

Response to reviewers

We thank the reviewers for their comments. Our responses are given below (in italics) with changes to the text in blue.

We have taken the opportunity to correct some count errors in the modern pollen database (SMPDSv2), which affects 2 % of the original samples. We also corrected errors in the elevation information (SMPDSv2), which affects 0.1 % of the original samples. We have also taken the opportunity to incorporate ca 2000 additional modern samples that were inadvertently omitted from the SMPDSv2. The cleaned dataset is now referred to in the MS as SMPDSv3. The corrections only slightly affect the training statistics. The previous performance is best for MTCO ($R^2 = 0.75$, RMSEP = 6.51, slope = 0.85) but is also good for MTWA ($R^2 = 0.59$, RMSEP = 3.68, slope = 0.71) and α ($R^2 = 0.65$, RMSEP = 0.182, slope = 0.71). The corrected performance is best for MTCO ($R^2 = 0.75$, RMSEP = 6.52, slope = 0.85) but is also good for MTWA ($R^2 = 0.60$, RMSEP = 3.58, slope = 0.72) and α ($R^2 = 0.65$, RMSEP = 0.183, slope = 0.71).

We have made one further correction by not applying dynamic time warping to the marine sites. It is not possible to define a pollen source area reliably for these records, and in any case the age models for marine records are generally well calibrated using the oxygen isotope.

These changes did not affect the conclusions of the paper. Δ MTCO was still found to be significantly larger than Δ MTWA in the northern extratropics, and larger but not significantly larger in the southern extratropics. There was still a significant positive relationship between $\Delta\alpha$ and Δ MTWA in all regions. The largest warming still occurs in the extratropics of Eurasia, while western North America and the southern extratropics are characterised by cooling. $\Delta\alpha$ still broadly follows the changes in temperature, with increased α in regions characterised by warming, and Europe is still characterised by a mixed signal of drying and wetting.

We have updated all the figures and tables to reflect all these changes. The updated figures and tables are included (but not marked since all of them have been replaced) in the track changes version of the MS.

Reviewer 1

1. In line 333 of the track-changes version, replace "Van Meersbeeck" with "Van Meerbeeck"
We have now corrected this.

2. In Table 1, replace "Burjachs & Ramon (1994) " with "Burjachs & Julià (1994)"
We have now corrected this.

Reviewer 2

In this revised version of the manuscript, Liu and colleagues have significantly improved their manuscript. Many elements that needed to be clarified before are much better explained. The results are very remarkable and cohesive. Still, I am worried that some results were inverted just by changing the scale of the study (site-based vs gridded). How data are processed or represented impacts conclusions. Hence, I must ask the authors to show me more of their intermediate data a second time. Your results appear coherent, but your data are hyper-processed, and risks exist at every step. I want to see what is happening at each step to be able to accept your conclusions. In particular, I want to see:

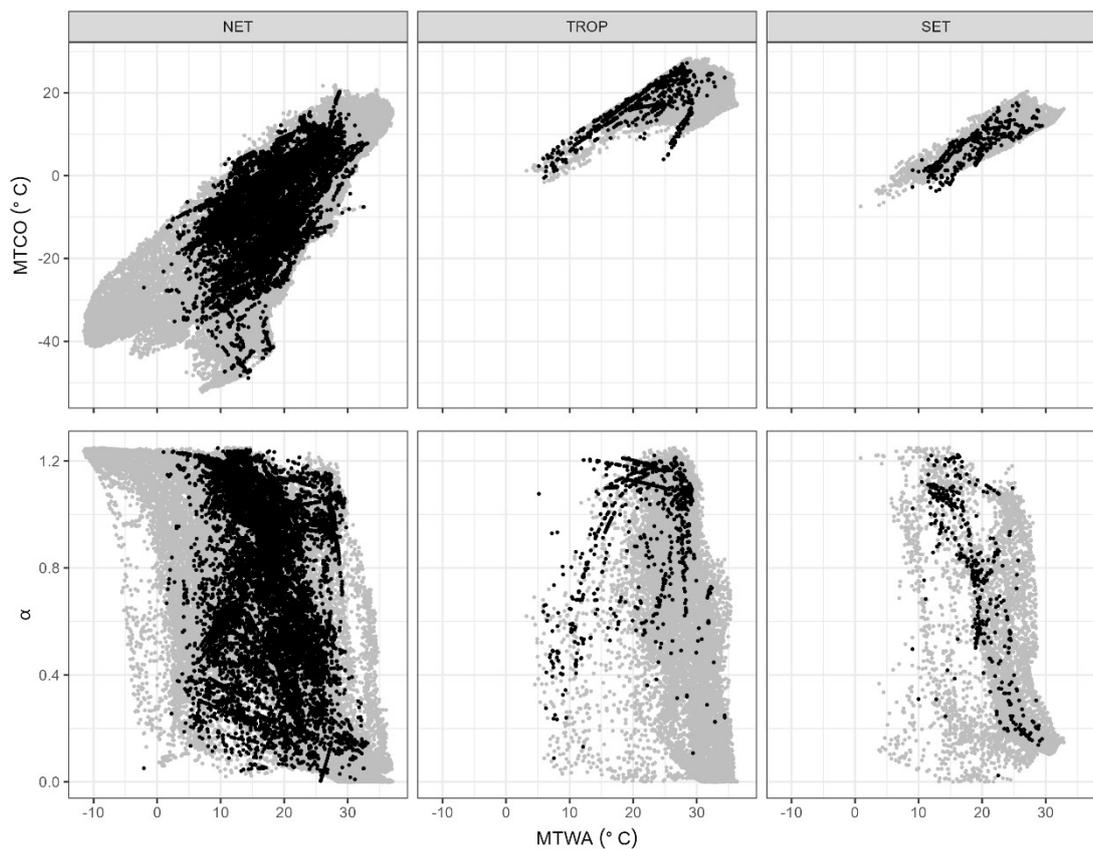
1. Where the validation errors are located and identify types of vegetation and/or climates that cannot be reasonably reconstructed with that dataset. I hardly believe you have enough calibration data to analyse African and Latin American records properly.

