LegacyClimate 1.0: A dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the late Quaternary

Response to comments of Anonymous Referee #1

1. General comments

Reviewer comment: (1) *The authors provide temperature and precipitation reconstructions based on pollen assemblage time series. They provide three different types of reconstructions and provide a clear description of the methods. The dataset is highly valuable and the manuscript is clearly written and the figures are of high quality (if sometimes a bit small).*

Response: Thank you for this encouraging comment.

Reviewer comment: (2) *The manuscript seems to be part of a set of articles (a trilogy?): a manuscript describing the raw pollen data, a manuscript dedicated exclusively to the chronology and the present manuscript about the pollen-derived climate reconstructions. I can to some degree follow the rationale of the sequence, but I think this (last?) article would benefit from a closer integration with the article describing the chronology. The chronology, and importantly its uncertainty, is an integral part of the climate reconstruction that the authors present here.*

Response: Thank you for this comment. We revised the method part and not more clearly indicate the rationale of the three manuscripts.

2. Major issues

2.1 Integration with chronology

Reviewer comment: (1) *This manuscript focuses entirely on the reconstruction of temperature and precipitation, yet the time series also have a chronology with associated uncertainty. By separating these two aspects into two manuscripts it becomes unclear how the full uncertainty of the paleoclimate time series can be derived. Looking at the data (on pangaea.de) it seems that the provided error only accounts for the reconstruction, not for the chronology. This is not the full story and the manuscript would be tremendously improved if the authors made this third manuscript of the sequence a true integration of the papers on the chronology and the climate reconstruction. In L341-343 the authors even touch on this possibility, but they refrain from taking the logical next step that would make the data product more useful for other researchers.*

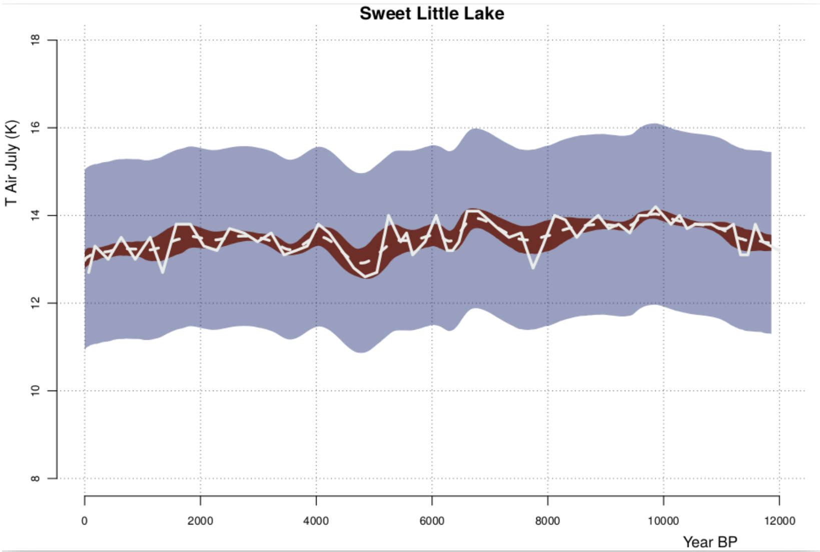
Response: We thank the reviewer for this suggestion. While it is possible to obtain the ensemble of age-models from the Bacon modelling in the chronology paper from Li et al. (2022, LegacyAge 1.0, https://doi.org/10.5194/essd-14-1331-2022), it would clearly simplify the task for users and foster better practices related to age-uncertainty if we also included it in the current manuscript along with the reconstructions data.

Therefore, we added to the database of reconstructions an ensemble of 1000 realizations of the age-models for each record based on the Bacon age modelling performed by Li et al. (2022, LegacyAge 1.0).

Reviewer comment: (2) *This means that the first order analysis of the time series as shown in figures 5 and 6 should include some combined error resulting from the reconstruction and the chronology and a clear description of the methodology to combine these errors. The provided data sets should also contain uncertainties that reflect both the chronological and the reconstructions errors. This is not a complicated step, but would massively improve the value of the data product.*

Response: Using these 1000-member ensembles, we recalculated the maps for figure 5 taking each time the average value over the ensemble of the interpolated values at 1 ka BP and 6 ka BP. The spread (standard deviation) of the ensemble is then used as a measure of the error related to age-uncertainty, or the chronological error. The standard reconstruction error (based on the RMSE from the transfer function) is likewise interpolated at 1ka BP and 6ka BP and the mean over the ensemble is taken as the reconstruction error for the given age. As the chronological and reconstruction errors are independent, they can be added in quadrature to obtain the combined error. This information was added to the source data for Figure 5 and the methodology described.

In the case of Figure 6 (now Figure 10), the figure is provided to show the overall spatio-temporal availability and variability of the timeseries and adding errors on the figure would make it illegible. The data product does include however all the necessary information to perform an analysis, namely the reconstruction error for every sample, and the ensemble of 1000 realizations of the age-model for each record. With this information, users can easily produce curves with all relevant uncertainties such as for example the following figure (title “Sweet Little Lake" on which is shown the reconstruction error (shaded blue) and the chronological error (shaded red) around the reconstruction smoothed by the time-uncertainty (i.e. when we interpolate at regular timesteps for the 1000 realizations and average over the ensemble, dashed white). As the reconstruction and chronological errors are independent, we could also show the combined error by adding them in quadrature, although it will be almost indistinguishable from the reconstruction error given it is much bigger. The original reconstruction with the median ages is also shown for comparison (solid white); this underlines that averaging over the age models only preserves the low-frequencies but (unrealistically) smooths out the high-frequencies. We see that showing all this information on Figure 6 for all the reconstructions would not be possible.



