A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years

Mario Krapp[1,2*]
Robert Beyer[1,3]
Stephen L. Edmundson[1,4]
Paul J. Valdes[5]
Andrea Manica[1]

1. Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom 2. GNS Science, PO Box 31312, Lower Hutt 5040, New Zealand 3. Potsdam Insitute for Climate Impact Research, Telegrafenberg A 31, 14473 Potsdam, Germany 4. Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands 5. School of Geographical Sciences, University of Bristol, BS8 1SS Bristol, United Kingdom *corresponding author(s): Mario Krapp (mariokrapp@gmail.com)

Abstract

Curated global climate data have been generated from climate model outputs for the last 120,000 years, whereas reconstructions going back even further have been lacking due to the high computational cost of climate simulations. Here, we present a statistically-derived global terrestrial climate dataset for every 1,000 years of the last 800,000 years. It is based on a set of linear regressions between 72 existing HadCM3 climate simulations of the last 120,000 years and external forcings consisting of CO2, orbital parameters, and land type. The estimated climatologies were interpolated to 0.5° resolution and bias-corrected using present-day climate. The data compare well with the original HadCM3 simulations and with long-term proxy records. Our dataset includes monthly temperature, precipitation, cloud cover, and 17 bioclimatic variables. In addition, we derived annual net primary productivity and global biome distributions using the BIOME4 vegetation model. The data are a relevant source for different research areas, such as archaeology or ecology, to study the long-term effect of glacial-interglacial climate cycles for periods beyond the last 120,000 years.

Background & Summary

Studying the ecology and environment throughout past climatic changes often involves environmental reconstructions that are either based on paleoclimate proxies or on paleoclimate simulations. Unfortunately, even by today's standards, simulating the climate over periods of thousands or hundreds of thousands of years in a continuous way can still be a costly and time-consuming endeavour. Present or future climate simulations are based on comprehensive Global Climate Models (GCMs) that resolve processes at high temporal and spatial resolution, such as those used in the fifth IPCC Assessment Report [1]. Climate model reconstructions for longer, continuous periods back in time are, therefore, challenging. They have to span a much longer period and are, thus, computationally too expensive. Instead,

GCMs provide snapshots for a specific time or short transients in the order of a few thousand years. For longer, transient simulations of tens or hundreds of thousands of years, we rely on simulations from Earth System Models of Intermediate Complexity (EMICs) [2, 3] but they come at the cost of lower spatial resolution and a simplified representation of the climate system [4].

Although there are a some high-resolution paleoclimate data sets readily available for download, for example, *WorldClim* [5] or *ecoClimate* [6], their temporal coverage is limited to a few snapshots of key periods in the past, for example, Mid-Holocene (6,000 before present (BP)) the Last Glacial Maximum (21,000 BP), or the Last Interglacial Period (130,000 BP). An exception is *PaleoView* [7] which covers the transient period of the last deglaciation, but this only goes back 21,000 years. Longer, continuous climate data sets of the past, based on HadCM3 [8] snapshots, have become available more recently, for example a Northern Hemisphere data set for the last 60,000 years [9] or a bias-corrected, high-resolution terrestrial climate data set of the last 120,000 years [10].

We used a linear regression model to extend existing HadCM3 climate simulations of the last 120,000 years to create climate reconstructions of the last 800,000 years. We then applied a bias correction [11] of the model output using present-day gridded observational data (CRU TS v. 4.04 [12]) to downscale climate output to a final horizontal resolution of 0.5° [10, 12]. This new data set is complementary to the aforementioned high-resolution terrestrial climate data set of the last 120,000 years [10] (which should be preferred for studies of the last glacial cycle, as they are based directly on the GCM output), and it is an extension for readers to explore the climate history for earlier periods of the past.

In this paper, we present annual and monthly mean climatologies for the last 800,000 years in 1000 year time steps (Table 1). The data set includes air temperature, precipitation, total cloud cover, 17 bioclimatic variables [13], as well as net primary productivity and biomes, the latter based on BIOME4 simulations [14] run on the debiased climatologies. We validated the long-term climate change signal using time series of various proxy records.

Methods

Our climate reconstructions are based on a set of linear regression models for each of the HadCM3 model grid boxes (N=96×73=7008). Each linear model predicts a climate variable, which can be either temperature, precipitation, or total cloud cover, i.e., the dependent variable. The independent variables, i.e., the forcing terms, of the model are three orbital parameters, atmospheric CO_2, and a surface type mask (land, ocean, or land ice), five variables in total.

Each linear model uses 72 data points, the HadCM3 snapshots throughout the past 120 ka [15, 16]. These snapshots cover both the Last Glacial Maximum, one of the coldest glacial stages, and the Last Interglacial, one of the warmest interglacial stages during the Middle and Late Pleistocene. By applying available long-term forcing to the solutions of the linear models, we reconstructed the climate for periods before 120 ka. The forcing consists of CO_2 [17], interpolated to 1 ka intervals for the last 800 ka, orbital parameters, taken from numerical solution to the Earth's orbit around the sun [18], and surface type masks based on numerical ice-sheet model output [2] and a global sea-level record [19]. At this stage, the reconstructed climate of the last 800,000 years has the same coarse spatial resolution as the underlying HadCM3 snapshots. In a last step, we applied a bias correction (including spatial downscaling) for the terrestrial climate to derive a spatially explicit data set that covers the last 800,000 years in 1000-year

intervals with a spatial resolution of $0.5°\times0.5°$. For each variable, these steps have been repeated for every monthly mean climatologies (Jan–Dec), as well as for the annual mean values (Table 1).

A comprehensive overview of our approach is shown in Fig. 1 and further details of our experimental setup are given below.

The HadCM3 climate model

HadCM3 is a fully coupled global climate model with an atmospheric component, HadAM3, which has a horizontal resolution of $3.75°\times2.5°$, 19 vertical levels, and a time step of 30 minutes. The ocean and sea-ice component of HadCM3 has a horizontal resolution of $1.25°\times1.25°$ and 20 vertical levels. HadCM3 simulations were run with a prescribed ice-sheet and continental geometry. We use the 72 available HadCM3 simulations covering the last 120,000 years in 2000-year intervals from 120,000 to 24,000 ka before present (BP) and in 1000-year intervals from 22,000 to present-day [15, 16].

Surface type mask: Ice-sheet extents, sea level and lakes

As ice sheet extents for the period outside the HadCM3 snapshots, we used model outputs from CLIMBER-2/SICOPOLIS simulations [2] for which Northern Hemisphere ice sheet extents and heights are available for the last 800 ka in 1 ka-year intervals. For the more recent period from 122–0 ka, we used the ice sheet configurations from the ICE-6G data set [20, 21]. Changes in the coast lines affecting the land–sea mask were derived from a global sea-level record [19]. We overlaid those changes on top of present-day coast lines, taken from the ETOPO1 data set [22, 23], while we preserved inland lakes which were taken from the *Global Lakes and Wetlands Database* [24, 25].

The linear regression model

For each HadCM3 grid box, we have fitted a linear regression model to a local series of each climatic variable of interest (temperature, precipitation, or cloud cover) with the following independent variables or forcings: atmospheric CO_2 concentrations (as a major greenhouse gas), three variables reflecting the orbital forcing [18], and the surface type, which is either ocean, land, or land ice. The orbital parameters are based on obliquity ε and two combinations of eccentricity e and precession ω: $e\sin\omega$, henceforth referred to as precession index I, and $e\cos\omega$ (precession index II), and they are a generally accepted set of orbital forcings [26, 27]. We chose temperature T, precipitation P, and total cloud cover C as dependent variables The independent variables, i.e., the predictors, are the normalised forcings.

More formally, let $Y(x,t)$ be a time series of a climate variable in a specific grid box x at time t. Our linear model should explain variations of Y, ΔY, around a mean value $\overline{Y}$:

$$\Delta Y = Y - \overline{Y}. \tag{1}$$

To make the linear model well-conditioned, all independent variables have been normalised. The mean has been subtracted and the result has then been divided by the standard deviation.

Precipitation and total cloud cover are bounded variables and that can lead to linear model predictions outside of allowed ranges. A common procedure to prevent these out of bounds predictions is to apply a transformation to the data beforehand. To prevent the linear model from predicting negative precipitation values, we therefore applied a logarithmic transformation to precipitation, which maps values

from [0,+∞] to [−∞,+∞]. Thus, in the case of precipitation, the linear regression coefficients predict the response in terms of anomalies in the exponent. For total cloud cover, expressed as fraction between 0 and 1, we choose a logit transformation which maps values from [0,1] to [−∞,+∞]. The decomposition of temperature T, precipitation P, and total cloud cover C, i.e., the ΔY on the left hand side of Eq. (1) is:

$$T = \overline{T} + \underbrace{\Delta T}_{\hat{=}\Delta Y} \tag{2}$$

$$\log(P) = \overline{\log(P)} + \underbrace{\Delta\log(P)}_{\hat{=}\Delta Y} \tag{3}$$

$$\log\left(\frac{C}{1-C}\right) = \overline{\log\left(\frac{C}{1-C}\right)} + \underbrace{\Delta\log\left(\frac{C}{1-C}\right)}_{\hat{=}\Delta Y} \tag{4}$$

The linear regression model for each (transformed) anomaly is:

$$\begin{aligned}\Delta Y(x,t) = &\underbrace{\beta_1(x)\Delta\varepsilon(t)}_{\text{obliquity forcing}} + \underbrace{\beta_2(x)\Delta(e\sin\omega)(t)}_{\text{precession index I forcing}} + \underbrace{\beta_3(x)\Delta(e\cos\omega)(t)}_{\text{precession index II forcing}} \\ &+ \underbrace{\beta_4(x)\Delta CO_2(t)}_{\text{greenhouse gas forcing}} + \underbrace{\beta_5(x)\Delta M(x,t)}_{\text{land/land ice/ocean effect}}\end{aligned} \tag{5}$$

Here, β_1 to β_5 are the regression coefficients for the respective predictors (see Fig. 2 for maps of β coefficients). Surface type changes are captured by the categorical variable $M(x,t) \in [\text{ocean}, \text{land}, \text{land ice}]$. For example, around coastlines land grid boxes can turn into ocean grid boxes when sea level is high. Similarly, expanding ice sheets turn land grid boxes into ice-covered grid boxes, and the climate variable Y may respond to different surface types in different ways.

We solved the linear model for each transformed variable, applied the extended forcing to generate the 800 ka climate reconstructions, and transformed the resulting data back to its original range according to Eq. (2)–(4).

