinke, A. Sarr, and P. Whetton: Regional Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.

Cocco, V., Joos, F., Steinacher, M., Froelicher, T. L., Bopp, L., Dunne, J., Gehlen, M., Heinze, C., Orr, J., Oschlies, A., Schneider, B., Segschneider, J., and Tjiputra, J.: Oxygen and indicators of stress for marine life in multi-model global warming projections, Biogeosciences, 10, 1849–1868, https://doi.org/10.5194/bg-10-1849-2013, wOS:000317010600040, 2013.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geoscientific Model Development, 4, 1051–1075, https://doi.org/https://doi.org/10.5194/gmd-4-1051-2011, https://www.geosci-model-dev.net/4/1051/2011/, 2011.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 Ocean Component, Journal of Climate, 25, 1361–1389, https://doi.org/10.1175/JCLI-D-11-00091.1, https://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00091.1, 2011.

Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H.-O., and Huey, R. B.: Climate change tightens a metabolic constraint on marine habitats, Science, 348, 1132–1135, https://doi.org/10.1126/science.aaa1605, http://science.sciencemag.org/content/348/6239/1132, 2015.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, Journal of Climate, 25, 6646–6665, https://doi.org/10.1175/JCLI-D-11-00560.1, https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-11-00560.1, 2012.

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J., and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation

Characteristics, Journal of Climate, 26, 2247–2267, https://doi.org/10.1175/JCLI-D-12-00150.1, https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-12-00150.1, 2013.

Enting, I. G.: On the use of smoothing splines to filter CO2 data, Journal of Geophysical Research: Atmospheres, 92, 10 977–10 984, https://doi.org/10.1029/JD092iD09p10977, http://onlinelibrary.wiley.com/doi/10.1029/JD092iD09p10977/abstract, 1987.

Frame, D., Joshi, M., Hawkins, E., Harrington, L. J., and de Roiste, M.: Population-based emergence of unfamiliar climates, Nature Climate Change, 7, 407–411, https://doi.org/10.1038/nclimate3297, https://www.nature.com/articles/nclimate3297, 2017.

Frölicher, T. L. and Paynter, D. J.: Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales, Environmental Research Letters, 10, https://doi.org/10.1088/1748-9326/10/7/075002, http://stacks.iop.org/1748-9326/10/i=7/a=075002, 2015.

Frölicher, T. L., Joos, F., Plattner, G.-K., Steinacher, M., and Doney, S. C.: Natural variability and anthropogenic trends in oceanic oxygen in a coupled carbon cycle–climate model ensemble, Global Biogeochemical Cycles, 23, https://doi.org/10.1029/2008GB003316, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008GB003316, 2009.

Frölicher, T. L., Rodgers, K. B., Stock, C. A., and Cheung, W. W. L.: Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors, Global Biogeochemical Cycles, 30, 1224–1243, https://doi.org/10.1002/2015GB005338, http://onlinelibrary.wiley.com/doi/10.1002/2015GB005338/full, 2016.

Garcia, H. E. and Gordon, L. I.: Oxygen solubility in seawater: Better fitting equations, Limnology and Oceanography, 37, 1307–1312, https://doi.org/10.4319/lo.1992.37.6.1307, http://onlinelibrary.wiley.com/doi/10.4319/lo.1992.37.6.1307/abstract, 1992.

Gattuso, J.-P., Magnan, A., Billé, R., Cheung, W. W. L., Howes, E. L., Joos, F., Allemand, D., Bopp, L., Cooley, S. R., Eakin, C. M., Hoegh-Guldberg, O., Kelly, R. P., Pörtner, H.-O., Rogers, A. D., Baxter, J. M., Laffoley, D., Osborn, D., Rankovic, A., Rochette, J., Sumaila, U. R., Treyer, S., and Turley, C.: Contrasting futures for ocean and society from different anthropogenic CO2 emissions scenarios, Science, 349, https://doi.org/10.1126/science.aac4722, http://science.sciencemag.org/content/349/6243/aac4722, 2015.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/jame.20038, 2013.

Gnanadesikan, A., Russell, J. L., and Zeng, F.: How does ocean ventilation change under global warming?, Ocean Science, 3, 43–53, https://hal.archives-ouvertes.fr/hal-00298310, 2007.

Gnanadesikan, A., Dunne, J. P., and John, J.: Understanding why the volume of suboxic waters does not increase over centuries of global warming in an Earth System Model, Biogeosciences, 9, 1159–1172, https://doi.org/https://doi.org/10.5194/bg-9-1159-2012, https://www.biogeosciences.net/9/1159/2012/, 2012.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., Gnanadesikan, A., Hurlin, W. J., Lee, H.-C., Liang, Z., Palter, J. B., Samuels, B. L., Wittenberg, A. T., Wyman, B. L., Yin, J., and Zadeh, N.: The GFDL CM3 Coupled Climate Model: Characteristics of the Ocean and Sea Ice Simulations, Journal of Climate, 24, 3520–3544, https://doi.org/10.1175/2011JCLI3964.1, https://journals.ametsoc.org/doi/full/10.1175/2011JCLI3964.1, 2011.

