PAPER • OPEN ACCESS

Improving biome and climate modelling for a set of past climate conditions: evaluating bias correction using the CDF-t approach

To cite this article: Anhelina Zapolska *et al* 2023 *Environ. Res.: Climate* **2** 025004

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
CLIMATE

PAPER

Improving biome and climate modelling for a set of past climate conditions: evaluating bias correction using the CDF-t approach

Anhelina Zapolska[1,*] , Mathieu Vrac[2] , Aurélien Quiquet[2,3] , Thomas Extier[4] , Frank Arthur[5] , Hans Renssen[5] and Didier M Roche[1,2]

[1] Earth and Climate Cluster, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[3] University Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, 38000 Grenoble, France
[4] University Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France
[5] Department of Natural Sciences and Environmental Health, University of South-Eastern Norway, Bø, Norway
[*] Author to whom any correspondence should be addressed.

E-mail: zapolskaanhelina@gmail.com

Abstract

Climate model simulations are inherently biased. It is a notably difficult problem when dealing with climate impact assessments and model-data integration. This is especially true when looking at derived quantities such as biomes, where not only climate but also vegetation dynamics biases come into play. To overcome such difficulties, we evaluate the performance of an existing methodology to correct climate model outputs, applied here for the first time to long past climate conditions. The proposed methodology relies on the 'Cumulative Distribution Function-transform' (CDF-t) technique, which allows to account for climate change within the bias-correction procedure. The results are evaluated in two independent ways: (i) using forward modelling, so that model results are directly comparable to reconstructed vegetation distribution; (ii) using climatic reconstructions based on an inverse modelling approach. The modelling is performed using the intermediate complexity model iLOVECLIM in the standard global and interactively downscaled over the Europe version. The combined effects of dynamical downscaling and bias correction resulted in significantly stronger agreement between the simulated results and pollen-based biome reconstructions (BIOME6000) for the pre-industrial (0.18 versus 0.44) and mid-Holocene (MH) (0.31 versus 0.40). Higher correlation is also observed between statistically modelled global gridded potential natural distribution and modelled biomes (0.36 versus 0.41). Similarly, we find higher correlation between the reconstructed and the modelled temperatures for the MH (0.02 versus 0.21). No significant difference is found for the Last Glacial Maximum when using temperature reconstructions, due to the low number of data points available. Our findings show that the application of the CDF-t method on simulated climate variables enables us to simulate palaeoclimate and vegetation distribution in better agreement with independent reconstructions.

1. Introduction

Earth's palaeoclimate research is a prominent source of knowledge about the climate system, its functions and interactions, as well as a tool for investigating potential human-made climate effects (Hansen and Sato 2012). Nowadays, various tools and methods allow us to assess the state of past climates based on a set of climate proxies (Guiot *et al* 1993, Peyron *et al* 1998, Mauri *et al* 2015, Kaufman *et al* 2020). Proxy-based climate reconstructions provide knowledge of past climatic conditions but do not readily give access to the

underlying mechanisms behind them. In addition, they are often spatially discontinuous and temporally limited. This brings the need for a methodology to physically interpolate them at a regional scale. In this view, climate modelling is a complementary approach that is both flexible in terms of spatial and temporal variability, and can be adapted to a wide range of studies (Haywood *et al* 2019). Data and models play a complementary role in understanding climate change. Models offer mechanisms to explain the data, while proxy data provides a knowledge base for evaluating the models.

Palaeoclimate model simulations using Earth system models of intermediate complexity (EMICs) can be widely used to improve the understanding of the mechanisms and dynamics of the climate system (Claussen *et al* 2002, Keery *et al* 2016). However, EMICs have very limited simulation power at finer geographical scales and need alternative methods to be compared to small-scale proxies (over land in particular).

To achieve higher spatial resolution, we can use downscaling approaches. In our approach, we rely on the interactive physical downscaling approach introduced by Quiquet *et al* (2018) within the model of intermediate complexity iLOVECLIM (Claussen *et al* 2002, Goosse *et al* 2010, Roche *et al* 2014). The downscaling technique significantly improved spatial resolution and overall performance of the model, including correction of several topography-related biases (Arthur *et al* 2023). However, large-scale model biases remain present in the outputs of the downscaled version of iLOVECLIM, considering that the downscaling technique was not designed to correct broader, regional scale, model biases (Quiquet *et al* 2018). Broad-scale iLOVECLIM biases, linked to the model formulation and underlying assumptions, were described by Goosse *et al* (2010), and include, among others: overestimation of temperatures in low latitudes and precipitations in subtropics (in particular in South America and West Africa), biases in atmospheric circulation, and too symmetric distribution of precipitation between the Northern and the Southern hemisphere.

Vrac *et al* (2012) have developed a bias correction methodology, based on cumulative distribution function transformations (CDF-t), aimed to reduce large-scale model biases. The findings indicate a significant improvement in the results of regional climate models with respect to observations for wind speed, temperature, and precipitation distributions (Vrac *et al* 2012). The CDF-t method is an advancement of the quantile mapping (QM) bias correction method (Guo *et al* 2020). The QM method was shown to have the potential to reduce biases in palaeoclimate simulations (Beyer *et al* 2020). The CDF-t method was previously applied as a downscaling technique (Michelangeli *et al* 2009, Lavaysse *et al* 2012, Vigaud *et al* 2013, Noël *et al* 2021) and as a bias correction methodology (Luu *et al* 2018, Mesta and Kentel 2022) to correct regional climate model outputs, used as an input for impact models. However, this methodology has not yet been tested for past time periods with available paleoenvironmental data synthesis, which is the goal of the present study. We are, therefore, assessing the performance of CDF-t as a bias correction technique only over three periods of the past with contrasting climatic conditions: pre-industrial (PI, 0 ka BP), mid-Holocene (MH, 6 ka BP or climatic optimum), and Last Glacial Maximum (LGM, 21 ka BP). These periods have been a focus for the synthesis of paleoenvironmental data (Prentice and Webb 1998, Prentice *et al* 2000, Harrison *et al* 2016, Harrison 2017), for intercomparison modelling exercises (Kageyama *et al* 2018), and to evaluate the climate models with data (Kohfeld and Harrison 2000, Bartlein *et al* 2011, Cleator *et al* 2020, Kaufman *et al* 2020), which created a basis for model evaluation and established them as key periods in palaeoclimate modelling (Brierley *et al* 2020).

The assessment of the performance of palaeoclimate modelling is typically done using proxy-based reconstructions as a reference (e.g. Ramstein *et al* 2007, Goosse *et al* 2010). This can be done in two ways, using a forward or an inverse approach. The forward approach implies modelling of observables (e.g. biomes, oxygen isotopes of water and air) directly within the climate model (Haywood *et al* 2019). During the inverse approach, proxies are processed to reconstruct physical climatic variables, analogous to typical climate model outputs (Haywood *et al* 2019). For example, applying reverse vegetation modelling or statistical modelling approaches (e.g. response surface and modern analogue approaches) to pollen data to estimate temperature and precipitation values (Guiot *et al* 2000, Chevalier *et al* 2020). The inverse approach was previously used to evaluate the performance of bias correction methods (the delta method, generalized additive models, and QM) for palaeoclimate simulations by Beyer *et al* (2020). To provide an in-depth analysis of the potential of the CDF-t bias correction method, we estimate its accuracy using the forward and inverse approaches of data-model comparison. Both approaches have considerable limitations. In the forward approach, data-model biome comparison is qualitative, and it is challenging to evaluate the similarity between data and model results. While, the inverse approach strongly depends on the vegetation model used. The use of both methods in combination provides an advantage of independent evaluation of the featured bias correction with two different methods. We will analyse the bias correction performance with a specific focus on terrestrial biomes at the global scale and over the European area at high resolution. If

the CDF-t method results in a better representation of the distributions of the tested variables (temperature and precipitation) in past climate contexts than non-corrected model results, we can be more confident that the biomes simulated by the model for these past periods are also more realistic. This allows potentially for a more accurate analysis of the climate dynamics, contributing to a better understanding of the conditions during these crucial benchmark periods in the past.

The main objectives of our study are: (i) to apply the CDF-t methodology for correcting climatic biases in the context of palaeoclimate modelling; (ii) to test the robustness of the developed technique on three time periods of the past with contrasting climatic conditions, using proxy-based climate and vegetation reconstructions; and (iii) to evaluate the advantages and shortcomings of the proposed approach.

2. Methodology

2.1. Climate input

2.1.1. iLOVECLIM model description and online dynamical downscaling
The climate simulations were performed with the iLOVECLIM model, version 1.1.5. This model is a code fork of the original LOVECLIM 1.2 model (Goosse *et al* 2010), revised by Roche (2013) and further expanded by Quiquet *et al* (2018). The applied version of iLOVECLIM includes an atmospheric component (ECBilt), an oceanic general circulation model (CLIO), and a simple land vegetation model (VECODE) (Roche *et al* 2014). ECBilt is a three-level, quasi-geostrophic model that consists of three vertical layers at T21 horizontal resolution, corresponding to $\sim$5.625° $\times$ $\sim$5.625° latitude–longitude (Opsteegh *et al* 1998). CLIO is an oceanic GCM with 20 unevenly spaced vertical layers at 3° $\times$ 3° latitude–longitude horizontal resolution (Goosse and Fichefet 1999). VECODE is a reduced-form dynamic global vegetation model (DGVM), which describes vegetation dynamics using two plant functional types (PFTs): trees and grass (Brovkin *et al* 1997). iLOVECLIM has been previously successfully applied to simulate the climate of the Holocene (Arthur *et al* 2023, 2016, Zhang *et al* 2018, Li *et al* 2019), the LGM (Lhardy *et al* 2021), the last deglaciation (Quiquet *et al* 2021) and the last interglacial (Li *et al* 2020).

In order to obtain higher spatial resolution climate model outputs, we make use of the online interactive downscaling method embedded in iLOVECLIM, first described by Quiquet *et al* (2018). It is a physical downscaling procedure that replicates the processes governing the computation of climatic parameters on a spatially refined grid to explicitly take into account the sub-grid topography in the computation of heat and precipitation. Here we use a global climatic simulation with a zoom at high spatial resolution grid of 0.25° $\times$ 0.25° over Europe. The downscaled segment is extracted and used for the analysis. This downscaling procedure is interactive in the sense that the large-scale temperature and precipitation are directly impacted by the sub-grid computations. This means that the large-scale climate outputs are different when using the downscaling or not.

2.1.2. Numerical experiments
To explore the performance of the CDF-t bias correction method under different climatic conditions, we carried out experiments for four distinct periods (present day (PD), PI, MH, and LGM) for the two iLOVECLIM grids: T21 (5.625° resolution, global simulations) and hi-res (0.25° resolution, simulations over Europe). In our simulations we used standardized boundary conditions for palaeoclimate simulations, provided by the Palaeoclimate Modelling Intercomparison Project Phase 4 (PMIP-4) (Kageyama *et al* 2017), using the forcings presented in table 1. The astronomic conditions in the simulations were derived from Berger (1978). Three equilibrium simulations were performed with a spinup of 3000 years: LGM, MH, and PI (taken at 1750 A.D.). From the PI, a transient PD simulation was then run transiently covering 1750–2014. During the transient simulations of 1750–2014 astronomical parameters were kept constant, under the assumption that at this time scale the changes were so minor that there was no significant effect on the climate. The concentrations of the atmospheric trace gases were derived from the consolidated datasets of historical atmospheric concentrations for the Climate Model Intercomparison Project (CMIP6) (Meinshausen *et al* 2017).

To be consistent with the length of the observation dataset, 38 years were extracted from each simulation: 1976–2014 for the transient PD simulation and 38 years at the end of the spin-up for each of the equilibrium simulations. For the analysis, we used climatological mean of the simulations. To ensure that the selection of a 38 year average does not distort our analysis, we tested sensitivity of our method to low-frequency climate variability within the iLOVECLIM model. The results (data not shown) indicated no significant impact of low-frequency climate variability in iLOVECLIM on studied variables following the CDF-t application.

