We thank the reviewers and editor for their useful comments, which have improved the manuscript. This document includes a response to all the Reviewer and Editor comments. This is then followed by a revised version of the manuscript in which all our proposed changes are clearly highlighted (including line numbers which are referenced by this document).

Reviewer I – Andrey Ganopolski

General comments

1. …I believe it would be useful for potential users of the methods it would be useful to present a more critical discussion of applicability of the methods and its potential limitations.

We agree, and have added a completely new section (Section 7; Conclusions are now Section 8) to clearly describe and discuss these limitations. A number of key limitations are also now included in the Abstract (lines 41-44).

Firstly, it should be stated very explicitly that the emulator is not applicable for simulations of transient climate change on time scales shorter than several millennia. However, on such long time scales, two major climate forcing – CO2 and ice sheets – strongly interact with each other, that cannot be accounted for in the method presented in the manuscript. On the page 11 the authors wrote that they "are able to simulate global climate development over long periods of time (several million years), provided that atmospheric CO2 level for the period is known, : : : ice sheets do not change outside the range considered... and the topography and land-see mask are unchanged". I believe the authors are too optimistic concerning "several million years" - even a much shorter time interval for which all these conditions are met would be difficult to find in the recent past or in the near future. Clearly, this method is not applicable to Quaternary.

We add a comment that the experiment design adopted here is not appropriate for simulations shorter than a few millennia, where complex models and transient simulations are most appropriate (lines 1006-1012). At the other extreme, "Several million years" has been changed to "several hundred thousand years or longer" (line 642) (see point below about the applicability of this emulator up to the next glacial inception, and new Section 7).

For Pliocene, CO2 concentration is not known sufficient accuracy. However, it is very likely that during the late Pliocene CO2 concentration experienced significant fluctuation at different time scales. It is also likely that during Pliocene, the extent of northern hemisphere ice sheets varied beyond the range used in this study (e.g. Willeit et al., 2015). I cannot see how all these problems can be circumvented without use of a comprehensive Earth system model. The less important but still not negligible problem is that according to the PRISM4 reconstruction, land-sea mask and orography during the late Pliocene in some regions (primarily North America and Europe) differed considerably from the modern ones.

Yes, we agree, and have made it clearer in the Pliocene section that (a) our approach is only appropriate for periods of the Pliocene with equivalent or less ice than modern, and (b) that we do not include palaeogeographic changes in our Pliocene simulations.

The situation is even more problematic for the future. It is not known how good is performance of existing carbon cycle models on such long time scales, but a reasonable agreement between results obtained with different models gives some hope. However, future simulations with the stand-alone carbon cycle models are only valid till the next glacial inception. For the medium emission scenarios, the next glacial inception is immanent (of course making a brave assumption that humans will not influence climate after the end of fossil fuel era) before or soon after 100,000 AD. Beyond that time, the methodology described in the manuscript is not applicable any more. For the extreme Business-as-usual type scenarios (5000 GtC and more), the situation is even worse. Under such scenarios, most of the Greenland ice sheet will melt completely already within the next 1000 years and most of the Antarctic ice sheet will also melt eventually (e.g. Winkelmann et al., 2016). And, according to recent study by DeConto and Pollard (2016), this "eventually" may occur already within one or two millennia. Such rate of ice sheet melt would strongly affect the ocean circulation and stratification with unknown but long-term consequences. In addition, 70 meter sea level rise resulting from melting of existing ice sheets would strongly affect global land-sea mask and regional climates. In addition, submersion of the large part of northern Europe would also have serious implications for the geological storage of nuclear wastes in this area. As the result, the conditions required for applicability of the proposed method can be violated already after the first few thousand years.

Thank you for raising these points. Whilst, as you say, we have mentioned some of the limitations at different points throughout the manuscript, it is better to have a section dedicated to describing the assumptions that the emulator is based on, its limitations and the conditions under which it may be applied. This includes the point made here about the applicability of the current emulator approach only up to the next glacial inception. We have added a new section (7; Conclusions are now Section 8) to clearly describe these limitations. We also explain that the emulator could be expanded to include glacial

states and therefore be applied to longer-term future (and the Quaternary), if the CO2 and ice volume were known (e.g. from a transient EMIC or conceptual model simulation).

Second, the emulator cannot be applied if the climate system possesses a strong nonlinearity. AMOC shutdown is the most natural example. The authors mentioned nonlinearity only once and assumed that "any non-linearities in the GCM response being absorbed by stochastic component of the Gaussian process" (p. 6). I am not sure I understand what this means. Please clarify.

Please see the response posted by Michel Crucifix in the Discussion.

2. The model reveals strong response in annual mean temperature on precessional forcing. Since annual mean precessional component of orbital forcing is zero, I wonder what causes such response. Is it really global or only regional phenomenon? May be it would be useful to add to the Fig. 4 annual SAT anomalies produced by other forcings: CO2, obliquity and precession (say difference between the maximum and minimum obliquity and difference between the "warm" and the "cold" orbits).

Figure 4 has been extended to show a larger range of forcings, as suggested. In addition to the SAT change due to reduced ice, plots are also now included showing the SAT change due to a doubling of CO_2, the difference between maximum and minimum obliquity and the difference between "warm" and "cold" orbital conditions. The following text has also been added to Section 3.5.1:

"Also shown in Fig. 4, for comparison, are mean annual SAT anomalies produced by the other forcings, including a doubling of CO_2, the difference between maximum and minimum obliquity and the difference between "warm" orbital conditions and "cold" orbital conditions. The warming caused by increased CO_2 is more widespread (Fig. 4b), with the largest warming occurring at high latitudes and for land regions, in agreement with typical future-climate simulations (IPCC, 2013, p. 1059). The temperature change due to obliquity and "warm" versus "cold" orbital conditions is less than that for either reduced ice (compared to pre-industrial) or increased CO_2. Changes in obliquity have the largest impact on temperatures in high latitude regions, since the exposure of these regions to the sun's radiation is most affected by changes in obliquity. Smaller temperature anomalies are observed over northern Africa and India and, since an increase in obliquity is indeed known to boost monsoon dynamics (e.g. Araya-Melo et al., 2015; Bosmans et al., 2015), changes in soil latent heat exchanges are therefore expected to contribute negatively to the temperature response. The comparison of "warm" versus "cold" orbital conditions, which highlights (annual mean) temperature changes primarily caused by precession, generally shows a warming trend, with the largest temperature changes occurring in monsoonal regions. Lower temperatures are observed in the Northern Hemisphere over northern Africa, India and, East Asia, whilst warmer temperatures occur in the Southern Hemisphere over South America, southern Africa and Australia. Figure 4 demonstrates that the temperature forcing caused by CO_2 affects mean annual temperatures on a global scale, whilst the forcing due to ice sheet and orbital changes affects mean annual temperatures in specific regions, having a limited impact on global mean temperatures. This is supported by the relatively high global mean SAT anomaly for the $2xCO_2$ scenario of 4.2°C, compared with the lower SAT anomalies that result from the obliquity and precession forcing of 0.4°C each (see caption of Figure 4)."

3. I found the attempt to reconstruct Pliocene CO2 from individual temperature records rather strange. These four temperature records are so poorly correlated with each other that it is hard to expect that any global factor (like CO2) can bring them in agreement with modelling results. As the result, all four CO2 "reconstructions" have very little in common. I wonder what one can learn from such exercise. Although I cannot be objective in this respect, but I do believe that using of stacked data (e.g. Willeit et al. (2015), Stap et al. (2016)) rather than individual records, is more appropriate approach to reconstruct past CO2 concentrations.

Whilst we agree that the use of stacked benthic oxygen isotope data is appropriate in many circumstances, we do not believe that it would be suitable in this instance. This is because the GCM experiments that the emulator is calibrated on were relatively short (500 years), and hence benthic ocean temperatures would not yet have spun-up fully and reached steady state, particularly in the high CO_2 experiments. Therefore, it would not be appropriate to compare deep ocean temperatures from the proxy data to the modelled temperatures. We have added some text to the paper to explain this (lines 766-770).

We also believe that there are benefits to using individual temperature records. For example, if the model and emulator are correct, then the analysis shows that the temperature records are not consistent with each other, which may not be obvious by just comparing the records visually. We have added the following paragraph to Section 5.3 to describe the possible sources of the inconsistencies between the reconstructions at different sites, and what we think are most likely to be the cause:

"There is substantial variation between our CO_2 estimates at different sites, and this may be attributed to a number of causes. It could be that there are errors in the GCM model used, in particular in its representation of the response of climate to CO_2 and/or orbital forcing. There could be inaccuracies associated with the SST data at one or more locations as, if the model was assumed to be correct, the estimated CO_2 should be similar across the four locations. The fact that they are not may indicate that the temperature records are not consistent with each other, which may not have been obvious by just comparing the records visually. This is one of the potential advantages to using individual temperature records rather than stacked records. It may also be that there is an issue with the dating of some of the proxy records; the data may be correct but there may be uncertainties/inaccuracies in the age models. Alternatively, the emulator may be wrong; for example, there may be non-linearities in the climate response simulated by the GCM that it is not capturing. Finally, there may be errors related to the modelled representation of the ice sheets, which are fixed at a constant configuration. In reality, of the possible sources of error that have been identified, the variations are less likely to be the result of errors in the emulator's estimates of the GCM output because validation diagnostics did not seem to suggest systematic failures. They are also less likely to be due to unrepresented changes in climate due to the ice sheets. Whilst some of the variation at the high latitude sites (982 and U1313) may be attributed to some regional climate processes not fully accounted for, e.g. involving the ice sheets and sea ice, two of the sites (722 and 662) are in tropical regions. Thus, SSTs at these sites would not be expected to be affected by changes in the ice sheets, and yet they show significantly different variations. Therefore, the inconsistencies are likely to be due to a combination of errors in the GCM model and inaccuracies in the SST data."

4. I would strongly suggest to not use expressions like "fossil fuel emission" or "anthropogenic fossil fuel emission". Unfortunately, this jargon is used in some publications related to energy and mitigation. However, I do not believe it is appropriate for climate modelling papers. In any case, burning of fossil fuel is the most important but not the only source of anthropogenic CO2. Land use and cement production also play a role in rising of atmospheric CO2 concentration.

This is a good point. These instances have been changed to "anthropogenic CO2 emissions" or similar throughout.

L. 54 Which "system" is meant here?

Inserted "climate" (line 60).

L. 74 Typo. "precessional"

Done.

L. 110 I would change "modern day" to "Quaternary"

Done.

L. 128 "input configuration"?

Inserted "(i.e. any set of orbital and CO_2 conditions)" (line 136).

L. 250 change "forcings" to "parameters"

Done.

L. 256 What about obliquity?

We had missed it out of the sentence, so thank you for pointing it out. It has now been added.

L. 265. This is not estimate of "remaining reserves". This is just "current estimate" of fossil fuel reserves which has a tendency to increase with time.

Removed "remaining".

L. 277 I do not believe that 20 ppm CO2 change during Holocene (which is primarily transient response to the deglaciation) has something to do with the natural CO2 variability during Anthropocene.

The sentence has been reworded to make it clear that we present variations during the Holocene as an example of natural variations, rather than the change that we expect to occur in the future (lines 289-291).

L. 209 Emission cannot be removed

Replaced with "taken up".

L. 298 CO2 will not return to preindustrial level because glacial cycles will resume before this will happen. But even without glacial cycles, it is unlikely that preindustrial level of 280 ppm is the true equilibrium CO2 concentration in the interglacial world.

Even small disbalance between volcanic outgassing and weathering would cause significant CO2 drift on time scale order of 100,000 years.

This sentence has been modified to make these assumptions clear (lines 310-314).

L. 353 Please specify initial conditions for model runs.

Inserted "All experiments were initiated from a pre-industrial spin-up experiment, with an atmospheric CO_2 concentration of 280 ppmv, and pre-industrial ice sheet extents and orbital conditions." (lines 370-371).

L. 367 Which positive feedback is meant here? I guess this is just an artefact of models with prescribed present day vertical ozone profile.

This sentence has been reworded and extended (lines 384-389):

"This is the result of a runaway positive feedback in the GCM caused, at least in part, by the vertical distribution of ozone in the model being prescribed for modern-day climate conditions. Consequently, the ozone distribution is not able to respond to changes in climate, meaning that when increased mean global temperatures result in an increase in altitude of the tropopause and hence an extension of the troposphere, relatively high concentrations of ozone, which were previously located in the stratosphere, enter the troposphere, resulting in runaway warming."

L. 502 Why "linear nature of the plot increases" confidence? In theory, this plot must not be necessarily linear.

This sentence has been removed.

L. 585 What is "SAT index"

It is the globally averaged mean annual SAT for each experiment, but it has not been adjusted for grid box area, therefore we refer to it as a "SAT index". The caption for Figure 8 has been amended to clarify this.

L. 751 "Across the four sites…" This sentence is not clear

Sentence has been reworded (lines 844-846).

L. 758 What is the meaning of "emulated uncertainty" and how it was defined?

Inserted "(defined as 1 standard deviation of the emulated grid box posterior variance)" (line 712 line 844).

L. 763 What is meant under other "human activities"?

Sentence has been reworded to include combustion of fossil fuels, land-use change and cement production (lines 856-858.

L. 776 "long atmospheric lifetime of fossil fuel emission"?

Sentence has been reworded to "CO_2 emissions" (lines 871-873).

L. 776 Reference to the original Archer (2005) paper would be much more appropriate

This reference has been added (line 871).

L. 813 -820. The authors try to argue here that the fact that they cannot model ice sheet evolution is not very important for the future 200,000 years climate projections. This is not true – see my general comments.

The following sentence has been added to clarify that on these timescales the inability to model ice sheet evolution may be an issue (lines 916-918):

"As will be discussed in Sect. 7, however, the emulator was not designed and calibrated to predict changes in ice sheets. This is a limitation that should be addressed when modelling future climate on timescales of tens of thousands of years or more (depending on the CO_2 scenario(s) being modelled)."

L. 899 "High latitude sites concentrations" Sounds like CO2 concentration is different in different sites

This sentence has been reworded (lines 1088-1090):

"Our CO_2 concentrations derived from tropical ODP/IODP sites show relatively similar concentrations to CO_2 proxy records for the same period, although the concentrations derived from higher latitude sites are generally significantly higher than the proxy data."

Fig. 9. I guess Fig9a shows annual SAT difference due to CO2 increase to 400 ppm. If so "modern annual SST" is misleading. What is shown in 9b is not clear to me.

Yes, it is correct that Fig. 9a shows the annual SAT anomaly due to CO2 being increased to 400 ppm. The caption states that this is "mean annual SAT for modern-day orbital conditions", not "modern annual SST", so we think that this is clear. We agree that Fig. 9b was not really adding anything and have therefore removed it, and amended the main text accordingly.

Reviewer II

1 – My main concern is about the limitations of the emulation strategy. They are not sufficiently stressed in the manuscript. Indeed, the authors have performed a very good job in developing and implementing the emulator technique, and the manuscript explains in details the methodology. To some extent, this is "the best that can be done" based on GCM tools. But, obviously this is also probably not entirely sufficient… Over all, the fundamental hypothesis is that "climate" responds very smoothly (as explained in the paper) to external forcing. This also makes the even stronger assumption that long-term components of the Earth system, in particular the deep ocean, the carbon cycle and ice-sheets, have no dynamic role. Though this is indeed a fairly usual assumption when studying century-scale changes, this is unlikely to be adequate for 100-kyr to million-year studies. I think the authors should clearly state that their strategy cannot account for : (for instance) deep ocean changes (as experienced during the Quaternary during cold and but also warm periods), CO2 dynamics, ice sheet dynamics. The authors make the hypothesis that it might be suitable for warmer climates (thus the Pliocene and the future) while it is clearly inadequate for the Pleistocene. This might be true, but it is also likely a perspective problem: we know quite well that the Pleistocene climate results from complex interactions between ice-sheets, deep ocean and CO2; with much fewer data, we may (or may not) assume that the Pliocene is simpler…

Thank you for these helpful suggestions. Please see the response to comment (1) of Reviewer I (André Ganopolski). In particular, this new section includes a discussion of the fact that we do not carry out truly transient simulations, but a series of snapshots, and as such our methodology is inappropriate for examining deep ocean trends, and becomes compromised if deep ocean transient changes are important for controlling surface climate evolution.

2 – On Pliocene results. In line with the above comment, the hypothesis of rather small ice-sheet changes in the late Pliocene is not very well founded. The authors mention that their chosen time window does not include the M2 glaciation at 3300 kyr BP (line 614). This is not quite correct since they investigate the 3300-2800 kyr BP time window, which starts precisely with the M2 glaciation, as can be clearly seen on the data of Fig.10. The M2 glaciation is estimated to correspond to a sea-level fall between 40 and 65 m (Miller et al. 2012; Dwyer & Chandler, 2009). The following cold events (KM2 at G20) are not so well characterized, but should correspond to roughly half the size of M2 (20 to 40 m of sea level drop). On the other side, the G17, K1 or KM3 time periods experienced significantly reductions in ice volume with sea level rise estimated to be +25±10 m (Miller et al. 2012). Overall, ice-sheet changes are certainly much larger than assumed in the manuscript, and not bounded by the lowice/modice configurations.

The following text has been added (lines 665-669). Please also see the response to comment (1) of Reviewer I.

"represents the warm phase of climate (interglacial conditions), and does not include major glaciations (though the M2 cooling event may persist to the very start of the simulation at 3300 kyr BP, and the simulated period does include periods of likely glaciation, such as KM2 (~3100 kyr BP) and G20 (~3000 kyr BP)). The emulator would not be appropriate to periods of extensive glaciation and may not be well-matched to the periods of lesser glaciation included within the simulated interval."

3 – The corresponding calculation of pCO2 (§6.3) probably illustrates the failure of these assumptions. In any case, the four "reconstructions" shown on Fig.12 have little in common, which certainly deserves some comments. The much higher variability seen in high-latitude data points to "polar" climatic processes not being accounted for by the emulator (like ice-sheets, incorrect sea-ice, …). Instead of presenting these curves as possible pCO2 reconstructions (something difficult to buy), I would rather use them to discuss the limitation of the overall strategy: if the model were perfect, the four curves should be identical… Most probably, the model-data strategy is furthermore inadequate: For instance, is it reasonable to use annual mean SAT to be compared with alkenone-based SST reconstructions?

Please see the response to comment (3) of Reviewer I, and new paragraph at the end of Section 5.3.

4 – On the future 200 ka results. I also have problems with the rather "conservative" assumption of small ice sheet changes. According to Pollard & DeConto (2016), the disappearance of WAIS (somewhat equivalent to lowice?) correspond to the rather mild RCP4.5 scenario, while an extended RCP8.5 results in more than 20 m of sea level rise for Antarctica alone. These ice-sheet changes might also impact the deep ocean circulation, something difficult to account with the emulator strategy.

These limitations have been discussed in a new section (7) describing the limitations of the methodology.

5 – Lines 808 + following are discussing the limitations of the overall strategy for the next glacial inception, since there is no ice-sheet model component. I would also add that the carbon cycle is prescribed here, not interactive. In other words, the long-term smooth decrease of CO2 is based on the assumption that nothing unexpected will happen in the Earth carbon cycle, and that the "silicate weathering" mechanism (or hypothesis) is a robust one, something far from being fully understood.

The following text has been added to this paragraph to highlight these assumptions (lines 919-924):

"Another caveat is that the carbon cycle in the emulator is also essentially prescribed, and thus not interactive. This means that the atmospheric CO_2 trajectory follows a smooth decline, as was projected using an impulse response function based on experiments using the *c*GENIE model (Lord et al., 2016), with long-term future climate being modelled as a series of snapshot simulations with the emulator. This smooth decline in CO_2 assumes that no non-linear or unexpected behaviour will be demonstrated by the long-term carbon cycle, and that the silicate weathering mechanism, which is associated with a substantial degree of uncertainty, is correct."

6 – On the experimental design, it could be useful to explain why the ice-sheet size (lowice/modice) has not been included in the emulation procedure.

The following text was added to the "Calibration and evaluation of the emulator" section (lines 563-566):

"This approach was adopted, rather than including the ice sheet extent as an active input parameter to the emulator, because only two ice sheet configurations have been simulated, which are not sufficient for an interpolation. One of the main benefits of including ice sheet extent as an active input parameter would be to emulate changing ice sheets over time, but this was beyond the scope of this study."

7 – The simulation of sea ice at high latitudes under high CO2 might be a problem, as explained in the text (lines 575-580). It could be useful to discuss rapidly how HadCM3 compares to other GCMs in terms of sea ice.

The section highlighted explains that the PCA, and therefore the emulator, may not be fully capturing high latitude variations, meaning that in the leave-one-out analysis some of the high CO_2 simulations include larger errors in these regions compared to the equivalent GCM simulation. It is true that there may also be underlying errors in the AOGCM representation of sea ice. These are discussed in Valdes et al (2017).

8 – Line 871. The comparison of model results with paleodata, or the projection of future impacts, is not so much a question of resolution. 1 - The GCM resolution is often not sufficient. 2 - Very often, this requires additional modelling (proxy modelling, impact models, …)

The following sentence has been added to state this (lines 1060-1062):

"However, further downscaling of the data may also be necessary or beneficial, via further modelling such as proxy modelling, impact models or regional climate models, or via statistical downscaling techniques."

9 – Fig.2: Simulations over 2000 ppm have been discarded (§3.4.1): the corresponding points should either be removed, or should be plotted with a different colour. These plots are not "slices" but "projections".

Fig. 2a has been modified as suggested (colour changed). Replaced with "projections".

10 – Fig.10: the comparison to data is poor. I believe just computing a correlation coefficient and/or explained variance ratio could be useful. See above comments on discussing the overall limitations.

Correlation coefficients have been computed and some text to describe the results has been added to Section 5.1 (lines 694-702).

In addition, we have made a small number of minor changes:

- **The affiliation of Charlotte O'Brien has been corrected**
- **The CO_2 reconstructions have been redone using a wider range of constant CO_2 scenarios (260, 300, 400, 500, 600, 700, and 800 ppmv) for the linear regression. Figure 12 has been updated with the new data.**
- **Figure 7 – CO_2 concentration has been added to the upper y axis**
- **Minor clarifications and rewordings throughout to improve clarity**
- **Ka/Ma has been changed to Myr/kyr where appropriate**

Emulation of long-term changes in global climate: Application to the late Pliocene and future

Natalie S. Lord[1,2], Michel Crucifix[3,4], Dan J. Lunt[1,2], Mike C. Thorne[5], Nabila Bounceur[3], Harry Dowsett[6], Charlotte L. O'Brien~~[6,]~~[7] and Andy Ridgwell[1,2,8]

[1]School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK.
[2]Cabot Institute, University of Bristol, Bristol, UK.
[3]Université catholique de Louvain, Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, 1348 Louvain-la-Neuve, Belgium.
[4]Belgian National Fund of Scientific Research, Brussels, Belgium.
[5]Mike Thorne and Associates Limited, Quarry Cottage, Hamsterley, Bishop Auckland, Co. Durham, DL13 3NJ, UK.
[6]Eastern Geology and Paleoclimate Science Center, U. S. Geological Survey, Reston, VA 20192, USA.
[7]Department of Geology and Geophysics, Yale University, New Haven, CT 06511, USA.
[8]Department of Earth Sciences, University of California, Riverside, CA 92521, USA.

Correspondence to: Natalie S. Lord (Natalie.Lord@bristol.ac.uk)

Abstract

Multi-millennial transient simulations of climate changes have a range of important applications, such as for investigating key geologic events and transitions for which high resolution palaeoenvironmental proxy data are available, or for projecting the long-term impacts of future climate evolution on the performance of geological repositories for the disposal of radioactive wastes. However, due to the high computational requirements of current fully coupled General Circulation Models (GCMs), long-term simulations can generally only be performed with less complex models and/or at lower spatial resolution. In this study, we present novel long-term "continuous" projections of climate evolution based on the output from GCMs, via the use of a statistical emulator. The emulator is calibrated using ensembles of GCM simulations which have varying orbital configurations and atmospheric CO_2 concentrations and enables a variety of investigations of long-term climate change to be conducted which would not be possible with other modelling techniques at the same temporal and spatial scales. To illustrate the potential applications, we apply the emulator to the late Pliocene (by modelling surface air temperature (SAT)), comparing its results with palaeo-proxy data for a number of global sites, and to the next 200 thousand years (kyr) (by modelling SAT and precipitation). A range of CO_2 scenarios are prescribed~~modelled~~ for each period. During the late Pliocene, we find that emulated SAT varies on an approximately precessional timescale, with evidence of increased obliquity response at times. A comparison of atmospheric CO_2 concentration for this period, estimated using the proxy sea surface temperature (~~SST~~SST) data from different sites and emulator results~~ and using proxy CO_2 records~~, finds that relatively similar CO_2 concentrations are estimated based on sites ~~produced ~~at lower latitudes, whereas~~although~~ higher latitude sites show larger discrepancies. In our second illustrative application, spanning the next 200 kyr into the future, we find that SAT oscillations appear to be primarily influenced by obliquity for the first ~120 kyr, whilst eccentricity is relatively low, after which precession plays a more dominant role. Conversely, variations in precipitation over the entire period demonstrate a strong precessional signal. Overall, we find that the emulator provides a useful and powerful tool for rapidly simulating the long-term evolution of climate, both past and future, due to its relatively high spatial resolution and relatively low computational cost. However, there are uncertainties associated with the approach used, including the inability of the emulator to ~~represent~~capture deviations from a quasi-stationary response to the forcing, such as ~~true ~~transient ~~changes~~adjustments ~~in the climate system, such as those associated with the~~of the deep ocean temperature and circulation, in addition to its limited range of fixed ice sheet configurations and its requirement for prescribed atmospheric CO_2 concentrations.