Hallberg, R.: Some aspects of the circulation in ocean basins with isopycnals intersecting sloping boundaries., University of Washington, https://books.google.ch/books?id=sdbENwAACAAJ, 1997.

Hameau, A., Mignot, J., and Joos, F.: Assessment of time of emergence of anthropogenic deoxygenation and warming: insights from a CESM simulation from 850 to 2100 CE, Biogeosciences, 16, 1755–1780, https://doi.org/10.5194/bg-16-1755-2019, https://www.biogeosciences.net/16/1755/2019/, 2019.

Han, W., Vialard, J., McPhaden, M. J., Lee, T., Masumoto, Y., Feng, M., and de Ruijter, W. P.: Indian Ocean Decadal Variability: A Review, Bulletin of the American Meteorological Society, 95, 1679–1703, https://doi.org/10.1175/BAMS-D-13-00028.1, https://journals.ametsoc.org/doi/full/10.1175/BAMS-D-13-00028.1, 2014.

Hauri, C., Gruber, N., McDonnell, A. M. P., and Vogt, M.: The intensity, duration, and severity of low aragonite saturation state events on the California continental shelf, Geophysical Research Letters, 40, 3424–3428, https://doi.org/10.1002/grl.50618, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50618, 2013.

Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Geophysical Research Letters, 39, https://doi.org/10.1029/2011GL050087, http://onlinelibrary.wiley.com/doi/10.1029/2011GL050087/abstract, 2012.

Henson, S. A., Beaulieu, C., and Lampitt, R.: Observing climate change trends in ocean biogeochemistry: when and where, Global Change Biology, 22, 1561–1571, https://doi.org/10.1111/gcb.13152, http://onlinelibrary.wiley.com/doi/10.1111/gcb.13152/abstract, 2016.

Henson, S. A., Beaulieu, C., Ilyina, T., John, J. G., Long, M., Séférian, R., Tjiputra, J., and Sarmiento, J. L.: Rapid emergence of climate change in environmental drivers of marine ecosystems, Nature Communications, 8, https://doi.org/10.1038/ncomms14682, http://www.nature.com/ncomms/2017/170307/ncomms14682/full/ncomms14682.html, 2017.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-12-00121.1, 2013.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., and Núñez-Riboni, I.: Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, Journal of Advances in Modeling Earth Systems, 5, 287–315, https://doi.org/10.1029/2012MS000178, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012MS000178, 2013.

IPCC: Summary for policymakers, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Pörtner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N., https://www.ipcc.ch/srocc/home/, in Press, 2019.

Joos, F., Plattner, G.-K., Stocker, T. F., Körtzinger, A., and Wallace, D. W. R.: Trends in marine dissolved oxygen: Implications for ocean circulation changes and the carbon budget, Eos, Transactions American Geophysical Union, 84, 197–201, https://doi.org/10.1029/2003EO210001, http://onlinelibrary.wiley.com/doi/10.1029/2003EO210001/abstract, 2003.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, Journal of Advances in Modeling Earth Systems, 5, 422–446, https://doi.org/10.1002/jame.20023, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/jame.20023, 2013.

Keller, K. M., Joos, F., and Raible, C. C.: Time of emergence of trends in ocean biogeochemistry, Biogeosciences, 11, 3647–3659, https://doi.org/10.5194/bg-11-3647-2014, 2014.

Keller, K. M., Joos, F., Lehner, F., and Raible, C. C.: Detecting changes in marine responses to ENSO from 850 to 2100 C.E.: Insights from the ocean carbon cycle, Geophysical Research Letters, 42, https://doi.org/10.1002/2014GL062398, http://onlinelibrary.wiley.com/doi/10.1002/2014GL062398/abstract, 2015.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369–373, https://doi.org/10.1038/nclimate1716, https://www.nature.com/articles/nclimate1716, 2013.

Lehner, F., Joos, F., Raible, C. C., Mignot, J., Born, A., Keller, K. M., and Frölicher, T. L.: Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, Earth System Dynamics, 6, 411–434, https://doi.org/10.5194/esd-6-411-2015, http://www.earth-syst-dynam.net/6/411/2015/, 2015.

Levitus, S., Antonov, J. I., Boyer, T. P., Locarnini, R. A., Garcia, H. E., and Mishonov, A. V.: Global ocean heat content 1955-2008 in light of recently revealed instrumentation problems, Geophysical Research Letters, 36, https://doi.org/10.1029/2008gl037155, 2009.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J. R., Seidov, D., Yarosh, E. S., and Zweng, M. M.: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051106, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051106, 2012.

Lombardozzi, D., Bonan, G. B., and Nychka, D. W.: The emerging anthropogenic signal in land–atmosphere carbon-cycle coupling, Nature Climate Change, 4, 796–800, https://doi.org/10.1038/nclimate2323, https://www.nature.com/articles/nclimate2323, 2014.

Long, M. C., Deutsch, C., and Ito, T.: Finding forced trends in oceanic oxygen, Global Biogeochemical Cycles, 30, https://doi.org/10.1002/2015GB005310, http://onlinelibrary.wiley.com/doi/10.1002/2015GB005310, 2016.

