Journal Title
XX(X):2–23

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Comparison of simulated and proxy-based climate reconstructions for mid-Holocene Europe reveals high uncertainty

Unformatted manuscript as accepted by The Holocene

DOI: 10.1177/09596836251366198

Wolfgang Traylor[1], Mark G. Thomas[2], Adrian Timpson[2], Edward Armstrong[3], Maria Ivanova-Bieg[4] and Thomas Hickler[1,5]

Prepared using sagej.cls

Abstract

Gridded and time-continuous paleoclimate model outputs are increasingly used to inform high-resolution models in biogeography and other disciplines. However, few studies quantitatively evaluate such outputs for bias, uncertainty, and consistency with other climate reconstructions of similar scale and scope of interest. Here we evaluate downscaled and bias-corrected output from two paleoclimate models against proxy-based reconstructions from European sites in the mid-Holocene (9000–5000 years before present) by comparing absolute values of annual precipitation sum, and mean January and July air temperature. We paired proxy-based climate values with simulated ones from the same time and place. For pairs within each site we checked for correlation. Then we pooled pair-wise differences of all sites together to check for systematic over- or underestimation. Correlation analysis revealed that climate models reproduce a winter warming trend in the northern half of Europe. Pair-wise differences indicate that all gridded datasets show both over- and underestimation bias compared to proxies in different climate variables, but without one dataset performing consistently better. Distinguishing performance between these datasets is complicated by high uncertainties in proxy-based reconstructions. Our results indicate that, at least on the whole-European scale, time-continuous model outputs resolve relevant mid-Holocene climatic changes only to a very limited degree. In order to account for current uncertainties in reconstructing paleoclimate we advise users not to rely on only one gridded dataset and to evaluate reconstructions from models against proxy data from the area of interest.

Keywords

paleoclimate model, general circulation model, pollen, chironomid, palaeoclimate, model evaluation, model validation

[1]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany
[2]Research Department of Genetics, Evolution and Environment, University College London, UK
[3]Department of Geosciences and Geography, University of Helsinki, Finland
[4]Department of Prehistoric and Early Historic Archaeology, Johannes Gutenberg University, Mainz, Germany
[5]Institute of Physical Geography, Goethe University, Frankfurt am Main, Germany

Corresponding author:
Wolfgang Traylor, Senckenberg Nature Research Society, Senckenberganlage 25, 60325, Frankfurt am Main, Germany
Email: wolfgang.traylor@posteo.de

Introduction

Understanding how ecosystems and prehistoric societies have responded to past climatic changes provides important background for understanding ongoing climatic, environmental, and cultural changes. Paleoclimate model simulations are valuable for understanding drivers and mechanisms of past climate. In addition, their output is increasingly used to inform analyses of biological or cultural changes: for example, in dynamic vegetation models (Miller et al. 2008; Allen et al. 2020; Huntley et al. 2023), in species distribution models (Franklin et al. 2015; Divíšek et al. 2022), in eco-cultural niche models (Banks et al. 2008, 2013), or as Bayesian priors in proxy-based climate reconstructions (Ön et al. 2023). Simulated paleoclimate datasets with continuous timeseries are particularly appealing because they can directly drive dynamic models or be matched to dated fossil or archaeological finds. However, few studies using such model outputs also evaluate them against proxy-based reconstructions of past climate. In particular, to users not specialized in paleoclimate modeling it can be difficult to assess how reliable simulation-derived datasets are and at which spatial and temporal scale they can be used.

Paleobiogeographers and archaeologists share an interest in linking biological or cultural remains with corresponding paleoclimate reconstructions. Based on these links, they can fit or evaluate models in order to investigate how climate shaped culture or the distribution and abundances of organisms (Franklin et al. 2015; d'Alpoim Guedes et al. 2016). Linking dated remains from specific sites to climate becomes particularly convenient with paleoclimate data that is gridded and time-continuous. Here, a high spatial resolution (e.g., a 1 km grid) promises to capture important local conditions, such as elevation effects. The temporal resolution should ideally range on the scale of decades to a few centuries in order to match (1) the speed of sociocultural processes (Shennan 2018) and (2) the dating precision of remains (Lacourse and Gajewski 2020). Paleoclimatic variable values can be given as relative numbers (e.g., in Kelvin difference to preindustrial) or absolute numbers (e.g., in °C). For biogeographers and archaeologists, absolute climate values are more useful (Myers et al. 2015) as they can drive mechanistic models (e.g., dynamic vegetation models), are comparable with modern observations (e.g., drought tolerance of crops), and can mark physiological limits of taxa (e.g., frost tolerance). If a paleoclimate dataset meets these desiderata on a technical level, it remains to be verified that (a) it actually resolves the spatiotemporal dynamics of interest and (b) it is not systematically biased towards climate values that are too high or too low.

Output from paleoclimate model runs can provide the spatiotemporal resolution needed for biogeographical analyses. However, computational constraints allow global climate models to run only on a rather coarse spatial resolution (typically 2–4° horizontally) on centennial-to-millennial timescales (Askjær et al. 2022). Downscaling is required to obtain more localized information.

Downscaling approaches can be dynamical or statistical. Dynamical downscaling may involve adding sub-grid cell calculations (Quiquet et al. 2018; Arthur et al. 2023) or running a finer-grained regional climate model that integrates with the coarser-grained global climate model (Ludwig et al. 2019). In statistical downscaling on the other hand, information from fine-grained modern-day climate data (e.g., 1° resolution) is used to correct the coarse-grained output of a global climate simulation (e.g., Lorenz et al. 2016). This way, modern-day climate helps to resolve topographic features, such as elevational effects, but it cannot improve temporal dynamics.

Statistical downscaling often goes hand in hand with bias correction. Bias refers to how much the climate-model output for the present deviates from observations on a grid cell basis (Maraun 2016). With present-day climate as the only overlap between simulations and observations, the relationship between them is assumed invariant over time (for critique see Maraun 2016). Bias correction can then be applied to the whole simulated timeseries on a grid-cell basis (Beyer et al. 2020). Notably, this procedure compensates for model shortcomings only on a very superficial level and cannot correct major model errors (Maraun et al. 2017). However, the advantage is that bias correction allows interpreting timeseries as absolute quantities rather than as relative to present. Therefore, downscaled and bias-corrected paleoclimate model output provides the spatial structure desired as input for biogeographical models.

Apart from spatial resolution, users of paleoclimate model outputs are interested in its temporal resolution and extent. Time-continuous simulation-based paleoclimate datasets can be created with different methods. One way is to perform actually transient climate simulations, which accounts for the transient nature of climate evolution but takes a long time to compute (Liu et al. 2009). Another way is to simulate equilibrium climate for time slices (e.g., every 1000 years) and then interpolate between them in order to generate a time-continuous output dataset (Singarayer and Valdes 2010). These “snapshots” can be simulated in parallel, which is much faster than transient simulations (Armstrong et al. 2019). Both approaches can yield datasets of multimillennial extent and of monthly, annual, or centennial resolution. However, it remains to be evaluated how the two approaches differ in resolving the spatiotemporal dynamics that are of interest to users.

Evaluation of paleoclimate model simulations is routinely done with independent, proxy-based reconstructions. Previous comparisons between simulated and proxy-based paleoclimate reconstructions have often focused on specific time slices (e.g., Bonfils et al. 2004) or on global-scale patterns (e.g., Zhang et al. 2022) and often only on mean surface air temperature (e.g., Kaufman et al. 2020b). Armstrong et al. (2019) and Karger et al. (2023) showed that statistical bias-correction of paleoclimate-model output improves the fit to Arctic ice-core temperature reconstructions, but gave little attention to temperate regions. Evaluations aimed at improving climate models typically focus on relative differences, for instance between mid-Holocene and pre-industrial conditions (e.g., Russo and Cubasch 2016; Russo et al. 2022). In contrast, our aim is to evaluate climate model data products as applied by end users.

Consequently, we chose time-continuous, downscaled, and bias-corrected gridded datasets from two different paleoclimate models: HadCM3B (Armstrong et al. 2019) and CHELSA-TraCE21k (Karger et al. 2023). We compared model-based with proxy-derived timeseries from across Europe (Kaufman et al. 2020a), checking for correlations as well as systematic over- or underestimation. Ideally, a good match between simulated and proxy-based paleoclimate values would justify using the full range of climate-model output variables with complete spatiotemporal coverage. However, an unsatisfactory match would be important for informing choice and interpretation of paleoclimate model outputs.