2.2 Meaning of reconstruction differences

Reviewer comment: (1) *The authors also mention other reconstruction methods (L372), which begs the question why MAT and WA-PLS were chosen. Only because they are widely used, or because they yield superior results?*

Response: Thank you for this comment. By providing the fully harmonized modern and fossil datasets as well as the climate data it is rather simple to adjust the R code to run customized reconstruction using further reconstruction methods. Given that many reconstruction methods were proposed in the last 20 years, we decided to provide here only the two most generally used methods. We scanned the literature before running the reconstructions to select the most commonly used methods and selected MAT and WA-PLS. To provide reconstructions from more methods would be beyond the focus of our manuscript. However, we added further discussion on the potential to use the framework for further reconstruction methods.

Reviewer comment: (2) *In addition, the authors provide three different reconstructions for each time series. What I miss is a discussion of how these different reconstructions can be used. Does the difference between them represent additional uncertainty on the reconstruction? How should the user include or use this information? Are certain reconstruction methods better than others? If so, which is to be preferred? If not, how can the (information from the) reconstructions be combined?*

Response: We revised the discussion in '**5.3 Reconstruction method and LegacyClimate 1.0 quality'** and addressed the questions raised in detail.

2.3 Reconstruction quality

Reviewer comment: *The CCA suggests that only some part of the variance in the training sets is explained by T and precip and the significance testing indicates that a shocking 60-70 % of the reconstructions are basically noise. Whilst the authors go some way and filter out the time series that do not pass the significance test, I feel that the authors hardly mention this, let alone discuss. I also realize that this manuscript should not analyze the data, but perhaps some discussion in place and the different ways in which (pollen) assemblages could be used in paleoclimate science, including forward modeling, could be highlighted.*

Response: We did not filter the dataset as we here only provide a dataset that could be used for climate analyses. However, we provide quality measures for each fossil pollen site including measures for the quality of the modern training set (e.g. CCA), for the transfer function (e.g RMSEP) and for the reconstruction (significance test) etc.. In particular, the significance test should rather be taken as additional information than as an exclusion criterion. The significance test tests whether the variation in the pollen data can be significantly explained by the reconstructed climate variable. If the reconstruction does not pass a significance test it indicates that either 1) the climate did not or only marginally change and hence variation in the pollen signal is small and the reconstructed climate variable does not explain a significant amount; or 2) the climate change signal in the pollen data was too small compared to non-climate related changes (e.g. taphonomic changes) or, 3) the changes in the pollen signal are not depicted by reconstructed variables e.g. because the modern data set is not appropriate. Only cases 2) and 3) indicate a failure of the reconstruction method.

We now highlight at the end of the discussion that further assessments and a more comprehensive uncertainty analyses would improve the quality of the dataset.

New text in the discussion part: Our assessments of the modern dataset (e.g. CCA), the transfer function (e.g. RMSEP) and the reconstruction (e.g. the significance test) revealed also the potential biases in the pollen-based reconstruction and pointed to limitations. Further validation and assessments of the results and a more comprehensive uncertainty analyses e.g. by applying forward modelling approaches (Izumi & Bartlein, 2016; Parnell et al., 2016) would be highly valuable.

2.4 Land use issues/human influence

Reviewer comment: *Some of the time series must bear an imprint of human influence. Can the authors briefly discuss to what degree and if and how this influences the reconstructions?*

Response: We added plots of typical land use indicators (as far as available from the harmonized pollen data LegacyPollen 1.0, Herzschuh et al., 2022).

New text in the methods part: We used *Plantaginaceae* (mostly representing *Plantago lanceolata*-type in Europe) and *Rumex*-type to assess human influence as an indicator for intense herding (Behre, 1988). In addition, we calculated the correlation between the WA-PLS reconstruction of T_{July}, T_{ann} and P_{ann} and the pollen percentages of *Plantaginaceae* and *Rumex* for 9000, 3000 and 1000 years BP.

New text in the result part: We used the abundance of *Plantaginaceae* and *Rumex* as indicators of grazing and such intense animal husbandry. Overall weak human impact is inferred for North America and Northern Asia. The indicators *indicate strong human impact* only in single records at 9000 years BP in China and the Mediterranean region (Fig. 7). The percentage values of *Plantaginaceae* and *Rumex* were high especially in Europe for 3000 year and 1000 years BP which indicates growing human impact on that region. High *Plantaginaceae* correlate with low T_{July} in Central Europe indicating potential biases in the temperature reconstructions i.e. too low temperatures become reconstructed (Fig. 8).

