Response Referee 1

We thank the referee 1 for his careful review and comments. They will be helpful for the revision of our paper. Below, we provide point-by-point responses to the referee comments (red text).

This manuscript by T. Caley and co-workers examines the results of a Last Glacial Maximum (LGM) climate simulation using the iLOVECLIM Earth system model of intermediate complexity (EMIC), which has been enhanced by a module for stable water isotope diagnostics. The presented analyses comprise a comparison of the simulated LGM climate anomalies (with respect to a pre-industrial reference simulation) with isotope-related paleo-records from a variety of different paleoclimate archives: polar ice cores, speleothems, as well as planktic and benthic foraminifera data. On a global scale, model results of the isotopic composition of precipitation agree with the available isotope records. Most depleted LGM values are found at the high-latitudinal polar regions, strong but spatially varying depletion is seen over the continents and very low depleted (or even slightly enriched) LGM anomalies are simulated for tropical precipitation. Data-model comparison for sea surface temperatures indicates that the iLOVECLIM results are in agreement with the latest proxy-based SST reconstructions. For the North Atlantic, the model results suggest a strong mean LGM annual cooling in agreement with foraminifera data. The data-model comparison also reveals large positive calcite O-18 anomalies in the Southern Ocean as well as a large positive (negative) O-18 anomaly in seawater for the north Indian Ocean (China Sea) between the LGM and present-day climate. According to the model results, LGM changes in d18O of deep ocean waters might have not been spatially homogenous. Although several studies on the LGM isotope distribution in precipitation and ocean sea water using either atmospheric (AGCM) or ocean general circulation models (OGCM) have been published in the past, so far only very few fully coupled AOGCMs or EMIC simulations with isotope diagnostics exist. Regarding the LGM climate, to my knowledge this is the second study with an isotope-enabled atmosphere-ocean EMIC model and thus the findings are certainly of high interest and worth to be published. However, the current presentation of the model results as well as the performed data-model comparison falls too short in several aspects. Thus, I unfortunately cannot support a publication of the manuscript in its present form but rather must ask the authors for some major revisions of their study regarding the following aspects:

(1) Model description: A description of the iLOVECLIM model is completely missing in this article. Although this paper is already the fourth manuscript in a series of recent papers describing and evaluating the O-18 isotope diagnostics scheme in the iLOVECLIM model, the authors may not assume that everybody is familiar with their model. Thus they should add a description of the iLOVECLIM model in the "Material and methods" chapter (p. 109ff). They should give a brief overview of main model components, representation of key physical processes, implementation of O-18 isotope diagnostics as well as spatial and temporal model resolution.

As mentioned by the referee 1, we have described and validated the iLOVECLIM model in details in a series of 3 recent papers (Roche, 2013; Roche and Caley, 2013; Caley and Roche, 2013). That's why we have not repeated the description in the present paper. However, we agree with the referee that not everybody is familiar with our model and so we give a brief overview of the model in the revised version in lines 169-185: “The iLOVECLIM (version 1.0) model is a derivative of the LOVECLIM-1.2 climate model extensively described in Goosse et al. (2010). From the original model, we retain the atmospheric (ECBilt), oceanic (CLIO), vegetation (VECODE) and land surface (LBM) components and developed a complete, conservative, water isotope cycle through all cited components. A detailed description of the method used to compute the oxygen isotopes in iLOVECLIM can be found in Roche (2013) and the validation of model results can be found in (Roche and Caley, 2013; Caley and Roche, 2013). With regards to water isotopes, the main development lies in the atmospheric component (approximately 5.6° resolution in latitude and longitude) in which evaporation, condensation and existence of different phases (liquid and solid) all affect the isotopic conditions of the water isotopes. In the ocean (approximately 3° resolution in latitude and longitude), the water isotopes are acting as passive tracers ignoring the small fractionation implied by the presence of sea-ice (Craig and Gordon, 1965). For the land surface model, the implementation in the bucket follows the same procedure as for the water except that equilibrium fractionation is assumed during phase changes. The isotopic fields simulated are shown to reproduce most expected δ^{18}O–climate relationships with the notable exception of the isotopic composition in Antarctica.”

