The relative role of orbital forcing and CO_2 and ice sheet feedbacks on Quaternary climate

C. J. R. Williams, N. S. Lord, A. T. Kennedy-Asser, D. A. Richards, M. Crucifix, A. Kontula, M. Thorne, P. J. Valdes, G. L. Foster, R. Brown, E. L. McClymont and D. J. Lunt.

Supplementary Material

TEXT

Brief Description of Previously-Published Emulator

The emulator presented in Lord *et al*. (2017) was calibrated on GCM simulations run using the HadCM3 climate model, a fully coupled atmosphere-ocean GCM developed by the UK Met Office. The specific model setup is denoted HadCM3B-M2.1aE and is described in Valdes *et al*. (2017). The GCM is coupled to the land surface scheme MOSES2.1 (Met Office Surface Exchange Scheme), which is in turn coupled to the dynamic vegetation model TRIFFID (Top-down Representation of Interactive Foliage and Flora Including Dynamics) (Cox *et al*. 2002). TRIFFID calculates the global distribution of vegetation based on five plant functional types: broadleaf trees, needleleaf trees, C3 grasses, C4 grasses and shrubs. The horizontal resolution of the atmosphere component is 2.5° latitude by 3.75° longitude with 19 vertical levels, whilst the ocean has a resolution of 1.25° by 1.25° and 20 vertical levels.

When compared with the latest generation of GCMs, such as those included in the IPCC Fifth Assessment Report (IPCC 2013), HadCM3 can no longer be considered as state-of-the-art. However, it is relatively computationally efficient which makes it useful for running experiments that cover relatively long periods of time (of several centuries or longer), as well as for running ensembles with a large number of ensemble members, as is required in this study. For this reason, the model is still widely used in climate research, both in palaeoclimatic studies (e.g. Prescott *et al*. 2014) and in projections of future climate (e.g. Armstrong *et al*. 2016, Lord *et al*. 2017). It has also previously been used in conjunction with a statistical emulator to investigate climate sensitivity (Araya-Melo *et al*. 2015).

Two 60-member ensembles of HadCM3 simulations were used by Lord *et al*. (2017), with input parameters varied across four dimensions: atmospheric CO_2 concentration and the three main orbital parameters of longitude of perihelion (ϖ), obliquity (ε) and eccentricity (e). The ranges of orbital and CO_2 values sampled in the ensembles (see Table S1) included those appropriate for wide range of palaeo and future climate states (Lord *et al*. 2016). Eccentricity and longitude of perihelion were combined under the forms $e\sin\varpi$ and $e\cos\varpi$ given that, in general at any point in the year, insolation can be approximated as a linear combination of these two terms and obliquity (Loutre 1993). Latin hypercube sampling was used in order to optimally fill the 4-dimensional input space.

Two different fixed ice sheet configurations were used, consisting of changes to the Greenland (GrIS) and West Antarctic (WAIS) ice sheets only. One represents their present-day extents, denoted the "*modice*" ensemble, and the other their reduced extents, based on the PRISM4 mid-Piacenzian reconstructions (Dowsett *et al*. 2016), denoted the "*lowice*" ensemble. An additional suite of simulations was also included, produced by (Singarayer and Valdes 2010) (denoted the "*LGC*" ensemble), and also run using HadCM3. These additional simulations are a series of snapshot simulations covering the last glacial cycle (LGC; 120 kyr BP to present day), forced by changes in orbit, atmospheric CO_2 concentration, and ice sheet evolution. Each of the 60 sampled combinations of orbital and CO_2 values therefore had two sets of GCM output; one with *modice* ice sheets and one with *lowice* ice sheets. For more information about the GCM ensembles, and the sampling and GCM simulation post-processing methods used, see Lord *et al*. (2017).

Methodology

To summarise the methodology (see Lord *et al*. 2017 for full details), the input parameter data (denoted $\boldsymbol{D}$ in Lord *et al*. (2017)) for the interglacial emulator was a 120×5 matrix ($n \times p$), consisting of 120 GCM simulations (n) and five input factors (ε, $e\sin\varpi$, $e\cos\varpi$, CO_2, and GSL; denoted p). On the other hand, the input data for the glacial emulator was a 122×5 matrix. The matrix containing the output data from the GCM (denoted $\boldsymbol{Y}$) had dimensions of $96 \times 73 \times 120$ (longitude $\times$ latitude $\times$ n) for the interglacial emulator, and $96 \times 73 \times 122$ for the glacial emulator. A PC analysis was performed on each set of output data, the results of which were then used to calibrate the two emulators. Five correlation length hyperparameters (δ) were used, one for each of the five input factors, which describe the smoothness of the climate response in the GCM data to the input conditions. A nugget term (υ) was also used, which accounts for any non-linearity in the output response to the inputs, non-explicitly specified inactive inputs, and the effects of lower-order PCs that are excluded from the emulator. The optimal values for these hyperparameters and the number of PCs retained were calculated during calibration and evaluation of the emulator, discussed in the next section. The GCM data used in this study are mean annual SAT and MAP, although these were each emulated separately using different emulators.

