

## Reply to comments by Referee #1

I think in general the study is worthwhile, and could be important to gas and biological studies. However, the study has some serious limitations, and the authors either need to extend their analysis or restrict some of their assertions.

**Response:** In the revised manuscript we will narrow our assertions where necessary to indicate that the suggested criterion is applicable for the Bellingshausen Sea region with potential application to other regions of the Southern Ocean and the global oceans.

My main concern is that the small scope of the work limits its applicability to other studies. The authors developed their method with 251 profiles from a coastal region near the Antarctic Peninsula, yet they asserted that their method could be used in the entire Southern Ocean. They need to provide evidence to support this assertion. Showing that their method works for a much wider range of profiles from the entire Southern Ocean would greatly expand the impact of the paper.

**Response:** The paper indeed focuses on the Bellingshausen Sea only, as the title implies. We will make this now even clearer at various places in the revised paper, as appropriate. It will be very interesting to test the new O<sub>2</sub>-based criterion in other regions of the world's oceans, and we suggest this in the paper, but this is beyond the scope of the present study.

The authors claim that because the O<sub>2</sub> profile depends on biology, it gives a more complete picture of all relevant processes occurring in the mixed layer (compared to temperature or density), therefore allowing for a more accurate MLD calculation. Couldn't biology also obscure the MLD?

**Response:** A situation when this would occur is difficult to imagine. The metabolic balance in the subsurface ocean is usually net heterotrophic, i.e. respiration exceeds production. Below the mixed layer, this leads to a general decrease of O<sub>2</sub> concentrations with depth. In some situations, the productive euphotic zone may extend below the mixed layer, leading to a build-up of O<sub>2</sub> below the mixed layer. In this case, our criterion of a 0.5 % relative change would still work, but in the opposite direction. Only if the concentration below  $z_{mix}$  was the same as in the mixed layer could a situation occur when biology could obscure the mixed layer for a short period of time when production, respiration and vertical mixing were in perfect balance.

In Figure 5, many of the MLDs look pretty close to 10 m depth. Does choosing a shallower reference depth change the MLD?

**Response:**

Yes, it changes  $z_{mix}$  slightly, due to presence of Winter Water (WW) intrusions in the top 10 m of the water column. However, these intrusions could mask the true seasonal  $z_{mix}$  and 10 m should therefore be considered an upper limit of  $z_{mix}$ . In the revised manuscript, we will add a new figure (Figure 3 in the revised manuscript and enclosed to this comments) where comparison between choosing a near-surface reference depth ( $\approx 2$  dbar) against the 10 m reference for the mixed layer-O<sub>2</sub> criterion will be depicted.

How does changing the relative difference criterion change the MLD distributions?

1511-15 by objective, numerically determined  $z_{mix}$  to identify a suitable O<sub>2</sub> . . . Do you mean you tested a number of different criterion, compared them to the visually identified MLDs, to settle on 0.5%?

**Response:** A range of threshold values was tested using 0.1 %, 0.5 % and 1 % relative differences with respect to  $c(O_2)$  nearest the surface ( $\approx 2$  dbar) and at 10 dbar. The results (i.e. subjective  $z_{mix}$ ) were compared to the mixed layer determined by visual inspection, and the 0.5 % criterion with a 10 dbar reference depth agreed best with the values obtained by visual inspection and was therefore adopted. We will add detailed information regarding this in the revised manuscript in section 2.2. Figure 3 will be replaced with a new version that includes the mean difference between  $z_{mix}$  determined by visual inspection and  $z_{mix}$  obtained

after applying different threshold O<sub>2</sub> criteria, both with respect to the near-surface value and 10 dbar.

1506-9 different than

**Response:** Will be corrected

1506-13 I don't think that I would say that the criteria was established by numerical analysis based on your description

**Response:** We will delete the word "numerical".

1507-21 has an extra period

1508-15 I might say near-surface reference value

1508-16 Previous studies have found that well-resolved vertical profiles are necessary to use gradient-based criteria successfully.

1510-3 need a space in O<sub>2</sub> profiles

1511-12 I would move some of the discussion of number of profiles/quality control into the CTD acquisition section

**Response:** All will be corrected as suggested.

1511-15

**Response:** See our response above.

1511-21 'allow comparing' sounds odd

**Response:** It is grammatically correct to use allow with gerund or infinitive and there is no difference in meaning.

1511-26 confusing description of the various comparisons in the introduction to this section.

1512-1 selected three-widely

**Response:** This sentence will be reformulated as follows: "To test our new  $z_{\text{mix}}$  criterion, we first compared it with conventional  $z_{\text{mix}}$  definitions based on temperature and potential density. These criteria were adopted from three widely used  $z_{\text{mix}}$  climatologies (Table 1) and applied to the 251 CTD profiles of this study. In a second step, we compared  $z_{\text{mix}}(\text{O}_2)$  to the climatological  $z_{\text{mix}}$  values, interpolated according to location and time of year."

