

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

Reviewer 1:

RC1: This paper looks into the comparison between climate models and proxies and to what extent the differences between them could be reduced. The authors use statistical methods to increase the resolution of the model data to make it more comparable to proxy data, which represent local conditions. The conclusion is that even though the downscaled model data has more details the comparison with proxies is not really improved.

Considering the assumptions made and the methods used in the paper I wonder why anyone should expect an improvement of the model data. I suppose the paper can be a valuable contribution if these methods are commonly used in their part of the field. In any case, I think the authors should make it clear that their results apply to one particular type of statistical downscaling. It's not possible to draw any general conclusions about downscaling from these findings. Especially since the authors completely fails to mention dynamical downscaling.

Dynamical downscaling is known to improve the description of processes in the climate system and improve the description of local climate (e.g. Rummukainen, 2016). Dynamical downscaling is not very common within the field of palaeoclimate, but there are studies, e.g. Strandberg et al., 2011; Russo and Cubash, 2016; Velasquez et al., 2021; Strandberg et al., 2022; Strandberg et al., 2023.

Statistical downscaling is also known to improve local climate data and successfully minimize biases in climate models (e.g. Francois et al., 2020, Berg et al., 2022) Bias adjustment methods (also more advanced methods like quantile mapping) build on the assumption that the relationship between model and observations is constant. This works for the present and future (coming 100 years or so) climate because climate change is not that large. For palaeoclimates, however, you cannot expect this relationship to hold. You can't expect the model biases to be the same in the present climate as in the LGM or in the early Eemian. In a climate different from today, and with different topography the weather regimes are not the same as today – and therefore you can't expect the model biases to be the same as today. If you in addition to these faulty assumptions use a very simplified method that only gives an offset of the model data, then I wonder why you at all expect your method to improve anything. Figure 2 clearly shows that your methods only slightly shifts model data. But you would like your method to also correct trends and variability.

AC1: We would like to thank the reviewer for drawing attention to our lack of discussion around dynamical downscaling, and for providing useful references. As the reviewer themselves suggests, dynamical downscaling is not very common within the field of palaeoclimate and associated fields (i.e. archaeology, palaeoecology etc.) who consume model outputs. This is because

this methodology is not accessible to researchers working with a large number of time steps due to the computational costs and time involved, particularly when exploring climatic variability over extended temporal or geographic spans. Yet tackling questions in archaeology, palaeoecology etc. often require finer levels of spatial resolution than typically provided by publicly available climatic model time series. We do not provide an overly extended discussion of these other methods of downscaling (i.e. dynamical downscaling), given they are not very relevant to our field, but add specific reference to them:

“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.”

And reiterate this point in the discussion:

“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.”

We have also stressed the importance of testing the delta method as one of the most accessible methods of downscaling for consumers of palaeoclimatic model outputs:

“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 classifications and estimations of ecologies (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.”

RC2: My point here is that the conclusions drawn in the paper are far too general. Statements like: “our results imply that downscaling to a very fine scale has minimal to no effect on the coherence of model data with pollen records.” (l 28-29) are simply wrong. Your conclusions only apply to the methods used in this study, not all varieties of downscaling and bias adjustment.

AC2: We thank the reviewer for pointing out that some of our statements are too generalised. We have corrected this throughout the manuscript, for example:

“Our results highlight that further downscaling models via statistical methods 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.”

And added further specification that we are testing the delta method *specifically*, including in the title:

“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”

RC3: I think that the authors could be a bit more critical towards proxies. It's a bit much to call it “golden standard”, and this comes from a modeller who is used to see all problems in models, and less so in proxies. Remember that proxies also have uncertainties. For example, Strandberg et al. (2011) come to the conclusion that the comparison between climate model and proxy data is mostly limited by the large errors bars in proxy data.

AC3: We have added further critique of proxies using the reference suggested by the reviewer, although we retain our stance that proxies are typically considered to be ‘gold standard’ by archaeologists, palaeontologists etc. when looking at climatic conditions at specific locations in the past:

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

“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.”

RC4: I would also like you to think about the distribution between figures in the paper and the supplementary material. The paper doesn’t include so many figures, and some of them are, to be honest, not that informative.

AC4: We have reworked all of the figures based on your specific suggestions (see below). Thank you.

