The impact of methodological decisions on climate reconstructions using WA-PLS

Mark G. Turner[a]* , Dongyang Wei[a], Iain Colin Prentice[b], Sandy P. Harrison[a]

[a]Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB, UK
[b]AXA Chair Programme in Biosphere and Climate Impacts, Department of Life Sciences, Imperial College London, Ascot, SL5 7PY, UK
* Corresponding author e-mail address: m.g.turner@pgr.reading.ac.uk

Abstract

Most techniques for pollen-based quantitative climate reconstruction use modern assemblages as a reference data set. We examine the implication of methodological choices in the selection and treatment of the reference data set for climate reconstructions using Weighted Averaging Partial Least Squares (WA-PLS) regression and records of the last glacial period from Europe. We show that the training data set used is important because it determines the climate space sampled. The range and continuity of sampling along the climate gradient is more important than sampling density. Reconstruction uncertainties are generally reduced when more taxa are included, but combining related taxa that are poorly sampled in the data set to a higher taxonomic level provides more stable reconstructions. Excluding taxa that are climatically insensitive, or systematically over-represented in fossil pollen assemblages because of known biases in pollen production or transport, makes no significant difference to the reconstructions. However, the exclusion of taxa overrepresented because of preservation issues does produce an improvement. These findings are relevant not only for WA-PLS reconstructions but also for similar approaches using modern assemblage reference data. There is no universal solution to these issues, but we propose a number of checks to evaluate the robustness of pollen-based reconstructions.

Keywords: WA-PLS; Quantitative climate reconstruction; Modern pollen assemblages; Late Quaternary

INTRODUCTION

Models of the coupled climate system are used to project how changes in natural and anthropogenic forcing will affect future climates (Collins et al., 2013; Kirtman et al., 2013). The expected changes in twenty-first-century forcing and climate are larger than those experienced during the recent historic period against which these models have been calibrated. Quaternary climate states in which the changes in forcing and climate are as large as those projected for the end of the twenty-first century are therefore now routinely used as an out-of-sample test of such models (Braconnot et al., 2012; Schmidt et al., 2014; Harrison et al., 2015). This evaluation is crucially dependent on the availability of reliable quantitative reconstructions of multiple climate variables (Harrison

et al., 2014). Indications of past climate can be found in many different marine and terrestrial archives, but pollen analysis provides by far the most geographically widespread source of data for terrestrial palaeoclimate reconstructions (Bartlein et al., 2011; Marsicek et al., 2018 and references therein). There are many methods to make quantitative reconstructions of climate variables from pollen data using both statistical relationships and process-based modelling. Most reconstructions of terrestrial palaeoclimate rely on statistical relationships between modern pollen abundances and modern climate. There are two basic approaches (see discussion in ter Braak and Juggins, 1993): analogue methods select the modern climate associated with the modern pollen samples whose abundance patterns are most like a given fossil sample; regression-based methods calculate a transfer coefficient for each pollen taxon and the climate variable, which is then applied to taxon abundances in fossil pollen samples to reconstruct climate through time.

Weighted averaging partial least squares (WA-PLS; ter Braak and Juggins, 1993) is the most widely used of the

regression-based methods. Weighted averaging is a form of regression adapted to the fact that taxon abundances generally show unimodal, rather than monotonic, responses to climate variables. The climate estimate for any given sample is the abundance-weighted mean of the estimated optima of all the taxa present. WA-PLS refines these optima by looking for further information in the residuals between the initial regression and the modern observations and repeats this until the incremental change ceases to be statistically significant. WA-PLS is considered more robust against spatial autocorrelation than modern analogue methods—that is, it is considered to be less affected by the fact that geographically clustered sites may inherently show similar taxon composition (Telford and Birks, 2005). WA-PLS has been widely used for climate reconstructions based on biotic assemblages, including pollen, diatoms, chironomids, and foraminifera (Lotter et al., 1997; Brooks and Birks, 2001; Seppä et al., 2009). In some more recent publications WA-PLS reconstructions have been presented alongside reconstructions using other statistical methods, such as modern analogue methods (Brewer et al., 2008; Peyron et al., 2011; Sinopoli et al., 2019).

A number of methodological choices have to be made in the application of WA-PLS, including the choice of a training data set and which taxa are included in the regression. Studies using WA-PLS generally rely on model performance statistics as a measure of the reliability of reconstructions. They are usually silent about how methodological decisions were reached, and the implications of specific choices for the quality of the reconstructions are rarely made explicit. However, a number of studies of quantitative reconstruction techniques (Birks and Seppä, 2004; Bjune et al., 2010; Telford and Birks, 2011; Juggins, 2013; Salonen et al., 2013; Juggins et al., 2015; Shennan et al., 2015; Jonkers and Kučera, 2018) have examined some of these choices and suggested that they can influence the quality of the reconstructions. None of the previous studies provides a comprehensive analysis of the issues or discriminates between the quality of the statistical model and the impact on the climate reconstructions.

Our goal in this paper is to increase confidence in the quantitative reconstructions of past climates that are used for model evaluation through an analysis of the implications of methodological decisions both on model performance measures under modern conditions and the resulting quantitative reconstructions of glacial climates in southern Europe. We propose ways in which the implications of specific methodological choices can be evaluated in order to document the reliability of the resulting reconstructions.

METHODS

We test the impact of methodological choices by making climate reconstructions for selected fossil pollen records using subsets of a continental-scale modern training data set as inputs to WA-PLS (see "Modern pollen data"). In order to run these tests under climate conditions substantially different from present, we use fossil pollen records from southern Europe covering the last glacial period (see "Fossil pollen"). We examine the effects of methodological choices on WA-PLS model parameters and on the resulting climate reconstructions and their uncertainties (see "Application of WA-PLS").

Modern pollen data

The modern pollen data set (Fig. 1, Supplementary Fig. 1), which we refer to as SMPDS, was constructed by combining records from the European Modern Pollen Database (EMPD) v3.0 (Davis et al., 2013), the Eastern Mediterranean-Black Sea-Caspian Corridor Biomes (EMBSeCBIO, which we abbreviate as EMB) database (Marinova et al., 2018), additional published records (see Supplementary Table 1) from the European Pollen Database (http://www.europeanpollen-database.net/) or taken from Pangaea (https://www.pangaea.de/), and 73 modern surface samples from northern Spain (Wei et al., 2019a). About two-thirds of the sites (4575) were derived from the EMPD v3.0 (Davis et al., 2013), and a further 1088 sites were derived from the EMBSeCBIO database (Marinova et al., 2018). Some of the sites in the EMPD also occur in the EMBSeCBIO database, and these duplicates have been removed. The final SMPDS data set consists of records from 6458 terrestrial sites (Harrison, 2019). We compare the EMPD and EMBSeCBIO subsets of the SMPDS data set to the full data set to examine the impact of the choice of training data set. The majority of the records were available as raw pollen counts; the remainder were percentages. The individual pollen records were taxonomically standardised and cleaned to remove obligate aquatics, insectivorous species, introduced species, and taxa that only occur as cultivars. The final data set expresses the counts as a percentage of the sum of all taxa remaining after this screening. There are 1558 taxa recorded in the SMPDS. Since some of these taxa are only recorded sporadically, we amalgamated them at higher taxonomic levels to produce a data set with 249 taxa (Harrison, 2019). We confined our analyses to taxa with ten or more occurrences in the SMPDS (n = 195).

Modern climate data

Modern climate data were derived from the Climate Research Unit CRU CL 2.0 data set, which provides monthly mean precipitation, monthly mean temperature, and fractional sunshine hours as long-term means for 1961 to 1990 at 10 minute spatial resolution (New et al., 2002). Geographically weighted regression (GWR) (Brunsdon et al., 2002) using latitude, longitude, and elevation as predictors was performed in ArcGIS (ESRI, 2014) to obtain the climate at the location and elevation of each modern pollen site. We then calculated three bioclimatic variables: (a) mean temperature of the coldest month (MTCO), (b) growing degree days above a baseline of 0°C (GDD_0), and (c) moisture index (MI), the ratio of annual precipitation to annual potential evapotranspiration. MTCO was taken directly from GWR output. Daily values
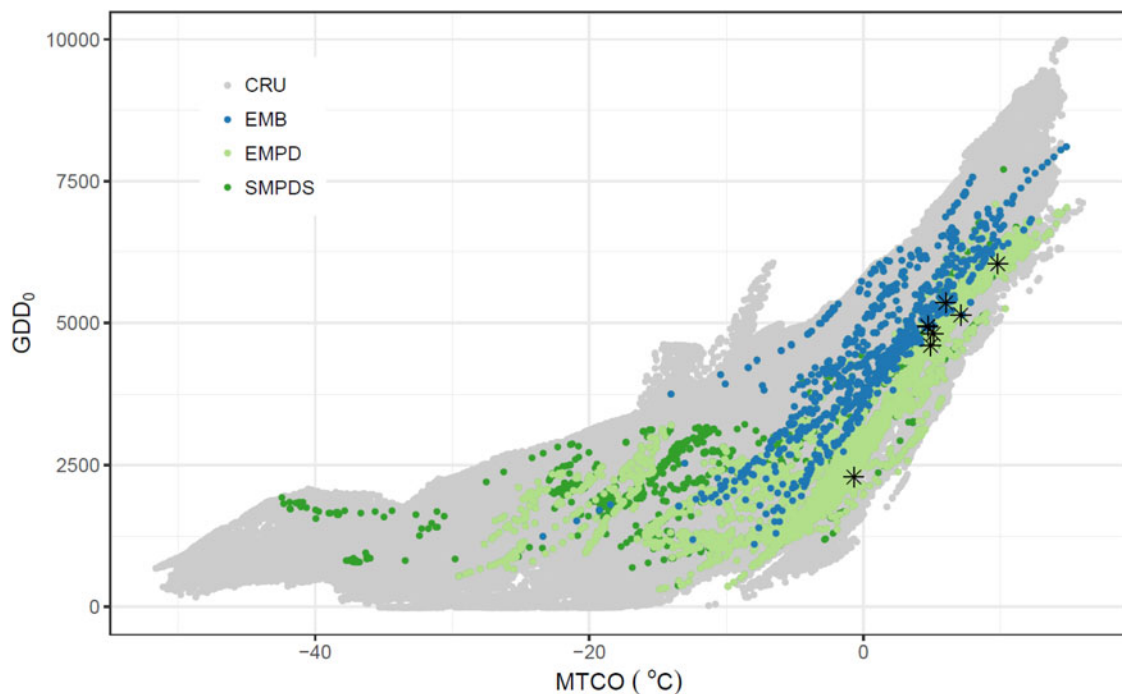
Figure 1. Distribution of modern pollen samples in climate space, represented by growing degree days above 0°C (GDD_0) and mean temperature of the coldest month (MTCO), sampled by Eastern Mediterranean-Black Sea-Caspian Corridor Biomes (EMBSeCBIO) (labelled EMB), the European Modern Pollen Database (EMPD), and the full SMPDS data sets; SMPDS includes EMB and EMPD. The background shows the climate space for a rectangular area (latitude 29° N to 82° N, longitude 21° W to 150° E) enclosing the SMPDS data set, derived from the Climate Research Unit CRU CL 2.0 database (New et al., 2002). Stars indicate the present climate at the eight fossil sites used as examples.

