We thank the referee for his comments. Below is our response. Blue text signifies a direct quote from the referee's comments.

1. Overall the method used to estimate water mass ages from 3He/3H by using a transient time distribution TTD approach is sound and well described. Tracer ages derived from different tracers present differences that depend on the tracer boundary conditions, sources and sink distribution, and the characteristics of the flow (e.g. Waugh et al., 2003). However, the first moment ('mean age') of the transit time distribution, is an intrinsic characteristic of the flow, and should be independent of the tracer considered. Stanley et al., adopt the 'mean age' to calculate AOUR from AOU (equation 7). Whereas I have no major concerns about this method, I think it would be worth clarifying the assumptions that allow the use of the 'mean age' in equation 7. Ideally – if it was available – one could use an 'AOU age' – depending on both circulation and oxygen sources and sinks – that is the tracer age that would be inferred from AOU measurements if one were to exactly know the sources and sinks of oxygen in the ocean interior. Put it another way – what is the correct age to be used in equation (7), and to what degree is the 'mean age' a good approximation to it? Similarly to the transient time, the AOU of a water parcel should be interpreted in a probabilistic way, and not necessarily the density distributions of transient times and AOU would coincide. These are points beyond the scope of the paper but it would be useful to see them discussed with more detail.

The referee has a very good point. What we are measuring is the probability distribution of transit times, not the probability distribution of AOU. However, we have no way of measuring the probability distribution of AOU directly. We have added to the discussion a statement to this effect.

2. Page 9982, equation (2): I am surprised that in the updated source function for tritium no confidence bounds are provided for the regression coefficients. This should be straightforward to include. Additionally, they could be included in Figure 1, which perhaps would benefit from being extended to the 1950s, to include Dreisigaker and Roether (1978) source for BATS. On a note, the empirical source function is not strictly speaking an exponential (page 9982, line 21), but the sum of an exponential and a linear trend.

We calculated the uncertainties in the parameters from a Monte Carlo simulation where we simulated data using the measurement uncertainty and performed the curve fit with the new suite of data. We reported those uncertainties in the paper. The confidence bounds are so similar to the

curve that they do not show up in Fig. 1, i.e. the width of the confidence bounds is the same as the width of the curve. A much larger uncertainty than that in the curve-fitting stems from the fact that the water at depth at BATS did not surface necessarily at BATS and there are small latitudinal differences in the source function. We modified the figure to have an inset showing the source function back to the 1950s. We changed the wording from exponential to combined exponential and linear.

3. Page 9983, line 19 and following. The TTD is by definition the Green's function of the advection-diffusion operator that propagates surface boundary conditions into the interior. The specific form for the Green's function chosen by Stanley et al. after Waugh et al. (2003) is an inverse Gaussian function. The assumption of inverse Gaussian TTDs for the u pper ocean in the subtropical gyre seems reasonable, given the patterns of water-mass circulation that characterize the region. Nonetheless, the section would benefit from a discussion of why this specific TTD has been chosen.

We have added some lines discussing the choice of an inverse Gaussian.

4. Page 9984, line 8-9: this sentence is incomplete. What is the assumption adopted here?

We fixed that.

5. Page 9984, line 22. It should be clarified why the 3He data give a more precise and robust determination of Gamma_best than 3H. I find it confusing since from Fig. 2 and section 3.2 it seems that in the upper water column the relative errors on 3H are smaller than the errors on 3He, and the equations used for the convolution should be similar. Also, why is it not possible (or worth) using 3He and 3H simultaneously to estimate Gamma_best?

For a given ^{3}H source function, model tritium concentrations are relatively insensitive to variations in mean age ($d^3H/d\tau$ is small) for ventilation time scales (i.e., $\tau < 20$ years or so) short relative to the time duration since the bomb-tritium injection event in the early 1960s. Thus, the major factor determining model ^{3}H concentrations in the upper few meters is the magnitude of the surface ^{3}H source function. In contrast, model 3He concentration varies with both mean age (tau) and ^{3}H concentration (and thus source function) and thus leads to more robust ages. We tried using ^{3}H and ^{3}He simultaneously to estimate Γ_best but again, because of the sensitivity of ^{3}H to the source function, this was not an effective approach. We have added an explanation along these lines to the paper.

