A comprehensive climate history of the last 800 thousand years

Mario Krapp*1, Robert Beyer1, Stephen L. Edmundson1,2, Paul J. Valdes3, and Andrea Manica1

1University of Cambridge
2Utrecht University
3University of Bristol
*contact: mariokrapp@gmail.com

Abstract

A detailed and accurate reconstruction of past climate is essential in understanding the drivers that have shaped species, including our own, and their habitats. However, spatially-detailed climate reconstructions that continuously cover the Quaternary do not yet exist, mainly because no paleoclimate model can reconstruct regional-scale dynamics over geological time scales. Here we develop a new approach, the Global Climate Model Emulator (GCMET), which reconstructs the climate of the last 800 thousand years with unprecedented spatial detail. GCMET captures the temporal dynamics of glacial-interglacial climates as an Earth System Model of Intermediate Complexity would whilst resolving the local dynamics with the accuracy of a Global Climate Model. It provides a new, unique resource to explore the climate of the Quaternary, which we use to investigate the long-term stability of major habitat types. We identify a number of stable pockets of habitat that have remained unchanged over the last 800 thousand years, acting as potential long-term evolutionary refugia. Thus, the highly detailed, comprehensive overview of climatic changes through time delivered by GCMET provides the needed resolution to quantify the role of long term habitat fragmentation in an ecological and anthropological context.

Current patterns of diversification within and between species, such as our own [1], and the structuring of whole ecosystems can only be studied in the context of past climatic changes that have shaped them through time [2]. A detailed understanding of such processes has become an urgent necessity in order to predict responses to global change. However, whilst predictions of climate change and their impacts over the next few tens or hundreds of years are based on comprehensive Global Climate Models (GCMs) that resolve processes at high temporal and spatial resolution, such as those used in the latest IPCC Assessment Report [3], reconstructions back in time are challenging as they have to span a much longer period. GCMs can provide snapshots for a specific time or short transients in the order of a few thousands of years, whilst periods of tens or hundreds of thousands of years can only be covered with Earth System Models of Intermediate Complexity (EMICs) [4, 5], at the cost of low spatial resolution and a simplified representation of the climate system [6]. Neither of those two types of models is intentionally designed for paleo-ecology or species evolution, disciplines that require appropriate temporal scales of up to hundreds of thousands of years and spatial scales down to tens of kilometres.

Here, we fill this gap for a long-term reconstruction of climate that resolves regional-scale dynamics by reconstructing the last 800 thousand years (ka) at an unprecedented spatial resolution of approximately
Unlike previous emulator approaches \cite{7, 8}, we explicitly focus on the local emulation of climate, which allows us to critically evaluate the reconstructed 800 ka of climate history against proxy records. Our approach consists of two steps (Fig. 1): a first reconstruction of the global climate at moderate spatial resolution followed by a more detailed representation of local dynamics using multiple snapshot simulations from the family of HadCM3 climate models \cite{9}. In the first step, we use 72 simulations covering the past 120 ka from the HadCM3 climate model \cite{10, 11}, and build a linear regression model that acts as a GCM emulator (GCMET). GCMET accurately predicts the output of HadCM3 given a set of boundary conditions that is representative of as observed in the Middle and Late Pleistocene (details about this approach are found in the \textit{Methods}). The logic behind our approach is that variations of a climate variable $X$ (e.g. temperature) at any given location can be explained by variations in the external forcings. For the HadCM3 snapshots, the most important forcings are atmospheric CO$_2$ and the orbital parameters, i.e., precession, obliquity, and eccentricity. Other boundary conditions are Northern Hemisphere ice sheets and respective global sea-level changes (see \cite{10} for details). The spatial model resolution after this first step is the same as of HadCM3, i.e., about 3° ($3.75^\circ \times 2.5^\circ$), henceforth referred to as GCMET-LO.

\textbf{Results and discussion}

We tested how well GCMET-LO matches HadCM3 snapshots by splitting them into a training and a test set (see \textit{Methods} for details). Predicted mean annual temperatures (MAT) for the test set were within ca. 2 K (estimated as root mean square error, RMSE) to the output of HadCM3 for most parts of the globe (Fig. 2a) (a more thorough discussion is provided in the \textit{Methods}). Mean annual precipitation (MAP) turned out to be less predictable but this was expected: temperature is a direct response to forcing whereas the precipitation response depends on multiple variables. We improved the MAP predictions substantially by using temperature and specific humidity as independent variables instead of CO$_2$ and orbital parameters (Extended Data Figure 1), and thereby reduced the average RMSE to from 9.3% down to 5.9%. Importantly, the discordance between HadCM3 and GCMET-LO are much smaller than the ensemble variability among different models in the Coupled Model Inter-comparison Project, Phase 5, and thus within the ranges of acceptable model uncertainties \cite{3}.