Training and test data

We divided the HadCM3 snapshots into a training (80%) and a test data set (20%). The training data set has been used to fit the linear model while the test data has been used for a comparison to the reconstructed snapshots. For a 80/20 division of the 72 time slices into training and test data, i.e., 58/14, there are $\binom{n}{k} = \binom{72}{58} \approx 3\times10^{14}$ possible combinations of snapshots. But instead of randomly dividing the snapshots into the training/test data, we followed an approach with the aim to preserve as much variance as possible in the training data, i.e. maximise the variance of the predictors. This is best illustrated by the phase plots of the predictors (Fig. 4). The training data set covers the edges of each phase plane and thus maximises the phase space covered by the linear regression model. This choice of training data ensured that the linear regression model interpolated within the phase space and did not need to extrapolate for the test data.

The procedure has been as follows. We calculated the covariance matrix of the full parameter set (n=72), $\mathscr{C}_{full}$. Then, we randomly created a sample training data set (k=58) for which we computed the

covariance matrix $\mathscr{C}_{sample}$. If the the eigenvalues of $\mathscr{C}_{sample}$ were larger than the eigenvalues of $\mathscr{C}_{full}$, then the training sample data set contained at least as much variance as the full data set and this sample training data set was marked as a candidate for the final training set. After several iterations (N=10,000), we summed up how many times each time slice had appeared within a candidate training set. We then ranked all time slices according to this number. In the final step, we picked the 80% top-ranked time slices as training data.

Note, the division into training and test data set had been made only for the validation of the linear model. For the actual climate reconstructions, we used all 72 snapshots to make full use of the whole available data.

Spatial downscaling to 0.5° and bias correction

The CRU TS climate data set

For the bias correction of the HadCM3 model output we used variables from the CRU TS (Climatic Research Unit Timeseries) data set (v. 4.04) [12]. Before the bias correction step, the data set has been bi-linearly interpolated from the HadCM3 resolution of 3.75°×2.5°to the final CRU TS resolution of 0.5°×0.5°. CRU TS v. 4.04 contains monthly time series fields of precipitation, daily maximum and minimum temperatures, cloud cover, and other variables covering all land areas (except Antarctica) for 1901 to 2019. As reference period for the bias correction we chose 1961–1990. We applied the additive "delta" method, which is the most effective bias correction with respect to paleoclimate reconstructions [11], to all climate variables as they have been predicted by the linear model (subscript LM) to create the final, bias-corrected climate data set $\widehat{Y}(x,t)$:

$$\widehat{Y}(x,t) = Y(x,t)_{\mathrm{LM}} + [Y(x,0)_{\mathrm{CRU\ TS}} - Y(x,0)_{\mathrm{LM}}]\,. \tag{6}$$

The BIOME4 vegetation model

We used the BIOME4 model [14] to calculate annual net primary productivity (NPP) and 28 different biomes. BIOME4 is coupled biogeography and biogeochemistry model that simulates competition of different plant functional types (PFTs). It optimises the leaf area of each PFT as a function of NPP. BIOME4 forcing consists of monthly values of temperature, precipitation, and sunshine percentage, as well as values of annual minimum temperature and atmospheric CO_2 concentrations. Further inputs, which were kept constant through time, are water holding capacity and percolation rate.

In its default setup, BIOME4 does not incorporate changes in insolation. To mimic variations of available sunshine due to past orbital variations, we adjusted total cloud cover by calculating the relative changes in insolation for the past [28] and then varying the cloud cover accordingly. Using the adjusted total cloud cover and insolation, we were then able to calculate sunshine percentage [29]. For atmospheric CO_2 we have used the same data set as described earlier [17].

Data Records

All data records are publicly available as NetCDF files in the project repository in the `data` directory [30].

Monthly mean and annual mean climatologies

The climate variables that are part of our data set are listed in Table 1 and can be downloaded as NetCDF files (Table 2) from the *Open Science Framework* data repository [30]. All climate variables are available on a $0.5° \times 0.5°$ resolution, i.e., the same regular grid as the CRU TS v4.04 data set [12].

Bioclimatic variables

For ecological applications, bioclimatic variables can be more relevant than the actual climate variables [13]. Many bioclimatic variables can be directly derived from monthly mean temperature and precipitation data, and are thus included in our data records (Table 1, 2).

Net primary productivity and biomes

Our data record contains annual net primary productivity as well as categorical biome data, both of which have been calculated by the BIOME4 vegetation model using the 800 ka of reconstructed climate (Table 1, 2).

Technical Validation

Comparison to original HadCM3 simulations

We validated our reconstruction from the linear model against the original HadCM3 snapshots. As comparison metric, we used R^2 values, a goodness of fit estimator that measures the proportion of variance explained by the linear model, and the root mean squared error (RMSE), an estimator of the goodness of the model that measures how far the linear model predictions are from the HadCM3 test data (Fig. 5).

Overall, our linear model is a better predictor for temperature than for precipitation or total cloud cover. Temperature responds more directly to local forcings than precipitation or cloud cover, because it is determined by the energy balance of downward and upward longwave and shortwave radiation and turbulent heat fluxes. The downward shortwave radiation depends on incoming solar radiation that is determined by orbital variations, whereas downward longwave radiation is determined by greenhouse gases such as CO_2 and water vapour, as well as cloud cover. Large-scale atmospheric circulation changes have a much smaller effect on temperature. The high R^2 and low RMSE values in most regions (Fig. 5A and B) mean that temperature is locally well constrained by global CO_2 and orbital variations and our linear model captures this effect well.

The matter is more complicated for precipitation and cloud cover. Both variables are directly affected by the hydrological cycle which itself depends on large-scale atmospheric dynamics, such as monsoonal systems in the tropics and subtropics, or mid-latitude storm systems. Local interactions between the atmosphere and the surface, such as evaporation and transpiration over the ocean, or deep convection over the tropics, matter to a lesser extent. Instead, processes and circulation features like moisture transport or the atmospheric Hadley cell dynamics determine the non-local response of precipitation (and cloud cover) to CO_2 or orbital variations to a much larger extent. Because of the larger dynamical component of the hydrological cycle, precipitation and cloud cover are much less constrained by the forcing than temperature. As a result, the linear model shows less predictive skill for precipitation and total cloud cover (Fig. 5C–F).

Our reconstruction represented only long-term climatologies of past climate changes, similar to other GCM snapshot of the past [15, 10]. Therefore, the data set does not contain sub-millennial scale variability.

The spatial covariance of the model output

Although the set of linear models for each grid box are independent, we can show that the reconstructed climate fields exhibit the same spatial structure as the original HadCM3 model output. First of all, the regression coefficient maps (Fig. 2) indicate that the climate response to external forcings is spatially coherent. We calculated this spatial coherence in terms of the spatial covariance matrix (Fig. 3). It consists of the covariance between the time series of any two grid points, i.e., it is a 7008×7008 matrix ($96 \times 73 = 7008$) of covariances over the 72 snapshots.

For both, HadCM3 and the linear model, the covariance matrices are similar in structure and magnitude, and their differences are relatively small (Fig. 3). The covariance matrices are much more similar for temperature than for precipitation or total cloud cover, because the linear regression works better for temperature than for the other two climate fields. Overall, the covariance matrices of the linear model reconstructions are so similar to HadCM3 that we can conclude that the spatial structure is indeed preserved.

Comparison to marine and terrestrial proxies

We compared the reconstructed climate data with marine proxies (before the downscaling/bias-correction step) as a means of highlighting how well the reconstructed long-term climatologies compare to empirical reconstructions.

Marine sediment cores are valuable archives of past sea surface temperature (SST) records. Because their associated bio-geochemistry is relatively straightforward, marine proxies can be utilised as paleothermometers and are thus well suited for a direct proxy–model comparison. For these proxies, we make a direct comparison between mean annual temperature (MAT) and SST quantified both in terms of correlation between the predicted and observed time series and the RMSE. We note that MAT and SST are not the same climatological quantities; SST is the temperature of the ocean surface and has a lower limit of about -1.8°C, the freezing point of saltwater. While we expect MAT and SST to co-vary in low and mid-latitudes, at higher latitudes, seasonal or perennial sea ice cover makes a comparison between both variables problematic.

For terrestrial proxies, for which a translation into climatic variables is not straightforward, we simply calculated the correlation between the two time series. However, the interpretation of terrestrial proxies from a climate perspective can also be problematic. For example, pollen-based vegetation reconstructions are suggested to be less reliable as climate proxies, particularly for interglacials [31]. Other land-based proxies such as dust deposits integrate long-term climatic changes over large regions and hence do not necessarily capture climatic effects at their specific location.

For the comparison of our climate reconstructions with proxy reconstruction, we assembled long-term marine SST and terrestrial climate proxy reconstructions (Figs. 6,7) which cover a period of at least 150 ka BP during the last 800 ka (Tables 3 and 4). Our climate reconstructions are representative of the modelled HadCM3 climate for both marine and terrestrial proxy locations.

Model bias

Because the linear model has been fitted to match HadCM3 model outputs, the resulting bias of our reconstructions is similar to the HadCM3 model bias. The bias of reconstructed temperature and precipitation with respect to the CRU TS data set has a similar spatial pattern (Fig. 8) and is as large as the bias shown for HadCM3 climate reconstructions of the last 60 ka [9]. This is also why our long-term reconstructions show the same bias towards the assembled paleo-climate proxy records as the original HadCM3 simulations (Figs. 6,7). From this, we concluded that this similarity means that our present-day climate reconstruction is of the same quality as the original HadCM3 simulation it is based on. As discussed earlier, the bias has been removed from our climate reconstructions using the "delta" method [11].

Code availability

Model code for the linear regression as well as the code for the analysis and visualisation of figures is publicly available in the project repository [30]. NetCDF files have been processed using `cdo` [32]. We used the Python language for most of our scripts with a few bash scripts as wrappers. The workflow for whole data generation process is managed by *Snakemake* [33]. The linear regression is based on the `statsmodels` package [34]. All visualisations are made with `matplotlib` [35] using `cartopy` [36] for maps. Other Python packages that have been used are (in alphabetical order): `adjustText` [37], `BeautifulSoup4` [38], `netCDF4` [39], `numpy` [40], `pandas` [41, 42], `scipy` [43], `scikit-image` [44] , and `tqdm` [45].

Usage Notes

Examples on how to access the NetCDF files of the reconstructed climate using *Python* or *R* are provided in the `examples` directory of the project repository [30].

Acknowledgements

This work was supported by an ERC Consolidator Grant to AM (Local Adaptation 647787). MK was also supported by the NZ Ministry for Business, Innovation and Employment through the Antarctic Science Platform (ANTA1801). We thank Eric Wolff, Max Holloway, Peter Hopcroft, Chris Brierley, and Michel Crucifix for commenting on early versions of this manuscript.

Author contributions

AM and MK devised the project. MK devised and implemented the linear regression model with input from RB and SLE. PJV provided additional *HadCM3* snapshot simulations. MK and AM wrote a first draft of the paper which was improved by input from all other authors. MK performed the analysis and prepared the figures.