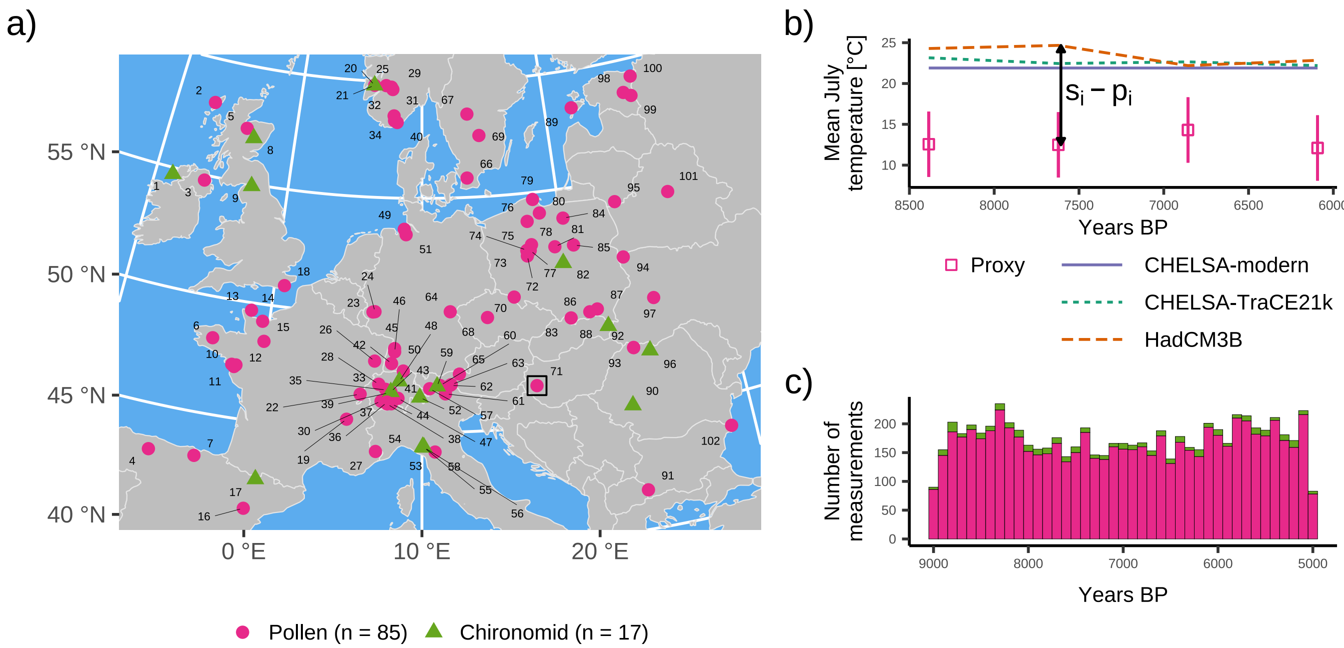

Figure 1. (a) Proxy sites used in this study. (b) Example timeseries for one climate variable in one proxy site (rectangle in map). Points show proxy-based reconstructions ($p_i \pm 2e$) and lines the values in the enclosing grid cell from simulated/modern climate (s_i). The pair-wise difference $s_i - p_i$ is the basis of comparison. (c) Temporal coverage in 100-year bins of proxy-based reconstructions (all climate variables pooled).

Methods

Study area and time

We compared outputs of paleoclimate models with proxy-based reconstructions in Europe (10° W–30° E and 38–60° N; Fig. 1a) in the range 9000–5000 years before present (BP): a period covering the introduction and spread of farming, and profound societal changes. Europe is exceptionally well covered by climate proxy records (Kaufman et al. 2020a). We excluded the early Holocene because it is very dissimilar to modern pollen and chironomid calibration datasets, rendering proxy-based reconstructions less reliable (Williams et al. 2001). In the late Holocene, pollen-based climate reconstructions become less reliable again because of increasing anthropogenic land use (Rao et al. 2022).

Climate model data

We selected the two time-continuous paleoclimate datasets by Armstrong et al. (2019) and Karger et al. (2023), which we refer to as HadCM3B and CHELSA-TraCE21k, respectively. A number of transient climate simulations have been performed for the

Holocene (Askjær et al. 2022, and references therein), but for typical biogeographic analyses their output needs to be downscaled, bias-corrected, and converted to an accessible data format. Armstrong et al. (2019) and Karger et al. (2023) have already done this postprocessing.

The two time-continuous paleoclimate model outputs originate from different climate models and experienced different post-processing. In a snapshot modeling approach, Armstrong et al. (2019) used the HadCM3B (Valdes et al. 2017) general circulation model to simulate 1000-year time slices, forced (during the Holocene) with orbital parameters, greenhouse gases, and ice sheets. Each simulation ran for at least 6000 years on a horizontal resolution of 3.75° × 2.5°. They splined these snapshot simulations together into a continuous monthly timeseries. Afterwards they added interannual variability produced by the original HadCM3B simulation. Finally they downscaled to 0.5° spatial resolution and bias-corrected using CRU CL v2.0 observations-based data (New et al. 2002). The CRU dataset covers the period 1901 to 1990 in ⅙th degree resolution. Armstrong et al. (2019) upscaled it to 0.5° in order to calculate the difference (or ratio, in the case of precipitation) between observations and modeled values for each month and each grid cell. Assuming a constant bias, they applied this difference as correction to all modeled values ("delta method"; Beyer et al. 2020). The resulting dataset has a monthly resolution.

The second paleoclimate model output, CHELSA-TraCE21k V1.0 (Karger et al. 2023), is based on the fully transient simulation TraCE-21k (Liu et al. 2009; He 2011), which used the Community Climate System Model 3. This 21,000-year simulation was not only forced with orbital parameters, greenhouse gases, and ice sheets but also with meltwater discharges. Karger et al. (2023) used the CHELSA V1.2 algorithm (Karger et al. 2017) to statistically downscale the original TraCE-21k output to a spatial resolution of 30 arcsec (c. 1 km) In a first step, they calculated annual mean air temperatures and annual precipitation sums of the original TraCE-21k output. Next, they downscaled these from their original resolution of 3.75° to 0.5°. On this 0.5° grid they calculated the modern-day bias ("delta") between simulated and observed values, using the upsampled CHELSA V1.2 data from 1980–1990 (Karger et al. 2017) as reference. Based on these deltas they bias-corrected the 21,000-year timeseries in 100-year average time steps. Finally, they processed all 100-year time steps with the CHELSA V1.2 algorithm (Karger et al. 2017), which calculates elevational temperature gradients and orographic precipitation. This algorithm used simulated atmospheric pressure and wind from TraCE-21k as well as paleo-orography. The final dataset provides a commonly used set of bioclimatic variables, similar to WorldClim (Hijmans et al. 2005), in 100-year slices at c. 1 km resolution.

Paleoclimate proxies

Although coverage of Holocene pollen records is exceptionally good in Europe, the number of quantitative climate reconstructions is limited, and of these only some are open access. Most proxy-based paleoclimate reconstructions in this study (102 sites; Fig. 1a) come from the Temp12k 1.1.0 database (Kaufman et al. 2020a), which itself drew from earlier compilations (Wanner et al. 2011; Marcott et al. 2013; Sundqvist et al. 2014; Routson et al. 2019; Marsicek et al. 2018). For quality control, Temp12k includes only records that span at least 4000 years, resolve sub-millennial patterns with a median sample spacing finer than 400 years, and have age control points spacing less than 3000 years (Kaufman et al. 2020a). Dating precision of individual samples varies with distance to age control point and age-depth model (Lacourse and Gajewski 2020). We supplemented Temp12k by data from Samartin et al. (2017) and Feurdean et al. (2008). All records stem from dated lake sediment or bog cores (Fig. 1a) and differ in number of data points and temporal coverage. Most records ($n = 85$) are based on pollen, which provide estimates of mean surface air temperature of the coldest and warmest months (here equated with January and July) and annual precipitation sum. The area size a pollen archive represents depends on the catchment size and other factors but typically corresponds to a radius of below 50 km (Theuerkauf et al. 2016). Some records ($n = 17$) are based on chironomid larvae, which only indicate temperature of the warmest month but on local scale (Brooks 2006). The modern analog technique (Overpeck et al. 1985) has been used for all reconstructions, which employs a transfer function calibrated with modern-day datasets (challenges of transfer functions are discussed by Brooks (2006) and Chevalier et al. (2020)).

Data analyses

For each proxy record, we extracted the corresponding simulated climate by identifying the grid cell of the respective dataset that encloses the proxy site. In each of these grid cells, we paired each proxy-based climate data point in time with the corresponding centennial mean from the simulated paleoclimate (Fig. 1b). CHELSA-TraCE21k is already aggregated to centennial means, and we aggregated the HadCM3B data to matching 100-year periods. For the HadCM3B data, we calculated annual precipitation as the twelve-month sum of monthly precipitation (dataset variable `pr`) and equated temperature of warmest/coldest month with July/January values of variable `tr`. In the CHELSA data, the annual precipitation sum is coded as the variable name `bio12`. As the CHELSA dataset lacks monthly mean temperature and only provides each month's minimum and maximum temperature (variables `tasmax` and `tasmin`), we approximated mean July and January temperature by taking the mean of the month's respective minimum and maximum temperature.