There are fewer sites from Africa, Latin America and Australia set than from other regions in the calibration data. However, the reasonableness of the reconstruction is determined by whether the calibration data are representative of the climate space rather than the geographic source. We have therefore checked that our training samples cover a reasonable range of the global climate space and therefore will provide robust reconstructions. We have now made this clear in the Methods (end of the 3rd paragraph in section 2.1):

The climate space occupied by SMPDSv3 (Figure S1) samples a reasonable range of global climate space and therefore should provide robust reconstructions of climate changes under glacial conditions.

We have included a diagram illustrating the climate space of our samples compared to the CRU data set in Supplementary.

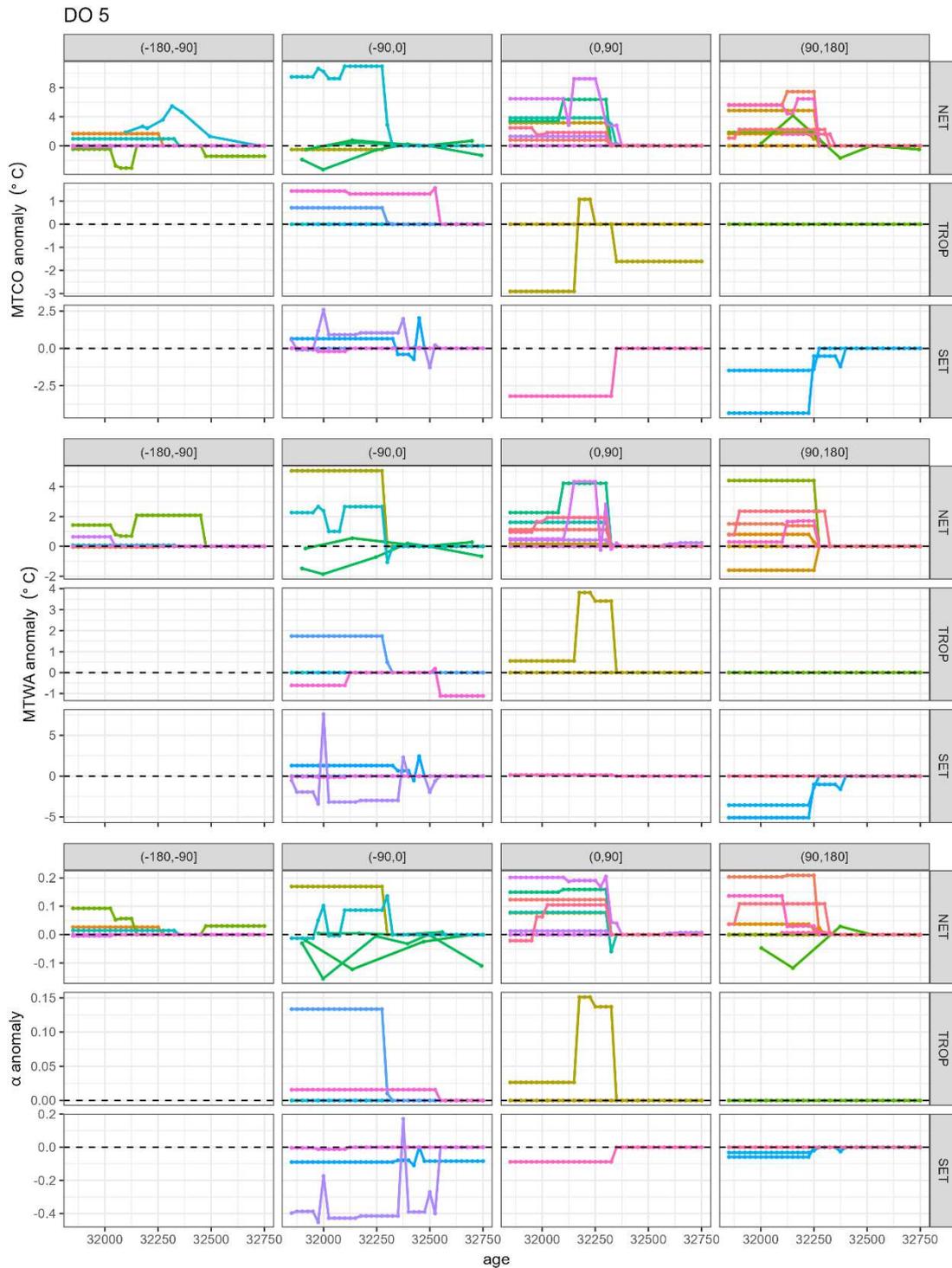
Figure. S1 : Climate space occupied by all SMPDSv3 sites (shown in black) and by all CRU TS4.04 (Climatic Research Unit Time-Series version 4.04) grids (shown in grey) in northern extratropics, tropics and southern extratropics.



2. (some of) the reconstructions. Calculating peak-to-peak distance between DO events is ok, but what do the general trends look like?

We suppose the reviewer means peak-to-trough distance here. We have now added the time series for individual sites into Supplementary Materials (Figs. S5.1~5.8). The example below is for D-O 5 (Figure S5.1). We have also added citations to these figures in the Results (second paragraph of section 3).

Figure S5.1: D-O 5 time series for individual sites (shown in different colours), segmented into different longitude + latitude combinations, using reconstructions with original ages for marine sites, and reconstructions with dynamic time warping adjusted ages for terrestrial sites. The anomalies are anomalies to the values at identified start for each site. The horizontal dashed lines are zero anomaly lines, to show whether a site has an increasing or decreasing trend more clearly.