(2) State of equilibrium: For many coupled AOGCM setups, it is a very challenging task to reach a defined state of equilibrium in paleoclimate simulations. For their LGM study the authors simply claim that “a new equilibrium was reached after 5000yr of integration” (p. 110, l. 4-5). The authors should address in more detail (i) how they define the LGM climate equilibrium in their study, and (ii) which trends and drifts, e.g. of temperature or O-18 in deep ocean waters, may still exist in their simulation. They should also explain in more detail how they have changed the modelled river routing in the LGM simulation as compared to the present-day experiment, as the redistribution of fresh water into the oceans might have a large impact on the simulated LGM equilibrium state.

To run the model to statistical equilibrium we used the following procedure: we ran the model long enough so that deep Pacific waters (bottom layer of our ocean model) do not show any visible trend on a millennial – timescale basis, that is that the modelled drift (if any) is, at millennial timescale, well within the simulated decadal variability of the model. For the current simulation, deep temperatures in the Pacific are changing by less than 0.001°C per millennium in the mean, while the decadal variability of the model in the same area and over the same part of the simulation is 0.01 °C (at 5000 meters depth). Trends in d18O are of the same order of magnitude. We have expanded slightly the text in the manuscript to account for that explanation, as follow in lines 205-210: *“We ran the model long enough so that deep Pacific waters (bottom layer of our ocean model) do not show any visible trend on a millennial – timescale basis. Deep temperatures in the Pacific are changing by less than 0.001°C per millennium in the mean, while the decadal variability of the model in the same area and over the same part of the simulation is 0.01 °C (at 5000 meters depth). Trends in δ18O are of the same order of magnitude.”*.

In iLOVECLIM, the river routing is computed as following the topography using the largest slope until the sea is reached. No lakes are allowed, meaning that in the case of a point lower than any neighboring point, we search the next neighbors until a lower one is found, and drained to the sea. With this method, any change in topography like imposing ice-sheets of the Last Glacial Maximum allow to compute a new river routing directly. We have added one sentence to the manuscript in lines 201-203 to precise that aspect as follow: *"River routing is computed following the topography using the largest slope until the sea is reached. No lakes are allowed and a depression is drained until it next neighbors to the sea."*

(3) Data compilation - ice cores: For EPICA Dome C and Vostok, the authors have chosen the global meteoric water line given by Rozanski et al. (1993) to convert dD measurements into d18O values. However, for Antarctica a more appropriate local meteoric water line has been published by Masson et al. (2008). Present-day values of d18O in snow at the EPICA Dome C and Vostok drill sites also exist in this data compilation by Masson and co-workers. The authors should correct their values in Table 1, accordingly.

We agree with the referee that the compilation of Masson et al., 2008 is more appropriate for conversion. Anyway, the results for Antarctica are not good as mentioned in the previous version of the manuscript already "iLOVECLIM results are not in good agreement ($R2 = 0.19$) (Fig. 3, right panel) with the small data set used." "Part of the mismatch between the model and the data at those sites may arise from the numerical humidity issue already outlined in Roche (2013)." As we have a numerical issue in the advection–diffusion scheme at very low humidity content in Antarctica (Roche, 2013) we decided to remove the data-model comparison results for Antarctica in the revised version (see also comment 9 of the referee) and we mention in the revised manuscript in lines 299-300: *"Concerning Antarctica, a numerical issue in the advection–diffusion scheme at very low humidity content (Roche, 2013) hampers a good data-model comparison."*

(4) Data compilation - speleothems: The authors need to explain in more detail their criteria for the selection of the speleothem d18O anomalies listed in Table 1. E.g., for China, why have the authors chosen to include data from the Kesang cave but not the well-known Hulu cave d18O record (Wang et al., 2001)?

Our criteria for the selection of speleothem d18O anomalies was to conserve only the speleothem records that cover both the LGM and Late Holocene time interval "Calcite d18O anomalies are reported as the difference between averaged d18O values computed over the period of 20 000–22 000 yr BP and over the last 1000 yr of each record, using published U/Th chronologies". For Hulu cave, the record cover the LGM but not the Late Holocene (Wang et al., 2001). Therefore it was not included in our compilation.