1 Introduction

Palaeoclimate natural archives reveal how the Earth's past climate has fluctuated between warmer and cooler intervals. Glacial periods, such as the Last Glacial Maximum (e.g. Lambeck et al., 2001; Yokoyama et al., 2000), exhibit relatively lower temperatures associated with extensive ice sheets at high northern latitudes (Herbert et al., 2010; Jouzel et al., 2007; Lisiecki and Raymo, 2005), whilst interglacials are characterized by much milder temperatures in global mean. Even warmer and sometimes transient ("hyperthermal") intervals, such as occurred during the Palaeocene-Eocene Thermal Maximum (e.g. Kennett and Stott, 1991), are~~occur~~ characterized by even higher global mean temperatures. Assuming that on glacial-interglacial timescales and across transient warmings and climatic transitions, tectonic effects can be neglected, the timing and rate of climatic change is at least partly controlled by the three main orbital parameters – precession, obliquity and eccentricity – which have cycle durations of approximately 23, 41, and both 96 and ~400 thousand years (kyr), respectively (Berger, 1978; Hays et al., 1976; Kawamura et al., 2007; Lisiecki and Raymo, 2007; Milankovitch, 1941). Further key drivers of past climate dynamics include changes in atmospheric CO_2 concentration and in respect of the glacial-interglacial cycles, changes in the extent and thickness of ice sheets.

In order to investigate the dynamics, impacts and feedbacks associated with the response of the climate system to orbital forcing and CO_2, long-term (>10^3 years (yr)) projections of changing climate are required. Transient simulations such as these are useful for investigating key past episodes of extended duration for which detailed palaeoenvironmental proxy data are available, such as through the Quaternary and Pliocene, allowing data-model comparisons. Simulations of long-term future climate change also have a number of applications, such as in assessments of the safety of geological disposal of radioactive wastes. Due to the long half-lives of potentially harmful radionuclides in these wastes, geological disposal facilities must remain functional for up to 100 kyr in the case of low- and intermediate-level wastes (e.g. Low Level Waste Repository, UK (LLWR, 2011)), and up to 1 ~~Ma~~ Myr in the case of high-level wastes and spent nuclear fuel (e.g. proposed KBS-3 facility, Sweden (SKB, 2011)). Projections of possible long-term future climate evolution are therefore required in order for the impact of potential climatic changes on the performance and safety of a repository to be assessed (NDA, 2010; Texier et al., 2003). Indeed, while the glacial-interglacial cycles are expected to continue into the future, the timing of onset of the next glacial episode is currently uncertain and will be fundamentally impacted by the increased radiative forcing from anthropogenic CO_2 emissions (Archer and Ganopolski, 2005; Ganopolski et al., 2016; Loutre and Berger, 2000b).

Making spatially-resolved past or future projections of changes in surface climate generally involves the use of fully coupled General Circulation Models (GCMs). However, a consequence of their high spatial and temporal resolution and structural complexity (and attendant computational resources) is that it is not usually practical to run them for simulations of more than a few millennia, and invariably, rather less than a single pre~~o~~cessional cycle. Even when run for several thousand years, only a limited number of runs can be performed. Previously, therefore, lower complexity models such as Earth system Models of Intermediate Complexity (EMICs) have been used to simulate long-term transient past (e.g. Loutre and Berger, 2000a; Stap et al., 2014) and future (e.g. Archer and Ganopolski, 2005; Eby et al., 2009; Ganopolski et al., 2016; Lenton et al., 2006; Loutre and Berger, 2000b) climate development. Where GCMs have been employed, generally only a relatively small number of snapshot simulations of particular climate states or time slices of interest have been modelled (Braconnot et al., 2007; Haywood et al., 2013; Marzocchi et al., 2015; Masson-Delmotte et al., 2011; Prescott et al., 2014).

In this study, we present long-term continuous projections of climate evolution based on the output from a GCM, via the use of a statistical emulator. Emulators have been utilised in previous studies for a range of applications, including sensitivity analyses of climate to orbital, atmospheric CO_2 and ice sheet configurations (Araya-Melo et al., 2015; Bounceur et al., 2015) and model parameterizations (Holden et al., 2010). However, to the best of our knowledge, this is the first time that an emulator has been trained on data from a GCM and then used to simulate long-term future transient climate change. It should be noted that, whilst other research communities may use different terms, we refer to the groups of climate model experiments as “ensembles”, and we refer directly to the GCM when discussing calibration of the emulator, rather than using the term “simulator” as has been used in a number of previous studies.

We calibrated an emulator using SAT data produced using the HadCM3 GCM (Gordon et al., 2000). Two ensembles of simulations were run, with varying orbital configurations and atmospheric CO_2 concentrations. Each ensemble was run twice, once with modern-day continental ice sheets and once (for a reduced number of members) with reduced-extent ice sheets. We adopted this approach because in at least two of the intended uses for the emulator (Pliocene, and long-term future climate for application to performance assessments for potential radioactive waste repositories), it is thought that the Greenland and West Antarctic ice sheets (GIS, WAIS) could be reduced relative to their current size. The implications and uncertainties associated with this approach are discussed in Sect. 7. The ensembles thus cover a range of possible future conditions, including the high atmospheric CO_2 concentrations expected in the near-term due to anthropogenic CO_2 ~~fossil fuel~~ emissions, and the gradual reduction of this CO_2 perturbation over timescales of hundreds of thousands of years by the long-term carbon cycle (Lord et al., 2015, 2016).

We go on to illustrate a number of different ways in which the emulator can be applied to investigate long-term climate evolution o~~f~~ver hundreds of thousands to millions of years. Firstly, the emulator is used to simulate SAT changes for the late Pliocene for the period 3300-2800 kyr before present (BP) for a range of CO_2 concentrations. This interval occurs in the middle part of the Piacenzian Age, and was previously referred to as the “mid-Pliocene” (e.g. Dowsett and Robinson, 2009). During this time, global temperatures were warmer than pre-industrial (e.g. Dowsett et al., 2011; Haywood and Valdes, 2004; Lunt et al., 2010), before the transition to the intensified glacial-interglacial cycles that are associated with ~~modern-day~~Quaternary climate (Lisiecki and Raymo, 2007). We then apply the emulator to future climate, simulating temperature and precipitation data for the next 200 kyr (AP – after present) for a range of ~~fossil fuel~~anthropogenic CO_2 emissions scenarios. Regional changes in climate at a number of European sites (grid boxes) are presented, selected either because they have been identified as adopted or proposed locations for the geological disposal of solid radioactive wastes, as in the cases of Forsmark, Sweden and El Cabril, Spain, or simply as reference locations where a suitable site has not yet been identified, as in the cases of Switzerland and the UK.

The paper is structured such that the theoretical basis of the emulator is described in Sect. 2, the GCM model description and simulations are presented in Sect. 3 and an account of how the emulator is trained and evaluated is given in Sect. 4. Section 5 presents illustrative examples of a number of potential applications of the emulator for the late Pliocene. Further examples of the application of the emulator to the next 200 kyr are described in Sect. 6. Section 7 includes a description and discussion of uncertainties associated with the methodology and tools, ~~,~~ and the conclusions of this study are presented in Sect. 8~~7~~.

2 Theoretical basis of the emulator

The emulator is a statistical representation of a more complex model, in this case a GCM. It works on the principle that a relatively small number of experiments are carried out using the GCM, which fill the entire multidimensional input space (in our case, four dimensions consisting of three orbital dimensions and a CO_2 dimension), albeit rather sparsely. The statistical model is calibrated on these experiments, with the aim of being able to interpolate the GCM results such that it can provide a prediction of the output that the GCM would produce if it were run using any particular input configuration (i.e. any set of orbital and CO_2 conditions). If successful (as can be tested by comparing emulator results with additional GCM results not included in the calibration), no further experiments are required using the GCM; the emulator can then be used to produce results for any set of conditions or sequence of sets of conditions within the range of conditions on which it has been calibrated. It should not~~cannot~~, of course, be used to extrapolate to conditions outside that range.

In this study, we use a principal component analysis (PCA) Gaussian Process (GP) emulator based on Sacks et al. (1989), with the subsequent Bayesian treatment of Kennedy and O'Hagan (2000) and Oakley and O'Hagan (2002), and ~~associated with~~ a principal component analysis approach associated with ~~by~~ Wilkinson (2010). All code for the GP package is available online at https://github.com/mcrucifix/GP. This principal component (PC) emulator is based on climate data for the entire global grid, as opposed to calibrating separate emulators based on data for individual grid boxes. This approach is taken because, for past climate, the global response overall is of interest, rather than just the response at specific locations individually. It also means that the results are consistent across all locations. For future climate, and in particular for application to nuclear waste, recommendations and results should be consistent across all sites, which would be especially relevant to a large country such as the US. Alternatively, for some countries and locations, it may be more appropriate to emulate specific grid boxes. The theoretical basis for the emulator and its calibration, is as follows.

Let $\boldsymbol{D}$ represent the design matrix of input data with n rows, where n is the total number of experiments performed with the GCM, here 60 (sum of the two ensembles). The number of columns, p, is defined by the number of dimensions in input parameter space. In this case, p = 4 representing the three orbital parameters and atmospheric CO_2 concentration. A more detailed explanation of the orbital input parameters is included in Sect. 3; however, briefly, they are longitude of perihelion (ϖ), obliquity (ε) and eccentricity (e), with longitude of perihelion and eccentricity being combined under the form $e\sin\varpi$ and $e\cos\varpi$. For a set of i=1,~~–~~n simulations, each simulation represents a point in input space, and is characterised by the input vector $\boldsymbol{x}_i$, i.e. a row of $\boldsymbol{D}$.

The corresponding GCM climate data output is denoted $f(\boldsymbol{x}_i)$, where the function f represents the GCM model. This output for all n experiments is contained in the matrix $\boldsymbol{Y}$. The raw output from the GCM is in the form of gridded data covering the Earth's surface, with 96 longitude by 73 latitude grid boxes. We perform a principal component analysis, to reduce the dimension of the output data before it is used to calibrate the emulator. Each column of $\boldsymbol{Y}$ contains the results for one experiment, i.e. $\boldsymbol{Y} = [y(x_1), \ldots, y(x_n)]$. Furthermore, the centred matrix $\boldsymbol{Y}^*$ can be defined as $\boldsymbol{Y} - \boldsymbol{Y}_{mean}$, where $\boldsymbol{Y}_{mean}$ is a matrix in which each row comprises a set of identical elements that are the row averages of $\boldsymbol{Y}$. The singular value decomposition (SVD) of $\boldsymbol{Y}^*$ is:

$$\boldsymbol{Y}^* = \boldsymbol{USV}^{T*}, \tag{1}$$

where $\boldsymbol{S}$ is the diagonal matrix containing the corresponding eigenvalues of $\boldsymbol{V}$, $\boldsymbol{V}$ is a matrix of the right singular vectors of $\boldsymbol{Y}$, and $\boldsymbol{U}$ is a matrix of the left singular vectors. $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal, and $V^{\mathrm{T}*}$ denotes the conjugate transpose of the unitary matrix $\boldsymbol{V}$. The columns of $\boldsymbol{US}$ represent the principal components, and the columns of $\boldsymbol{V}$ the principal directions/axes. Each column of $\boldsymbol{U}$ represents an eigenvector, $\boldsymbol{u}_k$, and $\boldsymbol{VS}$ provides the projection coefficients $\boldsymbol{\beta}_k$. Specifically, for experiment i, $a_k(\boldsymbol{x}_i) = \sum_k \boldsymbol{V}_{ik}\boldsymbol{S}_{kk}$ gives the projection coefficient for the kth eigenvector. The eigenvectors are ordered by decreasing eigenvalue, and in practice only a relatively small number of the eigenvectors will be retained (n'), typically selected on the basis of the largest values of $a_k(\boldsymbol{x})$. Thus:

$$y(\boldsymbol{x}) = \sum_{k=1}^{n'} a_k(\boldsymbol{x})\boldsymbol{u}_k, \tag{2}$$

We calibrate the emulator using the reduced dimension output data rather than the raw spatial climate data. However, for simplicity, we will first consider a simple GP emulator. For this, the model output $f(\boldsymbol{x})$ for the input conditions $\boldsymbol{x}$ is modelled as a stochastic quantity that is defined by a Gaussian process. Its distribution is fully specified by its mean function, $m(\boldsymbol{x})$, and its covariance function, $V(\boldsymbol{x}, \boldsymbol{x}')$, which may be written:

$$f(\boldsymbol{x}) = GP[m(\boldsymbol{x}), V(\boldsymbol{x}, \boldsymbol{x}')], \tag{3}$$

The mean and covariance functions take the form:

$$m(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta}, \tag{4}$$

$$V(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2[c(\boldsymbol{x}, \boldsymbol{x}')], \tag{5}$$

where $\boldsymbol{h}(\boldsymbol{x})$ is a vector of known regression functions of the inputs, $\boldsymbol{\beta}$ is a column vector of regression coefficients corresponding to the mean function, $c(\boldsymbol{x}, \boldsymbol{x}')$ is the GP correlation function and σ^2 is a scaling value for the covariance function. $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{\beta}$ both have q components and, as before, $^{\mathrm{T}}$ denotes the transpose operation.

A range of options are available for the regression functions $\boldsymbol{h}(\boldsymbol{x})$ and the GP correlation function c, the most suitable of which depends on the application of the emulator. Any existing knowledge that the user may have about the expected response of the GCM to the input parameters can be used to inform their function choices. However, if the emulator performs poorly, an alternative function can be selected which may prove to be more suitable.

We assume a linear model, $\boldsymbol{h}(\boldsymbol{x})^T = (1, \boldsymbol{x}^T)$, with any non-linearities in the GCM response being absorbed by the stochastic component of the GP. The correlation function is exponential decay with a nugget, a detailed discussion of which can be found in Andrianakis and Challenor (2012). Hence, for the input parameters *a*=1, *p*, the correlation function can be written as:

$$c(\boldsymbol{x}, \boldsymbol{x}') = exp\left[-\sum_{a=1}^{p}\left\{\frac{(x_a - x'_a)}{\delta_a}\right\}^2\right] + \nu I_{x=x'}, \tag{6}$$

where δ is the correlation length hyperparameter for each input, *v* is the nugget term, and *I* is an operator which is equal to 1 when $\boldsymbol{x} = \boldsymbol{x}'$, and 0 otherwise. The nugget term has a number of functions in this application, including accounting for any non-linearity in the output response to the inputs and for non-explicitly specified inactive inputs, such as initial conditions and experiment, and averaging length. It also represents the effects of lower-order PCs that are excluded from the emulator.

Now consider run *i*, which has inputs characterised by $\boldsymbol{x}_i$ and outputs by $\boldsymbol{y}_i$. Let $\boldsymbol{H}$ be the design matrix relating to the GCM output, where row *i* represents the regressors $\boldsymbol{h}(\boldsymbol{x}_i)$, making $\boldsymbol{H}$ an *n* by *q* matrix. The adopted modelling approach states that the prior distribution of $\boldsymbol{y}$ is Gaussian, characterised by $\boldsymbol{y} \sim N(\boldsymbol{H\beta}, \sigma^2 \boldsymbol{A})$, with $\boldsymbol{A}_{ij} = c(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Following the specification of the prior model above, a Bayesian approach is now used to update the prior distribution. The posterior estimate of the GCM output is described by:

$$m^*(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + t(\boldsymbol{x})\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{H}\widehat{\boldsymbol{\beta}}), \tag{7}$$

$$V^*(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2[c(\boldsymbol{x}, \boldsymbol{x}') - t(\boldsymbol{x})^T\boldsymbol{A}^{-1}t(\boldsymbol{x}') + \boldsymbol{P}(\boldsymbol{x})(\boldsymbol{H}^T\boldsymbol{A}^{-1}\boldsymbol{H})^{-1}\boldsymbol{P}(\boldsymbol{x}')^T], \tag{8}$$

where

$$\sigma^2 = (n - q - 2)^{-1}(\boldsymbol{y} - \boldsymbol{H}\widehat{\boldsymbol{\beta}})^T\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{H}\widehat{\boldsymbol{\beta}}), \tag{9}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{H}^T\boldsymbol{A}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{A}^{-1}\boldsymbol{y}, \tag{10}$$

and $t(\boldsymbol{x})_i = c(\boldsymbol{x}, \boldsymbol{x}_i)$ and $\boldsymbol{P}(\boldsymbol{x}) = h(\boldsymbol{x})^T - t(\boldsymbol{x})^T\boldsymbol{A}^{-1}\boldsymbol{H}$.

We follow the suggestion of Berger et al. (2001) and assume a vague prior ($\boldsymbol{\beta}$,σ^2) which is proportional to σ^2, an approach that has been adopted by several other studies, including Oakley and O'Hagan (2002), Bastos and O'Hagan (2009), Araya-Melo et al. (2015) and Bounceur et al. (2015). The posterior distribution of the GCM output is a student-t distribution with $n - q$ degrees of freedom, but is sufficiently close to being Gaussian for this application.

Now, taking the output from the PCA performed earlier, we apply the GP model to each basis vector ($a_k(\boldsymbol{x})$), which has been updated according to Eq. 7 and 8, in turn. Thus:

$$a_k(\boldsymbol{x}) = GP[m_k(\boldsymbol{x}), V_k(\boldsymbol{x}, \boldsymbol{x}')], \tag{11}$$

where mean and covariance functions take the form:

$$\boldsymbol{m}(\boldsymbol{x}) = \sum_{k=1}^{n'} m_k(\boldsymbol{x})\boldsymbol{u}_k, \quad (12)$$

$$\boldsymbol{V}(\boldsymbol{x},\boldsymbol{x}') = \sum_{k=1}^{n'} V_k(\boldsymbol{x},\boldsymbol{x}')\boldsymbol{u}_k\boldsymbol{u}_k^T + \sum_{k=n'+1}^{n} \frac{\boldsymbol{S}_{kk}^2}{n}\boldsymbol{u}_k\boldsymbol{u}_k^T, \quad (13)$$

The values of the hyperparameters are chosen by maximising the likelihood of the emulator, following Kennedy and O'Hagan (2000), and based on the following expression from Andrianakis and Challenor (2012):

$$logL(\nu,\delta) = -\frac{1}{2}(\log(|\boldsymbol{A}||\boldsymbol{H}^T\boldsymbol{A}^{-1}\boldsymbol{H}|) + (n-q)\log(\hat{\sigma}^2)) + K, \quad (14)$$

where K is an unspecified constant. On the recommendation of Andrianakis and Challenor (2012), a penalised likelihood is used, which limits the amplitude of the nugget:

$$logL^P(\nu,\delta) = logL(\nu,\delta) - 2\frac{\bar{M}(\nu,\delta)}{\epsilon\bar{M}(\infty)}, \quad (15)$$

where $\bar{M}(\nu,\delta)$ is the Mean Squared Error between the GCM's output data and the emulator's posterior mean at the design points, defined by $\bar{M}(\nu,\delta) = \nu^2/n(\boldsymbol{y}-\boldsymbol{H\beta})^T\boldsymbol{A}^{-2}(\boldsymbol{y}-\boldsymbol{H\beta})$. $\bar{M}(\infty)$ is its asymptotic value at $\delta_i \rightarrow \infty$, given by $\bar{M}(\infty) = 1/n(\boldsymbol{y}-\boldsymbol{H\beta})^T(\boldsymbol{y}-\boldsymbol{H\beta})$. ϵ is assigned a value of 1.

To summarise, in this study $\boldsymbol{D}$ is a 60 x 4 matrix (n x p) of input data, consisting of 60 GCM simulations and four input factors (ε, $e\sin\varpi$, $e\cos\varpi$, and CO_2). The matrix $\boldsymbol{Y}$ contains the output data from the GCM, with dimensions of 96 x 73 x 60 (longitude x latitude x n). A PC analysis is performed on this output data, which is then used to calibrate the emulator. Four hyperparameters (δ) are used, due to there being four input factors, along with a nugget term (ν). The optimal values for these hyperparameters and the number of PCs retained are calculated during calibration and evaluation of the emulator, discussed in Sect. 4. The GCM data used in this study are mean annual SAT~~,~~ and mean annual precipitation, although these are each emulated separately using different emulators.

3 AOGCM simulations

3.1 Model description

To run the GCM simulations, we used the HadCM3 climate model (Gordon et al., 2000; Pope et al., 2000) – a coupled atmosphere-ocean general circulation model (AOGCM) developed by the UK Met Office. Although HadCM3 can no longer be considered as state-of-the-art when compared with the latest generation of GCMs, such as those used in the most recent IPCC Fifth Assessment Report (IPCC, 2013), its relative computational efficiency makes it ideal for running experiments for comparatively long periods of time (of several centuries) and for running large ensembles of simulations, as performed in this study. As a result, this model is still widely used in climate research, both in palaeoclimatic studies (e.g. Prescott et al., 2014) and in projections of future climate (Armstrong et al., 2016). In addition, it has previously been employed in research into climate sensitivity using a statistical emulator (Araya-Melo et al., 2015). 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.

HadCM3 is coupled to the land surface scheme MOSES2.1 (Met Office Surface Exchange Scheme), which was developed from MOSES1 (Cox et al., 1999). It has been used in a wide range of studies (Cox et al., 2000; Crucifix et al., 2005), and a comparison to MOSES1 and to observations is provided by Valdes et al. (2017). MOSES2.1 in turn is 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. Further details of the overall model setup, denoted HadCM3M2.1E, can be found in Valdes et al. (2017).

3.2 Experimental design

In our simulations, four input parameters are varied: atmospheric CO_2 concentration and the three main orbital ~~forcings~~ parameters of longitude of perihelion (ϖ), obliquity (ε) and eccentricity (e). The extents of the GIS and WAIS are also modified, although only between two modes – their present-day configurations and their reduced-extent Pliocene configurations (Haywood et al., 2016). The extent and thickness of the East Antarctic Ice Sheet (EAIS) was not modified. A more detailed description of the continental ice sheet configurations is provided in Sect. 3.5.

We combined eccentricity and longitude of perihelion 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 (ε) (Loutre, 1993). The ranges of orbital and CO_2 values considered are appropriate for the next 1 ~~Ma~~ Myr and a range of anthropogenic emissions scenarios. For the astronomical parameters, calculated using the Laskar et al. (2004) solution, this essentially equates to their full ranges of -0.055 to 0.055 for $e\sin\varpi$ and $e\cos\varpi$, and 22.2° to 24.4° for ε.

For CO_2, an emissions scenario is selected from Lord et al. (2016) in which atmospheric CO_2 follows observed historical concentrations from 1750 CE~~AD~~ (Common Era~~Anno Domini~~) to 2010 CE~~AD~~ (Meinshausen et al., 2011), after which emissions follow a logistic trajectory, resulting in cumulative total emissions of 10,000 Pg C by year ~3200 CE. This experiment was run for 1 ~~Ma~~ Myr using the cGENIE Earth system model, and aims to represent a maximum total future CO_2 release. To put this into perspective: current estimates of ~~remaining~~ fossil fuel reserves are approximately 1000 Pg C, with an estimated ~4000 Pg C in fossil fuel resources that may be extractable in the future (McGlade and Ekins, 2015), and up to 20-25,000 Pg C in nonconventional resources such as methane clathrates (Rogner, 1997). The evolution of atmospheric CO_2 concentration over the next 200 kyr for this emissions scenario is show in Fig. 1. Although in the *c*GENIE simulation, atmospheric CO_2 reaches a maximum of 3900 parts per million (ppmv) within the first few hundred years, this concentration is not at equilibrium and only lasts for a couple of decades before decreasing. As a result, the concentration at 500 years into the experiment, 3600 ppmv, is chosen as the upper CO_2 limit, which means that the climatic effects of emissions of more than 10,000 Pg C cannot be estimated with the emulator.

By the end of the 1 ~~Ma~~ Myr emissions scenario, atmospheric CO_2 concentrations have nearly declined to pre-industrial levels, reaching 285 ppmv. However, this experiment does not account for natural variations in the carbon cycle, which result in periodic fluctuations in CO_2. For example, during the Holocene (11 kyr BP to

~1750 CE) ~~which resulted in~~ atmospheric CO_2 vari~~y~~ed~~ing~~ between 260 and 280 ppmv ~~during the Holocene (11 kyr BP to ~1750 AD)~~ (Monnin et al., 2004). A value of 250 ppmv is therefore deemed to be appropriate to account for these natural variations in an unglaciated world, in addition to possible uncertainties in the model and hence is assumed as the value of the lower CO_2 limit in the ensemble.

The orbital and CO_2 parameter ranges that have been selected are also applicable to unglaciated periods during the ~~the~~ late Pliocene, when atmospheric CO_2 was estimated to be higher than pre-industrial values (Martinez-Boti et al., 2015; Raymo et al., 1996). In this study, we do not consider or attempt to simulate past or future glacial episodes, which may be accompanied by larger continental ice sheets (see Sect. 7 for more discussion), although the conditions required to initiate the next glaciation, and extending the ensemble of GCM simulations to represent glacial states, are being investigated in a forthcoming~~separate~~ study. The underlying assumption of our ensemble is that it is suitable for simulating periods for which the CO_2 concentration is high enough to prevent entry into a glacial state.

Two ensembles were generated, each made up of 40 simulations, meeting the recommended 10 experiments per input parameter (Loeppky et al., 2009). One ensemble includes orbital values suitable for the next 1 ~~Ma~~ Myr and a relatively small range of lower CO_2 values, whereas the other ensemble represents the shorter-term future with a reduced range of orbital values and a larger range of higher CO_2 concentrations. This approach was adopted because various studies have shown that on geological timescales of thousands to hundreds of thousands of years, an emission of ~~fossil fuel~~anthropogenic CO_2 to the atmosphere is ~~removed~~ taken up by natural carbon cycle processes over different timescales (Archer et al., 1997; Lord et al., 2016). A relatively large fraction of the CO_2 perturbation is neutralised on shorter timescales of 10^3-10^4 years, but it takes 10^5-10^6 years for atmospheric CO_2 concentrations to very slowly return to pre-industrial levels (Colbourn et al., 2015; Lenton and Britton, 2006; Lord et al., 2016), if the effects of glacial-interglacial cycles and other natural variations, such as those due to imbalances between volcanic outgassing and weathering, are excluded. Hence, only a relatively short portion of the full million years has very high CO_2 concentrations under any emissions scenario, with the major part of the time having a CO_2 concentration no more than several hundred ppmv above pre-industrial, as demonstrated in Fig. 1.

