

Comparison of spatial downscaling methods of general circulation model results to study climate variability during the Last Glacial Maximum

Guillaume Latombe^{1,2}, Ariane Burke³, Mathieu Vrac⁴, Guillaume Levavasseur⁵, Christophe Dumas⁴,
5 Masa Kageyama⁴, Gilles Ramstein⁴

¹Centre for invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Matieland 7602, South Africa

²School of Biological Sciences, Monash University, Melbourne 3800, Australia

³Département d'Anthropologie, Université de Montréal, Montréal, QC, Canada

⁴Laboratoire des Sciences du Climat et de l'Environnement/Institut Pierre-Simon Laplace, Université Paris-Saclay, CE Saclay,
10 l'Orme des Merisiers, Bât.701, Gif-sur-Yvette, France

⁵Institut Pierre Simon Laplace (IPSL), Pôle de modélisation du climat, UPMC, Paris, France

Correspondence to: Guillaume Latombe (latombe.guillaume@gmail.com)

Abstract. The extent to which climate conditions influenced the spatial distribution of hominin populations in the past is highly
15 debated. General Circulation Models (GCMs) and archaeological data have been used to address this issue. Most GCMs are
not currently capable of simulating past surface climate conditions with sufficiently detailed spatial resolution to distinguish
areas of potential hominin habitat, however. In this paper we propose a Statistical Downscaling Method (SDM) for increasing
the resolution of climate model outputs in a computationally efficient way. Our method uses a generalized additive model
(GAM), calibrated over present-day climatology data, to statistically downscale temperature and precipitation time series from
20 the outputs of a GCM simulating the climate of the Last Glacial Maximum (19-23,000 BP) over Western Europe. Once the
SDM is calibrated, we first interpolate the coarse-scale GCM outputs to the final resolution and then use the GAM to compute
surface air temperature and precipitation levels using these interpolated GCM outputs and fine resolution geographical
variables such as topography and distance from an ocean. The GAM acts as a transfer function, capturing non-linear
relationships between variables at different spatial scales and correcting for the GCM biases. We tested three different
25 techniques for the first interpolation of GCM output: bilinear, bicubic, and kriging. The resulting SDMs were evaluated by
comparing downscaled temperature and precipitation at local sites with paleoclimate reconstructions based on paleoclimate
archives (archaeozoological and palynological data) and the impact of the interpolation technique on patterns of variability
was explored. The SDM based on kriging interpolation, providing the best accuracy, was then validated on present-day data
outside of the calibration period. Our results show that the downscaled temperature and precipitation values are in good
30 agreement with paleoclimate reconstructions at local sites, and that our method for producing fine-grained paleoclimate
simulations is therefore suitable for conducting paleo-anthropological research. It is nonetheless important to calibrate the
GAM on a range of data encompassing the data to be downscaled. Otherwise, the SDM is likely to over-correct the coarse-
grain data. In addition, the bilinear and bicubic interpolation techniques were shown to distort either the temporal variability

or the values of the response variables, while the kriging method offered the best compromise. Since climate variability is an aspect of the environment to which human populations may have responded in the past the choice of interpolation technique is therefore an important consideration.

1 Introduction

5 The extent to which past climate change influenced human population dynamics during the course of prehistory is a subject of lively debate. The Last Glacial period, including Marine Isotope Stages 3 (MIS3) and 2 (MIS2) and the Last Glacial Maximum (LGM), is particularly interesting in the context of this debate (van Andel, 2003). During MIS3 the archaeological record suggests that modern human populations originating in Africa expanded into Eurasia, while Neanderthal populations gradually contracted their range before becoming extinct ~27,000 years Before Present (BP) (Serangeli and Bolus, 2008). Progressively
10 colder and drier conditions, culminating in the LGM (19,000 - 23,000 years Before Present), are thought to have triggered further range contractions and the demographic decline of modern human populations in Europe. Climate affects the spatial behaviour of human populations directly (when conditions exceed human physiological limits) and indirectly (when it affects the distribution of resources upon which humans depend). The global climate during the Last Glacial period was characterised by a series of rapid oscillations (known as Dansgaard-Oeschger, or D-O events). These events may have acted as forcing
15 mechanisms, affecting the demographic processes described above (e.g. Müller et al., 2011; Schmidt et al., 2012; Sepulchre et al., 2007; Jimenez-Espejo et al., 2007; Banks et al., 2013; d'Errico and Sánchez Goñi, 2003; Gamble et al., 2004; Shea, 2008). While the timing of climate events relative to large-scale patterns in the archaeological record is suggestive, the mechanisms by which climate forcing acted on human populations are still imperfectly understood. More empirical evidence is needed to validate the hypothesis that climate forcing affected human population dynamics and explore the nature and scale of the effect.

20

The broad demographic patterns mentioned above are the result of smaller, local-scale patterns produced by mobile groups of hunter-gatherers distributing themselves on the landscape in order to exploit available resources. The availability of these resources fluctuated both predictably (on a seasonal basis) and unpredictably (as a result of climate variability). It is by gaining an understanding of these smaller-scale patterns, ultimately, that we will be able to understand how climate forcing affects the
25 spatial and cultural dynamics of prehistoric human populations. Previous analyses of climate forcing have used a variety of data to reconstruct the paleoclimate, such as ice-core or marine records (e.g., Bradtmöller et al., 2012; Jimenez-Espejo et al., 2007; Schmidt et al., 2012), present-day climate data (e.g., Jennings et al., 2011), and climate model simulations (e.g., Banks et al., 2008; Davies and Gollop, 2003; Sepulchre et al., 2007; Benito et al. 2017; Hughes et al. 2007; Tallavaara et al. 2015). These analyses were conducted at varying spatial resolutions, typically on the order of ~50 km x 50 km (= 2500 km²). Higher-
30 resolution climate simulations are nonetheless necessary for the quantification of climate variability at an inter-annual scale and a spatial scale which approximates the size of the catchments within which hunter-gatherer groups typically forage (~10 km from camp, or 314 km²; Vita-Finzi and Higgs, 1970) making this an ideal spatial scale at which to consider the impact of climate variability on human systems.

Global Climate Models (GCMs) are able to simulate climate conditions at various spatial and temporal scales, whereas climate proxy data are inherently limited by the uneven distribution of sample locations and taphonomic biases (i.e. biases in the fossil record, such as pollen preservation, location of archaeological sites, etc.). GCMs use physical equations, e.g. to represent atmospheric fluid dynamics, as well as parameterisations, e.g. for sub-grid scale phenomena, to simulate the Earth's climate. The major disadvantage of GCMs is that they are computationally intensive and are usually only used to model climate behaviour at relatively coarse spatial resolution, typically coarser than 100 km (cf. Flato et al., 2013 for the latest details on the CMIP5 models) especially for long paleoclimatic simulations. Their ability to simulate the small-scale physical processes that drive local surface variables, such as precipitation, is therefore limited (Wood et al., 2004).

10

Regional Climate Models (RCMs) represent a physically based approach to climate modelling at a finer spatial scale over a specific region of interest (e.g., Liang et al., 2006, Flato et al., 2013). However, RCMs use GCM outputs to set their boundary conditions. They therefore require the explicit modelling of physical processes at both coarse and fine scales globally and over the region of interest, respectively and are also computationally demanding. Statistical Downscaling Methods (SDM), on the other hand, are less computationally demanding. SDMs proceed by empirically associating local-scale variables with large-scale atmospheric variables produced by GCMs and are faster to compute than mechanistic RCMs (Vaithinada Ayar et al., 2015). SDMs fall into four main families: “transfer functions”, which directly link large-scale and local-scale variables; “weather typing” methods based on conditioning statistical models on recurrent weather states; “stochastic weather generators” that simulate downscaled values from their (potentially conditional) probability density functions; and “Model Output Statistics” (MOS) methods based on adjusting (i.e., correcting) the statistical distribution of the large-scale GCM simulations in order to generate local-scale variables with the correct statistical properties (e.g., Vaithinada Ayar et al., 2015).

In this study, we explore and refine the capacity of an SDM from the transfer functions family, based on Generalized Additive Modelling (GAM), to compute temperature and precipitation time series at a fine spatial and temporal resolution for the LGM over Western Europe, south of the Fennoscandian ice-sheets. GAM is a non-parametric statistical technique that has proven reliable for capturing non-linear relationships between local- and large-scale variables and correcting the biases specific to a given GCM (e.g., Vrac et al., 2007; Levavasseur et al., 2010). The SDM used here accurately downscales the climatology (i.e., the climate averages over several decades) of temperature and precipitation generated by a GCM for the LGM when calibrated using present-day data (Vrac et al. 2007). Its ability to generate projections of the small-scale temporal patterns necessary to explain the spatial dynamics of prehistoric human populations is untested, however. In the present study, therefore, we use present-day climate data (corresponding to the average of the 1961-1990 period) extracted from the IPSL-CM5A-LR GCM (Dufresne et al., 2013) to calibrate the SDM, applying it to a 50 year-long time series of climate simulations for the LGM (Kageyama et al., 2013a, b). Interpolated values of coarse-grain variables extracted from the GCM, as well as fine-scale geographical data, such as elevation and advective continentality, are used as predictors in the GAM. The result is the

production of downscaled monthly values over 50 years for temperature and precipitation, including local temporal variability in temperature and precipitation rates. In addition, we compare the impact of three different interpolation techniques (bilinear interpolation, bicubic interpolation and kriging) on the downscaling results, evaluating the resulting SDM outputs with the aid of climate proxies (palynological and archaeozoological data) and observing the impact of each technique on patterns of spatial and temporal variability in temperature and precipitation. The SDM using kriging interpolation is demonstrated to be a good compromise between computational complexity and accuracy. Although a longer timeframe could be used to validate the method, potentially reducing the influence of interdecadal internal variability, we validated on an 11-year present-day time series distinct from the calibration period due to computational constraints.

10 The Materials and Methods section is organised as follows. First, we present the GCM to be downscaled, detailing its characteristics and biases. Next, we detail how the GAMs are constructed and how they are used to perform the downscaling. The calibration data used to parameterise the GAMs are presented, distinguishing between the dependent variables, the large-scale predictors drawn from the GCM and the fine-scale geographical predictors. Since the large-scale predictor variables must be interpolated at fine-scale before applying the GAMs, we evaluate the relative merits of three interpolation techniques. The procedure used for the calibration of the GAM and the selection of predictor variables is then detailed, followed by a description of the procedure used to downscale the large-scale temperature and precipitation data for the LGM generated by the GCM for each interpolation technique, generating spatial measures of variability. We then describe how we obtained the independent data used to evaluate the downscaling outputs for the LGM. Finally, we describe the present-day data outside of the calibration period used for validating the downscaling method using kriging interpolation, since it provides the best results for the LGM.

20 Downscaling GCM outputs to a fine spatial scale using interpolation and a GAM alters the values and potentially introduces bias. The main focus of this paper is to fine-tune the downscaling method and, to this end, we focus on the relative impact of three different interpolation methods on the downscaled time series which are the main focus of this research. The Results section is therefore organised as follows. First we describe and compare the splines generated by the calibration of the GAM for the three interpolation techniques. Understanding these differences is important because the splines are used to correct the different biases introduced by the three interpolation techniques. We then describe and compare the maps generated by the SDMs using the three interpolation techniques for temperature and precipitation and assess the results using the evaluation data. The differences between the three interpolation techniques for generating measures of temperature and precipitation variability are then explored. Finally, the results of the SDM using kriging interpolation are compared to the present-day validation data. These results and the potential pitfalls associated with this downscaling method are then detailed in the Discussion section.

2 Materials and Methods

2.1 Global Climate Model

The GCM used in this study is the ocean-atmosphere coupled model IPSL-CM5A-LR (Dufresne et al., 2013) developed for the CMIP5 (Taylor et al., 2012) and PMIP3 (Braconnot et al., 2012) projects and the 5th IPCC report (IPCC, 2013). The IPSL-CM5A-LR model has a spatial resolution of 1.9° in latitude and 3.75° in longitude over Europe, which is the area of interest here (i.e. ~62 500 km²). The model performance and main biases are described in Dufresne et al. (2013) and Hourdin et al. (2013). This model version is known to have a cold (-1.4°C) bias in terms of globally averaged temperature and the bias in mean annual temperature over Europe is similar to the global value. In this low-resolution version of the model, the mid-latitude westerly winds are generally more equatorial than observed (Hourdin et al., 2013) and this is the case for the Northeast Atlantic and Europe too. The general pattern of extra-tropical precipitation over the North Atlantic and European sectors is satisfactory with respect to the annual mean (Dufresne et al., 2013). The equilibrium temperature response to a doubling CO₂ is 3.59°C (Dufresne et al., 2013), which is rather high compared to other CMIP5 models. Nevertheless, the model's response to LGM forcings somewhat underestimates the cooling over Europe, as reconstructed from pollen, and is satisfactory in terms of precipitation (Kageyama et al., 2013).

