We would like to thank the Anonymous Reviewer #1 and the Anonymous Reviewer #2 for their time and constructive comments. Their comments are reproduced in blue below. We reply in black.

Anonymous Reviewer #1

The manuscript "A new estimate of ocean oxygen utilization points to a reduced rate if respiration in the ocean interior" presents a method (the EOU method) of calculating the integrated oxygen utilization at any point in the interior ocean. Because the new method accounts for oxygen undersaturation at the ocean surface, it predicts about 25% weaker biological oxygen utilization than the AOU method.
That AOU overpredicts oxygen utilization due to surface O2 undersaturation is well known. The value of the present manuscript is that the authors present a method for quantifying this undersaturation, thereby deriving a less biased estimate of oxygen utilization, as supported by application of the EOU method to model output. However, this value is muted by the fact that the authors do not adequately address the limitations, uncertainties, and potential biases of their admittedly ad-hoc correction method. The method will be most useful applied to observations, and this is where the authors need to do a much better job of addressing the uncertainties. Modelers are unlikely to usethis method to determine oxygen utilization, as running a preformed oxygen tracer is straightforward and much more accurate.

I recommend that the manuscript be revised to better address uncertainties in the observationally-estimated EOU. Specifically, the following sources of error should be quantified and explicitly discussed in the revised manuscript:
1. Sensitivity of oxygen utilization predicted by the EOU method to the number of isopycnal layers. This is addressed for the models in Figure A1, but also needs to be addressed for the observations. Are the observational EOU estimates highly sensitive to the number of isopycnal layers (like UVIC or BLING models) or insensitive (like MIT or CSIRO models)? What mechanism determines the sensitivity of the EOU estimate to the number of isopycnal layers?

2. The effects of spatially and temporally variable surface saturations. The densest waters in the ocean outcrop in both the North Atlantic and the Southern Ocean, where surface saturations are very different. Also, surface saturations and water-mass formation rates vary seasonally. So the authors need to discuss how the assumption of a uniform (spatial and temporal average) surface saturation for the deepest waters affects the oxygen utilization inferred by the EOU method. The spatial differences in saturation are shown in Figure B1, but this is a significant source of error/bias that needs to be quantified and discussed in the main text.

We agree with the reviewer and we will include a more in-depth analysis of uncertainties of the EOU approach in the main text (Section 3.3) of the paper. A draft of the new text is given below. In this draft, we refer to Appendix B and Figure B1. For the revised version of the manuscript we plan to move both to the main text (into Section 3.1), below we keep the references to the BGD version for clarity. Please note also that, discussing assumptions and parameters of our method now in more detail in the main text, we plan to move Appendix A to main text as well (into Section 2).

3.3 Evaluation of the EOU approach in models

(p2250/l20 – p2251/l26 as in the original submission)

Sources of EOU biases (will replace and extend p2252/l1-l20 of the original submission)
In the EOU definition, the mean oxygen saturation of each of the considered isopycnal layers is estimated by computing the area-weighted mean oxygen concentration at the intercept of this layer with the 50 m depth horizon. This is an approximation as in reality waters in the interior of the ocean are variable mixtures of water masses formed in different source regions and having different specific surface oxygen disequilibrium (see Appendix, Fig. B1). On the same isopycnal, waters in for example the North Atlantic Ocean will origin mainly from the northern outcrops of this isopycnal, as will be their preformed oxygen content, while waters in the South Atlantic Ocean will mainly origin from the Southern Ocean outcrops. Resolving the complexity of variable water mass fractions Gebbie and Huybers, 2010) contributing to the actual preformed oxygen concentration explicitly is a non-trivial task, hence our ad hoc approximation.

This meridional gradient in the relative contributions of different source water masses contributes to explain negative EOU – TOU differences in the northern Atlantic Ocean and positive EOU – TOU differences in the Southern Ocean in UVIC and om1p7-BLINGv0. Both models are characterized by large differences in oxygen saturation between the Northern North Atlantic and the Southern Ocean (Fig. B1), with the latter being more strongly undersaturated in oxygen at the 50 m depth horizon. Assuming an isopycnal layer composed of a mean preformed oxygen quantity leads to an overestimate of the oxygen utilization in the Southern Ocean and to an underestimate of it in the North Atlantic Ocean. Models characterized by smaller differences between water mass formation regions also show smaller EOU – TOU differences throughout the ocean. Here the assumption of a mean preformed oxygen quantity for each isopycnal applies much better.

A critical parameter in the EOU computation is the choice of the intercept-depth horizon. Water masses form mostly in winter whereas annual mean model outputs have been used in our analysis for reasons discussed below (Section 4.1). In order to reduce the impact of this simplification, we compute the preformed oxygen used in the EOU method from the mean oxygen disequilibrium of a given isopycnal at a defined depth below the surface rather than the ocean surface itself. This is guided by the assumption that the upper permanent thermocline is mostly characterized by winter conditions (Stommel et al., 1979). We use here annual model outputs, as the goal of our study is to provide a practical and easy solution to better estimate the oxygen utilization from observations, in which high latitudes are biased seasonally, with very scarce observations during the winter season (see Section 4.1). Testing the sensitivity of the EOU to the depth of this intercept horizon, we find using an ensemble average, that the 50 m depth horizon leads, averaged over all models, to the best estimate of the TOU (Figure 1 of rebuttal). The EOU computed in models characterized by a weak seasonal cycle is less sensitive to the value of the intercept-depth horizon than the EOU computed in models where the seasonal cycle is strong, principally at high latitude.

Another critical parameter is the number of isopycnal layers considered. The main factor explaining the sensitivity to the number of isopycnal layers is the heterogeneity of oxygen disequilibrium in the near-surface. An ocean (model) displaying a spatially more constant surface oxygen saturation will not be very sensitive to a discretization in isopycnal layers. This is the case of, for instance, MIT28 or CSIRO (Fig. B1). Inversely, UVIC or om1p7-BLINGv0, characterized by a stronger heterogeneity of oxygen saturation at the intercept-depth display a large sensitivity to the number of isopycnal layers. The 'optimal' number of layers to take in account when computing the EOU also depends on the strength of diapycnal mixing. A model characterized by very high diapycnal diffusivity in the interior ocean will not conserve the outcrop properties in the ocean interior and the EOU – TOU bias will be minimized when only few isopycnal layers are considered. Inversely, a model characterized by weak diapycnal diffusivity requires a higher number of isopycnal layers in order to reproduce correctly the ocean interior. Using an ensemble average, we find that dividing the ocean into 10 isopycnal layers of equal density width (1kgm.-3) from 1020.5 to 1030.5 kg m-3 leads to the best estimate of the TOU (Fig. 1+2 of rebuttal). This is the number of layers which we apply in this study.

