Response to reviewer 2

We are grateful for the positive and constructive feedback provided by the reviewer 2 that helped us to significantly improve our manuscript. Below, we present a point-by-point response to the individual comments raised. A final version of our point-by-point response will be given at the end of the discussion phase with responses to all reviewers' comments.

Reviewer comments are shown in red color, with our responses in black.

General comment

Caley et al. investigate the use of planktonic foramifera $\delta^{18}O_c$ as a surface paleodensity proxy for the whole ocean. For that, the authors applied three Bayesian regression models on $\delta^{18}O_c$ datasets to reconstruct surface paleodensity for the late Holocene (LH) and the Last Glacial Maximum (LGM). Using isotope-enabled models of different complexities, Caley et al. investigated the additional uncertainties that are introduced by the potential evolution of the $\delta^{18}O_c$-density relationship with time (i.e., from LGM to LH). Except for the Nordic Seas, the authors demonstrated that additional uncertainties are weak globally (for a LGM to LH climate change).

The objectives and the method of this study correspond to the scope of CP. The study is easy to follow, and I took pleasure to read it. I have some minor comments related to the datasets, the evaluation, and the comparison with LGM results.

Major comments

- The description of the LGM d18O dataset lacks details, especially, on the additional data from more recent studies (lines 130-131), which are not available with the paper (or I missed them). For the revised paper, I suggest the reviewer to provide the compilation of all the data ($\delta^{18}O$ for LH and LGM + ocean datasets) they used for this study, with the appropriate references inside.

All the data and the code for the Bayesian calibrations will be available after publication. We refrained to share these data at the stage of the preprint because the data will be publicly available but still not validated and accepted for publication. We understand and agree that the reviewers should have access to these data before publication, so we will propose a link that will be shared with the editor if the reviewers want to have access to the code and additional dataset.

We will share the additional data from more recent studies but the previously published $\delta^{18}O$ data for LH and LGM are already available to download on the original repository in the "code and data availability" part: "LGM and LH $\delta^{18}Oc$ dataset are available at doi:10.5194/cp-10-1939-2014-supplement for Caley et al., 2014, at https://doi.org/10.1594/PANGAEA.894229 for Waelbroeck et al., 2014 and at https://doi.org/10.1594/PANGAEA.920596 for Tierney et al., 2020b. The additional LGM and LH $\delta^{18}Oc$ dataset will be available as a supplement."

We consider that it is important to keep the original datasets, references and citations of these previous studies as we did not change/reworked these datasets.

The full density dataset re-gridded onto a common 1° × 1° spatial grid will be published and available in the Barathieu et al. paper.

- For the evaluation of the residuals under LH climate (Figure 1), why is there a like a threshold in observed data at a value of 28. Is it a problem with the data? I think it should be discussed because it gives the largest density residuals. Moreover, the authors do not discuss the strongest residuals in the Mediterranean Sea, which are probably influenced by bias in net freshwater fluxes and thermoaline circulation. The authors could use this recent study from Ayache et al. (2024, https://doi.org/10.5194/gmd-17-6627-2024).

Thank you for bringing the interesting study of Ayache et al., 2024 to our attention. The "threshold" in observed data at a value of 28 for density is because we are close to the maximum of present day's annual ocean density. In some high latitudes regions (Nordic Seas and Austral Ocean), density is already high and temperature changes have a smaller effect. Cold water is already dense, so cooling it further doesn't increase density as much. Consequently, we observe a sensitivity decreases. The rate of change of density with respect to T flattens out, meaning the system becomes less responsive (see also our response to reviewer 3). However, the density residuals are not the largest in these specific regions for the poly1_hier Bayesian model as visible on (Fig. 1b) and on the Figure (a).

We added sentences in the text to clarify this point.