In the second step, we increase the resolution of our reconstructions to about 1° ($1.25^\circ \times 0.83^\circ$) using high resolution HadAM3H (Hadley Centre Atmospheric Model 3, High resolution) simulations covering the period of the last deglaciation. We computed high-resolution difference maps between equivalent HadAM3H and GCMET-LO snapshots (see \textit{Methods} for details) and then created interpolated maps for any level of CO$_2$ (given the limited number of observations, we focus on CO$_2$ as the main driver of those differences). Those maps were added to GCMET-LO to obtain high resolution reconstructions, which we henceforth refer to simply as GCMET. To illustrate the importance of higher spatial variability, we compared GCMET, GCMET-LO, and LOVECLIM (an EMIC with a horizontal resolution of ca $5.5^\circ \times 5.5^\circ$) to present-day observations (ERA-20C re-analysis 1961–1990 average \cite{12}). LOVECLIM and GCMET-LO fail to capture the observed continental climate patterns whereas GCMET resolves those spatial features well (Fig. 3b).
Proxy comparison

We tested the ability of GCMET to capture changes in climate over the last 800 ka by comparing its predictions to a number of proxies (for a detailed comparison with proxy records, we refer to the Methods), using LOVECLIM [5] as a benchmark. As forcings we used CO$_2$ estimates from EPICA Dome C (EDC3) ice core [13], numerical solutions for the orbital parameters [14], whilst global sea-level changes and the distribution of the major Northern Hemisphere ice sheets were taken from a transient CLIMBER-2 climate simulation [4] and from the ICE-6G data set [15] (see Methods for details). We compared MAT to terrestrial proxies and to sea surface temperatures (SST) estimates based on marine proxies: GCMET is in agreement with a number of marine records (Fig. 3a–c, time series of all used proxies are shown in Extended Data Figures 2 & 3), with a mean RMSE of 1.5 K for all SST proxies and a mean correlation of 0.54, significantly larger (paired t-test: $t_{38}=2.9$, $p=0.006$) than for LOVECLIM ($r=0.49$, Fig. 3b). Despite the diverse nature of the terrestrial proxies (e.g. speleothems, loess, pollen), GCMET performance was as good as for marine proxies ($r=0.52$, Fig. 3b & d).

GCMET can also be used to reconstruct the climate in the deeper past, for example, by going back 2 million years (Ma). For this deeper past only a point-wise CO$_2$ reconstruction is available [16] which can be used to complement the quasi-continuous EDC3 CO$_2$ record covering the last 800 ka. The GCMET reconstructed global average MAT over the last 2 Ma shows a remarkable agreement with a global mean temperature proxy record [17] (correlation $r=0.85$ & RMSE=1.0 K, Fig. 3a). The predictive power over the last 2 Ma may seems surprising given that we do not have any HadCM3 snapshots before 120 ka ago. However, it is important to note that the phase space of the external forcings, CO$_2$ and orbital variations, is well covered, especially over the last 800 ka, by the last glacial cycle (see Extended Data Figure 4) and thus, we are mostly interpolating in a statistical sense.

Past habitat stability

The spatially detailed reconstructions provided by GCMET allow us to explore the effect of climate on habitats and species over time. We investigated ecosystem stability (Fig. 4) over the last 800 ka, focussing on the 14 major terrestrial habitats as defined by the WWF Global 200 [18] (Fig. 4a). The reconstructions which are based on a random forest classifier [19] (see Methods for more details) show marked patterns in stability depending on location, with sparsely vegetated regions such as deserts among the most stable habitats in the world, the others being the core tropical rainforests along the equator. Large parts of Eurasia and North America are rendered unstable by the advancing and retreating Northern Hemisphere ice sheets with ecosystems alternating between vast forests during the warm interglacials and large tundras during the cold glacials (an animated version of the habitat changes throughout the last 800 ka is available as Supplementary Video). However, a few fragmented core boreal forest habitats remain. At the other end of the spectrum, unstable habitats as found in Subsaharan Africa support the idea that large scale habitat fragmentation have played a key role in the evolution of our species, *homo sapiens* [1].

