Response to the referee and community comments

Transient simulation of Holocene climatic and isotopic variations across global monsoon regions

Xiaoxu Shi, Xiya Liu, Jiping Liu, Martin Werner, Hu Yang, Francisco W. Cruz, Chaoyuan Yang, Gerrit Lohmann

Dear Editor, dear Reviewers,

We thank the Editor, two anonymous reviewers (RC1 and RC2) and Dr Giacomo Medici (community comment CC1) for their positive and constructive comments. We have revised the manuscript thoroughly. Below we respond to each comment point by point.

The comments are in black. Our responses follow each comment, and text quoted from the revised manuscript is shown in blue. A version with tracked changes will be provided as a separate file during a later revision stage.

The revision contains several new analyses that were motivated by your comments. A direct comparison with the iTRACE Holocene simulation and with proxy data is presented in a new supplementary figure (Fig. S2) and Section 4.3. The North African greening simulated by the dynamic vegetation is analyzed in a new supplementary figure and discussed in Sections 3.3 and 4.1. The variance that the regression cannot explain is quantified against an estimate of internal variability in a new Fig. 11 and Section 4.6, and the time dependence of the regression coefficients is discussed in Section 4.5, with confidence intervals in Fig. 8 that account for autocorrelation. The collinearity of the predictors and the robustness of the climate-index selection are assessed in new supplementary figures and tables (Figs. S3 and S6, Tables S2 and S3), the PI biases are mapped against a reanalysis (Fig. S4), the nonlinearity of the Holocene trends is examined with running trends (Fig. S5), the freshwater pathway of the 8.2 ka hosing is documented (Fig. S7), and a subsection on the limitations of the approach has been added (Section 4.7). The description of the regression method has been rewritten in full, the experimental design has been expanded in more details, the Discussion has been reorganized into subsections, and a separate Conclusion section has been added.

While addressing the comments on the regression method, we also corrected the calculation of the regional mean δ^{18}O and of the SACZ index. These corrections change several numbers in the Results, but the main conclusions of the study remain the same.

Community comment CC1 (Giacomo Medici)

General comments

Very good research on paleoclimate and stable isotopes. The specific comments can improve the manuscript.

Response. Thank you for the positive assessment and for the constructive comments, we have addressed your comments point by point below.

Specific comments

Lines 18-20. Insert more recent literature on the use of oxygen (δ 18O) or hydrogen (δD) for hydrological and paleo-hydrological studies.

- Medici, G., Marianelli, D., Cornacchia, I., Gori, F., Brandano, M. 2026. Multi-disciplinary approach to paleokarst occurrence in the Eocene–Oligocene succession of the Apulia Carbonate Platform (Salento, Italy). Facies, 72(2), 17.

- Barbieri, M. 2019. Isotopes in hydrology and hydrogeology. Water, 11(2), 291.

Response. Thank you for the suggestion. We have added a sentence on the use of δ^{18}O and δD in hydrological, hydrogeological and paleohydrological studies to the Introduction. It cites two recent reviews that cover these applications *(Bowen et al., 2019; Konecky et al., 2020).*

> *"For this reason, $\delta^{18}O$ and δD are widely used as tracers in hydrological, hydrogeological and paleohydrological studies. Applications range from tracing groundwater recharge and catchment water sources (Bowen et al., 2019) to compiling global paleoclimate networks (Konecky et al., 2020)."*

Line 72. The general goal of your research is clear. But, what about the specific objectives? You can describe the 3 to 4 specific objectives of your research by using numbers (e.g., i, ii, and iii) at the end of your introduction.

Response. We agree and have added 4 specific objectives at the end of the Introduction.

> *"Within this general goal, our study pursues 4 specific objectives. These are (i) to evaluate the simulated Holocene temperature and precipitation $\delta^{18}O$ against proxy reconstructions and another independent transient simulation, (ii) to characterize the climatic and isotopic responses to orbital forcing and to the 8.2 ka freshwater perturbation, (iii) to quantify the drivers of regional monsoon $\delta^{18}O$ with the MLR framework and to test whether their influence is stable through time, and (iv) to assess the spatial scale at which the amount effect holds and what this implies for the interpretation of speleothem records."*

Line 87. Provide more detail on the boundary conditions. You should at least describe the types.

Response. We have expanded the description of the experimental design (Section 2.2), including the astronomical forcing, greenhouse gases and ice sheets, lakes, river runoff, and the treatment of the vegetation.

> *"The transient simulation analyzed here is integrated for 8,300 model years without acceleration, covering the period from 8.3 ka to 1950 CE. Stable water isotopes and dynamic vegetation are switched on throughout the run. The simulation is initialized from an 8.3 ka equilibrium time-slice state, and its final model year corresponds to 1950 CE. Three types of time-varying boundary conditions are prescribed, namely astronomical forcing, atmospheric greenhouse gas concentrations, and ice-sheet boundary conditions. Orbital parameters (eccentricity, obliquity and longitude of perihelion) are computed following Berger (1978) and updated every model year. Atmospheric CO_2, CH_4 and N_2O concentrations are taken from the ice-core composite of Köhler et al. (2017) and are also updated every model year. CO_2 rises from 260 ppm at 8.2 ka to about 280 ppm in the late Holocene and reaches 311 ppm in the final model year, because the forcing record extends to 1950 CE. Ice-sheet extent and topography follow the GLAC-1D reconstruction (Tarasov et al., 2012; Briggs et al., 2014). The glacier mask, the surface orography and its subgrid-scale parameters, the lake fraction, and the river-routing parameters are updated every 100 model years using GLAC-1D fields at 100-yr intervals, whereas the land-sea mask and the ocean bathymetry are kept at their present-day configuration. Vegetation cover is computed interactively by the dynamic vegetation module of JSBACH (Brovkin et al., 2009; Reick et al., 2013). It evolves from the natural vegetation of the 8.3 ka initial state, and no anthropogenic land use is applied. Solar irradiance and volcanic forcing are not varied."*

Line 87. Provide reference for the types of boundary conditions of your model.

Response. Thanks. References are now given for every type of boundary condition, namely *Berger (1978)* for the orbital parameters, *Köhler et al. (2017)* for the greenhouse gases, *Tarasov et al. (2012)* and *Briggs et al. (2014)* for the GLAC-1D ice sheets, *Brovkin et al. (2009)* and *Reick et al. (2013)* for the dynamic vegetation, and *Barber et al. (1999)*, *LeGrande et al. (2008)* and *Carlson et al. (2008)* for the 8.2 ka freshwater forcing (please see the text quoted in our previous response).

Line 104. Table 1 needs a link with a map?

Response. Yes, thank you for this suggestion (see also the related comment on Table 1 below and the comment of Referee 1). Figure 6a already showed the precipitation regions as filled boxes. We have now added the moisture source regions as dashed boxes in the same colors, and the caption of Table 1 refers to this map (Figure R1). Moreover, Table 1 itself has been extended with information for the moisture source regions.

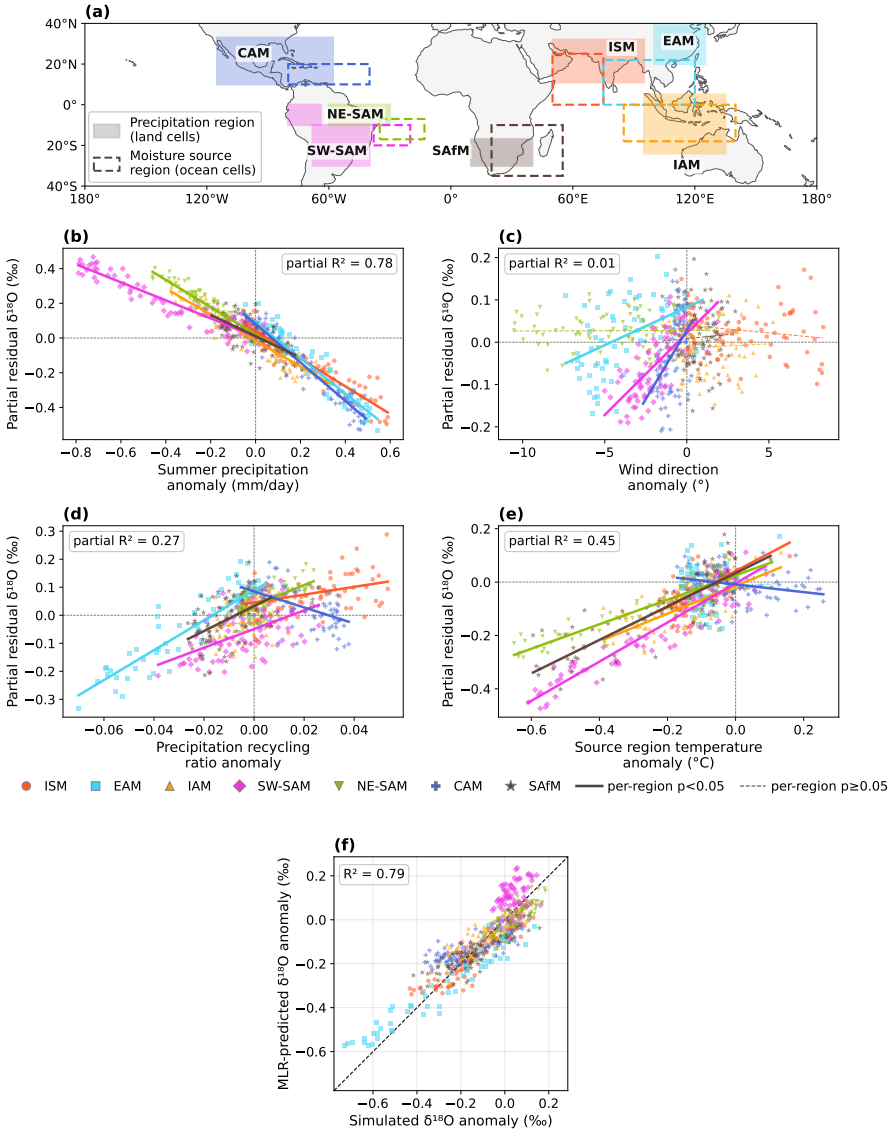

Figure R1. Revised Fig. 6. Panel (a) now shows the moisture source regions (dashed boxes) in addition to the precipitation regions (filled boxes). The caption also explains the solid and dashed within-region lines in panels (b)–(e). All regressions use the corrected, area- and amount-weighted regional δ^{18}O.

Line 233. Some issues with the significant figures after the comma. Please, fit it.

Response. Thank you for spotting the typo. The value "0,654" should be 0.654. We have corrected this. For consistency, now in the revised version, R^2 are now given with two decimals. Note that several regional values have changed slightly, because we fixed an error in our calculation.

Line 276. Here, I can see a paragraph with references and new concepts. I would keep it as a discussion creating a separate conclusion.

Response. We now organized the discussion in seven subsections. We also added a separate conclusion section that summary the main findings of our study.

Table 1. I would couple the table with a figure that shows the study regions.

Response. The caption of Table 1 is now explicitly linked to the map in Fig. 6a (Figure R1 in this letter), which shows both the monsoon regions and the moisture source regions.

Figure 3b. Add the associated equations.

Response. We were not sure which equations are meant. Panel (b) is a scatter plot that compares modeled and measured values, so it shows only the 1:1 line and the RMSE.

Figure 3. Make the legends larger and readable.

Response. Done. Please refer to Figure R2 below.

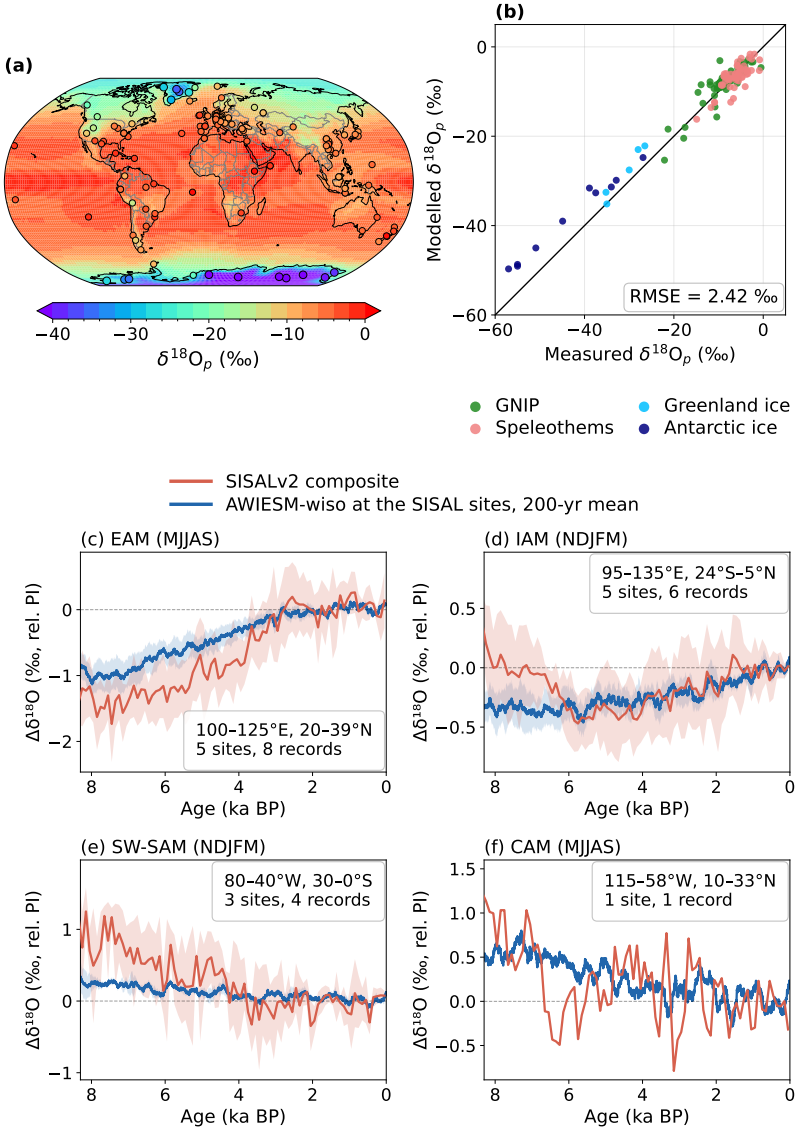

Figure R2. Revised Fig. 3 with larger legends.

Figure 4. Same here, make the legends larger and readable.

Response. Done. The colorbar of Fig. 4 are now horizontal and much larger (also refer to Figure R3 below).

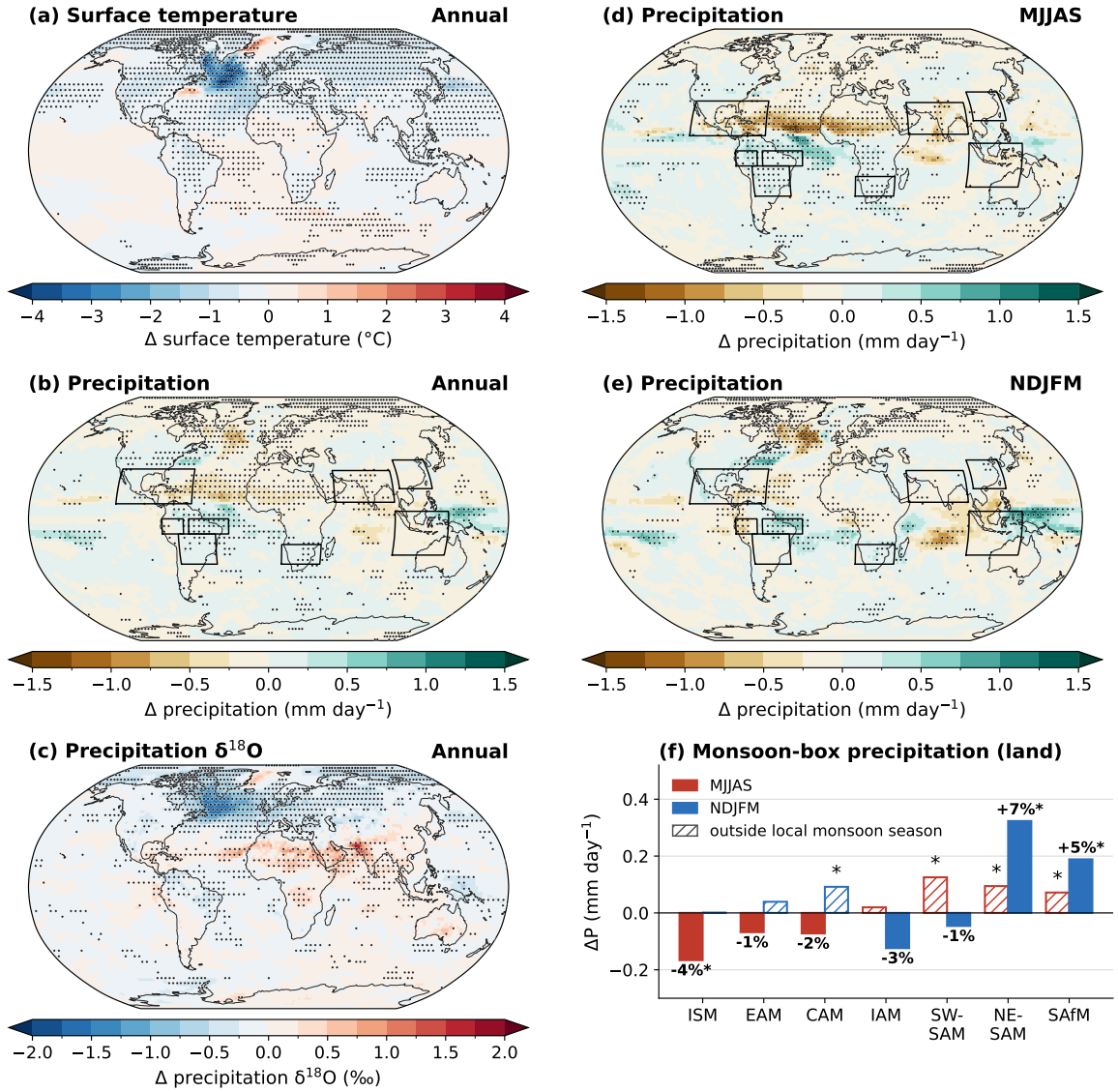

Figure R3. Revised Fig. 4 with enlarged colorbars and the new seasonal precipitation panels (d)–(f) requested by Referee 1.