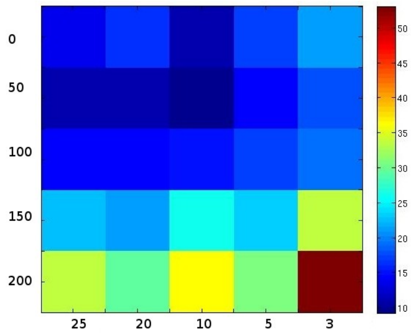


Figure 1: Model ensemble average of global EOU – TOU rms errors (mmol m-3) for a range of 'number of isopycnal layer' (abscisses) and 'intercept-depth horizon' (ordinates). The mean of the 6 models ensemble is displayed. The best fit (lowest global rms) is achieved combining 10 isopycnal layers and an intercept-depth of 50m.

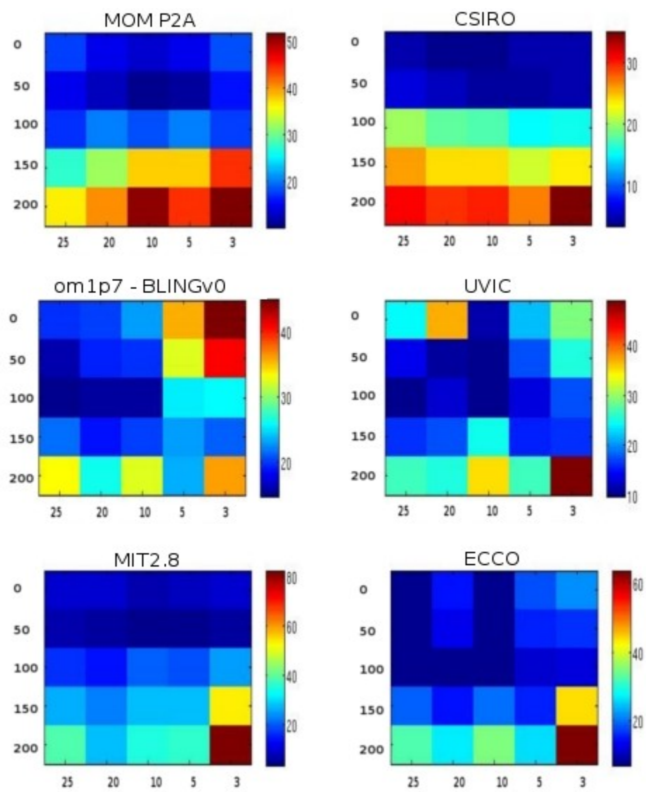


Figure 2 : Global EOU – TOU rms errors for each model and a range of 'number of isopycnal layer' (abscisses) and 'intercept-depth horizon' (ordinates). (Please note the differences in scale.)

[Concerning the reviewers suggestion to provide a more in depth analysis of the uncertainties of the EOU application to observations, we plan to extend the original section 4 as follows.]

4. Application of the EOU in observations.

p2252/l22 – p2253/l2 like in the original submission …

4.1 Uncertainties of the EOU approach in the observations (new section)

Unlike for the models, for which the preformed oxygen tracer provides a 'ground truth' information, there is nothing like that for the real ocean. It is, however, possible to detail the sources of EOU biases, discussed in Section 3.3 ('Evaluation of the EOU approach in models'), in order estimate the uncertainty of the EOU in the real ocean.

Spatial variability of oxygen disequilibrium

The most significant atmosphere/upper ocean oxygen disequilibrium occurs in the high latitude subduction regions, where deep water (considered here as the water with a potential density greater than 1027.5 kg.m-3) forms. This deep water subducts both in the southern ocean (forming the Circumpolar Deep Water, CDW) and in the northern North Atlantic (forming the North Atlantic Deep Water, NADW). In the observations, the annual mean oxygen field at 50 m depth (corresponding to the depth used to compute the preformed O2 in the EOU approach) of the CDW is significantly undersaturated (mean saturation of 86 percent, mean oxygen concentration of 325 mmol.m-3), particularly in the Ross and Weddell Sea (Gordon and Huber, 1990; Keeling et al., 2010), important sites of deep water formation. In the northern North Atlantic, the annual mean oxygen saturation at 50 m depth is higher and close to saturation (mean saturation of 96 percent, mean oxygen concentration of 300 mmol.m-3). Deep winter convection in this region usually brings relatively young waters with little oxygen debt back to the surface.

About one third of the outcrop area of dense (sigma0 > 1027.5 kg.m-3) water is located in the North Atlantic, whereas the other two thirds are located in the Southern Ocean (Weddell and Ross Sea). In our EOU approach, this implies that the dense water is assumed to consist of 36 percent of NADW and 64 percent of CDW, which is comparable to the results of Johnson (2008) and Khatiwala et al. (2012). The respective mean preformed oxygen concentration of this dense water mass is about 306 mmol.m-3. Note, however, that estimates of source water mass fractions are strongly sensitive to the choice of the water mass end members and the methods and tracers applied (Broecker (1998), Johnson (2008), Gebbie and Huybers (2010)).

Another source of bias is linked to the meridional variability of the water mass contribution on a given isopycnal. Dense waters observed in the North Atlantic indeed mainly origin from the northern outcrop, as does their oxygen disequilibrium, while waters in the South Atlantic mainly origin from the Southern Ocean outcrops. As shown by Khatiwala et al (2012), nearly 80 percent of the deep water of the Southern Ocean has been formed locally. Considering a mixture 20 (80) percent of CDW and 80 (20) percent of NADW leads to a mean preformed oxygen concentration of (305) 320 mmol.m-3. In our approach, this meridional gradient is not taken into account, leading to regionally varying errors with maximum values of +/- 10 mmol.m-3 close to the outcrop regions.

Seasonal variability of oxygen disequilibrium

Water masses form mostly in winter. Winter time observations are, however, very scarce, especially in the Southern Ocean, as shown by the figure 3 (of the rebuttal, upper right panel). This lack of data from the winter period is the principal reason why we do not use seasonal data for the EOU estimate, but an annual-averaged data product.