A major advantage of GCMET is that it is computationally inexpensive. Thus, GCMET can not only produce high quality reconstructions of the last 800 ka, but also quantify and explore uncertainties in the external forcings, e.g., atmospheric CO$_2$, as we did by going back to 2 Ma. In doing so, we reconstructed the equivalent of hundreds GCM snapshots, a prohibitive endeavour for the foreseeable future. A way to understand the excellent fit of GCMET predictions against time series of climate
proxies is that our approach captures the slow manifold of the stochastic climate system, thus allowing us to efficiently describe the behaviour over the longer, millennial, time scales. In turns, this implies that the glacial-interglacial climate of the Middle and Late Pleistocene responded in a consistent manner to orbital forcings and CO₂. It will interesting in the future to test whether this approximation holds for the Early Pleistocene with its faster ice age cyclicity of 41 ka; for this endeavour, we currently lack enough of estimates for CO₂ before the Mid-Pleistocene Transition, but GCMET is fully capable of covering the appropriate time periods if enough estimates become available. For the moment, we can offer a detailed, coherent reconstruction of the past 800 thousand years, which allowed us to pinpoint long-term potential refugia that have been characterised by the same habitat, and we expect that this will open up new ways to study the impact of past climate in a number of disciplines such as ecology and anthropology.
Methods

The global climate model emulator GCMET

The multiple linear regression model of GCMET-LO  
GCMET is derived from 72 available HadCM3 snapshot simulations [10, 11] (https://www.paleo.bristol.ac.uk/ummodel/data/tdwza/standard_html/tdwza.html, last accessed on 05 Oct 2018). It is a linear regression model for each individual model grid box with the following independent variables: atmospheric CO$_2$ concentrations as a major greenhouse gas, and eccentricity, obliquity, and precession as orbital parameters [14]. The sine function has been applied to the precession parameter which is expressed as longitude of the perihelion (in degrees) to make it a continuous function (was in degrees). Atmospheric CO$_2$ concentrations are the same as in the respective HadCM3 time slice simulation, e.g., 280 ppmv for 0 ka before present (BP). The available 72 HadCM3 simulations cover the last 120,000 years in 2,000-year intervals from 120,000 to 24,000 ka BP and in 1,000-year intervals from 22,000 to present-day.

The dependent variables are temperature $T$, precipitation $P$, or specific humidity $Q$. All independent variables, i.e., the predictors, are applied as normalised forcings. Thus, the resulting regression coefficients, or $\beta$ coefficients, can be compared across different climate variables, i.e., temperature and precipitation, and across each other (Extended Data Figures 5–7).

Variations of a climate variable $X$ based on a multiple linear regression model for the deviations from the mean, i.e., the anomalies $X'$, such that $X = \overline{X} + X'$ with $\overline{X}$ being the mean of $X$. The equation for $X'$ then is:

$$X'(x,y,t) = \beta_{CO_2}(x,y)CO_2'(t) + \beta_{e}(x,y)e'(t) + \beta_{\varepsilon}(x,y)\varepsilon'(t) + \beta_{\Omega}(x,y)\Omega'(t)$$

(1)

In this equation the $\beta$s are the regression coefficients for the respective predictor. CO$_2$ describes atmospheric CO$_2$ concentrations. $\varepsilon$ denotes obliquity, $e$ eccentricity, and $\Omega$ the sine of the longitude of the perihelion, i.e., the precessional component of Earth’s orbit around the sun. The prime ($'$) denotes the anomalies from the mean. The variables $x$, $y$, and $t$ represent the spatial, i.e., longitude and latitude, and the time coordinates. To make the linear regression well-conditioned, all independent variables have been normalised, i.e., the mean has been subtracted and the data has then been divided by their standard deviation. To prevent our linear regression model from predicting negative precipitation values, we apply a logarithmic transformation first. For bounded variables such as precipitation this is a common procedure. In the case of precipitation, the linear regression coefficients are predicting the response in terms of anomalies in the exponent. For similar reasons we transform specific humidity using the logit functions, $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$, which maps values from [0,1] to $[-\infty, +\infty]$; the units of specific humidity are [kg/kg] and its values fall in the range between 0 and 1. The decomposition of temperature $T$, precipitation $P$, and specific humidity $Q$ into anomalies, i.e., the $X'$ on the left hand side of Eq. 1 is:
We also consider changes in surface type, i.e., ocean, land, and ice. For example, around the coastlines, land can turn into ocean due to rising sea levels and vice versa, or the expanding ice sheets turn land into ice. Both, precipitation and temperature respond to different surface type in a different way. Therefore, each of the surface types (ocean, land, and ice) yields a distinct linear regression model.

For the improved precipitation model (as mentioned in the main text) we used temperature $T$ and specific humidity $Q$ as independent variables

$$\begin{align*}
X' &= (\log P)' = \beta_T T' + \beta_Q (\logit Q)'
\end{align*}$$

This leads to precipitation predictions with a lower root mean square error over land (it is also explained below and shown in Extended Data Figure 1). For the prediction of the climate before 120 ka BP this means that we first need to reconstruct $T$ and $Q$, and then we can use the $\beta$ coefficients for $T$ and $Q$ to reconstruct $P$.

In contrast to existing emulator approaches [7, 8, 9], we provide local-scale reconstructions which lead to reasonable agreement with existing palaeo-climate proxies as shown by the comprehensive model–data comparison. Furthermore, because the parameter sampling is based on realistic glacial cycle snapshot simulations, the obtained regression coefficients are good enough approximation to predict previous glacial–interglacial climate states well.