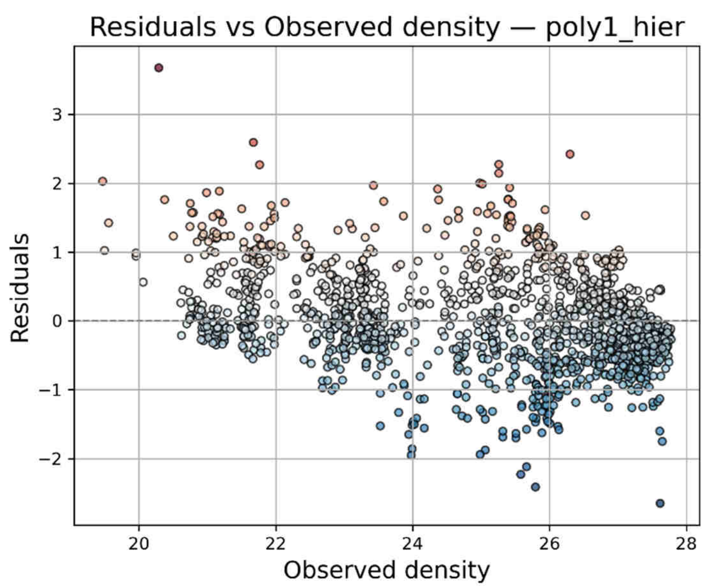

Figure a: Scatter plot between observed density and residuals for the poly1_hier Bayesian model. Density in kg/m^3.

Regarding "the strongest residuals in the Mediterranean Sea". According to the poly1_hier Bayesian model (Fig. 1b), there are a few points with strong residuals in its eastern part only.

We added a few sentences to explain that part of the Mediterranean Sea is characterized by high residuals, and propose future research improvements: “Strong negative residuals are also observed in the eastern part of the Mediterranean Sea. Malevich et al. 2019 already observed a reduced performance in the hierarchical seasonal model they developed to calibrate the relationship between δ^{18}Oc and SST. They attribute this to the unusual behavior of *G. ruber* data in the Mediterranean, which may reflect depth habitat migration. But estimation of seasonality for this region could also be problematic and play a role as highlighted in the study of Ayache et al., 2024: “The influence of seasonal temperature variability on δ^{18}Oc (Eq. 6) is important, particularly in the Mediterranean Sea because of marked seasonal thermal contrast.” And “Nonetheless, a dedicated study should be conducted to further elucidate the seasonal aspect.” Alternatively, bias in net freshwater fluxes and thermohaline circulation could affect late Holocene δ^{18}Oc values (Ayache et al., 2014). Future modelling developments, such as the use of high-resolution regional model (Ayache et al., 2024) in combination with the FAME module, could help to better understand the relation between δ^{18}Oc, density, temperature and δ^{18}Osw and past climate changes in the Mediterranean Sea.”

- Lines 328-329: I suggest to try with the yearly δ^{18}Osw values from ECHAM5/MPI-OM to really know the effect of seasonality on calculated d18Oc. With the current comparison, it be cannot excluded that the differences between the results using ECHAM5/MPI-OM and iLOVECLIM is due to lower resolution of iLOVECLIM or other missing/biased processed in this lower resolution model.

As suggested by the reviewer, we used the ECHAM5/MPI-OM yearly values of δ^{18}Osw to compute the δ^{18}Oc (see Figure b) and compared the results with our Figure 4 (a) to better assess the effect of seasonality. Results indicate a slight decrease of the R^2 of 0.02 and a slight increase in RMSE of 0.06 when seasonality is not taken into account. To assess whether the difference in predictive performance between the two models was statistically significant, a paired t-test was conducted using cross-validated R^2 and RMSE scores. The test yielded a t-statistic of −5.51 and a p-value of 0.0053 for R^2 (and of 5.27 with a p-value of 0.0062 for RMSE), indicating a significant difference between the models. Therefore, seasonality partly explains the weak difference between the results using ECHAM5/MPI-OM and iLOVECLIM. Lower resolution of iLOVECLIM or other missing/biased processes in this model could also contribute to this weak difference.

We added this in the text.