Table 1. Experimental set-up. The name of each experiment consists of an abbreviation of a studied period (PI, MH, LGM), and a grid name (T21, hi-res). T21 simulations are performed globally (g), hi-res simulations—over Europe (EU). For the LGM experiment, the ice-sheet forcing is the GLAC-1D reconstruction taken from the PMIP-4 protocol (Kageyama *et al* 2017). For a complete description of the experimental setup of the LGM simulation, the reader is referred to Lhardy *et al* (2021). For equilibrium (E) simulations, 3000 years of spinup were performed, and the last 38 years have been used for the analysis. For transient (T) simulations, 3000 years of spinup were performed, and then transient simulations for 264 years were performed, of which the last 38 years have been used for the analysis.

Simulation name	Resolution (°)	Simulation type	Astronomical parameters	Concentration of greenhouse gases
PD—T21 (g)	5.625	T	Eccentricity = 0.016 724	CO_2 = 333.3–395.5 ppm
PD—hi-res (EU)	0.25	T	Obliquity = 23.446°	CH_4 = 1527.7–1831.5 ppb
			Perihelion-180° = 102.04°	N_2O = 299.3–326.9 ppb
PI—T21 (g)	5.625	E		CO_2 = 280 ppm
PI—hi-res (EU)	0.25	E		CH_4 = 760 ppb
				N_2O = 270 ppb
MH—T21 (g)	5.625	E	Eccentricity = 0.018 682	CO_2 = 264.4 ppm
MH—hi-res (EU)	0.25	E	Obliquity = 24.105°	CH_4 = 597 ppb
			Perihelion-180° = 0.87°	N_2O = 262 ppb
LGM—T21 (g)	5.625	E	Eccentricity = 0.018 994	CO_2 = 190 ppm
			Obliquity = 22.949°	CH_4 = 375 ppb
			Perihelion-180° = 114.42°	N_2O = 200 ppb

2.1.3. Reference observation dataset
The observational reference dataset used for bias correction is the EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for the InterSectoral Impact Model Intercomparison Project (Lange 2018). The EWEMBI data covers the entire globe at 0.5° horizontal and daily temporal resolution from 1979 to 2014.

For consistency with the EWEMBI dataset, for bias correction we have extracted the 1979–2014 timespan from the 1750–2014 model simulation, both at T21 and on the high-resolution grid. The EWEMBI dataset grid has been regridded to the iLOVECLIM T21 and hi-res grids using the first-order conservative remapping approach (Jones 1999) to ensure that each iLOVECLIM grid cell time series corresponds to an EWEMBI grid cell. This approach allowed us to perform bias correction on a gridcell-by-gridcell basis.

2.2. CDF-t method
2.2.1. CDF-t description
The method we used for bias correction is the 'cumulative distribution function-transform' (CDF-t) (Vrac *et al* 2012). The CDF-t method is an extension of a widely-applied univariate bias-correction method, the 'quantile-quantile' (QQ) method (e.g. Haddad and Rosenfeld 1997, Déqué 2007). In contrast to the QQ method, the CDF-t approach allows accounting for the climate change signal provided by the climate simulations that need to be corrected. This allows for the transformation (i.e. correction) of the simulations by modifying its underlying statistical distribution while maintaining the modelled climate change signal (Vrac *et al* 2012). The QQ method performs a QM between the CDF F_{Sh} of the simulated data (subscript *S*) and the CDF F_{Rh} of the reference data (subscript *R*), both under a historical period (subscript *h*). Hence, when a simulated value, say x_p, over a projection period (subscript *p*) has to be bias-corrected, QQ first estimates its probability value in terms of CDF from historical simulations, before associating a quantile, say y_{QQ}, with the same probability in the 'observations' world:

$$F_{Rh}(y_{QQ}) = F_{Sh}(x_p) \Leftrightarrow y_{QQ} = F_{Rh}^{-1}(F_{Sh}(x_p)),$$

where F_{Rh}^{-1} is the inverse CDF function of F_{Rh}. Thus, even though x_p does not follow the distribution F_{Sh}, the associated probability $F_{Sh}(x_p)$ is computed anyway, which is problematic (Vrac *et al* 2012). The CDF-t approach avoids this potential mismatch between simulated values and their associated CDFs by considering the CDF change between observed and simulated climatic variables, in order to estimate the reference CDF F_{Rp} under the projection period, and then perform a QM between F_{Sp} and F_{Rp}. The transformation T from F_{Sh} to F_{Rh} is modelled as:

$$T(F_{Sh}(x)) = F_{Rh}(x), \tag{1}$$

which, by substituting x with $F_{Sh}^{-1}(u)$, where u represents any probability in [0, 1], is estimated as:

$$T(u) = F_{Rh}(F_{Sh}^{-1}(u)). \tag{2}$$

Based on this definition of T, and assuming that it is time-invariant, the transformation can be applied to F_{Sp}, the CDF of the variable of interest during a (future or past) projection period p, to estimate F_{Rp}, the CDF of the same variable during the projection period p:

$$T\left(F_{Sp}\left(x\right)\right) = F_{Rp}\left(x\right),\qquad(3)$$

that is

$$F_{Rp}\left(x\right) = F_{Rh}\left(F_{Sh}^{-1}\left(F_{Sp}\left(x\right)\right)\right).\qquad(4)$$

After obtaining F_{Rp} using equation (4), a QM can then be performed between F_{Sp} and F_{Rp} to obtain local bias-corrected time series. Unlike the QQ method that is applied directly between F_{Sh} and F_{Rh} (Déqué 2007), the CDF-t method corresponds to a QM between F_{Sp} and F_{Rp}. Subsequently, the values are generated based on the F_{Rp} in chronological agreement with simulated climate data. Previous studies with BIOME models (Prentice *et al* 1998, Harrison and Prentice 2003, Wohlfahrt *et al* 2004) had proposed an anomaly procedure to remove the inherent model biases for diagnostic purposes. In the anomaly procedure, the vegetation models are run using modern (baseline) climatology (e.g. CRU) with the added modelled climate change (anomaly) between the past and PI periods. Both CDF-t and anomaly methodologies preserve the climate change between two time periods. The main difference between CDF-t and the anomaly procedure is that using CDF-t, the temporal chronology (i.e. the sequence of events in its own variability) simulated by a climate model is preserved while in the anomaly procedure, chronology remains identical to reference (baseline) climate dataset variability. This is a fundamental advantage of CDF-t: a change in the temporal properties of the climate simulations is not accounted for in the anomaly procedure while it is also preserved by CDF-t.

A full description of the method can be found in Vrac *et al* (2012).

2.2.2. Application
In this study, we used CDF-t to correct three climatic variables simulated by iLOVECLIM and needed for the vegetation computations: near-surface air temperature, precipitation, and relative humidity. The calibration of the bias correction was based on these variables from the PD simulations 1979–2014 and the observation dataset for the same period. This bias correction was then applied to PI, MH, and LGM.

Seasonality of the geological past of the Earth differs from modern seasonality (Guiot *et al* 2000, Wu *et al* 2007, Koutavas and Joanides 2012) due to the change in the solar input's to the Earth over the year following long-term change of its orbit around the sun. Therefore, we tested our methodology by applying bias correction both on a monthly and yearly basis to evaluate the consequence of the choice made. For the monthly-based bias correction, we applied CDF-t, on a monthly basis, to the daily data ensemble of each grid cell, given that the days belong to the same month. During the yearly-corrected bias correction, we applied CDF-t to the full data ensemble of each grid cell, regardless of the month. Within the two options for bias correction with CDF-t (on a monthly or yearly basis) there is potential to test if one or the other methodology has a significant impact on the result. It must be recognized that bias correcting on a per month basis does not guarantee to maintain the seasonality changes at a period where the orbital configuration is significantly different. The performance of the bias correction was analysed by calculating the root-mean-square error (RMSE), Pearson's correlation coefficient (R), and standard deviation (SD) between the values of the bias-corrected iLOVECLIM simulations and the EWEMBI values.

An additional complexity arises for the LGM simulations: the land-sea distribution was different during that time period (supplementary figure 1), owing to a sea-level difference of about 125 m lower than that of the PD (e.g. Lambeck *et al* 2014). Hence, there are regions that were land 21 ka ago and that are oceans or seas today. Applying the CDF-t methodology on LGM land grid cells that are oceans or seas today would be inconsistent. Hence, we decided to restrict the bias correction of variables on locations that are land both during the LGM and today.

2.3. Method evaluation
2.3.1. Inverse approach: comparison with climate reconstructions
The inverse approach entails the comparison of modelled temperature and proxy-based reconstructions. The CDF-t approach preserves the evolutions of the input climate, and thus the anomalies (differences) with respect to the PD are constant before and after the CDF-t bias correction. These anomalies for the Holocene were previously analysed by Arthur *et al* (2023) with respect to proxy-based reconstructions from the Italian Alps (Harrison *et al* 1996, Mauri *et al* 2015, Furlanetto *et al* 2018), Scandinavia (Harrison *et al* 1996, Mauri *et al* 2015, Bjune *et al* 2005, Seppä and Birks 2001), and Mediterranean region (Kuhnt *et al* 2008, Bartlein *et al* 2011, Guiot and Kaniewski 2015, Mauri *et al* 2015, Peyron *et al* 2017, Finné *et al* 2019). Their findings

indicated that dynamically downscaled iLOVECLIM simulations result in a better agreement with proxy-based reconstructions and other climate model studies, compared to the coarse resolution grid, particularly in the Mediterranean and complex mountainous terrain. In the current study, the dataset used for the data-model comparison was extracted from the global database of Holocene paleotemperature records Temperature 12k at version 1.0.0 (Kaufman *et al* 2020), where data is represented as absolute temperature values, rather than anomalies. It consists of reconstructed through the inverse modelling absolute near-surface air temperature data points derived from lake sediment, marine sediment, peat, glacier ice, and other natural archives. Temperature reconstructions were done by numerous independent studies using a wide variety of proxy data (i.e. alkenones, d18O), associated with several uncertainties (see section 4.1). The design of the Temperature 12k database is further described by Kaufman *et al* (2020), while the sources of each data point are included in the metadata of the database.

For the purpose of this study, we extracted the Temperature 12k terrestrial mean annual values dated between 5.5–6.5 ka (MH), and 20.5–21.5 ka (LGM). For the sake of comparing with the iLOVECLIM model outputs, we resampled the Temperature 12k data points to T21 and hi-res grid, using the aggregated average resampling method. This method assigns to the new output cell the most common values of the cells in the input map within the limits of the output cell (Díaz-Pacheco *et al* 2018). We then analysed the differences between the modelled and reconstructed near-surface air temperatures before and after the CDF-t application at corresponding grid cells.

2.3.2. Forward approach: comparison with pollen-based vegetation reconstructions
As the second step of the evaluation, we used a forward approach to directly simulate the reconstructed vegetation from pollen proxies, so that we are not dependent on 'transfer functions', as in the inverse method. To evaluate the performance of the bias correction in palaeoclimate modelling with the forward evaluation approach, we performed a set of simulations of biome distribution of the studied periods. The embedded reduced-form DGVM in iLOVECLIM describes vegetation dynamics using only 2 PFTs (Brovkin *et al* 1997). Thus, for the current study, we use the more complex CARbon Assimilation In the Biosphere (CARAIB) dynamic vegetation model (Warnant *et al* 1994, François *et al* 1998, Otto *et al* 2002, Laurent *et al* 2008, Dury *et al* 2011), forced offline using iLOVECLIM climatic outputs. The CARAIB model is made of five modules: hydrological, canopy photosynthesis and stomatal regulation, carbon allocation-growth-respiration, heterotrophic respiration, and plant competition and biogeography. Vegetation is distributed in 26 PFTs, which, with respect to climatic parameters, are classified into 20 biomes. As an input, CARAIB requires six variables with a daily frequency: near-surface air temperature, precipitation, relative humidity, wind speed, sunshine hours and daily temperature amplitude. In all simulations of this study, wind speed was taken from the EWEMBI dataset, and sunshine hours were used directly from iLOVECLIM. The iLOVECLIM model does not have a daily cycle, thus daily temperature amplitudes were taken from observations (Climatic Research Unit (CRU) data, mean over 1901–2015) and kept constant for all time periods.

