More is not always better: delta-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], Andrea Vittorio Pozzi[3], 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. While proxies are thought to 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 of multiple bioclimatic variables. A key decision when using model output 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 for certain applications like species distribution models, 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 increasing the resolution of model output from 30 to 5 arc-minutes using the delta-downscaling method, which interpolates and applies the long-term difference between past and present model datasets to a higher resolution grid of observed present-day climate. We seek to determine to what extent further downscaling captures climatic trends at the site-level through direct comparison with proxy reconstructions, evaluating the different versions of the output from the HadCM3 Global Circulation model for annual temperature, mean temperature of July, and annual precipitation against a large empirical dataset of pollen-based reconstructions from across the Northern Hemisphere. Our results demonstrate that models tend to provide broadly similar accounts of past climate to that obtained from proxy reconstructions, with coherence tending to decline with age and at higher altitudes. However, our results imply that using the delta method to downscale to a very fine resolution has minimal net effect on the coherence of model output with pollen records in most cases. 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 variations and increasing potential error via unreliable interpolation.

Formatted: Font: 12 pt, Not Bold

Deleted:

Formatted: Font: 12 pt, Superscript

Deleted: st

Deleted: ;

Deleted: o

Deleted: s

Deleted: further downscaling

Deleted: s

Deleted: uses

Deleted:

Deleted: to increase spatial resolution of simulations

Deleted: ,

Deleted: in order to determine to

Deleted: ,

Deleted: . We use

Deleted:

Deleted: , which we evaluated

Deleted: , overall,

Deleted: ing

Deleted: scale

Deleted: to no

Deleted: data

1 Introduction

Realistic reconstructions of global paleoclimates are vital for modelling long-term evolutionary and ecological processes in fields like evolutionary biology, palaeoecology, palaeontology, and archaeology. Proxy records, such as those derived from pollen or other biomarkers, tend to be the preferred method for characterising past environments at specific locations; however, in order to extrapolate beyond the individual core sites and across wider regions, often it is necessary to rely on modelled or simulated climatic conditions. 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., Fordham *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). Openly accessible simulated datasets, such as those published by Beyer *et al.* (2020a), Krapp *et al.* (2021), Yun et al. (2023) and Barreto *et al.* (2023), and associated toolkits (e.g., the analytical package *pastclim* for manipulating and extracting modelled data; Leonardi *et al.,* 2023), are particularly useful for scientists interested in Pleistocene and Holocene timescales, facilitating continuous-time analyses at a high spatial resolution across a wide range of applications, such as habitat and species distribution modelling (SDM) and the quantitative analysis of climate change in relation to spatiotemporally diverse biological and behavioural phenomena (e.g. Beyer *et al.,* 2021; Padilla-Iglesias *et al.,* 2022; Blinkhorn *et al.,* 2022; Timmerman et al. 2022; Leonardi *et al.,* 2022; Zeller and Timmerman 2024; Mondanaro et al. 2025).

Proxy data, while allowing for detailed reconstructions of climatic conditions through time, are rarely 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. In this sense, they often provide relative estimates of past climate, 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 diverse nature of the data employed (i.e., pollen, lake sediments, ice cores etc.), which record climate in an inconsistent way that typically cannot be articulated as the bioclimatic indicators and environmental parameters that are routinely in species distribution models (SDMs) (e.g. Beyer *et al.* 2021; Blinkhorn *et al.* 2022; Leonardi *et al.* 2022). Model output have the potential to overcome these shortfalls, providing tangible values for parameters such as temperature, precipitation, and a range of derived bioclimatic indices (e.g., Hijmans *et al.,* 2005), that are consistent across variables for a more complete account of climatic conditions. Models additionally offer 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 proxy 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 and topographically diverse regions (Maraun and Widmann, 2018). Resultant differences can be in the order of several degrees for temperature and tens of percent for precipitation, which could lead to substantially different biome

Deleted: 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).

Deleted: remain the gold standard

Deleted: data

Deleted: analytical packages

Deleted: tool

Formatted: Font: Italic

Deleted: Middle-Late

Deleted: (e.g. Beyer *et al.,* 2021; Padilla-Iglesias *et al.,* 2022; Blinkhorn *et al.,* 2022; Leonardi *et al.,* 2022)

Deleted: .

Deleted: ¶

Deleted: The chronological resolution of proxy records has ... [1]

Deleted: p

Deleted:

Deleted:)

Deleted: Proxy records

Deleted: also

Deleted: rather than absolute parameters

Deleted: variable

Deleted: typically cannot be

Deleted: as

Deleted: applied when studying contemporary populations, such as

Deleted: ecological

Deleted: niche

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: ¶ ... [2]

Deleted: led data

Deleted: absolute

Formatted: Font: Italic

Deleted: , as well as

Deleted: ing

Deleted: observational

Deleted: regions and those in

Deleted: complex

Deleted: errors

classifications and estimations of ecologies experienced (Kottek *et al.*, 2006). Such variations can have important implications for the diverse fields employing model output for the reconstruction of past and present species distributions, dispersal and extinction processes, and biogeographic patterns.

High resolution simulations of multiple time slices are often desired by consumers of model output yet difficult to obtain due to computational costs. For example, dynamical downscaling allows for the detailed description of processes in the climatic system and can improve the capturing of localised climatic conditions (Rummukainen, 2016; Strandberg et al., 2023), however this method is rarely applied in fields like palaeoecology and archaeology due to the computational costs, particularly when a large number of time steps are required. 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 using statistical methods (Fordham et al. 2017; Beyer *et al.* 2020a; Krapp *et al.* 2021; Zeller and Timmerman 2024; Mondanaro et al. 2025), as these approaches can be more easily applied to several time periods. Within statistical downscaling, 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.

As a community, we are becoming increasingly aware of issues related to the scale and resolution of climate variables, yet it is currently unclear what level of downscaling is desirable for applications like SDM. Indeed, the ODMAP (Overview, Data, Model, Assessment, Prediction) protocol stresses the importance of spatial resolution and extent of environmental predictors, as well as processing and scaling (Fitzpatrick et al. 2021), yet there is still no universally agreed upon pipeline for SDM to help determine when downscaling may be important. Recently a resolution of 1km was obtained for the TRACE21K simulations using the CHELSA algorithm (Karger *et al.* 2023), interpolating very high-resolution climate for every 100 years for the last 21,000 years. Some studies support that much finer-scale simulations have higher predictive power in SDMs 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). Proxies offer a more localised account of climate in certain places, yet they too can be associated with high degrees of uncertainty, arising from multiple sources. Nonetheless, determining model

Deleted: on the ground

Deleted: . 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.

Deleted:

Deleted: .

Deleted: using statistical downscaling,

Deleted: is

Deleted: method

Deleted: D

Deleted: However,

Deleted: is a desirable level of downscaling

Formatted: Font: (Default) +Body (Times New Roman), 10 pt

Deleted:

Deleted: predicting

Deleted: species distribution models

Deleted: 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…;

;

Deleted: ;

Formatted: Font colour: Text 1

Deleted: the gold standard for capturing climate variability in specific locations through time.

Deleted: ¶

¶

agreement with empirical reconstructions from proxies remains a widely applied method for ground-truthing downscaled climatic output.

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 and methods of temperature reconstruction from proxies facing systematic deficiencies, though stronger emphasis is placed on the former. Strandberg *et al.* (2022) conversely suggest that comparisons between models and proxies are mostly limited by the large errors associated with proxy data.

Given the ever-increasing demand to produce more accurate models of past climate across extended timeframes, we tested whether downscaling climatic models from a relatively coarser (30-min) to a higher resolution (5-min) leads to increased agreement with empirical reconstructions of past climate from proxies. We applied a new suite of functions in the *pastclim* R package (Leonardi *et al.,* 2023) for delta-downscaling model output, and present new high-resolution climatic simulations for 17 bioclimatic variables for the HadCM3 Global Circulation Model (GCM) output (Huntley et al. 2022) and the Beyer *et al.* (2020a) model time series. We have provided 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 variables from these model outputs at varying spatial resolution. Our work has quantified the average divergence between the time series produced using modelled climate at varied spatial resolution and method of proxy reconstruction, with our results ultimately endorsing the use of model output in the absence of high-resolution proxies, though with careful consideration as to the most appropriate resolution for analysis.

2 Materials and methods

2.1 Climate models

To test the impact of delta-downscaling at different resolutions, we used two time series of model simulations. The first one is a set of raw temperature and precipitation outputs from the HadCM3 Global Circulation Model, at their native resolution of 3.275 x2.5 arc-degrees taken from Huntley *et al.* (2022). We consider a set of simulations in which the HadCM3 was run with appropriate boundary conditions for the last 120k years at 2,00 years intervals (the original set in that paper covered the last

Formatted: Font: Italic
Formatted: Font colour: Red
Deleted: Here
Deleted: further
Deleted: of
Deleted: models
Deleted: relatively
Deleted: than relatively coarser models (30-min).
Deleted: present new downscaled climatic models for 17 bioclimatic variables and
Deleted: y
Deleted: easily
Deleted: climatic simulations from Beyer *et al.* (2020a).
Deleted: our
Deleted: datasets
Deleted: coarse- and fine-grained
Deleted: s
Deleted: different
Deleted: data
Deleted: s
Deleted: led data
Deleted:
Deleted: s
Deleted: model
Deleted: (REF)
Formatted: Font: Italic

800k years). The second series comes from Beyer et al. (2020a) within the *pastclim* R package (Leonardi et al. 2023). This is based on an older series of runs of the HadCM3 Global Circulation Model (Singarayer and Valdes 2010, Singarayer and Burrough, 2015; Valdes et al. 2017) for the last 120k years, in 72 snapshots (2,000-year time steps between 120,000 BP and 22,000 BP; 1,000-year time steps between 22,000 BP and the pre-industrial modern era). As in the other set, the original model output of HadCM3 had a grid resolution of 3.75 x 2.5 arc-degrees.

These outputs were first downscaled using a series of runs of the higher resolution HadAM3H model, available at 1.25 x 0.83 arc-degrees for the last 21,000 years in 9 snapshots (2,000-year time steps between 12,000 BP and 6,000 BP; 3,000-year time steps otherwise) using an approached termed dynamic delta downscaling by Beyer et al (2020a). This method consists of generating a set of delta matrices based on the few time steps for which outputs were available from both HadCM3 and HadAM3H, and then using these matrices to downscale each time step in the full set by using a weighted interpolation of the two closest delta matrices based on CO_2 (see Beyer et al, 2020a, for details). This approach takes advantage of the higher resolution of local dynamics captured by HadAM3H, which is computationally too expensive to be run for all time steps. These outputs were then debiased and downscaled in Beyer et al. (2020a) to 0.5 x 0.5 arc-degrees with the delta method using the Climate Research Unit Global Climate Dataset (CRU) as the modern climatic reference (Mitchell and Jones, 2005).

For this study, we delta downscaled and debiased these two model outputs to a resolution of both 30 arc-minutes and 5 arc-minutes using modern observation from WorldClim2 (Fick and Hijmans, 2017). For the Beyer et al (2020a) model, as it was already at 30 arc-minutes, the delta downscaling at this resolution gives us a debiased version based on WorldClim2 rather than CRU. We used a global relief map from ETOPO2022 (NOAA National Center 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. All data manipulations were done using the R package *pastclim* (Leonardi et al. 2023).

