Firstly, the authors would like to thank the two reviewers for their thorough readings of our manuscript and feel that our revised draft manuscript represents a significant improvement to the original. We have tried to answer every point in as much detail as space allows and indeed there is only one point which we feel falls outside the scope of this short paper (point five from referee one).

We now address each of the referee's reports in turn, addressing specific points in the order given.

First referee's report

Major Comments:

1. The authors agree with this point and have added more points to the abstract in a revised draft manuscript.

2. We have added further references, variable dimensionalities and information as requested in this point, including the precise origin of the value of 138/106 as indicated by the referee. Also, to align our paper with OCMIP protocols (and the referee's suggestion), we have relabelled salt as *S* and Schmidt number as *Sc*.

3. All changes and additions from the recommendations in this bullet point have been taken into account and amendments made to the manuscript. Section 2 of the manuscript is now called 'Theory and model description' rather than simply 'Theory' in the original.

4. The reason for using the spatially-incomplete dataset (Helm et al. 2011) in the original manuscript was that a recent paper (Andrews et al. 2013) used it in studying deoxygenation over the last few decades both in observations and in the HadGEM2 model. It was the intent of the authors to keep the model development process fully traceable by using the same target dataset. We have added some text to make this point more explicit although we feel that the original pretext for using the Helm et al (2011) dataset remains valid. We have already begun preparing a revised manuscript and have found that when the oxygen concentrations are studied on the basin scale, using spatially complete oxygen dataset (from the World Ocean Atlas) is necessary because the dataset of Helm et al. does not give high enough quality figures due to the paucity of its coverage. We have noted however (in text and in a new figure in the draft manuscript) that although the World Ocean Atlas does indeed provide a complete latitude-longitude dataset, the constituent observations are far from complete, particularly in the southern hemisphere.

5. The points raised in this bullet point are certainly pertinent and apt but we feel that they lie outside the scope of this paper, particularly when discussing both past and future climate change. To do justice to these point would require a large additional section to be added, which, we feel, would take the paper outside its original remit of a model description paper.

For the *Specific Comments* section, we will address each point in turn using Word's 'track changes' facility, that is, commenting on quoted text from the first referee's report.

Specific comments: p. 1454: l.11-12: This sentence is not relevant for the abstract and can be deleted. p. 1455: l.1: Does the model include an interactive land and/or ocean carbon cycle? Please specify. p. 1455: l. 7: What do the authors mean with "up to date" models? Please specify. p. 1455: l. 12: There are several recent studies that have looked at changes in oceanic oxygen under future climate change. The authors may add some of the studies: e.g. Bopp et al. 2002 or Frölicher et al. 2009. p. 1456: l. 21: What are the units for T? p. 1457: l. 12: Unclear what 'simulator label' means. p. 1457: l. 8-13: This paragraph is unclear to me. Doesn't the HadCM3 represent relatively well observations? What was the reason to compare earlier FAMOUS versions with HadCM3 and not with observations? Please specify. p. 1458: l. 18-20: Why do the authors take an 1870-1880 SST pattern from Rayner et al. 2003 with low coverage instead of a present-day SST data-set? The overall zonal and meridional gradients are similar between present-day and preindustrial and differences between present-day and preindustrial SST shouldn't be an issue for this kind of comparison. p. 1459: l. 16: What's the motivation behind the use of a 2_C threshold? Please specify. p. 1460: l. 4-6: You may add AOU patterns here. AOU may help to further explain the simulated oxygen biases. p. 1461: l. 12-18: What's the reason for the large NPP overestimation in the equatorial Pacific and Atlantic? Please explain. Table 1 caption: Abbreviation for Atlantic Meridional Overturning is usually AMOC. Table 1 caption: Change to 'Note the lack of an error estimate for Talley et al. (2003)' Table 1: What is the difference between the two Atlantic MOC estimates? Please also specify in the Table caption.

Comment [JHTW1]: Done.

Comment [JHTW2]: Some text has been added expanding on this point.

Comment [JHTW3]: The authors have reworded this section and have added some further text below concerning the history and development of the FAMOUS model.

Comment [JHTW4]: The authors are very grateful for these specific recommendations, both of which have been incorporated. The review paper of Peña et al. (2010) has also been cited.

Comment [JHTW5]: The units are Celsius and this has now been made clearer in the text.

Comment [JHTW6]: As stated above in comment JHTW3 above, the text concerning the history and development of FAMOUS has been simplified. The authors are grateful for this being pointed out and we hope that this is now clearer for a more general readership.

Comment [JHTW7]: The authors acknowledge this point and have replotted the observed data to give a decadal mean which is coincident with the simulated data. As the referee states, the precise decadal mean of the observations which is used makes very little difference to the conclusions reached and text has been added to make this point.

Comment [JHTW8]: The original reason for choosing this threshold was that in Figure 2 in the original manuscript, the contour resolution is 2 degrees. As a sensitivity study, we have performed the same calculation using 1 degree and 3 degree thresholds and analogous conclusions were reached in all cases.