Section 3

6. Page 9987, line 5. The reference to mean age (tau) variations in figure 5b is misleading, since figure 5b shows AOUR and not mean age values.
We have removed reference to the figure.

7. Page 9987, line 13-15. The sentences 'The box model approach has an implicit exponential shape to the water mass probability distribution' and 'the TTD model . . . is mixing waters with a larger age spread and has a non-zero centroid' are confusing. Could these be clarified or rephrased?
We rephrased the sentences.

Section 4
8. Page 9991, line 18 and following. It should be clarified that the transport matrix method referenced in Kathiwala (2007) is based on model simulations, and the water mass contributions determined with this method depend on model-simulated circulations (albeit the transport matrix method has been applied to data-assimilating models). The total transport matrix method detailed in Gebbie and Huybers (2010) is perhaps a more relevant reference that is completely based on tracer observations. See for example figure 9 in Gebbie and Huybers (2010).

We thank the referee for pointing out the information in the Gebbie and Huybers (2010) paper. Indeed, Fig. 8 in that paper gives another way of arriving at the same conclusion. We have added a reference to the Gebbie and Huybers paper. We continue to have the Khatiwala reference since Khatiwala produced maps for us specifically at the BATS site to unambiguously show us where the water is coming from. We have added in the text that the Khatiwala approach is model-based whereas the Gebbie and Huybers approach is tracer-based.

9. Section 4.3. The central point of this section is the suggestion that methodological artifacts are responsible for the increase in AOUR from the 1970s-1980s to the 2003-2006 periods. Whereas I find the combination of figure 7.b and 8.b suggestive of the possibility of biases in earlier O2 measurements, I do not think the evidence is strong enough to conclude that ". . . the apparent differences in AOUR between 2003 and 2006 and the 1980s . . . is likely due to methodological artefacts". I do agree that the result calls for both caution in the interpretation of O2 time series and further analyses of earlier O2 measurements. In particular inspection of figure 8 alone does not fully convince me that the O2 difference is completely a methodological artifact. Figure 8 and the discussion in section 4.3 do not allow to assess whether there is a bias in the late 1980s measurements at Station S (blue circles), or a bias at Station 50 on the Endeavour 129-1 leg. Measurements at Station S show a decreasing O2 trend between mid 1980s and mid 1990s when they start to overlap with BATS measurements. Yet no indication is provided as to what methodological artifact could be responsible

for this decrease, and why BATS measurements should be trusted more than Station S measurements. Without a detailed intercomparison of O2 measurements from the different programs – including considerations on the analytical techniques, formal statistical time series analysis on O2 (for example change-point detection), and analysis of the variability of the regional hydrography – the hypothesis of methodological biases in early Station S O2 is speculative. See also comment 11.

Referee #1 thought our statement that our statements about the likelihood of methodological artifacts in the old Station S data were not strong enough, i.e. the data was certainly flawed and therefore we should not be using it at all. Referee #2 thinks that our statements are too strong – that perhaps the data is not flawed. Thus the paper's stance falls in the middle between those two although closer to the side of Referee #1. Referee #2 says that the figure is not convincing since there is no reason to believe that the BATS or Endeavor measurements are correct and the Station S is wrong. We disagree. First, Station S measurements are still being made and in the period after 1992, they agree with BATS measurements. We have modified the figure to include all the Station S data to show this (we had initially only plotted the early Station S data so the figure would not be too crowded but in light of the reviewer's concerns, we decided more data would make the point more definitively). So from the revised figure, it is clear that after 1992, the oxygen concentration on 2100 to 2700 m stabilized between approximately 258 and 265 umol/kg. In contrast, in the period between 1970 and 1990, the deep Station S data ranged from 252 to 290 umol/kg. The oxygen concentration of the deep waters should not change. Ocean deoxygenation usually refers to water in the upper thermocline. So the fact that the deep early data, i.e pre-1990 data, is different than the deep later data suggests that one set of data may be suspect. The early data is not only higher but it is more variable, suggesting it is likely the culprit. If the variability was real, it should be negatively correlated with temperature and salinity. We did a correlation analysis between the oxygen concentrations and salinity and show that prior to 1988, there is no significant correlation (R^2=0.03, P=0.144) whereas after 1992 there is a weak but significant negative correlation (R^2=0.1, P=0). Additionally, the Endeavor data agrees with the later deep O2 data, again suggesting that it is the early data that is flawed. The referee asks about what could be responsible for change in O2 and why BATS measurements should be trusted more. The answer is that in the early 1990s, an intercomparison study was done to try to standardize oxygen measurements (unfortunately this work was not published but I have talked to Jim Swift who was part of it and he has confirmed it). At that point, the protocols at BATS/Station S were revised and data from that point on at both BATS and Station S becomes much less variable. There are a number of known analytical problems that can easily introduce a positive bias in oxygen measurements. The early oxygen data could be flawed is that the samples were collected in Nansen bottles, a less gas-tight system than current-day Niskin bottles. Additionally, if bubbles got in during sampling or the solutions necessary for titration were not made correctly – or oxygen was entrained when dispensing those solutions, then the data is

