https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past Discussions

More is not always better: downscaling climate model outputs from 30 to 5-minute resolution has minimal impact on coherence with Late Quaternary proxies

Lucy Timbrell[1,2*], James Blinkhorn[1,2], Margherita Colucci[1,3], Michela Leonardi[3,4], Manuel Chevalier[5], Matt Grove[2], Eleanor Scerri[1,6,7] Andrea Manica[3]

[1] Human Palaeosystems Group, Max Planck Institute of Geoanthropology, Jena, Germany

[2] Department of Archaeology, Classics and Egyptology, University of Liverpool, U.K.

[3] Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge, U.K.

[4] Natural History Museum, London, U.K.

[5] Meteorology Department, University of Bonn, Bonn, Germany.

[6] Department of Classics and Archaeology, University of Malta, Malta

[7] Department of Prehistoric Archaeology, University of Cologne, Germany

Correspondence to: Lucy Timbrell (lucy.timbrell2@liverpool.ac.uk)

Abstract. Both proxies and models provide key resources to explore how palaeoenvironmental changes may have impacted diverse biotic communities and cultural processes. Whilst proxies provide the gold standard in reconstructing the local environment, they only provide point estimates for a limited number of locations; on the other hand, models have the potential to afford more extensive and standardised geographic coverage. A key decision when using model outputs is the appropriate geographic resolution to adopt; models are coarse scale, in the order of several arc degrees, and so their outputs are usually downscaled to a higher resolution. Most publicly available model time-series have been downscaled to 30 or 60 arc-minutes, but it is unclear whether such resolution is sufficient, or whether this may homogenise environments and mask the spatial variability that is often the primary subject of analysis. Here, we explore the impact of further downscaling model outputs from 30 to 5 arc-minutes using the delta method, which uses the difference between past and present model data sets to increase spatial resolution of simulations, in order to determine to what extent further downscaling captures climatic trends at the site-level, through direct comparison with proxy reconstructions. We use the output from the HadCM3 Global Circulation model for annual temperature, mean temperature of the warmest quarter, and annual precipitation, which we evaluated against a large empirical dataset of pollen-based reconstructions from across the Northern Hemisphere. Our results demonstrate that, overall, models tend to provide broadly similar accounts of past climate to that obtained from proxy reconstructions, with coherence tending to decline with age. However, our results imply that downscaling to a very fine scale has minimal to no effect on the coherence of model data with pollen records. Optimal spatial resolution is therefore likely to be highly dependent on specific research contexts and questions, with careful consideration required regarding the trade-off between highlighting local-scale variation and increasing potential error.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

1 Introduction

Realistic reconstructions of global paleoclimates are vital for modelling long-term evolutionary and ecological processes in fields like evolutionary biology, ecology, palaeontology, and archaeology. Recently, the production of high-resolution simulations, characterising climatic variables across vast time periods, have allowed for the production and analyses of time series similar to those produced using proxy data (e.g., Forham *et al.,* 2017; Armstrong *et al.,* 2019; Holden *et al.,* 2019; Beyer *et al.,* 2020; Brown *et al.,* 2020; Karger *et al.,* 2021; Krapp *et al.,* 2021; Timmerman *et al.,* 2022). Proxy records, such as those derived from pollen or other biomarkers, remain the gold standard for characterising past environments; however, in order to extrapolate beyond the core sites and across wider regions, often it is necessary to rely on modelled or simulated data. Openly accessible simulated datasets, such as those published by Beyer *et al.* (2020a), Krapp *et al.* (2021) and Barreto *et al.* (2023), and associated analytical packages (e.g., the analytical tool pastclim for manipulating and extracting modelled data; Leonardi *et al.,* 2023), are particularly useful for scientists interested in Middle-Late Pleistocene and Holocene timescales (e.g. Beyer *et al.,* 2021; Padilla-Iglesias *et al.,* 2022; Blinkhorn *et al.,* 2022; Leonardi *et al.,* 2022).

The chronological resolution of proxy records has dramatically improved in recent years, allowing for detailed reconstructions of climatic conditions through time. Yet rarely are proxy data in direct association with archaeological or palaeontological sites), nor do they consistently provide an absolute, linear, and standardised representation of past climate across large geographic areas. Proxy records also often provide relative estimates of past climate rather than absolute parameters, an issue highlighted in a synthesis of eastern African Late-Middle Pleistocene climate records by Timbrell *et al.* (2022), demonstrating that different proxy records – even from within a relatively spatiotemporally restricted region – can provide alternate ideas of relative ‘humidity’. This is the result of the variable nature of the data employed, which typically cannot be articulated as the climatic indicators and environmental parameters that are routinely applied when studying contemporary populations, such as in ecological niche models.

Modelled data have the potential to overcome these shortfalls, providing absolute values for parameters such as temperature, precipitation, and a range of derived bioclimatic indices (e.g., Hijmans *et al.,* 2005), as well as offering much wider spatial coverage of the landscape that can be directly related to specific study sites and the palaeoclimatic differences between them. However, the integration of modelled climate with observational data is not straightforward. For example, using simulations at a coarse resolution can produce biases when compared to on-site proxies due to the underlying complexity of the physical landscape, particularly in coastal regions and those in topographically complex regions (Maraun and Widmann, 2018). Resultant errors can be in the order of several degrees for temperature and tens of percent for precipitation, which could lead to substantially different biome classifications and estimations of ecologies experienced on the ground (Kottek *et al.,* 2006).

In many cases, increasing the spatial resolution of climatic models in better accordance with real-life environmental dynamics may be required for more accurate characterisations and to accommodate patterns of landscape diversity that affect climatic conditions at local scales.

High resolution simulations are desired yet difficult to obtain due to computational costs. Most of the recently produced time series of palaeoclimate outputs have been downscaled from the native resolution of the models (usually in the order of 2 or 3 arc-degrees) to a higher resolution of 30 arc-minutes (Beyer *et al.* 2020; Krapp *et al.* 2021). Different methods exist to increase the spatial resolution of model simulations; these include the delta method, generalised additive models (GAMs), and quantile mapping. These are all aimed at minimising biases in models, characterised as differences in statistical distributions between observed and simulated series. Analyses by Beyer *et al.* (2020b) comparing debiased simulation data and empirical reconstructions at 30-minute resolution indicate the effectiveness of the delta method, which generally produced the most accurate simulation, though with substantial spatial and temporal variation in model performance. To debias simulations, delta-downscaling uses a map of local differences between observed and modelled values in the present day to correct for bias in the past (Maraun and Widmann, 2018). In this sense, the method assumes that biases are location specific and constant over time. Delta-downscaling can account for some climatic variations in relation to the underlying landscape, such as capturing some of the effects of topography on temperature and rainfall, which can be useful in certain analyses of past processes and dynamics. However, it is currently unclear what is a desirable level of downscaling. Recently a resolution of 1km was obtained for the TRACE21K simulations using the CHELSA algorithm (Karger *et al.* 2023), predicting very high-resolution climate for every 100 years for the last 21,000 years. Some studies support that finer-scale simulations have higher predictive power in species distribution models of modern populations (Chauvier *et al*. 2022; Ozdemir 2024), though whether such accuracy can be extended to predicted distributions in the past or future is unclear, particularly due to the assumptions of the delta-downscaling method that local biases remain constant through time (Franklin *et al.* 2015). There is little consensus regarding the choice of the optimal spatial resolution for analysis of past climate, nor whether downscaling to very fine scales is indeed appropriate for capturing localised climatic dynamics to a similar level as that provided by proxies; the gold standard for capturing climate variability in specific locations through time.