Figure 9. Enlarge font size for the content of the two axes.

Response. Done. Please refer to Figure R4 below.

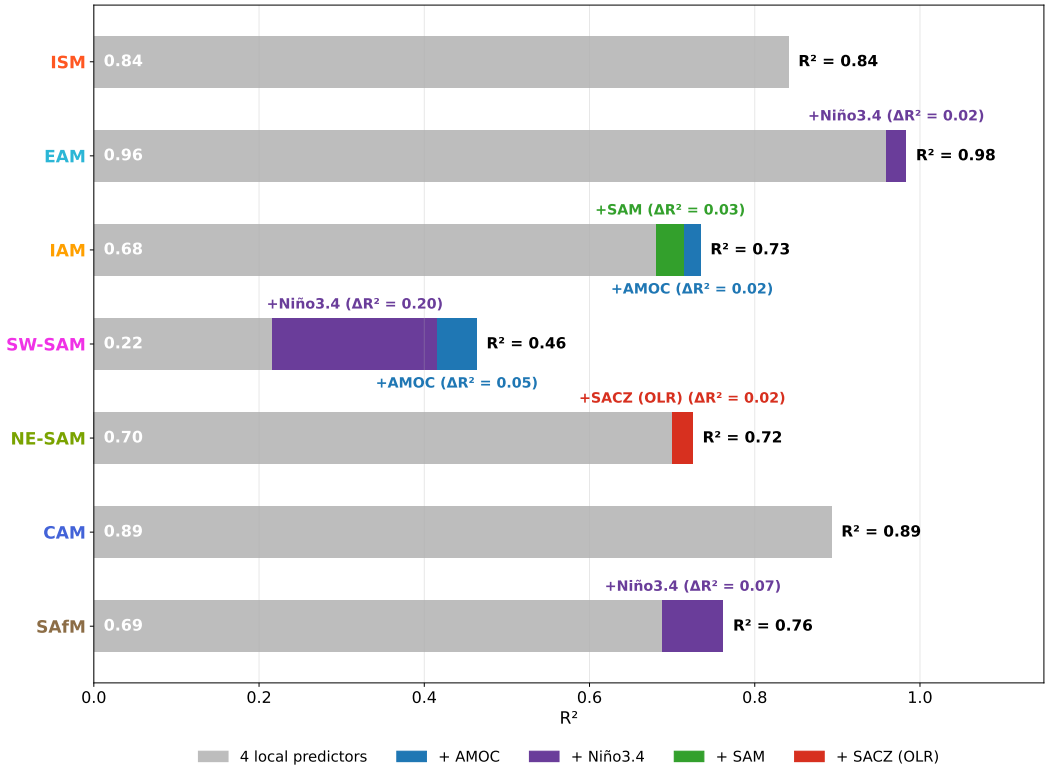

Figure R4. Revised Fig. 9 with enlarged axis labels and annotations. Its content was also slightly updated after some corrections.

Editor comment EC1 (Julien Emile-Geay)

Dear Dr Medici,

Thank you for offering some constructive feedback on the manuscript, which the authors can choose to incorporate into their revision. However, I find the suggestion to cite your own work (as well as that of a senior colleague of yours at La Sapienza) highly problematic. Without an explanation as to why either of those citations is essential to this manuscript, this reads as a naked attempt to bolster your citation statistics or those of a collaborator, neither of which the intent of a public review and comment process like this one. Please provide a well-argued justification or drop this suggestion altogether. If I am not convinced, I will advise the authors to ignore this suggestion.

Best regards,

Julien Emile-Geay

Response. We thank the Editor for this comment. We agree that the two references suggested in CC1 are not needed for the argument of the manuscript, and we have not included them in the revised version. Instead, we cite two reviews that cover these applications directly *(Bowen et al., 2019; Konecky et al., 2020)*.

Referee 1 (RC1)

Shi et al. present a new transient simulation of the Holocene using the isotope-enabled AWI-ESM2-wiso model. The simulation successfully captures some of the large-scale features associated with orbital and meltwater forcing, such as meridional displacements of the ITCZ and the bipolar seesaw. The authors find generally good agreement with proxy records and conduct a multiple

linear regression analysis to determine the principal drivers of δ^{18}O variability in several core monsoon regions.

Overall, this is an exciting and important new data product for the paleoclimate research community, particularly for understanding the climate drivers of water isotopes. Applying the revisions suggested below would help bolster this manuscript.

Response. Thank you for the very positive and constructive comments, which helped improve our manuscript. Below are our point by point response.

How does this simulation compare with the iTraCE Holocene run (Bao et al., 2025)? That is also a transient, isotope-enabled run featuring orbital, GHG, ice, and meltwater forcing. Are there any notable differences in how these models simulate the effects of these forcings during the Holocene? Is one of the models better correlated with observations?

Reference: Bao, Y., Liu, Z., Thompson, L. G., Mosley-Thompson, E., Wan, L., & Lu, J. (2025). Climate simulations and ice core data highlight the Holocene conundrum over tropical mountains. Communications Earth & Environment, 6(1), 205. https://doi.org/10.1038/s43247-025-02188-2

Response. We thank the reviewer for this excellent suggestion. The iTRACE Holocene output is publicly available *(Bao et al., 2024, Zenodo)*, so we performed a direct comparison of the two simulations with each other and with the proxy data. The results are shown in a new supplementary figure (Fig. S2, see also Figure R5 below) and discussed in a new Section 4.3.

> *"The iTRACE Holocene simulation (Bao et al., 2025a, 2024) provides an independent isotope-enabled transient run over 11–0 ka with a comparable set of forcings. It was performed with iCESM1.3, whose atmosphere has a horizontal resolution of about 1.9°×2.5° and whose ocean has a nominal resolution of 1°. It is driven by orbital forcing, greenhouse gases, ICE-6G ice sheets and bathymetry, and a meltwater flux derived from reconstructed sea-level change (Bao et al., 2025a; He et al., 2021). Unlike our simulation, iTRACE prescribes vegetation, dust and aerosols at their PI state. We compare the two simulations with each other and with the proxy data in Fig. S2 in the Supplement, using annual precipitation-weighted $\delta^{18}O$ at the SISAL sites for both models and JJA and DJF seasonal means for the regional comparison, which are the seasons available from iTRACE."*

The two simulations differ in model (iCESM1.3 versus AWI-ESM2-wiso), ice-sheet reconstruction (ICE-6G versus GLAC-1D), meltwater forcing (a flux converted from reconstructed sea-level change versus a two-stage freshwater release at 8.2 ka), and vegetation (prescribed at its PI state versus dynamic). They agree closely in the evolution of global-mean temperature, and both warm through the Holocene, by 0.42°C in AWI-ESM2-wiso and 1.12°C in iTRACE between 8–7 ka and the 0.5–1.5 ka reference period, whereas the reconstruction of *Marcott et al. (2013)* cools by 0.47°C. Part of the difference between the two models stems from centennial cold excursions in iTRACE after 8.2 ka, which *Bao et al. (2025)* related to AMOC variations in their simulation.

Regarding the second question, both models capture the long-term evolution of the SISAL speleothem composites in EAM and SW-SAM. On 100-yr bins over 8–0 ka, the correlations are 0.95 and 0.74 for AWI-ESM2-wiso and 0.89 and 0.47 for iTRACE. In IAM they are 0.53 and −0.25, and neither model reproduces the single CAM record. Both models underestimate the amplitude of the Holocene δ^{18}O trend. Because these correlations are dominated by the long-term trends and rest on few independent samples, we refrain from ranking the two models. The regional summer changes in precipitation and δ^{18}O agree in sign in 5 of the 7 regions. In EAM both models simulate a pronounced early-Holocene δ^{18}O depletion, while the JJA rainfall change differs in sign. This suggests that the driver hierarchy diagnosed with the MLR can be model dependent, which we now state in the manuscript.

> *"Both models capture the long-term evolution of the SISAL composites in EAM and SW-SAM over 8–0 ka (Fig. S2b–e). The agreement is weaker in IAM, and neither model reproduces the centennial*

structure of the single CAM record. Moreover, both models underestimate the amplitude of the Holocene trend.

The regional summer changes are broadly consistent between the models (Fig. S2f, g). The sign of the precipitation change from 8–7 ka to 1–0 ka agrees in 5 of the 7 monsoon regions, and the sign of the $\delta^{18}O$ change also agrees in 5 regions. Inside the EAM box both models simulate a pronounced early-Holocene depletion of about −1.0 and −0.8 ‰, although the JJA rainfall change differs in sign (+14 % in AWI-ESM2-wiso and −5 % in iTRACE). In iTRACE the early-Holocene rainfall maximum lies over northern and northeastern China, largely north of the EAM box, whereas rainfall over central eastern China increases through the Holocene (Ning et al., 2025, 2026). Speleothem records also show opposite early- to mid-Holocene rainfall changes in Northeast Asia and North China (Wang et al., 2025), so that a regional mean is sensitive to the position of the rain belt. The relative roles of local rainfall and of changes in moisture source and upstream rainout (Liu et al., 2014a; Battisti et al., 2014; Pausata et al., 2011) for EAM $\delta^{18}O$ may therefore differ between models, and repeating the MLR analysis with other isotope-enabled transient simulations would test the robustness of the driver hierarchy. The North African greening discussed in Sect. 4.1 does not occur in iTRACE, in which vegetation is prescribed at its PI state."

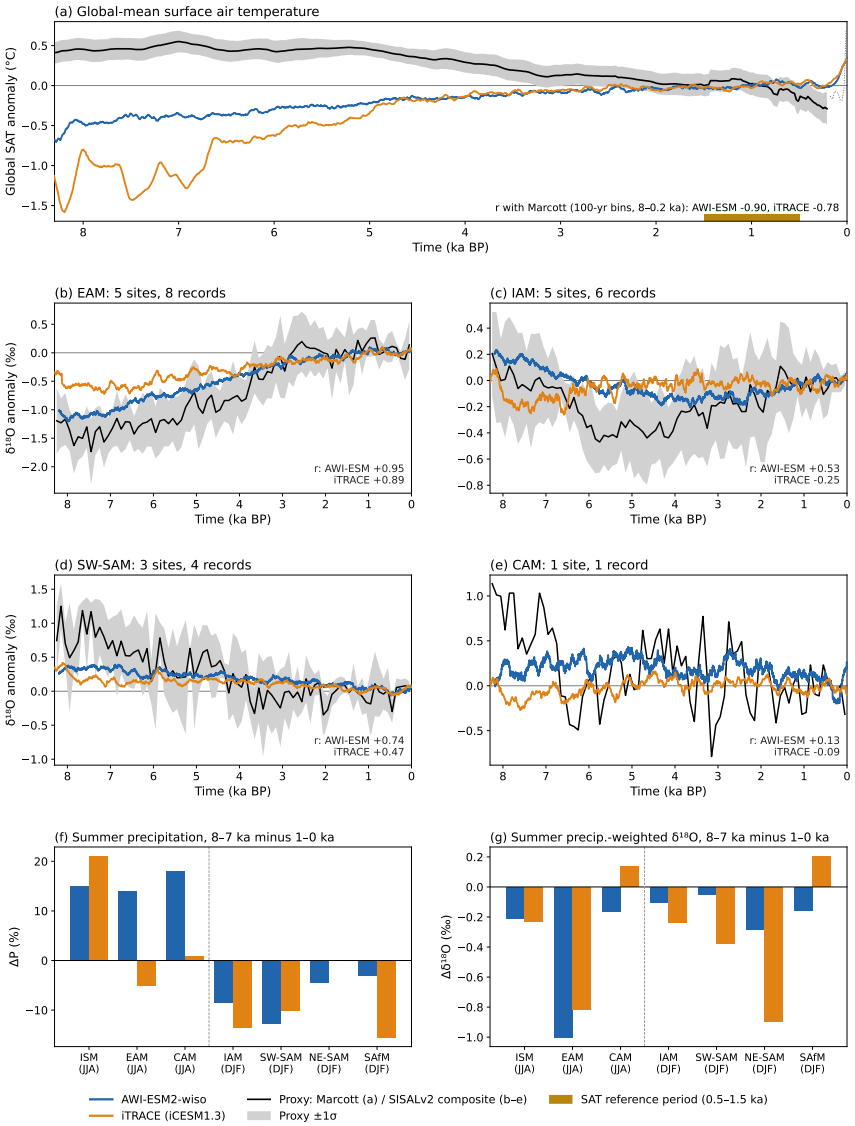

Figure R5. New Fig. S2 in the Supplement. Comparison of AWI-ESM2-wiso (blue) with the iTRACE Holocene simulation (orange) and proxy reconstructions (black). (a) Global-mean surface air temperature with the stack of Marcott et al. (2013). (b)–(e) Annual precipitation-weighted $\delta^{18}O$ at the SISAL sites of four monsoon regions together with the SISALv2 composite. (f, g) Early-minus-late Holocene change of summer precipitation and $\delta^{18}O$ in the seven monsoon regions.

There should be a mention of the Green Sahara, particularly because of the dynamic vegetation turned on in the model. This seems to be a driver of some of the largest simulated anomalies in the Holocene trend maps (Figure 5).

Response. We fully agree and thank the reviewer for pointing this out. We analyzed the vegetation output of the dynamic vegetation module (vegetated fraction, leaf area index) together with precipitation over North Africa and added a new supplementary figure (Fig. S1, Figure R6), a paragraph in Section 3.3 and a paragraph in the Discussion (Section 4.1).

The analysis shows that, over 15°W–35°E, 15–30°N, the vegetated fraction declines from 0.32 in the early Holocene to 0.08 at PI, and MJJAS precipitation from 1.37 to 0.15 mm day^{-1}. The region also shows a summer warming toward the present that opposes the insolation decline, consistent with reduced evaporative cooling and reduced cloud cover (more radiative flux) as the land surface dries.

The analysis also shows that the simulated greening is limited. Proxy records indicate that vegetation and lakes extended to about 23–30°N during the African Humid Period *(deMenocal et al., 2000; Shanahan et al., 2015; Tierney et al., 2017)*. While our simulation captures a clear northward expansion of early-Holocene vegetation relative to present conditions, the northern limit of this greening falls short of the maximum extent implied by proxy evidence. In the regional mean the simulated decline is gradual. Half of the vegetation loss has occurred by about 4.0 ka, and 10 to 90 % of it occurs between 6.2 and 1.9 ka. The decline occurs earlier in the Sahara (20–30°N, midpoint 5.0 ka) than in the Sahel (10–20°N, midpoint 3.1 ka). This southward progression is consistent with *Shanahan et al. (2015)*, who found that the African Humid Period ended progressively later at lower latitudes, and with the transient simulation of *Dallmeyer et al. (2020)*, in which it ends earlier in the north than in the south. *Shanahan et al. (2015)* also found the termination to be locally abrupt, which our regional means do not reproduce. We discuss these points as follows:

> *"Since the dynamic vegetation is active in our simulation, here we examine the North African vegetation response. Proxy records document an African Humid Period during which vegetation and lakes extended to about 23–30°N, and which ended between about 6 and 4 ka (deMenocal et al., 2000; Shanahan et al., 2015; Tierney et al., 2017). While our simulation captures a clear northward expansion of early-Holocene vegetation relative to present conditions (Fig. S1), the northern limit of this greening falls short of the maximum extent implied by proxy evidence. An underestimated northward extent of the monsoon rain belt and vegetation is common to most climate models and has been linked to missing or weak vegetation, soil and dust feedbacks (Kutzbach et al., 1996; Pausata et al., 2016, 2020). The simulated decline is gradual, with the midpoint of the vegetation loss at about 4.0 ka, while the loss progresses from 10 to 90 % of its total between 6.2 and 1.9 ka. It lags the precipitation decline, whose midpoint is at 5.0 ka, and it progresses southward from about 5.0 ka in the Sahara (20–30°N) to 3.1 ka in the Sahel (10–20°N). This time-transgressive retreat agrees with lake and leaf-wax records (Shanahan et al., 2015) and with the transient simulation of Dallmeyer et al. (2020), whereas the abrupt collapse of vegetation found in some models and marine records (Claussen et al., 1999; deMenocal et al., 2000; Hopcroft and Valdes, 2022) is not reproduced. Because North Africa has the strongest MJJAS $\delta^{18}O$ trends of all land regions, the limited greening suggests that the simulated Holocene $\delta^{18}O$ change over North Africa is a conservative estimate. None of the 7 monsoon regions analyzed with the MLR lies in North Africa, so this limitation does not enter the regional regressions directly."*

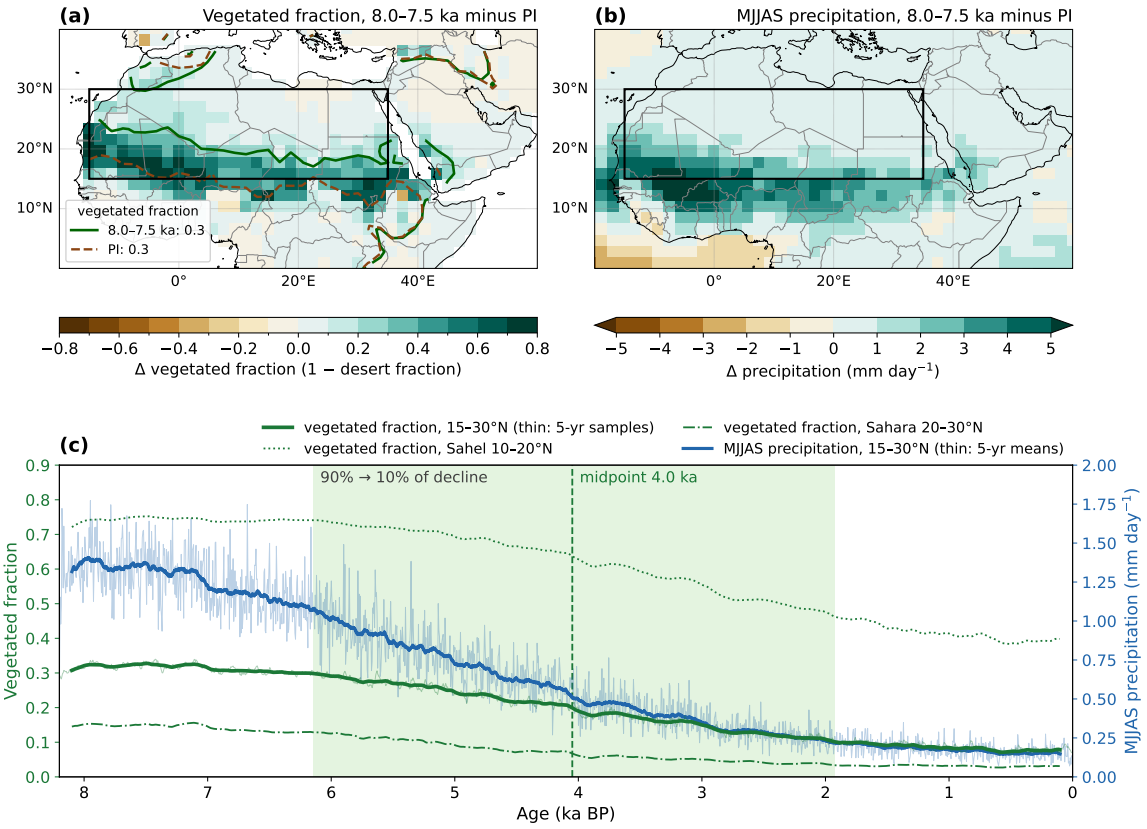

Figure R6. New Fig. S1. (a) Change in vegetated fraction and (b) change in MJJAS precipitation between the early Holocene (8.0–7.5 ka) and PI over North Africa. (c) Time series of the vegetated fraction and MJJAS precipitation over 15°W–35°E, 15–30°N.