of temperature, sunshine fraction, and precipitation were derived using a mean-conserving interpolation (Rymes and Myers, 2001) of the monthly data. MI was calculated from the daily temperature, precipitation, and sunshine data using modified Python code from the Simple Process-Led Algorithms for Simulating Habitats (SPLASH) model (Davis et al., 2017).

These three bioclimatic variables have been shown to provide a good prediction of vegetation distribution both at global (Wang et al., 2017) and regional (Wang et al., 2013) scales because they reflect important and distinct eco-physiological controls on plant growth (Harrison et al., 2010). MTCO is a surrogate for extreme winter temperatures and influences the survival of woody plants through a wide range of low temperature–tolerance mechanisms, GDD_0 is a combined measure of the length and warmth of the growing season that determines potential annual carbon accumulation, and MI is a measure of plant-available moisture (Harrison et al., 2010). Canonical correspondence analysis (CCA) applied to the SMPDS data set (Wei et al., 2019a) showed a strong correlation between species abundance and these climate variables. The modern climate data at each of the SMPDS sites (Harrison, 2020) are available online (doi: 10.5281/zenodo.3605003). Partial CCA carried out using each climate variable in turn, with the other two as covariates, shows that each variable has a highly significant effect ($p < 0.001$) that is independent of the others (Wei et al., 2019a).

Fossil pollen

The fossil pollen data were taken from the Abrupt climate Changes and Environmental Responses (ACER) database (Sanchez Goñi et al., 2017). The ACER database includes 93 pollen records globally covering part or all of the last glacial period (73–15 ka) with a temporal resolution better than 1000 years. Glacial-age pollen records provide a test of reconstruction techniques under climate conditions substantially different from present and therefore allow rigorous testing of the adequacy of the modern training data set. As examples, we use eight lacustrine records from southern Europe from the ACER database (Supplementary Table 2). The database provides standardised age models for all these cores, but the sampling resolution varies among cores. The Ioannina core (northwestern Greece, 39° 45' N, 20° 51' E) is used here as the primary example because it covers a long period (79.7–10.6 ka) and has relatively high temporal resolution (mean 237 years). However, as discussed below, the conclusions based on this core are supported by analysis of the other fossil data sets.

Climate space analysis

One assumption of WA-PLS is that the taxa used for reconstruction show a unimodal response to the climate variable being reconstructed. In order to test whether this assumption holds true, particularly for amalgamations to higher

taxonomic levels, the climate space occupied by individual pollen taxa was represented and visualised using Generalized Additive Models (GAM) (Guisan et al., 2002) implemented with the Mixed GAM Computation Vehicle (mgcv) R package (Wood, 2017). This approach fits a response surface to the concentration of the pollen abundance in 3D climate space. Abundance is taken as the pollen percentage based on a pollen sum that includes all of the 195 taxa. Convex hulls are used to delineate the area of climate space that contains sampling points and thus avoid representing taxon abundances in climates not closely constrained by the modern pollen data. Convex hulls were fitted using the alphahull and ggplot2 packages in R (Pateiro-Lopez and Rodriguez-Casal, 2016; Wickham, 2016). We used a square root transformation of MI, as differences between MI values at the low end of the MI scale are more important than differences at the high end in their effect on vegetation (Prentice et al., 2017); taking the square root "stretches" the lower values and "compresses" the higher ones. MTCO and GDD_0 were not transformed. We do not include interactions among the climate variables since previous analyses (Wei et al., 2019a) have shown that they each have a significant and independent influence on the distribution of plant taxa and vegetation types. For visualisation purposes, we show 2D slices (MTCO and MI) through the fitted 3D response surfaces at low, medium, and high values of GDD_0. Here we present a number of illustrative GAMs; Wei et al. (2019b, 2020) provide GAMs for all of the 195 pollen and pteridophyte spore taxa from the SMPDS data set.

Application of WA-PLS

The modern bioclimatic and pollen data were used to create pollen-climate transfer functions independently for MTCO, GDD_0, and $\sqrt{MI}$ using WA-PLS (ter Braak and Juggins, 1993). WA-PLS was implemented with the rioja R package (v0.9-15.1; Juggins, 2017). The performance of the calibration models was assessed through leave-one-out cross-validation. The number of components used in each model was estimated through a randomisation *t*-test on the results of this cross-validation (van der Voet, 1994). We selected the significant component with the lowest root mean square error (RMSE), but only if there was a significant improvement in RMSE relative to a lower number of components—since including more components can result in over-fitting of the data so that model predictive value decreases. When making direct comparisons between different subsets of the data and SMPDS, we used the component that was considered significant for that subset (see Supplementary Table 3) rather than the same component across all sets. This choice does not affect the analysis, and for completeness we include an example comparison using the same components for all training sets in the Supplementary Information (Supplementary Figs. 2, 3, and 4).

The (in)stability of transfer coefficients and the statistical uncertainty of reconstructions were assessed by bootstrapping the modern sample set with replacement 1000 times (Efron, 1979) and running WA-PLS each time to derive 1000 instances of both the taxon transfer coefficients and the reconstructed climate for each sample. The standard deviations (SDs) of the taxon transfer coefficients and the SDs of the reconstructions were then calculated from the 1000 bootstrap samples. The variability of a taxon coefficient, as measured by its SD, reflects how consistently the available modern samples including that taxon represent climate. The SD of the reconstruction calculated in this way represents the combined effect of the taxon coefficient variability across all taxa in a fossil sample. This approach differs from the standard method for calculating uncertainties in the rioja R package (Juggins, 2017), which is based only on bootstrapping of the modern data set, because it combines the uncertainties of individual taxa from the modern pollen data set. Data analysis and plotting were performed in R v3.5.1 (R Core Team, 2018).

RESULTS

WA-PLS provides a transfer coefficient for each taxon and each bioclimatic variable. The validity of this transfer coefficient depends on the taxon being sampled across the full range of its realised niche in climate space, and this in turn is determined by the climate space sampled in the modern pollen training data set. Our initial tests therefore focus on the impact of the sampling of climate space on WA-PLS performance metrics and the resulting reconstructions (see "Impact of choice of training data set" below). One of the factors that affects the width of the realised taxon niche is taxonomic resolution, with species in general occupying a more limited niche than genera or families. The use of higher taxonomic groupings increases the number of samples available and can improve the sampling of climate space. We therefore examine the impact of using higher taxonomic groupings on the sampling of climate (see "Impact of amalgamation of pollen taxa to higher taxonomic levels" below). The choice of taxa included in the WA-PLS model will also impact the overall width of the sampled climate space, and we also address the potential impact of including or excluding taxa on model performance and the resulting reconstructions (see "Impact of number of taxa" below).