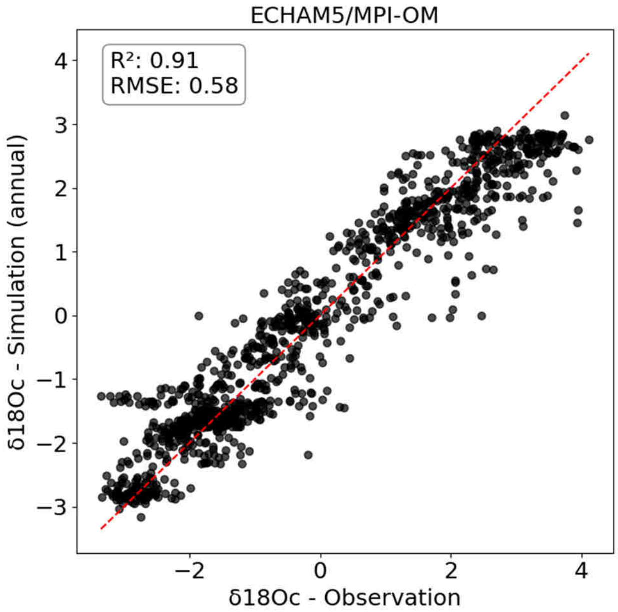

Figure b: comparison between PI simulated foraminifera δ^{18}Oc (‰) (FAME module forced with yearly ECHAM5/MPI-OM climate model hydrographic data) and observed core-top δ^{18}Oc (‰) data. The 1:1 line is indicated.

- For the evaluation for LGM results, the residuals in the Nordic Seas are stronger with ECHAM5/MPI-OM than with iLOVECLIM (Figure 5). This point should be discussed more in details by the authors.

The stronger difference in LGM – PI surface density anomaly residuals between ECHAM5/MPI-OM and iLOVECLIM in the Nordic Seas could be explained by different simulated sea ice coverage in ECHAM5/MPI-OM vs. iLOVECLIM (Figure c). Indeed, Nordic seas are the region with the largest difference of modeled annual SST below 0°C.

Temperature is used to calculate the δ^{18}Oc signal, ocean density and to force the FAME module. The PI δ^{18}Oc-density relationship is used to reconstruct LGM density based on LGM δ^{18}Oc. Then the reconstructed density is compared with modelled LGM density and so any temperature differences in the Nordic seas (as visible in Figure c) will affect density reconstructions and then the density residuals observed in Figure 5.

We added few sentences in the text to clarify this point.

