Dear Reviewer,

We highly appreciate your time and effort spent on reviewing our manuscript. We have prepared a new version of the manuscript with your comments taken into account. Below we include a point-by-point reply to each comment.

Comment:

“Part2.1 Ocean model in lines 28-30” Can the authors explained in more details the rescaling of the vertical coordinate?

Response:

Done.

Comment:

“Part 2.2 in lines 27, Note that the prescribed atmospheric forcing fields obtained from the PI ocean state estimate by Kurahashi-Nakamura et al., (submitted) and the corresponding isotopic fluxes are not entirely consistent and might introduce an error in our model simulation”. The authors refer to unpublished results here. They should show some results that indicate what could be the error and if the use of ratio of the isotopic content indeed minimize the uncertainty.

Response:

The paper by Kurahashi-Nakamura et al., (2017) has now been published, so we added the final citation. Using the ratio of the isotopic content of precipitation and water vapor inevitably leads to isotopic fluxes that are consistent with the optimized precipitation field. To show some results, we would have to run additional simulations.

Comment:

“Part 2.3.1 in line 30”. The authors compare long-term mean monthly value with GISS sample. This is indeed better than to compare with the annual mean isotope values. However, in the rest of the text it is difficult to know when the comparison is based on monthly value or annual. This is also not clear in the different figures and captions on the manuscript.

Response:

We changed both figure captions and parts of the text, to make it clearer whether we used annual mean or long-term monthly mean isotope values for the comparison.

Comment:

“Part 3.2 line21” The number of measurements for dD is rather small. According to the GISS database there is more than 1000 data points in dD. This is indeed more reduce than for the d18Osw but enough to realize a data-model comparison. Rather, the authors can mentioned that they choose to focus on the d18O and will work on the dD in the future.

Response:

We agree and rephrased the respective sentence. Due to the higher number of measurements we chose to focus on $\delta^{18}O$ to validate our model. However, since δD is now also used as a proxy in marine archives (i.e. Häggi et al. 2016) an implementation at this stage seemed reasonable.

Comment:

“Figure 3” Is it the annual or monthly value that are plotted for the model? Is it the surface data that are compared to the average 50m of the model or the data between 0 and 50 m? What could be the error associated if this is the surface data versus the average 50m?

Response:

The figure (now Fig. 4) shows the global annual mean surface (upper 50 m) $\delta^{18}O_w$ distribution of the model, while the GISS data are averaged over the upper 50 m. We rephrased the figure caption accordingly. The figure is just a first visualization of the model-data fit. For the statistical comparison (e.g. Fig. 6) we interpolated the GISS samples to the nearest tracer grid point of our model grid using inverse distance weighting. Hence, there is no error associated with the data either if it is surface data or data averaged over the upper 50 m.

Comment:

“Part 3.2 lines 3-4, the subtropical gyres are less enriched....” There is also a discrepancy for the Mediterranean Sea. What is the reason for such discrepancy in the subtropical gyres and Mediterranean region?

Response:

Based on the investigation of simulated E, P and $\delta^{18}O_w$ in P, we can conclude that the discrepancies in the subtropical gyres and the Mediterranean Sea are caused by enhanced P (having a dilutional effect on the surface water) and reduced E, whereby not enough ^{16}O is removed from the ocean surface. Even though the comparison with observed $\delta^{18}O_w$ in P is based on rather sparse data, the distribution of $\delta^{18}O_w$ in P seems to be reasonably well simulated. Further, one can assume that $\delta^{18}O_w$ in E is also slightly too enriched, but unfortunately, we cannot confirm this assumption because no observed data exists. We added these assumptions to section 4.2 and 4.3.

Comment:

“Figures 7 and 8” What is the depth used in the model (50 m?), is it annual or monthly?

Response:

The original figures showed the global annual mean $\delta^{18}O_c$ values of the surface (upper 50 m) simulated by the MITgcm. However, in response to the reviews we revised the discussion part (now section 4.4) and thus the figures (now Fig. 8 and 10) changed too. Now, we only compare modeled $\delta^{18}O_c$ values with foraminiferal calcite of plankton tow data (see last response). Therefore, we interpolated the plankton tow data to the nearest tracer grid point of our model grid using inverse distance weighting and thus compared them to the modeled $\delta^{18}O_c$ values of the respective month and depth level of sampling.

Comment:

“Part 4.1 lines 20 to 30 and Figure 9” A zoom on the artic region would be very helpful here. The isotopic values for rivers discussed in the text could eventually be added to this figure of the artic region.

Response:

We added a zoom on the Arctic Ocean to the figure (now Fig. 13), including the approximate location of discharge of the six rivers discussed in the text. Furthermore, we included a table (Table 2) to improve the comparison between simulated river values and observed river values by Cooper et al. (2008).

