

DOI: <https://doi.org/10.5194/egusphere-2026-1590>

Title: Advancing Last Glacial Maximum paleoclimate reconstructions in Europe using pollen data: a multi-method (mega)biomization approach

We sincerely thank the reviewers for the time, effort, and thoughtful consideration they devoted to reviewing our manuscript, as well as for their insightful comments and constructive suggestions, which greatly helped improve the quality of this work.

Reviewer comments are shown in *italic*. Our original responses are shown in **bold** and the updated descriptions of how we propose to modify the manuscript are indicated in **bold blue text**. Quotations from the original manuscript are shown in **bold italics** and quotations from the proposed manuscript are shown in **bold blue italics**.

Reviewer #1: RC1: 'Comment on egusphere-2026-1590', Anonymous Referee #1, 29 Apr 2026

This ms reconstructs LGM climate from pollen using a new method that begins with a biomisation step, followed by reconstructions based on modern samples from that each biome which are weighted according to the biome score to give the final reconstruction. I can understand the motivation for the proposed method. Reconstructions based on continental scale calibration sets can give poor reconstructions, perhaps because of large secondary climate gradients, perhaps because taxa with wide climatic tolerances create bad analogues.

One option would be to restrict the transfer function to base reconstructions only on calibration set samples from the same biome as the fossil sample. However uncertainty in assigning the fossil sample to a biome would lead to biases and instability in reconstructions from adjacent samples assigned to different biomes because of small differences in pollen assemblage composition.

The proposed method is analogous to model averaging with AIC weights. However, the AIC weights are the probability that each model is the best of a set of candidate models. Biome scores (even after transformation to have a unit sum) are not the probability of each biome occurring, so the weighted sum of reconstructions based on each biome may be a biased estimate of the climatic conditions. (If the authors believe biome affinity scores can be transformed into probabilities, they need to fully justify this). It might be worth trying the biomisation method devised by Cruz-Silva et al (2022; <https://doi.org/10.1111/jbi.14448>) who report that their method gives approximate probabilities of a pollen sample belonging to each biome.

In the section dedicated to method evaluation and cross-validation, we have substantially clarified the rationale underlying our multi-method approach, in which climatic reconstructions are weighted according to the affinity scores of the different (mega)biomes. We now explain more explicitly why this approach was preferred over a single global calibration, particularly in terms of methodological robustness and reconstruction performance. The cross-validation results demonstrate that calibrations performed within biome-specific subsets generally provide more reliable climatic estimates than a global calibration approach, especially when dealing with heterogeneous pollen assemblages and contrasting ecological contexts.

The revised manuscript would also place greater emphasis on the conceptual interest of retaining and integrating the full range of (mega)biome affinity scores rather than assigning each pollen assemblage to a single dominant biome. We argue that this strategy allows mixed or transitional

vegetation signals to be better represented and therefore preserves more ecological information contained in the pollen data. In this context, the weighted integration of all biome affinities provides a more flexible framework for capturing complex climatic signals than a strict categorical biome assignment.

Finally, following the reviewer's suggestion, we further detail the methodological links with Cruz-Silva et al. (2022), cited in the original manuscript, who highlight the sensitivity of reconstructed climates to small variations in biome scores. While our weighting procedure does not formally treat these scores as probabilities in the sense of AIC-based model averaging, it similarly relies on a continuous use of biome affinities rather than strict assignments. The validation results support the robustness and usefulness of our approach for climatic reconstructions.

"Then, our weighting procedure assumes that biome affinity scores approximate probabilities in the statistical sense. This makes our approach conceptually closer to the probabilistic framework of Cruz-Silva et al. (2022), who combine biomisation with a dissimilarity-based method to estimate biome membership likelihoods."

The methods, should be reorganised to start with the modern and then fossil pollen data, then the biomisation, followed by the transfer function methods. In the results, the cross-validation performance should be shown before the LGM reconstructions.

We agree that reorganizing the Methods section in this order improves the logical flow of the manuscript and makes the methodological framework easier to follow.

For a revised version of the manuscript, we propose a substantial restructuring of the text: i. data inputs from modern and fossil pollen dataset (Section II), ii. the calibration procedures (Section), and iii. finally the reconstruction methods.

Cross-validation

The cross-validation scheme should mimic the analysis used to reconstruct climate from the fossil data. As far as I can tell from the description and partial code (see below), this has not been done. The past climate estimated as the weighted mean of the reconstructions based on each biome, whereas the cross-validation is based only on one biome. This likely biases the model performance. Once consequence of this is that the uncertainties reported for the reconstructions are likely to be incorrect.

