https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

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

Ulrike Herzschuh[1,2,3], Thomas Böhmer[1], Chenzhi Li[1,2], Manuel Chevalier[4,5,6], Anne Dallmeyer[5], Xianyong Cao[1,7], Nancy H. Bigelow[8], Larisa Nazarova[1,9], Elena Y. Novenko[10,11], Jungjae Park[12,13], Odile Peyron[14], Natalia A. Rudaya[15,16], Frank Schlütz[17,18], Lyudmila S. Shumilovskikh[18], Pavel E. Tarasov[19], Yongbo Wang[20], Ruilin Wen[21,22], Qinghai Xu[23], Zhuo Zheng[24,25]

[1] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Telegrafenberg A45, 14473 Potsdam, Germany

[2] Institute of Environmental Science and Geography, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[3] Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[4] Institute of Geosciences, Sect. Meteorology, Rheinische Friedrich-Wilhelms-Universität Bonn, Auf dem Hügel 20, 53121 Bonn, Germany

[5] Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

[6] Institute of Earth Surface Dynamics IDYST, Faculté des Géosciences et l'Environnement, University of Lausanne, Batiment Géopolis, 1015 Lausanne, Switzerland

[7] Alpine Paleoecology and Human Adaptation Group (ALPHA), State Key Laboratory of Tibetan Plateau Earth System, and Resources and Environment (TPESRE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, 100101 Beijing, China

[8] Alaska Quaternary Center, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA

[9] Kazan Federal University, Kremlyovskaya str. 18, 420008 Kazan, Russia

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

[10] Lomonosov Moscow State University, Faculty of geography, Leniskie gory 1, 119991 Moscow, Russia

[11] Department of Quaternary Paleogeography, Institute of Geography Russian Academy of Science, Staromonrtny lane, 29, 119017, Moscow, Russia

[12] Department of Geography, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

[13] Institute for Korean Regional Studies, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

[14] Institut des Sciences de l'Evolution de Montpellier, Université de Montpellier, CNRS UMR 5554, Montpellier, France

[15] PaleoData Lab, Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Pr. Akademika 36 Lavrentieva 17, 630090 Novosibirsk, Russia

[16] Biological Institute, Tomsk State University, Pr. Lenina, 26, Tomsk, 634050, Russia

[17] Lower Saxony Institute for Historical Coastal Research, D-26382 Wilhelmshaven, Germany

[18] Department of Palynology and Climate Dynamics, Albrecht-von-Haller Institute for Plant Sciences, University of Göttingen, Untere Karspüle 2, 37073 Göttingen, Germany

[19] Freie Universität Berlin, Institute of Geological Sciences, Palaeontology Section, Malteserstrasse 74-100, Building D, 12249 Berlin, Germany

[20] College of Resource Environment and Tourism, Capital Normal University, 105 West 3rd Ring Rd N, 100048 Beijing, China

[21] Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, 19 Beitucheng West Road, Chaoyang District, 100029 Beijing, China

[22] CAS Center for Excellence in Life and Paleoenvironment, 100044 Beijing, China

[23] College of Geographical Sciences, Hebei Normal University, 050024 Shijiazhuang, China

[24] Guangdong Key Lab of Geodynamics and Geohazards, School of Earth Sciences and Engineering, Sun Yat-sen University, 519082 Zhuhai, China

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

[25] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), 519082 Zhuhai, China

Correspondence: Ulrike Herzschuh (Ulrike.Herzschuh@awi.de)

Abstract. Here we describe the LegacyClimate 1.0, a dataset of the reconstruction of mean July temperature (T_{July}), mean annual temperature (T_{ann}), and annual precipitation (P_{ann}) from 2594 fossil pollen records from the Northern Hemisphere spanning the entire Holocene with some records reaching back to the Last Glacial. Two reconstruction methods, the Modern Analogue Technique (MAT) and Weighted-Averaging Partial-Least Squares regression (WA-PLS) reveal similar results regarding spatial and temporal patterns. To reduce the impact of precipitation on temperature reconstruction and vice versa, we also provide reconstructions using tailored modern pollen data limiting the range of the corresponding other climate variable. We assess the reliability of the reconstructions using information from the spatial distributions of the root-mean squared error of prediction and reconstruction significance tests. The dataset is beneficial for climate proxy synthesis studies and to evaluate the output of climate models and thus help to improve the models themselves. We provide our compilation of reconstructed T_{July}, T_{ann}, and P_{ann} as open-access datasets at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.930512; Herzschuh et al., 2021). R code for the reconstructions is provided at Zenodo (https://doi.org/10.5281/zenodo.5910989; Herzschuh et al., 2022), including harmonized open-access modern and fossil datasets used for the reconstructions, so that customized reconstructions can be easily established.

1 Introduction

The evaluation of climate model outputs using climate data is essential for model improvements (Eyring et al., 2019). However, the period for which observations are available is only of limited use to validate simulations because it is short and characterized by strong changes in the climate driver. Climate proxy data derived from natural archives are therefore of great value. The extratropical Northern Hemisphere is of particular interest because it is known for complex spatial and temporal temperature and precipitation patterns.

Previous proxy-based climate inferences have contributed to major debates about Holocene climate change. For example, while simulations indicate a gradual warming of the Holocene, temperature proxy data syntheses rather support a mid-Holocene optimum which resulted in the “Holocene conundrum”

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

debate (Liu et al., 2014). Qualitative proxy-based inferences indicate that the mid-Holocene in the Northern Hemisphere mid-latitudes was rather dry and warm compared with present-day in agreement with modeling outputs (Routson et al., 2019). Also, quantitative precipitation reconstructions from Eastern and Central Asia unveiled the complex monsoon-westerlies interactions (Chen et al., 2019; Herzschuh et al., 2019).

Fossil pollen records are well-established in their use as a palaeoecological and palaeoclimatological proxy and of great value as indicators of past environmental and climatic change for many decades. Considerable efforts have been made to establish regional, continental and even global data repositories like the North American Pollen Database (http://www.ncdc.noaa.gov/paleo/napd.html), the European Pollen Database (http://www.europeanpollendatabase.net) and the Neotoma Paleoecology Database (https://www.neotomadb.org; Williams et al., 2018). Regarding the prevalence of pollen archives across multiple environmental settings such as lakes, wetlands, or marine sediments, fossil pollen records are widely used to quantitatively reconstruct past vegetation and climate variables (Birks, 2019; Chevalier et al., 2020). Pollen data are the only land-derived proxy data that have sufficient temporal and spatial coverage to allow for high-resolution climate model evaluation of the late Quaternary period. A number of methods have been proposed for making pollen-based climate reconstructions (Chevalier et al., 2020): among them, classification approaches like the Modern Analogue Technique (MAT) or regression approaches like Weighted-Averaging Partial-Least Squares regression (WA-PLS) are most commonly used.

For temperature reconstruction time-series, several broad-scale syntheses exist; however, either they originate from different proxies (Kaufman et al., 2020a and 2020b) or are restricted to certain continents or regions (Mauri et al., 2015; Marsicek et al., 2018; Routson et al., 2019). Temperature reconstructions from the large extratropical Asia are mostly lacking. Precipitation syntheses are available from Europe (Mauri et al., 2015), North America (Whitmore et al., 2005) and China and Mongolia (Herzschuh et al., 2019) but, hitherto, no global or hemispheric syntheses of quantitative precipitation changes are available for the Holocene.

In a recent effort, 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) were synthesized and taxonomically harmonized (Herzschuh et al., submitted). Furthermore, all chronologies of these records were recently revised using a Bayesian approach that allows for the inference of temporal uncertainties

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

(Li et al., 2022). Here 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.

2 Methods

2.1 Input data

The objective of this study is to create a dataset of quantitative reconstructions of T_{July}, T_{ann} and P_{ann} spanning the Holocene from a set of fossil pollen records. We used fossil data from the Neotoma Paleoecology Database (Williams et al., 2018; https://www.neotomadb.org; downloaded in July 2020), a dataset from Eastern and Central Asia (Cao et al., 2013; Herzschuh et al., 2019) and a dataset from Northern Asia (Cao et al., 2020). The harmonized dataset is stored on PANGAEA (LegacyPollen 1.0) and presented in Herzschuh et al. (submitted). Ages were taken from age-depth models presented in Li et al. (2022), who recently provided a set of harmonized chronologies under the "LegacyAge 1.0" framework, and applied to our fossil pollen synthesis. A modern pollen training dataset comprised of 15,379 sites includes datasets from Eurasia (EMPD1, Davis et al. 2013; EMPD2, Davis et al. 2020; Herzschuh et al., 2019; Tarasov et al., 2011) and North America (Whitmore et al., 2005). In order to reduce inconsistencies in pollen identification, the modern and fossil pollen datasets were taxonomically harmonized: major tree and shrub pollen were merged to genus level and most of the herbaceous taxa (except the most common ones such as *Artemisia*, *Thalictrum* or *Rumex*) to family level. We excluded aquatic pollen (with the exception of *Cyperaceae*), spores from ferns and fungi, as well as algae and calculated pollen percentages on the basis of the total number of terrestrial pollen grains. The site specific T_{ann}, T_{July}, P_{ann} were derived from WorldClim 2 (spatial resolution of 1 km, 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).