It is worth discussing what could account for the remaining variance after applying the predictors to the MLR framework. For instance, there is still ∼50 % unexplained variance in the SW-SAM region after accounting for forcings and internal modes. And the variance captured by the pooled regression is just a bit higher than that (63 %).

Response. We agree that this deserves a dedicated discussion, and we have added a quantitative analysis (new Fig. 11, Figure R7, new Methods paragraph, and new Section 4.6). Before summarizing it, we note that the R^2 values quoted by the reviewer have changed in the revised manuscript. After correcting the regional averaging of δ^{18}O (see our response to the comment on line 111), the pooled regression explains 79 % instead of 63 % of the variance. The final SW-SAM model, which includes ENSO and AMOC, explains 46 %.

To find out how much variance can be explained at all, we estimated the variance that interannual-to-multidecadal internal variability leaves in each 100-yr bin mean. We used four estimators, based on two high-pass filters and two ways of accounting for autocorrelation. This noise accounts for nearly all of the centennial δ^{18}O variance in SW-SAM, for 73 % in NE-SAM and 48 % in IAM, but only for 4 % in EAM and 14 % in ISM. Because the predictors come from the same simulation, they share part of this internal variability. The remaining, irreducible noise therefore sets an upper limit to R^2, which ranges from 0.38 in SW-SAM to 0.97 in EAM. In six of the seven regions, the residual variance of the final regression is within about 30 % of this irreducible noise, or below it (in SW-SAM, EAM and CAM the final R^2 slightly exceeds the estimated limit, which reflects the uncertainty of the estimate and some fitting of noise by the stepwise selection). The unexplained variance in these regions, including SW-SAM, is therefore consistent with internal variability rather than with a missing forced process. In SW-SAM the forced centennial signal is simply weak. Its simulated δ^{18}O shows little Holocene trend, and its centennial range is only about twice the noise

standard deviation, compared with about fifteen times in EAM.

ISM is the exception, as its residual is about twice the irreducible noise. We discuss physical processes that the four predictors represent only indirectly and that could account for these residual signals. These include changes in the distribution of rainfall among the months of the season, changes in convective organization and rain re-evaporation that modify the amount effect itself *(Risi et al., 2008; Moore et al., 2014; Konecky et al., 2019)*, upstream rainout and source changes beyond the fixed source regions, and nonlinear responses that a linear regression cannot represent.

> *"Several factors account for the variance that the regression leaves unexplained. Internal variability is the most important of them. Even after averaging over 100 years, interannual to multidecadal variability leaves noise in the bin means that no forced predictor can explain (Fig. 11). This noise accounts for nearly all of the centennial $\delta^{18}O$ variance in SW-SAM, for 73 % in NE-SAM and 48 % in IAM, but for only 4 % in EAM and 14 % in ISM. Because the predictors share part of this variability, the regression can explain more than the forced signal alone. The irreducible part of the noise sets an approximate upper limit on the attainable R^2, which ranges from 0.38 in SW-SAM to 0.97 in EAM (diamonds in Fig. 11a). The final R^2 of SW-SAM, EAM and CAM slightly exceeds this limit, which reflects the uncertainty of the noise estimate and the tendency of stepwise selection to fit part of the noise. In six of the seven regions, the residual variance of the final regression is within about 30 % of this irreducible noise or below it. The unexplained variance in these regions is therefore consistent with internal variability rather than with a missing forced process. The low R^2 of SW-SAM in particular reflects a weak forced signal. Its simulated $\delta^{18}O$ shows little Holocene trend, and its centennial range is only about twice the noise standard deviation (Fig. 11b), compared with about fifteen times in EAM (Fig. 11c). The 21 % left unexplained by the pooled regression combines this internal noise with regional differences in the secondary coefficients, which a pooled fit with common slopes cannot represent.*
>
> *ISM is the exception. Its residual variance is about twice the irreducible noise, and its largest residual occurs in the century after the 8.2 ka hosing. Processes that the four predictors represent only indirectly could account for these residual signals. Changes in the distribution of rainfall among the months of the monsoon season alter the amount-weighted $\delta^{18}O$ without changing the seasonal total. Changes in convective organization and in the re-evaporation of falling rain modify the strength of the amount effect itself (Risi et al., 2008; Moore et al., 2014; Konecky et al., 2019). Upstream rainout along the moisture trajectory and shifts of the moisture source beyond the fixed source regions are captured only partly by the box-mean source temperature and wind direction. Moreover, the linear form of the regression cannot represent nonlinear responses or interactions among the predictors. Quantifying these processes would require diagnostics such as water tagging, which is beyond the scope of this study."*

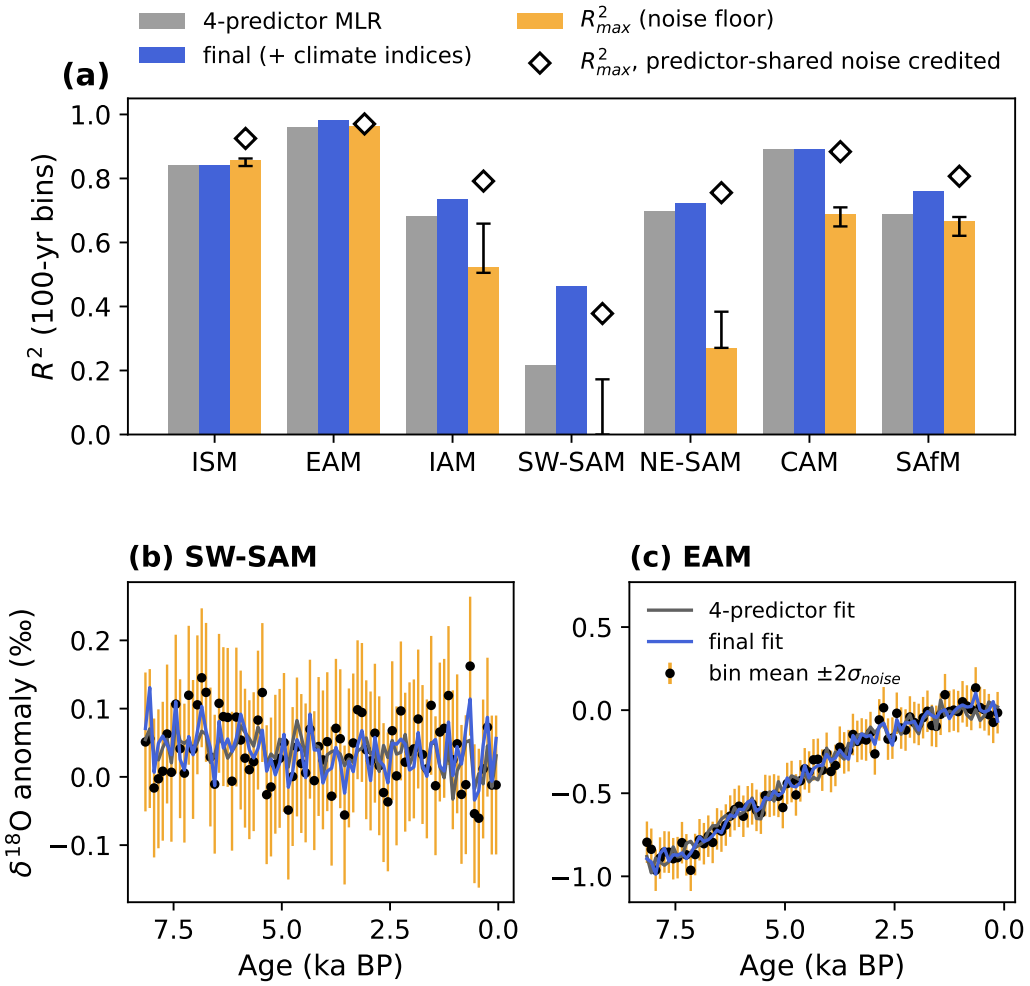

Figure R7. New Fig. 11. (a) R^2 of the four-predictor model (gray) and of the final model with climate indices (blue), compared with the fraction of the 100-yr bin variance that is not internal noise (orange, with the range of four noise estimates). Diamonds show the upper limit when the internal variability shared with the predictors is counted as explainable. (b, c) Bin means with $\pm 2\sigma$ noise bars and the regression fits for SW-SAM and EAM.

Line 104: Where are these source regions and how were they determined?

Response. The moisture source regions are taken without modification from the analysis code of *Parker et al. (2021)*, who defined them on the basis of published moisture-tracking studies and the simulated summer surface winds of GISS ModelE. We retained these definitions so that our transient results can be compared directly with their time-slice results. In the revised manuscript the coordinates of all source regions are listed in Table 1, they are drawn as dashed boxes in Fig. 6a (see also Figure R1 in this letter), and the text explains how they were determined and which ocean basin each one covers (Section 2.3).

> *"The moisture source region comprises the ocean grid cells (land fraction below 50 %) of an upstream area for each monsoon system, and the source temperature and wind direction predictors are extracted there. The source regions are taken without modification from the analysis code of Parker et al. (2021), who defined them on the basis of moisture-tracking studies and the simulated summer surface winds of their model. Retaining identical domains allows a direct comparison between their time-slice results and our transient results. The source regions cover the Arabian Sea for ISM, the Bay of Bengal and the South China Sea for EAM, the eastern Indian Ocean and the Maritime Continent seas for IAM, the tropical North Atlantic and Caribbean for CAM, the western tropical South Atlantic for SW-SAM and NE-SAM, and the southwestern Indian Ocean for SAfM."*

Line 107: There is too little detail in this section about the multi-linear regression method, the authors should further elaborate. What regression and fitting methods are employed? How is significance determined? In other words, for those that want to reproduce these results, what do they need to know?

Response. We agree. Section 2.4 has been rewritten and is now divided into 4 subsections (response variable, predictors, regression procedure, and large-scale climate indices). The full text of the revised section reads as follows.

*"**2.4.1 Response variable***

The response variable is the summer-season precipitation $\delta^{18}O$ of each monsoon region, computed for every model year in two steps from monthly output. The $\delta^{18}O$ of precipitation at each grid cell is weighted by the monthly precipitation over the months of the regional summer season,

$$\delta^{18}\mathrm{O}_i = \frac{\sum_m P_{i,m}\,\delta^{18}\mathrm{O}_{i,m}}{\sum_m P_{i,m}}, \tag{1}$$

where m runs over the 5 months of the season (MJJAS or NDJFM), $P_{i,m}$ is the monthly mean precipitation rate (mm day^{-1}) at grid cell i, and $\delta^{18}O_{i,m}$ is the monthly mean isotope ratio of precipitation (‰). The grid-cell values are then combined into a regional value weighted by grid-cell area and seasonal precipitation,

$$\delta^{18}\mathrm{O}_{\mathrm{precip}} = \frac{\sum_i A_i\,\bar{P}_i\,\delta^{18}\mathrm{O}_i}{\sum_i A_i\,\bar{P}_i}, \tag{2}$$

where the summation extends over all land grid cells of the precipitation region, A_i is the grid-cell area (m^2), and $\bar{P}_i = \frac{1}{5}\sum_m P_{i,m}$ is the seasonal mean precipitation rate. Equation (2) represents the isotopic composition of the total seasonal rainfall over the region.

2.4.2 Predictors

Four fixed predictors are included in every regional model. (1) `region_ppt` is the seasonal mean precipitation anomaly (mm day^{-1}), averaged with area weights over the land grid cells of the precipitation region. (2) `wind_dir` is the anomaly of the seasonal mean wind direction (°) over the moisture source region at the lowest model pressure level (1000 hPa). The direction is computed for each ocean grid cell with the two-argument arctangent of the zonal (u) and meridional (v) wind components, expressed in degrees clockwise from north, and then averaged over the source region. Anomalies are wrapped to the range −180 to 180°. This term is a proxy for large-scale atmospheric circulation and moisture transport pathways. The lowest level is used for consistency with the surface winds of Parker et al. (2021), and because specific humidity, and with it the moisture transport, is largest near the surface. The four-predictor regressions change little when the wind direction at 850 or 700 hPa is used instead (not shown). (3) `source_temp` is the seasonal mean surface temperature anomaly (°C) over the ocean grid cells of the moisture source region, which is equivalent to the sea surface temperature. (4) `ppt_rec` is the anomaly of the precipitation recycling index,

$$\mathrm{RI} = \frac{E}{E + 2\,F_{\mathrm{in}}}, \tag{3}$$

where E is the area-integrated evaporation over the land grid cells of the precipitation region and F_{in} is the column-integrated moisture flux entering the region across the four edges of its bounding box, both in kg s^{-1}, following Brubaker et al. (1993). The column-integrated moisture flux is computed as

$$\mathbf{F} = \frac{1}{g}\sum_k q_k\,\mathbf{u}_k\,|\Delta p_k|, \tag{4}$$

where g is the gravitational acceleration, q_k and $\mathbf{u}_k$ are the specific humidity and horizontal wind vector at pressure level k, and Δp_k is the layer thickness. The summation spans all 17 pressure levels from 1000 to 10 hPa. Seasonal mean q and $\mathbf{u}$ are used, so $\mathbf{F}$ represents the moisture transport by the seasonal mean flow. Only inward-directed flux components are counted in F_{in}.

2.4.3 Regression procedure

All climatic and isotopic variables spanning the full 8,300-year simulation are averaged in non-overlapping 100-year bins, which yields 83 bins per region. The bin length equals the 100-yr averaging period of the time slices of Parker et al. (2021), and it suits the comparison with speleothem records, which resolve centennial to orbital variations more robustly than interannual ones (Sect. 3.1). Averaging also suppresses the interannual variability, which is largely unforced. Regressions on annual values and on bins of 10 to 500 yr are compared in Fig. S3 in the Supplement. Anomalies are taken relative to the mean of the 5 youngest bins (0–500 yr BP). Bins in which any variable departs by more than three standard deviations from its mean are removed, using the regional distribution for the regional models and the pooled distribution for the pooled model. The regression is ordinary least squares with an intercept, implemented with the "statsmodels" package (Seabold and Perktold, 2010). Predictors are used in their physical units, so each coefficient gives the change in $\delta^{18}O$ per unit change of the predictor. Two configurations are fitted. The pooled model combines the bins of all seven regions (556 bins after outlier removal), in the same way as Parker et al. (2021). The regional models are fitted separately for each region with 81 or 82 bins. The goodness of fit is measured by the coefficient of determination R^2. The significance of each coefficient is assessed with a two-sided t test of the coefficient against its standard error, with a threshold of $p < 0.05$. In the partial-residual plots, the significance of the relationship within each region is assessed separately by a simple linear regression of the partial residuals on the predictor.

The temporal stability of the coefficients is examined with a rolling regression. The annual series of each region are first smoothed with a centred 100-yr running mean. The regional four-predictor model is then fitted in 1000-yr windows that move in steps of 100 yr over the 8.2–0 ka part of the run. In this analysis, wind direction anomalies are computed at each grid cell before the spatial average, and the recycling index uses the net moisture flux across each boundary of the region, following the procedure of Parker et al. (2021). Because the smoothed series are strongly autocorrelated, a 1000-yr window contains only about ten independent samples. Confidence intervals from ordinary least squares therefore underestimate the true uncertainty, and the effect of this underestimation is quantified in Sect. 4.5.