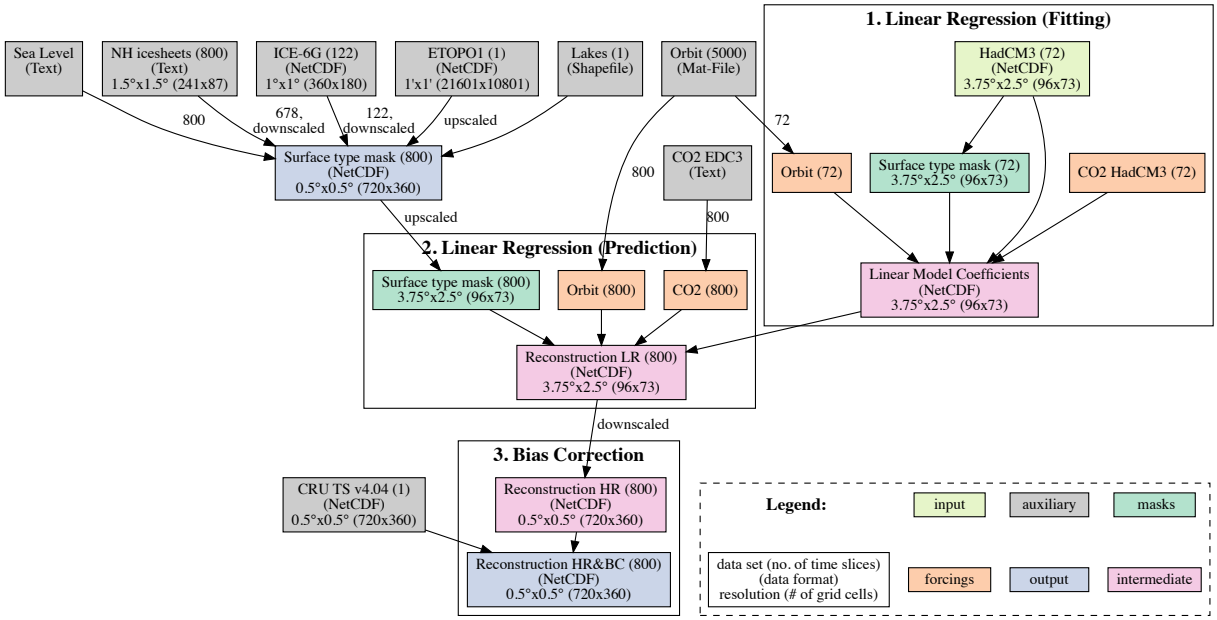

Figure 1: Flowchart showing how the different data sets have been used as input for the different stages of the paleo-climate data generation: A linear regression combines 72 low-resolution (LR) HadCM3 snapshot simulations with the external forcings, i.e., CO_2, orbital parameters, and surface type masks (ocean, land, land ice), which provides the basis of the long-term climate reconstructions using long-term forcings and surface type masks. The final bias correction procedure yields the high-resolution (HR), bias-corrected (BC) climate data set for the last 800 ka.

Competing interests

The authors declare no conflict of interest.

Figures

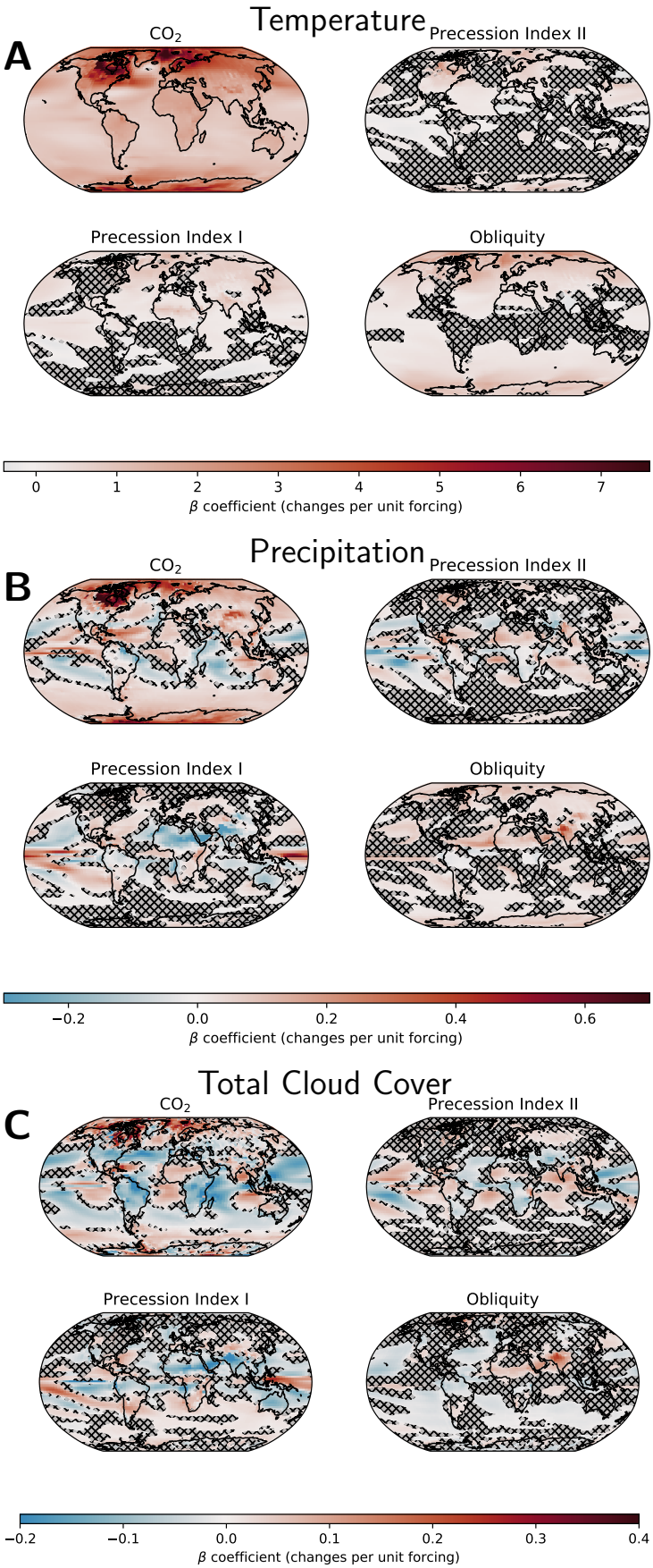

Figure 2: Regression coefficients, i.e., β coefficients, for (A) mean annual temperature, (B) precipitation, and (C) total cloud cover. Regions where the respective coefficient is not statistically significant ($p < 0.05$) are hatched and shaded.

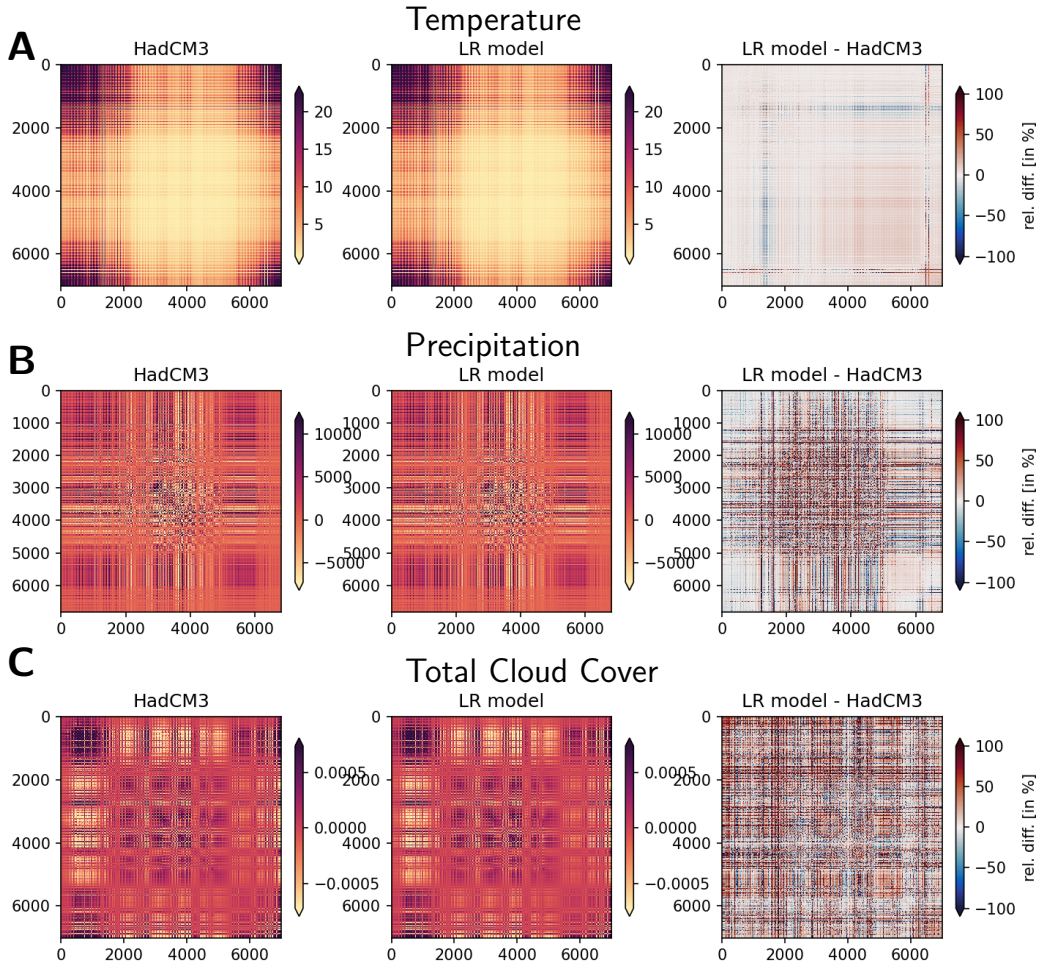

Figure 3: Spatial covariance matrix of (A) temperature (in units K^2), (B) precipitation (in units mm^2/a^2), and (C) total cloud cover (in units of 1^2) for HadCM3, our linear regression model (LR model), and the difference between the two. Each value represents a covariance matrix element from a flattened vector with the length of the total number of grid points (n=7008). The covariance matrices are symmetric and thus are their differences.