To assess the effect of the CDF-t application on the accuracy of simulated biome distribution, we used different sets of surface temperature, precipitations, and relative humidity taken (a) from the EWEMBI observation dataset (b) from the iLOVECLIM model, and (c) from the iLOVECLIM model after the CDF-t application to the outputs. The observation-enforced simulation was used as a control for the PI simulations.

The resulting computed biome distribution was compared to the BIOME6000 reconstruction (Harrison 2017). This reconstruction is based on the Paleovegetation Mapping Project (Prentice *et al* 1998, 2000). In BIOME6000, biomes are represented based on degrees of affinity of PFTs with each biome (Prentice *et al* 1996). In this work, we used megabiomes from BIOME6000, described by Harrison and Bartlein (2012). The 20 biomes in CARAIB were then reclassified into the nine megabiomes used in the BIOME6000 pollen database using the methodology of Extier (2019) (supplementary table 1).

Similarly to the inverse approach, we resampled BIOME6000 data points to the T21 and hi-res grids, using the majority resampling method (figure 6). The agreement was tested by calculating matching ratios between the resampled BIOME6000 and CARAIB-derived biomes before and after the CDF-t application. To account for the discrepancies in biome classification and spatial incoherency we also present the biome distribution in data and model in a form of contingency tables (supplementary tables 2–11), that outline frequency for particular combinations of biomes.

We found that the spatial irregularity of the data distribution of BIOME6000 and its absence in certain bioclimatic zones led to major regional inconsistencies. To overcome this data scarcity and irregularity across different regions and biomes (see section 4.1), we thus further compared our results with the statistically modelled gridded potential natural vegetation (PNV) distribution of Levavasseur *et al* (2012). The PNV is relying on BIOME6000 at the PI period as a predictand for a multinomial logistic regression on a high-resolution grid worldwide. Besides its regular grid spacing without missing data, it also has the

advantage of including a prediction of climatic zones where biomes remain are scarce or non-existent such as desertic areas (see section 4.1).

To test differences between the matching ratios before and after correction, we used the generalized Pearson chi-squared test of independence.

3. Results

In this section, we present the results of the bias correction of palaeoclimate modelling, starting with the outcomes of the test of monthly-based against yearly-based bias correction approach. These results are used to show the methodology for the bias correction. The test is followed by the results of inverse and forward accuracy evaluation approaches, which are used to assess the robustness of the developed technique.

3.1. Monthly-based and yearly-based bias correction approach

In order to evaluate the impact of methodological choices on our final result, we first apply both the monthly and yearly based CDF-t approaches, as highlighted above, to the PI climate. For the comparison, we used mean values of the temporal data ensemble for a grid cell, to evaluate the properties of the data ensemble in terms of spatial structure. The monthly-based approach results in higher Pearson correlation of precipitation to the observations. The SD of the monthly-based approach is also slightly better for the temperature and relative humidity than for yearly-based approach (figure 1). Additionally, the monthly-based approach allows for a better preservation of the raw monthly differences for temperature and precipitation between the different climate states (e.g. MH versus PI, not shown). We further tested monthly- and yearly-based approaches using the forward mode evaluation by comparison with pollen-based biome reconstructions (not included in the paper), and found consistently higher matching ratios using the monthly-based approach. Hence in the following, we report on results using the monthly-based approach.

### 3.2. Inverse mode. Comparison with paleotemperature record

The results of the comparison between the modelled climate parameters and proxy-reconstructed values from the Temperature 12k dataset are represented as scatterplots (figure 2). We observe significant improvement after the CDF-t application for the MH on the hi-res grid over Europe and on a T21 grid globally. No statistical difference was detected for the LGM.

In supplementary figures 2–7 we can observe a reduction in known large-scale iLOVECLIM biases, outlined in the literature (Goosse *et al* 2010). After the CDF-t application, North America is getting colder (supplementary figure 2) and, in parts, dryer (supplementary figures 3 and 4), which corrects the overestimation of temperatures in low latitudes in iLOVECLIM. Similarly, we observed a decrease in precipitations in subtropics, in particular in Asia and Australia (supplementary figure 3). The region of high precipitations is moved westward in Tropical Africa (supplementary figure 3), thus improving the representation of the recorded patterns of precipitations in these regions. Symmetry in the distribution of precipitation between the hemispheres is corrected with the increase in precipitations over the Northern hemisphere (supplementary figures 3 and 6).

The performance of the bias correction was further analysed using the RMSE, *R*, and SD of the modelled data with respect to Temperature 12k reconstructions, combined in Taylor diagrams (figure 3). Similarly to the figure 1, we used mean values of the temporal data ensemble for each grid cell.

We observe better *R* and RMSE in bias-corrected values for MH on a T21 grid globally ($R = 0.88$, RMSE = 6.32, SD = 9.37 before correction; $R = 0.91$, RMSE = 4.71, SD = 10.05 after correction) and on the hi-res grid over Europe ($R = 0.15$, RMSE = 4.13, SD = 3.46 before correction; $R = 0.46$, RMSE = 2.79, SD = 1.97 after correction). This indicates that temperatures after correction are closer to proxy-derived temperatures of Kaufman *et al* (2020), and thus confirms improvement of the temperature predictions after CDF-t bias correction for MH. The downscaled results show that after correction most of the data points of the simulated climate became colder than before correction (figure 4). These data points represent pollen data location, clustered in Scandinavia (the data points in figure 4 and supplementary figure 9 represent temperature values at site of the pollen reconstruction). Pollen-based reconstructions feature higher temperatures in Scandinavia than in our model and observations, which indicates one of the limitations of the proposed approach (see section 4.1). However, after correction, there are no more extreme outliers (8° and 9° warmer) relative to the reconstruction. The highest difference value after the correction is reduced to 5°. At the LGM, we observe no significant differences in *R* and an increased RMSE (figures 2 and 3), which indicates low similarity of the corrected model to the point reconstructions. The results are highly influenced by a low number of points with extremely negative values located in Greenland (figure 5, supplementary figure 9).
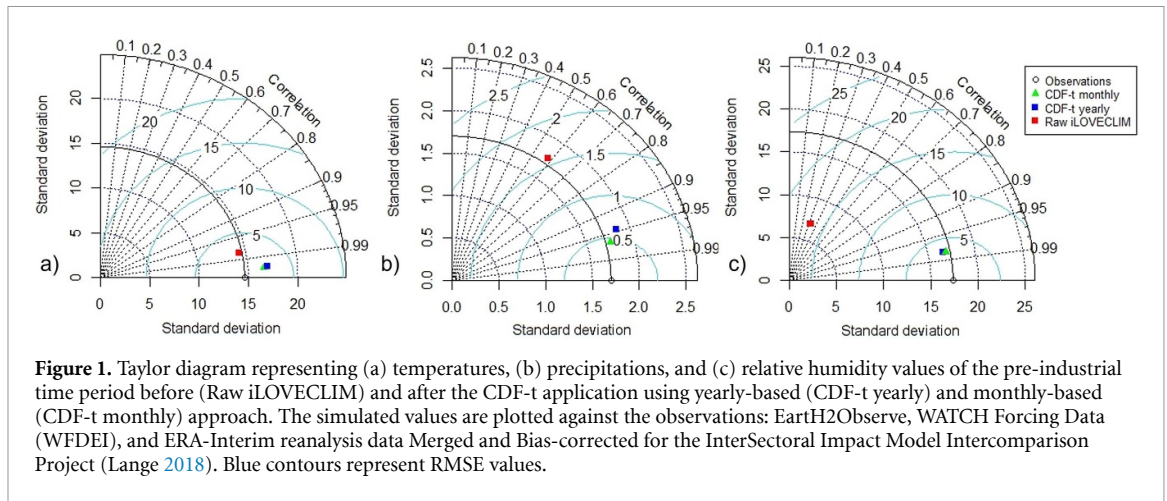
Figure 1. Taylor diagram representing (a) temperatures, (b) precipitations, and (c) relative humidity values of the pre-industrial time period before (Raw iLOVECLIM) and after the CDF-t application using yearly-based (CDF-t yearly) and monthly-based (CDF-t monthly) approach. The simulated values are plotted against the observations: EartH2Observe, WATCH Forcing Data (WFDEI), and ERA-Interim reanalysis data Merged and Bias-corrected for the InterSectoral Impact Model Intercomparison Project (Lange 2018). Blue contours represent RMSE values.
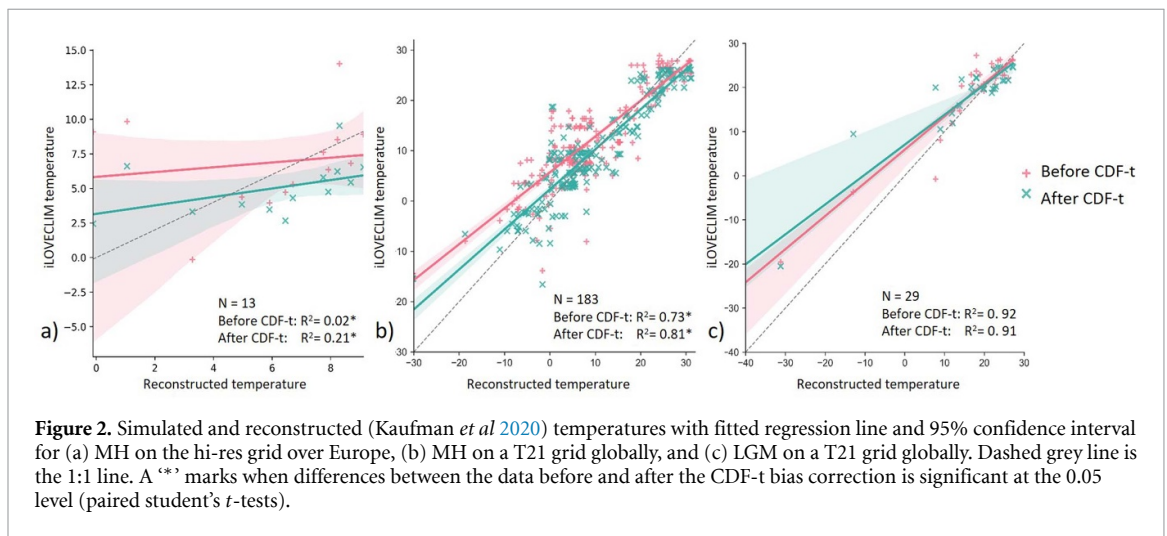
Figure 2. Simulated and reconstructed (Kaufman *et al* 2020) temperatures with fitted regression line and 95% confidence interval for (a) MH on the hi-res grid over Europe, (b) MH on a T21 grid globally, and (c) LGM on a T21 grid globally. Dashed grey line is the 1:1 line. A '*' marks when differences between the data before and after the CDF-t bias correction is significant at the 0.05 level (paired student's *t*-tests).
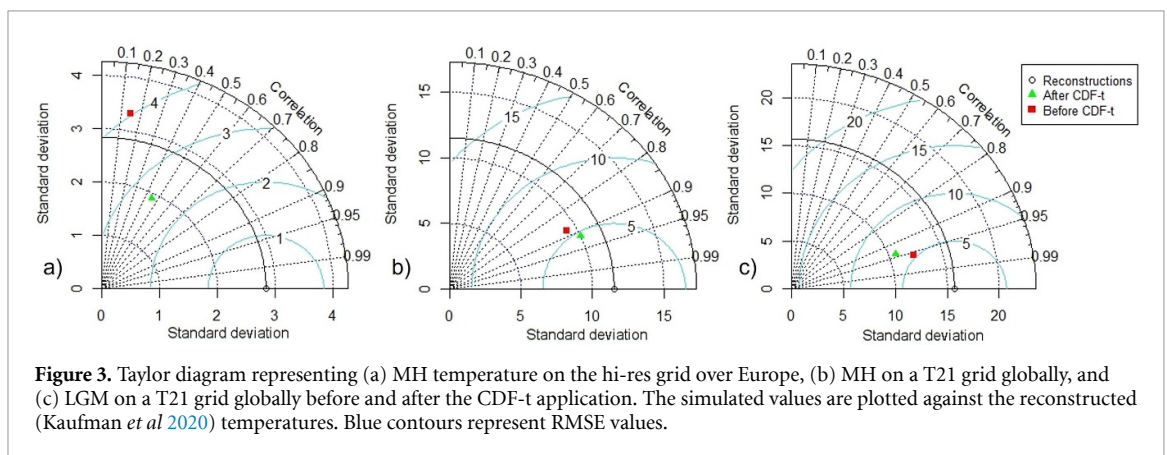
Figure 3. Taylor diagram representing (a) MH temperature on the hi-res grid over Europe, (b) MH on a T21 grid globally, and (c) LGM on a T21 grid globally before and after the CDF-t application. The simulated values are plotted against the reconstructed (Kaufman *et al* 2020) temperatures. Blue contours represent RMSE values.