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., Ethé, C., Flavoni, S., Graham, T., Harle, J., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Paul, J., Rousset, C., Storkey, D., Storto, A., and Vancoppenolle, M.: NEMO ocean engine, https://doi.org/10.5281/zenodo.3248739, https://zenodo.org/record/3248739#.XZXrtEYzaMo, 2017.

Magnan, A. K., Colombier, M., Billé, R., Joos, F., Hoegh-Guldberg, O., Pörtner, H.-O., Waisman, H., Spencer, T., and Gattuso, J.-P.: Implications of the Paris agreement for the ocean, Nature Climate Change, 6, 732–735, https://doi.org/10.1038/nclimate3038, https://www.nature.com/articles/nclimate3038, 2016.

Margolskee, A., Frenzel, H., Emerson, S., and Deutsch, C.: Ventilation Pathways for the North Pacific Oxygen Deficient Zone, Global Biogeochemical Cycles, 33, 875–890, https://doi.org/10.1029/2018GB006149, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018GB006149, 2019.

Matear, R. J. and Hirst, A. C.: Long-term changes in dissolved oxygen concentrations in the ocean caused by protracted global warming, Global Biogeochemical Cycles, 17, https://doi.org/10.1029/2002GB001997, http://doi.wiley.com/10.1029/2002GB001997, 2003.

Moore, J. K., Doney, S. C., Kleypas, J. A., Glover, D. M., and Fung, I. Y.: An intermediate complexity marine ecosystem model for the global domain, Deep-Sea Research Part II: Topical Studies in Oceanography, 49, 403–462, https://doi.org/10.1016/S0967-0645(01)00108-4, 2002.

Moore, J. K., Doney, S. C., and Lindsay, K.: Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, Global Biogeochemical Cycles, 18, https://doi.org/10.1029/2004GB002220, http://onlinelibrary.wiley.com/doi/10.1029/2004GB002220/abstract, 2004.

Nijsse, F. J. M. M., Cox, P. M., Huntingford, C., and Williamson, M. S.: Decadal global temperature variability increases strongly with climate sensitivity, Nature Climate Change, 9, 598–601, https://doi.org/10.1038/s41558-019-0527-4, https://www.nature.com/articles/s41558-019-0527-4, 2019.

Oschlies, A., Duteil, O., Getzlaff, J., Koeve, W., Landolfi, A., and Schmidtko, S.: Patterns of deoxygenation: sensitivity to natural and anthropogenic drivers, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375, https://doi.org/10.1098/rsta.2016.0325, http://rsta.royalsocietypublishing.org/content/375/2102/20160325, 2017.

Palmer, J. R. and Totterdell, I. J.: Production and export in a global ocean ecosystem model, Deep Sea Research Part I: Oceanographic Research Papers, 48, 1169–1198, https://doi.org/10.1016/S0967-0637(00)00080-7, http://www.sciencedirect.com/science/article/pii/S0967063700000807, 2001.

Plattner, G.-K., Joos, F., and Stocker, T. F.: Revision of the global carbon budget due to changing air-sea oxygen fluxes, Global Biogeochemical Cycles, 16, 1096, https://doi.org/10.1029/2001GB001746, http://onlinelibrary.wiley.com/doi/10.1029/2001GB001746/abstract, 2002.

Pörtner, H.-O., Karl, D. M., Boyd, P. W., Cheung, W., Lluch-Cota, S. E., Nojiri, Y., Schmidt, D. N., Zavialov, P. O., Alheit, J., and Aristegui, J.: Ocean systems, in: Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change, pp. 411–484, Cambridge University Press, 2014.

Resplandy, L., Séférian, R., and Bopp, L.: Natural variability of CO2 and O2 fluxes: What can we learn from centuries-long climate models simulations?, Journal of Geophysical Research: Oceans, 120, 384–404, https://doi.org/10.1002/2014JC010463, http://onlinelibrary.wiley.com/doi/10.1002/2014JC010463/abstract, 2015.

Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., Gulev, S., Johnson, G., Josey, S., Kostianoy, A., Mauritzen, C., Roemmich, D., Talley, L., and Wang, F.: Observations: Ocean, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 255–316, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.010, www.climatechange2013.org, 2013.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P.: RCP 8.5—A scenario of comparatively high greenhouse gas emissions, Climatic Change, 109, 33, https://doi.org/10.1007/s10584-011-0149-y, https://doi.org/10.1007/s10584-011-0149-y, 2011.

Rodgers, K. B., Lin, J., and Frölicher, T. L.: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model, Biogeosciences, 12, 3301–3320, https://doi.org/10.5194/bg-12-3301-2015, https://www.biogeosciences.net/12/3301/2015/, 2015.

Roemmich, D., Church, J., Gilson, J., Monselesan, D., Sutton, P., and Wijffels, S.: Unabated planetary warming and its ocean structure since 2006, Nature Climate Change, 5, 240–245, https://doi.org/10.1038/nclimate2513, https://www.nature.com/articles/nclimate2513, 2015.

Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J., and Manabe, S.: Simulated response of the ocean carbon cycle to anthropogenic climate warming, Nature, 393, 245, https://doi.org/10.1038/30455, https://www.nature.com/articles/30455, 1998.

Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Frölicher, T. L., Dunne, J. P., Ishii, M., and Slater, R.: Emergence of anthropogenic signals in the ocean carbon cycle, Nature Climate Change, pp. 1–7, https://doi.org/10.1038/s41558-019-0553-2, https://www.nature.com/articles/s41558-019-0553-2, citation Key Alias: schlunegger_emergence_2019, 2019.

Schmidtko, S., Stramma, L., and Visbeck, M.: Decline in global oceanic oxygen content during the past five decades, Nature, 542, 335–339, https://doi.org/10.1038/nature21399, http://www.nature.com/nature/journal/v542/n7641/full/nature21399.html, 2017.

Smith, R., Jones, P., Briegleb, P., Bryan, O., Danabasoglu, G., Dennis, M., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, R., Hecht, M., Jayne, S., Jochum, M., Large, G., Lindsay, K., Maltrud, M., Norton, J., Peacock, L., Vertenstein, M., and Yeager, S.: The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM), Tech. rep., Los Alamos National Laboratory (LANL), https://opensky.ucar.edu:/islandora/object/manuscripts%3A825/, 2010.

Stramma, L., Johnson, G. C., Sprintall, J., and Mohrholz, V.: Expanding Oxygen-Minimum Zones in the Tropical Oceans, Science, 320, 655–658, https://doi.org/10.1126/science.1153847, http://science.sciencemag.org/content/320/5876/655, 2008.

Stramma, L., Prince, E. D., Schmidtko, S., Luo, J., Hoolihan, J. P., Visbeck, M., Wallace, D. W. R., Brandt, P., and Körtzinger, A.: Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes, Nature Climate Change, 2, 33–37, https://doi.org/10.1038/nclimate1304, https://www.nature.com/articles/nclimate1304, 2012.

Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL052676, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL052676, 2012.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-11-00094.1, 2012.

Tjiputra, J. F., Goris, N., Lauvset, S. K., Heinze, C., Olsen, A., Schwinger, J., and Steinfeldt, R.: Mechanisms and Early Detections of Multidecadal Oxygen Changes in the Interior Subpolar North Atlantic, Geophysical Research Letters, 45, 4218–4229, https://doi.org/10.1029/2018GL077096, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077096, 2018.

Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years: An update from CMIP5 models, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL052868, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL052868, 2012.