Previous studies have produced varied results when comparing the climatic time series produced by model simulations with proxy-based reconstructions. Some find that simulations and reconstructions reproduce similar major changes in temperature at large spatial scales (Fernández-Donado *et al.,* 2013; Zhu *et al.,* 2019), whilst others suggest divergence (Laepple and Huybers 2014; Rehfeld *et al.,* 2018). A recent meta-analysis by Laepple *et al.* (2023) found that studies in the Northern Hemisphere (where data are more abundant) have mixed results, suggesting potential areas of mismatch at local and regional scales. These authors suggest that shortcomings in both model simulations and proxy reconstructions may contribute to this divergence with models being less efficient at simulating local and regional temperature variability at relatively long timescales

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

and methods of temperature reconstruction from proxies facing systematic deficiencies, though stronger emphasis is placed on the former.

Here we test whether further downscaling of climatic models to relatively high resolution (5-min) leads to increased agreement with empirical reconstructions from proxies than relatively coarser models (30-min). We present new downscaled climatic models for 17 bioclimatic variables and apply a new suite of functions in the *pastclim* R package (Leonardi *et al.,* 2023) for delta-downscaling climatic simulations from Beyer *et al.* (2020a). We provide an assessment of the 2,592 Northern Hemisphere records for the last 30,000 years available from LegacyClimate 1.0 (Herzschuh *et al.* 2023), a pollen-based database reconstructing past annual temperature and precipitation and July temperature, that can be directly compared to our model datasets at coarse- and fine-grained spatial resolution. Our work quantifies the divergence between the time series produced using different modelled data at varied spatial resolution, with our results ultimately endorsing the use of modelled data in the absence of high-resolution proxies with careful consideration as to the most appropriate resolution for analysis.

2 Materials and methods

2.1 Climate models

We applied the global temperature and precipitation simulations by Beyer *et al.* (2020a) within the *pastclim* R package (Leonardi *et al.,* 2023), published at 0.5° resolution in 1 to 2-thousand-year time slices spanning the past 120 thousand years. The original simulations were delta-downscaled and de-biased based on the HadCM3 general circulations model (Singarayer and Valdes, 2010; Singarayer and Burrough, 2015; Valdes *et al.,* 2017) using the Climate Research Unit Global Climate Dataset (CRU) as the modern climatic reference (Mitchell and Jones, 2005). We used *pastclim* to further downscale monthly temperature and precipitation variables using high-resolution (5 minute) modern observations from WorldClim2 (Fick and Hijmans, 2017) and a global relief map from ETOPO2022 (NOAA National Centres for Environmental Information, 2022) to reconstruct past coastlines following sea level change (Spratt and Lisiecki, 2016). We selected WorldClim2 as the modern reference as the transfer functions used in the LegacyClimate1.0 dataset were also derived from this dataset (at 30-minute resolution), allowing us to control for the effects of the modern data used for debiasing on our results.

Downscaling is performed one monthly variable at a time (i.e., January temperature) by taking the coarse simulations from Beyer *et al.* (2020a) with the corresponding set of high-resolution modern simulations from WorldClim2 (Fick and Hijmans, 2017) and equally high-resolution global relief map (NOAA National Centres for Environmental Information, 2022). Through integrating both bathymetric and topographic values for masking sea level changes, a delta raster is computed, adding the difference between past and present-day simulated climate to present-day observed climate (Beyer et al. 2020a). The delta method therefore assumes that local (i.e. grid-cell-specific) model biases are constant over time (Maraun and Widmann, 2018).

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past
Discussions

For temperature variables, the bias in a geographical location x is given by the difference between present-day observed and simulated temperature. Downscaled temperature in x at some time t in the past is thus estimated as

$$T_{sim}^{DM}(x,t) \coloneqq T_{sim}^{raw}(x,t) + \left(T_{obs}(x,0) - T_{sim}^{raw}(x,0)\right) \qquad (1)$$

Precipitation is lower bounded by zero and covers different orders of magnitude across different regions compared to temperature. Multiplying rather than adding the bias correction is therefore more common when applying the delta method for precipitation, which corresponds to applying the simulated relative change to the observations (Maraun and Widmann, 2018). This method can therefore be hypersensitive in drylands. Analogous to temperature, downscaled precipitation is estimated as

$$P_{sim}^{DM}(x,t) \coloneqq P_{sim}^{raw}(x,t) \cdot \frac{P_{obs}(x,0)}{P_{sim}^{raw}(x,0)} \qquad (2)$$

The resulting monthly datasets are then used within *pastclim* to recompute the 17 bioclimatic variables available in the original dataset (Supplementary Table S1), with mean annual temperature (bio01), mean temperature of the warmest quarter (bio10) and total annual precipitation (bio12) extracted here for further analysis given their relevance to the variables captured by the proxy reconstructions employed.

Considering that the original Beyer *et al.* (2020a) model was debiased using modern reconstructions from CRU, rather than WorldClim2 that was used here for the downscaling, we also tested the comparability of the modelled data at the native resolution, presented in Beyer *et al* (2020a), but debiased using WorldClim2. This ensured that any variability between the performance of models at different spatial resolutions was captured, and not compounded by that of differences between modern data used for debiasing, particularly considering that WorldClim2 was also used to calibrate the LegacyClimate1.0 dataset.

Interpolating over small spatial extents can lead to the introduction of artefacts due to the application of inverse distance weighted interpolation, which takes information from neighbouring cells to produce high-resolution reconstructions (Beyer *et al.* 2020b). Given the wide spatial distribution of the proxy dataset, we thus performed downscaling for the entire world. This is highly computationally expensive; however, the global downscaled bioclimatic variables have been made available on Zenodo (https://doi.org/10.5281/zenodo.7828454) for future use. Figure 1 show the different climatic models tested in this research and the geographic coverage of the proxy records.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