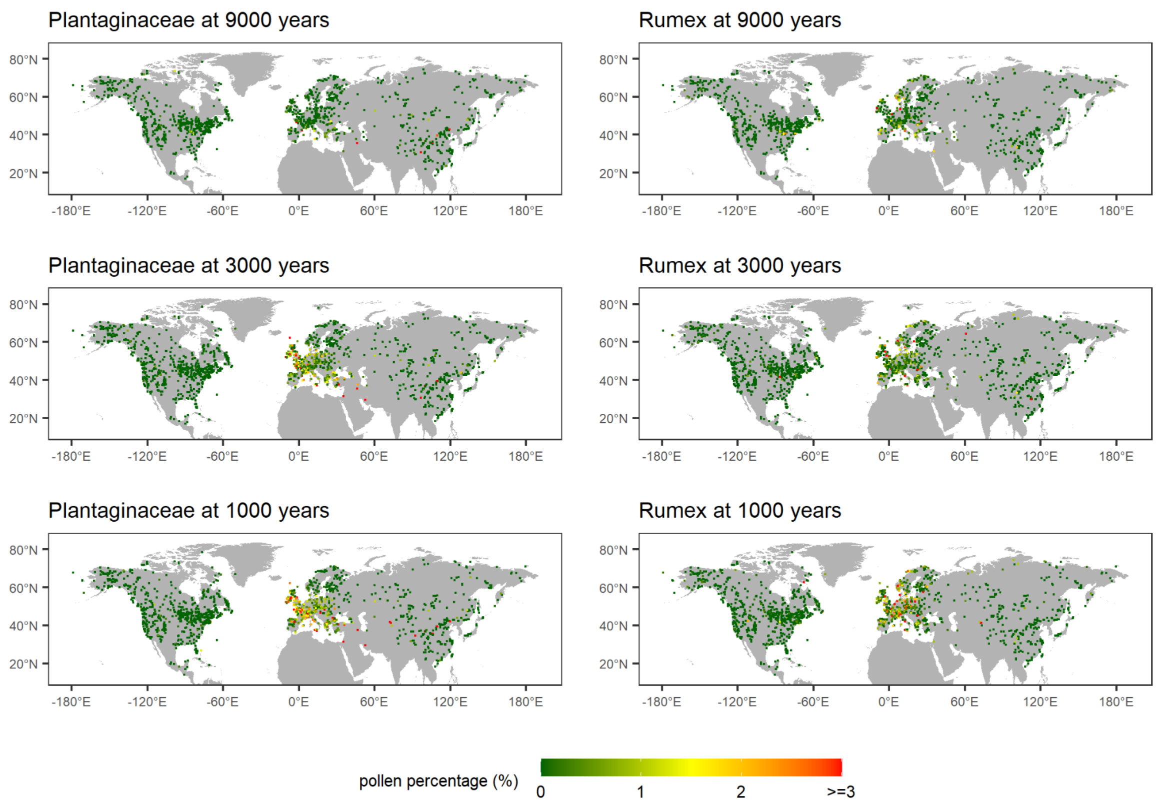


Figure 7. Abundance of *Plantaginaceae* (left) and *Rumex* (right) at 9000, 3000 and 1000 years BP. Colors indicate percentage values.

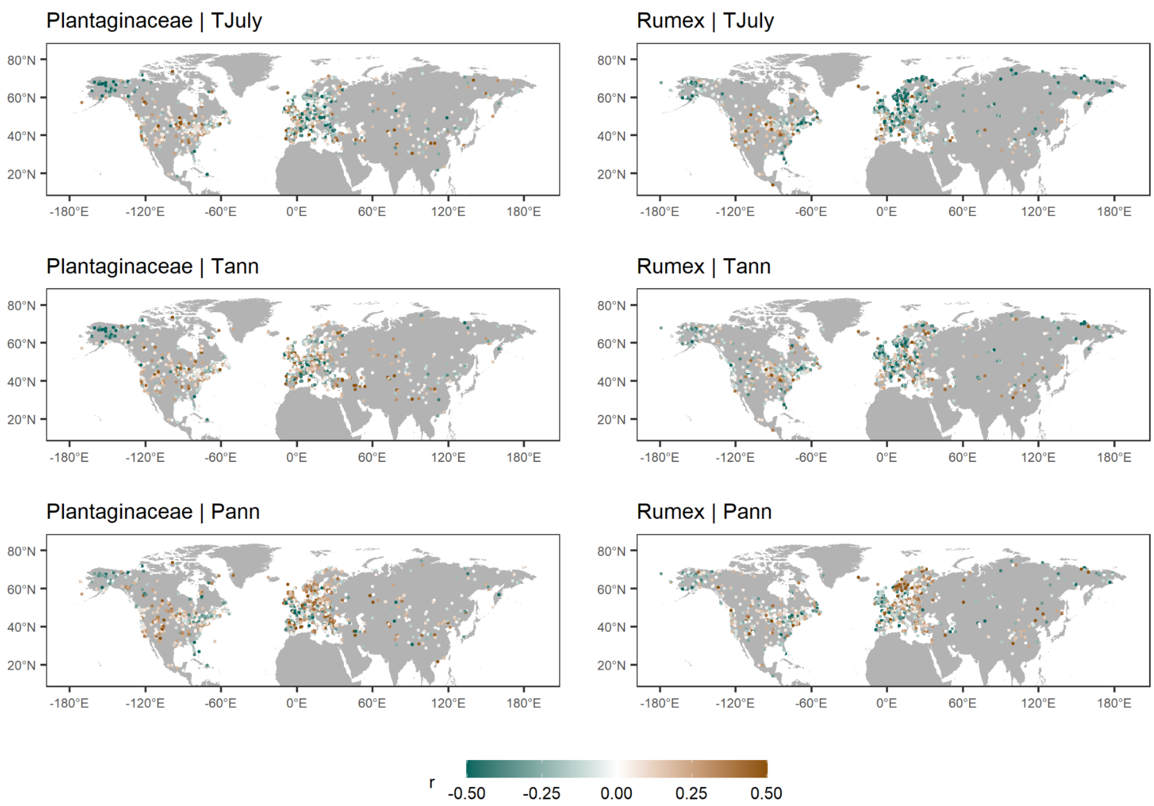


Figure 8. Correlation between the percentage of *Plantaginaceae* (left) and *Rumex* (right) and reconstructed T_{July}, T_{ann} and P_{ann} with WA-PLS.