Comment:

"Part 4.3: Planktonic foraminiferal d18Oc" When reading part 4.3 it seems that the main discrepancy between data and model results is because of the gametogenic calcification of foraminifera and so that paleotemperature equations derived from plankton-tow data are more appropriate to reconstruct surface water conditions that the commonly used paleotemperature equations like Shackleton (1974) or Kim and O'Neil (1997). This discussion is extremely interesting for paleoceanographic studies. Nonetheless I find that all the potential factors that can affect the d18Oc and so the data-model comparison and mismatch are not developed enough. Indeed, the temperature bias in the model (2° C or more in some regions, see figure 1) can affect significantly the d18Oc reconstruction with the model. Similarly, the bias in d18Osw could contribute significantly to this "biased towards lower values". For example, the d18Osw is 0.4‰ too depleted in the model in comparison to data in the tropics (see part 3.2) and 0.9‰ too enriched in the Arctic Ocean (see part 3.2). These biases can affect the d18Ocalcite reconstruction and comparison. Also, it seems that the shift on figure 8a is more important for tropical species than for polar species. The data-model agreement or disagreement seems different depending the oceanic region (or species considered). So I recommend to the authors to realize a data-model comparison for the d18Oc for the different species of foraminifera separately. This analyze is important not only to try to discuss the oceanic region separately but also because other factors can affect each species of foraminifera in a different way. The seasonality is one of this important factor. Although there is one sentence in the part 4.3 that mention that "seasonality could be a problem and is not considered" it would be interesting to estimate how much bias could be introduce by such inconsideration. One way to do that could be to calculate the simulated seasonal amplitude for ocean calcite d18O in the model. It could be that the "biased towards lower values" is partly or totally explained by a distortion of the foraminifera flux towards a specific season or period than the annual mean. Similarly, the effect of the vertical migration is not completely developed. The author discuss the gametogenic calcification that is indeed related to this effect of vertical migration but the different species that are grouped on Figure 8 have different depth habitats and this affect their d18Oc. They can also change their depth habitat (for example during upwelling conditions). Again a data-model comparison for each species separately and with a different mean depth habitat of calcification would be interesting. The data on figure 8 are only presented for the first 50 m (although not clearly indicated in the text or on the Figure 8 caption). Although it will be difficult to examine the results for the very surface only (because of the grid of the model), the authors can investigate how the integration of the results for deeper water depth affect the data-model comparison.

The authors also suggest that the more enriched d18Oc values obtained with the equation of Shackleton (1974) is because this equation is based on Uvigerina spp shells that are relatively enriched in 18O. In fact, Shackleton (1974) proposed that Uvigerina peregrina is in isotopic equilibrium with seawater contrary to Cibicides. On the contrary, Bemis et al. (1998) (not cited in the discussion) suggested that Cibicides might also calcify in isotopic equilibrium and that the heavier δ18O values of Uvigerina are due to calcification at lower porewater pH. More recently, Marchitto et al., 2014 (also not cited in the discussion) investigated this difference in more details. Their results agree with Bemis et al. (1998) that Cibicidoides and Planulina appear to be closer to isotopic equilibrium (as represented by the Kim and O'Neil (1997) inorganic precipitates, which is also a matter of debate) than Uvigerina, although scatter in the measurements limits their confidence in this statement. They also recommend that Uvigerina δ18O be adjusted to the Cibicidoides scale by subtracting 0.47‰ and not 0.64‰. They were also unable to discern an impact of bottom water pH on benthic foraminiferal δ18O, but they speculate that Uvigerina's deviation from equilibrium could be explained by admixture of rapidly-precipitated non-equilibrium CaCO3 that would be subject to a pH influence. So, to my knowledge, the question as to why the δ 18O of Uvigerina and Cibicides are different remains. The question of the pH influence is also not discussed for planktonic foraminifera whereas it could also have a significant effect on the oxygen isotopic composition (Bijma et al., 1999; Zeebe 1999). This pH effect could be a primary mechanism to explain the differences between the equations (Mulitza et al., 2004). Again, the pH effect will be different with the latitudes and so it is important to discuss the species (that are associated to different oceanic regions) separately.

To resume, I like the discussion in part 4.3, this is of strong interest for paleoceanographic studies and the gametogenic calcification is a factor that certainly need to be considered. Nonetheless, the authors do not discuss in details all the factors and biases that can affect the d18Oc of their data-model comparison. For each foraminifera specie, how the bias in d18Osw in the model, the depth use in the model to generate the d18Oc signal, the seasonality and vertical migration and the pH can affect the d18Oc signal modelled and the comparison with data? At the end, if we consider all these factors and potential biases for d18Oc and the data-model comparison, can the authors really conclude that the differences between data and model is mainly linked to gametogenic calcification? If the authors cannot confirm their hypothesis in a revised version, they should also reformulate this conclusion from the abstract and conclusion part.

Response:

We agree with reviewer 1 that our discussion of foraminiferal $\delta^{18}O_c$ was too far-reaching. Indeed, many other processes (i.e. seasonal and vertical calcification, dissolution) exist that influence the composition of foraminiferal shells recorded at the sea floor, besides $\delta^{18}O_w$ and temperature. Since our model does not have an ecosystem module, many of these processes are not simulated, and we feel that our model is not the right tool to either gain information on foraminiferal ecology or on model performance. For this reason, we refrain from comparing our model results to core top $\delta^{18}O_c$ in the revised version of the paper. Plankton tow data are better constrained with respect to the time and depth of calcification. In order to demonstrate how the combined simulation of seawater temperature and $\delta^{18}O_w$ reflects the isotopic composition of foraminiferal carbonate, we hence kept the comparison to plankton tow data.