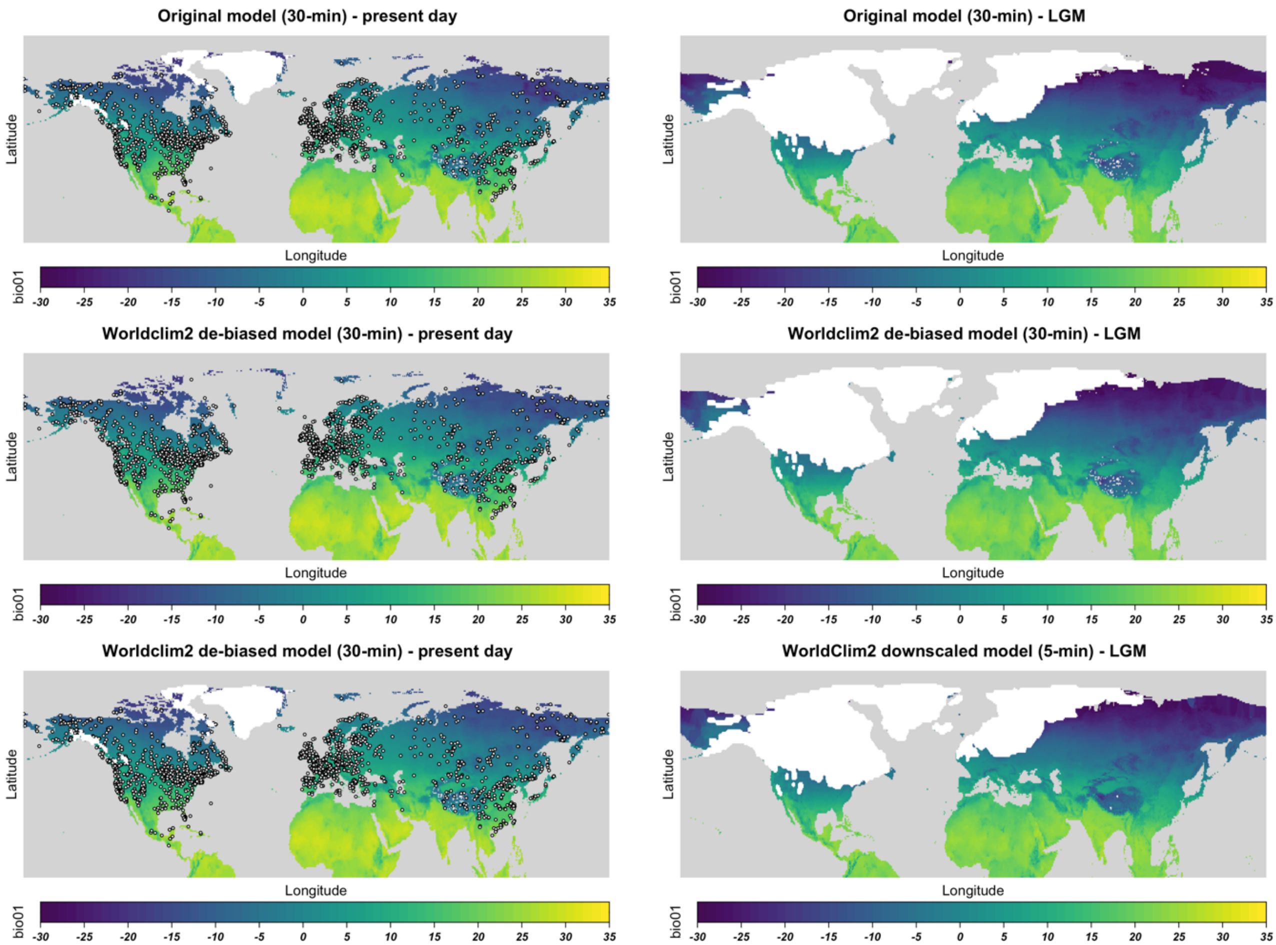


Figure 1. Site locations of proxy records studied in this analysis (left), against mean annual temperature (bio01) from Beyer *et al.* (2020a) for the present day and the Last Glacial Maximum (LGM), extracted from pastclim (Leonardi *et al.* 2023) at its original resolution (top), the model de-biased using a modern reference from WorldClim2 (Fick and Hijmans, 2017) to its original resolution (middle), the model further downscaled using WorldClim2 to higher resolution (bottom). Land mass in each time slice is masked by global ice sheets (plotted in white) and predicted sea level.

2.2 Proxy reconstructions

We employed the LegacyClimate 1.0 proxy dataset by Herzschuh *et al.* (2023) for direct validation. Mean annual temperature (T_{ann}), Mean July temperature (T_{july}) and total annual precipitation (P_{ann}) were reconstructed from fossil pollen data using the Weighted-Averaging Partial Least Squares (WA-PLS) and Modern Analogue Technique (MAT) methods, both of which are widely used and generate similar time series, though each method's performance vary in response to various factors, such as the quality and diversity of the calibration data, the time interval to be reconstructed, and the resolution of the pollen data (Sweeney *et al.,* 2018; Chevalier *et al.,* 2020). In LegacyClimate 1.0, the diverse pollen records are handled consistently

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past
Open Access
Discussions
EGU

through merging taxa into high-level harmonised taxonomic groups, increasing the possibility of matching modern climate analogues and fossil datasets. Its geographic coverage across the Northern Hemisphere is also much larger than other databases (e.g. Mauri *et al.,* 2015; Marsicek *et al.,* 2018; Routson *et al.,* 2019); the use of a single database based on a single climate proxy reduces inter-site variability resulting from the type of proxy utilised and allows the generation of analogous climatic parameters with direct relevance to bioclimatic variables available in the Beyer *et al.* (2020a) model. To facilitate comparison between the proxy reconstructions and the model simulations, we interpolated each proxy record via bilinear interpolation to the chronological resolution of the climatic model (1,000 years) to enable quantification; interpolating to regular time intervals ensures that periods of particularly dense sampling in the original cores do not exert undue influence on the results.

Following data-cleaning, we retained 2,420 records from LegacyClimate1.0. One record was removed as it did not have any proxy data associated with the MAT method (ID Dataset: 100127) and a further 170 records were removed as they fall under the cropped sea-level of the Beyer *et al.* (2020a) model; for example, the proxy sites are located on small islands not captured by the model or within lake margins. Table 1 summarises the records and models studied in this research.

Table 1. Summary of the proxy records selected from the LegacyClimate 1.0 (Herzschuh *et al.,* 2023) and the model simulations (Beyer *et al.,* 2020a) selected for analysis.

	Regions	N / Cell size	Type of data	Climatic variables extracted	Time min (1,000 years ago)	Time max (1,000 years ago)	Mean freq of records (per 1,000 years)	Reference (and doi)
Legacy Climate 1.0	Asia East North America West North America Europe	2420	Pollen reconstructions	T_{ann} T_{July} P_{ann}	0	30	0.67	Herzschuh, U. *et al.* (2023). *Scientific Data* (10.5194/essd-15-2235-2023)
Beyer *et al.* (2020a) simulations	Global	30- and 5-min	Simulations, de-biased using CRU (original)	Bio1 Bio10 Bio12	0	120	1/2	Beyer, R. *et al.* (2020a). Scientific Data (10.1038/s415

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

grid cells	and WordClim2 (this paper)	97-020-0552-1)

2.3 Analysis

To quantify the differences between time series, we calculated the bias, root mean square error (RMSE) and normalised RMSE (NRMSE). The RMSE measures the coherence between the model simulations and the proxy reconstructions, whilst the bias (calculated as the mean residual) highlights whether the raw or downscaled model overestimates (positive values) or underestimates (negative values) the proxy records. Standardising the RMSE using the mean allows comparing the coherence between variables. The bias can also be considered per proxy record to show which areas are over or underestimated for any given variable, facilitating comparability. Considering that downscaling to higher resolutions is thought to capture localised climatic dynamics, we tested the statistical significance of mean differences in coherence between lower resolution (30-min) and higher resolution (5-min) models, using a standard significance threshold of $p < 0.05$ via the Kruskal-Wallis non-parametric test. We also calculated the proportion of proxy records (reconstructed using the MAT and WA-PLS methods) that show higher RMSE with 30-min models compared to 5-min models. Instances where the proportion is higher than 0.5 highlight a positive effect of downscaling on model-data coherence.

These analyses allow us to evaluate both the output of the climate models and the reliability of the proxy data in predicting specific climatic parameters in the past, depending on geographic region, Marine Isotope Stage (chronology), method of climate reconstruction employed in the proxy datasets (MAT versus WA-PLS), elevation of site location (with sites above 1500 meter above sea level analysed as a subset) and topographic roughness (defined as the energetic cost of movement (see SOM. 1), with areas that require over 200 joules per meter to transverse deemed to have 'high roughness'). All these factors could potentially impact the articulation between the two types of time series.

3 Results

Our results demonstrate that overall proxy reconstructions and model simulations tend to highlight very similar climatic trends across variables, with average bias for both annual and July temperature time series remaining under 1 degree Celsius and annual precipitation under 25 mm across all records (Fig. 2, Supplementary Tables S1-3). Considering the NRMSE, the most divergent variable on average is reconstructed mean annual temperature, with annual precipitation and July temperature showing consistently lower values (Supplementary Tables S1-3). The latter two also show highly comparable results between different versions of the Beyer et al. (2020a) model, even at varying spatial resolution and when using different modern

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

reference datasets for downscaling. Overall, the difference in coherence between the two resolutions is minimal, particularly when controlling for the modern dataset used for de-biasing (Supplementary Table S4).

Our results highlight that the original model of annual temperature from Beyer *et al (*2020a) tends to estimate lower temperatures than the proxy reconstructions (as highlighted in the negative bias results reported in Supplementary Table S1), whereas the models de-biased using WorldClim2 (at both 30- and 5-min resolution) tend to predict higher annual temperatures compared to proxy records. Annual temperature time series from the original simulations (at 30-min resolution) tend to have more error in only around half the records, at 51% (MAT method, p = 0.4904) and 49% (WA-PLS method, p = 0.4904) of proxy sites when compared to the further downscaled simulations (at 5 min resolution), with the 30-min model de-biased by WorldClim2 having more error in slightly less than half of records compared to than the higher resolution model, at only 49% (MAT method, p = 0.4962) and 46% (WA-PLS method, p = 0.4962) (Supplementary Table S4).