We use a *historical* simulation run according to the CMIP5 protocol, and use model output for the period from 1961 to 1990 as our present-day reference climate. Outputs from this simulation are used in the calibration process (below). The simulation of LGM climate conditions follows the PMIP3 protocol (Braconnot et al., 2011, Braconnot et al., 2012). The concentrations of atmospheric greenhouse gases were lowered to their LGM values derived from ice core data (185 ppm for CO₂, 350 ppb for CH₄ and 200 ppb for N₂O) and the ice sheets are prescribed according to the product developed for PMIP3 (Abe-Ouchi et al., 2015). The model is run for several hundred years until the response to the LGM forcing in terms of surface climate variables is stabilised (Kageyama et al., 2013a, b). For this research, we extracted 50 years of monthly mean data (temperature 2m above the surface, precipitation, sea level pressure and relative humidity) from the stabilised part of the simulation. Next, we downscaled the data, calculated their average climatology and their temporal (interannual) variability.

2.2 Generalized additive models

Generalized additive models (GAM, Hastie and Tibshirani, 1990) are statistical models blending the properties of generalized linear models with additive models. Given a dependent variable Y and p predictor variables $[X_1, \dots, X_p]$, GAMs compute $E(Y|X_1, \dots, X_p)$, the expected value of Y , conditionally on the p predictors X_i , as a sum of non-parametric functions as follows:

$$E(Y|X_1, \dots, X_p) = \sum_{i=1}^p f_i(X_i), \quad (1)$$

Following Vrac et al. (2007), cubic spline functions were used for the f_i , represented by piece-wise third-order polynomial functions. For each function f_i , a number of knots are placed evenly throughout the predictor range, and the cubic polynomials that compose f_i are constrained to continuity conditions at each knot to ensure smooth transitions (Wood, 2000, 2004). GAMs

were calibrated using the mgcv package (Wood, 2011) in R, and the number of knots was determined automatically using generalised cross-validation.

Using a combination of geographical and physical predictor variables has been shown to improve spatial downscaling results (Vrac et al., 2007). The GAM uses these predictor variables, in addition to the original variables to be downscaled, to correct the biases of the GCM described above. This implies that a given GAM is only valid for the GCM for which it was calibrated, as it corrects its' specific biases. Using GAMs on climate variables requires the predictor and dependent variables to have the same spatial scale. The present-day dependent variables (precipitation and temperature) are at a fine spatial scale. The elevation variable is also at a fine spatial scale, whereas the predictor climate variables generated by the GCM are at a coarser scale. Thus, interpolation of the predictor climate variables is necessary. For this research we tested three interpolation techniques (see below).

Once the functions f_i have been fitted using the present-day data, the downscaling can be performed on the GCM outputs for the LGM. The downscaling uses fine-scale and interpolated predictor climate variables corresponding to the LGM to generate fine-scale dependent variables. Here, two GAMs are calibrated: one for temperature and one for precipitation.

2.3 Calibration data

2.3.1 Fine-scale dependent variables: the CRU climatology.

Fine-scale, present-day temperature and precipitation dependent data were obtained from the Climate Research Unit (CRU, New et al., 2002). The spatial resolution of the data is $10'$ (*i.e.* 1/6 degree), regularly gridded between 32.72° and 59.861° latitude (N = 164 values) and -11.578° and 24.738° longitude (N = 219 values) for a total of N = 35916 grid-points. We computed a monthly climatology for each gridpoint by averaging the variable of interest (temperature or precipitation) over 30 years (from 1961 to 1990) for each month (Fig. S1), resulting in 12 values for each cell. The GAM is calibrated over this 30-year climatology, because the GCM cannot be set up to generate predictor variables for a specific year.

We are specifically interested in downscaling temperature and precipitation over Western Europe, south of the ice-sheets during the LGM, a time when archaeological data indicates that human populations contracted in size and range. The GAMs were calibrated over a wider area than the region of interest in order to avoid edge effects and include the full range of climate conditions that prevailed during the LGM, which was much colder than the present day. As a result, the calibration domain extends further North and East (where more continental conditions prevail) than the region of interest. Preliminary simulations nonetheless showed that selecting too large a calibration region averaged-out the small-scale variation we are interested in.

2.3.2 Large-scale predictor variables.

We used the data from a CMIP5 historical simulation run with the IPSL-CM5A-LR model to produce the predictor climate variables used for the calibration of the GAMs. We calculated monthly climatological averages from the simulation outputs for the period from 1961 to 1990, i.e. the same years as the CRU data (see above). The predictor variables we used are: temperature (T), precipitation (P), atmospheric pressure at sea level (SLP) and relative humidity (RH). The variables were spatially interpolated to match the spatial resolution of CRU data, which is 10° ; each grid-point in the CRU data therefore matches a value for each of the predictor variables. Three interpolation methods were tested: bilinear, bicubic and kriging (Figs. S2-S4). For the kriging method, we used the “krig” function from the vacumm python package (<http://relay.actimar.fr/~raynaud/vacumm/>) using an exponential fit of the variogram, with the fit computed independently for every month and every variable. Different interpolation methods generate differences in the fine-scale predictor data. For example, the bicubic interpolation generates values outside of the initial range of values, contrary to the other two techniques. The bilinear interpolation generates abrupt changes in the slope of the values at the initial data points, whereas the other two techniques generate smooth surfaces. It is therefore important to assess the potential impact of the interpolation method on the output of the downscaling process.

2.3.3 Fine-scale predictor variables.

We extracted present-day elevation data from the CRU dataset gridded at the same fine-scale spatial resolution as the dependent variables. We computed the advective (Aco) and diffusive (Dco) continentalities, following Vrac et al. (2007). Dco is bounded between 0 and 1, and corresponds to the shortest distance to the ocean. A low value means that distance to the ocean is small, and *vice versa*. Aco is also bounded between 0 and 1, and takes the direction and intensity of prevailing winds into account along with the distance to the ocean. The change of Aco during a time dt is governed by Equation 2:

$$dAco = [-Aco(1 - i_{co}) + (1 - Aco)i_{co}] \frac{U}{U_0} \ln 2 \frac{dx}{U_0} \quad (2)$$

where i_{co} is 0 over sea and 1 over land, dx is the distance traveled by the air mass during dt , U is the mean wind norm, obtained from the GCM, and l_0/U_0 is the distance/wind ratio corresponding to a change of Aco of 1/2. Both variables are used to account for the fact that an air mass becomes more continental as it travels across land. Since Dco and the Aco proved to be highly correlated but Aco provided the best performance in the models, we only selected Aco for the present analysis (Figs. S2-S4).

2.4 Calibration of the GAMs

For each dependent variable (temperature and precipitation) and for each interpolation technique (bilinear, bicubic and kriging), we tested different combinations of physical and geographical predictor variables. To downscale temperature, we computed GAMs for all possible combinations of coarse-grain temperature values from the GCM interpolated at fine scale, with fine-grain elevation and advective continentality (Aco), resulting in seven possible models for each interpolation. To

downscale precipitation, we computed GAMs for all possible combinations of coarse-grain temperature, coarse-grain precipitation, sea-level pressure, and relative humidity values from the GCM interpolated at fine scale, with fine-grain elevation and advective continentality (Aco), resulting in 31 possible models for each interpolation. For each interpolation technique, the resulting GAMs were compared using the Akaike Information Criterion (AIC; Akaike, 1974), and the model with the lowest AIC was selected. The AIC is a measure of the relative goodness of fit of each of the models and penalizes the number of parameters, thus preventing overfitting. The significance of each variable was assessed using p-values, and verified by visual inspection of the spline 95% confidence intervals. Six GAMs were therefore retained after calibration (one for each response variable and for each interpolation technique; Table S1).

2.5 Downscaling of temperature and precipitation time series for the LGM

We computed downscaled temperature and precipitation values using the six GAMs resulting from the calibration process described above, i.e. for the same predictor variables. The large-scale climate variables were generated by the GCM using the PMIP3 protocol for the LGM prior to interpolation (Figs. S5-S7). The geographical variables are derived from a digital elevation model for the LGM (Levvasseur et al., 2011). In particular, the change in coastlines due to the lower sea-level at LGM is accounted for, which has an impact on the continentalities. The downscaling was performed for each month of the 50-year-long monthly output from the GCM, in order to obtain a long time series of fine-scale temperature and precipitation values over Europe and calculate climatological averages. We also calculated indices of variability, including measures of variance and inter-annual variability for the variables of interest. The standard deviation of monthly mean temperatures for each month was calculated for the 50-year run. The coefficient of variation of monthly mean precipitation values was calculated for the same period.

2.6 Evaluation data (palynological data and vertebrate remains)

To evaluate the performance of the SDM for the LGM we compared our temperature and precipitation outputs to local climate variables estimated on the basis of pollen and vertebrate fossils from 29 test locations (Table S1, Fig. 1). For each of our 29 test sites, we estimated the mean, minimum, and maximum temperature and precipitation rate on a monthly basis over the course of the 50 downscaled years, and compared the ranges of downscaled values to the ranges of temperature and precipitation values reconstructed using the palynological data and vertebrate remains.

Reconstruction of local temperature and precipitation values (annual mean temperature, mean temperature of the coldest month, mean temperature of the warmest month, mean annual precipitation, precipitation in January, precipitation in July) were obtained from pollen data reported in an independent study using inverse vegetation modelling for 14 sites located in Europe (Wu et al., 2007). For the remaining 19 sites, vertebrate remains from another study (Burke et al. 2014) were used to calculate bio-climate indices (BCI; ff. Hernandez Fernandez, 2001a, b). The method set forth by Hernandez-Fernandez uses

large and small vertebrates to compute the relative probability that a given assemblage reflects one of Walter's nine global zonobiomes (Walter and Box, 1976). The method is based on the "climate envelope" method commonly employed in biogeographical reconstructions. The BCI uses presence/absence data, thus avoiding the problems inherent with calculating the relative representation of species from the archaeozoological record, and all available taxa rather than one or two "indicator" species, thus avoiding the risk that changes in the distribution of a single taxon could bias the biogeographical reconstruction. Ranges of temperature and precipitation values (minimum, mean and maximum) for each zonobiome were estimated by mapping the modern distribution of zonobiomes in the northern hemisphere and compiling present-day temperature and precipitation data from the CRU data (see Burke et al. 2014). The zonobiomes (calculated using BCI) were then used to predict the climate ranges for each test location. Note that the range of values for each zonobiome corresponds to the minimum and maximum values for the zonobiome over Western Europe (see Burke et al. 2014). It is therefore not specific to a given location and encompasses a large range of values. The intervals generated by these two different climate reconstruction methods, therefore, are not equivalent. Nevertheless, they provide useful references for evaluating the values generated by the SDM.

2.7 Validation using present-day data (CRU 1950-1960 time series)

Due to computational constraints, the validation of the SDM was performed for the kriging interpolation technique only on an 11-year period outside of the calibration period, based on the comparison of the results of the downscaling for the LGM between the three interpolation techniques (see below). A GCM simulation was produced for the period from 1950-1960 and compared with a time series of temperature and precipitation at 10' resolution over Europe for the same period (Mitchell et al. 2004). Yearly averages and variability indices (standard deviation for temperature and coefficient of variation for precipitation) for the two sets of data were compared. The time series was based on the same original data used to create the 1961-1990 climatology, which forms the calibration set for the SDM (New et al. 1999). The time series was created by spatially interpolating data from irregularly spaced climate stations using thin-plate smoothing splines, however, and may be subject to its own bias. Potential differences between our results and the validation time series should therefore be interpreted with caution.