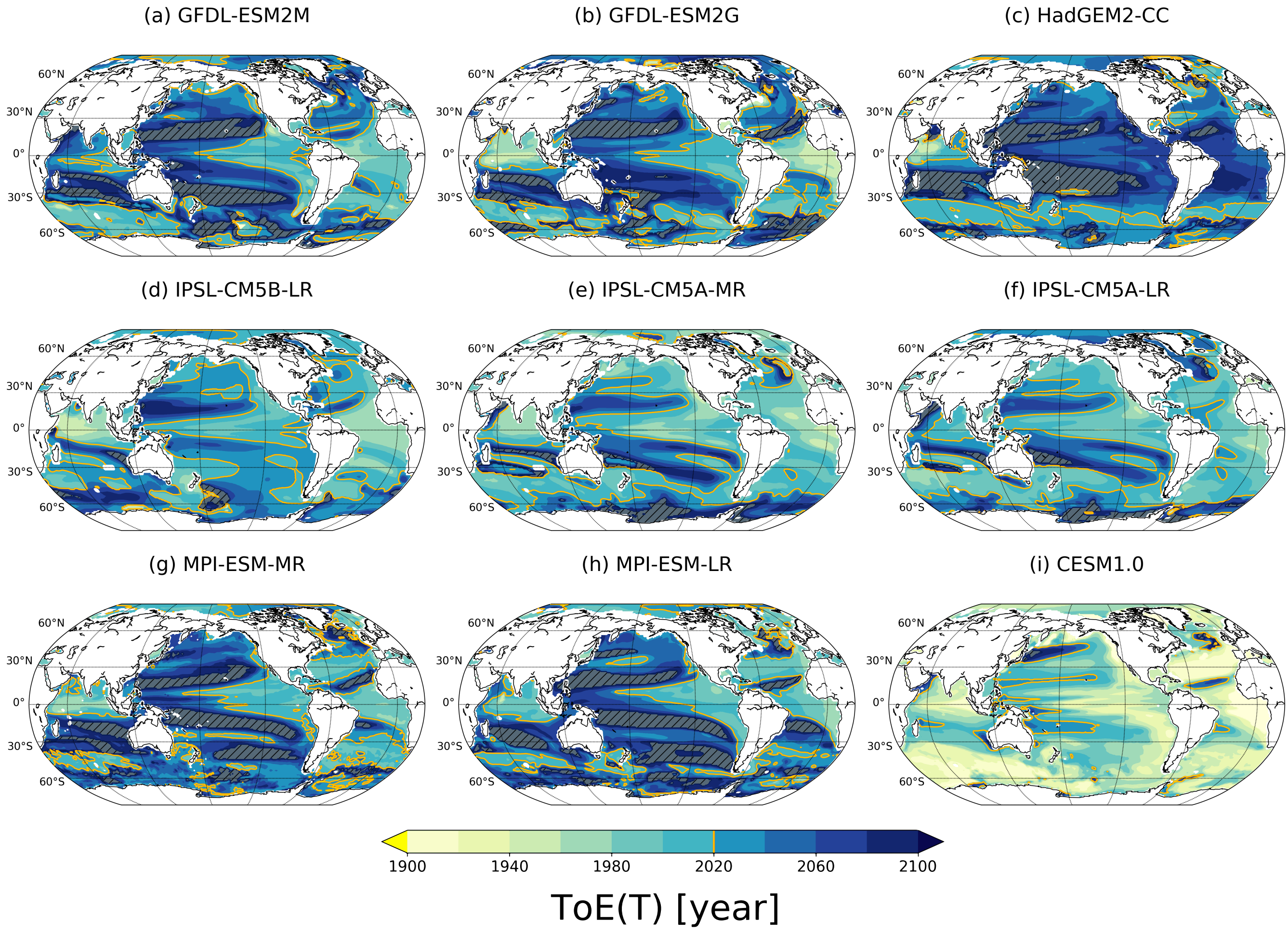

Figure S1. Time of Emergence (ToE) of T in the thermocline (200 – 600 m). The hatches areas represent the regions where the signal has not emerged by the end of the 21st century. The yellow contours highlight the present time (year: 2020).

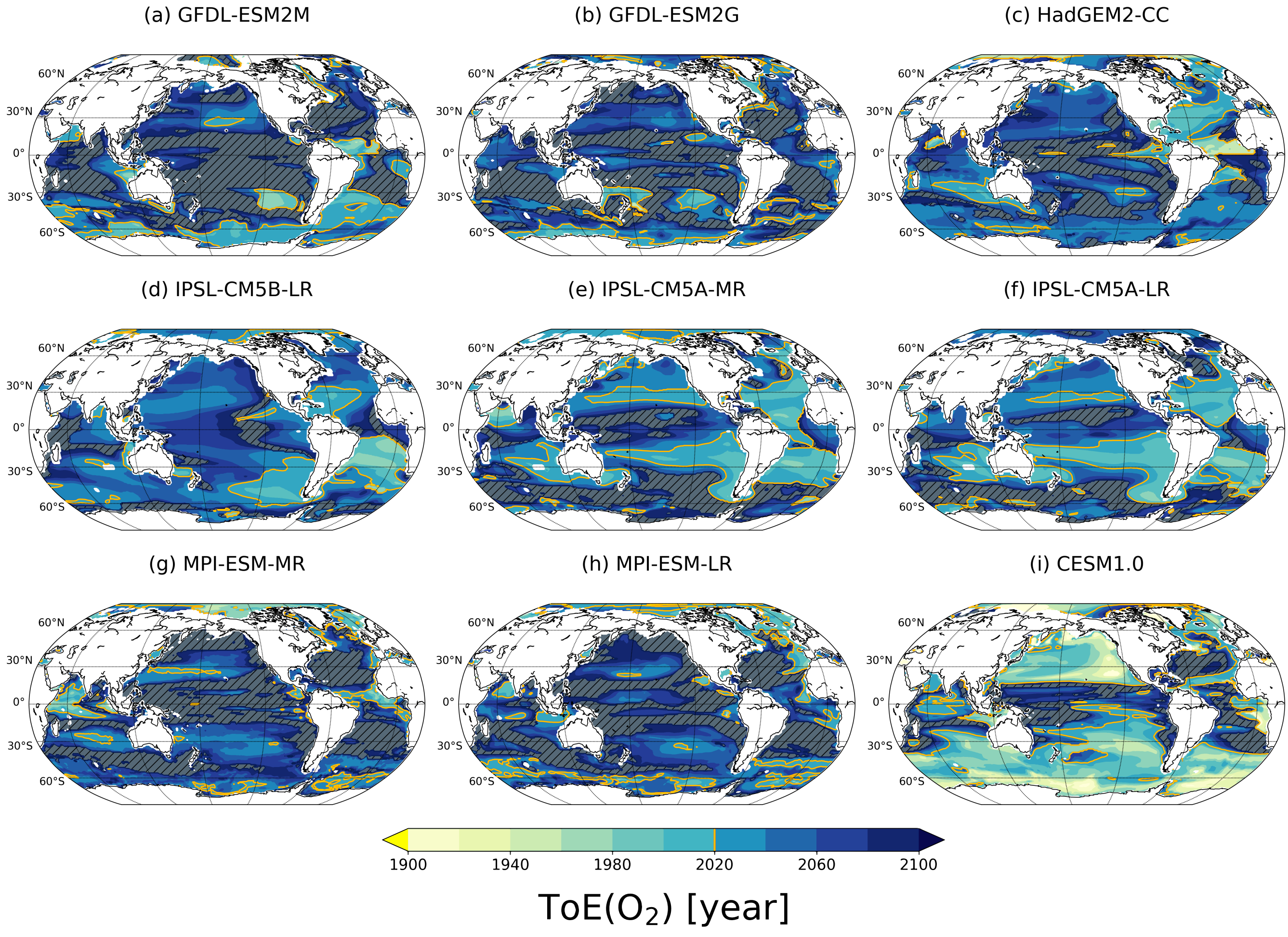

Figure S2. Time of Emergence (ToE) of O_2 in the thermocline (200 – 600 m). The hatches areas represent the regions where the signal has not emerged by the end of the 21st century. The yellow contours highlight the present time (year: 2020)