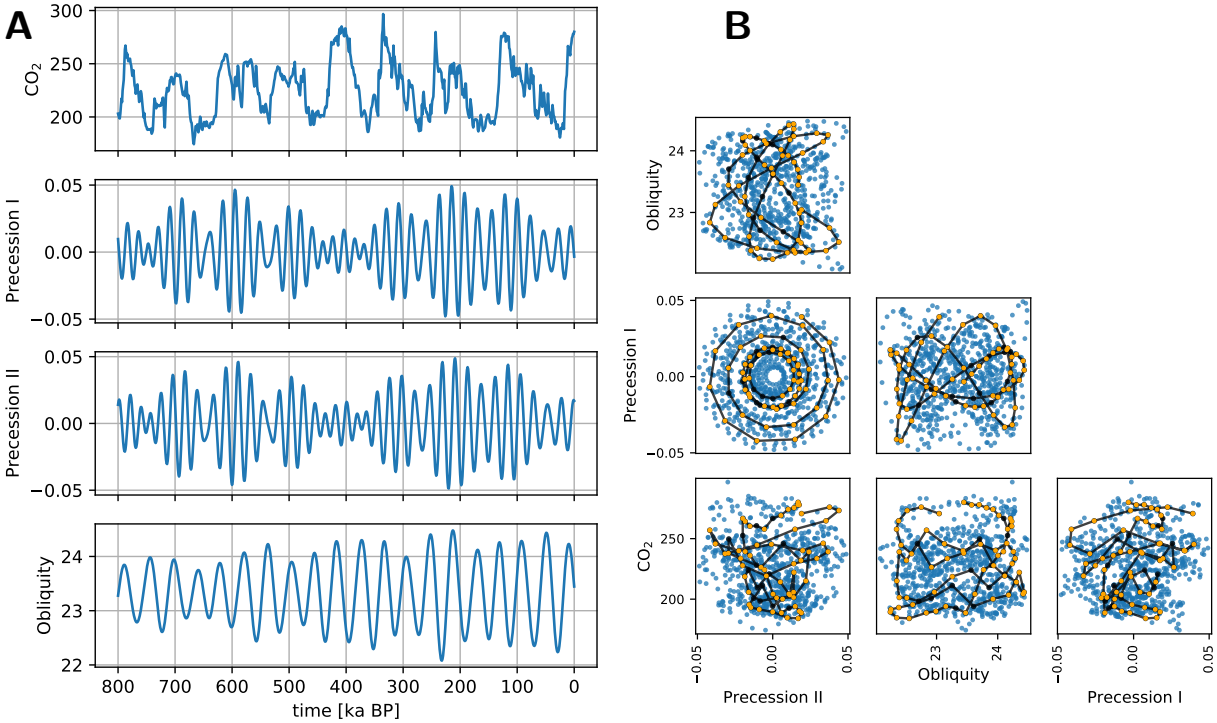

Figure 4: (A) Time series of the four external parameters: CO_2 and orbital parameters for the last 800 ka and (B) the associated parameter space as scatter plot matrix (blue dots). The continuous CO_2 record is from the EPICA Dome C ice core in Antarctica [17]. The orbital parameters are numerical solutions for the Earth's orbit and rotation in terms of eccentricity, precession, and obliquity [18]. In (B), black lines with black dots represent the total 72 parameter sets. Orange dots highlight the parameter sets of the 58 HadCM3 snapshot simulations which we used as training data (80% of the total 72) for the linear regression model.

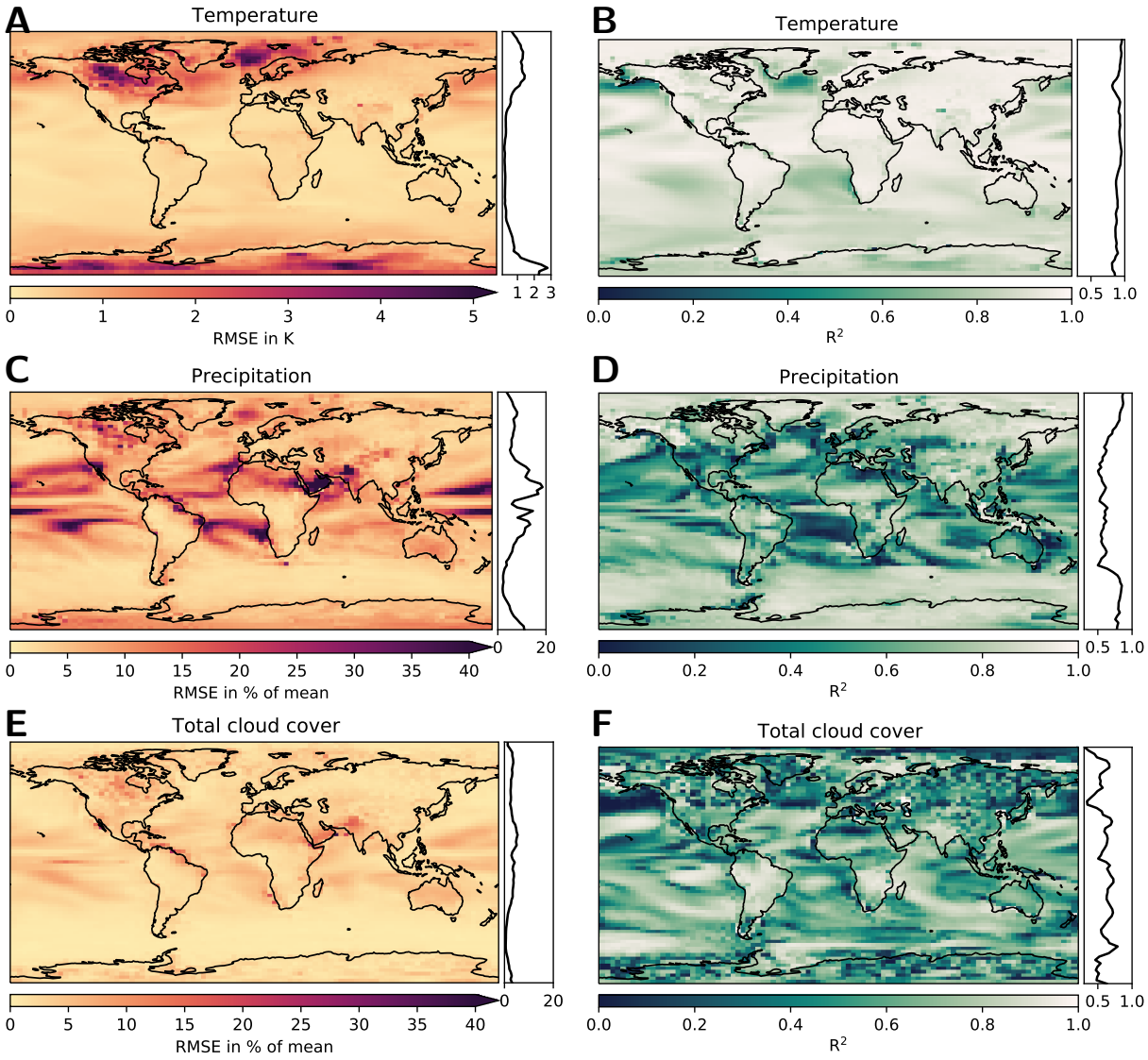

Figure 5: Left panel (A, C, E): Root mean square errors (RMSE) as estimators of the goodness of fit (lower is better) calculated using the test data. Right panel (B, D, F): R^2 values as estimator for the goodness of the model (higher is better) using the training data. Shown are the R^2 and RMSEs for (A,B) mean annual temperature, (C, D) mean annual precipitation and (E, F) mean annual total cloud cover.

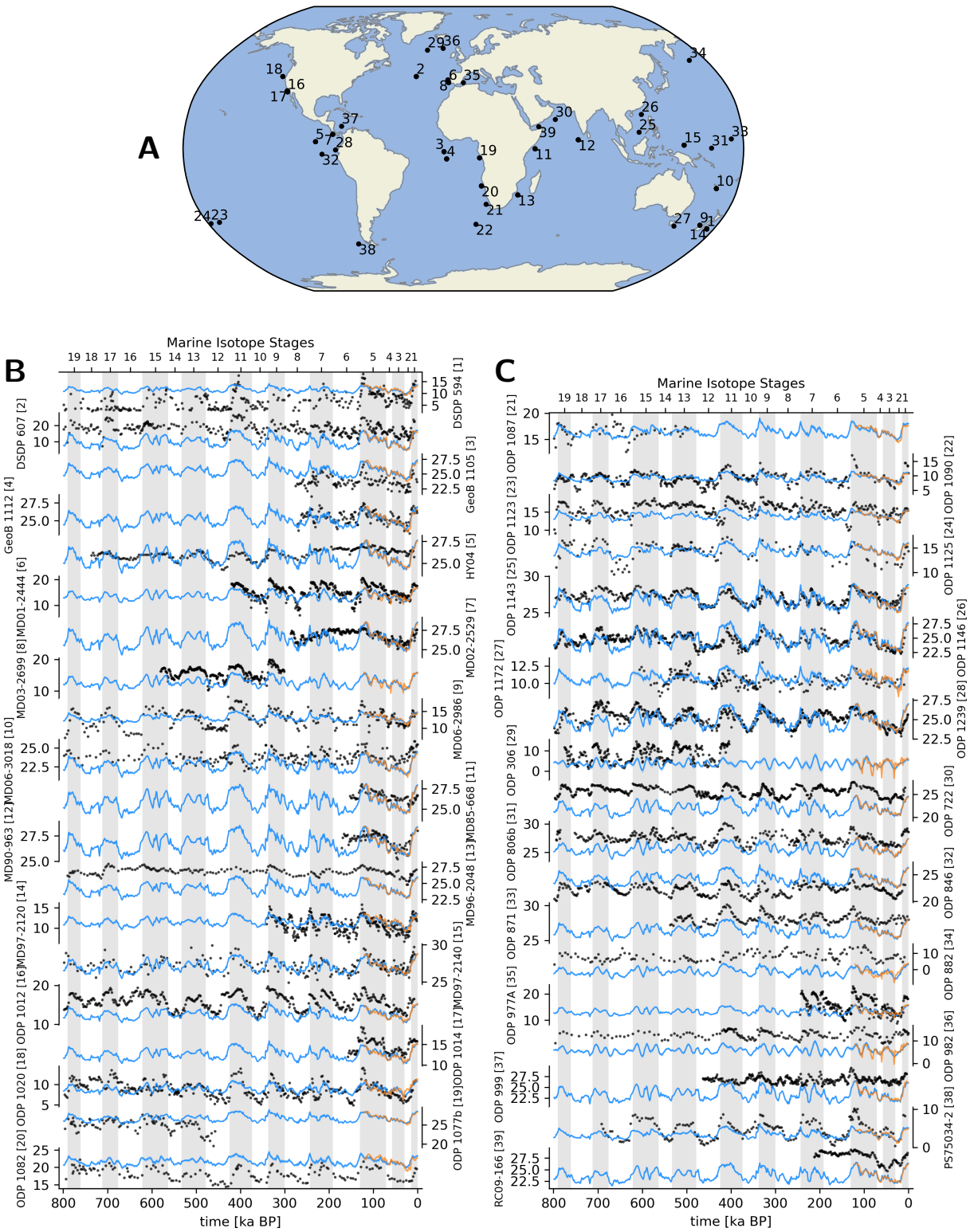

Figure 6: (A) Map of the 39 Middle and Late Pleistocene marine sea surface temperature proxies used in this study and their respective time series (B, C). Black dots indicate proxy sea surface temperature while blue lines indicate mean annual temperature as reconstructed. Proxy–derived and model temperature are on the same scale, in °C). Orange lines are original time series from HadCM3. Grey bars indicate glacial stages.