Downscaling was 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 was computed, adding the difference between past and present-day simulated climate to present-day observed climate, following Beyer et al. (2020a) and Krapp et al. (2021). The delta method therefore assumes that local (i.e. grid-cell-specific) model biases are constant over time (Maraun and Widmann, 2018). The resulting matrix only covers the land extent at the present. We then expanded this matrix to reach the largest land-extent in any of the times-steps under consideration using an inverse-distance-weighted interpolation. For most of the world, at the resolution of 30 and 5 arc-minutes, this only requires interpolating a small number of cells away

Deleted: ese reconstructions …sare ... [3]
Formatted: Font: Italic
Deleted: REFS
Deleted:
Deleted:
Formatted ... [4]
Deleted:
Deleted: …ear time steps between 120,000 BP and 22,000 BP; 1,000- ... [6]
Formatted ... [5]
Deleted: _…egrees for the the …ast 21,000 years in 9 snapshots (2,000- …ear time steps between 12,000 BP and 6,000 BP; 3,000-year time steps otherwise) using an approached termeds…dynamic delta downscaling by Beyer et al (2020aREF…. This method consists of in ... [7]
Deleted: .
Formatted: Subscript
Deleted: ¶
Deleted: We applyied
Deleted: the
Deleted: global temperature and precipitation bioclimatic simulations from the HadCM3 GCM (REF) and those produced byby Beyer *et al.* (2020a) available within the *pastclim* R package (Leonardi *et al.*, 2023), published at
Deleted: 0.5°
Deleted: 30-min resolution in 1 or 2- to 2-thousand-year time slices spanning the past 120 thousand years.
Deleted:
Deleted: The originalBeyer et al. (2020a) produced their climate time series simulations viaw
Deleted: ere
Deleted: delta-downscalinged and de-bias
Deleted: e
Deleted: ingd
Deleted: based
Deleted: of snapshot outputs fromn the HadCM3 general circulations model (Singarayer and Valdes, 2010; Singarayer and Burrough, 2015; Valdes *et al.*, 2017) using the Climate Resear ... [8]
Deleted: d
Deleted: *pastclim* to further downscale monthly temperatur ... [9]
Deleted: W… delta downscaled and debiased these two ... [10]
Deleted: REF
Formatted: Font: Italic
Deleted: is…performed one monthly variable at a time (i.e ... [11]

from the coastline; for higher resolutions, other interpolating algorithms might be more appropriate. We note that the delta-downscaling can also be obtained by creating first the difference between model outputs, which is then applied to the observational model. However, such a direction is more computationally expensive, as the interpolation outside the coastlines would have to be repeated for each time step.

For temperature variables, the bias in a geographical location x (a cell with a given latitude and longitude) is given by the difference between present-day observed $T_{obs}(x,0)$ and simulated $T_{sim}^{\oplus}(x,0)$ temperature, interpolated to the desired higher resolution grid via bilinear interpolation. Downscaled temperature (T_{sim}^{DD}) in x at time t is thus estimated as

$$T_{sim}^{DD}(x,t) := T_{sim}^{\oplus}(x,t) + \left(T_{obs}(x,0) - T_{sim}^{\oplus}(x,0)\right) \quad (1)$$

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 common when applying the delta method for precipitation, which corresponds to applying the simulated relative change to the observations (Maraun and Widmann, 2018). However, this method can therefore be hypersensitive in drylands, leading to overprediction of precipitation (and thus exacerbating the 'drizzling' bias of GCM). We have therefore adopted an additive approach for precipitation, analogous to the one used for temperature, with clamping within the range of observed maximum and minimum for current climate (Beyer et al. 2020a). Like temperature, downscaled precipitation is estimated as

$$P_{sim}^{DD}(x,t) := P_{sim}^{\oplus}(x,t) + \left(P_{obs}(x,0) - P_{sim}^{\oplus}(x,0)\right) \quad (2)$$

The resulting monthly datasets were then utilised within the *pastclim* framework 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.

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 for all of the time steps available in Beyer et al. (2020a) and the HadCM3 GCM (Huntley *et al.* 2022) for the last 120,000 years. The global downscaled bioclimatic variables have been made available on Zenodo (https://doi.org/10.5281/zenodo.7828453) for

Deleted: dataset

Deleted: For temperature variables, the bias in a geographical location x is given by the difference between present-day observed and simulated temperature, interpolated to the same grid via bilinear interpolation. Delta-Ddownscaled temperature inat location x at some time t in the past is thus estimated as¶
¶
$T_{sim}^{DM}(x,t) := T_{sim}^{raw}(x,t) + \left(T_{obs}(x,0) - T_{sim}^{raw}(x,0)\right)$→→→→→→(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¶
¶
$P_{sim}^{DM}(x,t) := P_{sim}^{raw}(x,t) \cdot \frac{P_{obs}(x,0)}{P_{sim}^{raw}(x,0)}$→→→→

Deleted: →→→→

Deleted: are

Deleted: used

Deleted:

Deleted: ¶
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.¶

Deleted: REF

Deleted: is is highly computationally expensive; however, the

Formatted: Not Highlight

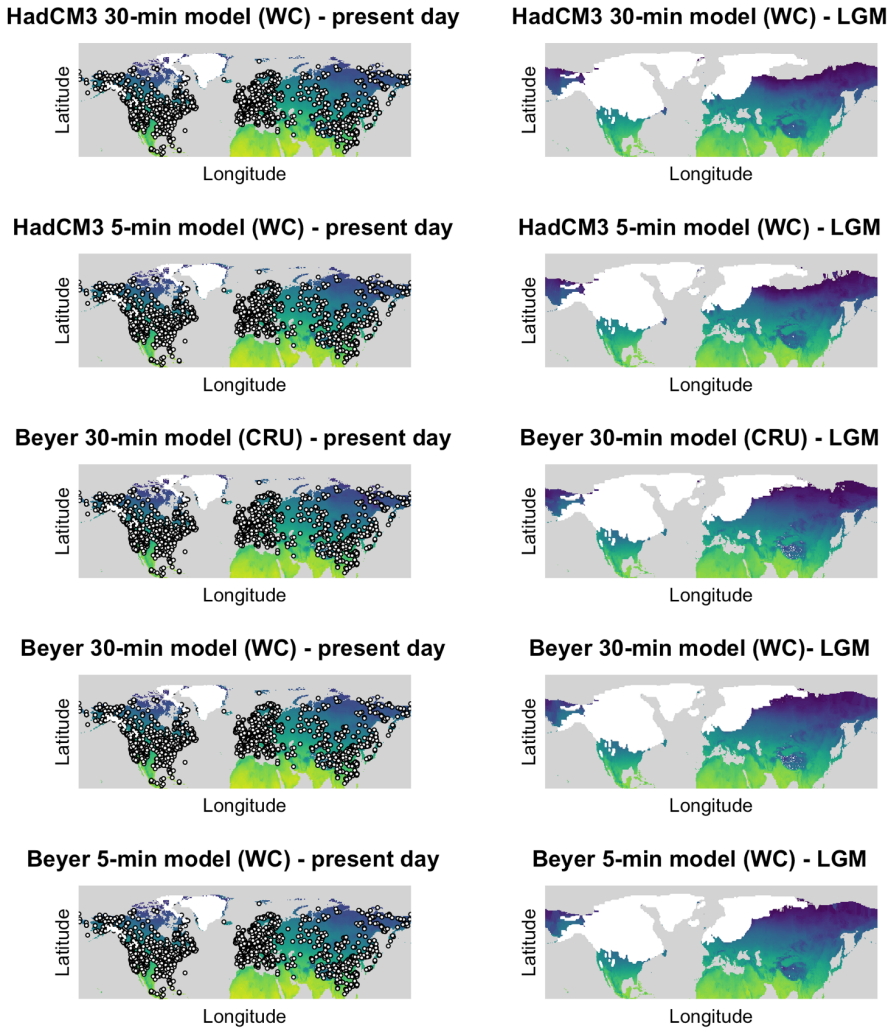
Deleted: ob

Deleted: .10.5281/zenodo.7828454

future use. Figure 1 shows the different climatic models tested in this research for both the present day and the Last Glacial Maximum (LGM) and the geographic coverage of the proxy records.

Deleted:) for future use.

Deleted: 10.5281/zenodo.7828454) for future use. Figure 1 shows the different climatic models tested in this research for both the present day and the Last Glacial Maximum (LGM) and the geographic coverage of the proxy records.


HadCM3 30-min model (WC) - present day
HadCM3 30-min model (WC) - LGM
HadCM3 5-min model (WC) - present day
HadCM3 5-min model (WC) - LGM
Beyer 30-min model (CRU) - present day
Beyer 30-min model (CRU) - LGM
Beyer 30-min model (WC) - present day
Beyer 30-min model (WC)- LGM
Beyer 5-min model (WC) - present day
Beyer 5-min model (WC) - LGM
Latitude
Longitude

525

Deleted:

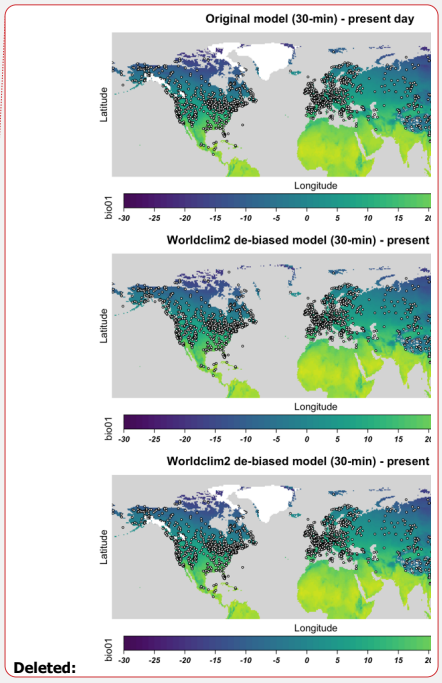

Original model (30-min) - present day
Worldclim2 de-biased model (30-min) - present
Worldclim2 de-biased model (30-min) - present
Latitude
Longitude
bio01
-30 -25 -20 -15 -10 -5 0 5 10 15

Figure 1. Site locations of proxy records studied in this analysis (left), against mean annual temperature (bio01) from the different model outputs for the present day and the Last Glacial Maximum (LGM), manipulated within pastclim (Leonardi *et al.* 2023). Land mass in each time slice is masked by global ice sheets (plotted in white) and predicted sea level.

Deleted: Beyer *et al.* (2020a) for the present day
Deleted: extracted
Deleted: from
Deleted: 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).

2.2 Proxy reconstructions

We employed the LegacyClimate 1.0 proxy dataset by Herzschuh *et al.* (2023) for direct validation of the model outputs. 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; Birks *et al.* 2010; Chevalier *et al.,* 2020). In LegacyClimate 1.0, the diverse pollen records are handled consistently 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). Our use of a single database reconstructing climate based on a single proxy reduces inter-site variability resulting from the type of data utilised and allows the generation of analogous climatic parameters with direct relevance to bioclimatic variables available in the Beyer *et al.* (2020a) model: T_{ann}, T_{july} and P_{ann} from LegacyClimate1.0 are the equivalent bioclimatic variables to bio01, bio10 and bio12 from HadCM3 GCM (Huntley *et al.* 2022) and Beyer et al. (2020a) model time series, which are standardly used in climatic modelling.