3.3. Forward mode. Comparison with pollen-based vegetation reconstructions

To quantify the performance of the CDF-t bias correction method to simulate the global biome distribution, we compare the resulting biome maps with biome reconstructions provided by the BIOME6000 project (Harrison 2017). Since there is an inherent bias related to the use of the CARAIB model as well as the reconstruction dataset, we first compute biomes from CARAIB vegetation output using the EWEMBI observations (control) as an input. This biome output represents the best attainable fields using the bias correction. The difference between that reference field and the other biome fields is indicative of the remaining error in the climatic field.

Table 2 shows that after the CDF-t correction, the modelled biomes better match the BIOME6000 pollen data in all experiments except for the T21 PI, as opposed to non-corrected values. The improvement is
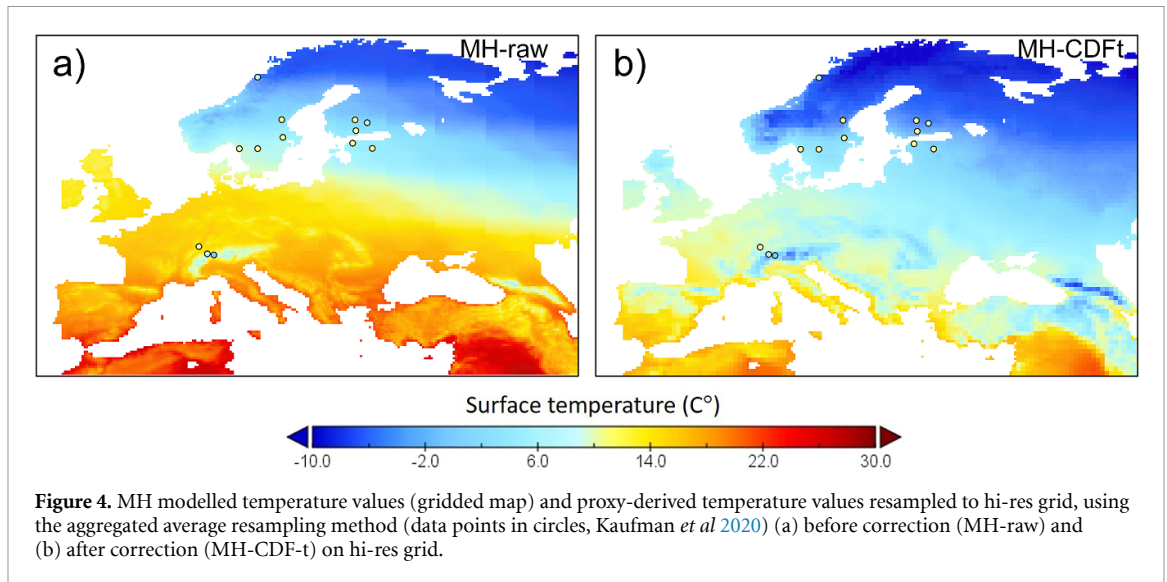
Figure 4. MH modelled temperature values (gridded map) and proxy-derived temperature values resampled to hi-res grid, using the aggregated average resampling method (data points in circles, Kaufman *et al* 2020) (a) before correction (MH-raw) and (b) after correction (MH-CDF-t) on hi-res grid.
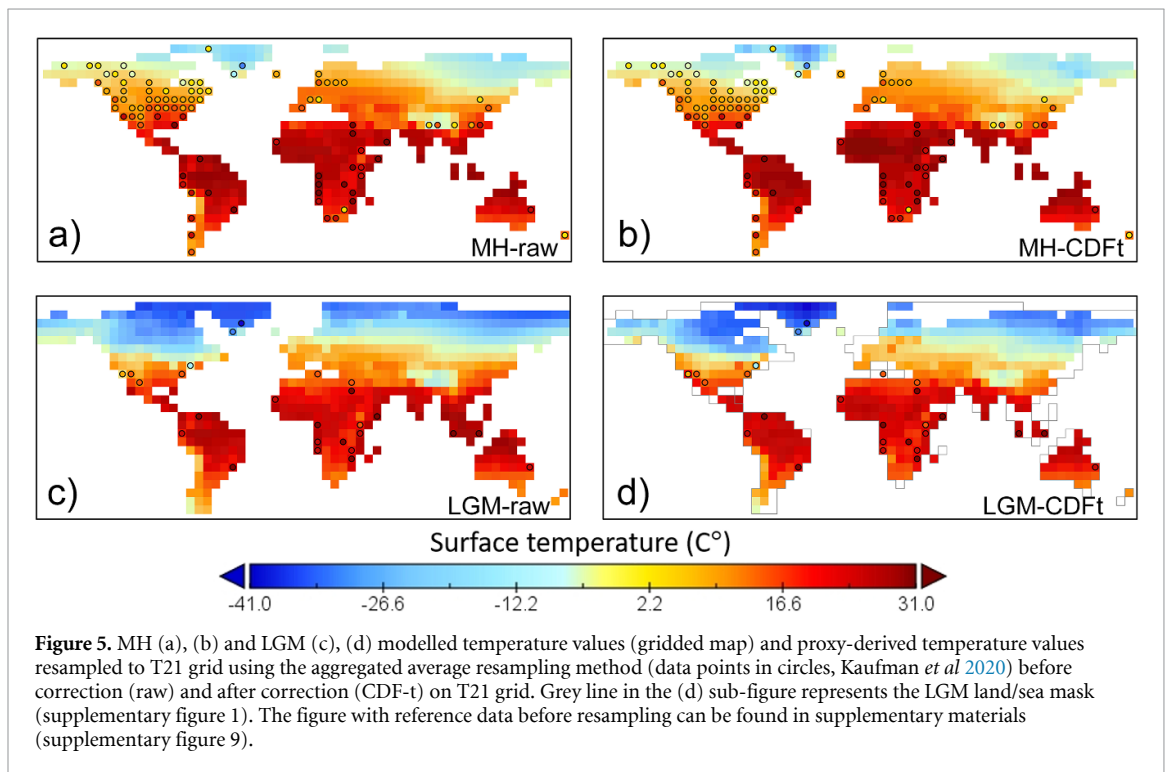
Figure 5. MH (a), (b) and LGM (c), (d) modelled temperature values (gridded map) and proxy-derived temperature values resampled to T21 grid using the aggregated average resampling method (data points in circles, Kaufman *et al* 2020) before correction (raw) and after correction (CDF-t) on T21 grid. Grey line in the (d) sub-figure represents the LGM land/sea mask (supplementary figure 1). The figure with reference data before resampling can be found in supplementary materials (supplementary figure 9).

Table 2. Matching ratios between megabiomes in pollen reconstructions (BIOME6000) and biomes predicted by the CARAIB model using: unaltered iLOVECLIM climatic variables (before correction); bias-corrected iLOVECLIM climatic variables using the CDF-t method on a monthly basis (after correction); and EWEMBI climatic variables (control). Hi-res: downscaled simulation over Europe; T21: standard global simulation. The megabiome classification consists of 9 biomes. A '*' marks when differences between the data before and after the CDF-t bias correction is statistically significant at the 0.05 level with a chi-squared test.

	hi-res EWEMBI (control)	T21 EWEMBI (control)	hi-res PI	hi-res MH	T21 PI	T21 MH	T21 LGM
Before CDF-t correction	0.39	0.47	0.18	0.31	0.36	0.38	0.22
After CDF-t correction	—	—	0.44*	0.40*	0.36	0.44*	0.25
N of grid cells	941	333	941	389	333	286	130

significant on the hi-res grid over Europe for PI and MH, and for MH on a T21 grid globally. Comparison with the PNV of Levavasseur *et al* (2012) (table 3) also indicates significant improvement at T21 grid globally and in Europe on the hi-res grid. Since the coverage in pollen samples and their accuracy varies widely among different continents (Harrison 2017, Chevalier *et al* 2020, Li *et al* 2022), we compare our results with

Table 3. Matching ratios per continent between megabiomes in Levavasseur *et al* (2012) pollen reconstructions (BIOME6000) and biomes predicted by the CARAIB model forced with: unaltered iLOVECLIM climatic variables (before correction); bias-corrected iLOVECLIM climatic variables using the CDF-t method on a monthly basis (after correction). A '*' marks when differences between the data before and after the CDF-t bias correction is statistically significant at the 0.05 level with a chi-squared test.

	Before CDF-t correction	After CDF-t correction	*N* of grid cells
Global	0.36	0.41*	286
North America	0.51	0.40	81
Australia	0.22	0.26	8
Asia	0.35	0.44*	94
South America	0.16	0.14	27
Africa	0.25	0.58*	34
Europe	0.51	0.39	42
Europe hi-res	0.39	0.49*	389

biome reconstructions for each continent separately (figures 6–8). The matching ratios for individual biomes can be found in supplementary tables 2–11.
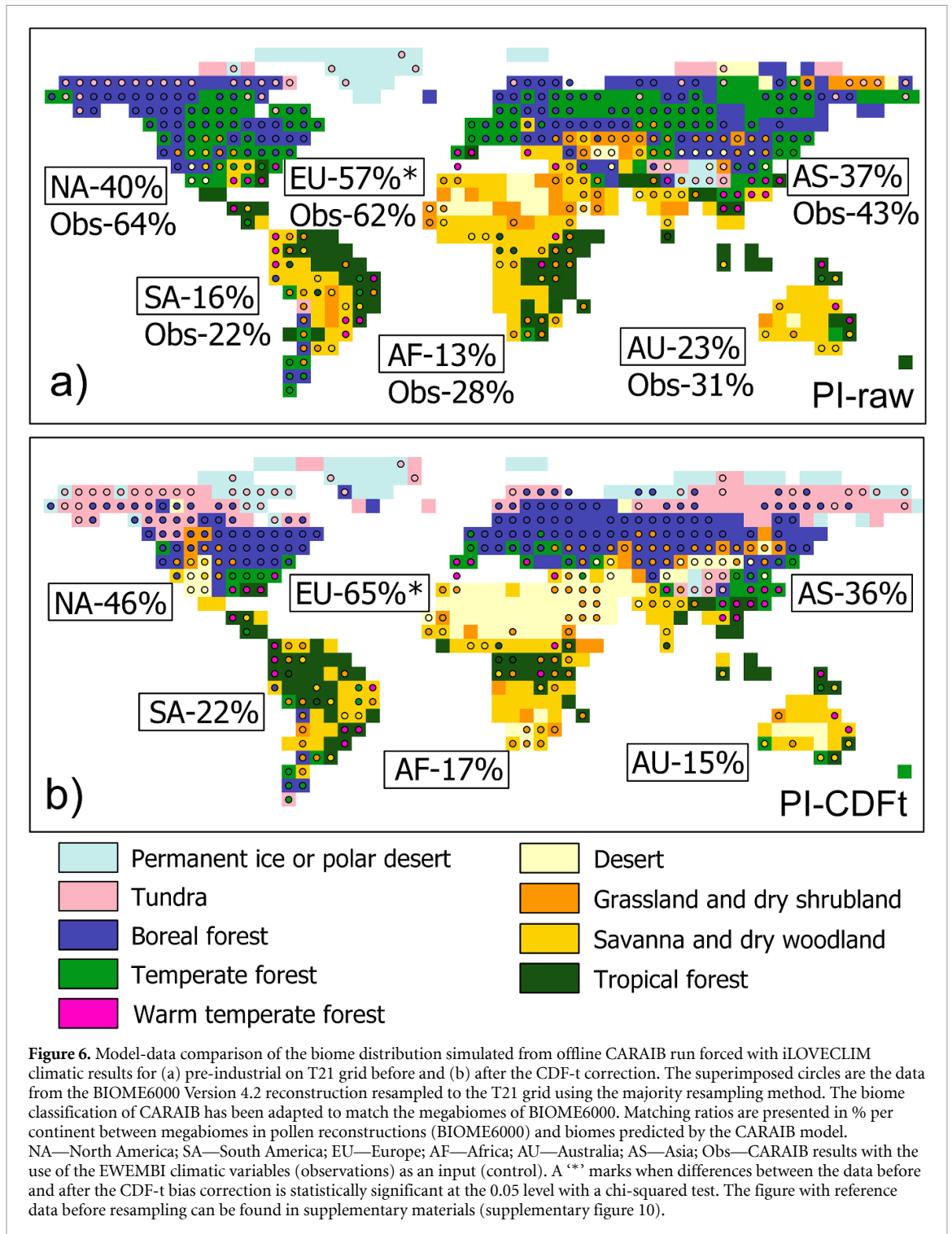
3.3.1. North America