To assess how much of the centennial $\delta^{18}O$ variance can be explained at all, the variance of internal noise in a 100-yr bin mean is estimated for each region. In the main estimate the forced signal is removed by subtracting the 100-yr bin mean from each annual value, and the variance of a 100-yr mean of these residuals is computed with an effective sample size based on their lag-1 autocorrelation (Zwiers and von Storch, 1995). Three further estimates test the sensitivity to these choices. One keeps the bin means but replaces the lag-1 autocorrelation by the scaling of block-mean variances with block length. The other two remove the forced signal with a 200-yr locally weighted smooth instead of the bin means. One of them uses the lag-1 autocorrelation as above. The other computes the variance directly from non-overlapping 100-yr means of the residuals and is corrected, with AR(1) surrogate series, for the part of the noise that the smooth removes. The fraction of bin-mean variance that is not noise gives an upper limit R^2_{max} for a regression on purely forced predictors. Its central value comes from the main estimate, and the range of all four estimates indicates its uncertainty. Because the predictors are taken from the same simulation, they share part of the internal variability with $\delta^{18}O$. A second, higher limit is therefore computed from the noise that remains after regressing the annual $\delta^{18}O$ anomalies on the annual predictor anomalies within each bin.

2.4.4 Large-scale climate indices

Five annually resolved climate indices are tested as additional predictors, to capture signals missing from the four-predictor framework. The Atlantic Meridional Overturning Circulation (AMOC) index is the maximum of the meridional overturning streamfunction at 27°N between 500 and 2000 m depth. The Niño 3.4 index is the area-weighted annual mean sea surface temperature over 5°S–5°N, 170°W–120°W. The Southern Annular Mode (SAM) index follows Marshall (2003) as the difference in annual zonal-mean sea-level pressure between 40°S and 65°S. The Arctic Oscillation (AO) index is the leading principal component of annual sea-level pressure north of 20°N, with positive values corresponding to reduced Arctic pressure. The strength of the South Atlantic Convergence Zone (SACZ) in austral

summer is represented by outgoing longwave radiation (OLR), which is widely used as a proxy for SACZ convection (Liebmann et al., 1999; Carvalho et al., 2004). The SACZ index is the area-weighted NDJFM OLR anomaly over the climatological SACZ band of the model, taken with reversed sign so that positive values indicate stronger convection. The band is defined as the grid cells within 60–20°W, 35–10°S where the PI NDJFM OLR is below 240 W m^{-2}, and it lies mainly over central Brazil. OLR output is available for 8.2–0 ka, which does not affect the analysis because the bin containing the 8.2 ka hosing is excluded (Sect. 2.4.3). All indices are expressed as anomalies relative to the PI mean, so that their Holocene trends are retained in the same way as for the local predictors, and they are averaged in the same 100-yr bins as the other variables. For each region, indices are selected by forward stepwise regression from all five candidates. At each step the candidate that raises R^2 most is added, provided the gain is at least 0.02, and no more than three indices are added per region to limit overfitting."

Line 111: Please clarify the amount-weighting scheme. Shouldn't this be monthly precipitation rates and δ^{18}O that you are summing and then weighting? The result of this equation is the seasonal precipitation rate, but what is going into it is monthly, no? Unless I am misunderstanding.

Response. We thank the reviewer for this question, which led us to an important correction. The reviewer is right that the weighting is applied to monthly values, and the original Eq. (1) did not make this clear. The computation is now described in detail in Section 2.4.1 (quoted in full in our response to the previous comment). Equation (1) weights the monthly δ^{18}O of each grid cell by its monthly precipitation over the 5 months of the season. Equation (2) then averages the grid cells, weighted by grid-cell area and seasonal precipitation.

While checking the analysis code against these equations, we found an error in the submitted results. The regional value had been computed over the grid cells in which δ^{18}O was defined. A grid cell without summer rain has no δ^{18}O, so it dropped out of the mean, and the number of cells entering the regional value changed through time. This should not happen, because the regional value should always represent the same area. The problem was largest in ISM, where 63 of the 155 land cells, in the Arabian and Thar deserts, receive no summer rain in some years. The number of ISM cells in the mean varied between 105 and 155 from year to year and decreased from about 143 in the early Holocene to about 128 in the late Holocene (Figure R8a). In the corrected calculation all land cells of a region enter every year, and cells without rain receive zero weight. With this correction, the ISM series shows an early-Holocene depletion of 0.34 ‰ relative to PI instead of almost no change (Figure R8b), in line with the weakening of the Indian monsoon through the Holocene recorded by speleothems *(Fleitmann et al., 2003)*. The pooled R^2 increases from 0.63 to 0.79, the ISM R^2 from 0.47 to 0.84 and the CAM R^2 from 0.75 to 0.89, whereas SW-SAM decreases from 0.30 to 0.22 and NE-SAM from 0.80 to 0.70 (Figure R8c). All affected numbers and figures (6-11) were updated.

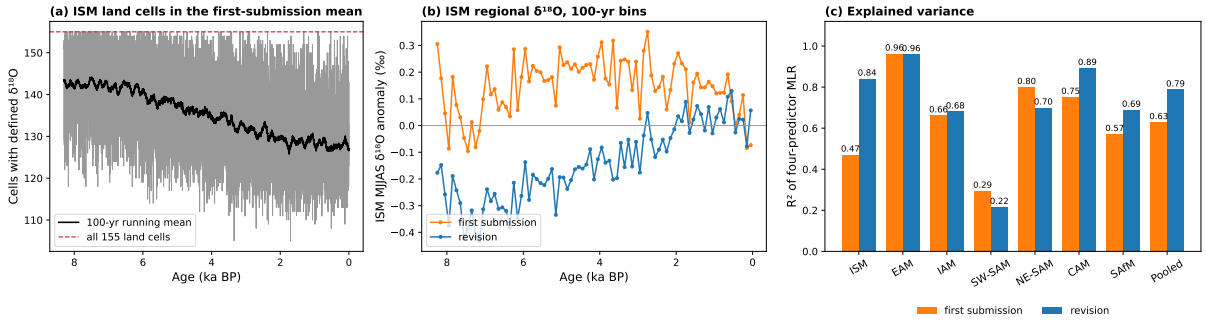

Figure R8. Effect of the regional averaging of δ^{18}O (provided for the referees). (a) Number of ISM land cells with defined MJJAS δ^{18}O in each year. Cells without summer rainfall have undefined δ^{18}O and dropped out of the regional mean of the first submission. (b) ISM regional δ^{18}O anomaly in 100-yr bins in the first submission (orange) and in the revision (blue). (c) R^2 of the four-predictor regression in the first submission and in the revision.

Lines 240–241: Why do you think these regional coefficient magnitudes are nonstationary?

Response. Thank you for this question. We see two possible reasons. The relationships could change physically, as the orbital configuration alters the seasonal cycle, the moisture sources and the transport pathways. However, within a 1000-yr window wind direction, recycling and source temperature vary little, so that their coefficients are poorly constrained. In addition, the 100-yr running mean makes the residuals strongly autocorrelated, so that a 1000-yr window contains only about 10 independent samples. The confidence intervals shown in the submitted Fig. 8 did not account for this and were too narrow. We have revised the corresponding statement of the Discussion (Section 4.5), which had concluded that a regression calibrated at one orbital state does not transfer to another. Figure 8 now shows confidence intervals corrected for the effective sample size, uses the correct units for each coefficient, and has a corrected time axis, which was offset by 100 years in the submitted version. The revised Section 4.5 reads as follows.

> *"The rolling regression shows a stable precipitation coefficient, whereas the coefficients of the secondary predictors vary through time and change sign in several regions (Fig. 8). Two explanations are possible. The relationships may change physically as the orbital configuration alters the seasonal cycle, the moisture source regions and the transport pathways of each monsoon. Alternatively, the coefficients may be poorly constrained within a single window. The 100-yr running mean makes the residuals within a 1000-yr window strongly autocorrelated, with a lag-1 autocorrelation of about 0.98, so that a window contains only about 10 independent samples (Bretherton et al., 1999). The confidence intervals in Fig. 8 are computed with this effective sample size.*
>
> *The apparent non-stationarity of the secondary coefficients is therefore most likely sampling noise. It arises because each window contains only about 10 independent samples, over which wind direction, recycling and source temperature vary little. Physical changes in these relationships cannot be excluded, but they are not detectable at this window length. In contrast, the precipitation coefficient is well constrained. It is significantly negative in about half of all windows and never significantly positive. A regression calibrated in one orbital state can therefore be transferred to another for the amount effect, whereas the secondary coefficients should be treated as uncertain."*

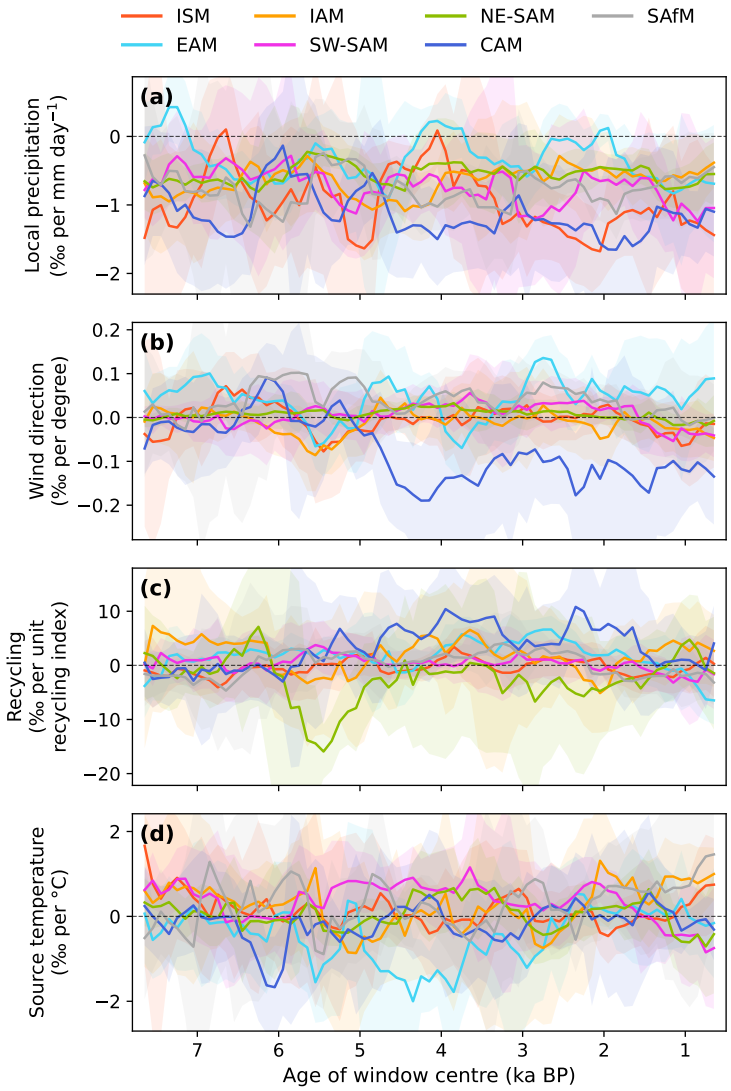

Figure R9. Revised Fig. 8. Rolling regression coefficients with 95 % confidence intervals corrected for the effective sample size, correct units and the corrected time axis.

Lines 345–346: A mechanism is not described here, but rather an observation about the spatial-scale dependence of the amount effect. The authors may wish to replace "This mechanism" with more precise wording.

Response. We agree. The sentence now reads:

> *"This scale dependence provides a physical explanation for the empirical consensus within the SISAL community that regional speleothem composites better represent large-scale monsoon variability than individual cave records (Comas-Bru et al., 2020; Lachniet, 2009). Studies that combine isotope-enabled simulations with speleothem records also show that $\delta^{18}O$ at a cave site need not follow local rainfall. During the last deglaciation, coherent $\delta^{18}O$ changes across the Asian monsoon region were accompanied by rainfall changes of opposite sign within the region (He et al., 2021), and in arid Central Asia precipitation $\delta^{18}O$ has been shaped by changes in moisture-source trajectories linked to the westerlies (Cai et al., 2026)."*

Table 1: To help the reader, mention in the caption that a map with boxes representing the monsoon study regions is shown in Figure 6.

Response. Done. The caption of Table 1 now refers to Fig. 6a, which shows both the precipitation regions and the newly added moisture source regions.

"A map of the precipitation regions (filled boxes) and source regions (dashed boxes) is shown in Fig. 6a."

Figure 2: How do these curves compare with the actual Marcott et al. 2013 reconstruction?

Response. This is a very useful suggestion. Figure 2 now includes (also shown in Figure R10 below) the global and latitudinal-band stacks of *Marcott et al. (2013)* (standard 5°×5° stack with $\pm 1\sigma$), taken from the supplementary database of that study. Because the reconstructed stacks lose most records in the last two centuries, all model and proxy curves are now expressed relative to their 0.5–1.5 ka mean, which is close to the window used by *Marcott et al. (2013)* to align their stack with the instrumental-era reconstruction. A 300-yr running mean of the seasonally biased site-stack is added, because the Monte Carlo procedure of *Marcott et al. (2013)* removes most variability on shorter timescales.

The comparison shows that the model area mean warms by 0.42°C from 8–7 ka to 0.5–1.5 ka, whereas the reconstruction cools by 0.47°C. In the northern extratropics, seasonal sampling turns the simulated warming into a cooling of 0.57°C, about half of the reconstructed 1.09°C. In the tropics and southern extratropics the reconstructed changes are small and within their 1σ uncertainty. We added this quantitative comparison to Section 3.1 and revised the corresponding statement in the Discussion, which now says that seasonal sampling explains part, but not all, of the model–proxy offset.

"To compare the model against temperature reconstructions, we follow the site-stack method of Liu et al. (2014b). Model temperature is bilinearly interpolated to the 73 proxy locations of Marcott et al. (2013). For the seasonally biased site-stack, each location is sampled in the season reported in the original database, and all Northern Hemisphere marine sites are sampled in boreal summer as in Liu et al. (2014b). Figure 2 compares the simulated temperature evolution with the reconstructed stacks of Marcott et al. (2013) for the globe and three latitudinal bands. All curves are expressed relative to their 0.5–1.5 ka mean, because the reconstructed stacks are not robust over the last two centuries. Both the model area mean and the annual site-stack warm toward the present in all bands. Between 8–7 ka and the reference period, the global area mean warms by 0.42°C, whereas the reconstruction cools by 0.47°C. This opposite sign is the Holocene temperature conundrum (Liu et al., 2014b). The seasonally biased stack yields markedly different behavior. Over the Northern Hemisphere extratropics, the simulated warming is reversed into a cooling of 0.57°C, which is about half of the reconstructed cooling of 1.09°C. Seasonal sampling therefore explains a substantial part, but not all, of the model-proxy discrepancy in this band. In the tropics and the Southern Hemisphere extratropics, the seasonal bias exerts a weak influence. The reconstructed changes there are small (+0.10 and −0.11 °C) and lie within the 1σ uncertainty of the stack, while the model warms by 0.54 and 0.20°C."

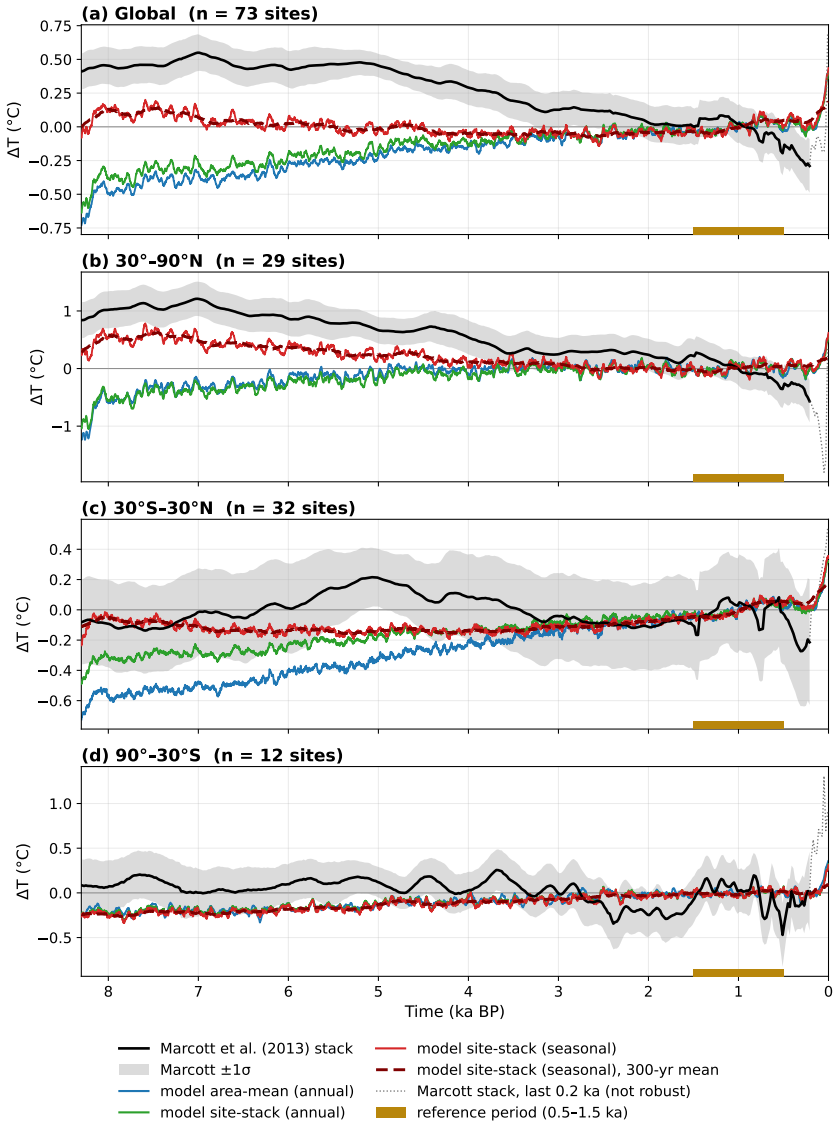

Figure R10. Revised Fig. 2 including the global and latitudinal-band stacks of Marcott et al. (2013) (black line with $\pm 1\sigma$ shading). All curves are anomalies relative to their 0.5–1.5 ka mean.

Figure 4: It would be useful to see the seasonal precipitation responses here, especially with the authors' focus on the monsoon regions.