The parameter ranges for the two ensembles, which are referred to as "*highCO₂*" and "*lowCO₂*", are given in Table 1. The cut-off point for the *highCO₂* ensemble is set at 110 kyr AP, as after this time eccentricity, which remained relatively low prior to this time, starts to increase more rapidly, and variability in $e\sin\varpi$ and $e\cos\varpi$ increases. This first ensemble therefore has CO_2 sampled up to 3600 ppmv, and the orbital parameters are sampled within the reduced range of values that will occur over the next 110 kyr. The *lowCO₂* ensemble samples the full range of orbital values and the upper CO_2 limit is set to 560 ppmv. This upper limit also covers the range of CO_2 concentrations that have been estimated for the late Pliocene (e.g. Martinez-Boti et al., 2015; Seki et al., 2010). At 110 kyr in the 10,000 Pg C emissions scenario, the atmospheric CO_2 concentration is 542 ppmv, which is rounded up to twice the pre-industrial atmospheric CO_2 concentration (560 ppmv = 2*280 ppmv), a common scenario used in future climate-change modelling studies.

The benefits of the approach of having separate ensembles for high and low CO_2 mean that both parameter ranges have sufficient sampling density, whilst also reducing the chance of unrealistic sets of parameters, in particular for the period of the next 110 kyr. During this time, CO_2 is likely to be comparatively high, while eccentricity remains relatively low, and $e\sin\varpi$ and $e\cos\varpi$ exhibit relatively low variability. Having a separate ensemble in which CO_2 and the orbital parameters are only sampled within the ranges experienced within the next 110 kyr avoids wasting computing time on parameter combinations that are highly unlikely to occur, such as very high CO_2 and very high eccentricity. This methodology also provides the additional benefit of the low CO_2 emulator being applicable to palaeo-modelling studies, as the ensemble encompasses an appropriate range of CO_2 and orbital values for many past periods of interest, such as the Pliocene.

3.3 Generation of experiment ensembles

We used the Latin hypercube sampling function from the MATLAB Statistics and Machine Learning Toolbox (LHC; (MATLAB, 2012b)) to generate the two ensembles, thereby ~~. This is a statistical method that~~ efficiently sampling~~es~~ the four-dimensional input parameter space (Mckay et al., 1979). Briefly, this method~~ works by~~ divides~~ing~~ the parameter space within the prescribed ranges into n equally probable intervals, n being the number of experiments required, which in this case is 40 per ensemble. n points are then selected for each input variable, one from each interval, without replacement. The sample points for the four variables are then randomly combined. The LHC sampling function also includes an option to maximize the minimum distance between all pairs of points (the maxi-min criterion~~a~~), which is utilised here to ensure the set of experiments is optimally space filling.~~ This is called the maxi-min criteria.~~

For each ensemble, 3000 sample sets were created, with each set consisting of an n by p matrix, $\boldsymbol{X}$, containing the four sampled input parameter values for each of the 40 experiments, and then the optimal sample set was selected as the final ensemble based on a number of criteria. Following Joseph and Hung (2008), we seek, in addition to the maxi-min criterion~~a~~, to maximise $det(\boldsymbol{X}^T\boldsymbol{X})$. Here, we will term this determinant the "orthogonality", because the columns of the design matrix will ~~indeed~~ approach orthogonality as this determinant is maximised (assuming that input factors are normalised). However, a limitation of the method of sampling the parameters $e\sin\varpi$ and $e\cos\varpi$, rather than eccentricity and longitude of perihelion directly, is that due to the nature of the $e\sin\varpi$ and $e\cos\varpi$ parameter space, the sampling process favours higher values of eccentricity over lower ones. This is not an issue for the longitude of perihelion, because~~as~~ when eccentricity is low the value of this parameter has little effect on insolation. However, the value of obliquity selected for a given eccentricity value could have a significant impact on climate, meaning that it is desirable to have a relatively large range of obliquity values for low (<0.01) and high (>0.05) eccentricity values, in order to sample the boundaries sufficiently. It was observed that the sample sets with the highest orthogonality had comparatively few, if any, values of low eccentricity, also meaning that a very limited number of obliquity values were sampled for low eccentricity. We therefore adopted the approach whereby all sample sets that demonstrated normalised orthogonality values that were more than 1 standard deviation above the mean orthogonality were selected. From these, the single sample set with the greatest range of obliquity values for low eccentricity, hence with maximal sampling coverage of the low eccentricity boundary, was selected as the final ensemble design. The input parameter values for the *high*CO_2 and *low*CO_2 ensembles are given in Table 2, and the distributions in parameter space illustrated in Fig. 2.

3.4 AOGCM simulations

The two CO_2 ensembles were initially run with constant modern-day GIS and WAIS configurations (*modice*). All experiments were initiated from a pre-industrial spin-up experiment, with an atmospheric CO_2 concentration of 280 ppmv, and pre-industrial ice sheet extents and orbital conditions. Atmospheric CO_2 and the orbital parameters were kept constant throughout each simulation, and each experiment was run for a total of 500 model years. This simulation~~run~~ length allows the experiments with lower CO_2 to reach near-equilibrium at the surface. Experiments with higher CO_2 have not yet equilibrated by the end of this period; the significance of this is addressed in Sect. 3.6. A number of the very high CO_2 experiments caused the model to become unstable and the interpretation of these experiments is discussed in Sect. 3.4.1. A control simulation was also run for 500 years, with the atmospheric CO_2 concentration and the orbital parameters set at pre-industrial values. All climate variable results for the model, unless specified, are an average of the final 50 years of the simulation. Anomalies compared with the pre-industrial control (i.e. emulated minus pre-industrial) are discussed and used in the emulator, rather than absolute values, to account for biases in the control climate of the model.

3.4.1 Very high CO_2 simulations

As mentioned previously, experiments in the *high*CO_2 ensemble with CO_2 concentrations of greater than 3100 ppmv become unstable. These experiments exhibit accelerating warming trends several hundred years into the simulation, which eventually cause the model to crash before completion. This is the result of a runaway positive feedback in the GCM caused, at least in part, by the vertical distribution of ozone in the model being prescribed for modern-day climate conditions. Consequently, ~~, rather~~ the ozone distribution~~it~~ is not ~~than being~~ able to respond to changes in climate,~~,~~ meaning that~~-~~ when increased mean global temperatures result in an increase in altitude of the tropopause and hence an extension of the troposphere, ~~resulting in runaway warming as r~~relatively high concentrations of ozone, which were previously located in the stratosphere, enter the troposphere, resulting in runaway warming.

All other experiments ran for the full 500 years. However, those with a CO_2 concentration of 2000 ppmv or higher also exhibited accelerating warming trends before the end of the simulation. Consequently, only simulations with CO_2 concentrations of less than 2000 ppmv (equivalent to a total ~~fossil fuel~~ CO_2 release of up to 6000 Pg C) are included in the rest of this study, meaning the methodology is not appropriate for CO_2 values greater than this. This equates to 20 experiments in total from the *high*CO_2 ensemble, with CO_2 concentrations ranging from 303 to 1901 ppmv. All 40 of the *low*CO_2 experiments were used.

3.5 Sensitivity to ice sheets

In addition to running the two ensembles with modern-day GIS and WAIS configurations, we also investigated the climatic impact of reducing the sizes of the ice sheets. Many of the CO_2 values sampled, particularly in the *high*CO_2 ensemble, are significantly higher than pre-industrial levels, and if the resulting climate were to persist for a long period~~s~~ of time ~~they~~ it could result in significant melting of the continental ice sheets over timescales of 10^3-10^4 years (Charbit et al., 2008; Stone et al., 2010; Winkelmann et al., 2015).

We therefore set up the *highCO₂* and *lowCO₂* ensembles with reduced GIS and WAIS extents (*lowice*), using the PRISM4 Pliocene reconstruction of the ice sheets (Dowsett et al., 2016). In this reconstruction, the GIS is limited to high elevations in the Eastern Greenland Mountains, and no ice is present over Western Antarctica. Similar patterns of ice retreat have been simulated in response to future warming scenarios for the GIS (Greve, 2000; Huybrechts and de Wolde, 1999; Ridley et al., 2005; Stone et al., 2010) and WAIS (Huybrechts and de Wolde, 1999; Winkelmann et al., 2015), equivalent to ~7 m (Ridley et al., 2005) and ~3 m (Bamber et al., 2009; Feldmann and Levermann, 2015) of global sea level rise, respectively. Large regions of the ~~East Antarctic ice sheet (~~EAIS~~)~~ show minimal changes or slightly increased surface elevation, although there is substantial loss of ice in the Wilkes and Aurora subglacial basins (Haywood et al., 2016).

The same CO_2 and orbital parameter sample sets were used for both ice configuration ensembles to allow the impact of varying the ice~~-~~ sheet extents on climate to be directly compared. Only the Greenland and Antarctic grid boxes were modified; the boundary conditions for all other grid boxes, as well as the land/sea mask, were the same as in the modern-day ice sheet simulations. For Greenland and Antarctica, the extent and orography of the ice sheets was updated with the PRISM4 data, as well as the orography of any grid boxes that are projected to be ice-free. Soil properties, land surface type and snow cover were also updated for these grid boxes. Figure 3 compares the orography for the *modice* and *lowice* ensembles, clearly showing the reduced extents for the ice sheets.

3.5.1 Pattern scaling of reduced ice simulations

It was expected that reducing the size of the continental ice sheets would have a relatively localised impact on climate (Lunt et al., 2004), and that the effect would be of a linear nature. Therefore, a subset of five simulations from the two ensembles were selected as reduced ice ~~-~~sheet simulations (*lowCO₂* – experiments 8, 19 and 29; *highCO₂* – experiments 21, and 34; see Table 2), covering a range of orbital and CO_2 values.

A comparison of the mean annual SAT anomaly for the five experiments showed that the largest temperature changes occur over Greenland and Antarctica, particularly in regions where there is ice in the *modice* ensemble but that are ice free in *lowice*. The spatial pattern of the change is also fairly similar across the simulations, suggesting that the response of climate to the extents of the ice sheets is largely independent of orbital variations or CO_2 concentration. The SAT anomaly for the five *lowice* experiments compared with their *modice* equivalents was calculated, and then averaged across the experiments, shown in Fig. 4a. The largest SAT anomalies occur locally to the GIS and Antarctic ice sheet (AIS), accompanied by smaller anomalies in some of the surrounding ocean regions (e.g. Barents and Ross Seas), with no significant changes in SAT elsewhere, in line with the results of Lunt et al. (2004); Toniazzo et al. (2004) and (Ridley et al., 2005). This SAT anomaly, caused by the reduced extents of the GIS and WAIS, was then applied (added) to the mean annual SAT anomaly data for all other *highCO₂* and *lowCO₂* *modice* experiments, to generate the SAT data for two *lowice* ensembles.

Also shown in Fig. 4, for comparison, are mean annual SAT anomalies produced by the other forcings, including a doubling of CO_2, the difference between maximum and minimum obliquity and the difference between "warm" orbital conditions and "cold" orbital conditions. The warming caused by increased CO_2 is more widespread (Fig. 4b), with the largest warming occurring at high latitudes and for land regions, in agreement with typical future-climate simulations (IPCC, 2013, p. 1059). The temperature change due to obliquity and "warm" versus "cold" orbital conditions is less than that for either reduced ice (compared to pre-industrial) or increased CO_2. Changes in obliquity have the largest impact on temperatures in high latitude regions, since the exposure of these regions to the sun's radiation is most affected by changes in obliquity. Smaller temperature anomalies are observed over northern Africa and India and, since an increase in obliquity is indeed known to boost monsoon dynamics (e.g. Araya-Melo et al., 2015; Bosmans et al., 2015), changes in soil latent heat exchanges are therefore expected to contribute negatively to the temperature response. The comparison of "warm" versus "cold" orbital conditions, which highlights (annual mean) temperature changes primarily caused by precession, generally shows a warming trend, with the largest temperature changes occurring in monsoonal regions. Lower temperatures are observed in the Northern Hemisphere over northern Africa, India and, East Asia, whilst warmer temperatures occur in the Southern Hemisphere over South America, southern Africa and Australia. Figure 4 demonstrates that the temperature forcing caused by CO_2 affects mean annual temperatures on a global scale, whilst the forcing due to ice sheet and orbital changes affects mean annual temperatures in specific regions, having a limited impact on global mean temperatures. This is supported by the relatively high global mean SAT anomaly for the 2xCO_2 scenario of 4.2°C, compared with the lower SAT anomalies that result from the obliquity and precession forcing of 0.4°C each (see caption of Fig. 4).

3.6 Calculation of equilibrated climate

Given the high values of CO_2 concentration in many of the experiments, particularly in the *highCO_2* ensemble, even by the end of the 500 yr running period the climate has not yet reached steady state. We therefore estimated~~calculated~~ the fully equilibrated climate response using the methods described below.

3.6.1 Gregory plots

In order to estimate the equilibrated response, we applied the method of Gregory et al. (2004) to the model results, regressing the net radiative flux at the top of the atmosphere (TOA) against the global average SAT change, as displayed in figures termed Gregory plots (Andrews et al., 2015; Andrews et al., 2012; Gregory et al., 2015). In this method, for an experiment which has a constant forcing applied (i.e. with no inter-annual variation) it can be assumed that:

$$N = F - \alpha \Delta T, \qquad (16)$$

where N is the change in the global mean net TOA radiative flux (W m^{-2}), F is the effective radiative forcing (W m^{-2}; positive downwards), α is the climate feedback parameter (W m^{-2} ^{o}C^{-1}), and ΔT is the global mean annual SAT change compared with the control simulation (^{o}C). This method works on the assumption that if F and α are constant, N is an approximately linear function of ΔT. By linearly regressing ΔT against N, both F (intercept of the line at $\Delta T = 0$) and -α (slope of the line) can be diagnosed. The intercept of the line at $N = 0$ provides an estimate of the equilibrium SAT change (relative to the pre-industrial SAT) for the experiment, denoted $\Delta T_{eq}{}^{g}$ to indicate it was calculated from the Gregory plots, and is equal to F/α. This is in contrast to the SAT change calculated directly from the GCM model data by averaging the final 50 years of the experiment (ΔT_{500}).

The Gregory plots for two *modice* experiments, *modice_lowCO2_13* (CO_2 555.6 ppmv) and *modice_highCO2_17* (CO_2 1151.6 ppmv), are shown in Fig. 5. These experiments were selected as they have CO_2 values nearest to the 2x and 4x pre-industrial CO_2 scenarios that are commonly used in idealised future climate experiments. For each experiment, mean annual data are plotted for years 1-20 of the simulation, and mean decadal data for years 21-500. The regression fits are to mean annual data in each case, and years 1-20 and 21-500 were fitted separately. The values for F and α estimated from Fig. 5 are presented in Table 3. These values are slightly lower than those identified in other studies using the same method. For example, Gregory et al. (2004) used HadCM3 to run experiments with 2x and 4xCO_2, obtaining values for years 1-90 of 3.9 ± 0.2 and 7.5 ± 0.3 W m^{-2} for F, and -1.26 ± 0.09 and -1.19 ± 0.07 W m^{-2} ^{o}C^{-1} for α, respectively. Andrews et al. (2015) calculated F to be 7.73 ± 0.26 W m^{-2} and α to be -1.25 W m^{-2} ^{o}C^{-1} for years 1-20 and -0.74 W m^{-2} ^{o}C^{-1} for years 21-100 for 4xCO_2 simulations using HadCM3. The differences between our results and theirs may be due to the fact that we used MOSES2.1 and the TRIFFID vegetation model, whereas they used MOSES1, which is a different land-surface scheme and does not account for vegetation feedbacks.

The decrease in the climate response parameter (α) as the experiment progresses suggests that the strength of the climate feedbacks changes as the climate evolves over time. Consequently, the ΔT intercept ($N = 0$) for the first 20 years of the simulation underestimates the actual warming of the model. Over longer timescales, the slope of the regression line becomes less negative, implying that the sensitivity of the climate system to the forcing increases (Andrews et al., 2015; Gregory et al., 2004; Knutti and Rugenstein, 2015). This non-linearity has been found to be particularly apparent in cloud feedback parameters, in particular shortwave cloud feedback processes (Andrews et al., 2015; Andrews et al., 2012). A number of studies have attributed this strengthening of the feedbacks to changes in the pattern of surface warming (Williams et al., 2008), mainly in the eastern tropical Pacific where an intensification of warming can occur after a few decades, but also in other regions such as the Southern Ocean (Andrews et al., 2015). The impact of variations in ocean heat uptake has also been suggested to be a contributing factor (Geoffroy et al., 2013; Held et al., 2010; Winton et al., 2010).

We take the ΔT intercept ($N = 0$) for years 21-500 to give the equilibrium temperature change ($\Delta T_{eq}{}^{g}$) for the experiments, equating to values of 4.3^{o}C and 8.9^{o}C for the 2x and 4xCO_2 scenarios in Fig. 5. A limitation of this approach is that it assumes that the response of climate to a forcing is linear after the first 20 years, which has been shown to be unlikely in longer simulations of several decades or centuries (Andrews et al., 2015; Armour et al., 2013; Winton et al., 2010). However, a comparison of the difference in temperature response to upper- and deep-ocean heat uptake and its contribution to the relationship between net radiative flux change (N) and global temperature change (ΔT) in Geoffroy et al. (2013) indicated that the method of Gregory et al. (2004) of fitting two separate linear models to the early and subsequent (N, ΔT) data gives a good approximation of $\Delta T_{eq}{}^{g}$, F and α as they have been calculated here. A study by Li et al. (2013) also found that, using the Gregory plot methodology, $\Delta T_{eq}{}^{g}$ was estimated to within 10% of its actual value, obtained by running the simulation very close to equilibrium (~6000 yr). However, this was using the ECHAM5/MPIOM model, meaning that it is not necessarily also true for HadCM3.

Given that the slope of the 21-500 yr regression line appears to become shallower with time, the estimates of ΔT_{eq}^{g} should be taken as a lower limit of the actual equilibrated SAT anomaly. However, this tendency to flatten, particularly as the CO_2 concentration is increased, further justifies our use of the Gregory methodology; by the end of 500 years the high CO_2 experiments have not yet reached steady state, and even in the lower CO_2 experiments SAT is increasing very slowly, so will likely take a long time to reach equilibrium. It would therefore not be feasible to run most of these experiments to steady state using a GCM, due to the associated computational and time requirements. Furthermore, on longer timescales the boundary conditions (orbital characteristics and, more importantly, atmospheric CO_2 concentrations) would have changed, such that, in reality, equilibrium would never be attained.

3.6.2 Equilibrated climate

The final estimates of ΔT_{eq}^{g} for the *low*CO_2 and *high*CO_2 *modice* ensembles range from a minimum of -0.4ºC (CO_2 264.5 ppmv) to a maximum of 12.5ºC (CO_2 1900.9 ppmv). Figure 6 illustrates the difference between global mean annual SAT anomaly calculated from the GCM model data (ΔT_{500}) and calculated using the Gregory plot (ΔT_{eq}^{g}). Experiments with CO_2 below or near to pre-industrial levels tended to reach equilibrium by the end of the 500 years making a Gregory plot unnecessary, hence ΔT_{eq}^{g} is taken to be the same as ΔT_{500} in these cases. As CO_2 increases, the data points in Fig. 6 deviate further from the 1:1 line. This is the result of the ratio between ΔT_{eq}^{g} and ΔT_{500} increasing, as the experiments grow increasingly far from equilibrium by the end of the GCM run with increasing CO_2.

We next calculated the ratio between ΔT_{eq}^{g} and ΔT_{500} for each experiment ($\Delta T_{eq}^{g}/\Delta T_{500}$), which represents the fractional increase in climate change still due to occur after the end of the 500-~~–~~year model run in order for steady state to be reached. To estimate the fully equilibrated climate anomaly, the spatial distribution of mean annual SAT anomaly was multiplied by the $\Delta T_{eq}^{g}/\Delta T_{500}$ ratio. The ratio identified for each experiment is assumed to be equally applicable to all grid boxes. The same scaling ratio was also applied to the precipitation anomaly data to estimate the equilibrated mean annual precipitation.

The equilibrated global mean annual SAT anomaly (ΔT_{eq}) for the *high*CO_2 and *low*CO_2 *modice* ensembles is plotted against ln~~og~~(CO_2) in Fig. 7, along with ΔT_{500} for reference.~~ The linear nature of the plot increases our confidence that the Gregory methodology is suitable for our uses, given the logarithmic relationship between SAT and CO2 concentration~~. Also plotted on Fig. 7 are a number of lines illustrating idealised relationships between ΔT_{eq} and CO_2 based on a range of climate sensitivities. The most recent IPCC report suggested that the likely range for equilibrium climate sensitivity is 1.5°C to 4.5°C (IPCC, 2013), hence sensitivities of 1.5°C, 3°C and 4.5°C have been plotted. The size of the correction required to calculate ΔT_{eq} from ΔT_{500} increases with increasing CO_2, and brings the final temperature estimates in line with the expected response (red lines), further increasing our confidence. The ΔT_{eq} estimated for the experiments generally follows the upper line, equivalent to an equilibrium climate sensitivity of 4.5°C, which is higher than a previous estimate of 3.3ºC for HadCM3

(Williams et al., 2001). This difference may be due to our simulations being "fully equilibrated" following the application of the Gregory plot methodology. In addition, Williams et al. (2001) used an older version of HadCM3 and prescribed vegetation (MOSES1), whilst in this study interactive vegetation is used (MOSES2.1 with TRIFFID).

4 Calibration and evaluation of the emulator

By considering different contributions of modern and low ice, high and low CO_2, different number of PCs, and different values for the correlation length hyperparameters, we generated an ensemble of emulators, in order to test their relative performance. The *modice* and *lowice* ensembles were treated as independent data sets that were used separately when calibrating the emulator, since ice extent is not defined explicitly as an input parameter in the emulator code. This approach was adopted, rather than including the ice sheet extent as an active input parameter to the emulator, because only two ice sheet configurations have been simulated, which are not sufficient for an interpolation. One of the main benefits of including ice sheet extent as an active input parameter would be to emulate changing ice sheets over time, but this was beyond the scope of this study.~~ and this methodology in its current form, as the glacial-interglacial cycles are not considered.~~ l~~L~~n~~og~~(CO_2) was used as one of the four input parameters, along with obliquity, $e\sin\varpi$ and $e\cos\varpi$. The performance of each emulator was assessed using a leave-one-out cross-validation approach, where a series of emulators is constructed, and used to predict one left-out experiment each time. For example, for the *lowCO₂ modice* ensemble (40 experiments), 40 emulators were calibrated with one experiment left out of each. This left-out experiment was then reproduced using the corresponding emulator, and the results compared with the actual experiment results. The number of grid boxes for each experiment calculated to lie within different standard deviation bands, and the root mean squared error (RMSE) averaged across all the emulators were used as performance indicators to compare the different input configurations and hyperparameter value selections. The results in this section are applicable to the *modice* emulator, unless otherwise specified, however the calibration and evaluation for the *lowice* emulator yielded similar trends and results.

4.1 Sensitivity to input data

We investigated the impact on performance of calibrating the emulator on the *high*CO_2 and *low*CO_2 *modice* ensembles separately, and combined. The *low*CO_2 *modice* emulator generally performs slightly better in the leave-one-out cross-validation exercise than the *high*CO_2 *modice* version, with a lower RMSE and fewer grid boxes with an error of more than 2 standard deviations. Combining the two ensembles into one emulator results in a similar RMSE to the *low*CO_2-only *modice* emulator but decreases the RMSE compared with the *high*CO_2-only *modice* emulator. As a consequence, we took the approach of calibrating the emulator on the combined ensembles for the rest of the study. This has the advantage that continuous simulations of climate with CO_2 levels that cross the boundary between the high and low CO_2 ensembles (~560 ppmv), such as may be appropriate for emulation of future climate, can be performed using one emulator, rather than having to calibrate separate emulators for different time periods based on CO_2 concentration. There is also no loss of performance in the emulator for either set of CO_2 ranges, but rather a slight improvement for the *high*CO_2 ensemble.

4.2 Optimisation of hyperparameters

We calibrated two separate emulators, the first using the *modice* data and the second using the *lowice* data, both with 60 experiments each (combined *highCO₂* and *lowCO₂*). The input factors (ε, $e\sin\varpi$, $e\cos\varpi$ and ln(CO_2)) 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 of the column. We tested different emulator configurations by varying the number of principal components retained, ranging from 5 to 20, and for each emulator configuration, the correlation length scales δ and nugget ν were optimized by maximization of the penalised likelihood. This optimisation was carried out in log-space, ensuring that the optimised hyperparameters would be positive. A leave-one-out validation was performed each time, and the *modice* and *lowice* configurations that performed best were selected as the final two optimised emulators. We found that a *modice* emulator retaining 13 principal components has the lowest RMSE and a relatively low percentage of grid boxes with errors of more than 2 standard deviations. The scales δ for the *modice* emulator are 7.509 (ε), 3.361 ($e\sin\varpi$), 3.799 ($e\cos\varpi$), 0.881 (CO_2), and the nugget is 0.0631. In contrast, a *lowice* emulator using 15 principal components exhibits the best performance, with length scales δ of 5.597 (ε), 2.887 ($e\sin\varpi$), 3.273 ($e\cos\varpi$), 0.846 (CO_2), and a nugget of 0.0925. In both cases, the scales for the three orbital parameters are larger than the range associated with the input factors, indicating that the response is relatively linear with respect to these terms.

The *modice* emulator was evaluated using the leave-one-out methodology and results are shown in Fig. 8. The results suggest that the emulator performs well. Figure 8a shows the percentage of grid boxes for each left-out experiment estimated by the corresponding emulator within different standard deviation bands, along with the RMSE. The mean percentage of grid boxes within 1 and 2 standard deviations is 80% and 97%, which roughly corresponds to the 68% and 95% ratios expected for a normal distribution, suggesting that the uncertainty in the prediction is being correctly captured.

