Response to reviewer 2

Our responses are given in *blue italics* and proposed changes to the text are given in normal blue font.

-My first point concerns the choice of the method to reconstruct past climate changes. You have selected the WAPLS (a modified version of the transfer function): why the WAPLS and not the MAT or BRT? This method is not appropriate here with because the size of your modern pollen dataset (S1) is high and it covers a wide range of biomes and taxa. The WAPLS is useful at local and regional scale but may not be optimal in continental or global scale studies, as the responses of some pollen taxa to the variable of interest can be multimodal (Chevalier et al., 2020) as it is here.

We do not use WA-PLS. As explained in the Methods section (lines 64-80), we use a modified version of fxTWA-PLS, which is explicitly designed to reduce compression biases that affect WA-PLS. The original version of fxTWA-PLS was published in Liu et al., 2020 and we provide detailed information about the new modification in the Appendix of the current paper. In the original paper (Liu et al., 2020), we compared the performance of fxTWA-PLS with the BRT approach as encoded in BUMPER. Although BUMPER performed better than the standard WA-PLS approach, it did not perform as well as fxTWA-PLS: the best BUMPER model had an RMSEP of 4.42, 882, 0.166 and R^2 of 0.74, 0.72, 0.71 for MTCO, GDD_0 and α compared to the fxTWA-PLS RMSEP of 4.37, 830, 0.148 and an R^2 of 0.76, 0.73, 0.72 for MTCO, GDD_0 and α in leave-one-out cross validation (BUMPER can't produce leave-pseudo-out cross validation as in Liu et al., 2020 and this paper, so we compared leave-one-out cross validation result); BUMPER model also has biased residuals in training (see Figure S5.2 in Liu et al., 2020). We provide an evaluation of the improved version of the fxTWA-PLS method using modern data in the current manuscript (Table A1, Figure A1, Figure A2), but we do not think it necessary to repeat the comparison with BUMPER here. In introducing this new methodology, we explain why it was used in preference to existing methods.

The importance of using a large and climatically extensive data set for pollen calibrations is increasingly recognised. The MAT-based reconstructions for Europe made by Mauri et al. (2015) includes 4700 samples covering Europe, the Middle East and Northern Africa. The Eurasian Modern Pollen Database (Chevalier et al., 2019; Davis et al., 2020), which includes over 8000 records and covers the much larger area of the Eurasian continent, was explicitly developed to serve as a calibration data set for pollen-based climate reconstructions. Analyses of the impact of the size of the training data set (Turner et al., 2020) show that training data sets that cover a more limited climate space result in poorer correlations between observed and reconstructed modern climates. Equation 2.14 in Liu et al. (2020) also shows that the standard error of the estimate to the climate will be reduced by increasing the number of taxa used. Small local calibration data sets can have better performance in some circumstances (see next response) but obviously make it impossible to reconstruct climate states outside the range of the modern climate in that locality.

It is difficult to make direct comparisons with the published MAT reconstructions for Europe by Mauri et al. (2015) because they reconstruct different climate variables. However, one of the known issues with the MAT approach is that it produces reconstructions that are more variable than e.g. Bayesian techniques and can also produce unrealistic jumps in the reconstructions because of switches between available analogues (see e.g. Brewer et al., 2008)

Multimodality in the response of pollen taxa to a specific climate variable often happens when there is inadequate sampling of the climate space (see e.g. Wei et al., 2020, Ecology), and to taxa with large tolerance and low abundance (see Supplementary Material 7 in Liu et al., 2020). In our original paper on fxTWA-PLS, we established that the multimodal peaks in abundance had almost no impact on the final climate reconstructions given the large number of taxa used (see Liu et al., 2020).

Moreover, it's better to use a local calibration than a global one: global versus local calibrations (WAPLS) have been investigated in Dugerdil et al (2021). They show that WAPLS performs better for the local database than for global databases.
For the reasons given above, we disagree that it is better to use a local calibration. The Dugerdil et al. (2021) paper shows that local calibration reduces the amplitude of the reconstructed climate changes compared to the global calibration. The reason that the local calibration gives a better result with WA-PLS is that the reconstructed climate is near to the 0-compression point in local calibration but far from the 0-compression point in global calibration. (see part 3c in Liu et al., 2020 for the explanation of 0-compression point.) The two 0-compression points can be very different in extreme climates, such as that examined by Dugerdil et al, and hence the difference between local calibration and global calibration shows up very clearly. However, the fact that they reduce the amplitude of climate changes indicates that the local calibration compresses the range of the reconstructions towards the central part of the climate range. This is exactly what out new method was designed to address and indeed the results shown in Appendix A show that there is reduced bias at the extremes of each climate variable. It is also interesting to note that the Dugerdil et al. paper shows that MAT performs less well than WA-PLS for the Mongolian and Baikal reconstructions.

Moreover, in your study, the relative contributions of individual taxa to the reconstructions of MTCO, MTWA and alpha (Table S2) raises some questions: most of these taxa are very rare in the Iberian Peninsula Holocene pollen records (Parrotia, Huperzia, Dryas, Zelkova....). These taxa can be recorded in the modern pollen dataset but are not representative of Holocene south Mediterranean pollen records. In this frame, I strongly recommend to resize your modern pollen dataset (by excluding biomes not recorded in the fossil assemblages or by spatial selection) and to recalibrate the WAPLS with a smaller but more appropriate training set.
The purpose of Table S2 is to illustrate the method since it shows the 10 most important taxa in the modern data set that contribute to making warmer/colder or wetter/drier reconstructions. It does not show the loadings of all the taxa that contribute to the reconstructions, only the top 10 in each category. We agree that some of the taxa in the current version are rare (although present in some samples) in the Holocene records for the Iberian Peninsula. For this reason, we have provided a new version in which we select the taxa that occur ≥ the median number of occurrences across samples and then show the top 10 of these taxa contributing to making reconstructions warmer/colder, wetter/drier. We have modified the caption describing this Figure, as follows:

Table S2. Relative contributions of individual taxa to the reconstructions of mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA) and plant-available moisture (α). The plots show the top 10 taxa for each end of the climate gradient after first screening out taxa that are relatively rare (i.e. occur < the median number of occurrences of all taxa in the fossil pollen record, which is 178 samples).

The new version is shown below:

	MTCO	MTWA		α
Increasing cold	*Abies*	*Ilex*	Increasing wet	*Myrica*
	Tilia	*Taxus*		*Taxus*
	Thalictrum	*Saxifragaceae*		*Ilex*
	Betula	*Myrica*		*Calluna*
	Onagraceae	*Orobanchaceae*		*Orobanchaceae*
	Ericaceae	*Calluna*		*Sorbus*
	Lycopodium	*Potentilla*		*Potentilla*
	Salix	*Onagraceae*		*Betula*
	Ulmus	*Sorbus*		*Onagraceae*
	Orobanchaceae	*Salix*		*Ericaceae*
		...		
Increasing warm	*Quercus.intermediate*	*Amaranthaceae*	Increasing dry	*Olea*
	Phillyrea	*Myrtaceae*		*Myrtaceae*
	Cistaceae	*Cistus*		*Quercus.evergreen*
	Oleaceae	*Amaryllidaceae*		*Pistacia*
	Cistus	*Phillyrea*		*Cistaceae*
	Arbutus	*Pistacia*		*Amaryllidaceae*
	Pistacia	*Ephedra*		*Cistus*
	Myrtaceae	*Thymelaeaceae*		*Ephedra*
	Ilex	*Tamarix*		*Tamarix*
	Thymelaeaceae	*Oleaceae*		*Thymelaeaceae*

Obviously, taxa that are not present in the fossil samples do not contribute to the climate reconstructions for these samples. The inclusion of these taxa in the training data set does not impact the reconstructions for the fossil samples from Iberia. Similarly, the rare taxa in the Iberian fossil samples, even if they are strongly weighted towards one end of a climate gradient, will not contribute significantly to the climate reconstructions.