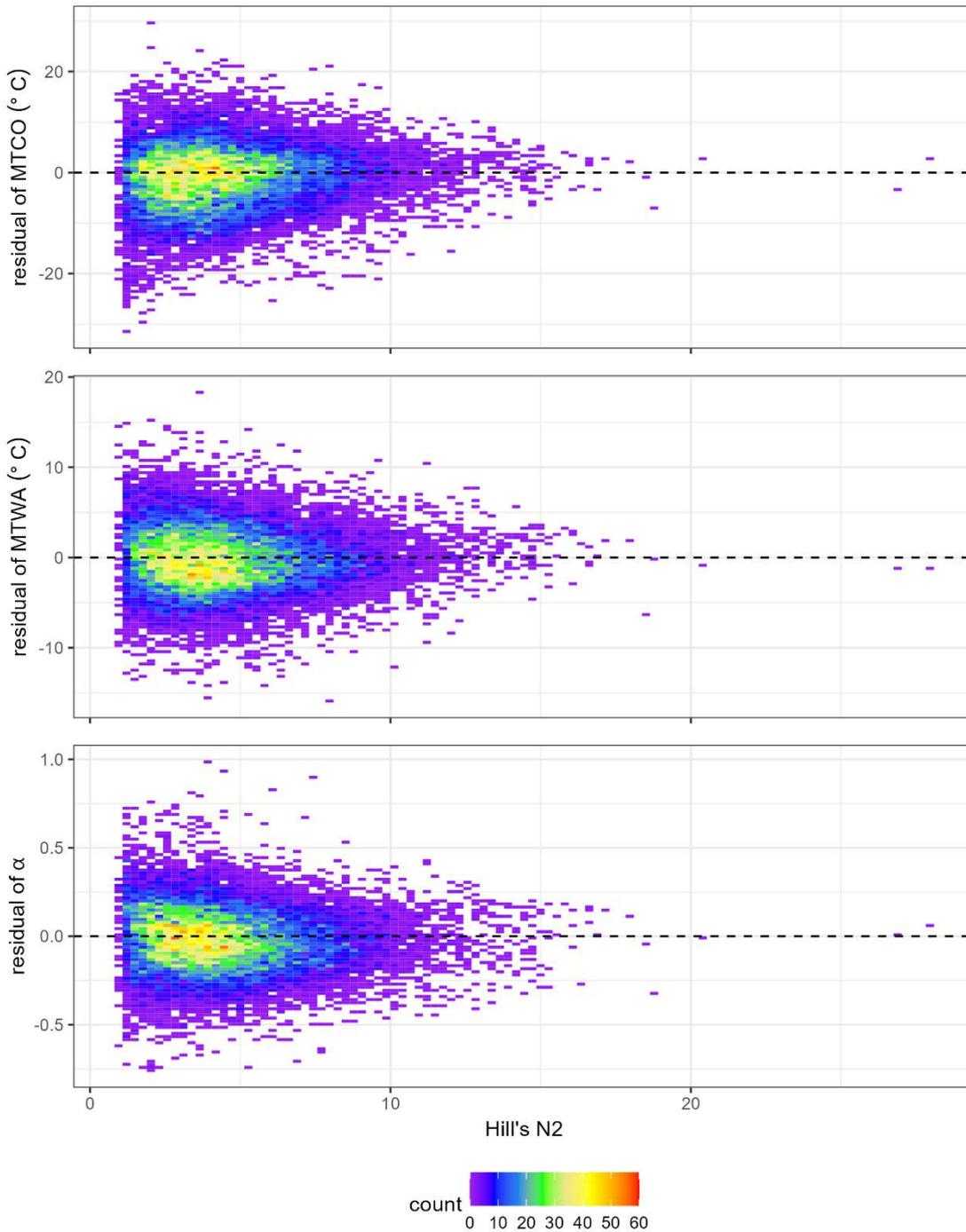
The plots of the adjusted time series have a square wave pattern because the reconstructions have fewer data points, while the reference model is much denser. This does not influence the magnitude of the identified climate change, since dynamic time warping only adjusts the age scale without affecting the climate value. However, this is why we calculated the magnitude based on trough-to-peak or peak-to-trough values, rather than providing a rate of change which will be influenced by the age scale. We have also added the following text in the Discussions (last paragraph in section 4):

The use of dynamic time warping (DTW: Belman and Kalaba, 1959; Burstyn et al., 2021; Alshehri et al., 2019) goes some way to improving the identification of potential D-O events. However, it precludes the calculation of a rate of change in climate. Thus, we have focused here on the absolute magnitude of the changes during specific warming events.

3. The RMSEP presented in Table 2 are – sorry for the term – horrible, and I am not convinced by your argument that “most” of the reconstructed differences between the warm and cold phases of the DO events are larger than this. You used the 1-sigma error, which is extremely conservative. If you used a more common 2-sigma interval, most of your reconstructions will fit within the “v-shaped” black lines. On average, the reconstructions are 6.5°C off target for MTCO. This is huge, especially for a leave-one-out cross-validation. If we propagate that error to the reconstructions and take the 2-sigma interval, your error bar will be +/- 13°C, a 26°C range. It is surely possible to do better than that. This takes me back to my original request to see the data this analysis is built on. Where are the samples with such high errors, and are they potential candidates for the D-O periods? If so, how do you deal with them in the analysis?

The large RMSEP for the modern training dataset reflects the fact that there are both sites with good representation and bad representation of taxa. All of the sites contribute to the final RMSEP. We have added a supplementary figure to show this. We can see that the higher the Hill's N2 (which means the better representation of the taxa), the smaller the residual. The RMSEP for the modern training data set cannot be interpreted as the error for the palaeo-reconstructions since this is determined by representation of the taxa in each fossil sample, with sample having a better representation giving more reliable reconstructions. This is why use sample specific errors for the D-O reconstructions (Supplementary Table 2).

Figure S4.2. Relationships between Hill's N2 and the residual of the training results using the last significant number of components, for mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA) and plant-available moisture (α). The colour shows the count of samples in a certain x-y grid. The warmer the colours, the denser the samples.



Regarding the 2-sigma interval, we have now also provided a version of Figure 5 only using those with amplitude of change twice that of the sample specific error (Figure S6). The spatial patterns are the same, indicating that our conclusions about the nature of the changes during D-O events are robust. This obviously results in a reduction of the number of points, and the choice of twice the sample specific error is conservative, so we prefer to use all the data for mapping but with the acknowledgement that users can choose the level of sample specific error that they tolerate.

Figure S6: Maps showing the median change of site-based reconstructions for Dansgaard-Oeschger (D-O) events 5~12, only using samples with $|\text{change}| \geq 2 \times |\text{sample specific error}|$ for each climate variable.