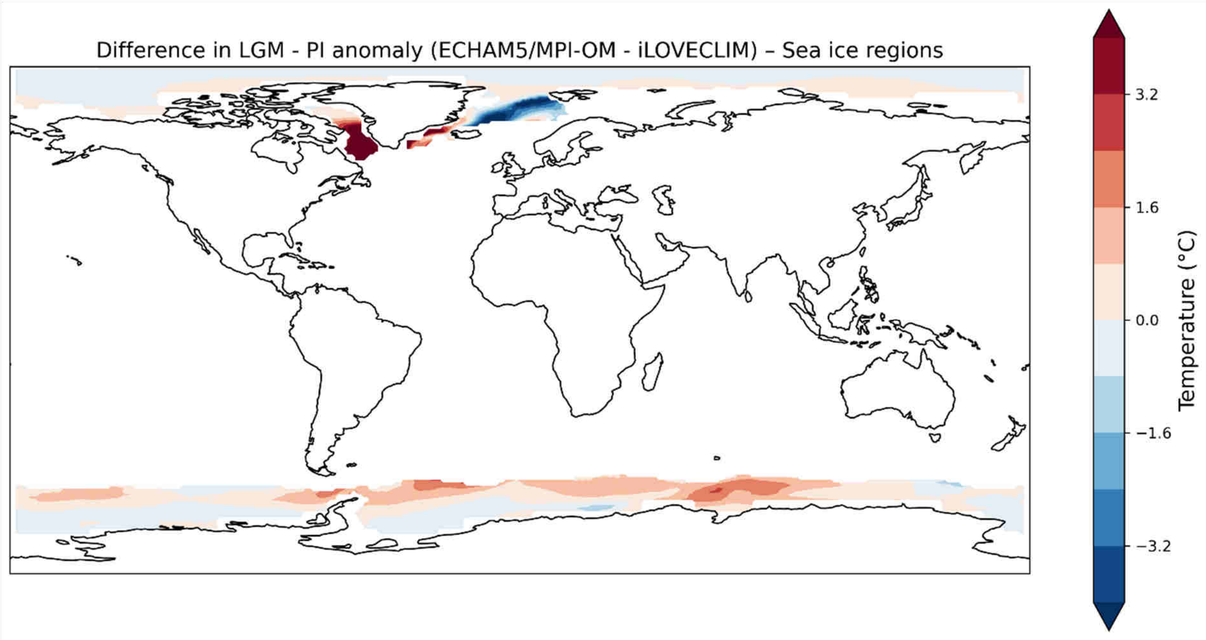

Figure c: Difference in LGM – PI anomaly between ECHAM5/MPI-OM and iLOVECLIM for SST. Only region with modeled annual SST below 0°C are shown to investigate differences linked to simulated sea ice coverage.

Moreover, I would like to see some evaluation of the reconstructed density anomalies between LGM and LH (Figure 7a). Are there other reconstructions? Or can the authors compare those results with modeled LGM-LH surface densities?

To our knowledge, there is no method that would provide a direct quantitative reconstruction of ocean density at the global scale. Some methods exist based on dinocyst assemblages, using transfer functions (see for example "Peyron, O., & Vernal, A. D. (2001). Application of artificial neural networks (ANN) to high‐latitude dinocyst assemblages for the reconstruction of past sea‐surface conditions in Arctic and sub‐Arctic seas. *Journal of Quaternary Science: Published for the Quaternary Research Association*, *16*(7), 699-709.") but these reconstructions are limited to the arctic regions, where we cannot evaluate our reconstructed densities.

This lack of method at the global scale to quantitatively reconstruct densities was one of our main motivations to develop our work.

Our results can be compared with modeled LGM-LH surface densities and this will be the focus of a follow-up study by Barathieu et al. that will be submitted very soon, as mentioned in our paper in line 439-440: "Further regional analyses of ocean surface density and comparison with numerical climate models are presented in Barathieu et al. in prep."

We think that it would be interesting to have this data-model comparison paper as a companion paper of this paper in CP if the editor agrees.

We present some results (Figure d) of data-model comparison from the paper of Barathieu et al., that will be submitted soon to show that the models used in our study (MPI and iLOVECLIM) show significant (p-value <0.05) and strong correlation (R^2>0.5) with our reconstructed density for the PI and LGM.