Actually, the current cross-validation does not reproduce the actual fossil reconstruction pipeline, and therefore may lead to biased performance estimates and potentially incorrect uncertainties. This cross-validation does not account at all for analyses performed using fossil data. It only evaluates the ability of MAT to reconstruct modern climate variables from modern pollen spectra.

This code is used to assess the ability of the MAT to correctly reconstruct a modern climatic variable from modern pollen assemblages, using an 8-fold cross-validation scheme.

The idea is as follows:

1. We have modern pollen assemblages (taxa) and observed modern climate data (modern_clim).
2. A subset of the modern data is removed.
3. Method is calibrated on the remaining data.
4. The climate of the removed samples is predicted.
5. The predictions are compared to the observed climate values.
6. This process is repeated until every sample has been used once as a test set.

The current cross-validation can be summarized as:

Modern pollen → one biome → reconstruction methods

To properly evaluate climatic reconstructions using the same datasets as those used for calibration, fossil data are not included as calibration data. Instead, the excluded modern samples are treated as pseudo-fossils.

The corrected CV framework should be:

Modern pollen → Biomisation → multiple biomes → reconstruction per biome → weighted mean reconstruction

The corrected idea is therefore as follows:

1. Sample modern pollen spectra by biome
2. Compute biome affinity scores
3. Apply MAT separately for each biome
4. Combine results using a multi-biome weighted scheme for each prediction
5. Compare the reconstructed values to observed modern climate

This revised methodology, which is consistent with the treatment of fossil pollen data, is formulated according to the following equation:

$$T_{site} = \sum_b weight_{site,b} \times Recon_{b,site}$$

*Where Recon is extracted from multi-method approach
b defines the biome affinity score of each modern pollen sample*

This approach assumes that modern biomes are representative of climatic responses and that biome–climate relationships are transferable through time.

We have revised the cross-validation framework to reproduce the full fossil reconstruction workflow. Each excluded modern sample is first biomised, then reconstructed using biome-specific MAT models, and finally combined through biome-weighted averaging before comparison with observed climate values.

These cross-validation procedures are provided in the Supplementary Material (“L-CV_Codes”). They are largely implemented in a matrix-based framework and structured similarly to the calibration (reconstruction) codes to ensure straightforward reproducibility.

The results of this revised approach are presented in Table 1 and are also reported in the potential revised manuscript prior to the presentation of the climatic reconstruction results. From this table, we draw three main conclusions:

- i. Global calibration approaches (using the full EMPD2 modern calibration dataset) perform less well than local calibration approaches, showing lower R^2 values and higher errors;
- ii. WA-PLS and CREST reconstruction methods exhibit lower or broadly comparable performance relative to MAT;
- iii. Winter temperature reconstructions (MTCO) are systematically less robust and less reliable, with errors 1-2°C higher than for annual and summer temperatures. In contrast, correlation coefficients (R^2) do not differ substantially across climatic variables.

Calibration	Reconstruction method	Climate variables		
		TANN (R^2 ; RMSEP)	MTWA (R^2 ; RMSEP)	MTCO (R^2 ; RMSEP)
Global	MAT	(84%; 2.96)	(87%; 2.87)	(81%; 3.86)
	WA-PLS	(58%; 4.89)	(41%; 4.54)	(50%; 5.98)
	CREST	(52%; 4.23)	(60%; 4.68)	(61%; 5.91)
Biome	MAT	(93%; 2.41)	(88%; 2.13)	(91%; 3.61)
	WA-PLS	(85%; 2.95)	(78%; 3.06)	(84%; 3.74)
	CREST	(87%; 3.11)	(88%; 3.34)	(79%; 4.36)
Megabiome	MAT	(95%; 2.40)	(90%; 2.09)	(94%; 3.51)
	WA-PLS	(89%; 2.79)	(87%; 2.94)	(77%; 4.77)
	CREST	(82%; 3.46)	(80%; 3.01)	(84%; 4.93)

Table #. Model performances for the reconstruction of three climate variables - TANN, MTWA and MTCO - using the Root Mean Squared Error of Prediction (RMSEP) and the squared correlation coefficient between observed and predicted values (R^2). k (optimal number) and $npls$ (optimal number of components) of modern analogues are selected to minimize prediction errors in the MAT and WA-PLS, respectively.

These results highlight the advantages of our methodological approach in terms of the reliability and robustness of climatic reconstructions compared to the more conventional framework without prior (mega)biomisation. They also reveal a consistent limitation across reconstruction methods in their ability to accurately reconstruct winter temperatures.

The validation results support the robustness and usefulness of our approach for climatic reconstructions. The potential revised manuscript includes these results in Section V, prior to the presentation of the climatic reconstructions of this study.