Whether models tend to predict higher or lower precipitation compared to proxies seems to vary between resolutions, regions and topographic complexity (Supplementary Table S2). However, again, the overall difference in performance between the two resolutions is marginal (Supplementary Table S4), with the annual precipitation time series from the original raw simulations having more error in 55% of records (both MAT method and WA-PLS methods, p = 0.4923 and p = 0.4936 respectively) than the higher resolution model. Yet the 30-min model de-biased by WorldClim2 shows higher RMSE in just 48% with proxy-based time series (both MAT and WA-PLS methods, p = 0.4962) compared to the further downscaled model (Supplementary Table S4). These results suggest that divergence between proxy reconstructions and models is unaffected by the spatial resolution of the model.

Models of bio10 (mean temperature of warmest quarter) almost always slightly underestimate temperatures compared to proxies of Mean July Temperature, regardless of resolution (Supplementary Table S3). This could be linked to the slight discrepancy in the climatic parameter predicted. Average difference in model-data coherence between the two spatial resolutions is not statistically significant, with the July temperature time series from the original simulations showing less coherence in 58% (MAT method, p = 0.4904) and 56% (WA-PLS method, p = 0.4904) of proxy reconstructions when compared to that from the further downscaled model, although again the 30-min model de-biased by WorldClim2 shows higher error with just 47% of the proxy reconstructions (MAT method, p = 0.4962) and 48% (WA-PLS method, p = 0.4962) (Supplementary Table S4).

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

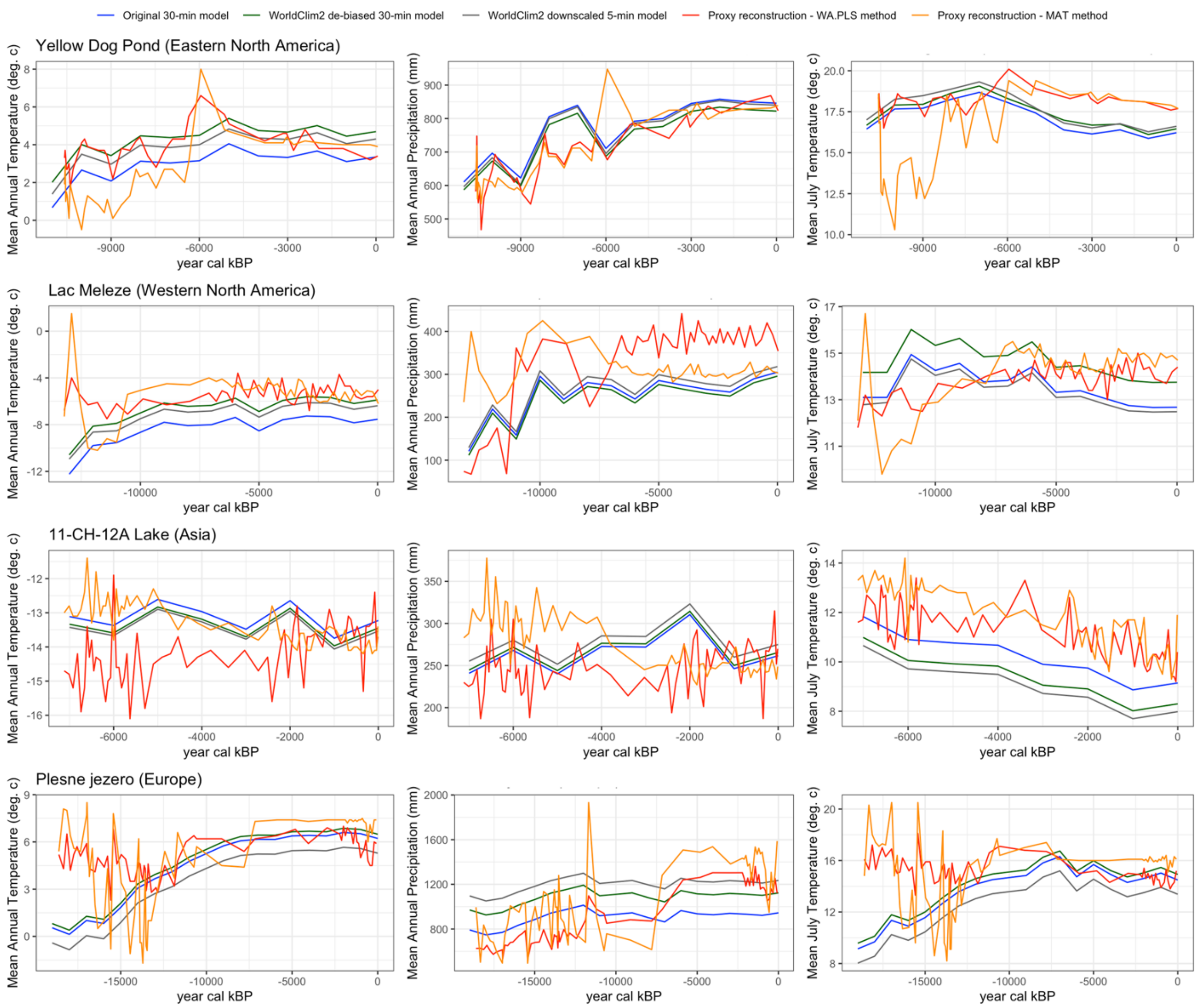


Figure 2. A sample from each regional group of reconstructed annual temperature (left), July temperature (middle) and annual precipitation (right) time series, comparing the original 30-min model, 30-min model bias-corrected using WorldClim2 and further downscaled 5-min model from Beyer *et al.* (2020a) with corresponding proxy reconstructions from LegacyClimate 1.0 (Herzshuch *et al.,* 2021), modelled using the MAT and WA-PLS approach.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

3.1 Regional differences

As highlighted in Fig. 3, our results demonstrate some key differences between regions. In Asia and East North America, both annual and July temperature proxy reconstructions are least divergent with the 5-min model and 30-min model de-biased with WorldClim2. These latter two models predict higher temperatures compared to the original 30 min CRU-debiased model, which tends to underestimate annual temperature compared to the proxy reconstructions (Supplementary Table S1, S3). A similar pattern is seen in Eastern North America for precipitation, with both models de-biased using WorldClim2 being the most coherent with proxy reconstructions. However, in Asia, we find that the original Beyer *et al.* (2020a) model is most coherent with proxy reconstructions, with the 5-min downscaled model having the highest RSME (Supplementary Table S2)). In West North America, the raw model de-biased with WorldClim2 and the downscaled model for annual temperature are also more coherent with the proxy records than the original model, with little difference between the two resolutions (Supplementary Table S1). For precipitation, however, proxy reconstructions from West North America show the lowest RMSE with the original model, like those in Asia (Supplementary Table S2). Yet for July temperatures in this region, there is some variability between proxy reconstruction methods, with the MAT method showing the lowest divergence with the downscaled model and the WA-PLS method that with the original model (Supplementary Table S3). Model-data coherence in Europe varies for the different climatic variables, with the 30-min model de-biased using WorldClim2 the most coherent for annual and July temperatures, though the former is closely followed by the original model whilst, for the latter the original model is the least coherent (Supplementary Table S1, S3). For precipitation proxy records in Europe, the 5-min model and 30-min model de-biased using WorldClim2 are similarly coherent, with the original model having the highest RSME (Supplementary Table S2)).