In a first step, we checked to what degree the simulated paleoclimate timeline correlates with the temporal changes in the proxies. To this end, we used R 4.2.2 (R Core Team 2023) to calculate Spearman's ρ rank correlation coefficient (Spearman 1904) for the pairs of proxy-based and simulated paleoclimate data points for each climate variable.

In the next step, we quantified the overall differences between simulated and proxy-based climate data. We formulated a model for the difference δ between the aforementioned pairs of simulated (s_i) and proxy-based (p_i) climate data. In order to account for differences between sites (sample count, study design, etc.), the model is hierarchical with "site" as levels (subscript j). While quantified uncertainty is unavailable for the simulated datasets, we included uncertainty (standard deviation e) in proxy-based reconstructions: for temperature $e = 1.4\,\mathrm{K}$ in chironomid-based and $e = 2.0\,\mathrm{K}$ in pollen-based values (Kaufman et al. 2020a), and for annual precipitation $e = 165\,\mathrm{mm\,a^{-1}}$ (Williams and Shuman 2008). The unknown "true" proxy-based climate values are denoted with φ_i. We fitted the following model once for each climate variable and gridded dataset.

$$
\begin{aligned}
p_i &\sim \text{Normal}(\varphi_i, e) \\
\delta_i &= (s_i - \varphi_i) \\
\delta_i &\sim \text{StudentT}(\nu, \alpha_j, \sigma_j) \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \epsilon) \\
\log(\sigma_j) &\sim \text{Normal}(\bar{\sigma}, \tau) \\
\bar{\alpha} &\sim \text{Normal}(0, 1) \\
3\epsilon &\sim \text{Exponential}(1) \\
\bar{\sigma} &\sim \text{Exponential}(1) \\
\tau &\sim \text{Exponential}(1) \\
\nu &\sim \text{Exponential}(1)
\end{aligned}
$$

Both the mean α_j and standard deviation σ_j are modeled hierarchically so that their pooled estimates $\bar{\alpha}$ and $\bar{\sigma}$ can be used to predict the proxy–simulation difference $\tilde{\delta}$ for a hypothetical new site (i.e., any grid cell in the study area). At the same time, $\bar{\alpha}$ indicates the "grand mean" of difference between the simulated and proxy-based climate. We chose weakly informative prior probability distributions. Note that for computational efficiency and easier prior selection we calculated δ as the z-score of $s - \varphi$ in the model implementation. We implemented and fitted the model in Stan, a probabilistic programming language for Bayesian inference (Carpenter et al. 2017). We used Stan version 2.34.0 (Stan Development Team 2023) through the R package `cmdstanr` (Gabry and Češnovar 2022). Please see Supplementary Information for the model source code.

Some biogeographical studies of the Holocene have used modern-day climate instead of paleoclimate (e.g., Conolly et al. 2008). This approach neglects any temporal changes. Consequently, such studies draw insights from spatial variance only, which has arguably remained fairly constant because it chiefly depends on topography (e.g., continental or elevational gradients). In order to evaluate the benefits of paleoclimate model outputs, we included the modern-day CHELSA 2.1 dataset (Karger et al. 2021) 1981–2010 in the analysis of pair-wise differences (Fig. 1b). It represents a 30-year average climatology at 30 arcsec resolution from statistically downscaling the observation-based ERA-Interim climatic reanalysis (Karger et al. 2017). Because of the shared CHELSA processing, the modern-day dataset differs from CHELSA-TraCE21k only in lacking any paleoclimate. In total, this yielded 9 model fits: 3 climate variables times 3 gridded climate datasets.

In order to test whether the estimate of $\bar{\alpha}$ may be biased by latitude or elevation we fitted one model with each of these variables as covariate added to α_j. We found no consistent improvement for out-of-sample estimates using Pareto Smoothed Importance Sampling (Vehtari et al. 2024), which indicates that latitude and elevation do not bias the estimate. Spatial autocorrelation do not appear to substantially bias the results either, as residuals do not show strong spatial patterns (Fig. SI 4).

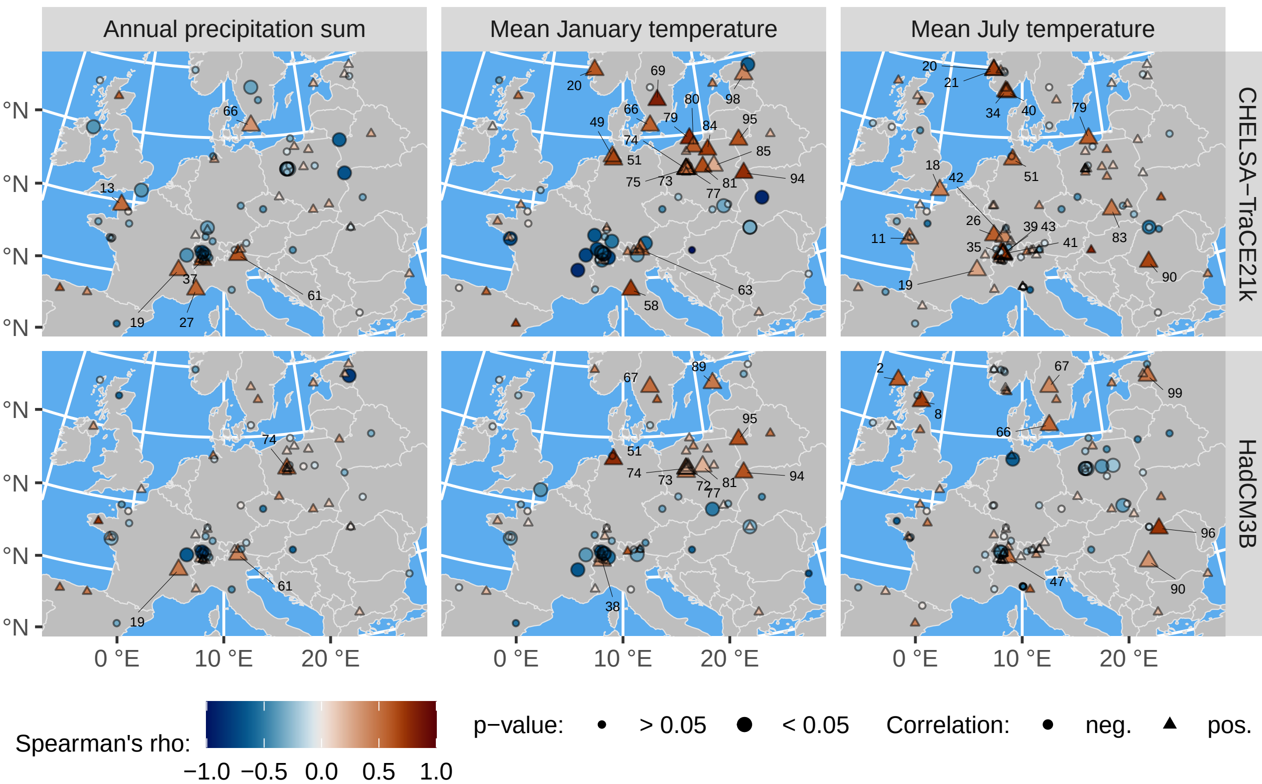

Figure 2. Rank correlation coefficients (Spearman's ρ) for each proxy site. Each point represents the correlation between the proxy-based and simulated paleoclimate timeseries (9000-5000 years before present) of the respective climate variable for one proxy site. Numbers on significant positive correlations correspond to timeseries plots in Figure 3 and the Supplementary Information. Among seemingly random distributions, both climate models capture proxy-derived trends in January temperature in the northern half of Europe (big triangles in center panels).

Results

Correlation coefficients between proxy-based and simulated paleoclimate timeseries (Fig. 2) are both positive and negative but mostly insignificant. Significant positive correlations suggest that simulated and proxy-based timeseries agree in temporal change. Studying the timeseries (Fig. SI 1–3) behind significant correlations we found no particular patterns except the following: Correlation coefficients (Fig. 2) for both models demonstrate a clear pattern of positive correlations for January temperatures of sites in the northern half of continental Europe ($>52°$ N). The timeseries plots for these sites in Figures 3 and SI 2 show a winter warming trend in proxies and model output.