Modern biomes

*I'm surprised to see that the biomisation scheme finds temperate forest on Svalbard. There are no trees on Svalbard; even *Betula nana* is scarce.*

The sample is Lomonosovfonna, this is a glacier sample, not a lake as the data report (the previous sample Vuoskkujávri is a lake not a glacier, and its elevation is wrong - please check if the data have further issues). Little of the pollen at Lomonosovfonna is from local sources (see the original paper), most has been transported from northern Scandinavia or further. Such extreme samples should probably be excluded to avoid them biasing the results. It should also be excluded because it, like a few other samples, has a very low pollen count (27 grains), and such low count sums cannot be expected to give a reliable reconstruction.

I'm also surprised that most of the Scandinavian + Western Siberian modern megabiomes are marked as temperate forest, with a few tundra locations.

This contradicts Dallmeyer et al 2019 which limits temperate forest to about 60°N in Scandinavia + Siberia with boreal forest covering most of the area. The ms should, at a minimum, explain why it extends temperate forest into the spruce and pine dominated northern forests. It's possible that there is an error somewhere in the analysis, the authors should make sure it does not also affect the LGM reconstructions.

We agree that the original pollen assemblage dataset, although already published and derived from the Eurasian Modern Pollen Database v2 (EMPD2; <https://empd2.github.io>, <https://www.pangaea.de/>), Davis et al., 2020), requires filtering and site selection prior to its use in climatic reconstructions.

In this dataset, the taxonomic and geographical information associated with some sites can be incomplete or of variable quality, with low pollen abundance and limited taxonomic diversity (e.g., EPDcoretop_E2539 from Lomonosovfonna, Svalbard and Jan Mayen).

Several corrections and filtering steps were applied to the EMPD2 calibration dataset during the revision of this manuscript:

1. A total of 281 Moroccan samples are flagged as “X” in the “locationReliabilities” category and are therefore not usable (Davis et al., 2020).
2. A further 202 samples have pollen counts of < 100 grains. Such low counts are generally considered insufficient for robust quantitative reconstructions and may disproportionately reflect stochastic pollen input, noise, or long-distance transport. These samples are associated with WAST and HODE biomes, as well as the TEFO and STEP megabiomes. This minimum threshold is consistent with previous studies (e.g., Salonen et al., 2012; Guiot et al., 2013).
3. Six modern “pollen trap” samples from Poland and Iceland (“Zimny”) can also be excluded; these correspond to COCO/COST biomes and the TEFO megabiome.

This filtering strategy is consistent with the issues of poor analogue quality described in the review by Chevalier et al. (2020). The excluded samples are marked as “Y” in the “Delete” column of the dataset available at the following URL: <https://filesender.renater.fr/?s=download&token=2882fef9-0b8a-49bb-bf40-02607f09705e>

The key step prior to normalisation and comparison of modern and fossil pollen assemblages is the strict exclusion of aquatic taxa and other local environmental proxies. All aquatic pollen types were removed from the calibration dataset in the revised version. For instance, the pollen spectrum from site EPDcoretop_E2539 at Lomonosovfonna (Svalbard and Jan Mayen) is dominated by *Lycopodium* spores, which reflect local environmental conditions. In such cases, terrestrial pollen is strongly underrepresented and the sample cannot be reliably used for robust climatic reconstructions.

We also identified and corrected several methodological aspects that explained discrepancies in the distribution of best modern megabiomes compared to previous studies (Dallmeyer et al., 2019; Li et al., 2019).

Four major issues affecting the implementation of the megabiomization scheme of Li et al. (2024) and Dallmeyer et al. (2019) were identified and corrected:

1. First, the taxonomic harmonisation table used to standardise taxon labels and homogenise pollen assemblages differed between studies. Li et al., 2024 provide harmonized pollen data from LegacyPollen 2.0 (https://pure.mpg.de/view/item_3594279_4).
2. In addition, we did not use the same taxon–PFT tables, although the overall algorithms were similar (<https://zenodo.org/records/7523423>, from Cao and Fang, 2021).
3. A major difference between the original and revised versions arises from the assignment of the dominant megabiome when two or more megabiomes have identical affinity scores. According to the biomization procedure, the dominant (mega)biome should be assigned to the megabiome with the lowest PFT richness in the case of equal affinity scores. However, this rule was not correctly implemented in the initial version due to the ordering of

megabiomes in the secondary PFT–megabiome lookup table provided in the supplementary material of Li et al. (2024).