Impact of choice of training data set

In order to test the impact of the choice of training data set on WA-PLS performance metrics and reconstructions, we compared results from the full SMPDS data set and those obtained using either the EMB or the EMPD subsets. The EMB samples warmer seasonal climates than the EMPD (see Fig. 1, Supplementary Fig. 5). The SMPDS samples much colder winter climates (as measured by MTCO) than either the EMB or EMPD data sets, and colder summers (as measured by GDD_0) than the EMB set. These differences in the sampled climate space are reflected in reconstructions of winter (Fig. 2, see Supplementary Fig. 2) and summer (see Supplementary Fig. 3) temperatures, and also in moisture (as measured by MI) (see Supplementary Fig. 4).
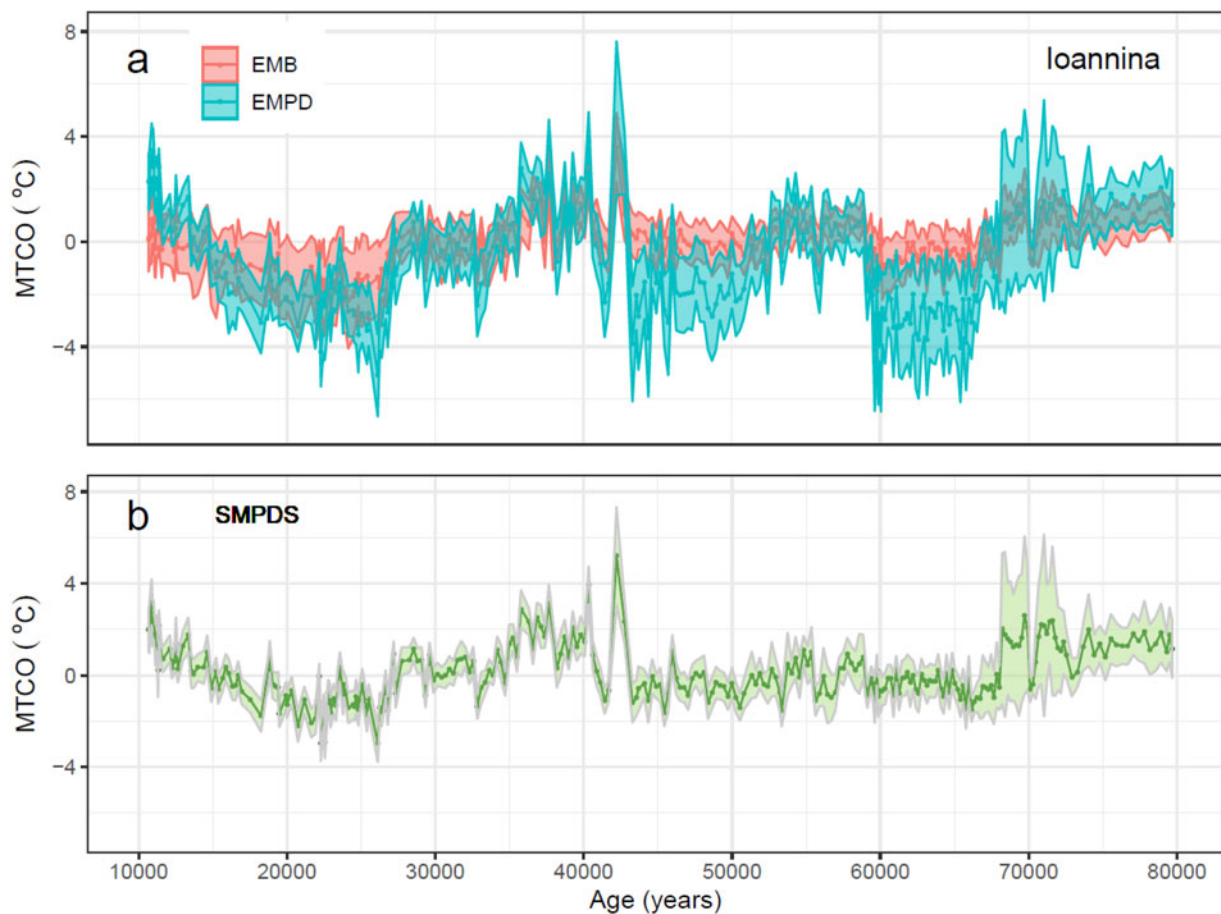
Figure 2. Reconstructions of mean temperature of the coldest month (MTCO) during the last glacial period (80,000 to 10,000 calendar years before 2000) using the pollen record from Lake Ioannina, (a) using the Eastern Mediterranean-Black Sea-Caspian Corridor Biomes (EMB-SeCBIO) (labelled EMB) and the European Modern Pollen Database (EMPD) as training data sets, and (b) using the full SMPDS data set. The reconstruction spread (±2σ) is obtained by resampling the training set 1,000 times.

Reconstructions based on the EMB data set are generally warmer than those based on the EMPD, and there is little overlap between the spread of the reconstruction estimates in the coldest intervals of the glacial period (see Fig. 2). The SMPDS data set produces temperature reconstructions similar to those based on the EMPD but has a considerably narrower reconstruction spread because it samples a wider climate space and contains more samples for each pollen taxon and thus a more complete sampling of the taxon's climatic range. The SDs of the coefficients are related to the number of occurrences (Fig. 3) of the taxon in the sampled data set.

The difference in the reconstructions and their stability is not reflected in differences in the WA-PLS model performance measures (Table 1). Model performance is poorer for EMB-based reconstructions than for SMPDS-based reconstructions, but all of the measures of model performance (r^2, RMSE, maximum bias) are better for EMPD than for the SMPDS data set because they are measures of how well the WA-PLS model replicates the sampled climate space, rather than whether the training data set encompasses a sufficiently large climate space. The standard performance measures are therefore an insufficient guide to the reliability of WA-PLS

reconstructions. Comparison of observed versus predicted values (and their residuals) for the SMPDS data set show that this data set adequately reproduces modern climate (Supplementary Fig. 6).

Both the range and the continuity of the sampled climate are important. This is a consequence of the underlying assumption of WA-PLS that response curves are unimodal and extend across the full realised niche of the taxon; an abundance distribution which is truncated or has gaps cannot be expected to give reliable coefficients. This is borne out by analyses in which the density of sampling is artificially reduced. A proportion of the modern samples was progressively removed while preserving the overall range of climate space (Fig. 4, Supplementary Figs. 7, 8, and 9), having permuted the samples to avoid artificial similarities when samples which are neighbours in the database are also geographically close. Even a 70% reduction in the number of samples does not change the reconstructed winter (see Fig. 4, Supplementary Fig. 10) or summer (Supplementary Figs. 11 and 12) temperature, providing that the range of the sampled climate is maintained (see Supplementary Figs. 7 and 8). It does, however, lead to larger uncertainties because
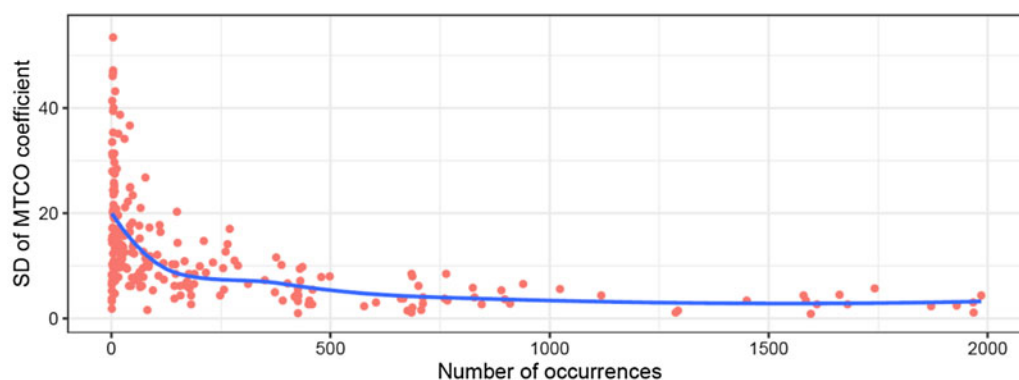
Figure 3. (color online) Relationship between the number of occurrences of a taxon and the standard deviation (SD) of the bootstrapped taxon coefficients. The analysis was made using the full SMPDS data set.

the number of times each taxon is sampled is reduced and the taxon coefficients are therefore less stable (Supplementary Fig. 9). Again, model performance measures do not discriminate between the quality of the reconstructions made with the reduced density or the full SMPDS data set (Table 2).

Tests in which samples from part of the climate range are systematically removed demonstrate the importance of continuous sampling of the climate range (Fig. 5). Removing blocks of the MTCO gradient representing equal numbers of samples shows that discontinuities result in greater uncertainties (as measured by the bootstrapped SDs of the taxon coefficients) than those of the full SMPDS set. With one marginal exception, the uncertainties are also greater than those of a set from which the same number of samples randomly selected has been removed. Width and continuity are both important for obtaining stable taxon coefficients. Thus, it is important to ensure that the modern pollen training data set encompasses a wide range of climate space, to avoid offsets in the climate reconstructions. It is also important that the training data set samples the bioclimatic gradient as continuously as possible, in order to provide the most precise coefficients.

Impact of amalgamation of pollen taxa to higher taxonomic levels

Large modern pollen data sets are created by combining data collected by many palynologists, often for different types of studies, and as a result the level of taxonomic discrimination varies among sites. Many pollen taxa are recorded at a limited number of sites in such data sets, either because they are not regularly identified or because they are genuinely rare. Using taxa that are only recorded rarely, or show geographic clustering suggesting that they have not been sampled across the whole of their potential climate range, leads to unstable WA-PLS coefficients because the bioclimatic range of the taxon is unlikely to have been sampled continuously. The principle of niche conservatism indicates that higher taxa commonly have coherent environmental distributions (Ackerly, 2003; Wake et al., 2009) (Fig. 6), and this principle provides a basis for amalgamating taxa that are only recorded at a few sites in the training data set. The coherency of the climate space occupied by the amalgamated taxon can be assessed using, for example, GAMs. Wei et al. (2019b) have tested the assumption that the taxa used here for reconstructions

Table 1. Weighted Averaging Partial Least Squares (WA-PLS) model parameters for the reconstructions of mean temperature of the coldest month (MTCO, 0°C), growing degree days above a baseline of 0°C (GDD$_0$, °day) and the square root of Moisture Index ($\sqrt{MI}$, unitless) using the Eastern Mediterranean-Black Sea-Caspian Corridor Biomes (EMBSeCBIO) (EMB), the European Modern Pollen Database (EMPD), and full SMPDS as training data sets, including the cross-validated r^2, number of significant components ($p < 0.05$), root mean square error (RMSE), maximum bias, and number of samples. Supplementary Table 2 provides a complete list of WA-PLS model parameters for all data sets.