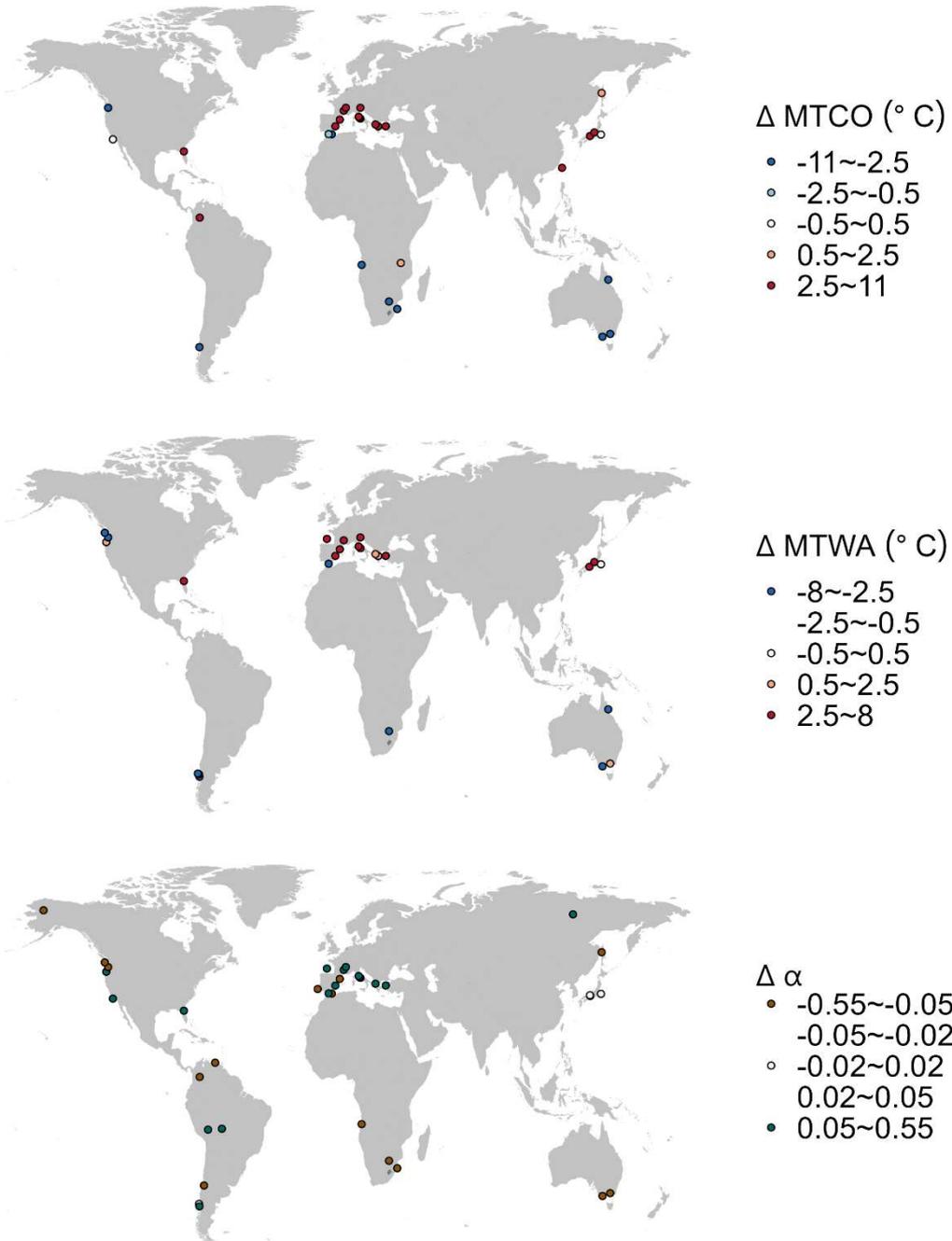


Table 3 and Table 4 are not influenced since maximum likelihood estimation of the ratio already considers the different errors between samples (both the values and the errors of the values are inputs when using this method, and both axes are considered as well).

We have also added the following text when describing the results in the revised version to include a consideration of the robustness taking sample specific errors into account (last paragraph in section 3):

The spatial patterns of ΔMTCO and ΔMTWA show consistent features across multiple D-O events (Figure 5), most noticeably that the largest warming occurs in the extratropics of Eurasia, while western North America and the southern extratropics are characterised by cooling. These patterns are also shown if only those reconstructions where the change is twice that of the sample specific error are considered (Figure S6), showing that the spatial patterns are robust to the choice of threshold.

4. L238-253: This string of arguments is very theoretical, and I don't see many convincing applications of this. You mention Turner et al. (2020). But even if the calibration dataset was extensive in that study, it was still confined to the Palearctic realm, which contains ecosystems that have co-evolved. Here, you are mixing data at an even larger scale and assuming a transitivity of the property relative to the spatial scale. I'm doubtful that mixing South American, European, and Japanese Quercus make much sense, especially in a regression model where all data are combined. As mentioned in the previous round by the other reviewer and myself, mixing the deciduous and evergreen Western European Quercus is already limiting, as these taxa bring very distinct climate information, even if not much can be done here. Finally, this is an assumption of all your results, not an element of discussion. Bring that part to the methods part.

We agree that the use of a global data set rests on the assumption of niche conservatism, and will clarify this in the Methods. However, since it is clear that niche conservatism and its consequences are not well understood, we will preserve and expand the text in the Discussion. We have added the following text to the Methods (second paragraph in section 2.1):

We have used a global pollen dataset for calibration of the pollen-climate relationships. The use of a global dataset, rather than region-specific training data, relies on the principle of phylogenetic niche conservatism (Harvey and Pagel, 1991; Qian and Ricklefs 2004; Wang et al., 2025), which states that traits tend to remain constant over time and that the climatic niches of specific genera are also conservative (Harrison et al., 2025). The use of a global dataset for calibration makes it possible to sample a large range of climates, and thus makes it more likely that the reconstructions of glacial climates are realistic and not confined to the limited climate range sampled in any one region (Turner et al., 2020).

New refs:

Qian, H. and Ricklefs, R. E.: Geographical distribution and ecological conservatism of disjunct genera of vascular plants in eastern Asia and eastern North America, *J. Ecol.*, 92(2), 253–265, doi:<https://doi.org/10.1111/j.0022-0477.2004.00868.x>, 2004.

Wang, C., Wang, M., Zhu, S., Wu, X., Yang, S., Yan, Y. and Wen, Y.: Multiple ecological niche modeling reveals niche conservatism and divergence in East Asian Yew (*Taxus*), *Plants*, 14(7), doi:10.3390/plants14071094, 2025.