**Training and test data**

To make useful predictions and to evaluate the skill of our model, we need to have an independent test data set. A sensible choice is to use 80% of the data for the training of a model and 20% for the aforementioned test of the model. For a 80%/20% division of the 72 time slices into training and test data, i.e., 14 or 58 out of 72, there are \( \binom{72}{14} = \binom{72}{58} \approx 3 \times 10^{14} \) possible combinations.

Instead of randomly dividing the data into the training/test data, we follow an approach with the aim to preserve as much variance as possible in the training data. The idea is to choose the parameter sets (i.e., the independent variables, not the dependent climate variables) in such a way that they retain the most variance. First, we derive the covariance matrix of the full parameter set ($n=72$) and calculate the eigenvalues of the covariance matrix. In the next step, we randomly create a training data set ($k=58$) for which we compute the covariance. If the covariance of this sample training set is larger than the full covariance matrix, i.e., the eigenvalues of the covariance matrix are larger than the eigenvalues of the covariance matrix of the full parameter set, this sample parameter set is marked as a candidate for the final training set. After several iterations ($N=10,000$), we sum up how many times each time slice has appeared within a candidate training set. We then rank all time slices according to this number. In the final step, we pick the 80% top-ranked time slices as training data.
Model validation

For the model validation, we use $R^2$ values, a goodness of fit estimator of the training data, and the root mean squared error (RMSE), an estimator of the goodness of the model for the prediction of the test data (see Extended Data Figure 8). Overall, our linear model is a better predictor for temperature than for precipitation.

Temperature responds more directly to local forcings because temperature is determined by the energy balance of downward and upward longwave and shortwave radiation and turbulent heat fluxes. The downward shortwave radiation depends on the incoming solar radiation, which is determined by orbital variations, whereas the downward longwave radiation is determined by greenhouse gases such as CO$_2$ and water vapour, as well as by cloud cover. Large-scale circulation changes have a much smaller effect on temperature. It is therefore locally far better constrained by global CO$_2$ and orbital variations. This increases the predictive skill of our linear regression model substantially leading to high $R^2$ values and low RMSEs.

The matter is more complicated for precipitation because it is a consequence of the hydrological cycle, which itself depends mainly on large-scale atmospheric dynamics, such as the monsoonal systems in the tropics and subtropics, or the midlatitude storm systems. To a lesser extend do local interactions between the atmosphere and the surface, i.e., ocean, land, or ice play a role. Examples are evaporation and transpiration over the ocean, or deep convection over the tropics. Processes and circulation features like moisture transport or the atmospheric Hadley cell dynamics determine to a much larger extent the non-local response of precipitation to CO$_2$ or orbital variations. Because of the larger dynamical component of the hydrological cycle, as compared to temperature, precipitation is much less constrained by external forcings than temperature. Therefore, the linear regression model has less predictive skill for precipitation than for temperature. However, it turns out that the predictive skill for precipitation can be improved by using temperature and specific humidity as predictors instead of the orbital parameters and CO$_2$. By doing so the RMSEs can be substantially reduces, especially over land (Extended Data Figure 1).

The regression coefficients

To get an idea of how reliable our estimate for predictors are, we calculate the p-values for each of the predictors, i.e., the beta coefficients. Here, the p-value tests the null hypothesis whether the coefficient is equal to zero, which means that the specific predictor has no effect. If the p-value is below a certain threshold—in our case below the 5% significance level: $p < 0.05$—the null hypothesis can be rejected. That means that the specific predictor is a meaningful addition to our linear regression model and any changes in the associated predictor are related to changes in the corresponding climate variable. Regions for which the null hypothesis cannot be rejected are displayed as shaded and hatched in Extended Data Figures 5–7. In these regions, we set the $\beta$ coefficients to zero and the associated forcing has no effect.

Increasing to high resolution in GCMET

Using nine high-resolution HadAM3H simulations, which cover the deglaciation since 21 ka BP (21, 18, 15, 12, 10, 8, 6, 3, and 0 ka BP), we are able to increase the spatial resolution from 3°, which is the spatial resolution of GCMET after the linear regression step (and the same as the coarse resolution of the original HadCM3 snapshots), to ca. 1°. We do so by calculating the difference between equivalent coarse- and high-resolution snapshots. For example, the difference at 10 ka BP is $\Delta_{10\text{kaBP}} = \text{HadAM3H}_{10\text{kaBP}} - \text{HadCM3}_{10\text{kaBP}}$. We choose to interpolate the differences linearly according to their CO$_2$ levels, e.g., 231 ppm at 10 ka BP, because any statistical model with more
than one variable would require more snapshots to adequately predict the differences. Thus, we simply assume that the differences between a coarse- and high-resolution climate can be explained as a function of the CO$_2$ forcing, i.e., $\Delta_{10,000}$ = $\Delta_{231}$ ppm. Now, for any period in the past, e.g., 300 ka BP, we add the high-resolution difference, i.e., the $\Delta$, which corresponds to the respective CO$_2$ level, to the coarse-resolution reconstruction. Note that the downscaling approach captures the regional-scale dynamics of the GCM in this step, which change over time. This is in contrast to commonly used "delta approach" for downscaling.