This implementation error particularly affected the distinction between boreal forest (BOFO) and temperate forest (TEFO), two ecologically similar megabiomes for which equal affinity scores occur relatively frequently. Following the original biomization scheme, BOFO should be preferentially assigned when BOFO and TEFO have identical affinity scores because BOFO contains fewer PFTs. In contrast, the original version of our manuscript incorrectly assigned TEFO as the dominant megabiome under these circumstances. These two megabiomes are often highly competitive and can be difficult to distinguish using pollen assemblages alone.

For example, modern pollen spectra from central and southern Scandinavia are characterized by relatively low taxonomic resolution, with pollen grains frequently identified only to the genus level rather than to species. This reduces the ecological information available for PFT assignment. When identification is restricted to the genus level, species belonging to different forest megabiomes may be assigned to the same key PFTs. In this region, pollen grains of *Betula pendula*, characteristic of temperate forests, and *Betula pubescens*, characteristic of boreal forests (Beck et al., 2016), cannot be reliably distinguished and are therefore both recorded as *Betula*. Consequently, these indicator species cannot be used to discriminate between temperate and boreal forests, which already share several PFTs. Under such conditions, *Larix* often becomes the only potentially abundant taxon capable of differentiating between these two megabiomes.

4. Furthermore, we applied different corrections for pollen taxa affected by long-distance aeolian transport, which are accounted for in the “variable_pollen” table provided in the Supp. Mat “G-Tools_MegabiomeScheme”.

These differences in preprocessing choices partly explain the discrepancies between our modern megabiome affinity patterns and those reported in previous studies.

Few major changes in megabiome distributions have been observed following the modifications listed above:

- A stronger dominance of tundra and boreal vegetation types in western Siberia, at the expense of the relatively closed temperate environments reconstructed in the original version of this study. As a result, the total European fraction of BOFO increases from approximately 9% to more than 20%, thereby reducing the extent of temperate conditions in modern Europe.
- Boreal forests, including taiga ecosystems, are now identified as the dominant megabiome across central and southern Scandinavia instead of temperate forests. This change results from the corrected tie-breaking rule, whereby BOFO contains one fewer PFT than TEFO and is therefore preferentially assigned when both megabiomes have identical affinity scores. Consequently, BOFO now occupies more than 90% of Scandinavia.
- A slight increase (+3%) in temperate forest cover across western Europe, at the expense of the warm-temperate forest megabiome (WTFO), despite the persistence of tundra-like or sparsely vegetated conditions around major European mountain ranges. Compared with the previous version of this study, this correction leads to a modest expansion of temperate vegetation in western Europe.

The LGM climate reconstructions obtained after this series of modifications are, on average, 0.2°C colder in TANN, 0.3°C colder in MTCO, and 2.3°C colder in MTWA than those from the previous version of this study. The results of the megabiomization approach are therefore more consistent with those of the megabiomization approach, and they reveal similar local-to-global differences between MTCO and MTWA anomalies.

The harmonisation tables are provided in the Supplementary Material (“E-HarmonizationTable”), and the associated processing codes are available in “H-(Mega)biomisationProcess”. These

corrections lead to modern affinity patterns that are now consistent with previous studies, reducing the dominance of the temperate megabiome (TEFO) and increasing the representation of tundra vegetation (TUND) in western Siberia and across Scandinavia from ~60°N. In the same way, modern pollen spectra from Russia are now predominantly assigned to the tundra megabiome, extending its dominance into Asia, using these new tables.

All climatic reconstructions (biomes and megabiomes) were therefore revised based both on the cleaned calibration dataset and on the updated pollen processing tables used in the megabiomisation scheme. However, the LGM climate reconstructions presented in the original manuscript are not significantly affected by the modifications applied to the modern training set.

To try to better understand what the authors have done, I looked at the provided R code. The code is not easy to read, having few comments and not being well formatted (the R package `styler` can ameliorate this latter problem instantly), and does not run as it cannot find the files it is trying to import. This is not simply a working directory problem; the files are missing. The problems seems to be that the provided code is only a fragment of the total code used. Some other code has been used to pre-process the provided raw data and split it into a multitude of files. This code should also be provided, as well the code use to make the maps, figures, tables etc. The code should be possible for the reader to run once and get all the results. The current code appears to need running repeatedly with different lines commented out. I would recommend making functions to encapsulate the logic so it can be run repeatedly. The biomisation code used is sufficient, but makes no check that pollen types are assigned to plant functional types. If pollen types are not assigned, the biomisation procedure will be biased.

We thank the reviewer for this detailed and constructive assessment of the code structure and usability.

Without modifying the overall structure of the original scripts, we propose to substantially improve their usability and readability for the revised version of the manuscript. Where possible, we propose converting the scripts to matrix-based computations, structured them as annotated functions, and minimized the number of user-defined variables required. Secondary, we propose adding extensive comments and annotations throughout the code to facilitate understanding and to support reuse with new datasets (e.g., alternative pollen records or applications beyond this study).