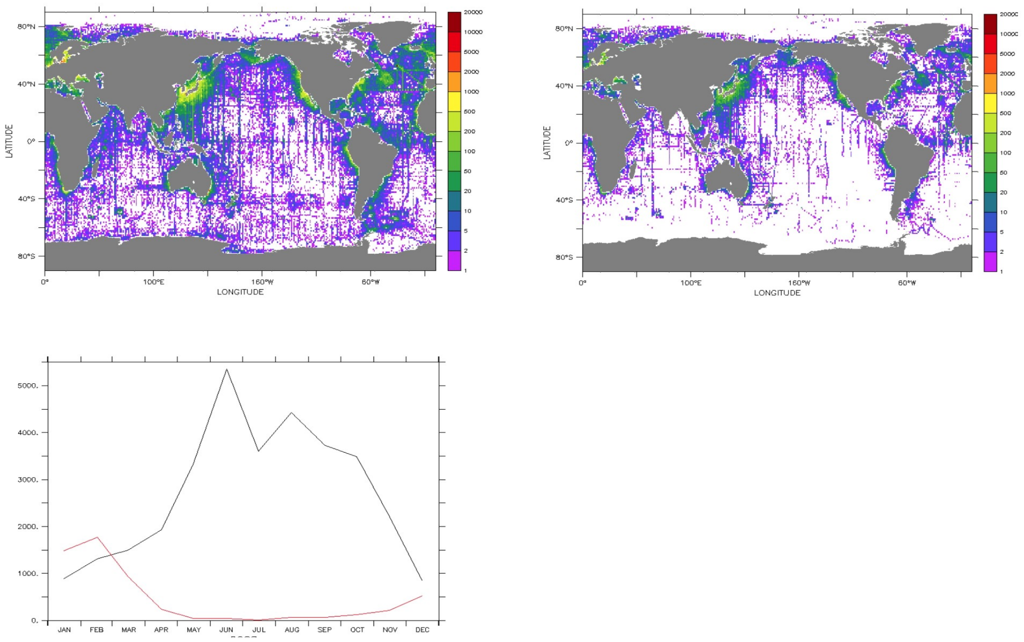


Figure 3: Surface distribution of observations of oxygen saturation in the World Ocean Atlas. Number of observations per 1x1 grid cell summed over the calendar year (Top left) and over the Southern Hemisphere winter (July-August-September) (Top right), respectively. (Bottom) Seasonal cycle of the total number of observations in the northern high latitudes (60N-90N) and the Southern Ocean (60S-90S), respectively.

In order to reduce the impact of this data bias, we compute the preformed oxygen used for EOU from the mean oxygen disequilibrium of a given isopycnal at the intercept with the 50 m depth layer rather than the ocean surface. The choice of this depth horizon is pragmatic, while guided by the model sensitivity analyses discussed in Section 3.1 (Figure 1 of the rebuttal). Choosing a shallower (deeper) depth horizon would result in a higher (lower) oxygen saturation and ultimately preformed concentration, as shown by the table below.

Using surface values would still lead to a deep water undersaturation by about 25 mmol.m-3, whereas considering 50 m as the intercept depth-horizon leads to an undersaturation of 35 mmol.m-3. When 100 m depth is used as the intercept depth, the undersaturation is on the order of 60 mmol.m-3. We used 50 m depth for pragmatical reasons, as it corresponds to the depth where the computed preformed oxygen tracer matches best the explicitly simulated preformed tracer in the 6 models (see Section 3.3). Furthermore, following Broecker et al., 1985, the concentration of phosphate in the NADW end member is 0.73 whereas it is 1.95 mmol.m-3 in the CDW. The mean concentration of phosphate of the dense water at the depth of 50 m in the Southern Ocean and 100 m in the North Atlantic Ocean matches well these values, suggesting that these depth layers are characteristic of the outcrops in winter, when water subducts. If we had used a different depth horizon in the Southern Ocean (50 m) and in the North Atlantic Ocean (100 m), the estimated preformed oxygen of the dense water would be lower by 10 mmol.m-3.

	End member						Area	Dense(>1027.5 kg.m-3) isopycnal	
	O2 CDW (mmol.m-3)	O2 NADW (mmol.m-3)	%Sat CDW	%Sat NADW	PO4 CDW (mmol.m-3)	PO4 NADW (mmol.m-3)	%CDW	O2 (mmol.m-3)	O2sat (mmol.m-3)
10	336	306	90	100	1.81	0.44	60	319	346
50	324	300	86	96	2.00	0.59	64	306	345
100	298	295	80	93	2.1	0.72	63	288	351
200	236	286	66	87	2.25	0.87	71	250	345

Table 1: Sensitivity of the densest water masses (sigma0 > 1027.5 kg.m-3) oxygen and phosphate concentrations with respect to chosen intercept-depth horizon.

Isopycnal discretization

In the observations, a preformed oxygen minimum of 310 mmol.m-3 is computed for the deep ocean (sigma0 > 1027.5 kg.m-3) when 10 isopycnal layers are considered, which is the number of layers leading to the optimal fit of EOU - TOU in the models that we considered (see Section 3.3). The maximal value is 320 mmol.m-3 when 5 layers are considered, corresponding to the minimum number of layers needed to discriminate the main water masses, which is significantly lower than the value of 345 mmol.m-3 derived from TS (AOU methodology). Increasing the number of layers further does not modify significantly the amount of preformed oxygen. We conclude that the preformed concentration is very likely to be between 310 and 320 mmol.m-3 in the deep ocean layer.

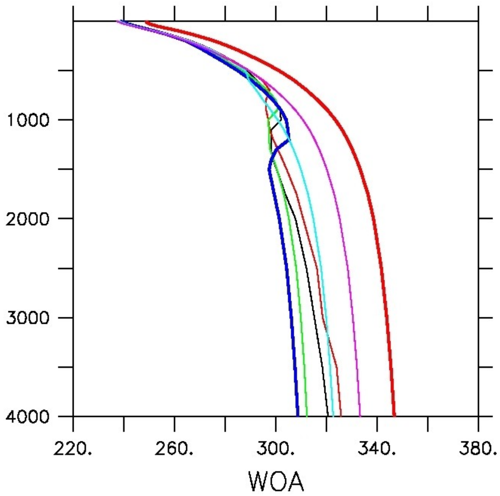


Figure 4: Global mean profiles of preformed oxygen concentrations (mmol O2 m-3) in the WOA corresponding to the preformed oxygen estimate of the AOU (bold red) and EOU (thin lines). EOU has been computed considering an increasing number of potential density layers ranging from 3 to 25 (thin black: 25 layers, thin red: 20, green: 15, blue: 10, light blue: 5, magenta: 3).