flawed. Oxygen is difficult to measure and in the early days, not enough was known about all the potential pitfalls of the measurement. We have added a discussion along these lines to the paper.

10. Section 4.4. I wonder if it would be possible to include the information provided by Delta, the width of the TTD distribution, in the estimate of the AOUR uncertainty. In a sense, the knowledge of the TTD should allow the estimate of a density distribution for AOUR (ignoring the issue of the density distribution of a water parcel AOU). For example, a TTD with a wider Delta should imply a larger range of uncertainties for the AOUR.

A TTD with a wider delta does not imply a larger range of uncertainties for AOUR. Instead, it implies a more diffusive ocean. In turn, this means that a water parcel would be exposed to a wider geographical area and thus potentially to more different biogeochemical regions. Hence perhaps a wider delta could be considered to add to uncertainty in the *interpretation* of AOUR if one tries to interpret AOUR in a strictly vertical framework. But it does not mean a greater uncertainty in terms of absolute numbers as is given in the error estimate.

Conclusions

11. Page 9999, lines 3-5. As noted in point 8., I do not feel that at this stage the C4822 evidence provided is sufficient to conclude that "this increase is due to an increase in AOU and is more likely associated with methodological artifacts in the oxygen data from the 1980s". However, this is an important possibility that the data do suggest and that should be further investigated. I suggest that the sentence be rephrased.

For the reasons outlined in response to comment 9, we firmly believe that the most likely cause is methodological artifacts. Thus we are leaving the statement as is. We have changed the text in section 4.3 to give more evidence why methodological artifacts are the most likely culprits. The statement in the conclusion does say "more likely" i.e. it does not claim that methodological artifacts are certainly the cause of the change. As we mentioned earlier, Referee #1 doesn't think the statement, as written, is strong enough. Referee #2 thinks it is too strong. So we are leaving it as we had it.

Figures

12. Fig. 3, Fig. 4, Fig. 7. The figures would be much easier to read with a different aspect ratio – that is a wider x-axis, as most of the information in the upper water column is squeezed to values close to zero. Could the authors re-plot them exanding the x-axis?

We have expanded the x-axis by 20% in those three figures.

Technical comments

1. Page 9979, line 24: I don't think Khatiwala et al., (2009) is the appropriate reference. Waugh et al. (2003) is sufficient.

We removed the Khatiwala reference.

2. Page 9980, line 21: delete additional "the".

Removed.

3. Page 9983, line 20: Green's and not Greens.

Fixed.

4. Page 9984, line 4: the Peclet number should be analogous to (Gammaˆ2)/(Deltaˆ2) and not to (Gamma/Delta) (e.g. Waugh et al., 2003).

Fixed.

5. Page 9988, line 24: sensitivity instead of sensitvty.

Fixed.

6. Page 9988, line 27. Remove the comma after AOUR-derived.

Fixed.

7. Page 9991, line 2-3. Should it be "differences between AOUR and OUR" instead of "Differences between AOUR and AOU"?

Yes it should. We fixed that.

8. Page 9991, line 16. Change "as well as sources from the southern ocean" to "as well as in the Southern Ocean".

Fixed.

9. Page 9994, line 23-24. 1977-1987 instead of 1977-1877.

Fixed.

10. Page 9996, line 4. Add 'the' before 'source function'.

Fixed.