Formatted: Font: Italic
Deleted: ; t
Deleted: he
Deleted: based
Deleted: climate
Deleted: proxy
Deleted: .
Deleted:
Deleted: REF
Formatted: Font colour: Text 1
Formatted: No underline, Font colour: Text 1

To facilitate comparison between the proxy reconstructions and the model outputs, we interpolate each proxy record via bilinear interpolation to the equivalent chronological resolution of the climatic models to enable quantification of differences between the time series; interpolating to regular time intervals ensures that periods of particularly dense sampling in the original cores do not exert undue influence on the results. For this, we extracted the climatic values from the model at the coordinates of the proxy site for the time steps captured in the proxy record. Following data-cleaning, we retain 2,385 records from LegacyClimate1.0. One record was removed as it did not have any proxy data associated with the MAT method (ID Dataset: 100127), a further 25 were omitted due to a lack of consistent time steps in the models being available, and an additional 170 records were removed as they fall under the cropped sea-level of the Beyer *et al.* (2020a) model. The latter includes some proxy sites that are located on small islands not captured by the model or within lake margins. Table 1 summarises the proxy records and climatic model outputs studied in this research.

Deleted: simulations
Deleted: d
Deleted: (1,000 years)
Deleted: ¶ ¶
Deleted: ed
Deleted: 420
Deleted:
Deleted: further
Deleted: ; for example, the
Deleted: p
Deleted: s
Deleted: simulations
Deleted: REF
Deleted:)

Table 1. Summary of the proxy records selected from the LegacyClimate 1.0 (Herzschuh *et al.,* 2023) and the model outputs (Beyer *et al.,* 2020a; Huntley et al. 2022) selected for analysis of mean annual temperature (bio01, T_{ann}), mean July temperature (bio10, T_{july}) and total annual precipitation (bio12, P_{ann}).

Formatted: Subscript
Deleted: .
Formatted: Subscript
Formatted: Subscript

	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 (years)	Reference (and doi)
Legacy Climate 1.0	Asia East North America West North America Europe	2385 proxy sites	Pollen reconstructions	T_{ann} T_{July} P_{ann}	0	30	670	Herzschuh, U. *et al.* (2023). *Scientific Data* (10.5194/essd-15-2235-2023)
HadCM3 Global Circulation model	Global	30- and 5-min grid cells	Simulations, de-biased and downscaled using WordClim2 (this paper)	Bio01 Bio10 Bio12	0	120	1,000 until 23,000 years ago and then every 4,000 years	Huntley, B. et al. (2023). *Journal of Biogeography*, (10.1111/jbi.14619)
Beyer *et al.* (2020a) simulations	Global	30- and 5-min grid cells	Simulations, de-biased and downscaled using CRU (original) and WordClim2 (this paper)	Bio01 Bio10 Bio12	0	120	1,000/2,000	Beyer, R. *et al.* (2020a). Scientific Data (10.1038/s41597-020-0552-1)

Formatted Table

Deleted: per 1,000 years)

Deleted: 420

Deleted: 0.

Deleted: REF

Deleted: https://doi.org/10.1111/jbi.14619

Deleted: https://doi.org

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Deleted: /2

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 coarse 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 spatial variations in climate, we tested the statistical significance of differences in model-data 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 net effect of downscaling on model-data coherence.

These analyses allow us to evaluate the coherence between the output of the climate models and the reconstructions of specific climatic parameters from proxy data, 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 climatic model outputs and the proxy records.

3 Results

Figure 2 highlights a sample of non-interpolated time series from proxy sites across the geographic span of the LegacyClim1.0 dataset, demonstrating the coherence through time between different models and empirical reconstructions (WA-PLS and MAT) of the three climatic parameters (annual temperature, July temperature and annual precipitation). Our results show that overall proxy reconstructions and model simulations tend to highlight very similar climatic trends across variables, with average bias across all comparisons for both annual and July temperature time series falling under 1 degree Celsius and annual precipitation less than 40 mm (Fig. 2, Appendix A Tables A1-3). Considering the NRMSE, the most divergent variable on average is mean annual temperature, particularly for the output of the HadCM3 30-min model (Appendix A Tables A1-3). This result contrasts with other large-scale studies (Bartlein et al. 2011; Chevalier et al. 2021), potentially due to the assumptions made for the proxy reconstructions employed that modern analogues should be utilised from within 2000km around each site. Precipitation should be less affected given that it is more variable through space however temperature tends to be much more autocorrelated, meaning that much colder/warmer temperatures occurring in the past may not occur within these geographic limits. We find that time series of annual precipitation and July temperature show consistently lower NRMSE

Deleted: raw

Deleted: localised climatic dynamics

Deleted: mean

Deleted: both the

Deleted: reliability of the

Deleted:

Deleted: in predicting specific climatic parameters in the past,

Deleted: movement

Deleted: (

Deleted: two types of time series

Deleted: demonstrate

Deleted: remaining under

Deleted: under

Deleted: 25

Deleted: across all records

Deleted: Supplementary Tables S1-

Deleted: reconstructed

Deleted: ,

Deleted: with

Deleted: a

Deleted: ing

values than mean annual temperature across our model-data comparisons (Appendix A Tables A1-3). These two variables also show highly comparable results between different versions of the model outputs, even at varying spatial resolution and when using different modern reference datasets for downscaling (Appendix A Tables A2-3). The output from the Beyer et al. (2020a) 30-min model (CRU) shows the most consistent net positive effect of downscaling (Supplementary Table S1), probably due to the difference in modern reference data used for debiasing. However, the overall difference in coherence between the two resolutions of both outputs is judged as minimal for all three variables, particularly when controlling for the modern dataset (Appendix A Tables A1-3), as none of the subsets of model-data comparisons highlighted statistically significant differences between models at 30-min and 5-min resolution (Supplementary Table S1).

Our results based on all of the comparisons in the dataset highlight that the 30-min model time series of annual temperature from Beyer *et al* (2020a) debiased using CRU as the modern reference tends to estimate slightly lower temperatures than those produced by proxy reconstructions (as highlighted in the negative bias results reported in Appendix A Table A1). All other model outputs de-biased using WorldClim2 (WC) at both 30- and 5-min resolution contrastingly tend to predict higher annual temperatures compared to proxy records. For the HadCM3 model output, the model-data coherence is not significantly different between the 30-min and 5-min model, with less than half of the proxy records seeing improvement in coherence in the 5-min model (49% MAT method, p = 0.4904: 46% WA-PLS method, p = 0.4961; Supplementary Table S1). Similarly, annual temperature time series from the Beyer et al. (2020a) 30-min (CRU) simulations tend to have more error in only around half the records compared to the higher resolution version, at 51% (MAT method, p = 0.4904) and 50% (WA-PLS method, p = 0.4961) of proxy sites, with the Beyer et al. (2020a) 30-min (WC) having more error in slightly less than half of records compared to the Beyer et al. (2020a) 5-min model, at only 49% (MAT method, p = 0.4904) and 47% (WA-PLS method, p = 0.4961) (Supplementary Table S1).

Whether models tend to predict higher or lower precipitation compared to proxy reconstructions varies for different subsets of the data, though negative bias is particularly prominent in the 30-min model outputs compared to the 5-min equivalents (Appendix A Table A2). However, again, the overall difference in performance between the two resolutions is marginal for both model time series. Model data-coherence for annual precipitation is not significantly different between the 30-min and 5-min HadCM3 model outputs, with less than half of the records (49%) returning higher RMSE at the coarser resolution (MAT and WA-PLS method, p = 0.4943 and p = 0.4961; Supplementary Table S1). Annual precipitation time series from the Beyer et al. (2020a) 30-min model (CRU) have more error in 55% of records (MAT method and WA-PLS methods, p = 0.4923 and p = 0.4961 respectively) than the higher resolution version (Supplementary Table S1), whereas the Beyer et al. (2020a) 30-min model (WC) shows higher RMSE in 48% of time series (MAT and WA-PLS methods, p = 0.4936 and p = 4961) (Supplementary Table S1).

Deleted: Supplementary Tables S1-3). …hese latter …wo variables also show highly comparable results between different versions of the Beyer et al. (2020a) model…odel outputs, even …t varying spatial resolution and when using different modern reference datasets for downscaling (Appendix A Tables A2-3). The output from the Beyer et al. (2020a) 30-min model (CRU) shows the most consistent net positive effect of downscaling (Supplementary Table S1), probably due to the difference in modern reference data used for debiasing. However, . Overall, …he …overall difference in coherence between the two resolutions of both outputs is judged as minimal for all three variables, particularly when controlling for the modern dataset used for de-biasing …Appendix A Tables A1-3Supplementary Table S4…, ….s none of the subsets of model-data comparisons highlighted statistically significant differences between models at 30-min and 5-min resolution (Supplementary Tabl … [12]

Deleted: original

Deleted: …ends to estimate slightly lower temperatures than those produced bye…proxy reconstructions (as highlighted in the negative bias results reported in Appendix A Table A1Supplementary Table S1…..…whereas…ll other the …odel outputss…de-biased using WorldClim2 (WC) (…t both 30- and 5-min resolution)…contrastingly tend to predict higher annual temperatures compared to proxy records. For the HadCM3 model output, the model-data coherence is not significantly different between the 30-min and 5-min model, with less than half of the proxy records seeing improvement in coherence in the 5-min model (49% MAT method, p = 0.4904: 46% WA-PLS method, p = 0.4961; Supplementary Table S1). Similarly, aA…nual temperature time series from the original Beyer et al. (2020a) 30-min (CRU) simulations (at 30-min resolution) resolution) …end to have more error in only …round half the records compared to the higher resolution version, at 51% (MAT method, p = 0.4904) and 5049… (WA-PLS method, p = 0.49612… of proxy sites, when compared to the further downscaled simulations (at 5 min resolution), …ith the 30-min model de-biased by WorldClim2…eyer et al. (2020a) 30-min (WC) …aving more error in slightly less than half of records compared to the than the higher resolution model…eyer et al. (2020a) 5-min model, at only 49% (MAT method, p = 0.4904) and 476… (WA-PLS method, p = 0.49612… (Supplementary Table S14 … [13]

Formatted: Font: Not Italic

Deleted:

Formatted: Font: (Default) Times New Roman

Deleted: ies…seems to vary…aries between resolutions, regions and topographic complexity…or different subsets of the data, though negative bias is particularly prominent in the 30-min model outputs compared to the 5-min equivalents (Appendix A Table A2Supplementary Table S2…. However, again, the overall difference in performance between the two resolutions is marginal for both model time series. Model data-coherence for annual precipitation is not significantly different between the 30-min and 5-min HadCM3 model outputs, with less than half of the records (49%) returning higher RMSE at the coarser resolution (MAT and WA-PLS method, p = 0.4943 and p = 0.4961; Supplementary Table S1). (Supplementary Table S4), with the …a…nual precipitation time series from the Beyer et al. (2020a) 30-min model (CRU) original raw simulations …aveing…more error in 55% of records (both …AT method and WA-PLS methods, p = 0.4923 and p = 0.496136…respectively) than the higher resolution mode…ersionl…(Supplementary Table S1), whereas the . Y … [14]

Models of mean temperature of warmest quarter almost always slightly underestimate temperatures compared to proxy reconstructions of mean July temperature, regardless of resolution (Appendix A Table A3). This could be linked to the slight discrepancy in the climatic parameter being captured between the models and the proxies. Average difference in model-data coherence between the two spatial resolutions is not statistically significant for either the HadCM3 or the Beyer et al. (2020a) model output, with the July temperature time series from the Beyer et al. (2020a) 30-min model (CRU) showing less coherence in 58% (MAT method, p = 0.4904) and 56% (WA-PLS method, p = 0.4961) of proxy reconstructions when compared to that from the Beyer et al. (2020a) 5-min model (WC), although again the Beyer et al. (2020a) 30-min model (WC) shows higher error in less than half of the proxies (47%, MAT method, p = 0.4904, WA-PLS method, p = 0.4961) (Supplementary Table S1). Results for the HadCM3 output mirror closely that of WC-debiased Beyer et al. (2020a) models (49% for the MAT method, p = 0.4904, and 47% for the WA-PLS method, p = 0.4961).