In the analysis of pair-wise differences to proxy-based climate variables, none of the gridded datasets appears to be clearly superior to any other. Figure 4 summarizes the posterior distributions of the grand mean of differences $\bar{\alpha}$, i.e., the systematic bias

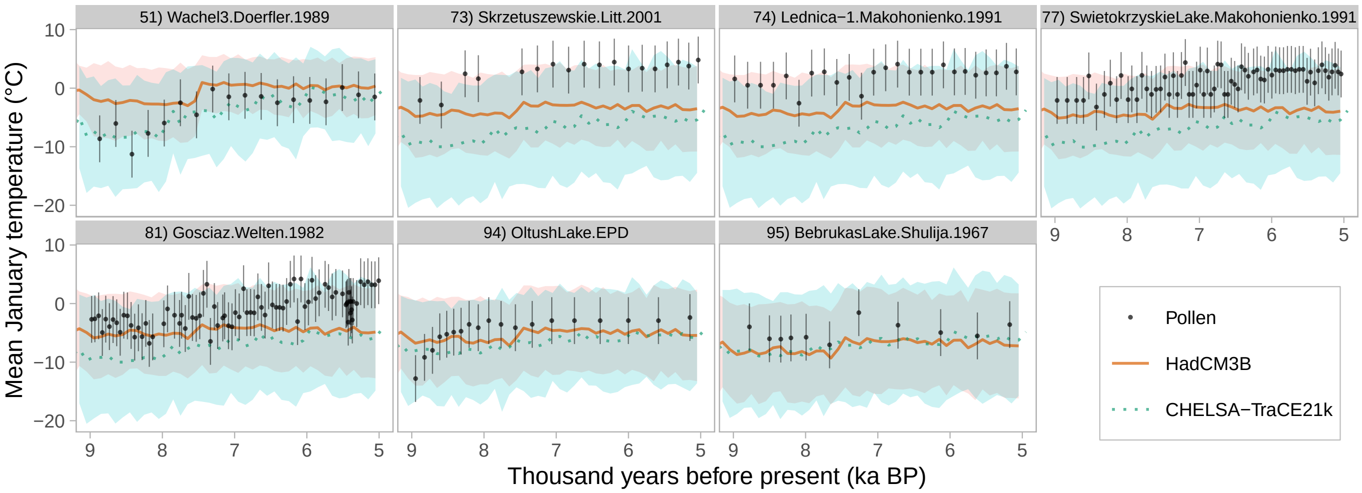

Figure 3. Timeseries for mean January temperature of those sites where *both* climate models (lines connecting 100-year means with ± 2 standard deviations from interannual variability) show significant ($p \leq 0.05$) correlation with pollen-based reconstructions (dots with ± 2 standard deviations). Most sites show a saturating temperature rise.

of the datasets. CHELSA-modern systematically overestimates precipitation (Fig. 4a), CHELSA-TraCE21k underestimates winter temperature (Fig. 4b), and HadCM3B overestimates summer temperature (Fig. 4c).

When using the gridded datasets for spatial analysis, the overall bias is of lesser interest than the deviation predicted for any single value in space and time. In principle, this predicted pair-wise difference $\tilde{\delta}$ can be used to correct values in the gridded dataset, based on proxy data. The lower subpanels in Figure 4 show the posterior distributions of the predicted differences. The distributions are very wide and overlapping, which suggests that there is no major distinction between the datasets in how well they may match *single* proxy-based reconstructions.

Discussion

Paleoclimate datasets from general circulation models provide gridded and time-continuous climate variables. We have focused on Holocene Europe, where such datasets hold potential to drive spatiotemporal analyses of ecological and cultural processes. By comparing two such available datasets with quantitative proxy-based reconstructions over the period 9000–5000 years BP we found that: (1) correlations between simulation-based and proxy-based reconstructions show model-data agreement for regional winter-warming trends, (2) simulation-based datasets do not match proxy-based reconstructions systematically better than does modern-day climate, and, (3) the evaluation is affected by high uncertainties. This is not to discard the value of paleoclimate simulation output for biogeographical studies.

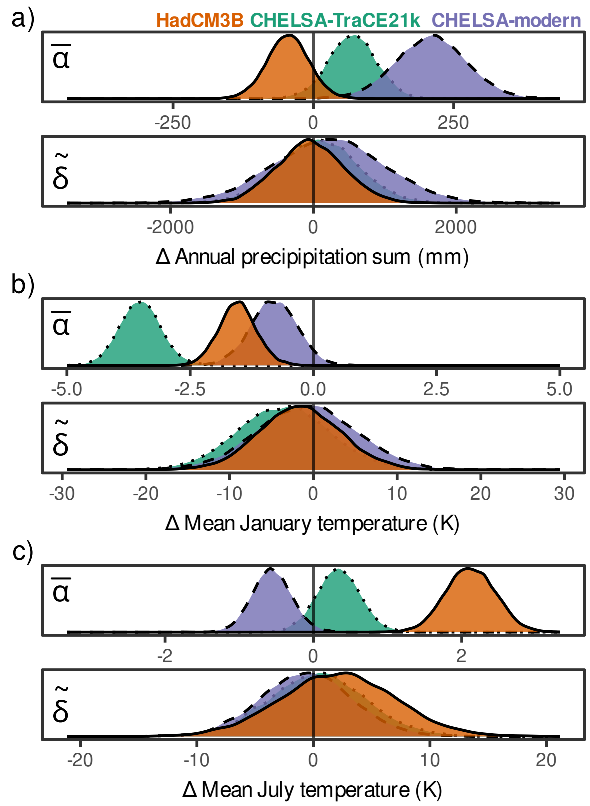

Figure 4. Posterior probability distributions of pair-wise differences ($s_i - \varphi_i$) between gridded datasets and European proxy-based paleoclimate reconstruction in 9000–5000 years before present. $\bar{\alpha}$ are mean pair-wise differences, indicating systematic mismatches. $\tilde{\delta}$ are predictions for $s_i - \varphi_i$ in a hypothetical new proxy site (i.e., any grid cell). Values above zero indicate an overestimation compared to proxies, and below zero underestimation.

Rather, our results will hopefully help to inform users of such products on where to apply them.

Previous evaluations of transient paleoclimate simulations compared temperature variability to proxies based on frequency bands. Model–proxy agreement strongly depends on scale (Laepple et al. 2023). On the regional scale, climate models generally struggle to reproduce the multidecadal and centennial variability seen in proxies of terrestrial temperatures (Hébert et al. 2022) and sea surface temperatures (Laepple and Huybers 2014; Weitzel et al. 2024). In contrast, global mean surface temperature variability aligns reasonably well between proxies and simulations on all timescales (Zhu et al. 2019; Askjær et al. 2022). Finally, at the local level, centennial-scale variability of temperature and precipitation is typically much higher in proxies than in simulations (Dee et al. 2017; Laepple et al. 2023). This mismatch on variability on local scale was not only found in frequency bands but also in direct timeseries comparison (Nicholson et al. 2019). Consequently, our results support previous findings from another methodological angle.

Possible reasons for model–proxy agreement

While mid-Holocene climate changes were undoubtedly large enough to be relevant for early farmers in Europe (Sánchez Goñi et al. 2016; Warden et al. 2017; Colledge et al. 2019; Betti et al. 2020), in comparison to late-glacial processes, they are of small magnitude and extent, with regionally varying impact (e.g., Seppä et al. 2007; Roberts et al. 2011; Samartin et al. 2017). The HadCM3B and TraCE21k simulations were not specifically designed to resolve these changes in mid-Holocene Europe but rather aimed at reproducing larger-scale patterns. Nonetheless, both models capture a proxy-derived winter warming trend at higher latitudes (Fig. 3). Notably, the winter warming trend we found in both proxies and models is completed around 6000 years BP (Fig. 3). From 6000 years BP to the pre-industrial time, studies show a winter cooling trend in northern Europe (Bonfils et al. 2004; Mauri et al. 2014; Strandberg et al. 2022). We therefore interpret the winter warming as largely a result of increasing winter insolation due to changes in the Earth's orbit (Zhang et al. 2022), a driving force well represented in climate models. In addition, the retreat of continental ice sheets might have contributed to warming winters up until 7000 years BP (Baker et al. 2017).

As paleoclimate simulation output is downsampled and bias-corrected with modern-day climate, it carries a strong topographic signal, which partially explains why there are no striking differences between simulated paleoclimate and modern-day climate in our analysis. Elevation of proxy sites did not help to statistically explain model–proxy differences. This indicates that the topographic signal is already in the gridded datasets from downscaling and bias-correcting post-processing. We did not, however, test this hypothesis by analyzing non-bias-corrected paleoclimate outputs.