Harrison, S.P., Bartlein, P.J., Cruz-Silva, E., Haas, O., Jackson, S.T., Kaushal, N., Liu, M., Magri, D., Robson, D., Vettoretti, G., Prentice, I.C.: Palaeoclimate perspectives on contemporary climate change, *Ann. Rev. Environ. Resour.*, 50, 2025.

As we pointed out in our earlier response, we combined deciduous and evergreen Quercus because in many parts of the world the pollen was only identified to genus level. Specifically, 63% of the training samples are only identified to genus level. We recognise that this could involve a loss of information, since in regions such as Europe evergreen oaks occupy a distinct climate space from deciduous oaks. We have now tested whether excluding Quercus from the training data set has an impact on the quality of the model. This test shows that excluding Quercus does not improve or degrade the fits. The spatial patterns are also not affected by excluding Quercus. ΔMTCO was still found to be significantly larger than ΔMTWA in the northern extratropics, and larger but not significantly larger in the southern extratropics. There was still a significant positive relationship between $\Delta\alpha$ and ΔMTWA in all regions. The largest warming still occurs in Eurasia, while western North America and the southern extratropics are characterised by cooling. $\Delta\alpha$ still broadly follows the changes in temperature, with increased α in regions characterised by warming, and Europe is still characterised by a mixed signal of drying and wetting. The robustness of the results shows that there are other plant taxa that are similarly diagnostic of the climate space occupied by deciduous or evergreen oaks, and thus the inclusion of Quercus at genus level does not impact the reliability of the reconstructions. We have included plots and tables documenting the analysis in the Supplementary (Section 4).

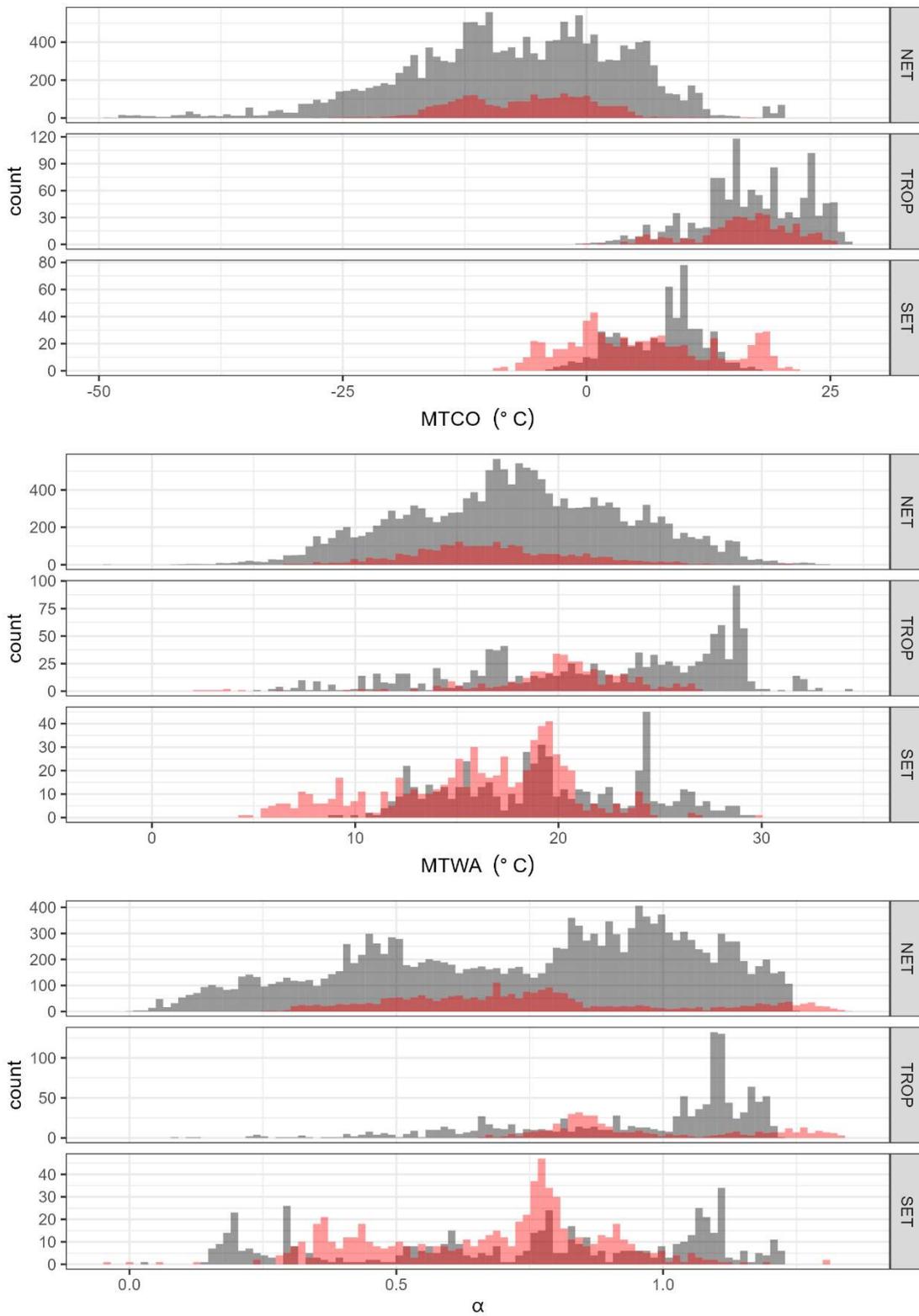
We have modified the Methods to indicate that we have tested the impact of including/excluding Quercus on the reconstructions (first paragraph in section 2.1):

Nevertheless, since there are other plant taxa that are similarly diagnostic of such regimes, the amalgamation of *Quercus* (deciduous) and *Quercus* (evergreen) should not have a major effect on the robustness of our climate reconstructions. We have tested this assumption by making reconstructions based on all taxa except *Quercus* (Supplementary Materials, Section 4).