3. Results

3.1 The GAM

The best models (i.e., the models with the lowest AIC value) for temperature and for precipitation were obtained by using the same sets of variables (one for temperature, one for precipitation) in the GAM for the three interpolation techniques. The predictors for temperature are: simulated temperature from the GCM, elevation and advective continentality (explaining 95.80%, 95.29% and 95.63% of the variance for the bilinear, bicubic, and kriging interpolations, respectively). For precipitation, the predictors are: simulated temperature, precipitation, sea-level pressure and relative humidity from the GCM,

elevation, and advective continentality (explaining 64.65%, 64.79% and 65.43% of the variance for the bilinear, bicubic, and kriging interpolations) (Table 1). The p-values for all variables for all models were < 0.001 .

The splines resulting from the calibration process for the downscaling of temperature values show that fine-scale temperature readings are related to the GCM temperature and to elevation in a linear fashion, and the differences between the three interpolation techniques tested are negligible (Fig. 2). The fine-scale temperature is proportional to the GCM temperature but it is inversely proportional to elevation, which means that the coarse-grain temperatures generated by the GCM are higher than fine grain observations in regions of high elevation. This is expected because temperature generally decreases with increasing altitude and because in the coarse grain GCM, it is the average elevation over the grid box that is considered. Although the model including all three predictor variables produced the lowest AIC, advective continentality has a very limited impact on temperature, as the values of $f(\text{Aco})$ remain close to 0. When applied outside the range of values for which they are calibrated, GAMs use a linear extrapolation of the splines. The range of values for elevation is similar for the present-day and the LGM (Fig. 3). Because of the increased land mass during the LGM (which correlates with a low sea-stand), there are more high values for advective continentality, but this difference has a small impact since the spline is relatively flat for this variable. As expected, temperature is lower during the LGM than for present-day. This has a limited impact on the projections because the linear interpolation of the spline outside of the range of values used for calibration is consistent with the fact that the spline is relatively linear for temperature and the few remaining values are within 10 degrees of the minimum temperature. For very low temperatures during the LGM, however, the SDM outputs should be interpreted carefully, as discussed below.

The splines for the downscaling of precipitation, in contrast, are non-linear (Fig. 4). The splines showing the influence of temperature on expected precipitation rates show larger variations due to a low expected precipitation for both low and high temperatures, but high expected precipitation for middle-range temperatures. Although the expected precipitation increases monotonically with the interpolated precipitation rates, the spline values are lower than the GCM precipitation values and the relationship is non-linear. The expected precipitation increases more rapidly for low than for high interpolated precipitation, in keeping with previous observations that GCMs (and even RCMs) overestimate drizzles, which may explain this correction (e.g. Gutowski et al., 2013). The three interpolation techniques produced similar splines for all variables, although the splines of the bicubic interpolation are slightly distinct from the other two interpolation techniques. The main difference is observed for the spline of the bicubic relative humidity, which indicates lower precipitation for low relative humidity than the other two interpolation techniques. Differences between the splines of the bicubic interpolation and the other two techniques were expected, since this interpolation generates the most divergent values (Figures S8, S11). As a GAM will adjust the splines to compensate for the potential biases of a GCM, it will also do so to compensate for the specificity of an interpolation technique.

Advective continentality is the variable with the least impact on precipitation rates. The ranges of values for simulated precipitation, relative humidity, and for elevation are similar for the present-day and the LGM periods, and the distributions

of the variable substantially overlap, indicating that the splines calibrated over the present-day period can apply for the LGM (Fig. 3). As for temperature, the spline of advective continentality is relatively flat, and the difference of range of values will have limited impact on the projections. The spline for the simulated atmospheric pressure at sea level has a positive slope for high values. This spline does not represent a causal relationship, but simply indicates that the GCM tends to underestimate precipitation at high atmospheric pressure. The simulated atmospheric pressure at sea level is also higher for the LGM than for the present-day period. Nonetheless, since the atmospheric pressure is mostly lower than 1045 hPa during the LGM, and given the low slope of the spline on the right-hand extremity, this discrepancy should have little impact on the results. For temperature, the splines are relatively linear near the lower end of the range of present-day values, and the linear interpolation of the spline at lower values for the LGM is therefore sensible.

10 3.2 Results for the LGM

3.2.1 Temperature

Downscaled annual mean temperature was very similar for the three interpolation techniques tested (Fig. 5). This was expected, since the splines for the GCM temperature and elevation for all three techniques are also very similar. Temperatures interpolated with the bilinear and kriging techniques were more similar to each other than to the temperature interpolated using the bicubic technique before (Fig. S8) and after (Fig. 5) applying the GAMs. The differences between the bilinear and kriging techniques show a pattern corresponding to the original coarse-grain cells from the GCM. This illustrates the difference between the two interpolation techniques: kriging generates smoother variations than the bilinear interpolation, which generates discontinuous variations at the original points. This difference remained after applying the GAMs, showing the impact of the interpolation technique used on the final outcome of the downscaling.

20 The main differences between the interpolated and downscaled temperatures occur in the northeast of Europe (Fig. S9), where downscaled temperatures are higher, especially in winter. This difference was also observed when comparing present-day CRU data with the interpolated GCM data (Fig. S10), although with a much lower amplitude. Northeast Europe is the coldest region of the study area for both the LGM and the present-day (Figs. S1-S7). Since the spline for temperature has a slope lower than 1 for low temperatures (Fig. 2), the GAM generates higher temperatures than the interpolated values, especially for very low temperatures which fall outside of the present-day range of values due to the linear interpolation of the spline. As a result, the difference between interpolated and downscaled temperatures during the LGM is lower in summer. The SDM also takes fine scale variations in topography into account, such as abrupt elevation changes in the Alps and Pyrenees (Fig. S9).

30 The range of temperatures for the 19 sites for which the BCIs were computed are in accordance with the temperature reconstructions, irrespective of the interpolation technique used (Fig. 6). Simulated temperature ranges fall within the reconstructed ranges corresponding to the BCIs and are within the reconstructed ranges from Wu et al. (2007) for all test sites,

as shown by the overlap of the red and blue error bars with the diagonal (Fig. 6a,c,e). As noted above, since the BCI reconstruction considers the minimum, mean and maximum values for each zonobiome over the whole of Western Europe (see Burke et al. 2014), some downscaled temperature values may differ from the mean, but as long as the error bars overlap with the diagonal are still in accordance with the BCI climate reconstruction.

5 3.2.2 Precipitation

The three interpolation techniques tested here produce similar distributions of precipitation rates but, compared to the two other interpolation techniques, the bicubic interpolation produced the most divergent results (Fig. S11). All three interpolation techniques nonetheless reflect the biases of the GCM (Fig. S12). Precipitation rates predicted using bicubic interpolation were substantially lower than those predicted using the other techniques in the South-West of Europe in winter and higher in the North of Europe (over the current North Sea) in summer (Fig. 7). This result is consistent with the observation that the splines for the bicubic interpolation differed from the other two interpolation techniques. Despite a general agreement between simulated and observed annual precipitation mean over Europe (Dufresne et al. 2013), comparing GCM projections interpolated at fine scale with the present-day CRU data shows that the GCM overestimates precipitation over the South of Europe in winter and underestimates them over the North in summer (Fig. S13), which results in the non-linear spline for precipitation (Fig. 4). Coarse-grain GCMs are known to perform poorly when simulating the small-scale physical processes that drive local surface variables such as precipitation (Wood et al., 2004). This explains the discrepancies between the present-day simulations and the CRU data and, by extension, explains the adjustments performed by the SDM.

The precipitation ranges for the 19 sites for which the BCIs were computed are in accordance with the precipitation ranges reconstructed for the LGM. Simulated precipitation ranges for all sites fall within the reconstructed ranges corresponding to the BCIs, as shown by the overlap of the horizontal error bars (in red) with the diagonal (Fig. 8). The SDM predicts higher precipitation values for 1 site (in North-West Iberia) than the reconstructions provided by Wu et al. (2007) (the simulated mean precipitation of the driest month was predicted to be higher than the maximum precipitation found by Wu et al.). However, while the BCI ranges correspond to minimum and maximum values over a relatively large spatial extent, the reconstructions offered by Wu et al. (2007) are site-specific and therefore produce smaller ranges of values. Moreover, Wu et al. (2007) based their reconstructions on local adjustments of the biome estimates from the BIOME4 model (Kaplan et al. 2003). They therefore used the same initial values for different sites, which may underestimate differences between sites and explain the lower range of precipitation values compared to the values generated by the SDM.

3.2.3 Variability

The temporal variability of temperature and precipitation rates highlights differences between the three interpolation techniques. When temporal variation was computed over the interpolated variables (Figs. S14, S15), bilinear interpolation

displays regular spatial patterns for both temperature and precipitation, especially in summer. This pattern was less apparent for kriging, and almost absent for bicubic interpolation.

5 With bilinear interpolation, the interpolated values will necessarily be less variable (whatever the index used) than the original values. For a simple linear interpolation, given 2 spatially consecutive values at 2 different points in time ($y(x_0, t_0)$, $y(x_0, t_1)$, $y(x_1, t_0)$ and $y(x_1, t_1)$), any linearly interpolated value $y(x_i)$ for a location x_i in $[x_0, x_1]$ will necessarily be comprised in $[y(x_0), y(x_1)]$. In addition, due to the linear interpolation $|y(x_i, t_1) - y(x_i, t_0)| < |y(x_0, t_1) - y(x_0, t_0)|$ and $|y(x_i, t_1) - y(x_i, t_0)| < |y(x_1, t_1) - y(x_1, t_0)|$. By contrast, since kriging does not impose linear interpolation between x_0 and x_1 , this relation does not necessarily hold, and even less for bicubic interpolation, which do not impose $y(x_0) \leq y(x_1) \leq y(x_1)$. However, because of this absence of restriction, bicubic
10 interpolation can generate values with a high variability, as for precipitation in summer in the South-West of the Iberian peninsula (Fig. S15; the high variability for kriging in winter only occurs at the boundary of the study area due to boundary conditions, and can therefore be discarded).

Computing temporal variation over the downscaled variables (Figs. 9, 10), showed that the GAMs attenuated this spatial
15 pattern, which nonetheless still occurred for the bilinear interpolation, and was almost non-existent for the other two interpolation techniques.

3.3 Validation on present-day data

The comparison of average downscaled (based on kriging interpolation) and observed (CRU) monthly temperature and daily precipitation for the 1950-1960 period shows that they are in good agreement (Figures S16, S17). For temperature, the main
20 difference occurs in the far North of the study area (Southern point of the Scandinavian peninsula), and on the Italian side of the Alps. The downscaled temperature is similar on both sides of the Alps, whereas there is some difference in the CRU data (Figure S16), suggesting the inclusion of orographic wind as a predictor may improve the SDM for specific areas. For precipitation, the main difference occurs in areas of high precipitation, especially the Western coast of Great Britain (Figure S17). Nonetheless, these areas were the areas with higher levels of precipitation for both datasets.

25 The overall patterns of variability were overall similar between the downscaled and CRU datasets (Figures S18, S19), with nonetheless some local differences. Temperature variability was higher in the North-East region of the study area and lower in the South-West region, especially in winter (Figure S18), and the range of values for the standard deviation were very similar for the different seasons. Downscaling tended to slightly overestimate temperature variability in the North of the study area, and underestimate it in the North-East compared to the CRU data. Some small-scale differences are nonetheless difficult to
30 interpret, since the CRU data showed some spatial artifact, for example in the center of the Iberia peninsula. The amplitude of the variability values was more different for precipitation, with the variability observed in the CRU data being higher (Figure

S19). Nonetheless, the general patterns were quite similar, with the Southern region of the study area having higher precipitation variability than the North for both the downscaled and the CRU datasets.

4 Conclusion and Discussion

We downscaled temperature and precipitation values produced by the IPSL-CM5 model for 50 simulated years over Western Europe during the last glacial maximum using a GAM, a computationally efficient method for downscaling GCMs (Vrac et al. 2007). A single GAM was used for each dependent variable, calibrated over an average of 30 years of present-day data. Comparing the outputs of the SDM with two different climate reconstructions showed that this method generates results that fall within the computed confidence intervals for the variables of interest. This enabled us to compute indices of climate variability for the LGM in Western Europe. In a separate study, we were then able to test a suite of environmental predictors and demonstrate that climate variability is a key factor governing the spatial distribution of prehistoric human populations during the LGM (Burke et al. 2014, 2017).

Downscaled time series for a present-day period (1950-1960) falling outside of the calibration period (1961-1990) were in good agreement with an independent time series for both averaged values and measures of variability. Our study, therefore, demonstrates that the SDM, originally designed to downscale climatology data (averaged over several decades), can be applied to a time series thus allowing us to compute spatio-temporal patterns at a fine scale appropriate for studying the spatial dynamics of prehistoric human populations. SDMs must be carefully parameterised, however, including selecting the appropriate size of the area used for calibration.