Several of the experiments performed considerably worse than others, exhibiting below the expected number of grid boxes with errors within 1 standard deviation (for reference, the mean value for 1 standard deviation across the left-out experiments is 0.3ºC), and/or higher than the expected number of grid boxes with errors of greater than 2 standard deviations, which is generally accompanied by a higher RMSE. However, the input conditions for these experiments are not particularly similar or unique. Experiments *modice_highCO2_43*, *modice_highCO2_45* and *modice_highCO2_46* all have a fairly low eccentricity and obliquity, and a CO_2 concentration of ~1000 ppmv, but there are multiple experiments with similar values that have lower RMSE values. A spatial map of the errors (not shown) indicates that the grid boxes with errors of 3 or more standard deviations are at high northern latitudes in these experiments. However, the signs of the anomalies are not the same across these experiments, as the emulator overestimates the Arctic SAT anomaly in *modice_highCO2_43* and underestimates it in *modice_highCO2_45* and *modice_highCO2_46*. This suggests that the emulator is perhaps not quite capturing the full model behaviour in high northern latitudes, particularly for low eccentricity values, but this is certainly not true for all experiments. The errors in the experiments are generally less than ±4ºC, and for most of the Arctic much lower than that. Note that the Arctic is a region in the model with high inter-annual variability, so one factor may be that the model simulations which are used to calibrate the emulator are not representative of the true stationary mean. There does not appear to be any obvious systematic error associated with the input parameters, suggesting that errors are less likely to be an issue resulting from the design of the emulator and more likely to arise from run-to-run variability in the behaviour of the underlying GCM.

Figure 8b compares the mean annual “SAT index” for each left-out experiment calculated by the GCM and the corresponding emulator (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 grid box area). There are no obvious outliers, and the emulated means are relatively close to their modelled equivalents. There also does not appear to be any significant loss of performance at very low or very high temperature, and therefore at very low or very high CO_2.

In summary, our calibration and evaluation shows that the emulator is able to reproduce the left-out ensemble simulations reasonably well, with no obvious systematic errors in its predictions. Using the emulator, calibrated on the full set of 60 simulations (*modice* or *lowice*), we are able to simulate global climate development over long periods of time (several ~~million~~ hundred thousand years or longer), provided that the atmospheric CO_2 levels for the period are known, and are within the limits of those used to calibrate the emulator, ice sheets do not change outside of the ~~range~~ two configurations considered in the two ensembles, and the topography and land-sea mask are unchanged.

In the next two sections, we present illustrative examples of a number of potential applications of the emulator, by applying it to the late Pliocene in Sect. 5, and the next 200 kyr in Sect. 6.

5 Application of the emulator to the late Pliocene

In addition to being able to rapidly project long-term climate evolution, the emulator also allows climatic changes to be examined and analysed using a range of different methods that may not be possible using other modelling approaches. To illustrate this, we applied the *lowice* emulator to the late Pliocene and compared the results to palaeo-proxy data for the period. The *lowice* emulator was used because the ice sheets in this configuration are the PRISM4 Pliocene ice sheets (Dowsett et al., 2016). It should be noted, however, that this approach is only appropriate for periods of the Pliocene with equivalent or less than modern ice sheet extents (i.e. not glacial conditions), and that palaeogeographic changes for the Pliocene are not included here (~~although~~ see Sect. 7 for further discussion ~~of the fact that we use modern geography~~). We also tested the *modice* emulator which, in agreement with the findings in Sect. 4, had a limited impact on the long-term evolution of global SSTs outside the immediate region of the ice sheets themselves. Potential applications of the emulator for palaeoclimate are described below.

5.1 Time series data

One application of the emulator is to produce a time series of the continuous evolution of climate for a particular time period, as is illustrated here where climate is simulated at 1 kyr intervals over the period 3300 – 2800 kyr BP. This period of the late Pliocene was selected because it has been extensively studied as part of a number of projects (e.g. PRISM (Dowsett et al., 2016; Dowsett, 2007), PlioMIP (Haywood et al., 2010; Haywood et al., 2016)), represents the warm phase of climate (interglacial conditions), and does not include major glaciations (though ~~like~~ the M2 cooling event may persist to the very start of the simulation at 3300 kyr BP, and the simulated period does include periods of likely glaciation, such as KM2 (~3100 kyr BP) and G20 (~3000 kyr BP)). T~~, for which t~~he emulator would not be appropriate to periods of extensive glaciation and may not be well-matched to the periods of lesser glaciation included within the simulated interval. Orbital data for each ~~of the time slices~~1 kyr (Laskar et al., 2004) were provided as input to the calibrated emulator, along with three representative CO_2 concentrations. Three CO_2 reference scenarios were initially emulated, with constant concentrations of 280, 350 and 400 ppmv (although note that in reality, CO_2 varied during this period on orbital timescales (Martinez-Boti et al., 2015)).

To illustrate the comparison of the emulator results to palaeo-proxy data, SST data for various locations were compared with the emulated SAT for the equivalent grid box. Specifically, alkenone-derived palaeo-SST estimates from four (Integrated) Ocean Drilling Program (IODP/ODP) sites were used: ODP Site 982 (North Atlantic; (Lawrence et al., 2009)), IODP Site U1313 (North Atlantic; (Naafs et al., 2010)), ODP Site 722 (Arabian Sea; (Herbert et al., 2010)) and ODP Site 662 (tropical Atlantic; (Herbert et al., 2010)). The locations of the sites are shown in Fig. 9a and detailed in Table 4. These Pliocene datasets were selected because they are all of sufficiently high resolution (≤4 kyr) for the impacts of individual orbital cycles on climate to be captured, whilst covering a range of locations and climatic conditions. Alkenone data are shown converted to SST using two commonly applied calibrations: Prahl et al. (1988) and Muller et al. (1998). All temperatures are presented as an anomaly compared with pre-industrial. The emulator results are compared with the SAT for the relevant grid box in the pre-industrial control experiment, whilst the proxy data are compared with SST observations for the relevant location taken from the HadISST dataset (Rayner et al., 2003). Observations are annual means and are averaged over the period 1870-1900.

Table 4 presents the mean SAT anomaly (compared with pre-industrial) ~~F~~for the modelled period as estimated by~~,~~ the emulator ~~estimates the mean SAT anomaly compared with the pre-industrial control in~~ for the 280 ppmv scenario ~~to be 0.6 ± 0.4°C, -0.8 ± 0.3°C, 0 ± 0.2°C, 0.2 ± 0.2°C for the two North Atlantic (982 and U1313), Arabian Sea, and equatorial Atlantic grid boxes, respectively (Table 4)~~for each of the four grid boxes. The~~is~~ mean increases with increasing CO_2, by ~1°C at low latitudes to 2-3°C at high latitudes for atmospheric CO_2 of 400 ppmv. Figure 10 illustrates the evolution of annual mean temperature variations through the late Pliocene as calculated using the various methods. For the equatorial and Arabian Sea sites (662 and 722), the SAT and SST estimates are relatively similar to each other in terms of the general estimated temperature, particularly for the higher CO_2 scenarios of 350 and 400 ppmv. However, the comparison of timings and variations between the SAT and SST data is fairly poor, and there was not found to be a significant correlation between the emulated and proxy data temperatures at these sites when correlation coefficients were calculated. In fact, Site 982 was the only location for which significant (negative) correlations were found for a confidence interval of 95%, although the correlation coefficient is still relatively low. These correlation coefficients were -0.22 (p-value 0.004) for the Prahl et al. (1988) proxy SST data compared with the emulated SAT for the 280 ppmv scenario, and -0.2 (p-value 0.007) for the same SST data compared with the emulated SAT for the 350 ppmv scenario.

The Muller et al. (1998) SST data demonstrated correlation coefficients that were essentially identical to those above when compared with the same emulated SATs.

At the higher latitudes, the simulated SAT estimate is generally lower than the proxy data SST. This is a common issue in GCM simulations of the late Pliocene, where temperatures at high latitudes under increased CO_2-induced radiative forcing are often underestimated (Haywood et al., 2013). It could also be that the alkenones are not recording mean annual temperature, and instead are being produced during peak warmth (e.g. during the summer months), especially at higher latitudes (Lawrence et al., 2009). This issue could be explored further by extending the methodology presented here to other variables. This seasonal bias could explain the large offset in temperature at the northernmost site (982), which exhibits a maximum difference in mean temperature anomaly for the period of 5.1ºC between data sets, and possibly also at Site U1313. The emulated uncertainty in SAT (defined as 1 standard deviation of the emulated grid box posterior variance) is also shown in Fig. 10, and average values for the period given in Table 4. This is slightly higher at the northernmost North Atlantic site (982) compared to the lower latitude sites, but overall the uncertainty is relatively small when compared with the effects of variations in the orbital parameters and atmospheric CO_2 concentration.

5.2 Orbital variability and spectral analysis

The emulator can also be used to identify the influence of orbital variations on long-term climate change. One approach is to assess the spatial distribution of orbital timescale variability, by plotting the standard deviation for a climate variable for each grid box, as illustrated for SAT in Fig. 9 for the 400 ppmv CO_2 scenario (blue lines in Fig. 10). Figure 9a shows mean annual SAT (compared with pre-industrial) produced by the emulator under modern-day orbital conditions. Anomalies over the majority of the Earth's surface are positive, due to the relatively high atmospheric CO_2 concentration of 400 ppmv. Warming is larger at high latitudes, primarily due to a number of positive feedbacks operating in these regions (known as polar amplification). The greatest warming is centred over parts of the GIS and WAIS, showing a similar spatial pattern to that in Fig. 4, and is a result of the reduced ice sheet extents in the emulated experiments compared with the pre-industrial simulation. Figure 9b shows the ~~difference between modern-day emulated mean annual SAT (Fig. 9a) and emulated mean annual SAT (compared with pre-industrial) averaged over the late Pliocene period (late Pliocene minus modern), whilst the~~ standard deviation of mean annual SAT for the late Pliocene~~-~~, with ~~is presented in Fig. 9c. In both Fig. 9b and 9c,~~ spatial variations primarily illustratinge differences in the impact of orbital forcing on climate. For example, the relatively higher values at high latitudes compared with low latitudes in Fig. 9b~~c~~ suggest that changes in the orbital parameters have a relatively large impact on SAT in these regions. This is consistent with astronomical theory, as changes in both obliquity and precession affect the distribution of insolation in space and time, with this effect being particularly significant at high latitudes. Monsoonal regions also demonstrate relatively large variations (Fig. 9b~~ and 9c~~), including Africa, India, and South America, in agreement with previous studies which suggest a link between orbital changes and monsoon variability (Caley et al., 2011; Prell and Kutzbach, 1987; Tuenter et al., 2003).

In order to visualise the effects of orbital forcing over time, a spectral wavelet analysis was performed on the SAT time series data produced by the emulator, for the scenario with constant CO_2 at 400 ppmv, shown in Fig.

10 (blue line). We used the standard MATLAB wavelet software of Torrence and Compo (1998) (available online at http://atoc.colorado.edu/research/wavelets). The wavelet power spectra for the four ODP/IODP sites are presented in Fig. 11, from which the dominant orbital frequencies influencing climate can be identified. For the late Pliocene up to ~2900 kyr, Fig. 11 suggests that changes in emulated SAT are paced by a combination of precession (longitude of perihelion) and eccentricity, with periodicities of approximately 21 and 96 kyr, respectively. The influence of precession is also supported by Fig. 4c, which demonstrates precessional forcing in the regions where the sites are located, as well as by the frequency of the SAT oscillations for this period shown in Fig. 10, and the observation that it appears to have a larger impact on SAT at higher latitudes (Fig. 10 and 11). After ~2900 kyr, obliquity appears to have an increased impact at the high latitude ~~s~~site 982, superimposing the precession-driven temperature variations with a periodicity of ~41 kyr (Fig. 10 and 11). This signal is also apparent to a lesser extent at Site 722, but not at Site U1313. Spectral analysis of palaeo-proxy data and June insolation at 65° N also finds a reduction in the influence of precession and an increase in 41 kyr obliquity forcing around this time (Herbert et al., 2010; Lawrence et al., 2009). SAT changes at the lower latitude sites generally continue to be dominated by variations in precession and eccentricity, although the relatively low eccentricity during this period is likely to reduce the impact that precession has on climate. It also significantly reduces the variability in temperature, which is also observed during the period of low eccentricity between approximately 3240 and 3200 kyr in both Fig. 10 and 11. The slightly higher amplitudes of the peaks in temperature around 3150 kyr, 3050 kyr and 2950 kyr in Fig. 10 coincide with periods of high eccentricity, when its impact on climate is increased (Fig. 11). It is more difficult to identify orbital trends in the proxy data, particularly in sections with lower resolution. This is due to there being significantly more variation, both on shorter timescales of several tens of thousands of years, and longer timescales of hundreds of thousands of years, likely caused in part by changes in atmospheric CO_2. However, the amplitude of variations in the palaeo data at all four sites is generally, though not always, lower during periods of low eccentricity, particularly for the period ~3225-3200 kyr.

5.3 Calculation of atmospheric CO_2

We also illustrate the use of the emulator for calculating a simple estimate of atmospheric CO_2 concentration during the late Pliocene, and its comparison to published palaeo CO_2 records obtained from proxy data. CO_2 is estimated from the four alkenone SST records presented in Table 4 and Fig. 10: Herbert et al. (2010) (Sites 662 and 722), Naafs et al. (2010) (Site U1313) and Lawrence et al. (2009) (Site 982). Individual records of SST, rather than stacked benthic oxygen isotope data, were used because the GCM experiments that the emulator is calibrated on were only run for 500 years, meaning that deep ocean conditions would not yet have spun-up sufficiently, particularly in the experiments with high CO_2. Thus, it would not be appropriate to compare deep ocean temperatures from- the experiments with those from the proxy data.

A linear regression is performed on the emulated grid box mean annual SAT data versus prescribed atmospheric CO_2 concentration, for ~~the three~~ constant CO_2 scenarios ~~of~~ ranging from 26~~8~~0~~, 350 and 400~~ ppmv up to 800 ppmv. The CO_2 concentration is then estimated from the palaeo SST data based on this linear relationship, and is presented in Fig. 12, along with the uncertainty. A number of CO_2 proxy records are also compared, derived from alkenone data at ODP Site 1241 in the east tropical Pacific (Seki et al., 2010) and Site 999 in the Caribbean

(Badger et al., 2013; Seki et al., 2010), and from boron (δ^{11}B) data at Site 662 (Martinez-Boti et al., 2015) and Site 999 (Bartoli et al., 2011; Martinez-Boti et al., 2015; Seki et al., 2010).

Our model-based CO_2 estimates suggest a mean atmospheric CO_2 concentration for the period of between approximately 350 ± 15~~4~~ and 629~~540~~ ± 3~~1~~7 ppmv (error represents the uncertainty taking into account the emulated g~~i~~rid box posterior variance for SAT), indicated at Sites 722 and 982, respectively. Our CO_2 estimates are generally higher than the CO_2 proxy records, particularly ~~using data fromat~~for the two North Atlantic sites (982 and U1313), where palaeo SST ~~temperatures~~ anomalies were also estimated to be high, compared with tropical ~~SSTs~~anomalies, by the proxy data (Fig. 10). However, CO_2 concentrations derived from SST data calibrated using the approach of Prahl et al. (1988) at the tropical sites of 722 and 662 shows greater similarity to the CO_2 proxy data, both in terms of mean concentration and variance (not shown). It is difficult to identify temporal similarities between our CO_2 estimates and the palaeo records. This is partly due to the high level of variability in our CO_2 time series, resulting from the variability in the SST records that they were derived from. In addition, the CO_2 proxy records have comparatively low resolutions, generally with intervals of 10 kyr or greater, and there is also considerable variation between them.

There is substantial variation between our CO_2 estimates at different sites, and this may be attributed to a number of causes. It could be that there are errors in the GCM model used, in particular in its representation of the response of climate to CO_2 and/or orbital forcing. There could be inaccuracies associated with the SST data at one or more locations as, if the model was assumed to be correct, the estimated CO_2 should be similar across the four locations. The fact that they are not may indicate that the temperature records are not consistent with each other, which may not have been obvious by just comparing the records visually. This is one of the potential advantages to using individual temperature records rather than stacked records. It may also be that there is an issue with the dating of some of the proxy records; the data ~~itself~~ may be correct but there ~~time data associated with ti may be wrong~~may be uncertainties/inaccuracies in the age models. Alternatively, the emulator may be wrong; for example, there may be non-linearities in the climate response simulated by the GCM that it is not capturing. Finally, there may be errors related to the modelled representation of the ice sheets, which are fixed at a constant configuration. In reality, of the possible sources of error that have been identified, the variations are less likely to be the result of errors in the emulator's estimates of the GCM output ~~, since it is based on the output from the GCM and the~~because validation diagnostics did not seem to suggest systematic failures. They are also less likely to be due to unrepresented changes in climate due to the ice sheets. Whilst some of the variation at the high latitude sites (982 and U1313) may be attributed to some regional climate processes not fully accounted for, e.g. involving the ice sheets and sea ice, two of the sites (722 and 662) are in tropical regions. Thus, SSTs at these sites would not be expected to be affected by changes in the ice sheets, and yet they show significantly different variations. Therefore, the inconsistencies are likely to be due to a combination of errors in the GCM model and inaccuracies in the SST data.

6 Application of the emulator to future climate

In addition to using the emulator to model past climates, it can also be applied to future climate, and in particular on the long timescales (>10^3 yr) that are of interest for the disposal of solid radioactive wastes.

Previous modelling of long-term future climate has involved the use of lower complexity models such as EMICs for transient simulations (Archer and Ganopolski, 2005; Eby et al., 2009; Ganopolski et al., 2016; Loutre and Berger, 2000b), or of GCMs to model a relatively small number of snapshot simulations of particular reference climate states of interest. The BIOCLIM (Modelling Sequential Biosphere Systems under Climate Change for Radioactive Waste Disposal) research programme (BIOCLIM, 2001, 2003), for example, utilised both of these approaches to investigate climatic and vegetation changes for the next 200 kyr, for use in performance assessments for radi~~a~~oactive waste disposal facilities.

Here, for the first time, a GCM has been used to project future long-term transient climate evolution, via use of the emulator. We provide illustrations of two possible applications of the emulator, including ~~to~~ production~~e~~ of a time series of climatic data and ~~to~~ assessing the impact of orbital variations on climate. This work has input to the International Atomic Energy Agency (IAEA) MOdelling and DAta for Radiological Impact Assessments (MODARIA) collaborative research programme (http://www-ns.iaea.org/projects/modaria).

6.1 Time series data

Similarly to the late Pliocene, snapshots of SAT and precipitation at 1 kyr intervals were produced using the *modice* emulator for the next 200 kyr, assuming modern day ice sheet configurations. The projected evolution of climate is a result of future variations in the orbital parameters and atmospheric CO_2 concentrations, which were provided as input data to the emulator (again, at 1 kyr intervals). Four CO_2 emissions scenarios were modelled, with the response of atmospheric CO_2 concentration to emissions and its long-term evolution calculated using the impulse response function of Lord et al. (2016). The scenarios adopted logistic CO_2 emissions of 500, 1000, 2000 and 5000 Pg C released over the first few hundred years, followed by a gradual reduction of atmospheric CO_2 concentrations by the long-term carbon cycle. These four scenarios cover the range of emissions that might occur given currently economic and potentially economic fossil fuel reserves, but not including other potentially exploitable reserves, such as clathrates.

Four single grid boxes are selected, shown in Fig. 13, which represent example locations that could potentially be relevant for nuclear waste disposal: Forsmark, Sweden (60.4° N latitude, 18.2° E longitude), Central England, UK (52.0° N latitude, 0° W longitude), Switzerland (47.6° N latitude, 8.7° E longitude) and El Cabril, Spain (38° N latitude, 5.4° W longitude). The evolution of SAT at these grid boxes is presented in Fig. 14, along with the emulated uncertainty (1 standard deviation of the emulated grid box posterior variance)~~1 standard deviation~~). ~~Across the four sites, the~~ In the 500 Pg C scenario, the ~~maximum~~ largest SAT increase of ~~is between~~ 4.1 ± 0.2°C occurs at the ~~(~~Switzerland grid box~~)~~, whilst the Spain grid box exhibits the largest warming in the 5000 Pg C scenario, of ~~and~~ 12.3 ± 0.3°C. ~~(Spain grid box) in the 500 Pg C and 5000 Pg C scenarios, respectively.~~ For comparison, when the *lowice* emulator is utilized, these values are reduced slightly to 3.9 ± 0.3°C (Spain grid box) and 12.2 ± 0.3°C (Spain grid box)~~,~~ in the 500 Pg C and 5000 Pg C scenarios, respectively. This peak in temperature occurs up to the first thousand years, when atmospheric CO_2 is at its highest following the emissions period, after which it decreases relatively rapidly with declining atmospheric CO_2 until around 20 kyr AP. By 200 kyr AP, SAT at all sites is within 2.6°C (2.2°C using the *lowice* emulator) of pre-industrial values, calculated by averaging the final 10 kyr of the 5000 Pg C scenarios. The emulated uncertainty for the next 200 kyr is of a similar magnitude to that for the late Pliocene and, similarly, is relatively small when compared with the fluctuations in SAT that result from orbital variations and changing atmospheric CO_2 concentration.

U~~p u~~ntil ~20 kyr AP, the behaviour of the climate is primarily driven by the high levels of CO_2 in the atmosphere ~~caused by~~ as a result of ~~fossil fuel~~ anthropogenic CO_2 emissions from a range of sources, including combustion of fossil fuels, ~~and other human activities~~land-use change and cement production. However, after this time, changes in orbital conditions begin to exert a relatively greater influence on climate, as the periodic fluctuations in SAT at all locations appear to be paced by the orbital cycles, which are shown in Fig. 14a.

The timing and relative amplitudes of the oscillations in future SAT are in good agreement with a number of previous studies. Paillard (2006) applied the conceptual model of Paillard and Parrenin (2004), previously mentioned in Sect. 5, to the next 1 Ma. The development of atmospheric CO_2 over the next 200 kyr, simulated by the model following emissions of 450 to 5000 Pg C and accounting for natural variations, shows a similar pattern of response to that of SAT presented here. Estimates of global mean temperature in Archer and Ganopolski (2005), derived by scaling changes in modelled ice volume to temperature, before applying anthropogenic CO_2 temperature forcing for a number of emissions scenarios, also demonstrate fluctuations in global mean annual SAT (not shown) of a similar timing and relative scale. The influence of declining CO_2 is still evident after 20 kyr, particularly for the higher emissions scenarios, in the slightly negative gradient of the general evolution of SAT. This is due to the long atmospheric lifetime of ~~fossil fuel~~ CO_2 emissions (Archer, 2005), and is also demonstrated in other studies (Archer and Ganopolski, 2005; Archer et al., 2009; Lord et al., 2016; Paillard, 2006). The impact of excess atmospheric CO_2 on the long-term evolution of SAT appears to be fairly linear, with only minor differences between the scenarios and sites, discounting the overall offset of SAT for different total emissions.

One of the key uncertainties associated with future climate change, which is of particular relevance to radioactive waste repositories located at high northern latitudes, is the timing of the next glacial inception. This is expected to occur during a period of relatively low incoming solar radiation at high northern latitudes, which, for the next 100 kyr, occurs at 0 kyr, 54 kyr and 100 kyr. A number of studies have investigated the possible timing of the next glaciation under pre-industrial atmospheric CO_2 concentrations (280 ppmv), finding that it is unlikely to occur until after 50 kyr AP (Archer and Ganopolski, 2005; Berger and Loutre, 2002; Paillard, 2001).

When ~~fossil fuel~~anthropogenic CO_2 emissions are taken into account, the current interglacial is likely to last significantly longer, until ~130 kyr AP following emissions of 1000 Pg C and beyond 500 kyr AP for emissions of 5000 Pg C (Archer and Ganopolski, 2005). A recent study by Ganopolski et al. (2016) using the CLIMBER-2 model found that emissions of 1000 Pg C significantly reduced the probability of a glaciation in the next 100 kyr, and that a glacial inception within the next 100 kyr is very unlikely for CO_2 emissions of 1500 Pg C or higher.

Our CO_2 emissions scenarios, modelled using the response function of Lord et al. (2016), suggest that atmospheric CO_2 will not have returned to pre-industrial levels by 100 ~~ka~~ kyr AP, equalling 298 and 400 ppmv for the 500 and 5000 Pg C emissions scenarios, respectively. We calculated the critical summer insolation threshold at 65° N using the logarithmic relationship identified between maximum summer insolation at 65° N and atmospheric CO_2 by Ganopolski et al. (2016). The evolution of atmospheric CO_2 concentration over the course of our emissions scenarios suggests that, for emissions of 1000 Pg C or less, Northern Hemisphere summer insolation will next fall below the critical insolation threshold in approximately 50 kyr~~a~~, and in ~100 ~~ka~~ kyr for emissions of 2000 Pg C. For the highest emissions scenario of 5000 Pg C, the threshold is not passed for considerably longer, until ~160 kyr~~a~~. However, the uncertainty of the critical insolation value is ±4 W m^{-2} (1 standard deviation), and often the difference between summer insolation at 65° N and the insolation threshold is less than this, potentially impacting whether the threshold has in fact been passed and therefore whether glacial inception is likely. For example, for the 1000 Pg C scenario, whilst insolation first falls below the critical threshold at ~50 k~~a~~yr, it does not fall below by more than the uncertainty value until ~130 kyr~~a~~.