Variable	Data Set	r^2	No. of components	RMSE	Maximum bias	p	No. of samples
MTCO	EMB	0.417	2	3.97	19.6	0.002	1088
	EMPD	0.743	3	3.91	6.1	0.001	4675
	SMPDS	0.691	4	4.82	10.3	0.019	6458
GDD$_0$	EMB	0.508	2	960	1727	0.001	1088
	EMPD	0.667	3	901	1898	0.001	4675
	SMPDS	0.660	3	910	2039	0.001	6458
$\sqrt{MI}$	EMB	0.592	2	0.128	0.439	0.010	1088
	EMPD	0.602	3	0.190	0.838	0.001	4675
	SMPDS	0.598	3	0.185	0.910	0.001	6458
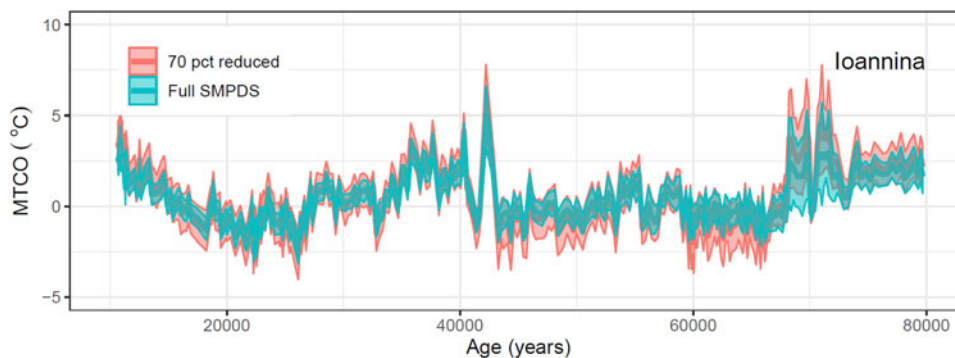
Figure 4. Impact of reducing the sampling density of the modern training data set on reconstructions of mean temperature of the coldest month (MTCO) during the last glacial period (80,000 to 10,000 calendar years before 2000) using the pollen record from Lake Ioannina. The plots show the impact of randomly removing 70% of the modern samples while preserving the overall range of climate space (see also Supplementary Figs. 7, 8, and 9) on the MTCO reconstructions compared to reconstructions made with the full SMPDS data set.

show a unimodal response to individual climate variables. They show that some taxa are unimodal with respect to one climate variable but are insensitive to others; the classic example is *Artemisia*, which appears to be insensitive to temperature but shows a clear optimum with respect to moisture. The assumption of unimodality holds true for most taxa across all climate variables; amalgamation to higher taxonomic levels does not modify this response substantially. As we show here, amalgamating rare taxa into higher taxa improves the bootstrapped SD of the transfer coefficients (Table 3, Supplementary Information Tables 5 and 6) by providing a more continuous sampling of bioclimatic gradients, and as a result leads to reduced uncertainties in the reconstructions.

Impact of number of taxa

Some taxa appear to be relatively uninformative either because they have a wide climatic tolerance or because they are rarely sampled, which raises the issue of whether such taxa should be excluded from the WA-PLS regression. However, analyses of randomly sampled taxa show that increasing the number of taxa monotonically narrows the reconstruction spread (Fig. 7a). In our analyses, the impact of including more

taxa is steep initially and becomes discernibly more gradual above about 80 taxa, after which the mean reconstruction remains relatively stable with minimal offsets (Fig. 7b).

The abundance of pollen does not directly correspond to the abundance of a taxon in the vegetation (Prentice and Parsons, 1983; Prentice, 1985; Sugita, 2007; Hellman et al., 2008). Some taxa are systematically overrepresented in pollen assemblages (e.g., *Pinus*) while others are systematically under-represented (e.g., *Larix*). Furthermore, the comparative ease of pollen transport from closed canopy vegetation such as forest into more open landscapes that produce and disperse less pollen means that arboreal pollen is systematically over-represented at sites in open tundra or steppe-type vegetation (e.g., Edwards et al., 2000; Bigelow et al., 2003; Marinova et al., 2018). However, the inclusion of taxa that are substantially overrepresented in modern assemblages compared to their abundance in the vegetation does not degrade model performance. Reconstructions made with the full SMPDS data set and with a data set that excludes *Pinus* (Supplementary Fig. 14) show little difference either in reconstructed MTCO or in reconstruction spread across multiple sites. Thus, although some studies (e.g., Sinopoli et al., 2019) have excluded such taxa, there is no general a priori reason to do so.

Table 2. Weighted Averaging Partial Least Squares (WA-PLS) model parameters for the reconstructions of mean temperature of the coldest month (MTCO), growing degree days above zero (GDD_0) and the square root of Moisture Index ($\sqrt{MI}$) at Lake Ioannina using the full SMPDS data set and a version of this data set randomly reduced by 70% for the training data sets, including the cross-validated r^2, number of significant components ($p < 0.05$), root mean square error (RMSE), maximum bias, and number of samples. Because the reduced sets are randomly chosen, different runs give different results; the standard deviations across 10 runs are given. The number of components is given as, for example, 2(7) meaning 7 runs allowed 2 significant components to be extracted; the balance allowed 3.

Variable	Data Set	r^2	No. of components	RMSE	Maximum bias	No. of samples
MTCO	Reduced	0.664 ± 0.022	2(6)	5.05 ± 0.07	12.7 ± 3.6	1959
	SMPDS	0.691	4	4.82	10.3	6458
GDD_0	Reduced	0.644 ± 0.010	2(7)	942 ± 17.8	2001 ± 257	1959
	SMPDS	0.660	3	910	2039	6458
$\sqrt{MI}$	Reduced	0.585 ± 0.013	2(8)	0.189 ± 0.004	0.898 ± 0.092	1959
	SMPDS	0.598	3	0.185	0.910	6458
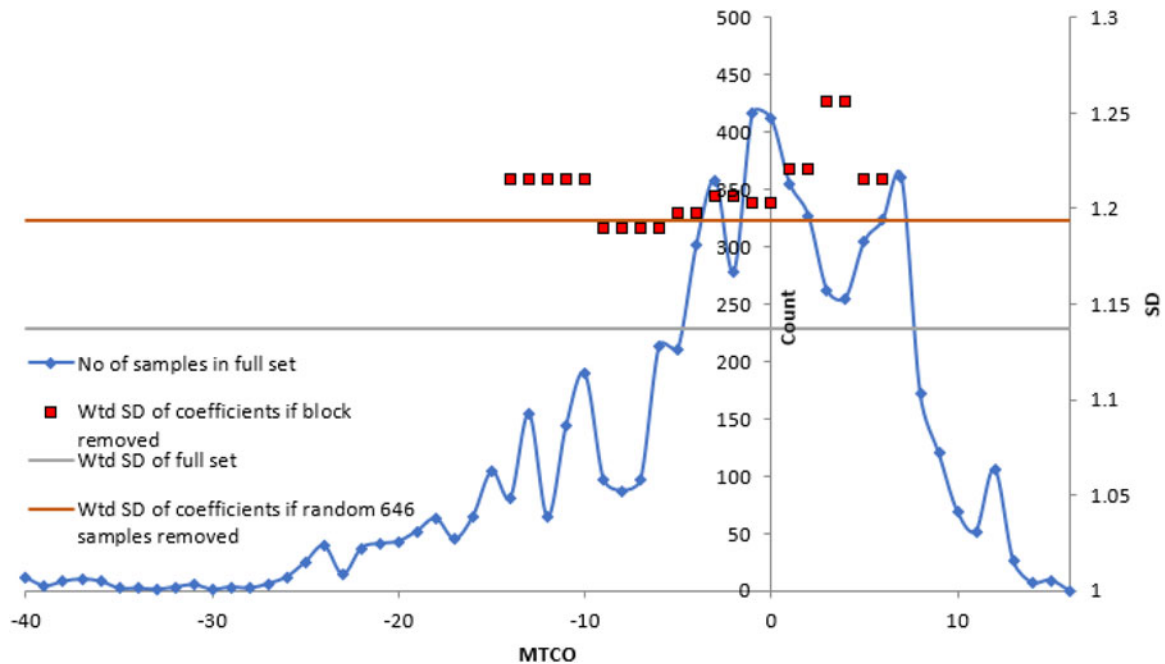
Figure 5. Impact of lack of continuity in sampling showing the effect on standard deviation (SD) of coefficients of the full set of taxa (n = 195), weighted by abundance, of removal of all samples in a given specified range of MTCO from the calibration data set. The red blocks (weighted [Wtd] SD of coefficients if block removed) show the impact of removing a specified number of samples (in this case 646 samples) contiguous in climate space. The low number of cold samples means that the first gap (samples 647–1293) starts at -14.6°C. The abundance-weighted SD using the full sample set (n = 6458) is indicated as a grey line for comparison, and the effect of removal of 646 randomly selected samples is indicated by a red line.

Fossil pollen assemblages almost always contain fewer taxa than the training data set. Thus, it might seem logical to base reconstructions solely on the taxa present in the fossil assemblages. However, the absence of taxa can also provide information about climate and thus can contribute usefully to the WA-PLS reconstruction. Reconstructions of MTCO using only the taxa that are present in the fossil pollen samples at Lake Ioannina are not distinguishable from those made with the full SMPDS, either in terms of mean values or reconstruction spread (Supplementary Fig. 15b). However, at other sites, the use of only taxa present in the fossil record results in a difference in the reconstructions. At Megali Limni, for example, the reconstructions based on the fossil-taxa-only model are ca 1°C colder than those based on the full SMPDS data set (Supplementary Fig. 15a). At both Lake Ioannina and at Megali Limni, the model performance indicators show a substantial improvement using the full SMPDS data set: the r^2 increases from ~0.55 to 0.7 at both sites, and the RMSE decreases from 5.80 to 4.82 at Ioannina and from 5.88 to 4.82 at Megali Limni (Supplementary Table 3). Using the full SMPDS data set is beneficial because, although taxon transfer coefficients are nearly independent of each other (and thus are little affected by the presence/absence of other taxa), WA-PLS uses information from other taxa to refine the coefficients.

The coefficients of rare taxa are not stable (see Fig. 3), and this can cause problems if these taxa are anomalously abundant, even at relatively low levels, in fossil samples. Changes

in reconstruction spread are one indication of what is essentially a poor-analogue problem. At Lake Ioannina, for example, the increased spread in reconstructed MTCO during the interval between 73 and 68 ka (see Fig. 2b) occurs because the abundance of *Ulmus/Zelkova* increases from ca 0.75% to 6–8% of the sample total. This taxon is recorded at only 31 sites in the modern data set and has a very large SD (14.7°C). Alternative methods, such as squared chord distance (Supplementary Fig. 16), confirm that this interval has poor analogues. Although little can be done about such poor-analogue situations, it is important to identify them through the use of a robust measure of reconstruction uncertainty for individual samples, such as the reconstruction spread measure used here.