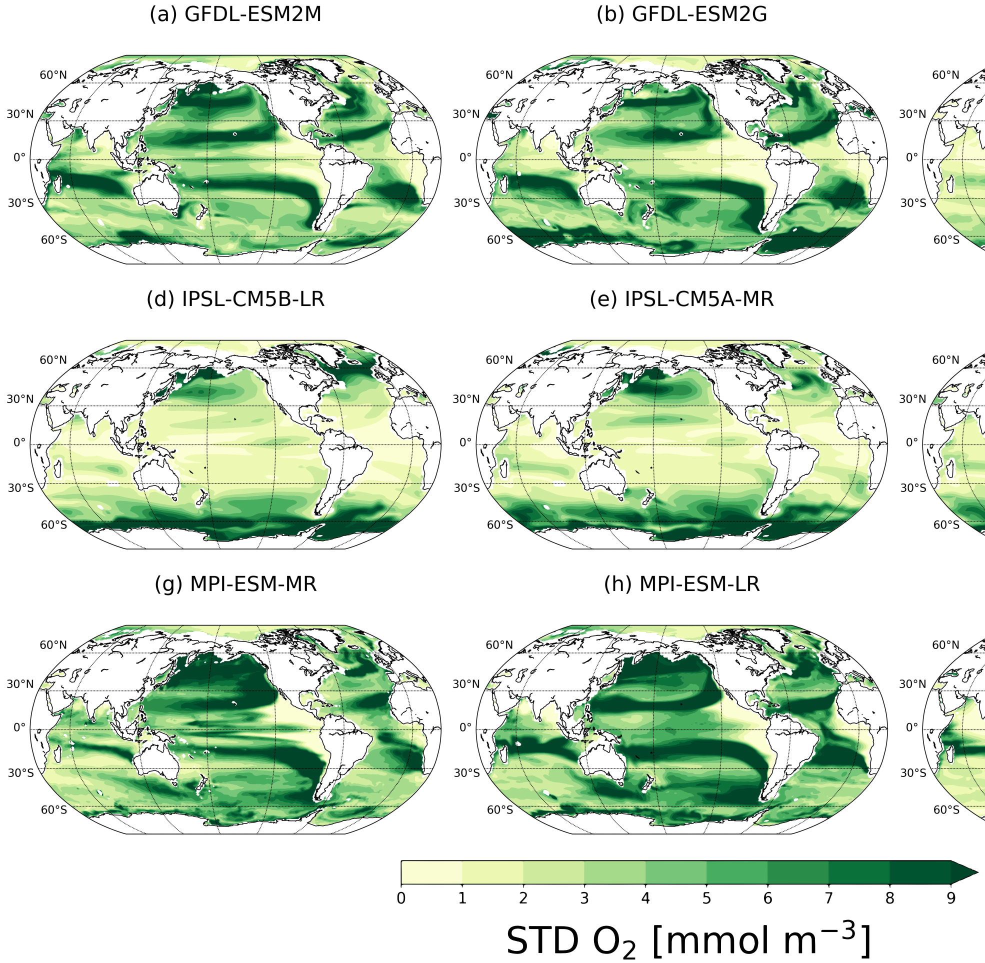

Figure S3. Standard deviation of temperature in the averaged layer 200 – 600 m of the associated control simulation.