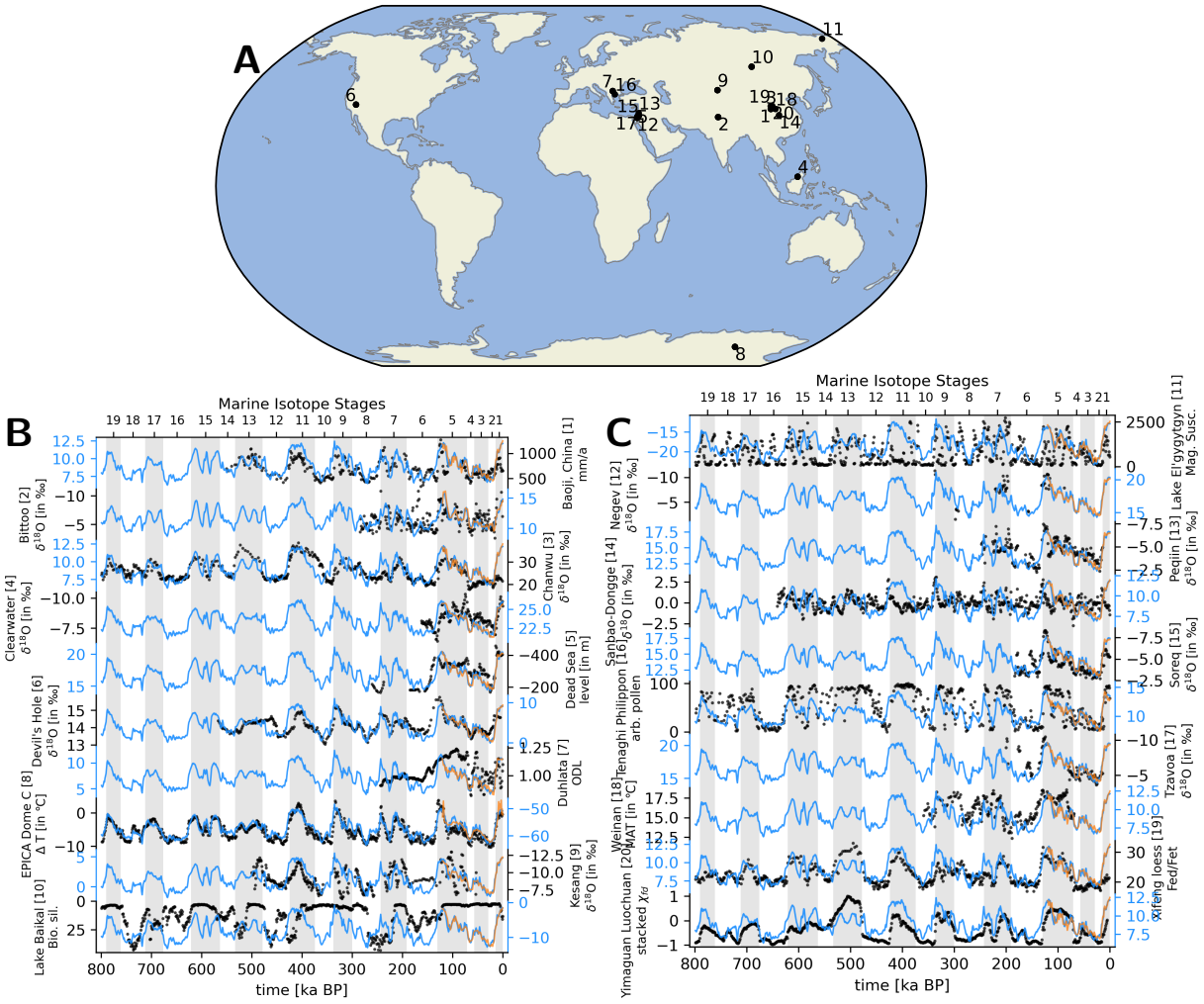

Figure 7: (A) Map of the 20 Middle and Late Pleistocene terrestrial climate proxies used in this study and their respective time series (B). Black dots indicate proxy variables (in different units) while blue lines indicate mean annual temperature as reconstructed (in °C). Orange lines are original time series from HadCM3. Grey bars indicate glacial stages.

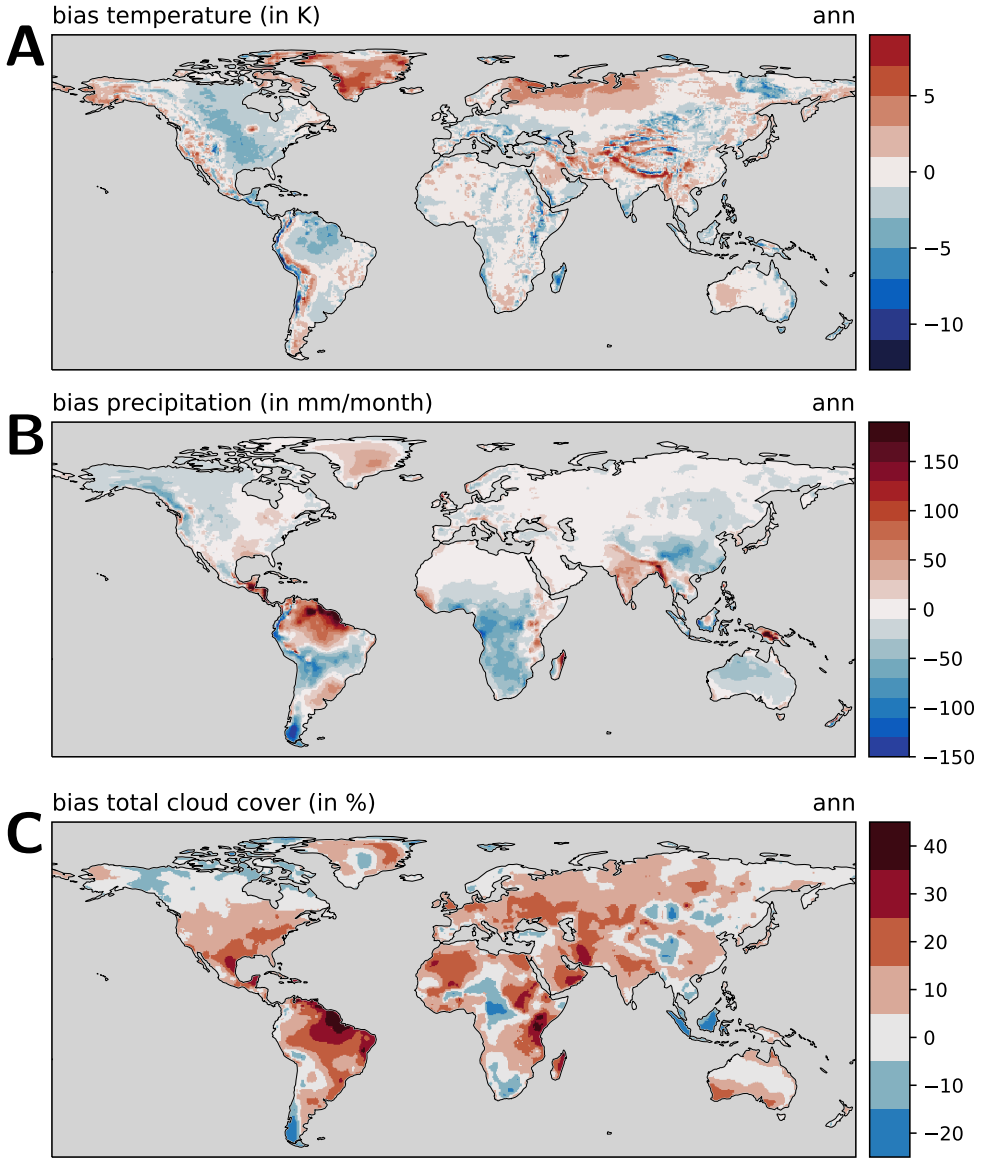

Figure 8: Model bias for (A) annual mean temperature, (B) annual mean precipitation, and (C) annual mean total cloud cover, as differences between the linear model reconstruction for 0 ka and present-day (average of 1961–1990) CRU TS data.

Tables

Table 1: Available reconstructions of environmental variables. All variables have the dimensions 720x360x800 (longitude, latitude, time). Temperature seasonality (BIO4) and precipitation seasonality (BIO15) are given by the standard deviation of monthly temperatures and by the coefficient of variation of monthly precipitation, respectively. Temperature annual range (BIO7) is given by the difference between maximum annual temperature (BIO5) and minimum annual temperature (BIO6). Unit abbreviations: mm (millimetres), m (metres), gC (grams carbon).

Variable	**Unit**
Dimensional variables	
longitude (720)	degrees east
latitude (360)	degrees north
time (800)	years before present
Climatic variables	
monthly temperature (Jan-Dec)	K
monthly precipitation (Jan-Dec)	mm $year^{-1}$
monthly cloudiness (Jan-Dec)	%
minimum annual temperature	K
Vegetation variables	
net primary productivity	gC m^{-2} $year^{-1}$
biome	categorical
Bioclimatic variables	
BIO1: annual mean temperature	°C
BIO4: temperature seasonality	°C
BIO5: minimum annual temperature	°C
BIO6: maximum annual temperature	°C
BIO7: temperature annual range	°C
BIO8: mean temperature of the wettest quarter	°C
BIO9: mean temperature of driest quarter	°C
BIO10: mean temperature of warmest quarter	°C
BIO11: mean temperature of coldest quarter	°C
BIO12: annual precipitation	mm $year^{-1}$
BIO13: precipitation of wettest month	mm $year^{-1}$
BIO14: precipitation of driest month	mm $year^{-1}$
BIO15: precipitation seasonality	–
BIO16: precipitation of wettest quarter	mm $year^{-1}$
BIO17: precipitation of driest quarter	mm $year^{-1}$
BIO18: precipitation of warmest quarter	mm $year^{-1}$
BIO19: precipitation of coldest quarter	mm $year^{-1}$
Land/land ice/ocean mask	
mask	categorical

Table 2: List of data sets that can be found in the *Open Science Framework* repository [30] under the project's `data` directory.