Calibration and Evaluation of Methodology

Four emulators were created in total, two of which were calibrated (and projected) on SAT (an interglacial and a glacial version), and two of which were calibrated (and projected) on MAP. Before being used, the two SAT emulator configurations were optimised via the method described in Lord *et al*. (2017). This involved generating an ensemble of emulators with different numbers of PCs retained (the majority of the PCs were discarded as they only account for a very small amount of total variation in the GCM data) and different values for a number of hyperparameters used in the emulator, including the correlation length hyperparameters (δ) and the nugget parameter (ν). The performances of the different emulators were compared in order to identify the optimal number of PCs to retain and the hyperparameter values to use. It was demonstrated by Lord (2017) that the optimisation can be carried out on SAT, and then the optimised configuration applied to other climate variables (e.g. precipitation) with no significant loss of performance. Hence, this approach was adopted here.

In order to optimise the two emulators, each was calibrated on the GCM SAT data from its respective ensemble(s) of GCM simulations. The input factors ($\ln(CO_2)$, ε, $e\sin\varpi$, $e\cos\varpi$, and GSL) were standardised prior to the calibration being performed; each was centred in relation to its column mean, and then scaled based on the standard deviation (SD) of the column. Different emulator configurations were tested by varying the number of PCs retained, ranging from 5 to 20, and for each emulator configuration, the correlation length scales δ and nugget ν were optimised by maximisation of the penalized likelihood (see Lord *et al*. 2017 for further details). This optimisation was carried out in log space, ensuring that the optimised hyperparameters would be positive.

The performance of the optimised emulators was evaluated, using a leave-one-out approach and by comparing global SAT produced by the emulator for the LGM to a reconstruction based on proxy data and GCMs. The performance of each emulator was assessed using a leave-one-out cross-validation approach, in which a series of emulators was constructed and used to predict one left out experiment each time. For example, for the interglacial emulator (120 experiments), 120 separate emulators were calibrated with one experiment left out of each. This left out experiment was then reproduced using the corresponding emulator and the results were compared with the actual experiment results. The number of grid boxes for each experiment calculated to lie within different SD bands, and the root mean squared error (RMSE) averaged across all the 120 emulators were used as performance indicators to compare the different selected values for retained PCs and hyperparameters.

The two emulator configurations that performed best were selected as the final two optimised emulators. It was found that the optimised interglacial emulator retained 15 PCs (accounting for 90% of the total variance), and had length scales δ of 2.792 (ε), 1.310 ($e\sin\varpi$), 1.664 ($e\cos\varpi$), 0.523 (CO_2), and 10.000 (GSL), and a nugget of 0.000. The optimised glacial emulator retained 15 PCs (accounting for 81% of the total variance), and had length scales δ of 6.908 (ε), 7.499 ($e\sin\varpi$), 5.460 ($e\cos\varpi$), 1.003 (CO_2), and 0.290 (GSL), and a nugget of 0.050.

The results of the evaluation of the emulators are shown in Figure S1. The results suggest that the emulators perform relatively well, similar to the results from Lord *et al*. (2017), the emulators in which did not include glacial GCM simulations. The calibration and evaluation shows that the emulators are able to reproduce the left out ensemble simulations reasonably well, with no obvious systematic errors in their predictions. These emulator configurations were optimised on SAT data, and these same configurations (i.e. same number of retained PCs and hyperparameter values) were then also applied to emulate the other variables. This approach ensures that the results for the variables are consistent with each other.

The SAT anomaly (compared to PI) at the LGM, predicted using the emulator is shown in Figure S2b, and compared to that produced directly by HadCM3 (Figure S2a) and to the SAT anomaly reconstruction of Annan and Hargreaves (2013) (Figure S2c). The latter reconstruction was produced by combining a range of proxy records of climate from different parts of the globe with an ensemble of simulations performed with various climate models as part of the PMIP2 project (Braconnot *et al*. 2007). It can be seen that the large-scale features of SAT are similar between the three projections, with the most significant regions of cooling located over the Laurentide ice sheet in North America, and to a lesser extent the Fennoscandian ice sheet, in line with current understanding. There are, however, some discrepancies between the two modelled (HadCM3 and emulator) projections and the Annan and Hargreaves (2013) reconstruction. Discrepancies include the extent of cooling over the Laurentide and Fennoscandian ice sheets being larger in the modelled projections than in the reconstruction, and high latitudes (Arctic and Antarctic regions) also being slightly colder in these simulations compared to that of Annan and Hargreaves (2013). This enhanced cooling observed in the emulator at this time is highlighted by a global mean cooling value of 5.1°C, more than one degree cooler than the 4°C estimated by Annan and Hargreaves (2013). However, the temperatures over ice sheets in the Annan and Hargreaves (2013) reconstruction are uncertain, and are largely based on model results rather than direct proxies. If this reconstruction is taken to be correct, the emulated SAT projections during glacial conditions could therefore be considered as being somewhat cold-biased at high latitudes.