Overall, the downscaled temperature and precipitation values produced by the SDM are in agreement with the climate reconstructions obtained from vertebrate remains and palynological data, with few exceptions (Figures 6, 8). The SDM results for the LGM differ from the interpolated data in the northeast of the study area, reflecting the adjustments made in the GAM to counter biases inherent in the IPSL-CM5A-LR GCM used in this study. These discrepancies have little consequence in the present study, since this region was covered by ice-sheets during the LGM. It was included for calibration because it represented present-day climate conditions that were close to those of the Southern part of the study area during the LGM. This region was also included in the downscaling to illustrate the fact that, since it was colder during the LGM than any present-day region of the study area, results for this region should be interpreted with caution.

The choice of interpolation technique used in the SDM also proved critical as it has a strong impact on the distribution of climate variability. We tested three different interpolation techniques. Since GAMs require the predictor and the dependent variables to have the same spatial grain, bilinear interpolation is commonly used to downscale the coarse-grain data generated by GCMs (Vrac 2007). However, as this research demonstrates, bilinear interpolation generates non-smooth surfaces which may cause spatial artifacts in the final output. We tested two other non-linear interpolation techniques which generate smoother

surfaces: bicubic interpolation and kriging. Bicubic interpolation generates values outside of the initial range of values (and therefore under- or overestimates the values) but is faster to apply than kriging. Kriging is more computationally demanding but offers the advantage of constraining the interpolated values within the range of initial values. The three interpolation techniques produced different results for both temperature and precipitation during the LGM (Figs. S8, S11). After applying the GAM, these differences were especially important for precipitation values (Fig. 7). Because the GCM generated coarse grain temperature values for present-day conditions which are highly correlated with the CRU data, all three interpolation techniques produced similar linear splines and led to similar results for this variable. In the case of precipitation, however, bicubic interpolation predicts drier environments than the other two techniques by up to 2 mm/day. Since GCMs operate at grains that are too coarse to accurately model small-scale physical processes driving local surface variables (Wood et al., 2004), the SDM for precipitation relies on more variables than are required to model temperature. The splines for these variables are non-linear (Figure 4), however, and may exacerbate the differences between the bicubic interpolation and the other two techniques. The cumulative impact of the interpolation and the GAM can therefore be non-negligible. This highlights the utility of the comparison presented in this research, especially for local phenomena such as precipitation.

The variability maps produced when using bilinear interpolation show the presence of a spatial artefact, in the form of a regular grid, for both temperature and precipitation (Figs. S14, S15). This artefact reflects the fact that bilinear interpolation generates lower variability between the points from which the interpolation is performed. Although slightly attenuated, this artefact remained after applying the GAMs (Figs. 9, 10). Prior to applying the GAMs, bicubic interpolation produced maps with the smallest level of artifacts, kriging was intermediate and bilinear interpolation produced the highest level of artifacts. However, bicubic interpolation sometimes generated unrealistically high variability for precipitation (Fig. S15) while the artifacts generated by kriging decreased after applying the GAMs (Figs. 9, 10). We conclude that although more computationally demanding than the other two techniques, kriging represents a good compromise between computational complexity and accuracy. Contrary to bicubic interpolation, kriging generates values within the range of the values generated by the GCM and generates variability indices with more realistic patterns than the bilinear interpolation. We therefore recommend using kriging for SDM applications based on the method presented here.

The IPSL-CM5A-LR GCM is known to predict lower temperatures than the values observed at high latitudes in winter (Dufresne et al., 2013). This bias was indeed observed when comparing the interpolated temperature with the CRU present-day data. As a result, the spline for temperature has a shallow slope at low temperatures (Fig. 2). The resulting correction applied by the GAM was emphasised for the LGM data generated by the GCM in winter in the North of Europe (Fig. S9), which lie outside of the range of present-day temperature and therefore relied on a linear interpolation of the spline. The large differences in temperature are therefore likely to be a combination of an underestimation of temperature by the GCM, and an over-correction of the very low temperature by the SDM. The spatial domain used to calibrate the GAM is larger than the domain of interest, namely Western Europe south of the ice sheets (the region occupied by human populations during the

LGM) for reasons discussed above. These include the necessity of avoiding edge effects and including the full range of climate conditions likely to have occurred during the LGM. The observed over-correction lies on the periphery of the calibration region and is not within the study region this SDM was designed for.

5 The calibration area selected should therefore be large enough to encompass a representative range of climate conditions but should overlap the study region in order to account for potential relationships between climate and geographical variables specific to the region. However, through trial and error we found that using too large a calibration region averages out these relationships and therefore runs counter to the objectives of the downscaling, which is to represent fine scale spatial variation. Further research into the impact of the size of the calibration region on the SDM would be an interesting avenue to pursue. In
10 addition, some results presented here are probably highly influenced by the specific GCM that was used. Especially, the variability in the SDM results is strongly influenced by the variability of the original GCM, in addition to the choice of the interpolation technique (Figure 9, S14, 10, S15). For temperature, for example, for a given interpolation technique, the downscaling adjusts the interpolated GCM temperature based on elevation, which is constant for a given location. The SDM will therefore not change variability compared with interpolation of the GCM. Using ensembles of models increases confidence
15 in climate projections by enabling a better quantification of such uncertainty (Tebaldi & Knutti 2007). Although the outputs of ensembles of models may be challenging to interpret, this is another promising avenue for improving the application of the SDM method presented here that should be pursued in the future, especially for the computation of variability indices.

Our goal in this research has been to develop and test tools for the production of climate simulations at suitable spatial and
20 temporal scales for investigating the mechanisms through which climate change and climate variability may have affected human populations in the past. Our aim is to help explain some of the broad evolutionary patterns visible in the archaeological record. Our results demonstrate the potential of GAMs for the production of climate simulations at a fine scale of resolution, both spatially and temporally, at relatively low computational cost. The resulting climate simulations can be used to test human decision-making at regional and local scales useful for investigating the spatial distribution of prehistoric populations against
25 a backdrop of inter and intra-annual climate variability (e.g., Burke et al. 2017; Burke et al. 2014).

Code and data availability. The code used for the downscaling and the input and output data are available at at <https://figshare.com/s/1b952e47ff274cc0687e> (DOI: 10.6084/m9.figshare.5487145).

30

Author contributions. GLa, AB, MK, MV and GLe conceived the study. GLa and GLe implemented the R code for the downscaling. MK, MV and GR ran the IPSL-CM5A-LR simulations and implemented the code for the interpolations. CD implemented the code for the computation of the continentality. GL wrote the paper and all commented on it.

5 *Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. This research was supported by the *Fonds Québécois de Recherche Société et Culture - Programme de soutien aux équipes* (demande no.179537). Jérôme Servonnat is thanked for his input on the IPSL-CM5A-LR model biases.

10 The LSCE team thanks Armel Martin for his patient and careful work on the downscaling method, quite some years ago.

References

- Abe-Ouchi, A., Saito, F., Kageyama, M., Braconnot, P., Harrison, S. P., Lambeck, K., Otto-Bliesner, B. L., Peltier, W. R., Tarasov, L., Peterschmitt, J.-Y., and Takahashi, K. Ice-sheet configuration in the CMIP5/PMIP3 Last Glacial Maximum experiments, *Geosci. Model Dev.*, 8, 3621-3637 2015.
- 15 Akaike, H. A new look at the statistical model identification. *IEEE T. Automat. Contr.*, 19(6), 716-723, 1974.
- Banks, W.E., d'Errico, F., Peterson, A.T., Kageyama, M., Sima, A., and Sánchez-Goñi, M.F. Neanderthal extinction by competitive exclusion. *PloS ONE*, 3, e3972, 2008a.
- Banks, W.E., d'Errico, F., Peterson, A.T., Vanhaeren, M., Kageyama, M., Sepulchre, P., Ramstein, G., Jost, A., and Lunt, D. Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. *J. Archaeol. Sci.*, 35, 481-491, 2008b.
- 20 Benito, BM., Svenning, J.-C., Kellberg-Nielsen, T., Riede, F., Gil-Romera, G., Mailund, T., Kjaergaard, P.C., and Sandel, B.S. The ecological niche and distribution of Neanderthals during the Last Interglacial. *J. Biogeogr.*, 44(1):51-61, 2017.
- Braconnot, P., Harrison, S. P., Otto-Bliesner, B., Abe-Ouchi, A., Jungclaus, J., and Peterschmitt, J.-Y. The Paleoclimate Modeling Intercomparison Project contribution to CMIP5, CLIVAR Exchanges No. 56, Vol. 16, No.2, May 2011, pp 15-19,
- 25 2011
- Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y. Evaluation of climate models using palaeoclimatic data, *Nat. Clim. Change*, 2(6), 417-424, 2012.
- Bradt Möller, M., Pastoors, A., Weninger, B., and Weniger, G. C. The repeated replacement model—rapid climate change and population dynamics in Late Pleistocene Europe. *Quatern. Int.*, 247, 38-49, 2012.

- Burke, A., Levvasseur, G., James, P.M.A., Guiducci, D., Izquierdo, M., Bourgeon, L., Ramstein, G., and Vrac, M. Exploring the impact of climate variability during the Last Glacial Maximum on the pattern of human occupation of Iberia. *J. Hum. Evol.*, 73:35-46, 2014.
- Burke, A., Kageyama, M., Latombe, G., Fasel, M., Vrac, M., Ramstein, G., and James, P.M.A. Risky business: The impact of climate and climate variability on human population dynamics in Western Europe during the Last Glacial Maximum. *Quaternary Sci. Rev.*, 164:217-229, 2017.
- Davies, W. and Gollop, P. The human presence in Europe during the Last Glacial Period II. Climate tolerance and climate preferences of Mid-and Late Glacial hominids. Neanderthals and modern humans in the European landscape of the last glaciation, 131-146, 2003.
- 10 Delgado Huertas, A., Iacumin, P. and Longinelli, A. A stable isotope study of fossil mammal remains from the Paglicci cave, southern Italy, 13 to 33 ka BP: palaeoclimatological considerations. *Chem. Geol.*, 141(3-4), p. 211-223, 1997.
- Dufresne, J.-L., Foujols, M.A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benschila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E.,
- 15 Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, F. G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., and Viovy, N.. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dynam.*, 40, 2123-2165, 2013.
- 20 Finlayson, C., Giles Pacheco, F., Rodriguez-Vidal, J., Fa, D.A., Maria Gutierrez Lopez, J., Santiago Perez, A., Finlayson, G., Allue, E., Baena Preysler, J., Caceres, I., Carrion, J.S., Fernandez Jalvo, Y., Gleed-Owen, C.P., Jimenez Espejo, F.J., Lopez, P., Antonio Lopez Saez, J., Antonio Riquelme Cantal, J., Sanchez Marco, A., Giles Guzman, F., Brown, K., Fuentes, N., Valarino, C.A., Villalpando, A., Stringer, C.B., Martinez Ruiz, F., and Sakamoto, T. Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* 443, 850-853, 2006.
- 25 Finlayson, C., Fa, D.A., Jiménez Espejo, F., Carrión, J.S., Finlayson, G., Giles Pacheco, F., Rodríguez Vidal, J., Stringer, C., and Martínez Ruiz, F. Gorham's Cave, Gibraltar--The persistence of a Neanderthal population. *Quatern. Int.*, 181, 64-71, 2008.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. Evaluation of Climate Models. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- Gamble, C., Davies, W., Pettitt, P., and Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Philos. T. Roy. Soc. B*, 359(1442), 243-254, 2004.