Reconstructed and bias-corrected 800 ka outputs
`temp_800ka_jan.nc`
`...`
`temp_800ka_dec.nc`
`temp_800ka_min.nc`
`temp_800ka_ann.nc`
`prec_800ka_jan.nc`
`...`
`prec_800ka_dec.nc`
`prec_800ka_ann.nc`
`tcc_800ka_jan.nc`
`...`
`tcc_800ka_dec.nc`
`tcc_800ka_ann.nc`
`bio01_800ka.nc`
`bio04_800ka.nc`
`bio05_800ka.nc`
`...`
`bio19_800ka.nc`
BIOME4 800 ka outputs
`biome4output_800ka.nc`
Land/land Ice/ocean masks
`icesheets_000-800_cru.nc`

Table 3: Marine proxy records that have been used in the validation of the climate reconstruction, their coordinates, correlation coefficients, types, and respective references.

core/name	lon	lat	corr coeff	type	reference(s)
DSDP 594	175.0	-45.5	0.58	SST	[46, 47]
DSDP 607	-33.0	41.0	0.48	SST	[46, 48]
GeoB 1105	-12.4	-1.7	0.64	SST	[49]
GeoB 1112	-10.7	-5.8	0.56	SST	[49]
HY04	-95.0	4.0	0.30	SST	[46, 50]
MD01-2444	-10.1	37.6	0.69	SST	[51]
MD02-2529	-84.1	8.2	0.34	SST	[52]
MD03-2699	-10.7	39.0	0.66	SST	[53]
MD06-2986	167.9	-43.4	0.71	SST	[46, 54]
MD06-3018	166.2	-22.6	0.42	SST	[55]
MD85-668	46.0	0.0	0.47	SST	[56]
MD90-963	73.9	5.1	0.53	SST	[57]
MD96-2048	36.0	-26.2	0.53	SST	[58]
MD97-2120	174.9	-45.5	0.74	SST	[59]
MD97-2140	141.5	2.0	0.56	SST	[46, 60]
ODP 1012	-118.4	32.3	0.60	SST	[46, 61]
ODP 1014	-118.9	32.8	0.81	SST	[62]
ODP 1020	-126.4	41.0	0.53	SST	[46, 63]
ODP 1077b	10.4	-5.2	0.18	SST	[64]
ODP 1082	11.8	-21.1	0.45	SST	[65]
ODP 1087	15.3	-31.5	0.13	SST	[66]
ODP 1090	8.9	-42.9	0.70	SST	[46, 67]
ODP 1123	-171.5	-41.8	0.37	SST	[46, 68]
ODP 1125	-178.2	-42.6	0.55	SST	[69]
ODP 1143	113.3	9.4	0.62	SST	[46, 70]
ODP 1146	116.3	19.5	0.53	SST	[46, 71]
ODP 1172	149.9	-44.0	0.32	SST	[72]
ODP 1239	-82.1	-0.7	0.52	SST	[73]
ODP 306	-27.9	56.4	0.35	SST	[74]
ODP 722	59.8	16.6	0.45	SST	[46, 71]
ODP 806b	159.4	0.3	0.57	SST	[46, 75]
ODP 846	-90.8	-3.1	0.53	SST	[46, 76]
ODP 871	172.3	5.6	0.65	SST	[77]
ODP 882	167.6	50.4	0.10	SST	[78]
ODP 977A	0.0	37.5	0.68	SST	[51]
ODP 982	-15.9	57.5	0.37	SST	[46, 79]
ODP 999	-78.7	12.8	0.14	SST	[80]
PS75034-2	-80.1	-54.4	0.79	SST	[46, 81]
RC09-166	48.8	12.5	0.36	SST	[82]

Table 4: Terrestrial proxy records that have been used in the validation of the climate reconstruction, their coordinates, correlation coefficients, types, and respective references.

core/name	lon	lat	corr coeff	type	reference(s)
Baoji, China	107.1	34.4	0.61	rainfall	[83]
Bittoo	77.8	30.8	-0.40	δ^{18}O	[84]
Chanwu	107.7	35.2	0.60	δ^{18}O	[85]
Clearwater	114.9	4.1	-0.49	δ^{18}O	[86]
Dead Sea	35.0	30.5	-0.63	lake level	[87]
Devil's Hole	-116.3	36.4	0.67	δ^{18}O	[88]
Duhlata	23.2	42.5	0.40	ODL	[89]
EPICA Dome C	123.4	-75.0	0.88	temperature	[90]
Kesang	81.8	42.9	-0.36	δ^{18}O	[91]
Lake Baikal	108.4	53.7	-0.15	Bio. sil.	[92]
Lake El'gygytgyn	172.0	67.5	0.18	mag. susc.	[93]
Negev	34.8	30.6	-0.68	δ^{18}O	[94]
Peqiin	36.0	32.6	-0.60	δ^{18}O	[95]
Sanbao-Dongge	110.4	31.7	0.12	δ^{18}O	[96]
Soreq	36.0	31.4	-0.71	δ^{18}O	[95]
Tenaghi Philippon	24.2	41.0	0.67	arb. pollen	[46, 97]
Tzavoa	35.2	31.2	-0.59	δ^{18}O	[98]
Weinan	109.6	34.4	0.49	temperature	[99]
Xifeng loess	107.6	35.7	0.60	Fed/Fet	[85]
Yimaguan Luochuan	108.5	35.8	0.60	mag. susc.	[46, 100]

References

[1] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Avery, M. Tignor, and H. Miller, editors. *IPCC: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2007.

[2] A. Ganopolski and R. Calov. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. *Clim. Past*, 7(4):1415–1425, December 2011.

[3] Axel Timmermann, Tobias Friedrich, Oliver Elison Timm, Megumi O. Chikamoto, Ayako Abe-Ouchi, and Andrey Ganopolski. Modeling Obliquity and CO2 Effects on Southern Hemisphere Climate during the Past 408 ka. *J. Climate*, 27(5):1863–1875, November 2013.

[4] M. Claussen, L. Mysak, A. Weaver, M. Crucifix, T. Fichefet, M.-F. Loutre, S. Weber, J. Alcamo, V. Alexeev, A. Berger, R. Calov, A. Ganopolski, H. Goosse, G. Lohmann, F. Lunkeit, I. Mokhov, V. Petoukhov, P. Stone, and Z. Wang. Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate Dynamics*, 18(7):579–586, March 2002.

[5] Robert J. Hijmans, Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15):1965–1978, 2005.

[6] Matheus Souza Lima-Ribeiro, Sara Varela, Javier González-Hernández, Guilherme de Oliveira, José Alexandre F. Diniz-Filho, and Levi Carina Terribile. EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics*, 10(0), August 2015.

[7] Damien A. Fordham, Frédérik Saltré, Sean Haythorne, Tom M. L. Wigley, Bette L. Otto-Bliesner, Ka Ching Chan, and Barry W. Brook. PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, 40(11):1348–1358, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecog.03031.

[8] P. J. Valdes, E. Armstrong, M. P. S. Badger, C. D. Bradshaw, F. Bragg, M. Crucifix, T. Davies-Barnard, J. J. Day, A. Farnsworth, C. Gordon, P. O. Hopcroft, A. T. Kennedy, N. S. Lord, D. J. Lunt, A. Marzocchi, L. M. Parry, V. Pope, W. H. G. Roberts, E. J. Stone, G. J. L. Tourte, and J. H. T. Williams. The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geosci. Model Dev.*, 10(10):3715–3743, October 2017.

[9] Edward Armstrong, Peter O. Hopcroft, and Paul J. Valdes. A simulated Northern Hemisphere terrestrial climate dataset for the past 60,000 years. *Sci Data*, 6(1):1–16, November 2019. Number: 1 Publisher: Nature Publishing Group.

[10] Robert M. Beyer, Mario Krapp, and Andrea Manica. High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Scientific Data*, 7(1):236, July 2020. Number: 1 Publisher: Nature Publishing Group.

[11] Robert Beyer, Mario Krapp, and Andrea Manica. An empirical evaluation of bias correction methods for palaeoclimate simulations. *Climate of the Past*, 16(4):1493–1508, August 2020. Publisher: Copernicus GmbH.

[12] Ian Harris, Timothy J. Osborn, Phil Jones, and David Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data*, 7(1):1–18, April 2020. Number: 1 Publisher: Nature Publishing Group.

[13] Michael S O'Donnell and Drew A Ignizio. Bioclimatic predictors for supporting ecological applications in the conterminous United States. *US Geological Survey Data Series*, 691(10), 2012.

[14] J. O. Kaplan, N. H. Bigelow, I. C. Prentice, S. P. Harrison, P. J. Bartlein, T. R. Christensen, W. Cramer, N. V. Matveyeva, A. D. McGuire, D. F. Murray, V. Y. Razzhivin, B. Smith, D. A. Walker, P. M. Anderson, A. A. Andreev, L. B. Brubaker, M. E. Edwards, and A. V. Lozhkin. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res.*, 108(D19):8171, October 2003.

[15] Joy S. Singarayer and Paul J. Valdes. High-latitude climate sensitivity to ice-sheet forcing over the last 120kyr. *Quaternary Science Reviews*, 29(1):43–55, 2010. 00092.

[16] T. Davies-Barnard, A. Ridgwell, J. Singarayer, and P. Valdes. Quantifying the influence of the terrestrial biosphere on glacial–interglacial climate dynamics. *Clim. Past*, 13(10):1381–1401, October 2017.

[17] Bernhard Bereiter, Sarah Eggleston, Jochen Schmitt, Christoph Nehrbass-Ahles, Thomas F. Stocker, Hubertus Fischer, Sepp Kipfstuhl, and Jerome Chappellaz. Revision of the EPICA Dome C CO2 record from 800 to 600 kyr before present. *Geophys. Res. Lett.*, 42(2):2014GL061957, January 2015.

[18] A. Berger and M. F. Loutre. Insolation values for the climate of the last 10 million years. *Quaternary Science Reviews*, 10(4):297–317, January 1991.

[19] R. M. Spratt and L. E. Lisiecki. A Late Pleistocene sea level stack. *Clim. Past*, 12(4):1079–1092, April 2016.

[20] W. R. Peltier, D. F. Argus, and R. Drummond. Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_c (VM5a) model. *Journal of Geophysical Research: Solid Earth*, 120(1):450–487, November 2014.

[21] W. R. Peltier. ICE-6G, May 2019. http://www.atmosp.physics.utoronto.ca/~peltier/data.php.

[22] C Amante and BW Eakins. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. *National Geophysical Data Center, NOAA*, NOAA Technical Memorandum NESDIS NGDC-24, 2009.

[23] NOAA National Geophysical Data Center. ETOPO1 Global Relief, 2009. https://ngdc.noaa.gov/mgg/global/global.html.