Linear factorisation

Before the linear-sum/shared-interaction factorisation was undertaken (discussed in the Results of the main manuscript), a linear factorisation (see Lunt *et al*. 2021) was initially conducted, in which each individual driver was added in turn, to transition from E_{00000} to E_{11111} (see 'Nomenclature and factorisation' in Methods of the main manuscript for more details). This is illustrated in Figure S5. Note that this is only one possible linear factorisation, or 'pathway'; this particular pathway involves adding one driving component incrementally i.e. E_{00000}, E_{10000}, E_{11000}, E_{11100}, E_{11110}, E_{11111} (see 'Nomenclature and factorisation' in Methods of the main manuscript for more details).

The results suggest that changes in CO_2 and, in particular, ice volume (using GSL as a proxy) generally result in significantly larger temperature variations, of several degrees or more, than those driven by orbital changes alone. This point is made more clearly, and indeed globally, by Figure S6, where the same single pathway is considered. Here, it is clear that the impacts of CO_2 (Fig. S6a) and ice (Fig. S6e) are having a much greater contribution to the overall signal (as measured by the M score compared to the 'all drivers' simulation) than any of the individual orbital components, with the latter (Fig. S6b-d) contributing less than half of the overall signal across much of the world. In certain places CO_2 and ice are contributing up to all of the overall signal, such as across the tropics for CO_2 (Fig. S6a). The ice component is providing the most contribution (accounting for almost all of the overall signal) over the polar latitudes (Fig. S6e). When compared to the proxy data, it is only when the CO_2 and ice components are added into the emulator does the temperature match the magnitudes suggested by the reconstructions, qualitatively suggesting that ice and CO_2 are providing the majority contribution to the 'all drivers' simulation (Fig. S5).

TABLES

Ensemble	Number of simulations	Input parameters				
		Orbital		Atmospheric CO_2 (ppmv)	Ice sheets
modice	60	ε	22.2 : 24.4	250 : 1901	Modern
		$E\sin\varpi$	-0.055 : 0.055		
		$E\cos\varpi$	-0.055 : 0.055		
lowice	60	ε	22.2 : 24.4	250 : 1901	PRISM4 Pliocene
		$E\sin\varpi$	-0.055 : 0.055		
		$E\cos\varpi$	-0.055 : 0.055		
LGC	62	LGC		LGC	ICE-5G LGC

Table S1. Input parameter set-ups for the modice, lowice and LGC (120-0 kyr BP) GCM ensembles, including sampling ranges.

		Emulated				
		Dome C	**ODP 1012**	**ODP 1123**	**ODP 1239**	**ODP 806**
Proxy	**Dome C**	620.018	143.853	87.3801	103.833	74.3982
	ODP 1012	209.391	327.926	321.399	360.423	273.541
	ODP 1123	98.5815	116.177	128.205	137.756	77.904
	ODP 1239	86.0269	150.248	197.004	200.483	157.456
	ODP 806	29.36	130.769	88.7556	121.533	128.86

Table S2. M scores between emulated and proxy data, for E_{11111}, at every possible combination of sites.