We have clarified this choice in the revised version of the manuscript that now states in lines 230-233: *"We compiled only the speleothem records that cover both the LGM and LH time interval. Calcite δ18O anomalies are reported as the difference between averaged δ18O values computed over the period 20,000-22,000 years BP and over the last 1000 years of each record, using published U/Th chronologies."*.

(5) Data compilation - marine calcite data: The data-model comparison of the simulated d18O glacial changes in ocean waters and marine sediments is based on a new compilation of calcite d18O measurements from both planktic and benthic foraminifera. I am not an expert on marine isotope data

and thus cannot evaluate the quality of the compiled new data set. As one of the co-authors of this manuscript is a well-known expert in this field and has participated in some previous key data compilation efforts, e.g. the MARGO project (2009), I assume that the compilation has been done in a scientifically sound manner. Nevertheless, even as a non-expert I rate it as somewhat problematic that the compilation contains a number of unpublished data points, which have not undergone any careful peer-review, yet. I suggest that these data points might be either omitted or reviewed by somebody with profound expertise on d18O foraminifera measurements before publication.

The topical editor (Dr. André Paul) is an expert on the topic and has already given comments and recommendations concerning the marine calcite data compilation. For the final version of the manuscript, we will provide a more complete documentation and an electronic version of the data set freely accessible from a database such as Pangaea and/or NOAA.

(6) General model evaluation #1: The authors have decided to present all their results in this study as anomalies of the LGM minus the present climate, only. They state as their motivation: “The use of anomaly renders absolute values irrelevant; it concerns a purely relative change. We can therefore ignore complications such as species-specific climate variable relationships, vital effect offsets, and calibrations of values measured relative to the VPDB standard to values on the VSMOW scale” (p. 109, l. 8-11). This statement is erroneous in any case, where the conversion includes not only a constant offset as an addend term, but also a multiplicand term. For example, the general conversion between PDB and SMOW is: d18O(SMOW) = 1.03091 * d18O(PDB) + 30.91. Thus, any LGM-present delta anomaly has to be converted between the two standards as: d18O_anom(SMOW) = 1.03091 * d18O_anom (PDB).

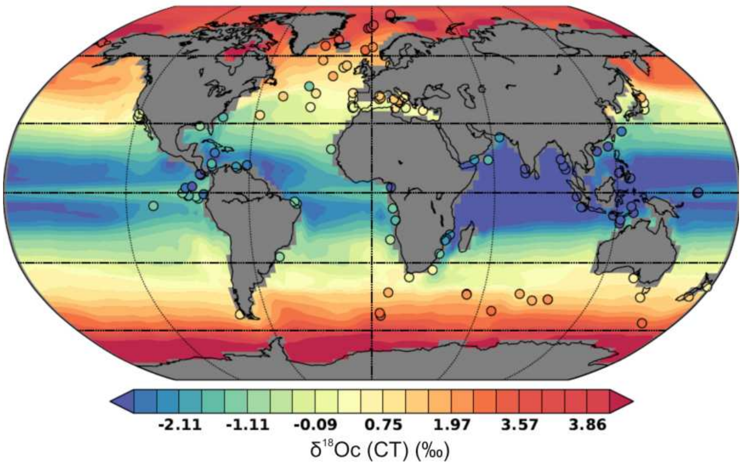
The multiplicand term mentioned by the referee is very close to 1 and so the error associated to this multiplicand term is negligible in comparison to error measurements (see table 1).
Anyway, all the data compiled in this study are expressed in PDB scale for carbonate samples and SMOW scale for ice core data. Therefore, all the anomaly calculated (LGM-present) do not suffer from calibration of value problems. In addition, the anomaly still limit complications such as species-specific climate variable relationships, vital effect offsets and non-calibration problems between laboratories. To lift any suspicion on the quality of our results, we now present in the revised version present day data-model comparison results (see after and new Figure 4 and 7).