[24] Bernhard Lehner and Petra Döll. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296(1):1–22, August 2004.

[25] WWF. Global Lakes and Wetlands Database, May 2019. https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database.

[26] P. A. Araya-Melo, M. Crucifix, and N. Bounceur. Global sensitivity analysis of the Indian monsoon during the Pleistocene. *Clim. Past*, 11(1):45–61, January 2015.

[27] Natalie S. Lord, Michel Crucifix, Dan J. Lunt, Mike C. Thorne, Nabila Bounceur, Harry Dowsett, Charlotte L. O'Brien, and Andy Ridgwell. Emulation of long-term changes in global climate: application to the late Pliocene and future. *Climate of the Past*, 13(11):1539–1571, November 2017.

[28] Peter Huybers and Ian Eisenman. Integrated Summer Insolation Forcing Data, 2006. https://www.ncdc.noaa.gov/paleo-search/study/5792.

[29] Douglas V. Hoyt. Percent of Possible Sunshine and the Total Cloud Cover. *Monthly Weather Review*, 105(5):648–652, May 1977. Publisher: American Meteorological Society Section: Monthly Weather Review.

[30] Mario Krapp. Terrestrial climate of the last 800,000 years. *Open Science Framework*, April 2020. https://osf.io/8n43x/.

[31] Ulrike Herzschuh, H. John B. Birks, Thomas Laepple, Andrei Andreev, Martin Melles, and Julie Brigham-Grette. Glacial legacies on interglacial vegetation at the Pliocene-Pleistocene transition in NE Asia. *Nature Communications*, 7:11967, June 2016.

[32] Uwe Schulzweida. CDO User Guide. October 2019. Publisher: Zenodo.

[33] Johannes Köster and Sven Rahmann. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, October 2012.

[34] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[35] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[36] Met Office. Cartopy: a cartographic python library with a Matplotlib interface, 2010.

[37] Ilya Flyamer, Samuël Weber/GwendalD, Zhuyi Xue, Colin, Andy Li, Christophe Van Neste, Josh L. Espinoza, Nader Morshed, Victor Vazquez, Ryan Neff, mski_iksm, and scaine1. Phlya/adjustText: 0.8 beta, June 2020.

[38] Leonard Richardson. beautifulsoup4: Screen-scraping library.

[39] Jeff Whitaker, Constantine Khrulev, David Huard, Christoph Paulik, Stephan Hoyer, Filipe, Lars Pastewka, Alexander Mohr, Christian Marquardt, Bas Couwenberg, Mike Taves, Matthias Cuntz, Sander Roet, Jeffrey Whitaker, Matthew Brett, Max Bohnet, Miloššš Korenčiak, Rob Hetland,

Andrew Barna, Joe Hamman, Jonathan J. Helmus, Kristjan Onu, barronh, Chris Barker, Erik Cederstrand, Jaka Smrekar, James Hiebert, Ryan May, Thomas Kluyver, and bekozi. Unidata/netcdf4-python: version 1.5.5 release, December 2020.

[40] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. Number: 7825 Publisher: Nature Publishing Group.

[41] Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, gfyoung, Sinhrks, Adam Klein, Matthew Roeschke, Simon Hawkins, Jeff Tratner, Chang She, William Ayd, Terji Petersen, Marc Garcia, Jeremy Schendel, Andy Hayden, MomIsBestFriend, Vytautas Jancauskas, Pietro Battiston, Skipper Seabold, chris b1, h vetinari, Stephan Hoyer, Wouter Overmeire, alimcmaster1, Kaiqi Dong, Christopher Whelan, and Mortada Mehyar. pandas-dev/pandas: Pandas 1.0.3, March 2020.

[42] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[43] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020. Number: 3 Publisher: Nature Publishing Group.

[44] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in Python. *PeerJ*, 2:e453, June 2014. Publisher: PeerJ Inc.

[45] Casper da Costa-Luis, Stephen Karl Larroque, Kyle Altendorf, Hadrien Mary, richardsheridan, Mikhail Korobov, Noam Yorav-Raphael, Ivan Ivanov, Marcel Bargull, Nishant Rodrigues, Guangshuo CHEN, Charles Newey, James, Martin Zugnoni, Matthew D. Pagel, mjstevens777, Mikhail Dektyarev, Alex Rothberg, Alexander, Daniel Panteleit, Fabian Dill, FichteFoll, HeoHeo, Hugo van Kemenade, Jack McCracken, Max Nordlund, Nikolay Nechaev, Orivej Desh, RedBug312, and Socialery. tqdm: A fast, Extensible Progress Bar for Python and CLI, February 2021.

[46] Past Interglacials Working Group of PAGES. Interglacials of the last 800,000 years. *Rev. Geophys.*, 54(1):2015RG000482, March 2016.

[47] Grace Schaefer, J. Stuart Rodger, Bruce W. Hayward, James P. Kennett, Ashwaq T. Sabaa, and George H. Scott. Planktic foraminiferal and sea surface temperature record during the last 1 Myr across the Subtropical Front, Southwest Pacific. *Marine Micropaleontology*, 54(3-4):191–212, March 2005.

[48] W. F. Ruddiman, M. E. Raymo, D. G. Martinson, B. M. Clement, and J. Backman. Pleistocene evolution: Northern hemisphere ice sheets and North Atlantic Ocean. *Paleoceanography and Paleoclimatology*, 4(4):353–412, August 1989.

[49] D. Nürnberg, A. Müller, and R. R. Schneider. Paleo-sea surface temperature calculations in the equatorial east Atlantic from Mg/Ca ratios in planktic foraminifera: A comparison to sea surface temperature estimates from U37K', oxygen isotopes, and foraminiferal transfer function. *Paleoceanography and Paleoclimatology*, 15(1):124–134, February 2000.

[50] Keiji Horikawa, Masafumi Murayama, Masao Minagawa, Yoshihisa Kato, and Takuya Sagawa. Latitudinal and downcore (0–750 ka) changes in n-alkane chain lengths in the eastern equatorial Pacific. *Quaternary Research*, 73(3):573–582, May 2010.

[51] Belen Martrat, Joan O. Grimalt, Nicholas J. Shackleton, Lucia de Abreu, Manuel A. Hutterli, and Thomas F. Stocker. Four Climate Cycles of Recurring Deep and Surface Water Destabilizations on the Iberian Margin. *Science*, 317(5837):502–507, July 2007.

[52] Daniel Rincón-Martínez and Guillaume Leduc. Sea surface temperature calculated from alkenones for the last 285 ka with high-reolution Holocene of sediment core MD02-2529, Panama Basin. *PANGAEA*, March 2012. type: dataset.

[53] T. Rodrigues, A. H. L. Voelker, J. O. Grimalt, F. Abrantes, and F. Naughton. Iberian Margin sea surface temperature during MIS 15 to 9 (580–300 ka): Glacial suborbital variability versus interglacial stability. *Paleoceanography*, 26(1):PA1204, March 2011.

[54] Bruce W. Hayward, Ashwaq T. Sabaa, Andrew Kolodziej, Martin P. Crundwell, Silke Steph, George H. Scott, Helen L. Neil, Helen C. Bostock, Lionel Carter, and Hugh R. Grenfell. Planktic foraminifera-based sea-surface temperature record in the Tasman Sea and history of the Subtropical Front around New Zealand, over the last one million years. *Marine Micropaleontology*, 82-83:13–27, January 2012.

[55] T. Russon, M. Elliot, A. Sadekov, G. Cabioch, T. Corrège, and P. De Deckker. Inter-hemispheric asymmetry in the early Pleistocene Pacific warm pool. *Geophysical Research Letters*, 37(11), June 2010.

[56] Edouard Bard, Frauke Rostek, and Corinne Sonzogni. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature*, 385:707–710, February 1997.

[57] Frauke Rostek, Götz Ruhlandt, Franck C. Bassinot, Peter J. Muller, Laurent D. Labeyrie, Yves Lancelot, and Edouard Bard. Reconstructing sea surface temperature and salinity using $\delta^{18}O$ and alkenone records. *Nature*, 364(6435):319, July 1993.

[58] T. Caley, J.-H. Kim, B. Malaizé, J. Giraudeau, T. Laepple, N. Caillon, K. Charlier, H. Rebaubier, L. Rossignol, I. S. Castañeda, S. Schouten, and J. S. Sinninghe Damsté. High-latitude obliquity as a dominant forcing in the Agulhas current system. *Clim. Past*, 7(4):1285–1296, November 2011.

[59] Katharina Pahnke, Rainer Zahn, Henry Elderfield, and Michael Schulz. 340,000-Year Centennial-Scale Marine Record of Southern Hemisphere Climatic Oscillation. *Science*, 301(5635):948–952, August 2003.

[60] Thibault de Garidel-Thoron, Yair Rosenthal, Luc Beaufort, Edouard Bard, Corinne Sonzogni, and Alan C. Mix. A multiproxy assessment of the western equatorial Pacific hydrography during the last 30 kyr. *Paleoceanography*, 22(3), 2005.

[61] Zhonghui Liu, Mark A. Altabet, and Timothy D. Herbert. Glacial-interglacial modulation of eastern tropical North Pacific denitrification over the last 1.8-Myr. *Geophysical Research Letters*, 32(23), December 2005.

[62] Masanobu Yamamoto, Masumi Yamamuro, and Yuichiro Tanaka. The California current system during the last 136,000 years: response of the North Pacific High to precessional forcing. *Quaternary Science Reviews*, 26(3-4):405–414, February 2007.

[63] T. D. Herbert. Collapse of the California Current During Glacial Maxima Linked to Climate Change on Land. *Science*, 293(5527):71–76, July 2001.

[64] Enno Schefuß, Jaap S. Sinninghe Damsté, and J. H. Fred Jansen. Forcing of tropical Atlantic sea surface temperatures during the mid-Pleistocene transition. *Paleoceanography*, 19(4), December 2004.

[65] Johan Etourneau, Philippe Martinez, Thomas Blanz, and Ralph Schneider. Pliocene–Pleistocene variability of upwelling activity, productivity, and nutrient cycling in the Benguela region. *Geology*, 37(10):871–874, October 2009.

[66] Erin L. McClymont, Antoni Rosell-Melé, Jacques Giraudeau, Catherine Pierre, and Jerry M. Lloyd. Alkenone and coccolith records of the mid-Pleistocene in the south-east Atlantic: Implications for the U37K' index and South African climate. *Quaternary Science Reviews*, 24(14):1559–1572, August 2005.

[67] Alfredo Martínez-Garcia, Antoni Rosell-Melé, Walter Geibert, Rainer Gersonde, Pere Masqué, Vania Gaspari, and Carlo Barbante. Links between iron supply, marine productivity, sea surface temperature, and CO2 over the last 1.1 Ma. *Paleoceanography*, 24(1), February 2009.