DISCUSSION

We have shown that the breadth and continuity of the climate space sampled by the modern training data set are important determinants of the quality of the reconstructions, although this may not be reflected in the WA-PLS model performance indicators. The number of samples in the data set does not provide a measure of the adequacy of sampling. There is a risk that training sets that rely on a very small number of samples (e.g., Xu et al., 2010; Salonen et al., 2012; Ding et al., 2017) do not sample a large enough range of climate space to encompass the true past climate if this is very different from present. However, as our comparison between
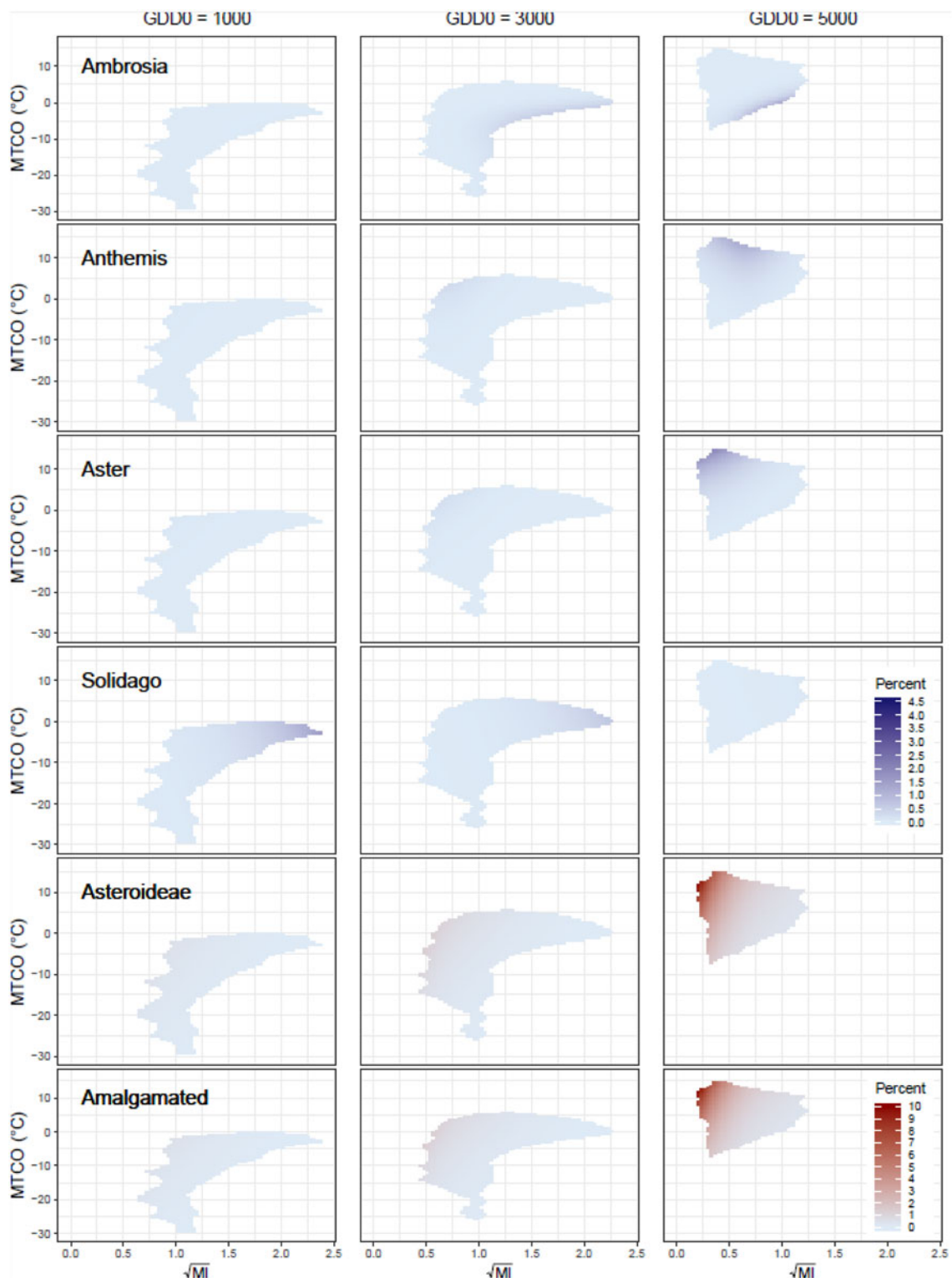
Figure 6. Climate space diagrams for individual taxa in the Asteroideae, based on Generalized Additive Models (GAMs). The plots are two-dimensional representations of a three-dimensional space defined by mean temperature of the coldest month (MTCO, °C), growing degree days above a baseline of 0°C (GDD$_0$, °C day), and the square root of a moisture index (MI, unitless), defined as the ratio of annual precipitation to annual potential evapotranspiration. The three columns show slices of the GAM at the points where GDD$_0$ is 1000°C, 3000°C, and 5000°C day. The plots have been trimmed so that they only represent the sampled areas using convex hulls. The top four rows show the relative abundance on climate space of four of the most important taxa in the Asteroideae, the fifth row shows the relative abundance of Asteroideae where this was explicitly recorded as a taxon in the original pollen counts, and the bottom panels show the relative abundance for the amalgamated taxon Asteroideae, created by summing all named component taxa together. Two scales are used, distinguished by colour: the scale for the top four panels is shown in the Solidago row, and for the bottom two the scale is shown in the Amalgamated row. Supplementary Fig. 13 shows the abundance in the same climate space at the sample level.

Table 3. Examples of the impact of amalgamating taxa to higher taxonomic levels on the bootstrapped standard deviation (SD) of transfer coefficients for mean temperature of the coldest month (MTCO). The mean SD of the component taxa (column 3) is weighted by abundance. Additional information on the component taxa and the transfer coefficients (climate optima) identified for the component taxa and the amalgamated taxon is given in Supplementary Tables 4 and 5.

Amalgamation level	Amalgamated taxon	Mean SD of component taxa	SD of amalgamated taxon
Family	Apiaceae	2.3	1.7
Family	Plantaginaceae	2.2	1.2
Subfamily	Asteroideae	2.5	1.3
Subfamily	Carduoideae	3.6	2.1
Subfamily	Cichoroideae	1.0	1.0
Genus	*Cistus*	2.5	1.7
Genus	*Ephedra*	4.8	3.5
Functional grouping	Deciduous Quercus	0.8	0.5
Functional grouping	Evergreen Quercus	0.7	0.4

reconstructions based on the EMPD and the full SMPDS shows, even large data sets may not have a sufficiently continuous sampling of climate space to provide stable coefficients and hence robust reconstructions. Furthermore, large data sets may still not encompass more extreme climates. This issue could be particularly important if the goal is to reconstruct climates very different from present-day climates, such as those of the last glacial period. Although most studies specify the training data set used, very few discuss how the training set was chosen or whether it samples an appropriate range of climate for the target reconstruction. Our analyses suggest that a minimum requirement is to plot the sampled climate space and thus to evaluate whether it is fit for purpose (Table 4). From the perspective of improving existing training data sets, the focus should be on filling gaps in climate space rather than adding samples in climate space already represented.

Amalgamating pollen taxa to higher taxonomic levels can provide a way of increasing the continuity of the sampling of climate and reducing the statistical error of the reconstructions. Huntley et al. (1989) were the first to demonstrate niche conservatism in surface pollen data, showing that the abundance distribution of *Fagus* as recorded in surface pollen samples in North America could be predicted from the
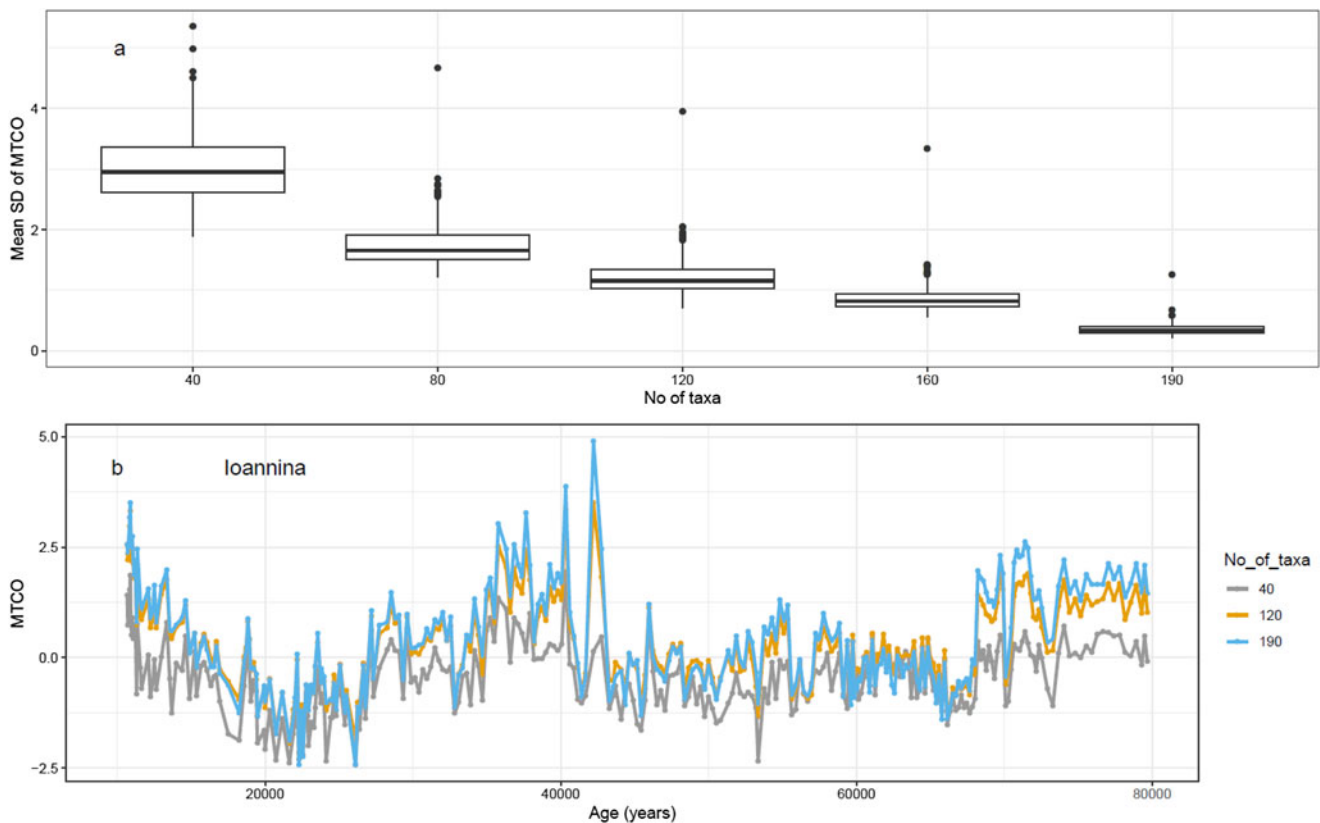
Figure 7. (a) Change in the standard deviation (SD) of the reconstruction of mean temperature of the coldest month (MTCO) as increasing numbers of taxa are used in the Weighted Averaging Partial Least Squares (WA-PLS) regression, taken across the whole time series, and **(b)** the mean reconstructions at three example points. The specified number of taxa is randomly selected without replacement 100 times, and the WA-PLS reconstruction run. The SD is taken across the 100 runs. Three WA-PLS components were used, but only results with $p < 0.05$ in a random t-test were included in the means and SDs.