2.5 Insufficient explanation and detail in the methods

Reviewer comment: (1) *2,000 km radius for training set. Please explain why this was done and why the distance is (globally) appropriate.*

Response: As part of this study we did not perform specific investigations to assess the optimal size of the modern training-set; which would go beyond the focus of this study. In a study from Eastern Asia 1000-1500 km was considered optimal (Cao et al., 2017). However, due to the low number of modern samples in some areas (e.g Northern Asia) we fixed the radius to 2000 km as a good compromise.

New text in the methods part: We fixed the radius to 2000 km, instead of 1500 km as suggested from a study in Eastern Asia by Cao et al. (2017), because the modern dataset density is rather low in Northern Asia.

Reviewer comment: (2) *Why were seven analogues used for MAT? Are the reconstructions weighted to analogue quality, or simply the arithmetic mean of the seven closest analogues?*

Response: We made some tests in advance of the analyses. We found that the results are not very sensitive to the number of analogues (i.e. we tested whether more or less records would pass the significance test). However, we refrained from a systematic study which would be computationally very expensive and go beyond the focus of this study. Accordingly, we decided to stay with the default parameters of the rioja R package used which is 7 analogues.

Reviewer comment: (3) *How is the calibration error determined? Was spatial autocorrelation taken into account? From the code it seems that this is not the case, why?*

Response: The calibration error was determined using the default leave-one-out cross-validation of the rioja package. We report the RMSEP from cross-validation for the models and RMSE for all samples. The *rioja* R-package is one of the most commonly used packages for climate reconstruction using proxy data.

Reviewer comment: (4) *What is the sample-specific error based on? Why is this provided and not the calibration error?*

Response: We provide the full model RMSE as well as the RMSEP derived from leave-one-out cross-validation.

Reviewer comment: (5) *If I am correct, the tailoring approach serves the purpose of reducing the effect of co-variation between T and P. Please mention this earlier in the methods. I understand the point and that this goes some way to alleviating the problem. But what is done in cases where the correlation is not reduced? After all, there still is a large proportion of the sites for which there is a marked correlation in the training set. Some discussion would be appropriate here.*

Response: We now provide more explanation on the rationale. We assume that information about temperature and precipitation cannot be separated from each other if all samples are almost located along a linear line in a temperature vs precipitation space i.e. if they are highly correlated.

New text in the methods part: In addition to the classic WA-PLS reconstruction, we also propose WA-PLS_tailored. This approach addresses the problem that co-variation of climate variables today in space is transferred to the reconstruction even if the past temporal relationship among the climate variables mechanistically differs. In fact, this approach aims to make use of the full climate space covered by the modern pollen samples avoiding those samples in the calibration set that cause the spatial covariation. This approach is based on the assumption that several climate variables can be reflected in one and the same pollen assemblage because different plant taxa have different optima in temperature and precipitation ranges and might therefore occur with different co-occurrence and abundance pattern.

Reviewer comment: (6) *Please provide more detail on the significance test. How were the random environmental fields generated? Simple permutation, or taking spatial correlation into account. Why?*
Response: The significance test is described in Telford and Birks (2011, https://doi.org/10.1016/j.quascirev.2011.03.002). We extended the text by a little more information.
New text in the methods part: A statistical significance test (Telford and Birks, 2011) was applied using the *randomTF* function in the *palaeoSig* R-package (version 2.0-3, Telford, 2019). In this test, the proportion of variance in the fossil pollen data explained by the reconstructed environmental variable is estimated from redundancy analysis (RDA) and tested against a null distribution generated from a total of 999 randomly generated environmental variables from the training data. A reconstruction is considered statistically significant if the reconstructed variable explains more of the variance than 95% of the random reconstructions (Telford and Birks, 2011). The reconstructed climate parameters were tested as introducing the environmental variable as a single variable in a run, as well as with partialling out the explained variance in the pollen data by the respective other variable.

Reviewer comment: (7) *Why were the tailoring and the significance testing not applied to the MAT reconstructions?*
Response: Significance testing is currently also applied to the MAT and summary will be reported in Table 2. From some tests we can see that the percentages of sites that pass the significance test are in the similar order of magnitude as for the WA_PLS. However, running this test is extremely computational time-consuming; accordingly we can provide the results only with the next review round. Tailoring would not make sense with MAT as here the same analogues for temperature and precipitation are used.

Reviewer comment: (8) *The CCA seems to be the first step in the development of the transfer function model to demonstrate that T and Precip really explain the variance in the assemblages. Would it not be better placed earlier in the description? And why are the implications barely discussed?*
Response: We agree and present the CCA now in the beginning of the results part and added some discussion text.

New text in the discussion part: We a priori selected T_{July}, T_{ann} and P_{ann} as target variables for our reconstructions. However, we provide λ_1/λ_2 (i.e. explained variance of the climate variable in the modern pollen data set relative to the variance explained by the unconstrained first axis; ter Braak, 1988), a commonly used proxy for the assessment of reconstructions. The higher λ_1/λ_2 in the spatial modern dataset desto higher the chance that this target climate variable has also impacted vegetation over time and is thus reflected in the variation of the fossil pollen dataset. As a rule of thumb a ratio of 1 is considered to indicate reliable reconstructions (Juggins, 2012) though useful reconstruction may also be obtained from datasets with lower values. .As expected, maps of RMSEPs reveal similar spatial pattern as the results of constrained ordination. Our results indicate that in particular calibration sets from Europe have low ratios and a high RMSEP for all climate variables (despite we have a high number of modern samples), likely related to the human impact on the modern and fossil data. Some areas that are known for its sensitivity to precipitation e.g. Eastern Asia show low RMSEPs as expected for P_{ann} but on the other hand show a low sensitivity to T_{ann} and T_{July}.