RC5: At the same time the paper is quite heavy on reference to the supplementary. Perhaps you would like to lift something from the supplementary to the main text? And while you’re at it rework some of the existing figures.

AC5: Thank you for this suggestion. We also apologise for the missing tables in the SOM. It was requested upon submission that that four tables from the manuscript be moved from the main text into the SOM due to CoP formatting issues. A new version of the SOM was submitted, including these 4 tables, but is unfortunate that this version was not shared with the reviewers nor uploaded online. We have however moved these large tables to an Appendix (Appendix A) so they are more easily accessible within the manuscript itself.

RC6: In conclusion, this paper has a very shallow description and discussion of downscaling and bias adjustment methods. This should be expanded. The conclusions should be reformulated to only apply to the methods used in the study, instead of all methods. If this is done, I think that the paper could be accepted (assuming that the methods are actually used in other projects). Otherwise I will recommend rejection.

AC6: Thank you for this summary. We believe we have sufficiently addressed all of your comments (see below) and would like to stress that accessible methods (i.e. that can be easily applied within a workflow, require manageable processing and accessible computational power) to downscale a large number of reconstructions are indeed very sought after in our field, who tend to be consumers of climatic model output as opposed to modellers.

Comments

RC7: L56-57 It could also be worth to mention that climate models also offer a picture that is also consistent across variables, thus giving a more complete picture of the climate.

AC7: We have amended the following:

“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.”

RC8: L60 what do you mean by “observational data” here? Do you mean proxies? In that case, say so. Proxies and observations are different things. If you mean observations, explain why it is relevant to mention here. The rest of the paragraph is about proxies.

AC8: We have changed this to say ‘proxy data’ for clarity.

RC9: L63 “errors” Perhaps it’s better to talk about “differences” since proxies also have errors.

AC9: Thank you for this suggestion; we have changed this to differences.

RC10: L71 “Different methods” -> “Different statistical methods”. Otherwise you should also mention dynamical downscaling.

AC10: We have edited the manuscript accordingly and added more discussion about dynamical downscaling, as suggested:

“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.”

RC11: Section 2.1 Here, I would like you to explain a bit more. It’s difficult to follow what is done and in which order. Consider a more linear description, like GCM run, bias adjustment, downscaling etc. For example I don’t understand what the Beyer et al simulation is. Is it a GCM run, a modification of the HadCM3 run or something else? Please also give some details about the HadCM3 run, for example regarding resolution and time span.

AC11: We now provide a detailed description of the output from Beyer et al. (2020), and the original HadCM3 model output (Huntley et al. 2022) we have subsequently added upon request from Reviewer 2:

“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 GCM, at their native resolution of 3.275 x 2.5 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 800k years). The second series comes from Beyer et al. (2020a) within the pastclim R package (Leonardi et al. 2023). These reconstructions are 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 approach 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 CO₂ (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).

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 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.”

RC12: L123-124 Is this the same simulation as in lines 112-113.

AC12: Yes, here we were referring to the Beyer et al. (2020a) output. We have adjusted the method sections to improve the clarity of our workflow (see above).

RC13: Eq. 1 Please explain what “DM”, “sim”, “raw” and “obs” denotes.

AC13: We have amended this section as follows:

“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

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

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 (see Beyer et al. 2020a). Like temperature, downscaled precipitation is estimated as

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

RC14: L161 Why do you use “bio01” here and “Tann” elsewhere? Use a consistent terminology. I would prefer abbreviations like Tann instead of bio01, because they are easier to understand.

AC14: We use ‘bio01’ and ‘Tann’ etc. as this is how mean annual temperature are abbreviated in the climatic model and proxy dataset respectively. We retain bio01, bio12 and bio10 when describing the model output in the Methods and in Figures of the modelled climatic layers, however we use the full variable names (e.g. mean annual temperature) throughout the manuscript when discussing our results to ensure consistency.

We have added an additional sentence explaining that these terms are equivalent variables:

“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. “

Moreover, in Table 1, we have provided an account of the equivalent climatic variables extracted, though we have added an explanation of their abbreviations into the Table caption for clarity:

“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}).”

RC15: L211-213 If this sentence is the only thing you write about Fig 2, why show it at all? I think it would be worth to describe also the differences between WAPLS and MAT.