Considering the biases in the EOU approach lead to a confidence estimate of the deep ocean preformed oxygen of about +/- 10 mmol.m-3.
[End of new section 4.1]

4.2 Regenerated vs total nutrients pool
p2253/l3-l19 like in the original submission …

Specific line items:

Page 2248, Line 15: need to be more clear about what is meant by "TS properties"

The original sentence:

'As a general remark, mixing of waters with different TS properties usually leads to an increase of the degree of oxygen saturation. The result of this non-linearity will then be an underestimation of oxygen utilization in both the AOU or EOU diagnostic. This bias is however small and does not exceed a few percent in the ocean interior (Dietze and Oschlies, 2005).'

will be replaced by :

'As a general remark, the oxygen content of a water mass obtained by mixing two saturated water masses of the same density but with different temperatures, is higher than the oxygen saturation of the resulting water mass, ultimately leading to an apparent increase of the degree of oxygen saturation. The result of this non-linearity will be an underestimation of oxygen utilization in both the AOU or EOU diagnostic. This bias is however small and does not exceed a few percent in the ocean interior (Dietze and Oschlies, 2005)'

Page 2249, Line 7: 2000 years is not long enough for the deep ocean (esp. Pacific) to reach equilibrium with the surface forcing. What are the model drifts in O2?

We agree that the statement of model spin-up being > 2000yrs was partly misguiding. We will be more specific and will provide the actual spinup-conditions and times in Table 1. In fact, 2000 yrs is well above the circulation age of the oldest waters observed in the North Pacific, see e.g. Matsumoto et al., 2007 (their Figure 4, maximum average circulation age below 1500m is < 1200 yrs). Given this, the minimum spin up times needed for our study, depends also very much on the initial conditions of the oxygen and preformed oxygen tracers. In some of the model runs (TMM MIT2.8, TMM ECCO) the status of the spin-up was further tested by implementing both a tracer for preformed oxygen and a tracer for utilized oxygen and running the model until the sum of both tracers equals the concentration of the oxygen tracer in basically any model grid box. UVIC and MOM3-P2A have been run for 5000 years. CSIRO has been run for 6000 years. om1p7-BLINGv0 has been ran for 2100 years and has been initialized by the WOA dataset, thus allowing a relative fast spinup. The models do not display a significant drift.

Page 2249, Line 19: The statement "The deep Pacific Ocean is solely filled by CDW" is not true. See e.g. Gebbie and Huybers, 2010; DeVries and Primeau, 2011.

The original statement 'The deep Pacific Ocean is solely filled by the CDW' will be replaced by: 'The deep Pacific Ocean is filled by a mixture of CDW and NADW'

Page 2249, Line 21: intermediate depth -> be more specific

The sentence:

'In both oceans, intermediate-depth AOU displays high values (exceeding 200 $mmolm^{-3}$) in the tropical regions, due to the sluggish circulation and the remineralization of large amounts of organic material exported from productive surface waters.'

will be replaced by :

'In both oceans, intermediate-depth waters display the highest values of AOU, due to the remineralization of large amounts of organic material exported from productive surface waters associated with limited contact with the atmosphere and high residence time. AOU reaches 200 mmol.m-3 between 200 - 1000 m in the tropical Atlantic Ocean and more than 280 mmol.m-3 between 200 - 2000 m in the tropical and North Pacific Ocean'

Page 2249, Line 22-23: "sluggish circulation" is debatable. Equatorial current systems are very strong.

The equatorial current system is indeed very strong. However, the thermocline is shallow in tropical regions compared to the gyres, which are well ventilated and characterized by a circulation extending down to 300-400 m depth. The term 'sluggish circulation' is possibly misleading. The 200-1000 m depth range in the tropical oceans may be better characterized by having 'limited contact with the atmosphere and high residence times' (see our reply above for the complete sentence).

Page 2253, Lines 12-15: Need to include uncertainty on EOU estimate of reg/tot nutrients. Examples of sources of uncertainty include number of isopycnal layers used to compute the EOU, variability in surface saturation, and ratio of O2:P. What ratio of O2:P is assumed?

The O2:P ratio of oxic remineralisation of organic matter assumed here is -170/1 (Takahashi et al.,1985; Anderson and Sarmiento, 1994; Li et al., 2000, Koertzinger et al., 2001). This ratio is commonly used and widely accepted. We therefore do not test for the sensitivity of the O2:P ratio on computed preformed and regenerated nutrients. The uncertainty of the estimate of the oxygen utilization itself has been evaluated in the new section (see 'Evaluation of the EOU approach in observations'). This uncertainty (+/- 10 mmol.m-3) in the oxygen utilization affects the estimate of the regenerated over total phosphate pool fraction (0.3 (+/- 0.03)).

Page 2255: Cut out the conceptual script as it is not needed. The calculation is straightforward.

We agree and delete the conceptual script.

Page 2254-5, Lines 24-2: This issue needs to be better addressed and contribution to EOU uncertainty quantified.

Page 2256, Lines 17-25: This issue needs to be better addressed and contribution to EOU uncertainty quantified.

We agree that both issues (number of isopycnal layers and location of surface disequilibrium) require a more in depth treatment. A draft of this section (3.3 /4.1) is presented earlier in this rebuttal.

Page 2256, Line 23: Are the values of diapycnal diffusivity in the interior ocean different in the models? If so, what are the values? What are the overturning strengths in the models?