This continent is well represented by high-quality pollen samples (Harrison 2017, Chevalier *et al* 2020, Li *et al* 2022), and thus has an exceptionally large number of BIOME6000 records. The matching ratios show an increase in agreement between the modelled and the reconstructed biomes of all three time periods post CDF-t application, which is significant for the MH (40 *versus* 53%). Figures 6 and 7 demonstrate a better representation of the southern extent of tundra over the continent for the PI and MH. The South-Western part of the continent became dryer after correction (supplementary figures 3 and 4), which led to a better representation of the biomes over the region.

3.3.2. Europe

The area is also well-represented in the BIOME6000 database. However, the number of records for LGM is quite low, and mostly represented by the grassland and dry shrubland data points near PD coastal areas, generally clustered around the Mediterranean. This lack of diversity in the reconstructions hampers obtaining a full picture for other biomes. At PI, in the Mediterranean region CARAIB simulates temperate (before CDF-t) and boreal forest (after CDF-t), thus no significant improvement in matching ratios is observed. However, the results indicate a significant improvement in biome representation for PI (57 *versus* 65%) and MH (50 *versus* 85%) after the CDF-t application. The lack of improvement in comparison with Levavasseur *et al* (2012) on a T21 grid (matching ratio 51% before and 39% after CDF-t) could be attributed to the coarse resolution of the model, since the downscaled simulation results in a significant increase in the matching ratios between the models after the CDF-t application (39 and 49% correspondingly) (table 3). Figure 9 indicates a better representation of the boreal forest in Central Europe in the corrected hi-res PI run. However, it still underestimates the Southward expansion of the boreal forest in Eastern Europe, especially under MH climate conditions. The CDF-t application decreases the temperature and increases annual precipitation levels of the Caspian regions (supplementary figures 5 and 6), which creates climatic conditions attributed to temperate forest biome by CARAIB and worsens the representation of biome distribution over the area in PI after bias correction. However, the above-mentioned changes in climate parameters improve the modelled biome distribution over the Balkans and Iberian Peninsula for both studied time periods. The parts of Southern Great Britain, identified as temperate forest in the BIOME6000 records, are classified by CARAIB before correction as tropical rainforest. The decrease in near-surface relative humidity after the CDF-t application (supplementary figure 7) transforms them into savannah and dry woodland in PI, and temperate and boreal forest in MH.

3.3.3. Asia

Asia appears to have lower matching ratios after the bias correction for each studied period, compared to non-bias corrected results, but these differences are not statistically significant (PI: 37 *versus* 36%, MH: 38 *versus* 35%, LGM: 33 *versus* 29%). Comparison with the PNV of Levavasseur *et al* (2012), however, indicates significant improvement in matching ratios after correction (35 *versus* 44%). This suggests that the results of comparison with reconstructions are negatively influenced by poor BIOME6000 data spatial coverage in Siberia, and highlights the role of spatial distribution and continuity of the reconstructions. CARAIB classifies cold desert as 'desert', and therefore, large areas of Siberia are assigned to this distribution. This negatively reflects on matching ratios for the region of Asia. Another aspect that contributes to the decrease in matching ratios over the entire continent, is the fact that the conditions of warm temperate forest are

Figure 6. Model-data comparison of the biome distribution simulated from offline CARAIB run forced with iLOVECLIM climatic results for (a) pre-industrial on T21 grid before and (b) after the CDF-t correction. The superimposed circles are the data from the BIOME6000 Version 4.2 reconstruction resampled to the T21 grid using the majority resampling method. The biome classification of CARAIB has been adapted to match the megabiomes of BIOME6000. Matching ratios are presented in % per continent between megabiomes in pollen reconstructions (BIOME6000) and biomes predicted by the CARAIB model. NA—North America; SA—South America; EU—Europe; AF—Africa; AU—Australia; AS—Asia; Obs—CARAIB results with the use of the EWEMBI climatic variables (observations) as an input (control). A '*' marks when differences between the data before and after the CDF-t bias correction is statistically significant at the 0.05 level with a chi-squared test. The figure with reference data before resampling can be found in supplementary materials (supplementary figure 10).

attributed to temperate forest in CARAIB. After CDF-t correction, the model overestimates the extent of tundra compared to reconstructions for all the studied time periods. However, the southern extent of the boreal forest after the CDF-t application is in better agreement with the BIOME6000 records and PNV of Levavasseur *et al* (2012).

3.3.4. South America

The extent and location of the Amazon rainforest after correction seem visually better represented (supplementary tables 3–7). However, this is not supported by the matching ratios. We observe a significant decrease in accuracy after bias correction for the MH (52 *versus* 26%). This is probably a consequence of the low number of pollen samples in the rainforest. The vast majority of data points from the region are found on the borders of the rainforest with other biomes and are, due to the coarse spatial resolution of our

Figure 7. Model-data comparison of the biome distribution simulated from offline CARAIB run forced with iLOVECLIM climatic results for (a) MH on T21 grid before and (b) after the CDF-t correction. The superimposed circles are the data from the BIOME6000 Version 4.2 reconstruction resampled to the T21 grid using the majority resampling method. The biome classification of CARAIB has been adapted to match the megabiomes of BIOME6000. Matching ratios are presented in % per continent between megabiomes in pollen reconstructions (BIOME6000) and biomes predicted by the CARAIB model. NA—North America; SA—South America; EU—Europe; AF—Africa; AU—Australia; AS—Asia; Obs—CARAIB results with the use of the EWEMBI climatic variables (observations) as an input (control). A '*' marks when d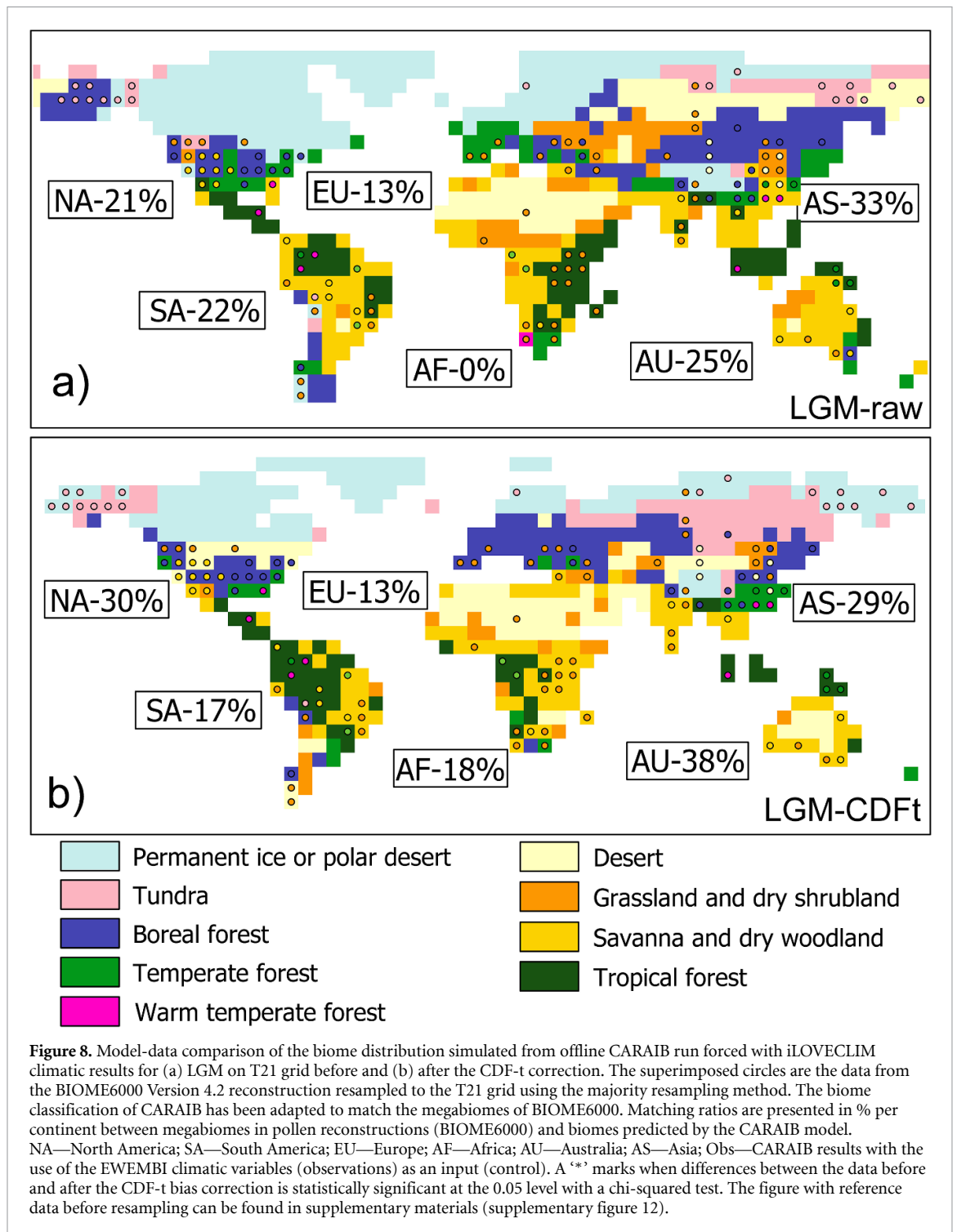ifferences between the data before and after the CDF-t bias correction is statistically significant at the 0.05 level with a chi-squared test. The figure with reference data before resampling can be found in supplementary materials (supplementary figure 11).

simulations, attributed to temperate, warm temperate forest, savannah, and dry woodland. Comparison with Levavasseur *et al* (2012) indicated no significant improvement (table 3), since the PNV of Levavasseur *et al* (2012) is based on BIOME6000 data, and thus tropical forest is largely replaced by warm-temperate forests in both datasets.