- Gutowski Jr, W. J., Decker, S. G., Donavon, R. A., Pan, Z., Arritt, R. W., and Takle, E. S. Temporal–spatial scales of observed and simulated precipitation in central US climate. *J. Climate*, 16(22), 3841-3847, 2003.
- Guttman, NB. Accepting the Standardized Precipitation Index: a calculation algorithm. *J. Amer. Water Resour. Assoc.* 35(2), 311-322, 1999.
- 5 Hastie, T. J. and Tibshirani, R. J., 1990. *Generalized Additive Models*, Chapman and Hall.
- Hernandez Fernandez, M. Análisis paleoecológico y paleoclimático de las sucesiones de mamíferos del Plio-Pleistoceno ibérico, *Biologie*. Universidad Complutense de Madrid, Madrid, p. 368, 2001a.
- Hernandez Fernandez, M. Bioclimatic discriminant capacity of terrestrial mammal faunas. *Global Ecol. Biogeogr.*, 10, 189-204, 2001b.
- 10 Hourdin, F., Foujols, M. A., Codron, F., Guemas, V., Dufresne, J. L., Bony, S., Denvil, S., Guez, L., Lott, F., Ghattas, J., Braconnot, P., Marti O., Meurdesoif, Y. And Bopp, L. Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model. *Climate Dynamics*, 40(9-10), 2167-2192, 2013.
- Hughes, J.K., Haywood, A., Mithen, S.J., Sellwood, B.W., and Valdes, P.J. Investigating early hominin dispersal patterns: developing a framework for climate data integration. *J. Hum. Evol.*, 53(5):465-474, 2007.
- 15 Huntley, B., Alfano, M. J., Allen, J. R., Pollard, D., Tzedakis, P. C., de Beaulieu, J. L., Gröger, E., and Watts, B.. European vegetation during marine oxygen isotope stage-3. *Quaternary Res.*, 59(2), 195-212, 2003.
- Iacumin, P. and Longinelli, A. Relationship between $\delta^{18}O$ values for skeletal apatite from reindeer and foxes and yearly mean $\delta^{18}O$ values of environmental water. *Earth Planet. Sc. Lett.*, 201(1), p. 213-219, 2002.
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 2013.
- 20 Jennings, R., Finlayson, C., Fa, D., and Finlayson, G. Southern Iberia as a refuge for the last Neanderthal populations. *J. Biogeogr.*, 38(10), 1873-1885, 2011.
- 25 Jiménez-Espejo, F. J., Martínez-Ruiz, F., Finlayson, C., Paytan, A., Sakamoto, T., Ortega-Huertas, M., Finlayson, C., Ijima, K., Gallego-Torres, D., and Fa, D. Climate forcing and Neanderthal extinction in Southern Iberia: insights from a multiproxy marine record. *Quaternary Sci. Rev.*, 26(7), 836-852, 2007.
- Kageyama, M., Braconnot, P., Bopp, L., Caubel, A., Foujols, M.-A., Guilyardi, E., Khodri, M., Lloyd, J., Lombard, F., Mariotti, V., Marti, O., Roy, T., and Woillez, M.-N. Mid-Holocene and Last Glacial Maximum climate simulations with the
- 30 IPSL model. Part I: comparing IPSL_CM5A to IPSL_CM4. *Clim. Dynam.*, 40, 2447-2468, 2013a.
- Kageyama, M., Braconnot, P., Bopp, L., Mariotti, V., Roy, T., Woillez, M.-N., Caubel, A., Foujols, M.-A., Guilyardi, E., Khodri, M., Lloyd, J., Lombard, F., and Marti, O. Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model. Part II: model-data comparisons. *Clim. Dynam.*, 40, 2469-2495, 2013b.

- Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and Lozhkin, A. V. Climate change and arctic ecosystems II: Modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res.*, 108(D19), 8171, 2003.
- 5 Levvasseur, G., Vrac, M., Roche, D. M., Paillard, D., Martin, A., and Vandenberghe, J. Present and LGM permafrost from climate simulations: contribution of statistical downscaling, *Clim. Past*, 7, 1225–1246, 2011.
- Liang, X.-Z., Pan, J., Zhu, J., Kunkel, K. E., Wang, J. X. L., and Dai, A. Regional climate model downscaling of the u.s. summer climate and future change, *J. Geophys. Res.: Atmospheres (1984–2012)*, 111(D10), 2006.
- Longinelli, A., Iacumin, P., Davanzo, S., and Nikolaev, V. Modern reindeer and mice: revised phosphate–water isotope equations. *Earth Planet. Sc. Lett.*, 214(3–4), p. 491-498, 2003.
- 10 López-García, J. M., Blain, H. A., Cuenca-Bescós, G., Ruiz-Zapata, M. B., Dorado-Valiño, M., Gil-García, M. J., Valdeolmillos, A., Ortega, A.I., Carretero, J.M., Arsuaga, J.L., Bermúdezde Castro, J.M., and Carbonell, E. Palaeoenvironmental and palaeoclimatic reconstruction of the latest Pleistocene of El Portalón site, Sierra de Atapuerca, northwestern Spain. *Palaeogeogr. Palaeocl.*, 292(3), 453-464, 2010.
- 15 López-García, J. M., Blain, H.-A., Bennàsar, M., Euba, I., Bañuls, S., Bischoff, J., López-Ortega, E., Saladié, P., Uzquiano, P. and Vallverdú, J. A multiproxy reconstruction of the palaeoenvironment and palaeoclimate of the Late Pleistocene in northeastern Iberia: Cova dels Xaragalls, Vimbodí-Poblet, Paratge Natural de Poblet, Catalonia. *Boreas*, 41(2), 235-249, 2011a.
- López-García, J. M., Cuenca-Bescós, G., Blain, H. A., Álvarez-Lao, D., Uzquiano, P., Adán, G., Arbizu, M. and Arsuaga, J. 20 L. Palaeoenvironment and palaeoclimate of the Mousterian–Aurignacian transition in northern Iberia: The small-vertebrate assemblage from Cueva del Conde (Santo Adriano, Asturias). *J. Hum. Evol.*, 61(1), 108-116, 2011b.
- López-García, J. M., Cuenca-Bescós, G., Finlayson, C., Brown, K., and Pacheco, F. G. Palaeoenvironmental and palaeoclimatic proxies of the Gorham’s cave small mammal sequence, Gibraltar, southern Iberia. *Quatern. Int.*, 243(1), 137-142, 2011c.
- 25 McKee, T.D., Doesken, N.J., and Kleist, J. The relationship of drought frequency and duration to time scales. . Proceedings of the 8th Conference of Applied Climatology. Anaheim, California: American Meteorological Society. p 179-184, 1993.
- Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., & New, M. (2004). A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). *Tyndall centre for climate change research working paper*, 55(0), 25.
- 30 Mix, A. C., Bard, E., and Schneider, R. Environmental processes of the ice age: land, oceans, glaciers (EPILOG), *Quaternary Sci. Rev.*, 20, 627-657, 2001.
- Müller, U.C., Pross, J., Tzedakis, P.C., Gamble, C., Kotthoff, U., Schmiedl, G., Wulf, S., and Christanis, K. The role of climate in the spread of modern humans into Europe. *Quaternary Sci. Rev.*, 30, 273-279, 2011.

- New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J Climate* 12:829–856
- New, M., Lister, D., Hulme, M., and Makin, I. A high-resolution data set of surface climate over global land areas, *Clim. Res.*, 21(1), 1-25, 2002.
- 5 Palombo, M. R., Filippi, M. L., Iacumin, P., Longinelli, A., Barbieri, M., and Maras, A. Coupling tooth microwear and stable isotope analyses for palaeodiet reconstruction: the case study of Late Middle Pleistocene *Elephas (Palaeoloxodon) antiquus* teeth from Central Italy (Rome area). *Quatern. Int.*, 126–128(0), p. 153-170, 2005.
- Sanchez Goñi, M. F., and Harrison, S. P. Millennial-scale climate variability and vegetation changes during the Last Glacial: Concepts and terminology. *Quaternary Sci. Rev.*, 29(21), 2823-2827, 2010.
- 10 Schmidt, I., Bradtmöller, M., Kehl, M., Pastoors, A., Tafelmaier, Y., Weninger, B., and Weniger, G. C. Rapid climate change and variability of settlement patterns in Iberia during the Late Pleistocene. *Quatern. Int.*, 274, 179-204, 2012.
- Sepulchre, P., Ramstein, G., Kageyama, M., Vanhaeren, M., Krinner, G., Sánchez-Goñi, M. F., and d'Errico, F.. H4 abrupt event and late Neanderthal presence in Iberia. *Earth Planet. Sc. Lett.*, 258(1), 283-292, 2007.
- Serangeli, J. and Bolus, M. Out of Europe - The dispersal of a successful European hominin form. *Quartar* 55, 83-98, 2008.
- 15 Shea, J.J. Transitions or turnovers? Climatically-forced extinctions of *Homo sapiens* and Neanderthals in the east Mediterranean Levant. *Quaternary Sci. Rev.*, 27, 2253-2270, 2008.
- Tallavaara, M., Luoto, M., Korhonen, N., Järvinen, H., and Seppä, H. Human population dynamics in Europe over the Last Glacial Maximum. *P. Natl. Acad. Sci. USA*, 112(27):8232-8237, 2015.
- Taylor, K.E., R.J. Stouffer, R.J. and Meehl, G.A. An Overview of CMIP5 and the experiment design.” *Bull. Amer. Meteor. Soc.*, 93, 485-498, 2012.
- 20 Tebaldi, C., and Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. A Math. Phys. Eng. Sci.*, 365(1857), 2053-2075, 2007.
- van Andel, T.H. and Davies, W. Neanderthals and Modern Humans in the European Landscape of the Last Glaciation: Archaeological results of the stage 3 project. McDonald Institute for Archaeological Research Monographs, Cambridge, 2003.
- 25 Vita-Finzi, C. and Higgs, E. Prehistoric Economy in the mount Carmel area of Palestine: Site catchment analysis. *P. Prehist. Soc.*, 36, 1970.
- Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C. Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: Present climate evaluations. *Clim Dyn*, 46(3) 1301-1329, 2016.
- 30 Vrac, M., Marbaix, P., Paillard, D., and Naveau, P. Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. *Clim. Past*, 3(4), 669-682, 2007.
- Walter, H. and Box, E. Global classification of natural terrestrial ecosystems. *Vegetatio*, 32(2), 75-81, 1976.
- Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S. O., and Svensson, A. Millennial-scale variability during the last glacial: The ice core record. *Quaternary Sci. Rev.*, 29(21), 2828-2838, 2010.

- Wood, A. W., Leung, L. R., SRidhar, V., and Lettenmaier, D. P. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Clim. Change*, 62, 189–216, 2004.
- Wood, S. N. Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties, *J. R. Stat. Soc. B*, 62(2), 413-428, 2000.
- 5 Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models, *J. Am. Stat. Assoc.*, 99(467), 2004.
- Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. (B)* 73(1), 3-36, 2011.
- Wu, H., Guiot, J., Brewer, S., and Guo, Z. Climatic changes in Eurasia and Africa at the last glacial maximum and mid-
10 Holocene: reconstruction from pollen data using inverse vegetation modelling, *Clim. Dynam.*, 29(2-3), 211-229, 2007.
- Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P., and Fifield, L. K. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, 406(6797), 713-716, 2000.

Table 1. Model selection among all possible combinations of variables for the temperature and the precipitations. Only the five models with the lowest AIC are presented. AIC scores, differences in AIC compared to the lowest scoring model (Δ_{AIC}), and AIC weights ($w_{AIC} = \exp(-0.5 \times \Delta_{AIC}) \in [0,1]$, representing the relative likelihood of the models) are reported.