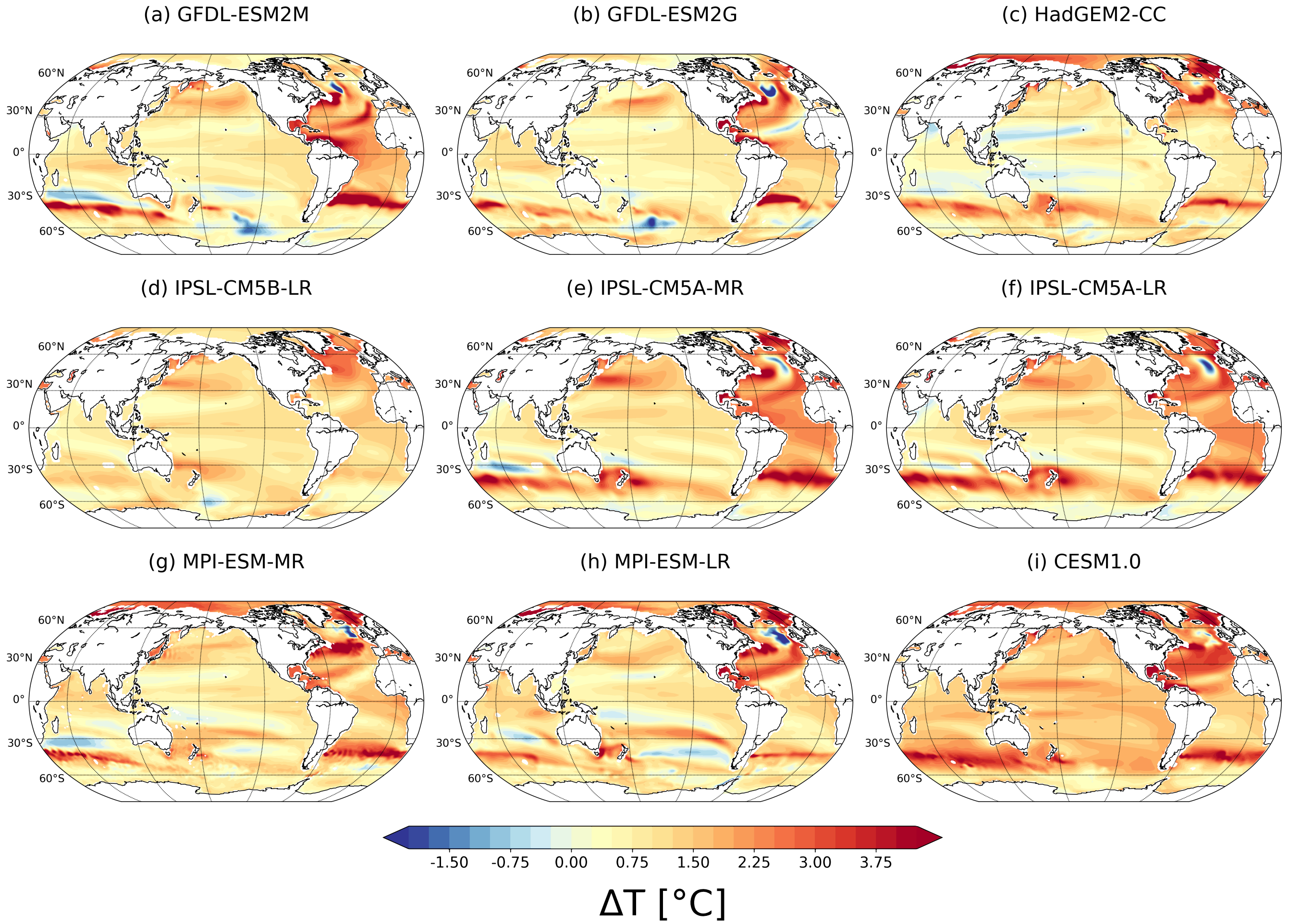

Figure S4. Anthropogenic changes ((2070-2099 CE) minus (1861-1959 CE)) in temperature.