A detailed analysis of the overturning strengths of the models is beyond the scope of this paper and detailed descriptions may be found in the cited original references presenting the models which we use. We will, however, augment the sentence the reviewer refers to:

'A possible cause of the differences between models and observations is linked to the intensity of diapycnal mixing in the ocean interior, which controls the overturning strength, and consequently the subduction rate (Gnanadesikan, 1999)'

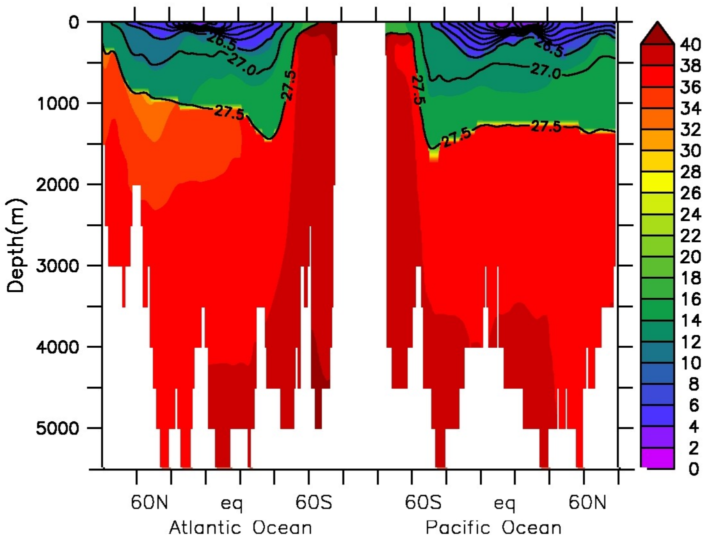
by adding:

'However, high mixing intensity implies high MOC and subduction rate but also a stronger impact of deoxygenated deep water on ocean surface. Ice cover plays also a significant role as well. Disentangling the mechanisms which lead to near surface disequilibrium of oxygen would need several sensitivity experiments, similar to those of Ito et al. 2004, but for each of the models. This is beyond the scope of this study.'

Table 2: Should add uncertainty to WOA EOU estimates (115.5 +- ? mmol/mˆ3 and 0.30 +- ?) due to variability in surface saturation, # of isopycnal layers used, etc.
The table has been complemented. The WOA EOU estimate is now 115.5 (+/- 10 mmol.m-3), for details see the new Section 4.1 'Evaluation of the EOU approach in the observations'. This leads to an uncertainty estimate of the regenerated to total phosphate ratio of 0.3 (+/- 0.03).

Figure 2: The authors could do a better job of addressing why the EOU-TOU differences vary so widely among the models. What do the discrepancies tell us about the circulations of the various models? What can they tell us about potential biases in observational EOU estimates?
The sources of the EOU – TOU errors and some possible causes of the inter-model variability are now discussed in more detail (see page 1 of this rebuttal).

Figure 4a, right panel: The colors seem to jump from pink (10-20) to orange (30-40) with little splotches of red along the 27.5 contour. Why?
A new figure has been issued with a sharper colorbar. The values are indeed characterized by a jump from about 30 to 20 mmol.m-3 along the sigma0 27.5 density (see below). This discontinuity corresponds to the discontinuity of the saturation value at 50 m depth between the denser water class (>27.5) and the lighter water class (26.5 – 27.5). There is no smoothing, as the figure corresponds to a section in the Atlantic and Pacific at a given longitude, and not to a zonal mean of the ocean.



WOA AOU – EOU

Anonymous Reviewer #2

Summary: Duteil and co-authors develop a new approach to estimate oxygen utilization rates in the ocean, the so-called Evaluated Oxygen Utilization (EOU). The new approach tries to account for the atmosphere-ocean oxygen disequilibrium in subduction regions. By using an modeled idealized preformed oxygen tracer, the authors show that the new approach outperforms the classical approach based on apparent oxygen utilization in six different ocean models. By applying the method to observational-based data, the authors suggest that the biological oxygen consumption rate is 25 percent lower than derived from AOU-based estimates.

Evaluation: It is well known that the oxygen concentration at the surface is not exactly at saturation and that AOU overestimates oxygen utilization rates. So far, to my knowledge, a 'simple' method to quantify the impact of the undersaturation on oxygen utilization rates is missing. Furthermore, no multi-model intercomparison of oxygen using an idealized preformed oxygen tracer has been performed. This paper addresses a clear gap in our understanding of the respiration in the open ocean, and therefore represent a welcomed and important contribution to the field. The paper is overall well written, and clearly organized. It addresses an important topic, and the method and results are of interest to a wider community.

We thank the reviewer for this positive evaluation.

Recommendation: I recommend acceptance of this manuscript after moderate revisions. I particularly recommend that the authors extend their introduction and discussion section to put the new findings into a broader context. Furthermore, the manuscript lacks of important details that have to be addressed before publication.

Major comments

1. Although I appreciate the effort to keep the paper short and dense, I recommend to extend the introduction and discussion section to put the new findings into context. Why is it important to have adequate oxygen consumption rates? What are the implications of a 25 percent lower OUR? Does the study change our understanding of the respiration processes in the open ocean? Instead of 'just' introducing the new method, discussing the implications of the new OUR estimate would be of interest to a broad audience.

Please note that we do not claim to re-evaluate the Oxygen Utilization Rate (OUR). OUR is defined as OUR = dAOU/dt (eg. Sarmiento and Gruber, 2006; chapter 5.2), where dAOU is an AOU gradient and dt is an age gradient, e.g. along a given isopycnal. The main uncertainty when estimating OUR of a water mass is the determination of age, i.e. of the time elapsed since the last contact with the atmosphere. For this reason, OUR estimates of North Atlantic Ocean thermocline waters are uncertain by even an order of magnitude (Sarmiento et al., 1990; Sarmiento and Gruber, 2006, Figure 5.2.4). Considering, however, that the age distribution in the ocean mainly follows isopycnals it is not clear whether an estimate of OUR, given a perfect estimate of age, but using either the gradient of AOU or EOU would be significantly different (see e.g. Ito et al., 2004).

Apart from this issue, we agree that the introduction needs to be extended in order to better put the presented research into context. For this purpose, we drafted the following new version of the introduction, which in particular highlights the use of AOU in a number of studies.

'Respiration is a key biological process in the ocean. Oxygen is the major electron acceptor used by microorganisms for the oxidation of organic matter. Classically, respiration in the interior of the ocean is quantified by the "Apparent Oxygen Utilization" (AOU). AOU is a time integrated measure of the amount of oxygen used since water left the ocean surface. This concept assumes that sea-surface oxygen is fully equilibrated with the atmosphere, i.e. concentrations are equivalent to 100% oxygen saturation, given temperature and salinity of the water (Weiss, 1970)

However significant areas of the surface ocean are undersaturated in high latitudes (e.g. Gordon and Huber, 1990; Koeve, 2001; Russel and Dickson, 2003; Koertzinger et al., 2005; Keeling et al., 2010), in particular during the winter season. These waters are subducted into the ocean interior. The surface ocean oxygen undersaturation is caused by a range of physical processes limiting air-sea exchange, including, for instance, the coverage with sea ice or strong subduction rates. Upwelling of and mixing with low-oxygen waters from below also contributes significantly to surface undersaturation (Ito et al., 2004). Inversely, surface ocean supersaturation is observed in the tropics, subtropical gyres and high latitudes in summer, principally due to photosynthesis (Craig and Hayward, 1986), but also from radiative warming below the mixed layer (Dietze and Oschlies, 2005). Waters affected by oxygen supersaturation, however, do not contribute much to the ventilation of the deep ocean due to their relative low density. As a consequence, AOU globally overestimates oxygen respiration in the water column due to subduction of dense undersaturated water in winter at high latitudes.