Deleted: bio10 (
Deleted:)
Deleted: ies
Deleted: M
Deleted: T
Deleted: Supplementary Table S3
Deleted: predicted
Deleted: original simulations
Deleted:
Deleted: 04
Deleted: further downscaled
Deleted: de-biased by WorldClim2
Deleted: with just
Deleted: of the proxy reconstructions (
Deleted: 62
Deleted:) and 48% (
Deleted: 2
Deleted: 4

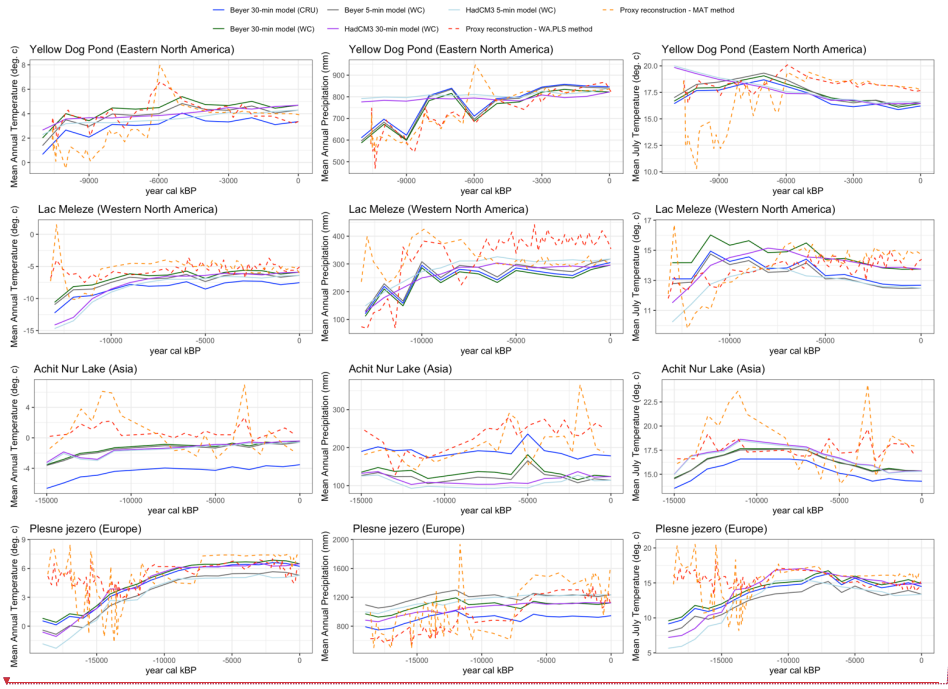


Figure 2. A sample from each regional group of simulated mean annual temperature (left), mean July temperature (middle) and total annual precipitation (right) time series, comparing different model outputs (solid lines) and corresponding non-interpolated proxy reconstructions from LegacyClimate 1.0 (Herzschuch *et al.,* 2021) (dashed lines).

Deleted: *<object>*

Deleted:

Deleted: reconstructed

Deleted: the original 30-min model, 30-min model bias-corrected using WorldClim2 and further downscaled 5-min model from Beyer *et al.* (2020a) with

Deleted: , modelled using the MAT and WA-PLS approach.

3.1 Regional differences

As highlighted in Fig. 3 and Supplementary Fig. S1-2, our results demonstrate some key differences between regions. Firstly, for annual temperature, average bias in European records is positive, suggesting model output in this region tends to overestimate temperatures compared to proxy records, whereas for all other regions annual temperature bias is negative (Appendix A Table A1). Asia and Europe have the highest NRMSE (indicating the highest divergence between proxy records and model outputs) for annual temperature in the Beyer et al. (2020a) 30-min model output (CRU) (Appendix A Table A1, Supplementary Fig. S1). However, Asia sees higher model-data coherence in both types of 30-min (WC) model outputs compared to their equivalent downscaled 5-min (WC) outputs, whereas the HadCM3 30-min model output produces very high NRMSE for European records (Appendix A Table A1; Supplementary Fig. S1). Downscaling the HadCM3 model output for annual temperature to a 5-min resolution has a positive impact on average coherence in Europe (Appendix A Table A1; Supplementary Figure S1), although this effect is reflected in less than half of the pair-wise comparisons (Supplementary Table S1). In East North America, average model-data coherence is improved by downscaling in the HadCM3 model output for annual temperature, however the Beyer et al. (2020a) 5-min model output has higher NRMSE than the equivalent 30-min model outputs (Appendix A Table A1; Supplementary Fig. S1). In West North America, the Beyer et al. (2020a) 30-min (WC) and 5-min (WC), as well as the HadCM3 5-min (WC), model outputs for annual temperature are more coherent with the proxy records than the Beyer et al. (2020a) 30-min (CRU) model and the HadCM3 3-min (WC) model outputs, with little difference between the two resolutions for the Beyer et al. (2020a) model debiased with WC (Appendix A Table A1; Supplementary Fig. S1).

Average model-data bias for precipitation varies regionally, with Europe, West North America and East North America showing consistently negative bias, suggesting that the models underestimate rainfall in these regions (Appendix A Table A1; Supplementary Fig. S1), in contrast to Asian localities where often average precipitation bias is positive. Model-data coherence for precipitation is highly similar across different resolutions of model output debiased using WC for East North America and Europe, whereas Asia and West North America has less coherence with proxy records in Beyer et al. (2020a) CRU 30-min model and the Beyer et al. (2020a) 5-min model (Appendix A Table A2; Supplementary Fig. 1). Precipitation proxy reconstructions from West North America show the highest NRMSE with the HadCM3 outputs, whereas for Asia the highest NRMSE model-data comparison is the Beyer et al. (2020a) CRU model, followed by the HadCM3 outputs (Appendix A Table A2; Supplementary Fig. S1)

July temperatures have negative model-data bias for all regions except in Asia for the WC-debiased Beyer et al. (2020a) 30-min and 5-min model output (Appendix A Table A3; Supplementary Fig. 1). In West North America, NRMSE is higher in the HadCM3 model outputs compared to that from Beyer et al. (2020a), with no differences between resolutions in the latter (when debiased using WC) and a slight improvement in coherence due to downscaling in the former (Appendix A Table A3;

Deleted:

Deleted: 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)).

Deleted: raw model de-biased with WorldClim2

Deleted: and the downscaled model

Deleted: also

Deleted: the original model,

Deleted:

Deleted: with

Deleted: Supplementary Table S1

Deleted: For precipitation, however, proxy reconstructions from West North America show the highest RMSE with the original model, like those in Asia (Supplementary Table S2).

Supplementary Fig. 1). There is no difference in average NRMSE between resolutions of model output for July temperatures in East North America, apart from the Beyer et al. (2020a) 30-min (CRU) model which has higher model-data divergence (Appendix A Table A3; Supplementary Fig. 1). In Asia, downscaling the Beyer et al. (2020a) 30-min (WC) and the HadCM3 model output improves coherence, whereas in Europe these higher resolution model outputs lead to slight decreases in coherence (Appendix A Table A3; Supplementary Fig. 1).

Fig. 3 and Supplementary Fig. S2 highlight these spatial heterogeneities in bias across the Northern Hemisphere, which could have many potential different sources, i.e. geographic variation in the performance of the model outputs, the quality of the present-day calibration data for LegacyClimate 1.0 or the modern reference used for debiasing, and/or the impact of confounding variables on the pollen-climate relationships. The East North American subset of proxy reconstructions appear to be the most coherent with the model outputs, generally showing the lowest NRSME values across all variables (Appendix A Table A1-A3; Supplementary Fig S1). Europe tends to show the lowest proportion of records where error is higher in the coarser models (30 min) compared to the higher-resolution models (5 min), with downscaling having the strongest impact on model-proxy divergence in East and North America, particularly when compared to the Beyer et al. (2020a) 30-min model (CRU) (Supplementary Table S1). Regions showing the least coherence varies depending on the climatic parameter, with Asia and East North America having the highest RMSE values for annual temperatures (Appendix A Table A1; Supplementary Fig. S1), Asia and West North America for precipitation (Appendix A Table A2; Supplementary Fig. S1) and East North America for July temperatures (Appendix A Table A3; Supplementary Fig. S1). Overall, no region shows a statistically significant difference in model-data coherence between models of different resolutions (Supplementary Table S1 and Figure S1). Indeed, often the coarser models have a higher proportion of proxy records with lower error than the 5-min models (Supplementary Table S1), particularly in Europe and Asia, suggesting higher resolutions could simply be adding noise in many scenarios.

Deleted: 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)).

Deleted: 1

Deleted: dataset

Deleted: .

Deleted: In LegacyClimate 1.0,

Deleted: t

Deleted: climate models,

Deleted: Supplementary Table S1, S3

Deleted: North

Deleted: the original model

Deleted: 4

Deleted: and July

Deleted: Supplementary Table S1, S3

Deleted: Europe

Deleted:

Deleted: that for annual precipitation

Deleted: Supplementary Table S2

Deleted: 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).

Deleted: 30-min model de-biased using WorldClim2

Deleted: s

Deleted: 4

Deleted: model

Formatted: Font colour: Text 1

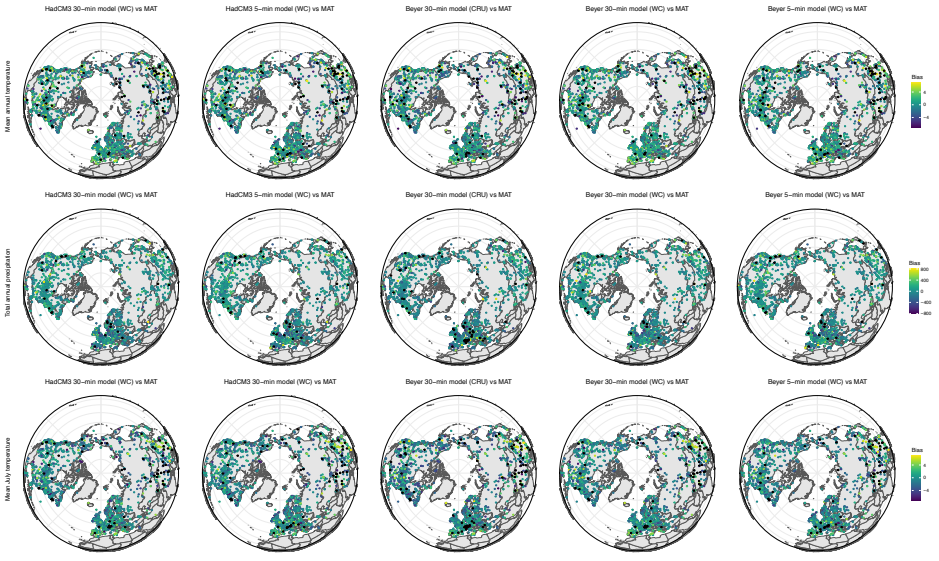
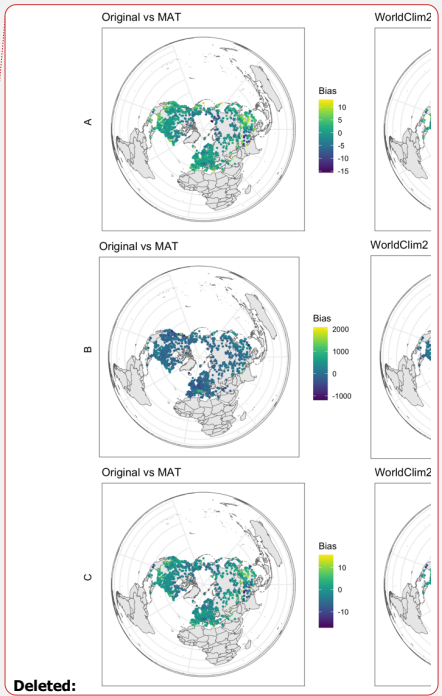


Figure 3. Absolute bias for mean annual temperature, mean annual precipitation, and mean July temperature for each proxy site, comparing the climatic values produced by the MAT method of proxy reconstruction against different versions of the HadCM3 GCM and Beyer *et al.* (2020a) model. Outliers have been highlighted in red, defined as =< -5 and => 5 degrees Celsius for mean annual temperature and July temperature, and =< -800 and => 800 millimetres for total annual precipitation. Visualisation of bias for the WA-PLS method is reported in Supplementary Figure S2.