Response. Thank you, this is a valuable suggestion. Figure 4 has been extended (please refer to Figure R3 in this letter) with the MJJAS and NDJFM precipitation responses (panels d and e) and a bar chart of the area-weighted precipitation change over each monsoon region in both seasons (panel f). The monsoon regions are outlined on all precipitation maps. The annual panels keep their labels (a)–(c). The seasonal view shows that the boreal-summer drying is concentrated along the Atlantic ITCZ, over the West African Sahel and the Caribbean, while the box-mean MJJAS precipitation decreases significantly only in ISM (−4 %). In austral summer, NE-SAM (+7 %) and SAfM (+5 %) become significantly wetter. We added this description to Section 3.2 and corrected an overstatement in the Discussion, which had described a drying across the entire Northern Hemisphere monsoon belt.

> *"The seasonal maps reveal that the drying is concentrated in boreal summer (Fig. 4d), along the Atlantic ITCZ, over the West African Sahel and over the Caribbean, while compensating moistening appears south of the equator over the Atlantic (Dixit and Kumari, 2026). In austral summer (Fig. 4e), precipitation increases over northeastern South America and southern Africa. Averaged over the monsoon*

regions, the summer-season response is modest (Fig. 4f). MJJAS precipitation decreases significantly only in ISM (−4 %), whereas EAM and CAM change by only −1 and −2 %. NDJFM precipitation increases significantly in NE-SAM (+7 %) and SAfM (+5 %) and changes little in IAM and SW-SAM."

Figure 6: To assist the reader, mention in the caption the difference between the solid and dashed lines of best fit.

Response. Done (see Figure R1). The caption now reads:

"Colored lines show the least-squares fit of the partial residuals on the predictor within each region. Solid lines mark regions where this within-region slope is significant at $p < 0.05$, and dashed lines mark regions where it is not significant ($p \geq 0.05$)."

Figure 7: "Observed" as in proxy records? Or is this also simulated data, but not MLR-predicted? It looks like it is all model data.

Response. The reviewer is right, all data in Fig. 7 are model output. The label "Observed" referred to the simulated δ^{18}O, as opposed to the MLR-predicted values. We have replaced it by "Simulated" in the figure legend (Figure R11), the caption and the text, and likewise in the axis label of Fig. 6f.

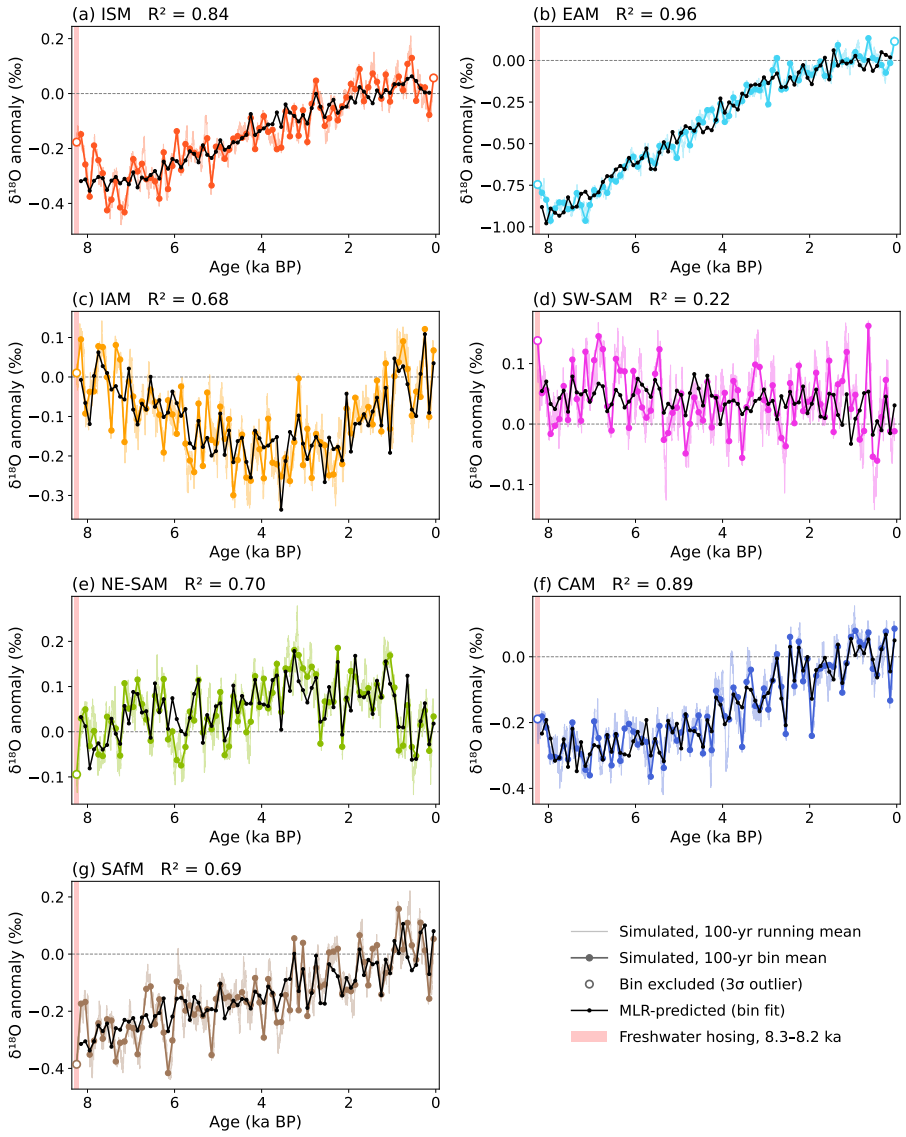

Figure R11. Revised Fig. 7. Simulated (colored) and MLR-predicted (black) δ^{18}O for each region. Markers show the 100-yr bins used in the regression, and the regression line and R^2 now come from the same fit.

Figure 10: I don't see the point of restricting the figure to the monsoon regions in 10a–b. But I understand why it is necessary in 10c to keep things restricted to the regions of interest. The authors should consider replotting the data for the whole tropics in a–b, and still have the monsoon region boxes overlaid.

Response. We agree and have redrawn Fig. 10 following this suggestion (Figure R12). Panel (a) now shows the grid-cell correlation over the whole tropical band (45°S–55°N), for land and ocean, with the monsoon regions overlaid. Inside the regions the local monsoon season is used, and outside them MJJAS is used north of the equator and NDJFM south of it. For panel (b), the regional precipitation is only defined for each monsoon region, so we extended the map to a 15° halo around each region. Where the halos of neighboring regions overlap, each grid cell is assigned to the nearest region. This shows how the correlation with the regional rainfall decays outside the region. Panel (c) remains restricted to the regions. A blue-to-red colorbar is now used because positive correlations appear outside the regions. The text of Section 3.4.3 was revised accordingly.

> *"At the grid-cell level, the correlation between $\delta^{18}O$ and precipitation is negative over almost the entire tropical band (Fig. 10a). Significant negative correlations cover 98 % of the tropical land cells between 40°S and 40°N, but their magnitude is patchy. The strongest correlations occur over the tropical oceans along the ITCZ, whereas weak correlations appear over the eastern equatorial Pacific and over parts of the ISM and CAM domains. Averaged over the monsoon regions, the grid-cell correlation ranges from* -0.36 *in ISM and CAM to* -0.58 *in NE-SAM. Replacing local precipitation with the regional precipitation of each monsoon region does not strengthen the relationship (Fig. 10b). The region-averaged correlation changes little and ranges from* -0.28 *in SW-SAM to* -0.57 *in NE-SAM. Outside the regions the correlation decays rapidly and even turns positive, for example over the eastern Pacific ITCZ west of CAM and over the western Pacific east of IAM. In stark contrast, regional $\delta^{18}O$ and regional precipitation correlate much more strongly, with r between* -0.51 *in SW-SAM and* -0.81 *in NE-SAM (Fig. 10c)."*

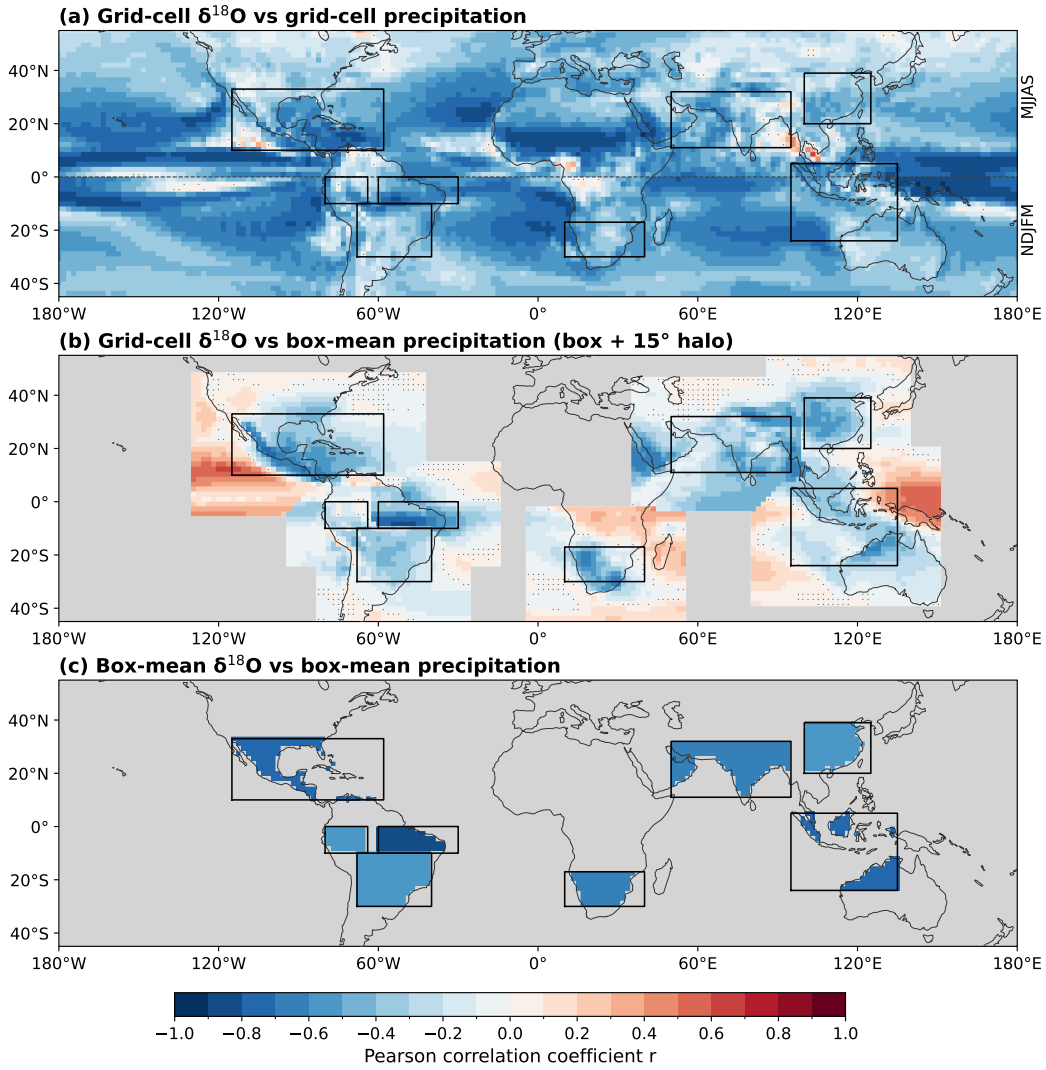

Figure R12. Revised Fig. 10. Panel (a) now covers the whole tropical band and panel (b) a 15° halo around each monsoon region, with the regions overlaid. Panel (c) remains restricted to the regions.

Referee 2 (RC2)

General

The manuscript investigates a transient simulation of the Holocene based on the isotope enabled Earth System Model AWI-ESM2-wiso for isotopic variations of global monsoon regions. To investigate the importance of different meteorological variables and indices authors setup multiple linear regression models to predict d18o for the various monsoon regions. Results indicate that for some regions especially precipitation is important concerning the amount effect, whereas other regions indicate less clear influence of specific variables. Authors also point to the importance of local effects included in speleothems, the main basis for comparisons.

The manuscript is well written and the data and methods are described to a degree that the reader has necessary background information to follow the line of arguments. Also the model system authors use is state-of-the-art and the experiment setup is carefully designed and implemented. The model shows an overall good performance for model validation against observations and in representing general Holocene near-surface temperature trends for specific seasons and latitudinal bands.

As such, results should be published. Below I listed some points that should be addressed in the revised version, especially related to the setup and interpretation of the MLR models.

Response. Thank you very much for your positive and insightful comments, below please see our point by point responses.

Abstract. The abstract is well written and comprehensively summarizes the main results of the manuscript. Authors might however also include some critical remarks on their conclusions concerning the MLR uncertainties, uncertainties involved in the proxies and maybe also the limits of interpreting results in the overall context on hydrological changes of global monsoons on Holocene time scales.

Response. We agree. Two sentences on these limitations now is added to the Abstract, and the statement on the role of precipitation has been qualified in view of the collinearity analysis (see our response to the comment on l. 217). Now the updated abstract is:

> *"Speleothem oxygen isotopes provide key insights into Holocene monsoon variability, but the scale-dependent and regionally variable controls on precipitation $\delta^{18}O$ remain poorly resolved. Here, a continuous 8.3 ka isotope-enabled transient simulation performed with AWI-ESM2-wiso is used to investigate Holocene climatic and isotopic changes, and the drivers of the isotopic variations are quantified with a four-predictor multiple linear regression (MLR) model. The simulation successfully reproduces the orbitally driven long-term weakening of Northern Hemisphere monsoons and the bipolar hydroclimatic response to the 8.2 ka freshwater perturbation, characterized by a southward Intertropical Convergence Zone (ITCZ) shift, Northern Hemisphere drying and isotopic enrichment, and Southern Hemisphere moistening and depletion. The simulated speleothem-site $\delta^{18}O$ follows the long-term trends of regional speleothem composites, although, as in the independent iTRACE simulation, the amplitude of these trends is underestimated. MLR results show that regional precipitation amount is the dominant and temporally stable control on monsoon $\delta^{18}O$ across regions. It alone explains more than 80 % of the centennial variance in the Indian, East Asian and Central American monsoons, whereas in the South American and southern African monsoons it acts only together with source temperature and recycling, which share the orbital trend. The secondary predictors exert regionally divergent influences, and their apparent changes through time mostly reflect sampling uncertainty. Supplementary large-scale climate indices improve model performance for specific monsoon domains. The El Niño–Southern Oscillation (ENSO) is the most widespread teleconnection, while the Atlantic Meridional Overturning Circulation (AMOC) acts as an additional control over the southwestern South American Monsoon, where the forced signal is weak compared with internal variability. Spatial correlation analyses further demonstrate that the precipitation amount effect is weak and spatially heterogeneous at local grid scales but becomes robust and uniform at the regional scale. This further indicates that individual speleothem records may misrepresent large-scale monsoon signals due to local spatial noise, while regional proxy ensembles effectively capture the dominant precipitation–isotope relationship. These conclusions rest on a single simulation with one model, and the collinearity of the local predictors limits the physical interpretation of the individual regression coefficients. The comparison with speleothems also neglects soil, karst and cave processes and is therefore robust mainly on centennial to orbital time scales."*

1 Introduction. l 18 ff: Authors start the intro with a very specific and technical section. I suggest to start-off with a more general paragraph on the monsoon and how it changed throughout the Holocene, based on previous modeling and proxy studies.

Response. Thank you for this suggestion. The Introduction now opens with a general paragraph on the Holocene evolution of the monsoons as seen in proxy records and simulations.

> *"Monsoon systems dominate the rainfall of large parts of Asia, Africa, Australia and the Americas, and their variability has been studied on time scales from years to millions of years (Wang et al., 2014; An et al., 2015). The Northern Hemisphere summer monsoons were stronger in the early and middle Holocene than today and weakened gradually toward the late Holocene, following the precession-driven decline of boreal summer insolation (Kutzbach, 1981; Wanner et al., 2008; Wang et al., 2014). Speleothem records from South America indicate an opposite evolution of the Southern Hemisphere monsoons, which strengthened as austral summer insolation increased (Cruz et al., 2005; Cheng et*

al., 2012). Insolation is the primary driver of these changes, while the distribution of land and sea, topography and ocean circulation shape their regional expression (Wang et al., 2017). Mid-Holocene simulations reproduce stronger Northern Hemisphere and weaker Southern Hemisphere monsoons, with ocean feedbacks amplifying the former and damping the latter (Zhao and Harrison, 2012), and vegetation feedbacks further enhance the African monsoon (Kutzbach et al., 1996). Most models, however, underestimate the magnitude of the reconstructed changes, in particular the northward extension of the North African monsoon (Zhao and Harrison, 2012; Brierley et al., 2020). Much of this knowledge comes from equilibrium time-slice experiments. Transient simulations follow the continuous climate evolution, for example TraCE-21ka since the Last Glacial Maximum (Liu et al., 2009) and MPI-ESM simulations of the deglaciation and the Holocene (Mikolajewicz et al., 2025; Dallmeyer et al., 2020; Bader et al., 2020). Isotope-enabled transient simulations, which also resolve the imprint of the monsoons in proxy archives, remain rare (He et al., 2021; Bao et al., 2025a)."

A second moment that might also be appreciated relates to the forward modeling of e.g. speleothems, taking into account proxy forward models (e.g. Dee et al. 2015; Hu et al., 2021) Authors could also use this framework as motivation and pre-requisite for their simulation activities using isotope-enabled simulations.

- Dee, S., Emile-Geay, J., Evans, M.N. et al. (2015): PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, J. Adv. Model. Earth Syst., 7, 1220–1247, doi:10.1002/2015MS000447.

- Hu, J., Dee, S. G., Corinne I. Wong, C.I. et al. (2021): Assessing proxy system models of cave drip-water δ^{18}O variability, Quaternary Science Reviews, 254, https://doi.org/10.1016/j.quascirev.2021.106799

Response. We agree that proxy system modeling provides a natural motivation for isotope-enabled transient simulations. We have added this argument to the Introduction and cite both suggested studies together with *Evans et al. (2013)*. Proxy system models are also discussed in Sections 3.1 and 4.7.