Table 4. Checklist of issues, steps to identify them, and some potential solutions.

Issue	Potential actions
Is the sampling of climate space sufficient to reconstruct very different climates from present?	1) Plot sampled climate space and determine whether this provides information that would allow reconstruction of different (e.g., much warmer or much colder than present) climates. 2) Expand training data set to include a wider sampling of climate, as determined by the expected past climate change (i.e., colder climate for glacial samples, warmer climates for early Holocene samples), by expanding the geographical coverage of the data set.
Is the sampling of climate space continuous or are there gaps which result in ambiguous or multiple optima for individual taxa?	1) Plot sampled climate space to identify regions that are underrepresented. 2) Plot taxon distributions in climate space to identify existence of multiple optima which might indicate discontinuous sampling of the climate niche of the taxon. 3) Expand training data sets by explicit field sampling of underrepresented areas of climate space.
Is the amalgamation of taxa into higher taxonomic groupings appropriate?	1) Use Generalized Additive Models (or similar approaches) to test whether the individual taxa included within a higher taxonomic grouping occupy a similar climate niche.
How many taxa should be included in the model?	1) Do not exclude taxa a priori. 2) Exclude taxa that are not represented at a sufficient number of sites in the training data set to sample their full climate range. The definition of the minimum number of sites could be based on whether these sites cover the modern climate niche of the taxon (e.g., as shown by presence/absence data) or from modern distribution maps. 3) Use independent evidence (e.g., sedimentological evidence) to exclude taxa that reflect local non-climate signals (e.g., erosion). 4) Check that the exclusion of rarely sampled taxa or taxa reflecting anomalous local signals has no impact on the resulting reconstruction.
Do taphonomic or preservation issues have an impact on the reconstructions?	1) Check the abundance of indeterminable (e.g., corroded) grains in the samples as a measure of the quality of the assemblage and the likelihood that there are preservation issues. 2) Calculate a biodiversity index (e.g., HillsN2) to assess the degree to which samples might be affected by preservation issues.
Is the choice of model appropriate?	1) Do not rely on the standard model diagnostics to select a model. 2) Use diagnostics that demonstrate the stability of transfer coefficients and consequent reconstruction spread to calculate uncertainties on the reconstructions for individual samples.

distribution of the genus in Europe and vice versa. Subsequent studies have emphasised the degree to which lineages are conservative in terms of climate preferences (Ackerly, 2003; Wake et al., 2009). Williams and Shuman (2008) have provided counterexamples, where congeneric species occur at different parts of the moisture gradient in eastern and western North America. However, they also point out that the impact of this on reconstructions of climate decreases as the number of taxa used increases. Nevertheless, it should not be assumed that taxon amalgamation is always or automatically justified. One way of testing whether taxon distributions are unimodal and whether amalgamations to higher taxonomic levels are defensible is to construct GAMs (Wei et al., 2019b).

The SMPDS includes sites from across the whole of the Palaearctic phytogeographic region (Wallace, 1876; Good,

1974; Kreft and Jetz, 2010; Dengler et al., 2014) in order to create a training data set that samples colder and more continental climates than would be available from a more limited region. Although the sampling of this region is incomplete, since there are large sampling gaps in Russia and northern China, the SMPDS provides samples from much colder winter climates and climates with a greater seasonal temperature range (see Fig. 1). The expansion to the Palaearctic phytogeographic region relies on the fact that this region is characterised by taxa that have diversified in relatively recent geological time, that is, during approximately the past 20 Ma. The climate preferences of these taxa are expected to be conservative, as has been demonstrated for many plant lineages (Ackerly, 2003; Williams and Shuman, 2008; Wake et al., 2009). The expansion of a training data set to increase the diversity of sampled climates is useful in

principle, because it increases the range of different climates that can be reconstructed. As in the case of amalgamations to higher taxonomic levels, however, it is both possible and desirable to test (e.g., using GAMs, see Table 4) whether the taxa considered retain a coherent, unimodal relationship to the climate variables of interest. There is no established rule about the choice of training data set, and the general hypothesis of climatic niche conservatism within phytogeographic regions would repay further research. However, the common practice of using training data sets with limited geographic coverage (Xu et al., 2010; Salonen et al., 2012; Ding et al., 2017) effectively limits the range of climates that can be reconstructed. Similarly, exclusion of modern pollen samples from particular vegetation types (Sinopoli et al., 2019) prevents the reconstruction of climates characterised by these vegetation types.

We have shown that WA-PLS model performance statistics and the stability of the coefficients improve as taxa are added. Several studies have examined the impact of using a subset of important taxa, defined in terms of abundance (Birks, 1994) or predictive importance (Racca et al., 2003), on reconstructions based on weighted averaging (WA). The results have been inconclusive. Juggins et al. (2015) showed that removing uninformative taxa resulted in better performance using artificially constructed data sets but did not have a similar impact on real data sets. Jonkers and Kučera (2018) found no impact from reducing the number of foraminifera species on WA model performance but showed that the resulting sea-surface temperature reconstructions were significantly different between reduced- and full-taxa reconstructions. A similar conclusion was reached by Bjune et al. (2010) in a study contrasting climate reconstructions based on 191 vs. 321 core-top pollen samples.

The impact of including taxa that do not have a significant relationship to the climate variable being reconstructed and therefore have limited predictive power is closely tied to whether they are influenced by other climate or environmental factors, which would result in a degradation of model performance as shown by the analysis in Juggins et al. (2015). Thus, the choice of appropriate variables to reconstruct, and accounting for interactions among variables, is probably more useful than removing taxa that are thought to be uninformative.

Some studies have excluded taxa that are overrepresented in pollen assemblages, such as *Pinus*, on the assumption that these suppress the expression of the signal from less abundant taxa (e.g., Sinopoli et al., 2019). Our analyses suggest there is no reason to exclude such taxa on a priori grounds. However, there may be circumstances that warrant exclusion of specific taxa. Wei et al. (2019a), for example, excluded both Poaceae and Polypodiales from the taxa used to reconstruct climate at El Cañizar de Villarquemado on the grounds that high abundances of these two taxa caused anomalous excursions in moisture reconstructions that were not seen in temperature reconstructions, and because independent sedimentological evidence supported the idea that anomalously high abundances of these two taxa in some

samples represented reeds (aquatic grasses, in the case of Poaceae) or erosional inputs (in the case of Polypodiales). They showed that the removal of these taxa had no impact except on the MI reconstructions of the anomalous samples. Insofar as anomalous results can be attributed to the high abundance of specific taxa that could reflect the influence of non-climatic factors, exclusions may be warranted. It is, however, important to test the impact of these exclusions on the final reconstructions (see Table 4).

It is often tacitly assumed that better WA-PLS model performance means better reconstruction, but we have shown that WA-PLS models of equivalent performance can produce very different reconstructions (e.g., see Fig. 2). WA-PLS model performance only reflects how well the model replicates the climate space of the modern training data set but says nothing about whether that climate space is adequate or appropriate. We conclude that model performance must be considered alongside an assessment of the representativeness of climate at the data set and taxon level (see Table 4). This could be done, for example, by plotting the sites and taxon abundances in climate space and expanding the calibration data set to ensure that it samples a range of climates commensurate with the climate that might have been experienced at a given location. Although the RMSE of the modern calibration is generally used as a measure of reconstruction uncertainty, it is in fact a measure of the average uncertainty for a randomly selected surface sample. Thus, it provides the same estimate of uncertainty for all fossil pollen samples, regardless of whether the taxa in the fossil samples are well represented in the modern data set and have stable coefficients. It provides no information about uncertainties arising because of mismatches between the sampling of modern taxa and their representation in fossil assemblages. We have used an alternative approach that propagates the calibration errors into the downcore reconstructions by assessing the calibration error associated with the particular assemblage in an individual pollen sample. This approach provides a way of identifying specific intervals downcore where the reconstruction errors are well constrained and differentiating these from intervals with larger uncertainties. As we show in the Lake Ioannina example (see Fig. 2b), it is then possible to identify the specific cause of these larger uncertainties and, in particular, a way of identifying poor-analogue situations that would be overlooked using the standard approach to representing uncertainty. Thus, the bootstrapped reconstruction spread provides an alternative, and we would argue better, measure of the consistency and redundancy of the calibration set and thus indirectly the robustness of the reconstructions than the standard use of RMSE.

There will be a continuing need for quantitative estimates of past climates in order to evaluate climate model simulations, and it is important to continue to explore ways to demonstrate and, if possible, improve their robustness. While there are alternative approaches to statistical reconstruction, such as vegetation-model inversion (Guiot et al., 2000; Wu et al., 2007, 2009; Izumi and Bartlein, 2016) or data assimilation (Goosse et al., 2006; Annan and Hargreaves, 2013;

Cleator et al., 2019), they do not provide a panacea—in the one case because they are heavily dependent on the specifics of the model, in the other because they require quantitative estimates as inputs. It has been suggested (Brewer et al., 2008) that more robust estimates of past climates can be obtained by using multiple statistical approaches, and this has been done in an increasing number of studies (Xu et al., 2010; Guiot and de Vernal, 2011; Peyron et al., 2011; Jonkers and Kučera, 2018; Sinopoli et al., 2019). However, many of the sampling issues described here apply equally to all analogue- or regression-based reconstruction methods. Thus, while using several methods together may provide an indication of the robustness of a reconstruction, such robustness does not necessarily demonstrate that the reconstructions are correct. Exploring and documenting methodological decisions, as suggested here, should lead to more transparent reconstructions and perhaps greater consistency between reconstructions at different sites.