FIGURES

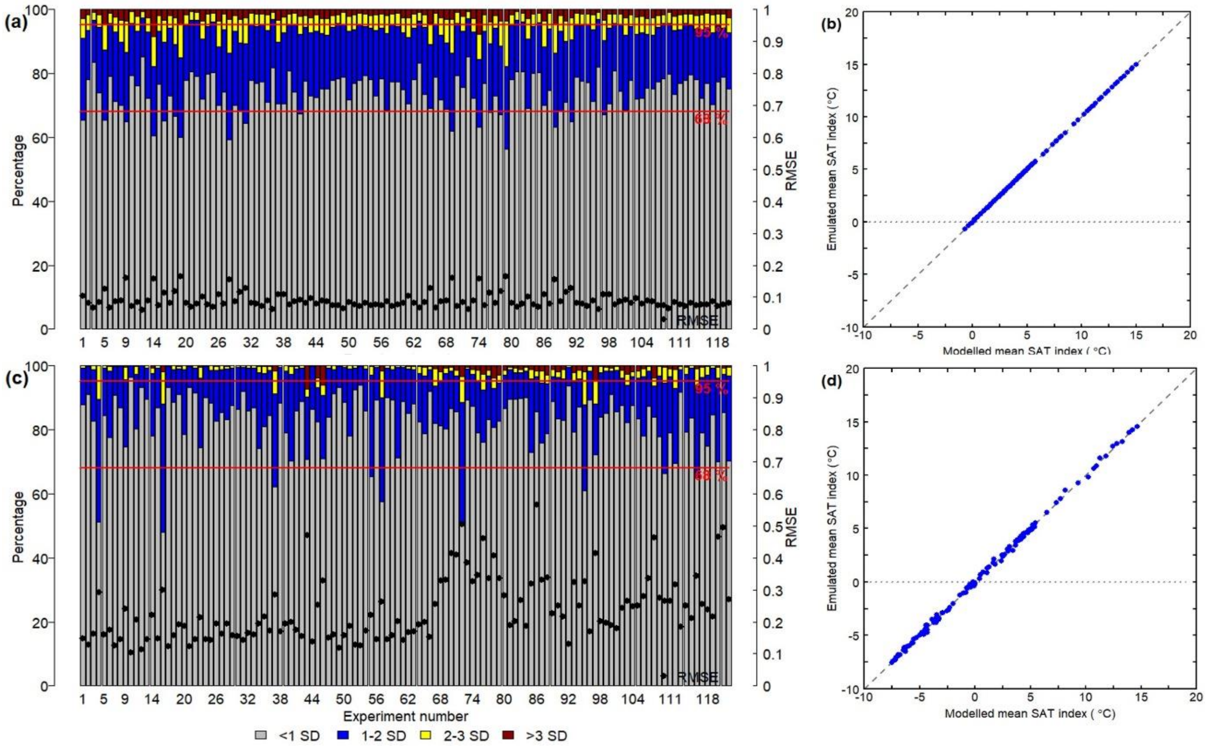

Figure S1. Evaluation of emulator performance for the interglacial emulator (top panel) and glacial emulator (bottom panel), both calibrated on SAT data. (a)+(c) Bars give the percentage of grid boxes for which the emulator predicts the SAT of the left-out experiment to within 1, 2, 3 and more than 3 SD. Also shown is the RMSE for the experiments (black circles). Red lines indicate 68 and 95 %. (b)+(d) Mean annual SAT index (°C) calculated by the emulator and the GCM. The 1:1 line (dashed) is included for reference. Note: this is the mean value for the GCM output data grid assuming all grid boxes are of equal size, hence not taking into account variations in grid box area. SAT is shown as an anomaly compared with the pre-industrial control simulation.

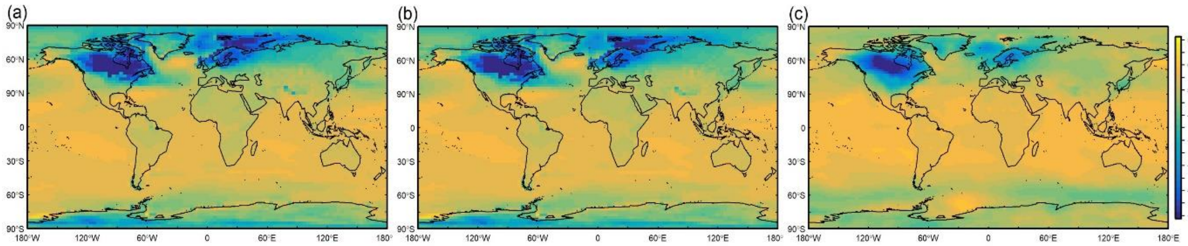

Figure S2. Maps of SAT anomaly (compared to pre-industrial; °C) at the Last Glacial Maximum (21 kyr BP) as projected by: a) HadCM3; b) the emulator; c) Annan and Hargreaves (2013), using a combination of proxy climate data and multi-model regression.