1512-21 is BM04 temperature criteria in relation to the surface value? Also maybe specify which criteria come from which study

**Response:** We will clarify this as requested. The following sentence will replace the previous one in the revised manuscript: " $\Delta\theta = 0.5$  °C and  $\Delta\sigma_\theta = 0.125$  kg m<sup>-3</sup> with respect to the surface value (ML97);  $\Delta\theta = 0.2$  °C and  $\Delta\sigma_\theta = 0.03$  kg m<sup>-3</sup> with respect to the 10 dbar value (BM04)."

1513-21 Is it fair to use the subjective MLD in the comparison with the other methods?

**Response:** We cannot think of an alternative way to establish potentially suitable  $z_{\text{mix}}$  criteria other than the human eye and brain. Presumably, the same route was chosen to define previous temperature- or density-based criteria. Of course, the detailed evaluation of potentially suitable criteria and their comparison to other methods then has to be (and indeed was) done using numerical algorithms. Whether the subjective or objective  $z_{\text{mix}}(\text{O}_2)$  values are used for comparison purposes is a matter of taste, but following similar queries from both reviewers, we are now using the objectively defined  $z_{\text{mix}}(\text{O}_2)$  values for comparison. Previous mean values will be updated by the comparison to objective  $z_{\text{mix}}(\text{O}_2)$  in the body text accordingly, and in Table 2.

1513-23 To me, this paragraph belongs in the next section.

**Response:** Paragraph will be moved as suggested.

1514-9 You extend your results to apply to the southern ocean, a much bigger area than the scope of the study. You could cite other studies that have shown this as well.

**Response:** As mentioned above, we decided to keep the paper as it is for the region of study only since it is not intended to describe a mixed layer depth criterion for the entire Southern Ocean.

1514-23 You've still had to define the 0.5% difference and the reference depth, so to me it is actually more similar to the threshold methods than Lorbacher

**Response:** The reviewer is correct that we have to define a threshold criterion and reference depth, but our point was that "Compared to difference criteria, [our] approach has the advantage of being independent of the actual value of the variable in question.", which appears to be valid, so there does not appear to be a need to change anything.

1515-11 maintain consistency

1515-16 remove 'were done.'

1515-26 In case of BM . . . sounds odd

1516-1 should be BM04. I would add that you are comparing density to density in this paragraph. It's a little confusing.

1516-18 than the other climatologies

1517-4 should be as instead of than

1518-28 lower, or deeper?

1519-18 influence production calculations

**Response:** All corrected as suggested.

1519-19 'resolution of the instrumental parameter used in the criterion to define it' is a confusing phrase. The second sentence could also use some simplifying/clarification.

1519-27 interpolation method, or the vertical resolution of the fields

**Response:** The paragraph will be reformulated as: "The accuracy of  $z_{\text{mix}}$  defined using a threshold criterion depends on the resolution of the hydrographic parameter chosen for the  $z_{\text{mix}}$  definition. Modern CTD observations provide sufficiently high enough resolution to resolve stratification in the upper water column. However, the coarser resolution of current climatologies and WOA data make them less suitable to establish  $z_{\text{mix}}$ . Furthermore, the low abundance of  $\text{O}_2$  profiles in Southern Ocean climatologies means that  $z_{\text{mix}}$  obtained from these data collections are unreliable when compared to results based on CTD- $\text{O}_2$  profiles. The difference between  $z_{\text{mix}}(\Delta\sigma_\theta)$  and  $z_{\text{mix}}(\text{O}_2)$  for CTD profiles and WOA05-profiles may often be due to the vertical resolution of the data and the interpolation method used to construct the temperature, salinity and  $\text{O}_2$  fields in WOA05."

1520-1 sentence needs work

**Response:** The sentence will be rewritten as: "For the present work, CTD- $\text{O}_2$  concentrations were calibrated against Winkler titrations before determining  $z_{\text{mix}}(\text{O}_2)$ . However, the latter criterion can be also applied to non-calibrated  $\text{O}_2$  profiles because sensor gain errors cancel out thanks to the relative calculation method. Of course, data affected by noise or ship motion still have to be removed first."

1520-25 I'm not convinced of this. You haven't shown evidence that this criterion would work outside of the very small area of study in the paper.

**Response:** This will be rephrased as: "For gas exchange studies,  $z_{\text{mix}}(\text{O}_2)$  has the advantage of being directly related to a species of interest. Moreover, the relative nature of  $z_{\text{mix}}(\text{O}_2)$  criterion proposed means that its applicability should be tested in many other parts of the worlds' oceans, including at other times of the year. The proposed criterion is more sensitive to reflect better upper mixed layer air-sea dynamics and influence of biological and physical processes, rather than the traditional criteria based on potential temperature or density, particularly in regions where weak vertical gradients of temperature and density in the upper waters are suspected. Thus, the criterion proposed here can be used as a baseline for  $z_{\text{mix}}$

definitions in other coastal areas of the Southern Ocean, especially for gas exchange studies."

Figure 4 - I'd say potential density differences, not salinity differences

**Response:** The reviewer is correct, this will be corrected in the revised manuscript.