CONCLUSIONS

Care needs to be exercised in the application of statistical reconstruction techniques, such as WA-PLS, in order to ensure that the results are robust and reasonable. We have demonstrated that WA-PLS model performance metrics are not a sufficient guide to the reliability of the reconstructions: models with similar performance metrics nevertheless provide different reconstructions of past climates. We argue that the modern pollen training data set should not only sample the full width of the climate niche of each taxon but also provide a continuous sampling of this range. The width and continuity of the sampling of climate space are more important than number of samples. Width and continuity can be improved by amalgamating taxa to higher taxonomic levels, providing the assumptions about the unimodality of these higher taxa with respect to individual climate variables are explicitly tested. We also argue that it is important for the modern training set to sample climates that encompass the full range of the climates that might be expected to have occurred in the past; for example, to sample colder and drier climates that might be found in the Mediterranean during the last glacial period. While it is difficult to know a priori how broad a sampling of climate space is necessary to ensure this for a specific site or region, ensuring that the training data set is broad enough to test explicit hypotheses about the expected climate change would be useful. These hypotheses could be based on alternative sources of palaeoclimate information, including model experiments.

Our analyses are not designed to provide a prescription for how to apply WA-PLS: we are not recommending the use of a specific modern pollen training data set or specific methods to test the usefulness of this data set. Instead, we are advocating that the robustness of statistical reconstructions should be explicitly tested and documented. We have suggested a number of straightforward checks (see Table 4) that should be made. These include: (1) testing the range and continuity of the sampling of climate gradients in the modern training data set by plotting the modern samples in climate space; (2) ensuring that the sampled climate covers a sufficient range to be able to reconstruct substantially different climates in the past, for example by ensuring that the training data set includes samples which are both much warmer/colder and wetter/drier than the modern climate of the region; (3) examining whether amalgamation to higher taxonomic levels provides a more continuous sampling of climate space by plotting the climate space of both individual taxa and the higher taxon into which they are placed; and (4) using the largest number of taxa in the reconstructions and avoiding a priori taxon exclusions. The reasons for any exclusions should be transparent and supported by independent evidence, and the impact of these exclusions on the reconstructed climate should be tested. The proposed checks do not necessarily identify whether specific methodological choices are correct, but they do provide insights into the uncertainties associated with the reconstructions.

AUTHOR CONTRIBUTIONS

All authors contributed to the conceptual development. Formal analysis, methods, and software: MT with contributions from DW on climate and GAMs. Writing—original draft: MT and SPH. Writing—editing and review: all authors.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/qua.2020.44.

REFERENCES

Ackerly, D.D., 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* 164, S165–S184. https://doi.org/10.1086/368401.

Annan, J.D., Hargreaves, J.C., 2013. A new global reconstruction of temperature changes at the Last Glacial Maximum. *Climate of the Past* 9, 367. http://dx.doi.org/10.5194/cp-9-367-2013.

Bartlein, P.J., Harrison, S.P., Brewer, S., Connor, S., Davis, B.A.S., Gajewski, K., Guiot, J., *et al*.., 2011. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics* 37, 775–802. https://doi.org/10.1007/s00382-010-0904-1.

Bigelow, N., Brubaker, L., Edwards, M.E., Harrison, S.P., 2003. Climate change and Arctic ecosystems: 1. Vegetation changes north of 55°N between the last glacial maximum, mid-Holocene, and present. *Journal of Geophysical Research* 108. https://doi.org/10.1029/2002JD002558.

Birks, H.J.B., 1994. The importance of pollen and diatom taxonomic precision in quantitative palaeoenvironmental reconstructions. *Review of Palaeobotany and Palynology* 83, 107–117. https://doi.org/10.1016/0034-6667(94)90062-0.

Birks, H.J.B., Seppä, H., 2004. Pollen-based reconstructions of late-Quaternary climate in Europe - Progress, problems, and pitfalls. *Acta Palaeobotanica* 44, 317–334.

Bjune, A.E., Birks, H.J.B., Peglar, S.M., Odland, A., 2010. Developing a modern pollen-climate calibration data set for Norway. *Boreas* 39, 674–688. https://doi.org/10.1111/j.1502-3885.2010.00158.x.

Braconnot, P., Harrison, S.P., Kageyama, M., Bartlein, P.J., Masson-delmotte, V., Abe-ouchi, A., Otto-bliesner, B., Zhao, Y., 2012. Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*; London 2, 417–424. http://dx.doi.org/10.1038/nclimate1456

Brewer, S., Guiot, J., Sánchez-Goñi, M.F., Klotz, S., 2008. The climate in Europe during the Eemian: a multi-method approach using pollen data. *Quaternary Science Reviews* 27, 2303–2315. https://doi.org/10.1016/j.quascirev.2008.08.029.

Brooks, S.J., Birks, H.J.B., 2001. Chironomid-inferred air temperatures from Lateglacial and Holocene sites in north-west Europe: progress and problems. *Quaternary Science Reviews, European Quaternary Biostratigraphy* 20, 1723–1741. https://doi.org/10.1016/S0277-3791(01)00038-5.

Brunsdon, C., Fotheringham, A.S., Charlton, M., 2002. Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems* 26, 501–524. https://doi.org/10.1016/S0198-9715(01)00009-6.

Cleator, S.F., Harrison, S.P., Nichols, N.K., Prentice, I.C., Roulstone, I., 2019. A method for generating coherent spatially explicit maps of seasonal palaeoclimates from site-based reconstructions. arXiv:1902.04973 [physics].

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G, 2013. Long-term climate change: projections, commitments and irreversibility. In: *Climate Change 2013—The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp. 1029–1136.

Davis, B.A.S., Zanon, M., Collins, P., Mauri, A., Bakker, J., Barboni, D., Barthelmes, A., *et al*.., 2013. The European Modern Pollen Database (EMPD) project. *Vegetation History and Archaeobotany* 22, 521–530. https://doi.org/10.1007/s00334-012-0388-5.

Davis, T.W., Prentice, I.C., Stocker, B.D., Thomas, R.T., Whitley, R.J., Wang, H., Evans, B.J., Gallego-Sala, A.V., Sykes, M.T., Cramer, W., 2017. Simple process-led algorithms for simulating habitats (SPLASH v.1.0): robust indices of radiation, evapotranspiration and plant-available moisture. *Geoscientific Model Development* 10, 689–708. http://dx.doi.org/10.5194/gmd-10-689-2017.

Dengler, J., Janišová, M., Török, P., Wellstein, C., 2014. Biodiversity of Palaearctic grasslands: a synthesis. In: *Agriculture, Ecosystems & Environment*, Biodiversity of Palaearctic grasslands: processes, patterns and conservation 182, 1–14. https://doi.org/10.1016/j.agee.2013.12.015.

Ding, W., Xu, Q., Tarasov, P.E., 2017. Examining bias in pollen-based quantitative climate reconstructions induced by human impact on vegetation in China. *Climate of the Past* 13, 1285.

Edwards, M.E., Anderson, P.M., Brubaker, L.B., Ager, T.A., Andreev, A.A., Bigelow, N.H., Cwynar, L.C., *et al.*, 2000. Pollen-Based Biomes for Beringia 18,000, 6000 and 0 14C yr BP. *Journal of Biogeography* 27, 521–554.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7, 1–26.

ESRI, 2014. ArcGIS Desktop. Environmental System Research Institute, Redlands, California.

Sanchez Goñi, M.F., Desprat, S., Daniau, A.-L., Bassinot, F.C., Polanco-Martínez, J.M., Harrison, S.P., Allen, J.R.M., *et al.*, 2017. The ACER pollen and charcoal database: a global resource to document vegetation and fire response to abrupt climate changes during the last glacial period. https://doi.org/10.1594/PANGAEA.870867. Supplement to: Sanchez Goñi, M.F., Desprat, S., Daniau, A.-L., Bassinot, F.C., Polanco-Martínez, J.M., Harrison, S.P., Allen, J.R.M., *et al.*, 2017. The ACER pollen and charcoal database: a global resource to document vegetation and fire response to abrupt climate changes during the last glacial period. *Earth System Science Data*, 9, 679–695, https://doi.org/10.5194/essd-9-679-2017.

Good, R., 1974. *The Geography of Flowering Plants*. 4th ed. Addison-Wesley-Longman.

Goosse, H., Renssen, H., Timmermann, A., Bradley, R.S., Mann, M.E., 2006. Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Climate Dynamics* 27, 165–184. https://doi.org/10.1007/s00382-006-0128-6.

Guiot, J., de Vernal, A., 2011. Is spatial autocorrelation introducing biases in the apparent accuracy of paleoclimatic reconstructions? *Quaternary Science Reviews* 30, 1965–1972. https://doi.org/10.1016/j.quascirev.2011.04.022.

Guiot, J., Torre, F., Jolly, D., Peyron, O., Boreux, J.J., Cheddadi, R., 2000. Inverse vegetation modeling by Monte Carlo sampling to reconstruct palaeoclimates under changed precipitation seasonality and CO_2 conditions: application to glacial climate in Mediterranean region. *Ecological Modelling* 127, 119–140. https://doi.org/10.1016/S0304-3800(99)00219-7.

Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157, 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1.

Harrison, S.P., 2019. Modern pollen data for climate reconstructions, version 1 (SMPDS). University of Reading. http://dx.doi.org/10.17864/1947.194.

Harrison, S.P., 2020. Climate reconstructions for the SMPDSv1 modern pollen data set. https://doi.org/10.5281/zenodo.3605003.