We compiled the fossil data into four sub-continental datasets for Eastern North America (<105°W; Williams et al., 2000), Western North America, Europe and Asia. For consistency with the amount of taxa in the North American training dataset, the fossil datasets were reduced to the 70 most common

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022


Open Access Earth System Science Data Discussions

taxa on the respective sub-continents, according to Hill´s N2 diversity index (i.e., the effective number of occurrences of a species in the dataset; Hill, 1973).

2.2 Reconstruction methods

Our reconstruction approach included MAT (Overpeck et al., 1985) and WA-PLS (ter Braak and Juggins, 1993) by applying the *MAT* and *WAPLS* functions from the *rioja* package (version 0.9-21, Juggins, 2019) for R (R Core Team, 2020) on our Northern Hemispheric fossil pollen synthesis. For each fossil location, we calculated the geographic distance between each modern sampling site and the fossil pollen record using the *rdist.earth* function from the *fields* R-package (version 10.3, Nychka et al., 2020) and selected a unique calibration set from modern sites within a 2000 km radius. For the reconstruction with MAT, we used the original pollen percentages of the selected fossil pollen taxa, looking for 7 analogues between the pollen data and the selected calibration dataset. The dissimilarity between the fossil samples and the modern pollen assemblages was determined by squared-chord distance metrics (Simpson, 2012; Cao et al., 2014). For the reconstruction with WA-PLS, we used the square-root transformed pollen percentages in a leave-one-out cross-validation approach (Cao et al., 2014). In addition to the classic WA-PLS reconstruction, we provide WA-PLS_tailored. To reconstruct T_{July} we "tailored" our modern training dataset with respect to the P_{ann} range. For this purpose, we identified the range of the reconstructed P_{ann} and extended it by 25% to both ends of the modern P_{ann} range in order to reduce the influence of P_{ann} on T_{ann} and T_{July} reconstruction due to co-variation. We applied the same method to the reconstruction of P_{ann}. T_{ann} and T_{july} were tailored by P_{ann}; P_{ann} was tailored by T_{July} and, additionally, by T_{ann} (illustrated for an example in Appendix Fig. 1). A statistical significance test (Telford and Birks, 2011) was performed for the reconstruction by using the *randomTF* function in the *palaeoSig* R-package (version 2.0-3, Telford, 2019). The reconstructed climate parameters were tested as single variables, as well as with partialling out the respective other variable. We applied a Canonical Correlation Analysis (CCA) to the modern training dataset in order 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 (λ_1) and unconstrained (λ_2) explained variance was determined for all modern training datasets.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

3 Dataset description LegacyClimate 1.0: input data, reconstructions and reconstruction model statistics

LegacyClimate 1.0 provides pollen-based reconstructions and sample-specific reconstruction errors of T_{ann}, T_{July} and P_{ann} for 2594 fossil pollen records (i.e., a total of 146,067 single pollen samples) from three reconstruction methods (WA-PLS, WA-PLS_tailored, MAT). Furthermore, we provide the method-specific model metadata and statistics for each record and each climate variable (Table 1). To ease data handling, the dataset files are separated into Western North America, Eastern North America, Europe and Asia.

Table 1. Structure and content of the LegacyClimate 1.0 data with details about the information contained in the input, datasets, in the reconstructions and the reconstruction model statistics.

Datasets	Content
Input datasets	Modern pollen dataset of 15,379 sites Modern dataset of T_{ann}, T_{July}, P_{ann} Fossil pollen data (LegacyPollen 1.0) for 2594 sites with a total of 146,067 samples
LegacyClimate 1.0: Climate reconstructions	Reconstructions and sample-specific reconstruction errors of T_{ann}, T_{July} and P_{ann} for 2594 sites using MAT, WA-PLS and WA-PLS_tailored
LegacyClimate 1.0: Reconstruction model statistics	**Site information** (Event label, Source, ID, Site name, Longitude, Latitude) **Modern pollen dataset information** (number of modern analogues, range of climate variables) **Model statistics for each site for MAT**, **WA-PLS, WA-PLS_tailored** (including r^2 observed vs. predicted, RMSEP, no. of WA-PLS components)

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

4 Dataset assessment

4.1 Spatial and temporal coverage of LegacyClimate 1.0

In total, we provide reconstructions for 2594 fossil pollen records, among them 670 records from Eastern North America, 361 records from Western North America, 1075 records from Europe and 488 Asian records (Fig. 1). The temporal coverage of the records is rather uneven: 119 and 289 records cover the periods before 30,000 years (Fig. 2) and the Last Glacial Maximum, respectively. A total of 1229, 1845, 2052 records are available for 12-11 ka, 6-5 ka BP and 2-1 ka BP, respectively.

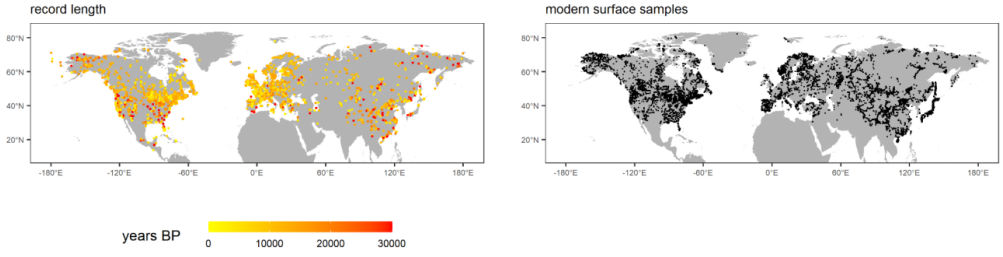


Figure 1. left: map indicating the spatial distribution and record lengths covered by the LegacyPollen 1.0 dataset (Herzschuh et al., submitted) for which climate reconstructions are provided in LegacyClimate 1.0 with a total of 2594 records; right: spatial distribution of modern pollen dataset used for reconstruction with a total of 15,379 sites.

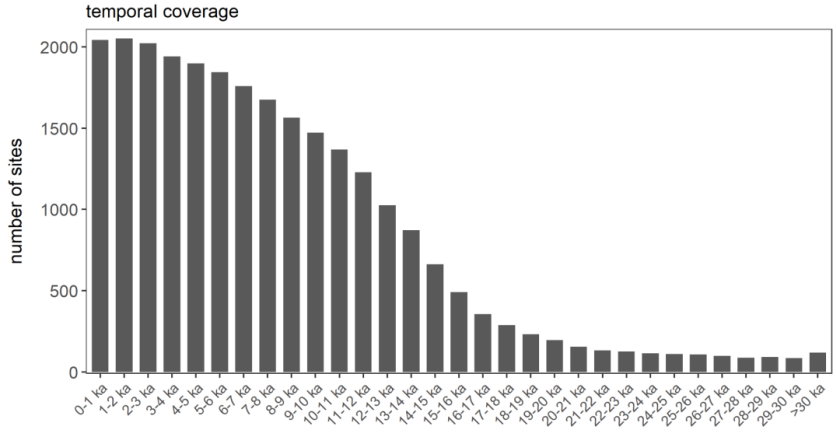


Figure 2. Number of records that cover certain millennia of the last 30 ka.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022


Open Access Earth System Science Data Discussions

4.2 Prediction errors of LegacyClimate 1.0

The mean RMSEPs and their standard deviations for T_{ann} are 1.98±0.52°C (MAT), 2.61±0.53°C (WA-PLS) and 2.24±0.61°C (WA-PLS_tailored) and mean RMSEPs as a percentage of modern T_{ann} range are 7.68±1.93% (MAT), 10.09±2.05% (WA-PLS) and 10.26±2.79% (WA-PLS_tailored). The largest mean RMSEP values are located in Central Asia in Kazakhstan, Mongolia and the north-western parts of the Tibetan Plateau and are consistent across all three reconstruction methods. Other areas with large mean RMSEP values are located in Western North America, Southern and Central Europe and south-east Asia. The smallest RMSEPs can be found along the east coast of North America. Relative to the modern temperature range, the RMSEP from this region also reveals the lowest fraction. In general, MAT has the lowest mean error fraction relative to the modern temperature range of all three methods.