(7) General model evaluation #2: Furthermore, I do not agree with the authors that their consequent use of anomalies in their performed model-data comparison “renders the absolute values as irrelevant”. For a profound evaluation of the model results, the overall performance of the model regarding different climate variables and regions of interest in terms of absolute values is of highly importance. Just to give a (slightly exaggerated) example: If a model overestimates present-day mean surface temperatures in a specific region by +20C as compared to observations, but simulates a LGM-present temperature cooling of -2C in agreement with some proxy data, such model results can certainly not be rated as a good model-data agreement! Only if one can demonstrate that a present-day bias between observational data and model results is rather small and most likely a climate-independent constant term related to some intrinsic model deficits, one might focus further model analyses on LGM anomaly values, only. Thus, I rate it as an absolute necessity that the authors present and discuss their key findings in much greater detail with respect to the related model performance of iLOVECLIM for the

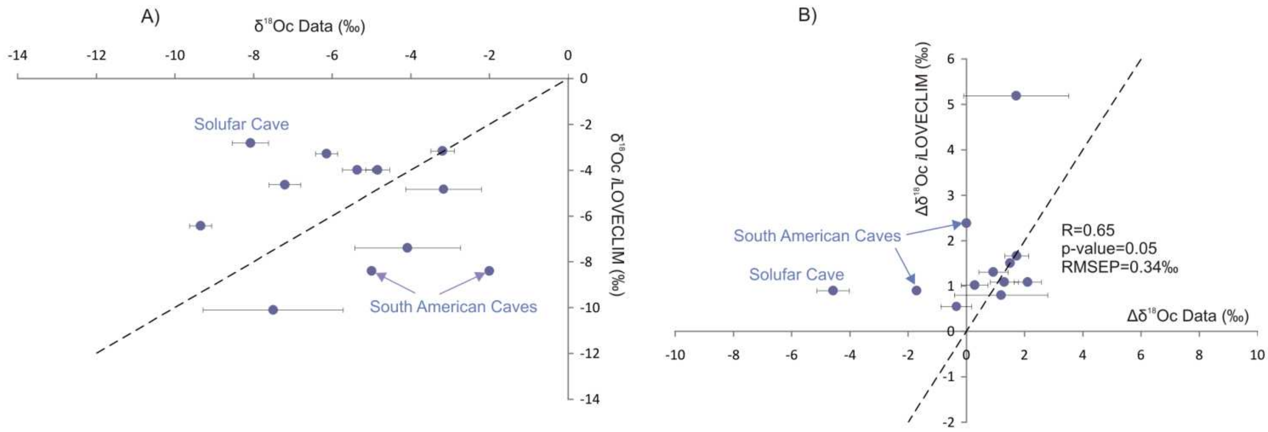
present-day climate. I am aware that some of these present-day iLOVECLIM results have already been published in previous articles of the authors (Roche, 2013; Roche and Caley, 2013; Caley and Roche, 2013) but I don't think that any knowledge of these older articles should be an essential prerequisite for a thorough understanding of the results presented in this new manuscript.

As mentioned by the referee, the present-day iLOVECLIM results have been extensively discussed in (Roche, 2013; Roche and Caley, 2013; Caley and Roche, 2013). To facilitate the understanding of the results presented in this new manuscript we will add a new figure 7 (see below) of data-model comparison for calcite d18O in surface ocean for the present day. We use only the data that contain also a LGM interval. Data and model results show an excellent agreement as already observed with a larger database (Caley and Roche, 2013) and therefore indicate no significant bias for the present day. We added in the revised version in lines 418-421: *"We computed annual mean calcite δ18O from simulated δ18Osw and SST and compared the results with deep-sea core data. Data and model results for the present day show an excellent agreement (Fig. 7) as already observed with a larger database (Caley and Roche, 2013)."*