Fig. 3 and Supplementary Fig. S1 highlight the spatial heterogeneities in bias across the dataset. In LegacyClimate 1.0, the East North American subset of proxy reconstructions appear to be the most coherent with the climate models, showing the lowest RSME values across all variables (Supplementary Table S1, S3). Europe tends to show the lowest proportion of records where error is higher in the coarser models (30 min) compared to the higher-resolution model (5 min), with downscaling having the strongest impact on model-proxy divergence in East North America, particularly when compared to the original model (Supplementary Table S4). Regions showing the least coherence varies depending on the climatic parameter, with Asia having the highest RMSE values for annual and July temperatures (Supplementary Table S1, S3) and Europe that for annual precipitation (Supplementary Table S2). Nonetheless, in many scenarios, a higher proportion of proxy reconstructions in these two regions show better coherence with the models when downscaling is performed, though this depends on variable, proxy method of reconstruction and the coarse model compared (Supplementary Table S4). Indeed, often the 30-min model de-biased using WorldClim2 has a higher proportion of proxy records with lower error than the 5-min model (Supplementary Table S4), suggesting higher resolution models could simply add noise in many scenarios.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past
Discussions
Open Access

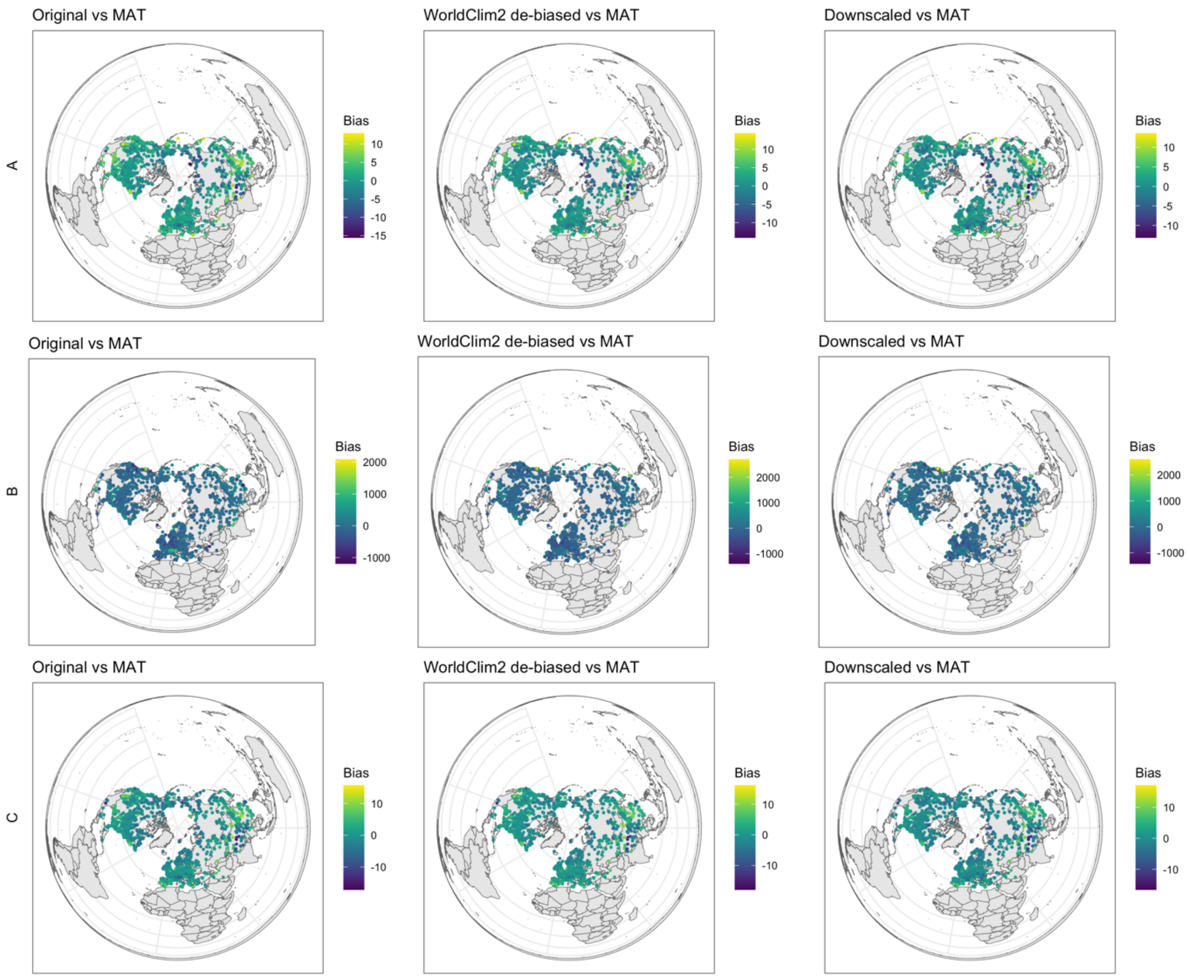


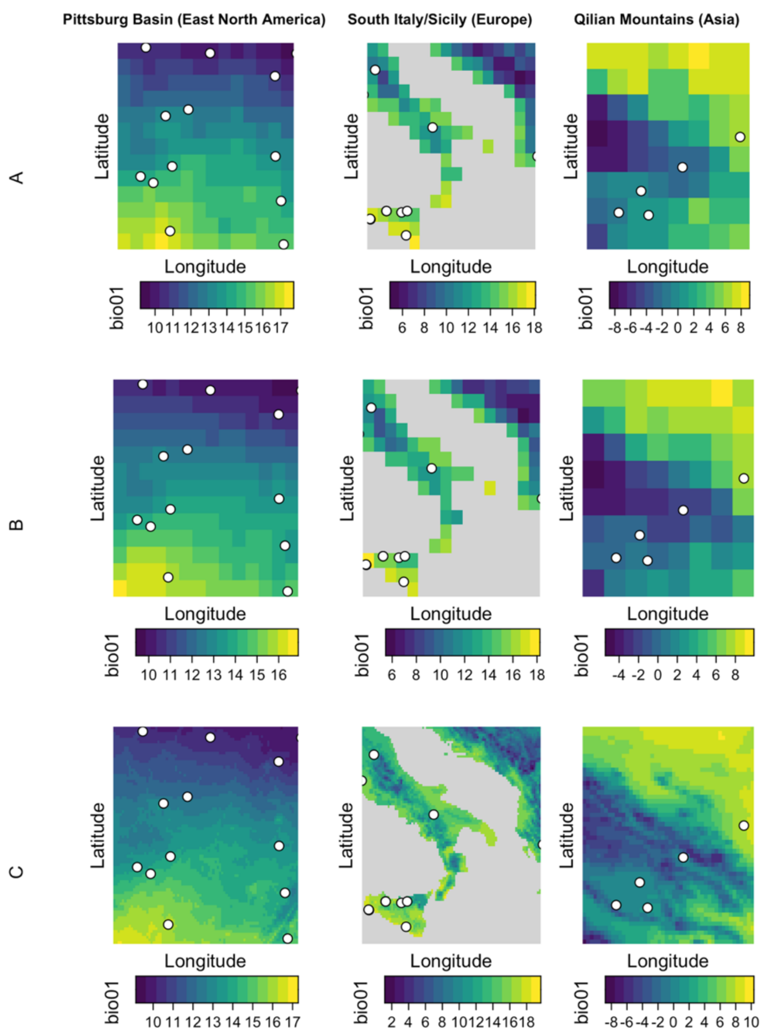
Figure 3. Absolute bias for mean annual temperature (A), mean annual precipitation (B), and mean July temperature (C) for each site, comparing the MAT method of proxy reconstruction against different versions of the Beyer *et al.* (2020)a model. Pair-wise comparisons of bias for the WA-PLS method are reported in Supplementary Figure S1.