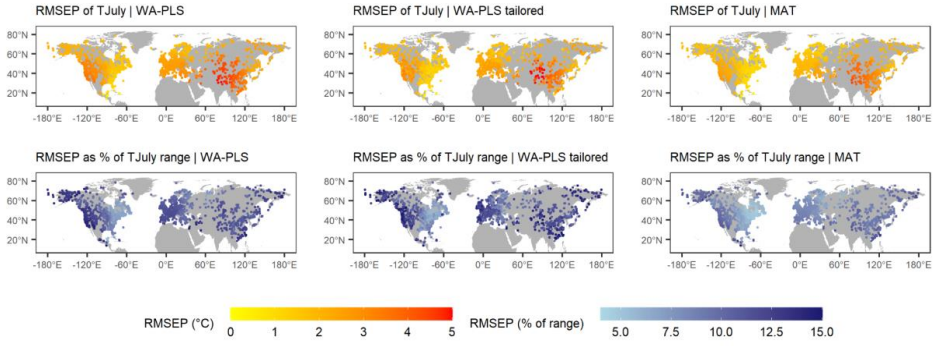
The mean RMSEPs of T_{July} are 1.90±0.63°C (MAT), 2.50±0.73°C (WA-PLS) and 2.21±0.75°C (WA-PLS_tailored) and mean percentages of T_{July} range are 8.11±1.64% (MAT), 10.71±1.94% (WA-PLS) and 10.70±2.60% (WA-PLS_tailored). Thus, they are slightly smaller than those of T_{ann} but slightly larger as a percentage of the range. The spatial patterns, however, are largely similar to those of T_{ann}.

The mean RMSEPs of P_{ann} are 176.38±51.40 mm (MAT), 244.48±75.84 mm (WA-PLS) and 232.71±98.57 mm (WA-PLS_tailored) and mean percentages of P_{ann} range are 6.78±1.48% (MAT), 9.27±1.70% (WA-PLS) and 10.26±2.67% (WA-PLS_tailored). High RMSEPs are found for Western North America, Europe and along the coastline of south-east Asia, while the lowest RMSEP values are found for Central Asia. A clear division in RMSEPs are found on the North American continent: while the western part of North America (with the exception of Alaska) has a rather high RMSEP, the eastern part of North America has a smaller RMSEP. This pattern is found for all three methods (Fig. 3).

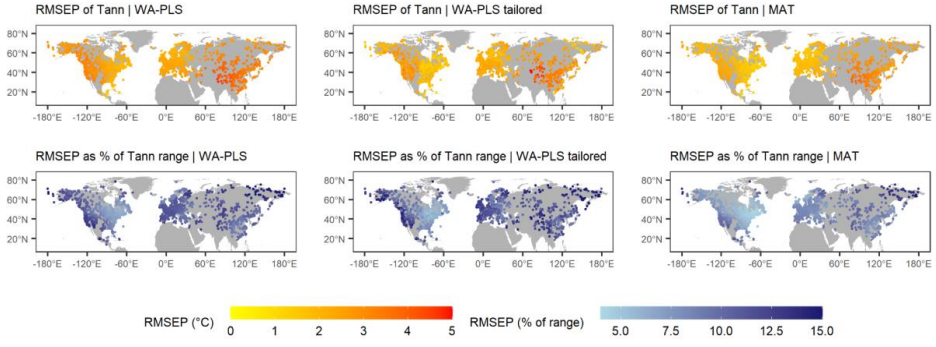
https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

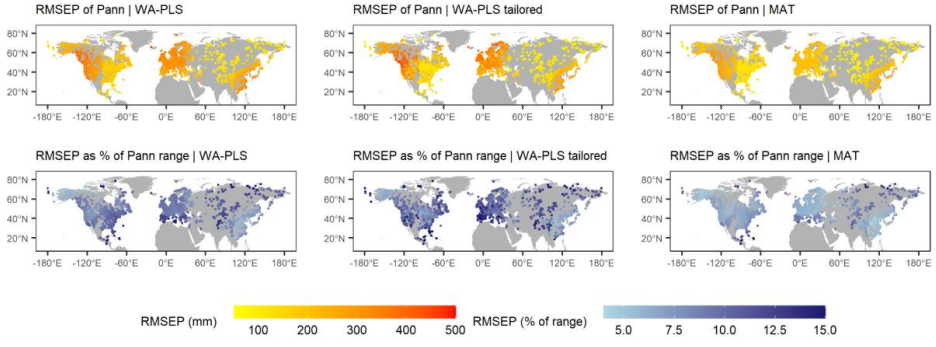
10



212



213



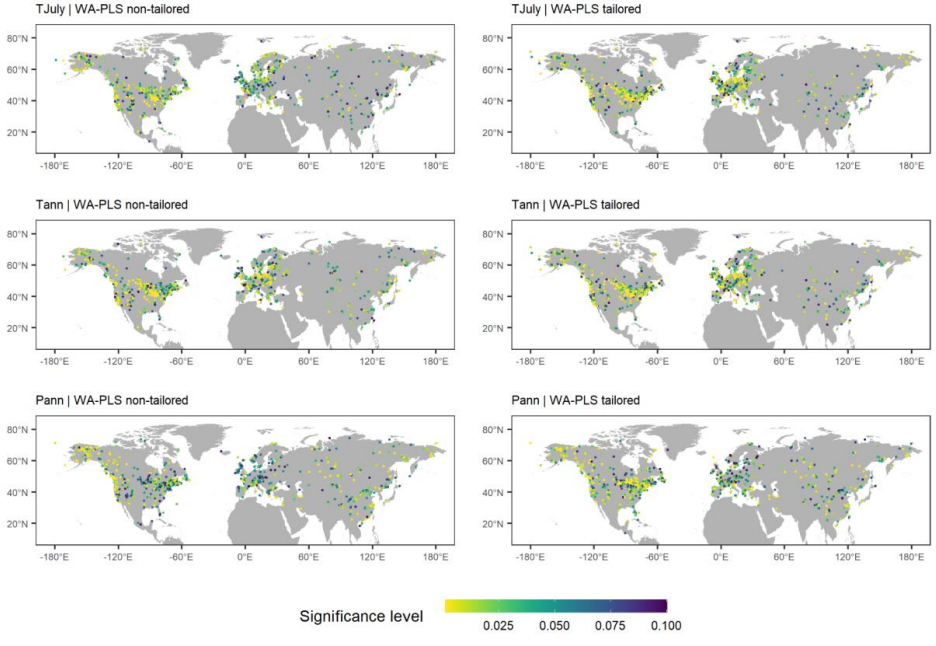
214

215 **Figure 3.** Spatial distribution of root mean squared error of prediction (RMSEP) as inferred from leave-
216 one out cross-validation presented as absolute values and as a percentage of the range of mean July
217 temperature (T_{July}), mean annual temperature (T_{ann}), mean annual precipitation (P_{ann}) in the modern
218 pollen data used for reconstruction for the three methods applied (weighted-averaging partial least

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

squares (WA-PLS), WA-PLS using a training set from within a limited climate range (WA-PLS_tailored) and modern analogue technique (MAT)).

A significance test ($p < 0.1$) according to Telford and Birks (2011) for the whole reconstructed time period was run for each record and for the reconstructions with WA-PLS and WA-PLS_tailored (Fig. 4; Table 2). The T_{July} reconstruction is significant for 30.9% (WA-PLS) and 35.2% (WA-PLS_tailored) when included as a single variable in the significance test. Partialling out precipitation as a conditional variable causes an increase in the amount of significant records to 35.5% for WA-PLS, but a decrease for WA-PLS_tailored to 33.6% of all records. For T_{ann}, 32.8% (WA-PLS) and 36.1% (WA-PLS_tailored) of all records pass the significance test when tested as a single variable. When partialling out precipitation, the amount of significant records decreases for both WA-PLS and WA-PLS_tailored. 32.1% (WA-PLS) and 33.4% (WA-PLS_tailored) of all records pass the significance test when testing P_{ann} as a single variable. In contrast to the significance tests for T_{ann}, partialling out the mean July temperature as a conditional variable increases the number of significant records for both WA-PLS and WA-PLS_tailored.



https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Figure 4. Maps showing mean July temperature (T_{July}), mean annual temperature (T_{ann}), mean annual precipitation (P_{ann}) records that passed the reconstruction significance test ($p<0.1$). Color indicates the significance level.

Table 2. Percentage of records that pass the reconstruction significance test ($p<0.1$) sensu Telford and Birks (2011).