"The comparison with proxy records can be sharpened further by proxy system models, which describe how an archive transforms the isotopic signal of precipitation (Evans et al., 2013). For speleothems, such forward models represent the transit of water through soil and karst and the formation of calcite (Dee et al., 2015; Hu et al., 2021). They require isotope-enabled simulations as input, which provides a further motivation for transient simulations that resolve the isotopic water cycle continuously."

2 Methods. 2.1 Model description. The authors should also state here some reference simulations and maybe add a section on the general performance of the model system when compared to observations, specifically for the isotope enabled output for regions where observations are available. This would be helpful to assess the quality of the model and maybe also would give rise to explain some limitations of the modeling approach, especially when interpreting results on Holocene time scales. In addition, a hint to the model validation chapter in 3.1 can also help to address this point.

Response. We have added a paragraph to Section 2.1 that summarizes the earlier evaluations of the model, including the PI and mid-Holocene reference simulations of *Shi et al. (2023)*, states the main limitation for Holocene applications, namely the underestimated amplitude of the simulated isotopic changes, and points to Section 3.1. In addition, Section 3.1 now quantifies the biases of the PI climate of the transient simulation against a reanalysis (see our response to the comment on Fig. 1).

"The coupled model without isotopes was evaluated by Sidorenko et al. (2019), who found its pre-industrial and historical climate to be within the range of other climate models. The isotope-enabled configuration used here was evaluated by Shi et al. (2023) with PI and mid-Holocene equilibrium simulations of 1500 model years each. Compared with GNIP station data, ice cores and speleothems, the simulated PI precipitation $\delta^{18}O$ has a root-mean-square error of 2.2‰, and the spatial $\delta^{18}O$–temperature slope of 0.71‰ per °C is close to the observed 0.69‰ per °C. The simulated mid-Holocene

changes, however, range only from −1.3 to +0.4 ‰, whereas the speleothem changes range from −3.4 to +2 ‰. An underestimated amplitude of the Holocene isotopic changes is common to isotope-enabled models (Comas-Bru et al., 2019; Cauquoin et al., 2019) and has to be kept in mind when the simulated changes are compared with proxy records. The PI mean state of the transient simulation is assessed in Sect. 3.1."

2.2 Experimental design. Knowing that isotope enabled simulations are very expensive in terms of computing resources, I was wondering if authors could add a section on carrying out an ensemble of simulations and how this could influence conclusions with respect to analyzing only a single simulation based on a single model system.

Response. This is an important point. Section 2.2 now states that the simulation is a single realization and refers to the new Section 4.7, whose first paragraph discusses what an ensemble would change. Using the noise estimates of Fig. 11, we show that an ensemble would change the results little for EAM and ISM, where internal noise contributes less than 15 % of the bin variance, but would substantially raise the attainable R^2 in IAM and NE-SAM. The conclusions on the secondary drivers and the climate indices in these regions are therefore the most affected by the use of a single realization. The paragraph also addresses the single realization of the 8.2 ka event and the model dependence revealed by the comparison with iTRACE.

"The results rest on a single realization of one model. Internal variability is therefore represented by one sample, and its effect on the regressions can be estimated only statistically (Fig. 11). In the mean of an ensemble the noise would decrease with the number of members, while the forced signal would be unchanged. For EAM and ISM, where noise contributes less than 15 % of the bin variance, an ensemble would change the results little. For IAM and NE-SAM, where it contributes 50–75 %, even a small ensemble would raise the attainable R^2 substantially. The secondary drivers and the climate indices of these regions, and the weak forced signal of SW-SAM, would be the conclusions most affected. The 8.2 ka response is also a single realization of a 100-yr event, and ensembles of hosing experiments have been used to separate its robust isotopic signal from internal variability (LeGrande and Schmidt, 2008). Given the cost of isotope-enabled transient simulations, short ensembles for selected periods such as the 8.2 ka event, together with transient simulations of other isotope-enabled models, are a practical way to separate forced from internal and model-specific signals (Deser et al., 2012). The comparison with iTRACE (Sect. 4.3) suggests that the attribution of the East Asian signal may depend on the model."

2.3 Study regions. In addition to Table 1, authors should point to Fig. 6a where the geographical regions are depicted.

Response. Done. The text of Section 2.3 now refers to Fig. 6a, and the caption of Table 1 also points to it. The revised Fig. 6a additionally shows the moisture source regions (Figure R1).

2.4 Multi-linear regression framework. l. 115: Using wind_dir as predictor in a monsoonal context is for sure a valid idea. I think using a higher atmospheric level, e.g. winds@850hPa could better represent large-scale conditions of monsoonal systems. At least authors might give some explaining words why they chose this predictor for their subsequent MLR setup.

Response. We chose the lowest model level (1000 hPa) to remain consistent with the surface winds used by *Parker et al. (2021)*, which allows a direct comparison with their time-slice results, and because the moisture transport is largest near the surface. Following the reviewer's suggestion, we recomputed the wind direction at 850 and 700 hPa with the same method and repeated all regressions (Table R1). The regional R^2 changes by at most 0.02. In the pooled model the wind-direction term becomes more significant at 850 hPa ($p = 0.0001$ instead of 0.02) and its partial R^2 increases from 0.01 to 0.03, while the precipitation coefficient remains negative and significant in all regions. The conclusions are therefore not sensitive to the level. In the index selection, three gains that lie close to the threshold of 0.02 cross it, which we now report in Section 4.6. We have added a short statement of the result in the Section 2.4.2.

"The lowest level is used for consistency with the surface winds of Parker et al. (2021), and because specific humidity, and with it the moisture transport, is largest near the surface. The four-predictor regressions change little when the wind direction at 850 or 700 hPa is used instead (not shown)."

Table R1. Sensitivity of the regressions to the level of the wind-direction predictor. R^2 of the regional and pooled four-predictor models with the wind direction at 1000 hPa (as in the manuscript), 850 hPa and 700 hPa, the p value and partial R^2 of the wind-direction term at 1000 and 850 hPa, and the climate indices selected for Fig. 9 at 1000 and 850 hPa.

	R^2			p (wind)		partial R^2 (wind)		Selected indices	
Region	1000	850	700	1000	850	1000	850	1000 hPa	850 hPa
ISM	0.84	0.84	0.85	0.30	0.09	0.014	0.037	none	none
EAM	0.96	0.96	0.96	0.94	0.005	0.000	0.101	Niño 3.4	none
IAM	0.68	0.68	0.68	0.44	0.25	0.008	0.017	SAM, AMOC	SAM, AMOC
SW-SAM	0.22	0.20	0.21	0.22	0.63	0.019	0.003	Niño 3.4, AMOC	Niño 3.4, AMOC
NE-SAM	0.70	0.72	0.69	0.08	0.007	0.039	0.092	SACZ	none
CAM	0.89	0.89	0.89	<0.001	<0.001	0.369	0.368	none	Niño 3.4
SAfM	0.69	0.70	0.70	0.91	0.12	0.000	0.031	Niño 3.4	Niño 3.4
Pooled	0.79	0.79	0.79	0.02	<0.001	0.010	0.026	–	–

l. 125: why did the authors proceed with binned 100 year mean values, rather than inter-annual values ?

Response. The 100-yr bins were chosen for three reasons. They match the 100-yr averaging of the time slices of *Parker et al. (2021)*, with whose results we compare. They correspond to the time scales on which speleothems and models can be compared robustly, given the dating uncertainty and the smoothing in the karst (see our response to the comment on l. 169). Moreover, the interannual variability is largely unforced and dominates the annual values. To show the effect of this choice, we fitted the regressions to annual values and to averages of 10 to 500 yr (new Fig. S3, Figure R13). The pooled R^2 increases from 0.46 for annual values to 0.85 for 100-yr bins, and the regional R^2 increases with averaging length in all regions except SW-SAM. The precipitation coefficient, however, changes little, with −0.71 and −0.75 ‰ per mm day^{-1} for annual values and 100-yr bins in the pooled model, and it is significant in all regions up to 100-yr averaging. The amount effect identified with the bins is therefore present at interannual time scales as well. We added this justification to Sections 2.4.3.

"The bin length equals the 100-yr averaging period of the time slices of Parker et al. (2021), and it suits the comparison with speleothem records, which resolve centennial to orbital variations more robustly than interannual ones (Sect. 3.1). Averaging also suppresses the interannual variability, which is largely unforced. Regressions on annual values and on bins of 10 to 500 yr are compared in Fig. S3 in the Supplement."

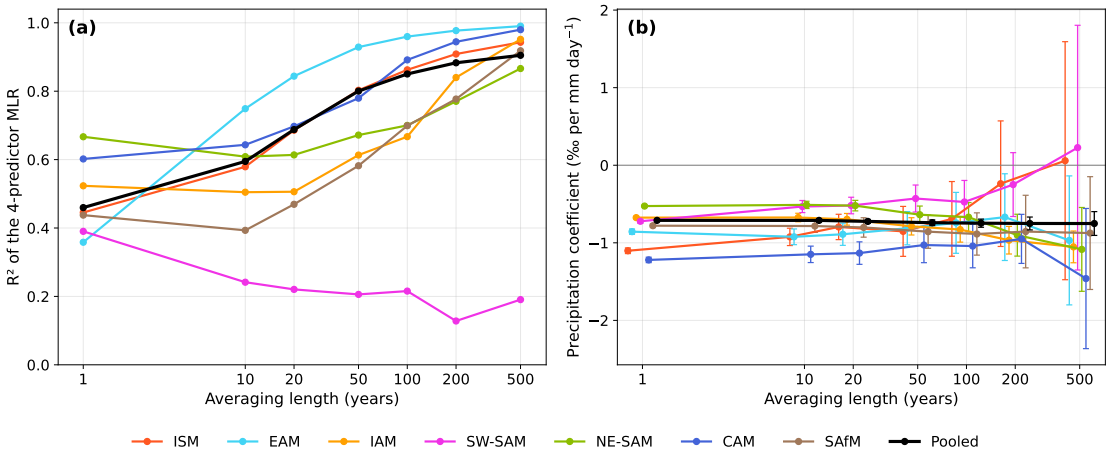

Figure R13. New Fig. S3. (a) R^2 of the regional (colored) and pooled (black) four-predictor regressions and (b) the precipitation coefficient with 95 % confidence intervals as a function of the averaging length.

ll 128 ff: It seems that the MLR uses the same set of predictors, irrespective of the target monsoon region. How about a pre-selection ? In addition, some of the indices might show a high degree of co-linearity. The authors try to address this by limiting the number of additional indices to three. However, also those three indices can show a high amount of co-linearity, weakening the robustness of the according MLR. A cross-correlation matrix among the different predictors prior to the setup of the MLR could help to elucidate which predictors are highly correlated and might not be used at the same time for a specific target region. For example, the AO and SAM will most likely show a high correlation with wind regimes over the Northern and Southern Hemisphere, respectively, and also with other initial predictors related to precipitation and moisture flux.

Response. Thank you, we followed this suggestion. The correlation matrices of all predictors and indices are shown for each region and for the pooled table in a new Fig. S6 (see also Figure R14 below), together with the largest variance inflation factor (VIF) of each four-predictor model.

As the reviewer anticipated, several indices are highly correlated with each other on the 100-yr bins. Niño 3.4 and the SACZ index correlate at 0.93, and the AMOC index correlates with both at -0.85 and -0.83. These correlations arise almost entirely from the common Holocene trend, because Niño 3.4, SACZ and AMOC correlate with age at -0.99, -0.93 and $+0.86$, whereas their correlations after removing a linear trend do not exceed 0.13 in magnitude. The SAM index correlates moderately with the source-region wind direction ($|r|$ about 0.5 in EAM, SW-SAM and NE-SAM), while the AO index is almost uncorrelated with all other variables ($|r| \leq 0.3$). The forward selection rarely combines collinear indices. Two indices are selected only in SW-SAM (Niño 3.4 and AMOC) and IAM (SAM and AMOC, $r = 0.41$), and in both cases the second index adds less than 0.05. We tested the selected models for overfitting with cross-validation and red-noise surrogates, as described in our response to the comment on Section 3.4.2.

Regarding a pre-selection of the local predictors, we kept the same four predictors in all regions by design. This keeps the regional coefficients comparable with each other and with *Parker et al. (2021)*. Moreover, the correlation matrices show that a pre-selection based on collinearity would remove different predictors in different regions (for example recycling in EAM and CAM, where it correlates with precipitation at $|r| = 0.97$–0.98, and source temperature in SW-SAM, where $r = 0.98$). Instead, we now report a decomposition of R^2 that is robust to collinearity and a regression on precipitation alone (Table S2), and we discuss the consequences for the interpretation of the coefficients (Sections 3.4.1 and 4.7, see our response to the comment on l. 217). The forward selection of the climate indices is itself a region-specific pre-selection, which adds an index only when it raises R^2 by at least 0.02.

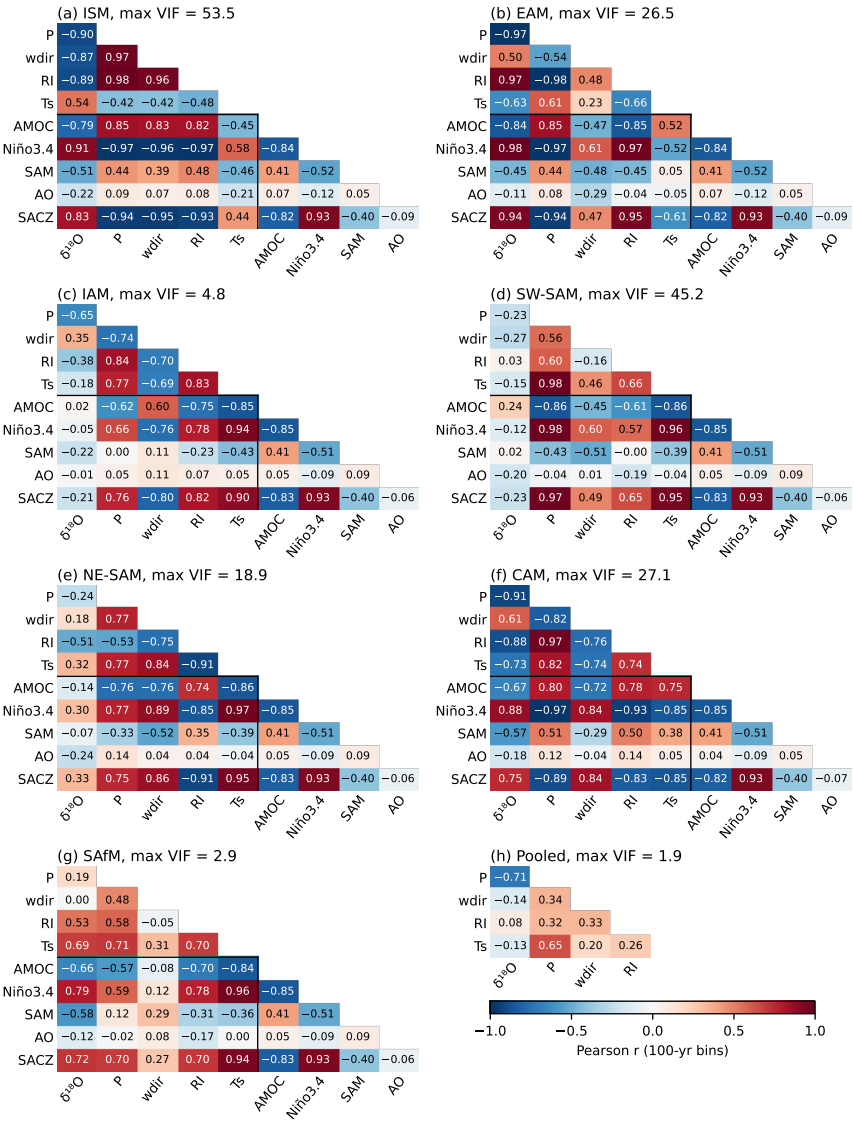

Figure R14. New Fig. S6. Pearson correlation coefficients between δ^{18}O, the four local predictors and the five climate indices on the 100-yr bins for each region, and between δ^{18}O and the four predictors for the pooled table. The panel titles give the largest variance inflation factor of the four-predictor model.

3 Results. 3.1 Model validation. l.169 Here authors should mention complications when comparing d18O from speleothems with d18O from precipitation in the context of forward modeling. For instance, which kind of factors control the amount of d18O that is finally recorded into the speleothem compared to the one that is included in the d18O of precipitation ?. (e.g. due to soil and other depletion/enrichment processes). Are the processes similar or different for the various regions and proxy archives in terms of speleothems. In addition, also some words concerning the time scale of robustly comparing results between isotope enabled models and speleothems would be helpful (i.e. the skillful time scale).

Response. We agree and have added a paragraph to Section 3.1. It describes the soil, karst and cave processes that separate speleothem δ^{18}O from precipitation δ^{18}O, explains how their importance differs between climates on the basis of the global drip-water compilation of *Baker et al. (2019)* and the proxy system model comparison of *Hu et al. (2021)*, and discusses the time scales on which models and speleothems can be compared. The implications are also discussed in the new Section 4.7.