Reviewer comment: (9) *How are poor analogues treated? Do they occur at all after the lumping? There is some discussion in L327-332, but it is unclear what the user of the data can do with this information.*

Response: We now calculated the analogue quality of the all samples and the thresholds (1%, 2.5%, 5% of modern calibration set) for single calibration sets. Results are presented in the manuscript now. However, we did not exclude reconstruction analogues without analogues because almost all samples had more than 7 analogues <5%.

New text in the methods part: To infer the analogue quality as an indicator of no-analogue situations we calculated the minimum dissimilarity (squared chord distance) between modern pollen assemblages and fossil pollen assemblages with probability thresholds of 1%, 2.5% and 5% using the *minDC* function from the *analogue* package (version 0.17-6, Simpson et al., 2021).

New text in the discussion part: We report the analogue distance for each sample to help identify such situations. From our assessments we revealed that analogues quality is overall rather good at least for the Holocene and except for Western Europe in particularly the British Isles (Fig. 4).

3. Minor issues

Reviewer comment: (1) *L3: reconsider the use of “late quaternary” in the title. The meaning is actually rather vague and something along the lines of 30,000 years would be more informative.*

Response: Done.

Reviewer comment: (2) *L108: not sure what the policy is to refer to submitted manuscripts.*

Response: Meanwhile, the manuscript about LegacyPollen 1.0 and LegacyAge 1.0 became accepted.

Reviewer comment: (3) *L131: please provide a bit more detail on WorldClim 2. For instance, what are the data based on, over what period are the data integrated, etc.*

Response: We added more detailed information about the WorldClim 2 dataset compilation to the text.

New text in the methods part: The site specific T_{ann}, T_{July}, P_{ann} were derived from WorldClim 2 version 2.1 (spatial resolution of 30 seconds (~1 km^2), https://www.worldclim.org, Fick and Hijmans, 2017) by extracting the climate data at the location of the modern sample sites using the *raster* package in R (version 3.5-11, Hijmans et al., 2021; R Core Team, 2020). The WorldClim 2 dataset provides spatially interpolated gridded climate data aggregated from weather stations as temporal averages between 1970-2000 (Fick and Hijmans, 2017). We used monthly average temperature data to extract the mean T_{July} and the "bioclimatic variables" bio1 (T_{ann}) and bio12 (P_{ann}).

Reviewer comment: (4) *L385: crucially, this manuscript does not describe a fossil pollen data set, but a data set of temperature and precip*

Response: We clarified this in the text.

Reviewer comment: (5) *L402-404: this seems a somewhat dangerous statement. Are the two reconstructions really independent?*

Response: We now refer to our tailoring approach where we target on the independent reconstruction to temperature and precipitation.

Reviewer comment: (6) *Why is the x axis of figure 6 on a log scale?*

Response: We now provide reconstruction on normal time-scale for the last 30 ka.

Reviewer comment: (7) *Whilst glancing through the code I missed the significance testing and the CCA. (But thumbs up for sharing the code.)*

Response: We used standard packages in climate reconstruction and reconstruction assessment. We decided to provide the code for the reconstruction in particular to show how the tailoring-approach is implemented which is methodologically new.

Response to comments of Referee #2 (P.J. Bartlein)

1. General comments

Reviewer comment: (1) *This paper describes a set of pollen-based climate reconstructions for the Northern Hemisphere from the LGM to present. The paper is obviously one of three, one describing the pollen data (Herzschuh et al., submitted, which I couldn't find), another describing the chronology (Li et al., 2022, ESSD-Disc), and this one, describing the reconstructions. There are obvious redundancies among the papers, and I think readers and potential users of the data will find it frustrating to have to track down three papers.*

Response: We revised the entire paper. We now refer to the other papers (which became published in the meantime) and kept redundancies only where necessary to understand the dataset and R.

Reviewer comment: (2) *Overall, the paper is not that well organized, with motivations for some of the analyses (e.g. CCA) not appearing until the results section (Section 4, titled "Dataset assessment"), and tutorial material on the nature of pollen data as a palaeo-archive appearing in the discussion, as opposed to the introduction (and presumably also in the first paper of the series, which, with good cross-referencing among the papers, would make it superfluous here). Perhaps this disorganization arose in parting-out the papers.*

Response: We restructured parts of the manuscript. The section about the reconstruction methods are now separated in climate reconstruction and the indication which additional measures were calculated as quality measures. In the results now one-by-one the results of the dataset quality measures are presented.

2. Major issues

Reviewer comment: (1) *Why were January and annual temperature and annual precipitation chosen as the targets for reconstruction? A more appropriate set of climate variables might be those that mechanistically control vegetation like winter cold, summer warmth, and moisture stress. A lot of the paper is devoted to dealing with the obvious correlation between annual temperature and precipitation, but it is never actually established why this is an issue.*

Response: We selected the target variables a priori. We reconstructed T_{July} and P_{ann} because they (or variables that are highly correlated with these variables) are known as major climate variables driving vegetation composition (e..g Cao et al., 2013) and Tann because this is used in many syntheses studies and data-model comparisons.