We also add and discuss the calcite d18O of speleothem for present day in the new table 1 and new figure 4 (see below). The results indicates that when error bars are considered, data and model are relatively close to the 1:1 regression line, except for Solufar cave and South American data. There is probably some bias due to limitation of the model together with the processes operating in the atmosphere, soil zone, epikarst and cave system that hamper a good quantitative data–model comparison for the calcite-d18O signal (Caley and Roche, 2013). However, we consider here that these biases are climate-independent constant term and we therefore calculate the anomaly with the LGM to test this hypothesis (see response to comments 10). We added in the revised version in lines 343-351: *"Simulated atmospheric temperature and precipitation δ18O are used to calculate calcite δ18O. Data and model results for the present day, when error bars are considered, are relatively close to the 1:1 regression line, except for Solufar cave and for South American speleothems (Fig. 4A). There is probably some bias due to limitation of the model together with the processes operating in the atmosphere, soil zone, epikarst and cave system that hamper a good quantitative data–model comparison for the calcite-δ18O signal as already observed with a larger database (Caley and Roche, 2013). To test if these biases are climate-independent constant term, the simulated annual mean anomaly between the LGM and present is then compared to speleothem data (Fig. 4B)."*



New Figure 7



Revised Figure 4

(8) General model evaluation #3: The authors consider their study as the first attempt of a coupled atmosphere-ocean EMIC simulation of the Last Glacial Maximum (p. 108, l.7-9). Apparently, the authors are not aware of the study by Brennan et al. (2012) who have already published a very similar study with the UVic model. A detailed comparison and discussion of their model results to this previous study needs to be included in a revised version of this manuscript.

In our understanding, there is a fundamental caveat in developing water isotopes in a Energy – Moisture Balance Model (EMBM) such as the one that is in the UViC model. Indeed, in an EMBM, the water is mainly diffused and not advected, which is exactly the opposite of the process of the Rayleigh fractionation, at the basis of the water isotope fractionation. The fact that in Brennan et al. (GMDD version) the water isotopes are obtain in relative coherence with the present-day state is mainly due to careful tuning of the model coefficients' for the water isotopes, which are different than for the water itself. Again, this is completely non-mechanistic since every process that is included in the water cycle has to be repeated in the water isotope cycle exactly identically to ensure a correct representation of the water isotopes cycle. In that view, the results obtained for the present-day are a wise regression of the model results to present-day data, not a predictive dynamical simulation of the water isotopes in the atmosphere. This was at the time noted by the reviewers (cf. http://www.geosci-model-dev-discuss.net/4/C1063/2011/gmdd-4-C1063-2011.pdf). That incoherence is particularly apparent in the results of Brennan et al. (GMDD) where the simulated LGM d18O in precipitation is largely at odd with the proxy data in the comparison shown (see Table in the manuscript). For all the reasons given above, we do not feel that including a comparison to that manuscript in ours is likely to improve our study and therefore decided to not to include a comparison to the GMDD manuscript.

Finally, we agree with the reviewers that our study is not the first coupled atmosphere – ocean simulation performed with an EMIC for the LGM: Roche et al., (EPSL, 2004) indeed performed it before with the CLIMBER-2 climate model and this was already mentioned in the previous version of the manuscript: *"One earlier coupled model study including oxygen isotopes used a spatially greatly simplified model (Roche et al., 2004b)."*

(9) Comparison to Antarctic ice core data: I am surprised to see a comparison of model results with Antarctic ice core data in this study. In Roche (2013) it has been clearly shown that iLOVECLIM fails to reproduce modern d18O values in precipitation over the Antarctic continent. The "absurd d18O values in precipitation" (quote from Roche, 2013, p. 1487) have been explained by a deficit of the advection module of iLOVECLIM, which is apparently not able to numerically conserve very low humidity content and related isotope values over Antarctica. This deficit is not mentioned at all in this new manuscript. Why not? Has the problematic advection scheme been replaced by a better one? If so, how? If not, why do the authors now rate their results for the LGM climate (with even lower LGM atmospheric moisture content and precipitation rates as compared to present-day) as more trustworthy than the results presented in Roche (2013)?

As mentioned previously, the results for Antarctica are not good and this is indicated in the manuscript "iLOVECLIM results are not in good agreement (R2 = 0.19) (Fig. 3, right panel) with the small data set used." "Part of the mismatch between the model and the data at those sites may arise from the numerical humidity issue already outlined in Roche (2013)." As we have a numerical issue in the advection–diffusion scheme at very low humidity content in Antarctica (Roche, 2013) we decided to remove the data-model comparison results for Antarctica in the revised version.