Transient vs. snapshot simulations

In contrast to the transient TraCE21k simulation by Liu et al. (2009), which ran continuously, the HadCM3B dataset by Armstrong et al. (2019) originates from snapshot simulations that were splined together in post-processing. Transient simulations are computationally very expensive but intended to represent continuous climate evolution more mechanistically. In contrast, snapshot simulations can be efficiently computed in parallel. In principle, one would expect a transient simulation to represent centennial-scale climate fluctuations better, in particular because of lag effects (Zhu et al. 2019) and drivers such as meltwater pulses, which are included in TraCE21k. The slightly stronger correlations with proxy-based reconstructions in CHELSA-TraCE21k (Fig. 2) could be attributable to its transience, but we consider the overall performance difference to HadCM3B to be marginal.

Possible reasons for model–proxy disagreement

Reasons for a lack of correlation can be sought in insufficiencies of the climate models, in inaccuracies from proxies, and in misalignment between the two. TraCE21k simulations were transient but their spatial resolution was probably too low to capture regional-scale climate dynamics. More recent studies using higher-resolution regional climate models promise to better capture spatial patterns (usually on 10–25 km grids; Ludwig et al. 2019): for example representing more realistic soil moisture availability (Russo and Cubasch 2016; Russo et al. 2022), topographic features (Arthur et al. 2023), or Neolithic land use effects (Strandberg et al. 2022). Also the temporal variability in Holocene climate remains enigmatic with regard to drivers and mechanisms on multi-centennial scale (Askjær et al. 2022). This variability may be caused by external drivers such as oscillations in solar irradiance (Mayewski et al. 2004; Wu et al. 2018), meltwater pulses (Gregoire and Morrill 2021), and anthropogenic land clearance (Hopcroft et al. 2023) or result from internal feedback mechanisms such as the Atlantic Meridional Overturning Circulation (Stuiver et al. 1995; McDermott et al. 2001) or Arctic sea ice (Hörner et al. 2017). These interactions and their regional-scale impacts are far from being fully understood.

Bias and uncertainty of proxy-based reconstructions may also explain their lack of correlation with model-based outputs. A number of assumptions are required for deriving absolute temperature and precipitation values from pollen and chironomid larvae (Brooks 2006; Chevalier et al. 2020): for example, stationarity of the transfer function, representativeness of the modern-day reference dataset, negligibility of human land-use effects, and so on. For pollen, it is often unclear which geographical extent a given sample represents (Theuerkauf et al. 2016). In addition, dating precision in proxies may be too low to resolve the climate change signals of interest (Mayewski et al. 2004; Blaauw 2012), which, if not taken into account, can skew comparison with climate simulations. All this makes it difficult to distinguish signal from noise in proxy-derived climate timeseries without considering the particularities of each proxy site. Consequently, a database of proxy-based timeseries such as Temp12k (Kaufman et al. 2020a) is undoubtedly able to demonstrate climate trends on multimillennial and continental-to-global scale (e.g., Kaufman et al. 2020b) but cannot necessarily be expected to represent climate dynamics on the scale relevant for cultural and ecological changes in mid-Holocene Europe.

Avenues for improving model–proxy agreement

The high uncertainty associated with single climate values ($\tilde{\delta}$ in Fig. 4) limits the usefulness of the gridded datasets as predictor variables in statistical models. The wide credibility intervals not only stem from accounting for measurement errors in proxy-based reconstructions but also from our aggregating over the whole of Europe and the mid-Holocene. Modeling pair-wise differences at a finer spatiotemporal resolution could potentially increase precision of estimating model–proxy differences. Erb et al. (2022) used a comparable data assimilation approach to fuse proxy data with information from climate models on a global scale. For such data assimilation on the regional scale, we believe it would not only need more proxy sites but also error quantifications of proxy transfer functions that are more granular than the conservative standard deviations we applied blanketly to all proxy reconstructions. The development of probabilistic transfer functions (Chevalier et al. 2014; Chevalier 2022) is a promising step in that direction.

On the other hand, paleoclimate simulations could be improved in different ways, for example by increasing spatial resolution (the models analyzed here operate on c. 3.75° resolution); by incorporating relevant forcings, such as meltwater fluxes (here only included in one climate model), volcanic eruptions, methane, or solar output (Ellerhoff et al. 2022; van Dijk et al. 2024); by refining atmosphere–biosphere interactions (Hopcroft et al. 2023); or by improving downscaling, for example using regional climate models (Maraun et al. 2017; Ludwig et al. 2019). In addition, the stationarity assumption of delta bias correction fails to account for any paleoclimate dynamics (Maraun 2016). Here, data assimilation can help to reduce model bias in a dynamic way (Fallah et al. 2018; Jonkers et al. 2021).

Reliability and scale

Returning to the introductory question of how reliable simulation-derived datasets are and at which scales they can be used. Pair-wise differences showed substantial discrepancies towards proxies for all gridded datasets (Fig. 4), but — on this level of data aggregation — we cannot say if proxies or bias-corrected simulation outputs are closer to true paleoclimatic conditions. Similar uncertainty remains with climate change signals. Here, evaluating climate dynamics on smaller scales than multimillennial protracted deglaciation processes requires scrutinizing the accuracy of proxy-based inference (Jacobson et al. 2024) as well as a better understanding of the physical mechanisms themselves (Wanner et al. 2008; Renssen et al. 2009). Regarding spatial scale, statistical downscaling helps resolve local-scale topographic features but modeling climate patterns specific to mid-Holocene Europe in simulations remains subject to ongoing research efforts (e.g., Bonfils et al. 2004; Samartin et al. 2017; Russo et al. 2022). For matching ecological or archaeological occurrence data to paleoclimate, it is probably more important to resolve relatively certain local-scale topography effects rather than uncertain climate change signals. For driving analyses of biogeographical or sociocultural dynamics in mid-Holocene Europe, we see little additional benefit in the analyzed time-continuous datasets over using high-resolution modern-day climate or downscaled output from relevant snapshot simulations (e.g., 6000 or 8500 years BP Brierley and Zhang 2021; Gregoire and Morrill 2021).

Conclusions

Downscaled and bias-corrected output of paleoclimate simulations are an accessible solution to drive biogeographical analyses. In the case of mid-Holocene Europe (9000–5000 years BP), the two model outputs analyzed (HadCM3B and CHELSA-TraCE21k) reproduce temporal changes of quantitative proxy-based reconstructions only partially: i.e., a multimillennial postglacial winter warming in the northern half of Europe. At this stage, we therefore discourage using them to drive analyses aimed at understanding the effects of climate *change* in mid-Holocene Europe, at least without further analyses. We have not analyzed in depth if there are parts or regions of Europe where proxies and climate models show similar patterns not apparent in correlation coefficients. In general, each dataset, including modern-day climate, systematically over- or underestimates certain paleoclimate variables, without any one being consistently superior in matching proxy-based reconstructions. Accordingly, we cannot recommend one over the other, and, in fact, modern-day climate may not necessarily be a worse choice than paleoclimatic model outputs, in light of the given uncertainties. In general, we advise users not to rely on only one gridded paleoclimatic dataset. Continuous improvement of climate models, increasing availability of proxy data, improved methods for proxy-based reconstructions, and data assimilation approaches promise to provide more precise and accurate datasets in the near future.

Acknowledgements

We thank Angelica Feurdean for generously sharing the data from Feurdean et al. (2008) with us and Dirk Karger for advice on deriving temperature variables. Thanks to Elena Marinova for helping to reach out to original authors of proxy publications in the beginning of the project. The constructive comments from two anonymous reviewers were of great help to improve the manuscript. We also thank M.E. Traylor for copy-editing.

Author Contributions

. . . filled out at the proofing stage. . .

Statements and Declarations

Ethical considerations Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Being part in the SUSTAIN project, this work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 865515—ERC-2019-CoG awarded to M.I.B.). M.G.T. and A.T. are also supported by ERC Horizon 2020 research and innovation programme grant agreement: no. 951385 (COREX) awarded to M.G.T.. M.G.T. is also supported by ERC Horizon 2020 grants: no. 324202 (NeoMilk), no. 788616 (YMPACT), and by a Wellcome Senior Research Fellowship Grant 100719/Z/12/Z.

Data Availability

Code and data to reproduce the analysis are available for download here: https://doi.org/10.5281/zenodo.10608482

References

Allen JRM, Forrest M, Hickler T, Singarayer JS, Valdes PJ and Huntley B (2020) Global vegetation patterns of the past 140,000 years. *Journal of Biogeography* DOI:10.1111/jbi.13930.

Armstrong E, Hopcroft PO and Valdes PJ (2019) A simulated Northern Hemisphere terrestrial climate dataset for the past 60,000 years. *Scientific Data* 6(1): 265. DOI:10.1038/s41597-019-0277-1.

Arthur F, Roche DM, Fyfe R, Quiquet A and Renssen H (2023) Simulations of the Holocene climate in Europe using dynamical downscaling within the iLOVECLIM model (version 1.1). *Climate of the Past* 19(1): 87–106. DOI:10.5194/cp-19-87-2023. URL `https://cp.copernicus.org/articles/19/87/2023/`.