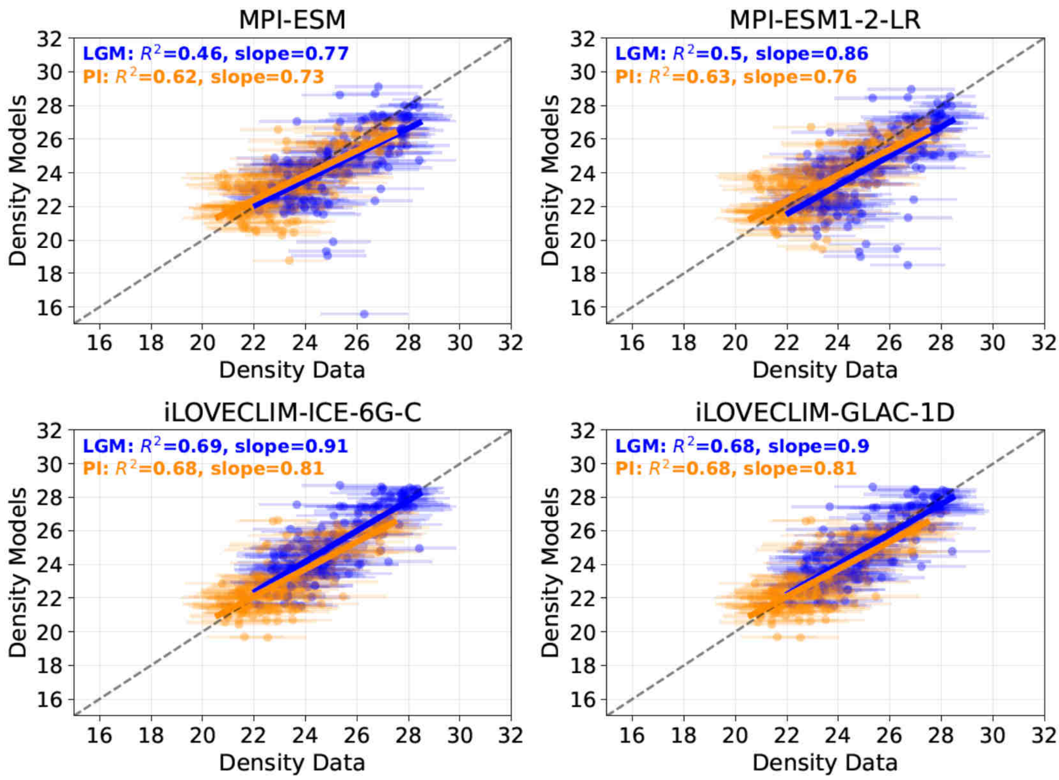

Figure d: Linear regressions between absolute surface density from proxy-based reconstructions (x-axis) and model simulations (y-axis), at the global scale. Results are shown for the LGM period (blue) and the PI period (orange). Error bars on the x-axis represent the 95% confidence intervals of the reconstructed values. Based on Barathieu et al. in preparation.

- I would like the authors to put into context their results regarding other climate periods. The authors state that additional uncertainties due to the evolution of the $\delta^{18}O_c$-density relationship with time are globally weak (lines 45-46). However, this is true, except for the Nordic Seas, for a LGM-to-LH change. It has not been proven for another period, such as the Last Interglacial (110-130 ka). Considering mid-Holocene period (6 ka) raised by the reviewer #1, the changes in $\delta^{18}O$ of seawater are rather small (+0.5‰ maximum, only, in the western Pacific Ocean according to Shi et al., 2023 and Cauquoin et al., 2019) compared to the LGM ones.

Our work is really focus on the LGM and so testing the stability of the $\delta^{18}O$/salinity relationships for the Last Interglacial (LIG) or MH (as asked by reviewer 1) and its potential effect on density predictions is rather out of the scope of this paper. Nonetheless, we agree that it is interesting to put into context our results regarding other climate periods. We already conducted some tests using isotope-enabled model runs representing the mid-Holocene (e.g., Shi et al. 2023) in order to demonstrate that additional uncertainties due to the evolution of the $\delta^{18}O_c$-density relationship with time are globally weak and that the new calibration has great potential to be applied to other past periods and to reconstruct the past temporal evolution of ocean surface density. Results indicate a strong stability of foraminifera δ^{18}Oc-density relations between MH and the PI and so very weak influence of δ^{18}O/Salinity relation instability on final density predictions.

We also test the LIG time period as asked by reviewer 2. We use isotope-enabled model runs representing the LIG at 125 kyr (e.g., corresponding to the maximum changes observed during the LIG period according to Figure 9 of Gierz et al., 2017) (Figure e).

Again results indicate a strong stability of foraminifera δ^{18}Oc-density relations between LIG and the PI and so very weak influence of δ^{18}O/Salinity relation instability on final density predictions (Figure e). **This confirms and reinforces our conclusion that the new calibration has great potential to be applied to other past periods and to reconstruct the past temporal evolution of ocean surface density.**

The additional tests we conducted during the review process for the MH and LIG time periods allow us to put our results regarding other climate periods into context. We will therefore include these results in the Appendix of the revised paper, together with an explanatory text to support and reinforce our conclusion about the new calibration and its application to other past periods.