(10) Comparison to speleothem data: Again, I am a little bit astonished about the author's choice of including a comparison of LGM-present speleothem isotope anomalies in this study. In Caley and Roche (2013) it was stated for the iLOVECLIM model that "limitation of the model together with the processes operating in the atmosphere, soil zone, epikarst and cave system hamper a good quantitative data–model comparison for the calcite-d18O signal." Fig. 2 in Caley and Roche (2013) clearly illustrates this

model deficit with an insignificant correlation (R2 = 0.1) between measured late-Holocene calcite-d18O values and model results. Now the authors argue: "The better agreement between data and model results in term of annual mean anomaly suggests that this approach allows us to reduce complications with the atmospheric, soil and cave processes and that the model is capable to reproduce the right amplitude of changes." (p. 115, l. 20ff). In line with my general model evaluation comment #2 (see above) I think that the authors have to present a much more profound speleothem data-model comparison for a convincing line of arguments. E.g., they should add a comparison and discussion of simulated present-day d18O values versus measured late-Hololcene calcite-18O for the identical speleothem data set as given in Table 1. Furthermore they should quantify any model-data agreement by calculating not only a correlation coefficient, but also root mean square errors as well as uncertainty and statistical significance of the found data-model consistency.

As mentioned by the referee even if the absolute values are not correct for the present days, if these differences are climate-independent constant term as processes operating in soil zone, epikarst and cave system, we can expect to obtained the right amplitude of changes between the LGM and present day. We also mentioned in Caley and Roche (2013) in the conclusion part that "It also remains to be tested if qualitative data–model comparison for the calcite-18O signal (i.e. anomaly between the past and present day) would give better results". We furnish the values of d18Oc for present day in the new Table 1 and Figure 4. Although we observe some bias due to limitation of the model together with the processes operating in the atmosphere, soil zone, epikarst and cave system that hamper a good quantitative data–model comparison we tested if this bias may be related to climate-independent constant term. We therefore calculate the anomaly between the LGM and present day for calcite d18O. Apart a problem at Solufar cave and for two speleothems in south America, the other anomalies significantly correlate. We quantify model-data agreement by calculating root mean square errors as well as uncertainty and statistical significance (R=0.65; p-value= 0.05; RMSEP=0.34‰) (Revised Figure 4). The better agreement between data and model results in term of annual mean anomaly suggests that this approach allows us to reduce complications with the atmospheric, soil and cave processes and that the bias in absolute calcite d18O values was probably climate-independent constant term.

(11) Comparison to marine isotope data: In Roche et al. (2007) it has been shown that the deep ocean circulation of an LGM simulation with the LOVECLIM model is at odds with the general accepted proxy-based view of LGM circulation. The model produces a stronger LGM overturning than for the Late Holocene, opposite to the expected weaker glacial overturning in the Atlantic Ocean. I am puzzled that this important aspect of the simulated LOVECLIM LGM climate is not discussed at all in this manuscript. Why not? In contrast to the study presented by Roche et al. (2007), the authors have now the unique opportunity to directly check if the simulated Atlantic overturning and its simulated isotopic signature is in agreement with available marine isotope data. Such analysis could lead to some new key insights about possible states of glacial overturning and its imprint in marine oxygen isotope data. It should obviously be included in a revised manuscript version.

We acknowledge that the discussion of the glacial overturning is missing in the first version of our manuscript. d18O is not the isotopic proxy that is most used for paleocirculation reconstruction, and hence, we did not discussed the implied changes at length. Contrary to what the reviewer implies ("general accepted proxy-based view of LGM circulation" – what study do they base that judgment on?) the LGM Atlantic deep ocean circulation is still debated today.