Comment [JHTW9]: We think that the addition of AOU figures (which have studied on a global and basin scale) have added a very useful extra dimension to the paper and we are grateful for this suggestion.

Comment [JHTW10]: Some text has been added to the draft manuscript explaining that in the original tuning of this model (Williams et al., 2013), it was surface nitrate concentration (not NPP) which was used as the tuning target in the ocean perturbed physics ensemble used. Fundamentally this NPP bloom is not ideal, however, with the size of the ensemble which was used in Williams et al. (2013) it was felt that this overestimation of NPP was acceptable given the significant improvement in model climatology obtained across many meteorological metrics.

Comment [JHTW11]: Changed.

Comment [JHTW12]: Changed.

Comment [JHTW13]: Changed.

Figure 3: Narrow the x-axis range. Reduce it to 150 to 400 umol/l or so, to highlight the important part of the oxygen range. Figure 4: Interestingly, both the HadGEM-ES and FAMOUS largely overestimate O2 concentrations from 0_ and 20_N and between 100m and 1500m depth. Any ideas why this might be the case? Figure 8: This figure can be deleted.

References: Anderson et al., 1994: Redfield ratios of remineralization determined by nutrient data analysis, Global Biogeochem. Cycles, 13, 337-349. Bopp et al., 2002: Climate-induced oceanic oxygen fluxes: Implications for the contemporary carbon budget, Global Biogeochem. Cycles, 16(2), 1022. Frölicher et al. 2009, Natural variability and anthropogenic trends in oceanic oxygen in a coupled carbon cycle-climate model, Global Biogeochem. Cycles, 23, GB1003. Garcia, et al., 2010, Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, Government Printing Office, Washington, DC, 344pp.

Comment [JHTW14]: Good suggestion, thank you. This has been changed.

Comment [JHTW15]: We have shown and explained (through basin-scale decomposition) that this overestimation is mostly due to the Pacific, where both models tend to underestimate mid-to-high latitude NPP.

Comment [JHTW16]: Deleted.

Comment [JHTW17]: The authors are very grateful the reviewer for point out these very helpful references, all of which have been incorporated into the latest draft of the manuscript. We have also added a reference to the review paper of Peña et al. (2010).

Second referee's report

SST observations from 1870-1880 are used but the time period that represent the oxygen observations is not mentioned. Please make sure that consistent time periods are used since a relation between SST bias and oxygen bias is invoked. I think it would be better to use more recent data where the coverage is better.

Comment [JHTW18]: This is analogous to the point made above by the first referee and has already been addressed. The authors agree with this point.

On page 1462 lines 5-7 the authors find "agreement" between FAMOUS and HadGEM2-ES circulations "encouraging" but I don't see much agreement. Also I think that the FAMOUS circulation is clearly inconsistent with observations. E.g. it does not display an Antarctic Bottom Water Cell, which is a fundamental property of the modern ocean circulation. I think this should be stated clearly and the comparison to observations should be extended to include AABW and flow of circumpolar deep water into the Indian and Pacific oceans.

Comment [JHTW19]: With hindsight, the authors agree with this comment and the wording of the manuscript has been amended accordingly. The further elucidate this, we have also added further discussion with reference to the observationally-based study of Lumpkin and Speer (2007). We have also added a figure to the draft manuscript showing the Pacific ocean circulation as suggested by the referee however we have not included the same for the Indian ocean since we feel that this would make the paper too long and would dilute the points being made here. The Indian ocean basin is considerably smaller that either the Pacific or Atlantic and we feel that we are able to answer the referees' point without invoking it.

I also recommend to show oxygen separately in the Atlantic and Indian/Pacific oceans since deep waters have large differences.

Comment [JHTW20]: We have added a new figure to the draft manuscript which shows the oxygen concentration in the Pacific Ocean but, as stated above, we do not feel that it is necessary to do the same for the Indian Ocean.

I'm not convinced by the author's attribution of low oxygen in the Southern Hemisphere to equatorial productivity bias. Why would this not affect the Northen Hemisphere equally?

Comment [JHTW21]: As stated above in our reply to the first referee, the authors are very grateful for the suggested inclusion of AOU distribution, which we feel enable us to answer this specific point (i.e. that the AOU shows that the waters of the north Atlantic are over-saturated with respect oxygen).

Another useful comparison would be horizontally averaged (in different basins) vertical profiles from the model(s) and observations. This could be done by using only model grid points where observations exist and would better show differences.

Comment [JHTW22]: Figures of AOU, oxygen, and circulation patterns are now present for the global ocean, the Pacific and the Atlantic. We feel that the addition of this level of detail makes our points more succinctly and completely so we are grateful for this suggestion.

Comment [JHTW23]: It was decided to use the spatially-complete World Ocean Atlas for the basin-scale oxygen decomposition (see above in the response to the first referee, point 4).

Apparent oxygen utilization (AOU) is another useful diagnostic that removes biases due to SST and solubility.

Comment [JHTW24]: See above comments to both referees.

I think the paper lacks in citing previous oxygen modeling work.

Comment [JHTW25]: See comment above with respect to the analogous comment from the first referee.