3.2 Effects of landscape heterogeneity

Downscaling model outputs to a very high resolution is often performed to account for smaller-scale landscape features that can locally impact climatic conditions, such as topography and coastlines (Fig. 4). Figure 4 highlights these effects of increasing model resolution in different areas of varying landscape complexity; for example, in the Pittsburg Basin (which is inland and flat) there is little change in the climate signal captured at proxy sites (white circles) following downscaling, whereas, in southern Italy and the Qillian Mountains, downscaling captures more localised details in climates associated with landscape-level variations. Proxy records at higher elevations and topographic complexity may therefore be expected to show stronger coherence with the higher resolution models compared to those at relatively lower resolution.



Deleted:
Deleted: (A)
Deleted: (B)
Deleted: (C)
Deleted:)
Deleted: Pair-wise comparisons of
Deleted: are
Deleted: 1
Deleted: Further model
Deleted: d
Deleted: We therefore tested whether
Deleted: p
Deleted: varying
Deleted: s
Deleted: ose

However, our analysis presents mixed results; for example, for annual temperature, subsets of proxy records at higher altitudes and in regions of higher topographic roughness both have higher NRMSE for the 30-min HadCM3 model compared to the equivalent 5-min version for the MAT method, yet for the WA-PLS method downscaling this output increases NRMSE for records in areas of higher roughness(App endix A Table A1; Supplementary Fig. S3). Similarly, a negative effect of downscaling on model-data coherence for locations of high roughness is observed for the Beyer et al. (2020a) 30-min model output (WC) for both the MAT and WA-PLS method, as well as proxy reconstructions using the MAT-method in high altitude areas (Appendix A Table A1; Supplementary Figure S3). Annual temperature at higher elevations and topographic complexity modelled based on Beyer et al. (2020a) 30-min (CRU) has consistently higher NRMSE compared to alternate versions of this model output, although the 30-min HadCM3 30-min model is the most divergent from proxy records, particularly for high altitude locations (Appendix A Table A1; Supplementary Fig. S3). In lower altitude and flat locations, downscaling the HadCM3 model shows modest improvements in NRMSE whereas the Beyer et al. (2020a) 5-min (WC) model output is less coherent for these subsets than the equivalent 30-min (WC) version (Appendix A Table A1; Supplementary Fig. S3). In terms of proportions of records that show more error at coarser resolutions, the high altitude subset consistently has a net positive impact of downscaling for annual temperature, yet no model-data comparisons highlight statistically significant differences in coherence (Supplementary Table S1). Our results also show that proxy reconstructions tend to indicate warmer temperatures at higher elevations and/or in areas of higher topographic roughness compared to model outputs and colder temperatures at lower elevations and/or lower topographic roughness (Appendix A Table A2). This is a known bias of transfer functions when constructing more 'extreme climates' from proxies, given that elevation negatively correlates with temperature and these functions rely on averages of data from modern calibration data sets (Chevalier *et al.,* 2020).

For precipitation, only in low altitude and/or flat areas does the Beyer et al. (2020a) 30-min model (CRU) produce lower values than the proxy reconstructions, indicated by negative bias (Appendix A Table A2; Supplementary Fig. S3). NRME tends to be higher in areas of high altitude (particularly) and areas of high topographic roughness (Appendix A Table A2; Supplementary Fig. S3), however the higher resolution versions of the models do not show an improvement in coherence. For these subsets, the Beyer et al. (2020a) WC model outputs show better average coherence than the Beyer et al. (2020a) CRU and the HadCM3 outputs (Appendix A Table A2; Supplementary Fig. S3). Our results highlight that subsets of low altitude and low roughness proxy records tend to show more instances of downscaling improving the model-data coherence compared to subsets of high altitude and high roughness records, although these are minimal and not statistically significant (Supplementary Table S1, Fig. S3).

Models of July temperatures always produce lower values than that of proxies, regardless of landscape properties (Appendix A Table A3; Supplementary Fig. S3). Our results suggest that, apart from downscaling the HadCM3 model output where minimal improvements in NRMSE are noted, model-data coherence for July temperature is not effected by model resolution

Deleted: 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 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).

Deleted: predict

Deleted: simulations

Deleted: Supplementary Table S2

Deleted: ,

Deleted: which

Deleted: , which

Deleted: and tend to overlook climate extremes

Deleted:

Deleted: original model

Deleted: tend to

Deleted: ; in all other scenarios, the models overestimate rainfall relative to pollen records

Deleted: Supplementary Table S2

Deleted: Supplementary Table S3

when controlling for the modern referenced used to debias (Appendix A Table A3; Supplementary Fig. S3). Overall, we find that the proportion of proxy records that show higher error (NRMSE) 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 (Supplementary Table S1, Fig. S3).

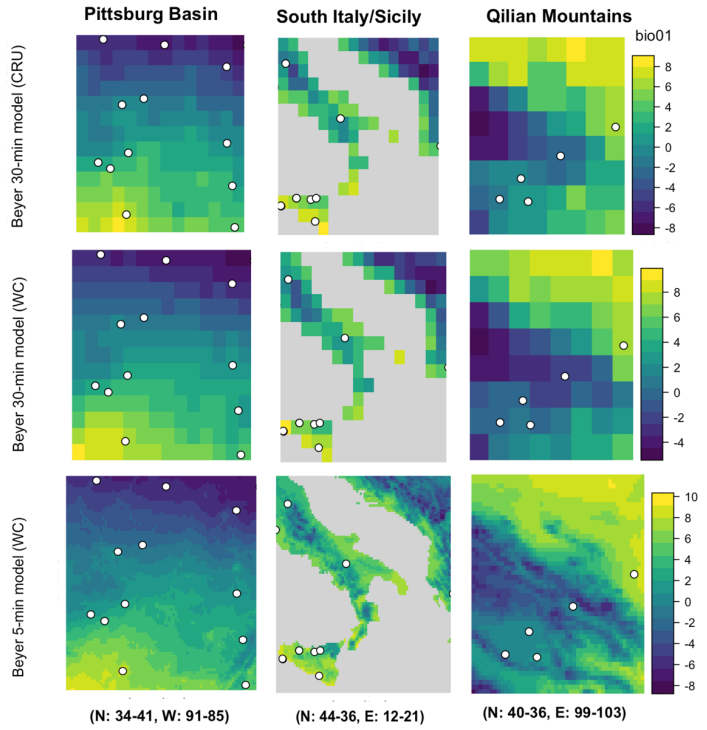


Figure 4. Three regional examples of modelled mean annual temperature for the present day (bio01), demonstrating how downscaling increases spatial resolution by capturing the effects of landscape dynamics through space on climate depending on the underlying topography. Geographic variability in temperature is shown, as simulated by the Beyer *et al.* (2020a) 30-min model output (CRU), Beyer *et al.* (2020a) 30-min model output (WC), and Beyer *et al.* (2020a) 5-min model output (WC). Locations of proxy locations from LegacyClimate 1.0 are shown as white circles.

Deleted: W

Deleted: *<object>*

Deleted: A) original

Deleted: at 30-min resolution

Deleted: B)

Deleted: WorldClim2 de-biased model at 30-min resolution, C)

Deleted: WorldClim2 downscaled model at 5-min resolution.

Deleted:

3.3 Glacial versus inter-glacial variability

We then examined discrepancies in model-data coherence through time, separating time slices from the model outputs covering the present day (i.e. timeslice 0), Marine Isotope Stage 1 (MIS 1; 0 – 14,000 years ago) and MIS 2 (14-29,000 years ago). In total, 1060 records were associated with the present day (44% of dataset), 2363 records captured time slices in MIS 1 (99% of dataset) whereas 473 spanned into MIS 2 (20%). Separate analysis of interpolated data points capturing the present day was performed, as the pollen proxies captured in these records 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 NRMSE (Appendix Tables A1-3; Fig. 5), though considerable error in some time series exists (Fig. 5). In contrast, the smaller subset of time series covering MIS 2 show the highest bias and NRMSE (Appendix Tables A1-3; Fig. 5), both across model outputs and resolutions, as well as methods of proxy reconstruction. Models capturing older time periods underestimate annual and July temperatures compared to proxy reconstructions and (often) overestimate annual precipitation (Appendix A Tables A1-A3; Fig. 5). 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 almost always over half for the present day, with annual temperature and July temperature during MIS 2 seeming to also see a slight benefit of downscaling, though this is not statistically significant for any comparison (Supplementary Table S1).

Figure 5 highlights the differences 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. Delta-downscaled models are also inherently designed 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 even after normalisation– there are many extreme values– confirming that age is just one contributing factor in the divergence between time series (Supplementary Fig. S1, S3).

Deleted: series

Deleted: 8

Deleted: timeslice 0

Deleted: only

Deleted: 5

Deleted: 98

Deleted: only

Deleted: 5

Deleted: Individual

Deleted: se

Deleted: 4

Deleted: between models and proxies

Deleted: model

Deleted: nd

Deleted: ,

Deleted: with models

Deleted: ing

Deleted:

Deleted: ing

Deleted: compared to proxy-based reconstructions

Deleted: Supplementary Tables S1-S3

Deleted: is less than half for most chronological subsets,

Deleted: only

Deleted: to

Deleted: (Supplementary Table S4),

Deleted: (p = 0.425 to 0.4962).