Askjær TG, Zhang Q, Schenk F, Ljungqvist FC, Lu Z, Brierley CM, Hopcroft PO, Jungclaus J, Shi X, Lohmann G, Sun W, Liu J, Braconnot P, Otto-Bliesner BL, Wu Z, Yin Q, Kang Y and Yang H (2022) Multi-centennial Holocene climate variability in proxy records and transient model simulations. *Quaternary Science Reviews* 296: 107801. DOI:10.1016/j.quascirev.2022.107801.

Baker JL, Lachniet MS, Chervyatsova O, Asmerom Y and Polyak VJ (2017) Holocene warming in western continental Eurasia driven by glacial retreat and greenhouse forcing. *Nature Geoscience* 10(6): 430–435. DOI:10.1038/ngeo2953.

Banks WE, Antunes N, Rigaud S and Francesco d'Errico (2013) Ecological constraints on the first prehistoric farmers in Europe. *Journal of Archaeological Science* 40(6): 2746–2753. DOI:10.1016/j.jas.2013.02.013.

Banks WE, d'Errico F, Peterson AT, Vanhaeren M, Kageyama M, Sepulchre P, Ramstein G, Jost A and Lunt D (2008) Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. *Journal of Archaeological Science* 35(2): 481–491. DOI:10.1016/j.jas.2007.05.011.

Betti L, Beyer RM, Jones ER, Eriksson A, Tassi F, Siska V, Leonardi M, Maisano Delser P, Bentley LK, Nigst PR, Stock JT, Pinhasi R and Manica A (2020) Climate shaped how Neolithic farmers and European hunter-gatherers interacted after a major slowdown from 6,100 BCE to 4,500 BCE. *Nature Human Behaviour* 4(10): 1004–1010. DOI:10.1038/s41562-020-0897-7.

Beyer R, Krapp M and Manica A (2020) An empirical evaluation of bias correction methods for palaeoclimate simulations. *Climate of the Past* 16(4): 1493–1508. DOI:10.5194/cp-16-1493-2020.

Blaauw M (2012) Out of tune: the dangers of aligning proxy archives. *Quaternary Science Reviews* 36: 38–49. DOI:e10.1016/j.quascirev.2010.11.012.

Bonfils C, de Noblet-Ducoudré N, Guiot J and Bartlein P (2004) Some mechanisms of mid-Holocene climate change in Europe, inferred from comparing PMIP models to data. *Climate Dynamics* 23(1): 79–98. DOI:10.1007/s00382-004-0425-x.

Brierley CM and Zhang Q (2021) Simulating the mid-Holocene in PMIP. *Past Global Changes Magazine* 29(2): 74–75. DOI:10.22498/pages.29.2.74.

Brooks SJ (2006) Fossil midges (Diptera: Chironomidae) as palaeoclimatic indicators for the Eurasian region. *Quaternary Science Reviews* 25(15): 1894–1910. DOI:10.1016/j.quascirev.2005.03.021. Quaternary beetle research: the state of the art.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P and Riddell A (2017) Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1): 1–32. DOI:10.18637/jss.v076.i01.

Chevalier M (2022) crestr: an R package to perform probabilistic climate reconstructions from palaeoecological datasets. *Climate of the Past* 18(4): 821–844. DOI:10.5194/cp-18-821-2022.

Chevalier M, Cheddadi R and Chase BM (2014) CREST (Climate REconstruction SofTware): a probability density function (PDF)-based quantitative climate reconstruction method. *Climate of the Past* 10(6): 2081–2098. DOI:10.5194/cp-10-2081-2014.

Chevalier M, Davis BA, Heiri O, Seppä H, Chase BM, Gajewski K, Lacourse T, Telford RJ, Finsinger W, Guiot J, Kühl N, Maezumi SY, Tipton JR, Carter VA, Brussel T, Phelps LN, Dawson A, Zanon M, Vallé F, Nolan C, Mauri A, de Vernal A, Izumi K, Holmström L, Marsicek J, Goring S, Sommer PS, Chaput M and Kupriyanov D (2020) Pollen-based climate reconstruction techniques for late Quaternary studies. *Earth-Science Reviews* 210: 103384. DOI:10.1016/j.earscirev.2020.103384.

Colledge S, Conolly J, Crema E and Shennan S (2019) Neolithic population crash in northwest Europe associated with agricultural crisis. *Quaternary Research* 92(3): 686–707. DOI:10.1017/qua.2019.42.

Conolly J, Colledge S and Shennan S (2008) Founder effect, drift, and adaptive change in domestic crop use in early Neolithic Europe. *Journal of Archaeological Science* 35(10): 2797–2804. DOI:10.1016/j.jas.2008.05.006.

Dee S, Parsons L, Loope G, Overpeck J, Ault T and Emile-Geay J (2017) Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability. *Earth and Planetary Science Letters* 476: 34–46. DOI:10.1016/j.epsl.2017.07.036.

Divíšek J, Večeřa M, Welk E, Danihelka J, Chytrý K, Douda J and Chytrý M (2022) Origin of the central European steppe flora: insights from palaeodistribution modelling and migration simulations. *Ecography* 2022(12): e06293. DOI:10.1111/ecog.06293.

d'Alpoim Guedes JA, Crabtree SA, Bocinsky RK and Kohler TA (2016) Twenty-first century approaches to ancient problems: Climate and society. *Proceedings of the National Academy of Sciences* 113(51): 14483–14491. DOI:10.1073/pnas.1616188113.

Ellerhoff B, Kirschner MJ, Ziegler E, Holloway MD, Sime L and Rehfeld K (2022) Contrasting state-dependent effects of natural forcing on global and local climate variability. *Geophysical Research Letters* 49(10): e2022GL098335. DOI:e10.1029/2022GL098335. E2022GL098335 2022GL098335.

Ellis EC (2021) Land use and ecological change: A 12,000-year history. *Annual Review of Environment and Resources* 46(1): 1–33. DOI:10.1146/annurev-environ-012220-010822.

Erb MP, McKay NP, Steiger N, Dee S, Hancock C, Ivanovic RF, Gregoire LJ and Valdes P (2022) Reconstructing Holocene temperatures in time and space using paleoclimate data assimilation. *Climate of the Past* 18(12): 2599–2629. DOI:10.5194/cp-18-2599-2022.

Fallah B, Russo E, Acevedo W, Mauri A, Becker N and Cubasch U (2018) Towards high-resolution climate reconstruction using an off-line data assimilation and COSMO-CLM 5.00 model. *Climate of the Past* 14(9): 1345–1360. DOI:10.5194/cp-14-1345-2018.

Feurdean A, Klotz S, Mosbrugger V and Wohlfarth B (2008) Pollen-based quantitative reconstructions of Holocene climate variability in NW Romania. *Palaeogeography, Palaeoclimatology, Palaeoecology* 260(3): 494–504. DOI:10.1016/j.palaeo.2007.12.014.

Franklin J, Potts AJ, Fisher EC, Cowling RM and Marean CW (2015) Paleodistribution modeling in archaeology and paleoanthropology. *Quaternary Science Reviews* 110: 1–14. DOI: 10.1016/j.quascirev.2014.12.015.

Gabry J and Češnovar R (2022) cmdstanr: R Interface to 'CmdStan'. URL `https://mc-stan.org/cmdstanr/`.

Githumbi E, Pirzamanbein B, Lindström J, Poska A, Fyfe R, Mazier F, Nielsen AB, Sugita S, Trondman AK, Woodbridge J and Gaillard MJ (2022) Pollen-based maps of past regional vegetation cover in Europe over 12 millennia—evaluation and potential. *Frontiers in Ecology and Evolution* 10. DOI:10.3389/fevo.2022.795794.

Gregoire LJ and Morrill C (2021) The 8.2 kyr event: Benchmarking climate model sensitivity to ice-sheet melt. *Past Global Changes Magazine* 29(2): 76–77. DOI:10.22498/pages.29.2.76.

He F (2011) *Simulating transient climate evolution of the last deglaciation with CCSM 3*. PhD Thesis, University of Wisconsin-Madison. URL `http://www.aos.wisc.edu/uwaosjournal/Volume15/He_PhD_Thesis.pdf`.

Hijmans RJ, Cameron SE, Parra JL, Jones PG and Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15): 1965–1978. DOI:10.1002/joc.1276.

Hopcroft PO, Valdes PJ, Shuman BN, Toohey M and Sigl M (2023) Relative importance of forcings and feedbacks in the Holocene temperature conundrum. *Quaternary Science Reviews* 319: 108322. DOI:10.1016/j.quascirev.2023.108322.