**Boundary conditions: CO$_2$, global sea-level, and Northern Hemisphere ice sheets** For realistic high-resolution reconstructions the model boundary conditions need to be known: atmospheric CO$_2$ levels, global sea levels (for the land-sea mask), and the extent of Northern Hemisphere (NH) ice sheets. The longest, quasi-continuous record of past CO$_2$ levels is the 800,000 years long CO$_2$ record from the EPICA Dome C ice core in Antarctica [13]. Before that we use point-wise CO$_2$ estimates that go back about 2 Ma [16], coinciding with the earliest time for which we are able to generate reasonable climate reconstructions (Extended Data Figure 9).

Because there are no self-consistent continuous reconstructions of NH ice sheets available that span the last 2 Ma, we use modelled NH ice sheet extents and heights which are available every 1 ka for the years from 800–123 ka BP from CLIMBER-2/SICOPOLIS simulations [4]. For the period from 122–0 ka BP we use the ice sheet configurations from the ICE-6G data set [15] (http://www.atmosphys.utoronto.ca/~peltier/data.php, last accessed 09.11.2018). For simplicity, we assume present-day ice sheets for any period before 800 ka BP. Topographic changes due to growing or shrinking ice sheets are derived from a global sea-level record [20] which have been added on top of present-day coast lines while preserving inland lakes.

**Comparison with proxy reconstructions**

Despite the increasing number of available paleoclimate proxies, only a small percentage can be used for a quantitative comparison to climate models because translating sediment core data into actual climate variables remains a difficult task. Marine sediment cores are the exception, as they are useful archives of sea surface temperature (SST). Because the associated biogeochemistry is relatively straightforward, marine proxies can be utilized as so-called paleo-thermometers and are thus well suited for a direct proxy–model comparison. For these proxies, we make a direct comparison between MAT and SST, quantified both in terms of correlation between the predicted and observed time series and the RMSE. 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 permanent sea ice could make a straightforward comparison between both variables problematic.

For terrestrial proxies, for which a translation into climatic variables is not straightforward, we simply quantify the correlation between the two standardized time series without a more detailed error quantification in terms of the RMSE. However, the interpretation of terrestrial of climate proxies can also be problematic. For example, pollen-based vegetation reconstructions are suggested to be less reliable as climate proxies, particularly for interglacials [21].

We have assembled existing long-term SST proxy reconstructions (see Extended Data Figure 2).
which cover at least a period of about 150 ka BP during the Middle and Late Pleistocene (specifically, the last 800 ka, for which we can reconstruct the climate continuously).

**Ecosystem reconstructions**

We use a random forest classifier [19, 22] which is trained by a set of four climate variables from GCMET: minimum and mean annual temperature, and minimum and mean annual precipitation, to reconstruct the present-day distribution of the 14 ecoregions. The required present-day data has been split into a training (80%) and a test data set (20%). The classification factors from this training data set were then applied to predict ecosystem changes of the last 800,000 years.

The goodness of the predictions by the random forest classifier can be estimated by the so-called receiver operating characteristic (ROC, see Extended Data Figure 10). A ROC curve displays the true positive rate against the false positive rate and the closer that curve is to the upper left corner, the better the prediction for a specific ecosystem is. For example, the point at coordinate (0,1) represents the best possible prediction with 100% sensitivity (i.e., no false negatives) and 100% specificity (i.e., no false positives). The diagonal line depicts a prediction by random guessing.

The random forest classification is very close to perfect classification for the average of all ecosystem types and the area under the curve (also given in the legend to Extended Data Figure 10) is an estimator for the goodness of the classification. Except for a few instances, such as for "Tropical & Subtropical Coniferous Forests" and "Mangroves", this value is larger than 0.9 (average 0.98).

**Data and model code availability**

**High-resolution climate data for the last 800 ka and 2 Ma** A continuous climate data set for the last 800,000 years are publicly available at [link to data repository]. We included the following variables in 1,000 year intervals and in at a 1° horizontal resolution:

- mean annual temperature
- minimum annual temperature
- mean annual precipitation
- minimum annual precipitation
- mean annual specific humidity (needed for mean annual precipitation)
- minimum annual specific humidity (needed for minimum annual precipitation)
- 14 major habitats according to the *global 200* defined by the WWF
- 3 aggregated ecosystem classifications (from the 14 major habitats): "open habitat", "forests", and "sparsely vegetated"

For the reconstruction if the 2 Ma with the sporadic CO$_2$ records (52 time steps) we provide the following variables:

- mean annual temperature (ensemble mean, n=50)
- mean annual temperature (ensemble standard deviation, n=50)
GCMET and proxy time series  The data generated for the individual time series comparisons of GCMET and proxy records (Extended Data Figures 2 & 3) is also publicly available as an MS Excel file at [link to data repository].

