Regional pollen-based Holocene temperature and precipitation patterns depart from the Northern Hemisphere mean trends

Response to comments of Referee #2

1. Major issues

1.1 Significance of the reconstructions

Reviewer comment: (3) *Significance of the reconstructions. Tests in H2022 show that only approximately â of the reconstructions show a temperature trend that deviates from noise (i.e. where the reconstruction shows a better correlation with the first principle component of the assemblages than 90 % of the reconstruction based on randomized temperatures; Table 2 in H2022). In the absence of any information about this in the method section, I assume that the same proportion holds for the selection of time series analyzed here. So, why did the authors not filter out these records, as was for instance done in previous work (Marsicek et al. 2018)? As it stands, the analysis presented here is based on reconstructions that are for approximately 66 % noise. Thus the authors really need to convince the reader why they ignore their own previous analyses and present the evidence they have that these reconstructions are valid. One obvious way to do so would be using sensitivity tests and to assess to what degree the observed trends are sensitive to the significance of the individual time series. (If on the other hand, the authors argue that these tests are not meaningful for assessing the robustness of the reconstruction, then that needs to be reflected in H2022.)*

Response: Results of a significance test sensu Telford & Birks (2011) are presented in Herzschuh et al., (2022a); the significance test shows rather low percentages of records that are significant. However, it is discussed in the literature that the Telford-Birks-test is rather conservative and that several other reasons could potentially cause a reconstruction to be flagged as non-significant (see Herzschuh et al., 2022a). A visual inspection of the latitudinal means between those reconstructions derived from WA-PLS, WA-PLS_tailored and WA-PLS_tailored with significant records revealed rather similar overall patterns, which suggests that non-significant records don't affect the outcome of our analyses presented in this study. We provide plots with the latitudinal means WA-PLS_tailored and WA-PLS_tailored with significant records similar to Fig. 3 in the Appendix (Appendix Figures 1 and 2).

1.2 co-variation of temperature and precipitation reconstructions

Reviewer comment: (4) *Here and in H2022 the authors discuss the independence of the temperature and precipitation reconstructions. This is an important issue as the second aim of this study is "What are the continental, latitudinal, and regional patterns of Holocene precipitation change and how do these changes co-vary with temperature trends?" (L111-112). In H2022 the authors use a method to reduce the influence of covariance between temperature and precipitation (tailoring). They conclude that "The tailoring successfully reduced the co-variation of temperature and precipitation in the modern dataset as indicated by the distribution of the correlation coefficient in Fig. 8. Nevertheless, the obtained reconstructions are largely consistent between WA-PLS and WA-PLS-tailored: a correlation of r >= 0.9 is found for 59.2% of all records for TJuly, 60.7% for Tann and 56.5% for Pann." (L292-296 H2022). Notwithstanding whether the r >= 0.9 is a good criterion or not, my conclusion is that the tailored reconstructions are superior because they suffer less from co-variation and that about*

40 % of the time series are markedly different from the non-tailored ones. So if independence of the temperature and precipitation reconstructions is a concern, I fail to understand why the authors ignore their own solution to this problem and not simply use the tailored reconstructions. Similarly, how independent are the annual and July temperature estimates and can one really interpret the difference between them?

Response: We compared the latitudinal mean curves derived from the reconstruction with WA-PLS (Fig. 3) with those curves derived from the reconstruction with WA-PLS_tailored (Appendix Figure 1) and found similar patterns. Hence, we decided to use the standard WA-PLS-derived reconstruction to be consistent with previous studies (see methods section). We conclude that co-variation between temperature and precipitation in the modern calibration dataset is not a major issue in the reconstructions.