We propose also providing all necessary input components to ensure reproducibility of the analyses, including the raw datasets, harmonization tables, calibration datasets, and all model outputs used in this study. These materials allow the full workflow of the climatic reconstructions to be reproduced as presented.

The main effort has therefore focused on improving documentation and clarity of use rather than restructuring the computational pipeline, in order to preserve consistency with the original analytical framework while making it more accessible to users.

In particular, we propose including pollen data, initial sources, and harmonized pollen names in “C-DatasetFoss_Int”, simplified the formatting of the modern training set and reconstruction functions in “H-ReconstructionMethods”, and streamlined the (mega)biomisation process in Appendix H. All relevant details and additional materials are provided in the Supplementary Material.

Citations and bibliography

The citations and bibliography need more attention.

I checked two authors and found errors with both.

We acknowledge that the citations and bibliography require careful revision. We have re-checked all references, including those highlighted by the reviewer, and corrected several inconsistencies and errors in author names, publication details, and formatting. We also revised the bibliography to improve consistency and completeness, and we cross-checked all in-text citations against the reference list to avoid further discrepancies. These corrections have been implemented in the revised version of the manuscript.

The ms cites:

- Birks et al., 2004 - perhaps Birks and Seppä, 2004
- Birks et al., 1005 - obviously wrong year
- Birks and Seppä, 2010 - wrong year
- Ter Braak and Looman, 1986 - not in bibliography (I also don't think this paper, written several years before the development of WAPLS, should be used to describe WAPLS).

Thank you for pointing this out. The reference by Ter Braak and Looman (1986) was cited in the methodological section describing weighted averaging (WA)-based regression approaches and the assumption that taxa data exhibit unimodal (Gaussian) responses along environmental gradients, rather than specifically to describe the development of Weighted Averaging Partial Least Squares (WAPLS). This paper presents the theoretical foundations of weighted averaging and species response models that underpin subsequent WA-derived transfer function methods, including WAPLS. To avoid confusion, we propose clarifying the text and ensured that the original reference is included in the bibliography.

Here, you find this paper cited in the methodological part.

“Ter Braak, C.J.F., Looman, W.N.C.: Weighted Averaging, Logistic Regression and the Gaussian Response Model. *Vegetation*, vol. 65, no. 1, pp. 3–11. JSTOR, 1986.”

The number of errors suggests that the authors are not using citation management software: zotero, jabref, endnote - take your pick, all are better than writing references by hand.

Figures

Figures 1 & 2 have STEP in yellow and DESE in beige.

Whereas figure 3 has STEP in beige, SAVA in yellow and no DESE.

Obviously, there should be a consistent colour scheme and choice of megabiomes.

I find the yellow and beige difficult to distinguish in figure 3 because of the low contrast.

Figure 3 just shows the dominant biomes.

For a revised version of the manuscript, we propose to homogenize the colour codes across all figures and improved their visibility for readers with colour vision deficiencies. In particular, we adjusted the colour shades used for the STEP and SAVA megabiomes in order to improve their distinction while remaining consistent with the biome colour conventions proposed by Prentice et al. (1992).

In addition, Figure 3 does not only display the dominant (mega)biomes (indicated by the outer colour of the pie charts for the LGM samples), but rather the relative proportions of the mean (mega)biome scores under LGM conditions (19–23 cal ka BP). Although the vast majority of sites are dominated by steppe vegetation (mainly cold steppe, COST) under the biomisation procedure, the remaining biome scores may collectively represent more than 50% of the total scores, highlighting the coexistence of several PFT assemblages (Fig 3). At the megabiome level, the scores

are more homogeneous and largely dominated by temperate vegetation (TEFO). By illustrating the competition between the main bioclimatic groups (generally between steppe and desert biomes), this figure supports and motivates our conservative score-integration approach for palaeoclimatic reconstruction.

Finally, the affinity scores for the DESE megabiome are equal to 0 under LGM conditions (Supp. Mat. A), unlike in the modern calibration dataset (Fig 2). Consequently, neither the pie charts nor the legend in Figure 3 include the DESE megabiome.