GCMET model code, analysis, and visualisation scripts  The model code for GCMET as well as the code for the analysis and visualisation of figures [23, 24] is publicly available at [link to model repository].

Author contributions
AM and MK devised the project; MK devised and implemented the emulator 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 wrote the methods and prepared the figures.

Competing interests
The authors declare no competing interests.

Acknowledgements
This work was supported by an ERC Consolidator Grant to AM (Local Adaptation 647787).

References


Figure 1: Schematic of the GCMET components: A linear regression combines 72 HadCM3 snapshot simulations with the external forcings, i.e., CO$_2$ and the three orbital parameters, which provides the basis of the long-term climate reconstructions of the last 800 thousand (or 2 million) years. Using 9 high-resolution snapshots covering the last deglaciation provides the basis of the downscaling approach based on CO$_2$ which yields the final high-resolution long-term climate reconstructions of GCMET.
Figure 2: Root mean square error of the GCMET-LO predictions for the 14 HadCM3 snapshots for (a) MAT and (b) MAP (lower is better). (c) Present-day, i.e., 0 ka BP, temperature–precipitation phase diagram for Asia, North America, Africa, Europe, South America, and Australia, as modelled by LOVECLIM and reconstructed by GCMET-LO and GCMET and compared to observed multi-annual mean values (grey contours) for the period from 1961-1990 [12]. The numbers in each plot indicate the number of grid points covering the respective continent. (d) Maps of present-day temperature (in K) and precipitation (in mm/a) as reconstructed by GCMET for the six continents.
Figure 3: (a) Global mean temperature for the last 2 Ma as predicted by GCMET based on different CO₂ records in comparison with a proxy-based global mean temperature reconstruction [17]. Furthermore, the time series from the 72 HadCM3 snapshots for the last 120 ka have also been added. Note the change in the spacing of the time axis at 800 ka and 140 ka BP. (b) Map of correlation coefficients between marine (in terms of sea surface temperature) and terrestrial climate proxy time series and mean annual temperatures as reconstructed by GCMET-LO for the respective locations. The individual time series and references for the proxies can be found in the Extended Data Table 1, Extended Data Figures 2 & 3 and in the Methods. (c) Box plots showing the range of correlations between GCMET (LOVECLIM) and the respective marine and terrestrial proxies. Time series of three selected (d) marine and (e) terrestrial proxies and the corresponding reconstructions by GCMET. While marine proxies are plotted on the same y axis, different scales have been used for terrestrial proxies.
Figure 4: (a) Map of 14 major terrestrial habitats as defined by the WWF [18] for present-day and (b) as reconstructed with GCMET inputs of minimum and annual temperature and minimum and annual precipitation. (c) Stability of open habitats, such as grasslands and savannahs, and forest habitats, and sparsely vegetated regions across the world through the last 800,000 years. Regions in which the habitats have been unstable, i.e., of different type, for more than 90% are coloured in grey.
<table>
<thead>
<tr>
<th>core/name</th>
<th>location</th>
<th>lon</th>
<th>lat</th>
<th>type</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY04</td>
<td>Eastern Equatorial Pacific</td>
<td>-95.0</td>
<td>4.0</td>
<td>SST</td>
<td>[25, 26]</td>
</tr>
<tr>
<td>MD06-2986</td>
<td>Tasman Sea</td>
<td>167.9</td>
<td>-43.4</td>
<td>SST</td>
<td>[25, 27]</td>
</tr>
<tr>
<td>ODP1125</td>
<td>Southwestern Pacific</td>
<td>-178.2</td>
<td>-42.6</td>
<td>SST</td>
<td>[28]</td>
</tr>
<tr>
<td>ODP1123</td>
<td>Southwestern Pacific</td>
<td>-171.5</td>
<td>-41.8</td>
<td>SST</td>
<td>[25, 29]</td>
</tr>
<tr>
<td>ODP846</td>
<td>Eastern Equatorial Pacific</td>
<td>-90.8</td>
<td>-3.1</td>
<td>SST</td>