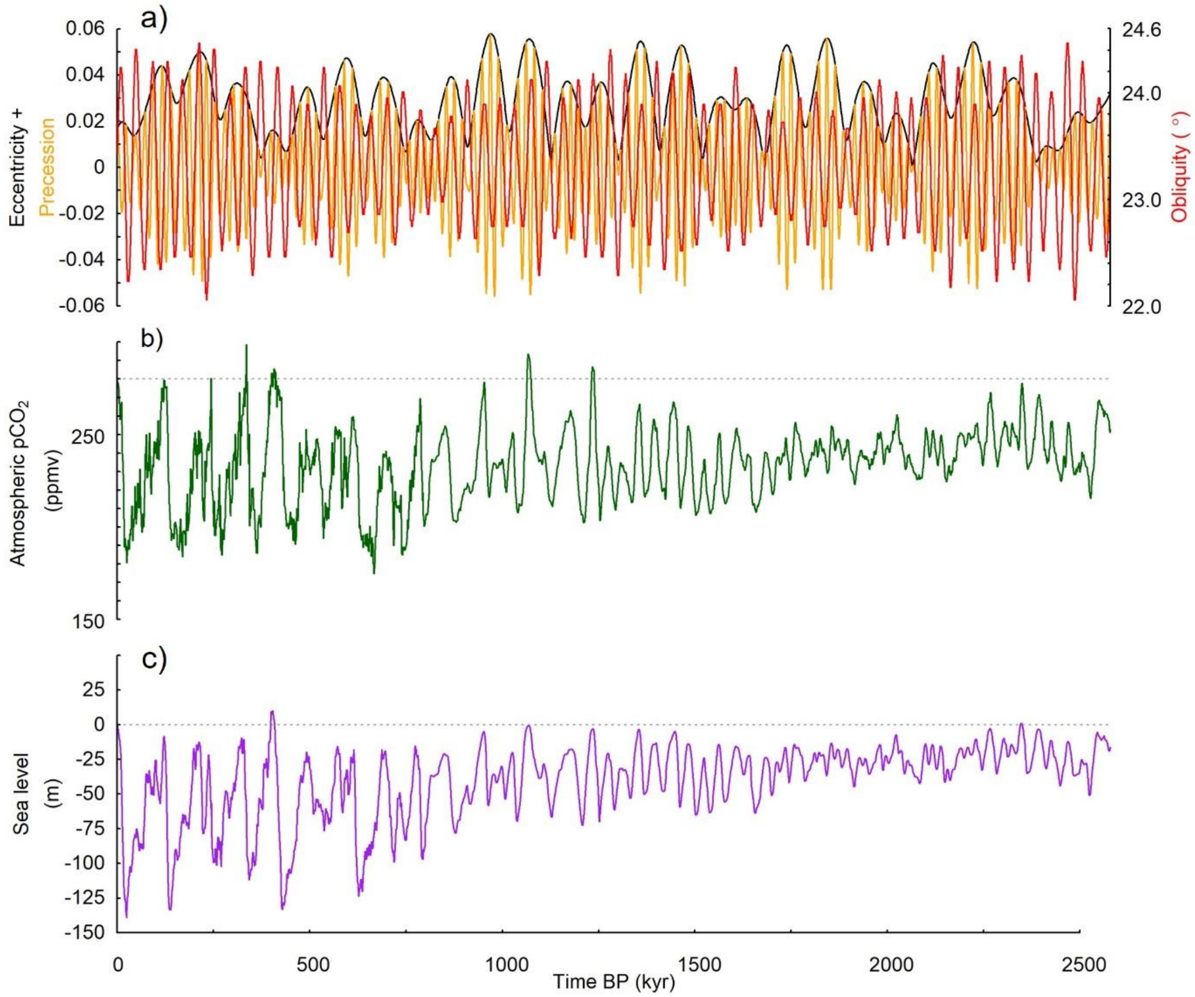

Figure S3. Climate forcing data used as input to the emulator for the last 2.58 Myr: a) Orbital variations (Laskar *et al*. 2004), showing eccentricity and precession on the left axis, and obliquity on the right axis; b) Atmospheric CO_2 concentrations, constructed from composite records from Antarctic ice cores (Bereiter *et al*. 2015) over the last 800 kya and based on a model-derived CO_2 signal for the remaining Pleistocene (van der Wal *et al*. 2011), where the PI CO_2 is also shown (grey dotted line); c) Reconstructed global sea level, derived from ocean sediment core $\delta^{18}O$ data (Spratt and Lisiecki 2016) over the last 800 kyr and model-derived sea level (De Boer *et al*. 2010) for the remaining Quaternary, both of which are shown as an anomaly compared with PI (grey dotted line).