Deleted: overlap

Deleted: M

Deleted: calibrated

Deleted: skewed

Deleted: at the right-hand side

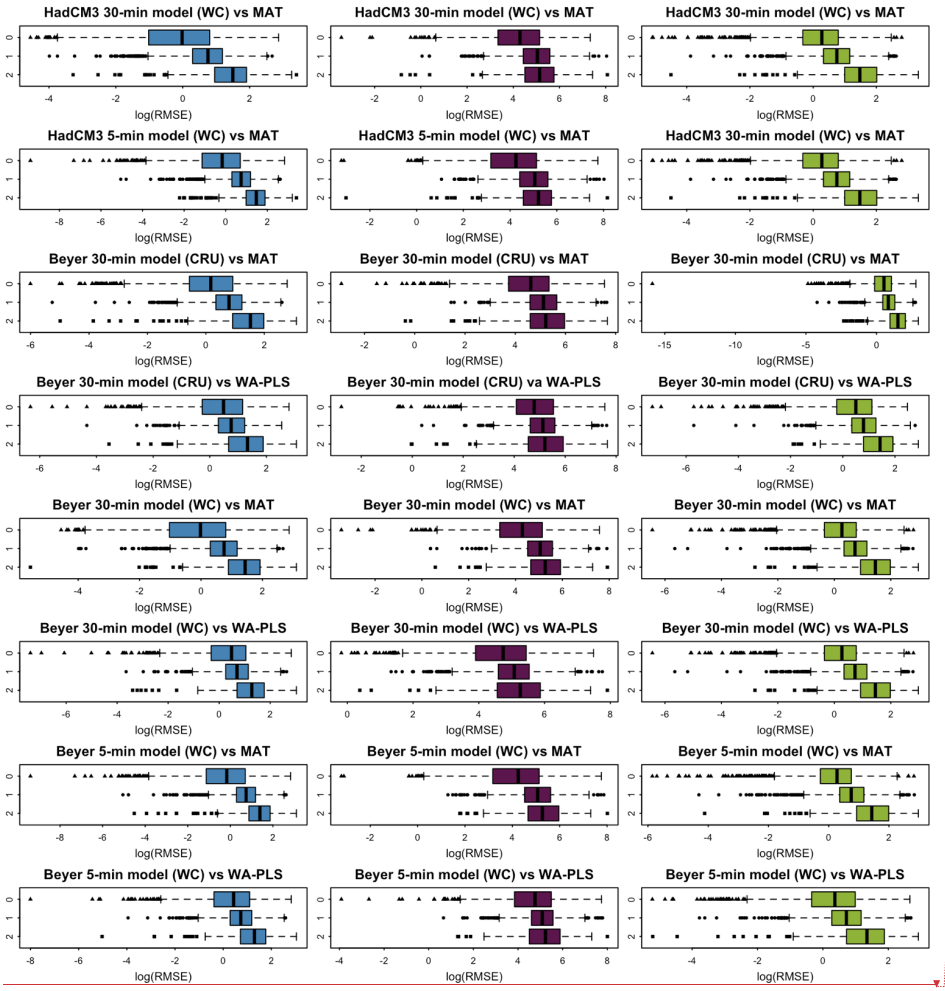


Figure 5. Boxplots of pair-wise log root mean square error (RMSE) results model-data comparisons of mean annual temperature (blue), mean annual precipitation (purple) and mean July temperature (green) from those representing the present (0), MIS 1 (1) and MIS 2 (2).

Deleted: <object>

Deleted:

Deleted: , 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 and Supplementary Figures S1 and S3, 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 (Appendix A Table A4). None of the individual records fall into the most divergent subset for all three variables studied, suggesting more extreme divergence is not related to any systematic issue in the model nor the proxy at specific locations. 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 (Appendix A Table A4).

To summarise, 44 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 cover the present day and/or MIS 2, and/or are located in Asia, areas of high altitude and/or low roughness (Appendix A Table A4). For mean annual precipitation, only 21 records consistently fall in the top 5% based on RMSE, demonstrating higher inconsistency in pairwise model-data coherence between different model versions and methods of proxy reconstruction compared to the temperature variables (Appendix A Table A4). We found that, for this parameter, significantly higher proportions of these outliers are located in Asia and West North America and/or in areas of high altitude and high roughness (Appendix A Table A4). Finally, for mean July temperature, 30 time series always fall into the most divergent 5%, significantly higher proportions of which date to the present and/or MIS 2, are located in Asia, areas of high altitude and/or areas of low topographic roughness than would be expected by chance (Appendix A Table A4).

Our results highlight that records spanning into MIS 2 consistently exhibit significantly higher proportions of divergent time series across all variables (Appendix A Table A4). This may specifically be a consequence of low CO_2 during MIS 2, which was not considered in LegacyClimate1.0, although this would mainly have an effect on moisture-related variables rather than temperature. Another potential source of divergence, leading to warmer reconstructions by proxies compared to the model outputs as well as significant deviations in precipitation, could derive from the geographic limits imposed on the LegacyClimate1.0 proxies for the modern samples used to perform reconstructions. This is particularly problematic for the LGM as comparable signals should be present in the modern climate space within the limit defined (2000km around each site), which is likely unreasonable for some areas (e.g. northerly areas of Europe, see Fig. 1). Similarly, we find sites in Asia and higher altitude areas, where modern calibration data tend to be more limited, also have more divergent time series than expected given the sample size of this subset for all three variables (Appendix A Table A4). Sites in flatter areas exhibit significantly higher proportions of divergent time series for annual and July temperatures than expected by random chance, whereas sites in higher roughness locations and West North America are more highly divergent than expected in precipitation (Appendix A

Deleted: Table 2

Deleted: individual

Deleted: sites

Deleted: Table 2

Formatted: Font colour: Text 1

Deleted:

Deleted: s

Deleted: in Asia and

Deleted: dating to MIS 2 consistently

Deleted: across climatic variables

Table A4). Interestingly, we find that proxy records that capture the present day also occur in the most divergent subset more often than expected for annual temperature and precipitation, however this is because many of these records also span into MIS 2 (Appendix A Table A4).

4 Discussion

Increasing the spatial resolution of model time-series is often thought to be required to more accurately capture the climatic conditions of specific places at specific times. But what is the optimal spatial resolution for adequately detailing finer-scale signals? We tackle this question by testing the agreement between different model outputs and empirical reconstructions from pollen proxies from the Late Quaternary for annual and July temperatures and annual precipitation. Ground-truthing modelled climate in this way is common, as proxies are considered to be the 'gold standard' for capturing more localised variations in climatic conditions in specific places. Our results highlight that further downscaling models via the delta method to much higher resolutions (5-minute) fails to *consistently* capture more of the climatic trend from pollen proxy records. Indeed, we were unable to demonstrate any statistically significant differences in model-data coherence between 30-min and 5-min model resolutions in any subset of this large dataset. Overall, this implies that more downscaling may not always be the best solution, with relatively coarser simulations (i.e. 30-minute) providing a similarly adequate representation of past climatic trends in many scenarios, even in areas of topographic complexity.

Regardless of resolution, we find that model-data coherence predictably decreases with age, with more divergent time series than expected by chance located in Asia, at higher altitudes and those capturing MIS 2. Annual precipitation and July temperature show consistently lower NRMSE than annual temperature, indicating good overall agreement between simulations and empirical reconstructions for these variables. Annual temperature shows low model-data convergence with greater disparity between model outputs and methods of proxy reconstructions, as well as in certain contexts. Variability in coherence between regions likely relates to spatial variability in the performance of the simulations, the quality of modern reference datasets and proxy data employed, and the complexity in relationships between pollen and temperature tolerances in different geographic areas. Moreover, greater divergence at high altitudes and at older time scales may reflect limitation in the calibration with modern conditions, with reduced modern reference data at higher elevations and a lack of good analogues of glacial/periglacial vegetations in the same areas as those in the past.

For this large-scale comparative analysis, we employed different de-biased and downscaled versions of the HadCM3 GCM output (Huntley et al. 2022) and Beyer *et al's.* (2020a) Late Pleistocene and Holocene climate simulations alongside harmonised pollen records from LegacyClimate1.0 (Herzschuh *et al.* 2023), providing corresponding estimates of three key climatic parameters for comparison between time series. Whilst the LegacyClimate 1.0 dataset provides an excellent standardised and spatiotemporal expansive resource to address whether downscaling to higher resolutions is effective in capturing local climatic details, it is worth noting that, because the type of proxy records employed tend to capture pollen from

Formatted: Font colour: Text 1

Deleted: (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 mod... [15]

Formatted: Font colour: Text 1

Formatted: Normal

Deleted: s

Deleted: better

Deleted: capture

Deleted: nuance in

Deleted: e

Deleted: capturing

Deleted: these

Deleted: dynamics of

Deleted: and climatic shifts

Deleted: signal

Deleted: , which are the gold standard for capturing localis... [16]

Formatted: Font: Italic

Deleted:

Deleted: T

Deleted: is

Deleted: tter

Deleted: seeming to

Deleted: e

Deleted: Nonetheless

Deleted: the

Deleted: the

Deleted: ,

Deleted: whereas

Deleted: a

Deleted: tends to

Deleted: , specifically

Deleted: in Asia

Deleted: ,

Deleted: (where modern calibration data tend to be more l... [17]

Deleted:)

Deleted: In our analysis,

Deleted: emulator

a broad catchment, they may represent geographically wide averages of 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.

Our results suggest that using statistical methods of downscaling simulated time series to much higher resolutions does not significantly improve the agreement between model output and pollen-proxy reconstructions, yet 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 (e.g. Boisard et al. 2025). 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 annual temperatures predicted by delta-downscaled models may better characterise the on-site environment than that also incorporating environmental trends in surrounding lower altitude landscape (Timbrell *et al.* 2022). Other methods of increasing model output, such as dynamical downscaling, may be better equipped for more localised applications, yet these are largely inaccessible for consumers of model output in fields like palaeoecology and archaeology where the computational costs are impractical. Overall, we present a streamlined pipeline for delta-downscaling climate model time series within the *pastclim* R package (Leonardi et al. 2023), though we stress that careful consideration is required to select the optimal method and spatial resolution, based on the scope of the research question at hand.

Deleted: Whilst o
Deleted: show
Deleted: necessarily
Deleted: simulations
Formatted: Font: Italic
Deleted: C
Deleted: therefore
Deleted: when using models,
Deleted:
Deleted: for

5 Conclusion

Paleoclimatic proxies and climate models constitute two contrasting yet complementary sources of information on past climates. Demand for high-resolution climatic simulations that characterise landscape-scale heterogeneities come from the multitude of fields that employ ecological data, such as those that wish to map species distributions through time and space or quantitatively test hypotheses about the impact of climatic change and/or variability on various biological or behavioural phenomena. We show that downscaling via the delta-method fails to consistently capture more signal from temperature and precipitation proxy reconstructions, though model time series at both median (30-arc minutes) and fine-grained (5-arc minutes) spatial resolutions characterise climatic variables in broadly similar ways to pollen proxies. Utilising model output for analyses of past climate therefore involves a careful balancing act between accentuating variations relevant to the study questions and the potential introduction of error by unreliable interpolation.

Deleted: C
Deleted: are needed to produce more accurate models within
Deleted: and/
Deleted: investigate
Deleted: and/
Deleted: additional
Deleted: from
Deleted: the
Deleted: s
Deleted: y reconstructions
Deleted: s
Deleted: nonetheless
Deleted: .

Code and data availability

The workflow to downscale climate model 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 models at 5-arc minutes resolution are stored on Zenodo: https://doi.org/10.5281/zenodo.7828453.

Deleted: s

Deleted: is

Deleted: https://doi.org/10.5281/zenodo.7828454.

Author contributions

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

Acknowledgements

LT, MCo and ES are supported by funding awarded by the Max Planck Society to the Human Palaeosystems Group. 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. AVP is supported by the Natural Environment Research Council grant number: NE/S007164/1.

Deleted:

Deleted: is

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

Bartlein, P.J., Harrison, S.P., Brewer, S. *et al.*: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. Clim Dyn 37, 775–802. https://doi.org/10.1007/s00382-010-0904-1 2011.

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.

Formatted: Font: Not Italic

Formatted: Font: Not Bold

Field Code Changed

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.

Birks, H.J., Heiri, O., Seppä, H., Bjune, A.E.: Strengths and Weaknesses of Quantitative Climate Reconstructions Based on Late-Quaternary Biological Proxies. The Open Ecolog. J. *3*, 68-110. http://dx.doi.org/10.2174/1874213001003020068 2010.

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.

Boisard, S., Wren, C., Timbrell, L. and Burke, A.: Climate frameworks for the Middle Stone Age and Later Stone Age in Northwest Africa. Quat. Int., 716, 109593. https://doi.org/10.1016/j.quaint.2024.109593

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.

Chevalier, M., B. A. S. Davis, M. J. A. B. de Beaulieu, R. A. DeGroot, S. P. Harrison, F. L. L. O. de Lima, and G. S. D. Fréville.: Pollen-based continental climate reconstructions at ~6 and ~0.7 ka for data–model comparison." Sci Rep 11(1), 11468. DOI: 10.1038/s41598-021-90866-0. 2021.