<td>[25, 30]</td>
</tr>
<tr>
<td>ODP1239</td>
<td>Eastern Equatorial Pacific</td>
<td>-82.1</td>
<td>-0.7</td>
<td>SST</td>
<td>[31]</td>
</tr>
<tr>
<td>ODP982</td>
<td>North Atlantic</td>
<td>-15.9</td>
<td>57.5</td>
<td>SST</td>
<td>[25, 32]</td>
</tr>
<tr>
<td>ODP1020</td>
<td>Northeastern Pacific</td>
<td>-126.4</td>
<td>41.0</td>
<td>SST</td>
<td>[25, 33]</td>
</tr>
<tr>
<td>ODP1146</td>
<td>South China Sea</td>
<td>116.3</td>
<td>19.5</td>
<td>SST</td>
<td>[25, 34]</td>
</tr>
<tr>
<td>ODP1143</td>
<td>Western Equatorial Pacific</td>
<td>113.3</td>
<td>9.4</td>
<td>SST</td>
<td>[25, 35]</td>
</tr>
<tr>
<td>ODP1090</td>
<td>Southeastern Atlantic</td>
<td>8.9</td>
<td>-42.9</td>
<td>SST</td>
<td>[25, 36]</td>
</tr>
<tr>
<td>ODP1012</td>
<td>Northeastern Pacific</td>
<td>-118.4</td>
<td>32.3</td>
<td>SST</td>
<td>[25, 37]</td>
</tr>
<tr>
<td>ODP1082</td>
<td>Southeastern Atlantic</td>
<td>11.8</td>
<td>-21.1</td>
<td>SST</td>
<td>[38]</td>
</tr>
<tr>
<td>MD06-3018</td>
<td>Tropical Western Pacific</td>
<td>166.2</td>
<td>-22.6</td>
<td>SST</td>
<td>[39]</td>
</tr>
<tr>
<td>ODP722</td>
<td>Arabian Sea</td>
<td>59.8</td>
<td>16.6</td>
<td>SST</td>
<td>[25, 34]</td>
</tr>
<tr>
<td>ODP882</td>
<td>Northwestern Pacific</td>
<td>167.6</td>
<td>50.4</td>
<td>SST</td>
<td>[40]</td>
</tr>
<tr>
<td>MD97-2140</td>
<td>Western Pacific Warm Pool</td>
<td>141.5</td>
<td>2.0</td>
<td>SST</td>
<td>[25]</td>
</tr>
<tr>
<td>MD96-2048</td>
<td>Mozambique Channel</td>
<td>36.0</td>
<td>-26.2</td>
<td>SST</td>
<td>[41]</td>
</tr>
<tr>
<td>ODP1172</td>
<td>Tasman Sea</td>
<td>149.9</td>
<td>-44.0</td>
<td>SST</td>
<td>[42]</td>
</tr>
<tr>
<td>DSDP594</td>
<td>Southwest Pacific</td>
<td>175.0</td>
<td>-45.5</td>
<td>SST</td>
<td>[25, 43]</td>
</tr>
<tr>
<td>DSDP607</td>
<td>North Atlantic</td>
<td>-33.0</td>
<td>41.0</td>
<td>SST</td>
<td>[25, 44]</td>
</tr>
<tr>
<td>PS75034-2</td>
<td>Southeastern Pacific</td>
<td>-80.1</td>
<td>-54.4</td>
<td>SST</td>
<td>[25, 45]</td>
</tr>
<tr>
<td>ODP871</td>
<td>Western Equatorial Pacific</td>
<td>172.3</td>
<td>5.6</td>
<td>SST</td>
<td>[46]</td>
</tr>
<tr>
<td>ODP806B</td>
<td>Western Equatorial Pacific</td>
<td>159.4</td>
<td>0.3</td>
<td>SST</td>
<td>[25, 47]</td>
</tr>
<tr>
<td>Baoji</td>
<td>China</td>
<td>107.1</td>
<td>34.4</td>
<td>rainfall</td>
<td>[48]</td>
</tr>
<tr>
<td>Soreq</td>
<td>Israel</td>
<td>36.0</td>
<td>31.4</td>
<td>δ¹⁸O</td>
<td>[49]</td>
</tr>
<tr>
<td>Lake El’gygytgyn</td>
<td>Russia</td>
<td>172.0</td>
<td>67.5</td>
<td>χ</td>
<td>[50]</td>
</tr>
<tr>
<td>Chanwu</td>
<td>China</td>
<td>107.7</td>
<td>35.2</td>
<td>δ¹⁸O</td>
<td>[51]</td>
</tr>
<tr>
<td>Dead Sea</td>
<td>Israel</td>
<td>35.0</td>
<td>30.5</td>
<td>lake level</td>
<td>[52]</td>
</tr>
<tr>
<td>Devil’s Hole</td>
<td>Nevada, USA</td>
<td>-116.3</td>
<td>36.4</td>
<td>δ¹⁸O</td>
<td>[53]</td>
</tr>
<tr>
<td>Tzavoa</td>
<td>Israel</td>
<td>35.2</td>
<td>31.2</td>
<td>δ¹⁸O</td>
<td>[54]</td>
</tr>
<tr>
<td>Yimaguan Luochuan</td>
<td>China</td>
<td>108.5</td>
<td>35.8</td>
<td>χ/f/d</td>
<td>[25, 55]</td>
</tr>
<tr>
<td>Weinan</td>
<td>China</td>
<td>108.8</td>
<td>34.4</td>
<td>MAT</td>
<td>[56]</td>
</tr>
<tr>
<td>Negev</td>
<td>Israel</td>
<td>34.8</td>
<td>30.6</td>
<td>δ¹⁸O</td>
<td>[54]</td>
</tr>
<tr>
<td>Peqin</td>
<td>Israel</td>
<td>36.0</td>
<td>32.6</td>
<td>δ¹⁸O</td>
<td>[49]</td>
</tr>
<tr>
<td>Xifeng</td>
<td>China</td>
<td>107.6</td>
<td>35.7</td>
<td>Fed/Fet</td>
<td>[51]</td>
</tr>
<tr>
<td>EPICA Dome C</td>
<td>Antarctica</td>
<td>123.4</td>
<td>-75.0</td>
<td>ΔT</td>
<td>[57]</td>
</tr>
<tr>
<td>Kesang</td>
<td>western China</td>
<td>81.8</td>
<td>42.9</td>
<td>δ¹⁸O</td>
<td>[58]</td>
</tr>
<tr>
<td>Clearwater</td>
<td>Borneo</td>
<td>114.9</td>
<td>4.1</td>
<td>δ¹⁸O</td>
<td>[59]</td>
</tr>
<tr>
<td>Tenaghi Philippon</td>
<td>Greece</td>
<td>24.2</td>
<td>41.0</td>
<td>pollen</td>
<td>[25, 60]</td>
</tr>
<tr>
<td>Sanbho-Dongge</td>
<td>China</td>
<td>110.4</td>
<td>31.7</td>
<td>δ¹⁸O</td>
<td>[61]</td>
</tr>
</tbody>
</table>