	WA-PLS	WA-PLS_tailored
T_{July}	30.9%	35.2%
T_{July} partialling out P_{ann}	35.5%	33.6%
T_{ann}	32.8%	36.1%
T_{ann} partialling out P_{ann}	32.6%	34.1%
P_{ann}	32.1%	33.4%
P_{ann} partialling out T_{July}	34.3%	36.5%

4.3 Assessment of major temporal patterns of LegacyClimate 1.0

For analyzing the temporal variation, we calculated means of all three climate variables for the time 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 especially in Eastern North America but also in Central and Northern Europe. The overall patterns are in good agreement for all three methods but show differences on a regional scale, especially when comparing the reconstructions with WA-PLS and MAT. For T_{July}, the reconstruction with MAT shows greater temperature differences in Western North America and south-east Asia. Compared to the reconstruction with WA-PLS, there is a reduced cooling in Eastern Europe and a warming in the Western Mediterranean region and along the south-eastern Asian coastline. Comparing the reconstructions of T_{ann}, more gradual patterns are seen in the reconstruction with WA-PLS: Western North America reveals a mid-Holocene warming, while Eastern North America shows a cooling. In Europe records that report a cooling are more concentrated in the northern and western parts of the continent. In the reconstruction with MAT, Eastern North America is divided into a reported cooling in the northern part and a warming

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

in the southern part. In Western North America, there is a mixture of locations with a warming and a cooling since the mid-Holocene. In Europe, only France and Southern Scandinavia show a cooling; in Central and parts of Southern Europe, a warming can be found in the reconstructions. For large areas in North America and Europe, the reconstructions with WA-PLS suggest an increase in precipitation since 6 ka BP. A shift to drier conditions can be found along the south-eastern coastline in North America, in the Mediterranean Region and especially in south-east Asia. The reconstruction with MAT reveals a gradient from increasing precipitation in south-western Europe to decreasing precipitation in north-eastern Europe. In contrast to the reconstructions with WA-PLS, records along the south-eastern Asian coastline suggest an increase in precipitation with MAT rather than a decrease (Fig. 5).

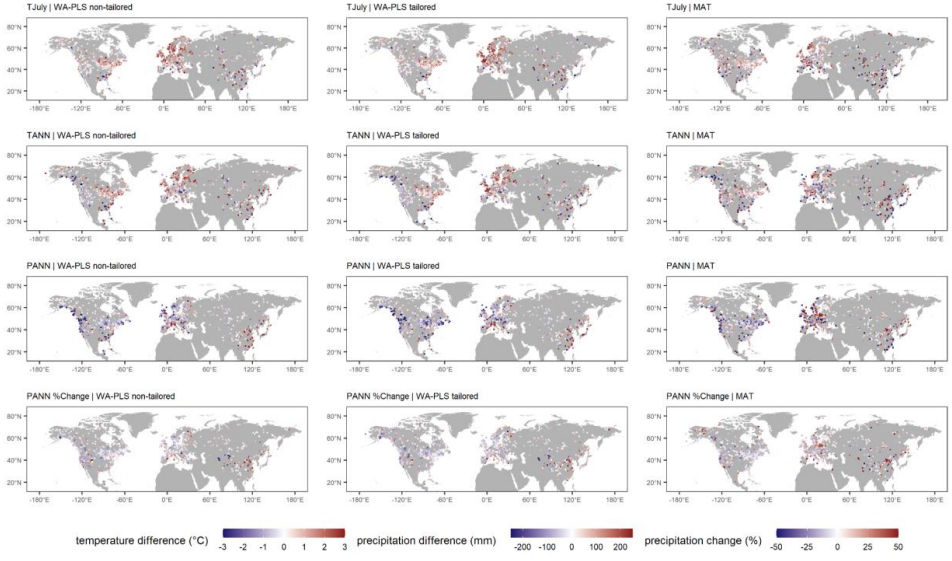


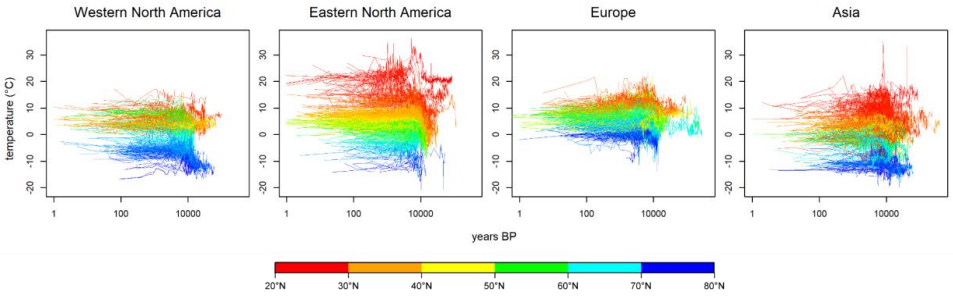
Figure 5. Difference from 6 ka to 1 ka for mean July temperature (T_{July}), mean annual temperature (T_{ann}), mean annual precipitation (P_{ann}) and P_{ann}% as reconstructed from weighted-averaging partial least squares (WA-PLS), WA-PLS using a training set from within a limited climate range (WA-PLS_tailored) and modern analogue technique (MAT).

Time-series of absolute T_{ann} reconstructions reveal temporal as well as latitudinal spatial variation on the single continents. Eastern North America and Asia show the most variation in the low latitudes. It is also Eastern North America which shows the most pronounced latitudinal gradient. In Western North America, the most variation takes place in the high latitudes, while the variation is concentrated to the

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

272 mid-latitudes in Europe. Especially in North America, the warming since the last deglaciation and the
273 beginning of the Holocene is well shown in the temporal variation of the time-series (Fig. 6).

274



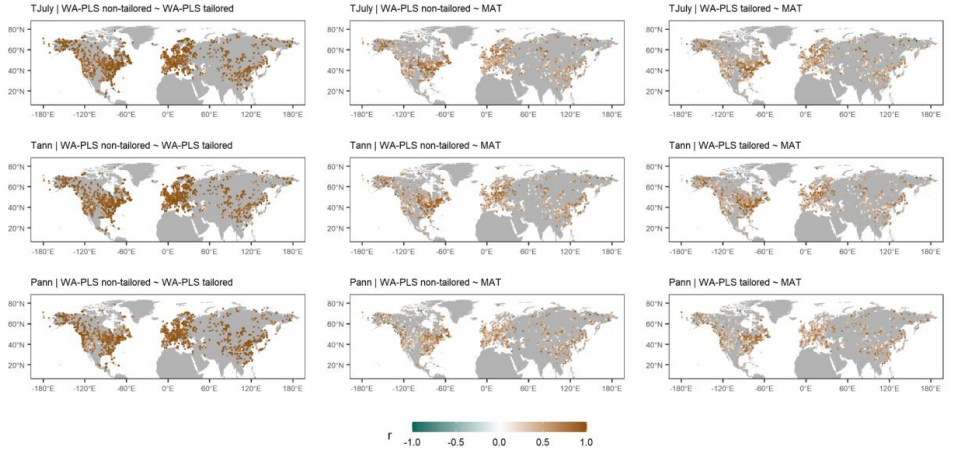
275

276 **Figure 6.** Time-series of absolute mean annual temperature (T_{ann}) reconstruction for each
277 (sub-)continent. Colors denote the latitude of record origin. Note logarithmic x-axis.

278

279 **4.4 Assessment of consistency among reconstruction methods**

280 Reconstructions with MAT are, in general, in good agreement with those derived from the WA-PLS.
281 Comparing MAT with WA-PLS, 37.3% (T_{July}), 38.9% (T_{ann}) and 30.4% (P_{ann}) of all records have a positive
282 correlation of r >= 0.6. Strong positive correlations (r >= 0.9) can mainly be identified in Eastern North
283 America, while weak correlation can be found for large areas in central North America and most of
284 Europe (Fig. 7).



285

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Figure 7. Correlation between time-series of the 3 different reconstruction methods used - weighted-averaging partial least squares using a global training set (WA-PLS), WA-PLS using a training set with a limited modern climate range (WA-PLS_tailored) and the modern analogue technique (MAT) for the three climate variables of mean July temperature (T_{July}), mean annual temperature (T_{ann}) and mean annual precipitation (P_{ann})

WA-PLS_tailored used a reduced modern training dataset (illustrated for an example in Appendix Fig. 1). 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 T_{July}, 60.7% for T_{ann} and 56.5% for P_{ann}.

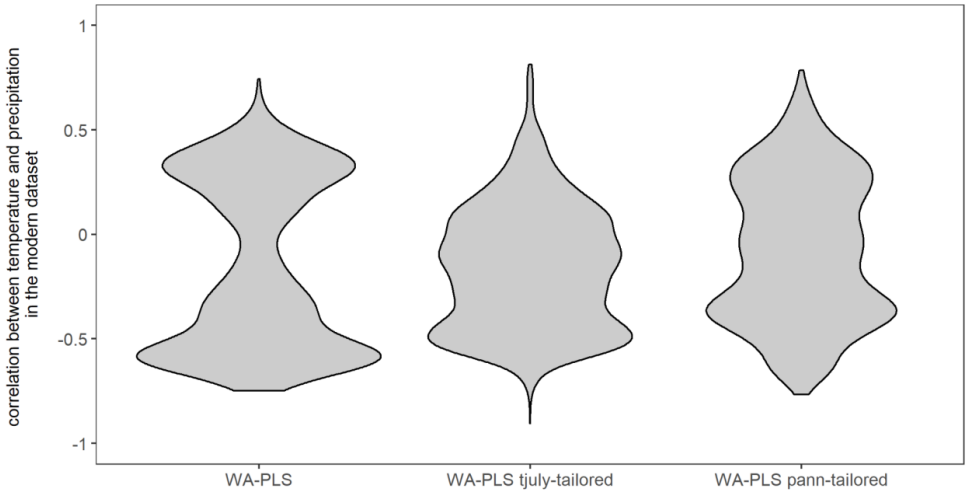


Figure 8. Violin plot of the correlation coefficients between T_{July} and P_{ann} in the 15,379 training datasets used for the reconstructions. Left: used for WA-PLS reconstructions; middle: WA-PLS tjuly-tailored (used for the reconstruction of P_{ann}); WA-PLS pann-tailored (used for the reconstruction of T_{July}).