Huntley B, Allen JRM, Forrest M, Hickler T, Ohlemüller R, Singarayer JS and Valdes PJ (2023) Global biome patterns of the Middle and Late Pleistocene. *Journal of Biogeography* 50(8): 1352–1372. DOI:10.1111/jbi.14619.

Hébert R, Herzschuh U and Laepple T (2022) Millennial-scale climate variability over land overprinted by ocean temperature fluctuations. *Nature Geoscience* 15(11): 899–905. DOI: 10.1038/s41561-022-01056-4.

Hörner T, Stein R and Fahl K (2017) Evidence for Holocene centennial variability in sea ice cover based on IP25 biomarker reconstruction in the southern Kara Sea (Arctic Ocean). *Geo-Marine Letters* 37(5): 515–526. DOI:10.1007/s00367-017-0501-y.

Jacobson MJ, Seguin J and Finné M (2024) Holocene hydroclimate synthesis of the Aegean: Diverging patterns, dry periods and implications for climate-society interactions. *The Holocene* 34(12): 1807–1823. DOI:10.1177/09596836241275028.

Jonkers L, Bothe O and Kucera M (2021) Preface: Advances in paleoclimate data synthesis and analysis of associated uncertainty: towards data–model integration to understand the climate. *Climate of the Past* 17(6): 2577–2581. DOI:10.5194/cp-17-2577-2021.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP and Kessler M (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122. DOI:10.1038/sdata.2017.122.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP and Kessler M (2021) Data from: Climatologies at high resolution for the earth's land surface areas. DOI:10.16904/envidat.228.

Karger DN, Nobis MP, Normand S, Graham CH and Zimmermann NE (2023) CHELSA-TraCE21k – high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past* 19(2): 439–456. DOI: 10.5194/cp-19-439-2023. URL `https://cp.copernicus.org/articles/19/439/2023/`.

Kaufman D, McKay N, Routson C, Erb M, Davis B, Heiri O, Jaccard S, Tierney J, Dätwyler C, Axford Y, Brussel T, Cartapanis O, Chase B, Dawson A, de Vernal A, Engels S, Jonkers L, Marsicek J, Moffa-Sánchez P, Morrill C, Orsi A, Rehfeld K, Saunders K, Sommer PS, Thomas E, Tonello M, Tóth M, Vachula R, Andreev A, Bertrand S, Biskaborn B, Bringué M, Brooks S, Caniupán M, Chevalier M, Cwynar L, Emile-Geay J, Fegyveresi J, Feurdean A, Finsinger W, Fortin MC, Foster L, Fox M, Gajewski K, Grosjean M, Hausmann S, Heinrichs M, Holmes N, Ilyashuk B, Ilyashuk E, Juggins S, Khider D, Koinig K, Langdon P, Larocque-Tobler I, Li J, Lotter A, Luoto T, Mackay A, Magyari E, Malevich S, Mark B, Massaferro J, Montade V, Nazarova L, Novenko E, Pařil P, Pearson E, Peros M, Pienitz R, Płóciennik M, Porinchu D, Potito A, Rees A, Reinemann S, Roberts S, Rolland N, Salonen S, Self A, Seppä H, Shala S, St-Jacques JM, Stenni B, Syrykh L, Tarrats P, Taylor K, van den Bos V, Velle G, Wahl E, Walker I, Wilmshurst J, Zhang E and Zhilich S (2020a) A global database of Holocene paleotemperature records. *Scientific Data* 7(1). DOI: 10.1038/s41597-020-0445-3.

Kaufman D, McKay N, Routson C, Erb M, Dätwyler C, Sommer PS, Heiri O and Davis B (2020b) Holocene global mean surface temperature, a multi-method reconstruction approach. *Scientific Data* 7(1): 201. DOI:10.1038/s41597-020-0530-7.

Lacourse T and Gajewski K (2020) Current practices in building and reporting age-depth models. *Quaternary Research* 96: 28–38. DOI:10.1017/qua.2020.47.

Laepple T and Huybers P (2014) Ocean surface temperature variability: Large model–data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences* 111(47): 16682–16687. DOI:10.1073/pnas.1412077111.

Laepple T, Ziegler E, Weitzel N, Hébert R, Ellerhoff B, Schoch P, Martrat B, Bothe O, Moreno-Chamarro E, Chevalier M, Herbert A and Rehfeld K (2023) Regional but not global temperature variability underestimated by climate models at supradecadal timescales. *Nature Geoscience* 16(11): 958–966. DOI:10.1038/s41561-023-01299-9.

Liu Z, Otto-Bliesner BL, He F, Brady EC, Tomas R, Clark PU, Carlson AE, Lynch-Stieglitz J, Curry W, Brook E, Erickson D, Jacob R, Kutzbach J and Cheng J (2009) Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. *Science* 325(5938): 310–314. DOI:10.1126/science.1171041.

Lorenz DJ, Nieto-Lugilde D, Blois JL, Fitzpatrick MC and Williams JW (2016) Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100AD. *Scientific data* 3. DOI:10.1038/sdata.2016.48.

Ludwig P, Gómez-Navarro JJ, Pinto JG, Raible CC, Wagner S and Zorita E (2019) Perspectives of regional paleoclimate modeling. *Annals of the New York Academy of Sciences* 1436(1): 54–69. DOI:10.1111/nyas.13865.

Maraun D (2016) Bias correcting climate change simulations - a critical review. *Current Climate Change Reports* 2(4): 211–220. DOI:10.1007/s40641-016-0050-x.

Maraun D, Shepherd TG, Widmann M, Zappa G, Walton D, Gutiérrez JM, Hagemann S, Richter I, Soares PMM, Hall A and Mearns LO (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change* 7(11): 764–773. DOI: 10.1038/nclimate3418.

Marcott SA, Shakun JD, Clark PU and Mix AC (2013) A reconstruction of regional and global temperature for the past 11,300 years. *Science* 339(6124): 1198–1201. DOI: 10.1126/science.1228026.

Marsicek J, Shuman BN, Bartlein PJ, Shafer SL and Brewer S (2018) Reconciling divergent trends and millennial variations in Holocene temperatures. *Nature* 554(7690): 92–96. DOI: 10.1038/nature25464.

Mauri A, Davis BAS, Collins PM and Kaplan JO (2014) The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data–model comparison. *Climate of the Past* 10(5): 1925–1938. DOI:10.5194/cp-10-1925-2014.

Mayewski PA, Rohling EE, Curt Stager J, Karlén W, Maasch KA, Meeker LD, Meyerson EA, Gasse F, van Kreveld S, Holmgren K and et al (2004) Holocene climate variability. *Quaternary Research* 62(3): 243–255. DOI:10.1016/j.yqres.2004.07.001.

McDermott F, Mattey DP and Hawkesworth C (2001) Centennial-scale Holocene climate variability revealed by a high-resolution speleothem δ^{18}O record from SW Ireland. *Science* 294(5545): 1328–1331. DOI:10.1126/science.1063678.

Miller PA, Giesecke T, Hickler T, Bradshaw RHW, Smith B, Seppä H, Valdes PJ and Sykes MT (2008) Exploring climatic and biotic controls on Holocene vegetation change in Fennoscandia. *Journal of Ecology* 96(2): 247–259. DOI:10.1111/j.1365-2745.2007.01342.x.

Myers CE, Stigall AL and Lieberman BS (2015) PaleoENM: applying ecological niche modeling to the fossil record. *Paleobiology* 41(2): 226–244. DOI:10.1017/pab.2014.19.

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

Nicholson C, Minckley TA and Shinker JJ (2019) Validating CCSM3 paleoclimate data using pollen-based reconstruction in the intermountain west. *Quaternary Science Reviews* 222: 105911. DOI:10.1016/j.quascirev.2019.105911.

Overpeck JT, Webb T and Prentice IC (1985) Quantitative interpretation of fossil pollen spectra: Dissimilarity coefficients and the method of modern analogs. *Quaternary Research* 23(1): 87–108. DOI:10.1016/0033-5894(85)90074-2.

Quiquet A, Roche DM, Dumas C and Paillard D (2018) Online dynamical downscaling of temperature and precipitation within the iLOVECLIM model (version 1.1). *Geoscientific Model Development* 11(1): 453–466. DOI:10.5194/gmd-11-453-2018.

R Core Team (2023) R: A language and environment for statistical computing. URL `https://www.r-project.org/`.

Rao Z, Tian Y, Guang K, Wei S, Guo H, Feng Z, Zhao L and Li Y (2022) Pollen data as a temperature indicator in the late Holocene: A review of results on regional, continental and global scales. *Frontiers in Earth Science* 10. DOI:10.3389/feart.2022.845650.

Renssen H, Seppä H, Heiri O, Roche DM, Goosse H and Fichefet T (2009) The spatial and temporal complexity of the Holocene thermal maximum. *Nature Geoscience* 2(6): 411–414. DOI:10.1038/ngeo513.