3.2 Effects of landscape heterogeneity

Further model downscaling is often performed to account for smaller-scale landscape features that can locally impact climatic conditions, such as topography (Fig. 4). We therefore tested whether proxy records at varying elevations and topographic complexity show stronger coherence with higher resolution models compared to those at relatively lower resolution. Our results highlight that in almost all scenarios the relatively coarser models de-biased using WorldClim2 show the lowest overall divergence with the proxy reconstructed variables (Supplementary Table S1, S3), outperforming both the original Beyer *et al.* (2020a) model and the further downscaled model; only the low altitude group of records for mean annual temperature

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

reconstructed using the MAT method show the highest coherence with the further downscaled model, but only by 0.01 which is not statistically significant (p = 0.4896). Our results also show that proxy reconstructions tend to predict warmer temperatures at higher elevations and/or in areas of higher topographic roughness compared to model simulations and colder temperatures at lower elevations and/or lower topographic roughness (Supplementary Table S2), which is a known bias of transfer functions, which rely on averages of data from modern calibration data sets and tend to overlook climate extremes (Chevalier *et al.,* 2020). For precipitation, only in low altitude areas does the original model tend to produce lower values than the proxy reconstructions; in all other scenarios, the models overestimate rainfall relative to pollen records (Supplementary Table S2). Models of July temperatures always produce lower values than that of proxies (Supplementary Table S3). We find that the proportion of proxy records that show higher error (RMSE) with lower resolution models than higher resolution is around half for all subsets according to landscape variations, indicating no statistically significant effect of further downscaling on data-model coherence, even in areas of landscape heterogeneity.



https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Figure 4. Three regional examples of mean annual temperature for the present day (bio01), demonstrating how downscaling increases spatial resolution by capturing the effects of landscape dynamics on climate depending on the underlying topography. A) original Beyer *et al.* (2020a) model at 30-min resolution, B) WorldClim2 de-biased model at 30-min resolution, C) WorldClim2 downscaled model at 5-min resolution.

3.3 Glacial versus inter-glacial variability

We then examined discrepancies in model-data coherence through time, separating time series covering the present day, Marine Isotope Stage 1 (MIS 1; 0 – 14,000 years ago) and MIS 2 (14-29,000 years ago). In total, 1080 records were associated with timeslice 0 only (45% of dataset), 2398 records captured time slices in MIS 1 (99% of dataset) whereas only 475 spanned into MIS 2 (20%). Individual analysis of interpolated data points capturing the present day was performed, as these pollen proxies should be highly representative of modern ecological communities whilst model data points are based on present-day observations as opposed to simulations into the past, thus providing somewhat of a baseline of model-data divergence.

Our results demonstrate that data points representative of the present have the lowest RMSE, though considerable error in some time series exists (Fig. 4). In contrast, the smaller subset of time series covering MIS 2 show the highest bias and RMSE between models and proxies, both across model resolutions and methods of proxy reconstruction, with models capturing older time periods underestimating annual and July temperatures and overestimating annual precipitation compared to proxy-based reconstructions (Supplementary Tables S1-S3). We find that the proportion of proxy records that show higher RMSE (and therefore are less coherent) with lower resolution models compared to those of higher resolution is less than half for most chronological subsets, with only annual temperature and July temperature during MIS 2 to seeming to see a slight benefit of downscaling (Supplementary Table S4), though this is not statistically significant for any comparison ($p = 0.425$ to 0.4962).

Figure 5 highlights the overlap between RMSE values from the present day, MIS 1 and MIS 2, confirming that data-model discrepancies tend to increase with age though not significantly so ($p > 0.05$). Chronological uncertainties in the proxy age model may complicate the comparison between climate simulations and pollen-based records, as well as the process of signal smoothing via interpolation to facilitate analysis. Models are also inherently calibrated to replicate current rather than past climate patterns, and proxy reconstructions rely on the identification of modern analogue species that may have a different link to climate than palaeoecological communities, likely further contributing to higher divergence in older time periods (Chevalier *et al.* 2020). Nonetheless, all of the distributions highlighted in Fig. 5 are highly positively skewed – there are many extreme values at the right-hand side – confirming that age is just one contributing factor in the divergence between time series.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past
Discussions
Open Access

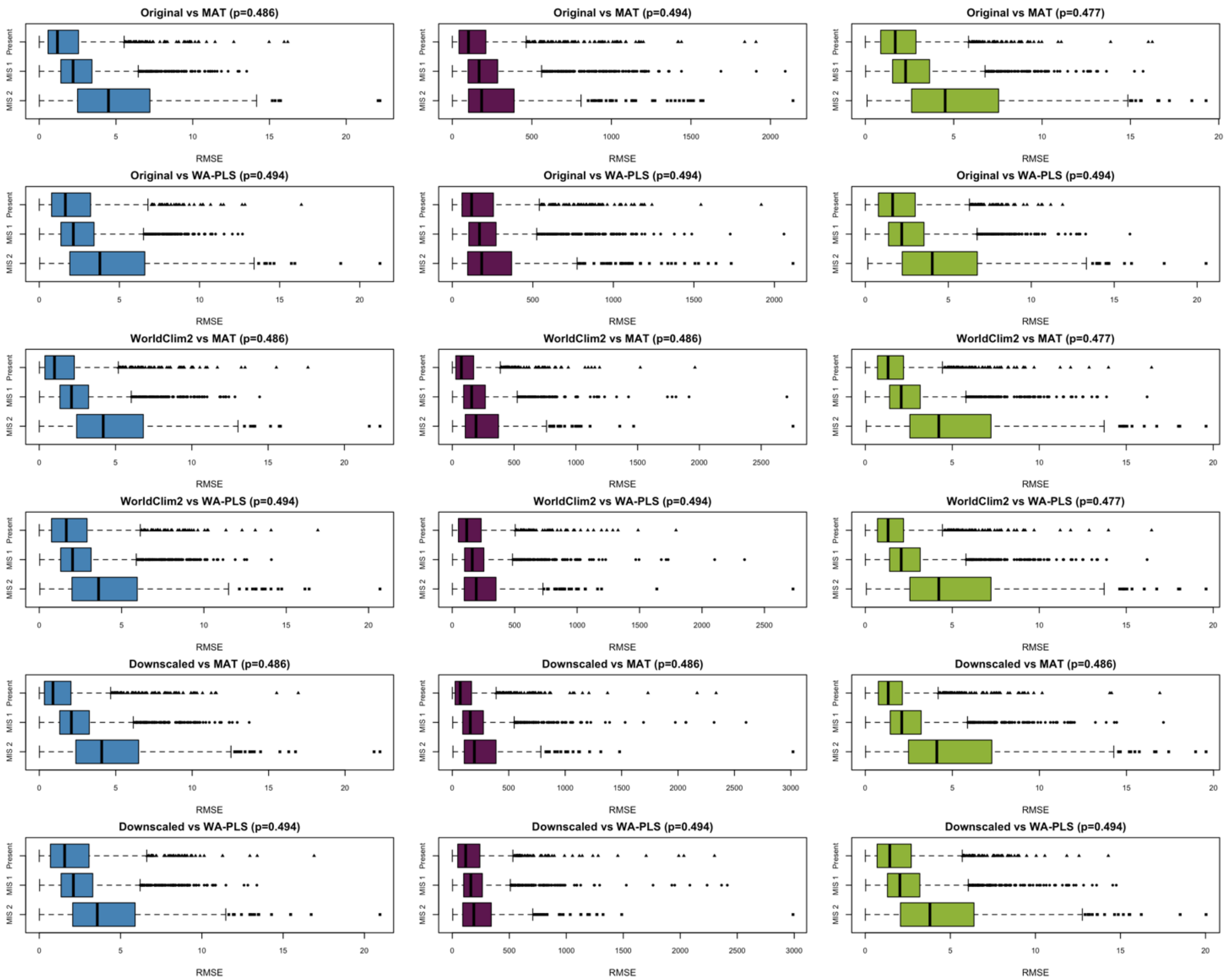


Figure 5. Boxplots of pair-wise root mean square error (RMSE) results model-data comparisons of mean annual temperature (blue), mean annual precipitation (purple) and mean July temperature (green) from those representing the present, MIS 1 and MIS 2, with p-values highlighting the significance of differences between period from Kruskal Wallis tests.