We have already explained why we do not think it necessary to use a smaller modern pollen data set. We would also like to point out that it is not possible to exclude modern samples from biomes that are not recorded in the fossil assemblages because this requires that the biomes present are known a priori.

Another way could be you need to validate your results by using another climate reconstruction method (MAT, BRT, RF for example) cf Salonen et al. works.
Comparison with other reconstruction methods does not provide a validation of the results. While several papers have used multiple reconstruction techniques at individual sites, these generally show that there are differences in the reconstructions based on different methods (see e.g. Brewer et al., 2008; Sinopoli et al., 2019) but cannot determine which is more correct except through comparison of goodness-of-fit and errors with the modern training data. As pointed out by Chevalier et al (2020) in their review of pollen-based climate reconstruction techniques, the primary purpose of using multiple techniques is to compare the methodologies rather than to determine which reconstructions are more accurate.

As we state in response to a previous comment, we compared the original version of fxTWA-PLS with BUMPER, which is a BRT approach, and have shown that our method produces better results. The modified version used in the current manuscript reduces compression bias even further and has better performance than our original version. Therefore, it's better to BUMPER reconstructions. Note that we compared our modified version and the original version based on leave-pseudo-out cross validation (see Appendix A), however, BUMPER can't produce leave-pseudo-out cross validation as in Liu et al. (2020) and this paper, so we compared leave-one-out cross validation result in Liu et al., 2020. Leave-one-out cross validation has inflated statistics than leave-pseudo-out cross validation, so the values of R^2, RMSEP and compression are not directly comparable between the leave-one-out BUMPER results in Liu et al. (2020) and the leave-pseudo-out modified fxTWA-PLS results in this paper.
We agree with the reviewer that it would be useful to make it very clear why we have chosen to use fxTWA-PLS instead of alternative methods, and we have therefore added text at the beginning of the methods section to clarify this and modified the current first paragraph to link it to the new text better, as follows:

Multiple techniques have been developed to make quantitative climate reconstructions from pollen (see reviews in Bartlein et al., 2011; Salonenen et al., 2011; Chevalier et al., 2020). Modern analogue techniques (MAT: Overpeck et al., 1985) tend to produce rapid shifts in reconstructed values corresponding to changes in the selection of the specific analogue samples, although this tendency is less marked in the conceptually analogous response surface technique (Bartlein et al., 1986). Regression-based techniques, including weighted averaging methods such as Weighted Average Partial Least-Square (WA-PLS: ter Braak and Juggins, 1993), do not produce step-changes in the reconstructions but suffer from the tendency to compress the reconstructions towards the central part of the sampled climate range. However, this tendency can be substantially reduced by accounting for the sampling frequency (fx) and the climate tolerance of the pollen taxa present in the training data set (fxTWA-PLS: Liu et al., 2020). Bayesian approaches have also been applied to derive climate reconstructions from pollen assemblages (Peyron et al., 1998). However, comparison of fxTWA-PLS with the Bayesian model BUMPER (Holden et al., 2017), shows that fxTWA-PLS performs better in capturing the climate of the modern training data set from Europe (Liu et al., 2020).

Although fxTWA-PLS has clear advantages over other quantitative reconstructions techniques, there is still a slight tendency towards compression. We have therefore made a further modification to the approach as described in Liu et al. (2020). In the original version of fxTWA-PLS, the fx correction is applied as a weight with the form of $1/fx^2$ in the regression (step 7 in Table 1 in Liu et al., 2020). Here (see Appendix A) we make a further modification of fxTWA-PLS by (a) applying the fx correction separately in both the taxon calculation and the regression (step 2 and 7 in Table 1 in Liu et al., 2020) as a weight with the form of 1/fx and (b) applying P-splines smoothing (Eilers and Marx, 2021) in order to reduce the dependence of the fx estimation on bin width. The modified version further reduces the biases at the extremes of the sampled climate range. We used this modified version of fxTWA-PLS to reconstruct three climate variables: mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA) and plant-available moisture represented by α, an estimate of the ratio of actual evapotranspiration to equilibrium evapotranspiration. The individual and joint effects of MTCO, MTWA and α were tested explicitly using canonical correspondence analysis (CCA).

Additional references:

Bartlein PJ, Prentice IC, Webb T III (1986) Climatic response surfaces from pollen data for some eastern North American taxa. J Biogeogr 13:35–57

Chevalier, M., Davis, B.A.S., Heiri, O., Seppä, H., Chase, B.M., Gajewski, K., Lacourse, T., Telford, R.J., Finsinger, W., Guiot, J., Kühl, N., Maezumi, S.Y., Tipton, J.R., Carter, V.A., Brussel, T., Phelps, L.N., Dawson, A., Zanon, M., Vallé, F., Nolan, C., Mauri, A., de Vernal, A., Izumi, K., Holmström, L., Marsicek, J., Goring, S., Sommer, P.S., Chaput, M., Kupriyanov, D., 2020. Pollen-based climate reconstruction techniques for late Quaternary studies, Earth-Science Reviews, 210, 103384, https://doi.org/10.1016/j.earscirev.2020.103384.

Holden PB, Birks HJB, Brooks SJ, Bush MB, Hwang GM, Matthews-Bird F, Valencia BG, van Woesik R. 2017 BUMPER v1.0: a Bayesian user-friendly model for palaeo-environmental reconstruction. *Geosci. Model Dev.* **10**, 483–498. (doi:10.5194/gmd-10-483-2017)

Overpeck JT, Webb T III, Prentice IC (1985) Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. Quat Res 23:87–108

Peyron O, Guiot J, Cheddadi R, Tarasov P, Reille M, de Beaulieu J-L, Bottema S, Andrieu V (1998) Climatic reconstruction in Europe for 18000 yr BP from pollen data. Quat Res 49:183–196

Salonen JS, Ilvonen L, Seppä H, Holmström L, Telford RJ, Gaidamavičius A, Stančikaitė M, Subetto D. 2011 Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative climate reconstruction. *Holocene* **22**, 413–424. (doi:10.1177/0959683611425548)

ter Braak CJF, Juggins S. 1993 Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* **269**, 485–502. (doi:10.1007/BF00028046)

- I strongly recommend to add also regional composite panels (north, central, south?) of temperature and alpha changes instead of a unique composite curve (fig. 3). Regional climate patterns are important (fig 2) and the signal is too averaged if you only look at composite curves. You may miss important signal, so add a discussion about the additional panels and the regional patterns.