We have also modified the Discussion section to comment on the impact of this exclusion, as follows (second paragraph of section 4):

The use of a global dataset for calibration makes it possible to sample a large range of climates, and specifically to reconstruct climate variables that might be very different from the range experienced in a region in the modern day. This is particularly important when reconstructing changes in the fundamentally different climate of the last glacial. Reconstructed glacial climates at some sites were indeed found to exceed the climate ranges sampled for the region under modern conditions, most noticeably MTCO and MTWA in the southern extratropics (Figure S8). However, the use of a global data set can create issues because of inconsistencies in taxonomic resolution between regions. The necessity for treating all species of *Quercus* as a single taxon, despite the fact that evergreen and deciduous species may occupy distinct climate niches in some regions such as Europe, is a consequence of this. However, we have shown (Supplementary Materials, Section 4) that this has little impact on our reconstructions – largely because the climatic distinction that would be conveyed through separating deciduous and evergreen *Quercus* is also registered by the presence of other taxa. Although the use of a global training data set for climate reconstructions has not been a common practice, it also facilitates making reconstructions for sites from regions with limited modern pollen data or where the modern samples do not capture the very different climates that might have occurred in that region during glacial times.

Figure S8: Histogram of reconstructed climates (shown in red) vs modern climates (shown in grey) in the northern extratropics, tropics and southern extratropics.



5. Finally, at L125, I would like to understand the rationale behind excluding climatically close samples. While I appreciate the need to remove spatially close samples to account for spatial autocorrelation, even if regression techniques are not the most sensitive to it, I'm unclear about the decision to exclude geographically distant yet climatically close samples. This exclusion may be contributing to the high RMSEP and I believe it's important to justify this decision.

Excluding climatically close samples will lead to a higher RMSEP than not excluding them. The leave-(pseudo-)out cross validation used in this paper will have higher RMSEP than the traditional leave-one-out cross validation. We use this approach to make sure that we could reconstruct climates that are not in the training dataset and to check the predictive power of the model. This ensures that we are not just tuning to the training dataset, and that we can reconstruct climates even when our training set does not completely cover the climate to be reconstructed because there are gaps in the climate space. We have added text to explain this in the Methods (second paragraph in section 2.2):

We evaluated the \hat{x} TWA-PLSv2 models by comparing the reconstructions with observations using leave-out cross-validation, where one site at a time was randomly selected as a test site and sites that are both geographically close (within 50 km horizontal distance from the site) and climatically close (within 2% of the full range of each climate variable in the dataset) were also removed from the training set, to prevent redundancy in the climate information from inflating the cross-validation goodness of fit, following Liu et al. (2020). This ensures that we are not just tuning to the training dataset, and that we can reconstruct climates even when the training set does not completely cover the climate to be reconstructed because there are gaps in the climate space.

Reviewer 3

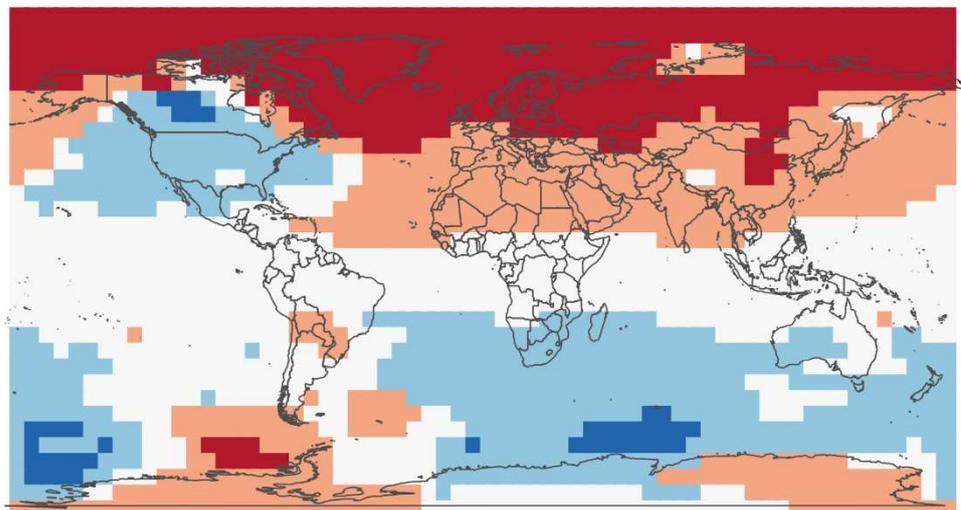
1. The authors repeat 3 times that this dataset will be used to compare with upcoming PMIP meltwater exp. But why don't they already assess existing exp? I understand that there are only a handful of modelling exp are available for MIS3 but there are also some meltwater exp. performed under LGM conditions (eg Kageyama et al. 2013). Please also compare your moisture estimates to changes in precipitation (if that's what it means). For example, the most interesting aspect of your compilation is that the data suggests a cooling over western North America during DO events. That is only consistent with AMOC shutdown exp. in which the NPIW formation strengthens, as in Menviel et al. 2014 for example.

We agree with the reviewer that it would be useful to compare our reconstructions with climate model simulations, and to investigate the mechanisms involved in these changes. However, our chief goal here was to make robust global reconstructions of the changes during D-O events that could subsequently be used for model comparisons. Detailed comparisons with existing simulations, which use different protocols and methods for triggering D-O events is an entirely different endeavour and beyond the scope of this paper. In addition to the comparisons with model findings already included in the Discussion, however, we have added a preliminary comparison of the temperature reconstructions with the LOVECLIM simulations for D-O events 5-12 (Menviel et al., 2014). We have not made comparisons for the moisture signals because α is plant-available moisture, and is calculated as the ratio of actual evapotranspiration to equilibrium evapotranspiration. It is therefore not directly comparable to precipitation. We have added the following text, figures and tables in the Discussions (4th paragraph in section 4):

The LOVECLIM model is used as a reference to adjust the age scale in the reconstructions using MAT, but this does not preclude comparison of the reconstructed and simulated seasonal temperatures. The general spatial pattern of simulated changes in MTCO and MTWA (Figure S9.1) is consistent with the reconstructions, with largest warming in Eurasia, and cooling over most of the southern extratropical land. However, there are important differences. The reconstructions show cooling over western North America in both seasons, for example, but only in winter in the simulations. The relationship between Δ MTCO versus Δ MTWA is also different (Figure S9.2): the simulated Δ MTCO is shown to be significantly larger than Δ MTWA in the northern extratropics, but significantly smaller than Δ MTWA in the southern extratropics (Supplementary Table 3). This comparison illustrates the usefulness of the reconstructions for model evaluation and to investigate the mechanisms that may not be adequately captured by current models.