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

Fitzpatrick, M.C., Lachmuth, S., Haydt, N.T.: The ODMAP protocol: a new tool for standardized reporting that could revolutionize species distribution modelling. Ecograph. 44(7), 1067-1070. https://doi.org/10.1111/ecog.05700. 2021

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T.M.L., Otto-Bliesner, B.L., Chan, K.C. and Brook, B.W.: PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. Ecograph 40(11), 1348-1358. https://doi.org/10.1111/ecog.03031, 2017.

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.

Huntley, B., Allen, J.R., Forrest, M., Hickler, T., Ohlemüller, R., Singarayer, J.S. and Valdes, P.J.: Global biome patterns of the Middle and Late Pleistocene. *J. Biogeog, 50*(8), pp.1352-1372. 2023.

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. pastclim 1.2: an R packages to easily access and use paleoclimatic reconstructions. Ecography. e06481. https://doi.org/10.1111/ecog.06481, 2023

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

Deleted: 2023.

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.

Mondanaro, A., Girardi, G., Castiglione, S., Timmermann, A., Zeller, E., Venugopal, T., Serio, C., Melchionna, M., Esposito, a., Di Febbrano, M. and Raia, P.: EutherianCoP. An integrated biotic and climate database for conservation paleobiology based on eutherian mammals. Sci. Data., 12, 6 https://doi.org/10.1038/s41597-024-04181-4, 2025

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

Rummukainen, M., 2016. Added value in regional climate modeling. Wire Clim. Change 7, 145e159. https://doi.org/10.1002/wcc.378, 2016.

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.

Strandberg, G., Lindström, J., Poska, A., Zhang, Q., Fyfe, R., Githumbi, E., Kjellström, E., Mazier, F., Nielsen, A.B., Sugita, S., Trondman, A.-K., Woodbridge, J., Gaillard, M.-J., 2022: Mid-Holocene European climate revisited: New high-resolution regional climate model simulations using pollen-based land-cover, Quaternary Science Reviews, Volume 281, 107431, ISSN 0277-3791,

Strandberg, G., Chen, J., Fyfe, R., Kjellström, E., Lindström, J., Poska, A., Zhang, Q., and Gaillard, M.-J.: Did the Bronze Age deforestation of Europe affect its climate? A regional climate model study using pollen-based land cover reconstructions, Clim. Past, 19, 1507–1530, https://doi.org/10.5194/cp-19-1507-2023, 2023.

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.

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.

Yun, K-S., Timmerman, A., Lee, S-S., Willeit, M., Ganopolski, A. and Jadhav, J.: A transient coupled general circulation model (CGCM) simulation of the past 3 million years. Clim. Past, 19, 1951–1974, https://doi.org/10.5194/cp-19-1951-2023, 2023.

Zeller, E., and Timmermann, A.: The evolving three-dimensional landscape of human adaptation. Sci. Adv. 10, eadq3613, https://doi.org/10.1126/sciadv.adq3613, 2024

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.

Appendix A:

Table A1a. Summary of results for mean annual temperature from Legacy.Climate 1.0 using the modern analogue technique (MAT). Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated annual temperature for each records and averaged over each subset, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased and downscaled using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the modern analogue (MAT) techniques.

Mean annual temperature - Modern analogue technique (MAT)

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	2.86	2.73	2.73	2.77	2.78	2.86	2.17	2.66	3.65	2.05	-0.09	0.50	0.36	0.24	0.12
Asia (n = 455)	3.95	3.74	3.74	3.85	3.85	3.35	1.68	2.43	1.65	3.82	-0.11	0.48	0.63	0.16	0.31

East N America (n = 613)	2.61	2.39	2.39	2.44	2.44	1.71	1.90	2.99	2.22	1.70	-0.28	0.31	0.21	0.22	0.11
West N America (n = 328)	3.00	2.82	2.73	2.92	2.82	2.82	2.28	2.21	2.83	1.71	-0.03	0.62	0.35	0.43	0.17
Europe (n = 989)	2.47	2.44	2.48	2.43	2.49	3.37	2.51	2.72	5.72	1.58	0.01	0.58	0.33	0.22	0.01
Present (n = 1060)	1.90	1.73	1.60	1.73	1.60	0.64	0.65	0.67	0.65	0.67	0.44	1.05	0.80	1.05	0.80
MIS 1 (n = 2363)	2.70	2.57	2.58	2.59	2.61	2.15	2.10	2.46	3.64	1.85	0.11	0.71	0.57	0.40	0.28
MIS 2 (n = 473)	5.22	4.94	4.93	5.13	5.11	3.83	3.08	53.25	5.15	7.17	-3.28	-2.76	-2.70	-2.21	-2.16
High altitude (n = 362)	3.35	3.10	2.97	3.13	3.00	7.28	3.43	3.29	10.17	2.74	-0.62	0.02	-0.18	-0.12	-0.32
Low altitude (n = 2058)	2.78	2.67	2.69	2.70	2.74	2.13	1.96	2.54	2.56	1.95	-0.01	0.58	0.45	0.30	0.19
High roughness (n = 412)	2.94	2.74	2.71	2.78	2.76	6.42	1.61	1.88	2.18	2.55	-0.37	0.29	-0.05	0.07	-0.26

Low roughness (n =2008)	2.85	2.73	2.74	2.76	2.78	2.16	2.29	2.84	3.97	1.96	-0.04	0.54	0.44	0.27	0.19

Table A1b. Summary of results for mean annual temperature from Legacy.Climate 1.0 using weighted average partial least squares (WA-PLS) technique. Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated annual temperature, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the weighted average partial least squares (WA-PLS) technique.

Mean annual temperature - Weighted average partial least squares (WA-PLS)

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	2.77	2.64	2.71	2.64	2.72	3.20	1.75	2.66	3.22	1.90	-0.11	0.48	0.35	0.22	0.10
Asia (n = 455)	3.82	3.63	3.68	3.70	3.73	4.01	1.48	2.09	1.58	3.23	-0.05	0.54	0.70	0.23	0.38
East N America (n = 613)	2.59	2.38	2.41	2.37	2.40	1.61	1.76	2.65	2.12	1.57	-0.44	0.16	0.06	0.06	-0.04

West N America (n = 328)	2.93	2.76	2.77	2.81	2.81	2.67	2.34	2.22	3.20	2.03	-0.12	0.53	0.26	0.33	0.08
Europe (n = 989)	2.35	2.30	2.44	2.27	2.43	3.98	1.66	3.06	4.65	1.44	0.07	0.64	0.40	0.29	0.07
Present (n = 1060)	2.24	2.18	2.14	2.17	2.14	0.96	0.98	1.34	0.97	1.34	0.47	1.08	0.82	1.07	0.82
MIS 1 (n = 2363)	2.64	2.51	2.59	2.51	2.61	2.37	1.73	2.53	3.24	1.83	0.09	0.69	0.55	0.38	0.26
MIS 2 (n = 473)	4.60	4.34	4.39	4.22	4.22	4.33	2.74	36.52	8.22	5.56	-3.06	-2.53	-2.48	-1.98	-1.94
High altitude (n = 346)	3.11	2.90	2.93	2.91	2.93	9.10	3.33	4.12	9.52	2.95	-0.60	0.04	-0.16	-0.10	-0.30
Low altitude (n = 2023)	2.72	2.59	2.68	2.59	2.69	2.21	1.48	2.42	2.16	1.73	-0.02	0.56	0.43	0.28	0.17
High roughness (n = 398)	2.76	2.55	2.66	2.59	2.70	8.02	1.48	1.75	1.89	2.38	-0.36	0.30	-0.04	0.08	-0.25
Low roughness (n =1971)	2.78	2.65	2.72	2.65	2.73	2.24	1.81	2.86	3.50	1.81	-0.06	0.52	0.42	0.25	0.17

Table A2a. Summary of results for mean total annual precipitation from Legacy.Climate 1.0 using the modern analogue technique (MAT). Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated annual precipitation, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the modern analogue (MAT) techniques.

Mean total annual precipitation - Modern analogue technique (MAT)

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	236.64	211.62	217.62	206.27	212.50	0.32	0.32	0.33	0.30	0.31	2.56	7.61	22.58	8.84	21.15
Asia (n = 455)	201.45	206.20	214.40	195.44	204.16	0.39	0.53	0.57	0.43	0.44	21.91	41.24	42.91	39.73	42.52
East N America (n = 613)	177.85	172.77	169.42	169.30	166.46	0.26	0.23	0.22	0.22	0.21	-26.50	27.16	31.93	46.42	49.29

West N America (n = 328)	222.22	210.40	225.75	205.67	221.81	0.31	0.27	0.27	0.26	0.27	18.01	57.35	81.34	68.14	92.02
Europe (n = 989)	294.05	238.59	246.28	234.36	241.79	0.33	0.30	0.30	0.31	0.31	6.53	-36.48	-12.05	-48.33	-29.63
Present (n = 1060)	175.39	137.21	137.47	137.47	137.47	0.24	0.19	0.18	0.19	0.18	-14.81	-3.60	17.00	-4.02	17.00
MIS 1 (n = 2363)	230.94	205.23	211.05	203.15	208.71	0.31	0.29	0.29	0.29	0.29	-1.72	2.83	17.52	7.77	19.64
MIS 2 (n = 473)	299.17	279.95	283.03	240.27	247.04	0.46	0.60	0.65	0.79	0.65	142.10	138.68	142.04	-24.60	73.24
High altitude (n = 346)	297.38	219.76	225.67	212.06	219.56	0.36	0.33	0.34	0.35	0.36	90.02	24.08	36.36	21.34	33.76
Low altitude (n = 2023)	226.47	210.34	216.28	205.31	211.31	0.32	0.28	0.33	0.29	0.30	-12.55	4.46	19.94	6.58	18.92
High roughness (n = 398)	309.53	226.49	245.27	221.11	239.13	0.36	0.31	0.31	0.31	0.31	60.16	1.88	36.07	-1.90	29.94
Low roughness (n =1971)	222.14	208.73	212.08	203.31	207.14	0.31	0.32	0.33	0.30	0.29	-9.22	8.43	19.56	10.89	19.30

Table A2b. Summary of results for mean total annual precipitation from Legacy.Climate 1.0 using weighted average partial least squares (WA-PLS) technique. Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated annual precipitation, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the weighted average partial least squares (WA-PLS) technique.