Figure 2 provides a summary of the results at individual sites. We have provided composite curves for the Iberian Peninsula to test explicit hypotheses about the controls of climate changes during the Holocene. While we agree that there might be interesting local patterns, there are good reasons for not making composites for smaller arbitrarily defined areas. Firstly, there is the question of how to divide Iberia into coherent regions: the current climate does not show straightforward north-south, east-west patterns. Furthermore, as we show in our analysis of the changes in moisture gradients, the appropriate coherent regions today would not necessarily be coherent in the past. Secondly, composites based on a limited number of sites will be inherently noisier. We will expand the section describing the construction of the composite curves in the Methods to clarify the purpose of these, as follows:

In addition to examining the reconstructions for individual sites, we constructed composite curves for the Iberian Peninsula as a whole. The composite curves provide a way of comparing the relationship between trends in the reconstructed climate changes and insolation changes. The curves were constructed after binning the site-based reconstructions using ± 500-year bins. We did 1000 bootstrap resampling of the reconstructed climate values in each ± 500-year bin to avoid the influence of a single value or a single site on the mean climate value in this bin, and use the standard deviation of the 1000 values to represent the uncertainty of the mean climate value. We constructed linear regression plots to examine the longitudinal and elevational patterns in the reconstructed climate variables, and assessed the significance of differences in these trends through time compared to the most recent bin (0.5 ka ± 500 years) based on *p* values, with the customary threshold of 0.05. We then compared the climate trends with changes in summer and winter insolation.

- I first suggest to better highlight the innovative side of this study. Your work and those of Tarroso et al (2016) (not cited in your paper!) focus on the reconstruction of the climate (temperature and precipitation) in Iberian Peninsula during the last 15000 years from pollen data. What's new in your paper?
The differences between our paper and the Tarroso et al. (2016) paper are (a) that we use a larger number of sites for the reconstructions, (b) we use a better reconstruction technique, and (c) they show no Holocene signal in temperature after 9 ka although they do show a trend in precipitation. In response to another comment, we have now added a comparison of the Tarroso et al reconstructions with our reconstructions and the other reconstructions available for Iberia. The principle focus of our paper, however, is to use reconstructions to investigate postulated changes in the west-east gradient of temperature and moisture (here represented by α, ratio of actual evapotranspiration to equilibrium evapotranspiration, rather than precipitation) through time. This focus is encapsulated in the title (Holocene climates of the Iberian Peninsula: pollen-based reconstructions of changes in the west-east gradient of temperature and moisture) and we discuss the nature of this gradient and the evidence for changes in the Holocene in the first paragraph of the Introduction. However, since this was obviously insufficiently clear to the reviewer, we will modify the final paragraph of the Introduction to be more explicit about this, as follows:

Here we re-examine the trends in summer and winter temperature and plant-available moisture through the Holocene across Iberia, using a new and relatively comprehensive compilation of pollen data (Shen et al., 2021) with age models based on the latest radiocarbon calibration curve (IntCal20: Reimer et al., 2020). We explicitly test whether there are significant differences in the west-east gradient of moisture and temperature through time. We then analyse the relationships between the changes in the three climate variables and how trends in these variables are related to external climate forcing. These analyses allow us to confirm that the west-east gradient in moisture was less steep during the mid-Holocene and indicate the importance of changes in atmospheric circulation in explaining observed patterns of climate change across the region.

- The paragraph on the modern pollen dataset is too short given that the accuracy of the modern pollen dataset is very important in transfer functions. The ref given for the modern pollen dataset (Harrison, 2019) is not a paper, so more details are needed; how do you calculate the climate parameters? Wordclim1, 2? Chelsea? How do you calculate alpha, which ref?
The Harrison, 2019 reference is to the modern pollen data set that we used, and the contents of that data set are described in the readme file. Please see response below about the

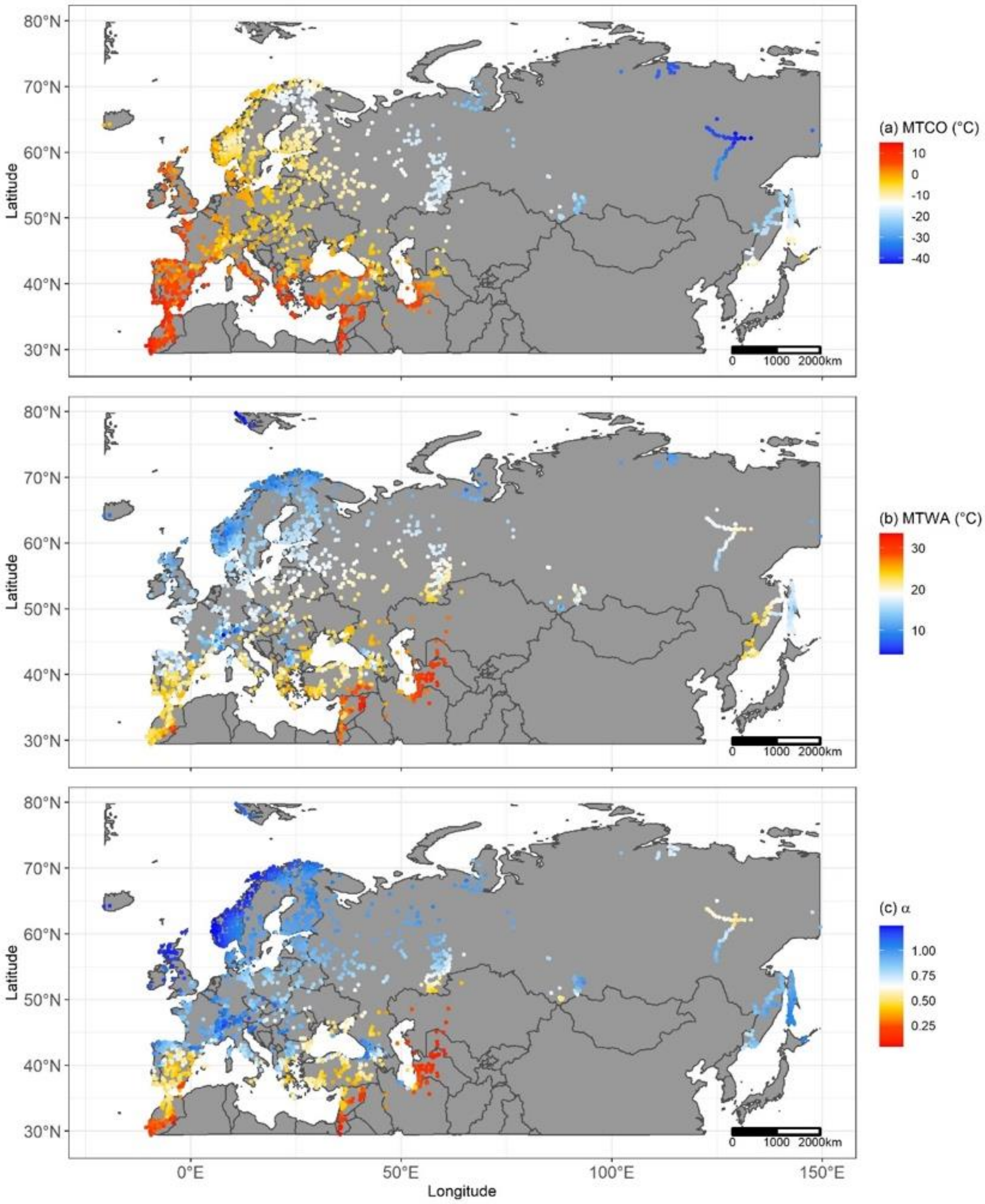
calculation of α, and also the expanded text describing the climate data set. We will also expand the description of the SMPDS as follows:

The modern pollen training dataset was derived from the SPECIAL Modern Pollen Data Set (SMPDS: Harrison, 2019). The SMPDS consists of relative abundance records from 6458 terrestrial sites from Europe, northern Africa, the Middle East and northern Eurasia (SI Figure S1) assembled from multiple different published sources. The pollen records were taxonomically standardized, and filtered (as recommended by Chevalier et al, 2020) to remove obligate aquatics, insectivorous species, introduced species, and taxa that only occur in cultivation. Taxa (mainly herbaceous) with only sporadic occurrences were amalgamated to higher taxonomic levels (genus, sub-family or family) after ensuring consistency with their distribution in climate space. As a result of these amalgamations, the SMPDS contains data on 247 pollen taxa. For our analysis, we use the 195 taxa that occur at more than 10 sites.

New reference:
Chevalier, M., Davis, B.A.S., Heiri, O., Seppä, H., Chase, B.M., Gajewski, K., Lacourse, T., Telford, R.J., Finsinger, W., Guiot, J., Kühl, N., Maezumi, S.Y., Tipton, J.R., Carter, V.A., Brussel, T., Phelps, L.N., Dawson, A., Zanon, M., Vallé, F., Nolan, C., Mauri, A., de Vernal, A., Izumi, K., Holmström, L., Marsicek, J., Goring, S., Sommer, P.S., Chaput, M., Dmitry Kupriyanov, D., 2020. Pollen-based climate reconstruction techniques for late Quaternary studies. Earth-Science Reviews 210: 103384

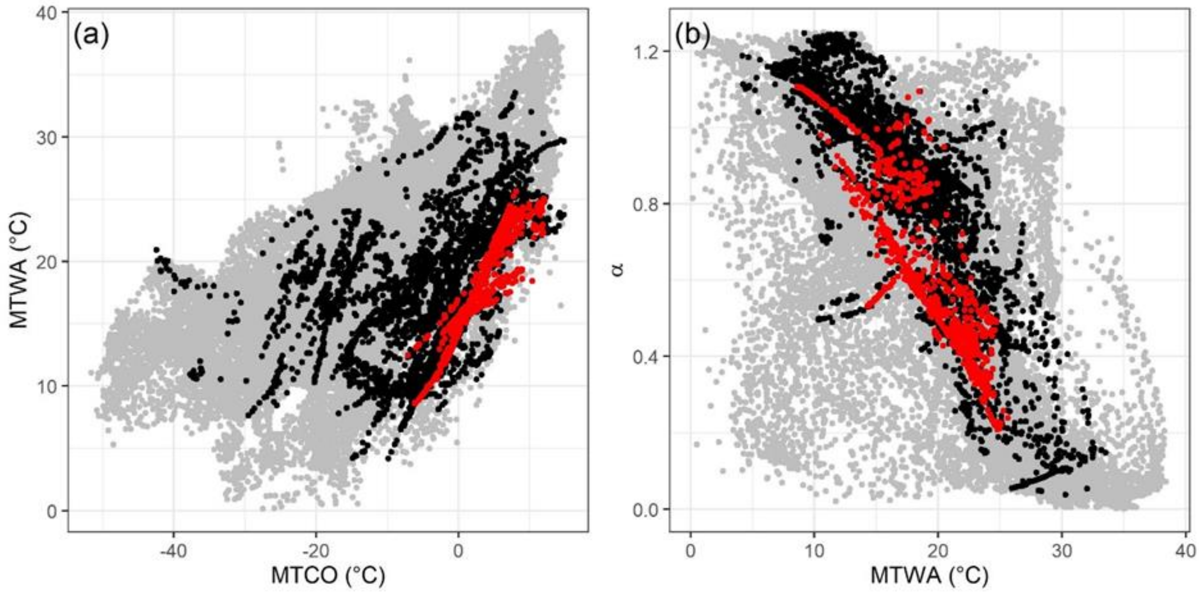
Please add modern values of MTCO and MTWA as you did for alpha (S1). Moreover, the figure with climate values of the training set must be included in the text, not in the Supplementary.

We have added two new panels to the Supplementary figure showing the modern values of MTCO and MTWA at the sites in the training data set.



Rather than moving these figures into the main text, which would not be appropriate given the focus of our paper, we have added a new two-panel figure showing the sites in climate space described by MTCO and MTWA, and MTWA and α respectively. This will now be Figure 1 and we will re-number the other figures accordingly.

Figure 1. Climate space represented by mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA), and plant-available moisture as represented by α, an estimate of the ratio of actual evapotranspiration to equilibrium evapotranspiration. The grey points show climate values for a rectangular area (21° W ~ 150° E, 29° N ~ 82° N) enclosing the SMPDS data set, derived from the Climate Research Unit CRU CL 2.0 database (New et al., 2002). The black points show climate values of the SMPDS dataset. The red points show climate values of the Iberian Peninsula region in the SMPDS dataset.



We have expanded the text describing the climate data as follows:

Modern climate data at each of the sites in the training data set were obtained from Harrison (2019). This data set contains climate reconstructions of MTCO, growing degree days above a baseline of 0° C (GDD_0) and a moisture index (MI), defined as the ratio of annual precipitation to annual potential evapotranspiration. The climate at each site was obtained using geographically-weighted regression of the CRU CL v2.0 gridded dataset of modern (1961-1990) surface climate at 10 arc minute resolution (New et al., 2002) in order to correct for elevation differences between each pollen site and the corresponding grid cell. The geographically-weighted regression used a fixed bandwidth kernel of 1.06 ° (~140km) to optimize model diagnostics and reduce spatial clustering of residuals relative to other bandwidths. The climate of each pollen site was then estimated based on its longitude, latitude, and elevation. MTCO and GDD_0 was taken directly from the GWR regression and MI was calculated for each pollen site using code modified from SPLASH v1.0 (Davis et al., 2017) based on daily values of precipitation, temperature and sunshine hours again obtained using a mean-conserving interpolation of the monthly values of each. For this application, we used MTCO directly from the data set but calculated MTWA from MTCO and GDD_0, based on the relationship between MTCO, MTWA and GDD_0 given by Appendix 2 of Wei et al. (2021). We derived α from MI following Liu et al. (2020). The modern training data set provides records spanning a range of MTCO from – 42.4 °C to 14.8 °C, of MTWA from 4.2 °C to 33.5 °C, and of α from 0.04 to 1.25 (Figure 1, SI Figure 1).