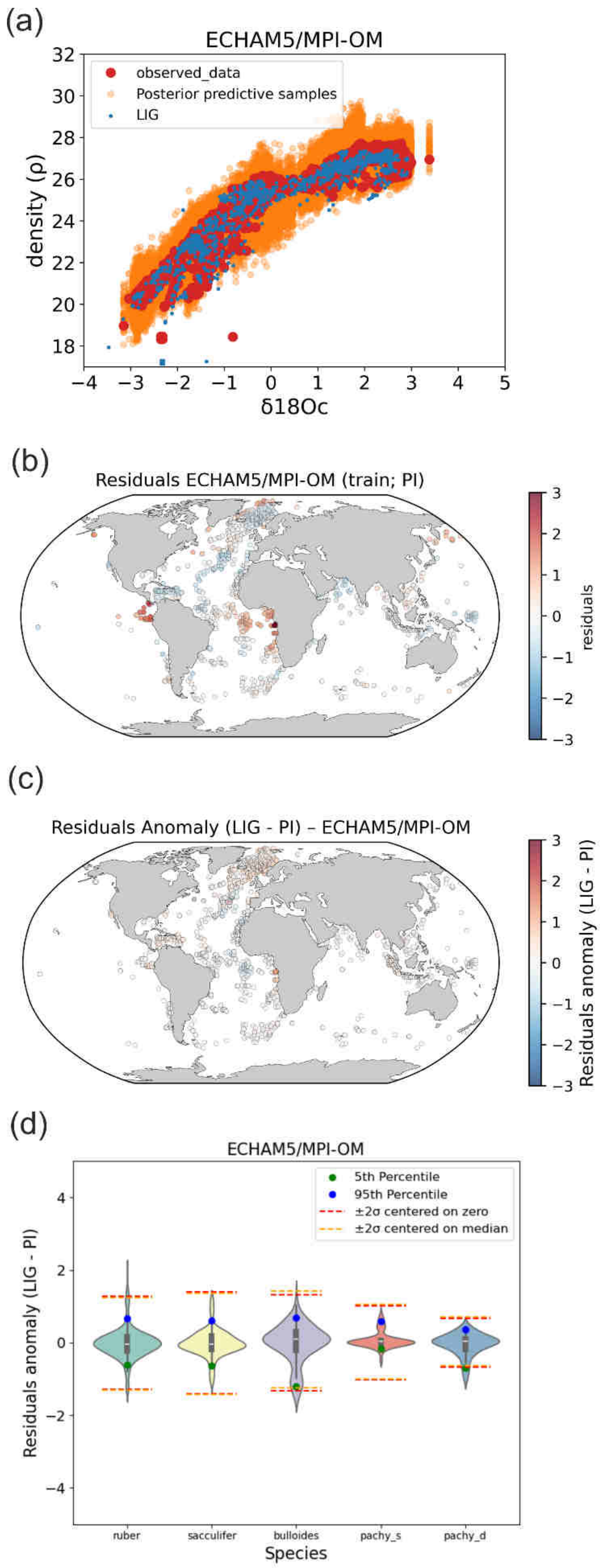

Figure e: Stability of foraminifera δ^{18}Oc-density relations between PI and the LIG calculated with FAME and forced by global ECHAM5/MPI-OM (Gierz et al., 2017) hydrographic data.

(a) PI Bayesian regression models between foraminifera δ^{18}Oc and annual surface density. Posterior predictive samples and LIG δ^{18}Oc-density relations (LIG) are visible. (b) Density residuals (predicted - observed) for the PI experiments. (c) Density residuals anomaly between LIG and PI. (d) Probability distributions of surface density residuals anomaly (LIG - PI) without Nordic Seas (north of 40°N).

- Generally, the units are missing in the labels of figures' axes. Please check all the figures. Also, the panel labels are used on one complete column or row in Figures 1, 3, 5, 6 or are absent for Figure 4. Please add a letter label for each panel in the figures.

Ok, we corrected this in the revised version.

Specific comments

- Lines 74-75: to quantify past ocean density and dynamics.

We corrected in the revised version.

- Lines 148: give the units for $\delta^{18}O_c$ (and relative to which standard) and ρ (sigma-theta relative to a density of 1029 kg/m3?).

We changed for: "δ^{18}Oc (‰ VPDB), and annual mean surface density, ρ (kg/m^3 relative to water density of 1000 kg/m^3)"

- Section 2.4.1: specify that iLOVECLIM is an intermediate-complexity model, whereas ECHAM5/MPI-OM is an Earth System Model.