In Herzschuh et al. (2022a) we applied a Canonical Correlation Analysis (CCA) to the modern training dataset to infer the relationship between the modern pollen assemblages and climate. A high ratio (>= 1) of constrained (λ_1) and unconstrained (λ_2) explained variance indicate ecologically important determinants. We found the spatial pattern of λ_1/λ_2 for T_{ann} overall similar to T_{July}, but with slightly higher values. We reconstructed both, T_{ann} and T_{July}, as authors use either mean annual temperatures or seasonal (e.g. T_{July}) temperatures for synthesis studies and model-data comparisons. Therefore, we provide both temperature estimates so that the authors could choose which variable they want to use. For our analyses in this study, we also applied our Monte-Carlo test comparison to assess if the linear trends between the reconstructed climate variables are significantly different.

1.3 age uncertainty of the reconstructions

Reviewer comment: (5) *Furthermore, the authors mention the reconstruction uncertainty in the method section and refer to LegacyAge 1.0 (https://doi.org/10.5194/essd-2021-212) for the chronology (and its uncertainties). It remains nevertheless unclear how these uncertainties are treated or if they are considered at all in the analyses presented here. This is important as the inferred changes in temperature and precipitation are small relative to the stated error and because LegacyAge 1.0 indicates that age uncertainties of the time series have a median uncertainty of about 500 years (but reach to over 1,000 years). So, are the regional reconstructions really different from each other?*

Response: Using the full reconstruction error would be over-conservative as the errors are not independent and a large part of the stated error will be in the form of a constant bias for all the samples in a given record, which will then vanish when taking the anomalies. It remains an unresolved issue in the field to our knowledge. Regarding the chronological errors, as they are independent between sites, their overall contribution to a regional average will be small. The same will be true for the reconstruction errors for large enough regions. We now show the regional reconstructions with the standard error computed from the spread between the records. In addition, we applied a Monte-Carlo test comparison to examine linear trends of the latitudinal means and test if they are significantly different from each other (see methods section). We tested the linear trends for both, the zonal means within the continents as well as the weighted means between the continents (see Appendix Tables 2-5).

1.4 reconstruction errors

Reviewer comment: (6) *L156-160: "As it has already been shown in previous comparisons, WA-PLS can have higher RMSEPs than MAT but these do not necessarily reflect a less reliable reconstruction but methodological differences (Cao et al., 2014)." This is an interesting statement and it would be good to repeat some of the reasoning presented in Cao et al here. More importantly, if the estimate of the error is method dependent, how useful then is the error? Would one not get a better, more meaningful, estimate of the reconstruction uncertainty if the difference between various methods is accounted for (see e.g. (Kaufman, McKay, Routson, Erb, Dätwyler, et al. 2020)).*

Response: We added a sentence to the text with some of the reasoning presented in Cao et al. (2014). As for the estimation of the reconstruction uncertainties, we presented quantitative pollen-based reconstructions with 3 different methods (i.e., WA-PLS, WA-PLS_tailored and MAT) in Herzschuh et al. (2022a) and made assessments about estimating the reconstruction uncertainty. However, for this study, we focus on WA-PLS and WA-PLS_tailored, so it would make little sense to derive the reconstruction errors from a comparison of those two methods.

New text: As it has already been shown in previous comparisons, WA-PLS can have higher RMSEPs than MAT but these do not necessarily reflect a less reliable reconstruction but methodological differences. MAT is known to be more sensitive to spatial autocorrelation, which causes the model performance to be over-optimistic compared to WA-PLS (Cao et al., 2014).

Reviewer comment: (7) L160-161: "*Besides, the reconstruction errors are likely much smaller when only the trends and the relative changes are assessed, as in this study.*" This may be true to some extent, but it would be good if the authors provided some explanation for this statement.

Response: Arguments are mainly that the same transfer function (based on the same modern dataset) is used and errors are not independent from each other. However, a quantification of the dependence remains, to our knowledge, an unsolved problem. We expand the explanation.

New text: Besides, trends and the relative changes, as interpreted in this study, are less sensitive to methodological biases than absolute values.