What is apparent from d13C distribution (e.g. Curry and Oppo, 2007) is the fact that the glacial Atlantic ocean was more stratified than today. Whether this means a stronger Antarctic input in the deep (e.g. Stephens and Keeling, 2001, Paleoceanography) or a very weak deep circulation (like in Bouttes et al., 2011) is quite unclear. Similarly, though it seems indeed accepted that the upper part of the deep water mass has shoaled by ~ 1 km at the LGM, its strength remains unclear, the values in d13C giving no constrain on the speed. Inversion of the ocean state is not providing a stronger constraint (cf. Marchal & Curry, 2008, doi:10.1175/2008JPO3895.1). The only indication that could be built for the speed is the integrated Pa/Th proxy that shows a slightly reduced ocean speed (Gherardi et al., 2009), integrated over the whole column. This could be indicative of a stronger surface and very inactive deep for example, in addition to all the complex effects that biological fluxes add onto Pa/Th. The picture is not clear for this particular aspect.

Turning back to the iLOVECLIM model, we do simulate a slightly enhanced overturning in iLOVECLIM, as was discussed in Roche et al., 2007. At the time, Roche and co-workers discussed all the available data and concluded that whether the circulation was in agreement or not was unclear.

Using Fig. 12.A. Of the current manuscript (d18Osw) one could infer from the (very) sparse proxy data coverage that the negative d18Osw anomaly we simulate from the north Atlantic is not propagating deep enough. This, if applied to a circulation issue would lead to the conclusion that an even MORE active deep oceanic circulation is needed. Alternatively, it could be caused by not depleted enough deep waters at the formation site. However, two data points in the northern North Atlantic are indicating a correct range of values there.

Turning to Fig. 13. A, it is quite hard to decipher from the data spread whether the anomaly simulated is in accordance with the data. We obtain values that are quite close to the one obtained in data evidence (within 0.3 per mil) with some notable discrepancies in the northern and tropical Atlantic. There is however no real consistent pattern, and one should remember the large imprint of temperature on d18Ocalcite and the fact that the simulated very deep temperature anomalies are not in very good agreement with the reconstructions. Finally, where the data – model comparison could be evaluated (Fig13, C), it seems that we do agree to a large extent in the central Atlantic between 2 and 3.5 kilometers, though this region is the most affected by too strong northern sourced deep water in our model.

The comparison is thus not very conclusive as to whether or not a stronger deep Atlantic circulation (as in our model) is truly at odd with proxy data evidence or still compatible (as we show here) with them. Only future sensitivity experiments to different glacial circulations can tell whether the obtained pattern is robust to circulations changes.

We have added part of that discussion into the revised version of the manuscript in lines 546-556 that now reads: *"Considering the uncertainties on pore fluids reconstructions (0.1‰) (Adkins et al., 2002; Schrag et al., 2002), the data point that exhibits a negative δ18Osw anomaly in the deep Atlantic*

(~4500m) is the only measurement in significant disagreement with the model results (Figure 12A). The value of this data point is surprising as it would imply a vigorous and/or deeper extension of the north Atlantic deep water at the LGM, in disagreement with reconstructions based on δ13C and Cd/Ca proxies (Lynch-Stieglitz et al., 2007 and references therein) and with our model results. It could also be caused by not depleted enough deep waters at the formation site. However, two data points in the northern North Atlantic are indicating a correct range of values there. Future sensitivity experiments to different glacial circulation are necessary to determine if the obtained pattern is robust to circulations changes."

(12) Figs. 2,3,4,6,8,11,12,13: The chosen asymmetric, highly non-linear colour code of most contour map plots makes it unnecessary difficult to determine absolute values of difference between individual isotope measurements and related simulation results. I highly recommend that the authors change the colour scale of the contour maps to a more conventional linear (or logarithmic, if necessary) one.

The color scale we use is automatically generated using a statistical treatment of the available proxy or model data, to ensure that the values are clearly seen in all plots. The rationale is that the center of the color scale is the mean of the data and that 95% of the range of the data is contained in the color scale, using constant increments in the distribution. We have added that explanation in the legend of the first figure using it as: *"The color scale we used is based on the distribution of the values in the dataset used: 95% of the proxy data are meant to be appropriately represented in that color scale. It is centered around the mean of the dataset, hence red means heavier values than the mean and blue values lighter than the mean of the dataset.".* It is further referred to in the other figures.