Additional references:

Davis, T. W., I. C. Prentice, B. D. Stocker, R. T. Thomas, R. J. Whitley, H. Wang, B. J. Evans, A. V. Gallego-Sala, M. T. Sykes, and W. Cramer. 2017 Simple process-led algorithms for simulating habitats (SPLASH v.1.0): Robust indices of radiation, evapotranspiration and plant-available moisture. *Geoscientific Model Development* **10**: 689-708, https://doi.org/10.5194/gmd-10-689-2017

New M., Lister D., Hulme M., Makin I., 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21, 1–25. https://doi.org/10.3354/cr021001.

- The paragraph on the fossil pollen dataset is also too short. In the ref cited for the fossil dataset (Shen et al., 2021 CPD) I just found a list of the taxa in the supplementary. It's not enough. Data have been extracted from Neotoma, Pangea, EPD?
We should have provided a full reference for the data set rather than saying where we obtained it. The data set description now provides information on whether the data were obtained from the original authors or from a public-access data set.

The fossil pollen data from the Iberian Peninsula were compiled by Shen et al. (2021) and the data set (Harrison et al., 2021) was obtained from https://doi.org/10.17864/1947.000343. The taxonomy used by Shen et al. (2021) is consistent with that employed in the SMPDS. Shen et al. (2021) provides consistent age models for all the records based on the IntCal20 calibration curve (Reimer et al., 2020) and the BACON Bayesian age-modelling tool (Blaauw et al., 2021; Blaauw and Christeny, 2011) using the supervised modelling approach implemented in the ageR package (Villegas-Diaz et al, 2021). We excluded individual pollen samples with large uncertainties (standard error larger than 100 years) on the attributed in the new age model. As a result, the climate reconstructions are based on a fossil data set of 7294 pollen samples from 117 records covering part or all of the last 12,000 years (Figure 2), with 42 individual records provided by the original authors, 73 records obtained from the European Pollen Database (EPD, www.europeanpollendatabase.net) and 2 records from PANGAEA (www.pangaea.de/). Details of the records are given in SI Table S1. The average temporal resolution of these records is 101 years. We then excluded a few samples where the reconstructed values of α exceed the natural limit of 0 and 1.26. Finally, 7121 samples from 117 records are used for the analyses of the climate reconstructions.

The revised version of Table S1 in this paper and the Supplementary to the Shen et al. paper now provide a list of sites, the source for each site and the original references.

The description of the data sources of fossil pollen used to reconstruct the climate in the Iberian Peninsula (table S1) must be included directly here in the text and not in supplementary material. Table S1 must be updated with the origin of fossil pollen records: for each site, please add the references of the papers, information about the number of 14C date available, and the temporal range covered as for example, 8000-2000 cal yrs BP (not clear as it is in table S1: what does length mean?).
We do not think it is necessary to move this large table from the Supplementary into the main text, particularly since we have expanded it as suggested by the reviewer. We have added the source of each record and the publications to Table S1 (as in the Supplementary to Shen et al., 2021). We have also added the number of dating points used to construct the age models - noting that some of these sites have other types of date than radiocarbon. We emphasised the length of the period covered and the number of samples available in the original version of this table because this is important for temporal resolution. However, we have now added the start and end dates of each record.

There is a lack of comparison of your results with the climate parameters available in the Mediterranean area: the study of Tarroso et al (2016) for Iberian Peninsula of course, Dormoy et al (2009), Combourieu-Nebout et al., (2013), Di Rita et al (2018), Jalali et al. (2016) for south Spain and western Mediterranean. It's important to add the curves of

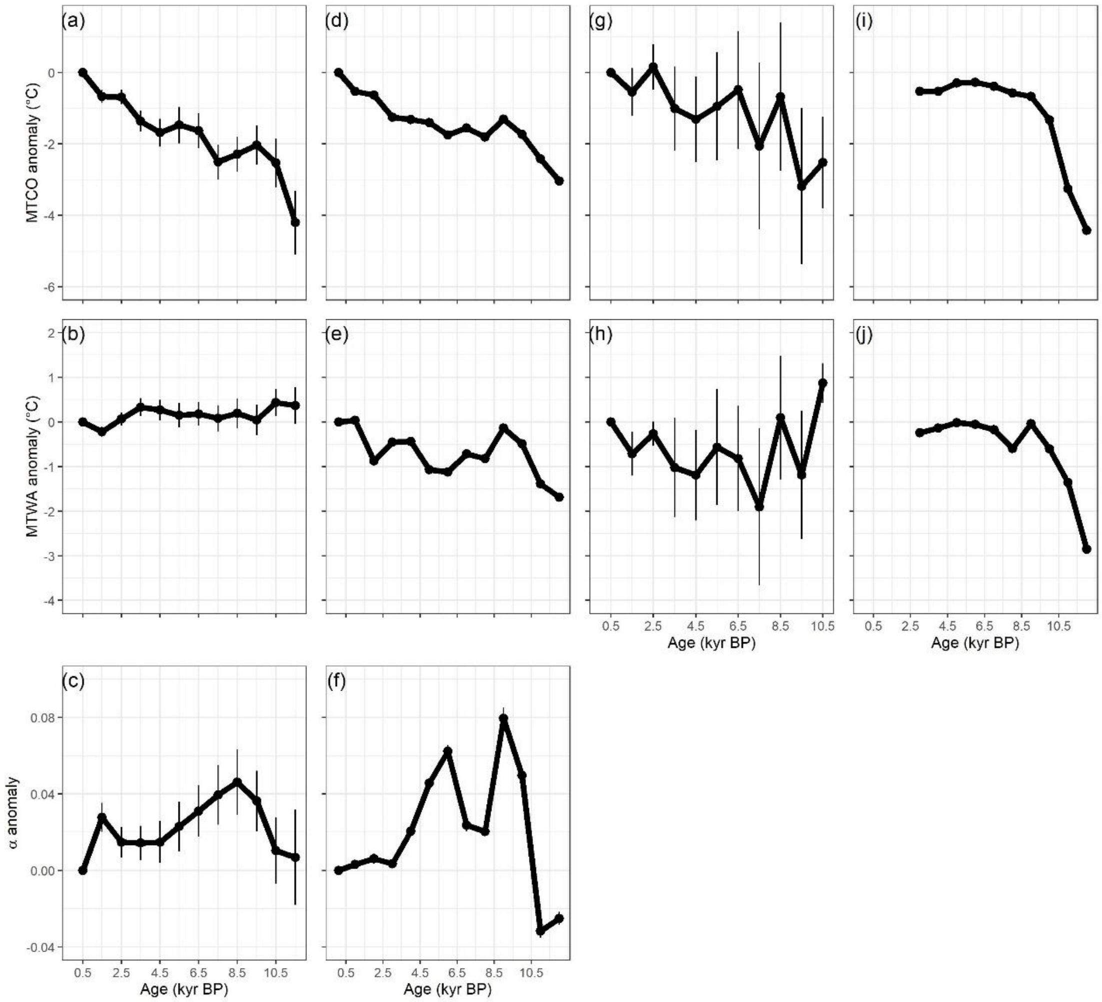
Tarroso et al., (2016) which are based on another climate reconstruction method (the PDF) in your figure to discuss regional patterns.