"Such a comparison involves processes that the model does not simulate. Speleothem calcite does not

record precipitation $\delta^{18}O$ directly. Rainwater is modified by evaporation in the soil and the upper karst, by the seasonal selection of the water that infiltrates, and by mixing of water of different ages in the karst, and the calcite–water fractionation depends on cave temperature (Lachniet, 2009; Wong and Breecker, 2015). A global compilation of cave drip waters shows that drip-water $\delta^{18}O$ equals the amount-weighted precipitation $\delta^{18}O$ at sites with a mean annual temperature below 10°C. At warmer sites with seasonal rainfall it is weighted toward the recharge season, and at warm and dry sites it can be enriched by evaporation (Baker et al., 2019). Most monsoon caves belong to the warmer group, so their records are expected to be weighted toward the wet season, which is the season used for the model values here. Evaporative enrichment is most likely in semi-arid settings, whereas in humid regions the karst mainly smooths the signal. In wet climates with short residence times a running mean of rainfall $\delta^{18}O$ already reproduces drip-water $\delta^{18}O$, whereas semi-arid caves with long residence times require more complex proxy system models (Hu et al., 2021). This smoothing and the dating uncertainty determine the time scales on which models and speleothems can be compared. In a comparison of last-millennium simulations with SISAL records, speleothem $\delta^{18}O$ was less variable than the simulations on decadal time scales and more variable on centennial time scales, and neighboring records correlated only weakly (Bühler et al., 2021). The comparison here therefore focuses on the long-term evolution. Expressing both data sets as anomalies removes constant offsets, for example those due to the mean cave temperature, but not changes of these processes through time."

3.2 The 8.2 ka event. For the 8.2 ka event the routing of rivers and the according pathway of the fresh water masses seems to be critical (c.f. also Younger Dryas event). Is the resolution of the FESOM model high enough to allow for a realistic pathway along the coast of North America (southward displacement due to Labrador current) or does the fresh water directly enter the core regions of the NADW production sites, leading to changes in the AMOC and subsequent consequences ?

Response. Thank you for raising this point. In our set-up the freshwater is added as a surface volume flux directly over a box that covers the Labrador Sea and the western Irminger Sea (70–35°W, 50–70°N). It therefore enters the deep-convection region of the Labrador Sea directly and does not follow the Hudson Strait and the Labrador Current. We diagnosed the spreading of the freshwater and the AMOC response from the model output (new Fig. S7, Figure R15). The ocean mesh has a resolution of 24–27 km in the Labrador Sea, on the Labrador shelf and in the Hudson Strait, and of 45–50 km on the shelf south of 45°N, so the Labrador Current is represented by only a few grid points and is not eddy resolving. A coastal pathway is nevertheless present, with a surface freshening of about 3 psu on the shelf at 35–41°N in the first decade, but it carries only a small fraction of the added freshwater. Most of the freshwater spreads into the subpolar gyre and then into the open North Atlantic. The Labrador Sea convection collapses (March mixed layer depth from 2370 m to less than 500 m in the first decade), and the AMOC at 26.5°N weakens from 17.9 to an 11-yr minimum of 10.4 Sv and recovers about 13 yr after the hosing ends.

As the reviewer suggests, this direct release probably overestimates the AMOC response for the given freshwater volume. In earlier AWI-ESM experiments, freshwater released over the Labrador Sea caused a major decline of the AMOC, whereas an injection along the coast of the Laurentide Ice Sheet had little effect on its strength *(Shi et al., 2020)*. High-resolution experiments showed that a flood through the Hudson Strait can remain on the shelf as a narrow coastal current and enter the subtropical gyre *(Condron and Winsor, 2011, 2012)*, although other experiments found the AMOC response to Labrador Sea forcing insensitive to resolution *(Spence et al., 2008)*. We have added a paragraph on this issue to Section 4.2, with the conclusion that the simulated response should be seen as the fingerprint of a pronounced AMOC slowdown rather than as a quantitative reconstruction of the event.

"The strength of this response depends on the pathway of the freshwater. In earlier AWI-ESM experiments, freshwater released evenly over the Labrador Sea caused a major decline of the AMOC, whereas an injection along the coast of the Laurentide Ice Sheet had little effect on its strength (Shi et al., 2020). In the present simulation the freshwater is added directly over the Labrador Sea, where it caps the deep convection. The March mixed layer in the Labrador Sea shoals from about 2400 m

before the hosing to less than 500 m during the first decade, and the AMOC at 26.5°N weakens from 17.9 to about 11 Sv (Fig. S7 in the Supplement). The ocean mesh has a resolution of about 25 km in the Labrador Sea and on the Labrador shelf and of about 50 km on the shelf south of 45°N, so the Labrador Current, which is 50–100 km wide, is resolved by only a few grid points. A coastal pathway is nevertheless present. Surface salinity on the shelf between 35 and 41°N decreases by about 3 psu during the first decade, but this route carries only a small fraction of the added freshwater. Most of the freshwater spreads into the subpolar gyre and from there into the open North Atlantic south of 45°N. An ocean model with 1/6° resolution showed that a flood released through the Hudson Strait can remain trapped on the shelf as a narrow coastal current that carries the water into the subtropical gyre, which leaves the convection regions largely untouched (Condron and Winsor, 2011, 2012). Other experiments found the AMOC response to Labrador Sea forcing to be insensitive to resolution (Spence et al., 2008), and the long duration of the event appears to require the sustained meltwater flux from the Hudson Bay ice dome (Wagner et al., 2013), which corresponds to the background flux applied here. Releasing the freshwater directly over the convection region therefore probably yields an upper estimate of the AMOC weakening for the given freshwater volume. The simulated 8.2 ka response should accordingly be seen as the hydroclimatic and isotopic fingerprint of a pronounced AMOC slowdown rather than as a quantitative reconstruction of the event."

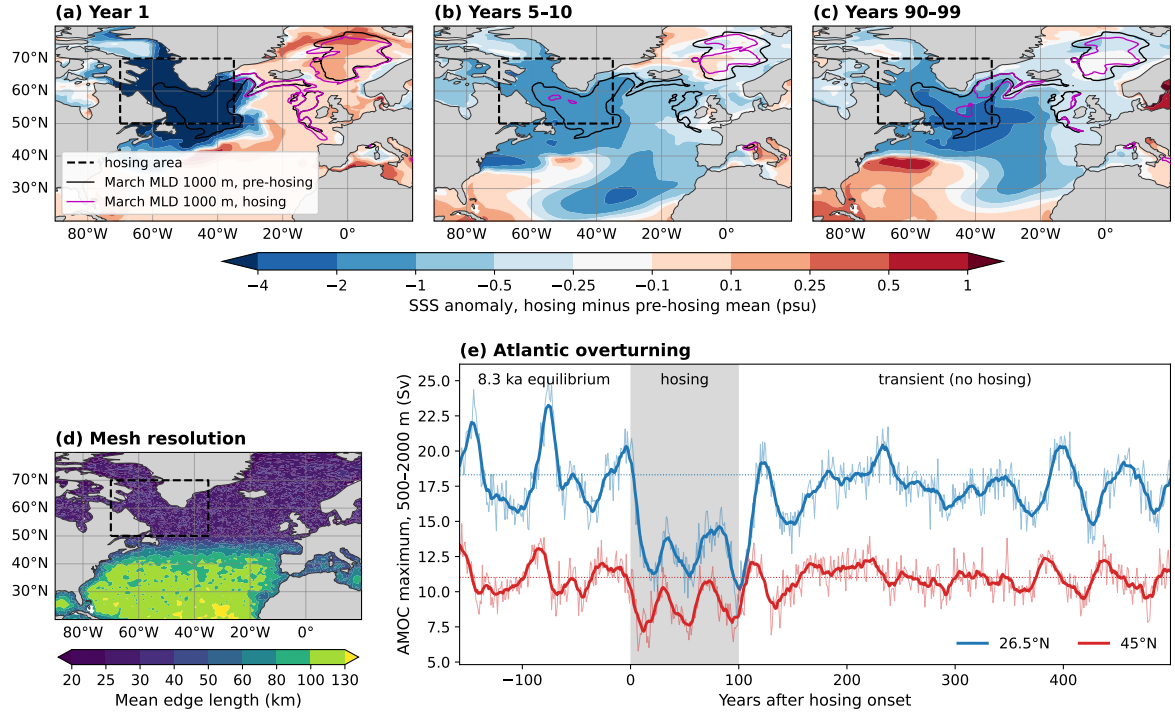

Figure R15. New Fig. S7. (a)–(c) Sea surface salinity anomaly during the 8.2 ka hosing for the first year, years 5–10 and years 90–99, with the hosing box (dashed) and the 1000 m contour of the March mixed layer before (black) and during (magenta) the hosing. (d) Resolution of the ocean mesh. (e) AMOC at 26.5°N and 45°N before, during and after the hosing.

3.3 Holocene climate trend. Are the trends linear ? Maybe the authors could also present a Figure representing trends based on different time chunks and/or running trends for the different latitudinal bands for different temporal filters.

Response. Thank you for this suggestion. The trends are not linear in all cases. We computed running linear trends of zonal-band means of surface temperature, precipitation and δ^{18}O in windows of 1, 2 and 4 kyr (the noisy 1-kyr trends are not plotted) (new Fig. S5, also in Figure R16 below). The boreal winter warming of the northern high latitudes and the associated δ^{18}O enrichment occur mainly before 4 ka, the boreal summer cooling at 30–60°N is strongest between 5 and 4 ka, and the boreal summer drying of the northern tropics weakens through the Holocene. The austral summer precipitation trends of the southern tropics and mid-latitudes change sign. It also

shows that the greenhouse gas rise of the last century affects trends over the last two millennia but hardly changes the 8.2–0 ka trends. We added a paragraph on these results to Section 3.3.

> *"These trends are not constant through the Holocene (Fig. S5 in the Supplement). For the 100-yr means of zonal-band averages, the boreal winter warming of the Northern Hemisphere occurs mainly before 4 ka. At 60–90°N its rate decreases from 0.49 °C kyr^{-1} in 8.2–6 ka to −0.09 °C kyr^{-1} in 2–0.1 ka, and the boreal winter $\delta^{18}O$ enrichment north of 30°N likewise occurs mostly before 4 ka. The boreal summer cooling at 30–60°N is strongest between 5 and 4 ka (−0.16 °C kyr^{-1}) and reverses in the last 2 kyr, and the boreal summer drying of the northern tropics weakens from −0.033 mm day^{-1} kyr^{-1} in 8.2–6 ka to nearly zero after 2 ka. The austral summer precipitation trends at 0–30°S and 30–60°S change sign during the Holocene. The linear trends in Fig. 5 are therefore averages over periods with different rates of change. The greenhouse gas rise of the last century affects the late-Holocene temperature trends but hardly changes the 8.2–0 ka trends."*

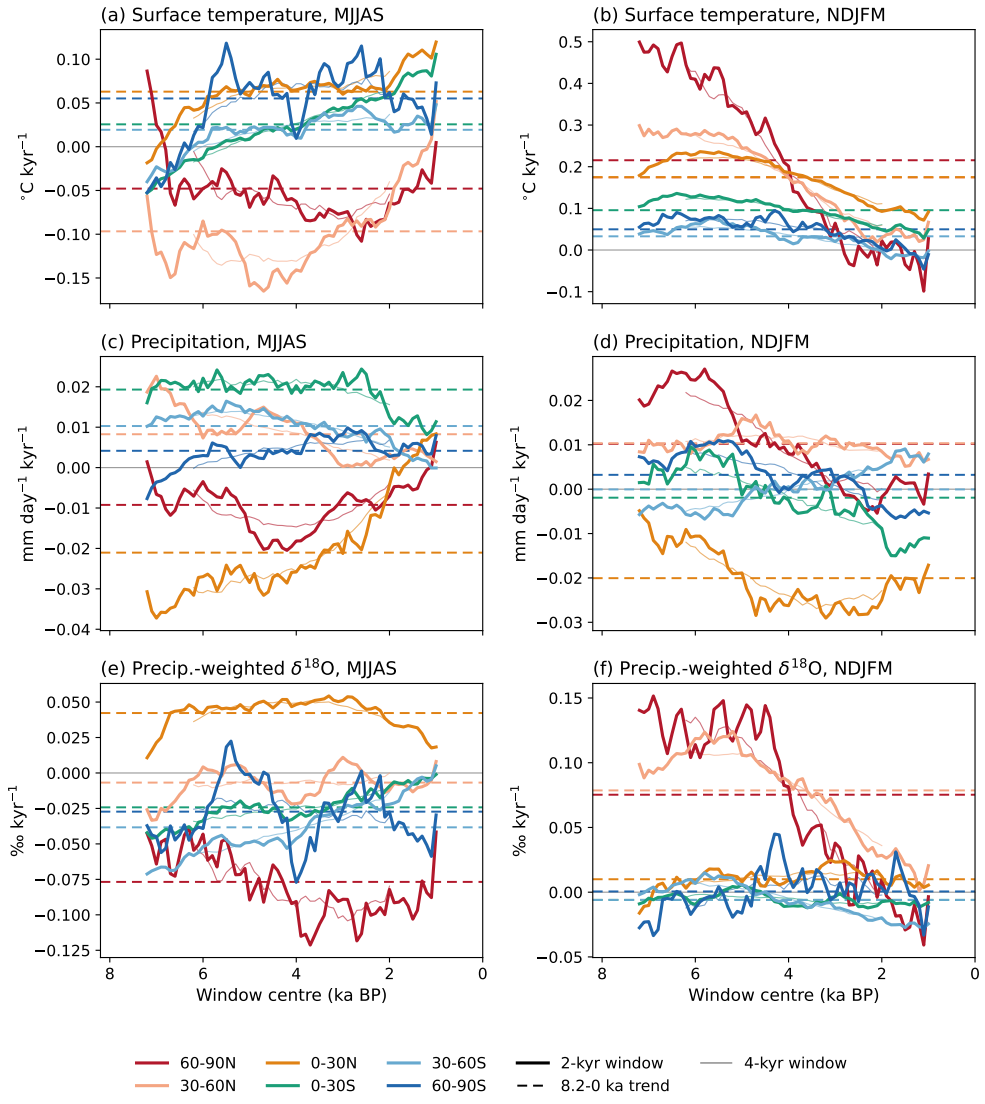

Figure R16. New Fig. S5. Running linear trends in 2-kyr windows (thick lines) and 4-kyr windows (thin lines) of zonal-mean surface temperature, precipitation and precipitation-weighted $\delta^{18}O$ for MJJAS and NDJFM, with the 8.2–0 ka linear trends (dashed).

3.4 Multiple linear regression analysis. 3.4.1 Pooled regression and the dominance of the amount effect. l. 217: Are the four predictors statistically independent? Again, a cross-correlation matrix among the predictors will help to elucidate the amount of co-linearity and might help to construct more robust MLR models for the different monsoon regions (and might also help to explain some

of the effects described in the performance and regression slopes of the different MLR setups). The (near-) independence of predictors used for the MLR is essential for a physically robust interpretation of the according regression slopes/coefficients.

Response. This is a important point. The four predictors are nearly independent in the pooled table, with pairwise correlations of at most 0.65 in magnitude and VIF below 2 (Fig. S6h). A decomposition of R^2 over all orderings of the predictors (LMG method, *Grömping, 2006*) attributes 80 % of the explained variance to precipitation, and precipitation alone explains 51 % of the pooled variance. The pooled result on the dominance of the amount effect is therefore robust.

Within individual regions, however, the predictors are strongly collinear, because they share the orbital trend (Fig. S6a–g, Table S2). The magnitude of the correlation between precipitation and recycling is 0.97–0.98 in ISM, EAM and CAM, and precipitation and source temperature correlate at 0.98 in SW-SAM, so that the largest VIF reaches 19 to 54 in five regions. In ISM, EAM and CAM precipitation alone explains 81–95 % of the bin variance, so the amount effect dominates regardless of how the remaining variance is shared. In SW-SAM, NE-SAM and SAfM precipitation alone explains less than 6 % of the variance, and its negative coefficient emerges only when source temperature and recycling are included. In SAfM the coefficient of precipitation alone is even weakly positive. In these regions the regional coefficients describe a joint fit to the shared trend rather than separable physical sensitivities. We have added these results to Section 3.4.1 and qualified the corresponding statements in the Abstract and the Conclusions. The implications for the physical interpretation are discussed in Section 4.7 (see our response to the comments on the Discussion).

> *"The interpretation of these coefficients requires the predictors to be nearly independent. In the pooled table they are, with pairwise correlations of at most 0.65 in magnitude and variance inflation factors (VIF) below 2 (Fig. S6 in the Supplement). A decomposition of R^2 that averages over all orderings of the predictors (Grömping, 2006) attributes 80 % of the explained variance to precipitation, and precipitation alone explains 51 % of the pooled variance (Table S2). Within individual regions the predictors are much more collinear, because they share the orbital trend, and the largest VIF of the regional models reaches 19 to 54 in ISM, EAM, SW-SAM, NE-SAM and CAM. Precipitation alone explains 81–95 % of the bin variance in ISM, EAM and CAM and 42 % in IAM. In SW-SAM, NE-SAM and SAfM it explains less than 6 %, and its negative coefficient emerges only when source temperature and recycling are included. The amount effect therefore dominates the pooled regression and the Northern Hemisphere regions, whereas in the Southern Hemisphere monsoons of South America and southern Africa the centennial $\delta^{18}O$ evolution is explained only by the combination of predictors. In these regions the individual regional coefficients describe a joint fit to the shared trend rather than separable physical sensitivities."*

l. 233: How should the model reproduce short-lived oscillations when the MLR was constructed based on 100 year mean values ?

Response. The reviewer is right that our wording was misleading. The "short-term oscillations" referred to the fluctuations of the SW-SAM δ^{18}O from one 100-yr bin to the next. The sentence has been rewritten and now states that these centennial fluctuations are mostly internal variability, as quantified in Section 4.6 and Fig. 11.

> *"Its simulated $\delta^{18}O$ shows little long-term trend and is dominated by centennial fluctuations of about ±0.1 ‰, which the regression reproduces only in part. These fluctuations from one 100-yr bin to the next are mostly internal variability of the simulation (Sect. 4.6)."*

3.4.2 Contribution from climate indices. The inclusion of additional predictors should be taken out with care in order to not artificially increase the performance metrics due to over-fitting because of the presence of co-linear predictors.