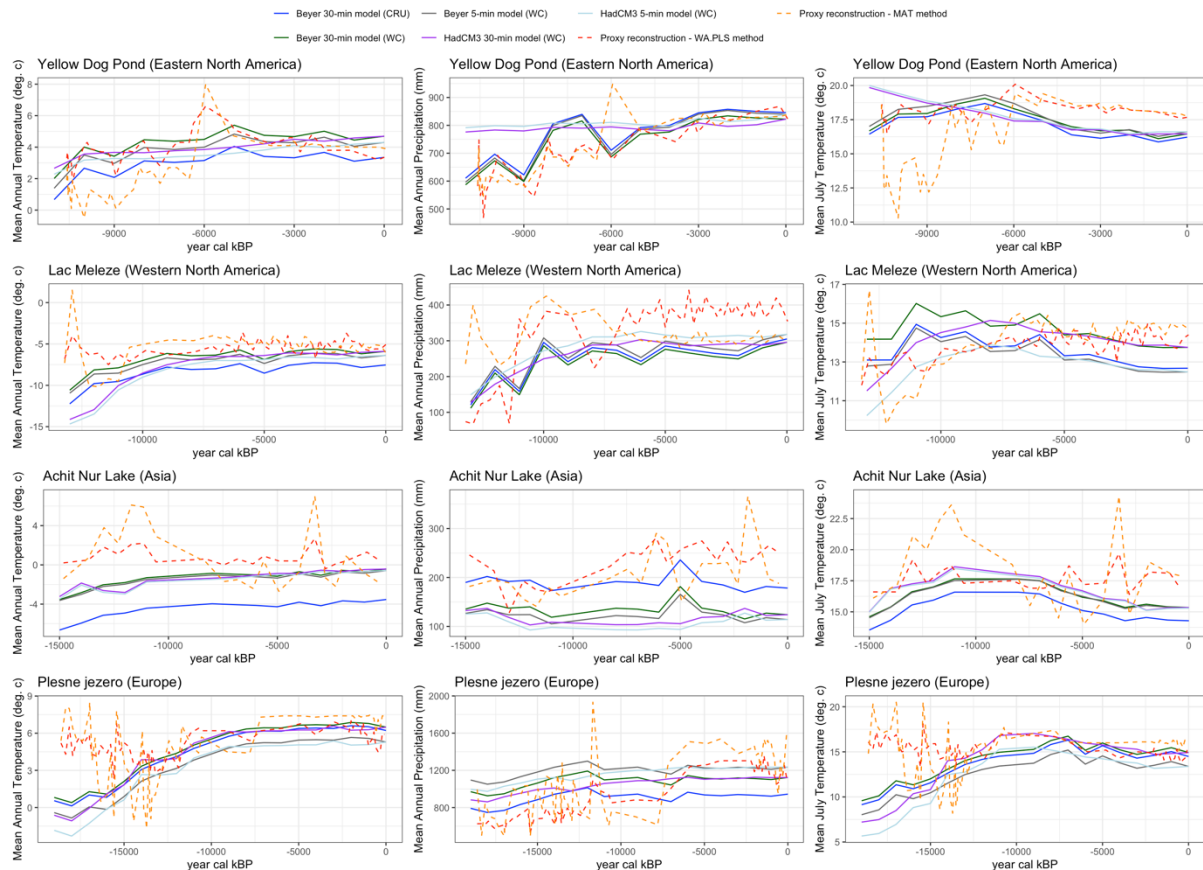
AC15: We show Figure 2 as it visually captures the comparisons between time series that we are quantifying in this paper. We have added an additional sentence:

“Figure 2 highlights a sample of non-interpolated time series from proxy sites across the geographic span of the LegacyClim1.0 dataset, highlighting 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).”

We do not think it is relevant to this paper to extensively describe the differences between the WA-PLS and MAT methods. These two state-of-the-art analytical methods have been commonly used in the field for over 3 decades, and there is ample documentation on how they work and how they perform in different situations. We feel that entering into technicalities would not add anything significant to the paper. However, and to guide interested readers, we have added three important references that correspond to extensive reviews of the field of pollen-based climate reconstructions that clearly highlight that the relative strengths and weaknesses of each of the methods (Sweeney *et al.*, 2018; Birks *et al.* 2010; Chevalier *et al.*, 2020). If the reviewer is referring here to the differences in *results* between WA-PLS and MAT, these are reported throughout Section 3, with limited variations between methods.