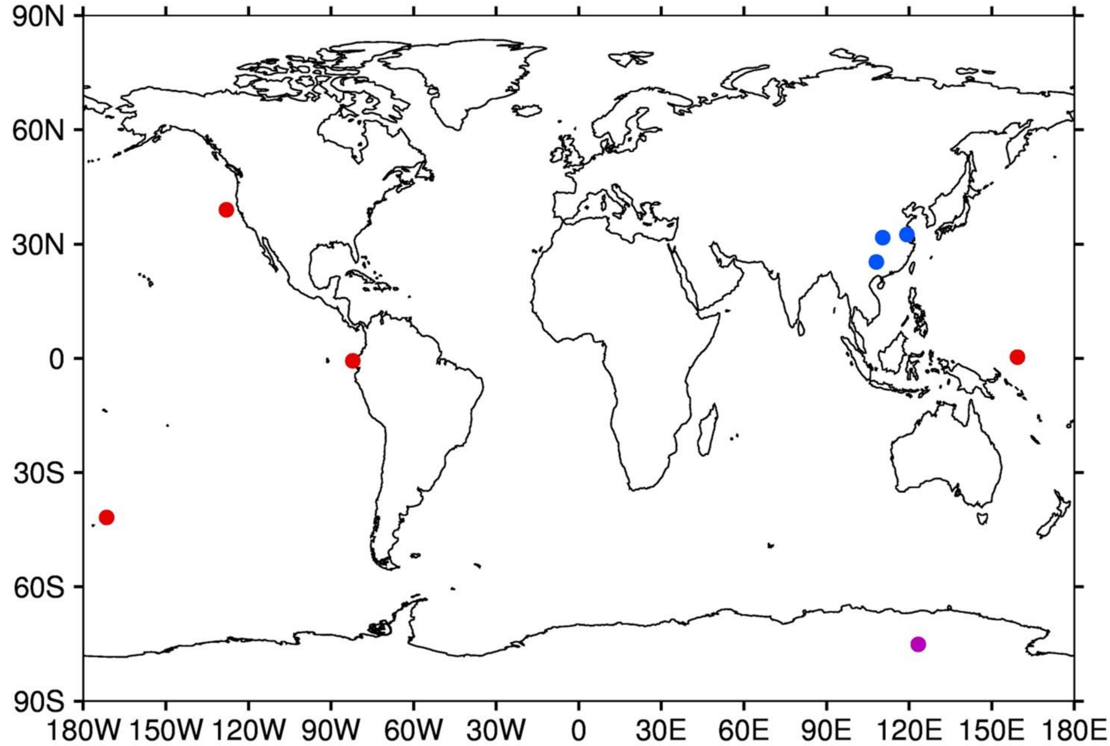

Figure S4. Map of the source locations of the proxy data records: red dots = SST, blue dots = precipitation, purple dot = SAT. See Table 1 for details.

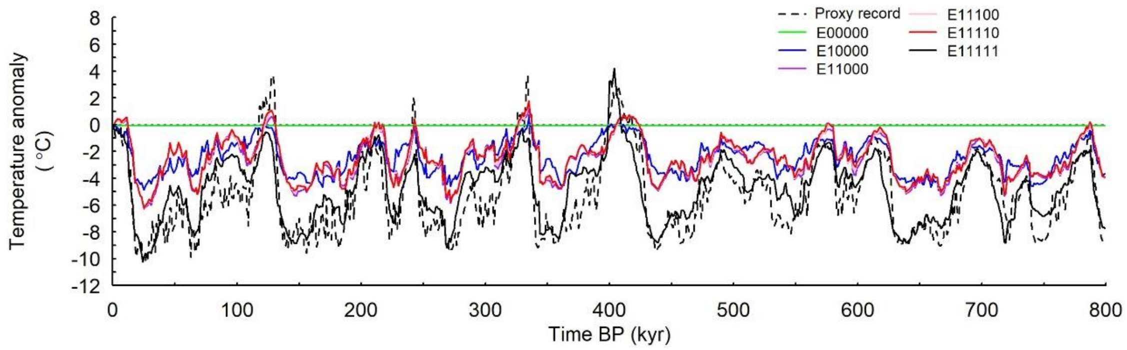

Figure S5. Timeseries of SAT anomaly (°C) for the last 800 kyr at Dome C, Antarctica, reconstructed from proxy data (dashed black lines) and modelled every 1 kyr using the emulator, which was forced by each simulation (solid lines) from the first pathway in the linear factorisation (in which each driving component is added incrementally, see 'Nomenclature and factorisation' in Methods).