3.3.5. Australia

Due to the low number of BIOME6000 data points and model resolution, CDF-t has no significant impact on the matching ratios on the continent compared both to reconstructions and to PNV of Levavasseur *et al* (2012). Due to the lack of representation of biomes, not specific to pollen-based climate reconstructions, the desert is not reflected in the BIOME6000 dataset. Thus, here, as well as in North Africa and the Himalayas, the matching ratios are negatively biased by the absence of an important biome class in one of the compared

Figure 8. Model-data comparison of the biome distribution simulated from offline CARAIB run forced with iLOVECLIM climatic results for (a) LGM on T21 grid before and (b) after the CDF-t correction. The superimposed circles are the data from the BIOME6000 Version 4.2 reconstruction resampled to the T21 grid using the majority resampling method. The biome classification of CARAIB has been adapted to match the megabiomes of BIOME6000. Matching ratios are presented in % per continent between megabiomes in pollen reconstructions (BIOME6000) and biomes predicted by the CARAIB model. NA—North America; SA—South America; EU—Europe; AF—Africa; AU—Australia; AS—Asia; Obs—CARAIB results with the use of the EWEMBI climatic variables (observations) as an input (control). A '*' marks when differences between the data before and after the CDF-t bias correction is statistically significant at the 0.05 level with a chi-squared test. The figure with reference data before resampling can be found in supplementary materials (supplementary figure 12).
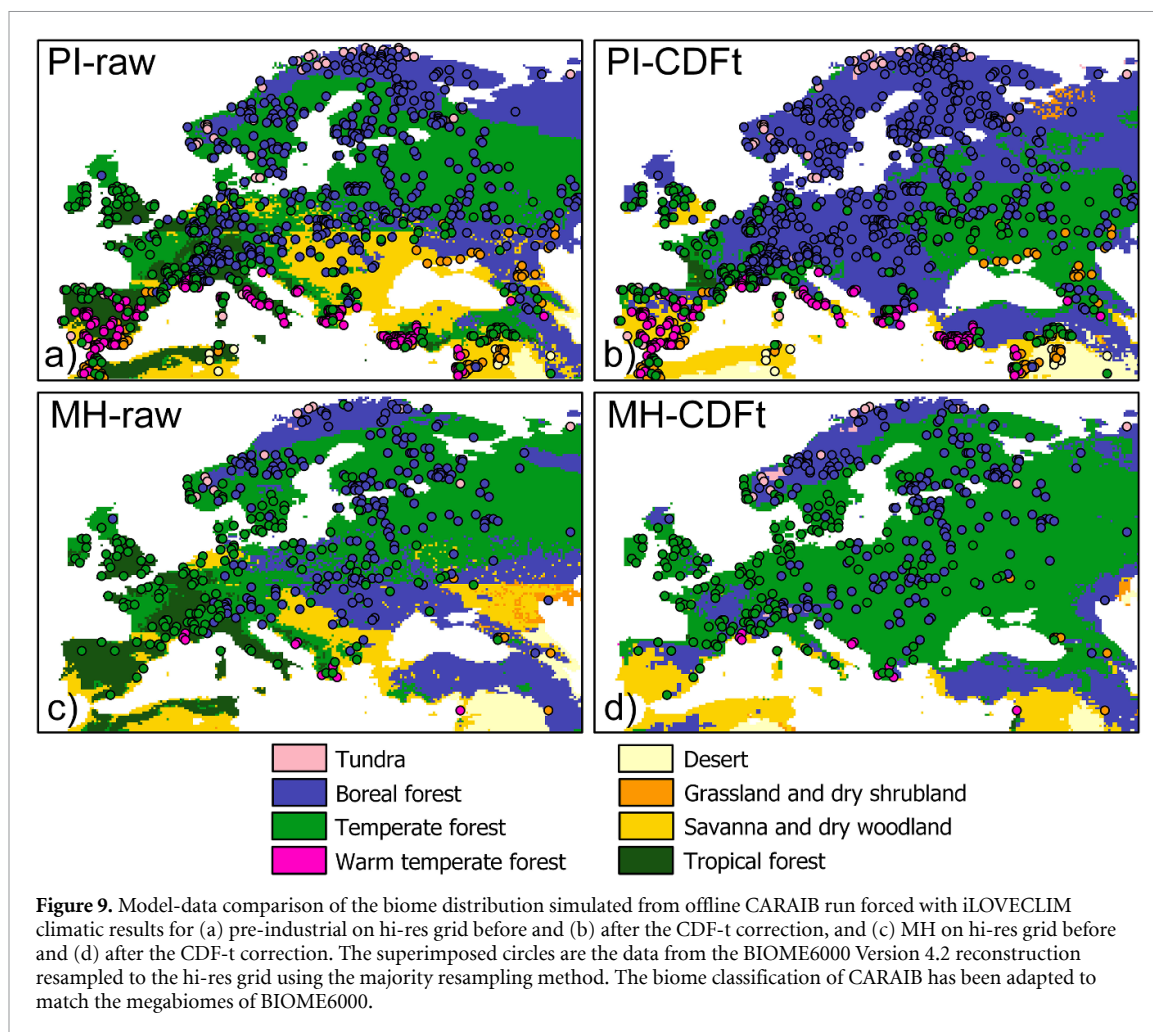
datasets. Taking this aspect into account, it is evident from figure 6 and supplementary figure 10 that the expansion of the Australian desert resulting from the bias correction in the model leads to a spatial distribution in better agreement with its PD size and location, and the biome distribution modelled by Levavasseur *et al* (2012) (supplementary figure 8).

3.3.6. Africa

The accuracy of modelled biome distribution increases with the CDF-t application for all three studied periods, but the improvement is not statistically significant (figures 6–8). The location of the Congo rainforest before correction is conditioned by the precipitation patterns of the region. In iLOVECLIM, the African precipitation bias is one of the large-scale biases, described by Goosse *et al* (2010). The CDF-t method reduces the bias over this region (supplementary figure 3, section 3.2), which results in a shift of the African rainforest to a location that is more consistent with its PD position and maps of Levavasseur *et al*

Figure 9. Model-data comparison of the biome distribution simulated from offline CARAIB run forced with iLOVECLIM climatic results for (a) pre-industrial on hi-res grid before and (b) after the CDF-t correction, and (c) MH on hi-res grid before and (d) after the CDF-t correction. The superimposed circles are the data from the BIOME6000 Version 4.2 reconstruction resampled to the hi-res grid using the majority resampling method. The biome classification of CARAIB has been adapted to match the megabiomes of BIOME6000.

(2012). Similarly to the Amazon rainforest, a large number of data points are located in the bordering regions with other biomes, which contributes to a decrease in matching ratios due to the model resolution. As previously described, the desert range is marked by the absence of pollen data and thus does not contribute to the matching ratios displayed in the figure 6. Similarly to the observation made for the Australian desert, an observation of the PD extent of the desertic areas in the Sahara and in our bias-corrected result shows that the desert is better represented overall. This is confirmed by comparing our biome distribution with the gridded PNV maps of Levavasseur *et al* (2012), which resulted in significantly better matching ratios after bias correction (25 *versus* 58%).

4. Discussion

By simulating the climate and the biome distributions for different past conditions and comparing outcomes before and after bias correction, our goal is to evaluate whether the CDF-t approach can significantly improve modelled fields with respect to reconstructions and field data. We additionally test the robustness of the method on different time periods and resolution grids. As within any modelling approach, the models used in our study (iLOVECLIM and CARAIB) entail biases that we do not try to evaluate since this has been done elsewhere (Warnant *et al* 1994, Otto *et al* 2002, Laurent *et al* 2008, Goosse *et al* 2010, Dury *et al* 2011). For the method to work, we also assume a form of stationarity of the biases that are being corrected relative to the PD. For the time periods analysed, the significant improvement in representing the PNV distribution across regions and climate regimes (see table 3) seems to indicate it is a valid assumption. There are however a number of limitations that are worth discussing in the following.

4.1. Uncertainty of CDF-t approach application and evaluation
While analysing our results, it is essential to understand the potential drawbacks and shortcomings of the methodology. Among the important limitations to consider when using this approach are those related to (1)

vegetation modelling procedure, (2) uncertainties of paleovegetation reconstructions, (3) uncertainties of biomization procedure, and (4) uncertainties of paleotemperature reconstructions.

4.1.1. Limitations related to multivariate correlations in vegetation modelling

As an input, CARAIB model requires six variables: near-surface air temperature, precipitation, relative humidity, wind speed, sunshine hours, and daily temperature amplitude. However, due to data availability of the observational dataset, we were only able to apply the CDF-t method to three of them (surface temperature, precipitation, relative humidity). CFD-t bias correction has been applied independently for each of the three variables. However, the application of univariate calibrations by CDF-t should not disrupt existing spatial coherency and dependence among variables. Indeed, as it is a QQ-based technique, CDF-t respects the dependence (i.e. coherency) between the variables of the simulations to be corrected. This means that a high (or low) simulated value is transformed to a high (respectively, low) corrected value and, more generally, that a simulation value corresponding to the p% percentile is transformed to the p% percentile of the correction values. As a consequence, the separate applications of CDF-t to various variables and gridcells do not modify the Spearman (i.e. rank) correlations from the simulations (Vrac 2018, François *et al* 2020). Therefore, the statistical spatial coherence and inter-variable dependencies are mostly preserved and come from the model simulations.

4.1.2. Limitations related to paleovegetation reconstructions

While being a quantitative approach to denote a similarity between datasets, there are several limitations to consider while interpreting paleovegetation reconstructions. Firstly, the BIOME6000 is an amalgamation of multiple data sets, each containing significant reconstruction uncertainties (Hengl *et al* 2018) related to pollen sampling and analysis methods (Prentice and Webb 1998, Harrison 2017). There are relatively little data for tropical South America, Africa, and Asia (Hengl *et al* 2018), where lower confidence in the results is expected. This is also true for other regions for the LGM because of the very few data points. BIOME6000 is a point-based dataset, which represents biomes of specific locations, often influenced by the local micro-climate and geo-environmental parameters. In this study, we resampled the BIOME6000 dataset to match the native grid of iLOVECLIM, which led to the loss of spatial detail in the reconstructions. Additionally, it created a bias in the regions, poorly represented in the BIOME6000 database, since scarcely located data points were considered representative of entire grid cells. This bias is particularly strong on a coarser resolution T21 grid, at bordering regions between the biomes, and in cases where the BIOME6000 data points were located near an edge of a grid cell. An example of such bias can be seen on the Western border of the Sahara desert at PI (figure 6, supplementary figure 10), where several reconstruction data points associate the region with grassland and dry woodland. While being in close vicinity with the corresponding biome class in CARAIB, a few of them fall at the edge of a desert grid cell, and therefore, do not contribute to the increase in matching ratios. This brings to our attention another limitation: a lack of representation of biomes, not specific to pollen-based climate reconstructions, i.e. desert and permanent ice or polar desert. Thus, the grid cells of the mentioned biomes in CARAIB are often attributed to BIOME6000 grassland and dry woodland, and tundra data points respectively (supplementary tables 2–11). To overcome this limitation, we used the gridded PNV distribution of Levavasseur *et al* (2012) which is constrained by biome reconstructions but is also providing biomes in regions where no pollen data is available such as desert. The analysis of the results alongside the PNV is, therefore, a promising approach, which provides an added benefit to interpretation of the modelled vegetation distribution.

4.1.3. Limitations related to biomization

As expected, a large caveat of the methodology arises due to differences in biomization procedures, as no universally acknowledged guidelines regulate the biomization of PFTs (Dallmeyer *et al* 2019). To further investigate the relationship between BIOME6000 and modelled biomes we analysed contingency tables of data distribution across different biomes (supplementary tables 2–11). The analysis revealed that the majority of data points of BIOME6000 were attributed to biomes with similar ternary coordinates (Pierce *et al* 2017) by the CARAIB model, and are, therefore, attributed to similar environmental and climatic conditions. For example, the megabiome of the warm temperate forest is not well reproduced by CARAIB under the applied biomization procedure. It is often misclassified as savanna and dry woodland, or tropical forest, with which it shares similar key species on a genus or family level (Ni *et al* 2010). The functional diversity of savannas (fire tolerance, functional ecology shade intolerance) is also not usually well-represented in DGVMs (Ratnam *et al* 2011, Dallmeyer *et al* 2019). This biome is unstable and vulnerable to grazing, fire regime, and climate, which transform savannas into forests or grasslands (Franco *et al* 2014). This feature is reflected in our findings since BIOME6000 savannas are largely represented by grasslands, dry shrublands, and tropical forests in CARAIB (supplementary tables 2–11). The limitations of

confinement of warm temperate forest and savanna biomes via PFT mixtures are further discussed by Dallmeyer *et al* (2019). Moreover, the biomization procedure is based on PD bioclimatic limits, which may not be applicable to glacial vegetation (Dallmeyer *et al* 2019). The biomization technique has been applied for the LGM but the conversion from PFT to biomes based on bioclimatic limits for the PI is considered constant over time which may not be the case. This may contribute to lower matching ratios at LGM compared to PI and MH (tables 2 and 3). The abovementioned patterns persist after the CDF-t application, as well as for the control simulation, which indicates that differences in biomization methods of BIOME6000 compared to CARAIB significantly influenced the inter-comparison of the two datasets.