Despite this well known but often neglected bias, the AOU is widely used in biogeochemical studies. The contribution of dissolved organic carbon as the source of reduced carbon to total respiration in the ocean interior has for instance been evaluated based on AOU measurements (Ogura, 1970; Doval and Hansell, 1999; Aristegui et al., 2002; Carlson et al., 2010). The observed correlation between nitrous oxide disequilibrium, a long-lived natural greenhouse gas, and AOU provides a means to estimate nitrous oxide production as a function of oxygen consumption, even if the exact mechanism of production is not resolved (Najjar, 1992, Nevison et al., 2003). Further it has been suggested that AOU has increased in parts of the ocean during the last decades, due to a complex interplay of carbon export and circulation changes (Emerson et al., 2004). Finally, nutrients may be partitioned in a regenerated and preformed pool using the AOU metric (Ito and Follows, 2005; Duteil et al., 2012) supporting the distinction of different water masses (Broecker et al., 1985) as well as the analysis of biogeochemical models. For example, the ratio of regenerated over total nutrients has also been used to characterize the relative strength of the biological and physical pathways in the return of nutrients to the ocean interior (Ito and Follows, 2005).

Here, we develop a new diagnostic that aims to correct the bias in AOU estimates described above. We refer to this diagnostic as ``Evaluated Oxygen Utilization'' (EOU). The EOU is based on the assumption of an isopycnal propagation of the surface ocean oxygen disequilibrium into the interior of the ocean. 'Our new approach is tested in an ensemble of six different coupled biogeochemical physical models in order to obtain a better confidence of the methods validity. In these models, the oxygen respired since the water's last contact with the atmosphere (True Oxygen Utilization, TOU) is computed explicitly using an idealized passive tracer of preformed oxygen (Ito et al., 2004). Finally, AOU and EOU diagnostics are computed from a global observational dataset, allowing the concentrations of regenerated and preformed phosphate to be determined.'

2. The result section ('Computations of AOU, EOU and TOU') lacks of details and the description of the figures could be much sharper. There is enough room in a journal such as Biogeosciences to discuss the results in more detail (see specific comments below).
We agree with the reviewer. The discussion of the results will be extended (see specific comment below p2250) and the Appendix moved to the main text. We will extend Section 3.3 'Sources of the EOU biases' where we discuss the differences EOU – TOU in the models. See the draft of this text in our response to reviewer 1 (page 1ff of this rebuttal).

3. While I appreciate the effort to come up with a new approach to estimate the oxygen utilization rate, I am not really convinced about the robustness of the results when applying to observational-based data. As pointed out by the anonymous reviewer #1, the uncertainties in the observational-based EOU estimates have to be addressed in much more detail. Even if the EOU method works well in a model framework, it does not mean that it also works in the real world.
We agree. As mentioned under (2) we will provide an evaluation of the EOU uncertainties in the models as a new Section 3.3, and concerning the observations as a Section 4.1. See the drafts of these texts in our response to reviewer 1 (pages 1-6 of the rebuttal).

4. The models used in this study are very coarse, and more up-to-date CMIP5-type (ocean) models are available. While I know, that the CMIP5 model output does not include an idealized preformed oxygen tracer, it might be helpful to discuss in more detail the limitations of the six ocean models used in this study. How do they compare with other models (e.g. used in Cocco et al. 2013)? The authors claim that the six models realistically represent the combined effects of circulation and biology (p. 2250 l. 4). Having a closer look at Fig. 1, you need to convince me. What is the most critical thing to get right in the models, to successfully represent OUR with EOU? Why are the EOU estimates so different in the different models when comparing with TOU?

We agree with the reviewer that it would have been a great opportunity if the CMIP5 models would have included an idealized preformed oxygen tracer, as suggested in fact in an earlier model inter comparison paper by Najjar et al, 2007. Such a tracer, together with a reasonably long spin-up, could have helped a lot in evaluating CMIP5 models and their respective deficiencies in ocean biogeochemistry. In addition it would have allowed us to use these models for our work. Concerning our study, however, it is important to note that the principal aim is not that of a model intercomparison study of some models. It is hence outside the scope of our work to quantify in detail model biases using different metrics, as it has been done in Cocco et al., 2013 concerning the CMIP5 models.
The scope of our work is, as well outlined in the introduction, to develop and test a new approach for the quantification of oxygen utilization in the ocean, with the ultimate goal of applying this tested method to observations. We think that for such a method development and testing a reasonably sized ensemble of different models is a very suitable approach. Perhaps, one could state, that the more different the models (and their O2 and AOU distribution), the better, i.e. the more rebust the result of method testing. Given all the discussed variability between our models, it is an important result of our study that in ALL cases the new method is a clear improvement compared to the AOU in estimating TOU. The sources of EOU - TOU differences in the models will be however discussed in more detail in the extended section 3.3 (see page 1 of this rebuttal)

Specific comments:

p.2246 l. 7. 'twice as well' please specify if globally or regionally.
We specified that the EOU approximation performs twice as well globally AND regionally. The final sentence becomes: 'We develop a new approach that we call Evaluated Oxygen Utilization (EOU), which approximates globally and regionally the TOU at least twice as well as AOU in all 6 models'

p.2246 l. 13. 'respiration is a key biological process in the ocean' . Please extend the introduction to put your study into a broader context.
The introduction will be extended (see draft text above)

p. 2246. l. 18-22. Be more specific. What about supersaturation? Vast areas in mid-to-low latitudes are slightly supersaturated (see Fig. 3.1.1 in Sarmiento and Gruber 2006).
The reviewer is right that vast surface areas in mid to low latitudes are supersaturated, in particular during summer. Photosynthesis and radiative warming below the mixed layer are driving this supersaturation. However, these waters (temporarily) affected by oxygen supersaturation contribute very little to the ventilation of the intermediate and deep ocean. We will discuss this in the extended introduction, see the draft text above.

p. 2246 l. 19. What sort of physical and chemical processes?
As outlined above (see drafted new version of the introduction), this will be presented in more detail.