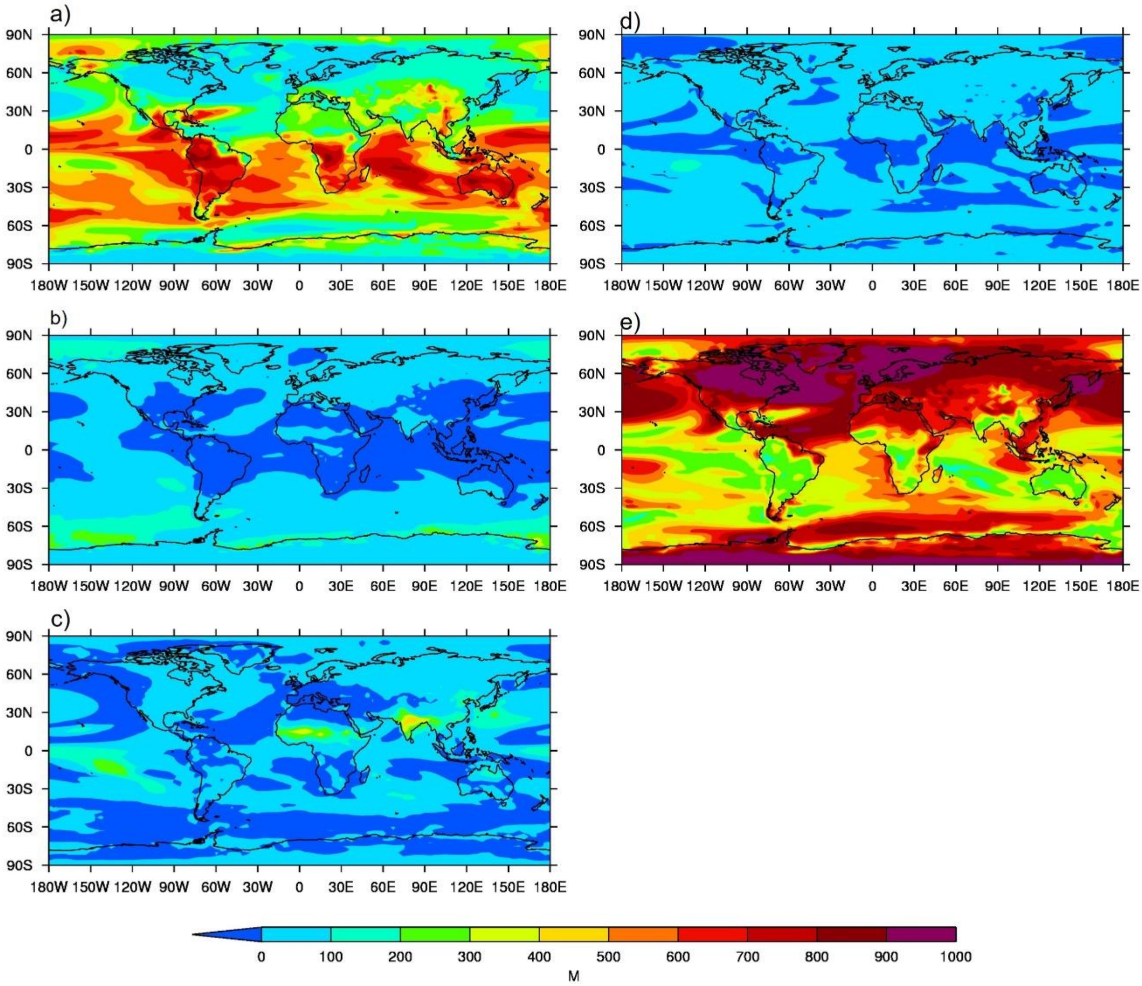

Figure S6. M scores between emulated temperature from the 'all drivers' simulation (E_{11111}) and emulated temperature from the first pathway in the linear factorisation, for each driving component: a) CO_2; b) Obliquity; c) Eccentricity; d) Precession; e) Ice.