To clarify this point, we propose adding the following sentence to Section V.1.2:

“While the mean DESE biome score remains null throughout the LGM, the combined scores of arid megabiomes (i.e. grassland and dry shrubland (STEP) + desert (DESE) + tundra and polar desert (TUND)) generally represent a proportion comparable to that of the TEDE megabiome (points with black outlines in Appendix 5).” (line 586-589)

Figure 4 seems to show that few of the samples from Lac du Bouchet (please don't switch from Lac to Lake) have highest biome affinity to temperate forest during the LGM. I'm not sure which of the site on fig 3 is Lac du Bouchet, but many of the sites show a lot of temperate forest, even near the glacial limits. I'm missing the information I need to understand this surprising result, and really don't want to have to run the partial code in its current state. Would it be possible to make, at least for some lakes a plot of biome affinity against time, especially for the LGM.

As shown in Figure 4, during the LGM the dominant cold-steppe (COST) and warm-steppe (WAST) conditions fluctuate at the Lake Bouchet site (Fig R1a), in agreement with the results obtained from the La Grande Pile peat bog record (Fig R1b). The other biome affinity scores remain comparatively low and clearly separated from the dominant steppe conditions. Ecologically, the warm-steppe biome (WAST), represented through the WAGS PFT, includes mainly thermophilous and xerophytic vegetation, generally associated with Mediterranean-type environments and locally with shrub formations. In contrast, the cold-steppe biome (COST) includes the COGS PFT, which is indicative of dry open environments, including periglacial settings. The pollen assemblages shared between these two biomes may therefore reflect mixed environments composed of taxa with broad ecological tolerances. During the Holocene, the biome results shown in Figure R1 instead indicate more competitive bioclimatic constraints, with biome scores clustering around ~20%, despite an overall dominance of the TEDE biome.

Conversely, at the megabiome level, both sites display scores much more clearly dominated by temperate conditions (TEFO) during the Holocene, whereas under LGM conditions the pollen data are distributed among several PFT groups. TUND scores (which include steppe biomes; Appendix 3) nevertheless remain approximately 5-10% higher than the temperate megabiome TEFO during the LGM (Fig R1). Tundra-like conditions thus appear to dominate almost continuously throughout the LGM at both sites, before the maximum expansion of the temperate megabiome (TEFO) during the Holocene from ~10 cal ka BP. **More specifically, the La Grande Pile peat bog record exhibits a stronger contrast between arid megabiomes (STEP and TUND) and arboreal megabiomes (TEFO/BOFO/WAST) during the LGM than the Lake Bouchet record.**