p. 2246. l. 26. What was the motivation to use six different ocean models? Please specify
The original sentence 'Our approach is tested in six coupled biogeochemical physical models' has been rewritten to be more explicit (see last paragraph of new introduction; our response to Rev. 2, Major Comments, 1.). Concerning the motivation of this study see also our response to Rev. 2, Major Comments, 4.

p. 2248 l. 12 Why do the authors take annual-mean model output? I assume that monthly output should be available for at least a subset of models.
Our ultimate goal is to apply the EOU approach to observations. These, however, are very scarce in winter in high latitudes where most of the waters inhabiting the deep ocean forms. This will be presented in more details in the new Section 4.1 (see draft in response to Rev. 1), where we will provide global maps of the availability of oxygen saturation data per 1x1° grid cell on an annual base and for Southern Ocean winter months only. From the latter map it becomes clear that the monthly-interpolated oxygen saturation database of WOA is based on a lot of seasonal and regional interpolation when it comes to the Southern Ocean in winter. Hence monthly output might help with the models, but would provide a misguiding test since the monthly gridded climatologies (data) are unreliable in the Southern Ocean outcrop regions during winter. That is in fact why we decided to combine an intercept-depth horizon of 50m and the annual database when we estimate the winter-time oxygen disequilibrium. The annual database comprises much more data, even in the SO, and the chosen depth gives a fair estimate of winter conditions. These aspects of our method will be discussed in more detail in the new sections 3.3 and 4.1 (see draft page 1-6 of this rebuttal)

p. 2249. l. 7. How large is the drift in the control simulations of the forward models?

The models do not display a significant drift. We agree that the statement of model spin-up being > 2000yrs was partly misguiding. We will be more specific and will provide the actual spinup-conditions and times in Table 1. In fact, 2000 yrs is well above the circulation age of the oldest waters observed in the North Pacific, see e.g. Matsumoto et al., 2007 (their Figure 4, maximum average circulation age below 1500m is < 1200 yrs). Given this, the minimum spin up times needed for our study, depends also very much on the initial conditions of the oxygen and preformed oxygen tracers. In some of the model runs (TMM MIT2.8, TMM ECCO) the status of the spin-up this was further tested by implementing both a tracer for preformed oxygen and a tracer for utilized oxygen and running the model until the sum of both tracers equals the concentration of the oxygen tracer in basically any model grid box. UVIC and MOM3-P2A have been run for 5000 years. CSIRO has been run for 6000 years. om1p7-BLINGv0 has been ran for 2100 years and has been initialized by the WOA dataset, thus allowing a relative fast spinup.

p. 2249. l. 10. Why not using the improved O2 data set from Bianchi et al. 2012 (GBC)?
The main differences between the climatology of Bianchi et al. (2012) and the WOA are located in the suboxic regions, which are not discussed here. The concentration of oxygen in the water-mass formation regions are very similar in both datasets.

p. 2249. l. 22. 'sluggish' be more specific.
Sluggish is possibly misleading and debatable, as pointed out also by the reviewer 1. The equatorial current system is indeed very strong. However, the thermocline is shallow in tropical regions compared to the gyres, which are well ventilated and characterized by a circulation extending down to 300-400 m depth. The 200-1000 m depth range in the tropical oceans can be better characterized by having 'limited contact with the atmosphere and high residence time' (see our reply to reviewer 1 for the complete reworded sentence).

p. 2250. l.1-5. What about the deficiencies in the models? Please specify them and discuss the implications for the EOU estimates.
The sources of EOU - TOU differences in the models are discussed in Section 3.3 (see page 1ff of this rebuttal) and uncertainties associated with the EOU approach are presented in the new section 4.1 (again see response to reviewer 1 for a draft text).
Concerning the overall deficiencies of the different models we plan to add the following paragraph at the end of Section 3.1 'AOU patterns in observations and models'

"However, some regional differences exist between the models and observations. The most striking difference occurs in the tropical and north Pacific Ocean. In MOM P2A and MIT28, the AOU is significantly lower than in WOA (220 to 240 mmol.m-3) in this region, whereas the volume of high AOU (greater than 280 mmol.m-3) is larger in om1p7-BLING than in WOA. The other models (CSIRO, UVIC and ECCO2) display intermediate values. These differences may be due to a variety of processes, including inter-model differences of oxygen disequilibrium at the outcrops, model circulation (i.e. circulation ages), export production patterns and carbon decay length scales implemented in the individual models. Disentangling these processes would require a suite of sensitivity experiments for each model and with additional tracers (e.g. ideal age) and is beyond the scope of this work."

It is important to note here that deficiencies in the models do not necessarily affect our method testing since we have ground truth information (i.e. the explicit tracer of preformed oxygen) for each model and evaluate the model estimates of EOU always against this ground truth information. Having models with variable deficiencies of ocean circulation, export production and surface ocean processes in the outcrop regions (for neither of which a good ground truth is available), but all providing EOU being at least twice better than AOU in estimating TOU, in fact, demonstrates the robustness of our method.

p. 2250 l 14-15 'O2 pre is always less than O2 sat ' What about the negative values in the low latitudes?
The sentence will be re-worded: 'Because O2pre is less than O2sat in most of the ocean interior, TOU is less than AOU and the AOU – TOU difference is positive. Slightly negative values of AOU-TOU are observed in tropical-subtropical waters of the upper thermocline of some of the models corresponding to very limited sinking of oversaturated water masses. '
Note also that we will explicitly mention the issue of supersaturation in the new introduction.

p. 2250. l. 16. Why does om1p7- BLINGv0 poorly represent AOU in the Southern Ocean and the deep Pacific Ocean
From the AOU-TOU differences it is obvious that om1p7-BLINGv0 is characterized by a particularly large Southern Ocean oxygen disequilibrium. One possible explanation for this is that om1p7-BLINGv0 has a too strong organic matter production at the surface and remineralization of it at depth, which in combination with mixing of surface and subsurface waters may explain the strong surface disequilibrium. Alternatively, a slow subduction rate would also favour high AOU, but would allow for more time for surface waters to equilibrate, possibly compensating for the effect of surface oxygen disequilibrium. The EOU – TOU difference is larger in om1p7-BLINGv0 than in the other models, however the ratio of correction is similar to other models (see figure 3 of the original submission to BGD). For instance the improvement in the cost function is three times at global scale, both for om1p7-BLING and CSIRO.
As a note of caution, we stress that a large difference between AOU-TOU in any model does not mean that that model poorly represents oxygen utilization compared to that of the real ocean. It rather indicates that in that model the winter-time disequilibrium is particularly large. However, since we do not know well how large the disequilibrium in the real ocean is (see the limited data coverage during winter in the SO), we can't say for sure whether om1p7-BLINGv0 poorly represents the oxygen utilization in the Southern Ocean or the deep Pacific, or not.