Mean total annual precipitation - Weighted average partial least squares (WA-PLS)

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	228.91	211.12	217.76	206.27	210.54	0.32	0.32	0.33	0.30	0.31	-8.04	-2.99	11.98	-1.76	10.55
Asia (n = 455)	187.28	192.43	201.02	180.10	189.80	0.37	0.49	0.53	0.41	0.43	5.66	24.98	26.66	23.47	26.27
East N America (n = 613)	177.35	173.37	172.03	164.04	164.24	0.27	0.25	0.24	0.22	0.22	-39.19	14.47	19.23	33.73	36.60

West N America (n = 328)	217.52	217.75	234.28	205.28	222.60	0.32	0.29	0.30	0.27	0.27	18.95	58.28	82.27	69.07	92.95
Europe (n = 989)	283.79	240.90	284.33	239.28	244.78	0.33	0.31	0.31	0.32	0.32	-3.99	-47.00	-22.57	-58.85	-40.15
Present (n = 1060)	194.19	172.78	178.90	172.27	178.90	0.27	0.25	0.25	0.25	0.25	-22.28	-11.07	9.53	-11.49	9.53
MIS 1 (n = 2363)	222.29	204.37	210.98	201.43	207.38	0.31	0.30	0.30	0.29	0.30	-13.05	-8.49	6.19	-3.56	8.31
MIS 2 (n = 473)	295.46	273.61	272.93	224.25	228.86	0.45	0.57	0.61	0.87	0.67	154.24	150.82	154.18	84.65	85.63
High altitude (n = 346)	287.16	229.44	237.83	217.33	227.33	0.35	0.34	0.35	0.35	0.36	80.14	14.20	26.48	11.46	23.89
Low altitude (n = 2023)	219.33	208.27	214.56	202.00	207.97	0.31	0.32	0.33	0.30	0.30	-23.43	-6.24	9.05	-4.30	8.04
High roughness (n = 398)	289.89	234.98	253.52	228.46	246.67	0.35	0.32	0.32	0.32	0.32	43.22	-15.06	19.13	-18.84	13.00
Low roughness (n =1971)	216.99	206.59	210.78	199.34	203.55	0.31	0.32	0.33	0.30	0.31	-18.71	-1.06	10.07	1.40	9.81

Table A3a. Summary of results for mean July temperature from Legacy.Climate 1.0 using the modern analogue technique (MAT). Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated July temperature, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the modern analogue (MAT) techniques.

Mean July temperature - Modern analogue technique (MAT)

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	3.01	2.72	2.74	2.74	2.75	0.27	0.20	0.20	0.23	0.23	-0.76	-0.27	-0.41	-0.29	-0.40
Asia (n = 455)	3.95	3.72	3.76	3.70	3.68	0.42	0.28	0.28	0.35	0.32	-0.16	0.09	0.24	-0.19	-0.03
East N America (n = 613)	2.82	2.41	2.38	2.33	2.31	0.23	0.17	0.17	0.18	0.18	-0.98	-0.47	-0.55	-0.44	-0.52

West N America (n = 328)	3.03	2.60	2.56	2.75	2.69	0.27	0.19	0.19	0.33	0.30	-0.78	-0.22	-0.49	-0.34	-0.57
Europe (n = 989)	2.68	2.48	2.55	2.56	2.62	0.21	0.17	0.19	0.19	0.20	-0.88	-0.32	-0.59	-0.22	-0.45
Present (n = 1060)	2.21	1.78	1.77	1.79	1.77	0.19	0.13	0.14	0.13	0.14	-0.95	-0.50	-0.78	-0.52	-0.78
MIS 1 (n = 2363)	2.84	2.55	2.57	2.55	2.56	0.23	0.18	0.18	0.20	0.20	-0.53	-0.04	-0.18	-0.06	-0.17
MIS 2 (n = 473)	5.42	5.22	5.17	5.39	5.35	1.25	0.74	0.76	2.72	0.79	-3.53	-3.19	-3.14	-3.26	-3.21
High altitude (n = 346)	3.60	3.18	3.10	3.15	3.04	0.35	0.26	0.25	0.31	0.26	-1.23	-0.69	-0.89	-0.67	-0.86
Low altitude (n = 2023)	2.91	2.64	2.68	2.68	2.71	0.25	0.19	0.19	0.22	0.22	-0.67	-0.19	-0.32	-0.22	-0.33
High roughness (n = 398)	3.21	2.86	2.86	2.85	2.85	0.40	0.23	0.24	0.24	0.24	-1.09	-0.48	-0.85	-0.42	-0.77
Low roughness (n =1971)	2.97	2.70	2.72	2.72	2.74	0.24	0.19	0.19	0.23	0.23	-0.69	-0.22	-0.32	-0.26	-0.33

Table A3b. Summary of results for mean July temperature from Legacy.Climate 1.0 using weighted average partial least squares (WA-PLS) technique. Mean bias, root mean square error (RMSE) and normalised root mean square error (RMSE) is calculated for interpolated July temperature, comparing the outputs from the Beyer et al. (2020a) 30-min model debiased using Climate Research Unit Global Climate Dataset (CRU), Beyer et al. (2020a) 30-min model debiased using WorldClim2 (WC) data, Beyer et al. (2020a) 5-min model debiased using WorldClim2 data, HadCM3 30-min model debiased using WorldClim2 data and HadCM3 5-min model debiased using WorldClim2 data. These are compared against the chronologically equivalent proxy data reconstructed by Herzshuch *et al.* (2021) via the weighted average partial least squares (WA-PLS) technique.

Mean July Temperature - Weighted average partial least squares (WA-PLS)

Formatted Table

	RMSE					NRMSE					Bias				
	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)	Beyer 30-min (CRU)	Beyer 30-min (WC)	Beyer 5-min (WC)	HadCM3 30-min (WC)	HadCM3 5-min (WC)
All records (n = 2395)	2.89	2.63	2.72	2.59	2.67	0.26	0.19	0.20	0.23	0.23	-0.78	-0.29	-0.43	-0.31	-0.43
Asia (n = 455)	3.92	3.72	3.79	3.68	3.70	0.42	0.28	0.28	0.36	0.33	0.02	0.27	0.41	-0.01	0.15
East N America (n = 613)	2.71	2.34	2.35	2.19	2.21	0.23	0.17	0.17	0.17	0.18	-1.09	-0.58	-0.66	-0.55	-0.63
West N America (n = 328)	2.87	2.50	2.57	2.62	2.66	0.23	0.19	0.20	0.33	0.31	-1.03	-0.47	-0.74	-0.59	-0.82
Europe (n = 989)	2.54	2.37	2.51	2.34	2.48	0.26	0.17	0.19	0.17	0.19	-0.87	-0.31	-0.58	-0.21	-0.44
Present (n = 1060)	2.22	1.93	2.04	1.94	2.04	0.20	0.15	0.17	0.15	0.17	-0.99	-0.53	-0.81	-0.55	-0.81
MIS 1 (n = 2363)	2.74	2.49	2.58	2.43	2.51	0.22	0.17	0.18	0.19	0.20	-0.56	-0.06	-0.21	-0.08	-0.20
MIS 2 (n = 473)	4.90	4.65	4.65	4.67	4.69	1.13	0.70	0.79	3.05	0.73	-3.35	-3.02	-2.96	-3.08	-3.03
High altitude (n = 346)	3.42	3.05	3.14	3.03	3.06	0.34	0.26	0.26	0.31	0.28	-1.28	-0.75	-0.95	-0.72	-0.92

Low altitude (n = 2023)	2.81	2.57	2.66	2.52	2.61	0.24	0.18	0.19	0.21	0.22	-0.69	-0.21	-0.34	-0.24	-0.35
High roughness (n = 398)	3.01	2.68	2.83	2.64	2.77	0.38	0.22	0.24	0.23	0.24	-1.08	-0.48	-0.84	-0.41	-0.76
Low roughness (n =1971)	2.87	2.63	2.71	2.59	2.65	0.23	0.19	0.20	0.23	0.22	-0.72	-0.25	-0.35	-0.29	-0.36

Table A4. Results from bootstrapping of climatic records that routinely fall into the worse performing 5% in terms of model-data coherence, representing the most divergent time series of the dataset. O = observed proportion, P = mean of bootstrapped proportion, U = upper 95% confidence interval. Statistically significant ($p < 0.05$) results are highlighted in bold, indicating where higher proportions are observed than expected by chance.

	Present			MIS 1			MIS 2			Asia			East North America			West North America			Europe			High altitude			Low altitude			High roughness			Low roughness		
	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U	O	P	U
Mean annual temperature (N = 44)	**0.75**	**0.44**	**0.46**	0.37	0.99	0.99	**0.61**	**0.20**	**0.21**	**0.77**	**0.19**	**0.20**	0.11	0.26	0.27	0.07	0.14	0.15	0.06	0.41	0.43	**0.32**	**0.14**	**0.16**	0.68	0.44	0.86	0.09	0.17	0.18	**0.91**	**0.82**	**0.86**
Mean annual precipitation (N = 21)	**0.81**	**0.44**	**0.45**	0.76	0.98	0.98	**0.24**	**0.20**	**0.21**	**0.24**	**0.19**	**0.20**	0.14	0.25	0.27	**0.29**	**0.14**	**0.15**	0.33	0.41	0.42	**0.19**	**0.14**	**0.15**	0.81	0.44	0.85	**0.24**	**0.16**	**0.18**	0.76	0.81	0.83
Mean July	0.13	0.44	0.46	0.6	0.99	0.99	**0.37**	**0.20**	**0.21**	**0.77**	**0.19**	**0.20**	0.10	0.26	0.27	0.07	0.14	0.15	0.07	0.41	0.43	**0.37**	**0.14**	**0.16**	0.63	0.44	0.86	0.10	0.17	0.18	**0.90**	**0.82**	**0.86**

tempe
rature
(N =
30)

Formatted: Left, Line spacing: single

Page 2: [1] Deleted **Timbrell, Lucy [timbrell]** **10/10/2024 15:44:00**

Page 2: [2] Deleted **Timbrell, Lucy [timbrell]** **10/10/2024 15:45:00**

Page 5: [3] Deleted **Timbrell, Lucy [timbrell]** **08/03/2025 18:08:00**

Page 5: [3] Deleted **Timbrell, Lucy [timbrell]** **08/03/2025 18:08:00**

Page 5: [4] Formatted **Timbrell, Lucy [timbrell]** **08/03/2025 18:07:00**

Font: (Default) Times New Roman, 10 pt

Page 5: [4] Formatted **Timbrell, Lucy [timbrell]** **08/03/2025 18:07:00**

Font: (Default) Times New Roman, 10 pt

Page 5: [5] Formatted **Timbrell, Lucy [timbrell]** **08/03/2025 18:07:00**

Font: (Default) Times New Roman, 10 pt

Page 5: [5] Formatted **Timbrell, Lucy [timbrell]** **08/03/2025 18:07:00**

Font: (Default) Times New Roman, 10 pt

Page 5: [6] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [6] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted **Lucy Timbrell** **25/02/2025 08:55:00**

Page 5: [7] Deleted	Lucy Timbrell	25/02/2025 08:55:00
Page 5: [8] Deleted	Andrea Manica	21/02/2025 20:55:00
Page 5: [9] Deleted	Andrea Manica	21/02/2025 20:55:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [10] Deleted	Timbrell, Lucy [timbrell]	08/03/2025 18:09:00
Page 5: [11] Deleted	Lucy Timbrell	24/02/2025 11:29:00
Page 5: [11] Deleted	Lucy Timbrell	24/02/2025 11:29:00
Page 5: [11] Deleted	Lucy Timbrell	24/02/2025 11:29:00
Page 5: [11] Deleted	Lucy Timbrell	24/02/2025 11:29:00
Page 5: [11] Deleted	Lucy Timbrell	24/02/2025 11:29:00
Page 12: [12] Deleted	Timbrell, Lucy [timbrell]	14/10/2024 10:14:00
Page 12: [12] Deleted	Timbrell, Lucy [timbrell]	14/10/2024 10:14:00

Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [12] Deleted	**Timbrell, Lucy [timbrell]**	**14/10/2024 10:14:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**

Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [13] Deleted	**Timbrell, Lucy [timbrell]**	**22/01/2025 10:29:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**

Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**

Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 12: [14] Deleted	**Timbrell, Lucy [timbrell]**	**08/03/2025 18:15:00**
Page 23: [15] Deleted	**Timbrell, Lucy [timbrell]**	**24/01/2025 16:48:00**
Page 23: [16] Deleted	**Timbrell, Lucy [timbrell]**	**24/01/2025 17:10:00**
Page 23: [17] Deleted	**Timbrell, Lucy [timbrell]**	**24/01/2025 17:16:00**