Reviewer comment: (2) *What was the role of the canonical correlation analysis? To simply explore the data perhaps, but in fact it represents an alternative reconstruction approach. In any case, it's neighther clear what the purpose of the analysis is, nor are the results fully explained.*

Response: We applied CCA to the modern calibration set to explore the modern relationship between the pollen spectra and the climate variables. High values of lambda1 (explained variance of the constrained) vs lambda2 (the unconstrained axis) is commonly considered as a measure how well the target environmental variable is related to the modern pollen data set (e.g. Juggins 2013). We provide this information for users of our data set. They can use it make decisions which sites should be include in their analyses. We explained the rationale in the method part and added further discussion in the discussion part.

New text Method part: We applied a Canonical Correlation Analysis (CCA) to the modern training dataset in order to explore the modern relationship between the pollen spectra and the climate variables and to infer the explained variance in the modern dataset by using the cca function in the vegan R-package (version 2.5-7, Oksanen et al., 2020). The ratio between constrained ($\lambda 1$) and unconstrained ($\lambda 2$) explained variance was determined for all modern training datasets. High values of $\lambda 1$ vs $\lambda 2$ are commonly considered as a measure how well the target environmental variable is related to the modern pollen data set (e.g. Ter Braak, 1988; Juggins 2013).

New Text Discussion part: We a priori selected T_{July}, T_{ann} and P_{ann} as target variables for our reconstructions. However, we provide λ_1/λ_2 (i.e. explained variance of the climate variable in the modern pollen data set relative to the variance explained by the unconstrained first axis; ter Braak, 1988), a commonly used proxy for the assessment of reconstructions. The higher λ_1/λ_2 in the spatial modern dataset higher the chance that this target climate variable has also impacted vegetation over time and is thus reflected in the variation of the fossil pollen dataset. As a rule of thumb a ratio of 1 is considered to indicate reliable reconstructions (Juggins, 2012) though useful reconstruction may also be obtained from datasets with lower values. As expected, maps of RMSEPs reveal similar spatial pattern as the results of constrained ordination. Our results indicate that in particular calibration sets from Europe have low ratios and a high RMSEP for all climate variables (despite we have a high number of modern samples), likely related to the human impact on the modern and fossil data. Some areas that are known for its sensitivity to precipitation e.g. Eastern Asia show low RMSEPs as expected for P_{ann} but on the other hand show a low sensitivity to T_{ann} and T_{July}.

Reviewer comment: (3) *The two reconstruction approaches, weighted-averaging – partial-least-squares (WA-PLS) and the modern analogue technique (MAT) may be frequently applied, but they are not without issues themselves. WA-PLS, as is the case with some other methods, tends* *to "compress" the reconstructions toward the center of the distribution of the climate data (see Liu et al, 2020, Proc. Royal Soc. A, https://doi.org/10.1098/rspa.2020.0346). This will reduce the amplitude of the time series of the reconstructions. MAT suffers from the no-analogue problem, typically diagnosed by looking at the dissimilarities. The performance of the two approaches are examined in Fig. 3, but there is no attempt to account for the obvious spatial patterns.*

Response: This manuscript provides and describes the dataset of the reconstruction and in addition of quality measures. Depending on the purpose of the studies that the users implement they can account for potential biases taking our quality measures as guidance.

Reviewer comment: (4) *A number of the analysis steps are not explained much at all, with the results just briefly described before moving on. In particular, the significance testing in Section 4.2 isn't fully explained: What is the "take-home message"? What does this analysis say about the usefulness of the reconstructions.*

Response: See our response before. Whether or not our reconstructions are useful depends much on the purpose of the study that the user of the data will implement. With our quality measures, they can make their decisions based on common numerical quality measures.

Reviewer comment: (5) *The results are described in terms of mid-Holocene minus present (1.5 to 0.5 ka) long-term mean differences, and some unusual time series plots, but there is no attempt to assess the reasonableness of the reconstructions with respect to paleoclimatic first principles or to compare them with simulations or independent observations.*

Response: This manuscript describes the data set and provides measures to assess the data quality. The full assessment of data e.g. by comparing it to other data sets is beyond the objective of this manuscript.

3. Specific comments

Reviewer comment: (1) *line 62: "climate proxy synthesis studies". Do you mean "syntheses of climate reconstructions" or "syntheses of climate proxies" (i.e. the pollen data)? It's the former that can be directly compared with climate-model output.*

Response: We followed your suggestion and changed the phrasing in the text.

Reviewer comment: (2) *line 71: "The evaluation of climate model outputs…" It's actually the climate models that are being evaluated in data-model comparisons (of simulations and observations or reconstructions).*

Response: Revised.

New text: The evaluation of climate observation data and/or the output of simulations or reconstructions, which can be used for data-model comparisons with climate model outputs, is essential for model improvements (Eyring et al., 2019).

Reviewer comment: (3) *line 73: "strong changes in the climate driver" Are you alluding to changes in GHGs during the instrumental record? Changes in insolation, ice-sheet distribution and size, and GHGs between the LGM and present are much larger. For example, the companion CMIP experiment to the LGM is the 4xCO2 experiment. CO2 has yet to double from pre-industrial levels yet.*

Response: Revised.

New text in the introduction: The comparison of climate model outputs with climate data is essential for model improvements (Eyring et al., 2019). The extratropical Northern Hemisphere is of particular interest because it is known for complex spatial and temporal temperature and precipitation patterns. However, the period for which instrumental observations are available is only of limited use to validate simulations in particular when assessing climate response to natural climate drivers because it is too short and because it is impacted by human-induced greenhouse gas forcing. . -. Climate proxy data derived from natural archives are therefore of great value.