A CCA was performed to infer the ratio between constrained and unconstrained explained variance for all modern training datasets (λ_1/λ_2) for the modern datasets used for WA-PLS and WA-PLS_tailored. Modern datasets used for WA-PLS constrained by T_{July} reveal a concentration of high ratios in Eastern North America while low ratios can be found in Central Asia. While the spatial pattern of λ_1/λ_2 constrained

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

by T_{ann} is similar, the ratios are slightly higher for T_{ann} than for T_{July}. Reconstructions for P_{ann} show low ratios in Europe and Eastern North America. Areas with high ratios are concentrated in Alaska and East Asia.

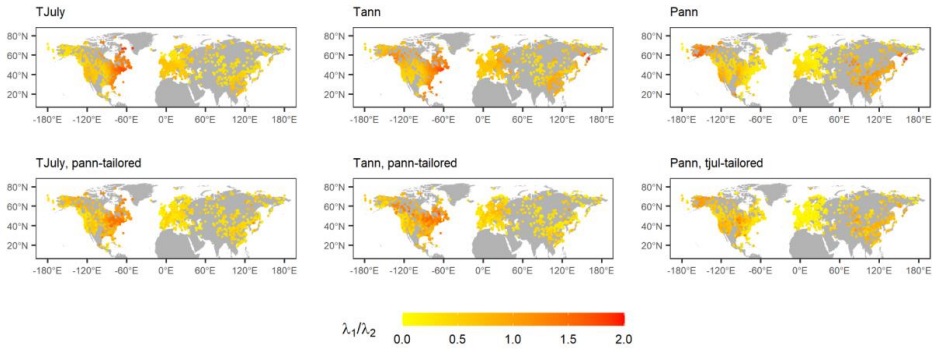


Figure 9. Maps showing λ_1/λ_2, representing the ratio of explained variance of first axis (constrained) vs. second (unconstrained) axis as revealed by applying a CCA to all modern training datasets that were used for the reconstructions. Constraining variable as well as tailoring of the dataset (see methods) is indicated in the map captions.

5 Discussion

5.1 Impact of the fossil pollen data source on LegacyClimate 1.0 quality

LegacyClimate 1.0 contains reconstructions of climate variables from fossil pollen data derived from open-access data repositories. The fossil records were derived from multiple natural archives, most commonly, assemblages from continuous lacustrine and peat accumulations (Herzschuh et al., submitted). Different sizes of lakes and peat areas result in varying sizes of pollen source areas and thus the spatial representativeness of a record, as small lakes and peatlands are considered to provide information about the (extra-)local scale, while pollen assemblages from large lakes are considered as a regional signal (Jackson, 1990; Sugita, 1993). However, such signals might be impacted by taphonomy of the record, for example pollen from azonal riverine vegetation might be over-represented in fluvially impacted pollen records.

Our dataset is based on taxonomically harmonized modern and fossil pollen datasets using a restricted number of taxa (i.e., the most common 70 taxa on each (sub-)continent). Such an approach guarantees

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

that all records are handled consistently. Although losing taxonomic information when merging taxa together into a higher taxonomic level, it also increases the possibility of matching climate analogues in the modern and the fossil datasets. However, one needs to keep in mind that species with different ecological requirements may be merged together into one genus or family, for example, *Pinus* species that are restricted to tropical or subtropical areas in China or ones that grow in boreal forests (Cao et al., 2013; Tian et al., 2017).

Along with the pollen assemblages, data repositories also provide chronological information for fossil records. The quality of such chronologies varies strongly with respect to dating methods, calibration and numerical algorithms for determining an age-depth relationship (Blois et al., 2011; Trachsel and Telford, 2017). Having accurate and precise chronologies is thus of pivotal importance for reconstructing past climate in order to identify temporo-spatial patterns and therefore in helping to evaluate climate model outputs. The advantage of the fossil pollen dataset used for the reconstruction presented here (i.e., LegacyPollen 1.0; Herzschuh et al., submitted) is that it has harmonized chronologies (LegacyAge 1.0) along with information about uncertainties as well as related metadata and scripts that allow a customized re-establishment of the chronologies (Li et al., 2022). This, for example, allows the calculation of the temporal uncertainties when presenting reconstruction uncertainties of a specific time-slice.

5.2 Modern pollen and climate data sources and LegacyClimate 1.0 quality

Palaeoclimate reconstruction methods such as MAT and WA-PLS rely on extensive collections of modern training data. Designing a robust calibration dataset from modern pollen assemblages is a crucial part of the reconstruction process. A suitable calibration dataset should cover a wide range of climatic and environmental gradients in order to represent an empirical relationship between pollen assemblages and climate (Birks et al., 2010; Chevalier et al., 2020). Like with fossil pollen records, data syntheses and repositories also exist for modern surface pollen data. Most of the records in our modern dataset were compiled from well-established pollen assemblages from North America (Whitmore et al., 2005), Eurasia (Davis et al., 2013 and 2020) and China (Cao et al., 2013; Herzschuh et al., 2019). 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

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

for reconstructions with a classification approach such as MAT. Our routine uses the modern pollen data from within a radius of 2000 km around the site of the fossil record. The information provided in the reconstruction metadata including number of modern pollen samples and ranges of reconstructed variables, allows an assessment of the modern dataset used for reconstruction.

5.3 Reconstruction method and LegacyClimate 1.0 quality

Climate reconstruction methods all have different strengths and weaknesses. MAT and WA-PLS used in this study heavily rely on extensive collections of modern assemblage data covering diverse climatic and environmental gradients and are applicable on a broad spatial scale. However, both methods may struggle with complex species responses, are sensitive to spatial autocorrelation, can only deal with a certain extent of non-analogous situations and may produce poor results in so-called “quantification deserts” (Chevalier, 2019), where fossil pollen is hardly preserved or nearby modern surface pollen samples are missing (Chevalier et al., 2020). Nonetheless, for reconstructions on a local or regional scale, MAT and WA-PLS are most commonly used in climate reconstructions. The format of the modern and fossil datasets as well as the provided scripts could also be easily adapted to apply to other reconstruction methods such as CREST, a Bayesian approach that combines presence-only occurrence data and modern climatologies to estimate the conditional response of a given taxon to a climate variable (Chevalier et al., 2014 and 2020).

Through numerous physical processes that vary with both location and time, temperature and precipitation are interconnected, especially within the extratropical regions (Adler et al., 2008; Trenberth, 2011) and thus temperature and precipitation may not be treated as independent variables. Due to the numerical mechanisms in the transfer function, the correlation between both climate variables may reduce the reliability of the reconstructions. This is especially true for regions with a temperature-moisture driven circulation system such as the East Asian Summer Monsoon (EASM) that can heavily affect precipitation patterns in certain regions (Herzschuh et al., 2019). With our tailoring approach we are able to reduce the influence of co-variation of these two climate variables for the reconstruction and increase the number of records that pass a significance level of $p < 0.1$ (Telford and Birks, 2011).

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

5.4 Potential use of LegacyClimate 1.0

Our fossil pollen synthesis contains records from all over the Northern Hemisphere extratropics and thus can be used to infer spatio-temporal patterns in climate reconstructions that are not only limited to a local or regional scale. Although several hemispheric or global reconstruction studies exist, they have been largely restricted to temperature or have included relatively few records (Marcott et al., 2013; Marsicek et al., 2018; Routson et al., 2019; Kaufman et al., 2020a and 2020b). Our dataset is therefore a valuable addition. It may be used in a multi-proxy approach, synthesizing marine and terrestrial records in order to assess temperature development during the Holocene and can help to highlight possible interdependencies between oceans and land masses and such contribute to the "Holocene conundrum" debate (Liu et al., 2014). Temperature reconstructions from proxy data indicate peak temperatures during the Holocene Thermal Maximum around 6000 years BP followed by a pronounced cooling trend toward the late Holocene (Liu et al., 2014; Bova et al., 2021), which is also visible in our pollen-based reconstructions (Fig. 6). In contrast, climate models simulate a monotonic warming throughout the Holocene, which resulted in the "Holocene conundrum" debate (Liu et al., 2014). Temperature reconstructions are often derived from sea-surface temperatures as either mean annual temperatures (Birks, 2019; Bova et al., 2021) or global mean surface temperature (Marcott et al., 2013; Marsicek et al., 2018; Kaufman et al., 2020a and 2020b). However, it is argued that proxy-based climate reconstructions are seasonally biased and therefore might be the reason for the observed proxy-model divergence (Liu et al., 2014; Rehfeld et al., 2016; Bova et al., 2021). In this respect, it might help that we provide T_{july} along with T_{ann} reconstructions, which provides the opportunity to assess seasonal impacts on the reconstruction.