4.1.4. Limitations related to paleotemperature reconstructions

Most published studies present the reconstructed, as well as modelled climate data in a form of anomalies (differences) to modern values (Mauri *et al* 2015, Cleator *et al* 2020). The CDF-t approach, however, preserves the evolutions of the input climate values by design, since the CDF-t corrects stationary biases under different mean states and preserves simulated by climate models anomalies. Therefore, useful data for the evaluation of the CDF-t methodology are scarce. Among the few available options, we selected the Temperature 12k dataset (Kaufman *et al* 2020) as a reference dataset for paleotemperature reconstructions. While being a compilation of non-continuous point data and featuring exclusively temperatures, it presents the data in a form of absolute values, which allows direct comparison with the modelled values prior to and following the CDF-t correction. In addition, the Temperature 12k is composed of consistent, quality-controlled paleotemperature records derived from the lake and marine sediments, peat, glacier ice, and other natural archives (Kaufman *et al* 2020). High standards of quality control of the featured data and consistency throughout a large number of records allow the use of the dataset as a reference in the context of the current study. Being largely based on pollen samples, the Temperature 12k dataset shares a similar spatial bias as the BIOME6000 dataset regarding quantification of deserts and other not specific to pollen-based reconstructions biomes (Chevalier *et al* 2020). Other uncertainties of the dataset are related to calibration and proxy biases, chronology, and spatiotemporal coverage (Kaufman *et al* 2020). In addition, a vast majority of the Temperature 12k data points were derived from marine proxies and thus were not used in this study, which is focused on the land surface. The effect of the limitations of the reconstruction datasets is more likely to be significant when the number of data points is low. This can be seen in the example of the downscaled MH. In figure 3, we compare the results of our simulations to the observation-based simulation. Temperature 12k indicates higher temperatures in Scandinavia than iLOVECLIM. This was also noted in comparison between the temperature reconstructions and PMIP4-CMIP6 simulations, where annual mean temperatures of the reconstructions for the MH were warmer at most latitudes, with maximum warming in the Arctic (Brierley *et al* 2020). However, the EWEMBI data show that this region is colder than in iLOVECLIM at the PD, thus it further lowers the temperatures over the region when applying the CDF-t. A similar phenomenon is observed in the Northern Hemisphere for the LGM. This indicates the significance of the impact of sampling biases in the data compilations used as a reference for calibration and evaluation of our methodology and contributes to the worsening of *R* and RMSE values for downscaled MH and LGM (figures 2 and 3).

4.2. Benefit of bias correction

Despite the above-mentioned limitations, the forward approach indicated significant improvement in the representation of biomes across different timesteps (tables 2 and 3), as well as across different regions (figures 6–8) and model resolutions (tables 2 and 3, figures 6–9). A lower matching ratio after CDF-t is systematically associated with non-significant results. This means that the correction is in fact not significantly deteriorating the results although the matching ratio is sometimes decreasing. Furthermore, we observe statistically significant improvement in bias-corrected biome representation of well-studied regions with high confidence in pollen reconstructions—Europe and North America (Harrison 2017, Chevalier *et al* 2020, Li *et al* 2022) (figures 6–8). Better representation of these regions in pollen databases, compared to rather poorly-covered South America, Africa, and Asia (Hengl *et al* 2018), attributes to higher confidence in the results of the comparison for the aforementioned regions. The T21 coarse grid is rather disadvantageous for capturing regionally confined biomes (i.e. savannah and dry woodland), which are better represented at finer downscaled resolution. However, large biome belts (such as tropical forest) are well captured following the bias correction procedure on both studied resolution grids (figures 6–9).

The use of the CDF-t bias correction improves the agreement for the PI with predicted PNV distribution based on multinomial logistic models (Levavasseur *et al* 2012) and the BIOME1 model results (Dallmeyer *et al* 2019).

Taylor diagrams (figures 1 and 3) indicate that the climatic parameters after the CDF-t application are statistically closer to observations and reconstructions in most cases, compared to non-corrected model

outputs. Moreover, they better agree with the temperature and precipitation patterns from the palaeoclimate simulations included in the current phase of PMIP for MH (Brierley *et al* 2020) and LGM (Kageyama *et al* 2017).

The results of the inverse accuracy evaluation approach indicated no major increase in accuracy in the PI period on the T21 grid after the CDF-t application since the iLOVECLIM model was calibrated with observations at PI (table 2). However, the downscaled simulations for this time period indicate a significant improvement in the agreement with the reconstructed data following the bias correction procedure. Thus, by combining the CDF-t method with the interactive downscaling, we can observe the added value of CDF-t. While acting as bias correction methods, they both target principally different biases. The design of interactive downscaling highly relies on topography, and thus it reduces biases related to topographic forcing (Quiquet *et al* 2018). The CDF-t method, when used for bias correction purposes, is designed to correct large-scale systemic model biases from one time period to another in a consistent manner (Vrac *et al* 2012). Combining the two methods into a joined workflow synthesises the aforementioned increase in large-scale model biases with the reduction in topography-related biases, provided by the embedded downscaling (Quiquet *et al* 2018).

5. Conclusions

This study was aimed: (i) to establish a methodology for correcting climatic biases in the context of palaeoclimate modelling; (ii) to test the robustness of the developed technique on three time periods of the past with contrasting climatic conditions, using proxy-based climate and vegetation reconstructions; and (iii) to evaluate the advantages and shortcomings of the proposed approach. We presented a CDF-t methodology for correcting biases of palaeoclimate modelling and evaluated its ability to perform bias correction of the modelled palaeoclimate variables of the iLOVECLIM climate model. The robustness of the developed technique was tested on three time periods (PI, MH, and LGM) and two spatial resolutions (5.625° and 0.25°).

Despite both forward (comparison with paleovegetation reconstructions) and inverse (comparison with paleotemperature reconstructions) accuracy evaluation approaches having limitations, we achieved a significant improvement in the agreement between the modelled and reconstructed data by the application of CDF-t to the modelled values. We have demonstrated that placing the CDF-t technique as an additional post-processing step of modelled data is a clear advantage to correct palaeoclimate model data for systematic biases. Our conclusion is supported by a set of statistical parameters (matching ratios, R, RMSE) showing better agreement between the models and the reconstruction datasets for the PI and MH. We found higher correlation between the reconstructed (Kaufman *et al* 2020) and the modelled temperatures after bias correction on both 5.625° ($R = 0.88$, RMSE $= 6.32$, SD $= 9.37$ before correction; $R = 0.91$, RMSE $= 4.71$, SD $= 10.05$ after correction) and 0.25° grid ($R = 0.15$, RMSE $= 4.13$, SD $= 3.46$ before correction; $R = 0.46$, RMSE $= 2.79$, SD $= 1.97$ after correction). Comparison with BIOME6000 reconstructions resulted in significant increase in matching ratios after bias correction for the MH at T21 grid at the 0.05 level (0.38 before and 0.44 after CDF-t).

The application of the proposed methodology to the downscaled climate variables further improved the accuracy and the level of detail captured by the model (matching ratios 0.31 before and 0.40 after CDF-t for the MH, and 0.18 and 0.44 for the PI respectively) and outlined an added value of the combination of the embedded downscaling with the bias correction that could be further extended to other regions.

At the LGM we observed no significant improvement in comparison with pollen-based reconstructions, which may suggest limitations in the use of the PD bioclimatic limits to describe a glacial environment. However, the results are largely determined by a low number of data points in the reconstructions, so further research is needed.

Our study reports the first attempt of bias correcting palaeoclimate model outputs using the CDF-t approach. To further assess the usefulness of the CDF-t application in palaeoclimate modelling the proposed methodology remains to be tested on more time periods, using climate models of different complexity and conducting perfect model experiments.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Anhelina Zapolska ⦿ https://orcid.org/0000-0002-8045-8016
Mathieu Vrac ⦿ https://orcid.org/0000-0002-6176-0439
Aurélien Quiquet ⦿ https://orcid.org/0000-0001-6207-3043
Thomas Extier ⦿ https://orcid.org/0000-0002-9531-8753
Frank Arthur ⦿ https://orcid.org/0000-0002-9217-2058
Hans Renssen ⦿ https://orcid.org/0000-0002-5104-0526
Didier M Roche ⦿ https://orcid.org/0000-0001-6272-9428

References

Arthur F, Roche D M, Fyfe R, Quiquet A and Renssen H 2023 Simulations of the Holocene climate in Europe using an interactive downscaling within the iLOVECLIM model (version 1.1) *Clim. Past* **19** 87–106

Bartlein P J *et al* 2011 Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis *Clim. Dyn.* **37** 775–802

Berger A L 1978 Long-term variations of daily insolation and Quaternary climatic changes *J. Atmos. Sci.* **35** 2361–7

Beyer R, Krapp M and Manica A 2020 An empirical evaluation of bias correction methods for palaeoclimate simulations *Clim. Past* **16** 1493–508

Bjune A E, Bakke J, Nesje A and Birks H J B 2005 Holocene mean July temperature and winter precipitation in western Norway inferred from palynological and glaciological lake-sediment proxies *The Holocene* **15** 177–89

Brierley C M *et al* 2020 Large-scale features and evaluation of the PMIP4-CMIP6 mid Holocene simulations *Clim. Past* **16** 1847–72

Brovkin V, Ganopolski A and Svirezhev Y 1997 A continuous climate-vegetation classification for use in climate-biosphere studies *Ecol. Modell.* **101** 251–61

Chevalier M *et al* 2020 Pollen-based climate reconstruction techniques for late Quaternary studies *Earth-Sci. Rev.* **210** 103384

Claussen M *et al* 2002 Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models *Clim. Dyn.* **18** 579–86

Cleator S F, Harrison S P, Nichols N K, Colin Prentice I and Roulstone I 2020 A new multivariable benchmark for Last Glacial Maximum climate simulations *Clim. Past* **16** 699–712

Dallmeyer A, Claussen M and Brovkin V 2019 Harmonising plant functional type distributions for evaluating Earth system models *Clim. Past* **15** 335–66

Déqué M 2007 Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values *Glob. Planet. Change* **57** 16–26

Díaz-Pacheco J, van Delden H and Hewitt R 2018 The importance of scale in land use models: experiments in data conversion, data resampling, resolution and neighborhood extent *Geomatic Approaches for Modeling Land Change Scenarios* (Cham: Springer) pp 163–86

Dury M, Hambuckers A, Warnant P, Henrot A, Favre E, Ouberdous M and François L 2011 Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity *IForest* **4** 82–99

Extier T 2019 Variations climatiques et variations du cycle hydrologique aux basses latitudes au cours du Quaternaire : une approche combinant modèle et données *PhD Thesis* Université Paris-Saclay (available at: www.theses.fr/2019SACLV056)

Finné M, Woodbridge J, Labuhn I and Roberts C N 2019 Holocene hydro-climatic variability in the Mediterranean: a synthetic multi-proxy reconstruction *The Holocene* **29** 847–63

Franco A C, Rossatto D R, de Carvalho Ramos Silva L and da Silva Ferreira C 2014 Cerrado vegetation and global change: the role of functional types, resource availability and disturbance in regulating plant community responses to rising CO_2 levels and climate warming *Theor. Exp. Plant Physiol.* **26** 19–38

François B, Vrac M, Cannon A J, Robin Y and Allard D 2020 Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth Syst. Dyn.* **11** 537–62