3.4 Exploring the most divergent time series

Observing the distribution of the data in Fig. 5, we decided to segment the highest 5% of RMSE values for each pair-wise model-data comparison for further investigation. We then amalgamated those that routinely fall into this category for each climatic variable, representing the most divergent time series of the overall dataset for the three parameters studied here (Table 2). None of the records fall into the most divergent subset for all variables studied, suggesting more extreme divergence is not

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past Discussions

related to any systematic issue in the model nor the proxy at individual sites. We then produced 1000 bootstrapped samples (without replacement) of corresponding sample size, ascertaining whether the observed proportion of time series in this highly divergent subset is greater than expected by random chance (Table 2).

Our results highlight that sites in Asia and dating to MIS 2 consistently exhibit significantly higher proportions of divergent time series across climatic variables (Supplementary Table S5). Overall, 53 records of mean annual temperature fall into the most divergent 5% of time series based on RMSE, of which statistically significantly higher proportions of these than expected are in areas of high altitude and/or low roughness (Supplementary Table S5). For mean annual precipitation, only 25 records consistently fall in the top 5% based on RMSE, demonstrating higher inconsistency in pairwise model-data coherence compared to the temperature variables (Supplementary Table S5). We found that, for this parameter, significantly higher proportions of these are located in West North America, in areas of low altitude and/or high roughness, and date to MIS 1 (Supplementary Table S5). Finally, for mean July temperature, 41 time series always fall into the most divergent 5% of reconstructions, significantly higher proportions of which are located in areas of high altitude than would be expected by random chance (Supplementary Table S5.

4 Discussion

Increasing the spatial resolution of models is often thought to be required to better capture nuance in the climate of specific places. But what is the optimal spatial resolution for adequately capturing these finer-scale signals? Our results highlight that further downscaling models to higher resolutions (5-minute) fails to consistently capture more signal from proxy records, which are the gold standard for capturing localised ecological dynamics. This implies that more downscaling is not always better, with relatively coarser simulations (i.e. 30-minute) seeming to provide a similarly adequate representation of past climatic trends in many scenarios, even in areas of topographic complexity. Nonetheless, we find that model-data coherence predictably decreases with age, with the more divergent time series than expected by chance located in Asia. Annual precipitation and July temperature show consistently low NRMSE, indicating good overall agreement between the simulations and empirical reconstructions, whereas annual temperature tends to show greater disparity in certain contexts, specifically in West North America and Asia, at high altitudes (where modern calibration data tend to be more limited), and at older time scales (likely due to a lack of good analogues of glacial/periglacial vegetations).

In our analysis, we employed different de-biased and downscaled versions of the Beyer *et al.* (2020a) climate emulator alongside harmonised pollen records from LegacyClimate1.0 (Herzschuh *et al.* 2023), providing corresponding estimates of three key climatic parameters for comparison. Whilst the LegacyClimate 1.0 dataset provides an excellent resource to address whether downscaling to higher resolutions is effective in capturing local climatic details, it is worth noting that, because the proxy records employed tend to capture pollen from a broad catchment, they may represent geographically wide averages of

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

past climate. This could inherently make them more compatible with coarser-level model simulations, which also capture broader landscape rather than local-level trends. Future work should seek to expand systematic model-data comparisons on other types of harmonised proxies, as well as different climatic models and modern references, ensuring that the equivalent bioclimatic variables are being predicted by both sources.

Whilst our results show that downscaling to much higher resolutions does not necessarily improve the agreement between model simulations and pollen-proxy reconstructions, we note that there is a trade-off between enhancing spatial resolution and increasing potential error. Such error in a given location could either be caused by using too coarse a resolution on the one hand or by unreliable interpolation on the other. For this reason, there are likely to be many circumstances in which it is still better to use downscaled models (with caveats), particularly when variability within 30-min cells (~55km on each side) is important. For example, the identification of conditions at specific locations within climatic extremes may be overlooked when using a model at a broader scale, such as at Late Pleistocene archaeological site Fincha Habera in the Bale Mountains of southern Ethiopia (Groos *et al.* 2021). Here, lower temperatures predicted by downscaled models may better characterise the on-site environment than that also incorporating environmental trends in surrounding lower altitude landscape (Timbrell *et al.* 2022). Careful consideration is therefore required to select the optimal spatial resolution when using models for the research question at hand.

5 Conclusion

Climatic simulations that characterise landscape-scale heterogeneities are needed to produce more accurate models within the multitude of fields that employ ecological data, such as those that wish to map species distributions through time and space and/or investigate the impact of climatic change on various biological and/or behavioural phenomena. We show that additional downscaling fails to consistently capture more signal from the proxy reconstructions, though models at both median (30-arc minutes) and fine-grained (5-arc minutes) spatial resolutions characterise climatic variables in broadly similar ways to pollen proxy reconstructions. Utilising models for analyses of past climate nonetheless involves a careful balancing act between accentuating variations relevant to the study questions and the introduction of error.

Code and data availability

The workflow to downscale climate models outputs with the delta method has been made publicly available as functions in *pastclim*. Code and data relating to this analysis, as well as a vignette for downscaling in *pastclim*, was made available during the peer review of this article and can be found here: https://osf.io/duq3j/. The global downscaled model at 5-arc minutes resolution is stored on Zenodo: https://doi.org/10.5281/zenodo.7828454.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Author contributions

Conceptualisation: LT, JB, MG, ES, AM; Data curation: LT, JB, MCh, AM; Formal analysis: LT, JB, MG, AM; Methodology: LT, JB, MG, MCh, AM; Software: LT, JB, MCo, ML, AM; Visualisation: LT; Writing – original draft preparation: LT, Writing – reviewing and editing: LT, JB, MCo, ML, MCh, MG, ES, AM.

Acknowledgements

MCh is supported by the German Federal Ministry of Education and Research (BMBF) with the Research for Sustainability initiative (FONA) through the PalMod Phase III project (grant no. FKZ: 01LP2308B). ML and AM were funded by the Leverhulme Research Grant RPG-2020-317

Competing interests

The authors declare that they have no conflict of interest.

References

Armstrong, E., Hopcroft, P. O. and Valdes, P. J.: A simulated Northern Hemisphere terrestrial climate dataset for the past 60000 years. Sci. Data., 6, 265, https://doi.org/10.1038/s41597-019-0277-1, 2019.

Barreto, E., Holden, P.B., Edwards, N.R., Rangel, T.F.: PALEO-PGEM-Series: A spatial time series of the global climate over the last 5 million years (Plio-Pleistocene). Glob. Eco. Bio, 32, 7, 1034-1045. https://doi.org/10.1111/geb.13683. 2023

Beyer, R. M., Krapp, M. and Manica, A.: High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. Sci, Data., 7, 236, https://doi.org/10.1038/s41597-020-0552-1, 2020a.

Beyer, R., Krapp, M., and Manica, A.: An empirical evaluation of bias correction methods for palaeoclimate simulations. Clim. Past., 16, 1493-1508 https://doi.org/10.5194/cp-16-1493-2020, 2020b.

Beyer, R. M., Krapp, M., Eriksson, A. and Manica, A.: Climatic windows for human migration out of Africa in the past 300 000 years. Nat. Commun., 12, 4889, https://doi.org/10.1038/s41467-021-24779-1, 2021.

Blinkhorn, J., Timbrell, L., Grove, M., Scerri, E. M. L.: Evaluating refugia in recent human evolution in Africa. Phil. Trans. B., 377, 20200485, https://doi.org/10.1098/rstb.2020.0485, 2022.

Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L. and Fordham,D. A.: StableClim, continuous projections of climate stability from 21 000 BP to 2100 CE at multiple spatial scales. Sci. Data., 7, 335, https://doi.org/10.1038/s41597-020-00663-3, 2020.