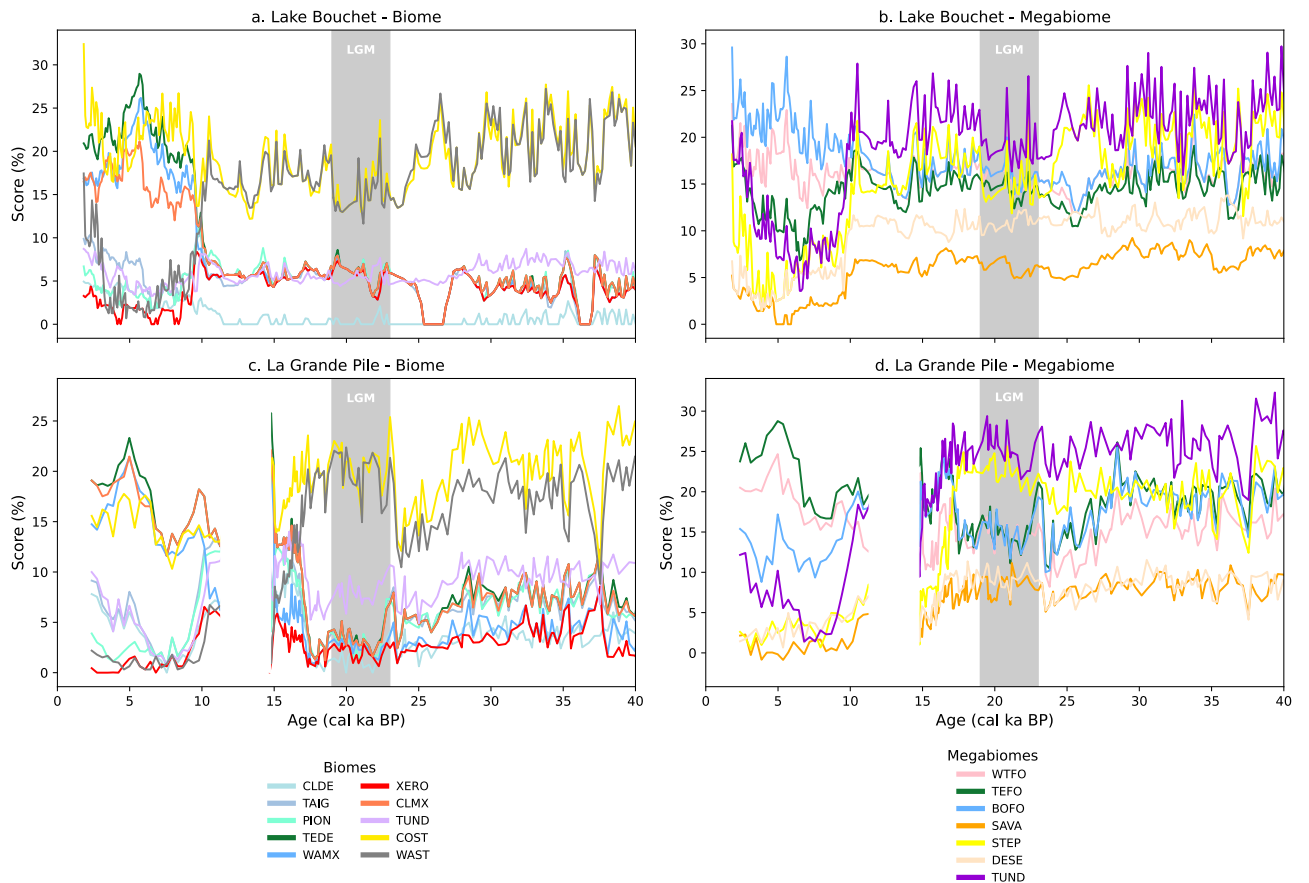


Figure R1. Affinity scores of biomes (a & c) and megabiomes (b & d) through time at the Lake Bouchet and La Grande Pile sites. Scores are normalized relative to one another in order to subsequently apply weighted-average calculations of climatic variables based on these score results. The grey-shaded interval corresponds to the LGM. Periods without scores correspond to a hiatus in the pollen record at the La Grande Pile site.

To facilitate comparison between the equilibrium LGM (mega)biome patterns shown in Figure 3 and the transient score series through time, we added the position of the Lake Bouchet site to Figure 3. Furthermore, the purpose of using this temporal sequence in the present study is not to reconstruct climatic gradients throughout the last deglaciation, but rather to demonstrate the added value of our score-integration approach within pollen-based climate reconstruction methodologies.

Appendix 1C I would recommend a continuous colour scale for this plot. I doubt the current colour scale is particularly accessible for people with colour deficient vision.

Corrected. Done

Appendix 4 Why are some points outlined in black and others not?
I don't see a red contour.

We agree that the original figure caption was not sufficiently clear to adequately describe the multiple types of information presented in this figure. We propose to modify the caption to explicitly identify the points outlined in black, which indicate the sites where the TEFO megabiome dominates under the LGM condition.

Mostly Minor points

I'm sure the two parts of table 2 can be merged.

We agree and have revised Table 2 accordingly. The redundant label row has been removed, and an additional column has been added to clearly distinguish biomes from megabiomes.

241. All the methods used here could be described as statistical.

We recognize that the CREST method reconstructs climatic variables from taxon probability distributions using statistical tools. From a mathematical point of view, any probabilistic method is also a statistical method. The distinction intended here is thus not between “statistical” and “non-statistical” approaches, but rather between “calibration-based statistical and regression methods” and “probabilistic niche-based statistical methods”. We propose the following correction: “*than traditional statistical methods*”.

524 "WGS84 projection" the projection used to plot the data is not especially relevant (and figure 2 is not a WGS84 projection).

We fully agree that a simple WGS84 geographic representation is not suitable for the visualization of these data. Although not incorrect per se, it does not provide an appropriate cartographic framework for the clear and consistent representation of vegetation and climate spatial patterns. All map figures have therefore been updated using a Miller Cylindrical projection, selected for its readability and visual suitability at continental scales.

544 "The lack of data in Eastern Europe results in an underrepresentation of cold modern climate conditions." I'm no longer sure what is meant by Eastern Europe. Its a term with a diverging geographic and political meanings. What I think of as Eastern Europe, isn't particularly cold compared to Finland. Perhaps Western Siberia instead.

We acknowledge that the geographical and geopolitical meaning of “Eastern Europe” is imprecise, non-quantitative, and ambiguous in this context. We therefore propose replacing this term with the more geographically and climatically appropriate designation “Western Siberia”.

560 "biomization also helps reduce the impact of the non-analogue effect" This statement needs supporting evidence.

We agree that the original statement was too strong and insufficiently supported. For the revised version of the manuscript, we have revised the text to clarify that biomization reduces the impact of the non-analogue problem indirectly by aggregating pollen taxa into PFTs and biomes constrained by climatic tolerances, thereby reducing taxonomic noise and dependence on the composition of the modern calibration dataset.