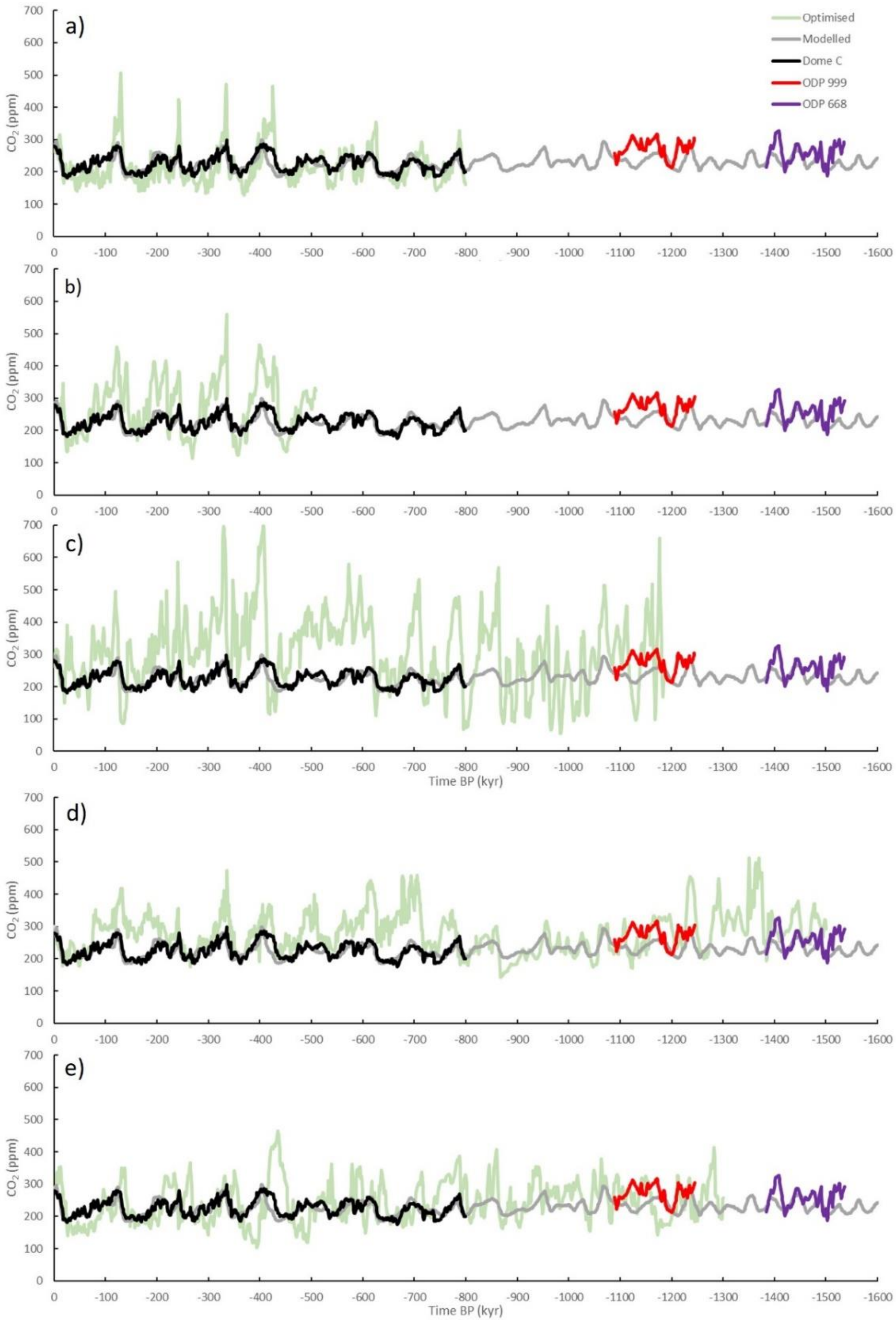

Figure S7. Timeseries of CO_2 (ppm) for the last 1.6 Myr, showing proxy CO_2 (black, red and purple lines), modelled CO_2 (grey line) and new CO_2 (green line) created by interpolating from proxy temperature reconstructions at five locations: a) SAT from the Dome C ice core, Antarctica; b) SST at ODP 1012, north-east Pacific; c) SST at ODP 1123, central South Pacific; d) SST at ODP 1239, Equatorial East Pacific; e) SST at ODP 806, Equatorial West Pacific.

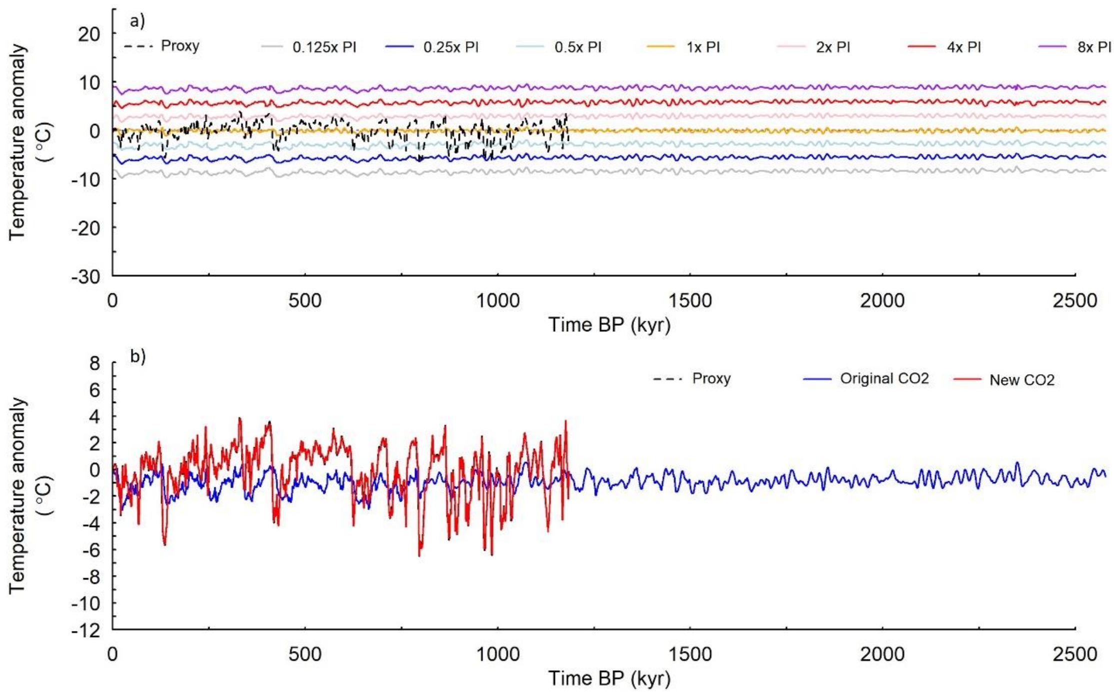

Figure S8. Examples from the CO_2 generation process: a) Timeseries of emulated (coloured lines) and reconstructed (dashed line) SAT anomalies (°C) for the last 2.58 Myr in the central South Pacific, using the constant CO_2 inputs from the idealised simulations; b) Timeseries of emulated (coloured lines) and reconstructed (dashed line) SAT anomalies (°C) for the last 2.58 Myr in the central South Pacific, showing the emulated SAT resulting from the original (i.e. reconstructed) CO_2 and the new CO_2. Note that here the proxy temperatures are mostly invisible, because the emulated temperatures (as a result of running with the optimised CO_2) are identical.