Reviewer comment: (4) *line 74: "The extratropical Northern Hemisphere … complex spatial and temporal … patterns." Well, yes, but it's also where most of the pollen data is from. I don't think you need to motivate focusing on the Northern Hemisphere extratropics.*

Response: We shifted this sentence. So the context became a little different and it reads like an introduction sentence now.

Reviewer comment: (5) *line 90: "Regarding the prevalence…". Just say "Pollen data from … have been used…"*

Response: We followed your suggestion and changed the phrasing in the text.

Reviewer comment: (6) *line 94: "high resolution". Temporal? Spatial? Also, the last millennium is part of the Holocene, and the late-Quaternary, so you might get some push-back from dendroclimatologists about this notion.*

Response: Deleted high-resolution.

Reviewer comment: (7) *line 102: delete "the large" (I think we know extratropical Asia is large area.)*

Response: We followed your suggestion and deleted "the large" in the text.

Reviewer comment: (8) *line 103: Whitmore et al. (2005) describes the modern pollen (and climate) data set for North America, not (paleo) precipitation reconstructions.*

Response: We refer to a different reference.

Reviewer comment: (9) *line 108: If "Herzschuh et al., submitted" is "LegacyPollen 1.0: A taxonomically harmonized global…" then how is that different from this paper (and the data sets on Zenodo)? Does it describe just the fossil-pollen data, or the modern data set too?*

Response: This does describe the fossil pollen data set with focus on the harmonization of the taxa and spatio-temporal coverage.

Reviewer comment: (10) *line 110: "Li et al., 2022). So there are three papers, 1) the pollen data set, 2) new chronologies, and 3) this paper, right? Why not just say that?*

Response: We clarified in the text, that this study is the third part of interconnected studies.

New text: In a recent effort, we synthesized and taxonomically harmonized pollen records available in the Neotoma Paleoecology Database (Williams et al., 2018) and additional records from China and Siberia (Cao et al., 2013 and 2020) into a global Late Quaternary fossil pollen dataset (LegacyPollen 1.0; Herzschuh et al., submitted) and revised all chronologies of those records using a Bayesian approach that allows for the inference of temporal uncertainties (LegacyAge 1.0; Li et al., 2022). Here, in the third part of interconnected studies, we present the pollen-based reconstruction of mean July temperature (T_{July}), mean annual temperature (T_{ann}) and annual precipitation (P_{ann}) from these 2594 records from the Northern Hemisphere using WA-PLS and MAT.

Reviewer comment: (11) *line 116: Why reconstruct temperature and precipitation, as opposed to climate variables that are mechanistically related to vegetation?*

Response: We assume that the very specific variable that drives vegetation change at each site is mostly not known and might change temporarily. However, at most sites in the Northern Hemisphere extratropics, this or these specific variable(s) are typically highly correlated to either T_{July} or annual precipitation which we selected for reconstruction. In numerous studies of modern pollen-environmental relationships, this has been confirmed. These variables and, Tann in addition, are also typically used by synthesis studies and proxy-model comparison studies. Accordingly, we selected these three variables. However, our dataset provides more climate variables and the fossil and modern pollen data, so further customized reconstructions could be implemented using our data and coding environment.

Reviewer comment: (12) *line 136: "For consistency with the amount (number?) of taxa…". This needs to be a little better explained. Why 70 taxa (except for tradition)?*

Response: Explanation added.

New text in method part: We restricted the reconstruction to the 70 most common taxa on each continent to reduce computational power after making sure that higher taxa number would not substantially improve model statistics.

Reviewer comment: (13) *line 147: "2000 km radius". Why 2000 km?*

Response: We fixed the radius to 2000 km, instead of 1500 km as suggested from a study in Eastern Asia by Cao et al. (2017), because the modern dataset density is rather low in northern Asia.

Reviewer comment: (14) *line 150: "metrics". Meaning something other than just the squared-chord distance?*

Response: Revised.

Reviewer comment: (15) *line 151: "square-root transformed pollen percentages". It might be worth pointing out that the same transformation is embedded in the use of the squared-cord distance dissimilarity measure in the MAT approach.*

Response: Revised.

New Text: For all analyses square-root percentages were used if not indicated otherwise.

Reviewer comment: (16) *line 156: "co-variation". Why is this an issue? It might be the case that covariation among predictands wouldn't be an issue if they were mechanistically related to vegetation, as in the case of variables like MTCO and GDD (Wei, et al., 2020, Ecology http://dx.doi.org/ 10.17864/1947.194.*

Response: Explanation added.

New text: In addition to the classic WA-PLS reconstruction, we also use provide WA-PLS_tailored, to address the problem that co-variation of climate variable today in space is directly transferred to the reconstruction because we assume that spatial relationships of in particular temperature and climate is mechanistically not necessarily linked to temporal relationships. We assume that this can be reflected in the composition of the plant assemblages because different taxa have optima in temperature and precipitation and might therefore occur in different co-occurrences and abundances.

Reviewer comment: (17) *line 161: "… partialling out the respective other variable". Please explain.*

Response: Explanation added.

New text: The reconstructed climate parameters were tested as introducing the environmental variable as a single variable in a runs, as well as with partialling out the explained variance in the pollen data by the respective other variable.

Reviewer comment: (18) *line 161: "We applied a Canonical Correlation Analysis…". What were the community, constraining, and conditioning matrices in this analysis? More to the point, what was the objective of this analyisi?*

Response: Explanation added.