[68] Martin Crundwell, George Scott, Tim Naish, and Lionel Carter. Glacial–interglacial ocean climate variability from planktonic foraminifera during the Mid-Pleistocene transition in the temperate Southwest Pacific, ODP Site 1123. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 260(1):202–229, April 2008.

[69] Bruce W. Hayward, George H. Scott, Martin P. Crundwell, James P. Kennett, Lionel Carter, Helen L. Neil, Ashwaq T. Sabaa, Kate Wilson, J. Stuart Rodger, Grace Schaefer, Hugh R. Grenfell, and Qianyu Li. The effect of submerged plateaux on Pleistocene gyral circulation and sea-surface

temperatures in the Southwest Pacific. *Global and Planetary Change*, 63(4):309–316, October 2008.

[70] Li Li, Qianyu Li, Jun Tian, Pinxian Wang, Hui Wang, and Zhonghui Liu. A 4-Ma record of thermal evolution in the tropical western Pacific and its implications on climate change. *Earth and Planetary Science Letters*, 309(1):10–20, September 2011.

[71] T. D. Herbert, L. C. Peterson, K. T. Lawrence, and Z. Liu. Tropical Ocean Temperatures Over the Past 3.5 Million Years. *Science*, 328(5985):1530–1534, June 2010.

[72] Dirk Nürnberg and Jeroen Groeneveld. Pleistocene variability of the Subtropical Convergence at East Tasman Plateau: Evidence from planktonic foraminiferal Mg/Ca (ODP Site 1172A). *Geochemistry, Geophysics, Geosystems*, 7(4), April 2006.

[73] K. A. Dyez, A. C. Ravelo, and A. C. Mix. Evaluating drivers of Pleistocene eastern tropical Pacific sea surface temperature. *Paleoceanography*, 31(8):2015PA002873, August 2016.

[74] M. Alonso-Garcia, F.J. Sierro, M. Kucera, J.A. Flores, I. Cacho, and N. Andersen. Ocean circulation, ice sheet growth and interhemispheric coupling of millennial climate variability during the mid-Pleistocene (ca 800–400ka). *Quaternary Science Reviews*, 30(23-24):3234–3247, November 2011.

[75] Martín Medina-Elizalde and David W Lea. The Mid-Pleistocene Transition in the Tropical Pacific. *Science (New York, N.Y.)*, 310:1009–12, December 2005.

[76] Zhonghui Liu. *Pleistocene climate evolution in the eastern Pacific and implications for the orbital theory of climate change*. Ph.D., Brown University, United States – Rhode Island, 2004.

[77] K. A. Dyez and A. C. Ravelo. Late Pleistocene tropical Pacific temperature sensitivity to radiative greenhouse gas forcing. *Geological Society of America*, 41(1):23–26, January 2013.

[78] Alfredo Martínez-Garcia, Antoni Rosell-Melé, Erin L. McClymont, Rainer Gersonde, and Gerald H. Haug. Subpolar Link to the Emergence of the Modern Equatorial Pacific Cold Tongue. *Science*, 328(5985):1550–1553, June 2010.

[79] Kira T. Lawrence, Timothy D. Herbert, Catherine M. Brown, Maureen E. Raymo, and Alan M. Haywood. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography*, 24(2):PA2218, June 2009.

[80] Matthew W. Schmidt, Maryline J. Vautravers, and Howard J. Spero. Western Caribbean sea surface temperatures during the late Quaternary. *Geochemistry, Geophysics, Geosystems*, 7(2), February 2006.

[81] Sze Ling Ho, Gesine Mollenhauer, Frank Lamy, Alfredo Martínez-Garcia, Mahyar Mohtadi, Rainer Gersonde, Dierk Hebbeln, Samuel Nunez-Ricardo, Antoni Rosell-Melé, and Ralf Tiedemann. Sea surface temperature variability in the Pacific sector of the Southern Ocean over the past 700 kyr. *Paleoceanography*, 27(4), October 2012.

[82] Jessica E. Tierney, Peter B. deMenocal, and Paul D. Zander. A climatic context for the out-of-Africa migration. *The Geological Society of America*, 45(11):1023–1026, 2017.

[83] J. Warren Beck, Weijian Zhou, Cheng Li, Zhenkun Wu, Lara White, Feng Xian, Xianghui Kong, and Zhisheng An. A 550,000-year record of East Asian monsoon rainfall from 10Be in loess. *Science*, 360(6391):877–881, May 2018.

[84] Gayatri Kathayat, Hai Cheng, Ashish Sinha, Christoph Spötl, R. Lawrence Edwards, Haiwei Zhang, Xianglei Li, Liang Yi, Youfeng Ning, Yanjun Cai, Weiguo Lui Lui, and Sebastian F. M. Breitenbach. Indian monsoon variability on millennial-orbital timescales. *Scientific Reports*, 6:24374, April 2016.

[85] Z. T. Guo, A. Berger, Q. Z. Yin, and L. Qin. Strong asymmetry of hemispheric climates during MIS-13 inferred from correlating China loess and Antarctica ice records. *Clim. Past*, 5(1):21–31, February 2009.

[86] Stacy A. Carolin, Kim M. Cobb, Jean Lynch-Stieglitz, Jessica W. Moerman, Judson W. Partin, Syria Lejau, Jenny Malang, Brian Clark, Andrew A. Tuen, and Jess F. Adkins. Northern Borneo stalagmite records reveal West Pacific hydroclimate across MIS 5 and 6. *Earth and Planetary Science Letters*, 439:182–193, April 2016.

[87] Nicolas Waldmann, Adi Torfstein, and Mordechai Stein. Northward intrusions of low- and mid-latitude storms across the Saharo-Arabian belt during past interglacials. *Geology*, 38(6):567–570, June 2010.

[88] Jurate M Landwehr, Warren D Sharp, Tyler B Coplen, Kenneth R Ludwig, and Isaac J Winograd. The Chronology for the δ^{18}O Record from Devils Hole, Nevada, Extended Into the Mid-Holocene. Technical report, US Geological Survey, 2011.

[89] D. A. Stoykova, Y. Y. Shopov, D. Garbeva, L. T. Tsankov, and C. J. Yonge. Origin of the climatic cycles from orbital to sub-annual scales. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(2-4):293–302, February 2008.

[90] J. Jouzel, V. Masson-Delmotte, O. Cattani, G. Dreyfus, S. Falourd, G. Hoffmann, B. Minster, J. Nouet, J. M. Barnola, J. Chappellaz, H. Fischer, J. C. Gallet, S. Johnsen, M. Leuenberger, L. Loulergue, D. Luethi, H. Oerter, F. Parrenin, G. Raisbeck, D. Raynaud, A. Schilt, J. Schwander, E. Selmo, R. Souchez, R. Spahni, B. Stauffer, J. P. Steffensen, B. Stenni, T. F. Stocker, J. L. Tison, M. Werner, and E. W. Wolff. Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years. *Science*, 317(5839):793–796, August 2007.

[91] H. Cheng, P. Z. Zhang, C. Spötl, R. L. Edwards, Y. J. Cai, D. Z. Zhang, W. C. Sang, M. Tan, and Z. S. An. The climatic cyclicity in semiarid-arid central Asia over the past 500,000 years. *Geophysical Research Letters*, 39(1), January 2012.

[92] Alexander A. Prokopenko, Linda A. Hinnov, Douglas F. Williams, and Mikhail I. Kuzmin. Orbital forcing of continental climate during the Pleistocene: a complete astronomically tuned climatic record from Lake Baikal, SE Siberia. *Quaternary Science Reviews*, 25(23):3431–3457, December 2006.

[93] Martin Melles, Julie Brigham-Grette, Pavel S. Minyuk, Norbert R. Nowaczyk, Volker Wennrich, Robert M. DeConto, Patricia M. Anderson, Andrei A. Andreev, Anthony Coletti, Timothy L. Cook, Eeva Haltia-Hovi, Maaret Kukkonen, Anatoli V. Lozhkin, Peter Rosén, Pavel Tarasov, Hendrik Vogel, and Bernd Wagner. 2.8 Million Years of Arctic Climate Change from Lake El'gygytgyn, NE Russia. *Science*, 337(6092):315–320, July 2012.

[94] Anton Vaks, Miryam Bar-Matthews, Alan Matthews, Avner Ayalon, and Amos Frumkin. Middle-Late Quaternary paleoclimate of northern margins of the Saharan-Arabian Desert: reconstruction from speleothems of Negev Desert, Israel. *Quaternary Science Reviews*, 29(19):2647–2662, September 2010.

[95] Miryam Bar-Matthews, Avner Ayalon, Mabs Gilmour, Alan Matthews, and Chris J. Hawkesworth. Sea–land oxygen isotopic relationships from planktonic foraminifera and speleothems in the Eastern Mediterranean region and their implication for paleorainfall during interglacial intervals. *Geochimica et Cosmochimica Acta*, 67(17):3181–3199, September 2003.

[96] Hai Cheng, R. Lawrence Edwards, Ashish Sinha, Christoph Spötl, Liang Yi, Shitao Chen, Megan Kelly, Gayatri Kathayat, Xianfeng Wang, Xianglei Li, Xinggong Kong, Yongjin Wang, Youfeng Ning, and Haiwei Zhang. The Asian monsoon over the past 640,000 years and ice age terminations. *Nature*, 534(7609):640–646, June 2016. 00000.

[97] P. C. Tzedakis, H. Hooghiemstra, and H. Pälike. The last 1.35 million years at Tenaghi Philippon: revised chronostratigraphy and long-term vegetation trends. *Quaternary Science Reviews*, 25(23):3416–3430, December 2006.

[98] A. Vaks, M. Bar-Matthews, A. Ayalon, A. Matthews, A. Frumkin, U. Dayan, L. Halicz, A. Almogi-Labin, and B. Schilman. Paleoclimate and location of the border between Mediterranean climate region and the Saharo–Arabian Desert as revealed by speleothems from the northern Negev Desert, Israel. *Earth and Planetary Science Letters*, 249(3):384–399, September 2006.

[99] Elizabeth K. Thomas, Steven C. Clemens, Youbin Sun, Warren L. Prell, Yongsong Huang, Lee Gao, Shannon Loomis, Guangshan Chen, and Zhengyu Liu. Heterodynes dominate precipitation isotopes in the East Asian monsoon region, reflecting interaction of multiple climate factors. *Earth and Planetary Science Letters*, 455, October 2016.

[100] Qingzhen Hao, Luo Wang, Frank Oldfield, Shuzhen Peng, Li Qin, Yang Song, Bing Xu, Yansong Qiao, Jan Bloemendal, and Zhengtang Guo. Delayed build-up of Arctic ice sheets during 400,000-year minima in insolation variability. *Nature*, 490(7420):393–396, October 2012.