We did not originally include the Tarroso et al. (2016) paper because they show almost no change in either MTWA or MTCO during 9 ~ 3 ka. We suspect that this is a problem with the methodology - in that the modern distribution data is based on species occurrence data from Atlas Florae Europeae or the Global Biodiversity Information Facility (GBIF), both of which have incomplete coverage. Although they validated the PDFs using core top data from the 31 sites, they did not perform a wider validation using surface samples e.g. from the European Modern Pollen Dataset. Nevertheless, we are happy to cite this paper as a source of quantitative climate reconstructions for the region (and we will also take the opportunity to mention the Kaufman et al data set that we refer to later), and will expand the text in the Introduction to do so as follows:

However, much of the evidence for Holocene climates of the Iberian Peninsula is based on qualitative interpretations of vegetation changes, generally interpreted as reflecting changes in moisture availability (Morellón et al., 2018). These records are extensive and they seem to indicate fairly complex spatial patterns of change. Kaufman et al. (2020) provides quantitative reconstructions of summer and winter temperature in their compilation of Holocene climate information, but there are only 5 terrestrial sites from the Iberian Peninsula. Iberia was also included in the quantitative pollen-based reconstructions of European climate through the Holocene (Mauri et al., 2015). However, the geographical distribution of sites included is uneven and a large fraction of the records were from the Pyrenees and the Cantabrian mountains, with additional clustering of sites in coastal regions. Thus, the inferred patterns of climate over most of the central part of the Peninsula are therefore largely extrapolated. Tarroso et al. (2016) has provided reconstructions of summer and winter temperature and mean annual precipitation since the Last Glacial Maximum for the Iberian Peninsula, by using modern species distribution data to develop climate probability distribution functions (PDFs) and applying these to 31 fossil records. However, although they identified trends in precipitation during the Holocene, the temperature reconstructions do not seem to be reliable since they show no changes through time (9 ~ 3 ka), either for the Iberian Peninsula as a whole or for individual sub-regions, in contra-distinction to the other reconstructions. The current state of uncertainty about Holocene climate changes in Iberia is further exacerbated because quantitative reconstructions of summer temperature made at individual sites using chironomid data (Muñoz Sobrino et al., 2013; Tarrats et al., 2018) are not consistent with reconstructed summer temperatures based on pollen for the same sites.

We have included plots of the reconstructed MTCO and MTWA from Tarroso et al. in Supplementary Figure S8. However, α can't be directly transformed from precipitation due to a lack of other parameters.

Figure S8. Comparison between reconstructed composite changes in climate anomalies. The first column represents this paper, the second column represents Mauri et al. (2015), the third column represents Kaufman et al. (2020), the fourth column represents Tarroso et al. (2016). The composite curves from this paper and Kaufman et al. (2020) are calculated from individual reconstructions, using anomalies to 0.5 ka and a bin of ± 500 years (time slices are 0.5, 1.5, …, 11.5 ka). The composite curves from Mauri et al. (2015) are converted directly from the gridded time slices which are provided with anomalies to 0.1 ka and a bin of ± 500 years (time slices are 1, 2, …, 12 ka). The composite curves from Tarroso et al. (2016) are also converted directly from the gridded time slices provided, with anomalies to 0.5 ka and a bin of ± 500 years (time slices are 3, 4, …, 12 ka). Note that Tarroso et al. (2016) applied a smoothing to the data such that the plots in the paper do not show the excursion in MTWA at 8 ka. In all of the plots, the black lines show mean values across sites, with vertical line bars showing the standard deviation of mean values using 1000 bootstrap cycles of site/grid resampling.



We will revise the Discussion section to include a comparison of these results with our reconstructions, as follows:

We have shown that there was a gradual increase in MTCO over the Holocene, both for most of the individual sites represented in the data set and for Iberia as a whole. Colder winters in southern Europe during the mid-Holocene (6 ka) are a feature of many earlier reconstructions (e.g. Cheddadi et al., 1997; Wu et al., 2007). A general warming trend over the Holocene is seen in gridded reconstructions of winter season (December, January, February) temperatures as reconstructed using the modern analogue approach by Mauri et al. (2015), although there is somewhat less millennial-scale variability in these reconstructions (SI Fig. S8). Nevertheless, their reconstructions show a cooling of 3°C in the early Holocene comparable in magnitude to the ca 4°C cooling at 11.5 ka reconstructed here. Although they show conditions slightly cooler than present persisting up to 1 ka, the differences are very small (ca 0.5°C) after 2 ka, again consistent with our reconstructions of MTCO similar to present by 2.5 ka. Quantitative reconstructions of winter temperature for the 5 terrestrial sites from the Iberian Peninsula in the Kaufman et al. (2020) compilation all show a general trend of winter warming over the Holocene, but the magnitude of the change at some of the individual sites is much larger (ca 10°C) and there is no assessment of the uncertainty on these reconstructions. The composite curve of Kaufman et al. (2020) shows an increasing trend in MTCO through the Holocene although with large uncertainties (SI Fig. S8). In contrast to the consistency of the increasing trend in MTCO during the Holocene between our reconstructions and those of Mauri et al. (2015) and Kaufman et al. (2020), there is no discernible trend in MTCO during the Holocene reconstruction of Tarroso et al. (2016). Indeed, there is no significant change in their MTCO values after ca 9 ka, either for the Peninsula as a whole (SI Fig. S8) or for any of the four sub-regions they considered.

When discussing the MTWA trends, we will add:

The differences between the three data sets probably reflect differences in the number of records used, but the lack of coherency points to there not being a strong, regionally coherent signal of summer temperature changes during the Holocene. Tarroso et al (2016) also showed no significant changes in MTWA after ca 9 ka (SI Fig. S8).

The Dormoy et al (2009), Combourieu-Nebout et al., (2013), Jalali et al. (2016) and Di Rita et al (2018) papers do not provide reconstructions from terrestrial sites from the Iberian Peninsula, although Demoy et al (2009) and Di Rita et al (2018) include reconstructions respectively for one/two marine records south of Iberia. Given our focus on the climate gradients across the Iberian Peninsula, it does not seem appropriate to cite these papers. A pan-Mediterranean analysis of changes in temperature and moisture gradients during the Holocene is beyond the scope of the current paper.

The synthesis figure (S8) must be updated and added in the text not in supplementary.
We have updated this Figure to include the Tarroso et al. (2016) curves for MTWA and MTCO. However, we do not think it is necessary to move this Figure into the main text. Our purpose here is to discuss the degree to which our reconstructions are consistent (or not) with previous reconstructions, but we are not aiming to provide detailed comparisons of the methods used or to evaluate which of these reconstructions is most accurate.

The discussion part on the CO_2 impact must be removed, as you work on the Holocene not on the Late glacial or LGM.