Relevant references have been added. “*Biomization may help mitigate the impact of non-analogue situations by reducing taxonomic complexity and grouping taxa into climatically constrained plant functional types (Prentice et al., 1996; Birks, 1998).*”

603 I have failed to parse this long sentence.

We propose rewriting and split this sentence with: “*By comparing biome and megabiome dominance scores, we show that present-day steppe conditions (COST and WAST biomes) do not systematically correspond to the steppic (STEP) and tundra (TUND) megabiomes. As defined in Appendix 3, these results suggest that the (mega)biome classifications (based on pollen and PFTs) in the two schemes reflect distinct climatic conditions (Fig 1).*”

736 "increase" - from the context, I was expecting "decrease", but see above comments on the cross-validation scheme.

From the k-fold cross-validation results shown in Figure 5, we demonstrate that the RMSE values obtained using the global modern calibration (i.e., the full EMPD2 dataset) are higher than those obtained when grouping taxa into (mega)biomes (local calibration), as evaluated individually for each (mega)biome calibration set (pink bars above the boxplot medians, Fig 5). This result therefore suggests that structuring the modern calibration dataset into (mega)biomes contributes to reducing errors in LGM climate reconstructions. In line with Prentice et al. (1996), biome classification reduces taxonomic complexity and thereby decreases noise in the reconstructions.

Table 3 would be more useful in an appendix. At the moment it is very large, and will be difficult to extract data from it.

Indeed, this table includes a very large number of cross-referenced results and is therefore better suited for the Appendix. It is now presented as Appendix 7 in the revised version of the manuscript.

865 I suspect the seasonal anomalies have little skill. This doubt could be assuaged by showing that this method has skill in a robust cross-validation analysis. It would probably help to explain how a seasonal anomaly should be interpreted - positive anomalies indicate increased seasonality wrt modern?

We note that the predictive performance of the method has already been evaluated using k-fold cross-validation for both MTWA and MTCO, demonstrating its robustness and reconstruction skill. While the reconstruction methods perform relatively well in reproducing summer temperatures during the LGM (mean RMSE of $\sim 3^{\circ}\text{C}$ for MTWA), winter temperatures (MTCO) show larger errors when using MAT and CREST (RMSE $> 4.5^{\circ}\text{C}$). This pattern is consistent across both calibration datasets ((mega)biome).

To better interpret the seasonal anomaly results, we add the following sentence at the beginning of Section V2.4: "Negative LGM–modern seasonal anomalies indicate that seasonality during the LGM was weaker than under present-day conditions, while positive values indicate a stronger seasonal contrast relative to the modern climate." (line 783-786)

952 Any discussion of the apparently good performance of MAT without mentioning that MAT is extremely susceptible to spatial autocorrelation and its performance statistics can be very misleading is flawed.

Even though the MAT method is applicable at broad spatial scales, it shows limited performance in no-analogue situations. In addition, it is highly sensitive to the choice of dissimilarity metric and threshold, as well as to spatial autocorrelation, which may artificially reduce uncertainty estimates. It also provides a relatively poor characterization of uncertainties (Overpeck et al., 1985; Guiot, 1990). We therefore suggest reporting these limitations more explicitly to provide a more comprehensive overview of the method: "The MAT method may therefore be suitable for large-scale spatial studies using global datasets to reconstruct annual temperatures across Europe, although its performance statistics should be interpreted with caution given its sensitivity to spatial autocorrelation and the potential for over-optimistic skill estimates." (line 881-884).

1077 "performs well for pollen taxa with low taxonomic resolution". This statement needs supporting evidence.

We consider that this statement is too strong and insufficiently justified.

Actually, the MAT method may better tolerate a low taxonomic resolution because it relies on the overall structure of assemblages, which are compared to modern samples using dissimilarity measures. Replacing species-level information by broader groups does reduce information content, but the assemblage largely retains its global structure, which is then used to identify modern analogues.

We therefore suggest tempering and refining this statement, and putting it into perspective as follows: « *it is generally less sensitive to reduced taxonomic resolution than approaches relying on individual taxon–climate relationships (WA-PLS) or climatic niche estimates (CREST), because it relies on the overall similarity among assemblages rather than on the climatic response of individual taxa (Chevalier et al., 2020).* » (line 1018-1021).

The authorship contribution statement is inadequate, describing only the writing and editing stages. Please see the CREDIT taxonomy for other roles which can be recorded.

Thank you for this comment. We propose revising it according to the CRediT taxonomy to include all relevant roles (e.g., conceptualization, methodology, data curation, formal analysis, visualization, and writing, including original draft writing and review/editing). **The updated statement now provides a complete description of each author’s contribution beyond writing and editing.**