So far, reconstructions of precipitation have not been implemented on a hemispheric scale. The interconnection between temperature and precipitation (Trenberth, 2011) and its spatio-temporal variation across the Northern Hemisphere is therefore an important aspect of evaluating climate models (Wu et al., 2013; Hao et al., 2019; Herzschuh et al., submitted). A broad-scale quantitative reconstruction of temperature and precipitation would therefore be of great value for evaluating transient model runs performed by climate models such as TraCE 21k (He, 2010).

6 Data and code availability

The compilation of reconstructed T_{July}, T_{ann}, and P_{ann}, is open access and available at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.930512; in the "*Other version*" section; Herzschuh et al.,

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

2021). The dataset files are stored in machine-readable data format (.CSV), which are already separated into Western North America, Eastern North America, Europe, and Asia for easy access and use.

The R code to run the reconstructions for single sites is available at Zenodo (https://doi.org/10.5281/zenodo.5910989; Herzschuh et al., 2022) including harmonized open-access modern and fossil pollen datasets so that customized reconstructions can be easily established.

Author contributions. UH designed the study design and reconstruction dataset. CL and TB compiled the metadata and the harmonized pollen dataset. TB wrote the R scripts and ran the analyses under the supervision of UH. UH, TB and MC wrote the first draft of the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to express our gratitude to all the palynologists and geologists who, either directly or indirectly by providing their work the Neotoma Paleoecology Database, contributed pollen data and chronologies to the dataset. The work of data contributors, data stewards, and the Neotoma community is gratefully acknowledged. We thank Andrej Andreev, Mareike Wieczorek, and Birgit Heim from AWI for providing information on pollen records and data uploads. We also thank Cathy Jenks for language editing on a previous version of the paper.

Financial support. This research has been supported by the European Research Council (ERC Glacial Legacy 772852 to UH) and the PalMod Initiative (01LP1510C to UH). TB and MC is supported by the German Federal Ministry of Education and Research (BMBF) as a Research for Sustainability initiative (FONA; https://www.fona.de/en) through the PalMod Phase II project (grant no. FKZ: 01LP1926D). CL holds a scholarship from the Chinese Scholarship Council (grant no. 201908130165). NR work was supported by the Russian Science Foundation (Grant No. 20-17-00110).

References

Adler, R. F., Gu, G., Wang, J.-J., Huffman, G. J., Curtis, S., and Bolvin, D.: Relationships between global precipitation and surface temperature on interannual and longer timescales (1979–2006), J. Geophys. Res. Atmos., 113, https://doi.org/10.1029/2008JD010536, 2008.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

441 Birks, H. J. B.: Contributions of Quaternary botany to modern ecology and biogeography, Plant Ecol.
442 Divers., 12, 189–385, https://doi.org/10.1080/17550874.2019.1646831, 2019.

443 Birks, H. J. B., Heiri, O., Seppä, H., and Bjune, A. E.: Strengths and Weaknesses of Quantitative Climate
444 Reconstructions Based on Late-Quaternary, Open Ecol. J., 3, 68–110,
445 http://dx.doi.org/10.2174/1874213001003020068, 2010.

446 Blois, J. L., Williams, J. W., Grimm, E. C., Jackson, S. T., and Graham, R. W.: A methodological
447 framework for assessing and reducing temporal uncertainty in paleovegetation mapping from late-
448 Quaternary pollen records, Quat. Sci. Rev., 30, 1926–1939,
449 https://doi.org/10.1016/j.quascirev.2011.04.017, 2011.

450 Bova, S., Rosenthal, Y., Liu, Z., Godad, S. P., and Yan, M.: Seasonal origin of the thermal maxima at
451 the Holocene and the last interglacial, Nature, 589, 548–553, https://doi.org/10.1038/s41586-020-
452 03155-x, 2021.

453 Cao, X., Ni, J., Herzschuh, U., Wang, Y., and Zhao, Y.: A late Quaternary pollen dataset from eastern
454 continental Asia for vegetation and climate reconstructions: Set up and evaluation, Rev. Palaeobot.
455 Palynol., 194, 21–37, https://doi.org/10.1016/j.revpalbo.2013.02.003, 2013.

456 Cao, X., Herzschuh, U., Telford, R. J., and Ni, J.: A modern pollen–climate dataset from China and
457 Mongolia: Assessing its potential for climate reconstruction, Rev. Palaeobot. Palynol., 211, 87–96,
458 https://doi.org/10.1016/j.revpalbo.2014.08.007, 2014.

459 Cao, X., Tian, F., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe,
460 A., Wieczorek, M., and Herzschuh, U.: A taxonomically harmonized and temporally standardized fossil
461 pollen dataset from Siberia covering the last 40 kyr, Earth Syst. Sci. Data, 12, 119–135,
462 https://doi.org/10.5194/essd-12-119-2020, 2020.

463 Chen, F., Chen, J., Huang, W., Chen, S., Huang, X., Jin, L., Jia, J., Zhang, X., An, C., Zhang, J., Zhao,
464 Y., Yu, Z., Zhang, R., Liu, J., Zhou, A., and Feng, S.: Westerlies Asia and monsoonal Asia:
465 Spatiotemporal differences in climate change and possible mechanisms on decadal to sub-orbital
466 timescales, Earth Sci. Rev., 192, 337–354, https://doi.org/10.1016/j.earscirev.2019.03.005, 2019.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

467 Chevalier, M.: Enabling possibilities to quantify past climate from fossil assemblages at a global scale,
468 Glob. Planet. Change, 175, 27–35, https://doi.org/10.1016/j.gloplacha.2019.01.016, 2019.

469 Chevalier, M., Cheddadi, R., and Chase, B. M.: CREST (Climate REconstruction SofTware): a
470 probability density function (PDF)-based quantitative climate reconstruction method, Clim. Past, 10,
471 2081–2098, https://doi.org/10.5194/cp-10-2081-2014, 2014.

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

478 Davis, B. A. S., Zanon, M., Collins, P., Mauri, A., Bakker, J., Barboni, D., Barthelmes, A., Beaudouin, C.,
479 Bjune, A. E., Bozilova, E., Bradshaw, R. H. W., Brayshay, B. A., Brewer, S., Brugiapaglia, E., Bunting,
480 J., Connor, S. E., de Beaulieu, J.-L., Edwards, K., Ejarque, A., Fall, P., Florenzano, A., Fyfe, R., Galop,
481 D., Giardini, M., Giesecke, T., Grant, M. J., Guiot, J., Jahns, S., Jankovská, V., Juggins, S., Kahrmann,
482 M., Karpińska-Kołaczek, M., Kołaczek, P., Kühl, N., Kuneš, P., Lapteva, E. G., Leroy, S. A. G., Leydet,
483 M., Guiot, J., Jahns, S., Jankovská, V., Juggins, S., Kahrmann, M., Karpińska-Kołaczek, M., Kołaczek,
484 P., Kühl, N., Kuneš, P., Lapteva, E. G., Leroy, S. A. G., Leydet, M., López Sáez, J. A., Masi, A.,
485 Matthias, I., Mazier, F., Meltsov, V., Mercuri, A. M., Miras, Y., Mitchell, F. J. G., Morris, J. L., Naughton,
486 F., Nielsen, A. B., Novenko, E., Odgaard, B., Ortu, E., Overballe-Petersen, M. V., Pardoe, H. S.,
487 Peglar, S. M., Pidek, I. A., Sadori, L., Seppä, H., Severova, E., Shaw, H., Święta-Musznicka, J.,
488 Theuerkauf, M., Tonkov, S., Veski, S., van der Knaap, W. O., van Leeuwen, J. F. N., Woodbridge, J.,
489 Zimny, M., and Kaplan, J. O.: The European Modern Pollen Database (EMPD) project, Veg. Hist.
490 Archaeobot., 22, 521–530, https://doi.org/10.1007/s00334-012-0388-5, 2013.