RC16: Fig 2 It's difficult to see the difference between the lines representing models. Consider using colours that are more different from each other, and to use dashes and dots to separate them even more.

AC16: We have made these suggested amendments by changing to a divergent colour scheme and using line representations to differentiate proxy from model time series in Figure 2.



RC17: Fig 2 How large are the areas shown here? How is the comparison between model and proxies made? Is it one model grid point vs. One proxy data point? If you average model data over a larger area some of the point of downscaling will disappear.

AC17: We have added further description of the comparison in the methods section:

“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.”

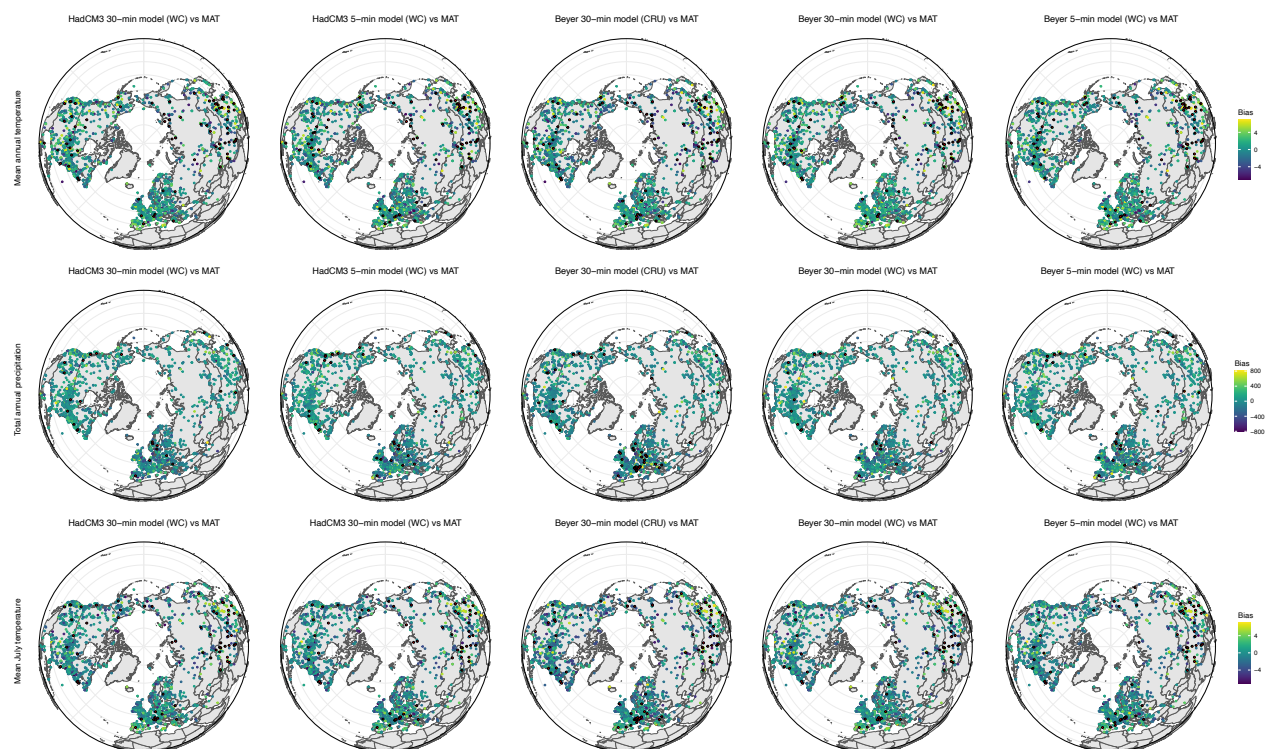
Figure 2 shows the climatic time series produced by the proxy reconstruction and the model output at the coordinates of the proxy sites.

RC18: Fig 3 Add units to the panels. Add temperature, precipitation etc to the leftmost panel in every row.

AC18: We have made these suggested amendments to Figure 3 (see AC19).

RC19: Fig 3 This could be presented much better. The panels are small, the data only covers a part of the panels, the colours are difficult to distinguish. I cannot draw any conclusions from looking at Fig 3. Think about alternative ways to show this. Perhaps you could collect the point in regions and do boxplots show the differences per region. That would give you a quantitative comparison.