A limitation of our study relates to the continental ice sheets in HadCM3 being prescribed rather than responsive to changes in climate. A consequence of this is that an increase in the extent or thickness of the ice sheets, and hence the onset of glaciation, cannot be explicitly projected, but this also means that a regime shift of the ice sheets to one of negative mass balance, which may be expected to occur under high CO_2 emissions scenarios (Ridley et al., 2005; Stone et al., 2010; Swingedouw et al., 2008; Winkelmann et al., 2015), cannot be modelled. However, the results of the sensitivity analysis to ice sheets described in Sect. 3.5., for which a number of simulations were run again with reduced GIS and WAIS extents, suggest that the reduction in continental ice results in relatively localised increases in SAT in regions that are ice free, in addition to some regional cooling at high latitudes. Consequently, this does not act as a significant restriction on the glaciation timings put forward in this study considering their radioactive waste disposal application; given that the earliest timing of the next glaciation is of significant interest, smaller continental ice sheets and therefore higher local SATs would likely inhibit the build-up of snow and ice, delaying glacial inception further. As such, the estimates presented here should be viewed as conservative. As will be discussed in Sect. 7, however, the emulator was not designed and calibrated to predict changes in ice sheets. This is a limitation that should be addressed when modelling future climate on timescales of tens of thousands of years or more (depending on the CO_2 scenario(s) being modelled). Another caveat is that the carbon cycle in the emulator is also essentially prescribed, and thus not interactive. This means that the atmospheric CO_2 trajectory follows a smooth decline, as was projected using an impulse response function based on experiments using the *c*GENIE model (Lord et al., 2016), with long-term future climate being modelled as a series of snapshot simulations with the emulator. This smooth decline in CO_2 assumes that no non-linear or unexpected behaviour will be demonstrated by the long-term carbon cycle, and that the silicate weathering mechanism, which is associated with a substantial degree of uncertainty, is correct.

The emulator can also be used to project the evolution of a range of other climate variables, providing that they were modelled as part of the initial GCM ensembles. Figure 15 illustrates the development of mean annual precipitation and emulated uncertainty over the next 200 kyr at the four sites. The maximum increase in precipitation is between 0.3 ± 0.1 mm day^{-1} (Switzerland grid box) and 0.6 ± 0.1 mm day^{-1} (Sweden grid box) in the 500 Pg C and 5000 Pg C scenarios, respectively. Precipitation increases with increasing atmospheric CO_2 at all sites apart from the Spain grid box, where it decreases by up to 0.9 ± 0.1 mm day^{-1}. Regional differences in the sign of changes in precipitation, including an increase at high latitudes and a decrease in the Mediterranean, are consistent with modelling results included in the International Panel on Climate Change (IPCC) Fifth Assessment Report, for simulations forced with the Representative Concentration Pathway (RCP) 8.5 scenario (Collins et al., 2013). In contrast to SAT, precipitation appears to be more closely influenced by precession, illustrated by its periodicity of slightly less than 25 kyr. There appears to be~~;~~ an increase in the intensity of precipitation fluctuations from approximately 140 kyr onwards, ~~suggest~~ implying that the modulation of precession by eccentricity also has an impact~~, as expected~~.

6.2 Orbital variability and spectral analysis

The impact of orbital forcing was assessed by performing a spectral wavelet analysis on the SAT and precipitation time series data produced by the emulator for the Central England grid box for the 5000 Pg C emissions scenario, represented by blue lines in Fig. 14c and 15c, respectively. As for the late Pliocene, the wavelet software of Torrence and Compo (1998) was utilized. The analysis was performed on the data for 20-200 kyr AP, because the climate response up until ~20 kyr AP is dominated by the impact of elevated atmospheric CO_2 concentrations, which masks the orbital signal and affects the results of the wavelet analysis.

For future SAT, Fig. 16a suggests that, up until ~160 kyr, the obliquity cycle acts as the dominant influence, resulting in temperature oscillations with a periodicity of approximately 41 kyr. This is confirmed by Fig. 14c, which shows that the major peaks in SAT generally coincide with periods of high obliquity. Over this period, precession has a far more limited influence, likely due to eccentricity being relatively low until ~110 kyr (Fig. 14a). However, from ~120 kyr AP onwards, concurrently with increasing eccentricity, precession becomes a more significant forcing on climate, resulting in SAT peaks approximately every 21 kyr. In contrast, precession appears to be the dominant forcing on precipitation for the Central England grid box for the entire 20-200 kyr AP period (Fig. 15c and 16b). This signal is particularly strong after ~120 kyr AP, due to higher eccentricity.

7 Limitations

There are a number of limitations associated with the methodology outlined above, ~~emulator,~~ particularly relating to the assumptions that it is based on and its application to different periods of time. Although these have mostly been discussed briefly in the preceding sections, here we summarise them together.~~,~~

- A~~Firstly, a~~s noted previously, the carbon cycle in the emulator is not coupled to the climate, ~~essentially fixed~~, since the atmospheric CO_2 concentration is prescribed. ~~It~~ The methodology thus assumes that there will be no unexpected non-linearities in the carbon cycle, and that changes in climate that are different from those in *c*GENIE do not feed back to the carbon cycle. This may be of particular importance when simulating future climates, when the natural carbon cycle is expected to be significantly perturbed due to ongoing anthropogenic emissions of CO_2, in a way that may not be fully represented in *c*GENIE. There is also uncertainty surrounding the dynamics of the carbon cycle over long periods of time, such as the role of the silicate weathering mechanism, although the observation that different carbon cycle models generally produce fairly similar results increases our confidence (Archer et al., 2009).

- T~~Secondly, t~~he ice sheets in the emulator are also fixed, at either modern–-day or reduced extents, although expanding the range of ice sheets that can be modelled is currently being undertaken in ongoing research. This means that care needs to be taken when simulating very long periods of time. For example, neither Quaternary nor future~~the~~ glacial-interglacial cycles can~~not~~ be ~~accounted for~~~~captured~~simulated using the current version of the emulator~~, which are known to have occurred in the past (e.g. Petit et al., 1999) and are expected to continue in the future~~. Furthermore, even d~~D~~uring the Pliocene, it is likely that the extent of ice sheets in the Northern Hemisphere varied beyond the range simulated in this study (Willeit et al., 2015), and the emulator in its current form cannot represent this.

- In the context of the ~~Plicoene~~Pliocene, t~~T~~he land-sea mask and orography used in the simulation of Pliocene climate are ~~also~~ fixed and appropriate to modern-day conditions, whereas the PRISM4 reconstruction of paleogeography~~Pliocene ice sheets~~ suggests that there~~y~~ may have been considerably different in some regions, for example the region of the Hudson Bay is thought to have been land in the Pliocene (Dowsett et al., 2016).

- Due to both the carbon cycle and ice sheets being prescribed, interactions between these components of the climate system can also not be simulated. These include natural changes in CO_2 which have been found to accompany past glacial-interglacial cycles, with glacial periods over the last 800 kyr exhibiting CO_2 concentrations of approximately 180 to 200 ppmv (Petit et al., 1999), whereas interglacial periods demonstrated concentrations of 240 to 290 ppmv (Luthi et al., 2008). Changes in the ice sheets in response to atmospheric CO_2 can also not be modelled, such as the likely future melting of the GIS and AIS in response to anthropogenic CO_2 emissions. Various studies have modelled the response of the ice sheets to future climate warming, finding that the ice sheets may experience significantly increased melt. In fact, for scenarios with high CO_2 emissions (>~5000 Pg C), it has been suggested that the GIS and AIS may be almost entirely melted within the next few thousand years (e.g. DeConto and Pollard, 2016; Huybrechts et al., 2011; Winkelmann et al., 2015), which would cause significant changes in deep ocean circulation and ocean stratification. These ocean changes cannot be captured by the current version of the methodology~~emulator~~ and, whilst their impacts on global and regional climate are uncertain, they are expected to be long-term. The melting of the ice sheets would also cause significant increases in global sea level, of approximately 70 m if both the GIS and AIS melted, which would strongly affect the global land-sea mask and regional climates, and which cannot be represented using the current methodology. This sea level rise would also have serious implications for radioactive waste repositories located in relatively low-lying coastal regions that are vulnerable to sea level rise, such as in northern Europe.
- Since the emulator models climate via a series of snapshots rather than a truly transient simulation, it is not able to capture deviations fro~~r~~m a stationary response~~trends in deep ocean conditions~~. As a consequence, the methodology becomes inappropriate if such transient changes in the deep ocean are found to be important for controlling surface climate evolution.

- The emulator presented in this study is only suitable for modelling transient climate changes on timescales of several millennia or longer, as a number of shorter-term processes in the climate and carbon cycle are not represented. These include internal variability in the climate system, such as interannual variability, North Atlantic Oscillation (NAO), and El Niño – Southern Oscillation (ENSO), as well as radiative forcing occurring on shorter timescales (e.g. volcanic activity), and terrestrial carbon cycle processes. On these timescales, transient simulations run using ~~fully~~ complex models such as GCMs or EMICS are most appropriate.

As a consequence of these limitations, care needs to be taken when applying the emulator to ensure that its application is appropriate. For example, ~~Also,~~ when considering future climate, the way in which future carbon dioxide concentration have been modelled, and the ice sheet configurations modelled, mean that this methodology is only applicable on timescales up until~~ntil~~ the next glacial inception. After this, atmospheric CO_2 would be expected to change in response to the initiation of glacial conditions, accompanied by the expansion of the ice sheets, decreasing sea level, and the climatic changes that would results from these changes. A number of studies have modelled the possible timing of the next glacial inception, finding that for CO_2 scenarios with medium emissions the current interglacial period may end in approximately 130 kyr (Archer and Ganopolski, 2005; Ganopolski et al., 2016). However, for high emissions of 5000 Pg C, glacial inception may be delayed for more than 500 kyr (Archer and Ganopolski, 2005). A study by Brandefelt et al. (2013) estimated that for permafrost development to occur at Forsmark, Sweden during the insolation minima at 17 and 54 kyr AP, atmospheric CO_2 concentrations of ~210 ppmv or less and ~250 ppmv or less would be required, respectively. In light of the long atmospheric lifetime of CO_2 emissions that has been discussed, low concentrations such as these are unlikely in the next few tens of thousands of years; however, they cannot be entirely excluded. In order to account for this limitation, the emulator could be extended to include glacial states, meaning that it could be applied to future climate on a longer timescale, as well as to the Quaternary, if the evolution of CO_2 and ice volume were known (e.g. from a transient EMIC of conceptual model simulation). Thus, when emulating long-term climate, careful consideration should be given to what assumptions are being made and whether the methodology is appropriate for the conditions being modelled.

Bearing in mind these limitations, ~~we believe~~ the methodology described in this paper ~~emulator~~is ~~to be~~ a useful and powerful tool for simulating long-term past and future climatic changes, as well as for exploring the dynamics and sensitivities of the climate system.

8~~7~~ Summary and Conclusions

In this study, we present long-term continuous projections of future climate evolution at the spatial resolution of a GCM, via the use of a statistical emulator. The emulator was calibrated on two ensembles of simulations with varied orbital and atmospheric CO_2 conditions and modern day continental ice sheet extents, produced using the HadCM3 climate model. The method presented by Gregory et al. (2004) to calculate the steady-state global temperature change for a simulation, by regressing the net radiative flux at the top of the atmosphere against the change in global SAT, was utilised to calculate the equilibrated SAT data for these ensembles, as it was not feasible to run the experiments to equilibrium due to the associated time and computer resources needed. A number of simulations testing the sensitivity of SAT to the extent of the GIS and WAIS suggest that the response of SAT is fairly linear regardless of orbit, and that the largest changes are generally local to regions that are ice free. The mean SAT anomaly identified across these experiments was then applied to the equilibrated SAT results of the modern-day ice sheet extent ensembles, to generate two equivalent ensembles with reduced ice sheets.

Output data from the modern-day and reduced ice sheet ensembles were then used to calibrate separate emulators, which were optimised and then validated using a leave-one-out approach, resulting in satisfactory performance results. We discuss a number of useful applications of the emulator, which may not be possible using other modelling approaches at the same temporal and spatial resolution. Firstly, a particular benefit of the emulator is that it can be used to produce time series of climatic variables that cover long periods of time (i.e. several thousand years or more) at a GCM resolution, accompanied by an estimation of the uncertainty in the form of the posterior variance. This would not be feasible using GCMs due to the significant time and computational requirements involved. The global grid coverage of the data also means that the evolution of a climate variable at a particular grid box can be examined, allowing for comparisons to data at a regional or local scale, such as palaeo-proxy data, or for the evolution of climate at a specific site to be studied. However, further downscaling of the data may also be necessary or beneficial, via further modelling such as proxy modelling, impact models or regional climate models, or via statistical downscaling techniques. Secondly, the influence of orbital forcing on climate can be assessed. This effect may be visualised with a continuous wavelet analysis on the time series data for a particular CO_2 emissions scenario, which will identify the orbital frequencies dominating at different times. The spatial distribution of orbital timescale variability can also be simulated, by plotting the standard deviation for a climate variable for each grid box, taking into account the emulator posterior variance. Finally, the emulator can be used to back-calculate past atmospheric CO_2 concentrations based on proxy climate data. Through an inversion, atmospheric CO_2 concentrations can be estimated using SST proxy data, based on a linear relationship between emulated grid box mean annual SAT and prescribed CO_2 concentration. Estimated CO_2 can then be compared with palaeo CO_2 concentration proxy records.

To illustrate these potential applications, we applied the emulator at 1 kyr intervals to the late Pliocene (3300-2800 kyr BP) for atmospheric CO_2 concentrations of 280, 350 and 400 ppmv, and compared the emulated SATs at specific grid boxes to SSTs determined from proxy data from a number of ODP/IODP sites. The wavelet power spectrum for SAT at each site was also produced, and the dominant orbital frequency assessed. In addition, we used the SST proxy data to estimate atmospheric CO_2 concentrations, based on a linear relationship between emulated grid box mean annual SAT and prescribed CO_2 concentration. We find that:

- Temperature estimates from the emulator and proxy data show greater similarity at the equatorial sites than at the high latitude sites. Discrepancies may be the result of ~~biases in the GCM, errors in the emulator,~~ seasonal biases in the proxy data, unknown changes in the climate and/or carbon cycle,~~ or~~ issues with the tuning of parts of the record, biases in the GCM, or errors in the emulator~~,~~.

- The response of emulated SAT appears to be dominated by a combination of precessional and eccentricity forcing from 3300 kyr to approximately 2900 kyr, after which obliquity begins to have an increased influence.
- Regions with a particularly large response to orbital forcing include the high latitudes and monsoon regions (Fig. 9b~~ and 9c~~).
- Our CO_2 ~~reconstructions~~ concentrations derived from tropical ODP/IODP sites show relatively similar concentrations to CO_2 proxy records for the same period, although ~~for~~ the ~~higher latitude sites~~ concentrations derived from higher latitude sites are generally significantly higher than the proxy data.

The emulator was also applied to the next 200 kyr, as long-term future simulations such as these have relevance to the geological disposal of solid radioactive wastes. The continuous evolution of mean annual SAT and precipitation at a number of sites in Europe are presented, for four scenarios with ~~fossil fuel~~anthropogenic CO_2 emissions of 500, 1000, 2000 and 5000 Pg C. A spectral wavelet analysis was also performed on the SAT and precipitation data for the Central England grid box. The data suggests that:

- SAT and, to a lesser extent, precipitation exhibit a relatively rapid decline back towards pre-industrial values over the next 20 kyr, as excess atmospheric CO_2 is removed by the long-term carbon cycle.
- Following this, SAT fluctuates due to orbital forcing on an approximate 41 kyr obliquity timescale until ~160 kyr AP, before the influence of precession increases with increasing eccentricity from ~120 kyr AP.
- Conversely, precipitation variations over the entire 200 kyr period demonstrate a strong precessional signal.

Overall, we find that the emulator provides a useful and powerful tool for rapidly simulating the long-term evolution of climate, both past and future, due to its relatively high spatial resolution and relatively low computational cost. We have presented illustrative examples of a number of different possible applications, which we believe make it suitable for tackling a wide range of climate questions.

Code availability

Code for the Latin hypercube sampling function is available from the MATLAB Statistics and Machine Learning Toolbox. The wavelet software of Torrence and Compo (1998) is available online at http://atoc.colorado.edu/research/wavelets.

Data availability

The data used in this paper are available from Natalie S. Lord (Natalie.Lord@bristol.ac.uk).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research is funded by RWM Limited via a framework contract with Amec Foster Wheeler, who are being supported by Quintessa. It contributes to the MODARIA international research programme, sponsored and coordinated by the International Atomic Energy Agency (IAEA). The ensembles of AOGCM simulations were run using the computational facilities of the Advanced Computing Research Centre, University of Bristol – http://www.bris.ac.uk/acrc/. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

References

Andrews, T., Gregory, J. M., and Webb, M. J.: The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models, J Climate, 28, 1630-1648, doi: 10.1175/Jcli-D-14-00545.1, 2015.

Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys Res Lett, 39, L09712, doi: 10.1029/2012gl051607, 2012.

Andrianakis, I. and Challenor, P. G.: The effect of the nugget on Gaussian process emulators of computer models, Comput Stat Data An, 56, 4215-4228, doi: 10.1016/j.csda.2012.04.020, 2012.

Araya-Melo, P. A., Crucifix, M., and Bounceur, N.: Global sensitivity analysis of the Indian monsoon during the Pleistocene, Clim Past, 11, 45-61, doi: 10.5194/cp-11-45-2015, 2015.

Archer, D.: Fate of fossil fuel CO_2 in geologic time, J Geophys Res-Oceans, 110, doi: 10.1029/2004jc002625, 2005.

Archer, D. and Ganopolski, A.: A movable trigger: Fossil fuel CO_2 and the onset of the next glaciation, Geochem Geophy Geosy, 6, Q05003, doi: 10.1029/2004gc000891, 2005.

Archer, D., Kheshgi, H., and Maier-Reimer, E.: Multiple timescales for neutralization of fossil fuel CO_2, Geophys Res Lett, 24, 405-408, doi: 10.1029/97gl00168, 1997.

Archer, D., Eby, M., Brovkin, V., Ridgwell, A., Cao, L., Mikolajewicz, U., Caldeira, K., Matsumoto, K., Munhoven, G., Montenegro, A., and Tokos, K.: Atmospheric lifetime of fossil fuel carbon dioxide, Annu Rev Earth Pl Sc, 37, 117-134, doi: 10.1146/annurev.earth.031208.100206, 2009.

Armour, K. C., Bitz, C. M., and Roe, G. H.: Time-Varying Climate Sensitivity from Regional Feedbacks, J Climate, 26, 4518-4534, doi: 10.1175/Jcli-D-12-00544.1, 2013.

Armstrong, E., Valdes, P., House, J., and Singarayer, J.: The role of CO_2 and dynamic vegetation on the impact of temperate land-use change in the HadCM3 coupled climate model, Earth Interact, 20, 10, doi: 10.1175/Ei-D-15-0036.1, 2016.

Badger, M. P. S., Schmidt, D. N., Mackensen, A., and Pancost, R. D.: High-resolution alkenone palaeobarometry indicates relatively stable pCO_2 during the Pliocene (3.3-2.8 Ma), Philos T R Soc A, 371, 20130094, doi: 10.1098/rsta.2013.0094, 2013.

Bamber, J. L., Riva, R. E. M., Vermeersen, B. L. A., and LeBrocq, A. M.: Reassessment of the potential sea-level rise from a collapse of the West Antarctic ice sheet, Science, 324, 901-903, doi: 10.1126/science.1169335, 2009.

Bartoli, G., Honisch, B., and Zeebe, R. E.: Atmospheric CO_2 decline during the Pliocene intensification of Northern Hemisphere glaciations, Paleoceanography, 26, Pa4213, doi: 10.1029/2010pa002055, 2011.

Bastos, L. S. and O'Hagan, A.: Diagnostics for Gaussian Process Emulators, Technometrics, 51, 425-438, doi: 10.1198/Tech.2009.08019, 2009.

Berger, A.: Long-term variations of daily insolation and Quaternary climatic changes, J Atmos Sci, 35, 2362-2367, doi: 10.1175/1520-0469(1978)035<2362:Ltvodi>2.0.Co;2, 1978.

Berger, A. and Loutre, M. F.: An exceptionally long interglacial ahead?, Science, 297, 1287-1288, doi: 10.1126/science.1076120, 2002.

Berger, J. O., De Oliveira, V., and Sanso, B.: Objective Bayesian analysis of spatially correlated data, J Am Stat Assoc, 96, 1361-1374, doi: 10.1198/016214501753382282, 2001.

BIOCLIM: Deliverable D3: Global climatic features over the next million years and recommendation for specific situations to be considered, Agence Nationale pour la Gestion des Dechets Radioactifs (ANDRA), Parc de la Croix Blanche, 1/7 rue Jean Monnet, 92298, Châtenay-Malabry, France, Available from: www.andra.fr/bioclim/pdf/d3.pdf, 2001.

BIOCLIM: Deliverable D4/5: Global climatic characteristics, including vegetation and seasonal cycles over Europe, for snapshots over the next 200,000 years, Agence Nationale pour la Gestion des Dechets Radioactifs (ANDRA), Parc de la Croix Blanche, 1/7 rue Jean Monnet, 92298, Châtenay-Malabry, France, Available from: www.andra.fr/bioclim/pdf/d45.pdf, 2003.

Bosmans, J. H. C., Drijfhout, S. S., Tuenter, E., Hilgen, F. J., and Lourens, L. J.: Response of the North African summer monsoon to precession and obliquity forcings in the EC-Earth GCM, Clim Dynam, 44, 279-297, doi: 10.1007/s00382-014-2260-z, 2015.

Bounceur, N., Crucifix, M., and Wilkinson, R. D.: Global sensitivity analysis of the climate-vegetation system to astronomical forcing: an emulator-based approach, Earth Syst Dynam, 6, 205-224, doi: 10.5194/esd-6-205-2015, 2015.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J. Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C. D., Kageyama, M., Kitoh, A., Laine, A., Loutre, M. F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., Yu, Y., and Zhao, Y.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features, Clim Past, 3, 261-277, doi, 2007.

Brandefelt, J., Näslund, J.-O., Zhang, Q., and Hartikainen, J.: The potential for cold climate and permafrost in Forsmakr in the next 60,000 years, SKB Report, Svensk Kärnbränslehantering AB, Stockholm, Sweden, TR-13-04. Available from: www.skb.com/publication/2652151/TR-13-04.pdf, 2013.

Caley, T., Malaize, B., Revel, M., Ducassou, E., Wainer, K., Ibrahim, M., Shoeaib, D., Migeon, S., and Marieu, V.: Orbital timing of the Indian, East Asian and African boreal monsoons and the concept of a 'global monsoon', Quaternary Sci Rev, 30, 3705-3715, doi: 10.1016/j.quascirev.2011.09.015, 2011.

Charbit, S., Paillard, D., and Ramstein, G.: Amount of CO_2 emissions irreversibly leading to the total melting of Greenland, Geophys Res Lett, 35, L12503, doi: 10.1029/2008gl033472, 2008.

Colbourn, G., Ridgwell, A., and Lenton, T.: The time scale of the silicate weathering negative feedback on atmospheric CO_2, Global Biogeochem Cy, 29, 583-596, doi: 10.1002/2014GB005054, 2015.

Collins, M., Knutti, R., Arblaster, J. M., Dufresne, J. L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., and Wehner, M.: Long-term climate change: projections, commitments and irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (Eds.), Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184-187, doi: 10.1038/35041539, 2000.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Modelling vegetation and the carbon cycle as interactive elements of the climate system. In: Meteorology at the Millennium, Pearce, R. (Ed.), Academic Press, San Diego CA, USA, 2002.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Clim Dynam, 15, 183-203, doi: 10.1007/s003820050276, 1999.

Crucifix, M., Braconnot, P., Harrison, S. P., and Otto-Bliesner, B.: Second phase of paleoclimate modelling intercomparison project, Eos, Transactions American Geophysical Union, 86, 264-264, doi: 10.1029/2005EO280003, 2005.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591-597, doi: 10.1038/nature17145, 2016.

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K., and Haywood, A.: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction, Clim Past, 12, 1519-1538, doi: 10.5194/cp-12-1519-2016, 2016.

Dowsett, H. J.: The PRISM palaeoclimate reconstruction and Pliocene sea-surface temperature. In: Deep-time perspectives on climate change: marrying the signal from computer models and biological proxies, Williams, M., Haywood, A. M., Gregory, J., and Schmidt, D. N. (Eds.), Micropalaeontological Society (Special Publication), Geol. Soc., London, UK, 2007.

Dowsett, H. J. and Robinson, M. M.: Mid-Pliocene equatorial Pacific sea surface temperature reconstruction: a multi-proxy perspective, Philos T R Soc A, 367, 109-125, doi: 10.1098/rsta.2008.0206, 2009.

Dowsett, H. J., Haywood, A. M., Valdes, P. J., Robinson, M. M., Lunt, D. J., Hill, D., Stoll, D. K., and Foley, K. M.: Sea surface temperatures of the mid-Piacenzian Warm Period: A comparison of PRISM3 and HadCM3, Palaeogeogr Palaeocl, 309, 83-91, doi: 10.1016/j.palaeo.2011.03.016, 2011.

Eby, M., Zickfeld, K., Montenegro, A., Archer, D., Meissner, K. J., and Weaver, A. J.: Lifetime of anthropogenic climate change: millennial time scales of potential CO_2 and surface temperature perturbations, J Climate, 22, 2501-2511, doi: 10.1175/2008jcli2554.1, 2009.

Feldmann, J. and Levermann, A.: Collapse of the West Antarctic ice sheet after local destabilization of the Amundsen Basin, P Natl Acad Sci USA, 112, 14191-14196, doi: 10.1073/pnas.1512482112, 2015.

Ganopolski, A., Winkelmann, R., and Schellnhuber, H. J.: Critical insolation-CO_2 relation for diagnosing past and future glacial inception, Nature, 529, 200-203, doi: 10.1038/nature16494, 2016.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivie, D. J. L., and Tyteca, S.: Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs, J Climate, 26, 1859-1876, doi: 10.1175/Jcli-D-12-00196.1, 2013.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, Clim Dynam, 16, 147-168, doi: 10.1007/s003820050010, 2000.

Gregory, J. M., Andrews, T., and Good, P.: The inconstancy of the transient climate response parameter under increasing CO_2, Philos T R Soc A, 373, 20140417, doi: 10.1098/rsta.2014.0417, 2015.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys Res Lett, 31, L03205, doi: 10.1029/2003gl018747, 2004.

Greve, R.: On the response of the Greenland ice sheet to greenhouse climate change, Climatic Change, 46, 289-303, doi: 10.1023/A:1005647226590, 2000.