p. 2250 l. 15 'Maximum AOU-TOU differences ...' Do you mean along the transects investigated in Fig. 2?
The maximum AOU – TOU differences described here are indeed along the transects investigated in Fig 2. In deep waters the zonal components of the currents are generally stronger than the meridional components, at least away from continental margins, and are more effective at homogenizing water properties. This will be specified in the text.

p. 2250 l.18. Be more specific where Ito et al. (2004) found their maximum O2 disequilibrium.
The sentence: 'This range of values encompasses the maximum oxygen disequilibrium of 73 mmol.m-3 found by Ito et al., 2004' will be complemented by: 'These authors found a maximum oxygen disequilibrium in the Southern Ocean region (at least 50 mmol.m-3 south of 60S) and the bottom waters of the ocean (at least 30 mmol.m-3 deeper than 2500 m). The location of their maximum disequilibrium is coherent with the results of our study'.

p. 2252 l. 12-20. This can clearly be tested by using monthly model output instead of

annual mean output.
p. 2254. l. 12: How sensitive are the results to the choice of the depth horizon? Give some uncertainty estimates.
Both aspects will be discussed in the extended/new section 3.3 and 4.1 (see page 1ff of this rebuttal).

p. 2256. l. Do the results shown in Fig B1 represent single year values? If yes, the results may just show one single large convection event in the Southern Ocean.
The results shown here represent the annual mean of the last year of the spinup period. The spinup has been conducted using climatological forcing. The only source of interannual variability would be internal oscillations, which we do not notice in the models studied here.

p. 2265. Fig 4a (right) Please adjust color bar to make differences visible.
The color bar has been adjusted (see the new figure in the reply to the reviewer 1) .

References

Only the references that we added in this rebuttal are displayed below. Please refer to the original manuscript for the remaining references.

Aristegui, J; Duarte, C. M; Agusti, S; Doval, M.D ; Alvarez-Salgado, X. A.; Hansell, D. A. (2002): Dissolved organic carbon support of respiration in the dark ocean. Science, 298(5600), 1967, doi:10.1126/science.1076746

Broecker, W. S., Takahashi, T., & Takahashi, T. (1985). Sources and flow patterns of deep-ocean waters as deduced from potential temperature, salinity, and initial phosphate concentration. Journal of Geophysical Research: Oceans (1978–2012), 90(C4), 6925-6939.

Broecker, W. S., Peacock, S. L., Walker, S., Weiss, R., Fahrbach, E., Schroeder, M., Mikolajewicz, U., Heinze, C., Key, R., Peng, T. H., and Rubin, S.: How much deep water is formed in the Southern Ocean? J. Geophys. Res., 103(C8), 15833--15843, 1998.

Carlson, C.A.,Hansell, D.A.,Nelson, N.B.,Siegel, D.A.,Smethie, W.M.J.,Khatiwala, S.,Meyers, M.M.,Halewood, E.,(2010). Dissolved organic carbon export and subsequent remineralization in the mesopelagic and bathypelagic realms of the North Atlantic basin. Deep-Sea Research II57(16),1433–1445.

Craig, H., T. L. Hayward (1986). Oxygen supersaturation in the ocean: biological versus physical contributions, Science, 235, 199–206

Doval, M. D. and Hansell D. A. (1999). Organic carbon and apparent oxygen utilisation in the western South Pacific and the central Indian Oceans. Marine Chemistry, 68(3), 249-264, doi:10.1016/S0304-4203(99)00081-X

Emerson, S., Y. W. Watanabe, T. Ono, and S. Mecking (2004), Temporal trends in apparent oxygen utilization in the upper pycnocline of the North Pacific: 1980– 2000, J. Oceanogr., 60, 139–147.

Johnson, G. C. (2008), Quantifying Antarctic Bottom Water and North Atlantic Deep Water volumes, J. Geophys. Res., 113, C05027, doi:10.1029/2007JV004477

Khatiwala, S, Primeau, F., and Holzer, M.(2012). Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age. Earth and Planet. Sci. Lett., doi: 10.1016/j.epsl.2012.01.038.

Koertzinger, A., Hedges, J., and Quay, P.: Redfield ratios revisited: Removing the biasing effect of anthropogenic CO2 , Limnol. Oceanogr., 46, 964--970, 2001.

Li, Y. H., Karl, D. M., Winn, C. D., Mackenzie, F. T., and Gans, K. (2000) Remineralization ratios in the subtropical north Pacific gyre, Aquatic Geochem., 6, 65—86.

Matsumoto, K. (2007), Radiocarbon-based circulation age of the world oceans, J. Geophys. Res., 112, C09004, doi:10.1029/2007JC004095

Najjar, R. (1992). Marine Biogeochemistry. In: Climate System Modeling, Trenberth, K.

(ed.), Cambridge University Press, Cambridge, England, 241-280.

Najjar, R., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S., Dutay, J., Follows, M., Gruber, N., Joos, F., Lindsay, K., Maier-Reimer, E., Matear, R., Matsumoto, K., Monfray, P., Mouchet, A., Orr, J., Plattner, G., Sarmiento, J., Schlitzer, R., Slater, R., Weirig, M., Yamanaka, Y., and Yool, A.(2007). Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2), Global Biogeochem. Cy., 21(3), GB3007, doi:10.1029/2006GB002857

Nevison, C., Butler, J. H., Elkins, J. W. (2003), Global distribution of N2O and the Delta N2O-AOU yield in the subsurface ocean. Global Biogeochemical Cycles, 17(4): 1119.

Ogura, N., 1970. The relation between dissolved organic carbon and apparent oxygen utilization in the Western North Pacific. Deep-Sea Res 17, 221–231.

Sarmiento, J.L., G Thiele, R.M Key, and W S Moore (1990), Oxygen and nitrate new production and remineralizaion in the North Atlantic subtropical gyre. Journal of Geophysical Research, 95(C10), 18,303-18,315.

Takahashi, T., Broecker, W. S., and Langer, S. (1985), Redfield ratio based on chemical data from isopycnal surfaces, J. Geophys. Res.-Oceans, C4(90), 6907—6924.