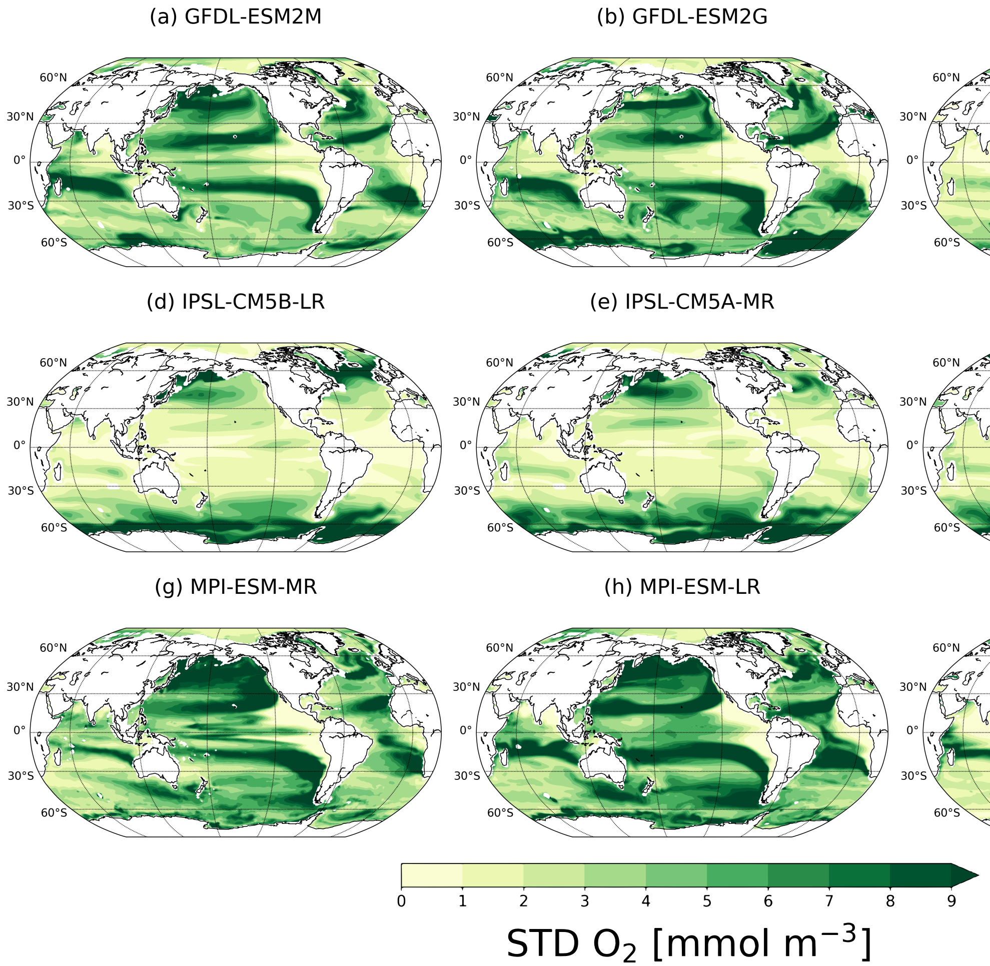

Figure S5. Standard deviation of O_2 in the averaged layer 200 – 600 m of the associated control simulation.

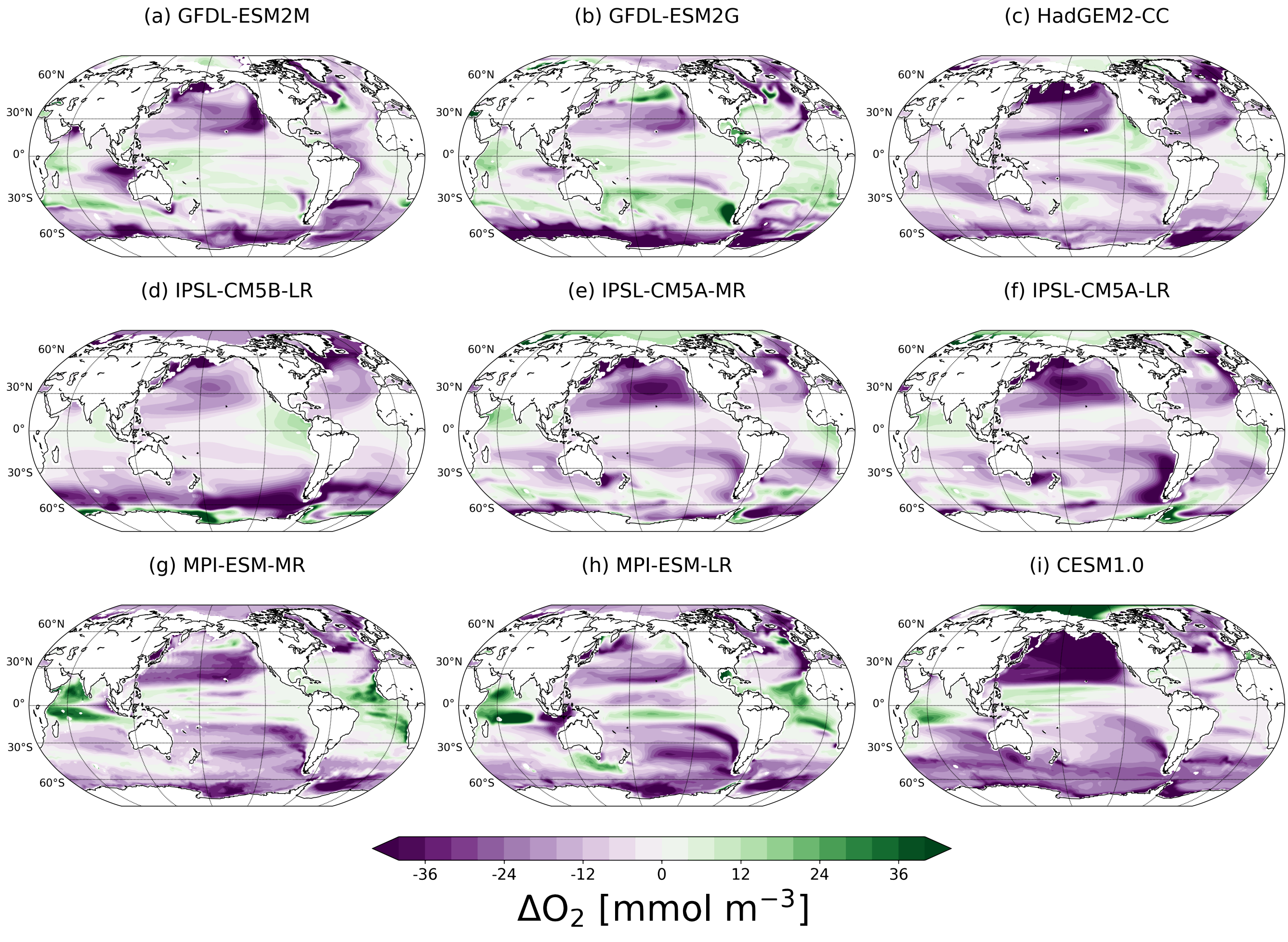

Figure S6. Anthropogenic changes ((2070-2099 CE) minus (1861-1959 CE)) in O_2.