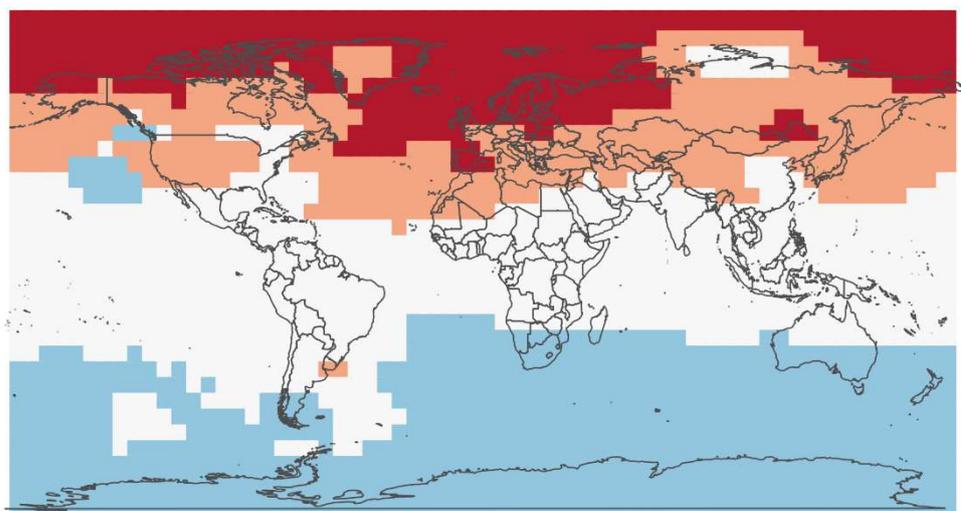
Figure S9.1: Median change of (a) MTCO and (b) MTWA for Dansgaard-Oeschger (D-O) events 5~12 simulated by the LOVECLIM model, using the same way to identify changes as for the reconstructions. MTCO and MTWA values are approximated using the winter and summer temperatures, respectively, since monthly temperature were not available.

(a)



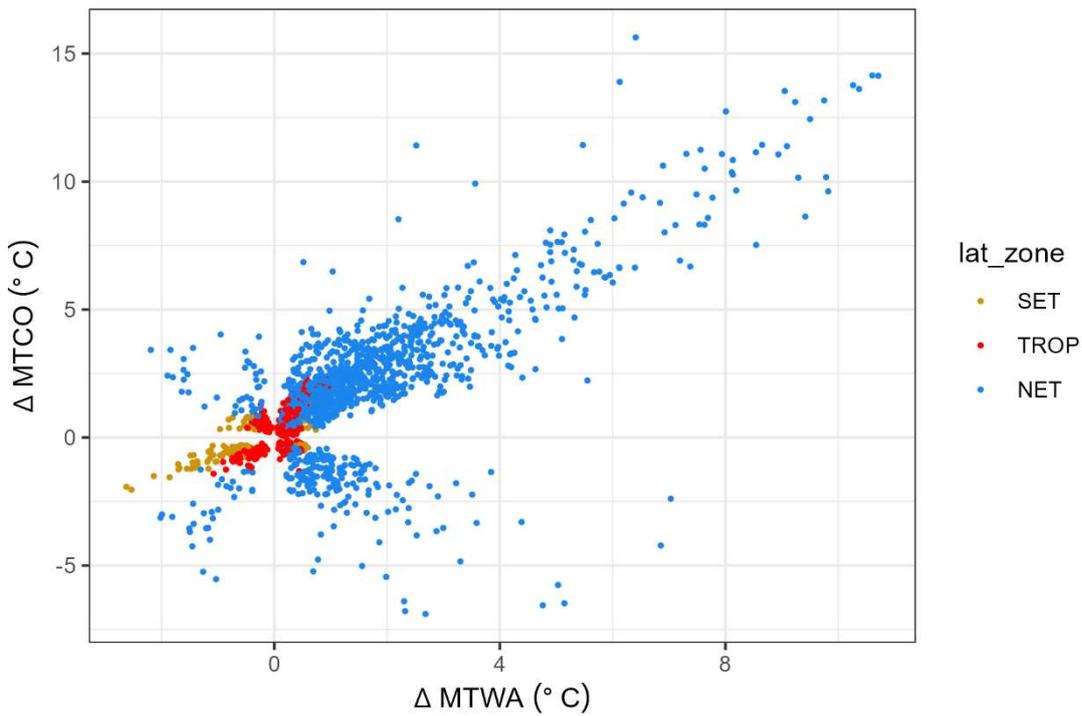
Δ MTCO (°C) ■ -10~-2.5 ■ -2.5~-0.5 ■ -0.5~0.5 ■ 0.5~2.5 ■ 2.5~15

(b)



Δ MTWA (°C) ■ -2.5~-0.5 ■ -0.5~0.5 ■ 0.5~2.5 ■ 2.5~10

Figure S9.2: Scatter plot of the change in mean temperature of the coldest month (ΔMTCO) versus the change in mean temperature of the warmest month (ΔMTWA) during individual Dansgaard-Oeschger (D-O) events at individual land gridcells from the LOVECLIM model. The points are colour-coded to indicate whether the grids are from the northern extratropics (NET, north of 23.5°N), the tropics (TROP, between 23.5°N and 23.5°S) or southern extratropics (SET, south of 23.5°S). MTCO and MTWA values are approximated using the winter and summer temperatures, respectively, since LOVECLIM since monthly temperature were not available.



Supplementary Table 3: Maximum likelihood estimates of the relationship between the change in mean temperature of the coldest month (ΔMTCO) and the change in mean temperature of the warmest month (ΔMTWA) over land in the LOVECLIM model, for the northern extratropics (NET, north of 23.5°N), tropics (TROP, between 23.5°N and 23.5°S) and southern extratropics (SET, south of 23.5°S). The intercepts were set to zero since both variables are changes. MTCO and MTWA values are approximated using the winter and summer temperatures, respectively, since monthly temperatures were not available.