491 Davis, B. A. S., Chevalier, M., Sommer, P., Carter, V. A., Finsinger, W., Mauri, A., Phelps, L. N., Zanon,
492 M., Abegglen, R., Åkesson, C. M., Alba-Sánchez, F., Anderson, R. S., Antipina, T. G., Atanassova, J.
493 R., Beer, R., Belyanina, N. I., Blyakharchuk, T. A., Borisova, O. K., Bozilova, E., Bukreeva, G., Bunting,
494 M. J., Clò, E., Colombaroli, D., Combourieu-Nebout, N., Desprat, S., Di Rita, F., Djamali, M., Edwards,
495 K. J., Fall, P. L., Feurdean, A., Fletcher, W., Florenzano, A., Furlanetto, G., Gaceur, E., Galimov, A.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

496 T., Gałka, M., García-Moreiras, I., Giesecke, T., Grindean, R., Guido, M. A., Gvozdeva, I. G.,
497 Herzschuh, U., Hjelle, K. L., Ivanov, S., Jahns, S., Jankovska, V., Jiménez-Moreno, G., Karpińska-
498 Kołaczek, M., Kitaba, I., Kołaczek, P., Lapteva, E. G., Latałowa, M., Lebreton, V., Leroy, S., Leydet,
499 M., Lopatina, D. A., López-Sáez, J. A., Lotter, A. F., Magri, D., Marinova, E., Matthias, I., Mavridou,
500 A., Mercuri, A. M., Mesa-Fernández, J. M., Mikishin, Y. A., Milecka, K., Montanari, C., Morales-Molino,
501 C., Mrotzek, A., Muñoz Sobrino, C., Naidina, O. D., Nakagawa, T., Nielsen, A. B., Novenko, E. Y.,
502 Panajiotidis, S., Panova, N. K., Papadopoulou, M., Pardoe, H. S., Pędziszewska, A., Petrenko, T. I.,
503 Ramos-Román, M. J., Ravazzi, C., Rösch, M., Ryabogina, N., Sabariego Ruiz, S., Salonen, J. S.,
504 Sapelko, T. V., Schofield, J. E., Seppä, H., Shumilovskikh, L., Stivrins, N., Stojakowits, P., Svobodova
505 Svitavska, H., Święta-Musznicka, J., Tantau, I., Tinner, W., Tobolski, K., Tonkov, S., Tsakiridou, M.,
506 et al.: The Eurasian Modern Pollen Database (EMPD), version 2, Earth Syst. Sci. Data, 12, 2423–
507 2445, https://doi.org/10.5194/essd-12-2423-2020, 2020.

508 Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier,
509 B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P.,
510 Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C.,
511 Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and
512 Williamson, M. S.: Taking climate model evaluation to the next level, Nat. Clim. Chang., 9, 102–110,
513 https://doi.org/10.1038/s41558-018-0355-y, 2019.

514 Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land
515 areas, Int. J. Climatol., 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

516 Hao, Z., Phillips, T. J., Hao, F., and Wu, X.: Changes in the dependence between global precipitation
517 and temperature from observations and model simulations, Int. J. Climatol., 39, 4895–4906,
518 https://doi.org/10.1002/joc.6111, 2019.

519 He, F.: Simulating transient climate evolution of the last deglaciation with CCSM3, Ph.D. thesis,
520 University of Wisconsin-Madison, USA, 185 pp., 2010.

521 Herzschuh, U., Böhmer, T., Li, C., and Cao, X.: Northern Hemisphere temperature and precipitation
522 reconstruction from taxonomically harmonized pollen data set with revised chronologies using WA-

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

PLS and MAT (LegacyClimate 1.0), PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.930512, 2021.

Herzschuh, U., Böhmer, T., Li, C., Chevalier, M., Dallmeyer, A., Cao, X., Bigelow, N. H., Nazarova, L., Novenko, E. Y., Park, J., Peyron, O., Rudaya, N. A., Schlütz, F., Shumilovskikh, L. S., Tarasov, P. E., Wang, Y., Wen, R., Xu, Q., and Zheng, Z.: LegacyClimate 1.0: A dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the late Quaternary, Zenodo, https://doi.org/10.5281/zenodo.5910989, 2022.

Herzschuh, U., Cao, X., Laepple, T., Dallmeyer, A., Telford, R. J., Ni, J., Chen, F., Kong, Z., Liu, G., Liu, K.-B., Liu, X., Stebich, M., Tang, L., Tian, F., Wang, Y., Wischnewski, J., Xu, Q., Yan, S., Yang, Z., Yu, G., Zhang, Y., Zhao, Y., and Zheng, Z.: Position and orientation of the westerly jet determined Holocene rainfall patterns in China, Nat. Commun., 10, 2376, https://doi.org/10.1038/s41467-019-09866-8, 2019.

Herzschuh, U., Li, C., Böhmer, T., Postl, A. K., Heim, B., Andreev, A. A., Cao, X., and Wieczorek, M.: LegacyPollen 1.0: A taxonomically harmonized global Late Quaternary pollen dataset of 2831 records with harmonized chronologies, submitted.

Hijmans, R. J., van Etten, J., Sumner, M., Cheng, J., Baston, D., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Hingee, K., Ilich, A., Institute for Mathematics Applied Geosciences, Karney, C., Mattiuzzi, M., Mosher, S., Naimi, B., Nowosad, J., Pebesma, E., Lamigueiro, O. P., Racine, E. B., Rowlingson, B., Shortridge, A., Venables, B., and Wueest, R.: Raster: Geographic Data Analysis and Modeling, R package version 3.5-11, https://cran.r-project.org/web/packages/raster, 2021.

Hill, M. O.: Diversity and Evenness: A Unifying Notation and Its Consequences, Ecology, 54, 427–432, https://doi.org/10.2307/1934352, 1973.

Jackson, S. T.: Pollen source area and representation in small lakes of the northeastern United States, Rev. Palaeobot. Palynol., 63, 53–76, https://doi.org/10.1016/0034-6667(90)90006-5, 1990.

Juggins, S.: rioja: Analysis of Quaternary Science Data, R package version 0.9-21, https://cran.r-project.org/web/packages/rioja, 2019.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

550 Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler,
551 C., Axford, Y., Brussel, T., Cartapanis, O., Chase, B., Dawson, A., de Vernal, A., Engels, S., Jonkers,
552 L., Marsicek, J., Moffa-Sánchez, P., Morrill, C., Orsi, A., Rehfeld, K., Saunders, K., Sommer, P. S.,
553 Thomas, E., Tonello, M., Tóth, M., Vachula, R., Andreev, A., Bertrand, S., Biskaborn, B., Bringué, M.,
554 Brooks, S., Caniupán, M., Chevalier, M., Cwynar, L., Emile-Geay, J., Fegyveresi, J., Feurdean, A.,
555 Finsinger, W., Fortin, M.-C., Foster, L., Fox, M., Gajewski, K., Grosjean, M., Hausmann, S., Heinrichs,
556 M., Holmes, N., Ilyashuk, B., Ilyashuk, E., Juggins, S., Khider, D., Koinig, K., Langdon, P., Larocque-
557 Tobler, I., Li, J., Lotter, A., Luoto, T., Mackay, A., Magyari, E., Malevich, S., Mark, B., Massaferro, J.,
558 Montade, V., Nazarova, L., Novenko, E., Pařil, P., Pearson, E., Peros, M., Pienitz, R., Płóciennik, M.,
559 Porinchu, D., Potito, A., Rees, A., Reinemann, S., Roberts, S., Rolland, N., Salonen, S., Self, A.,
560 Seppä, H., Shala, S., St-Jacques, J.-M., Stenni, B., Syrykh, L., Tarrats, P., Taylor, K., van den Bos,
561 V., Velle, G., Wahl, E., Walker, I., Wilmshurst, J., Zhang, E., and Zhilich, S.: A global database of
562 Holocene paleotemperature records, Sci. Data, 7, 115, https://doi.org/10.1038/s41597-020-0445-3,
563 2020a.

564 Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P. S., Heiri, O., and Davis, B.:
565 Holocene global mean surface temperature, a multi-method reconstruction approach, Sci. Data, 7,
566 201, https://doi.org/10.1038/s41597-020-0445-3, 2020b.

567 Li, C., Postl, A. K., Böhmer, T., Cao, X., Dolman, A. M., and Herzschuh, U.: Harmonized chronologies
568 of a global late Quaternary pollen dataset (LegacyAge 1.0), Earth Syst. Sci. Data Discuss. [preprint],
569 https://doi.org/10.5194/essd-2021-212, in review, 2022.

570 Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann,
571 G., Zheng, W., and Timm, O. E.: The Holocene temperature conundrum, PNAS, 111, E3501–E3505,
572 https://doi.org/10.1073/pnas.1407229111, 2014.

573 Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A Reconstruction of Regional and Global
574 Temperature for the Past 11,300 Years, Science, 339, 1198–1201,
575 https://doi.org/10.1126/science.1228026, 2013.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

576 Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., and Brewer, S.: Reconciling divergent trends
577 and millennial variations in Holocene temperatures, Nature, 554, 92–96,
578 https://doi.org/10.1038/nature25464, 2018.