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

Chauvier, Y., Descombes, P., Guéguen, M., Boulangeat, L., Thuiller, W., Zimmermann, N.E.: Resolution in species distribution models shapes spatial patterns of plant multifaceated diversity. Ecograph. e05973. https://doi.org/10.1111/ecog.05973, 2022.

Chevalier, M., Davis, B.A.S, Heiri, O., Seppa, H., Chase, B.M., Gajewski, K., Lacourse, T., Telford, R.J., Finsinger, W., Guiot, J., Kuhl, N., Maezumi, S.Y., Tipton, J.R., Carter, V.A., Brussel, T., Phelps, L.N., Dawson, A., Zanon, M., Vallé, F., Nolan, C., Kupriyanov, D.: Pollen-based climate reconstruction techniques for late Quaternary studies. Earth-Sci. Rev., 210(103384), https://doi.org/10.1016/j.earscirev.2020.103384, 2020.

Fernández-Donado, L., Gonzalez-Rouco, J.F., Raible, C.C., Ammann, C.M., Barriopedro, D., Garcia-Bustamante, E., Junglcaus, J.H., Lorenz, S.J., Luterbacher, J., Phipps, S.J., Servonnat, J., Swingedouw, S., Tett, S.F.B., Wagner, S., Yiou, P., Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. Clim. Past, 9, 393–421, https://doi.org/10.5194/cp-9-393-2013, 2013.

Fick, S. E. and Hijmans, R. J.: Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. Inter. J. Climatolog. 2013, Available: http://www.worldclim.com/version2

Franklin, J., Potts, A.J., Fisher, E.C., Cowling, R.M., Marean, C.W.: Palaeodistribution modelling in archaeology and paleoanthropology. Quat. Sci. Rev. 110, 1-14. https://doi.org/10.1016/j.quascirev.2014.12.015. 2015.

Herzschuh U., Böhmer T., Li C., Chevalier M., Hebert R., 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., Qinghai, Zheng Z.: LegacyClimate 1.0: a dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the last 30 kyr and beyond. Earth Syst. Sci. Data., 15 (6), 2235–2258, https://doi.org/10.5194/essd-15-2235-2023, 2023

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. Inter. J. Climatolog., 25, 1965–1978, 2005.

Holden, P. B., Edwards, N. R., Rangel, T. F., Pereira, E. B., Tran, G. T. and Wilkinson, R. D.: PALEO-PGEM v1.0: a statistical emulator of Pliocene–Pleistocene climate. – Geosci. Model Dev., 12, 5137–5155, 2019.

Karger, D. N., Nobis, M. P., Normans, S., Graham, C. H. and Zimmermann, N. E.: CHELSA-TraCE21k v1.0. down-scaled transient temperature and precipitation data since the last glacial maximum. Clim. Past., 19, 2, 439-456, https://doi.org/10.5194/cp-2021-30, 2021

Krapp, M., Beyer, R. M., Edmundson, S. L., Valdes, P. J. and Manica, A.: A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. Sci Data 8, 228 https://doi.org/10.1038/s41597-021-01009-3, 2021.

Kottek, M., Grieser, J. and Beck, C. World Map of the Köppen-Geiger climate classification updated. Gebrüder Borntraeger, Berlin, Stuttgart. https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Laepple, T. and Huybers, P.: Global and regional variability in marine surface temperatures. Geophys. Res. Lett., 41, 2528–2534, 2014.

Laepple, T., Ziegler, E., Weitzel, N., Hebert, R., Ellerhoff, B., Schoch, P., Martrat, B., Bothe, O., Moreno-Chamarro, E., Chevalier, M., Herbert, A. and Rehfeld, K.: Regional but not global temperature variability underestimated by climate models at supradecadal timescales. Nat. Geosci., 16, 958–966, https://doi.org/10.1038/s41561-023-01299-9, 2023.

Leonardi, M., Boschin, F., Boscato, P. and Manica, A.: Following the niche: the differential impact of the last glacial maximum on four European ungulates. Commun. Biol., 5, 1038, https://doi.org/10.1038/s42003-022-03993-7, 2022.

Leonardi, M., Hallet, E. Y., Beyer, R., Krapp, M. and Manica, A. 2023. pastclim 1.2: an R packages to easily access and use paleoclimatic reconstructions. Ecography. e06481. https://doi.org/10.1111/ecog.06481

Maraun, D. and Widmann, M.: Statistical downscaling and bias correction for climate research, Cambridge University Press, Cambridge, UK., 2018.

Mitchell, T.D. and Jones, P.D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Inter. J. Climatolog., 25, 693-712, https://doi.org/10.1002/joc.1181, 2005.

NOAA National Centers for Environmental Information. 2022: ETOPO 2022 15 Arc-Second Global Relief Model. NOAA National Centers for Environmental Information. https://doi.org/10.25921/fd45-gt74, 2022, Avaliable here: https://www.ncei.noaa.gov/products/etopo-global-relief-model

Ozdemir, S.: Testing the Effect of Resolution on Species Distribution Models Using Two Invasive Species. Pol. J. Environ. Stud. 33, 2, 1325-1335. https://doi.org/10.15244/pjoes/166353. 2024

Padilla-Iglesias, C., Atmore, L. M., Olivero, J., Lupo, K., Manica, A., Isaza, E. A., Vinicius, L. and Migliano, A. B.: Population interconnectivity over the past 120,000 years explains distribution and diversity of Central African hunter-gatherers. PNAS, 119(21), e2113936119. https://doi.org/10.1073/pnas.2113936119 , 2022

Rehfeld, K., Münch, T., Ho, S. L. and Laepple, T.: Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene. Nat., 554, 356–359, https://doi.org/10.1038/nature25454, 2018

Singarayer, J. S. and Valdes, P. J.: High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. Quat. Sci. Rev., 29, https://doi.org/10.1016/j.quascirev.2009.10.011, 2010.

Singarayer, J. S. and Burrough, S. L.: Interhemispheric dynamics of the African rainbelt during the late Quaternary. Quat. Sci. Rev., 124 , https://doi.org/10.1016/j.quascirev.2015.06.021, 2015.

Spratt, R. M. and Lisiecki, L. E.: A Late Pleistocene sea level stack, Clim. Past, 12, 1079–1092, https://doi.org/10.5194/cp-12-1079-2016, 2016.

Sweeney, J., Salter-Townshend, M., Edwards, T. Buck, C. E. and Parnell, A. C.: Statistical Challenges in Estimating Past Climate Changes. WIRES Computational Statistics, 10 (5), e1437, https://doi.org/10.1002/wics.1437, 2018

Timbrell, L., Grove, M., Manica, A., Rucina, S. and Blinkhorn, J.: A spatiotemporally explicit paleoenvironmental framework for the Middle Stone Age of eastern Africa. Sci. Rep.12: 3689, https://doi.org/10.1038/s41598-022-07742-y, 2022.

https://doi.org/10.5194/cp-2024-53
Preprint. Discussion started: 31 July 2024

Climate of the Past
Discussions

Timmermann, A., Yun, K. S., Raia, P., Ruan, J., Mondanaro, A.,Zeller, E., Zollikofer, C., Ponce de León, M., Lemmon, D.,Willeit, M. and Ganopolski, A.: Climate effects on archaic human habitats and species successions. Nat. 604: 495–501, https://doi.org/10.1038/s41586-022-04600-9, 2022.

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J. J., Farnsworth, A., Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., Roberts, W. H. G., Stone, E. J., Tourte, G. J. L., and Williams, J. H. T.: The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0, Geosci. Model Dev., 10, 3715–3743, https://doi.org/10.5194/gmd-10-3715-2017, 2017.

Zhu, F., Emile-Geay, J., McKay, N.P., Hakim, G. J., Khider, D., Ault, T.R., Steig, E.J., Dee, S. and Kirchner, J.W.: Climate models can correctly simulate the continuum of global-average temperature variability. Proc. Natl Acad. Sci. USA., 116, 8728–8733, https://doi.org/10.1073/pnas.1809959116, 2019.