To be precise, we mention in the revised version "The iLOVECLIM (version 1.1.3) earth system model of intermediate-complexity" and "use the ECHAM5/MPIOM coupled GCM".

- Figure 1: This is for LH period I suppose?

Yes, we revised for "Bayesian calibration models for late Holocene core-top samples against observed density"

- Lines 235: explain a bit more that is ELPD.

The ELPD measures the expected predictive accuracy of a Bayesian model. It is defined as the sum over all data points of the expected log posterior predictive density (see Equation (2) in Gelman et al. (2014)). In plain words, one could say that the ELPD is the average log probability that a Bayesian model assigns to new data, summed across observations. So, in our case, a higher ELPD means the model makes sharper and more accurate density predictions. More details can be found in Gelman et al. (2014).

- Figure 3: give the p-values.

p-values have been added in the revised Figure 3 and revised text.

- Figure 4: Only for LH period?

We specified this in the revised version: “comparison between simulated PI foraminifera $\delta^{18}Oc$ (FAME module forced with ECHAM5/MPI-OM and iLOVECLIM climate model hydrographic data) and observed LH core-top $\delta^{18}Oc$ data. The 1:1 line is indicated.”

- Row (a) of Figure 5: the legend for the LGM values is not clear.

We specified in the revised legend of the Figure 5: “the LGM $\delta^{18}Oc$-density relations (LGM) are visible”

- Line 421: 1 or 1.05‰?

Yes, we keep 1.0 ‰ and we added references of Schrag et al., 2002 and Labeyrie et al. (1987) in agreement with reviewer 1’s comment.

- Lines 455-456: By applying a Bayesian regression hierarchical model to LGM and LH $\delta^{18}O_c$ foraminifera databases, we reconstructed LGM and LH annual surface density and found stronger LGM density…

We corrected in the revised version

References

Ayache, M., Dutay, J.-C., Mouchet, A., Tachikawa, K., Risi, C., and Ramstein, G.: Modelling the water isotope distribution in the Mediterranean Sea using a high-resolution oceanic model (NEMO-MED12-watiso v1.0): evaluation of model results against in situ observations, *Geosci. Model Dev.*, **17**, 6627–6655, https://doi.org/10.5194/gmd-17-6627-2024, 2024.

Cauquoin, A., Werner, M., and Lohmann, G.: Water isotopes – climate relationships for the mid-Holocene and preindustrial period simulated with an isotope-enabled version of MPI-ESM, *Clim. Past*, **15**, 1913–1937, https://doi.org/10.5194/cp-15-1913-2019, 2019.

Shi, X., Cauquoin, A., Lohmann, G., Jonkers, L., Wang, Q., Yang, H., Sun, Y., and Werner, M.: Simulated stable water isotopes during the mid-Holocene and pre-industrial periods using AWI-ESM-2.1-wiso, *Geosci. Model Dev.*, **16**, 5153–5178, https://doi.org/10.5194/gmd-16-5153-2023, 2023.

References

Gelman, A., Hwang, J. and Vehtari, A.: Understanding predictive information criteria for Bayesian models. *Stat. Comput.*, Springer US, 24, 997-1016, 2014.

Gierz, P., Werner, M., & Lohmann, G. (2017). Simulating climate and stable water isotopes during the Last Interglacial using a coupled climate-isotope model. *Journal of Advances in Modeling Earth Systems*, *9*(5), 2027-2045.

Malevich, S. B., Vetter, L., and Tierney, J. E.: Global core top calibration of $\delta^{18}O$ in planktic foraminifera to sea surface temperature, Paleoceanogr. Paleoclimatol., 34, 1292–1315, 2019.

Peyron, O., & Vernal, A. D. (2001). Application of artificial neural networks (ANN) to high-latitude dinocyst assemblages for the reconstruction of past sea-surface conditions in Arctic and sub-Arctic seas. *Journal of Quaternary Science: Published for the Quaternary Research Association*, *16*(7), 699-709.