AC19: Thank you for this suggestion. We have made edits to Figure 3 to improve the readability of this figure (namely cropped the map so the data fill the frame, and highlighted outliers in red). Boxplots are however a good suggestion, and we have added these for the regional subgroups and landscape subgroups to the SOM as alternative ways of displaying the results presented in the tables in Appendix 2 and Figure 3.



RC20: L297 Is “predict” the right word here? The proxy data do not predict temperatures.

AC20: We have changed this to ‘indicate’.

RC21: Fig 4 It’s obvious that Fig 4 shows the effect of the resolution. I’m, however, not sure that it shows the “effects of landscape dynamics”. What do you mean by that. Furthermore, I think you could make your point by showing just one region in one line. This is a lot of figure space for little information.

AC21: This figure demonstrates how increasing the resolution of the model better captures more fine-scale detail of the landscape, such as coastlines and topographic differences. We believe that this figure effectively highlights the impact that downscaling can have in different types of landscapes (i.e. in the

Pittsburg Basin where it is very flat and inland, there is little change, whereas in South Italy there is much more detail captured in localised climate at coastlines and areas of diverse topography). We have added further detail to this effect:

“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.”

RC22: Fig 4 What do the dots represent?

AC22: We have added to the caption of Figure 4:

“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.”

RC23: L322 Is it correct to refer to Fig 4 here?

AC23: Thank you for pointing this out – we were referring to Figure 5 here. This has been amended.

RC24: L334 “Models are also inherently calibrated ...” This is a very general statement that doesn’t apply to all climate models. Please specify which models you refer to.

AC24: We have specified that here we are referring to delta-downscaled models:

“Delta-downscaled models are also inherently tuned 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).”

RC25: L364 I don't think this is a question well posed. How do you know that the downscaling is the problem, and not the methods you used to do the downscaling. Again, this is a very general statement that doesn't apply to all downscaling techniques.

AC25: We have edited the phrasing of this section:

“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.

We have also specified that we are referring to the methods that we tested in the paper:

“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 proxy records.”

RC26: L364-369 I think this is a testament of the poor methods you use.

AC26: It may be the case that other methods, such as dynamical downscaling, would produce better results, however unfortunately, these are not accessible methods to many researchers who use climatic models. We have stressed this in the discussion:

“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.”

RC27: L376 You have not mentioned that Beyer et al is a climate emulator. Please add this to section 2.1.

AC27: This was a mistake and has been removed.

RC28: L401-403 This is simply wrong. You only show that the downscaling method used in this paper fails. Based on that you should not dismiss all different ways to do

downscaling. It would be unfortunate if the community thought that all downscaling is pointless.

AC28: We have amended the conclusion to specify that we are referring to the method we have tested in the paper:

“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.”

As highlighted in AC26, we have added further discussion of other methods that may be better equipped than the ones tested in this paper, albeit more inaccessible.

Minor comments

RC29: L49 missing “(“ somewhere before this “)”

AC29: We have removed this error.

References

- Berg, P., Bosshard, T., Yang, W., and Zimmermann, K.: MIdASv0.2.1 – Multi-scale bias AdjuStment, *Geosci. Model Dev.*, 15, 6165–6180, <https://doi.org/10.5194/gmd-15-6165-2022>, 2022.
- François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, *Earth Syst. Dynam.*, 11, 537–562, <https://doi.org/10.5194/esd-11-537-2020>, 2020.
- Rummukainen, M., 2016. Added value in regional climate modeling. *Wire Clim. Change* 7, 145e159. <https://doi.org/10.1002/wcc.378>, 2016.
- Russo, E., Cubasch, U., 2016. Mid-to-late Holocene temperature evolution and atmospheric dynamics over Europe in regional model simulations. *Clim. Past* 12, 1645e1662. <https://doi.org/10.5194/cp-12-1645-2016>.
- Strandberg, G., Brandefelt, J., Kjellström, E. and Smith, B. 2011: High-resolution regional simulation of last glacial maximum climate over Europe. *Tellus* 63A, 107-125. DOI: 10.1111/j.1600-0870.2010.00485.x
- 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, <https://doi.org/10.1016/j.quascirev.2022.107431>.
- 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.

Velasquez, P., Kaplan, J.O., Messmer, M., Ludwig, P., Raible, C.C., 2021. The role of land cover in the climate of glacial Europe. *Clim. Past* 17, 1161e1180.
<https://doi.org/10.5194/cp-17-1161-2021>.