Hays, J. D., Imbrie, J., and Shackleton, N. J.: Variations in the earth's orbit: Pacemaker of the Ice Ages, Science, 194, 1121-1132, doi, 1976.

Haywood, A. M. and Valdes, P. J.: Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere, Earth Planet Sc Lett, 218, 363-377, doi: 10.1016/S0012-821x(03)00685-X, 2004.

Haywood, A. M., Dowsett, H. J., Otto-Bliesner, B., Chandler, M. A., Dolan, A. M., Hill, D. J., Lunt, D. J., Robinson, M. M., Rosenbloom, N., Salzmann, U., and Sohl, L. E.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1), Geosci Model Dev, 3, 227-242, doi: 10.5194/gmd-3-227-2010, 2010.

Haywood, A. M., Dowsett, H. J., Dolan, A. M., Rowley, D., Abe-Ouchi, A., Otto-Bliesner, B., Chandler, M. A., Hunter, S. J., Lunt, D. J., Pound, M., and Salzmann, U.: The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: scientific objectives and experimental design, Clim Past, 12, 663-675, doi: 10.5194/cp-12-663-2016, 2016.

Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F., Chan, W. L., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Abe-Ouchi, A., Pickering, S. J., Ramstein, G., Rosenbloom, N. A., Salzmann, U., Sohl, L., Stepanek, C., Ueda, H., Yan, Q., and Zhang, Z.: Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project, Clim Past, 9, 191-209, doi: 10.5194/cp-9-191-2013, 2013.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. R., and Vallis, G. K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, J Climate, 23, 2418-2427, doi: 10.1175/2009jcli3466.1, 2010.

Herbert, T. D., Peterson, L. C., Lawrence, K. T., and Liu, Z. H.: Tropical ocean temperatures over the past 3.5 million years, Science, 328, 1530-1534, doi: 10.1126/science.1185435, 2010.

Holden, P. B., Edwards, N. R., Oliver, K. I. C., Lenton, T. M., and Wilkinson, R. D.: A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, Clim Dynam, 35, 785-806, doi: 10.1007/s00382-009-0630-8, 2010.

Huybrechts, P. and de Wolde, J.: The dynamic response of the Greenland and Antarctic ice sheets to multiple-century climatic warming, J Climate, 12, 2169-2188, doi: 10.1175/1520-0442(1999)012<2169:Tdrotg>2.0.Co;2, 1999.

Huybrechts, P., Goelzer, H., Janssens, I., Driesschaert, E., Fichefet, T., Goosse, H., and Loutre, M. F.: Response of the Greenland and Antarctic Ice Sheets to Multi-Millennial Greenhouse Warming in the Earth System Model of Intermediate Complexity LOVECLIM, Surv Geophys, 32, 397-416, doi: 10.1007/s10712-011-9131-5, 2011.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (Eds.), Cambridge University Press, Cambridge, UK and New York, USA, 2013.

Joseph, V. R. and Hung, Y.: Orthogonal-maximin Latin hypercube designs, Stat Sinica, 18, 171-186, doi, 2008.

Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., Minster, B., Nouet, J., Barnola, J. M., Chappellaz, J., Fischer, H., Gallet, J. C., Johnsen, S., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, F., Raisbeck, G., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J. P., Stenni, B., Stocker, T. F., Tison, J. L., Werner, M., and Wolff, E. W.: Orbital and millennial Antarctic climate variability over the past 800,000 years, Science, 317, 793-796, doi: 10.1126/science.1141038, 2007.

Kawamura, K., Parrenin, F., Lisiecki, L., Uemura, R., Vimeux, F., Severinghaus, J. P., Hutterli, M. A., Nakazawa, T., Aoki, S., Jouzel, J., Raymo, M. E., Matsumoto, K., Nakata, H., Motoyama, H., Fujita, S., Goto-Azuma, K., Fujii, Y., and Watanabe, O.: Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years, Nature, 448, 912-914, doi: 10.1038/Nature06015, 2007.

Kennedy, M. C. and O'Hagan, A.: Predicting the output from a complex computer code when fast approximations are available, Biometrika, 87, 1-13, doi: 10.1093/biomet/87.1.1, 2000.

Kennett, J. P. and Stott, L. D.: Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Paleocene, Nature, 353, 225-229, doi: 10.1038/353225a0, 1991.

Knutti, R. and Rugenstein, M. A. A.: Feedbacks, climate sensitivity and the limits of linear models, Philos T R Soc A, 373, 20150146, doi: 10.1098/rsta.2015.0146, 2015.

Lambeck, K., Yokoyama, Y., Johnston, P., and Purcell, A.: Global ice volumes at the Last Glacial Maximum and early Lateglacial, Earth Planet Sc Lett, 190, 275-275, doi: 10.1016/S0012-821x(01)00386-7, 2001.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A. C. M., and Levrard, B.: A long-term numerical solution for the insolation quantities of the Earth, Astron Astrophys, 428, 261-285, doi: 10.1051/0004-6361:20041335, 2004.

Lawrence, K. T., Herbert, T. D., Brown, C. M., Raymo, M. E., and Haywood, A. M.: High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period, Paleoceanography, 24, Pa2218, doi: 10.1029/2008pa001669, 2009.

Lenton, T. M. and Britton, C.: Enhanced carbonate and silicate weathering accelerates recovery from fossil fuel CO_2 perturbations, Global Biogeochem Cy, 20, GB3009, doi: 10.1029/2005gb002678, 2006.

Lenton, T. M., Williamson, M. S., Edwards, N. R., Marsh, R., Price, A. R., Ridgwell, A. J., Shepherd, J. G., Cox, S. J., and Team, T. G.: Millennial timescale carbon cycle and climate change in an efficient Earth system model, Clim Dynam, 26, 687-711, doi: 10.1007/s00382-006-0109-9, 2006.

Li, C., von Storch, J. S., and Marotzke, J.: Deep-ocean heat uptake and equilibrium climate response, Clim Dynam, 40, 1071-1086, doi: 10.1007/s00382-012-1350-z, 2013.

Lisiecki, L. E. and Raymo, M. E.: A Pliocene-Pleistocene stack of 57 globally distributed benthic d18O records, Paleoceanography, 20, PA1003, doi: 10.1029/2004pa001071, 2005.

Lisiecki, L. E. and Raymo, M. E.: Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics, Quaternary Sci Rev, 26, 56-69, doi: 10.1016/j.quascirev.2006.09.005, 2007.

LLWR: Environmental Safety Case - Main Report, LLW Repository Limited, LLWR/ESC/R(11) 10016. Available from: www.llwrsite.com/wp-content/uploads/2013/04/Environmental-Safety-Case-%E2%80%93-Full-Report.pdf, 2011.

Loeppky, J. L., Sacks, J., and Welch, W. J.: Choosing the sample size of a computer experiment: A practical guide, Technometrics, 51, 366-376, doi: 10.1198/Tech.2009.08040, 2009.

Lord, N. S., Ridgwell, A., Thorne, M. C., and Lunt, D. J.: The 'long tail' of anthropogenic CO_2 decline in the atmosphere and its consequences for post-closure performance assessments for disposal of radioactive wastes, Mineralogical Magazine, 79, 1613-1623, doi: 10.1180/minmag.2015.079.6.37, 2015.

Lord, N. S., Ridgwell, A., Thorne, M. C., and Lunt, D. J.: An impulse response function for the "long tail" of excess atmospheric CO_2 in an Earth system model, Global Biogeochem Cy, 30, 2-17, doi: 10.1002/2014gb005074, 2016.

Loutre, M. F.: Paramètres orbitaux et cycles diurne et saisonnier des insolations, PhD, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 1993.

Loutre, M. F. and Berger, A.: No glacial-interglacial cycle in the ice volume simulated under a constant astronomical forcing and a variable CO_2, Geophys Res Lett, 27, 783-786, doi: 10.1029/1999gl006081, 2000a.

Loutre, M. F. and Berger, A.: Future climatic changes: Are we entering an exceptionally long interglacial?, Climatic Change, 46, 61-90, doi: 10.1023/A:1005559827189, 2000b.

Lunt, D. J., de Noblet-Ducoudre, N., and Charbit, S.: Effects of a melted Greenland ice sheet on climate, vegetation, and the cryosphere, Clim Dynam, 23, 679-694, doi: 10.1007/s00382-004-0463-4, 2004.

Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J., and Dowsett, H. J.: Earth system sensitivity inferred from Pliocene modelling and data, Nat Geosci, 3, 60-64, doi: 10.1038/Ngeo706, 2010.

Luthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J. M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., and Stocker, T. F.: High-resolution carbon dioxide concentration record 650,000-800,000 years before present, Nature, 453, 379-382, doi: 10.1038/Nature06949, 2008.

Martinez-Boti, M. A., Foster, G. L., Chalk, T. B., Rohling, E. J., Sexton, P. F., Lunt, D. J., Pancost, R. D., Badger, M. P. S., and Schmidt, D. N.: Plio-Pleistocene climate sensitivity evaluated using high-resolution CO_2 records, Nature, 518, 49-54, doi: 10.1038/nature14145, 2015.

Marzocchi, A., Lunt, D. J., Flecker, R., Bradshaw, C. D., Farnsworth, A., and Hilgen, F. J.: Orbital control on late Miocene climate and the North African monsoon: insight from an ensemble of sub-precessional simulations, Clim Past, 11, 1271-1295, doi: 10.5194/cp-11-1271-2015, 2015.

Masson-Delmotte, V., Braconnot, P., Hoffmann, G., Jouzel, J., Kageyama, M., Landais, A., Lejeune, Q., Risi, C., Sime, L., Sjolte, J., Swingedouw, D., and Vinther, B.: Sensitivity of interglacial Greenland temperature and δO^{18}: ice core data, orbital and increased CO_2 climate simulations, Clim Past, 7, 1041-1059, doi: 10.5194/cp-7-1041-2011, 2011.

MATLAB: Statistics and Machine Learning Toolbox Release, The MathWorks, Inc., Natick, Massachusetts, United States, 2012b.

McGlade, C. and Ekins, P.: The geographical distribution of fossil fuels unused when limiting global warming to 2°C, Nature, 517, 187-190, doi: 10.1038/Nature14016, 2015.

Mckay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239-245, doi: 10.2307/1268522, 1979.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213-241, doi: 10.1007/s10584-011-0156-z, 2011.

Milankovitch, M.: Canon of insolation and the Ice-Age problem, Royal Serbian Academy Special Publication, 132, doi, 1941.

Monnin, E., Steig, E. J., Siegenthaler, U., Kawamura, K., Schwander, J., Stauffer, B., Stocker, T. F., Morse, D. L., Barnola, J. M., Bellier, B., Raynaud, D., and Fischer, H.: Evidence for substantial accumulation rate variability in Antarctica during the Holocene, through synchronization of CO_2 in the Taylor Dome, Dome C and DML ice cores, Earth Planet Sc Lett, 224, 45-54, doi: 10.1016/j.epsl.2004.05.007, 2004.

Muller, P. J., Kirst, G., Ruhland, G., von Storch, I., and Rosell-Mele, A.: Calibration of the alkenone paleotemperature index $U_{37}^{K'}$ based on core-tops from the eastern South Atlantic and the global ocean (60°N-60°S), Geochim Cosmochim Ac, 62, 1757-1772, doi: 10.1016/S0016-7037(98)00097-0, 1998.

Naafs, B. D. A., Stein, R., Hefter, J., Khelifi, N., De Schepper, S., and Haug, G. H.: Late Pliocene changes in the North Atlantic Current, Earth Planet Sc Lett, 298, 434-442, doi: 10.1016/j.epsl.2010.08.023, 2010.

NDA: Geological disposal. An overview of the generic Disposal System Safety Case, Nuclear Decomissioning Agency, Report no. NDA/RWMD/010. 2010.

Oakley, J. and O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs, Biometrika, 89, 769-784, doi: 10.1093/biomet/89.4.769, 2002.

Paillard, D.: Glacial cycles: Toward a new paradigm, Rev Geophys, 39, 325-346, doi: 10.1029/2000rg000091, 2001.

Paillard, D.: What drives the Ice Age cycle?, Science, 313, 455-456, doi: 10.1126/science.1131297, 2006.

Paillard, D. and Parrenin, F.: The Antarctic ice sheet and the triggering of deglaciations, Earth Planet Sc Lett, 227, 263-271, doi: 10.1016/j.epsl.2004.08.023, 2004.

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C.,

Pepin, L., Ritz, C., Saltzman, E., and Stievenard, M.: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, Nature, 399, 429-436, doi: 10.1038/20859, 1999.
Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A.: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, Clim Dynam, 16, 123-146, doi: 10.1007/s003820050009, 2000.
Prahl, F. G., Muehlhausen, L. A., and Zahnle, D. L.: Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions, Geochim Cosmochim Ac, 52, 2303-2310, doi: 10.1016/0016-7037(88)90132-9, 1988.
Prell, W. L. and Kutzbach, J. E.: Monsoon Variability over the Past 150,000 Years, J Geophys Res-Atmos, 92, 8411-8425, doi: 10.1029/JD092iD07p08411, 1987.
Prescott, C. L., Haywood, A. M., Dolan, A. M., Hunter, S. J., Pope, J. O., and Pickering, S. J.: Assessing orbitally-forced interglacial climate variability during the mid-Pliocene Warm Period, Earth Planet Sc Lett, 400, 261-271, doi: 10.1016/j.epsl.2014.05.030, 2014.
Raymo, M. E., Grant, B., Horowitz, M., and Rau, G. H.: Mid-Pliocene warmth: Stronger greenhouse and stronger conveyor, Mar Micropaleontol, 27, 313-326, doi: 10.1016/0377-8398(95)00048-8, 1996.
Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J Geophys Res-Atmos, 108, 4407, doi: 10.1029/2002jd002670, 2003.
Ridley, J. K., Huybrechts, P., Gregory, J. M., and Lowe, J. A.: Elimination of the Greenland ice sheet in a high CO_2 climate, J Climate, 18, 3409-3427, doi: 10.1175/Jcli3482.1, 2005.
Rogner, H. H.: An assessment of world hydrocarbon resources, Annu Rev Energ Env, 22, 217-262, doi: 10.1146/annurev.energy.22.1.217, 1997.
Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and analysis of computer experiments, Statistical Science, 4, 409-423, doi: 10.1214/ss/1177012413, 1989.
Seki, O., Foster, G. L., Schmidt, D. N., Mackensen, A., Kawamura, K., and Pancost, R. D.: Alkenone and boron-based Pliocene pCO_2 records, Earth Planet Sc Lett, 292, 201-211, doi: 10.1016/j.epsl.2010.01.037, 2010.
SKB: Long-term safety for the final repository for spent nuclear fuel at Forsmark. Main report of the SR-Site project, Svensk Kärnbränslehantering AB, Stockholm, Sweden, SKB Report TR-11-01. Available from: www.skb.com/publication/2345580, 2011.
Stap, L. B., van de Wal, R. S. W., de Boer, B., Bintanja, R., and Lourens, L. J.: Interaction of ice sheets and climate during the past 800 000 years, Clim Past, 10, 2135-2152, doi: 10.5194/cp-10-2135-2014, 2014.
Stone, E. J., Lunt, D. J., Rutt, I. C., and Hanna, E.: Investigating the sensitivity of numerical model simulations of the modern state of the Greenland ice-sheet and its future response to climate change, Cryosphere, 4, 397-417, doi: 10.5194/tc-4-397-2010, 2010.
Swingedouw, D., Fichefet, T., Huybrechts, P., Goosse, H., Driesschaert, E., and Loutre, M. F.: Antarctic ice-sheet melting provides negative feedbacks on future climate warming, Geophys Res Lett, 35, doi: 10.1029/2008gl034410, 2008.
Texier, D., Degnan, P., Loutre, M. F., Paillard, D., and Thorne, M. C.: Modelling sequential BIOsphere systems under CLIMate change for radioactive waste disposal. Project BIOCLIM, Las Vegas, Nevada, 30 March – 2 April 2003, 2003.
Toniazzo, T., Gregory, J. M., and Huybrechts, P.: Climatic impact of a Greenland deglaciation and its possible irreversibility, J Climate, 17, 21-33, doi: 10.1175/1520-0442(2004)017<0021:Cioagd>2.0.Co;2, 2004.
Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, B Am Meteorol Soc, 79, 61-78, doi: 10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2, 1998.

Tuenter, E., Weber, S. L., Hilgen, F. J., and Lourens, L. J.: The response of the African summer monsoon to remote and local forcing due to precession and obliquity, Global Planet Change, 36, 219-235, doi: 10.1016/S0921-8181(02)00196-0, 2003.

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Davies-Barnard, T., Day, J. J., Farnsworth, A., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Roberts, W. H. G., Stone, E. J., Tourte, G. J. L., and Williams, J. H. T.: The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0, Geosci Model Dev, doi: 10.5194/gmd-2017-16, 2017. doi: 10.5194/gmd-2017-16, 2017.

Wilkinson, R. D. (Ed.): Bayesian calibration of expensive multivariate computer experiments, John Wiley & Sons, Ltd, 2010.

Willeit, M., Ganopolski, A., Calov, R., Robinson, A., and Maslin, M.: The role of CO_2 decline for the onset of Northern Hemisphere glaciation, Quaternary Sci Rev, 119, 22-34, doi: 10.1016/j.quascirev.2015.04.015, 2015.

Williams, K. D., Senior, C. A., and Mitchell, J. F. B.: Transient climate change in the Hadley Centre models: The role of physical processes, J Climate, 14, 2659-2674, doi: 10.1175/1520-0442(2001)014<2659:Tccith>2.0.Co;2, 2001.

Williams, K. D., Ingram, W. J., and Gregory, J. M.: Time variation of effective climate sensitivity in GCMs, J Climate, 21, 5076-5090, doi: 10.1175/2008jcli2371.1, 2008.

Winkelmann, R., Levermann, A., Ridgwell, A., and Caldeira, K.: Combustion of available fossil-fuel resources sufficient to eliminate the Antarctic ice sheet, Science Advances, 1, e1500589, doi: 10.1126/sciadv.1500589, 2015.

Winton, M., Takahashi, K., and Held, I. M.: Importance of ocean heat uptake efficacy to transient climate change, J Climate, 23, 2333-2344, doi: 10.1175/2009jcli3139.1, 2010.

Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P., and Fifield, L. K.: Timing of the Last Glacial Maximum from observed sea-level minima, Nature, 406, 713-716, doi: 10.1038/35021035, 2000.

Table 1. Ensembles setup: sampling ranges for input parameters (obliquity, *e*sinϖ, *e*cosϖ and CO_2) for the *highCO_2* and *lowCO_2* ensembles.

Ensemble	Time covered from present day (AP)	Parameter	Sampling range	
			Minimum	Maximum
highCO_2	110 kyr	ε (º)	22.3	24.3
		*e*sinϖ	-0.016	0.016
		*e*cosϖ	-0.016	0.015
		CO_2 (ppmv)	280	3600
lowCO_2	1 Ma	ε (º)	22.2	24.4
		*e*sinϖ	-0.055	0.055
		*e*cosϖ	-0.055	0.055
		CO_2 (ppmv)	250	560

Table 2. Experiment setup: Orbital parameters (obliquity, eccentricity and longitude of perihelion) and atmospheric CO_2 concentration for simulations in the *highCO_2* and *lowCO_2* ensembles. All experiments in both ensembles were run with modern ice sheet (*modice*) configurations. Experiments shown in bold were also run with reduced ice sheet (*lowice*) configurations. The experiment number is given, and the experiment name is constructed using the ice sheet configuration, the ensemble name and the experiment number, for example: modice_lowCO2_1.

Ensemble	#	ε (º)	*e* -	ϖ (º)	CO_2 (ppmv)	Ensemble	#	ε (º)	*e* -	ϖ (º)	CO_2 (ppmv)
highCO_2	1	23.53	0.0093	240.3	3348.2	*lowCO_2*	1	22.99	0.0481	320.1	375.7
	2	24.24	0.0135	212.6	2159.3		2	23.02	0.0323	63.7	516.9
	3	22.38	0.0110	260.0	1645.0		3	22.81	0.0481	334.2	470.4
	4	24.07	0.0044	101.8	800.8		4	24.03	0.0537	84.9	390.3
	5	23.07	0.0203	313.0	1999.9		5	23.09	0.0294	293.8	325.3
	6	24.03	0.0087	184.9	3049.0		6	23.58	0.0098	325.1	337.5
	7	22.53	0.0163	162.0	900.9		7	23.72	0.0133	74.3	489.2
	8	23.57	0.0158	21.0	1746.3		**8**	**24.17**	**0.0066**	**174.1**	**346.0**
	9	23.34	0.0131	113.5	996.8		9	23.82	0.0400	48.2	260.6
	10	23.37	0.0198	220.2	3139.3		10	23.39	0.0412	53.8	409.5
	11	22.73	0.0187	236.1	1081.9		11	22.89	0.0531	115.2	436.6
	12	22.63	0.0121	184.8	2451.5		12	23.34	0.0281	133.9	504.4
	13	22.41	0.0131	192.8	3372.4		13	22.65	0.0473	102.6	555.6
	14	22.78	0.0137	299.3	448.2		14	23.20	0.0368	180.9	385.1
	15	22.97	0.0111	14.1	1225.7		15	23.96	0.0232	40.0	403.4
	16	22.90	0.0087	62.2	1841.9		16	24.27	0.0460	298.1	341.1
	17	23.63	0.0151	200.6	1151.6		17	22.35	0.0391	265.9	522.1
	18	23.77	0.0134	78.7	2101.7		18	23.91	0.0361	343.2	318.6
	19	23.73	0.0159	323.7	1526.6		**19**	**22.33**	**0.0484**	**324.2**	**264.5**
	20	24.29	0.0082	164.6	2890.4		20	22.94	0.0350	268.7	540.8

21	**22.31**	**0.0038**	**299.1**	**1389.5**	21	22.68	0.0323	332.4	531.5
22	23.42	0.0117	122.5	397.3	22	24.28	0.0387	118.7	446.7
23	24.00	0.0101	206.6	303.4	23	23.60	0.0484	282.0	310.5
24	22.48	0.0146	294.9	2845.7	24	24.19	0.0337	346.3	548.3
25	22.57	0.0067	81.2	1341.2	25	24.14	0.0423	11.6	425.4
26	22.93	0.0171	114.4	3516.0	26	22.20	0.0035	85.2	303.0
27	24.13	0.0143	257.3	2951.8	27	22.78	0.0070	212.1	480.4
28	23.00	0.0062	272.2	2274.6	28	22.72	0.0526	239.9	280.0
29	23.95	0.0103	114.7	564.7	**29**	**23.65**	**0.0543**	**30.3**	**362.0**
30	23.17	0.0169	56.7	1900.9	30	23.24	0.0351	200.4	411.9
31	23.70	0.0122	1.4	773.0	31	23.87	0.0276	156.5	287.5
32	23.24	0.0021	310.2	2582.1	32	22.25	0.0499	208.9	365.3
33	22.81	0.0121	66.3	2386.5	33	22.54	0.0510	103.4	471.1
34	**24.18**	**0.0145**	**36.6**	**668.2**	34	22.58	0.0404	292.2	544.5
35	23.82	0.0075	10.8	2244.8	35	22.87	0.0530	20.9	498.2
36	23.14	0.0141	314.1	3588.9	36	23.53	0.0414	147.0	507.0
37	23.49	0.0121	101.5	2760.4	37	22.39	0.0165	149.1	393.9
38	22.66	0.0162	69.5	2623.9	38	22.43	0.0537	175.0	484.8
39	23.28	0.0146	207.5	1484.8	39	24.38	0.0482	342.9	418.3
40	23.89	0.0092	21.1	3188.8	40	23.76	0.0504	127.0	528.1

Table 3. Parameter values estimated from Gregory plots for the 2x and 4x pre-industrial CO_2 simulations. Shown are the effective radiative forcing (*F*; W m^{-2}) and the climate feedback parameter (*α*; W m^{-2} $^{o}C^{-1}$) for years 1-20 and years 21-100. The uncertainties are the standard error from the linear regression.

Simulation		*F* (W m^{-2})		*α* (W m^{-2} $^{o}C^{-1}$)	
		yr 1-20	yr 21-100	yr 1-20	yr 21-100
2xCO_2	*modice_lowCO2_13*	4.24 ± 0.4	-	-1.30 ± 0.2	-0.68 ± 0.05
4xCO_2	*modice_highCO2_17*	6.88 ± 0.3	-	-0.99 ± 0.1	-0.56 ± 0.02

Table 4. Mean temperature anomalies and uncertainties (1 standard deviation) for the period 3300-2800 kyr BP estimated by the emulator and alkenone proxy data for the four ODP/IODP sites.

ODP/IODP Site	Location			Emulated SAT anomaly (°C)			Proxy data SST anomaly (°C)	
		Lat	Lon	280 ppmv	350 ppmv	400 ppmv	Prahl et al. (1988)	Muller et al. (1998)
982[1]	North Atlantic	57.5° N	15.9° W	0.6 ±0.4	2.4 ±0.3	3.3 ±0.3	5.4	5.7
U1313[2]	North Atlantic	41.0° N	33.0° W	-0.8 ±0.3	0.0 ±0.2	0.8 ±0.2	1.6	2.0
722[3]	Arabian Sea	16.6° N	59.8° E	0.0 ±0.2	1.0 ±0.2	1.7 ±0.2	1.0	1.7
662[3]	Tropical Atlantic	1.4° S	11.7° W	0.2 ±0.2	0.9 ±0.2	1.3 ±0.2	1.3	1.9

[1]Lawrence et al. (2009); [2]Naafs et al. (2010); [3]Herbert et al. (2010).