Harrison, S.P., Bartlein, P.J., Brewer, S., Prentice, I.C., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., Willis, K., 2014. Climate model benchmarking with glacial and mid-Holocene climates. *Climate Dynamics* 43, 671–688. https://doi.org/10.1007/s00382-013-1922-6.

Harrison, S.P., Bartlein, P.J., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P., Kageyama, M., 2015. Evaluation of CMIP5 palaeo-simulations to improve climate projections.

Nature Climate Change 5, 735–743. https://doi.org/10.1038/nclimate2649.

Harrison, S.P., Prentice, I.C., Barboni, D., Kohfeld, K.E., Ni, J., Sutra, J.-P., 2010. Ecophysiological and bioclimatic foundations for a global plant functional classification. *Journal of Vegetation Science* 21, 300–317. https://doi.org/10.1111/j.1654-1103.2009.01144.x.

Hellman, S., Gaillard, M.-J., Broström, A., Sugita, S., 2008. The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes: validation in southern Sweden. *Journal of Quaternary Science* 23, 21–42. https://doi.org/10.1002/jqs.1126.

Huntley, B., Bartlein, P.J., Prentice, I.C., 1989. Climatic Control of the Distribution and Abundance of Beech (*Fagus L.*) in Europe and North America. *Journal of Biogeography* 16, 551–560. https://doi.org/10.2307/2845210.

Izumi, K., Bartlein, P.J., 2016. North American paleoclimate reconstructions for the Last Glacial Maximum using an inverse modeling through iterative forward modeling approach applied to pollen data. *Geophysical Research Letters* 43, 10,965–10,972. https://doi.org/10.1002/2016GL070152.

Jonkers, L., Kučera, M., 2018. Sensitivity to species selection indicates the effect of nuisance variables on marine microfossil transfer functions. *Climate of the Past Discussions* 1–19. https://doi.org/10.5194/cp-2018-107.

Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? *Quaternary Science Reviews* 64, 20–32. https://doi.org/10.1016/j.quascirev.2012.12.014.

Juggins, S., 2017. rioja: Analysis of quaternary science data. CRAN.

Juggins, S., Simpson, G.L., Telford, R.J., 2015. Taxon selection using statistical learning techniques to improve transfer function prediction. *The Holocene* 25, 130–136. https://doi.org/10.1177/0959683614556388.

Kirtman, B., Power, S.B., Adedoyin, A.J., Boer, G.J., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A.M., Kimoto, M., and Meehl, G., 2013. Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013—The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp. 953–1020.

Kreft, H., Jetz, W., 2010. A framework for delineating biogeographical regions based on species distributions: Global quantitative biogeographical regionalizations. *Journal of Biogeography* 37, 2029–2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x.

Lotter, A.F., Birks, H.J.B., Hofmann, W., Marchetto, A., 1997. Modern diatom, cladocera, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps. I. *Climate. Journal of Paleolimnology* 18, 395–420. https://doi.org/10.1023/A:1007982008956.

Marinova, E., Harrison, S.P., Bragg, F., Connor, S., Laet, V. de, Leroy, S.A.G., Mudie, P., *et al.*, 2018. Pollen-derived biomes in the Eastern Mediterranean–Black Sea–Caspian-Corridor. *Journal of Biogeography* 45, 484–499. https://doi.org/10.1111/jbi.13128.

Marsicek, J., Shuman, B.N., Bartlein, P.J., Shafer, S.L., Brewer, S., 2018. Reconciling divergent trends and millennial variations in Holocene temperatures. *Nature* 554, 92–96. https://doi.org/10.1038/nature25464.

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

Pateiro-Lopez, B., Rodriguez-Casal, A., 2016. alphahull: Generalization of the Convex Hull of a Sample of Points in the Plane. CRAN.

Peyron, O., Goring, S., Dormoy, I., Kotthoff, U., Pross, J., de Beaulieu, J.-L., Drescher-Schneider, R., Vannière, B., Magny, M., 2011. Holocene seasonality changes in the central Mediterranean region reconstructed from the pollen sequences of Lake Accesa (Italy) and Tenaghi Philippon (Greece). *The Holocene* 21, 131–146. https://doi.org/10.1177/0959683610384162.

Prentice, I.C., 1985. Pollen Representation, Source Area, and Basin Size: Toward a Unified Theory of Pollen Analysis. *Quaternary Research* 23, 76–86. https://doi.org/10.1016/0033-5894(85)90073-0.

Prentice, I.C., Cleator, S.F., Huang, Y.H., Harrison, S.P., Roulstone, I., 2017. Reconstructing ice-age palaeoclimates: Quantifying low-CO_2 effects on plants. *Global and Planetary Change* 149, 166–176. https://doi.org/10.1016/j.gloplacha.2016.12.012.

Prentice, I.C., Parsons, R.W., 1983. Maximum Likelihood Linear Calibration of Pollen Spectra in Terms of Forest Composition. *Biometrics* 39, 1051–1057. https://doi.org/10.2307/2531338.

Racca, J.M.J., Wild, M., Birks, H.J.B., Prairie, Y.T., 2003. Separating wheat from chaff: Diatom taxon selection using an artificial neural network pruning algorithm. *Journal of Paleolimnology* 29, 123–133.

R Core Team, 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Rymes, M.D., Myers, D.R., 2001. Mean preserving algorithm for smoothly interpolating averaged data. *Solar Energy* 71, 225–231. https://doi.org/10.1016/S0038-092X(01)00052-4.

Salonen, J.S., Ilvonen, L., Seppä, H., Holmström, L., Telford, R.J., Gaidamavičius, A., Stančikaitė, M., Subetto, D., 2012. Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative climate reconstruction. *The Holocene* 22, 413–424. https://doi.org/10.1177/0959683611425548.

Salonen, J.S., Seppä, H., Birks, H.J.B., 2013. The effect of calibration data set selection on quantitative palaeoclimatic reconstructions. *The Holocene* 23, 1650–1654. https://doi.org/10.1177/0959683613496295.

Schmidt, G.A., Annan, J.D., Bartlein, P.J., Cook, B.I., Guilyardi, E., Hargreaves, J.C., Harrison, S.P., *et al.*, 2014. Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate of the Past* 10, 221–250. https://doi.org/10.5194/cp-10-221-2014.

Seppä, H., Bjune, A.E., Telford, R.J., Birks, H.J.B., Veski, S., 2009. Last nine-thousand years of temperature variability in Northern Europe. *Climate of the Past* 13.

Shennan, I., Long, A.J., Horton, B.P., 2015. *Handbook of Sea-Level Research*. American Geophysical Union, Washington, United Kingdom.

Sinopoli, G., Peyron, O., Masi, A., Holtvoeth, J., Francke, A., Wagner, B., Sadori, L., 2019. Pollen-based temperature and precipitation changes in the Ohrid Basin (western Balkans) between 160 and 70ka. *Climate of the Past* 15, 53–71. https://doi.org/10.5194/cp-15-53-2019.

Sugita, S., 2007. Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. *The Holocene* 17, 229–241. https://doi.org/10.1177/0959683607075837.

Telford, R.J., Birks, H.J.B., 2005. The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quaternary Science Reviews* 24, 2173–2179. https://doi.org/10.1016/j.quascirev.2005.05.001.

Telford, R.J., Birks, H.J.B., 2011. A novel method for assessing the statistical significance of quantitative reconstructions inferred

from biotic assemblages. *Quaternary Science Reviews* 30, 1272–1278. https://doi.org/10.1016/j.quascirev.2011.03.002.

ter Braak, C.J.F., Juggins, S., 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269/270, 485–502.

van der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics and Intelligent Laboratory Systems* 25, 313–323. https://doi.org/10.1016/0169-7439(94)85050-X.

Wake, D.B., Hadly, E.A., Ackerly, D.D., 2009. Biogeography, Changing Climates, and Niche Evolution. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19631–19636.

Wallace, A.R., 1876. *The geographical distribution of animals*. Harper and Brothers, New York.

Wang, H., Prentice, I.C., Keenan, T.F., Davis, T.W., Wright, I.J., Cornwell, W.K., Evans, B.J., Peng, C., 2017. Towards a universal model for carbon dioxide uptake by plants. *Nature Plants* 3, 734. https://doi.org/10.1038/s41477-017-0006-8.

Wang, H., Prentice, I.C., Ni, J., 2013. Data-based modelling and environmental sensitivity of vegetation in China. *Biogeosciences* 10, 5817. http://dx.doi.org/10.5194/bg-10-5817-2013.

Wei, D., González-Sampériz, P., Gil-Romera, G., Harrison, S.P., Prentice, I.C., 2019a. Climate changes in interior semi-arid Spain from the last interglacial to the late Holocene. *Climate of the Past* Discussions 1–31. https://doi.org/10.5194/cp-2019-16.

Wei, D., Harrison, S., Prentice, I.C., 2019b. *The climate space of European pollen taxa*. University of Reading.

Wei, D., Prentice, I.C., Harrison, S., 2020. *The climatic space of European pollen taxa*. Ecology/Ecological Archives (in press).

Wickham, H., 2016. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. CRAN.

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

Wood, S, 2017. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.

Wu, H., Guiot, J., Brewer, S., Guo, Z., 2007. Climatic changes in Eurasia and Africa at the last glacial maximum and mid-Holocene: reconstruction from pollen data using inverse vegetation modelling. *Climate Dynamics* 29, 211–229. https://doi.org/10.1007/s00382-007-0231-3.

Wu, H., Guiot, J., Peng, C., Guo, Z., 2009. New coupled model used inversely for reconstructing past terrestrial carbon storage from pollen data: validation of model using modern data. *Global Change Biology* 15, 82–96. https://doi.org/10.1111/j.1365-2486.2008.01712.x.

Xu, Q., Xiao, J., Li, Y., Tian, F., Nakagawa, T., 2010. Pollen-Based Quantitative Reconstruction of Holocene Climate Changes in the Daihai Lake Area, Inner Mongolia, China. *Journal of Climate* 23, 2856–2858,2861–2868.