Region		Coefficient	Standard error (SE)	Lower 95%	Upper 95%
NET	Slope	1.278	0.032	1.215	1.342
TROP	Slope	1.515	0.056	1.405	1.625
SET	Slope	0.728	0.023	0.683	0.774

2. It would be good to clarify what alpha represents for modellers: is it precipitation or evaporation-precipitation or is it more related to soil moisture?

α is plant-available moisture, calculated as the ratio of actual evapotranspiration to equilibrium evapotranspiration. It is a transformation of the commonly used moisture index MI, to emphasize the differences at the dry end of the climate range, which have a more pronounced effect on vegetation distribution than differences at the wet end. MI is calculated using SPLASH based on daily values of precipitation, temperature and sunshine hours obtained using a mean-conserving interpolation of the monthly values of each. We have now modified the sentence in the Methods (3rd paragraph in section 2.1):

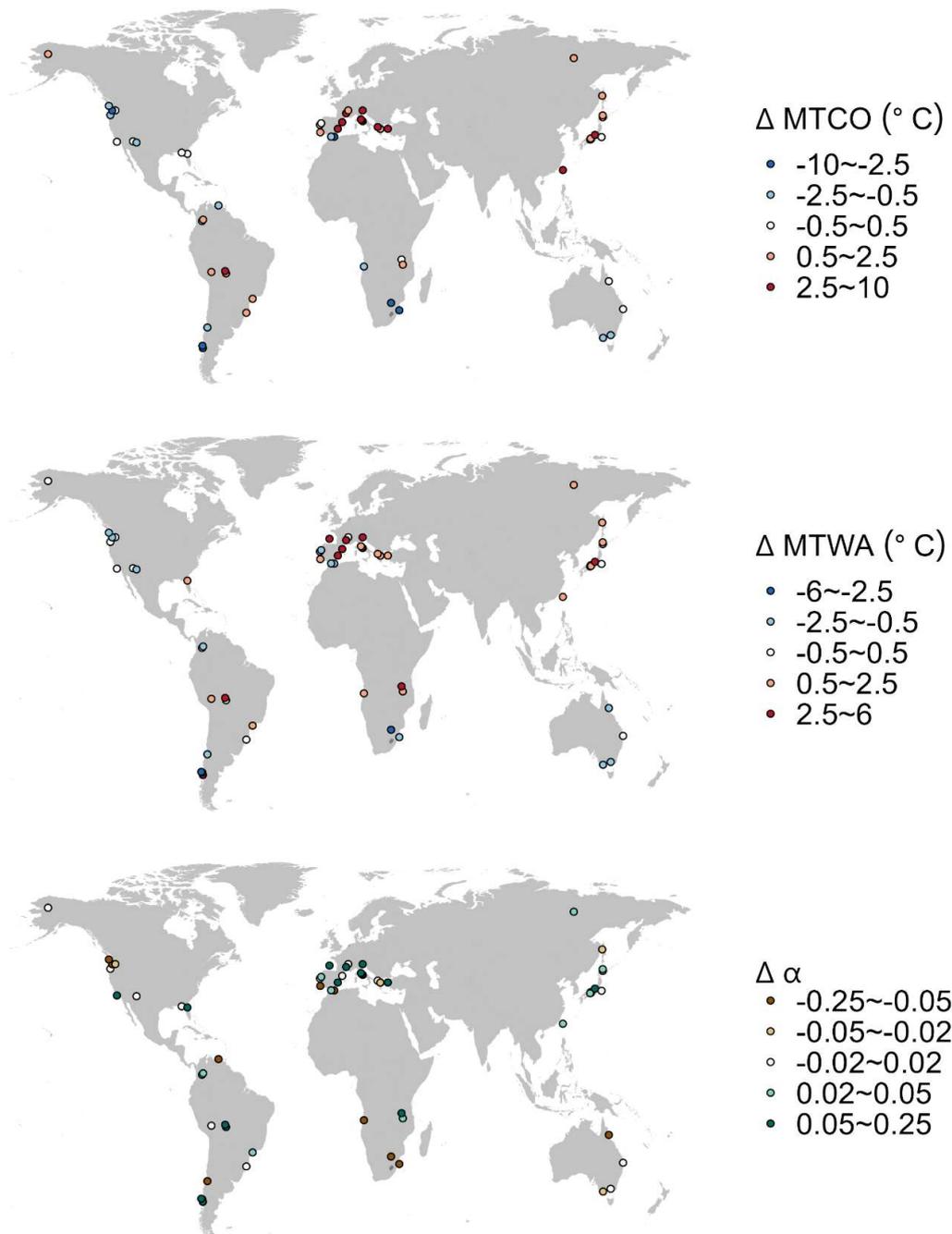
The SMPDSv3 also provides climatic information at each pollen site, specifically the mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA), and a plant-available moisture (α) calculated as the ratio of actual evapotranspiration to equilibrium evapotranspiration. These bioclimate variables reflect mechanistically distinct controls on plant growth. α is a transformation of the commonly used moisture index MI (Supplementary Material, Section 2) to emphasize the differences at the dry end of the climate range, which have a more pronounced effect on vegetation distribution than differences at the wet end (Prentice et al., 2017).

*Prentice, I. C., Cleator, S. F., Huang, Y. H., Harrison, S. P. and Roulstone, I.: Reconstructing ice-age palaeoclimates: Quantifying low- CO_2 effects on plants, *Glob. Planet. Change*, 149, 166–176, doi:<https://doi.org/10.1016/j.gloplacha.2016.12.012>, 2017.*

3. Figure 5 is the main figure with important information, but it is a bit difficult to see the colors and differentiate between the two blues and two reds. Maybe you could make the markers bigger and/or choose reds and blue shades that are more different.

We have now redrawn this.

Figure 5: Maps showing the median change of site-based reconstructions for Dansgaard-Oeschger (D-O) events 5~12.



4. L 185-186: all the equal signs are missing.

We have now added equal signs.

The performance is best for MTCO ($R^2 = 0.75$, RMSEP = 6.52, slope = 0.85) but is also good for MTWA ($R^2 = 0.60$, RMSEP = 3.58, slope = 0.72) and α ($R^2 = 0.65$, RMSEP = 0.183, slope = 0.71).

5. L 293-294: you already mention that at least in 2 other parts of the manuscript. Please remove at least one occurrence.

We have mentioned PMIP in the abstract and in the paragraphs above, so we will remove the statement here.