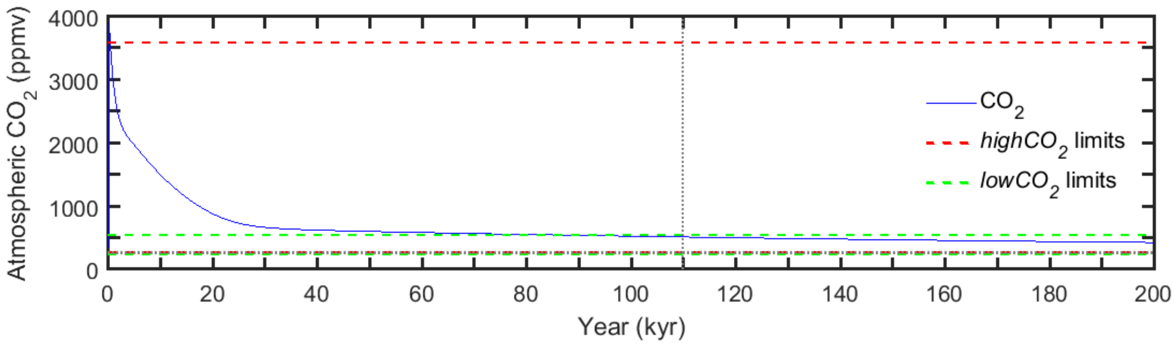


Figure 1. Time series of atmospheric CO_2 concentration (ppmv) for the next 200 kyr following logistic CO_2 emissions of 10,000 PgC, run using the cGENIE model (Lord et al., 2016). Also shown are the upper and lower CO_2 limits of the *high*CO_2 (red dashed lines) and *low*CO_2 (green dashed lines) ensembles. The pre-industrial CO_2 concentration of 280 ppmv (horizontal grey dotted line), and the 110 kyr cut-off for the highCO_2 ensemble (vertical grey dotted line) are included for reference.

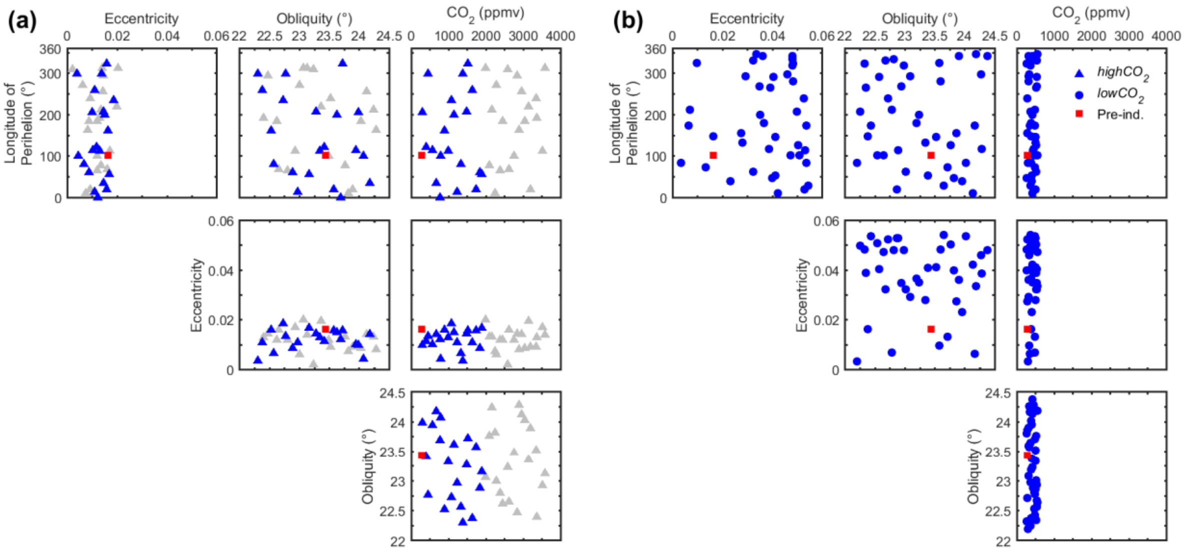


Figure 2. Distribution of 40 experiments produced by Latin hypercube sampling, displayed as two-dimensional ~~slices~~ projections through four-dimensional space. (a) *highCO$_2$* ensemble, (b) *lowCO$_2$* ensemble. The variables are eccentricity (*e*), longitude of perihelion (*ϖ*; degrees), obliquity (*ε*; degrees), and atmospheric CO$_2$ concentration (ppmv). A pre-industrial control simulation is shown in red. In the *highCO$_2$* ensemble, experiments with CO$_2$ concentrations of more than 2000 ppmv, shown in grey, were excluded from the emulator.

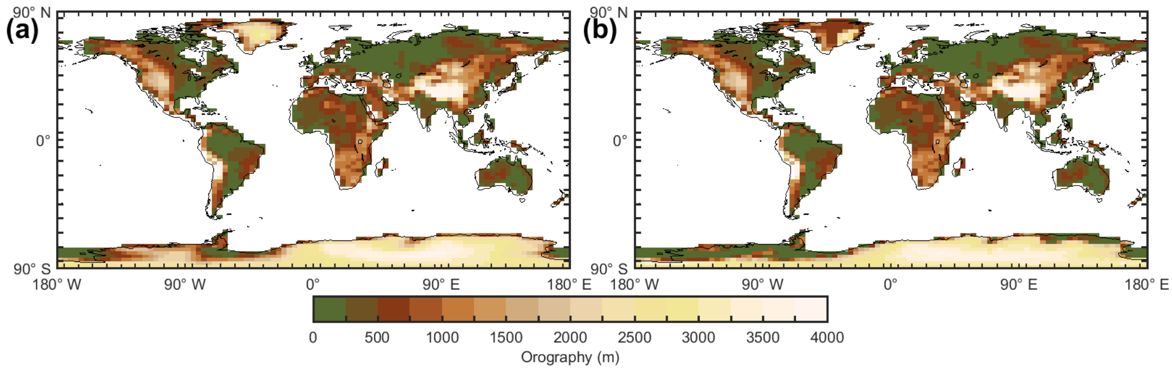
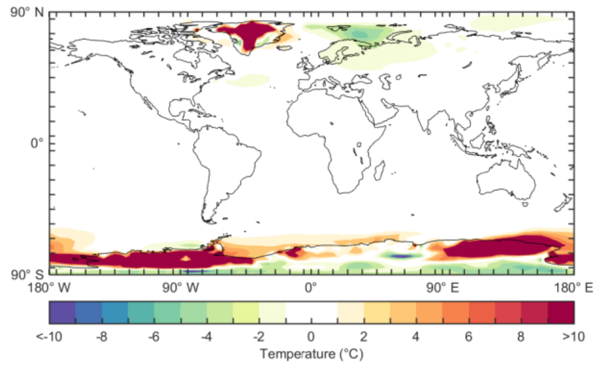
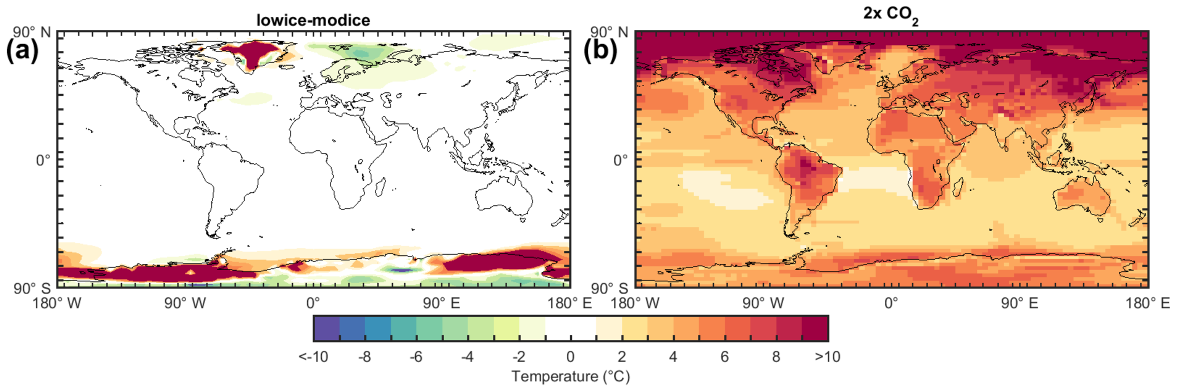


Figure 3. Orography (m) in the two ice sheet configuration ensembles. (a) *modice* ensemble, (b) *lowice* ensemble. Differences only occur over Greenland and Antarctica.





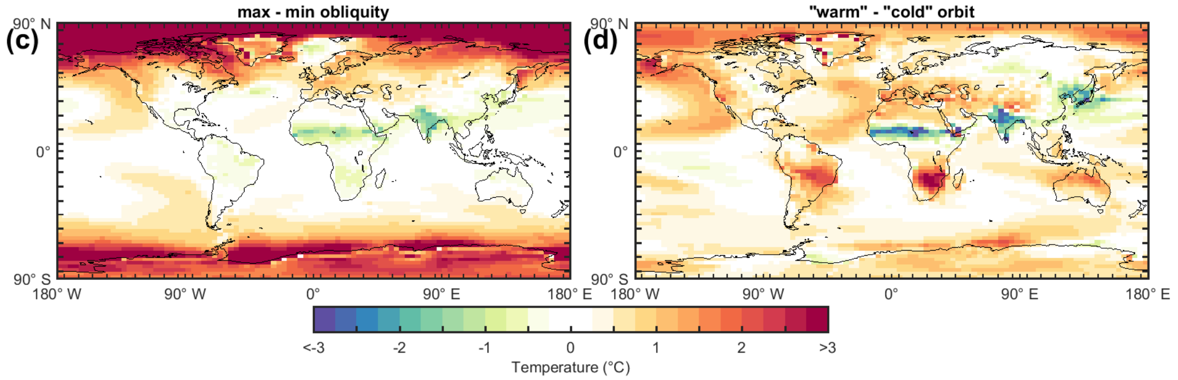


Figure 4. Mean annual SAT (ºC) anomalies produced by the various climate forcings. (a)~~y~~ ~~for t~~The *lowice* experiments compared with their *modice* equivalents, averaged across the five *lowice* experiments. (b) ~~to~~ :(d) Idealized experiments performed using the *modice* emulator. All orbital and CO_2 conditions are set to pre-industrial values unless specified: (b) 2x pre-industrial CO_2, (c) maximum obliquity compared to minimum obliquity, (d) "warm" orbital conditions (high eccentricity, NH summer at perihelion) compared to "cold" orbital conditions (low eccentricity, NH summer at aphelion). ~~All SAT anomalies have been calculated compared with the pre-industrial control simulation.~~ The different forcings result in global mean SAT anomalies of: (b) 4.2°C, (c) 0.4°C, and (d) 0.4°C.

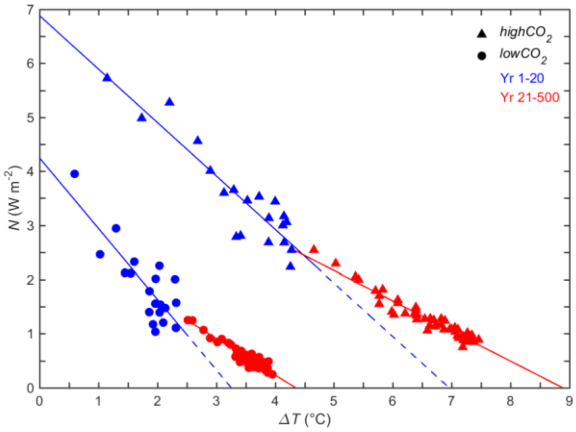


Figure 5. Gregory plot showing change in TOA net downward radiation flux (*N*; W m^{-2}) as a function of change in global mean annual SAT (*ΔT*; ^{o}C) for approximate 2xCO_2 (modice_lowCO2_13; circles) and 4xCO_2 (modice_highCO2_17; triangles) experiments. Lines show regression fits to the global mean annual data points for years 1-20 (blue) and years 21-500 (red). Data points are mean annual data for years 1-20 (blue) and mean decadal data for years 21-500 (red). The *ΔT* intercepts (*N*=0) of the red lines give the estimated equilibrated SAT (ΔT_{eq}^{g}) for the two experiments. The *ΔT* intercepts of the dashed blue lines represent the equilibrium that the experiment would have reached if the feedback strengths in the first 20 years had been maintained. SAT is shown as an anomaly compared with the pre-industrial control simulation.

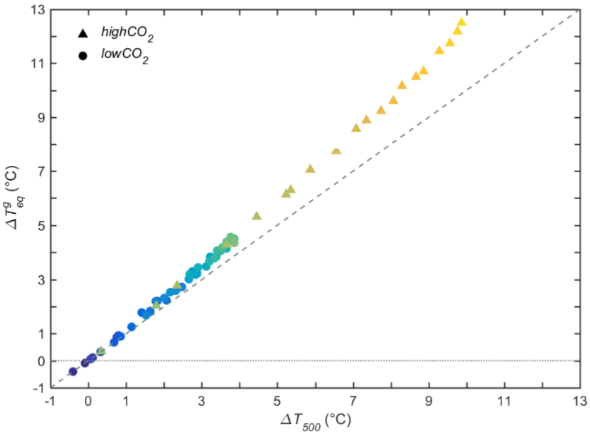


Figure 6. Equilibrated global mean annual change in SAT (ΔT_{eq}^{g}; ^{o}C) estimated using the methodology of Gregory et al. (2004) against global mean annual change in SAT (ΔT_{500}; ^{o}C) at year 500 (average of final 50 years) for the *lowCO2* (circles) and *highCO2* (triangles) *modice* ensembles. The colours of the points indicate the CO_2 concentration of the experiment, from low (blue) to high (yellow). The 1:1 line (dashed) is included for reference. SAT is shown as an anomaly compared with the pre-industrial control simulation.

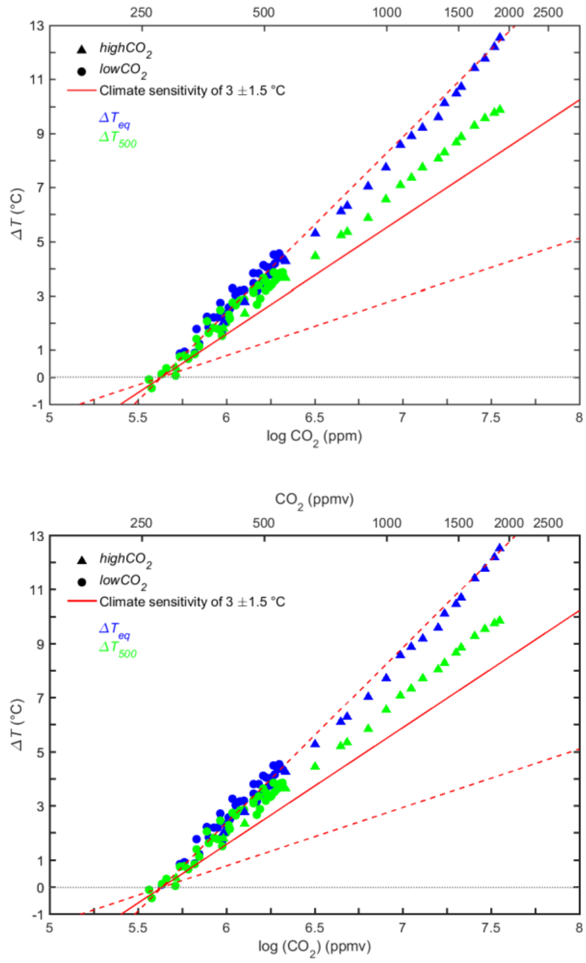


Figure 7. Equilibrated global mean annual change in SAT (ΔT_{eq}; °C; blue), estimated by applying the $\Delta T_{eq}^{g}/\Delta T_{500}$ ratio identified using the Gregory methodology to the GCM data, against atmospheric CO_2 (ppmv) for the *lowCO₂* (circles) and *highCO₂* (triangles) *modice* ensembles. Also shown is ΔT_{500} (green), along with the idealized relationship between ln~~og~~(CO_2) and ΔT (red lines) for a climate sensitivity of 3°C (solid), 1.5°C (lower dashed) and 4.5°C (upper dashed) (IPCC, 2013). SAT is shown as an anomaly compared with the pre-industrial control simulation.

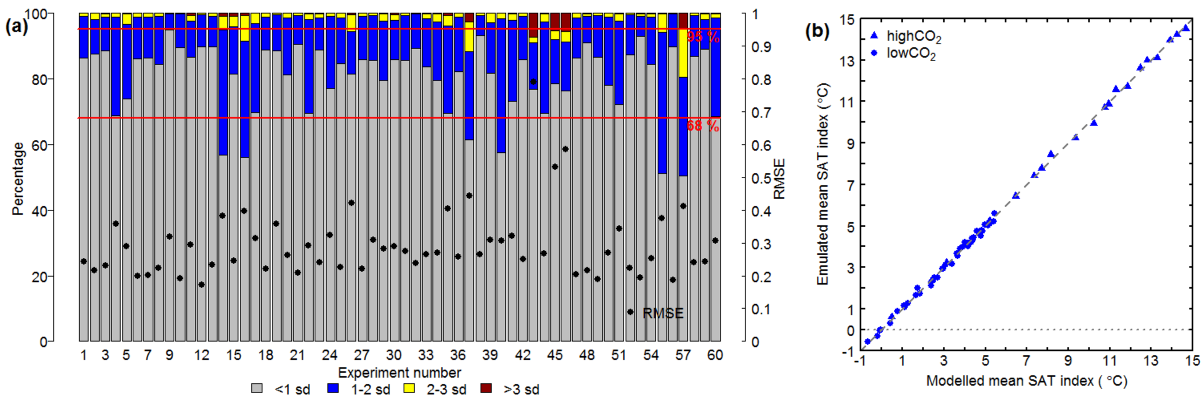
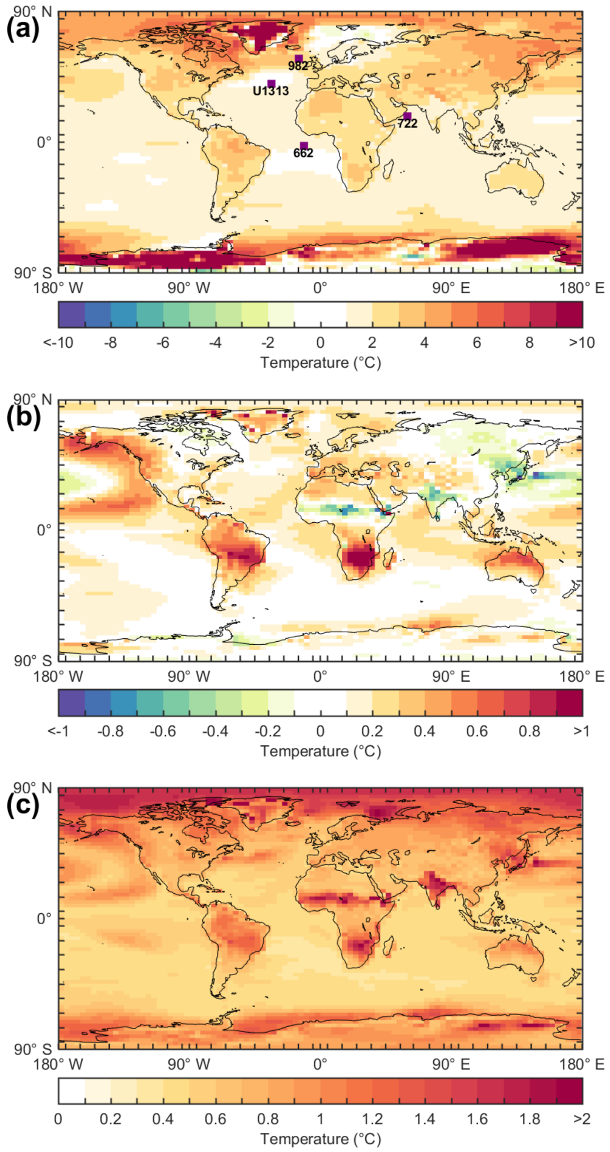


Figure 8. Evaluation of emulator performance. (a) Bars give the percentage of grid boxes for which the emulator predicts the SAT of the left-out experiment to within 1, 2, 3 and more than 3 standard deviations (sd). Also shown is the RMSE for the experiments (black circles). Red lines indicate 68% and 95%. (b) ~~M~~Global mean annual SAT index (°C) calculated by the emulator and the GCM for the *lowCO₂* (circles) and *highCO₂* (triangles) *modice* ensembles. 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: we therefore refer to it as a SAT index. SAT is shown as an anomaly compared with the pre-industrial control simulation.


(a)
982
U1313
722
662
90° N
0°
90° S
180° W
90° W
0°
90° E
180° E
<-10
-8
-6
-4
-2
0
2
4
6
8
>10
Temperature (°C)
(b)
<-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
>1
Temperature (°C)
(c)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
>2
Temperature (°C)

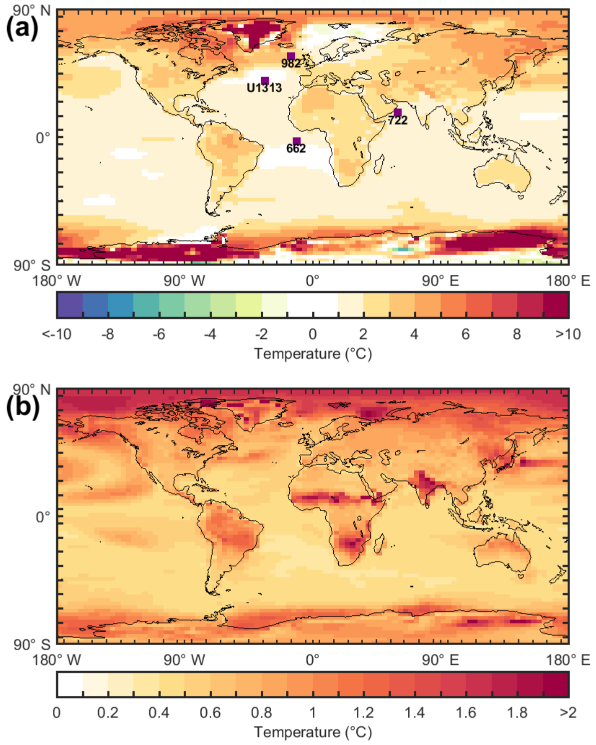


Figure 9. Emulated mean annual SAT (ºC) for the 400 ppmv CO_2 scenario, modelled using the *lowice* emulator. SAT is shown as an anomaly compared with the pre-industrial control simulation. (a) Mean annual SAT for modern-day orbital conditions. Also shown are the locations of the four ODP/IODP sites (purple squares): Site 982 (North Atlantic; (Lawrence et al., 2009)), Site U1313 (North Atlantic; (Naafs et al., 2010)), Site 722 (Arabian Sea; (Herbert et al., 2010)) and Site 662 (tropical Atlantic; (Herbert et al., 2010)). (b) ~~Anomaly in mean annual SAT averaged over the period 3300-2800 kyr BP (late Pliocene) compared to that produced under modern-day orbital conditions (Fig. 9a). (c)~~ Standard deviation of mean annual SAT for the period 3300-2800 kyr BP (late Pliocene), also taking into account the emulator posterior variance.

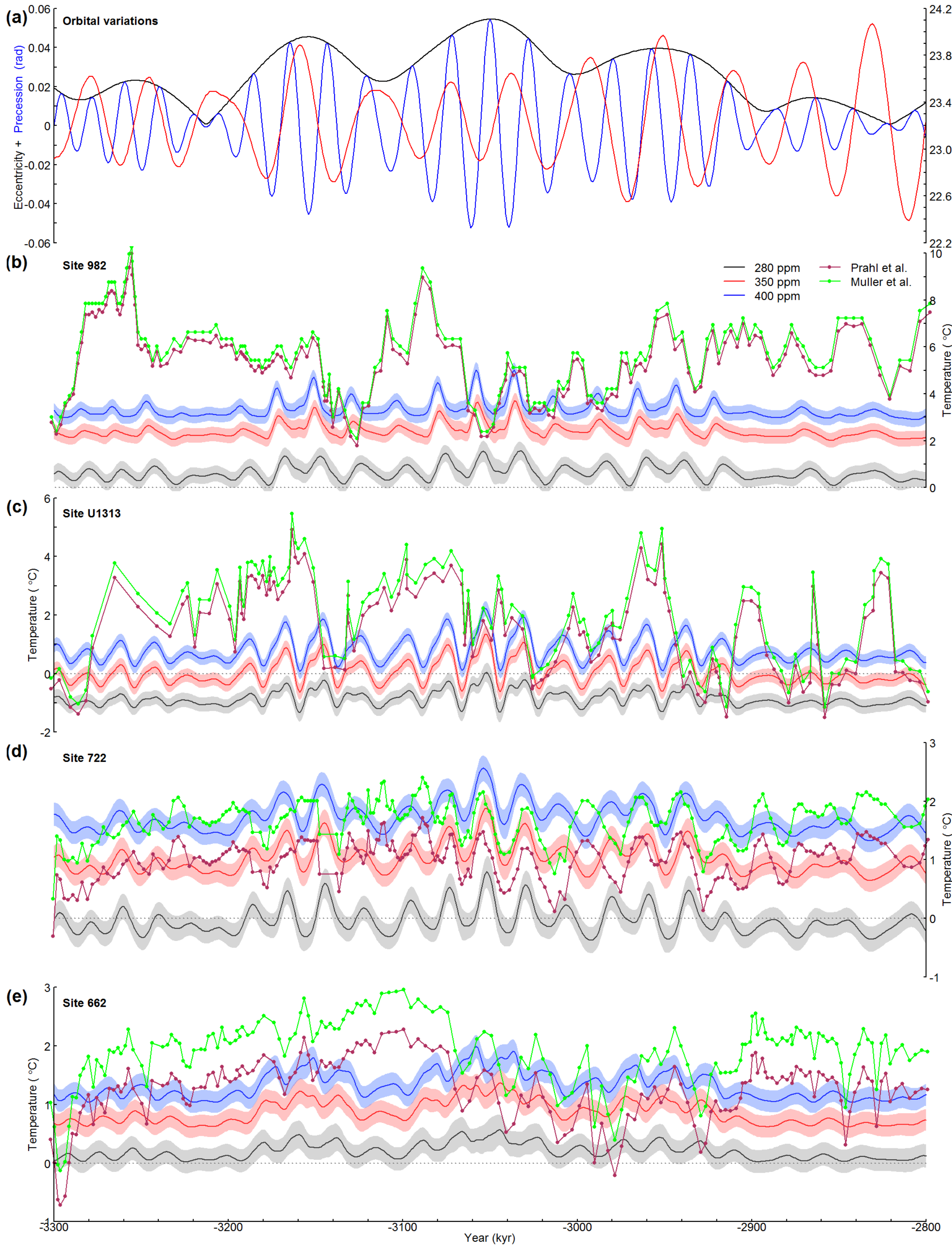


Figure 10. Data-model comparison of temperature anomaly for the period 3300-2800 kyr BP (late Pliocene). (a) Time series of orbital variations (Laskar et al., 2004), showing eccentricity (black) and precession (radians; blue) on the left axis, and obliquity (degrees; red) on the right axis. (b):(e) Time series of emulated grid box mean annual SAT (°C; plain lines), modelled every 1 kyr, for three constant CO_2 scenarios; 280 ppmv (black), 350 ppmv (red) and 400 ppmv (blue). Modelled using the *lowice* emulator. Error bands represent the emulated grid box posterior variance (1 standard deviation). Also shown is SST proxy data (°C; dotted lines) calibrated using the method of Prahl et al. (1988) (maroon), and the method of Muller et al. (1998) (green). SSTs for four ODP/IODP sites are compared: Site 982 (North Atlantic; (Lawrence et al., 2009)), Site U1313 (North Atlantic; (Naafs et al., 2010)), Site 722 (Arabian Sea; (Herbert et al., 2010)) and Site 662 (tropical Atlantic; (Herbert et al., 2010)). SAT is shown as an anomaly compared

with the pre-industrial control simulation, SST is shown as an anomaly compared with SST observations for the period 1870-1900 taken from the HadISST dataset (Rayner et al., 2003). Note the different vertical axis scales.