François L M, Delire C, Warnant P and Munhoven G 1998 Modelling the glacial-interglacial changes in the continental biosphere *Glob. Planet. Change* **16–17** 37–52

Furlanetto G, Ravazzi C, Pini R, Vallè F, Brunetti M, Comolli R, Novellino M D, Garozzo L and Maggi V 2018 Holocene vegetation history and quantitative climate reconstructions in a high-elevation oceanic district of the Italian Alps. Evidence for a middle to late Holocene precipitation increase *Quat. Sci. Rev.* **200** 212–36

Goosse H *et al* 2010 Description of the Earth system model of intermediate complexity LOVECLIM version 1.2 *Geosci. Model Dev.* **3** 603–33

Goosse H and Fichefet T 1999 Importance of ice-ocean interactions for the global ocean circulation: a model study *J. Geophys. Res. Ocean* **104** 23337–55

Guiot J, Harrison S P and Colin Prentice I 1993 Reconstruction of Holocene precipitation patterns in Europe using pollen and lake-level data *Quat. Res.* **40** 139–49

Guiot J and Kaniewski D 2015 The Mediterranean Basin and Southern Europe in a warmer world: what can we learn from the past? *Front. Earth Sci.* **3** 28

Guiot J, Torre F, Jolly D, Peyron O, Boreux J J and 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 *Ecol. Modell.* **127** 119–40

Guo L, Jiang Z, Chen D, Le Treut H and Li L 2020 Projected precipitation changes over China for global warming levels at 1.5 °C and 2 °C in an ensemble of regional climate simulations: impact of bias correction methods *Clim. Change* **162** 623–43

Haddad Z S and Rosenfeld D 1997 Optimality of empirical Z-R relations *Q. J. R. Meteorol. Soc.* **123** 1283–93

Hansen J E and Sato M 2012 Paleoclimate implications for human-made climate change *Climate Change: Inferences from Paleoclimate and Regional Aspects* (Vienna: Springer) pp 21–47

Harrison S P 2017 BIOME 6000 DB classified plotfile version 1 *Univ. Reading Dataset* (https://doi.org/10.17864/1947.99)

Harrison S P, Bartlein P J and Prentice I C 2016 What have we learnt from palaeoclimate simulations? *J. Quat. Sci.* **31** 363–85

Harrison S P and Bartlein P 2012 Records from the past, lessons for the future: what the palaeorecord implies about mechanisms of global change *Future World's Climate* (Amsterdam: Elsevier) pp 403–36 (available at: https://researchers.mq.edu.au/en/publications/records-from-the-past-lessons-for-the-future-what-the-palaeorecor)

Harrison S P and Prentice C I 2003 Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: Analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations *Glob. Change Biol.* **9** 983–1004

Harrison S P, Yu G E and Tarasov P E 1996 Late Quaternary lake-level record from northern Eurasia *Quat. Res.* **45** 138–59

Haywood A M *et al* 2019 What can Palaeoclimate modelling do for you? *Earth Syst. Environ.* **3** 1–18

Hengl T, Walsh M G, Sanderman J, Wheeler I, Harrison S P and Prentice I C 2018 Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential *Peer J.* **6** e5457

Jones P W 1999 First- and second-order conservative remapping schemes for grids in spherical coordinates *Mon. Weather Rev.* **127** 2204–10

Kageyama M *et al* 2017 The PMIP4 contribution to CMIP6—part 4: scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments *Geosci. Model Dev.* **10** 4035–55

Kageyama M *et al* 2018 The PMIP4 contribution to CMIP6—part 1: overview and over-arching analysis plan *Geosci. Model Dev.* **11** 1033–57

Kaufman D *et al* 2020 A global database of Holocene paleotemperature records *Sci. Data* **7** 1–34

Keery J, Holden P, Edwards N, Monteiro F, Ridgwell A, Keery J, Holden P, Edwards N, Monteiro F and Ridgwell A 2016 Insights into the paleoclimate of the PETM from an ensemble of EMIC simulations *EGU General Assembly* vol 18 pp EPSC2016–13948 (available at: https://ui.adsabs.harvard.edu/abs/2016EGUGA.1813948K/abstract)

Kohfeld K E and Harrison S P 2000 How well can we simulate past climates? Evaluating the models using global palaeoenvironmental datasets *Quat. Sci. Rev.* **19** 321–46

Koutavas A and Joanides S 2012 El Niño-Southern oscillation extrema in the Holocene and Last Glacial Maximum *Paleoceanography* **27** 4208

Kuhnt T, Schmiedl G, Ehrmann W, Hamann Y and Andersen N 2008 Stable isotopic composition of Holocene benthic foraminifers from the Eastern Mediterranean Sea: past changes in productivity and deep water oxygenation *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **268** 106–15

Lambeck K, Rouby H, Purcell A, Sun Y and Sambridge M 2014 Sea level and global ice volumes from the Last Glacial Maximum to the Holocene *Proc. Natl Acad. Sci. USA* **111** 15296–303

Lange S 2018 Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset *Earth Syst. Dyn.* **9** 627–45

Laurent J M, François L, Bar-Hen A, Bel L and Cheddadi R 2008 European bioclimatic affinity groups: data-model comparisons *Glob. Planet. Change* **61** 28–40

Lavaysse C, Vrac M, Drobinski P, Lengaigne M and Vischel T 2012 Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario *Nat. Hazards Earth Syst. Sci.* **12** 651–70

Levavasseur G, Vrac M, Roche D M and Paillard D 2012 Statistical modelling of a new global potential vegetation distribution *Environ. Res. Lett.* **7** 44019–30

Lhardy F *et al* 2021 A first intercomparison of the simulated LGM carbon results within PMIP-Carbon: role of the Ocean boundary conditions *Paleoceanogr. Paleoclimatol.* **36** e2021PA004302

Li C, Postl A K, Böhmer T, Cao X, Dolman A M and Herzschuh U 2022 Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0) *Earth Syst. Sci. Data* **14** 1331–43

Li H, Renssen H and Roche D M 2019 Global vegetation distribution driving factors in two dynamic global vegetation models of contrasting complexities *Glob. Planet. Change* **180** 51–65

Li H, Renssen H and Roche D M 2020 Modeling climate-vegetation interactions during the last interglacial: The impact of biogeophysical feedbacks in North Africa *Quat. Sci. Rev.* **249** 106609

Luu L N, Vautard R, Yiou P, van Oldenborgh G J and Lenderink G 2018 Attribution of extreme rainfall events in the South of France using EURO-CORDEX simulations *Geophys. Res. Lett.* **45** 6242–50

M R D 2013 Delta O-18 water isotope in the iLOVECLIM model (version 1.0)—part 1: implementation and verification *Geosci. Model Dev.* **6** 1481–91

Mauri A, Davis B A S, Collins P M and Kaplan J O 2015 The climate of Europe during the Holocene: a gridded pollen-based reconstruction and its multi-proxy evaluation *Quat. Sci. Rev.* **112** 109–27

Meinshausen M *et al* 2017 Historical greenhouse gas concentrations for climate modelling (CMIP6) *Geosci. Model Dev.* **10** 2057–116

Mesta B and Kentel E 2022 Superensembles of raw and bias-adjusted regional climate models for Mediterranean region, Turkey *Int. J. Climatol.* **42** 2566–85

Michelangeli P A, Vrac M and Loukos H 2009 Probabilistic downscaling approaches: application to wind cumulative distribution functions *Geophys. Res. Lett.* **36** L11708

Ni J, Yu G, Harrison S P and Prentice I C 2010 Palaeovegetation in China during the late Quaternary: biome reconstructions based on a global scheme of plant functional types *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **289** 44–61

Noël T, Loukos H, Defrance D, Vrac M and Levavasseur G 2021 A high-resolution downscaled CMIP5 projections dataset of essential surface climate variables over the globe coherent with the ERA5 reanalysis for climate change impact assessments *Data Brief* **35** 106900

Opsteegh J D, Haarsma R J, Selten F M and Kattenberg A 1998 ECBILT: a dynamic alternative to mixed boundary conditions in ocean models *Tellus* A **50** 348–67

Otto D, Rasse D, Kaplan J, Warnant P and François L 2002 Biospheric carbon stocks reconstructed at the Last Glacial Maximum: comparison between general circulation models using prescribed and computed sea surface temperatures *Glob. Planet. Change* **33** 117–38

Peyron O *et al* 2017 Precipitation changes in the Mediterranean basin during the Holocene from terrestrial and marine pollen records: a model-data comparison *Clim. Past* **13** 249–65

Peyron O, Guiot J, Cheddadi R, Tarasov P, Reille M, De Beaulieu J L, Bottema S and Andrieu V 1998 Climatic reconstruction in Europe for 18,000 YR B.P. from pollen data *Quat. Res.* **49** 183–96

Pierce S *et al* 2017 A global method for calculating plant CSR ecological strategies applied across biomes world-wide *Funct. Ecol.* **31** 444–57

Prentice C I, Guiot J, Huntley B, Jolly D and Cheddadi R 1996 Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka *Clim. Dyn.* **12** 185–94

Prentice C I and Webb T 1998 BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records *J. Biogeogr.* **25** 997–1005

Prentice I C *et al* 2000 Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa *J. Biogeogr.* **27** 507–19

Prentice I C, Harrison S P, Jolly D and Guiot J 1998 The climate and biomes of Europe at 6000 yr BP: comparison of model simulations and pollen-based reconstructions *Quat. Sci. Rev.* **17** 659–68

Quiquet A, Dumas C, Paillard D, Ramstein G, Ritz C and Roche D M 2021 Deglacial ice sheet instabilities induced by proglacial lakes *Geophys. Res. Lett.* **48** e2020GL092141

Quiquet A, Roche D, Dumas C, Paillard D and Roche D M 2018 Online dynamical downscaling of temperature and precipitation within the iLOVECLIM model (version 1.1) *Geosci. Model Dev.* **11** 453–66

Ramstein G, Kageyama M, Guiot J, Wu H, Hély C, Krinner G and Brewer S 2007 How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison *Clim. Past* **3** 331–9

Ratnam J, Bond W J, Fensham R J, Hoffmann W A, Archibald S, Lehmann C E R, Anderson M T, Higgins S I and Sankaran M 2011 When is a "forest" a savanna, and why does it matter? *Glob. Ecol. Biogeogr.* **20** 653–60

Roche D M, Dumas C, Bügelmayer M, Charbit S and Ritz C 2014 Adding a dynamical cryosphere to iLOVECLIM (version 1.0): coupling with the GRISLI ice-sheet model *Geosci. Model Dev.* **7** 1377–94

Seppä H and Birks H J B 2001 July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: pollen-based climate reconstructions *The Holocene* **11** 527–39

Vigaud N, Vrac M and Caballero Y 2013 Probabilistic downscaling of GCM scenarios over southern India *Int. J. Climatol.* **33** 1248–63

Vrac M 2018 Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (R2D2) bias correction *Hydrol. Earth Syst. Sci.* **22** 3175–96

Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L and Somot S 2012 Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment *Nat. Hazards Earth Syst. Sci.* **12** 2769–84

Warnant P, François L, Strivay D and Gérard J-C 1994 CARAIB: a global model of terrestrial biological productivity *Glob. Biogeochem. Cycles* **8** 255–70

Wohlfahrt J, Harrison S P and Braconnot P 2004 Synergistic feedbacks between ocean and vegetation on mid- and high-latitude climates during the mid-Holocene *Clim. Dyn.* **22** 223–38

Wu H, Guiot J, Brewer S and 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 *Clim. Dyn.* **29** 211–29

Zhang Y, Renssen H and Seppä H 2016 Effects of melting ice sheets and orbital forcing on the early Holocene warming in the extratropical Northern Hemisphere *Clim. Past* **12** 1119–35

Zhang Y, Renssen H, Seppä H and Valdes P J 2018 Holocene temperature trends in the extratropical Northern Hemisphere based on inter-model comparisons *J. Quat. Sci.* **33** 464–76