Candidate model	AIC	Δ_{AIC}	w_{AIC}
<u>Bilinear interpolation</u>			
<i>Temperature</i>			
T+elv+Aco	869468.4	0	1
T+elv	878862.9	9394.52	0
T+Aco	951357.5	81889.14	0
T	955986.9	86518.50	0
elev+Aco	1605682.9	736214.55	0
<i>Precipitations</i>			
T+P+elev+Aco+RH+SLP	275068.7	0	1
T+P+elev+Aco+SLP	281222.3	6153.56	0
T+P+elev+Aco+RH	283266.4	8197.70	0
T+P+elev+RH+SLP	288574.6	13505.85	0
T+P+elev+Aco	288864.1	13795.35	0
<u>Bicubic interpolation</u>			
<i>Temperature</i>			
T+elv+Aco	896517.0	0	1
T+elv	904724.6	8207.52	0
T+Aco	952109.7	55592.63	0
T	955079.5	58562.43	0
elev+Aco	1605402.4	708885.36	0
<i>Precipitations</i>			
T+P+elev+Aco+RH+SLP	274137.9	0	1
T+P+elev+Aco+SLP	278178.8	4040.881	0
T+P+elev+Aco+RH	283162.3	9024.412	0
T+P+elev+Aco	286857.5	12719.589	0
T+P+elev+RH+SLP	286926.3	12788.379	0
<u>Kriging</u>			
<i>Temperature</i>			
T+elv+Aco	878970.1	0	1
T+elv	888326.4	9356.24	0
T+Aco	952033.1	73062.92	0
T	956903.8	77933.70	0
elev+Aco	1607626.2	728656.08	0
<i>Precipitations</i>			
T+P+elev+Aco+RH+SLP	269767.4	0	1
T+P+elev+Aco+SLP	276171.0	6403.59	0
T+P+elev+Aco+RH	277692.0	7924.60	0
T+P+elev+Aco	283099.2	13331.79	0
T+P+Aco+RH+SLP	286495.6	16728.23	0

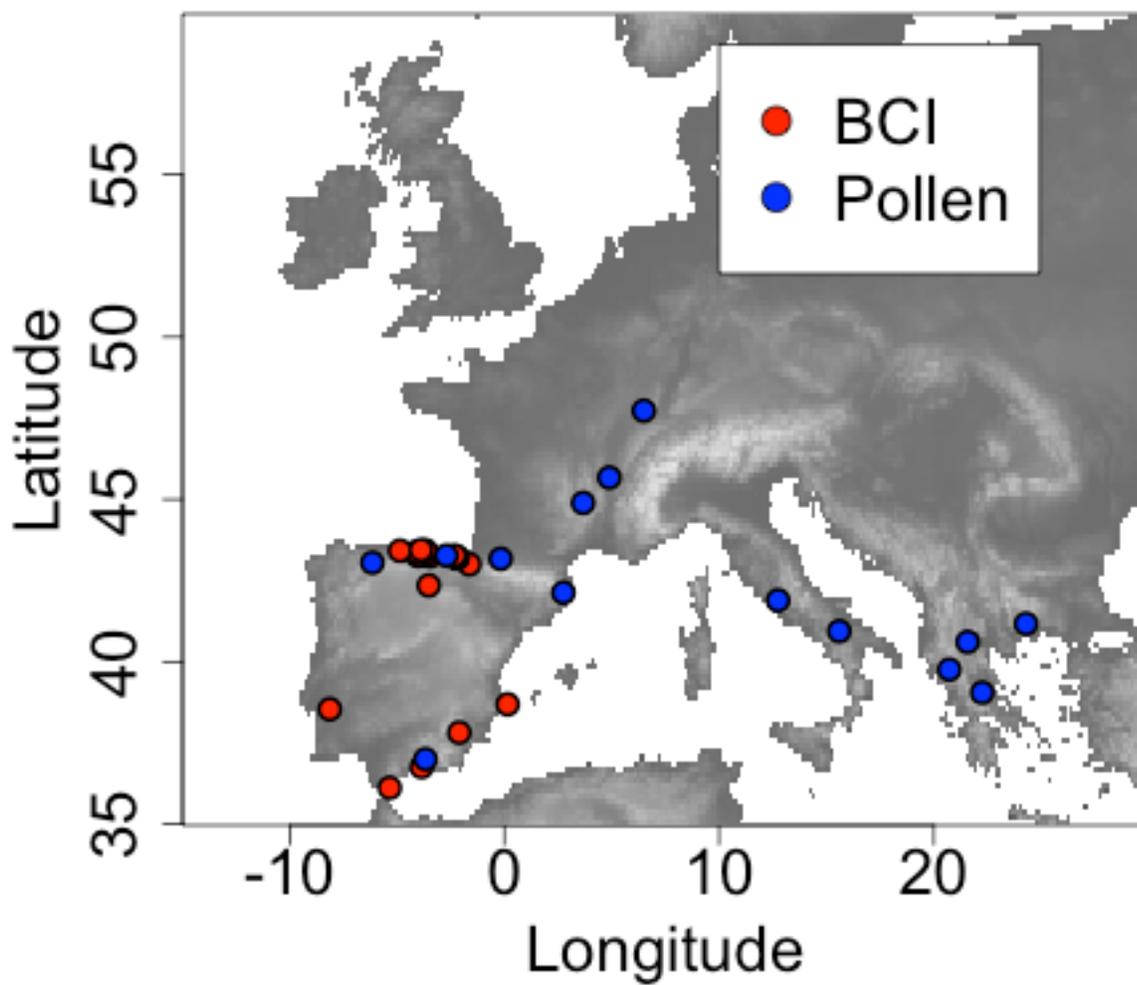


Figure 1. Study area and locations of the sites used for reconstructing local climate variables estimated on the basis of pollen and vertebrate fossils, used for the evaluation of the method. The grey scale represents elevation.

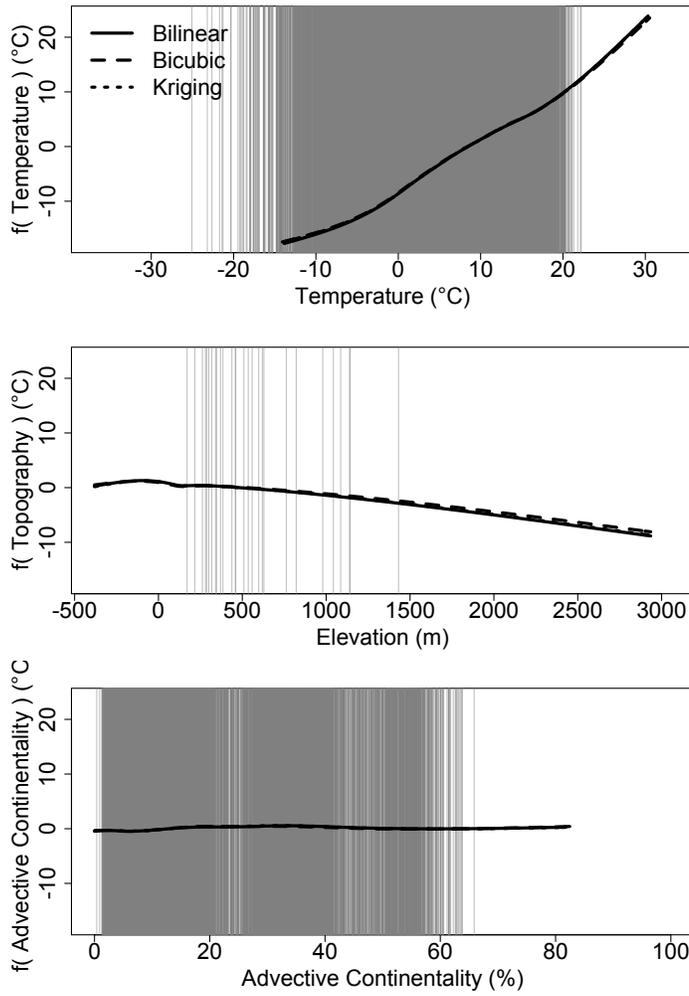


Figure 2. Splines of the GAM for temperature. The splines are scaled to the same range to allow for visual estimation of their relative importance. The range of the x-axes combines the ranges of values for the present-day period and the LGM. The grey lines indicate the values for the 12 months over the 50 years during the LGM at the archaeological sites of Figure 1 (except for the elevation, for which there is only one value per site).

5

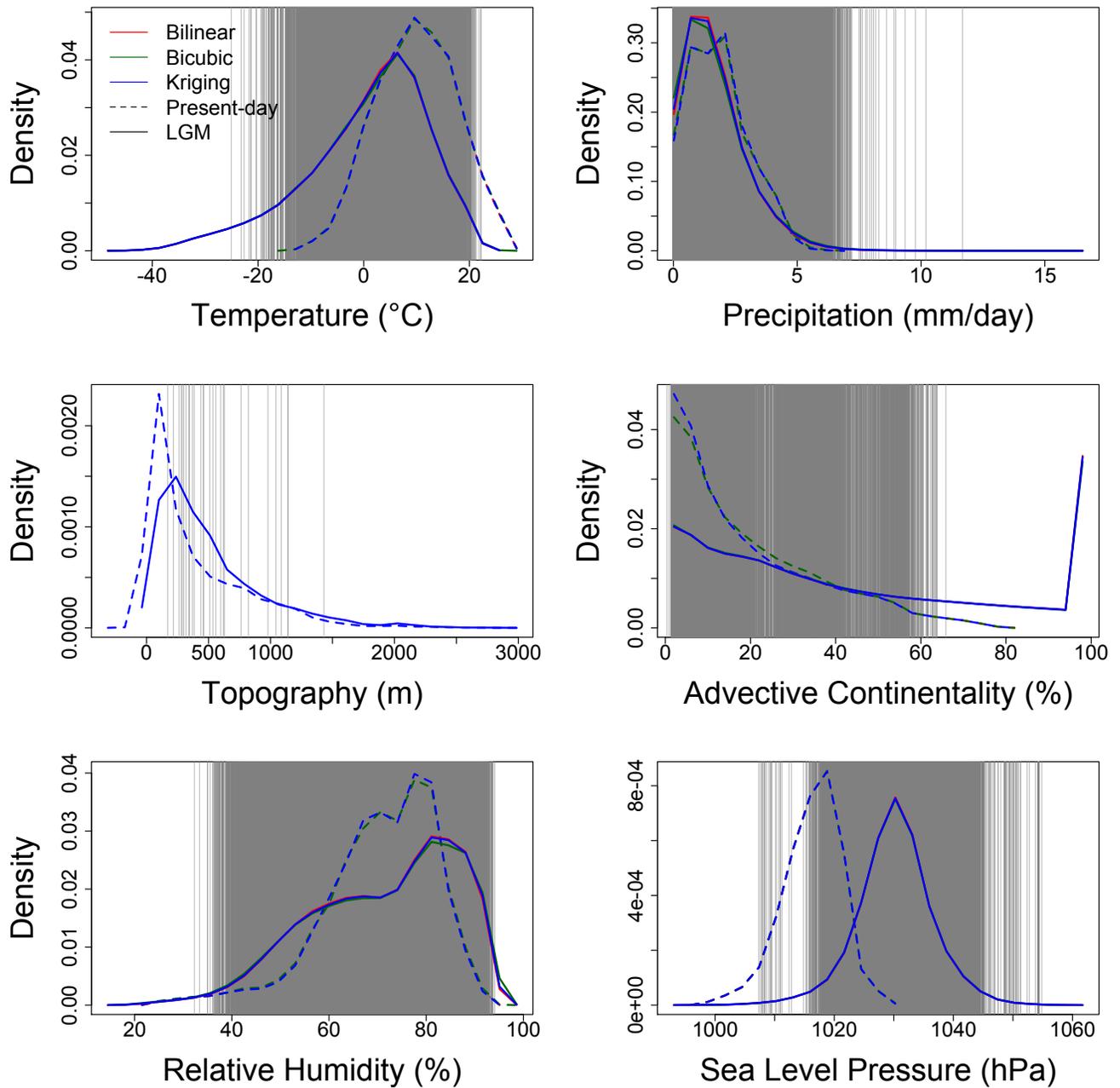


Figure 3. Histograms of the predictor variables for the present-time (1961-1990; dashed lines) and for the LGM (solid lines) over Western Europe, using the bilinear (red), bicubic (green) and kriging (blue) interpolations. The grey lines indicate the values for the 12 months over the 50 years during the LGM at the archaeological sites of Figure 1 (except for the elevation, for which there is only one value per site).