579 Mauri, A., Davis, B. A. S., Collins, P. M., and Kaplan, J. O.: The climate of Europe during the Holocene:
580 a gridded pollen-based reconstruction and its multi-proxy evaluation, Quat. Sci. Rev., 112, 109–127,
581 https://doi.org/10.1016/j.quascirev.2015.01.013, 2015.

582 Nychka, D., Furrer, R., Paige, J., Sain, S., Gerber, F., and Iverson, M.: fields: Tools for Spatial Data, R
583 package version 10.3, https://cran.r-project.org/web/packages/fields/index.html, 2020.

584 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara,
585 R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.: Vegan:
586 Community Ecology Package, R package version 2.5-7, https://cran.r-
587 project.org/web/packages/vegan, 2020.

588 Overpeck, J. T., Webb, T., and Prentice, I. C.: Quantitative Interpretation of Fossil Pollen Spectra:
589 Dissimilarity Coefficients and the Method of Modern Analogs, Quat. Res., 23, 87–108,
590 https://doi.org/10.1016/0033-5894(85)90074-2, 1985.

591 R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical
592 Computing, Vienna, Austria, available online at: https://www.R-project.org/, 2020.

593 Rehfeld, K., Trachsel, M., Telford, R. J., and Laepple, T.: Assessing performance and seasonal bias of
594 pollen-based climate reconstructions in a perfect model world, Clim. Past, 12, 2255–2270,
595 https://doi.org/10.5194/cp-12-2255-2016, 2016.

596 Routson, C. C., McKay, N. P., Kaufman, D. S., Erb, M. P., Goosse, H., Shuman, B. N., Rodysill, J. R.,
597 and Ault, T.: Mid-latitude net precipitation decreased with Arctic warming during the Holocene, Nature,
598 568, 83–87, https://doi.org/10.1038/s41586-019-1060-3, 2019.

599 Simpson, G. L.: Analogue Methods in Palaeolimnology, in: Tracking Environmental Change Using Lake
600 Sediments: Data Handling and Numerical Techniques, edited by: Birks, H. J. B., Lotter, A. F., Juggins,
601 S., and Smol, J. P., Springer Netherlands, Dordrecht, 495–522, https://doi.org/10.1007/978-94-007-
602 2745-8_15, 2012.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

603 Sugita, S.: A Model of Pollen Source Area for an Entire Lake Surface, Quat. Res., 39, 239–244,
604 https://doi.org/10.1006/qres.1993.1027, 1993.

605 Tarasov, P. E., Nakagawa, T., Demske, D., Österle, H., Igarashi, Y., Kitagawa, J., Mokhova, L.,
606 Bazarova, V., Okuda, M., Gotanda, K., Miyoshi, N., Fujiki, T., Takemura, K., Yonenobu, H., and Fleck,
607 A.: Progress in the reconstruction of Quaternary climate dynamics in the Northwest Pacific: A new
608 modern analogue reference dataset and its application to the 430-kyr pollen record from Lake Biwa,
609 Earth Sci. Rev., 108, 64–79, https://doi.org/10.1016/j.earscirev.2011.06.002, 2011.

610 Telford, R. J.: palaeoSig: Significance Tests for Palaeoenvironmental Reconstructions, R package
611 version 2.0-3, https://cran.r-project.org/web/packages/palaeoSig, 2019.

612 Telford, R. J. and Birks, H. J. B.: A novel method for assessing the statistical significance of quantitative
613 reconstructions inferred from biotic assemblages, Quat. Sci. Rev., 30, 1272–1278,
614 https://doi.org/10.1016/j.quascirev.2011.03.002, 2011.

615 ter Braak, C. J. F. and Juggins, S.: Weighted averaging partial least squares regression (WA-PLS): an
616 improved method for reconstructing environmental variables from species assemblages,
617 Hydrobiologia, 269, 485–502, https://doi.org/10.1007/BF00028046, 1993.

618 Tian, F., Cao, X., Dallmeyer, A., Zhao, Y., Ni, J., and Herzschuh, U.: Pollen-climate relationships in time
619 (9 ka, 6 ka, 0 ka) and space (upland vs. lowland) in eastern continental Asia, Quat. Sci. Rev., 156, 1–
620 11, https://doi.org/10.1016/j.quascirev.2016.11.027, 2017.

621 Trachsel, M. and Telford, R. J.: All age–depth models are wrong, but are getting better, Holocene, 27,
622 860–869, https://doi.org/10.1177/0959683616675939, 2017.

623 Trenberth, K. E.: Changes in precipitation with climate change, Clim. Res., 47, 123–138,
624 https://doi.org/10.3354/cr00953, 2011.

625 Whitmore, J., Gajewski, K., Sawada, M., Williams, J. W., Shuman, B., Bartlein, P. J., Minckley, T., Viau,
626 A. E., Webb, T., Shafer, S., Anderson, P., and Brubaker, L.: Modern pollen data from North America
627 and Greenland for multi-scale paleoenvironmental applications, Quat. Sci. Rev., 24, 1828–1848,
628 https://doi.org/10.1016/j.quascirev.2005.03.005, 2005.

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

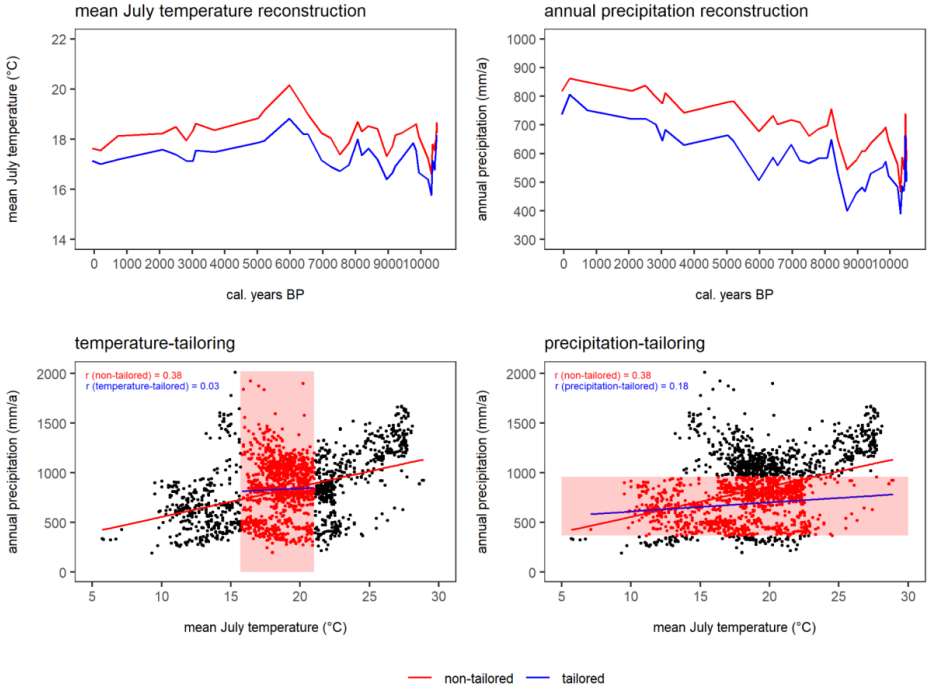
629 Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W.,
630 Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth,
631 R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., Nichols, J., Purdum, T.,
632 Roth, R. E., Stryker, M., and Takahara, H.: The Neotoma Paleoecology Database, a multiproxy,
633 international, community-curated data resource, Quat. Res., 89, 156–177,
634 https://doi.org/10.1017/qua.2017.105, 2018.

635 Williams, J. W., Webb III, T., Richard, P. H., and Newby, P.: Late Quaternary biomes of Canada and the
636 eastern United States, J. Biogeogr., 27, 585–607, https://doi.org/10.1046/j.1365-2699.2000.00428.x,
637 2000.

638 Wu, R., Chen, J., and Wen, Z.: Precipitation-surface temperature relationship in the IPCC CMIP5 models,
639 Adv. Atmos. Sci., 30, 766–778, https://doi.org/10.1007/s00376-012-2130-8, 2013.

640

641 **Appendix Figures**



642

https://doi.org/10.5194/essd-2022-38
Preprint. Discussion started: 10 February 2022

Open Access Earth System Science Data Discussions

643 **Appendix Figure 1.** Example to illustrate the effect of tailoring the modern dataset for the location
644 "Yellow Dog Pond" in Eastern North America. Upper part: reconstruction of T_{July} and P_{ann} with WA-PLS
645 (red) and WA-PLS_tailored (blue); lower part: correlation of T_{July} and P_{ann} in the modern dataset and the
646 effect of tailoring the modern dataset (indicated with the red box). Correlations are given for non-tailored
647 (red) and tailored (blue) data.