1.5 methodology

Reviewer comment: (8) *Finally, the section on methodology to calculate the time series of temperature and precipitation is descriptive, but I have some additional questions and, crucially, miss some explanation of the rationale. Why were the time series 500-year smoothed and resampled at 100 year resolution and spatially averaged (at 2x2 deg) prior to analysis and why is that the best method if one aims to investigate spatial variability? How was the value of 500 years chosen? How close is the 100 years to the actual resolution of the time series? How were gaps in the time series treated (looking at the data at https://doi.pangaea.de/10.1594/PANGAEA.930500 it appears that the sampling was not done continuously in depth and the time series therefore contain gaps). How spatially representative are the averaged time series for the different subsets? I.e. how many time series (or 2x2 grid cells) used for each regional reconstruction? And how does data availability affect the (un)certainty of the reconstructions and the differences among them?*

Response: The focus of this study is to assess temporal variability on a multi-decadal to centennial-scale and therefore we are interested in long-term trends. To infer those long-term trends, we applied a 500-year smoothing to the time series, which is the typical resolution of the time series used in this study. The function that we used for the smoothing is designed to resample irregularly sampled time

series to an equidistant spacing (Reschke et al., 2019) so that we can read climate values for each record at the exact same time-slice. To address how many time series contribute to a single grid cell, we added a map to the manuscript (Fig. 2).

2. Minor issues

Reviewer comment: (9) *L79-80: "despite the existence of many Holocene pollen records" this seems an odd comment given that some of the syntheses referred to in this sentence post-date the "previous reconstructions". Moreover, some of the authors of this study were also involved in the temperature 12k project, raising the question why they did not include these records at that time.*

Response: The phrasing of this sentence was misleading, indeed. We rephrased the sentence.

New text: Synthesis studies hitherto included rather few records from the large non-glaciated Asian continent (Andreev et al., 2004; Leipe et al., 2015; Melles et al., 2012; Nakagawa et al., 2002; Stebich et al., 2015; Tarasov et al., 2009 and 2013). The inclusion of recently compiled Holocene pollen records (Cao et al., 2019; Herzschuh et al., 2019) and high-quality modern pollen datasets (Tarasov et al., 2011; Cao et al., 2014; Davis et al., 2020; Dugerdil et al., 2021) from Asia now allows for higher quality quantitative reconstructions.

Reviewer comment: (10) *L154: rather than referring to a map that shows the error for the entire LegacyClimate dataset, it would be helpful to present a map of the reconstruction error for the subset of ~1600-900 time series analyzed here.*

Response: We think it is sufficient to refer to the error maps in Herzschuh et al. (2022a) as the 957 records presented here are only a subset of the LegacyClimate 1.0 dataset and the spatial pattern would not change if we present such maps with only the subset. Therefore, we don't see the necessity to add an additional figure (given that our main text and Appendix is already so figure-heavy).

Reviewer comment: (11) *L176: please define the boundary between Asia and Europe.*

Response: We defined the boundary between Asia and Europe at 43°E between Black Sea and Caspian Sea. We indicated the boundaries between the continents in the text now.

New text: To calculate zonal, (sub-)continental (i.e., Asia (>43°E), Europe (<43°E), Eastern North America (<104°W; Williams et al., 2000) and Western North America), and hemispheric means we selected all 957 smoothed and resampled time-series of T_{July}, T_{ann}, and P_{ann} that cover the full period between 11 and 1 ka and calculated climate anomalies for all three climate variables.

Reviewer comment: (12) *L247: "fewer" instead of "less".*

Response: Thank you, we changed the wording.

Reviewer comment: (13) *L250: "values outside the range" please show (or mention) the entire range and what proportion of the data points falls within the restricted range. This sentence raises suspicion about the reconstructions that can easily be avoided.*

Response: We added a table with the entire ranges and the proportions of values that fall within the restricted range in the Appendix (Appendix Table 1), as requested.

Reviewer comment: (14) *The maps, especially those in Fig. 3, are really small and difficult to read. The individual panels can be made bigger by removing white space and redundant labeling without the need to increase the overall figure size.*

Response: We revised the maps in Figures 1, 4 and 5, made them bigger, changed the arrangement of the panels, reduced white space between the panels, removed redundant labeling and improved the color contrast (see also Reviewer comment (14) of Referee #1).