Roberts N, Brayshaw D, Kuzucuoğlu C, Perez R and Sadori L (2011) The mid-Holocene climatic transition in the Mediterranean: Causes and consequences. *The Holocene* 21(1): 3–13. DOI: 10.1177/0959683610388058.

Routson CC, McKay NP, Kaufman DS, Erb MP, Goosse H, Shuman BN, Rodysill JR and Ault T (2019) Mid-latitude net precipitation decreased with Arctic warming during the Holocene. *Nature* 568(7750): 83–87. DOI:10.1038/s41586-019-1060-3.

Russo E and Cubasch U (2016) Mid-to-late Holocene temperature evolution and atmospheric dynamics over Europe in regional model simulations. *Climate of the Past* 12(8): 1645–1662. DOI:10.5194/cp-12-1645-2016.

Russo E, Fallah B, Ludwig P, Karremann M and Raible CC (2022) The long-standing dilemma of European summer temperatures at the mid-Holocene and other considerations on learning from the past for the future using a regional climate model. *Climate of the Past* 18(4): 895–909. DOI:10.5194/cp-18-895-2022.

Samartin S, Heiri O, Joos F, Renssen H, Franke J, Brönnimann S and Tinner W (2017) Warm Mediterranean mid-Holocene summers inferred from fossil midge assemblages. *Nature Geoscience* 10(3): 207–212. DOI:10.1038/ngeo2891.

Seppä H, Birks HJB, Giesecke T, Hammarlund D, Alenius T, Antonsson K, Bjune AE, Heikkilä M, MacDonald GM, Ojala AEK, Telford RJ and Veski S (2007) Spatial structure of the 8200 cal yr BP event in northern Europe. *Climate of the Past* 3(2): 225–236. DOI: 10.5194/cp-3-225-2007.

Shennan S (2018) *The First Farmers of Europe*. Cambridge University Press. ISBN 978-1-108-43521-5. DOI:10.1017/9781108386029.

Singarayer JS and Valdes PJ (2010) High-latitude climate sensitivity to ice-sheet forcing over the last 120kyr. *Quaternary Science Reviews* 29(1): 43–55. DOI:10.1016/j.quascirev.2009.10.011. Climate of the Last Million Years: New Insights from EPICA and Other Records.

Spearman C (1904) The proof and measurement of association between two things. *The American Journal of Psychology* 15(1): 72–101. DOI:10.2307/1412159.

Stan Development Team (2023) *Stan Modeling Language Users Guide and Reference Manual, 2.32*. URL `https://mc-stan.org`.

Strandberg G, Lindström J, Poska A, Zhang Q, Fyfe R, Githumbi E, Kjellström E, Mazier F, Nielsen AB, Sugita S, Trondman AK, Woodbridge J and Gaillard MJ (2022) Mid-Holocene European climate revisited: New high-resolution regional climate model simulations using pollen-based land-cover. *Quaternary Science Reviews* 281: 107431. DOI:e10.1016/j.quascirev.2022.107431.

Stuiver M, Grootes PM and Braziunas TF (1995) The GISP2 δ^{18}O climate record of the past 16,500 years and the role of the sun, ocean, and volcanoes. *Quaternary Research* 44(3): 341–354. DOI:e10.1006/qres.1995.1079.

Sundqvist HS, Kaufman DS, McKay NP, Balascio NL, Briner JP, Cwynar LC, Sejrup HP, Seppä H, Subetto DA, Andrews JT, Axford Y, Bakke J, Birks HJB, Brooks SJ, de Vernal A, Jennings AE, Ljungqvist FC, Rühland KM, Saenger C, Smol JP and Viau AE (2014) Arctic Holocene proxy climate database – new approaches to assessing geochronological accuracy and encoding climate variables. *Climate of the Past* 10(4): 1605–1631. DOI: 10.5194/cp-10-1605-2014.

Sánchez Goñi MF, Ortu E, Banks WE, Giraudeau J, Leroyer C and Hanquiez V (2016) The expansion of Central and Northern European Neolithic populations was associated with a multi-century warm winter and wetter climate. *The Holocene* 26(8): 1188–1199. DOI: 10.1177/0959683616638435.

Theuerkauf M, Couwenberg J, Kuparinen A and Liebscher V (2016) A matter of dispersal: REVEALSinR introduces state-of-the-art dispersal models to quantitative vegetation reconstruction. *Vegetation History and Archaeobotany* 25(6): 541–553. DOI:10.1007/s00334-016-0572-0.

Valdes PJ, Armstrong E, Badger MPS, Bradshaw CD, Bragg F, Crucifix M, Davies-Barnard T, Day JJ, Farnsworth A, Gordon C, Hopcroft PO, Kennedy AT, Lord NS, Lunt DJ, Marzocchi A, Parry LM, Pope V, Roberts WHG, Stone EJ, Tourte GJL and Williams JHT (2017) The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geoscientific Model Development* 10(10): 3715–3743. DOI:10.5194/gmd-10--3715-2017.

van Dijk EJC, Jungclaus J, Sigl M, Timmreck C and Krüger K (2024) High-frequency climate forcing causes prolonged cold periods in the Holocene. *Communications Earth & Environment* 5(1): 242. DOI:10.1038/s43247-024-01380-0.

Vehtari A, Simpson D, Gelman A, Yao Y and Gabry J (2024) Pareto smoothed importance sampling. *Journal of Machine Learning Research* 25(72): 1–58. URL `http://jmlr.org/papers/v25/19-556.html`.

Wanner H, Beer J, Bütikofer J, Crowley TJ, Cubasch U, Flückiger J, Goosse H, Grosjean M, Joos F, Kaplan JO, Küttel M, Müller SA, Prentice IC, Solomina O, Stocker TF, Tarasov P, Wagner M and Widmann M (2008) Mid- to Late Holocene climate change: an overview. *Quaternary Science Reviews* 27(19): 1791–1828. DOI:10.1016/j.quascirev.2008.06.013.

Wanner H, Solomina O, Grosjean M, Ritz SP and Jetel M (2011) Structure and origin of Holocene cold events. *Quaternary Science Reviews* 30(21): 3109–3123. DOI:e10.1016/j.quascirev.2011.07.010.

Warden L, Moros M, Neumann T, Shennan S, Timpson A, Manning K, Sollai M, Wacker L, Perner K, Häusler K, Leipe T, Zillén L, Kotilainen A, Jansen E, Schneider RR, Oeberst R, Arz H and Sinninghe Damsté JS (2017) Climate induced human demographic and cultural change in northern Europe during the mid-Holocene. *Scientific Reports* 7(1): 15251. DOI: 10.1038/s41598-017-14353-5.

Weitzel N, Andres H, Baudouin JP, Kapsch ML, Mikolajewicz U, Jonkers L, Bothe O, Ziegler E, Kleinen T, Paul A and Rehfeld K (2024) Towards spatio-temporal comparison of simulated and reconstructed sea surface temperatures for the last deglaciation. *Climate of the Past* 20(4): 865–890. DOI:10.5194/cp-20-865-2024.

Williams J and Shuman B (2008) Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quaternary Science Reviews* 27(7): 669–687. DOI:10.1016/j.quascirev.2008.01.004.

Williams JW, Shuman BN and Webb III T (2001) Dissimilarity analyses of late-Quaternary vegetation and climate in eastern North America. *Ecology* 82(12): 3346–3362. DOI: 10.1890/0012-9658(2001)082[3346:DAOLQV]2.0.CO;2.

Wu CJ, Usoskin IG, Krivova N, Kovaltsov GA, Baroni M, Bard E and Solanki SK (2018) Solar activity over nine millennia: A consistent multi-proxy reconstruction. *Astronomy and Astrophysics* 615: A93. DOI:10.1051/0004-6361/201731892.

Zhang W, Wu H, Cheng J, Geng J, Li Q, Sun Y, Yu Y, Lu H and Guo Z (2022) Holocene seasonal temperature evolution and spatial variability over the Northern Hemisphere landmass. *Nature Communications* 13(1): 5334. DOI:10.1038/s41467-022-33107-0.

Zhu F, Emile-Geay J, McKay NP, Hakim GJ, Khider D, Ault TR, Steig EJ, Dee S and Kirchner JW (2019) Climate models can correctly simulate the continuum of global-average temperature variability. *Proceedings of the National Academy of Sciences* 116(18): 8728–8733. DOI:10.1073/pnas.1809959116.

Ön ZB, Macdonald N, Akçer-Ön S and Greaves AM (2023) A novel Bayesian multilevel regression approach to the reconstruction of an eastern Mediterranean temperature record for the last 10,000 years. *The Holocene* DOI:10.1177/09596836231163508.