Response. We agree and have tested the index selection for overfitting in three ways (new Table S3, summarized in Table R2 below). The adjusted R^2 and the corrected Akaike information criterion

favor the final model in all five regions with selected indices. In a cross-validation with 8 contiguous blocks of about 1000 yr, in which the selection is repeated for each training set, the out-of-sample R^2 rises from 0.94 to 0.97 in EAM and from 0.61 to 0.69 in SAfM, and Niño 3.4 is selected in 7 and 8 of the 8 folds. In IAM and NE-SAM the out-of-sample gain is negligible. In a null test with 1000 sets of surrogate indices that have the same power spectra as the real indices, the observed gain exceeds the 95th percentile of the surrogate gains in all regions except NE-SAM, whose SACZ gain is therefore indistinguishable from that of random red-noise indices. In SW-SAM the same two indices are selected in every fold, but neither the baseline nor the final model has out-of-sample skill, because the fitted relationships fail when extrapolated to the oldest and youngest blocks.

These results are now reported in Section 3.4.2, and the summary of the index analysis has been revised accordingly. The Niño 3.4 gains in EAM and SAfM are robust, whereas the gains of AMOC, SAM and SACZ are small and less robust. The Discussion (Section 4.6) additionally notes that the SACZ gain in NE-SAM is indistinguishable from red noise and that gains close to the threshold change with the wind level (see our response to the comment on l. 115).

> *"Because the indices share the Holocene trend with each other and with the local predictors, the gains in R^2 may partly reflect overfitting. Niño 3.4, the SACZ index and the AMOC index highly correlate with each other, whereas their correlations after removing a linear trend do not exceed 0.13 (Fig. S6). The selected models are therefore tested in three ways (Table S3). The adjusted R^2 and the corrected Akaike information criterion favor the final model in all 5 regions with selected indices. In a cross-validation with 8 contiguous blocks of about 1000 yr, in which the selection is repeated for each training set, the out-of-sample R^2 rises from 0.94 to 0.97 in EAM and from 0.61 to 0.69 in SAfM, and Niño 3.4 is selected in 7 and 8 of the 8 folds. In IAM and NE-SAM the out-of-sample gain is negligible. When the 5 indices are replaced by surrogate series with the same power spectra (Theiler et al., 1992), the observed gain exceeds the 95th percentile of the surrogate gains in all regions except NE-SAM, whose SACZ gain is therefore indistinguishable from that of random red-noise indices. In SW-SAM the same two indices are selected in every fold, but neither the baseline nor the final model has out-of-sample skill, because the fitted relationships fail when they are extrapolated to the oldest and youngest blocks.*
>
> *Overall, the four baseline predictors capture most $\delta^{18}O$ variability in most monsoon regions. ENSO, represented by the Niño 3.4 index, provides a robust additional control in EAM and SAfM and a significant but not predictive one in SW-SAM, whereas the gains from the AMOC, SAM and SACZ indices are small and less robust."*

Table R2. Robustness of the climate-index selection in Fig. 9 (identical to the new Table S3). For each region the baseline model contains the four local predictors and the final model adds the indices selected by forward selection ($\Delta R^2 \geq 0.02$, at most three). $\bar{R}^2$ is the adjusted R^2. CV R^2 is the out-of-sample R^2 of leave-one-block-out cross-validation with 8 contiguous blocks of about 1000 yr, with the index selection repeated inside each training fold. The null columns give the 95th percentile of the total R^2 gain obtained when the five indices are replaced by 1000 sets of phase-randomized surrogates with the same power spectra, and the fraction of surrogate sets whose gain reaches the observed one (p, with a floor of $1/(N+1)$ for N surrogates).

		R^2		$\bar{R}^2$		CV R^2			Null test	
Region	Selected indices	base	final	base	final	base	final	ΔR^2	95th pct.	p
ISM	none	0.84	0.84	0.83	0.83	0.80	0.79	0.00	0.03	–
EAM	Niño3.4	0.96	0.98	0.96	0.98	0.94	0.97	0.02	0.00	≤0.001
IAM	SAM, AMOC	0.68	0.73	0.66	0.71	0.61	0.62	0.05	0.03	0.004
SW-SAM	Niño3.4, AMOC	0.22	0.46	0.18	0.42	−0.03	0.02	0.25	0.13	≤0.001
NE-SAM	SACZ	0.70	0.72	0.68	0.71	0.61	0.60	0.02	0.06	0.534
CAM	none	0.89	0.89	0.89	0.89	0.88	0.88	0.00	0.00	–
SAfM	Niño3.4	0.69	0.76	0.67	0.75	0.61	0.69	0.07	0.04	0.003

4 Discussion and conclusions. The discussion and conclusions are well structured, summarize the main results and also put new results in context of current literature and hypotheses. I suggest however to add a dedicated paragraph on the limitations of the model system, the difficulties

interpreting and comparing d18O based on modeled precipitation vs. d18o based on speleothem proxies.

Response. Thank you for this suggestion. We have added a new Section 4.7 "Limitations". Its first paragraph discusses the single realization and the model dependence (see our response to the comment on Section 2.2). The second paragraph addresses the limitations of the model system, namely the smoothed orography at T63 resolution, the mean-state biases quantified in Section 3.1, the dependence of the amount effect on the convection scheme, the missing solar, volcanic and land-use forcing, and the underestimated amplitude of the simulated isotopic trends. The third paragraph addresses the comparison of simulated precipitation δ^{18}O with speleothem δ^{18}O.

> *"The model itself introduces further uncertainty. At T63 resolution the orography of the Andes and of the Tibetan Plateau is smoothed, which weakens the altitude effect and modifies the upstream rainout that shapes $\delta^{18}O$ in SW-SAM and EAM. The mean-state biases described in Sect. 3.1, such as the dry bias in ISM and CAM and the wet bias in SAfM, alter the background against which the Holocene changes develop. The strength of the amount effect also depends on the representation of convection and of the re-evaporation of falling rain (Risi et al., 2008; Moore et al., 2014), which differs among models. Solar and volcanic forcing and land use are not included, so the simulated centennial variability may be lower than in reality. Moreover, the simulated amplitude of the Holocene $\delta^{18}O$ trend is smaller than in the speleothem records (Sect. 3.1)."*

In addition also some critical words on the physical interpretation of the individual predictors in the MLR based models should be mentioned, including a discussion on the effect of using co-linear predictors for the MLR models.

Response. We agree. Now the last paragraph of the new Section 4.7 addresses this point.

> *"The regression is a statistical description and does not by itself establish causal links. Its coefficients measure associations between centennial anomalies that share the orbital trend. In the pooled table the predictors are nearly independent, so the dominance of the amount effect is robust. Within individual regions the local predictors are strongly collinear (Sect. 3.4.1), so that the variance can be distributed among them in different ways with similar skill. The regional coefficients of the secondary predictors, and their signs in particular, should therefore not be treated as physical sensitivities. The same caution applies to the climate indices. The box-mean predictors also represent the underlying processes only approximately. Water tracing diagnostics, which are available in ECHAM6-wiso (Gao et al., 2024), could attribute the isotopic changes to moisture sources and transport pathways more directly."*

Figure and Tables. Fig 1: The Figures show the absolute values for the last 500 years of the simulation. As such it is difficult to see any model bias compared to observations. Maybe authors could provide difference maps using a gridded re-analysis product for the pre-industrial era (e.g Franke et al., 2017; Steiger et al., 2018).

- Franke, J., Brönnimann, S., Bhend, J. et al. (2017): A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. Sci Data 4, 170076, `https://doi.org/10.1038/sdata.2017.76`.

- Steiger, N., Smerdon, J., Cook, E. et al. (2018): A reconstruction of global hydroclimate and dynamical variables over the Common Era. Sci Data 5, 180086. `https://doi.org/10.1038/sdata.2018.86`.

Response. Thank you for this suggestion. We have added difference maps of seasonal surface temperature and precipitation between the simulated PI climate and version 3 of the Twentieth Century Reanalysis (20CRv3) for 1836–1880 (new Fig. S4, Figure R17), together with a table of the biases in each monsoon region (Table S1). We chose 20CRv3 because it covers the pre-industrial 19th century with monthly temperature and precipitation fields. The PHYDA reconstruction of *Steiger et al. (2018)* provides temperature and drought indices but no precipitation. Like 20CRv3 before 1880, the paleo-reanalysis of *Franke et al. (2017)* derives its precipitation largely from its

atmospheric model. As an independent check of land precipitation we added the gauge-based GPCC data for 1891–1920. The biases are described in Section 3.1. Seasonal surface temperature correlates with the reanalysis at 0.98–0.99, with global mean biases of −0.4 and −0.2°C, and tropical land is 1–3°C too cold. Precipitation shows a double-ITCZ structure in the tropical Pacific, a dry bias of 16–32 % in ISM and CAM against both references, and a wet bias in SAfM. In EAM, IAM and NE-SAM the two references differ from each other by about 30 %, which is as large as the model bias.

> *"Difference maps against version 3 of the Twentieth Century Reanalysis (20CRv3; Slivinski et al., 2019) for 1836–1880 quantify the biases of this PI climatology (Fig. S4 and Table S1 in the Supplement). Before 1880 this reanalysis assimilates only surface pressure observations, so its precipitation and land temperature are largely determined by its own model. The simulated seasonal surface temperature correlates with the reanalysis at $r = 0.98$ in MJJAS and 0.99 in NDJFM, with global mean biases of −0.4 and −0.2 °C. The largest temperature biases occur along the winter sea-ice edges. Tropical and subtropical land is 1–3°C too cold, most strongly over southern Africa and northern Australia, whereas the eastern-boundary upwelling regions are too warm. The pattern correlation of precipitation is 0.71 in MJJAS and 0.67 in NDJFM. The model shows a double-ITCZ structure in the tropical Pacific, too little boreal summer rainfall over Central America, northern India and eastern China, and too much austral summer rainfall over the Maritime Continent, northern Australia and southern Africa. Averaged over the monsoon regions in their own season, precipitation is 16–32 % too low in ISM and CAM compared with both 20CRv3 and the gauge-based GPCC data set for 1891–1920 (Schneider et al., 2014b, 2018), 12–15 % too high in SW-SAM, and 60–95 % too high in SAfM. In EAM, IAM and NE-SAM the two references differ from each other by about 30 %, which is as large as the model bias. These mean-state biases do not enter the anomalies analyzed in this study directly, but they can modify the simulated responses, for example where a rain belt is displaced or where the seasonal rainfall is strongly overestimated as in SAfM."*

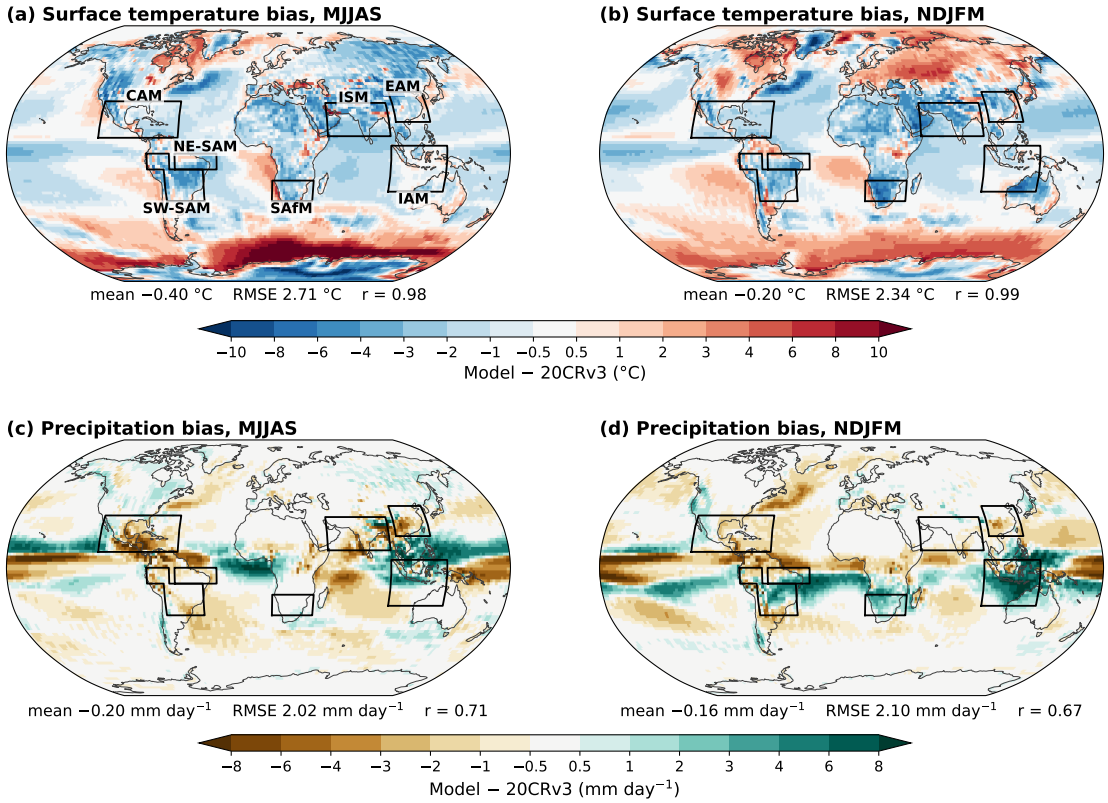

Figure R17. New Fig. S4. Simulated PI climate minus 20CRv3 (1836–1880) for (a, b) surface temperature and (c, d) precipitation in MJJAS and NDJFM, with the monsoon regions outlined.

Fig. 4: The areas showing stat. significant differences (i.e. alpha =5%) should be marked.

Response. Stippling for $p < 0.05$ had been added to Fig. 4 in the first round of revision, but is was too small and was hardly visible. The dots are now larger and drawn above the coastlines in all map panels (Figure R3).

Fig. 5: The stippling is hardly visible.

Response. We agree. The stippling of Fig. 5 has been redrawn with larger dots on every second grid cell (Figure R18). The significance test now uses an effective sample size that accounts for the lag-1 autocorrelation of the residuals *(Santer et al., 2000)*. Because the trend is significant over 86–96 % of the global area in every panel, stippling all significant cells would cover almost the entire map. The dots therefore now mark the cells where the trend is not significant, as stated in the revised caption.

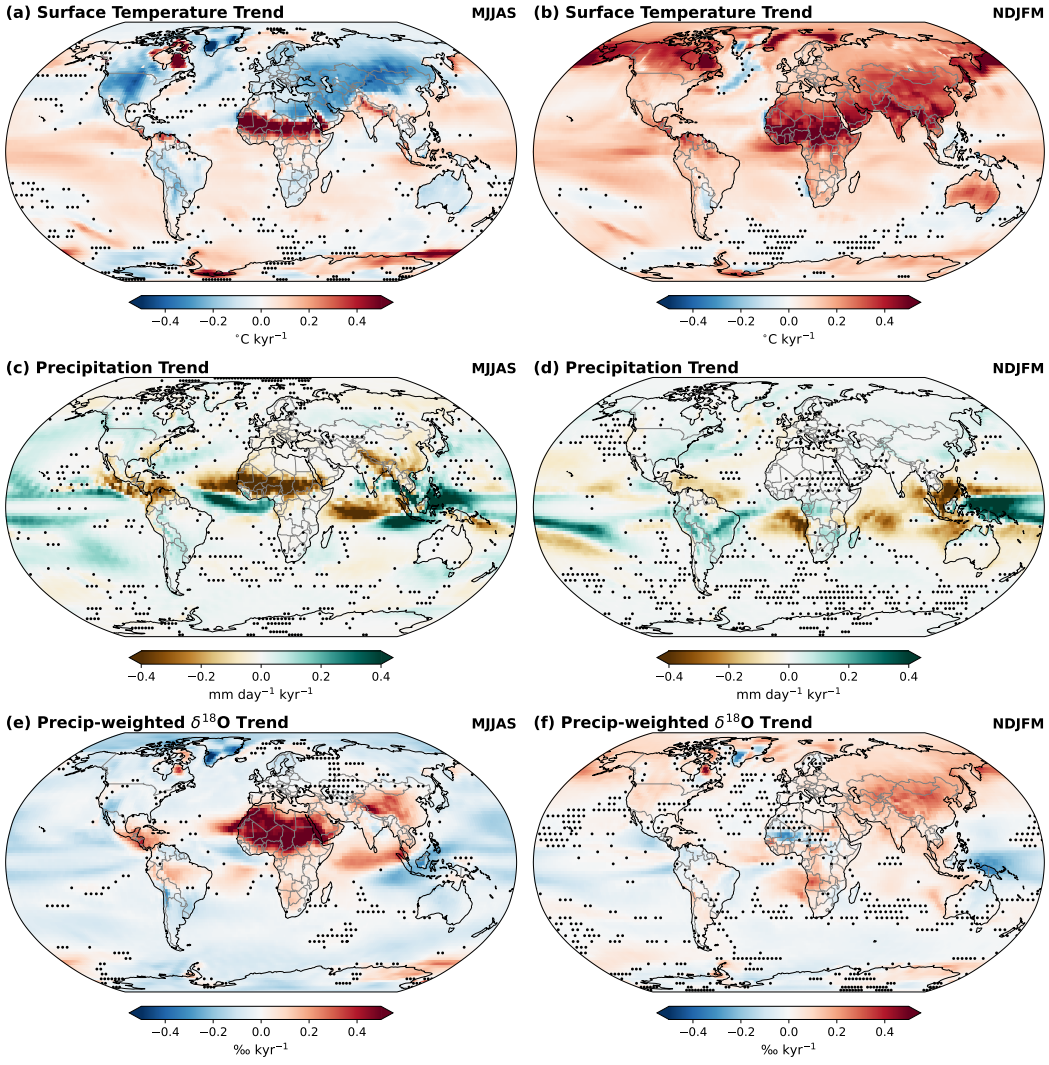

Figure R18. Revised Fig. 5. Stippling now marks grid cells where the trend is not significant at $p < 0.05$, using a t test with an effective sample size.

Fig. 7: Please include a dashed line at "0" for better reference in all subplots.

Response. Done. A dashed gray line at zero is now drawn in all panels (Figure R11), and the caption mentions it.