Extended Data Table 1: Marine and terrestrial proxy records that have been used in this study, their location, coordinates and the respective reference.
Extended Data Figure 1: Root mean square error of the GCMET-LO predictions for the 14 HadCM3 snapshots for mean annual precipitation with (a) CO$_2$ and orbital parameters as independent variables and (b) mean annual temperature and specific humidity as independent variables (lower values are better).
Extended Data Figure 2: Time series of 39 Middle and Late Pleistocene marine sea surface temperature proxies (black dots) and modelled mean annual temperature at their closest location (blue lines). Proxy-derived and model temperature are on the same scale, in °C).
Extended Data Figure 3: Time series of 20 Middle and Late Pleistocene terrestrial proxies (black dots) and modelled mean annual temperature at their closest location (blue lines), in °C.
Extended Data Figure 4: Parameter space of the four independent variables (i.e., external forcing or regressors) as scatter plot matrix for last 800 ka (blue dots). The black dots highlight the location of the independent variables er sets of the 58 HadCM3 snapshot simulations which we used as training data (80% of the total 72) for the linear regression model.
Extended Data Figure 5: Regression coefficients for mean annual temperature. Regions where the respective coefficient is not statistically significant (p < 0.05) are hatched and shaded.
Extended Data Figure 6: Regression coefficients for mean annual precipitation. Regions where the respective coefficient is not statistically significant ($p < 0.05$) are hatched and shaded.

Extended Data Figure 7: Regression coefficients for mean annual precipitation with alternative independent variables temperature and specific humidity. Regions where the respective coefficient is not statistically significant ($p < 0.05$) are hatched and shaded.
Extended Data Figure 8: $R^2$ values as estimator for the goodness of the model (higher is better) using the training data, and root mean square errors (RMSE) as estimators of the goodness of fit (lower is better) using the test data. Shown are the $R^2$ and RMSEs for mean annual temperature, precipitation, and the alternative model for precipitation—based on temperature and specific humidity.
Extended Data Figure 9: Time series of external parameters: CO$_2$ and orbital parameters for the last 2 million years. The continuous CO$_2$ record is from the EPICA Dome C ice core in Antarctica [62]. The point-wise CO$_2$ record is based on boron isotopes from planktonic foraminifera [63]. The orbital parameters are numerical solutions for the Earth’s orbit and rotation in terms of eccentricity, precession, and obliquity [64].
Extended Data Figure 10: A receiver operating characteristic curve for the random forest classifier of the WWF 14 major habitats. The upper left corner represents a perfect prediction of an ecosystem, while the diagonal line represents a prediction made by random guessing. The closer the ROC curve is to the perfection point (0,1) the better the random forest classification is.
References