5

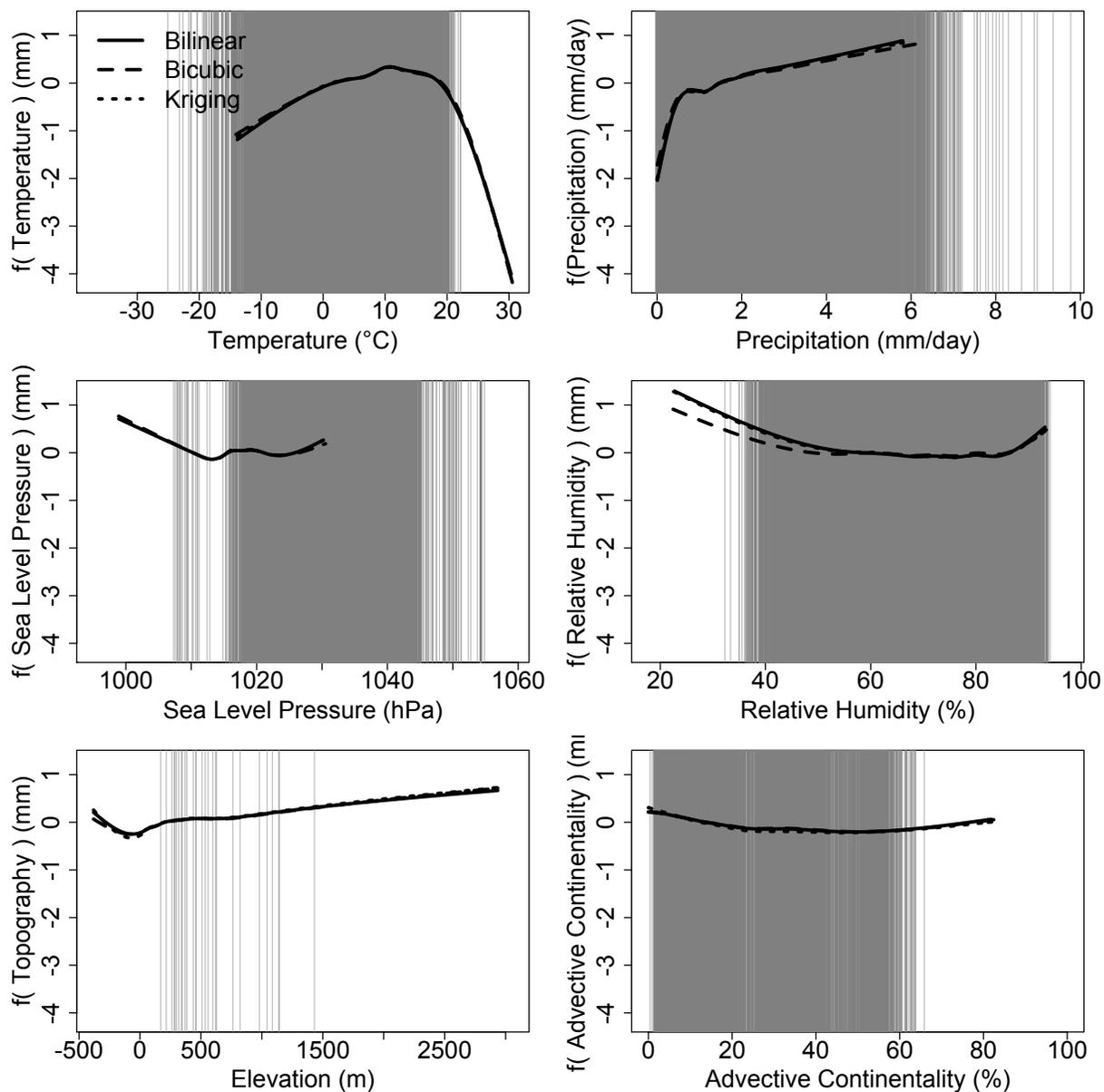
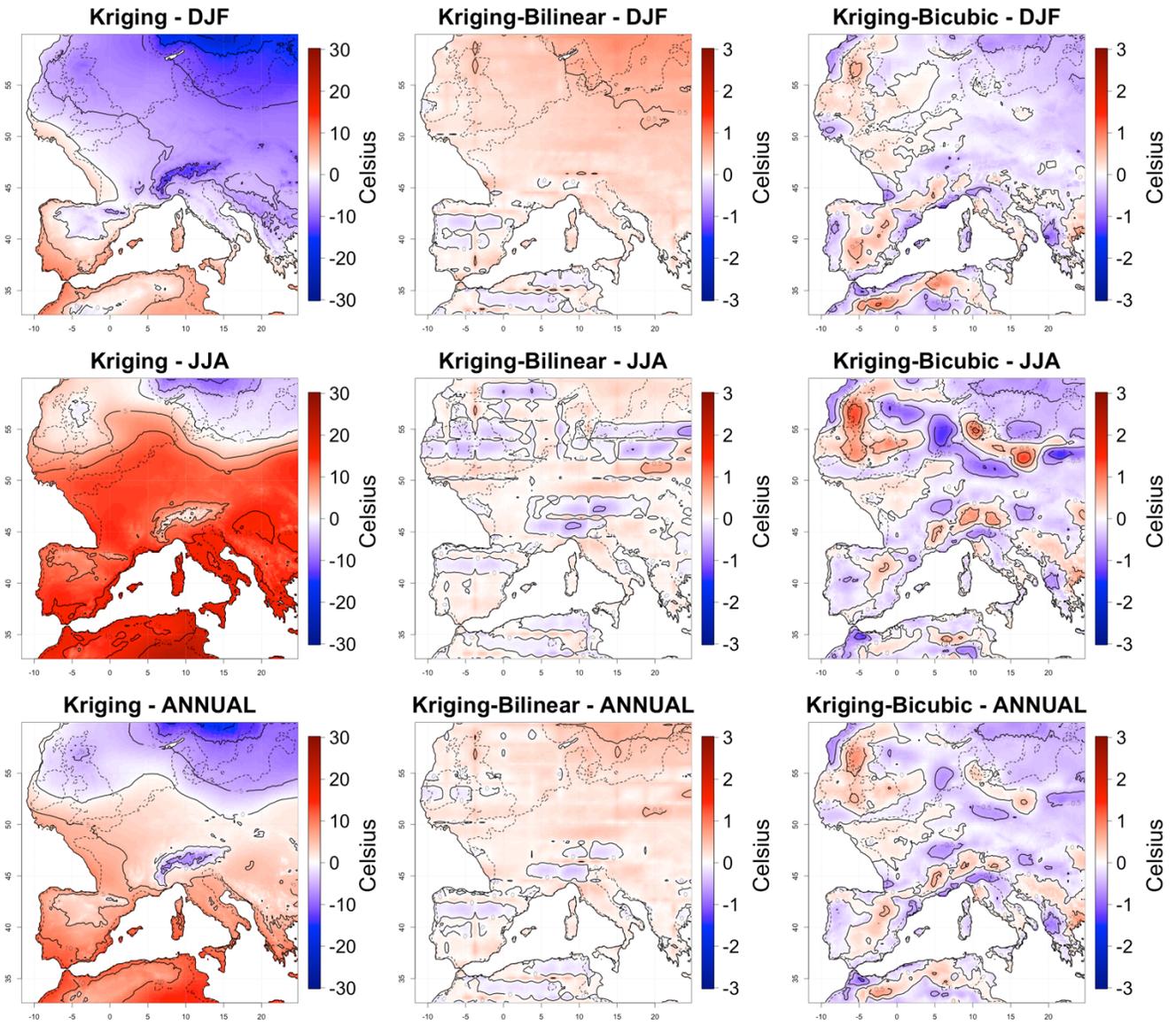


Figure 4. Splines of the GAM for precipitations. The splines are scaled to the same range to allow for visual estimation of their relative importance. The range of the x-axes combines the ranges of values for the present-day period and the LGM. The grey lines indicate the values for the 12 months over the 50 years during the LGM at the archaeological sites of Figure 1 (except for the elevation, for which there is only one value per site).

5



5 **Figure 5. Mean distributions of monthly mean downscaled temperatures over Western Europe during the LGM for winter (December, January, February), summer (June, July, August), and the whole year, computed over 50 years for the kriging interpolation technique, and difference between the kriging and the other two techniques. Downscaling was performed for each month independently, but results are combined into seasons to summarise the results.**

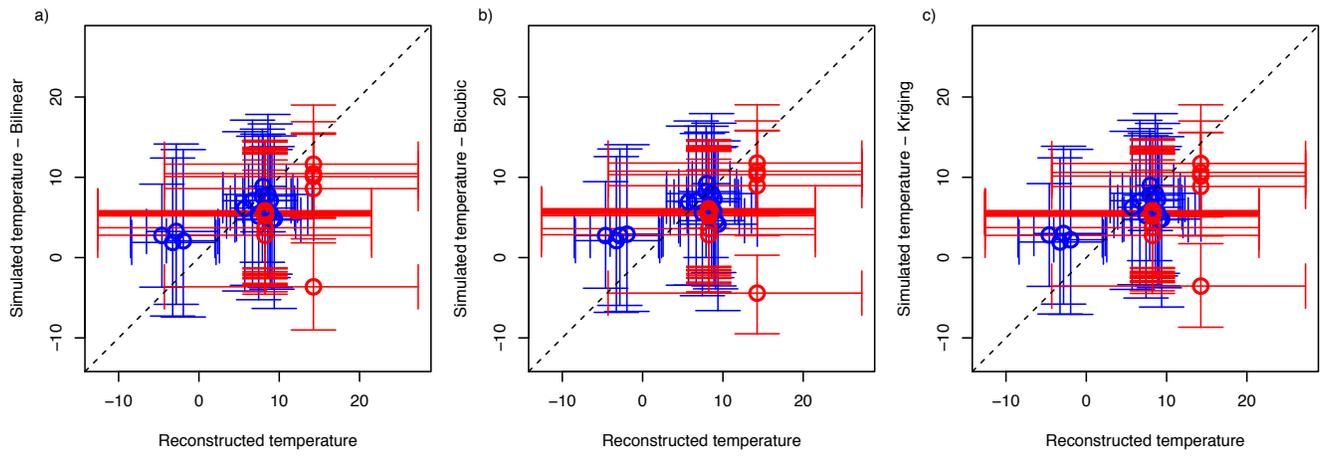


Figure 6. Comparison of reconstructed vs. downscaled temperatures for the LGM based on the BCI indices (red), and from Wu et al. (2007)'s reconstructions (blue) for a) the bilinear, b) the bicubic and c) the kriging interpolations. The circles represent the mean temperature values for the two reconstruction methods (x-axis) and the downscaled values over the 50 simulated years (y-axis). The horizontal error bars represent the range of temperature values for the reconstruction method (minimum and maximum over the whole zonobiome for the BCI indices, mean temperature of the coldest and warmest month for Wu et al. 2007). The vertical error bars correspond to the mean temperature of the coldest and warmest month over the 50 simulated years.

5

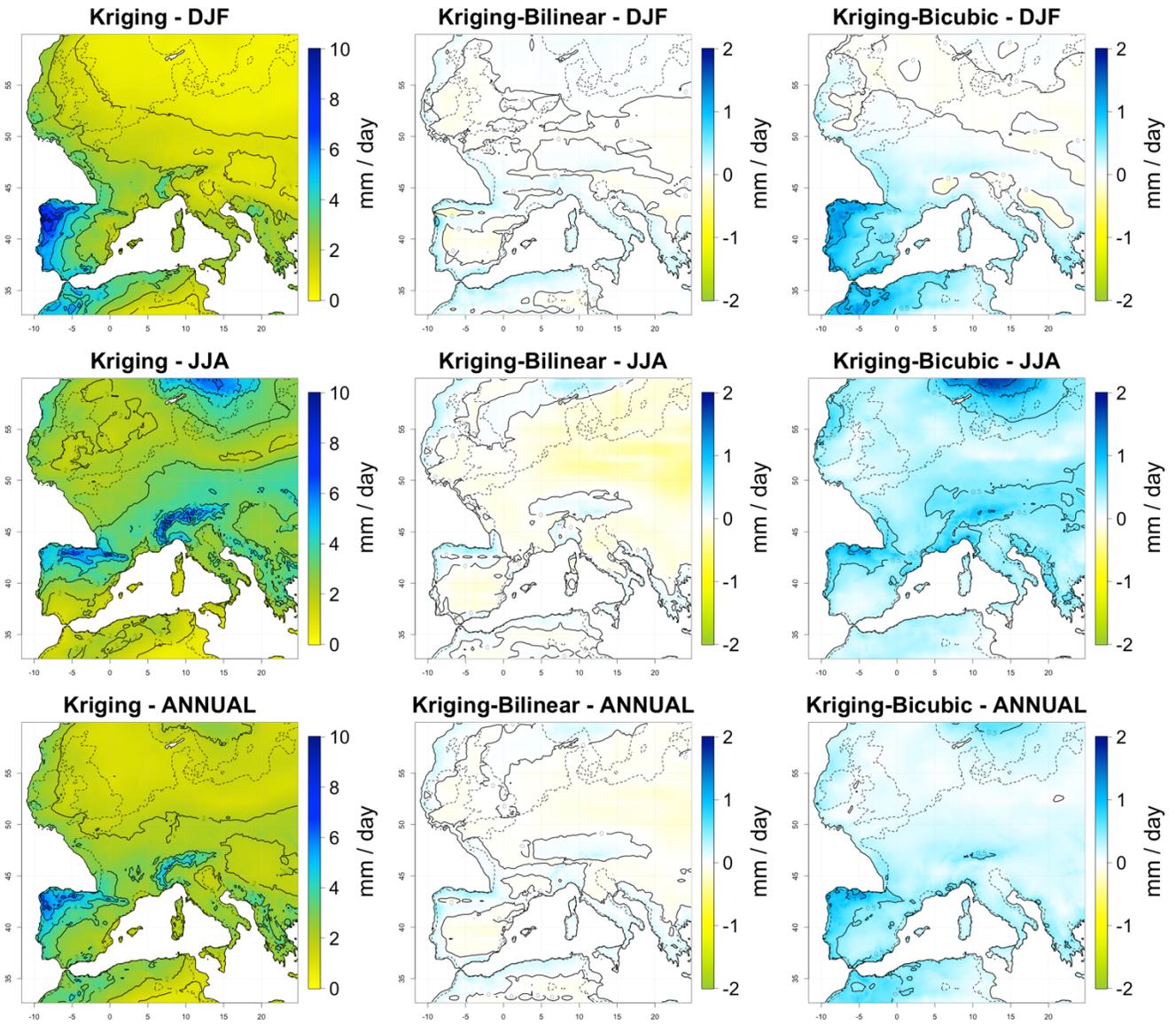


Figure 7. Mean distributions of downscaled daily precipitations over Western Europe during the LGM for winter (December, January, February), summer (June, July, August), and the whole year, computed over 50 years for the kriging interpolation technique, and difference between the kriging and the other two techniques. Downscaling was performed for each month independently, but results are combined into seasons to summarise the results.