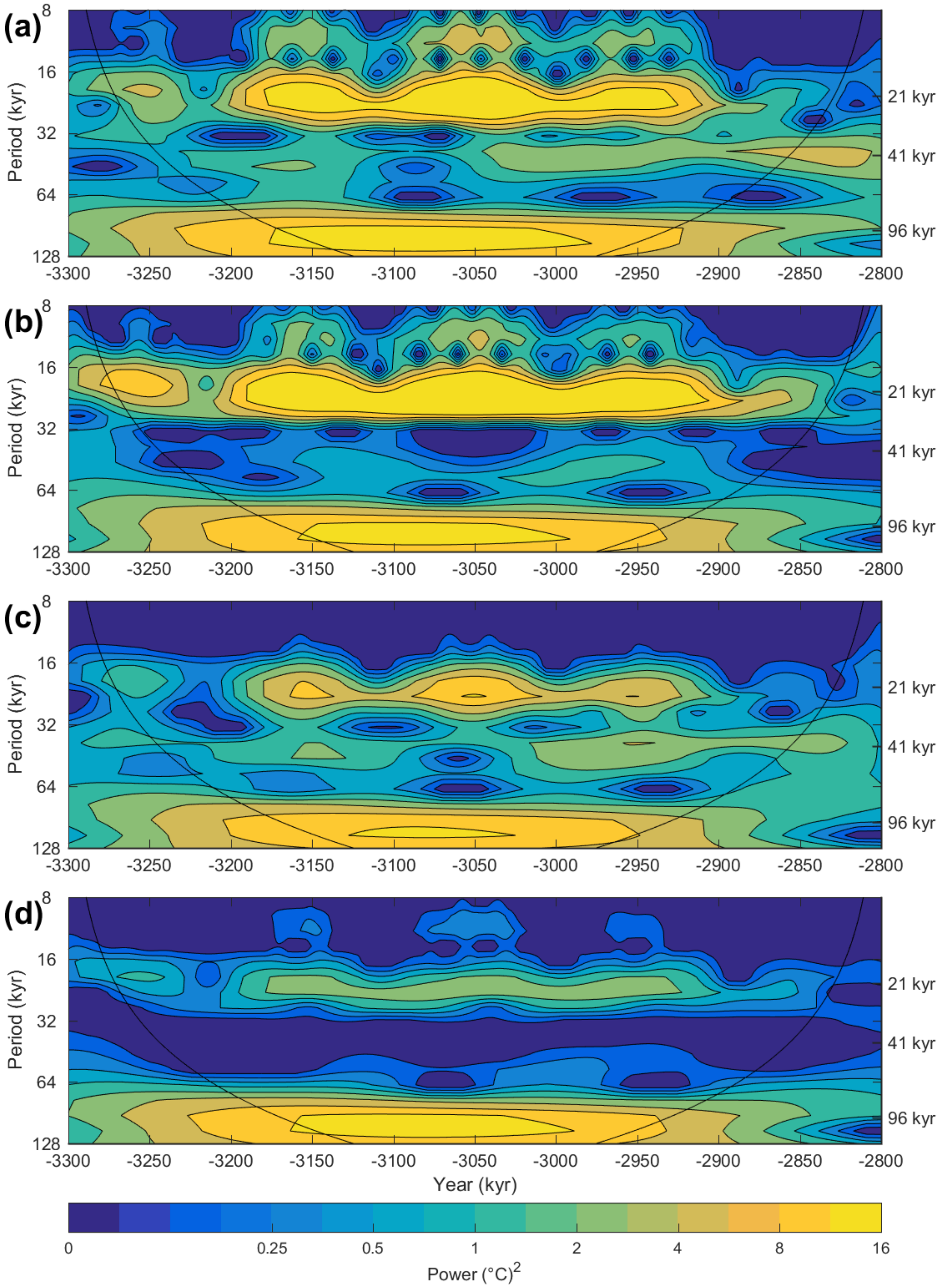


Figure 11. The wavelet power spectrum for 3300-2800 kyr BP (late Pliocene). Wavelet analysis was performed on emulated grid box mean annual SAT (°C), modelled every 1 kyr using the *lowice* emulator, for constant CO_2 of 400 ppmv (blue line in Fig. 10b to 10e). The data are normalized by the mean variance for the analysed SAT data (σ^2 =

0.14°C). Four ODP/IODP sites are compared: (a) Site 982 (North Atlantic; (Lawrence et al., 2009)), (b) Site U1313 (North Atlantic; (Naafs et al., 2010)), (c) Site 722 (Arabian Sea; (Herbert et al., 2010)) and (d) Site 662 (tropical Atlantic; (Herbert et al., 2010)).

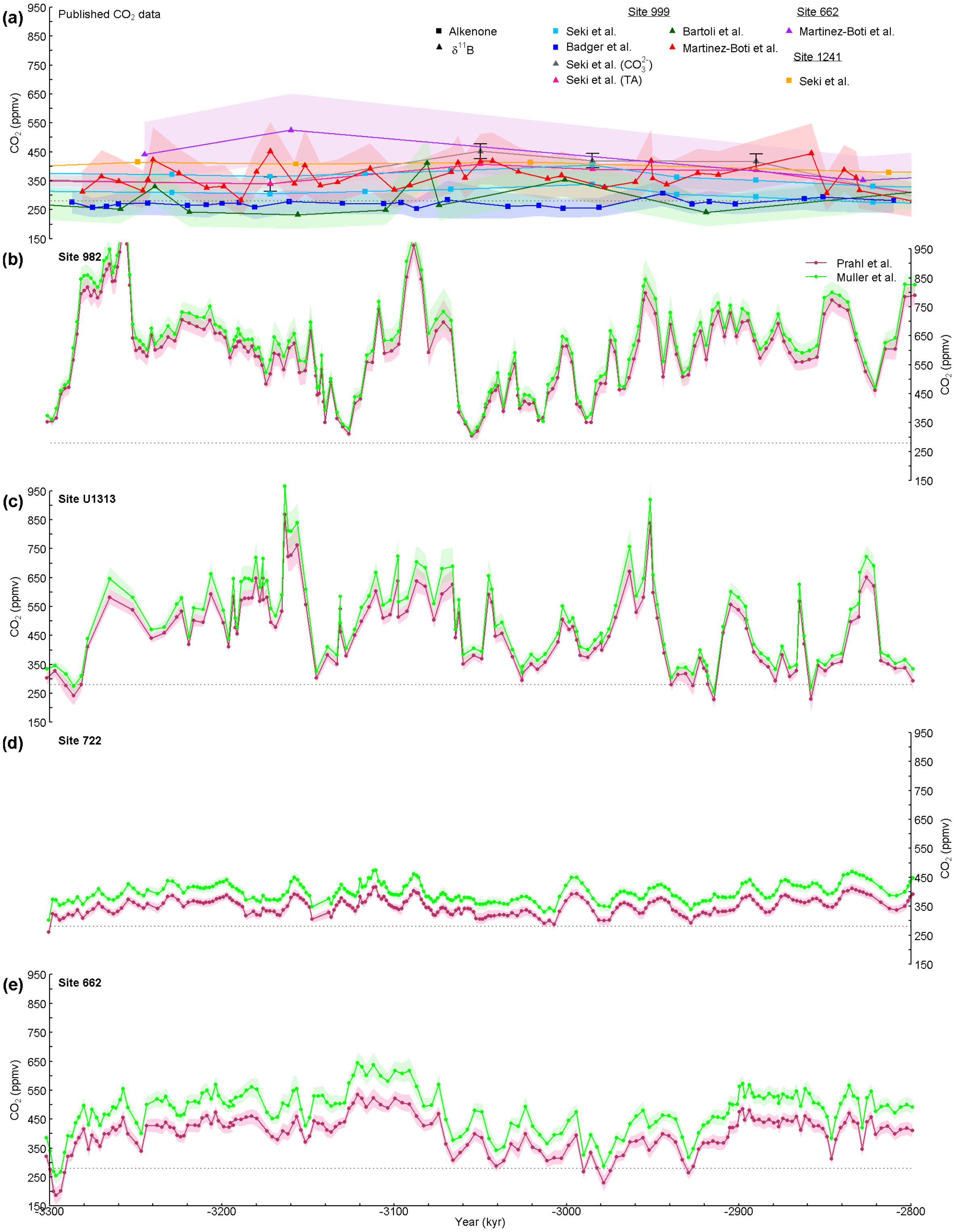


Figure 12. Data-model comparison of atmospheric CO_2 concentration (ppmv) for the period 3300-2800 kyr BP (late Pliocene) for six ODP/IODP sites: Site 982 (North Atlantic), Site U1313 (North Atlantic), Site 722 (Arabian Sea), Site 999 (Caribbean), Site 662 (tropical Atlantic), and Site 1241 (east tropical Pacific). (a) Time series of atmospheric CO_2 concentration from selected proxy data records. Shown is CO_2 estimated from alkenone (squares) for Site 999 by Seki et al. (2010) (light blue), Badger et al. (2013) (dark blue) and for Site 1241 by Seki et al. (2010) (orange), and estimated from $\delta^{11}B$ (triangles) for Site 999 by Seki et al. (2010) based on modelled carbonate concentration ($[CO_3^{2-}]$)

(grey) and assuming modern total alkalinity (TA; pink), Bartoli et al. (2011) (dark green), Martinez-Boti et al. (2015) (red) and for Site 662 by Martinez-Boti et al. (2015) (purple). For the Seki et al. (2010) δ^{11}B records, error bars are ±25 ppmv and the error band is the result of varying the modern TA by ±5%, whilst for Martinez-Boti et al. (2015) the error band represents the 95% confidence interval for a 10,000 member Monte Carlo analysis. (b):(e) Time series of atmospheric CO_2 concentration estimated from SST proxy data (circles; Herbert et al. (2010) – Sites 662 and 722, Naafs et al. (2010) – Site U1313, Lawrence et al. (2009) – Site 982) calibrated using the method of Prahl et al. (1988) (maroon), and the method of Muller et al. (1998) (light green). CO_2 is calculated based on a linear relationship between emulated grid box mean annual SAT (modelled using the *lowice* emulator) and CO_2, for three constant CO_2 scenarios of 280, 350 and 400 ppmv. Error bands represent estimated atmospheric CO_2 concentration taking into account the emulated grid box posterior variance (1 standard deviation). Where the error appears to be very low, this is generally an artefact of the way that the data has been plotted. The pre-industrial CO_2 concentration of 280 ppmv (grey dotted line) is included for reference.

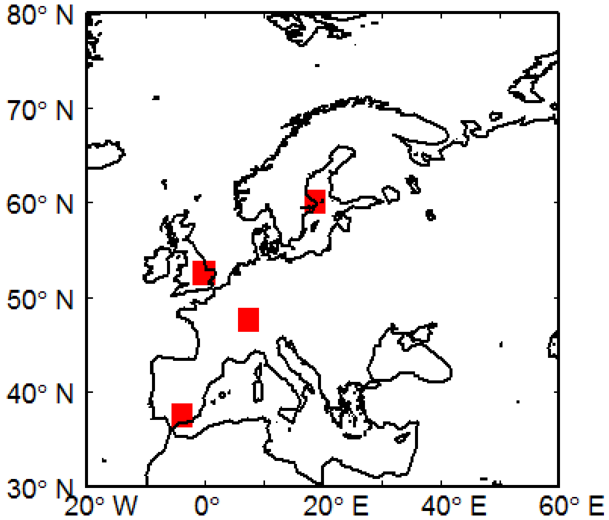


Figure 13. Map of Europe highlighting the grid boxes that represent the four case study sites. From north to south: Sweden, Central England, Switzerland and Spain.

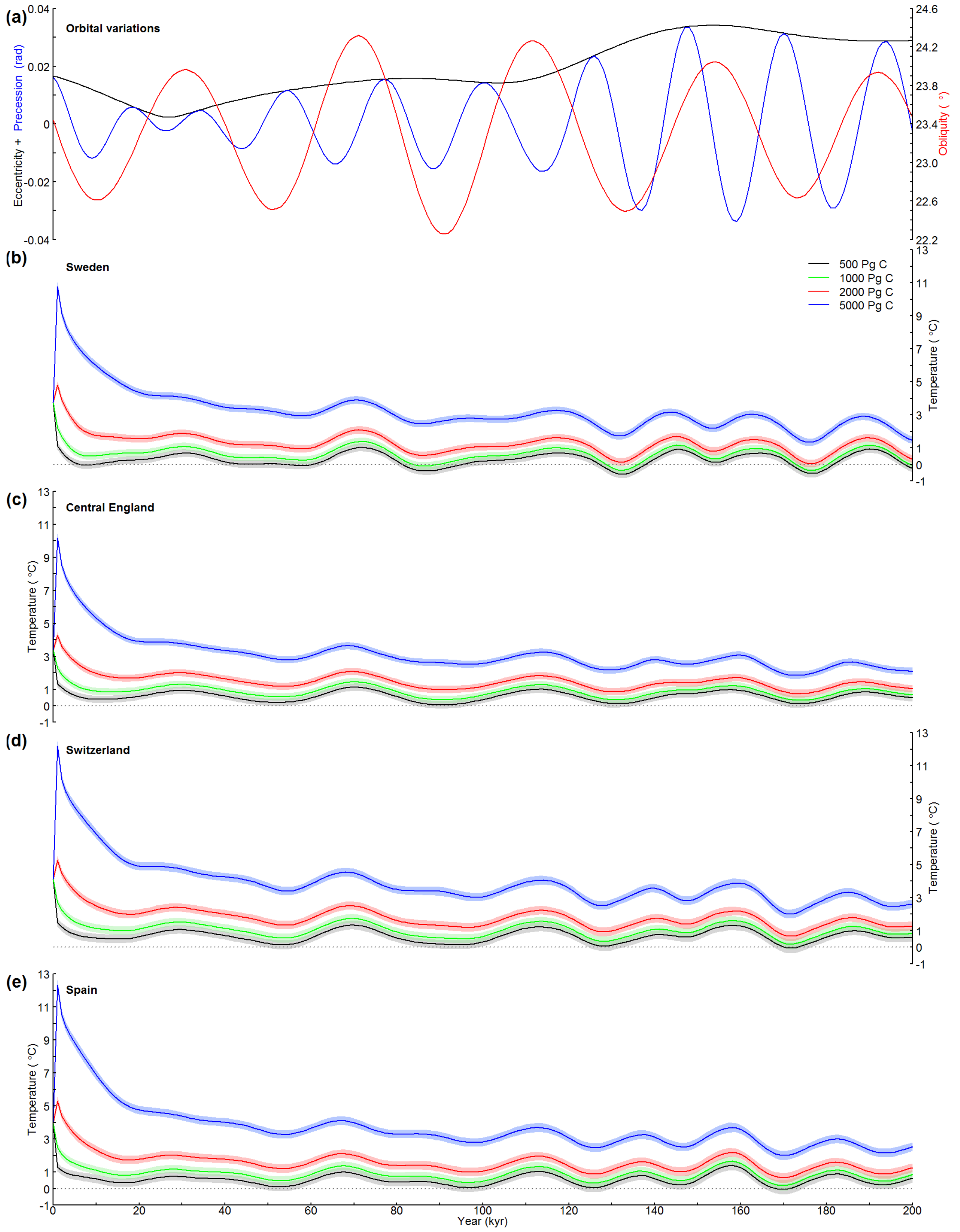


Figure 14. Emulation of SAT anomaly for the next 200 kyr. (a) Time series of orbital variations (Laskar et al., 2004), showing eccentricity (black) and precession (radians; blue) on the left axis, and obliquity (degrees; red) on the right axis. (b):-(e) Time series of emulated grid box mean annual SAT (ºC), modelled every 1 kyr, for four CO_2 emissions scenarios; 500 Pg C (black), 1000 Pg C (green), 2000 Pg C (red) and 5000 Pg C (blue). Modelled using the *modice* emulator. Error bands represent the emulated grid box posterior variance (1 standard deviation). Four sites are presented, representing grid boxes in Sweden, Central England, Switzerland and Spain. SAT is shown as an anomaly compared with the pre-industrial control simulation.

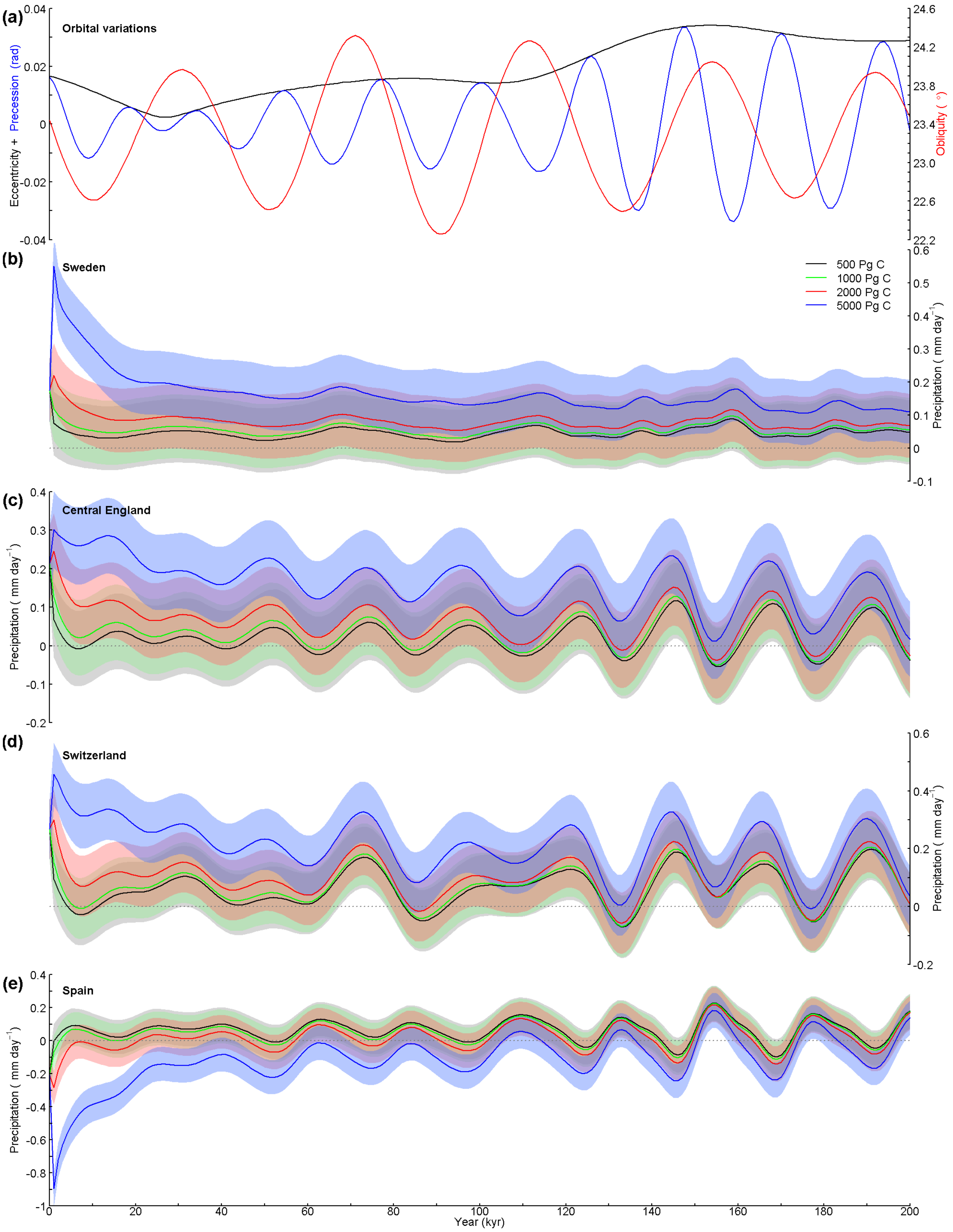


Figure 15. Emulation of precipitation anomaly for the next 200 kyr. (a) Time series of orbital variations (Laskar et al., 2004), showing eccentricity (black) and precession (radians; blue) on the left axis, and obliquity (degrees; red) on the right axis. (b)-:-(e) Time series of emulated grid box mean annual precipitation (mm day^{-1}), modelled every 1 kyr, for four CO_2 emissions scenarios; 500 Pg C (black), 1000 Pg C (green), 2000 Pg C (red) and 5000 Pg C (blue). Modelled using the *modice* emulator. Error bands represent the emulated grid box posterior variance (1 standard deviation). Four sites are presented, representing grid boxes in Sweden, Central England, Switzerland and Spain. Precipitation is shown as an anomaly compared with the pre-industrial control simulation. Note the different vertical axis scales.

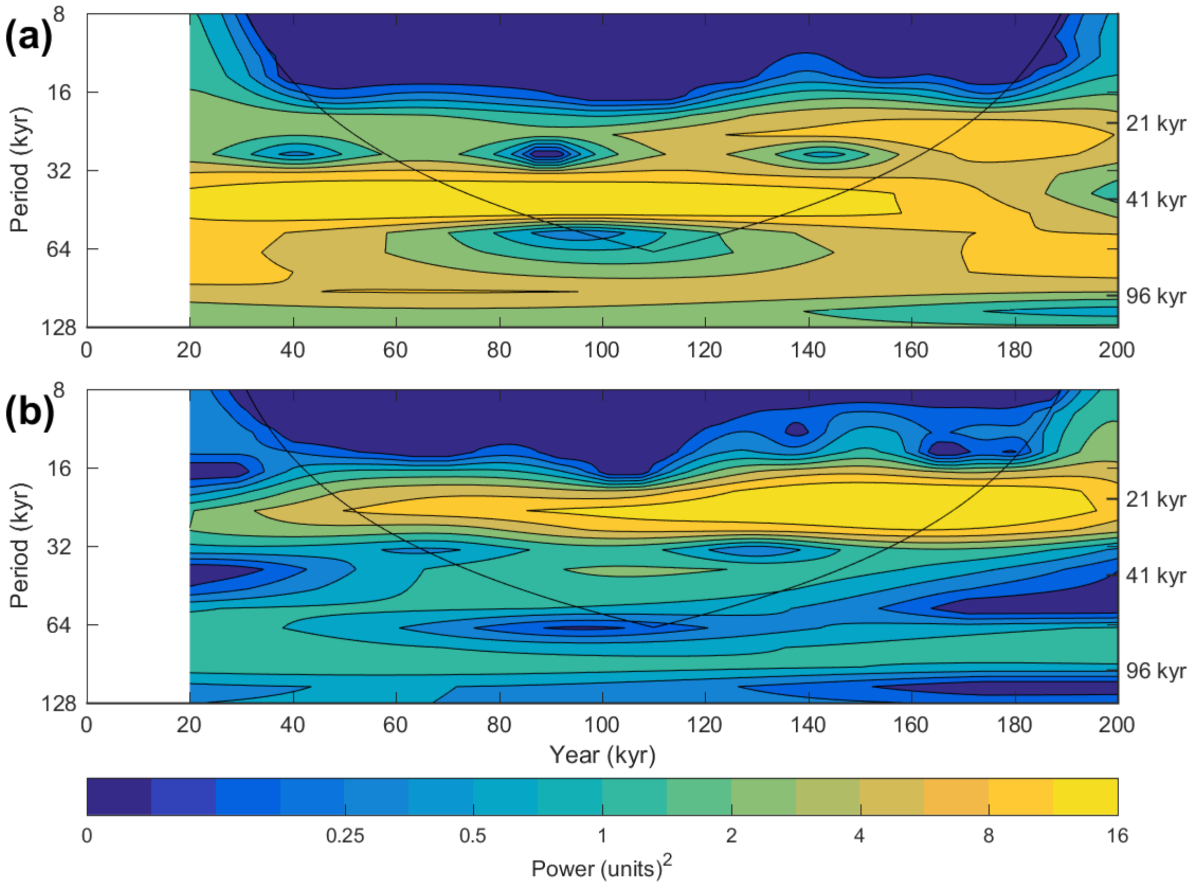


Figure 16. The wavelet power spectrum for the next 200 kyr for the Central England grid box. Wavelet analysis was performed on data for 20 kyr AP onwards, for: (a) emulated grid box mean annual SAT (°C; blue line in Fig. 14c), and (b) emulated grid box mean annual precipitation (mm day^{-1}; blue line in Fig. 15c). Both variables were modelled every 1 kyr using the *modice* emulator, for the 5000 Pg C emissions scenario. The data are normalized separately by: (a) the mean variance for the analysed SAT data ($\sigma 2$ = 0.14°C), and (b) the variance for the analysed precipitation data (σ^2 = 0.003°C).