We disagree. We include this because we have previously published on the potential impact of CO_2 on quantitative climate reconstructions based on modern training data sets, and furthermore have developed a robust method to account for this based on known plant physiology responses linking ambient CO_2 levels with changing water use efficiency. Our point here is that this will have an impact, even during the Holocene (see e.g. Figure 6 in Wei et al, 2019). However, the impact of a 40 ppm reduction in CO_2 on reconstructed moisture is less than the uncertainties in our Holocene reconstructions, and will not affect the reconstruction of changes in the west-east gradient, and this is why we do not make this correction in the current analyses. Rather than removing this text, we will clarify why it could be an issue and why we do not think it important for the conclusions of the current paper as follows:

Nevertheless, climate is not the only driver of vegetation changes. On glacial-interglacial timescales, changes in CO_2 have a direct impact on plant physiological processes and reductions in plant water-use efficiency at low CO_2 result in vegetation appearing to reflect drier conditions than were experienced in reality (Farquhar, 1997; Gerhart and Ward, 2010; Prentice et al., 2017; Prentice and Harrison, 2009). The difference between post- and pre-industrial CO_2 levels could also influence the reliability of moisture reconstructions based on modern training data sets. However, the change in CO_2 over the Holocene was only 40 ppm. Prentice et al. (2022) shows that this change relative to modern levels has only a small impact on pollen-based reconstructed moisture indices. The magnitude of this impact is within the uncertainties on our reconstructions. Furthermore, accounting for the effect of this change in CO_2 or not won't affect the reconstructed west-east gradient through time. Therefore, we have not accounted for the impact of changing CO_2 in our reconstructions of α, although there are techniques to do this (Prentice et al., 2011, 2017; Wei et al., 2021).

We will update the Prentice et al. reference, originally given as 2021, to 2022.

You may replace this part by a more in depth discussion on data model comparison (too short!) and atmospheric circulation process.

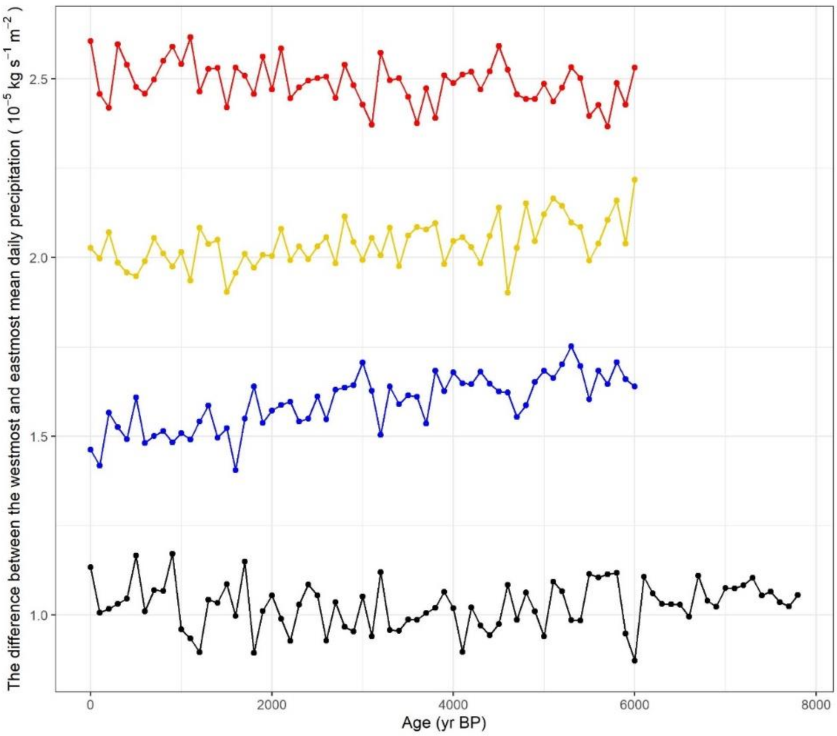
The key point that we want to make in referring to climate model simulations is that they do not show increased moisture advection and this is why the simulated changes in summer temperature are inconsistent with the reconstructions. This point has been made before in the papers we cite with respect to mid-Holocene changes across Europe and in central Eurasia. We do not need to make detailed data-model comparisons for this. Nevertheless, we will expand the Discussion to make the evidence clearer, as follows:

We have shown that stronger moisture advection is not a feature of transient climate model simulations of the Holocene, which may explain why these simulations do not show a strong modification of the insolation-driven changes in summer temperature (Fig. S9). Although the amplitude differs, all of the models show a general decline in summer temperature. The failure of the current generation of climate models to simulate the observed strengthening of moisture transport into Europe and Eurasia during the mid-Holocene has been noted for previous versions of these models (e.g. Bartlein et al., 2017; Mauri et al., 2014) and also shown in Fig. S10. Mauri et al. (2014), for example, showed that climate models participating in the last phase of the Coupled Model Intercomparison Project (CMIP5/PMIP3) were unable to reproduce reconstructed climate patterns over Europe at 6000 yr B.P. and indicated that this resulted from over-sensitivity to changes in insolation forcing and the failure to simulate increased moisture transport into the continent. Bartlein et al. (2017) showed that the CMIP5/PMIP3 models simulated warmer and drier conditions in mid-continental Eurasia at 6000 yr B.P., inconsistent with palaeo-environmental reconstructions from the region, as a result of the simulated reduction in the zonal temperature gradient which resulted in weaker westerly flow and reduced moisture fluxes into the mid-continent. They also pointed out the strong feedback between drier conditions and summer temperatures. The drying of the mid-continent is also a strong feature of the mid-Holocene simulations made with the current generation of CMIP6/PMIP4 models (Brierley et al., 2020). The persistence of these data-model mismatches highlights the need for better modelling of land-surface feedbacks on atmospheric circulation and moisture.

New reference

Brierley, C., Zhao, A., Harrison, S.P., Braconnot, P., Williams, C., Thornalley, D., Shi, X., Peterschmitt, J-Y., Ohgaito, R., Kaufman, D.S., Kagayama, M., Hargreaves, J.C., Erb, M., Emile-Geay, J., D'Agostino, R., Chandan, D., Carré, M., Bartlein, P.J., Zheng, W., Zhang, Z., Zhang, Q., Yang, H., Volodin, E.M., Routsen, C., Peltier, W.R., Otto-Bliesner, B., Morozova, P.A., McKay, N.P., Lohmann, G., LeGrande, A.N., Guo, C., Cao, J., Brady, E., Annan, J.D., Abe-Ouchi, A., 2020. Large-scale features and evaluation of the PMIP4-CMIP6 *midHolocene* simulations. *Climate of the Past* 16: 1847-1872.

Fig S10. The difference between the westmost and eastmost simulated mean daily precipitation in Iberian Peninsula between 8 ka and 0 ka, smoothed using 100 year bins. Here BP means before 1950 AD. The black lines represent Max Planck Institute Earth System Model (MPI) simulations, the red lines represent Alfred Wagner Insitute Earth System Model (AWI) simulations, the blue lines represent Institut Pierre Simon Laplace Climate Model (IPSL-CM5) TR5AS simulations, the orange lines represent Institut Pierre Simon Laplace Climate Model (IPSL-CM6) TR6AV simulations.



-How do you calculate alpha? A ref is needed. How do you explain values above 1?

The calculation of α is based on the Priestley-Taylor formulation and as such has a range between 0 and 1.26 (see Davis et al., 2017 and Supplementary Material 3 in Liu et al., 2020). As explained in the revised text about the climate data given in response to an earlier comment about these data, we derived site-based climate values from Harrison et al. (2019) and we calculated α from the MI values provided in that data set, using the following equation:

$$\alpha = 1.26 \cdot MI \cdot \left(1 + \frac{1}{MI} - \left(1 + \left(\frac{1}{MI}\right)^{\omega}\right)^{\frac{1}{\omega}}\right)$$

using a value for ω of 3, as in Liu et al. (2020).