5

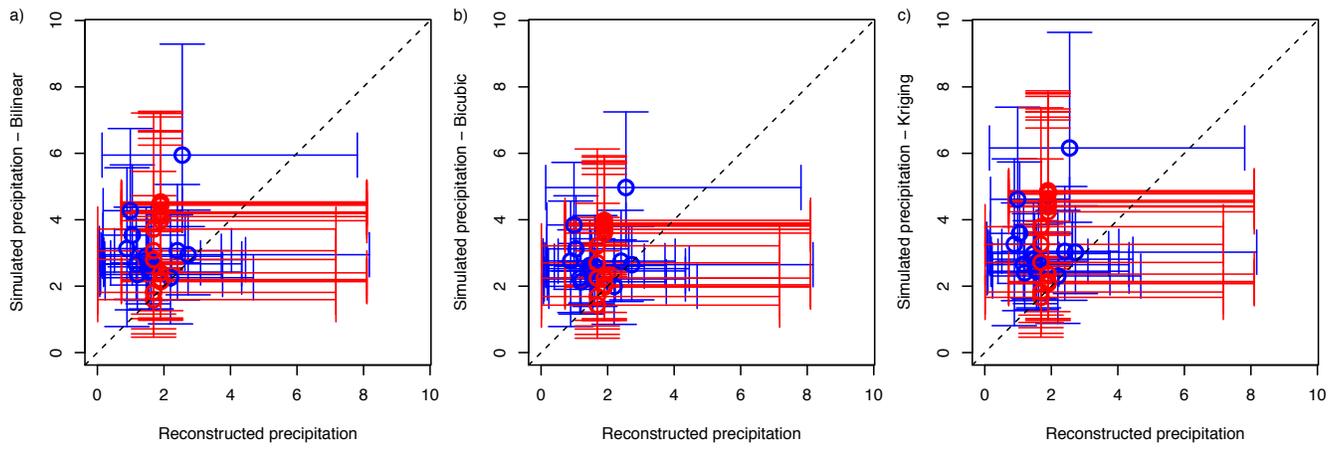


Figure 8. Boxplot of reconstructed vs. downscaled precipitations for the LGM based on the BCI indices (red), and from Wu et al. (2007)'s reconstructions (blue) for a) the bilinear, b) the bicubic and c) the Kriging interpolations. The circles represent the mean temperature values for the two reconstruction methods (x-axis) and the downscaled values over the 50 simulated years (y-axis). The horizontal error bars represent the range of precipitation values for the reconstruction method (minimum and maximum over the whole zonobiome for the BCI indices, mean precipitation of the coldest and warmest month for Wu et al. 2007). The vertical error bars correspond to the mean precipitation of the driest and wettest month over the 50 simulated years.

5

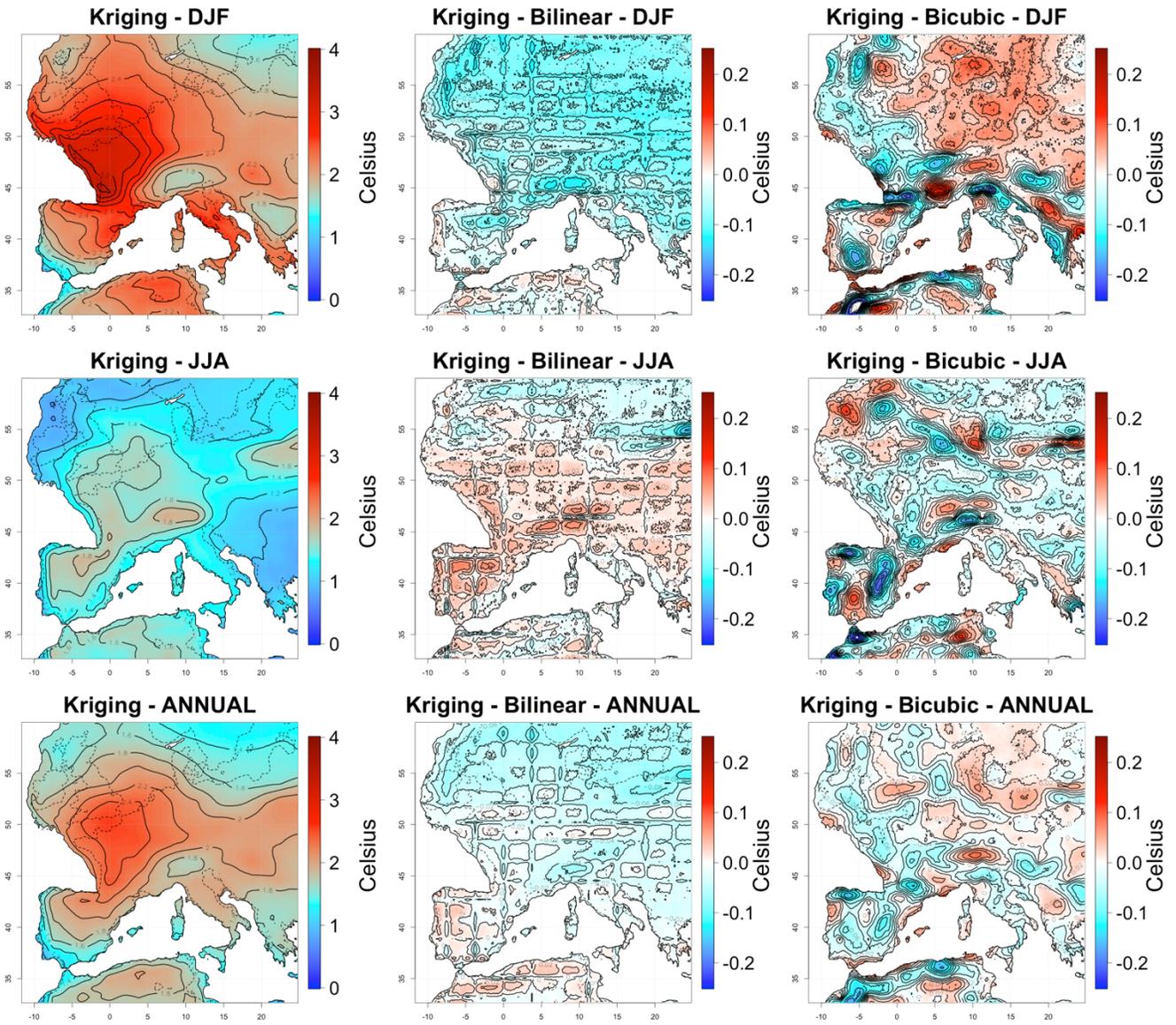


Figure 9. Maps of temporal variations (standard deviation of each month across 50 years) of downscaled monthly mean temperatures over Western Europe during the LGM averaged over winter (December, January, February), summer (June, July, August), and the whole year for the kriging interpolation technique, and difference between the kriging and the other two techniques. Variability was computed for each month independently, but results are combined into seasons to summarise the results.

5

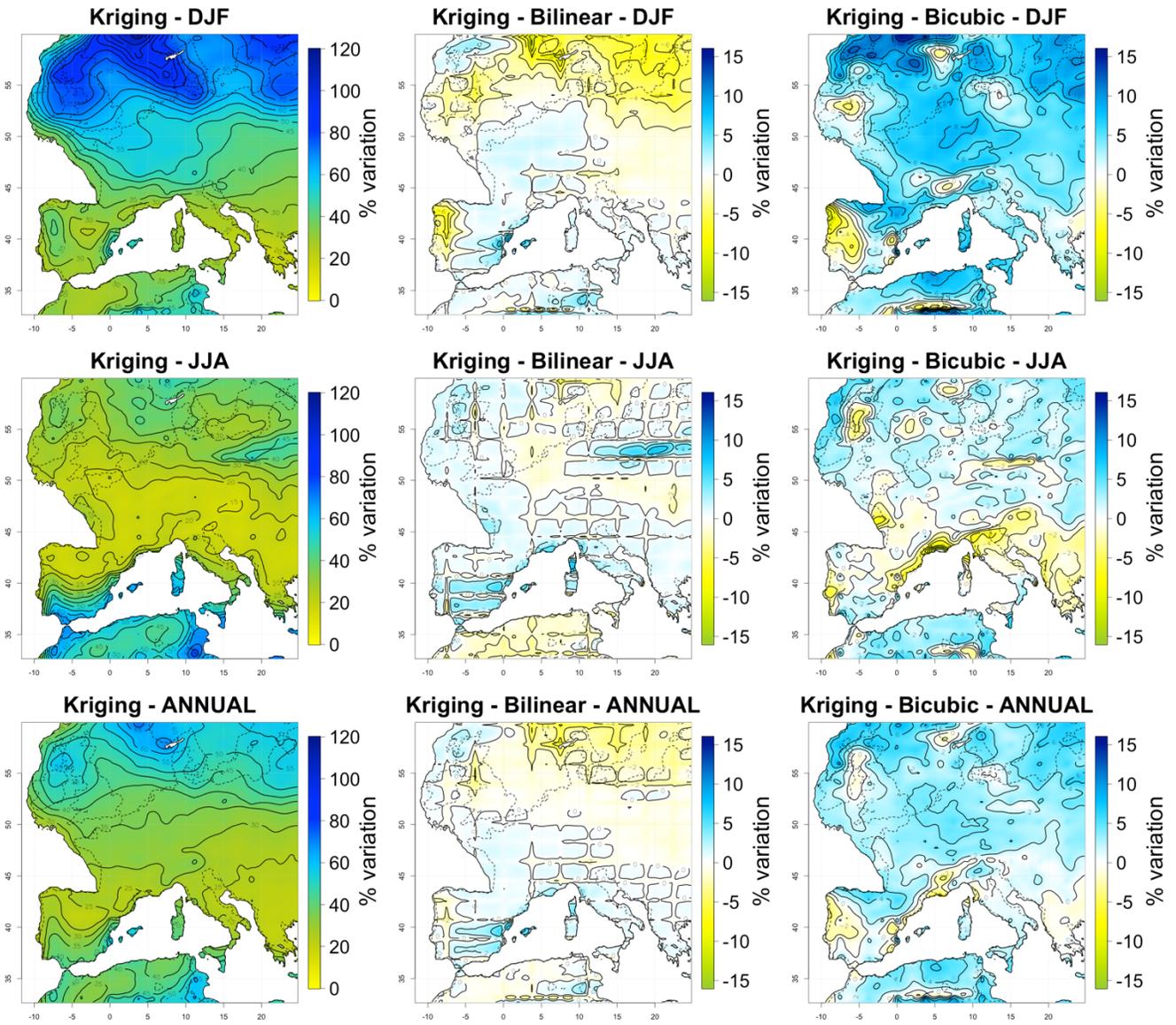


Figure 10. Maps of temporal variations (coefficient of variation of each month across 50 years) of downscaled daily precipitations over Western Europe during the LGM averaged over winter (December, January, February), summer (June, July, August), and the whole year, computed over 50 years for the kriging interpolation technique, and difference between the kriging and the other two techniques. Variability was computed for each month independently, but results are combined into seasons to summarise the results.

5