New text: "We applied a Canonical Correlation Analysis (CCA) to the modern training dataset in order to explore the modern relationship between the pollen spectra and the climate variables and to infer the explained variance in the modern pollen dataset by the target climate variables…."

Reviewer comment: (19) *line 164: "the ratio … was determined…". Why and for what purpose?*

Response: Explanation added.

New text: The ratio between constrained ($\lambda 1$) and unconstrained ($\lambda 2$) explained variance was determined for all modern training datasets used for climate reconstructions. High values of $\lambda 1$ vs $\lambda 2$ are commonly considered as an indicator measure that how well the target environmental variable is strongly related to the variation in the modern pollen data set (e.g. Juggins 2013).

Reviewer comment: (20) *line 191: Define "RMSEP" on first use in the text.*

Response: Done.

New text: Reconstruction error is provided as root mean square errors (RMSE) derived from the output in the *MAT* and *WAPLS* functions. Model errors of WA-PLS and MAT are reported as root mean square error of prediction (RMSEP) derived from leave-one-out cross validation.

Reviewer comment: (21) *lines 190-220: What accounts for the spatial variations in RMSEPs? Data density? Data quality (of both the pollen and climate data)? Confounding environmental factors?*

Response: We added some discussion in the discussion part.

New text: For fossil pollen records in areas with an insufficient coverage of modern surface pollen samples (e.g., Central Asia or Western Siberia), it might be difficult to create a calibration dataset that maps the required variety of environmental and climatic gradients and therefore find enough modern analogues for reconstructions with a classification approach such as MAT. This is indicated by the high RMSEPs as percentages of gradient length in these areas.

As expected, maps of RMSEPs reveal show similar spatial pattern as the results of constrained ordination. Our results indicate that in particular calibration sets from Europe have low ratios and a high RMSEP for all climate variables (despite we have a high number of modern samples), likely related to the human impact on the modern and fossil data. Some areas that are known for its sensitivity to precipitation e.g. Eastern Asia show low RMSEPs as expected for Pann but on the other hand show a low sensitivity to Tann and T_{July}.

Reviewer comment: (22) *line 221: "significance test". Of what? What hypothesis does the Telford and Birks test address?*

Response: Explanation added.

Reviewer comment: (23) *line 241: "we subtracted those means from every record". There are two mean values (6.5 to 5.5 ka and 1.5 to 0.5 ka), and "every record" implies to me the whole data set, LGM to present. Aren't you just looking at the difference between those two mean values? (And why 1.5 to 0.5 ka?)*

Response: See response to reviewer 2. We now provide results for 6 minus 1 ka including error estimates. 1 ka was chosen as example as more reocrds were available for 1 ka compare with records covering the last 100 years representin 0 ka. The map here is only an example to illustrate the results.

Reviewer comment: (24) *line 243: "warmer and drier" Than what? (Which time period is the warmer and drier one?). Throughout this paragraph, the sense of change in climate has to be made explicit. For parallelism, you should adopt a standard way of expressing the changes, e.g. "warmer than present in the mid-Holocene" or "cooling from the mid-Holocene to present" but don't mix states and trend.*

Response: Revised.

New text: To illustrate analyzing the temporal variation between mid and late Holocene, we calculated means of all three climate variables for the time slices at 6 and 1 ka periods between 6.5 and 5.5 ka BP and between 1.5 and 0.5 ka BP and subtracted those means from every record in order to evaluate the changes between the reconstructed mid-Holocene conditions and those of modern times. Differences between these time periods reveal warmer and drier conditions during mid-Holocene compared with late Holocene conditions especially in Eastern North America but also in Central and Northern Europe.

Reviewer comment: (25) *line 250: What's a "more gradual pattern"?*

Response: We revised this paragraph.

Reviewer comment: (26) *Figure 6: What exactly is plotted here? Why use a log age axis? An alternative depiction of all of the reconstructions, and their temporal and latitudinal varliations would be a Hovmöller diagram.*

Response: This manuscript describes a dataset of reconstructed climate variables and explores the quality of the data while assessing past climate change, as could be obtained with Hovmöller diagrams is not in the focus of this study.

Reviewer comment: (27) *Figure 8: I guess we're supposed to see that there are more correlation coefficients between temperature and precipitation close to zero in the "tailored" analyses. I've got nothing against violin plots, but I think a standard histogram would work a lot better.*

Response: We prefer to stay with a violin plot.

Reviewer comment: (28) *line 301+: What are the implications of these statistics and their spatial patterns?*

Response: No added further discussion on the potential limitation of the data set as indicated by the quality measures. See discussion part.

Reviewer comment: (29) *lines 315-343: This tutorial on pollen data, chronologies, etc. should probably be in the introduction, not the discussion.*

Response: We revised the entire discussion and always related comments to our data set and tried to avoid general "tuturial" comments.

Reviewer comment: (30) *line 378: "numerical mechanisms … reduce the reliability" Please explain.*

Response: We rephrased the entire section.

Reviewer comment: (31) *line 410: "TraCE 21k" is a transient experiment. The model used was CCM 3.*

Response: Revised.

4. Code and data

Reviewer comment: *I was able to run the example R code without problems. However, the data sets, described and labelled (via the extension) as .csv files (comma-separated values), are instead tab-separated files, which usually have the extension ".tab", or sometimes ".txt". This situation prevents a user from getting a quick look at the data using a spreadsheet program.*

Response: Csv-files can use other separators than comma (also tab-separator). When opening in a spreadsheet program like Excel, it would be necessary to import the file and specify the separator rather than open it by double-click. However, we will provide the dataset now in addition as txt file on Pangaea.