- I don't agree with your sentence p 2, line 47 "much of the evidence of the Holocene climates is based on qualitative interpretations of vegetation changes...". A lot of other proxies are available: speleothems, chironomids, alkenones... all give independent **values** of climate parameters.

We were not precise enough here. We are in fact referring to the evidence for Holocene climates of the Iberian Peninsula. There are two quantitative chironomid reconstructions from Iberia, and we do indeed compare our reconstructions with these (lines 186 et seq). There are speleothem records from Iberia, but these provide information about oxygen isotopic changes. While these are used to infer changes in precipitation and (in some cases) temperature, they are not a direct quantitative estimate of the climate parameters, and indeed in some cases it is difficult to infer what specifically is driving the changes in isotopic composition (see e.g. Parker et al., 2021). There are alkenone records from the seas around Iberia, but these provide estimates of sea-surface temperature and so are not directly comparable with our reconstructions. We will make the meaning of this sentence clearer as follows:

However, much of the evidence for Holocene climates of the Iberian Peninsula is based on qualitative interpretations of vegetation changes, generally interpreted as reflecting changes in moisture availability (Morellón et al., 2018). These records are extensive, they seem to indicate fairly complex spatial patterns of change.

-- I don't agree with your sentence p 2, line 51 "most of the ca 50 sites from Iberia (Mauri et al 2015) were from the Pyrenees...". Please check and correct: in the Mauri's paper, at least 25 sites of the Iberian Peninsula are not from Pyrenean area and are not extrapolated!

Since this is a gridded data set, and there are large areas of the Peninsula which are not represented in the Mauri et al data set, the inferred patterns of change are indeed extrapolated. A substantial proportion of the sites in the data set are from the Pyrenees and the Cantabrian mountains. The rather "blobby" reconstructions for Iberia in this paper compared to other parts of Europe suggest that individual sites are playing a large role in *the extrapolated surfaces. We will be more precise in our description of the data set and the importance of site distribution in creating gridded surfaces, as follows:*

Iberia was also included in the quantitative pollen-based reconstructions of European climate through the Holocene (Mauri et al., 2015). However, the geographical distribution of sites included is uneven and a large fraction of the records were from the Pyrenees and the Cantabrian mountains, with additional clustering of sites in coastal regions. Thus, the inferred patterns of climate over most of the Peninsula are largely extrapolated.

- Some MTWA and MTCO anomalies values are very low for the Holocene period, especially for the last 6 ka: for example, some sites indicate -7° for MTWA (figs S5, S7), it's too low. Could you check your reconstructions?

There are indeed three individual data points that indicate values of -7° for MTWA (figs S5, S7). These individual samples may be depauperate or otherwise unreliable. In previous work (e.g. Wei et al., 2021) we have removed suspect samples of this sort. In the absence of evidence to exclude these specific samples, we have not excluded them here. However, their contribution to the composite curve is negligible.

- How do you take into account human impact in your modern and fossil pollen data? Usually, we consider that the reconstruction of past climate for the last 2000 years is biased by human impact (check the IPA).
We have removed introduced and cultivated species from our training data set (see revised text describing this data set above) in order to focus on species that can be expected to be diagnostic of climate. We do not otherwise take account of potential human impact on the pollen assemblages. Attempts to quantify human impacts on the vegetation of Europe (e.g. Marquer et al., 2017; Roberts et al., 2018) have only limited coverage for Iberia, although they do imply that major anthropogenic changes in forest cover in northern Iberia occurred only in the last 2-3000 years. There is no obvious break in our reconstructions at this time that would suggest they are less reliable because of human influence.

- fig S9: what is PACMEDY, please explain or add a reference.
PACMEDY was the "PAleao-Constraints on Monsoon Evolution and Dynamics" project which coordinated the transient climate models simulations. Rather than adding a reference to the project, we will add the references to the publications describing the individual simulations to Supplementary, as follows:

We compared our reconstructions to outputs from four transient climate model simulations run as part of the "PAleao-Constraints on Monsoon Evolution and Dynamics" (PACMEDY) project (https://pacmedy.lsce.ipsl.fr/wiki/doku.php): version 1.2 of the MPI (Max Planck Institute) Earth System model (Dallmeyer et al., 2020), version 2 of the AWI (Alfred Wegener Institute) Earth System model (Sidorenko et al., 2019), a version of the IPSL (Institut Pierre Simon Laplace) Earth system model with prescribed vegetation (IPSL-CM5, TR5AS), and one with a dynamic vegetation module (IPSL-CM6, TR6AV) (Braconnot et al., 2019a, 2019b). The four simulations were forced by evolving orbital parameters and greenhouse gas concentrations. The four models have different spatial resolution, with the finest resolution being 1.875 x 1.875° (AWI, MPI) and the coarsest resolution being 1.875 x 3.75° (IPSL-CM5, TR5AS).

Fig S9. Simulated mean values of mean temperature of the coldest month (MTCO), mean temperature of the warmest month (MTWA) and mean daily precipitation in Iberian Peninsula between 8 ka and 0 ka, smoothed using 100 year bins. Here BP means before 1950 AD. The black lines represent Max Planck Institute Earth System Model (MPI) simulations, the red lines represent Alfred Wagner Insitute Earth System Model (AWI) simulations, the blue lines represent Institut Pierre Simon Laplace Climate Model (IPSL-CM5) TR5AS simulations, the orange lines represent Institut Pierre Simon Laplace Climate Model (IPSL-CM6) TR6AV simulations.

Additional references

Braconnot, P., Crétat, J., Marti, O., Balkanski, Y., Caubel, A., Cozic, A., Foujols, M.-A. and Sanogo, S.: Impact of multiscale variability on last 6,000 years Indian and West African monsoon rain, Geophys. Res. Lett., 46(23), 14021–14029, doi:https://doi.org/10.1029/2019GL084797, 2019a.

Braconnot, P., Zhu, D., Marti, O. and Servonnat, J.: Strengths and challenges for transient Mid- to Late Holocene simulations with dynamical vegetation, Clim. Past, 15(3), 997–1024, doi:10.5194/cp-15-997-2019, 2019b.

Dallmeyer, A., Claussen, M., Lorenz, S. J. and Shanahan, T.: The end of the African humid period as seen by a transient comprehensive Earth system model simulation of the last 8000 years, , doi:10.5194/cp-2019-86, 2020.

Sidorenko, D., Goessling, H. F., Koldunov, N. V, Scholz, P., Danilov, S., Barbi, D., Cabos, W., Gurses, O., Harig, S., Hinrichs, C., Juricke, S., Lohmann, G., Losch, M., Mu, L., Rackow, T., Rakowsky, N., Sein, D., Semmler, T., Shi, X., Stepanek, C., Streffing, J., Wang, Q., Wekerle, C., Yang, H. and Jung, T.: Evaluation of FESOM2.0 coupled to ECHAM6.3: Preindustrial and HighResMIP simulations, J. Adv. Model. Earth Syst., 11(11), 3794–3815, doi:https://doi.org/10.1029/2019MS001696, 2019.